Four essays on applied and methodological issues in the study of subjective life satisfaction



Caspar Kaiser

Department of Social Policy and Intervention, Nuffield College, Wellbeing Research Centre

University of Oxford

Trinity Term 2020

Thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Abstract

This thesis is comprised of four empirical papers. Paper 1, entitled "**People do not adapt. New analyses of the dynamic effects of own and reference income**", uses yearly GSOEP and UKHLS data to show that there is no significant adaptation to the effects of own and reference income on life satisfaction. The paper gives an in-depth explanation of why almost all previous studies found adaptation to income and explores the implications of this finding for the estimation of equivalence scales. Paper 2 (with Nhat An Trinh), entitled "**Positional, mobility and reference effects: How does social class affect life satisfaction in Europe?**", assesses the effects of social class on life satisfaction. Using ESS data, the paper gives an account of the separate effects of class position, intergenerational class mobility, and reference effects of social class on life satisfaction. The paper also gives an extended discussion of the assumptions underlying Sobel's (1981) influential 'diagonal reference model', which is often used in seeking to separate the effects of class destination and origin from those of class mobility. Paper 3 (with Maarten Vendrik), entitled "**How threatening are transformations of happiness scales to subjective wellbeing research?**", investigates the possibility and plausibility of reversing the signs of estimated effects of explanatory variables on happiness and life satisfaction by means of positive monotonic transformations of the response scale. The paper responds to previous papers that purport to show that such reversals are possible and plausible, which would undermine the validity of most subjective wellbeing research. Fortunately for this literature, we show that such reversals are in most cases either impossible or highly implausible. Paper 4, entitled "**Memories as anchors: Theoretical reflections and empirical analyses on the intrapersonal comparability of wellbeing reports**", explores the consequences of explanatory variables affecting how respondents report their wellbeing. The paper first shows that such scale-shift effects are to be expected and subsequently analyses the biases that result from them. The paper then proposes a method based on recollections of past wellbeing to correct for intrapersonal scale shifts. The proposed method is then applied to BHPS data, showing that unemployment and bereavement are likely to be associated with strong scale-shift effects.

Acknowledgements

I thank my doctoral supervisors - Brian Nolan and Maarten Vendrik - for their continuous support. In their very different ways, both have been wonderful teachers.

I particularly thank Brian for helping me to navigate academia, for structuring my thoughts, and for allowing me to see the bigger issues when lost in minutiae. I particularly thank Maarten for the amount of time and thought he gave to my work, and for pushing me to be much more rigorous than I would otherwise have been. I also thank him for being a co-author, and for initially sparking my interest in the topic of this thesis.

Writing a DPhil at the University of Oxford is a privilege. I therefore thank Nuffield College, the Department of Social Policy and Intervention, and the Wellbeing Research Centre for funding this project. I also thank the members of these institutions for their feedback, the intellectual exchanges, and for the ideas that were sparked by these.

Ironically - given the topic of this thesis - the last four years have been psychologically taxing. I therefore thank Anders Malthe Bach-Mortensen, Zep Kalb, Alistair Leitch, Michael Plant, Anton Wille, and Xiaochu Wu for lifting my spirits. I am especially grateful to Nhat An Trinh, who, as well as being a magnificent co-author, is a great partner and friend.

Finally, I thank Inge Schuett, Angela Schuett, and Matthias Kaiser. I would not be without them.

Contents

Introduction 4

Paper 1: People do not adapt. New analyses of the dynamic effects of own and reference income 24

Paper 2: Positional, mobility and reference effects: How does social class affect life satisfaction in Europe? 91

Paper 3: How threatening are transformations of happiness scales to subjective wellbeing research? 158

Paper 4: Memories as anchors: Theoretical reflections and empirical analyses on the intrapersonal comparability of wellbeing reports 207

Conclusions 286

Introduction

1 Background and motivation

This thesis contributes to the literature that uses responses to variants of the question *'How satisfied are you with your life, all things considered?'* as the main dependent variable. Such responses are called measures of 'life satisfaction'. Life satisfaction is a primarily cognitive component of subjective wellbeing. It is thought to be a reflective assessment of respondents about their lives as a whole (Pavot et al. 1991; Frey and Stutzer 2010; Luhmann et al. 2012).[1]

Using this dependent variable, paper 1 is concerned with short- and long-run dynamics in the effects of income. Paper 2 investigates how social class and social mobility may affect life satisfaction. Papers 3 and 4 consider more general methodological aspects of how reports on life satisfaction are analysed. Section 3 of this introduction provides a more detailed overview of the thesis.

All four papers are united in their concern with the question of what determines people's subjective life satisfaction. One motivation for this focus is the belief that evidence on the signs and magnitudes of determinants of reports of these judgements could, in principle, generate unambiguous policy recommendations.[2] In the context of social policy as an academic discipline, adding to the evidence that enables such robust policy recommendations is of paramount importance.

[1] In turn, subjective wellbeing may be defined as *'Good mental states, including all of the various evaluations, positive and negative, that people make of their lives, and the affective reactions of people to their experiences'* (OECD 2013). Next to life satisfaction, the literature also postulates the existence of an 'affective' (Pavot and Diener 2008; Arthaud-Day et al. 2005) and a 'eudaimonic' (Huppert et al. 2009; Huppert and So 2013; Deci and Ryan 2008; Clark and Senik 2011) component. Measures of affective wellbeing seek to directly gauge the valence and intensity of mental states at the time of their occurrence (Diener et al. 1999), while eudaimonic measures seek to capture wider (and arguably less well defined) notions of 'flourishing' (Deci and Ryan 2008).

[2] In addition, there were two practical reasons: (1) life satisfaction appears to be the most commonly available type of data on subjective wellbeing, (2) life satisfaction is more strongly correlated with socio-economic variables of the sort I was interested in (Kahneman and Deaton 2010; Luhmann, Schimmack, and Eid 2011).

The belief that life satisfaction can be used for such purposes is founded on two independent strands of literature – one empirical, and one philosophical. The empirical strand of the literature showed that responses to life satisfaction question are reliable, valid, and can be treated as measuring satisfaction cardinally. Regarding 'reliability', that literature showed that different ways of asking about a respondent's life satisfaction strongly correlate (Pavot et al. 1991; Pavot and Diener 2008) and that responses by the same respondent to the same question over relatively short time intervals are moderately correlated, too (Krueger and Schkade 2008; Michalos and Kahlke 2010; Lucas and Donnellan 2012). Regarding 'validity', the literature showed that individuals' reports correlate with ratings made by others (Schneider and Schimmack 2009), and that such reports correlate with objectively observable variables for which we expect such correlations. Among others, these variables include the numbers of positive and negative events respondents can recall about their lives (Pavot et al. 1991), local amenities (Oswald and Wu 2010), neural activity (Urry et al. 2004; Kong et al. 2015), age patterns in anti-depressant use (Blanchflower and Oswald 2016), or suicide risk (Koivumaa-Honkanen et al. 2001).[3] For an overview of this type of work, see Diener, Inglehart, and Tay (2013). Furthermore, a (small) literature showed that OLS regressions, which treat life satisfaction as cardinally measured, yield very similar results compared to more sophisticated techniques that (purport to)[4] only use the ordinal information contained in life satisfaction responses (Ferrer-i-Carbonell and Frijters 2004; Riedl and Geishecker 2014). It therefore seems that individuals' numbered *reports* about their life satisfaction can be innocuously employed as cardinal measures of individual's internally *felt* life satisfaction.

Some literature in philosophy suggests that life satisfaction, unlike other measures of subjective wellbeing or other variables, reliably captures the level of welfare of individuals. Sumner (1996) forcefully motivates this view. The core argument goes as follows. Assume that a subject's welfare

[3] However, Daly et al. (2011) find that places with higher average life satisfaction tend to be associated with higher suicide rates.

[4] I write 'purport to' for a reason that is explored in the third empirical chapter of the thesis: Ordered/probit logit models indeed do not assume any particular cardinal response scale. However, they do assume that respondents use the same cardinal scale - whichever it might be.

is fully determined by the subject in question. Call this subjectivism about welfare (Hawkins 2010; Tiberius 2015; Plant 2020).[5] The main motivation for subjectivism is that it is not paternalistic: Any theory of welfare that is not subjectivist may come to the repugnant conclusion that some candidate X is good for a subject S despite S vehemently denying that X is in fact good for S (Railton 1986). Put differently, if it is conceivable for some S to deny that some candidate X is good for S, then X cannot constitute welfare for all S. Most candidates fail this test. For example, the capabilities approach, as theorized by Sen (1985) and Nussbaum (2000), states that when the set of capabilities available to S is expanded, then S's welfare is increased. But, under subjectivism, if S were to deny that some particular expansion of the capability set indeed benefits S, then S cannot be benefited. As another example, hedonism states that welfare depends on the valence and intensity of a subject's mental states (Crisp 2006).[6] But it is conceivable for some S to have more positively valenced mental states and yet judge her welfare as worse. Thus, if subjectivism is true, hedonism cannot be a true theory of welfare. We may therefore be less interested in measures of affect.

However, if a respondent states that she is more satisfied with her life, she cannot also think that her life is worse for her (unless she uses language in an odd manner). Thus, life satisfaction, and only life satisfaction, seems to pass the subjectivist test (Plant 2020).[7] Of course, this does not imply that other variables do not track welfare at all. But, under subjectivism, such other variables only track welfare indirectly via their potential effects on life satisfaction.

[5] Arneson (1999) calls this the 'endorsement constraint' when discussing the views of Griffin (1986) and Dworkin (1990).

[6] Or, more classically, on quantities of pleasure and pain as in Bentham and Mill's versions of utilitarianism (Driver 2014).

[7] However Plant (2020) raises two serious objections against subjectivism. The first objection notes that under subjectivism, there is no logical constraint on subjects arbitrarily maximising (or otherwise altering) their level of welfare by merely deciding that they are maximally well off. The second objection is that subjectivism restricts the class of welfare subjects too tightly. Under subjectivism, any being that is incapable of judging the quality of her life as a whole, would not have a welfare level. However, we think that e.g. animals or young children do have some level of welfare despite being incapable of such judgements. Moreover, in a recent working paper Benjamin et al. (2020) show that respondents also take into account the welfare of others when responding to life satisfaction questions. Thus, measures of life satisfaction cannot only capture individual welfare.

If one further assumes that the sole goal of government is maximising expected welfare,[8] then evidence on the determinants of life satisfaction may indeed generate reliable policy recommendations. Although this intuition provides the overarching motivation for focusing on life satisfaction here, the thesis focuses on several more detailed aspects of the study of life satisfaction. The thesis therefore does not seek to offer a description of how such policy recommendations could be generated under what precise assumptions.[9] Nevertheless, each of the empirical chapters, which individually seek to answer more focused questions, offer considerations highly relevant to this wider endeavour. Papers 1 and 2 offer pertinent empirical conclusions regarding over-time dynamics and reference effects, while papers 3 and 4 offer more methodological considerations regarding the peculiarities of subjective reports about welfare.

2 Determinants of life satisfaction

Each of the empirical chapters reviews the literature relevant to that chapter's more specialised subject matter. However, a brief overview on the determinants of life satisfaction, and where this thesis seeks to contribute, may be helpful at the introductory stage. General literature reviews are given by e.g. Diener et al. (1999), Dolan, Peasgood, and White (2008), MacKerron (2012), Clark (2018), and Diener, Lucas, and Oishi (2018). Some important determinants of life satisfaction, which are reported in these reviews, are the following.

On the individual level, higher incomes increase satisfaction, albeit at a diminishing rate and with a potential satiation point (e.g. Layard, Mayraz, and Nickell 2008; Kahneman and Deaton 2010; Jebb et al. 2018).[10] Being female (Stevenson and Wolfers 2009; Helliwell, Nicole, and Wang 2015;

[8] Which is controversial. Among other things, that assumption entails that (I) overall goodness (economists might call this 'social welfare') only depends on the welfare of relevant subjects in a separable manner, and (II) that governments follow expected utility theory and are thus neither risk-seeking nor risk-averse.
By 'separable' I mean that the contribution of the welfare of one subject to overall goodness does not influence the contribution of another subject's welfare. Separability does not rule out that individuals' welfare levels causally depend on others' welfare levels (e.g. via altruistic or sadistic channels). Instead, separability only requires that overall goodness can be computed from individual welfare levels using an additive aggregation rule. For further details, see Broome (1991).

[9] See Frijters et al. (2020) for an attempt in that direction.

[10] However, see Oswald (2008) on the cautionary point that diminishing marginal returns of income to *reported* satisfaction need not imply diminishing returns to *felt* satisfaction. A related point is developed in paper 3 of this thesis.

Montgomery 2017), being married (Lucas 2005; Stutzer and Frey 2006), having close social relationships (Lelkes 2006), being physically healthy (Steptoe, Deaton, and Stone 2015; Ngamaba, Panagioti, and Armitage 2017), and being religious (Diener, Tay, and Myers 2011) are robustly associated with higher reported satisfaction. Having children also tends to increase satisfaction, though the evidence is less decisive (Cetre, Clark, and Senik 2016). Unemployment (Winkelmann and Winkelmann 1998; Lucas et al. 2004), commuting times (Ingenfeld, Wolbring, and Bless 2019), living in cities (as countries grow rich; Burger et al. 2020; Easterlin, Angelescu, and Zweig 2011), not being white (Stevenson and Wolfers 2008a), care responsibilities (Borg and Hallberg 2006), and disability (Lucas 2007; Powdthavee 2009) are variables consistently seen to be associated with lower satisfaction. Non-cognitive personality traits like the Big Five are also associated with satisfaction, with greater conscientiousness and extraversion having a positive association, and neuroticism having a negative effect (Schimmack et al. 2004; Weiss, Bates, and Luciano 2008). Recent evidence further shows that personality traits are an important moderator of the effects of other explanatory variables (Boyce, Wood, and Brown 2010; Budría and Ferrer-I-Carbonell 2019). The effect of age is U-shaped with highs in satisfaction in young and old age, and a trough in mid-life (Blanchflower and Oswald 2019).

There is some evidence that people adapt to individual circumstances like marriage, having children, incomes, and health (Clark and Georgellis 2013). However, paper 1 shows this to be unlikely in the case of income, which places some doubt on adaptation to other domains, too. Paper 4 argues that observed adaptation may in fact be driven by lagged scale-shift effects. Furthermore, social comparisons are also a determinant of life satisfaction. Most of this work is concerned with social comparisons in the domain of incomes (Clark and Oswald 1996; McBride 2001; Ferrer-i-Carbonell 2005; Vendrik and Woltjer 2007) and little is known about over-time dynamics in social comparison effects (see, however, Vendrik, 2013). In this context, paper 1 explores over-time dynamics in the effects of reference incomes. Paper 2 extends the study of

social comparisons to social class. Paper 4 studies the potential effects of social comparison on scale use.

On the macro-level, more generous welfare states (Di Tella, MacCulloch, and Oswald 2003; Radcliff 2013), greater availability of green spaces (Krekel, Kolbe, and Wüstemann 2016), clean air (Luechinger 2009; Levinson 2012), generalized social trust (Bjørnskov 2003), and low crime rates (Ambrey, Fleming, and Manning 2014) are among those variables found to improve satisfaction. Higher unemployment and inflation rates[11] tend to be robustly associated with lower levels of satisfaction (Di Tella, MacCulloch, and Oswald 2001; Blanchflower et al. 2014). Surprisingly, the evidence on the effects of income inequality is mixed. For example Alesina, Di Tella, and (MacCulloch 2004) find a negative effect of inequality (that is stronger in Europe), Knight, Song, and Gunatilaka (2009) find a positive effect for China (which they interpret as current inequality signalling future prosperity), and Bjørnskov, Dreher, and Fischer (2007) find no significant effect in global data (from before the 2008 crisis). Ifcher, Zarghamee, and Graham (2019) find that the effects of inequality depend on the size of geographical units under consideration. See Clark and d'Ambrosio (2015) for a comprehensive review.

A much-debated topic, to which most of the empirical chapters in the thesis refer in some form, is the effect of GDP per capita on average life satisfaction. In a seminal paper, Easterlin (1974) noticed that mean satisfaction in the USA did not increase over a time span of 40 years despite sustained increases in GDP per capita. Together with the observation that incomes tend to be associated with higher satisfaction in the cross-section at a point in time, this observation has become known as the 'Easterlin Paradox'. This paradox is a much debated (and yet unresolved) topic of research. While Veenhoven and Vergunst (2014) and Stevenson and Wolfers (2008) present evidence showing GDP per capita to be associated with higher mean levels of satisfaction, Easterlin (2016) presents evidence to the contrary. De Neve et al. (2018) show that growth in GDP

[11] With inflation having a much smaller effect than unemployment.

per capita has a much smaller effect than recessions. Kaiser and Vendrik (2019) note that there are several distinct versions of the paradox, and that the most plausible version of the paradox is more likely to be satisfied for richer than for poorer countries.

3 Thesis overview

The thesis is comprised of two primarily applied and two primarily methodological papers. On the applied level, the thesis is concerned with the interpersonal and intertemporal effects of socio-economic variables on subjective life satisfaction. On the methodological level, the thesis is concerned with exploring, investigating, and developing appropriate statistical techniques for the study of subjective data, particularly life satisfaction reports.

Paper 1: **People do not adapt. New Analyses of the dynamic effects of own and reference income** uses yearly data from the German Socio-Economic Panel and the UK Household Longitudinal Study to investigate whether, with respect to subjective life satisfaction, people adapt to the effects of own and reference household income. In the paper I define 'adaptation' to mean that the long-run effect of a variable is significantly smaller than the short-run effect. The opposite of adaptation is defined to be 'reinforcement', which occurs when the long-run effect is larger than the short-run effect. In order to identify the income effect, and following a previous study by Vendrik (2013), I instrument own income with yearly occupation- and industry-wide variation in labour earnings.

Surprisingly, I do not find any credible evidence of adaptation to own income. I further find weak evidence suggesting reinforcement of the reference income effect. This contrasts with much of the previous literature. I therefore attempt to give an in-depth explanation of why previous studies found evidence of adaptation. That explanation has two parts. First, in contrast to e.g. Di Tella, Haisken-De New, and MacCulloch (2010), I control not only for contemporaneous values of explanatory variables, but also their lagged values. It turns out that the effects of the number of adults and children in the household are subject to reinforcement and are negative in the long run.

When failing to control for such lags, lags for household income pick up these negative effects. This yields results which indicate adaptation to household income when in fact there is none. Second, and in contrast to Vendrik (2013), I control for whether a respondent's partner is unemployed. Partner unemployment is strongly correlated with household income. Moreover, partner unemployment has a strong negative effect that is subject to reinforcement, too. Controlling for such reinforcement results in equal short-run and long-run estimates of the effect of own income.

The paper also demonstrates that previous studies which estimated equivalence scales on the basis of life satisfaction data are likely misleading. This is because these studies exclusively estimated equivalence scales on the basis of contemporaneous values of household composition and income. When taking into account reinforcement of the negative household composition effect, I show that equivalence scales like those recommended by the OECD probably overestimate the long-run benefits of resource pooling within households.

A version of this paper has been published in the *Journal of Economic Behavior & Organization* and is publicly available.[12]

Paper 2 (with Nhat An Trinh): **Positional, mobility and reference effects: How does social class affect life satisfaction in Europe?** is more sociological in orientation. We investigate the possible pathways by which social class affects life satisfaction across European regions. To do so, we use the first five waves of the European Social Survey. That survey includes detailed information on respondents' and their parents' occupation. On the basis of that information we are able to assign each respondent and their parents to one of nine social classes distinguished by the European Socio-Economic Classification (ESeC), which is designed for cross-country class analyses. With these assignments, we can observe each respondent's trajectory in intergenerational social mobility. In addition, we construct measures of 'reference class position' and 'reference class mobility'. The

[12] See: https://osf.io/preprints/socarxiv/qtgbn/.

former measures which class destination (i.e. own class) people from the same class origin (i.e. parents' class) as the respondent obtain on average. The latter measures measure the average amount of class mobility experienced by people in the same class destination as the respondent.

Having thus obtained measures of class origin, class destination, own class mobility, reference class position, and reference class mobility, we then attempt to give an integrated account of how each of these may affect life satisfaction. We distinguish between Eastern and Western Europe in our analysis. We find that class destination and class origin are both major determinants of life satisfaction across both European regions. Compared to other social classes, classes related to self-employment are associated with relatively higher satisfaction in Eastern than in Western Europe. Reference class mobility has a robust negative effect on life satisfaction across both Western and Eastern Europe, though reference class position yields no significant effects. Own class mobility has a positive effect, which is larger and more robust in the case of Eastern Europe.

Methodologically we note that when attempting to jointly estimate the effects of class origin, class destination and social mobility, we face the issue that social mobility is perfectly collinear with origins and destinations. In order to get around this problem, we make use of so-called 'diagonal reference models' (Sobel 1981). Since the assumptions underlying that approach are somewhat opaque, and not well understood in the literature, a portion of the paper is devoted to explicating and analysing the potentially strong assumptions incorporated into this model. Finally, since no relevant and accessible Stata package existed, a by-product of this paper was the creation of '*drm*'. That Stata package allows for the estimation of different variants of the diagonal reference model and is publicly accessible to the research community.[13]

A version of this paper has been accepted for publication at the *European Sociological Review* and is publicly available.[14]

[13] See: https://ideas.repec.org/c/boc/bocode/s458506.html.
[14] See: https://osf.io/preprints/socarxiv/4adhe/.

Paper 3 (with Maarten Vendrik): **How threatening are transformations of happiness scales to subjective wellbeing research?** was initially motivated by a paper published in the *Journal of Political Economy* (Bond and Lang 2019, 'The Sad Truth about Happiness Scales'). Another paper by Schröder and Yitzhaki (2017, 'Revisiting the evidence for cardinal treatment of ordinal variables', *European Economic Review*) provided further motivation. These papers show that transformations of the underlying wellbeing scales can lead to reversals of the signs of estimated effects of explanatory variables on wellbeing. If their arguments held generally, much of the empirical literature on the determinants of subjective wellbeing would be fatally undermined. In essence, both papers are based on noting that even when assuming that all respondents use the response scale in the same manner, typical wellbeing data may be seen to only contain ordinal information. As a consequence, so the authors argue, any positive monotonic transformation of that data is equally permissible. Bond & Lang then prove that in the context of their ordered probit model, there always exists a positive monotonic transformation of the latent index that leads to a reversal of the sign of the effect of an explanatory variable. Schröder and Yitzhaki (2017) show that such reversals are also possible in the context of OLS regressions.

In response to these arguments, paper 3 makes four contributions. First, it provides a clearer explanation of the mechanics and intuitions behind Bond & Lang's approach to achieving reversals of signs of effects of explanatory variables. Second, it shows that the transformations required to achieve reversals imply that respondents use the response scale in a strongly non-linear fashion. It then sets out theoretical arguments and empirical evidence which suggest that such behavior is implausible, and that respondents are more likely to use the response scale in a roughly linear fashion. Second, it shows that these reversals are fundamentally driven by heterogeneities in the effects of explanatory variables across the distribution of reported wellbeing. In the context of OLS regressions, the paper then derives a simple test of whether reversals are possible by relabelling the

scores of reported wellbeing.[15] In addition it deduces bounds for ratios of coefficients under any labelling scheme. Fourth, using German Socio-Economic Panel data, it provides additional empirical evidence on whether reversals of effects of standard demographic variables are both possible and plausible. It turns out that reversals by either relabelling or by using Bond & Lang's approach are impossible or implausible for almost all variables of interest.

An expanded version of the paper is publicly available[16].

Paper 4: **Memories as anchors: Theoretical reflections and empirical analyses on the intrapersonal comparability of wellbeing reports** explores the consequences of violating the assumption that the response scale is used in a common manner across respondents and over time. Call this the assumption of *common scale use*. That assumption, like the assumption of *linearity of scale use* explored in paper 3, is commonly adopted in the literature without much argument about its validity and with little exploration of its importance.

Against this background, Paper 4 makes four contributions. First, it analyses the direction and extent of bias that results from violations of the *common scale use* assumption. It shows that violations of that assumption make it possible that coefficient estimates obtained from standard analyses have incorrect signs. Moreover, when violations of *common scale use* do not lead to incorrectly estimated signs of coefficients, coefficient magnitudes as well their ratios with other coefficients (which are needed for e.g. cost-benefit analyses) are shown to nevertheless be biased in most circumstances.

Second, it argues that violations of *common scale use* are likely to occur and notes that in the case of contemporaneous personal characteristics, it is most likely that variables which increase (decrease)

[15] In most empirical research (including papers 1 and 2), the ordinal data are recorded in their rank order. In other words, the first response option is coded as '1', the second response option is coded as '2', and so on. In our use of the term, to 'relabel' is to give different numerical values to each response option. We might for example give a '1' to the first option, a '4', to the second, an '9' to third, a '16' to the fourth, and so on. If we call the variable under the original labelling hr, then this relabelling would be a positive monotonic transformation given by $\widetilde{hr} = hr^2$.

[16] See: https://osf.io/gzt7a/.

satisfaction, are also likely to cause respondents to be more (less) stringent in their scale use.[17] Absolute magnitudes of such variables are therefore likely to be biased towards zero in analyses that do not account for scale-shift effects. In contrast, variables relating to reference groups and lagged values of covariates may be expected to cause scale shift effects with signs opposite to these variables' effects on latent wellbeing. Analyses that do not account for such shifts may therefore yield coefficients that are biased away from zero, thus potentially overstating the importance of adaptation and reference effects.

Third, I propose a method to correct for differences in scale use. That method attempts to identify how respondents' scale use depends on contemporaneous covariates by investigating inconsistencies between respondents' memories and their reports of past satisfaction. I discuss the plausibility of the method's identifying assumptions and derive the direction of bias in cases where these assumptions are violated.

Fourth, using that proposed method, in conjunction with a little-known method advanced by Van Praag (2015) to account for fixed effects, I utilize British Household Panel Survey data to investigate whether intrapersonal scale shifts effects indeed occur. I find that the direction in which contemporaneous explanatory variables affect latent satisfaction is indeed typically the same as the direction in which scale use is affected. Unemployment and bereavement appear to have particularly strong effects on scale use. Nevertheless, compared to an analysis that maintains common scale use, differences in results are never so large as to change the qualitative conclusions. This is good news for previous and future wellbeing research, as it suggests a minimum level of stability in the way people use response scales.

An expanded version of the paper is publicly available[18].

[17] By 'more stringent' I here mean that respondents answer with a comparatively lower response option for a given level of latent satisfaction than somebody who is less stringent in their scale use.
[18] See: https://osf.io/xpghn.

References

Alesina, Alberto, Rafael Di Tella, and Robert MacCulloch. 2004. 'Inequality and Happiness: Are Europeans and Americans Different?' *Journal of Public Economics* 88 (9–10): 2009–2042. https://doi.org/10.1016/j.jpubeco.2003.07.006.

Ambrey, Christopher L., Christopher M. Fleming, and Matthew Manning. 2014. 'Perception or Reality, What Matters Most When It Comes to Crime in Your Neighbourhood?' *Social Indicators Research* 119 (2): 877–96. https://doi.org/10.1007/s11205-013-0521-6.

Arneson, Richard J. 1999. 'Human Flourishing Versus Desire Satisfaction'. *Social Philosophy and Policy* 16 (1): 113–42. https://doi.org/10.1017/S0265052500002272.

Arthaud-Day, Marne L., Joseph C. Rode, Christine H. Mooney, and Janet P. Near. 2005. 'The Subjective Well-Being Construct: A Test of Its Convergent, Discriminant, and Factorial Validity'. *Social Indicators Research* 74 (3): 445–76.

Benjamin, Daniel J, Jakina Debnam, Marc Fleurbaey, Ori Heffetz, and Miles Kimball. 2020. 'What Do Happiness Data Mean? Evidence from a Survey of Happiness Respondents'. *Working Paper*, 1–41.

Bjørnskov, Christian. 2003. 'The Happy Few. Cross-Country Evidence on Social Capital and Life Satisfaction'. *Kyklos* 56 (1): 3–16.

Bjørnskov, Christian, Axel Dreher, and Justina A. V. Fischer. 2007. 'Cross-Country Determinants of Life Satisfaction: Exploring Different Determinants across Groups in Society'. *Social Choice and Welfare* 30 (1): 119–73. https://doi.org/10.1007/s00355-007-0225-4.

Blanchflower, David G., David NF Bell, Alberto Montagnoli, and Mirko Moro. 2014. 'The Happiness Trade-off between Unemployment and Inflation'. *Journal of Money, Credit and Banking* 46 (S2): 117–141.

Blanchflower, David G., and Andrew J. Oswald. 2016. 'Antidepressants and Age: A New Form of Evidence for U-Shaped Well-Being through Life'. *Journal of Economic Behavior & Organization* 127: 46–58.

———. 2019. 'Do Humans Suffer a Psychological Low in Midlife? Two Approaches (with and without Controls) in Seven Data Sets'. In *The Economics of Happiness*, edited by Mariano Rojas, 439–453. Springer.

Bond, Timothy N., and Kevin Lang. 2019. 'The Sad Truth about Happiness Scales'. *Journal of Political Economy* 127 (4): 1629–40. https://doi.org/10.1086/701679.

Borg, Christel, and Ingalill R. Hallberg. 2006. 'Life Satisfaction among Informal Caregivers in Comparison with Non-Caregivers'. *Scandinavian Journal of Caring Sciences* 20 (4): 427–38. https://doi.org/10.1111/j.1471-6712.2006.00424.x.

Boyce, Christopher J., Alex M. Wood, and Gordon D A Brown. 2010. 'The Dark Side of Conscientiousness: Conscientious People Experience Greater Drops in Life Satisfaction Following Unemployment'. *Journal of Research in Personality* 44 (4): 535–539. https://doi.org/10.1016/j.jrp.2010.05.001.

Broome, John. 1991. *Weighing Goods: Equality, Uncertainty and Time*. John Wiley & Sons.

Budría, Santi, and Ada Ferrer-I-Carbonell. 2019. 'Life Satisfaction, Income Comparisons and Individual Traits'. *Review of Income and Wealth* 65 (2): 337–357.

Burger, Martijn, Philip Morrison, Martijn Hendriks, and Marloes Hoogerbrugge. 2020. 'Urban-Rural Happiness Differentials across the World'. *World Happiness Report*, 67–94.

Cetre, Sophie, Andrew E. Clark, and Claudia Senik. 2016. 'Happy People Have Children: Choice and Self-Selection into Parenthood'. *European Journal of Population* 32 (3): 445–473.

Clark, Andrew E. 2018. 'Four Decades of the Economics of Happiness: Where Next?' *Review of Income and Wealth* 64 (2): 245–69. https://doi.org/10.1111/roiw.12369.

Clark, Andrew E., and Conchita d'Ambrosio. 2015. 'Attitudes to Income Inequality: Experimental and Survey Evidence'. In *Handbook of Income Distribution*, 2:1147–1208. Elsevier.

Clark, Andrew E., and Yannis Georgellis. 2013. 'Back to Baseline in Britain: Adaptation in the British Household Panel Survey'. *Economica* 80 (319): 496–512. https://doi.org/10.1111/ecca.12007.

Clark, Andrew E., and Andrew J. Oswald. 1996. 'Satisfaction and Comparison Income'. *Journal of Public Economics* 61 (3): 359–381. https://doi.org/10.1016/0047-2727(95)01564-7.

Clark, Andrew E., and Claudia Senik. 2011. 'Is Happiness Different from Flourishing? Cross-Country Evidence from the ESS'. *Revue d'economie Politique*, 17–34.

Crisp, Roger. 2006. 'Hedonism Reconsidered'. *Philosophy and Phenomenological Research* 73 (3): 619–45. https://doi.org/10.2307/40041013.

Daly, Mary C., Andrew J. Oswald, Daniel Wilson, and Stephen Wu. 2011. 'Dark Contrasts: The Paradox of High Rates of Suicide in Happy Places'. *Journal of Economic Behavior & Organization* 80 (3): 435–42. https://doi.org/10.1016/j.jebo.2011.04.007.

De Neve, Jan-Emmanuel, George Ward, Femke De Keulenaer, Bert Van Landeghem, Georgios Kavetsos, and Michael I. Norton. 2018. 'The Asymmetric Experience of Positive and Negative Economic Growth: Global Evidence Using Subjective Well-Being Data'. *Review of Economics and Statistics* 100 (2): 362–375.

Deci, Edward L., and Richard M. Ryan. 2008. 'Hedonia, Eudaimonia, and Well-Being: An Introduction'. *Journal of Happiness Studies* 9 (1): 1–11.

Di Tella, Rafael, John Haisken-De New, and Robert MacCulloch. 2010. 'Happiness Adaptation to Income and to Status in an Individual Panel'. *Journal of Economic Behavior and Organization* 76 (3): 834–852. https://doi.org/10.1016/j.jebo.2010.09.016.

Di Tella, Rafael, Robert J. MacCulloch, and Andrew J. Oswald. 2001. 'Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness'. *American Economic Review* 91 (1): 335–341.

———. 2003. 'The Macroeconomics of Happiness'. *Review of Economics and Statistics* 85 (4): 809–827. https://doi.org/10.1162/003465303772815745.

Diener, Ed, Ronald Inglehart, and Louis Tay. 2013. 'Theory and Validity of Life Satisfaction Scales'. *Social Indicators Research* 112 (3): 497–527. https://doi.org/10.1007/s11205-012-0076.

Diener, Ed, Richard E. Lucas, and Shigehiro Oishi. 2018. 'Advances and Open Questions in the Science of Subjective Well-Being'. *Collabra: Psychology* 4 (1): 15. https://doi.org/10.1525/collabra.115.

Diener, Ed, Eunkook Suh, Richard Lucas, and Heidi Smith. 1999. 'Subjective Well-Being: Three Decades of Progress.' *Psychological Bulletin* 125 (2): 276.

Diener, Ed, Louis Tay, and David G. Myers. 2011. 'The Religion Paradox: If Religion Makes People Happy, Why Are so Many Dropping Out?' *Journal of Personality and Social Psychology* 101 (6): 1278.

Dolan, Paul, Tessa Peasgood, and Mathew White. 2008. 'Do We Really Know What Makes Us Happy? A Review of the Economic Literature on the Factors Associated with Subjective Well-Being'. *Journal of Economic Psychology* 29 (1): 94–122. https://doi.org/10.1016/j.joep.2007.09.001.

Driver, Julia. 2014. 'The History of Utilitarianism'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2014. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2014/entries/utilitarianism-history/.

Dworkin, Ronald. 1990. 'Foundations of Liberal Equality'. In *The Tanner Lectures on Human Values Vol. 11*, 1–119. Salt Lake City: University of Utah Press.

Easterlin, Richard. 1974. 'Does Economic Growth Improve the Human Lot? Some Empirical Evidence'. *Nations and Households in Economic Growth* 89: 89–125.

———. 2016. 'Paradox Lost?' *IZA Discussion Paper*, no. 9676.

Easterlin, Richard A., Laura Angelescu, and Jacqueline S. Zweig. 2011. 'The Impact of Modern Economic Growth on Urban–Rural Differences in Subjective Well-Being'. *World Development* 39 (12): 2187–2198.

Ferrer-i-Carbonell, Ada. 2005. 'Income and Well-Being: An Empirical Analysis of the Comparison Income Effect'. *Journal of Public Economics* 89 (5–6): 997–1019. https://doi.org/10.1016/j.jpubeco.2004.06.003.

Ferrer-i-Carbonell, Ada, and Paul Frijters. 2004. 'How Important Is Methodology for the Estimates of the Determinants of Happiness?' *The Economic Journal* 114 (497): 641–659.

Frey, Bruno S., and Alois Stutzer. 2010. *Happiness and Economics: How the Economy and Institutions Affect Human Well-Being*. Princeton University Press. https://books.google.co.uk/books?hl=en&lr=&id=XD8-H35mspoC&oi=fnd&pg=PP1&dq=frey+stutzer+happiness+and+economics&ots=SRUyKTWZLq&sig=lEIFEMw4xmZaytKvmnjoztOhLxY.

Frijters, Paul, Andrew E. Clark, Christian Krekel, and Richard Layard. 2020. 'A Happy Choice: Wellbeing as the Goal of Government'. *Behavioural Public Policy* 4 (2): 126–165.

Griffin, James. 1986. *Well-Being: Its Meaning, Measurement and Moral Importance*. Clarendon Press.

Hawkins, Jennifer S. 2010. 'The Subjective Intuition'. *Philosophical Studies* 148 (1): 61–68. https://doi.org/10.1007/s11098-010-9505-4.

Helliwell, John F., Fortin Nicole, and Shun Wang. 2015. 'How Does Subjective Wellbeing Vary around the World by Gender and Age?' *World Happiness Report Ch.3* 2015: 42–75.

Huppert, Felicia A., Nic Marks, Andrew E. Clark, Johannes Siegrist, Alois Stutzer, Joar Vittersø, and Morten Wahrendorf. 2009. 'Measuring Well-Being across Europe: Description of the ESS Well-Being Module and Preliminary Findings'. *Social Indicators Research* 91 (3): 301–315.

Huppert, Felicia A., and Timothy T C So. 2013. 'Flourishing Across Europe: Application of a New Conceptual Framework for Defining Well-Being'. *Social Indicators Research* 110 (3): 837–861. https://doi.org/10.1007/s11205-011-9966-7.

Ifcher, John, Homa Zarghamee, and Carol Graham. 2019. 'Income Inequality and Well-Being in the U.S.: Evidence of Geographic-Scale- and Measure-Dependence'. *The Journal of Economic Inequality* 17 (3): 415–34. https://doi.org/10.1007/s10888-018-9404-z.

Ingenfeld, Julia, Tobias Wolbring, and Herbert Bless. 2019. 'Commuting and Life Satisfaction Revisited: Evidence on a Non-Linear Relationship'. *Journal of Happiness Studies* 20 (8): 2677–2709. https://doi.org/10.1007/s10902-018-0064-2.

Jebb, Andrew T., Louis Tay, Ed Diener, and Shigehiro Oishi. 2018. 'Happiness, Income Satiation and Turning Points around the World'. *Nature Human Behaviour* 2 (1): 33. https://doi.org/10.1038/s41562-017-0277-0.

Kahneman, Daniel, and Angus Deaton. 2010. 'High Income Improves Evaluation of Life but Not Emotional Well-Being.' *PNAS* 107 (38): 16489–93. https://doi.org/10.1073/pnas.1011492107.

Kaiser, Caspar, and Maarten CM Vendrik. 2019. 'Different Versions of the Easterlin Paradox: New Evidence for European Countries'. In *The Economics of Happiness*, edited by Mariano Rojas, 27–55. Springer.

Knight, John, Lina Song, and Ramani Gunatilaka. 2009. 'Subjective Well-Being and Its Determinants in Rural China'. *China Economic Review*, Symposium on Health Economics Issues in China, 20 (4): 635–49. https://doi.org/10.1016/j.chieco.2008.09.003.

Koivumaa-Honkanen, Heli, Risto Honkanen, Heimo Viinamäki, Kauko Heikkilä, Jaakko Kaprio, and Markku Koskenvuo. 2001. 'Life Satisfaction and Suicide: A 20-Year Follow-Up Study'. *American Journal of Psychiatry* 158 (3): 433–39. https://doi.org/10.1176/appi.ajp.158.3.433.

Kong, Feng, Ke Ding, Zetian Yang, Xiaobin Dang, Siyuan Hu, Yiying Song, and Jia Liu. 2015. 'Examining Gray Matter Structures Associated with Individual Differences in Global Life Satisfaction in a Large Sample of Young Adults'. *Social Cognitive and Affective Neuroscience* 10 (7): 952–60. https://doi.org/10.1093/scan/nsu144.

Krekel, Christian, Jens Kolbe, and Henry Wüstemann. 2016. 'The Greener, the Happier? The Effect of Urban Land Use on Residential Well-Being'. *Ecological Economics* 121: 117–127.

Krueger, Alan B., and David A. Schkade. 2008. 'The Reliability of Subjective Well-Being Measures'. *Journal of Public Economics* 92 (8–9): 1833–1845.

Layard, Richard, Guy Mayraz, and Stephen Nickell. 2008. 'The Marginal Utility of Income'. *Journal of Public Economics* 92 (8): 1846–1857.

Lelkes, Orsolya. 2006. 'Knowing What Is Good for You: Empirical Analysis of Personal Preferences and the "Objective Good"'. *The Journal of Socio-Economics*, The Socio-Economics of Happiness, 35 (2): 285–307. https://doi.org/10.1016/j.socec.2005.11.002.

Levinson, Arik. 2012. 'Valuing Public Goods Using Happiness Data: The Case of Air Quality'. *Journal of Public Economics* 96 (9): 869–80. https://doi.org/10.1016/j.jpubeco.2012.06.007.

Lucas, Richard E. 2005. 'Time Does Not Heal All Wounds: A Longitudinal Study of Reaction and Adaptation to Divorce'. *Psychological Science* 16 (12): 945–50. https://doi.org/10.1111/j.1467-9280.2005.01642.x.

———. 2007. 'Adaptation and the Set-Point Model of Subjective Well-Being: Does Happiness Change after Major Life Events?' *Current Directions in Psychological Science* 16 (2): 75–79.

Lucas, Richard E., Andrew E. Clark, Yannis Georgellis, and Ed Diener. 2004. 'Unemployment Alters the Set Point for Life Satisfaction'. *Psychological Science* 15 (1): 8–13.

Lucas, Richard E., and M. Brent Donnellan. 2012. 'Estimating the Reliability of Single-Item Life Satisfaction Measures: Results from Four National Panel Studies'. *Social Indicators Research* 105 (3): 323–331.

Luechinger, Simon. 2009. 'Valuing Air Quality Using the Life Satisfaction Approach'. *The Economic Journal* 119 (536): 482–515.

Luhmann, Maike, Louise C. Hawkley, Michael Eid, and John T. Cacioppo. 2012. 'Time Frames and the Distinction between Affective and Cognitive Well-Being'. *Journal of Research in Personality* 46 (4): 431–441. https://doi.org/10.1016/j.jrp.2012.04.004.

Luhmann, Maike, Ulrich Schimmack, and Michael Eid. 2011. 'Stability and Variability in the Relationship between Subjective Well-Being and Income'. *Journal of Research in Personality* 45 (2): 186–197.

MacKerron, George. 2012. 'Happiness Economics from 35 000 Feet'. *Journal of Economic Surveys* 26 (4): 705–35. https://doi.org/10.1111/j.1467-6419.2010.00672.x.

McBride, M. 2001. 'Relative-Income Effects on Subjective Well-Being in the Cross-Section'. *Journal of Economic Behavior & Organization* 45 (3): 251–278. https://doi.org/10.1016/S0167-2681(01)00145-7.

Michalos, Alex C., and P. Maurine Kahlke. 2010. 'Stability and Sensitivity in Perceived Quality of Life Measures: Some Panel Results'. *Social Indicators Research* 98 (3): 403–434.

Montgomery, Mallory. 2017. 'Reversing the Gender Gap in Happiness: Validating the Use of Life Satisfaction Self-Reports Worldwide'. *Working Paper*.

Ngamaba, Kayonda Hubert, Maria Panagioti, and Christopher J. Armitage. 2017. 'How Strongly Related Are Health Status and Subjective Well-Being? Systematic Review and Meta-Analysis'. *European Journal of Public Health* 27 (5): 879–85. https://doi.org/10.1093/eurpub/ckx081.

Nussbaum, Martha Craven. 2000. *Women and Human Development: The Capabilities Approach*. Cambridge University Press.

OECD. 2013. *OECD Guidelines on Measuring Subjective Well-Being*. https://doi.org/10.1787/9789264191655-en.

Oswald, A. J., and S. Wu. 2010. 'Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A.' *Science* 327 (5965): 576–79. https://doi.org/10.1126/science.1180606.

Oswald, Andrew J. 2008. 'On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings'. *Economics Letters* 100 (3): 369–72. https://doi.org/10.1016/j.econlet.2008.02.032.

Pavot, William, E. D. Diener, C. Randall Colvin, and Ed Sandvik. 1991. 'Further Validation of the Satisfaction with Life Scale: Evidence for the Cross-Method Convergence of Well-Being Measures'. *Journal of Personality Assessment* 57 (1): 149–161.

Pavot, William, and Ed Diener. 2008. 'The Satisfaction With Life Scale and the Emerging Construct of Life Satisfaction'. *The Journal of Positive Psychology* 3 (2): 137–52. https://doi.org/10.1080/17439760701756946.

Plant, Michael. 2020. 'Life Satisfaction and Its Discontents'. *Happier Lives Institute Working Paper*. https://www.happierlivesinstitute.org/uploads/1/0/9/9/109970865/life_satisfaction_and_its_discontents_jul2020.pdf.

Powdthavee, Nattavudh. 2009. 'What Happens to People before and after Disability? Focusing Effects, Lead Effects, and Adaptation in Different Areas of Life'. *Social Science & Medicine*, Part Special Issue: New approaches to researching patient safety, 69 (12): 1834–44. https://doi.org/10.1016/j.socscimed.2009.09.023.

Radcliff, Benjamin. 2013. *The Political Economy of Human Happiness How Voters ' Choices Determine the Quality of Life*. https://doi.org/10.1017/CBO9781139344371.

Railton, Peter. 1986. 'Facts and Values'. *Philosophical Topics* 14 (2): 5–31.

Riedl, Maximilian, and Ingo Geishecker. 2014. 'Keep It Simple: Estimation Strategies for Ordered Response Models with Fixed Effects'. *Journal of Applied Statistics* 41 (11): 2358–74. https://doi.org/10.1080/02664763.2014.909969.

Schimmack, Ulrich, Shigehiro Oishi, R. Michael Furr, and David C. Funder. 2004. 'Personality and Life Satisfaction: A Facet-Level Analysis'. *Personality and Social Psychology Bulletin* 30 (8): 1062–75. https://doi.org/10.1177/0146167204264292.

Schneider, Leann, and Ulrich Schimmack. 2009. 'Self-Informant Agreement in Well-Being Ratings: A Meta-Analysis'. *Social Indicators Research* 94 (3): 363.

Schröder, Carsten, and Shlomo Yitzhaki. 2017. 'Revisiting the Evidence for Cardinal Treatment of Ordinal Variables'. *European Economic Review* 92 (February): 337–58. https://doi.org/10.1016/j.euroecorev.2016.12.011.

Sen, Amartya. 1985. *Commodities and Capabilities*. Amsterdam: North-Holland.

Sobel, Michael E. 1981. 'Diagonal Mobility Models: A Substantively Motivated Class of Designs for the Analysis of Mobility Effects'. *American Sociological Review* 46 (6): 893–906. https://doi.org/10.2307/2095086.

Steptoe, Andrew, Angus Deaton, and Arthur A. Stone. 2015. 'Subjective Wellbeing, Health, and Ageing'. *The Lancet* 385 (9968): 640–48. https://doi.org/10.1016/S0140-6736(13)61489-0.

Stevenson, Betsey, and Justin Wolfers. 2008a. 'Happiness Inequality in the United States'. *The Journal of Legal Studies* 37 (S2): S33–S79.

———. 2008b. 'Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox'. Working Paper 14282. National Bureau of Economic Research. https://doi.org/10.3386/w14282.

———. 2009. 'The Paradox of Declining Female Happiness'. *American Economic Journal: Economic Policy* 1 (2): 190–225. https://doi.org/10.1257/pol.1.2.190.

Stutzer, Alois, and Bruno S. Frey. 2006. 'Does Marriage Make People Happy, or Do Happy People Get Married?' *The Journal of Socio-Economics* 35 (2): 326–347.

Sumner, L. W. 1996. *Welfare, Happiness, and Ethics*. Oxford University Press.

Tiberius, Valerie. 2015. 'Prudential Value'. *The Oxford Handbook of Value Theory*, 158–174.

Urry, Heather L., Jack B. Nitschke, Isa Dolski, Daren C. Jackson, Kim M. Dalton, Corrina J. Mueller, Melissa A. Rosenkranz, Carol D. Ryff, Burton H. Singer, and Richard J. Davidson. 2004. 'Making a Life Worth Living: Neural Correlates of Well-Being'. *Psychological Science* 15 (6): 367–372.

Van Praag, Bernard M. S. 2015. 'A New View on Panel Econometrics: Is Probit Feasible After All?' *IZA Discussion Paper*, no. 9345.

Veenhoven, Ruut, and Floris Vergunst. 2014. 'The Easterlin Illusion: Economic Growth Does Go with Greater Happiness'. *International Journal of Happiness and Development* 1 (4): 311–343.

Vendrik, Maarten C. M. 2013. 'Adaptation, Anticipation and Social Interaction in Happiness: An Integrated Error-Correction Approach'. *Journal of Public Economics* 105: 131–149. https://doi.org/10.1016/j.jpubeco.2013.06.009.

Vendrik, Maarten C. M., and Geert B. Woltjer. 2007. 'Happiness and Loss Aversion: Is Utility Concave or Convex in Relative Income?' *Journal of Public Economics* 91 (7–8): 1423–1448. https://doi.org/10.1016/j.jpubeco.2007.02.008.

Weiss, Alexander, Timothy C. Bates, and Michelle Luciano. 2008. 'Happiness Is a Personal(Ity) Thing: The Genetics of Personality and Well-Being in a Representative Sample'. *Psychological Science* 19 (3): 205–10. https://doi.org/10.1111/j.1467-9280.2008.02068.x.

Winkelmann, Liliana, and Rainer Winkelmann. 1998. 'Why Are the Unemployed so Unhappy? Evidence from Panel Data'. *Economica* 65 (257): 1–15.

People do not adapt. New analyses of the dynamic effects of own and reference income on life satisfaction†

Caspar Kaiser

Abstract

Do people adapt to changes in income? In contradiction to much of the previous literature, I find no evidence of adaptation to income in GSOEP (1984–2015) and UKHLS (1996–2017) data. Furthermore, I find that people also do not adapt to changes in reference income. Instead, reference income effects may be subject to reinforcement over time. Following the empirical approach of Vendrik (2013), I obtain these findings by estimating life satisfaction equations in which contemporaneous and lagged terms for a respondent's own household income and their estimated reference income are simultaneously entered. Additionally, I instrument for own income and include lags of a large set of controls. What was found to be adaptation to raw household income in previous studies turns out to have been driven by reinforcement of an initially small negative effect of household size that grows large over time. Implications of this result for the estimation of equivalence scales with subjective data are discussed.

Keywords: *subjective wellbeing; life satisfaction; adaptation; income comparisons; equivalence scales; GSOEP; BHPS; Understanding Society; UKHLS*

† This paper is based on an earlier manuscript available on the Open Science Framework (see https://osf.io/qtgbn/). A version of this paper has been published in the Journal of Economic Behavior & Organization (see: https://doi.org/10.1016/j.jebo.2020.06.003).

1 Introduction

When people earn more, they tend to become more satisfied with their lives. However, with time, people may grow accustomed to higher incomes. As a consequence, people's life satisfaction tends to decline. This phenomenon is commonly called adaptation. In the present study, I show that no significant adaptation to income can be observed in two of the datasets (GSOEP and UKHLS[1]) that are most commonly used in the study of life satisfaction.

With this paper I follow a large literature that evaluates whether people adapt to various life events more generally (for studies on adaptation in non-pecuniary domains, see e.g. Angeles 2010; Clark et al. 2008; Hanglberger and Merz 2011). There are some previous studies that investigate whether we adapt to income in life satisfaction (Di Tella, Haisken-De New, and MacCulloch 2010; Di Tella and MacCulloch 2010; Vendrik 2013; Ferrer-i-Carbonell and Van Praag 2008; Wolbring, Keuschnigg, and Negele 2013). However, those studies do not form a consensus. Despite this lack of consensus, a number of authors assume that adaptation to income in life satisfaction has been empirically established (Graham and Nikolova 2015; Jayawickreme and Pawelski 2012; Nussbaum 2000; Qizilbash 2006; Schokkaert 2007; Sen 1985). Furthermore, adaptation has previously been put forward as a possible explanation for the well-known and much debated Easterlin Paradox (Clark 2016; Clark, Frijters, and Shields 2008; Easterlin 2016). Thus, there is some importance to assessing whether adaptation to income in life satisfaction does in fact occur.

Before doing so, it is useful to clarify my terminology. By adaptation to income I mean that the absolute size of the long run[2] effect of income on life satisfaction is significantly smaller than the absolute size of the contemporaneous effect of income on life satisfaction. I further distinguish between partial and full adaptation. Partial adaptation refers to a situation in which the absolute size of the long run effect is indeed significantly smaller, but where the long run effect is

[1] Which is a combination of the older British Household Panel Study (BHPS) and its successor Understanding Society.
[2] I take the long run to be a situation in which income and life satisfaction are unchanging and in equilibrium with respect to each other.

nevertheless significant and of the same sign. In contrast, full adaptation occurs in situations where the contemporaneous effect is significantly larger than the long run effect, and where the long run effect is close to zero and insignificant. Following Vendrik (2013), I call the opposite of adaptation 'reinforcement'. Reinforcement occurs when long- and short run effects have the same sign and when the absolute size of the long run effect of income on life satisfaction is significantly larger than the absolute size of the short run effect. An illustration of reinforcement, as well as full and partial adaptation is given in figure 1.

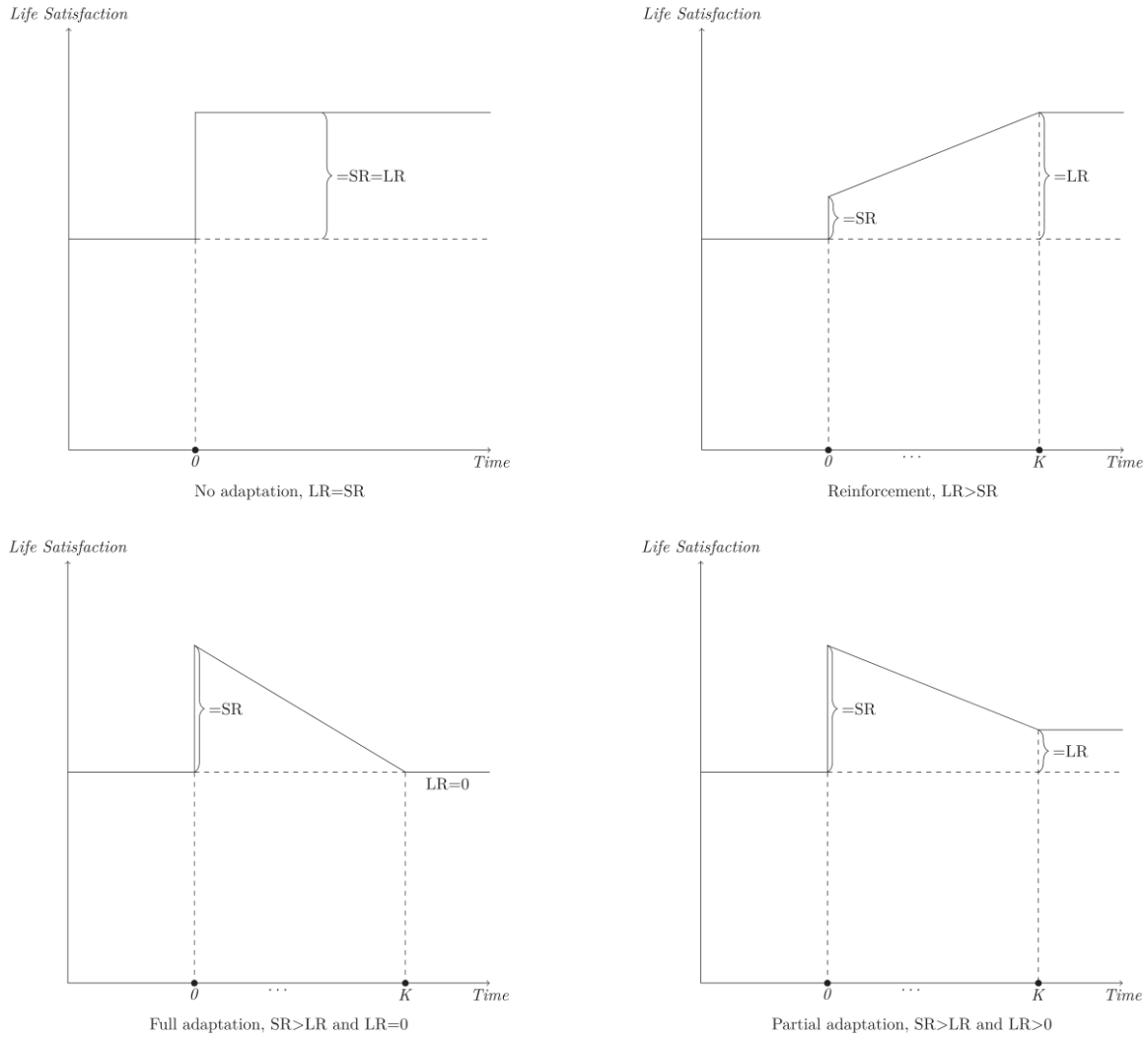


Figure 1. Adaptation types. In all cases a positive and persistent income shock occurs at time zero. I assume that higher incomes have a positive effect on life satisfaction and that the adaptation process ends after K periods.

In most studies adaptation to income is exclusively taken to be adaptation to own income. However, this need not be so. Given that we know that the incomes of peers (henceforth, 'reference incomes') tend to have negative contemporaneous effects on life satisfaction (e.g. McBride 2001; Ferrer-i-Carbonell 2005; Vendrik and Woltjer 2007), it may well be that such reference incomes could also be subject to either adaptation or reinforcement. If so, it is useful to estimate the combined short- and long run effects of both own and reference income. Such estimates can be obtained by simply taking sums of the effects of own and reference income (see figure 2). Substantively, these effects can be interpreted as the short- and long run effects of simultaneously increasing (or decreasing) an individual's, as well as her reference group's income by the same amount. Only if this combined long run effect is positive would an increase in the income of all also increase the happiness of all. Crucially, the oft-debated Easterlin paradox denies that this combined long run effect is positive.

The present paper builds on the empirical approach of Vendrik (2013), which seems to be the most robust study on adaptation to income to date. I give a much-needed attempt at corroboration of that paper's main ideas with updated GSOEP data (with ten additional waves). This attempt at corroboration is well justified, since the results of Vendrik (2013) appear to suffer from a crucial omitted variable bias, namely the omission of partner employment status. With the addition of this variable, all evidence of adaptation disappears.

Beyond that, as the first in the literature on adaptation to income, I explicitly compare estimates obtained from the two main surveys used in the literature on subjective well-being, i.e. the GSOEP and UKHLS. Given that it often seems in the literature as though results obtained from surveys that are representative of the population of one particular country are generalizable to all industrialized countries, it is especially important to investigate whether this implicit assumption does in fact hold. This is particularly true today, when it has become apparent that many results in psychology and economics are not replicable (Earp and Trafimow 2015; Chang and Li 2018), while there is still a lack of attempts at replication (Hamermesh 2007).

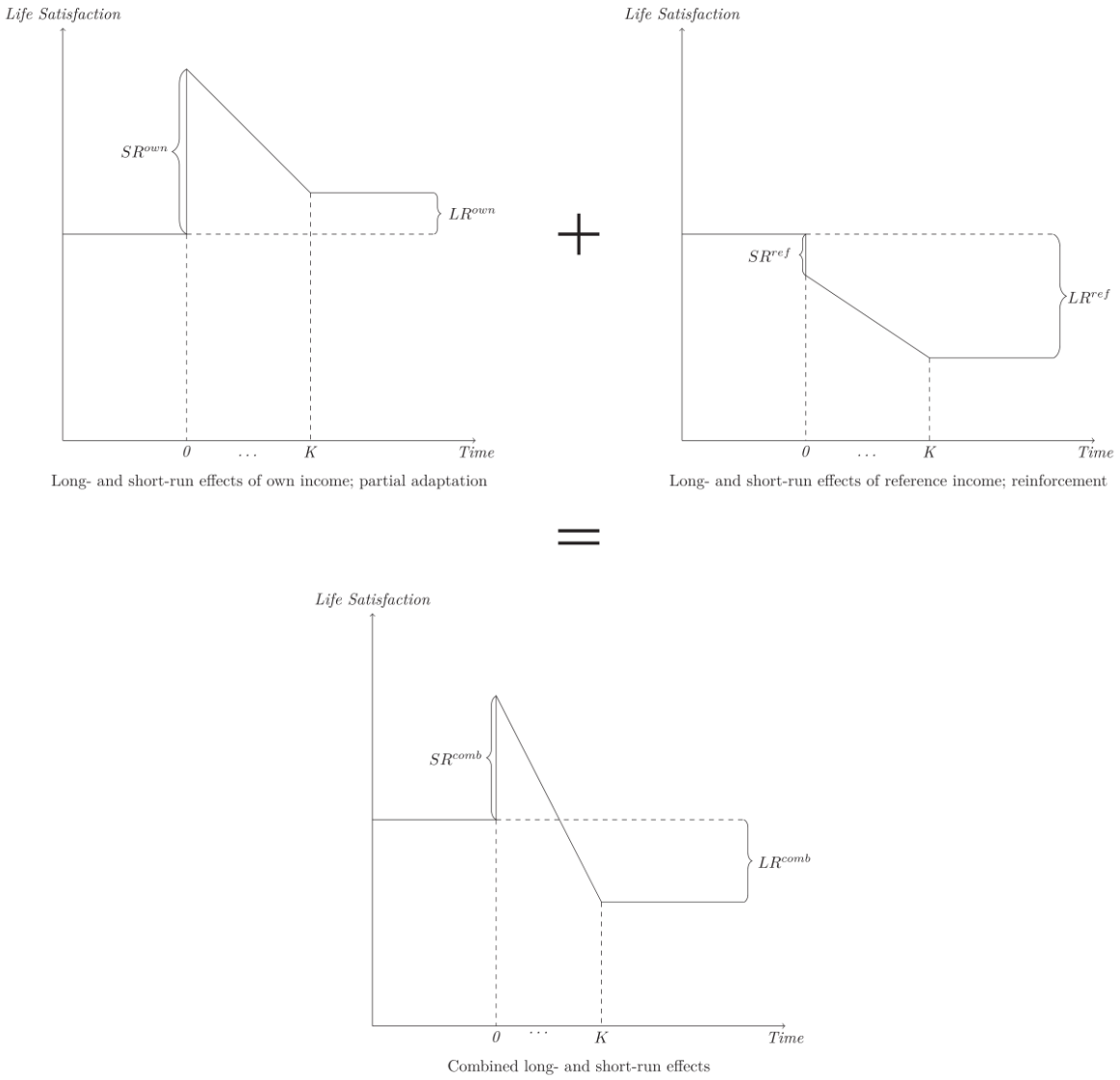


Figure 2. Illustration of how different adaptation/reinforcement patterns in own and reference income may yield a distinct pattern for the combined effects. I assume partial adaptation in own income and reinforcement in reference income. Combining these effects results in a positive short run and a negative long run effect.

My key substantive findings are summarized as follows: Firstly, no adaptation to income can be observed in any of the preferred specifications. This holds true independently of the country considered and whether an instrument for own income was used or not. Secondly, all previous evidence of adaptation was driven by omitted variable bias, most crucially via the omission of lags of controls for household composition and partner employment status. Thirdly, my estimates of the combined long run effects are always insignificant, thus corroborating the Easterlin Paradox.

The next section provides a review of the literature. Section 3 outlines my empirical approach and data. In section 4, I present my results. I make some concluding remarks in Section 5.

2 Literature Review

The study by Di Tella, Haisken-De New, and MacCulloch (2010) is the most well-known and widely cited[3] work on adaptation to income in life satisfaction. Because of that I will take their approach as my baseline for the analysis to follow in Section 4. Using data from GSOEP (1984-2000; restricted to West Germany), they regress life satisfaction on the contemporaneous level of (the ln of) real post-tax household income, as well as up to 4-year lags of this income measure. To test for adaptation, they perform tests on the significance of the sum of these lags as a well as the sum of lags and the contemporaneous coefficient. They obtain a large positive contemporaneous effect ($= 0.23$), find clear evidence for adaptation ($\sum lags = -0.15$), and cannot reject the hypothesis of full adaptation. In a follow-up, Di Tella and MacCulloch (2010) replicate these findings with up to 7 rather than 4 year lags. They only find evidence of adaptation for homeowners, showing that adaptation to income levels may be moderated by levels of wealth.

Ferrer-i-Carbonell and Van Praag (2008) follow a different strategy. Also using GSOEP data (2000–2004), they use first differences to account for individual fixed effects. They also use post-tax household income as their measure of income and include up to 4-year lags. In total, they run four specifications, one without any controls, one with levels of a standard set of contemporaneous demographic controls, one using the first differences of all controls, and finally one with their lagged first differences. When using no controls, they find adaptation of about 60%. However, when using contemporaneous terms of their controls, they obtain evidence of full adaptation. In contrast, when using all lags of the first differences, observed adaptation is reduced to as little as 40%. Ferrer-i-Carbonell and Van Praag (2008) do not offer an explanation for these strange results.

[3] Google Scholar citation count: 535 (31.10.2019).

Pfaff (2013) makes use of both GSOEP (1992–2010, separating East- and West- Germany) and BHPS (1996–2008) data. In the main, his estimation strategy is identical to that of Di Tella, Haisken-De New, and MacCulloch (2010). He obtains effect sizes that are much larger for Germany than for the UK but does not comment on this puzzling result. After experimenting with different income measures, it turns out that he only finds evidence of (complete) adaptation when either using labour income or raw real net household income; but not when using equivalized household income. Like Ferrer-i-Carbonell and Van Praag in the case of their peculiar results, Pfaff does not give an explanation for this result. He simply notes that equivalized incomes are to be preferred for what he views as theoretical reasons (see p.7). Binder and Coad (2010), who use a somewhat different estimation strategy, face the same puzzlement as Pfaff (2013): when using raw household income they obtain evidence of adaptation, but when using equivalized household income they fail to do so. Further studies on adaptation with similarly inconclusive results include those by Van Praag (1971), Clark (1999), Stutzer (2004), Easterlin (2005), Frijters, Johnston, and Shields (2011), Paul and Guilbert (2013).

There is a large literature on reference incomes. However, almost all of them simply focus on the contemporaneous effects of reference incomes. Although there is a small literature that measures reference incomes directly via the subjective assessments of respondents (e.g. Senik 2009; Mayraz, Wagner, and Schupp 2009; Goerke and Pannenberg 2015; Dumludag, Gokdemir, and Giray 2016; Clark, Senik, and Yamada 2017), most studies compute reference incomes on the basis of cell means in income of a presumed reference group. Such groups may either be defined in spatial terms (e.g. Luttmer 2005; Kingdon and Knight 2007; Ifcher, Zarghamee, and Graham 2018), or in terms of similar demographic characteristics, particularly age and education (e.g. McBride 2001; Ferrer-i-Carbonell 2005; Vendrik and Woltjer 2007; Vendrik 2013). However, this choice seems

not to generally matter, yielding negative effects for contemporaneous reference incomes that tend to roughly equal the contemporaneous effects of own income.[4]

To my knowledge, only D'Ambrosio and Frick (2012) and Vendrik (2013) study adaptation/reinforcement to reference income. Unfortunately, D'Ambrosio and Frick (2012) only include one-year lags and estimate the effects of reference income in a way that makes it difficult to explicitly recover the dynamics and relative strengths of the (combined) long run, short run and adaptation effects of own and reference income.

Vendrik (2013) may be the most robust investigation on both adaptation to own as well as reference income. He also uses GSOEP data (1984-2007; restricted to West Germany). In the main, he combines the approaches of Ferrer-i-Carbonell and Van Praag (2008) and Di Tella, Haisken-De New, and MacCulloch (2010) and regresses life satisfaction on the current ln of equivalized real net household income and reference income as well as lags of both variables. He also includes lags of all his control variables. Moreover, he includes a term for one-year lagged life satisfaction, yielding an error correction model that implicitly models effects of unincluded lags of income in the further past. In the main analyses he instruments income using the approach of Luechinger (2009) and Luttmer (2005). When simply using observed income, Vendrik is unable to obtain any evidence of adaptation to own income. However, in his (seemingly) more robust regressions where he instruments for own income, he does obtain evidence of adaptation. He further finds a negative long run effect of reference income that is larger than that of own income.

Four key points may be taken from this literature review. First, whether we observe adaptation to own income depends on whether we use equivalized or raw household income, and whether we instrument for own income. Only in the case of raw household income or instrumented income, do we observe adaptation (e.g. Binder and Coad, 2010; Pfaff, 2013). Second, we observe less adaptation to own income when including lagged terms of controls (Ferrer-i-Carbonell and Van

[4] Especially in volatile economies this negative effect may not hold. It may instead be dominated by a positive 'signaling' effect. See (Senik 2004; 2008).

Praag, 2008; Vendrik, 2013). Third, with respect to adaptation to reference income and from today's viewpoint, the only (to my knowledge) viable study is the one of Vendrik (2013). Fourth, there seem to be almost no studies doing cross-country comparative work. The working paper of Pfaff (2013) is the only exception, but he does not comment on the comparative aspect. The present study attempts to address each of these points.

3 Empirical strategy and data

3.1 Econometric specification

I start with a distributed lag model very similar to that of Di Tella, Haisken-De New, and MacCulloch (2010), with the main difference being my inclusion of reference incomes:

$$LS_{it} = \sum_{k=0}^{K} \beta_{0-k} \ln(y_{it-k}) + \sum_{k=0}^{K} \gamma_{0-k} \ln(y_{it-k}^{ref}) + \phi X_{it} + \mu_i + \tau_t + \epsilon_{it} \qquad (1)$$

Here LS_{it} is the (reported) life satisfaction of individual i at time t, y_{it-k} is a measure of own household (HH) income, y_{it-k}^{ref} is a measure of reference income, μ_i is an individual fixed effect, τ_t is a year fixed effect and ϵ_{it} is an error term. X_{it} is a vector of time-varying controls thought to be correlated with both income and life satisfaction. These include: dummies for numbers of children and adults in the household, employment status of the respondent, employment status of the respondent's partner, dummy for child birth, age, age-squared, marital status, housing tenure, job hours, dummy for illness, region dummies, and a region-specific linear time trend. These controls are similar to those chosen by Di Tella, Haisken-De New, and MacCulloch (2010) and Vendrik (2013). A precise description of each of these variables is given in table D1 of the appendix. In this set-up, the contemporaneous, or short run (SR), effect of own income on life satisfaction is given by the contemporaneous effect:

$$SR = \beta_0 \qquad (2)$$

In general, the long run (LR) effect is given by the contemporaneous effect, plus the sum of all lags of own income:

$$LR = \sum_{k=0}^{\infty} \beta_{0-k} \tag{3}$$

However, I assume $\beta_{0-k} = 0 \; \forall \; k > K$ in the special case of equation (1). Thus, in equation (1), the LR effect is given by $\sum_{k=0}^{K} \beta_{0-k}$. The total effect of adaptation, i.e. the degree to which the effect of income dissipates with time, is given by the sum of all lags of own income:[5]

$$ADAP = \sum_{k=1}^{\infty} \beta_{0-k} \tag{4}$$

We may also generally say that adaptation is given by the difference between the long run and the contemporaneous effect (i.e. $ADAP = LR - SR$). All of this is illustrated in figure 3.[6]

The same reasoning applies to the estimates of the SR, LR, and adaptation effects of reference income. When looking for the respective combined effects, we may simply take the sums of the effects of own and reference income.

The review in Section 2 showed that controlling for lagged terms of controls can drastically change results. The reason for this may be as follows: controlling for X_{it} stems from the belief that X_{it} is correlated with y_{it} and LS_{it}. If this is so, we must also believe that X_{it-k} is correlated with y_{it-k}. Given the literature on adaptation to various life events (Angeles 2010; Clark et al. 2008b; Hanglberger and Merz 2011), we may have further reason to believe that at least some variables in

[5] Or $\sum_{k=1}^{K} \beta_{0-k}$ in the case of equation (1).

[6] The study of Wolbring, Keuschnigg, and Negele (2013), where they also attempt to study adaptation to income in life satisfaction, pursues an altogether different approach. They exclusively use lagged first differences (up to three) and no term for the level of income, i.e. $LS_{it} = \gamma_0 \Delta \ln(y_t) + \gamma_1 \Delta \ln(y_{t-1}) + \gamma_2 \Delta ln(y_{t-2}) + \varphi X_{it} + \mu_i + \epsilon_{it}$. They state that if the effect of the first differences declines as one moves to higher order lags (i.e. if $\gamma_0 > \gamma_1 > \gamma_2$), there is evidence of adaptation. This is indeed what they find. However, such an interpretation is not warranted. All this would imply is that more recent changes in income have a stronger effect on life satisfaction. In fact, their model presumes that full adaptation will occur. This can be seen by rewriting the model as $LS_{it} = \gamma_0 \ln(y_t) + (\gamma_1 - \gamma_o)\ln(y_{t-1}) + (\gamma_2 - \gamma_1)\ln(y_{t-2}) + \gamma_2 \ln(y_{t-3}) \varphi X_{it} + \mu_i + \epsilon_{it}$. Applying the formula above gives $LR = 0$ (all terms cancel), $SR = \gamma_0$, and $ADAP = -\gamma_0$.

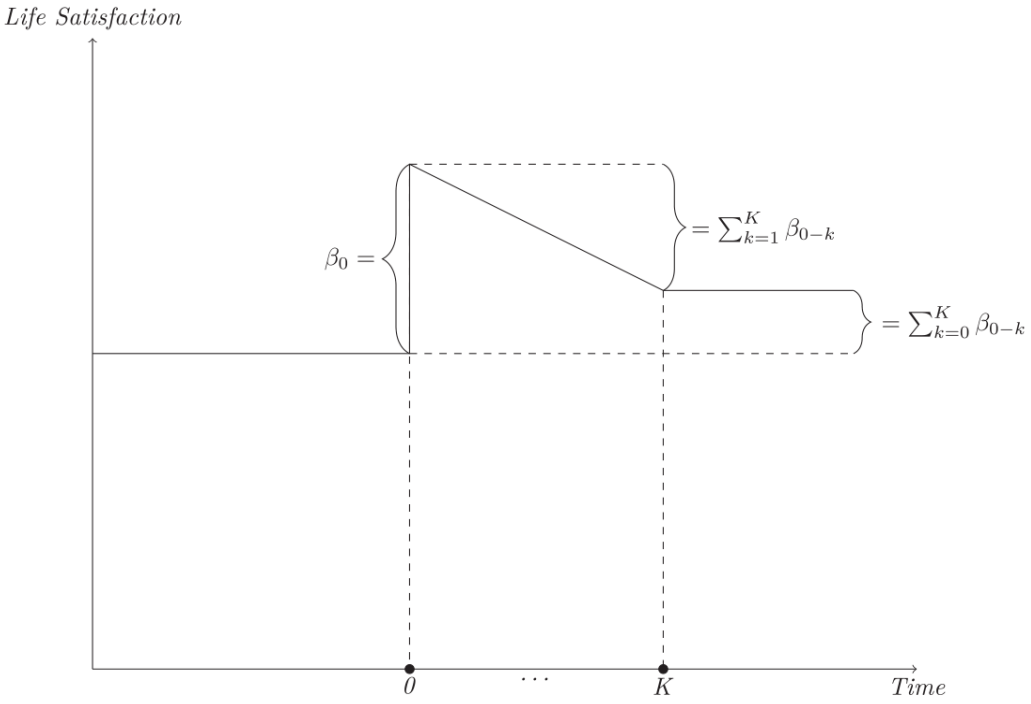


Figure 3. Illustration of the process modelled in equations (1) and (5). A positive and persistent income shock in own income is experienced at time zero. I assume partial adaptation.

X_{it-k} will also be correlated with LS_{it}.[7] If this is so, my estimates of the long run, short run and adaptation effects would have to be biased. In particular, any observed adaptation to income may be driven by unobserved adaptation to changes in any of the covariates in X. Hence, I should further extend the model by adding lags of all control variables that are not perfectly collinear with its contemporaneous values.[8] This is similar to the approach of Ferrer-i-Carbonell and Van Praag (2008) and Vendrik (2013). When adding such lags, I always include as many lags as used for the own and reference income terms (indicated by K). The model thus becomes:

$$LS_{it} = \sum_{k=0}^{K} \beta_{0-k} \ln(y_{it-k}) + \sum_{k=0}^{K} \gamma_{0-k} \ln(y_{it-k}^{ref}) + \sum_{k=0}^{K} \phi_{0-k} X_{it-k} + \mu_i + \tau_t + \epsilon_{it} \qquad (5)$$

When attempting to compare effects obtained from UKHLS and GSOEP and to account for the limited coverage of the UKHLS, I restrict the observation window for GSOEP to start in 1996.

[7] It may further be that elements in an appropriately constructed vector X_{it-k}^{ref}, e.g. levels of unemployment in the reference group, would behave similarly. I omit this complication from the present investigation.
[8] This turns out to make a large difference to my results. The aforementioned studies investigating adaptation to non-monetary life events may therefore also be affected by this problem.

Moreover, I use the same income concept (at baseline: real net household income)[9] and use comparable controls (see appendix table D1) across both countries. Furthermore, to make sure that lags are comparable and so that only yearly adaptation dynamics are indeed captured, I exclusively include observations that occur within a time-frame of 10 to 14 months in succession.[10] Finally, I treat life satisfaction as cardinal. Ferrer-i-Carbonell and Frijters (2004), Van Praag and Ferrer-i-Carbonell (2008) and Riedl and Geishecker (2014) showed that ratios of coefficients and levels of significance from OLS regressions are not much different from ordered probit regressions.[11]

3.2 Instrumenting

There are two main objections to directly using observed own income for estimation. First, income is measured with error. Such error may be driven by individuals not being able to accurately report on their income, the need to impute income in some cases, or inaccuracies in the tax-simulation routines used to arrive at net incomes. Such measurement error may lead to attenuation bias towards zero of the estimated effects. Second, there may unobserved costs to income generation, e.g. stress or increased commuting times, which may present a downward bias. To alleviate these concerns, I instrument income following the approach first pursued by (Luttmer 2005) and later adopted by Luechinger (2009) and Vendrik (2013). This approach goes as follows: For each wave, I predict real individual labour earnings using the respondent's age, age-squared, gender and a large set of occupation and industry dummies. Thus, for each wave, I run a regression of the form:

$$y_i^{labour} = \alpha + \beta_1 Male_i + \beta_2 Age_i + \beta_3 \mathrm{Age}_i^2 + \delta_{ind} IND_i + \theta_{occ} OCC_i + \epsilon_i \qquad (6)$$

[9] I choose not to convert Pounds into Euros. Under the assumption that the separate series for inflation used to obtain a household income measure in constant prices in each country are comparable, my use of ln*s* for all income terms in all estimations, as well as my use of year fixed effects, renders such a conversion unnecessary.

[10] I thank an anonymous reviewer for this point. Unfortunately, enforcing that all observations must strictly occur within 12 months intervals leads to an exceedingly large drop in panel observations and was thus not possible.

[11] However, papers 3 and 4 of this thesis sheds some doubt on this assumption, which ought to be pursued in future work.

Here δ_{ind} and θ_{occ} are industry- and occupation-specific parameters. I use 3-digit ISCO88 codes for occupation, and 2-digit NACE codes for industry. Finally, I sum the predicted earnings of each individual in the household to obtain a measure of predicted household incomes. This constructed measure is then used to instrument for observed household income.

There are two main threats to the exclusion restriction of this instrument. The first concern is that there may be other time-varying effects on life satisfaction of industry or occupation beyond those of time-varying pecuniary returns. For example, it may be that individuals in the same industry/occupation form part of the individual's reference group. Although I control for reference incomes defined over education and age, there may nevertheless be an uncontrolled-for negative effect that would downwardly bias the IV estimate. Alternatively, it may be that some generally better paying industries/occupations have more pleasant work environments than others. This would yield an upward bias. I partially guard against the latter concern by adding contemporaneous and lagged dummies for each industry and occupation to the model. In line with the rest of the empirical approach, the number of lags used is always equal to the number of lags used for the income terms and all other controls. Note that the predicted incomes from which the IV is constructed are estimated separately at each wave. This is why they are not perfectly collinear with the industry/occupation dummies. The inclusion of these dummies controls for all time-invariant effects of industry and occupation. The aforementioned biases are therefore restricted to occur only via time-varying non-pecuniary effects of industry and occupation. However, work conditions in some industries may for example have gotten better compared to others, while pay has increased too. Such a bias would not be captured by the added dummies. The second concern is that respondents may choose to select into particular industries or occupations in anticipation of changes in income or other benefits that may impact life satisfaction. Pischke and Schwandt (2012) note that industry choice is indeed correlated with fixed characteristics of individuals. If individuals had no option to switch industries or occupations during the estimation period, such selection would not be a concern (since such differences would be captured by the individual fixed effects).

However, clearly some individuals have the option (or are forced) to switch industries/occupations. In about 23% (Germany) and 28% (UK) of all observations such switches occur. This may be problematic. As noted above, the addition of (lags of) dummies for industries and occupations capture the effects emanating from time-invariant pecuniary and non-pecuniary differences between industries/occupations. However, these dummies do not capture over-time changes in the differences between industries/occupations. Therefore, if people actively switch industries or occupations in anticipation of changes in industry- or occupation-specific pecuniary or non-pecuniary returns, and when the propensity to switch is correlated with unobservables, my estimates may still be biased. In a robustness test, I therefore exclude observations where a switch of industry/occupation occurred (see table R9 of the appendix). However, since failing to switch is also potentially endogenous, this is not a complete test. Notice that all of these potential biases, and more are discussed in Section 4.3, apply equally to both contemporaneous and lagged own income. Therefore, although the absolute sizes of the short- and long run effects may be biased in an unknown direction, their difference, i.e. the degree of adaptation/reinforcement, is potentially less biased.[12]

3.3 Determining the number of lags

A tricky question is what the appropriate number of lags K is. On the one hand, one would wish to include very many lags in order to make the assumptions made for equation (5) as plausible as possible. On the other hand, adding more lags is problematic for two reasons: First the addition of each lag causes the loss of a large number of panel observations. Second, levels of income at t and $t-1$ are highly collinear, thus also leading to a loss in efficiency of the estimates. Vendrik (2013) determines the number of lags by first starting with some large number of lags. He then conducts F- and t-tests of the (joint) significance of the highest lag of own and reference income. If all tests fail to reject the null hypothesis of (joint) equality to zero, he removes all highest order lags and

[12] I thank an anonymous reviewer for this point.

repeats the process until he obtains at least one rejection. At baseline, this leads him to use a maximum lag length of 3, while when including lagged life satisfaction and instrumenting for income, he finds an appropriate lag length of 1.

When I apply this approach, I obtain a lag length of 1 for both the UK and Germany. Given that I have no need to use such a short lag length, I chose a more conservative lag-length of $K = 3$ for my baseline estimates. This choice of $K = 3$ is much closer to the choice of Di Tella et al. (2010), Pfaff (2013) and Ferrer-i-Carbonell and Van Praag (2008) and many others in the literature who use a lag length of four. However, as it turns out these considerations turn out not to be very important, since my IV results on the full samples are largely robust to lag lengths between 1 and 6 for both the UK and Germany (see section 4.3). At higher order lags my instruments turn weak.

3.4 Data

For Germany I use waves 1–32 of the German Socio-Economic Panel (GSOEP), spanning the time 1984–2015. For the UK, I use all available waves of the UKHLS, which combines waves 6-18 of the BHPS and waves 1–8 of Understanding Society.[13] This translates to the years 1996–2017. Previous work has indicated that incomes at the very top and the very bottom of the income distribution suffer from large measurement error (Berthoud and Bryan 2011; Figari 2012). Moreover, self-employed individuals are known to either under- or over-report their incomes (Hurst, Li, and Pugsley 2013). For these reasons, I exclude the top and bottom 1% within each wave of household income, as well as all self-employed individuals from the analysis. Since very few individuals younger than 20 years or older than 70 years are observed to have labour earnings (which is needed to instrument household income), I also exclude all individuals that fall outside of this range.

[13] I exclude waves 1–5, as life satisfaction is not observed for these waves. Life satisfaction is also not observed in wave 11. Hence, this wave is excluded, too.

In the UKHLS, life satisfaction is measured on a 1–7 scale, with respondents first being presented the statement "*Here are some questions about how you feel about your life. Please tick the number which you feel best describes how dissatisfied or satisfied you are with the following aspects of your current situation. 1 = Not satisfied at all 7 = Completely satisfied*". After a set of domain-specific questions, respondents are presented with the question: *"Using the same scale, how dissatisfied or satisfied are you with your life overall?"*. Within the GSOEP, life satisfaction is measured on a 0–10 scale, with respondents answering the question *"In conclusion we would like to ask you about your satisfaction with your life in general, please answer according to the following scale: 0 means completely dissatisfied and 10 means completely satisfied: How satisfied are you with your life, all things considered?*". As can be discerned from figure 4, the distribution of responses to these questions is very similar across countries.

Regarding own income, unless otherwise noted and for both Germany and the UK, I use annual post-government (net of taxes and benefits) household income. This income variable is generated by Levy and Jenkins (2012) for the BHPS. It is directly provided in Understanding Society and the German SOEP. I deflate both income variables using the consumer price index (base = 2005). In

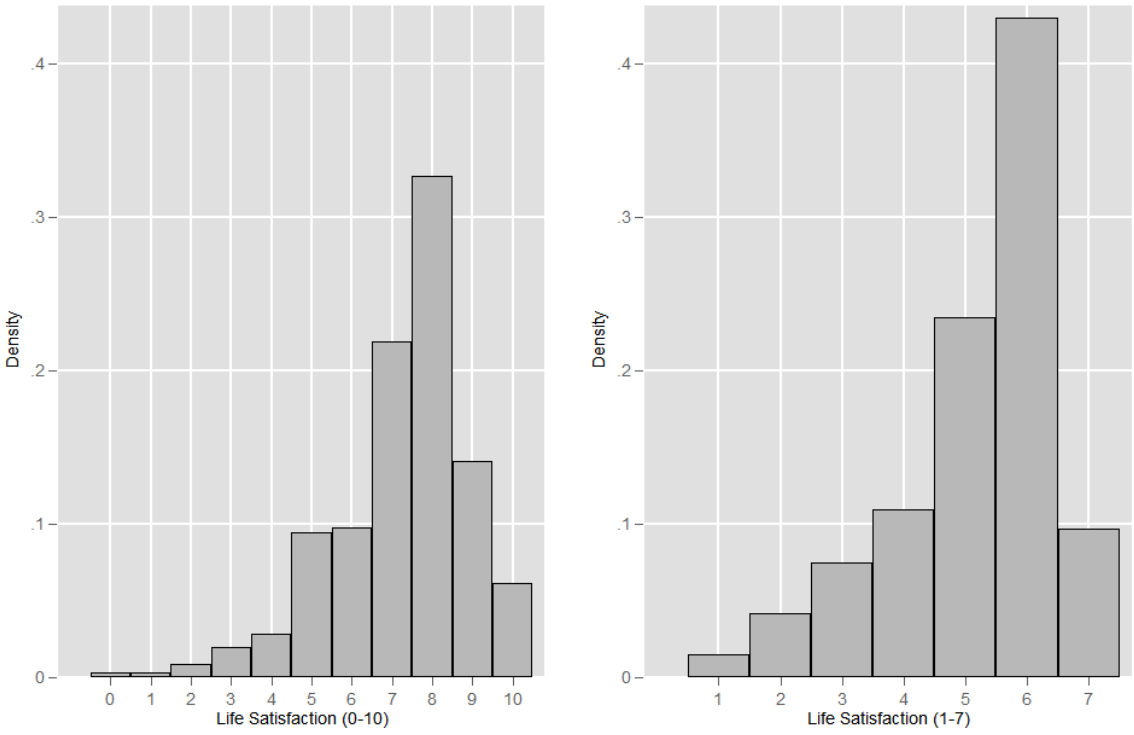


Figure 4. Distribution of life satisfaction in Germany (left) and the UK (right).

the main specifications, I do not adjust own household incomes for household size by means of an equivalization scale. This makes my specification most comparable to that of Di Tella et al. (2010). However, as made clear in section 4.4, including (lags of) controls for household composition is equivalent to equivalizing own incomes. Robustness regressions in which I use equivalized own household income verify this (see section 4.3). I define reference incomes by the mean annual equivalized net post-government household income of the set of individuals up to 5 years older or younger than the respondent, which have the same educational attainment as the respondent (as measured by a collapsed version of the CASMIN classification). Reference incomes are equivalized since explicitly controlling for household composition of the reference group would not be easily feasible. Generally, whenever an equivalization scale is used, I utilize the modified OECD scale. This scale assigns a weight of 1 to the first member of the household, a weight of 0.5 to every subsequent adult, and a weight of 0.3 for every child in the household. Definitions for all control variables can be found in table D1 of the appendix.

Table 1 provides summary statistics for my main variables of interest - household income and life satisfaction. The overall standard deviation of life satisfaction is roughly 1.63 in the German sample, whereas for the British sample it is roughly 1.30. Thus, the ratio of standard deviations ($1.63/1.30 \approx 1.28$) is reasonably similar to the ratios of the response scales ($11/7 \approx 1.57$). This gives further assurance for my comparative purposes.[14]

The IV strategy primarily relies on variation in returns within occupations and industries. Therefore, tables D2–D5 in the appendix show levels and year-to-year changes of observed and predicted household income across major occupational and industry groupings. Levels of income are ordered as one would expect across industry and occupation groupings (e.g. incomes for the occupation grouping "professionals" are highest, and those for "agricultural workers" and "elementary occupations" are lowest).

[14] These ratios need not be equal: It may be that there is less variance in life satisfaction in Germany than in the UK.

Table 1. Descriptive Statistics

	N	Mean	Median	SD	Min	Max
Germany (1984-2015)						
Life satisfaction	128,850	7.179	8.000	1.629	0	10
HH income (2005 €)	128,850	39,497	36,790	16,023	6,086	124,905
Ln of HH income (2005 €)	128,850	10.503	10.513	0.408	8.714	11.735
Ln of ref. eq. HH income (2005 €)	128,850	9.844	9.815	0.185	9.292	10.382
Germany (1996-2015)						
Life satisfaction	91,776	7.156	7.000	1.610	0	10
HH income (2005 €)	91,776	40,576	37,756	16,797	6,100	124,905
Ln of HH income (2005 €)	91,776	10.526	10.539	0.420	8.716	11.735
Ln of ref. eq. HH income (2005 €)	91,776	9.876	9.853	0.193	9.292	10.382
UK (1996-2017)						
Life satisfaction	88,170	5.195	6.000	1.269	1	7
HH income (2005 €)	88,170	31,409	29,148	14,093	2,602	114,766
Ln of HH income (2005 €)	88,170	10.255	10.280	0.458	7.864	11.651
Ln of ref. eq. HH income (2005 €)	88,170	9.670	9.668	0.207	8.626	10.082

Unfortunately, no industry or occupation codes are observed for respectively 19.1% and 19.6% of included observations in Germany. For the UK, 11.3% (industry) and 10.4% (occupation) of codes are missing. Most of these observations are of persons that are either unemployed, retired or otherwise outside the workforce. Since these individuals do not have labour incomes, they do not contribute to the construction of the instrument. They are nevertheless retained in the sample as other members of the household will have a labour income. In section 4.3 I show that my key conclusions do not depend on the inclusion of these observations.

Another source of variation in the instrument comes from individuals switching industry/occupation combinations. Tables D2-D5 of the appendix also show levels and changes of observed and predicted income across observations in which such switches occurred and where they did not. In general, (predicted) household income levels do not differ much between stayers and leavers. However, and as may be expected, there is much greater variation in predicted income changes for switchers compared to leavers. This variation might primarily drive results. In section 4.3 I therefore discuss results which exclude these observations.

We may be concerned that the reference group (i.e. those of similar age and education) and the set of people with the same industry and occupation overlaps. Table D6 of the appendix provides summary statistics on the size of the reference group as well as its overlap with respondent's

industries and occupations. The average size of the reference group is 702 and 848 persons in Germany and the UK, respectively. Although there is a small number of very small reference groups, most reference groups have between 233 (1st percentile) and 1264 (99th percentile) members in Germany and between 185 and 1986 members in the UK.[15] On average, only 6.4% (Germany) and 3.4% (UK) of members in a reference group share the same industry and occupation combination as the respondent.

4 Results

Sections 4.1 and 4.2 present my main results for Germany and the UK, respectively. Section 4.3 provides extensive robustness checks which qualify these results. In Section 4.4, I show how my estimates are relevant to the construction of equivalence scales that are sensitive to the short- and long run.

4.1 Results for Germany

In this section, I present my results for my baseline estimations of variants of equations (1) and (5) using GSOEP data for the period 1984–2015. Thereafter, in Section 4.2, I present results for the UK.

To provide a baseline, the estimates in column 1 of table 2 are obtained using equation (1) (i.e. no lags of controls) and do not instrument for own income. Throughout, I use 3-year lags for income. The estimates are thus closest to Di Tella, Haisken-De New, and MacCulloch (2010). I find significant adaptation to own income that is roughly equal in magnitude to that which they find (-0.127 *versus* -0.150; see their table 1 column 2). The contemporaneous effect that I obtain is similar to theirs (0.304 *versus.* 0.230). However, especially given my much larger sample (using an extra 15 years), I obtain a significant long run effect, implying partial rather than full adaptation to own income. However, in column 2, I estimate equation (5), which thus now includes lags of all

[15] There are 92 and 4 observations with reference groups smaller than 10 members in Germany and the UK, respectively.

controls. This inclusion has the important consequence that the long run effect of own income is now slightly larger than the short run effect. Consequently, all evidence of adaptation observed in column 1 vanishes in column 2. As explained in Section 3.1, this suggests that significant adaptation observed in column 1 is the result of omitted variable bias, i.e. adaptation to one of the control variables which are correlated with income.

Table 2. Main results for West Germany

	(1) HH income, contemp. controls	(2) HH income, lagged controls	(3) Instr. HH income, contemp. controls	(4) Instr. HH income, lagged controls	(5) Instr. HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.304***	0.250***	0.773***	0.727***	0.435**
	(0.025)	(0.025)	(0.149)	(0.199)	(0.185)
LR effect of own inc.	0.177***	0.280***	0.370***	0.481***	0.408***
	(0.034)	(0.039)	(0.107)	(0.130)	(0.131)
Adaptation to own inc.	-0.127***	0.030	-0.403***	-0.246+	-0.028
	(0.027)	(0.034)	(0.088)	(0.183)	(0.171)
SR effect of ref. inc.	-0.099	-0.055	-0.088	-0.126	-0.095
	(0.158)	(0.166)	(0.160)	(0.170)	(0.169)
LR effect of ref. inc.	-0.297**	-0.341**	-0.322**	-0.361**	-0.342**
	(0.140)	(0.151)	(0.147)	(0.157)	(0.158)
Adaptation to ref. inc.	-0.198+	-0.286+	-0.234+	-0.235	-0.248+
	(0.150)	(0.185)	(0.152)	(0.190)	(0.190)
Combined SR effect	0.205+	0.195	0.685***	0.601**	0.341+
	(0.159)	(0.166)	(0.210)	(0.234)	(0.228)
Combined LR effect	-0.120	-0.061	0.048	0.120	0.065
	(0.140)	(0.153)	(0.152)	(0.172)	(0.177)
Combined Adaptation	-0.325**	-0.256+	-0.637***	-0.480**	-0.275
	(0.151)	(0.187)	(0.176)	(0.238)	(0.233)
Kleibergen-Paap F-stat.			156.149	41.943	51.873
Serial Correlation	0.027	0.023	0.022	0.018	0.019
Average T	7.456	7.456	7.456	7.456	7.456
# of Persons	17,282	17,282	17,282	17,282	17,282
# of Observations	128,850	128,850	128,850	128,850	128,850

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

However, a crucial problem with these baseline results is that the sources of variation in own income are largely unknown and potentially endogenous. In particular, these estimates may suffer from reverse causality (more satisfied persons earn more), which may upwardly bias my estimates of both the short- and the long run effects. Moreover, there may be unobserved hedonic costs to income generation (like stress), which negatively bias my effect estimates. Finally, measurement error in the income variable may present a further bias towards zero. To alleviate these concerns,

columns 3–5 instrument for own income using the approach given in Section 3.2. Generally, the effect estimates for own income are increased when instrumenting.

This difference is in line with Powdthavee (2010), Luechinger (2009) and Knight and Gunatilaka (2010) who also find effect increases when instrumenting.[16] In column 3, no lags of controls are included. There, I again find significant adaptation. However, as was the case previously, adaptation turns marginally insignificant when including lags of all controls ($p = 0.18$) in column 4. Finally, in column 5 I include contemporaneous and lagged values of all industry and occupation dummies to the regression. This addresses the concern raised in section 3.2, namely that the instrument may pick up further time-constant non-pecuniary effects of industry and occupation choice. The inclusion of these dummies depresses the estimates of the short- and long run effects of own income. Importantly, the estimate of the short run effect is depressed much more than the estimate of the long run effect. Consequently, my estimate of adaptation to own income is also much diminished and insignificant ($p = 0.87$). Changes in occupation and industry typically come about via job changes. It therefore appears that most of the (insignificant) adaptation observed in column 4 was driven by adaptation to changing jobs. I therefore take the estimates of column 5 to be the preferred specification. Nevertheless, further robustness tests are given in section 4.3.

My results contradict the finding of Vendrik (2013) who finds significant adaptation to own income when instrumenting. In table R17 of the appendix, I investigate the sensitivity of my estimates to the exclusion of each control. That exercise reveals that when I fail to include partner's employment, I do obtain significant estimates of adaptation. It turns out that there is a large positive contemporaneous effect of partner employment that is subject to full adaptation after three periods. Especially for women, partner employment is strongly correlated with (instrumented) household income. As Vendrik (2013) does not include partner's employment status (nor does any

[16] Knight and Gunatilaka (2010) and Powdthavee (2010) use different instruments to the one employed here. Vendrik (2013) and Luechinger (2009) use the same kind of instrument.

other study that I am aware of), it seems that the estimates of Vendrik (2013) also suffer from omitted variable bias.

Concerning reference incomes, I find an insignificant short run effect throughout. However, the long run effect is sizable and significant in all specifications. As a consequence, I observe (marginally) insignificant reinforcement of the reference income effect. Thus, we observe that the short- and long run effects are roughly constant for own income, but that the effect of reference income may grow over time. This has the consequence that the combined effect of own and reference income is positive, significant, and sizable in the short run, but becomes insignificant and close to zero in the long run. These combined estimates may be interpreted as the marginal short- and long run effects of equally sized and permanent income changes of both the respondent and all members of the respondent's reference group.[17]

Therefore, as everybody grows richer, mean satisfaction does not increase in the long run. This is in line with the Easterlin Paradox.

4.2 Results for the UK

I now turn to comparative results for the UK and a comparison with Germany. I am not aware of any previous paper in the literature on adaptation in life satisfaction that has been strongly comparative. The working paper by Pfaff (2013) uses the BHPS and GSOEP simultaneously, but is not very comparative. This general lack of cross-country comparison is problematic. Much of the previous happiness literature seems to implicitly view results that are obtained with data from only one country (i.e. Germany with GSOEP) as indicative of facts beyond this particular country (i.e. rich industrialized democracies). Pursuing a comparative approach is an informal test of this presumed external validity. Put differently, in response to a lack of replication and/or replicability in economics and psychology (Earp and Trafimow 2015; Chang and Li 2018), using UKHLS data

[17] In some of the specifications the combined adaptation estimate is significant. We can think of this as collective adaptation to a common income shock.

alongside GSOEP data can be viewed as a replication of what is thought to be known from GSOEP data. To make results comparable I use the same set of controls for the UK as I did for Germany. Additionally, I provide further estimates in which I restrict the estimation period for the German sample to the British sample as closely as possible.

Table 3 presents my results. Columns 1–5 mirror the econometric specifications of table 3. My substantive conclusions are the same for the UK as for Germany. Both in the baseline and in the instrumented case, I find significant adaptation to own income when failing to include lags of controls (columns 1 and 3). However, when adding lags of controls, I find that the short run effect of own income is smaller than the long run effect (columns 2, 4, 5). This indicates (insignificant) reinforcement rather than adaptation. Unlike the case for Germany, the addition of dummies for industries and occupations makes little difference to my effect estimates (column 5).

As was the case previously, I find smaller effect estimates when using observed own income rather than instrumenting own income. These non-instrumented estimates for the UK are also smaller than the equivalent results for Germany (see table 2), even accounting for the fact that the UK response variable has fewer response options (7 *versus* 11), and a smaller standard deviation.[18] Similarly small estimates were obtained by Pfaff (2013). However, my instrumented long run effect estimates are, also when accounting for differences in the response variable, reasonably close to those obtained for Germany (see column 6).[19] Given that the BHPS does not directly provide net income data and had to be generated with external information by Levy and Jenkins (2012), it may be that the large difference in estimates between observed and instrumented income is caused by attenuation bias due to measurement error.

[18] Using results from column 2 of tables 2 and 3, standardizing effect sizes via the number of response categories yields that the UK long run effect equals 71.3% $((11/7) * 0.127)/0.280 = 0.713$ of the German effect. Standardizing via the respective standard deviations yields that UK's effect equals 58.2% $((1.629/1.269) * 0.127)/0.280 = 0.713$ of the German effect.

[19] Using results from column 5 of tables 2 and 3, the UK long run effect equals 137.9% ($((11/7) * 0.382)/0.408 = 1.379$) of the German effect adjusting for the number of response categories. Adjusting via standard deviations, yields the UK effect to equal 112.6% of the German effect ($((1.629/1.269) * 0.382)/0.408 = 1.379$).

Table 3. Main results for the UK

	(1) HH inc., contemp. controls	(2) HH inc., lagged controls	(3) Instr. HH inc., contemp. controls	(4) Instr. HH inc., lagged controls	(5) Instr. HH inc., lagged controls, indust. & occup.	(6) Instr. HH inc., lagged controls, indust. & occup., (W.Ger., 1996-2015)
SR effect of own inc.	0.103***	0.090***	0.352***	0.285***	0.302***	0.278
	(0.019)	(0.019)	(0.085)	(0.095)	(0.102)	(0.239)
LR effect of own inc.	0.047+	0.130***	0.171**	0.329***	0.382***	0.368**
	(0.029)	(0.033)	(0.081)	(0.107)	(0.115)	(0.170)
Adaptation to own inc.	-0.056**	0.040+	-0.181***	0.044	0.080	0.090
	(0.024)	(0.029)	(0.056)	(0.103)	(0.108)	(0.220)
SR effect of ref. inc.	0.108	0.139	0.058	0.106	0.115	-0.105
	(0.135)	(0.138)	(0.136)	(0.141)	(0.142)	(0.223)
LR effect of ref. inc.	-0.113	-0.195	-0.153	-0.218	-0.234+	-0.140
	(0.154)	(0.176)	(0.156)	(0.177)	(0.179)	(0.190)
Adaptation to ref. inc.	-0.221*	-0.334*	-0.211+	-0.324*	-0.349*	-0.035
	(0.129)	(0.174)	(0.130)	(0.177)	(0.179)	(0.242)
Combined SR effect	0.212+	0.229*	0.410***	0.391**	0.417***	0.173
	(0.135)	(0.139)	(0.150)	(0.157)	(0.160)	(0.295)
Combined LR effect	-0.065	-0.065	0.018	0.111	0.148	0.228
	(0.155)	(0.177)	(0.164)	(0.198)	(0.201)	(0.205)
Combined Adaptation	-0.277**	-0.294*	-0.392***	-0.280+	-0.269+	0.056
	(0.131)	(0.175)	(0.136)	(0.189)	(0.192)	(0.296)
Kleibergen-Paap F-stat.			293.629	120.649	109.005	32.193
Serial Correlation	-0.174	-0.176	-0.175	-0.177	-0.182	-0.030
Average T	4.669	4.669	4.669	4.669	4.669	6.396
# of Persons	18,884	18,884	18,884	18,884	18,884	14,348
# of Observations	88,170	88,170	88,170	88,170	88,170	91,776

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Surprisingly, all short- and long run estimates for reference income in the UK are insignificant. The short run effect is even positive, which may be driven by a "tunnel" effect as proposed by Senik (2004; borrowing from Hirschman and Rothschild 1973). Individuals may initially welcome increased incomes of others as signalling future increases in their own income. However, as time passes, the classical envy effect starts to dominate. Moreover, my measure of reference income can only imperfectly proxy for true perceived reference income.

As a consequence, these estimates are biased toward zero and will likely be larger in reality.[20] Finally, despite the statistical insignificance of the negative long run effect of reference income, I also find

[20] I thank an anonymous reviewer for this remark.

that the combined long run effects of own and reference income are again insignificant and close to zero. Thus, increasing the incomes of all fails to increase the satisfaction of all in the UK, too.

In column 6, I provide results for Germany when restricting the observation period to 1996–2015 for comparability. My estimates for own income are very similar to those obtained in table 2 (column 5) on the unrestricted period. Again, there is no evidence of adaptation. Instead, I find insignificant reinforcement. Furthermore, my estimated reference income effects are much diminished and now insignificant. Consequently, the long run effect of own and reference income combined is more positive but remains insignificant.

4.3 Robustness tests

This section reports on a number of robustness checks against which the main results of the previous sections are tested. All tables relating to these tests can be found in the appendix. Recall that previous studies which found a significant estimate of adaptation used raw household income (Vendrik, 2013 only found adaptation when instrumenting). Pfaff (2013) and Binder and Coad (2010) in particular observed adaptation when using raw household income but did not do so when using equivalized household income. To evaluate the robustness of my own results, in tables R1 (Germany) and R2 (UK), I re-estimated the specifications of tables 2 (Germany) and 3 (UK), using equivalized rather than raw household income. When doing so, I mostly find no adaptation to own income even when failing to include lags of all controls. It can be shown that equivalizing income amounts to controlling for (lags of) changes in household composition (see section 4.4). Therefore, as hinted at by Vendrik (2013), the significant estimate of adaptation to income observed in columns 1 and 3 of tables 2 and 3 was largely driven by adaptation to changes in household composition. This explains the puzzling findings of Pfaff (2013) and Binder and Coad (2010). Some implications of this finding to the construction of equivalence scales are discussed in the next section.

I also analyse how using non-equivalized reference income affects my estimates. See table R3. Doing so raises the short- run reference income effect to also become positive in Germany. As a consequence, the negative long run effect is no longer significant for Germany. A potential explanation may be analogous to that for own income: In the short run, individuals in Germany react positively to household formation in the reference group but adapt to such a positive effect over time. However, this does not appear to hold in the UK where estimates are not much affected.

While Di Tella et al. (2010) use four-year lags, I only use three-year lags. One may therefore worry that the main results presented here are an artifact of the chosen number of lags. In general, it is useful to evaluate how the number of lags impacts results. Therefore, in tables R4 (Germany) and R5 (UK), I use the preferred specification in column 5 of tables 2 and 3 but vary the number of lags from zero to six (with more lags, the instrument becomes weak). In tables R6 and R7, I apply the same procedure to the specification of column 4 in tables 2 and 3 (i.e. without the industry and occupation dummies). In general, results are quite robust to varying the number of lags. In Germany, the long run effects of both own and reference income become larger with the addition of each lag, suggesting reinforcement rather than adaptation. In the UK, for $K > 4$, the long run effect of own income becomes insignificant and is sometimes smaller than the short run effect. However, the adaptation estimate remains insignificant. Thus overall, my key conclusions are robust to the choice of the number of lags.

It may be that individuals anticipate income changes. If so, the adaptation process may begin prior to the actual income change. As a consequence, the estimate of the short run effect of own income would be biased downward and the adaptation estimate would be biased towards zero. To evaluate whether this is a concern, in table R8 I replicate tables columns 4 and 5 of tables 2 and 3, but add a lead of own income and leads of controls to each specification. For both own and reference income, I here define the 'short run effect' as the sum of the contemporaneous and the anticipation effect. Despite an increase in magnitude, I continue not to find a significant adaptation effect for own income, though the anticipation term is always positive. With respect to reference income, I

also find a positive anticipation effect for both countries, with the total short run effect always being smaller than the anticipation effect. This is in line with the hypothesized signalling function of reference income.

My instrument relies on over-time variation in industry/occupation premia that accrue to individuals. As already noted in section 3.2, variation in the instrument may either occur via variation within an industry/occupation combination, or via individuals switching into other industries/occupations (i.e. changing jobs). Such job switching behaviour may be endogenous. To gauge the severity of this potential issue, table R9 replicates columns 4 and 5 of tables 2 and 3 but restricts the sample to those that did not switch industries/occupations in the year they are observed. Doing so reduces the short run effect, but generally increases the long run effect of own income in both countries (but much more so in Germany). The long run reference income effect for Germany is also diminished. Nevertheless, none of the key conclusions are affected. However, since the choice not to switch jobs may also be endogenous, these estimates cannot be regarded as a complete test of this concern.

The instrument also relies on being able to assign (where available) a unique industry and occupation to each respondent.[21] However, when people have a second job this may not be possible. This is the case in 6.4% and 6.6% of observations in Germany and the UK, respectively. In table R10, I exclude these observations and thereby evaluate whether my results depend on them. Moreover, and as a separate concern, since the instrument is estimated from within the sample, individuals to whom these estimates are assigned contribute to these estimations. To address this endogeneity concern, table R11 excludes observations in which the number of people

[21] Unfortunately, the GSOEP documentation provides no information on how such cases are handled, i.e. which occupation is assigned when a respondent works more than one job. In the UK, the highest paying job is assigned.

with the same occupation or industry in the same wave as the respondent is smaller than 50.[22] The main results of tables 2 and 3 are not substantially affected by either of these exclusions.

Some individuals will be included in the sample who do not have an income of their own (due to e.g. unemployment or retirement). For these individuals, any variation in predicted household income will depend on other household members. While the inclusion of these individuals need not bias my estimates, these respondents may nevertheless react differently to variation in predicted household income than those who have a labour income of their own. Therefore, table R.12 excludes these individuals from my estimations. For Germany, their exclusion results in a much diminished short run effect of own income, which is in line with the previous hypothesis that adaptation to job changes may drive the negative (though insignificant) estimate for adaptation in column 5 of table 2.

My long run effect estimates for reference income are also somewhat diminished and now insignificant. The instrument for own income and (true) reference income may be positively correlated. The inclusion of reference income could then move the estimates for own income upwards (though this would not constitute a bias). In tables R13 and R14, it is shown that removing reference income from the estimation impacts the estimates for own income little, implying a weak correlation between own and reference income. Thus, if we believe that the reference group is partly constituted by individuals from the same industry/occupation (such individuals may approximate colleagues), this small correlation implies that reference income is measured with considerable error. In turn, if this measurement error is random, this also implies that the true effects of reference income are larger than here estimated, and that my own income effects are biased downward. Moreover, if we believe that the true reference income effect is subject to

[22] A better approach would be to simply exclude each individual from the computation of the instrument or reference income assigned to her. However, since this would amount to a separate regression for each observation (i.e. ~100,000 regressions) this was not feasible

reinforcement over time, my long run effect of own income should be biased downward more than my short run effect estimates, implying even less adaptation than currently observed.

The addition of dummies for industry and occupation in column 5 of tables 2 and 3 may be viewed as problematic in so far as a large portion of the variation in predicted incomes may be explained by these dummies, potentially making my IV estimates less stable. In table R15, I check how the non-instrumented results react to the addition of these dummies. Results are virtually unchanged with these dummies. As a further test on the instrument and its reaction to the addition of industry/occupation dummies, in table R16 I investigate the sensitivity of my results to constructing the instrument in a slightly different fashion: Instead of regressing labour incomes on industry and occupation dummies (plus sex and age(-squared)) in each wave, I regress labour income on the interactions of these dummies in each wave and construct the instrument from these estimates.[23] For both Germany and the UK, this results in an increase in the estimated short run effect of own income. For Germany, and without separately adding industry/occupation dummies this increase results in a marginally significant adaptation estimate. My estimate of adaptation also remains negative when adding industry/occupation dummies. However, effects from merely changing jobs may now also be picked up by the instrument via the industry-occupation interactions and are therefore not entirely controlled for.

Vendrik (2013) includes a lag of the dependent variable - life satisfaction - to his estimated equation. This is to further guard against possible omitted variables biases and to implicitly model lags of income in the further past. A more in-depth account of this justification and the econometric approach is given in appendix A. My results in table A1 verify that my key results do not depend on such an alternative approach.

[23] In other words, for each wave, I estimate: $y_i^{labour} = \alpha + \beta_1 Male_i + \beta_2 Age_i + \beta_3 \text{Age}_i^2 + \delta_{ind,occ}(INDUSTRY_i * OCCUPATION_i) + \epsilon_i$, where $\delta_{ind,occ}$ is an estimated parameter specific to each industry-occupation combination.

Finally, I assessed how robust my key estimates are against the removal of any of my controls. These results can be found in tables R17 and R18. This exercise shows that the estimates for own income are strongly diminished when excluding controls for the number of adults. This is because household income mechanically depends on the number of adults in the household, and each additional adult household member has a negative effect on life satisfaction (see section 4.4 below). This is consistent with the relative stability of results when using equivalized household income, but their relative instability when using raw household income. Moreover, concerning Germany, my estimate of adaptation also strongly depends on the employment status of the partner, yielding a sizable and significant estimate of adaptation when excluding this control. Lastly, in both countries, my estimate of the long run effect of reference income is very sensitive to the inclusion of wave dummies. This is probably due to a negative time-trend in life satisfaction across waves which is picked up by reference income, which is increasing with time. The exclusion of any other control variable does not change my substantive conclusions.

4.4 Investigating adaptation/reinforcement to household composition

In Section 4.3, I asserted that results which show significant adaptation to raw household income are likely driven by adaptation to changes in household composition. Such adaptation to household composition is not accounted for when failing to include lags of controls. However, using equivalized household income amounts to an implicit control for lags of household composition. Let y_{it}^{eq} be equivalized HH income and y_{it}^{defl} be non-equalized HH income. Let $EQ_{it} \equiv (1 + \gamma_1(nadults_{it} - 1)) + \gamma_2 nchildren_{it}$ denote the equivalization scale which depends on the number of adults and children. In the OECD equivalence scale, γ_1 and γ_2 are set to 0.5 and 0.3, respectively. Thus, $y_{it}^{eq} = y_{it}^{defl}/EQ_{it}$. We can therefore, if we were using equivalized income for estimation, write equation (1) as:

$$LS_{it} = \sum_{k=0}^{K} \beta_{0-k} \ln(y_{it-k}^{defl}/EQ_{it-k}) + \sum_{k=0}^{K} \gamma_{0-k} \ln(y_{it-k}^{ref}) + \phi X_{it} + \mu_i + \tau_t + \epsilon_{it}$$

Which can be rewritten to give:

$$LS_{it} = \sum_{k=0}^{K} \beta_{0-k}^{inc} \ln(y_{it-k}^{defl}) + \sum_{k=0}^{K} \beta_{0-k}^{comp} \ln(EQ_{it-k}) + \dots \quad (7)$$

Hence, terms for own equivalized income implicitly include controls for the ln of the number of adults and children in the household. It is assumed that for all k , $\beta_{0-k}^{comp} = -\beta_{0-k}^{inc}$. If one thinks that household income should be equivalized with respect to life satisfaction, this is a substantive assumption. In so far as many in the literature view life satisfaction as a proxy for utility (e.g. Layard, Mayraz, and Nickell 2008; Stutzer 2004; Bayer and Juessen 2015), and view equivalence scales as a means by which to assess - loosely speaking - how much more income one would need in an n person household to attain the same utility as living alone (Rojas 2007; Van Praag and Ferrer-i-Carbonell 2008; Deaton and Muellbauer 1986), this appears plausible. Only when it is in fact true that $\beta_0^{comp} = -\beta_0^{inc}$, is the contemporaneous life satisfaction effect of own equivalized income comparable across household compositions. Similarly, only when it is true that $\sum_{k=0}^{\infty} \beta_{0-k}^{comp} = \sum_{k=0}^{\infty} -\beta_{0-k}^{inc}$, is the long run life satisfaction effect of own equivalized income comparable across household compositions. When it is for example the case that $\beta_0^{inc} + \beta_0^{comp} > 0$, then the negative short run effect of an increase in household size is smaller than is assumed by the equivalization scale. In that case, and if one would wish to equivalize on short run life satisfaction effects, one ought to decrease the weight on either the number of adults (γ_1) or the number of children (γ_2).

I pursue the question of under- vs over-compensation in columns 1 and 2 of table 4. There, I estimate equation (7) , but let all β_{0-k}^{inc} and β_{0-k}^{comp} be estimated freely (and remove the control for child birth). For both countries, the negative effect of household composition is much larger in the long run than in the short run, implying significant reinforcement. Furthermore, in Germany the OECD equivalence scales overcompensates (i.e. $\beta_0^{inc} + \beta_0^{comp} > 0$) in the short run, but undercompensates (i.e. $\sum_{k=0}^{\infty} \beta_{0-k}^{inc} + \beta_{0-k}^{comp} < 0$) in the long run. In the UK, the OECD scale

neither under- nor overcompensates in the short run, but severely undercompensates in the long run.

Table 4. Undercompensation or overcompensation?

	(1) W.Germany, Instr.HH income, lagged controls, indust. & occup.	(2) UK, Instr.HH income, lagged controls, indust. & occup.	(3) W.Germany, Instr.HH income, lagged controls, indust. & occup., lagged ln(adults) & ln(children)	(4) UK, Instr.HH income, lagged controls, indust. & occup., lagged ln(adults) & ln(children)
SR effect of own income	0.281+	0.281***	0.326*	0.284***
	(0.177)	(0.096)	(0.179)	(0.098)
LR effect of own income	0.312***	0.388***	0.352***	0.381***
	(0.113)	(0.098)	(0.128)	(0.113)
Adaptation to own income	0.031	0.107	0.026	0.097
	(0.167)	(0.101)	(0.167)	(0.106)
SR effect of HH size	-0.062	-0.319***		
	(0.175)	(0.105)		
LR effect of HH size	-0.428***	-0.591***		
	(0.107)	(0.096)		
Adaptation to HH size	-0.366**	-0.272***		
	(0.163)	(0.105)		
SR Over-/under-compensation	0.219***	-0.038		
	(0.043)	(0.041)		
LR Over-/under-compensation	-0.116**	-0.203***		
	(0.047)	(0.053)		
SR effect of # of adults			-0.094	-0.212***
			(0.117)	(0.069)
LR effect of # of adults			-0.319***	-0.373***
			(0.086)	(0.079)
Adaptation to # of adults			-0.225**	-0.161**
			(0.108)	(0.075)
SR effect of # of children			0.077*	-0.023
			(0.046)	(0.034)
LR effect of # of children			-0.045+	-0.123***
			(0.028)	(0.027)
Adaptation to # of children			-0.121***	-0.100***
			(0.044)	(0.034)
Kleibergen-Paap F-stat.			156.149	41.943
Serial Correlation	0.027	0.023	0.022	0.018
Average T	7.456	7.456	7.456	7.456
# of Persons	17,282	17,282	17,282	17,282
# of Observations	128,850	128,850	128,850	128,850

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Several prior studies have derived equivalence scales using subjective satisfaction data. Examples of this are Schwarze (2003), Borah, Keldenich, and Knabe (2019), Biewen and Juhasz (2017), and Van Praag and Ferrer-i-Carbonell (2008) in the German case, Bollinger, Nicoletti, and Pudney

(2012) and Van Praag and Ferrer-i Carbonell (2008) in the British case, or Rojas (2007) in the Mexican case. All these studies find that the OECD equivalence scale overcompensates. However, every study exclusively estimated short run equivalence scales, failing to take into account reinforcement in the negative effect of household composition. My results imply that an equivalence scale that takes into account adaptation and reinforcement dynamics to both income and household composition yields much smaller economies of scale in the long run. Importantly these would be even smaller than implied by the OECD.

It is a further question whether it is the number of adults or the number of children that drives these results. It would have been nice to estimate $\beta_{0-k}^{comp}, \gamma_1$ and γ_2 freely (as do Biewen and Juhasz 2017), but this would not be estimable linearly and thus not be easily accomplished in my panel-IV context. I therefore estimated the following, simpler equation in columns 3 and 4 of table 4:

$$LS_{it} = \sum_{k=0}^{K} \beta_{0-k}^{inc} \ln(y_{it-k}^{defl}) + \sum_{k=0}^{K} \beta_{0-k}^{adu} \ln(nadults_{it-k}) + \sum_{k=0}^{K} \beta_{0-k}^{kid} \ln(1 + nkids_{it-k}) + \sum_{k=0}^{K} \gamma_{0-k} \ln(y_{it-k}^{ref}) + \sum_{k=0}^{K} \phi_{0-k} X_{it-k} + \mu_i + \tau_t + \epsilon_{it} \quad (8)$$

Independent of the country considered, the negative effect of additional adults in the household always far exceeds that of additional children, even beyond the difference implied by the OECD scale. For Germany, I even find positive short- run effects of additional children, which turn negative in the long run. These results for children are in line with previous results that show full adaptation to child-birth (see Angeles 2010; Clark and Georgellis 2013; Clark et al. 2008a). Finally, the specification of equation (8) allows me to derive short- and long run equivalence scales (cf. Rojas 2007). Taking a single adult living alone as the reference case, I can find the amount of additional income necessary to obtain the same life satisfaction as when living with additional n^{adu} adults and n^{kid} children. This is given by:

$$R_{SR}(Y_0 \rightarrow Y_{n^{adu}, n^{kid}}) = \exp\left[\ln\left(\frac{1}{n^{adu}}\right)\frac{\beta_0^{adu}}{\beta_0^{inc}} + ln\left(\frac{1}{1+n^{kid}}\right)\frac{\beta_0^{kid}}{\beta_0^{inc}}\right] \quad (9)$$

To see this, solve for y_{it}^{defl} in equation (8). Then take the ratio of this expression when $nadults = n^{adu}$ to when $nadults = 1$, and when $nkids = n^{kid}$ to when $nkids = 0$ and simplify. For the long run, replace the short run coefficients β_0^{inc}, β_0^{adu}, and β_0^{kid} by their long run equivalents $\sum_{k=0}^{3}\beta_{0-k}^{inc}$, $\sum_{k=0}^{3}\beta_{0-k}^{adu}$, and $\sum_{k=0}^{3}\beta_{0-k}^{kid}$ to obtain:

$$R_{SR}\left(Y_0 \rightarrow Y_{n^{adu},n^{kid}}\right) = \exp\left[\ln\left(\frac{1}{n^{adu}}\right)\frac{\sum_{k=0}^{3}\beta_{0-k}^{adu}}{\sum_{k=0}^{3}\beta_{0-k}^{inc}} + ln\left(\frac{1}{1+n^{kid}}\right)\frac{\sum_{k=0}^{3}\beta_{0-k}^{kid}}{\sum_{k=0}^{3}\beta_{0-k}^{inc}}\right] \qquad (10)$$

I plot results of that exercise in figure 5. Full tables are given in appendix B. From figure 5 it becomes evident that the OECD equivalence scales completely overcompensates for additional children, both in the short- and in the long run. However, the OECD scale only overcompensates for additional adults in the short run for Germany. In the long run, the OECD scale severely undercompensates for the effect of additional adults. Indeed, in the British case there appear to be no economies of scale in the long run at all.[24] In so far as the German welfare state is much more

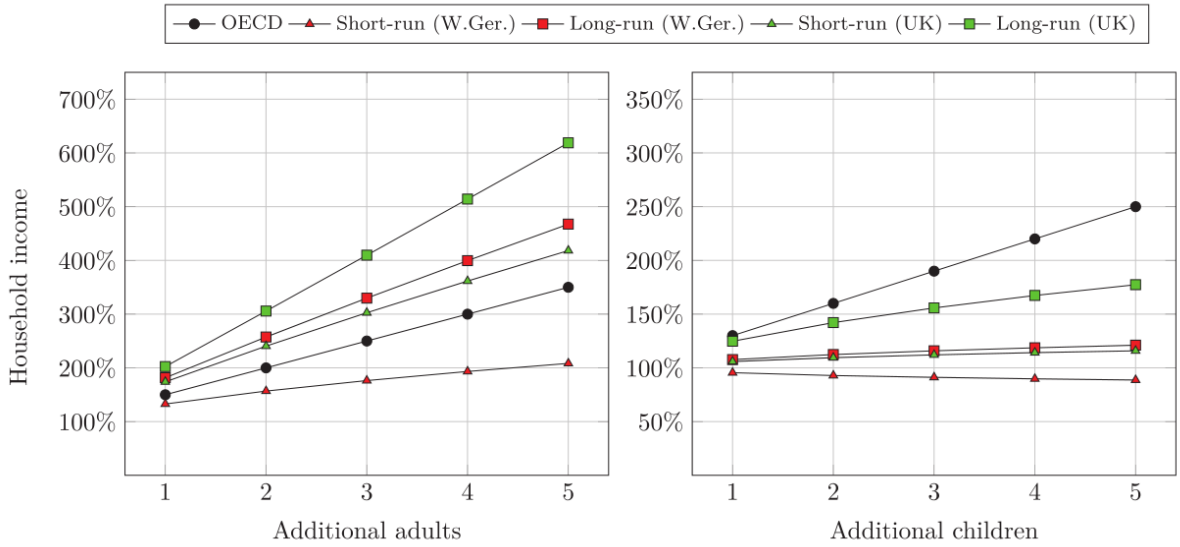


Figure 5. Illustration of equivalence scales computed from equations (9) and (10) on the basis of parameter estimates from Table 4, columns 3 (Germany) and 4 (UK). In both panels a single adult living alone is taken to be the reference case. The panel on the left shows the amount of household income (in % of the reference case) necessary to compensate (i.e. to obtain the same level of life satisfaction as in the reference case) for each additional adult in the household. The right panel show the amount of household income necessary to maintain constant life satisfaction with each additional child in the household.

[24] Note that I control for marital status. Therefore, much of the positive effect of being in a multi-adult household will be captured by this variable.

generous with respect to child-care than the British (thus lessening the burden of additional children), and more willing to support partnered households, these observed country differences may be driven by the institutional structure of these countries.

5 Conclusion

Substantively I make four main points. First, I find no significant evidence of adaptation to own income. All cases in which my estimate of adaptation is significant and negative occur when omitting variables relating to household composition. This finding is in direct disagreement with most of the previous literature and implies that higher own incomes have lastingly positive effects on life satisfaction. Second, the effects of reference income have been consistently negative across countries and did not dissipate with time. To the contrary, I find some (albeit insignificant) indication that in the short run the effects of reference incomes are positive. This may indicate that a signalling mechanism of the sort described by Senik (2004) dominates in the short run, while a social comparison mechanism[25] dominates in the long run. Third, in the long- run the effects of own and reference income are of similar absolute magnitude in most specifications. As a consequence, the combined effects of own and reference income are always insignificant.[26] Thus, if everybody gets richer, nobody becomes happier. Although this is a sad result, it validates the much discussed Easterlin Paradox. Finally, previous endeavours that derived equivalence scales from life satisfaction data did not take into account adaptation to household size. The resulting bias appears to be substantial, indicating that the OECD equivalence scale severely undercompensates in the long run.

Methodologically two major points are worth noting. First, instrumenting for own income is paramount. I find much larger income effects when instrumenting, especially with respect to the UK. The pronounced rise in effect sizes when instrumenting is in line with the IV results of Vendrik

[25] Vendrik and Woltjer (2007) propose an alternative mechanism where greater reference incomes may stifle the ability of individuals to participate in social activities.

[26] Except when using non-equivalized reference income and small lag lengths ($K < 2$) in Germany.

(2013), Powdthavee (2010), Luechinger (2009) and Knight and Gunatilaka (2010). Second, the use of lags of controls was crucial. In particular, what was previously taken to be adaptation to income, appears in fact to have been reinforcement of a negative household composition effect. In the present application, such adaptation can be controlled for via the use of equivalized income. However, this is not generally true. It may therefore be the case that similar problems apply to the effects of other socio-economic variables for which adaptation processes are thought to occur.

An important limitation of this study is the use of annual income data. If adaptation to income occurs much faster than within the first year, such adaptation could not be detected (Ferrer-i-Carbonell and Van Praag, 2008). Tackling this problem with temporally more fine-grained data (perhaps the Australian data Frijters, Johnston, and Shields (2011) use) or a combination of annual and monthly household income, seems like a useful avenue for future research. However, given that I do find sizable short run effects of income, it seems that adaptation would have to follow from a very large, and possibly implausibly large, initial effect of income. Moreover, my significant long run effects imply that any adaptation must be partial at best. Nevertheless, and to quote Ferrer-i-Carbonell and Van Praag (2008), *the discussion is not finished yet.*

References

Angeles, Luis. 2010. 'Adaptation and Anticipation Effects to Life Events in the United Kingdom'. *Working Paper*. http://repo.sire.ac.uk/handle/10943/134.

Bayer, Christian, and Falko Juessen. 2015. 'Happiness and the Persistence of Income Shocks'. *American Economic Journal: Macroeconomics* 7 (4): 160–187.

Berthoud, Richard, and Mark Bryan. 2011. 'Income, Deprivation and Poverty: A Longitudinal Analysis'. *Journal of Social Policy* 40 (1): 135–156. https://doi.org/10.1017/S0047279410000504.

Biewen, Martin, and Andos Juhasz. 2017. 'Direct Estimation of Equivalence Scales and More Evidence on Independence of Base'. *Oxford Bulletin of Economics and Statistics* 79 (5): 875–905. https://doi.org/10.1111/obes.12166.

Binder, Martin, and Alex Coad. 2010. 'An Examination of the Dynamics of Well-Being and Life Events Using Vector Autoregressions'. *Journal of Economic Behavior & Organization* 76 (2): 352–371.

Bollinger, Christopher, Cheti Nicoletti, and Stephen Pudney. 2012. 'Two Can Live as Cheaply as One...But Three's a Crowd'. *University of Kentucky Center for Poverty Research Discussion Paper Series*. https://uknowledge.uky.edu/ukcpr_papers/31.

Borah, Melanie, Carina Keldenich, and Andreas Knabe. 2019. 'Reference Income Effects in the Determination of Equivalence Scales Using Income Satisfaction Data'. *Review of Income and Wealth* 65 (4): 736–70. https://doi.org/10.1111/roiw.12386.

Chang, Andrew C., and Phillip Li. 2018. 'Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Often Not"'. *Critical Finance Review* 7.

Clark, Andrew E. 1999. 'Are Wages Habit-Forming? Evidence from Micro Data'. *Journal of Economic Behavior & Organization* 39 (2): 179–200. https://doi.org/10.1016/S0167-2681(99)00031-1.

———. 2016. 'Adaptation and the Easterlin Paradox'. In *Advances in Happiness Research*, 75–94. Springer.

Clark, Andrew E., Ed Diener, Yannis Georgellis, and Richard E. Lucas. 2008a. 'Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis'. *The Economic Journal* 118 (529): F222–F243.

Clark, Andrew E., Ed Diener, Yannis Georgellis, and Richard E Lucas. 2008b. 'Lags and Leads in Life Satisfaction : A Test of the Baseline Hypothesis'. *The Economic Journal* 118 (529): 222–243.

Clark, Andrew E, Paul Frijters, and Michael A Shields. 2008. 'Relative Income, Happiness, and Utility: An Explanation for the Easterlin Paradox and Other Puzzles'. *Journal of Economic Literature* 46 (461): 95–144. https://doi.org/10.1257/jel.46.1.95.

Clark, Andrew E., and Yannis Georgellis. 2013. 'Back to Baseline in Britain: Adaptation in the British Household Panel Survey'. *Economica* 80 (319): 496–512. https://doi.org/10.1111/ecca.12007.

Clark, Andrew E., Claudia Senik, and Katsunori Yamada. 2017. 'When Experienced and Decision Utility Concur: The Case of Income Comparisons'. *Journal of Behavioral and Experimental Economics* 70: 1–9.

D'Ambrosio, Conchita, and Joachim R. Frick. 2012. 'Individual Wellbeing in a Dynamic Perspective'. *Economica* 79 (314): 284–302. https://doi.org/10.1111/j.1468-0335.2011.00896.x.

De Vos, Ignace, Ilse Ruyssen, and Gerdie Everaert. 2015. 'XTBCFE: Stata Module to Perform Bootstrap-Corrected Fixed Effects Estimation and Inference in Dynamic Panel Models'. *Statistical Software Components*.

Deaton, Angus S., and John Muellbauer. 1986. 'On Measuring Child Costs: With Applications to Poor Countries'. *Journal of Political Economy* 94 (4): 720–44.

Di Tella, Rafael, John Haisken-De New, and Robert MacCulloch. 2010. 'Happiness Adaptation to Income and to Status in an Individual Panel'. *Journal of Economic Behavior and Organization* 76 (3): 834–852. https://doi.org/10.1016/j.jebo.2010.09.016.

Di Tella, Rafael, and Robert MacCulloch. 2010. 'Happiness Adaptation to Income beyond "Basic Needs"'. In *International Differences in Wellbeing*, edited by Ed Diener, John Helliwell, and Daniel Kahneman. Oxford University Press.

Dumludag, Devrim, Ozge Gokdemir, and Selay Giray. 2016. 'Income Comparison, Collectivism and Life Satisfaction in Turkey'. *Quality & Quantity* 50 (3): 955–980.

Earp, Brian D., and David Trafimow. 2015. 'Replication, Falsification, and the Crisis of Confidence in Social Psychology'. *Frontiers in Psychology* 6. https://doi.org/10.3389/fpsyg.2015.00621.

Easterlin, Richard. 2016. 'Paradox Lost?' *IZA Discussion Paper*, no. 9676.

Easterlin, Richard A. 2005. 'A Puzzle for Adaptive Theory'. *Journal of Economic Behavior and Organization* 56 (4 SPEC. ISS.): 513–521. https://doi.org/10.1016/j.jebo.2004.03.003.

Ferrer-i-Carbonell, Ada. 2005. 'Income and Well-Being: An Empirical Analysis of the Comparison Income Effect'. *Journal of Public Economics* 89 (5–6): 997–1019. https://doi.org/10.1016/j.jpubeco.2004.06.003.

Ferrer-i-Carbonell, Ada, and Paul Frijters. 2004. 'How Important Is Methodology for the Estimates of the Determinants of Happiness?' *The Economic Journal* 114 (497): 641–659.

Ferrer-i-Carbonell, Ada, and Bernard M. S. Van Praag. 2008. 'Do People Adapt to Changes in Income and Other Circumstances? The Discussion Is Not Finished Yet'.

Figari, Francesco. 2012. 'Cross-National Differences in Determinants of Multiple Deprivation in Europe'. *Journal of Economic Inequality* 10 (3): 397–418. https://doi.org/10.1007/s10888-010-9157-9.

Frijters, Paul, David W. Johnston, and Michael A. Shields. 2011. 'Happiness Dynamics with Quarterly Life Event Data'. *Scandinavian Journal of Economics* 113 (1): 190–211. https://doi.org/10.1111/j.1467-9442.2010.01638.x.

Goerke, Laszlo, and Markus Pannenberg. 2015. 'Direct Evidence for Income Comparisons and Subjective Well-Being across Reference Groups'. *Economics Letters* 137: 95–101. https://doi.org/10.1016/j.econlet.2015.10.012.

Graham, Carol, and Milena Nikolova. 2015. 'Bentham or Aristotle in the Development Process? An Empirical Investigation of Capabilities and Subjective Well-Being'. *World Development* 68 (1): 163–179. https://doi.org/10.1016/j.worlddev.2014.11.018.

Hamermesh, Daniel S. 2007. 'Viewpoint: Replication in Economics'. *Canadian Journal of Economics/Revue Canadienne d'economique* 40 (3): 715–33. https://doi.org/10.1111/j.1365-2966.2007.00428.x.

Hanglberger, Dominik, and Joachim Merz. 2011. 'Are Self-Employed Really Happier Than Employees? An Approach Modelling Adaptation and Anticipation Effects to Self-Employment and General Job Changes'. *IZA Discussion Paper*, no. 5629. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1884784.

Hirschman, Albert O., and Michael Rothschild. 1973. 'The Changing Tolerance for Income Inequality in the Course of Economic Development* With A Mathematical Appendix'. *The Quarterly Journal of Economics* 87 (4): 544–566.

Hurst, Erik, Geng Li, and Benjamin Pugsley. 2013. 'Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed'. *The Review of Economics and Statistics* 96 (1): 19–33. https://doi.org/10.1162/REST_a_00363.

Ifcher, John, Homa Zarghamee, and Carol Graham. 2018. 'Local Neighbors as Positives, Regional Neighbors as Negatives: Competing Channels in the Relationship between Others' Income, Health, and Happiness'. *Journal of Health Economics* 57: 263–276.

Jayawickreme, Eranda, and James O Pawelski. 2012. 'Positivity and the Capabilities Approach'. *Philosophical Psychology* 5089 (December): 1–18.

Kingdon, Geeta Gandhi, and John Knight. 2007. 'Community, Comparisons and Subjective Well-Being in a Divided Society'. *Journal of Economic Behavior and Organization* 64 (1): 69–90. https://doi.org/10.1016/j.jebo.2007.03.004.

Knight, John, and Ramani Gunatilaka. 2010. 'Great Expectations? The Subjective Well-Being of Rural-Urban Migrants in China'. *World Development* 38 (1): 113–124.

Layard, Richard, Guy Mayraz, and Stephen Nickell. 2008. 'The Marginal Utility of Income'. *Journal of Public Economics* 92 (8): 1846–1857.

Levy, Horacio, and Stephen P. Jenkins. 2012. 'Documentation for Derived Current and Annual Net Household Income Variables, BHPS Waves 1-18'. *UK Data Archive Study Number 3909*.

Luechinger, Simon. 2009. 'Valuing Air Quality Using the Life Satisfaction Approach'. *The Economic Journal* 119 (536): 482–515.

Luttmer, Erzo FP. 2005. 'Neighbors as Negatives: Relative Earnings and Well-Being'. *The Quarterly Journal of Economics* 120 (3): 963–1002.

Mayraz, Guy, Gert Wagner, and Juergen Schupp. 2009. 'Life Satisfaction and Relative Income : Perceptions and Evidence'. *CEP Discussion Paper*, no. 938. http://cep.lse.ac.uk/pubs/download/dp0938.pdf.

McBride, M. 2001. 'Relative-Income Effects on Subjective Well-Being in the Cross-Section'. *Journal of Economic Behavior & Organization* 45 (3): 251–278. https://doi.org/10.1016/S0167-2681(01)00145-7.

Nussbaum, Martha Craven. 2000. *Women and Human Development: The Capabilities Approach*. Cambridge University Press.

Paul, Satya, and Daniel Guilbert. 2013. 'Income-Happiness Paradox in Australia: Testing the Theories of Adaptation and Social Comparison'. *Economic Modelling* 30 (Supplement C): 900–910. https://doi.org/10.1016/j.econmod.2012.08.034.

Pfaff, T. 2013. 'Income Comparisons, Income Adaptation, and Life Satisfaction: How Robust Are Estimates from Survey Data?' *SOEPpapers*, no. 555. http://www.econstor.eu/handle/10419/76664.

Pischke, Jörn-Steffen, and Hannes Schwandt. 2012. 'A Cautionary Note on Using Industry Affiliation to Predict Income'. *IZA Discussion Paper*, no. 6840.

Powdthavee, Nattavudh. 2010. 'How Much Does Money Really Matter? Estimating the Causal Effects of Income on Happiness'. *Empirical Economics* 39 (1): 77–92. https://doi.org/10.1007/s00181-009-0295-5.

Qizilbash, Mozaffar. 2006. 'Well-Being, Adaptation and Human Limitations'. *Royal Institute of Philosophy Supplements* 59: 83–109.

Riedl, Maximilian, and Ingo Geishecker. 2014. 'Keep It Simple: Estimation Strategies for Ordered Response Models with Fixed Effects'. *Journal of Applied Statistics* 41 (11): 2358–74. https://doi.org/10.1080/02664763.2014.909969.

Rojas, Mariano. 2007. 'A Subjective Well-Being Equivalence Scale for Mexico: Estimation and Poverty and Income-Distribution Implications'. *Oxford Development Studies* 35 (3): 273–293.

Schokkaert, Erik. 2007. 'Capabilities and Satisfaction with Life'. *Journal of Human Development* 8 (3): 415–430. https://doi.org/10.1080/14649880701462239.

Schwarze, Johannes. 2003. 'Using Panel Data on Income Satisfaction to Estimate Equivalence Scale Elasticity'. *Review of Income and Wealth* 49 (3): 359–72. https://doi.org/10.1111/1475-4991.00092.

Sen, Amartya. 1985. *Commodities and Capabilities*. Amsterdam: North-Holland.

Senik, Claudia. 2004. 'When Information Dominates Comparison: Learning from Russian Subjective Panel Data'. *Journal of Public Economics* 88 (9): 2099–2123.

———. 2008. 'Ambition and Jealousy: Income Interactions in the "Old" Europe versus the "New" Europe and the United States'. *Economica* 75 (299): 495–513. https://doi.org/10.1111/j.1468-0335.2007.00629.x.

———. 2009. 'Direct Evidence on Income Comparisons and Their Welfare Effects'. *Journal of Economic Behavior and Organization* 72 (1): 408–424. https://doi.org/10.1016/j.jebo.2009.04.019.

Stutzer, Alois. 2004. 'The Role of Income Aspirations in Individual Happiness'. *Journal of Economic Behavior and Organization* 54 (1): 89–109. https://doi.org/10.1016/j.jebo.2003.04.003.

Van Praag, Bernard M. S. 1971. 'The Welfare Function of Income in Belgium: An Empirical Investigation'. *European Economic Review* 2 (3): 337–69.

Van Praag, Bernard M. S., and Ada Ferrer-i-Carbonell. 2008. *Happiness Quantified: A Satisfaction Calculus Approach*. 2nd ed. Oxford University Press.

Vendrik, Maarten C. M. 2013. 'Adaptation, Anticipation and Social Interaction in Happiness: An Integrated Error-Correction Approach'. *Journal of Public Economics* 105: 131–149. https://doi.org/10.1016/j.jpubeco.2013.06.009.

Vendrik, Maarten C. M., and Geert B. Woltjer. 2007. 'Happiness and Loss Aversion: Is Utility Concave or Convex in Relative Income?' *Journal of Public Economics* 91 (7–8): 1423–1448. https://doi.org/10.1016/j.jpubeco.2007.02.008.

Wolbring, Tobias, Marc Keuschnigg, and Eva Negele. 2013. 'Needs, Comparisons, and Adaptation: The Importance of Relative Income for Life Satisfaction'. *European Sociological Review* 29 (1): 86–104. https://doi.org/10.1093/esr/jcr042.

Appendix

A Addition of lagged life satisfaction

Vendrik (2013) uses an error-correction model as his primary specification. An alternative representation of such a model is an autoregressive distributed lag model, which may be written as:

$$LS_{it} = \sum_{k=0}^{K} \beta_{0-k} \ln(y_{it-k}) + \sum_{k=0}^{K} \gamma_{0-k} \ln(y_{it-k}^{ref}) + \sum_{k=0}^{K} \phi_{0-k} X_{it-k} + \rho LS_{it} + \mu_i + \tau_t + \epsilon_{it} \quad (11)$$

Using repeated substitutions, equation (11) can be rewritten as follows (controls omitted for brevity):

$$\begin{aligned} LS_{it} = {} & \beta_0 \ln(y_{it}) + (\beta_{-1} + \rho\beta_0) \ln(y_{it-1}) + (\beta_{-2} + \rho\beta_{-1} + \rho^2\beta_0) \ln(y_{it-2}) + \cdots \\ & + (\beta_{-K} + \cdots + \rho^K \beta_0) \ln(y_{it-K}) + \sum_{m=1}^{\infty} \rho^m \left(\sum_{l=0}^{K} \rho^l \beta_{-K-l} \right) \ln(y_{it-K-m}) + \cdots \end{aligned} \quad (12)$$

Defining $\beta_{0-k}^*, \gamma_{0-k}^*$, etc. appropriately, this can be written as:

$$LS_{it} = \sum_{k=0}^{\infty} \beta_{0-k}^* \ln(y_{it-k}) + \sum_{k=0}^{\infty} \gamma_{0-k}^* \ln(y_{it-k}^{ref}) + \sum_{k=0}^{\infty} \phi_{0-k}^* X_{it-k} + \mu_i^* + \tau_t^* + \epsilon_{it}^* \quad (13)$$

The addition of lagged life satisfaction therefore implicitly models the effects of indefinitely long lags of own and reference income, as well as all controls. In this model, we must assume that life satisfaction, beyond the part that is explicitly modelled, is an autoregressive process of order 1, with $0 \leq \rho < 1$. The effects of income and all controls at $t - K - m$ are assumed to decay at rate ρ^m for $m > 0$. This homogeneous rate of decay is also assumed for the effects of the unobserved time-varying controls. To obtain the SR, LR, and adaptation effects the same formulas as those presented in Section 3.1 can be applied to equation (13). The sum $\sum_{k=0}^{\infty} \beta_{0-k}^*$, i.e. the long run effect of own income, can be expressed as $\sum_{k=0}^{K} \beta_{0-k} / (1 - \rho)$ (the same goes for reference incomes).[27] In a fixed effects setting, the OLS estimate of ρ will be downwardly biased for finite T

[27] To see this, set LS and $\ln(y)$ to their unchanging long run values in equation (11) and rearrange.

(Nickell Bias). I therefore apply the bootstrap bias correction of De Vos, Ruyssen, and Everaert (2015). Vendrik (2013) finds a further small-T bias in the estimate of ρ that goes beyond the Nickell Bias. This is why he fixates ρ to an estimate from a balanced panel with $T = 21$. I observe the same: with the full German sample I obtain $\rho = 0.13$. When only selecting respondents who are observed in every wave (which yields an effective T of 29), the estimated ρ increases to 0.26, which is close to Vendrik's estimate ($= 0.27$). Therefore, I also fixate ρ to the estimate with largest possible T, making these estimates most comparable to Vendrik (2013).[28] See table A1 for results.

Table A1. Main specifications after adding lags of life satisfaction

	(1) W.Germany, instr. HH income, lagged controls	(2) W.Germany, instr. HH income, lagged controls, indust. & occup.	(3) UK, instr. HH income, lagged controls	(4) UK, instr. HH income, lagged controls, indust. & occup.
Lagged life satisfaction	0.266	0.266	0.168	0.168
SR effect of own inc.	0.671***	0.369*	0.206**	0.201*
	(0.206)	(0.191)	(0.100)	(0.107)
LR effect of own inc.	0.437***	0.357**	0.317**	0.342**
	(0.146)	(0.148)	(0.128)	(0.137)
Adaptation to own inc.	-0.235	-0.012	0.111	0.141
	(0.194)	(0.182)	(0.122)	(0.127)
SR effect of ref. inc.	-0.198	-0.157	0.199+	0.217+
	(0.168)	(0.167)	(0.155)	(0.157)
LR effect of ref. inc.	-0.345**	-0.321*	-0.177	-0.186
	(0.168)	(0.169)	(0.206)	(0.210)
Adaptation to ref. inc.	-0.148	-0.164	-0.376*	-0.403*
	(0.201)	(0.201)	(0.206)	(0.209)
Combined SR effect	0.474**	0.212	0.405**	0.418**
	(0.238)	(0.230)	(0.170)	(0.173)
Combined LR effect	0.091	0.036	0.140	0.156
	(0.188)	(0.194)	(0.229)	(0.233)
Combined adaptation	-0.382+	-0.176	-0.265	-0.262
	(0.249)	(0.244)	(0.220)	(0.224)
Kleibergen-Paap F-stat.			156.149	41.943
Serial Correlation	0.027	0.023	0.022	0.018
Average T	7.456	7.456	7.456	7.456
# of Persons	17,282	17,282	17,282	17,282
# of Observations	128,850	128,850	128,850	128,850

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

[28] Moreover, when selecting the same subset of respondents, but randomly selecting an observation window of seven periods (which is closest to mean T when using the full set of respondents) for each of these respondents, I get an estimate of $\rho = 0.14$. This is thus very close to the estimate obtained from the full set of respondents. It may therefore be concluded that the smaller estimate of ρ from the full sample indeed arises from a further, unaccounted-for small-T bias, and that the larger estimate for ρ obtained when $T = 29$ is less biased and not driven by selection effects.

B Full results for equivalization scales

Table B1. Short- and long run equivalence scales for Germany.

		Number of children						
		0	1	2	3	4	5	6
Number of Adults	1	1.000/1.000	0.956/1.077	0.931/1.124	0.913/1.160	0.900/1.188	0.889/1.211	0.880/1.231
	2	1.328/1.814	1.270/1.953	1.236/2.039	1.213/2.103	1.196/2.154	1.182/2.196	1.170/2.233
	3	1.569/2.569	1.499/2.767	1.460/2.889	1.433/2.979	1.412/3.051	1.395/3.111	1.381/3.163
	4	1.765/3.289	1.687/3.542	1.643/3.699	1.612/3.814	1.589/3.906	1.570/3.983	1.554/4.049
	5	1.934/3.984	1.848/4.290	1.800/4.480	1.766/4.620	1.741/4.731	1.720/4.825	1.703/4.905
	6	2.084/4.660	1.992/5.018	1.939/5.240	1.903/5.403	1.876/5.534	1.853/5.642	1.835/5.736

Note: Calculated with parameter estimates from the specification of table 4, column 3. First element of each cell refers to the short run, second element refers to the long run.

Table B2. Short- and long run equivalence scales for the UK.

		Number of children						
		0	1	2	3	4	5	6
Number of Adults	1	1.000/1.000	1.058/1.249	1.094/1.422	1.120/1.559	1.141/1.675	1.158/1.775	1.173/1.865
	2	1.740/2.024	1.842/2.528	1.904/2.879	1.949/3.156	1.985/3.390	2.015/3.594	2.041/3.776
	3	2.406/3.058	2.547/3.819	2.633/4.349	2.695/4.768	2.745/5.122	2.786/5.430	2.822/5.705
	4	3.028/4.099	3.205/5.118	3.313/5.827	3.392/6.390	3.455/6.864	3.507/7.276	3.551/7.645
	5	3.619/5.143	3.830/6.422	3.960/7.313	4.054/8.019	4.129/8.613	4.191/9.131	4.244/9.594
	6	4.187/6.192	4.431/7.731	4.581/8.804	4.690/9.654	4.777/10.369	4.849/10.993	4.910/11.549

Note: Calculated with parameter estimates from the specification of table 4, column 4. First element of each cell refers to the short run, second element refers to the long run.

C Descriptives tables

Table D1. Description of Controls

Concept	GSOEP		UKHLS	
Number of children	Dummies for:		Dummies for:	
	0	-55.10%	0	-57.30%
	1	-20.40%	1	-18.10%
	2	-17.90%	2	-18.90%
	3	-5.17%	3	-4.87%
	4	-1.05%	4	-0.86%
	5	-0.26%	5	-0.10%
	6 or more	-0.10%	6 or more	-0.01%
Number of adults	Dummies for:		Dummies for:	
	1	-10.00%	1	-11.10%
	2	-59.80%	2	-61.40%
	3	-18.90%	3	-17.70%
	4	-8.89%	4	-7.91%
	5	-1.88%	5	-1.45%
	6 or more	-0.53%	6 or more	-0.44%
Employment status	Dummies for:		Dummies for:	
	Non-working	-10.50%	Lt.sick or disabled	-1.52%
	Retired	-1.20%	Retired	-3.34%
	Unemployed	-2.16%	Unemployed	-1.34%
	Student	-0.96%	Student	-0.53%
	Maternity leave	-1.58%	Family Care	-4.83%
	Working	-81.40%	Working	-88.20%
Partner employment status	Dummies for:		Dummies for:	
	Working	-60.40%	Working	-65.30%
	No partner / NA	-22.30%	No partner / NA	-23.30%
	Unemployed	-1.92%	Unemployed	-1.20%
	Non-working	-15.40%	Non-working	-10.40%
Age	Mean=43.64	SD=11.07	Mean=44.06	SD=11.08
	Min=23	Max=70	Min=22	Max=70
Age squared	Mean=2026	SD=981.0	Mean=2064	SD=997.4
	Min=529	Max=4900	Min=484	Max=4900
Marital status	Dummies for:		Dummies for:	
	Married/Couple	-73.20%	Married	-66.80%
	Single	-16.50%	Single	-11.70%
	Widowed	-1.51%	Widowed	-1.26%
	Divorced	-6.27%	Divorced	-7.37%
	Separated	-2.32%	Living as couple	-12.90%
	Other	-0.22%		
Housing	Dummies for:		Dummies for:	
	Own	-52.10%	Own	-82.30%
	Rented	-47.90%	Rented privately	-8.43%
			Social housing	-9.29%
Working hours	ln(yearly working hours)		ln(weekly working hours)	
	Mean=6.30	SD=2.73	Mean=3.09	SD=1.16
	Min=0	Max=8.80	Min=0	Max=4.61
Ill health	Dummy over 0-10 subj. health		Dummy over 1-7 subj. health variable	
	=1 if subj.health < 5	(13.4%)	=1 if subj.health < 4	(21.4%)
	=0 if subj.health > 4	(86.5%)	=1 if subj.health > 3	(78.5%)
Birth of children	Dummies for		Dummies for	
	Birth	-3.99%	Birth	-3.67%
	No birth	-96.00%	No birth	-96.30%
Education	Dummies for:		Dummies for:	
	Inadequate/elementary	-16.30%	No Qualification	-6.92%
	Mittlere Reife	-32.70%	GCSE	-22.90%
	Abitur & Ausbildung	-33.50%	A-levels	-21.80%
	Tertiary education	-16.00%	Tertiary degree	-26.90%
	Other/missing	-1.45%	Other higher	-12.50%
Region	Bundesländer dummies	-11	Govt. region dummies	-11
Wave	Waves dummies	-28	Waves dummies	-18
Regional time trend	Interactions of region and waves		Interactions of region and waves	

Table D2. Descriptive statistics for observed ln household income across groups, Germany

	N	Mean	SD	Min	Max	Mean Δ	SD Δ	Min Δ	Max Δ
Leavers	29,588	10.50	0.41	8.72	11.74	0.01	0.25	-2.64	1.96
Stayers	99,262	10.50	0.41	8.71	11.74	0.01	0.21	-2.11	1.82
Occupation:									
Legislators, senior officials & managers	5,068	10.73	0.41	8.93	11.73	0.02	0.20	-1.59	0.97
Professionals	16,097	10.75	0.40	8.80	11.74	0.02	0.21	-2.10	1.74
Technicians & associate professionals	23,898	10.54	0.40	8.74	11.71	0.01	0.22	-2.32	1.77
Office Clerks	13,555	10.51	0.42	8.72	11.70	0.01	0.22	-2.05	1.58
Service, shop & market sales workers	10,432	10.39	0.41	8.71	11.68	0.00	0.24	-2.11	1.82
Skilled agricultural & fishery workers	534	10.32	0.46	8.89	11.58	0.00	0.23	-1.19	1.08
Craft & related trades workers	16,262	10.41	0.35	8.99	11.70	0.01	0.21	-1.59	1.96
Plant & machine operators & assemblers	10,075	10.38	0.35	8.72	11.56	0.00	0.22	-1.79	1.77
Elementary occupations	8,284	10.32	0.40	8.73	11.71	0.00	0.24	-2.64	1.62
Missing	24,645	10.48	0.37	8.74	11.74	0.00	0.21	-2.05	1.36
Industry:									
Agriculture, hunting & forestry	726	10.40	0.45	8.89	11.44	0.00	0.21	-1.52	0.88
Fishing	7	10.05	0.84	9.20	11.27	0.07	0.48	-0.55	1.06
Mining & quarrying	474	10.44	0.40	9.22	11.74	0.00	0.22	-1.32	1.00
Manufacturing	32,990	10.51	0.39	8.80	11.71	0.01	0.21	-1.85	1.73
Electricity, Gas & water supply	1,239	10.63	0.39	9.35	11.64	0.02	0.22	-1.27	1.44
Construction	5,842	10.43	0.36	8.74	11.73	0.01	0.22	-1.59	1.63
Wholesale & retail trade; repairs	11,689	10.42	0.41	8.78	11.70	0.01	0.24	-2.05	1.96
Hotels & restaurants	1,733	10.32	0.44	8.71	11.62	0.00	0.28	-1.55	1.64
Transport, storage & communication	5,326	10.45	0.41	8.72	11.64	0.01	0.22	-2.09	1.77
Financial intermediation	4,762	10.68	0.41	8.90	11.70	0.02	0.21	-1.54	1.24
Real estate, renting & business activities	6,267	10.54	0.47	8.74	11.71	0.01	0.25	-2.32	1.74
Public admin. & defence; social security	10,255	10.57	0.39	8.92	11.66	0.01	0.20	-1.85	1.35
Education	7,028	10.69	0.42	8.83	11.74	0.01	0.22	-2.10	1.73
Health & Social Work	11,133	10.45	0.44	8.72	11.69	0.01	0.24	-2.11	1.82
Other community, social & personal services	3,566	10.52	0.43	8.78	11.71	0.01	0.22	-2.64	1.48
Private households with employed persons	376	10.26	0.40	8.86	11.38	0.01	0.24	-1.34	1.09
Extra-territorial organisations & bodies	119	10.36	0.38	9.34	11.28	0.00	0.25	-1.31	0.71
Missing	25,318	10.48	0.37	8.74	11.74	0.00	0.21	-2.05	1.36

'Stayers' are observations in which industry/occupation combination did not change since last observation. 'Leavers' are observations in which industry/occupation combination did change since last observation. Rows for occupation are major groupings from ISCO88. Rows for industry are major groupings from NACE1.1. Δ means change between years.

Table D3. Descriptive statistics for observed ln household income across groups, UK

	N	Mean	SD	Min	Max	Mean Δ	SD Δ	Min Δ	Max Δ
Leavers	24,729	10.21	0.45	7.94	11.61	0.01	0.31	-2.12	2.10
Stayers	63,441	10.27	0.46	7.86	11.65	0.02	0.28	-2.40	2.23
Occupation:									
Legislators, senior officials & managers	12,512	10.44	0.42	8.47	11.59	0.03	0.26	-2.34	2.10
Professionals	11,985	10.46	0.41	8.25	11.63	0.02	0.27	-2.12	2.12
Technicians & associate professionals	12,273	10.33	0.43	8.22	11.65	0.02	0.27	-1.97	1.86
Office Clerks	12,456	10.19	0.44	8.18	11.57	0.02	0.29	-2.30	2.11
Service, shop & market sales workers	12,442	10.13	0.48	7.86	11.59	0.01	0.32	-2.40	1.74
Skilled agricultural & fishery workers	355	9.95	0.47	8.96	11.08	0.04	0.30	-1.48	1.37
Craft & related trades workers	5,116	10.19	0.39	8.43	11.61	0.03	0.25	-1.78	1.53
Plant & machine operators & assemblers	5,075	10.12	0.42	8.07	11.64	0.02	0.26	-2.09	2.02
Elementary occupations	5,999	10.05	0.47	7.94	11.65	0.02	0.31	-2.02	2.23
Missing	9,957	10.17	0.44	7.86	11.64	-0.02	0.31	-1.95	2.23
Industry:									
Agriculture, hunting & forestry	162	10.28	0.49	9.20	11.54	0.04	0.32	-0.83	2.02
Fishing	521	10.14	0.50	8.19	11.51	0.05	0.26	-1.48	1.06
Mining & quarrying	11,365	10.23	0.42	8.07	11.54	0.03	0.25	-2.09	2.09
Manufacturing	852	10.26	0.46	8.43	11.52	0.03	0.32	-1.91	1.50
Electricity, Gas & water supply	2,566	10.31	0.43	8.47	11.61	0.03	0.28	-1.78	2.10
Construction	8,585	10.16	0.46	8.07	11.59	0.01	0.31	-2.34	2.06
Wholesale & retail trade; repairs	1,822	10.08	0.49	8.45	11.65	0.02	0.35	-2.30	2.23
Hotels & restaurants	6,077	10.26	0.46	8.18	11.64	0.02	0.28	-1.74	1.92
Transport, storage & communication	5,963	10.30	0.48	7.86	11.59	0.03	0.28	-1.98	2.12
Financial intermediation	5,334	10.34	0.46	7.94	11.57	0.03	0.28	-1.97	1.73
Real estate, renting & business activities	6,517	10.35	0.43	8.45	11.57	0.02	0.26	-1.70	2.09
Public admin. & defence; social security	9,140	10.38	0.45	8.38	11.58	0.01	0.29	-2.40	1.69
Education	12,082	10.29	0.45	8.21	11.65	0.02	0.30	-2.00	2.04
Health & Social Work	5,491	10.20	0.45	8.03	11.54	0.02	0.28	-1.95	2.11
Other community, social & personal services	954	10.08	0.45	8.39	11.07	0.03	0.25	-1.58	1.13
Private households with employed persons	1,539	10.13	0.48	8.11	11.61	0.04	0.29	-2.07	1.40
Extra-territorial organisations & bodies	9,198	10.14	0.43	7.86	11.55	-0.02	0.32	-1.95	2.23
Missing	24,729	10.21	0.45	7.94	11.61	0.01	0.31	-2.12	2.10

'Stayers' are observations in which industry/occupation combination did not change since last observation. 'Leavers' are observations in which industry/occupation combination did change since last observation. Rows for occupation are major groupings from ISCO88. Rows for industry are major groupings from NACE1.1. Δ means change between years.

Table D4. Descriptive statistics for predicted ln household income across groups, Germany

	N	Mean	SD	Min	Max	Mean Δ	SD Δ	Min Δ	Max Δ
Leavers	29,588	10.67	0.56	4.47	12.27	0.00	0.50	-5.37	4.21
Stayers	99,262	10.64	0.55	2.77	12.27	0.00	0.28	-7.13	5.25
Occupation:									
Legislators, senior officials & managers	5,068	11.11	0.39	8.03	12.17	0.09	0.29	-2.23	2.98
Professionals	16,097	11.02	0.40	8.71	12.27	0.03	0.26	-2.07	3.60
Technicians & associate professionals	23,898	10.74	0.50	7.52	12.27	0.02	0.30	-2.90	3.05
Office Clerks	13,555	10.66	0.57	5.05	12.07	0.01	0.34	-5.93	4.21
Service, shop & market sales workers	10,432	10.46	0.65	4.22	12.21	0.00	0.40	-7.13	4.21
Skilled agricultural & fishery workers	534	10.42	0.69	6.03	11.51	0.01	0.46	-4.55	3.60
Craft & related trades workers	16,262	10.63	0.42	5.81	12.27	0.02	0.29	-3.50	3.79
Plant & machine operators & assemblers	10,075	10.59	0.44	6.69	12.17	0.02	0.30	-3.11	2.94
Elementary occupations	8,284	10.34	0.69	2.77	11.86	-0.01	0.46	-5.37	5.25
Missing	24,645	10.44	0.53	4.47	12.27	-0.08	0.38	-5.23	3.83
Industry:									
Agriculture, hunting & forestry	726	10.53	0.59	7.36	11.81	0.03	0.39	-2.48	3.60
Fishing	7	9.93	0.43	9.48	10.50	-0.06	0.10	-0.28	0.03
Mining & quarrying	474	10.72	0.50	6.80	12.27	0.04	0.34	-3.11	2.83
Manufacturing	32,990	10.76	0.46	6.69	12.27	0.03	0.28	-3.48	3.79
Electricity, Gas & water supply	1,239	10.94	0.40	9.30	11.86	0.04	0.28	-1.97	1.95
Construction	5,842	10.62	0.44	7.99	11.99	0.03	0.29	-1.89	3.05
Wholesale & retail trade; repairs	11,689	10.52	0.64	4.22	12.08	0.00	0.40	-7.13	3.79
Hotels & restaurants	1,733	10.33	0.70	6.71	12.21	0.01	0.45	-3.67	2.83
Transport, storage & communication	5,326	10.65	0.51	7.29	12.05	0.03	0.32	-2.62	2.94
Financial intermediation	4,762	11.00	0.40	8.16	12.21	0.03	0.25	-1.47	2.98
Real estate, renting & business activities	6,267	10.71	0.62	7.14	12.17	0.02	0.37	-3.34	3.60
Public admin. & defence; social security	10,255	10.76	0.50	4.55	12.21	0.01	0.27	-3.70	2.90
Education	7,028	10.82	0.54	6.70	12.05	0.01	0.30	-3.13	2.74
Health & Social Work	11,133	10.57	0.63	5.68	12.15	0.01	0.37	-5.37	5.24
Other community, social & personal services	3,566	10.64	0.63	6.41	12.12	0.00	0.38	-4.15	4.68
Private households with employed persons	376	10.16	0.98	2.77	11.68	-0.08	0.73	-5.25	5.25
Extra-territorial organisations & bodies	119	10.63	0.48	9.17	12.17	-0.01	0.34	-1.65	1.19
Missing	25,318	10.43	0.53	4.47	12.27	-0.08	0.38	-5.23	3.83

'Stayers' are observations in which industry/occupation combination did not change since last observation. 'Leavers' are observations in which industry/occupation combination did change since last observation. Rows for occupation are major groupings from ISCO88. Rows for industry are major groupings from NACE1.1. Δ means change between years.

Table D5. Descriptive statistics for predicted ln household income across groups, UK

	N	Mean	SD	Min	Max	Mean Δ	SD Δ	Min Δ	Max Δ
Leavers	24,729	10.28	0.56	4.45	11.87	0.01	0.45	-4.16	3.49
Stayers	63,441	10.27	0.59	3.40	11.87	0.01	0.30	-3.87	3.63
Occupation:									
Legislators, senior officials & managers	12,512	10.64	0.37	8.95	11.87	0.06	0.29	-2.27	3.19
Professionals	11,985	10.61	0.39	8.78	11.72	0.03	0.28	-1.44	3.14
Technicians & associate professionals	12,273	10.41	0.46	8.57	11.87	0.02	0.31	-1.72	2.82
Office Clerks	12,456	10.17	0.58	5.24	11.68	-0.01	0.38	-3.81	3.49
Service, shop & market sales workers	12,442	10.00	0.68	4.45	11.65	0.01	0.42	-3.26	3.35
Skilled agricultural & fishery workers	355	9.94	0.49	8.68	11.24	0.00	0.37	-1.70	1.22
Craft & related trades workers	5,116	10.31	0.39	8.09	11.56	0.02	0.28	-2.34	2.20
Plant & machine operators & assemblers	5,075	10.19	0.44	6.92	11.59	0.02	0.30	-2.08	3.33
Elementary occupations	5,999	9.95	0.68	3.40	11.56	0.00	0.40	-4.16	3.33
Missing	9,957	9.95	0.59	4.82	11.87	-0.09	0.43	-4.16	3.63
Industry:									
Agriculture, hunting & forestry	162	10.22	0.47	9.25	11.22	0.03	0.33	-1.45	1.09
Fishing	521	10.29	0.59	7.30	11.56	0.05	0.37	-1.67	2.36
Mining & quarrying	11,365	10.38	0.44	6.75	11.83	0.03	0.28	-2.34	2.12
Manufacturing	852	10.40	0.48	7.96	11.49	0.03	0.29	-1.54	1.94
Electricity, Gas & water supply	2,566	10.43	0.43	8.77	11.52	0.03	0.28	-1.34	2.90
Construction	8,585	10.10	0.65	6.50	11.65	0.01	0.39	-3.62	3.49
Wholesale & retail trade; repairs	1,822	9.88	0.77	6.16	11.63	0.00	0.45	-2.77	2.52
Hotels & restaurants	6,077	10.35	0.53	7.49	11.63	0.02	0.32	-2.44	2.67
Transport, storage & communication	5,963	10.44	0.57	3.40	11.80	0.03	0.34	-4.16	2.95
Financial intermediation	5,334	10.47	0.50	7.81	11.63	0.04	0.33	-2.11	3.35
Real estate, renting & business activities	6,517	10.46	0.47	8.74	11.87	0.02	0.30	-1.79	3.13
Public admin. & defence; social security	9,140	10.39	0.57	6.67	11.87	0.01	0.35	-2.16	3.14
Education	12,082	10.27	0.59	6.95	11.80	0.02	0.35	-2.15	3.19
Health & Social Work	5,491	10.21	0.59	5.24	11.61	0.02	0.36	-3.81	2.55
Other community, social & personal services	954	10.13	0.65	4.45	11.61	0.03	0.38	-3.26	1.99
Private households with employed persons	1,539	10.17	0.64	7.02	11.63	0.04	0.42	-2.46	3.33
Extra-territorial organisations & bodies	9,198	9.90	0.57	4.82	11.87	-0.10	0.44	-4.16	3.63
Missing	24,729	10.28	0.56	4.45	11.87	0.01	0.45	-4.16	3.49

'Stayers' are observations in which industry/occupation combination did not change since last observation. 'Leavers' are observations in which industry/occupation combination did change since last observation. Rows for occupation are major groupings from ISCO88. Rows for industry are major groupings from NACE1.1. Δ means change between years.

Table D6. Descriptive statistics on reference groups and overlap with industry/occupation and reference group

	Mean	1st pct.	25th pct.	Median	75th pct.	99th pct.	SD	Min	Max
Germany									
Size of reference group	701.8	233.0	394.0	619.0	903.0	1,264.0	429.9	1.0	2,246.0
Same indust. & occup.	0.064	0.001	0.003	0.008	0.038	0.240	0.127	0.000	1.000
Same occupation	0.103	0.010	0.022	0.055	0.129	0.280	0.127	0.000	1.000
Same industry	0.081	0.006	0.014	0.032	0.074	0.244	0.124	0.000	1.000
UK									
Size of reference group	847.7	185.0	325.0	580.0	1261.0	1,986.0	657.5	7.0	2,524.0
Same indust. & occup.	0.034	0.001	0.002	0.006	0.026	0.084	0.079	0.000	0.802
Same occupation	0.079	0.007	0.016	0.051	0.108	0.200	0.091	0.000	0.802
Same industry	0.058	0.007	0.015	0.035	0.069	0.118	0.078	0.000	0.802

Note: "pct." refers to percentile.

D Robustness tables

Table R1. Results for West Germany using equivalized HH income

	(1) Eq.HH income, contemp. controls	(2) Eq.HH income, lagged controls	(3) Instr. Eq.HH income, contemp. controls	(4) Instr. Eq.HH income, lagged controls	(5) Instr. Eq.HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.266***	0.247***	0.682***	0.699***	0.409**
	(0.025)	(0.025)	(0.199)	(0.197)	(0.183)
LR effect of own inc.	0.275***	0.284***	0.396***	0.484***	0.411***
	(0.038)	(0.039)	(0.108)	(0.128)	(0.129)
Adaptation to own inc.	0.009	0.038	-0.286*	-0.215	0.002
	(0.033)	(0.034)	(0.170)	(0.182)	(0.171)
SR effect of ref. inc.	-0.133	-0.058	-0.159	-0.129	-0.096
	(0.158)	(0.166)	(0.160)	(0.170)	(0.169)
LR effect of ref. inc.	-0.358**	-0.346**	-0.382***	-0.369**	-0.349**
	(0.140)	(0.151)	(0.148)	(0.158)	(0.158)
Adaptation to ref. inc.	-0.225+	-0.288+	-0.222+	-0.240	-0.253+
	(0.150)	(0.185)	(0.153)	(0.190)	(0.190)
Combined SR effect	0.133	0.188	0.523**	0.570**	0.312+
	(0.159)	(0.166)	(0.239)	(0.233)	(0.227)
Combined LR effect	-0.083	-0.062	0.015	0.115	0.061
	(0.141)	(0.153)	(0.151)	(0.171)	(0.175)
Combined Adaptation	-0.215+	-0.251+	-0.508**	-0.455*	-0.251
	(0.152)	(0.187)	(0.216)	(0.238)	(0.233)
Kleibergen-Paap F-stat.			47.028	42.531	52.232
Serial Correlation	0.027	0.023	0.022	0.018	0.020
Average T	7.456	7.456	7.456	7.456	7.456
# of Persons	17,282	17,282	17,282	17,282	17,282
# of Observations	128,850	128,850	128,850	128,850	128,850

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R2. Results for UK using equivalized HH income

	(1) HH income, contemp. controls	(2) HH income, lagged controls	(3) Instr. Eq.HH income, contemp. controls	(4) Instr. Eq.HH income, lagged controls	(5) Instr. Eq.HH income, lagged controls, indust. & occup.	(6) Instr. Eq.HH income, lagged controls, indust. & occup., (W.Ger., 1996-2015)
SR effect of own inc.	0.095***	0.088***	0.332***	0.279***	0.295***	0.269
	(0.019)	(0.019)	(0.093)	(0.096)	(0.103)	(0.237)
LR effect of own inc.	0.118***	0.129***	0.255***	0.329***	0.383***	0.368**
	(0.031)	(0.034)	(0.087)	(0.109)	(0.117)	(0.168)
Adaptation to own inc.	0.023	0.041+	-0.077	0.050	0.088	0.099
	(0.028)	(0.029)	(0.084)	(0.105)	(0.110)	(0.219)
SR effect of ref. inc.	0.110	0.139	0.061	0.107	0.116	-0.105
	(0.135)	(0.138)	(0.136)	(0.141)	(0.142)	(0.223)
LR effect of ref. inc.	-0.135	-0.195	-0.176	-0.219	-0.236+	-0.144
	(0.155)	(0.176)	(0.158)	(0.177)	(0.179)	(0.190)
Adaptation to ref. inc.	-0.245*	-0.334*	-0.236*	-0.326*	-0.352**	-0.039
	(0.129)	(0.174)	(0.133)	(0.177)	(0.179)	(0.242)
Combined SR effect	0.205+	0.227+	0.392**	0.386**	0.411**	0.164
	(0.135)	(0.139)	(0.153)	(0.157)	(0.160)	(0.294)
Combined LR effect	-0.017	-0.066	0.079	0.110	0.147	0.225
	(0.155)	(0.177)	(0.164)	(0.198)	(0.202)	(0.203)
Combined Adaptation	-0.222*	-0.293*	-0.313**	-0.276+	-0.264+	0.061
	(0.131)	(0.175)	(0.141)	(0.189)	(0.192)	(0.296)
Kleibergen-Paap F-stat.			152.017	120.690	108.802	32.746
Serial Correlation	-0.174	-0.176	-0.175	-0.177	-0.182	-0.030
Average T	4.669	4.669	4.669	4.669	4.669	6.396
# of Persons	18,884	18,884	18,884	18,884	18,884	14,348
# of Observations	88,170	88,170	88,170	88,170	88,170	91,776

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R3. Main results, using non-eq. reference income

	(1) W.Germany, instr. HH income, lagged controls	(2) W.Germany, instr. HH income, lagged controls, indust. & occup.	(3) UK, instr. HH income, lagged controls	(4) UK, instr. HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.714***	0.424**	0.285***	0.302***
	(0.199)	(0.185)	(0.095)	(0.102)
LR effect of own inc.	0.473***	0.401***	0.328***	0.383***
	(0.130)	(0.131)	(0.107)	(0.115)
Adaptation to own inc.	-0.242+	-0.023	0.043	0.081
	(0.183)	(0.171)	(0.103)	(0.108)
SR effect of ref. inc.	0.202	0.251+	0.123	0.145
	(0.169)	(0.168)	(0.131)	(0.132)
LR effect of ref. inc.	-0.138	-0.123	-0.246+	-0.268*
	(0.113)	(0.112)	(0.161)	(0.162)
Adaptation to ref. inc.	-0.340**	-0.374**	-0.369**	-0.414**
	(0.168)	(0.168)	(0.163)	(0.164)
Combined SR effect	0.916***	0.674***	0.408***	0.448***
	(0.229)	(0.222)	(0.147)	(0.150)
Combined LR effect	0.335**	0.278*	0.082	0.115
	(0.143)	(0.147)	(0.190)	(0.194)
Combined adaptation	-0.582***	-0.397*	-0.326*	-0.333*
	(0.224)	(0.218)	(0.184)	(0.187)
Kleibergen-Paap F-stat.	42.042	51.956	120.672	109.162
Serial Correlation	0.018	0.019	-0.177	-0.182
Average T	7.456	7.456	4.669	4.669
# of Persons	17,282	17,282	17,282	17,282
# of Observations	128,850	128,850	128,850	128,850

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R4. Results for varying lag lengths, West Germany, industry and occupation dummies included

	(1) Instr. HH income, indust. & occup, K=0	(2) Instr. HH income, indust. & occup, K=1	(3) Instr. HH income, indust. & occup, K=2	(4) Instr. HH income, indust. & occup, K=3	(5) Instr. HH income, indust. & occup, K=4	(6) Instr. HH income, indust. & occup, K=5	(7) Instr. HH income, indust. & occup, K=6
SR effect of own inc.	0.434***	0.384**	0.481***	0.435**	0.386*	0.532**	0.412+
	(0.091)	(0.150)	(0.168)	(0.185)	(0.206)	(0.224)	(0.257)
LR effect of own inc.	0.434***	0.501***	0.523***	0.408***	0.456***	0.662***	0.679***
	(0.091)	(0.094)	(0.110)	(0.131)	(0.156)	(0.187)	(0.217)
Adaptation to own inc.		0.116	0.042	-0.028	0.071	0.130	0.267
		(0.135)	(0.151)	(0.171)	(0.199)	(0.224)	(0.263)
SR effect of ref. inc.	-0.328***	-0.139	-0.049	-0.095	-0.169	-0.014	-0.029
	(0.115)	(0.144)	(0.156)	(0.169)	(0.181)	(0.193)	(0.204)
LR effect of ref. inc.	-0.328***	-0.319**	-0.366***	-0.342**	-0.541***	-0.674***	-0.698***
	(0.115)	(0.124)	(0.140)	(0.158)	(0.179)	(0.201)	(0.229)
Adaptation to ref. inc.		-0.180	-0.317*	-0.248+	-0.371*	-0.660***	-0.669***
		(0.144)	(0.165)	(0.190)	(0.211)	(0.234)	(0.259)
Combined SR effect	0.106	0.246+	0.432**	0.341+	0.216	0.518*	0.383
	(0.127)	(0.189)	(0.208)	(0.228)	(0.252)	(0.272)	(0.307)
Combined LR effect	0.106	0.182+	0.157	0.065	-0.084	-0.012	-0.019
	(0.127)	(0.136)	(0.156)	(0.177)	(0.203)	(0.237)	(0.273)
Combined Adaptation		-0.064	-0.275+	-0.275	-0.301	-0.530*	-0.402
		(0.175)	(0.203)	(0.233)	(0.265)	(0.294)	(0.336)
Kleibergen-Paap F-stat.	1,707.390	349.375	104.163	51.873	31.651	19.694	12.324
Serial Correlation	0.030	0.028	0.022	0.019	0.015	0.011	0.008
Average T	7.264	7.264	7.290	7.456	7.831	7.886	7.690
# of Persons	25,236	25,236	21,202	17,282	13,741	11,607	10,244
# of Observations	183,311	183,311	154,573	128,850	107,603	91,527	78,774

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R5. Results for varying lag lengths, UK, industry and occupation dummies included

	(1) Instr. HH income, indust. & occup, K=0	(2) Instr. HH income, indust. & occup, K=1	(3) Instr. HH income, indust. & occup, K=2	(4) Instr. HH income, indust. & occup, K=3	(5) Instr. HH income, indust. & occup, K=4	(6) Instr. HH income, indust. & occup, K=5	(7) Instr. HH income, indust. & occup, K=6
SR effect of own inc.	0.241***	0.244***	0.261***	0.302***	0.325***	0.471***	0.315**
	(0.061)	(0.072)	(0.084)	(0.102)	(0.108)	(0.125)	(0.152)
LR effect of own inc.	0.241***	0.212***	0.272***	0.382***	0.329**	0.243+	0.335+
	(0.061)	(0.070)	(0.092)	(0.115)	(0.142)	(0.177)	(0.221)
Adaptation to own inc.		-0.032	0.011	0.080	0.003	-0.228+	0.020
		(0.067)	(0.090)	(0.108)	(0.130)	(0.157)	(0.204)
SR effect of ref. inc.	-0.071	0.024	-0.019	0.115	0.029	-0.139	-0.369*
	(0.104)	(0.112)	(0.128)	(0.142)	(0.156)	(0.171)	(0.205)
LR effect of ref. inc.	-0.071	-0.132	-0.193+	-0.234+	-0.100	-0.187	-0.114
	(0.104)	(0.123)	(0.150)	(0.179)	(0.210)	(0.244)	(0.307)
Adaptation to ref. inc.		-0.156+	-0.174	-0.349*	-0.129	-0.048	0.255
		(0.110)	(0.148)	(0.179)	(0.211)	(0.244)	(0.295)
Combined SR effect	0.170+	0.268**	0.242*	0.417***	0.355**	0.331*	-0.053
	(0.110)	(0.122)	(0.142)	(0.160)	(0.172)	(0.189)	(0.225)
Combined LR effect	0.170+	0.081	0.079	0.148	0.229	0.056	0.221
	(0.110)	(0.130)	(0.165)	(0.201)	(0.238)	(0.293)	(0.379)
Combined Adaptation		-0.188+	-0.163	-0.269+	-0.126	-0.276	0.275
		(0.115)	(0.157)	(0.192)	(0.225)	(0.270)	(0.344)
Kleibergen-Paap F-stat.	2,109.554	577.368	222.517	109.005	55.381	35.320	20.168
Serial Correlation	-0.136	-0.138	-0.161	-0.182	-0.207	-0.243	-0.285
Average T	5.150	5.150	4.887	4.669	4.423	4.100	3.795
# of Persons	29,778	29,778	23,779	18,884	15,000	12,163	8,950
# of Observations	153,353	153,353	116,211	88,170	66,352	49,867	33,967

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R6. Results for varying lag lengths, West Germany

	(1) Instr. HH income, indust. & occup, K=0	(2) Instr. HH income, indust. & occup, K=1	(3) Instr. HH income, indust. & occup, K=2	(4) Instr. HH income, indust. & occup, K=3	(5) Instr. HH income, indust. & occup, K=4	(6) Instr. HH income, indust. & occup, K=5	(7) Instr. HH income, indust. & occup, K=6
SR effect of own inc.	0.553***	0.642***	0.764***	0.727***	0.699***	0.830***	0.708**
	(0.091)	(0.158)	(0.179)	(0.199)	(0.224)	(0.246)	(0.283)
LR effect of own inc.	0.553***	0.582***	0.605***	0.481***	0.502***	0.733***	0.735***
	(0.091)	(0.093)	(0.110)	(0.130)	(0.154)	(0.182)	(0.210)
Adaptation to own inc.		-0.061	-0.160	-0.246+	-0.197	-0.097	0.027
		(0.144)	(0.160)	(0.183)	(0.214)	(0.241)	(0.286)
SR effect of ref. inc.	-0.333***	-0.155	-0.061	-0.126	-0.186	-0.018	-0.047
	(0.116)	(0.145)	(0.157)	(0.170)	(0.182)	(0.194)	(0.205)
LR effect of ref. inc.	-0.333***	-0.307**	-0.364***	-0.361**	-0.538***	-0.671***	-0.708***
	(0.116)	(0.124)	(0.140)	(0.157)	(0.178)	(0.198)	(0.224)
Adaptation to ref. inc.		-0.152	-0.303*	-0.235	-0.352*	-0.653***	-0.660***
		(0.144)	(0.165)	(0.190)	(0.211)	(0.232)	(0.255)
Combined SR effect	0.220*	0.487**	0.703***	0.601**	0.513**	0.812***	0.660**
	(0.126)	(0.193)	(0.212)	(0.234)	(0.262)	(0.285)	(0.323)
Combined LR effect	0.220*	0.274**	0.241+	0.120	-0.036	0.063	0.027
	(0.126)	(0.134)	(0.153)	(0.172)	(0.195)	(0.226)	(0.256)
Combined Adaptation		-0.213	-0.462**	-0.480**	-0.549**	-0.750**	-0.634*
		(0.179)	(0.208)	(0.238)	(0.272)	(0.301)	(0.347)
Kleibergen-Paap F-stat.	1,832.014	325.406	89.602	41.943	24.675	14.839	9.245
Serial Correlation	0.032	0.028	0.021	0.018	0.010	0.007	0.008
Average T	7.264	7.264	7.290	7.456	7.831	7.886	7.690
# of Persons	25,236	25,236	21,202	17,282	13,741	11,607	10,244
# of Observations	183,311	183,311	154,573	128,850	107,603	91,527	78,774

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R7. Results for varying lag lengths, UK

	(1) Instr. HH income, indust. & occup, K=0	(2) Instr. HH income, indust. & occup, K=1	(3) Instr. HH income, indust. & occup, K=2	(4) Instr. HH income, indust. & occup, K=3	(5) Instr. HH income, indust. & occup, K=4	(6) Instr. HH income, indust. & occup, K=5	(7) Instr. HH income, indust. & occup, K=6
SR effect of own inc.	0.232***	0.244***	0.230***	0.285***	0.292***	0.432***	0.318**
	(0.058)	(0.068)	(0.079)	(0.095)	(0.102)	(0.120)	(0.145)
LR effect of own inc.	0.232***	0.195***	0.206**	0.329***	0.265**	0.212	0.166
	(0.058)	(0.066)	(0.087)	(0.107)	(0.133)	(0.167)	(0.210)
Adaptation to own inc.		-0.049	-0.024	0.044	-0.027	-0.220+	-0.151
		(0.064)	(0.086)	(0.103)	(0.124)	(0.152)	(0.199)
SR effect of ref. inc.	-0.068	0.019	-0.012	0.106	0.028	-0.119	-0.356*
	(0.104)	(0.112)	(0.127)	(0.141)	(0.153)	(0.167)	(0.198)
LR effect of ref. inc.	-0.068	-0.132	-0.191	-0.218	-0.086	-0.191	-0.122
	(0.104)	(0.122)	(0.149)	(0.177)	(0.206)	(0.239)	(0.294)
Adaptation to ref. inc.		-0.151+	-0.179	-0.324*	-0.115	-0.072	0.234
		(0.110)	(0.147)	(0.177)	(0.207)	(0.238)	(0.283)
Combined SR effect	0.164+	0.263**	0.218+	0.391**	0.321*	0.313*	-0.038
	(0.110)	(0.121)	(0.140)	(0.157)	(0.168)	(0.185)	(0.216)
Combined LR effect	0.164+	0.063	0.015	0.111	0.179	0.021	0.045
	(0.110)	(0.128)	(0.162)	(0.198)	(0.236)	(0.290)	(0.369)
Combined Adaptation		-0.200*	-0.203+	-0.280+	-0.142	-0.293	0.082
		(0.115)	(0.155)	(0.189)	(0.221)	(0.265)	(0.334)
Kleibergen-Paap F-stat.	2,345.833	649.550	252.539	120.649	60.882	38.042	21.296
Serial Correlation	-0.135	-0.136	-0.158	-0.177	-0.198	-0.230	-0.269
Average T	5.150	5.150	4.887	4.669	4.423	4.100	3.795
# of Persons	29,778	29,778	23,779	18,884	15,000	12,163	8,950
# of Observations	153,353	153,353	116,211	88,170	66,352	49,867	33,967

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R8. Check anticipation

	(1) W.Germany, Instr. HH income, lagged controls	(2) W.Germany, Instr. HH income, lagged controls, indust. & occup.	(3) UK, Instr. HH income, lagged controls	(4) UK, Instr. HH income, lagged controls, indust. & occup.
Anticipation of own inc.	0.212	0.206	0.059	0.005
	(0.213)	(0.196)	(0.110)	(0.116)
Contemp. effect of own inc.	0.494*	0.249	0.338***	0.362***
	(0.287)	(0.253)	(0.128)	(0.136)
SR effect of own inc.	0.706***	0.455**	0.397***	0.367***
	(0.207)	(0.197)	(0.132)	(0.141)
LR effect of own inc.	0.471***	0.397***	0.417***	0.471***
	(0.153)	(0.154)	(0.143)	(0.156)
Adaptation to own inc.	-0.235	-0.058	0.020	0.105
	(0.192)	(0.180)	(0.123)	(0.130)
Anticipation of ref inc.	0.325*	0.285+	0.153	0.171
	(0.176)	(0.176)	(0.162)	(0.165)
Contemp. effect of own inc.	-0.307+	-0.264+	-0.088	-0.092
	(0.204)	(0.203)	(0.172)	(0.176)
SR effect of ref. inc.	0.018	0.021	0.065	0.079
	(0.196)	(0.196)	(0.198)	(0.205)
LR effect of ref. inc.	-0.363**	-0.362**	-0.156	-0.185
	(0.179)	(0.180)	(0.219)	(0.224)
Adaptation to ref. inc.	-0.381*	-0.382*	-0.220	-0.264
	(0.208)	(0.209)	(0.202)	(0.207)
Combined anticipation	0.537**	0.491**	0.212	0.176
	(0.250)	(0.241)	(0.179)	(0.184)
Combined contemp. effect	0.187	-0.015	0.250+	0.270+
	(0.306)	(0.286)	(0.189)	(0.195)
Combined SR effect	0.724***	0.476*	0.461**	0.446*
	(0.259)	(0.257)	(0.223)	(0.230)
Combined LR effect	0.108	0.035	0.261	0.286
	(0.195)	(0.202)	(0.251)	(0.258)
Combined adaptation	-0.616**	-0.441*	-0.201	-0.159
	(0.256)	(0.251)	(0.215)	(0.221)
Kleibergen-Paap F-stat.	24.768	32.222	60.005	54.699
Serial Correlation	0.018	0.017	-0.216	-0.223
Average T	7.653	7.653	4.285	4.285
# of Persons	14,227	14,227	14,931	14,931
# of Observations	108,882	108,882	63,975	63,975

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R9. Main results, exclude industry/occupation switches

	(1) W.Germany, instr. HH income, lagged controls	(2) W.Germany, instr. HH income, lagged controls, indust. & occup.	(3) UK, instr. HH income, lagged controls	(4) UK, instr. HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.381*	0.348+	0.271**	0.332**
	(0.217)	(0.219)	(0.133)	(0.136)
LR effect of own inc.	0.528***	0.503***	0.325**	0.455***
	(0.144)	(0.145)	(0.143)	(0.153)
Adaptation to own inc.	0.147	0.155	0.053	0.123
	(0.201)	(0.201)	(0.139)	(0.143)
SR effect of ref. inc.	0.002	-0.010	0.109	0.120
	(0.190)	(0.190)	(0.189)	(0.191)
LR effect of ref. inc.	-0.191	-0.174	-0.269	-0.236
	(0.170)	(0.171)	(0.238)	(0.240)
Adaptation to ref. inc.	-0.193	-0.164	-0.379+	-0.356+
	(0.213)	(0.214)	(0.240)	(0.241)
Combined SR effect	0.383+	0.338+	0.381*	0.452**
	(0.260)	(0.263)	(0.218)	(0.219)
Combined LR effect	0.337*	0.329*	0.055	0.219
	(0.193)	(0.199)	(0.268)	(0.271)
Combined adaptation	-0.046	-0.009	-0.325	-0.233
	(0.263)	(0.265)	(0.255)	(0.257)
Kleibergen-Paap F-stat.	42.583	48.852	67.879	63.790
Serial Correlation	-0.001	-0.003	-0.263	-0.269
Average T	6.489	6.489	3.792	3.792
# of Persons	15,006	15,006	16,222	16,222
# of Observations	97,381	97,381	61,508	61,508

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R10. Main results, excluding respondent observations with second job

	(1) W.Germany, instr. HH income, lagged controls	(2) W.Germany, instr. HH income, lagged controls, indust. & occup.	(3) UK, instr. HH income, lagged controls	(4) UK, instr. HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.699***	0.418**	0.258**	0.284***
	(0.209)	(0.194)	(0.100)	(0.108)
LR effect of own inc.	0.444***	0.382***	0.363***	0.406***
	(0.133)	(0.135)	(0.115)	(0.123)
Adaptation to own inc.	-0.255+	-0.036	0.105	0.121
	(0.194)	(0.181)	(0.108)	(0.114)
SR effect of ref. inc.	-0.161	-0.134	0.061	0.058
	(0.175)	(0.175)	(0.147)	(0.149)
LR effect of ref. inc.	-0.343**	-0.332**	-0.184	-0.198
	(0.163)	(0.164)	(0.183)	(0.186)
Adaptation to ref. inc.	-0.182	-0.198	-0.244+	-0.256+
	(0.197)	(0.197)	(0.184)	(0.187)
Combined SR effect	0.538**	0.285	0.319*	0.342**
	(0.244)	(0.236)	(0.165)	(0.169)
Combined LR effect	0.101	0.050	0.180	0.207
	(0.179)	(0.185)	(0.207)	(0.212)
Combined adaptation	-0.437*	-0.235	-0.140	-0.135
	(0.249)	(0.242)	(0.197)	(0.200)
Kleibergen-Paap F-stat.	36.159	45.745	113.373	101.357
Serial Correlation	0.013	0.015	-0.191	-0.197
Average T	7.258	7.258	4.525	4.525
# of Persons	16612	16612	18047	18047
# of Observations	120576	120576	81671	81671

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R11. Main results, exclude observations with industries/occupations with ≤ 50 observations

	(1) W.Germany, instr. HH income, lagged controls	(2) W.Germany, instr. HH income, lagged controls, indust. & occup.	(3) UK, instr. HH income, lagged controls	(4) UK, instr. HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.653***	0.396*	0.262**	0.284**
	(0.237)	(0.221)	(0.104)	(0.111)
LR effect of own inc.	0.374**	0.287*	0.309**	0.400***
	(0.147)	(0.148)	(0.123)	(0.132)
Adaptation to own inc.	-0.279	-0.109	0.047	0.116
	(0.219)	(0.204)	(0.119)	(0.125)
SR effect of ref. inc.	-0.087	-0.089	0.175	0.200
	(0.193)	(0.192)	(0.160)	(0.162)
LR effect of ref. inc.	-0.282+	-0.318*	-0.240	-0.268+
	(0.182)	(0.182)	(0.202)	(0.205)
Adaptation to ref. inc.	-0.194	-0.229	-0.414**	-0.468**
	(0.217)	(0.218)	(0.204)	(0.207)
Combined SR effect	0.565**	0.307	0.436**	0.483***
	(0.271)	(0.263)	(0.177)	(0.180)
Combined LR effect	0.092	-0.031	0.070	0.132
	(0.197)	(0.202)	(0.226)	(0.230)
Combined adaptation	-0.473*	-0.338	-0.367*	-0.352+
	(0.274)	(0.268)	(0.217)	(0.222)
Kleibergen-Paap F-stat.	32.282	40.773	101.808	91.746
Serial Correlation	-0.003	-0.003	-0.208	-0.212
Average T	6.722	6.722	4.322	4.322
# of Persons	15462	15462	17302	17302
# of Observations	103940	103940	74787	74787

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R12. Main results, sample restricted to those with own labour income

	(1) W.Germany, instr. HH income, lagged controls	(2) W.Germany, instr. HH income, lagged controls, indust. & occup.	(3) UK, instr. HH income, lagged controls	(4) UK, instr. HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.569**	0.288+	0.244**	0.220**
	(0.224)	(0.203)	(0.108)	(0.109)
LR effect of own inc.	0.391***	0.300**	0.273**	0.273**
	(0.147)	(0.146)	(0.111)	(0.118)
Adaptation to own inc.	-0.178	0.011	0.029	0.053
	(0.200)	(0.184)	(0.110)	(0.112)
SR effect of ref. inc.	-0.216	-0.182	0.022	0.040
	(0.190)	(0.189)	(0.149)	(0.150)
LR effect of ref. inc.	-0.379**	-0.344**	-0.239	-0.280+
	(0.174)	(0.173)	(0.188)	(0.190)
Adaptation to ref. inc.	-0.163	-0.163	-0.261+	-0.320*
	(0.210)	(0.210)	(0.187)	(0.188)
Combined SR effect	0.352+	0.107	0.266+	0.260+
	(0.259)	(0.249)	(0.169)	(0.170)
Combined LR effect	0.012	-0.044	0.034	-0.007
	(0.187)	(0.192)	(0.208)	(0.212)
Combined adaptation	-0.341+	-0.151	-0.233	-0.267+
	(0.259)	(0.252)	(0.200)	(0.201)
Kleibergen-Paap F-stat.	35.289	46.075	113.270	103.420
Serial Correlation	0.005	0.007	-0.180	-0.186
Average T	7.141	7.141	4.631	4.631
# of Persons	15372	15372	17128	17128
# of Observations	109776	109776	79313	79313

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R13. Results for West Germany, reference income dropped

	(1) HH income, contemp. controls	(2) HH income, lagged controls	(3) Instr. HH income, contemp. controls	(4) Instr. HH income, lagged controls	(5) Instr. HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.302***	0.249***	0.755***	0.708***	0.421**
	(0.025)	(0.025)	(0.148)	(0.197)	(0.184)
LR effect of own inc.	0.169***	0.273***	0.346***	0.455***	0.386***
	(0.034)	(0.039)	(0.104)	(0.127)	(0.128)
Adaptation to own inc.	-0.133***	0.024	-0.409***	-0.253+	-0.035
	(0.027)	(0.034)	(0.088)	(0.182)	(0.171)
Kleibergen-Paap F-stat.			155.974	42.139	52.098
Serial Correlation	0.027	0.023	0.022	0.018	0.020
Average T	7.456	7.456	7.456	7.456	7.456
# of Persons	17,282	17,282	17,282	17,282	17,282
# of Observations	128,850	128,850	128,850	128,850	128,850

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R14. Results for UK, reference income dropped

	(1) HH income, contemp. controls	(2) HH income, lagged controls	(3) Instr. HH income, contemp. controls	(4) Instr. HH income, lagged controls	(5) Instr. HH income, lagged controls, indust. & occup.	(6) Instr. HH income, lagged controls, indust. & occup., (W.Germany, 1996-2015)
SR effect of own inc.	0.103***	0.090***	0.350***	0.285***	0.302***	0.270
	(0.019)	(0.019)	(0.085)	(0.095)	(0.101)	(0.238)
LR effect of own inc.	0.045+	0.128***	0.165**	0.324***	0.376***	0.358**
	(0.029)	(0.033)	(0.081)	(0.107)	(0.115)	(0.166)
Adaptation to own inc.	-0.058**	0.038+	-0.185***	0.039	0.074	0.089
	(0.024)	(0.029)	(0.056)	(0.103)	(0.108)	(0.219)
Kleibergen-Paap F-stat.			294.770	121.078	109.483	32.320
Serial Correlation	-0.174	-0.176	-0.175	-0.177	-0.182	-0.030
Average T	4.669	4.669	4.669	4.669	4.669	6.396
# of Persons	18884	18884	18884	18884	18884	14348
# of Observations	88170	88170	88170	88170	88170	91776

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R15. Regressions using observed income and controlling for industry and occupation dummies

	(1) W.Germany, HH income, indust. & occup.	(2) W.Germany, HH income, lagged controls, indust. & occup.	(3) UK, HH income, indust. & occup.	(4) UK, HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.298***	0.245***	0.102***	0.089***
	(0.025)	(0.025)	(0.019)	(0.019)
LR effect of own inc.	0.168***	0.267***	0.047+	0.128***
	(0.034)	(0.039)	(0.029)	(0.034)
Adaptation to own inc.	-0.129***	0.022	-0.055**	0.038+
	(0.027)	(0.034)	(0.024)	(0.030)
SR effect of ref. inc.	-0.118	-0.066	0.112	0.149
	(0.158)	(0.166)	(0.135)	(0.140)
LR effect of ref. inc.	-0.313**	-0.336**	-0.098	-0.200
	(0.140)	(0.152)	(0.155)	(0.178)
Adaptation to ref. inc.	-0.195+	-0.271+	-0.211+	-0.349**
	(0.150)	(0.186)	(0.129)	(0.175)
Combined SR effect	0.179	0.179	0.214+	0.239*
	(0.159)	(0.167)	(0.136)	(0.140)
Combined LR effect	-0.145	-0.069	-0.052	-0.072
	(0.141)	(0.155)	(0.155)	(0.179)
Combined adaptation	-0.325**	-0.248+	-0.266**	-0.311*
	(0.152)	(0.188)	(0.131)	(0.177)
Kleibergen-Paap F-stat.	0.025	0.021	-0.176	-0.181
Serial Correlation	7.456	7.456	4.669	4.669
Average T	17282	17282	18884	18884
# of Persons	128850	128850	88170	88170
# of Observations	109776	109776	79313	79313

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R16. Main results, using interactions of industry and occupations

	(1) W.Germany, instr. HH income, lagged controls	(2) W.Germany, instr. HH income, lagged controls, indust. & occup.	(3) UK, instr. HH income, lagged controls	(4) UK, instr. HH income, lagged controls, indust. & occup.
SR effect of own inc.	0.693***	0.539***	0.254***	0.269***
	(0.132)	(0.127)	(0.078)	(0.083)
LR effect of own inc.	0.464***	0.421***	0.185**	0.210**
	(0.105)	(0.106)	(0.090)	(0.096)
Adaptation to own inc.	-0.228*	-0.117	-0.069	-0.059
	(0.128)	(0.125)	(0.088)	(0.091)
SR effect of ref. inc.	-0.138	-0.124	0.096	0.104
	(0.168)	(0.168)	(0.140)	(0.142)
LR effect of ref. inc.	-0.397**	-0.382**	-0.197	-0.204
	(0.154)	(0.155)	(0.176)	(0.179)
Adaptation to ref. inc.	-0.259+	-0.258+	-0.293*	-0.308*
	(0.188)	(0.188)	(0.176)	(0.177)
Combined SR effect	0.554***	0.415**	0.350**	0.373**
	(0.198)	(0.197)	(0.151)	(0.153)
Combined LR effect	0.067	0.039	-0.012	0.007
	(0.165)	(0.169)	(0.192)	(0.195)
Combined adaptation	-0.488**	-0.376*	-0.362**	-0.367*
	(0.212)	(0.212)	(0.185)	(0.187)
Kleibergen-Paap F-stat.	136.948	151.517	167.525	158.156
Serial Correlation	0.021	0.020	-0.177	-0.182
Average T	7.455	7.455	4.669	4.669
# of Persons	17,264	17,264	18,879	18,879
# of Observations	128,707	128,707	88,152	88,152

Note: Estimated with Correia's 'reghdfe' command. Individual and year fixed effects & (lagged) controls included. Clustered standard errors in parentheses. + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R17. Sensitivity to removal of controls, Germany

	Own income			Reference income		
Variable omitted	Short run	Long run	Adaptation	Short run	Long run	Adaptation
Number of adults	0.219**	0.088+	-0.131+	-0.150	-0.364**	-0.215
Number of children	0.408**	0.346***	-0.062	-0.067	-0.288*	-0.221
Employment status	0.565***	0.464***	-0.100	-0.117	-0.377**	-0.260+
Partner employment status	0.792***	0.460***	-0.332**	-0.134	-0.344**	-0.210
Age	0.435**	0.408***	-0.028	-0.095	-0.342**	-0.248+
Age-squared	0.398**	0.363***	-0.035	-0.209	-0.291*	-0.082
Marital status	0.422**	0.410***	-0.012	-0.101	-0.341**	-0.240
Housing tenure	0.453**	0.424***	-0.029	-0.097	-0.343**	-0.245+
Working hours	0.440**	0.406***	-0.033	-0.096	-0.339**	-0.243
Ill health	0.487**	0.334**	-0.154	-0.080	-0.400**	-0.320+
Child birth	0.364**	0.387***	0.023	-0.079	-0.333**	-0.253+
Education	0.426**	0.396***	-0.030	0.012	-0.234*	-0.247+
Wave	0.411**	0.391***	-0.021	-0.407***	-0.617***	-0.210*
Region	0.434**	0.407***	-0.027	-0.090	-0.340**	-0.250+
Regional time trend	0.435**	0.407***	-0.027	-0.080	-0.325**	-0.245+

Note: + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Table R18. Sensitivity to removal of controls, UK

	Own income			Reference income		
Variable omitted	Short run	Long run	Adaptation	Short run	Long run	Adaptation
Number of adults	0.131**	0.132*	0.001	0.110	-0.192	-0.302*
Number of children	0.289***	0.352***	0.063	0.200+	0.006	-0.194
Employment status	0.316***	0.408***	0.092	0.084	-0.282+	-0.366**
Partner employment status	0.377***	0.311***	-0.066	0.080	-0.223	-0.303*
Age	0.303***	0.383***	0.080	0.115	-0.233+	-0.349*
Age-squared	0.285***	0.346***	0.060	0.091	-0.316*	-0.407**
Marital status	0.257**	0.353***	0.096	0.131	-0.238+	-0.369**
Housing tenure	0.309***	0.390***	0.082	0.114	-0.243+	-0.357**
Working hours	0.298***	0.374***	0.076	0.109	-0.255+	-0.364**
Ill health	0.364***	0.428***	0.064	0.128	-0.272+	-0.399**
Child birth	0.299***	0.381***	0.082	0.116	-0.240+	-0.356**
Education	0.301***	0.380***	0.079	0.072	-0.165+	-0.236*
Wave	0.306***	0.343***	0.037	0.146	-0.642***	-0.788***
Region	0.305***	0.380***	0.075	0.122	-0.232+	-0.355**
Regional time trend	0.300***	0.378***	0.078	0.096	-0.228	-0.325*

Note: + p<0.2, * p<0.10, ** p<0.050, *** p<0.010.

Positional, mobility and reference effects: How does social class affect life satisfaction in Europe?†

Caspar Kaiser and Nhat An Trinh*

Abstract

In this study, we analyse the effects of social class on life satisfaction. We develop a theoretical framework that shows how social class affects life satisfaction through five different pathways. Informed by this framework, we estimate the direct effects of class destination and class origin, the effect of own intergenerational class mobility as well as the effects of others' class position and mobility. To do so, we utilize European Social Survey (ESS) waves 1 to 5 (2002-2010) and obtain information on life satisfaction as well as destination and origin class for about 80,000 respondents in 32 European countries. We find (1) class destination consistently and strongly structures life satisfaction across Europe, (2) own class mobility has a significant impact on life satisfaction in Eastern Europe, as does (3) the class mobility of others. The last finding points to the hitherto neglected importance of reference effects when considering the impact of social class on life satisfaction.

Keywords: *subjective well-being, social class, ESeC, diagonal reference models, social mobility, reference effects*

---

† This paper is based on an earlier manuscript available on the Open Science Framework (see: https://osf.io/4adhe/).

* Department of Social Policy & Intervention, University of Oxford.

1 Introduction

The strong association between social class and life outcomes has been recognised as an enduring feature of advanced market societies. Past research shows that the upper classes enjoy a clear advantage over those towards the bottom of the class distribution: They are more economically successful (Goldthorpe and McKnight, 2006), better educated (Bernardi and Ballarino, 2016; Shavit and Blossfeld, 1993), lead healthier and longer lives (White et al., 2007), and offer their children greater opportunities for a self-determined life (Bukodi and Goldthorpe, 2019). Consequently, occupying a sufficiently high class position, by either being born into it or attaining it through upward mobility, should truly matter for individuals' lives and their own personal evaluation thereof.

Nevertheless, quantitative scholarship has had only limited interest in the effects of social class, including the effect of class mobility, on life satisfaction. Few notable exceptions provide evidence for a class gradient in life satisfaction, finding that across various countries those in the working classes are less satisfied than those in the upper classes (Chan, 2018; Hadjar and Samuel, 2015; Haller and Hadler, 2006; Iveson and Deary, 2017; Lipps and Oesch, 2018; Zhao, Li, Heath, and Shryane, 2017). By contrast, results on the effect of intergenerational class mobility on life satisfaction appear mixed. Marshall and Firth (1999), Iveson and Deary (2017) as well as Zhao et al. (2017) do not find a robust relationship, while Hadjar and Samuel (2015) argue that intergenerational downward mobility decreases life satisfaction at least in the UK. Studies conceptualising mobility in terms of income (Dolan and Lordan, 2013), education (Nikolaev and Burns, 2014; Schuck and Steiber, 2017) or status (Clark and D'Angelo, 2015; Nikolaev and Burns, 2014; Zang and de Graaf, 2016) also report a negative effect of downward and a positive effect of upward mobility.

Moreover, theoretical as well methodological shortcomings appear to limit the ability of existing research to shed light on the complex relationship. An encompassing theorisation of how class position and mobility might affect life satisfaction is still lacking. Studies also struggle to distinguish

between the effects of class position and mobility in their empirical analyses, while the direct effects of own and parental class position that go beyond the effect of class mobility are still to be examined.

Against this background, the aim of this study is to offer a comprehensive account of how social class affects life satisfaction in Europe. Until now, the effect of social class on life satisfaction has been conceptualized narrowly, focusing on the effect of *own* social class. However, when societies change their class structure or become more mobile over time, these macro-level changes may have effects that go beyond the individual's own experience of social class. To allow for such macro-level impacts, we include so-called *reference effects* into our framework, which relate to outcomes of others. Since the notion of 'relative deprivation' was introduced by Stouffer (1949), social comparison theory (Festinger 1954) and reference group theory (Merton and Rossi, 1968) have become highly influential in sociological scholarship, and feature prominently in other domains of the wellbeing literature (Alderson and Katz-Gerro 2016; Präg, Mills, and Wittek 2014; van Deurzen, van Ingen, and van Oorschot 2015).

We analyse Western and Eastern Europe separately. Given the different political and economic trajectories of these two country groups, social class may structure inequalities in these societies differently and consequently exert heterogeneous effects on life satisfaction. Past research indicates a higher class gradient in mortality rates (Bessudnov et al., 2012; Vågerö, 2010), greater willingness to accept economic inequality (Kelley and Zagorski, 2004), and a greater dependence of political attitudes on the experience of social mobility (Gugushvili 2016) in Eastern than in Western Europe. Using subjective social class, Haller and Hadler (2006) further show that effects on life satisfaction appear stronger in post-socialist compared to traditionally capitalist countries. A joint account of East and West would hence mask regional variation, translating into an inaccurate picture of the relationship between social class and life satisfaction in Europe.

The remainder of the paper continues as follows: The next section outlines the theoretical framework for our analysis. Section 3 elaborates on our empirical approach. In section 4, we present our results before concluding in the final section.

2 Theoretical framework

We understand life satisfaction as a subjective and cognitive evaluation of a person's life (Pavot and Diener, 2008). Regarding social class, we employ a conception based on employment relations (Erikson et al., 1979, Bukodi and Goldthorpe, 2019). According to this conception, the service relationship and wage contract constitute ends of a class schema with the salariat at the top and the working class at the bottom. Self-employment is a third ideal type of employment relations (Rose et al., 2009).

As summarised in Figure 1, social class may affect life satisfaction through five pathways: (1) one's own social class (*class destination*), (2) the social class of one's parents (*class origin*), (3) the difference between class destination and origin, i.e. one's experience of intergenerational *class mobility*, (4) other's class position (*reference class position*), and (5) other's class mobility (*reference class mobility*). These aspects of social class are interrelated. Class origins influence destinations, and class origin and destination jointly define class mobility. Nevertheless, each aspect may also independently determine life satisfaction.

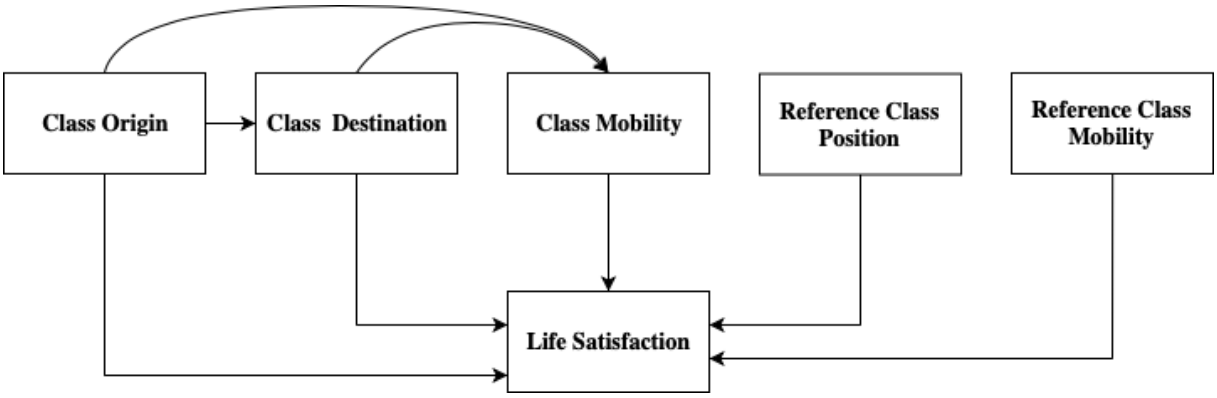


Figure 1. Theoretical framework of the relationship between social class and life satisfaction

Members of the working class are exposed to greater economic instability, experience lower income increases over the life course, and are more likely to be unemployed than those in the salariat (Goldthorpe and McKnight, 2006; Lucchini and Schizzerotto, 2010). Income (Kahneman and Deaton, 2010) and unemployment (Luechinger et al., 2010; Wulfgramm, 2014) are important socio-economic outcomes in determining life satisfaction. *Class destination* should hence affect life satisfaction through these outcomes.

Social class is further related to material deprivation (Nolan and Whelan, 2011), and predicts exposure to health hazards (Melchior et al. 2005). Higher class destinations may also increase individuals' social status, especially when considered in terms of the level of respect and admiration that individuals receive from others (Alderson and Katz-Gerro, 2016; Anderson, et al., 2012; Connolly and Sevä, 2018). As elaborated below, this suggests a role for social comparisons in explaining the relationship between social class and life satisfaction. We thus hypothesise:

H1: Individuals whose *class destination* is better characterised by a service relationship (higher social class) report higher levels of life satisfaction than those whose social class is better described by a wage contract (lower social class).

Parents tend to pass on class-based inequalities to their children and thereby shape their opportunities and outcomes in life. A strong association between class origin and destination has been found in many countries (Erikson and Goldthorpe 1992; Breen 2004). Class origin hence likely affects life satisfaction positively through its impact on class destination. It is nevertheless conceivable that parental class exerts effects that go beyond its effect on class destination.

Children of the working classes are more likely to grow up in an unstable home environment and experience stressful events such as financial hardship, parental unemployment or lone parenthood (Bukodi and Goldthorpe, 2019; Rowlingson and McKay, 2002). These have detrimental effects on children's wellbeing (Bubonya et al. 2017; Ermisch et al. 2004) that will likely figure in later life. As parental resources are available to compensate for undesirable outcomes, individuals with higher

class origins can afford to take greater risks and are more encouraged to pursue their aspirations. Parents also directly transmit health-related behaviours such as smoking (Gugushvili 2016).

Direct effects of class origin on life satisfaction are thus possible. However, since some class origin effects run via class destination, we expect class origin effects on life satisfaction to be smaller once controlling for the latter. Therefore:

> **H2:** Individuals with higher *class origins* report higher life satisfaction than individuals with lower *class origins*, even when controlling for class destination. When controlling for class destination, the effect of class origin is smaller than the effect of class destination.

Most theoretical advances have so far been made concerning the effects of intergenerational class mobility on life satisfaction. Two hypotheses received particular attention: Sorokin's (1959) 'dissociative hypothesis' and Newman's (1999) 'falling from grace hypothesis' (Houle, 2011; Houle and Martin, 2011; Zang and de Graaf, 2016; Zhao et al., 2017).[1] According to the former, moving between classes regardless of the direction is an inherently stressful experience as one leaves contexts of cultural familiarity that one has grown used to since childhood. Thus, ending up in a class that is different from one's origin may lead to disorientation, distress, and lower life satisfaction.

By contrast, Newman's hypothesis predicts that only individuals who experience downward mobility evaluate their life as worse. Underlying this conception is the notion of loss aversion: Individuals prefer avoiding losses over making gains and therefore gear efforts towards maintaining what they already have – such as the class position they were born and raised in (see also Breen and Goldthorpe 1997). Falling down to a lower class than one's parents should thus exert a negative effect.

[1] Although Sorokin and Newman were concerned with *intra*generational mobility, we take their arguments to also apply to the *inter*generational case.

Although Newman remains silent on upward mobility, it is then also conceivable that attaining a class destination that is higher than one's origin leads to increases in life satisfaction. Exceeding the class of one's parents may install a positive sense of achievement or pride, making the upwardly mobile more content. We therefore formulate two competing hypotheses about the effect of upward class mobility on life satisfaction:

H3a: Individuals who experience upward or downward *class mobility* report lower levels of life satisfaction than individuals who remain in their class origin.

H3b: Individuals who experience downward *class mobility* report lower levels of life satisfaction, while individuals who experience upward *class mobility* report higher levels of life satisfaction than individuals who remain in their class origin.

People compare themselves and derive value from having more advantageous positions relative to others (Festinger 1954). In the context of life satisfaction, the presence of such *reference effects* has been most notably discussed with respect to income (Delhey and Kohler, 2006; Ferrer-i-Carbonell, 2005; McBride, 2001), and unemployment (Clark, 2003; Heggebø and Elstad, 2018). Social class has not been considered so far. Yet, processes of social comparison are likely to operate here too. If those in one's reference group predominantly occupy higher class destinations, one can be expected to evaluate one's life as worse. This is because seeing others doing better sets a higher bar for one's own achievement. Likewise, if those in one's reference group predominantly occupy lower class destinations, one may evaluate one's life as better, since the criteria for this judgement are less demanding. We thus state:

H4a: Higher levels of *reference class position* negatively affect life satisfaction, while lower levels of *reference class position* positively affect life satisfaction.

Analogously, if those within one's reference group have been more upwardly mobile, one may evaluate one's own performance more poorly. For example, after observing that most individuals

in one's class destination come from lower origins, one might consider one's own class mobility less impressive, and vice versa. Therefore, we expect:

H4b: Higher levels of upward *reference class mobility* negatively affect life satisfaction, while higher levels of downward *reference class mobility* positively affect life satisfaction.

Important differences between Eastern and Western Europe lead us to believe that the strength of the association between social class and life satisfaction varies across Europe. First, most Eastern European countries experienced large increases in income inequality (Grosfeld and Senik, 2010; Solt, 2019), have smaller welfare states (Kuitto 2018) and are less economically developed. Thus, Eastern Europeans face greater economic uncertainty and a greater risk of poverty. As explained above, class destination and origin mitigate or exacerbate such economic risks. Hence:

H5: The effects of *class destination* and *class origin* on life satisfaction are larger in Eastern than in Western Europe.

Second, prior to Eastern Europe's economic transition to capitalism, working in white-collar professions did not result in markedly better economic outcomes than blue-collar jobs. However, during the transition, blue-collar and manual labour became relatively more disadvantaged over time (Gerber and Hout, 1998; Slomczyński and Shabad, 1996). Thus, social classes are currently more stratified than in the past, making differences between one's class destination and one's class origin more salient. Consequently:

H6: The effect of intergenerational *class mobility* on life satisfaction is larger in Eastern than in Western Europe.

Third, Eastern European countries today exhibit lower relative social mobility ("social fluidity") than the West (Paskov and Bukodi, 2018). In such environments, reference class mobility gives individuals a stronger signal about their relative performance than environments with weaker origin-destination associations. Our last hypothesis therefore states:

H7: The effect of *reference class mobility* on life satisfaction is larger in Eastern than in Western Europe.

3 Empirical approach

3.1 Data

We use waves 1-5 of the European Social Survey (ESS), covering the period 2002-2010 with repeated cross-sections. Later waves cannot be used, since data to construct class origin is unavailable. The 32 countries included are roughly evenly split between Eastern and Western Europe.[2] A list of countries is given in appendix Table A1.

Our dependent variable are responses to the question *"All things considered, how satisfied are you with your life as a whole nowadays?"*, which are scored on a scale from 0 (dissatisfied*)* to 10 (satisfied*)*. See Diener et al. (2013) for a review on the validity of this measure. For social class, we use the European Socio-Economic classification (ESeC; Rose et al., 2014). Unlike previous studies, we do not collapse the 9-category class scheme, allowing for more fine-grained analyses. We assume that social classes are hierarchically ordered across employment relations. Occupations with the same employment relation are assumed to be ordered from blue-collar to white-collar work. However, regarding life satisfaction, the self-employed classes do not easily fall into a hierarchical ordering. See Section 4 for descriptive evidence on this. We therefore exclude the self-employed from analyses that require assumptions about the ordering of classes.

For both class destination and class origin, we construct the "full" version of ESeC based on 3-digit ISCO-88 occupational codes, employment status, number of employees and supervisory functions. We use parental ISCO codes as provided in Ganzeboom (2014). When variables other

[2] Blasius and Thiessen (2018) note that data quality is worse in Eastern than in Western Europe. This is problematic if data quality is correlated with the association between class and life satisfaction, which may occur via interviewers/interviewees providing answers at random. If so, the association between class and life satisfaction is biased towards zero. Our results for Eastern Europe may therefore be conservative compared to Western Europe.

Table 1. Description and hierarchical ordering of ESeC classes (adapted from Rose et al., 2009)

	ESeC Class	Common term	Employment relation
I	Large employers, higher grade professional, administrative and managerial occupations	Higher salariat	Service relationship
II	Lower grade professional, administrative and managerial occupations and higher grade technician and supervisory occupation	Lower salariat	Service relationship (modified)
III	Intermediate occupations	Higher grade white collar workers	Mixed
IV	Lower supervisory and lower technician occupations	Higher grade blue collar workers	Mixed
V	Lower services, sales and clerical occupations	Lower grade white collar workers	Labour contract (modified)
VI	Lower technical occupations	Skilled workers	Labour contract (modified)
VII	Routine occupations	Semi- and non-skilled workers	Labour contract
SEI	Small employers and self-employed occupations (excl. agriculture.)	Petit bourgeoisie or independents	-
SEII	Self-employed occupations (agriculture)	Petit bourgeoisie or independents	-

Note: Classes SEI and SEII are assumed to fall outside any hierarchical ordering. Those that never worked are not included in this study.

than ISCO codes are missing, we assign the weighted mode of ESeC class within that ISCO code. As a refinement of the official "simplified" version, we determine modes separately for destinations and origins based on respondents for whom the "full" version is available. Class origin refers to parental class when respondents were 14 years old. We use the "dominance approach" (Erikson 1984), which assigns the higher class when parental class positions differ. To accurately capture intergenerational mobility trajectories, we only include respondents of working age, and who can be assumed to have reached "occupational maturity", i.e. those between ages 35 and 65 (Bukodi, Paskov, and Nolan 2019).

3.2 Model choice

A general model for the effects of class destination, origin and mobility, including further covariates such as reference position and reference mobility, is given by:

$$LS_{idom} = \delta_d dest_{id} + \omega_o orig_{io} + \gamma_m mob_{im} + \beta' X_i + \varepsilon_i \,, \qquad (1)$$

with parameters δ_d for class destinations d, ω_o for class origins o, and γ_m for mobility levels m.[3] $dest_{id}$, $orig_{io}$, mob_{im} are sets of dummies, indicating origin, destination, and mobility of respondent i. Destination class d and origin class o are reversely coded from 1 to 7. Thus, an index

[3] An even more general model not considered here would further allow for interactions of origins and destinations with levels of mobility.

of 1 refers to Class VII and an index of 7 refers to Class I. Mobility levels are given by the difference between destination and origin, i.e. $m = d - o$. Therefore, higher levels of m indicate greater upward mobility. X_i is a vector of covariates, including a constant and terms for reference mobility and position. ε_i denotes the error term.

Unfortunately, (1) is not identified due to the perfect collinearity of m with d and o (Blalock, 1966). We must therefore place constraints on the relationships between δ_d, ω_o, and γ_m. One possibility is to assume $\omega_o = 0$ for all levels of o, i.e. that there are no origin effects (see e.g. Hadjar and Samuel, 2015). However, if any ω_o is in fact different from zero, δ_d and γ_m will pick up these effects, and hence be biased. Similar reasoning applies to excluding terms for destination.

Models without mobility terms may yet be instructive. Origin may have a direct effect, an effect via destination, and an effect via mobility. Likewise, class destination might impact life satisfaction directly, as well as indirectly via class mobility. Since mobility is defined by the difference between destination and origin, a one-step increase in origin deterministically decreases mobility by one step[4], whereas a one-step increase in destination deterministically increases mobility by one step. Thus, in a model that only contains terms for origins, its coefficients provide the total origin effects, i.e. direct effects and the effects that run via destination and mobility. By adding terms for destinations, the origin effects that run via destinations are removed from the origin coefficients. In such a model, where both sets of terms are included, the coefficients for origins and destinations consequently contain their respective direct effects and the indirect effects via mobility.[5]

Still, such models cannot fully separate the mobility from the origin and destination effects. We hence employ a model proposed by Sobel (1981; 1985):

$$LS_{idom} = p\lambda_d dest_{id} + (1-p)\lambda_o orig_{io} + \gamma_m mob_{im} + \beta' \boldsymbol{X_i} + \varepsilon_i \qquad (2)$$

[4] This only holds when controlling for class destination, since higher origins may also cause higher destinations.
[5] Due to space constraints, we do not perform a formal mediation analysis for the degree to which the origin effect runs via destination. Besides challenges posed by mobility effects, this is because the effect of origin on entering particular destinations is probably non-linear (exploratory analyses are available upon request). Therefore, how destinations mediate the effect of origin differs across origins.

Here, p is a parameter that is constrained to take values between 0 and 1. It may be interpreted as indicating the importance of destinations *versus* origins (Hendrickx et al., 1993). In baseline specifications and to ease interpretation, we replace the full set of mobility parameters γ_m by a simpler linear term:

$$LS_{idom} = p\lambda_d dest_{id} + (1-p)\lambda_o orig_{io} + \dot{\gamma}m + \beta' X_i + \varepsilon_i \quad (3)$$

Such models are called *diagonal reference models* (DRMs), because the λs are interpreted as positional effects of the immobile, i.e. individuals who remain on the diagonal of a mobility table (e.g. Zhao et al. 2017). Substantively, we assume that the life satisfaction of intergenerationally mobile individuals is a weighted sum of three main components: (1) the life satisfaction of the immobile in individuals' class origin (λ_o), (2) the life satisfaction of the immobile in individuals' class destination (λ_d), and (3) a further effect of having moved between origin and destination by a certain number of steps (γ_m or $\dot{\gamma}$).

Mathematically, the DRM is a constrained version of (1). It constrains the destination effects to be positive linear transformations of the origin effects. To see this, note that the effect of being in class destination k rather than $k-1$ is given by $p(\lambda_k - \lambda_{k-1})$. Likewise, the effect of being in class origin k rather than $k-1$ is given by $(1-p)(\lambda_k - \lambda_{k-1})$. Thus, origin and destination effects may only differ due to parameter p. So long as p lies between 0 and 1, destination effects are positive transformations of origin effects.[6] Consequently, relative differences between each step among class origins and destinations are assumed to be the same, while the absolute magnitude of these steps may differ. See section 4.3 and appendix C for tests on the sensitivity of our results to these assumptions.

[6] When $p = 1$, DRMs are equivalent to an OLS regressions that exclude origin terms. Likewise, when $p = 0$, DRMs are equivalent to OLS models without destination terms.

3.3 Selection of controls

We always control for NUTS 1 region and wave dummies, age(-squared), ethnic minority status, and whether a respondent was born in the country of residence. The ESS provides additional controls. Some of these are pre-treatment variables with respect to class destination, and thus potential confounders, while also being post-treatment variables with respect to class origin. Their inclusion could therefore induce post-treatment bias (Montgomery et al. 2018). Education is an example. Origin determines education, which may directly affect life satisfaction. Including education thus blocks the effect of origin that may run via education. Yet, education is also a strong predictor of class destination (Bukodi and Goldthorpe, 2016), so its exclusion confounds the destination effect.

Similarly, religiosity, marital status, residence in rural versus urban areas, and household size are post-treatment variables with respect to class origin. Regarding class destination, the causal direction is ambiguous, making these variables either confounders or terms inducing post-treatment bias. We therefore report results with and without these variables.

To gauge the degree to which origin and destination effects are mediated by unemployment and income, we additionally provide results with these variables. Household income is always entered as its natural logarithm.[7]

One key limitation is that our data is observational and cross-sectional. Our estimates could therefore be subject to further omitted variable bias (e.g. personality traits (Boyce 2010)). Nevertheless, these may provide a useful baseline against which to compare past and future evidence. Even if one was reluctant to engage in causal interpretation, the estimates still provide information on conditional mean differences in life satisfaction between origin, destination and mobility levels. Given the normative significance and vigorous debate surrounding class-based

[7] Due to space constraints and reasons similar to footnote 5, we leave a full-fledged mediation analysis to future research.

inequalities, such differences are likely to have worthwhile independent political or moral implications.

3.4 Operationalizing reference effects

When constructing measures for reference effects, we must first choose a relevant reference group and then decide how outcomes of members of the reference group are aggregated.

Regarding the first choice, the previous literature suggests that individuals mostly compare themselves with others who are similar to themselves, particularly with colleagues, friends, and family members (Alderson and Katz-Gerro, 2016; Clark and Senik, 2010; Goerke and Pannenberg, 2015). They therefore tend to compare themselves to those with similar jobs, educational backgrounds, or parents, implying that reference groups likely consist of people of the same class destination and origin. As it is impossible to estimate the effect of reference class position (mobility), if individuals of the same destination (origin) are included in the reference group, we take respondents with the same class origin to constitute the relevant reference group for reference position, and respondents occupying the same class destination to form the relevant reference group for reference mobility.[8]

Concerning the second choice, we use average class destination and class mobility as measure for others' 'typical' levels of class destination and mobility. Following Vendrik and Woltjer (2007) and Ferrer-i-Carbonell (2005), we further restrict reference groups to those up to five years older or younger than the respondent. See appendix E for a more formal definition of reference mobility and position.

Some individual-level controls will be correlated with these measures. For example, individuals with higher incomes are likely situated in environments where others also have higher incomes.

[8] Defining the reference group for reference position as those within a respondent's class destination, would mean that respondents of the same class destinations are compared to each other, thus resulting in no variation. Taking respondents with the same class origin to form the reference group for reference mobility, yields a term for reference mobility that is given by reference position minus own origin. Reference mobility would then be perfectly collinear with reference position and class origin, and thus not identifiable.

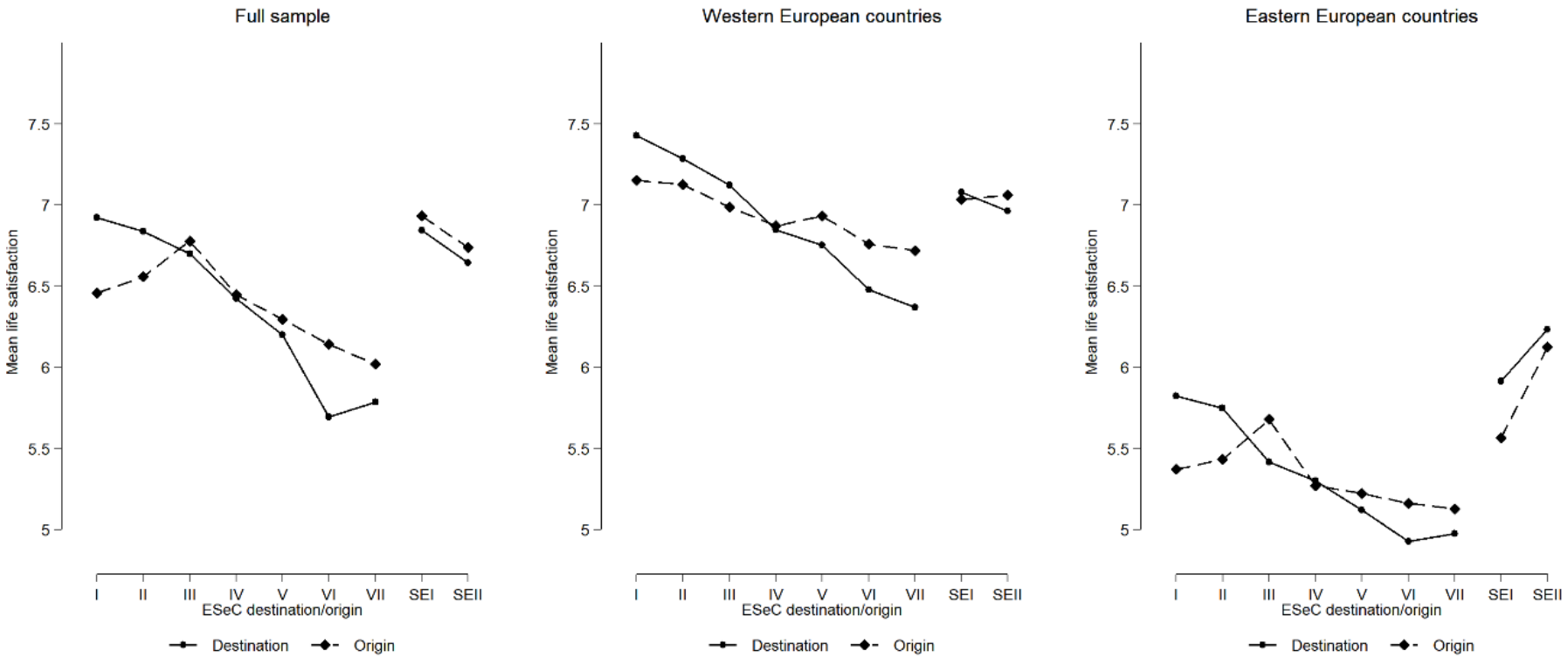
However, apart from network effects, it is unlikely that variables like income are strongly affected by our measures of reference position and mobility. Instead, such variables may proxy for other reference group characteristics (e.g. average incomes). They should therefore be included to prevent reference mobility and position from picking up spurious effects from correlated individual-level characteristics.

Finally, although **H7** is motivated by differences in relative social mobility, our measure of reference mobility exclusively captures absolute mobility. Contrasting our approach with measures of relative mobility, perhaps using the individual-level M-Index by Seiler and Jann (2019), seems like a fruitful avenue for future work.

4 Results

Figure 2 shows mean life satisfaction by destination and origin classes. In the full sample, we observe, for class destination, a negative gradient that follows the ordering of ESeC classes. Respondents in the salaried classes have the highest level of life satisfaction and respondents in the working classes have the lowest. However, the pattern is different when origin is considered. Children of the intermediate classes are on average slightly more satisfied than respondents of both working class and salariat origins. Splitting the sample into Eastern and Western Europe reveals that this pattern is pre-dominantly driven by the East. There, mean life satisfaction is substantially lower overall. Those in origin and destination class I are even less satisfied than Western members of origin and destination class VII. Strikingly, the self-employed (SEI) and farmers (SEII) are the most satisfied in Eastern Europe. One possible reason is that following the collapse of socialism, many who were in state-managerial positions entered careers as entrepreneurs (Slomczyński and Shabad, 1996). This vindicates our view that distinguishing between East and West is crucial, and that the self-employed classes cannot be straightforwardly placed in a hierarchical ordering for our present purposes. Unfortunately, these figures can neither directly show mobility effects nor indicate the presence of reference effects. Effects of confounders are also not cancelled out. We therefore turn to our regression results.

Figure 2. Mean life satisfaction by destination and origin classes in Europe



4.1 Western Europe

Table 2 presents OLS regression results, in which terms for class mobility are omitted. For both class destination and origin, we define Class I, i.e. the upper salariat, as the reference category and analyse whether life satisfaction differs between classes (see also top panels of figure 3).

In column 1, we only enter class origin and our set of basic controls. As explained in section 3.2, these estimates are informative about the total effect of origin, i.e. its direct and indirect effects via destination and mobility. We observe a clear and substantial gradient across origins, with a difference between Class I and VII of 0.406 on our 11-point life satisfaction scale (see appendix Tables B1-B6 for all coefficients). We do not observe a significant difference between Class I and II, and Classes IV and V do not follow the hierarchical ordering. The two self-employed classes SEI and SEII fall between Classes III and IV.

Table 2. Origin and Destination estimates for Western Europe (OLS)

	(1) Basic model	(2) Class destination added	(3) Controls for education	(4) Controls for educ., further demographics, UE & income
Orig II	-0.044 (0.046)	0.021 (0.046)	0.032 (0.047)	0.041 (0.046)
Orig. III	-0.110* (0.064)	-0.025 (0.062)	-0.002 (0.063)	0.020 (0.061)
Orig. IV	-0.244*** (0.062)	-0.081 (0.061)	-0.049 (0.062)	-0.050 (0.059)
Orig. V	-0.191*** (0.070)	0.005 (0.068)	0.040 (0.069)	0.062 (0.066)
Orig. VI	-0.312*** (0.058)	-0.046 (0.055)	0.005 (0.058)	0.039 (0.055)
Orig. VII	-0.406*** (0.059)	-0.117** (0.056)	-0.057 (0.058)	-0.021 (0.056)
Orig. SEI	-0.123** (0.057)	0.020 (0.056)	0.050 (0.057)	0.051 (0.055)
Orig. SEII	-0.107 (0.070)	0.129* (0.069)	0.183*** (0.069)	0.076 (0.067)
Dest. II		-0.149*** (0.043)	-0.125*** (0.044)	-0.025 (0.040)
Dest. III		-0.247*** (0.054)	-0.195*** (0.056)	-0.026 (0.055)
Dest. IV		-0.529*** (0.064)	-0.436*** (0.067)	-0.218*** (0.067)
Dest. V		-0.624*** (0.071)	-0.531*** (0.072)	-0.259*** (0.067)
Dest. VI		-0.773*** (0.072)	-0.644*** (0.077)	-0.297*** (0.073)
Dest. VII		-0.932*** (0.066)	-0.799*** (0.069)	-0.398*** (0.069)
Dest. SEI		-0.303*** (0.068)	-0.222*** (0.069)	-0.071 (0.063)
Dest. SEII		-0.520*** (0.117)	-0.399*** (0.117)	-0.224** (0.113)
Log likelihood	-124485.221	-123899.862	-123842.369	-121736.466
AIC	249156.442	248001.724	247890.737	243704.933
N	57,868	57,868	57,868	57,868

Note: * p<0.10, ** p<0.05, *** p<0.01. Reference categories are Origin and Destination class I. Region- and wave-clustered standard errors in parentheses. All models include controls for region and wave dummies, age(-squared), ethnic minority status, and whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

In column 2, we add class destination. The class origin estimates should now be interpreted as representing the sum of the direct effect of origin plus the indirect effect via class mobility. Similarly, the coefficients for destination are interpreted as direct effects plus the indirect effect via mobility. The indirect effect of origin via destination is blocked by controlling for the latter.

The estimates for origin are now largely insignificant. Recall that when controlling for destination, a one-step higher origin implies a one-step lower level of upward mobility. Thus, assuming that mobility effects themselves are positive (i.e. the higher upward mobility, the higher life satisfaction, and the higher downward mobility, the lower life satisfaction), the indirect effect of origin via mobility is negative. Therefore, these results suggest that origin and mobility effects cancel.

However, we obtain significant coefficients for origin SEII (=0.129). This is striking, since it implies that children of farmers are more satisfied than those who grew up in the presumably most socio-economically advantaged social class. Regarding destination, we observe a pronounced gradient in life satisfaction. Differences roughly equal 0.1, although there is a larger gap between Class III and IV (=0.27).

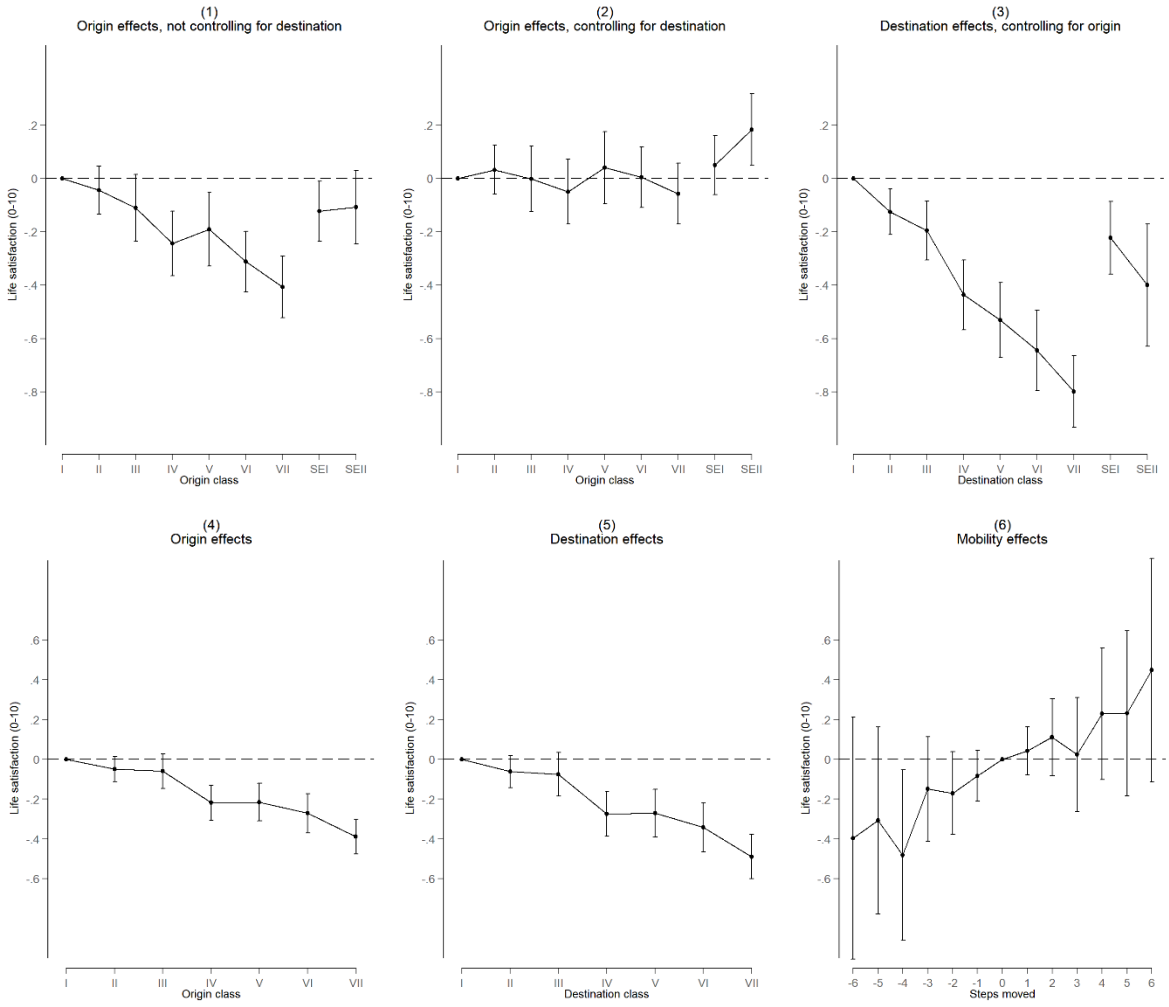
Given the discussion in section 3.3, these estimates for destination are likely confounded by failing to control for education. Column 3 rectifies this issue, which results in a minor reduction of the destination effect. Our inclusion of further "ambiguous" demographic controls also makes little difference to our results (omitted from table 2, see column 4 of appendix Table B1).

In column 4, we add unemployment and log household income. These are mediators of class destination. Compared to column 3, the overall gradient in class destination is reduced by about 50%. Importantly, there is no statistically significant difference between Classes II and III versus Class I. Therefore, differences between Classes I to III seem to be driven by income and unemployment rather than other benefits of Class I.[9]

[9] Our coefficients on unemployment and income are respectively 0.979 and 0.373 in the specification of Table 2 (see table B1 for all coefficients). In column 3 of appendix table D1, we find that the coefficients of Classes II and III on

Given the observed gradient in life satisfaction across class destinations, these results appear to support **H1**. Our failure to find a clear gradient for class origin when controlling for destination suggests a contradiction of **H2**. However, in both cases our origin and destination estimates include their indirect effects via class mobility.

Figure 3. Origin, destination, and mobility effects in Western Europe with 95% confidence intervals



Note: Panels 1-3 are based on OLS estimates of table 2. Panel 1 uses coefficients from column 1. Panels 2-3 use coefficients form column 3. Panels 4-6 are based on DRM estimates of Table 3 column 4.

income are respectively 0.216 and 0.318. The corresponding coefficients for unemployment are -0.001 for Class II and 0.021 for Class III. Taking products of coefficients, this suggests that more of the indirect effect of these class destinations runs via income than unemployment. Similar statements apply to the other classes.

We therefore attempt to disentangle the separate positional and mobility effects by using DRMs. Since we now require assumptions about the ordering of social classes, we exclude the self-employed. DRMs constrain the shape, but not the magnitude, of the origin and destination effects to be equal. Consequently, Table 3 only presents one set of coefficients for class position. The relative magnitudes of destination and origin effects are determined by p. To discern the effect of e.g. being in destination VII rather than I, we need to take the size of the coefficient on class VII and multiply it by p. Similarly, for origin effects, we take the coefficient of a particular class and multiply by $(1-p)$.

Table 3. Mobility estimates for Western Europe (DRM)

	(1) Linear mobility	(2) Linear mobility, controls for education	(3) Linear mobility, controls for educ., further demographics, UE & income	(4) Non-linear mobility, controls for educ.	(5) Non-linear mobility, controls for educ., further demographics, UE & income
Class II	-0.160** (0.071)	-0.109 (0.073)	0.004 (0.063)	-0.110 (0.073)	0.007 (0.066)
Class III	-0.266*** (0.082)	-0.167* (0.092)	0.028 (0.089)	-0.134 (0.101)	0.061 (0.094)
Class IV	-0.661*** (0.089)	-0.510*** (0.095)	-0.286*** (0.092)	-0.493*** (0.102)	-0.261** (0.105)
Class V	-0.663*** (0.100)	-0.506*** (0.102)	-0.203** (0.095)	-0.484*** (0.110)	-0.177* (0.101)
Class VI	-0.833*** (0.118)	-0.628*** (0.118)	-0.245** (0.108)	-0.611*** (0.113)	-0.228** (0.108)
Class VII	-1.096*** (0.088)	-0.879*** (0.098)	-0.420*** (0.098)	-0.877*** (0.101)	-0.420*** (0.100)
Steps mobile	0.056 (0.064)	0.059 (0.047)	0.031 (0.019)		
6 Steps down				-0.396 (0.310)	-0.142 (0.215)
5 Steps down				-0.306 (0.240)	-0.162 (0.150)
4 Steps down				-0.482** (0.218)	-0.323** (0.146)
3 Steps down				-0.149 (0.134)	-0.076 (0.086)
2 Steps down				-0.169 (0.106)	-0.101 (0.080)
1 Step down				-0.082 (0.065)	-0.048 (0.050)
1 Step up				0.044 (0.062)	0.012 (0.049)
2 Steps up				0.112 (0.098)	0.055 (0.070)
3 Steps up				0.024 (0.146)	-0.044 (0.094)
4 Steps up				0.229 (0.168)	0.132 (0.097)
5 Steps up				0.233 (0.212)	0.103 (0.114)
6 Steps up				0.449 (0.287)	0.288* (0.157)
p	0.601* (0.355)	0.560* (0.321)	0.575** (0.243)	0.557** (0.271)	0.578** (0.227)
Log likelihood	-83710.896	-83670.845	-82212.744	-83657.160	-82201.708
AIC	167609.793	167533.689	164643.488	167528.320	164643.416
N	40,342	40,342	40,342	40,342	40,342

Note: Reference category is Class I. All other notes as in Table 2.

In all specifications, we observe that p is larger than 0.5, indicating that destination effects are somewhat larger in magnitude than origin effects. Again, the inclusion of education in column 2 marginally reduces the positional class effects compared to column 1. Adding further demographic controls also does little to change our results. The inclusion of unemployment and log household

income roughly halves the coefficients on class position. As shown in bottom panels of figure 3, we find significant gradients in life satisfaction for both class destination and origin. This offers evidence in favour of hypotheses **H1** and **H2**.

Columns 1-3 contain a simple linear term for mobility (see equation (3)). This mobility term is not precisely estimated and therefore insignificant in all specifications. In columns 4-5, we use the full specification as given in equation (2). Most of the dummies for each mobility level are not significantly different from non-mobility. The only exceptions are those indicating moving down by four steps and moving up by six steps. Further tests on whether moving down by six steps is significantly different from moving up by six steps are insignificant.

With these estimates, we reject both **H3a** and **H3b** for Western Europe. However, given the coefficient signs of column 4 and 5, we reject **H3a** (mobility in any direction is harmful) more firmly than **H3b** (upward mobility is beneficial and downward mobility is harmful). Indeed, our evidence to reject **H3b** is marginal and depends on a p-value of 0.1 to determine significance.

Table 4 presents results on reference effects to evaluate **H4a** and **H4b**. As discussed earlier, it seems implausible that our measures of reference position and mobility strongly affect our individual-level controls. Instead these controls may act as proxies for wider macro-level factors. We hence prefer models with a full set of controls (columns 2 and 4) when considering reference effects.

Reference position is insignificant and positive throughout, thus contradicting **H4a**. By contrast, although reference mobility is initially insignificant in column 1, it turns significant in column 2.[10] The negative coefficient implies that greater upward (downward) mobility in individuals' reference group has a negative (positive) effect on their life satisfaction. This is similar to the negative reference effects found in the literature on relative income. Given seven hierarchically ordered

[10] Appendix table B3 which includes additional specifications and coefficients for all controls shows that this difference in coefficients is primarily driven by the inclusion of unemployment and income. In column 2 of table 4, we also find that own mobility is just significant, underlining our earlier rejection of **H3b** as marginal.

classes, our measure of reference mobility is theoretically bounded between -6 and 6. In practice, we observe a range of -5.02 and 5.13, and a standard deviation of 1.31. Thus, moving from the lowest observed value of reference mobility to the highest value implies a decrease in life satisfaction by -0.65 points on a 0-10 scale. Alternatively, a one standard deviation increase implies a 0.08 point decrease in life satisfaction.

In columns 3 and 4, we define separate terms for predominant upward or downward mobility[11] to disentangle which of these drive the observed effects. It turns out that cases with predominant downward mobility drive the result of column 2. Consequently, individuals who observe greater downward mobility around them are more satisfied, while those who observe greater upward mobility are not less satisfied. We thus obtain evidence in favour of **H4b**.

Table 4. Reference estimates for Western Europe (DRM)

	(1) Basic controls	(2) Controls for educ., further demographics., UE & income	(3) Basic controls; positive. vs. negative ref. mobility	(4) Controls for educ., further demographics., UE & income; pos. vs. neg. ref. mobility
Class II	-0.151** (0.076)	-0.021 (0.065)	-0.138* (0.071)	-0.006 (0.063)
Class III	-0.269*** (0.099)	-0.025 (0.096)	-0.235** (0.102)	0.017 (0.094)
Class IV	-0.653*** (0.121)	-0.361*** (0.117)	-0.611*** (0.118)	-0.297** (0.126)
Class V	-0.674*** (0.146)	-0.337** (0.139)	-0.629*** (0.143)	-0.286** (0.140)
Class VI	-0.830*** (0.183)	-0.404** (0.178)	-0.805*** (0.180)	-0.386** (0.184)
Class VII	-1.123*** (0.196)	-0.633*** (0.185)	-1.125*** (0.194)	-0.651*** (0.184)
Steps mobile	0.082 (0.069)	0.049* (0.030)	0.076 (0.055)	0.039 (0.030)
Ref. position	0.053 (0.040)	0.014 (0.038)	0.051 (0.040)	0.011 (0.038)
Ref. mobility	-0.036 (0.040)	-0.064* (0.037)		
Ref. mob. up			-0.004 (0.043)	-0.022 (0.043)
Ref. mob. down			0.071 (0.056)	0.116** (0.058)
p	0.570 (0.347)	0.592** (0.233)	0.596** (0.284)	0.685*** (0.247)
Log likelihood	-83707.526	-82208.695	-83705.955	-82205.913
AIC	167607.053	164639.389	167605.910	164635.826
N	40,342	40,342	40,342	40,342

Note: All notes as in Table 3.

4.2 Eastern Europe

Table 5 presents specifications for Eastern Europe (see also top panels of figure 4). These are analogous to those in table 2 and omit terms for class mobility. Only class origin is included in column 1. Overall, the estimated size of the coefficients is similar to those obtained for Western

[11] See appendix E for formal definitions.

Europe, with a difference of -0.338 between Classes I and VII. Notably, the coefficient of Class VI is even more negative (=-0.414) than for Class VII.

Table 5. Origin and destination estimates for Eastern Europe (OLS)

	(1) Basic model	(2) Class destination added	(3) Controls for education	(4) Controls for educ., further demographics, UE & income
Orig. II	-0.115 (0.101)	-0.028 (0.098)	-0.011 (0.103)	-0.027 (0.095)
Orig. III	-0.178 (0.141)	-0.085 (0.141)	-0.054 (0.135)	-0.059 (0.127)
Orig. IV	-0.275** (0.111)	-0.129 (0.108)	-0.088 (0.116)	-0.062 (0.096)
Orig. V	-0.269* (0.158)	-0.048 (0.150)	-0.003 (0.155)	-0.046 (0.149)
Orig. VI	-0.414*** (0.110)	-0.149 (0.101)	-0.077 (0.114)	-0.082 (0.099)
Orig. VII	-0.338*** (0.062)	-0.031 (0.055)	0.052 (0.058)	0.115** (0.049)
Orig. SEI	-0.370 (0.256)	-0.179 (0.257)	-0.125 (0.256)	-0.117 (0.231)
Orig. SEII	-0.058 (0.118)	0.215* (0.117)	0.329*** (0.123)	0.298*** (0.102)
Dest. II		-0.118* (0.060)	-0.045 (0.064)	-0.040 (0.078)
Dest. III		-0.519*** (0.099)	-0.364*** (0.097)	-0.340*** (0.104)
Dest. IV		-0.590*** (0.130)	-0.382*** (0.126)	-0.325** (0.135)
Dest. V		-0.751*** (0.182)	-0.512** (0.204)	-0.385* (0.212)
Dest. VI		-0.933*** (0.075)	-0.648*** (0.103)	-0.457*** (0.126)
Dest. VII		-0.915*** (0.091)	-0.607*** (0.091)	-0.437*** (0.097)
Dest. SEI		-0.132 (0.137)	0.047 (0.144)	0.027 (0.130)
Dest. SEII		-0.392** (0.161)	-0.072 (0.175)	-0.042 (0.180)
Log likelihood	-51280.739	-51050.728	-50996.867	-50405.175
AIC	102645.478	102201.457	102097.734	100940.350
N	22,308	22,308	22,308	22,308

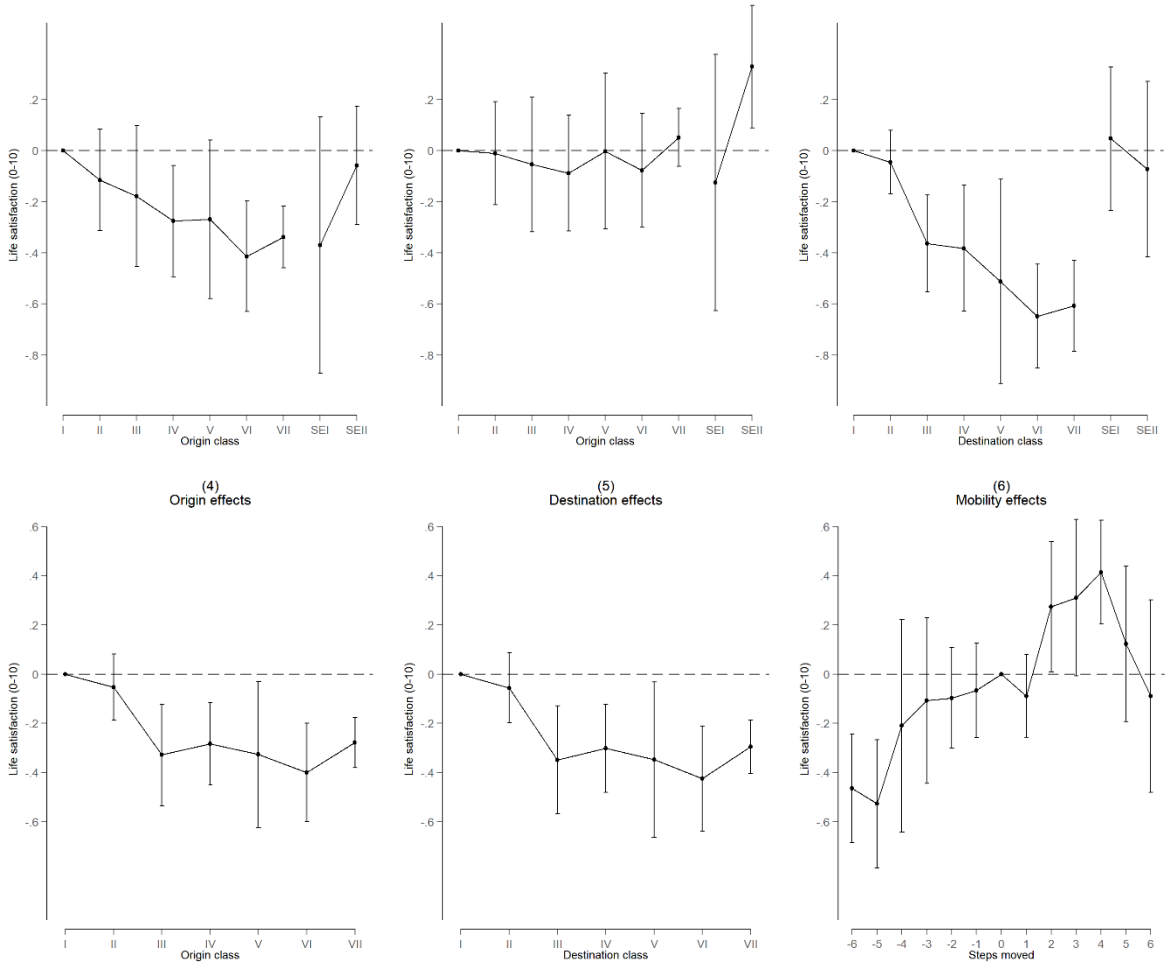
Note: All notes as in table 2.

In column 2, we enter terms for class destination. As was the case for Western Europe, almost all origin terms are now insignificant and stay insignificant in the subsequent specifications. This suggests again that origin and mobility effects may move in opposite directions. The only exception is the robustly significant positive coefficient of the class of farmers (SEII), which is also positive for Western Europe. As it is difficult to place the self-employed in an overall hierarchical class ordering, it is also difficult to establish whether individuals of SEII origin are upwardly, downwardly or horizontally mobile when moving to a different class. Our results may therefore either be driven by origin class itself (perhaps via benefitting from growing up in a rural, more socially integrated environment), or via a SEII specific mobility effect.

Regarding class destination, the gradient in life satisfaction is similar to Western Europe. Coefficients reported in column 2 are likely confounded by education. Indeed, after adding education in column 3, our destination estimates are reduced by about a third, with the largest

difference of 0.648 between Classes I and VI. Unlike Western Europe, we observe a large difference of -0.365 between Classes II and III and no significant difference between Classes I and II. It therefore appears that Classes I and II are particularly more advantaged than white- and blue-collar workers, resulting in higher life satisfaction. Including unemployment and income as mediators reduces the magnitudes of the destination coefficients by much less than in Western Europe. Characteristics of class destination other than current income and unemployment may hence matter more in the East.[12]

Figure 4. Origin, destination, and mobility effects in Eastern Europe with 95% confidence intervals



Note: Panels 1-3 are based on OLS estimates of table 5. Panel 1 uses coefficients from column 1. Panels 2-3 use coefficients form column 3. Panels 4-6 are based on DRM estimates of table 6 column 4.

[12] Coefficients on unemployment and income in Eastern Europe are respectively 0.839 and 0.367 (see appendix table B4), and hence slightly smaller than for Western Europe. Columns 4 and 6 of tables D1 and D2 reveal that the effects of class destination on income are stronger in Western than in Eastern Europe, while the effects of destination on unemployment are similar. It therefore appears that differences in income effects drive differences in mediation across Eastern and Western Europe.

Since the estimates of table 5 include mobility effects of unknown size, we cannot make conclusions regarding **H1** and **H2**. Table 6 presents our results using DRMs (see bottom panels of Figure 4). We again find that Class VI rather than VII shows the most negative coefficient. As also shown in table 5, Classes I and II are more satisfied than any other class, even when controlling for unemployment and income as potential mediators.

Table 6. Mobility estimates for Eastern Europe (DRM)

	(1) Linear mobility	(2) Linear mobility, controls for education	(3) Linear mobility, controls for educ., further demographics, UE & income	(4) Non-linear mobility, controls for educ.	(5) Non-linear mobility, controls for educ., further demographics, UE & income
Class II	-0.127 (0.113)	-0.030 (0.132)	-0.046 (0.150)	-0.108 (0.141)	-0.118 (0.148)
Class III	-0.661*** (0.181)	-0.462** (0.183)	-0.459*** (0.164)	-0.677*** (0.217)	-0.639*** (0.202)
Class IV	-0.611*** (0.144)	-0.357** (0.170)	-0.300* (0.168)	-0.585*** (0.177)	-0.512*** (0.181)
Class V	-0.797*** (0.284)	-0.501 (0.326)	-0.416 (0.325)	-0.674** (0.312)	-0.593* (0.311)
Class VI	-1.113*** (0.155)	-0.753*** (0.207)	-0.588*** (0.208)	-0.824*** (0.211)	-0.658*** (0.212)
Class VII	-0.932*** (0.123)	-0.538*** (0.117)	-0.314*** (0.092)	-0.573*** (0.107)	-0.346*** (0.089)
Steps mobile	0.060*** (0.018)	0.053*** (0.015)	0.054*** (0.012)		
6 Steps down				-0.464*** (0.112)	-0.461*** (0.113)
5 Steps down				-0.527*** (0.133)	-0.448*** (0.130)
4 Steps down				-0.209 (0.220)	-0.129 (0.202)
3 Steps down				-0.107 (0.172)	-0.096 (0.174)
2 Steps down				-0.097 (0.104)	-0.103 (0.098)
1 Step down				-0.066 (0.098)	-0.089 (0.107)
1 Step up				-0.089 (0.086)	-0.081 (0.084)
2 Steps up				0.275** (0.135)	0.288** (0.138)
3 Steps up				0.311* (0.163)	0.286 (0.177)
4 Steps up				0.414*** (0.108)	0.311*** (0.108)
5 Steps up				0.124 (0.161)	0.130 (0.154)
6 Steps up				-0.089 (0.199)	-0.197 (0.252)
p	0.627*** (0.101)	0.599*** (0.121)	0.475*** (0.116)	0.515*** (0.064)	0.483*** (0.068)
Log likelihood	-52056.414	-51999.934	-51447.841	-51975.353	-51423.802
AIC	104198.828	104089.869	103013.683	104062.706	102985.604
N	19,152	19,152	19,152	19,152	19,152

Note: All notes as in Table 3.

The effect of class mobility, entered linearly in columns 1 to 3, is of similar magnitude as for Western Europe. However, the coefficient is estimated more precisely and significant at the 1% level. Our non-linear estimates in columns 4 and 5 yield a striking inverted U-shaped pattern for upward mobility. We find large positive effects for moving up by two to four steps, yet a slightly negative (albeit insignificant) effect for moving up by six steps. Although this does not corroborate Sorokin's dissociative hypothesis, it nevertheless shows that long-range upward mobility in Eastern Europe is less beneficial than medium-range mobility.

Concerning downward mobility, we find a significant effect of -0.464 for moving down by six steps. Importantly, since we have -0.573*(1-0.515)=-0.278 for the origin effect of Class VII (column 4), those who move down from Class I to Class VII are expected to be less satisfied than those who started in Class VII and remained there.

Overall, in terms of positional effects, we find slightly smaller coefficients than for Western Europe. Typically, p is larger than 0.5, indicating that destination effects are greater in magnitude than origin effects. Parameter p only falls below 0.5 when income and unemployment are entered, indicating that these mediate destination effects more than origin effects.

Table 7 provides reference effect estimates. Given our reasoning in sections 3.4 and 4.1., we take columns 2 and 4 as the preferred specifications. Surprisingly, the coefficients for reference position are significant in columns 1 and 3 at the 10%-level. However, since a positive sign implies that higher reference class positions yield higher life satisfaction, this contradicts **H4a**.

Our estimate of reference mobility is consistently significant at the 1%-level. Importantly, the size of the mobility effect is roughly three times the effect for Western Europe (-0.176 *versus* -0.064). We thus find support for **H4b** and **H7**. As becomes evident in columns (3) and (4), cases where upward rather than downward reference mobility dominates drive the effect. This contrasts the result for Western Europe and may be driven by those in higher destination classes feeling devalued by an influx of those from lower origins in an otherwise less socially fluid environment.

The effect of own class mobility is larger upon entering the reference mobility term (compare tables 6 and 7). Both reference and own mobility are positively correlated with class destination. The negative reference mobility effect may therefore partly confound the positive effect of own mobility. However, the own mobility estimate is still smaller than the reference mobility coefficient (-0.141 *versus* -0.176). Hence, if every member of the reference group were upwardly mobile to the same degree, no one would benefit from such mobility. This conclusion on the relativity of mobility

effects stands in line with previous findings which showed that life satisfaction gains from income are largely relative (e.g McBride, 2001).

Table 7. Reference estimates for Eastern Europe (DRM)

	(1) Basic controls	(2) Controls for educ., further demographics., UE & income	(3) Basic controls; positive. vs. negative ref. mobility	(4) Controls for educ., further demographics., UE & income; pos. vs. neg. ref. mobility
Class II	-0.199 (0.134)	-0.153 (0.178)	-0.215* (0.124)	-0.143 (0.167)
Class III	-0.748*** (0.180)	-0.563*** (0.155)	-0.767*** (0.172)	-0.584*** (0.164)
Class IV	-0.917*** (0.175)	-0.622*** (0.184)	-0.980*** (0.220)	-0.658*** (0.190)
Class V	-1.175*** (0.303)	-0.843** (0.342)	-1.254*** (0.305)	-0.872*** (0.334)
Class VI	-1.576*** (0.203)	-1.098*** (0.269)	-1.604*** (0.201)	-1.103*** (0.259)
Class VII	-1.545*** (0.260)	-0.973*** (0.244)	-1.500*** (0.231)	-0.921*** (0.210)
Steps mobile	0.157*** (0.040)	0.141*** (0.043)	0.131*** (0.042)	0.120*** (0.036)
Ref. position	0.110* (0.062)	0.032 (0.060)	0.112* (0.059)	0.034 (0.057)
Ref. mobility	-0.203*** (0.051)	-0.176*** (0.047)		
Ref. mob. up			-0.253*** (0.090)	-0.214*** (0.075)
Ref. mob. down			0.122 (0.104)	0.110 (0.098)
p	0.538*** (0.108)	0.358** (0.171)	0.626*** (0.126)	0.468*** (0.129)
Log likelihood	-52036.254	-51435.163	-52034.249	-51433.651
AIC	104162.509	102990.326	104160.498	102991.303
N	19,152	19,152	19,152	19,152

Note: All notes as in Table 3.

4.3 Robustness

Since the DRM of equation (2) is a constrained version of the general model given in equation (1), we may be worried about the degree to which the DRM deviates from an unconstrained OLS model. To assess this, we follow two recent papers by Fosse and Winship (2019b; 2019a) that were written in the context of age-period-cohort models.[13] Borrowing from these and an extended abstract by Fosse and Pfeffer (2019), we show in the appendix how an OLS model without mobility terms already includes mobility effects in the coefficients on destination and origin (informally we already used this point in our interpretation of tables 2 and 5). This analysis also allows us to distinguish between linear trend effects of destination, origin and mobility on the one hand, and non-linear deviations on the other. We then investigate how far the estimated linear trends of the DRM are inconsistent with the implied estimates of an unconstrained OLS regression. In all

[13] These papers are largely concerned with outlining an altogether different empirical strategy, which relies on finding bounds for age, period and cohort effects on the basis of *a priori* beliefs. In general, such a strategy could also be applicable to the case of estimating the effects of origin, destination and mobility. However, pursuing this strategy here would be beyond the scope of the present paper.

specifications, the DRM model deviates only marginally from the OLS model. The largest deviation occurs for the result of table B6 column 2, where the constrained DRM estimate for origin minus mobility is 0.004 points smaller than the implied unconstrained OLS estimate.[14] See tables C1-C6 and related explanations in appendix C.

We also examine to what extent our results from the DRM depend on constraining the deviation of each origin and destination class from an overall trend to be equal. Using the preferred specifications of columns 2 and 4 of tables 3 and 6 as our baseline, we do so by sequentially relaxing this constraint for each class position. See tables C7-C10 and the related explanations in appendix C. Our results are largely robust to this exercise. However, when relaxing the constraint on Class III we obtain a significant positive mobility effect for Western Europe. Hence, our rejection of **H3b** from section 4.2 remains doubtful.

Regarding the robustness of our reference position and mobility estimates, we first run DRMs with non-linear own mobility terms rather than the linear specifications presented in tables 4 and 7. To investigate whether our results depend on DRMs in particular, we also run OLS specifications. Moreover, as an alternative to the asymmetric specifications of columns 3 and 4 in tables 4 and 7, we run models with separate terms for the mean number of classes travelled by the downwardly and upwardly mobile in the reference group. Our main results are robust to each of these tests. See appendix tables C11-C16 for full results.

Furthermore, our reference mobility measure weighs each difference in mobility levels equally. The robustness test of tables C15-C16 already relaxed this equal-weight assumption by allowing for different effect sizes for reference upward and downward mobility. To gauge the sensitivity of our results to this assumption, tables C17-C18 present results where we define alternative measures of upward (downward) reference mobility as the proportion of individuals in the reference group that

[14] In most other cases, these deviations are even smaller. This suggests that there may be a tighter relationship between the DRM and OLS model that we are unable to discern. If so, future research would strongly benefit from analysing such relationships.

moved up (down) by at least 1, 2,…, or 6 steps. For Western Europe, these alternative downward reference mobility measures all yield positive coefficients. For measures with proportions of people that are upwardly mobile by 4 to 6 steps, we find negative coefficients. However, for proportions of people that are upwardly mobile by 1-3 steps, we find positive coefficients, which may explain the insignificance of the coefficients in table 5. For Eastern Europe, all of our alternative reference mobility specifications have the expected sign.

With respect to reference position, we perform a similar robustness test. See appendix tables C15-C16. In columns 1 and 2, we define alternative reference position measures as the share of individuals in the reference group with destination I *versus* VII, and the combined share of individuals in the reference group occupying destinations I+II *versus* VI+VII. Coefficient signs in all of these specifications are in line with the main results (indicating, contrary to **H4a**, a positive reference position effect), but remain insignificant.

It may be that reference mobility effects also run via own mobility, i.e. that higher (lower) levels of reference mobility dampen any positive (negative) effects of own upward (downward) mobility. Although interaction terms have the expected negative sign, none are statistically significant. See appendix tables C19-C20.

Finally, from table A1 we observe that relatively more countries in Western Europe are observed in ESS waves 1-3 than Eastern European countries. Our conclusions concerning **H5**-**H7** on differences across Europe may therefore be driven by a period effect. To assess this concern, we re-estimate tables 2-7, but restrict the observation window to waves 4-5 for both regions (see tables C21-C26). For Eastern Europe we obtain slightly smaller and sometimes marginally insignificant mobility estimates. Regarding Western Europe, we generally find smaller coefficients on origin, mobility and reference mobility. In particular, in the case of column 2 of table C23 (corresponding to the original table 4), our estimate of reference mobility is no longer significant, and for columns 2-4 parameter p is constrained to 1, implying a zero origin effect. Overall, we thus find a larger difference between Eastern and Western Europe for this restricted time period.

5 Conclusion

This paper makes three contributions: First, we develop a theoretical framework bringing together all the pathways through which social class may impact life satisfaction. Second, we consider how reference class position and mobility affect life satisfaction. Third, we comparatively assess the relationship between social class and life satisfaction across two major European regions.

In terms of our hypotheses, our results can be summarised as follows. Destinations and origins structure life satisfaction in both East and West (**H1** and **H2** are confirmed). Own upward mobility improves life satisfaction and downward mobility reduces life satisfaction (**H3b** is confirmed and **H3a** is rejected). Contrastingly, higher reference mobility leads to declines in life satisfaction (corroborating **H4b**). For these hypotheses, our evidence is much stronger for Eastern than for Western Europe.

Our finding that there are own intergenerational mobility effects at all stands in contrast to previous studies that use diagonal reference models such as Marshall and Firth (1999) or Zhao et al. (2017). Furthermore, regarding our rejection of **H4a** and confirmation of **H4b**, we note that individuals are mostly in contact with other people that share the same destination class (i.e. colleagues and friends). Therefore, observing the class mobility of one's direct peers may be easier and thus appear more salient than the outflux from one's origin class into other destinations. Indeed, in both European regions, the magnitude of the negative reference mobility effect is similar or even larger than the magnitude of the positive own class mobility effect. Hence, while mobility may increase an *individual*'s satisfaction with her life, there is no *societal* benefit to class mobility per se.

Reference mobility effects are much stronger in Eastern Europe, confirming **H7**. We also find some support that own mobility effects are stronger in the East, endorsing **H6**. Surprisingly, destination and origin effects are stronger in the West. **H5** is therefore rejected. Since hypotheses **H5**-**H7** depended on macro-level differences between Eastern and Western Europe, future research may attempt to further tests these hypotheses using multi-level models.

Several further limitations of this paper may also be usefully addressed in future research. First, even though our results largely remain robust to several tests, the DRM rests on strong assumptions. Supplementing DRMs with other strategies to separate origin, destination and mobility effects may therefore be informative. Adopting the bounding approach of Fosse and Winship (2019b) could be one such avenue.

Second, irrespective of using DRMs, our estimated mobility and destination effects could be contaminated by unobserved heterogeneity. It would hence be useful to either extend the present analysis using fixed effects panel regressions (although analysing time-invariant *inter*generational mobility is then not possible) or using cross-sectional data with detailed information on personality traits (Boyce 2010), which may alleviate problems of selection into mobility. Yet, unless clear exogenous variation is available for *both* origin and destination, concerns regarding selection may persist.

Third, we showed only informally how origin effects are mediated by education and destination, and how both effects are potentially mediated by income and unemployment. Extending this by means of a formal mediation analysis could further elucidate all the mechanisms by which social class affects life satisfaction.

Despite these limitations, we nevertheless conclude that in Europe social class matters greatly to life satisfaction and in more complex ways than previously considered. Not only does current class position enter into such evaluations. For being satisfied with one's life, it also matters in which class one was raised, whether one has been intergenerationally mobile, and where those with whom one now shares a class destination originate from.

References

Alderson, Arthur S., and Tally Katz-Gerro. 2016. 'Compared to Whom? Inequality, Social Comparison, and Happiness in the United States'. *Social Forces* 95 (1): 25–54.

Anderson, Cameron, Michael W. Kraus, Adam D. Galinsky, and Dacher Keltner. 2012. 'The Local-Ladder Effect: Social Status and Subjective Well-Being'. *Psychological Science* 23 (7): 764–71. https://doi.org/10.1177/0956797611434537.

Bernardi, Fabrizio, and Gabriele Ballarino. 2016. *Education, Occupation and Social Origin*. Edward Elgar Publishing.

Bessudnov, Alexey, Martin McKee, and David Stuckler. 2012. 'Inequalities in Male Mortality by Occupational Class, Perceived Status and Education in Russia, 1994–2006'. *European Journal of Public Health* 22 (3): 332–37. https://doi.org/10.1093/eurpub/ckr130.

Blalock, Hubert M. 1966. 'The Identification Problem and Theory Building: The Case of Status Inconsistency'. *American Sociological Review* 31 (1): 52–61.

Blasius, Jörg, and Victor Thiessen. 2018. 'Perceived Corruption, Trust, and Interviewer Behavior in 26 European Countries'. *Sociological Methods & Research*, July. https://doi.org/10.1177/0049124118782554.

Boyce, Christopher J. 2010. 'Understanding Fixed Effects in Human Well-Being'. *Journal of Economic Psychology* 31 (1): 1–16. https://doi.org/10.1016/j.joep.2009.08.009.

Breen, Richard, ed. 2004. *Social Mobility in Europe*. Oxford ; New York: Oxford University Press.

Breen, Richard, and John Goldthorpe. 1997. 'Explaining Educational Differentials: Towards a Formal Rational Action Theory'. *Rationality and Society* 9 (3): 275–305.

Bubonya, Melisa, Deborah A. Cobb-Clark, and Mark Wooden. 2017. 'Mental Health and Productivity at Work: Does What You Do Matter?' *Labour Economics* 46 (June): 150–65. https://doi.org/10.1016/j.labeco.2017.05.001.

Bukodi, Erzsébet, and John Goldthorpe. 2016. 'Educational Attainment-Relative or Absolute-as a Mediator of Intergenerational Class Mobility in Britain'. *Research in Social Stratification and Mobility* 43: 5–15.

———. 2019. *Social Mobility and Education in Britain: Research, Politics and Policy*. Cambridge University Press.

Bukodi, Erzsébet, Marii Paskov, and Brian Nolan. 2019. 'Intergenerational Class Mobility in Europe: A New Account'. *Social Forces*, 1–32. https://doi.org/10.1093/sf/soz026.

Chan, Tak-Wing. 2018. 'Social Mobility and the Well-Being of Individuals'. *The British Journal of Sociology* 69 (1): 183–206. https://doi.org/10.1111/1468-4446.12285.

Clark, Andrew. 2003. 'Unemployment as a Social Norm: Psychological Evidence from Panel Data'. *Journal of Labor Economics* 21 (2): 323–351. https://doi.org/10.1086/345560.

Clark, Andrew, and Emanuela D'Angelo. 2015. 'Upward Social Mobility, Well-Being and Political Preferences: Evidence from the BHPS'. *Unpublished Working Paper*.

Clark, Andrew, and Claudia Senik. 2010. 'Who Compares To Whom ? The Anatomy of Income Comparisons in Europe'. *The Economic Journal* 120: 573–594. https://doi.org/10.1111/j.1468-0297.2010.02359.x.

Connolly, Filip Fors, and Ingemar Johansson Sevä. 2018. 'Social Status and Life Satisfaction in Context: A Comparison between Sweden and the USA'. *International Journal of Wellbeing* 8 (2): 110–134.

Delhey, Jan, and Ulrich Kohler. 2006. 'From Nationally Bounded to Pan-European Inequalities? On the Importance of Foreign Countries as Reference Groups'. *European Sociological Review* 22 (2): 125–140. https://doi.org/10.1093/esr/jci047.

Deurzen, Ioana van, Erik van Ingen, and Wim J. H. van Oorschot. 2015. 'Income Inequality and Depression: The Role of Social Comparisons and Coping Resources'. *European Sociological Review* 31 (4): 477–89. https://doi.org/10.1093/esr/jcv007.

Diener, Ed, Ronald Inglehart, and Louis Tay. 2013. 'Theory and Validity of Life Satisfaction Scales'. *Social Indicators Research* 112 (3): 497–527. https://doi.org/10.1007/s11205-012-0076.

Dolan, Paul, and Grace Lordan. 2013. 'Moving Up and Sliding Down: An Empirical Assessment of the Effect of Social Mobility on Subjective Wellbeing'. *CEP Discussion Paper*.

Erikson, Robert. 1984. 'Social Class of Men, Women and Families'. *Sociology* 18 (4): 500–514. https://doi.org/10.1177/0038038584018004003.

Erikson, Robert, and John Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford University Press, USA.

Erikson, Robert, John Goldthorpe, and Lucienne Portocarero. 1979. 'Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden'. *The British Journal of Sociology* 30 (4): 415–441.

Ermisch, John, Marco Francesconi, and David J. Pevalin. 2004. 'Parental Partnership and Joblessness in Childhood and Their Influence on Young People's Outcomes'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 167 (1): 69–101.

Ferrer-i-Carbonell, Ada. 2005. 'Income and Well-Being: An Empirical Analysis of the Comparison Income Effect'. *Journal of Public Economics* 89 (5–6): 997–1019. https://doi.org/10.1016/j.jpubeco.2004.06.003.

Festinger, Leon. 1954. 'A Theory of Social Comparison Processes'. *Human Relations* 7 (2): 117–40. https://doi.org/10.1177/001872675400700202.

Fosse, Ethan, and Fabian T Pfeffer. 2019. 'Beyond the Diagonal Reference Model: Critiques and New Directions'. *Unpublished Abstract*.

Fosse, Ethan, and Christopher Winship. 2019a. 'Analyzing Age-Period-Cohort Data: A Review and Critique'. *Annual Review of Sociology* 45 (1): 467–92. https://doi.org/10.1146/annurev-soc-073018-022616.

———. 2019b. 'Bounding Analyses of Age-Period-Cohort Effects'. *Demography* 56 (5): 1975–2004. https://doi.org/10.1007/s13524-019-00801-6.

Ganzeboom, Harry. 2014. 'ESS Developmental Project'. 2014. http://www.harryganzeboom.nl/ESS-DEVO/index.htm.

Gerber, Theodore P., and Michael Hout. 1998. 'More Shock than Therapy: Market Transition, Employment, and Income in Russia, 1991-1995'. *American Journal of Sociology* 104 (1): 1–50. https://doi.org/10.1086/210001.

Goerke, Laszlo, and Markus Pannenberg. 2015. 'Direct Evidence for Income Comparisons and Subjective Well-Being across Reference Groups'. *Economics Letters* 137: 95–101. https://doi.org/10.1016/j.econlet.2015.10.012.

Goldthorpe, John. 2007. *On Sociology*. Vol. 2. Stanford University Press.

Goldthorpe, John, and Abigail McKnight. 2006. 'The Economic Basis of Social Class'. In *Mobility and Inequality Frontiers of Research in Sociology and Economics*, edited by David Grusky and Paula England. Stanford University Press.

Grosfeld, Irena, and Claudia Senik. 2010. 'The Emerging Aversion to Inequality'. *Economics of Transition* 18 (1): 1–26. https://doi.org/10.1111/j.1468-0351.2009.00376.x.

Gugushvili, Alexi. 2016. 'Intergenerational Social Mobility and Popular Explanations of Poverty: A Comparative Perspective'. *Social Justice Research* 29 (4): 402–28. https://doi.org/10.1007/s11211-016-0275-9.

Hadjar, Andreas, and Robin Samuel. 2015. 'Does Upward Social Mobility Increase Life Satisfaction? A Longitudinal Analysis Using British and Swiss Panel Data'. *Research in Social Stratification and Mobility* 39 (March): 48–58. https://doi.org/10.1016/j.rssm.2014.12.002.

Haller, Max, and Markus Hadler. 2006. 'How Social Relations and Structures Can Produce Happiness and Unhappiness: An International Comparative Analysis'. *Social Indicators Research* 75 (2): 169–216. https://doi.org/10.1007/s11205-004-6297-y.

Heggebø, Kristian, and Jon Ivar Elstad. 2018. 'Is It Easier to Be Unemployed When the Experience Is More Widely Shared? Effects of Unemployment on Self-Rated Health in 25 European Countries with Diverging Macroeconomic Conditions'. *European Sociological Review* 34 (1): 22–39. https://doi.org/10.1093/esr/jcx080.

Hendrickx, John, Nan Dirk De Graaf, Jan Lammers, and Wout Ultee. 1993. 'Models for Status Inconsistency and Mobility: A Comparison of the Approaches by Hope and Sobel with the

Mainstream Square Additive Model'. *Quality and Quantity* 27 (4): 335–52. https://doi.org/10.1007/BF01102497.

Houle, Jason N. 2011. 'The Psychological Impact of Intragenerational Social Class Mobility'. *Social Science Research* 40 (3): 757–72. https://doi.org/10.1016/j.ssresearch.2010.11.008.

Houle, Jason N., and Molly A. Martin. 2011. 'Does Intergenerational Mobility Shape Psychological Distress? Sorokin Revisited'. *Research in Social Stratification and Mobility* 29 (2): 193–203. https://doi.org/10.1016/j.rssm.2010.11.001.

Hout, Michael. 2004. 'Getting the Most out of the GSS Income Measures'.

Iveson, Matthew H., and Ian J. Deary. 2017. 'Intergenerational Social Mobility and Subjective Wellbeing in Later Life'. *Social Science & Medicine* 188 (September): 11–20. https://doi.org/10.1016/j.socscimed.2017.06.038.

Kahneman, Daniel, and Angus Deaton. 2010. 'High Income Improves Evaluation of Life but Not Emotional Well-Being.' *PNAS* 107 (38): 16489–93. https://doi.org/10.1073/pnas.1011492107.

Kelley, Jonathan, and Krzysztof Zagorski. 2004. 'Economic Change and the Legitimation of Inequality: The Transition from Socialism to the Free Market in Central-East Europe'. *Research in Social Stratification and Mobility*, The Shape of Social Inequality, 22 (January): 319–64. https://doi.org/10.1016/S0276-5624(04)22011-X.

Kuitto, Kati. 2018. 'Measuring Welfare Entitlement Generosity in Transitional Welfare States: The Case of Post-Communist Countries in Central and Eastern Europe'. *Social Indicators Research* 136 (1): 203–24. https://doi.org/10.1007/s11205-016-1520-1.

Lipps, Oliver, and Daniel Oesch. 2018. 'The Working Class Left behind? The Class Gap in Life Satisfaction in Germany and Switzerland over the Last Decades'. *European Societies* 20 (4): 549–71. https://doi.org/10.1080/14616696.2018.1448106.

Lucchini, Mario, and Antonio Schizzerotto. 2010. 'Unemployment Risks in Four EU Countries: A Validation Study of the ESeC'. In *Social Class in Europe An Introduction to the European Socio-Economic Classification*, edited by David Rose and Eric Harrison. Routledge.

Luechinger, Simon, Stephan Meier, and Alois Stutzer. 2010. 'Why Does Unemployment Hurt the Employed? Evidence from the Life Satisfaction Gap Between the Public and the Private Sector'. *Journal of Human Resources* 45 (4): 998–1045. https://doi.org/10.3368/jhr.45.4.998.

Marshall, Gordon, and David Firth. 1999. 'Social Mobility and Personal Satisfaction: Evidence from Ten Countries'. *The British Journal of Sociology* 50 (1): 28–48. https://doi.org/10.1111/j.1468-4446.1999.00028.x.

McBride, M. 2001. 'Relative-Income Effects on Subjective Well-Being in the Cross-Section'. *Journal of Economic Behavior & Organization* 45 (3): 251–278. https://doi.org/10.1016/S0167-2681(01)00145-7.

Melchior, Maria, Nancy Krieger, Ichiro Kawachi, Lisa F. Berkman, Isabelle Niedhammer, and Marcel Goldberg. 2005. 'Work Factors and Occupational Class Disparities in Sickness Absence: Findings From the GAZEL Cohort Study'. *American Journal of Public Health* 95 (7): 1206–12. https://doi.org/10.2105/AJPH.2004.048835.

Merton, R., and A. Rossi. 1968. 'Contributions to the Theory of Reference Group Behavior'. In *Social Theory and Social Structure.*, edited by R. Merton. Macmillan.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. 'How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It'. *American Journal of Political Science* 62 (3): 760–75. https://doi.org/10.1111/ajps.12357.

Newman, Katherine S. 1999. *Falling from Grace: Downward Mobility in the Age of Affluence*. University of California Press.

Nikolaev, Boris, and Ainslee Burns. 2014. 'Intergenerational Mobility and Subjective Well-Being-Evidence from the General Social Survey'. *Journal of Behavioral and Experimental Economics* 53: 82–96. https://doi.org/10.1016/j.socec.2014.08.005.

Nolan, Brian, and Christopher T. Whelan. 2011. 'Poverty and Deprivation in Europe'. *OUP Catalogue*.

Paskov, Marii, and Erzsébet Bukodi. 2018. 'Income Inequality, Living Standards and Intergenerational Social Mobility'. In *Generating Prosperity for Working Families in Affluent Countries*, edited by Brian Nolan. Oxford University Press.

Pavot, William, and Ed Diener. 2008. 'The Satisfaction With Life Scale and the Emerging Construct of Life Satisfaction'. *The Journal of Positive Psychology* 3 (2): 137–52. https://doi.org/10.1080/17439760701756946.

Präg, Patrick, Melinda Mills, and Rafael Wittek. 2014. 'Income and Income Inequality as Social Determinants of Health: Do Social Comparisons Play a Role?' *European Sociological Review* 30 (2): 218–29. https://doi.org/10.1093/esr/jct035.

Rose, David, Eric Harrison, and D. J. Pevalin. 2014. 'The European Socio-Economic Classification: A Prolegomenon'. In *Social Class in Europe*, edited by David Rose and Eric Harrison. Routledge.

Rowlingson, K., and S. McKay. 2002. *Lone Parent Families: Gender, Class, and State*. New York: Prentice Hall.

Schuck, Bettina, and Nadia Steiber. 2017. 'Does Intergenerational Educational Mobility Shape the Well-Being of Young Europeans? Evidence from the European Social Survey'. *Social Indicators Research* 139 (3): 1237–1255. https://doi.org/10.1007/s11205-017-1753-7.

Shavit, Yossi, and Hans-Peter Blossfeld. 1993. *Persistent Inequality*. Westview Press.

Slomczyński, Kazimierz M., and Goldie Shabad. 1996. 'Systemic Transformation and the Salience of Class Structure in East Central Europe'. *East European Politics and Societies* 11 (1): 155–189.

Sobel, Michael E. 1981. 'Diagonal Mobility Models: A Substantively Motivated Class of Designs for the Analysis of Mobility Effects'. *American Sociological Review* 46 (6): 893–906. https://doi.org/10.2307/2095086.

———. 1985. 'Social Mobility and Fertility Revisited: Some New Models for the Analysis of the Mobility Effects Hypothesis'. *American Sociological Review* 50 (5): 699–712. https://doi.org/10.2307/2095383.

Solt, Frederick. 2019. 'Measuring Income Inequality Across Countries and Over Time: The Standardized World Income Inequality Database.' SWIID Version 8.0.

Sorokin, Pitirim A. 1959. *Social and Cultural Mobility*. The Free Press.

Stouffer, S.A. 1949. *The American Soldier*. John Wiley and Sons, Inc.

Vågerö, Denny. 2010. 'The East–West Health Divide in Europe: Growing and Shifting Eastwards'. *European Review* 18 (1): 23–34. https://doi.org/10.1017/S106279870999010X.

Vendrik, Maarten C. M., and Geert B. Woltjer. 2007. 'Happiness and Loss Aversion: Is Utility Concave or Convex in Relative Income?' *Journal of Public Economics* 91 (7–8): 1423–1448. https://doi.org/10.1016/j.jpubeco.2007.02.008.

White, Chris, Myer Glickman, Brian Johnson, and Tania Corbin. 2007. 'Social Inequalities in Adult Male Mortality by the National Statistics Socio-Economic Classification, England and Wales, 2001-03'. *Health Statistics Quarterly*, no. 36: 6–23.

Wulfgramm, Melike. 2014. 'Life Satisfaction Effects of Unemployment in Europe: The Moderating Influence of Labour Market Policy'. *Journal of European Social Policy*, no. 3: 258–272. https://doi.org/10.1177/0958928714525817.

Zang, Emma, and Nan Dirk de Graaf. 2016. 'Frustrated Achievers or Satisfied Losers? Inter- and Intragenerational Social Mobility and Happiness in China'. *Sociological Science* 3: 779–800. https://doi.org/10.15195/v3.a33.

Zhao, Yizhang, Yaojun Li, Anthony Heath, and Nick Shryane. 2017. 'Inter- and Intra-Generational Social Mobility Effects on Subjective Well-Being – Evidence from Mainland China'. *Research in Social Stratification and Mobility* 48 (April): 54–66. https://doi.org/10.1016/j.rssm.2017.02.002.

Appendix

A Descriptives

Table A1. List of countries and observed years

ESS wave (year)	1 (2002)	2 (2004)	3 (2006)	4 (2008)	5 (2010)
			Western Europe		
Austria	x	x	x	x	
Belgium	x	x	x	x	x
Cyprus			x		x
Denmark	x	x	x	x	x
Finland	x		x	x	x
France		x	x	x	x
Germany	x	x	x	x	x
Greece	x	x		x	x
Iceland		x			
Ireland		x	x	x	x
Italy	x	x			
Luxemburg	x	x			
Netherlands	x	x	x	x	x
Norway	x	x	x	x	x
Portugal	x	x	x	x	
Spain	x	x	x	x	x
Sweden	x	x	x	x	x
Switzerland	x	x	x	x	x
United Kingdom	x	x	x	x	x
			Eastern Europe		
Bulgaria			x		x
Croatia				x	x
Czech Republic	x	x		x	x
Estonia				x	x
Hungary		x		x	x
Latvia			x	x	
Lithuania				x	x
Poland	x	x	x	x	x
Romania				x	
Russia			x	x	x
Slovakia		x	x		x
Slovenia	x		x	x	x
Ukraine				x	x

Table A2. Number of observations and shares of ESeC class origin and destination combinations, Western Europe

		Destination									
		I	II	III	IV	V	VI	VII	SEI	SEII	Total
Origin	I	1711 (2.9%)	2010 (3.3%)	659 (1.2%)	432 (0.8%)	381 (0.7%)	147 (0.3%)	324 (0.6%)	435 (0.9%)	19 (0.0%)	6118 (10.6%)
	II	1830 (3.0%)	3288 (5.2%)	1091 (1.9%)	792 (1.6%)	817 (1.4%)	344 (0.6%)	756 (1.5%)	663 (1.3%)	55 (0.1%)	9636 (16.5%)
	III	588 (1.1%)	1064 (1.8%)	529 (1.0%)	299 (0.6%)	357 (0.6%)	168 (0.4%)	288 (0.6%)	262 (0.5%)	13 (0.0%)	3568 (6.6%)
	IV	507 (0.9%)	1188 (2.2%)	521 (1.0%)	679 (1.4%)	539 (1.0%)	312 (0.6%)	587 (1.2%)	352 (0.8%)	15 (0.0%)	4700 (9.2%)
	V	304 (0.5%)	751 (1.2%)	410 (0.7%)	365 (0.7%)	443 (0.8%)	247 (0.4%)	457 (1.0%)	205 (0.4%)	13 (0.0%)	3195 (5.7%)
	VI	575 (0.9%)	1382 (2.5%)	858 (1.6%)	934 (2.0%)	980 (1.9%)	1045 (2.0%)	1703 (3.7%)	516 (1.0%)	43 (0.1%)	8036 (15.8%)
	VII	517 (0.9%)	1283 (2.1%)	761 (1.4%)	853 (1.7%)	1080 (1.9%)	931 (1.7%)	2255 (4.6%)	586 (1.2%)	40 (0.1%)	8306 (15.6%)
	SEI	922 (1.4%)	1661 (2.5%)	794 (1.3%)	672 (1.0%)	742 (1.2%)	447 (0.8%)	916 (1.7%)	1071 (1.8%)	107 (0.2%)	7332 (11.9%)
	SEII	538 (0.6%)	1251 (1.2%)	523 (0.6%)	597 (0.7%)	684 (0.7%)	688 (0.8%)	1294 (1.7%)	646 (0.9%)	756 (0.9%)	6977 (8.1%)
	Total	7492 (12.0%)	13878 (22.1%)	6146 (10.8%)	5623 (10.6%)	6023 (10.2%)	4329 (7.6%)	8580 (16.7%)	4736 (8.8%)	1061 (1.4%)	57,868 (100%)

Note: Design and population weights are applied to shares.

Table A3. Number of observations and shares of ESeC class origin and destination combinations, Eastern Europe

		Destination									
		I	II	III	IV	V	VI	VII	SEI	SEII	Total
Origin	I	541 (3.2%)	712 (3.5%)	231 (1.0%)	168 (1.3%)	203 (1.3%)	150 (0.8%)	275 (1.7%)	163 (0.7%)	8 (0.0%)	2451 (13.4%)
	II	518 (2.7%)	986 (4.3%)	302 (1.1%)	268 (1.2%)	338 (1.5%)	392 (2.2%)	572 (2.6%)	195 (0.8%)	20 (0.1%)	3591 (16.5%)
	III	105 (0.3%)	212 (0.6%)	97 (0.2%)	59 (0.2%)	74 (0.2%)	89 (0.2%)	119 (0.4%)	60 (0.2%)	2 (0.0%)	817 (2.5%)
	IV	165 (0.8%)	337 (1.3%)	119 (0.5%)	203 (0.9%)	138 (0.6%)	222 (1.0%)	291 (1.1%)	83 (0.4%)	12 (0.0%)	1570 (6.6%)
	V	108 (0.6%)	259 (1.0%)	133 (0.7%)	105 (0.6%)	189 (0.9%)	226 (0.9%)	352 (1.6%)	77 (0.2%)	7 (0.0%)	1456 (6.7%)
	VI	316 (1.4%)	730 (3.0%)	345 (1.5%)	342 (1.1%)	550 (2.2%)	1001 (3.9%)	1503 (5.9%)	226 (0.7%)	29 (0.1%)	5042 (19.9%)
	VII	281 (1.3%)	687 (2.9%)	330 (1.4%)	342 (1.8%)	592 (3.0%)	1000 (4.9%)	1875 (8.0%)	193 (0.8%)	37 (0.1%)	5337 (24.3%)
	SEI	53 (0.2%)	94 (0.3%)	40 (0.1%)	54 (0.2%)	31 (0.1%)	86 (0.3%)	104 (0.4%)	39 (0.1%)	5 (0.0%)	506 (1.8%)
	SEII	82 (0.4%)	165 (0.9%)	94 (0.5%)	103 (0.6%)	123 (0.7%)	246 (1.3%)	422 (2.1%)	66 (0.4%)	237 (1.7%)	1538 (8.4%)
	Total	2169 (10.9%)	4182 (18.0%)	1691 (7.0%)	1644 (7.9%)	2238 (10.4%)	3412 (15.4%)	5513 (23.8%)	1102 (4.4%)	357 (2.1%)	22,308 (100%)

Note: Design and population weights are applied to shares.

Table A4. Variable descriptions

Continuos variables	Mean	SD	Min-Max
Life satisfaction	6.41	2.42	0
Own mobility	0.30	2.48	-6-6
Reference position	3.09	1.85	0-7
Reference mobility	0.22	1.31	-5.02 -5.13
Age	49.13	8.57	35-65
Age squared	2,487.69	853.75	1,225-4,225
Years of education	12.71	3.70	0-56
Years of education squared	175.32	104.25	0-3,136
Household size	2.99	1.35	1-15
Household size squared	10.75	10.20	1-255
Ln real HH income	10.62	1.42	6.15-18.56

Categorical variables	Category	N	Share **(%)**
Unemployed	No	74,905	93.63
	Yes	5,271	6.37
Gender	Female	38,445	49.28
	Male	41,731	50.72
Ethnic minority	No	75,737	93.01
	Yes	3,578	6.18
	Missing/No answers	861	0.81
Born in country	No	73511	91.82
	Yes	6585	8.14
	Missing/No answers	80	0.04
Religious	No	32,139	40.37
	Yes	46,795	58.75
	Missing/No answers	1242	0.89
Married/cohabitation	No	26,196	28.59
	Yes	52,211	70.73
	Missing/No answers	1,769	0.67
Place of domicile	Big city	14,807	19.81
	Suburbs	10,109	11.37
	Town	24,742	34.59
	Country/Village	24,599	30.25
ESS round	2002	5,776	3.86
	2004	12,824	13.97
	2006	15,286	16.26
	2008	15,061	21.06
	2010	18,661	24.45
NUTS I region	102 fixed effects		

Note: Overall N=80,176. Mean, SD and shares are computed using population and design weights. Ns are

B Main results with coefficients on controls

Table B1. Origin and destination estimates for Western Europe (OLS), all coefficients

	(1) Basic model	(2) Class destination added	(3) Controls for education	(4) Controls for educ. & further demographics	(5) Controls for educ., further demographics, UE & income
Orig.=II	-0.044 (0.046)	0.021 (0.046)	0.032 (0.047)	0.023 (0.047)	0.041 (0.046)
Orig.=III	-0.110* (0.064)	-0.025 (0.062)	-0.002 (0.063)	0.004 (0.061)	0.020 (0.061)
Orig.=IV	-0.244*** (0.062)	-0.081 (0.061)	-0.049 (0.062)	-0.085 (0.060)	-0.050 (0.059)
Orig.=V	-0.191*** (0.070)	0.005 (0.068)	0.040 (0.069)	0.022 (0.067)	0.062 (0.066)
Orig.=VI	-0.312*** (0.058)	-0.046 (0.055)	0.005 (0.058)	-0.033 (0.055)	0.039 (0.055)
Orig.=VII	-0.406*** (0.059)	-0.117** (0.056)	-0.057 (0.058)	-0.088 (0.058)	-0.021 (0.056)
Orig.=SEI	-0.123** (0.057)	0.020 (0.056)	0.050 (0.057)	0.020 (0.056)	0.051 (0.055)
Orig.=SEII	-0.107 (0.070)	0.129* (0.069)	0.183*** (0.069)	0.060 (0.068)	0.076 (0.067)
Dest.=II		-0.149*** (0.043)	-0.125*** (0.044)	-0.114*** (0.042)	-0.025 (0.040)
Dest.=III		-0.247*** (0.054)	-0.195*** (0.056)	-0.167*** (0.055)	-0.026 (0.055)
Dest.=IV		-0.529*** (0.064)	-0.436*** (0.067)	-0.418*** (0.065)	-0.218*** (0.067)
Dest.=V		-0.624*** (0.071)	-0.531*** (0.072)	-0.496*** (0.070)	-0.259*** (0.067)
Dest.=VI		-0.773*** (0.072)	-0.644*** (0.077)	-0.600*** (0.075)	-0.297*** (0.073)
Dest.=VII		-0.932*** (0.066)	-0.799*** (0.069)	-0.717*** (0.068)	-0.398*** (0.069)
Dest.=SEI		-0.303*** (0.068)	-0.222*** (0.069)	-0.210*** (0.067)	-0.071 (0.063)
Dest.=SEII		-0.520*** (0.117)	-0.399*** (0.117)	-0.458*** (0.113)	-0.224** (0.113)
ESS=2	-0.072 (0.048)	-0.074 (0.046)	-0.077* (0.046)	-0.068 (0.047)	0.081 (0.050)
ESS=3	0.023 (0.045)	0.007 (0.044)	-0.011 (0.045)	-0.004 (0.043)	0.283*** (0.048)
ESS=4	-0.087** (0.041)	-0.100** (0.040)	-0.121*** (0.041)	-0.099** (0.041)	0.384*** (0.053)
ESS=5	-0.003 (0.051)	-0.012 (0.049)	-0.034 (0.049)	0.006 (0.049)	0.582*** (0.056)
Gender=female	0.054* (0.029)	0.063* (0.032)	0.069** (0.032)	0.077** (0.032)	0.099*** (0.031)
Age	-0.114*** (0.021)	-0.121*** (0.021)	-0.122*** (0.021)	-0.144*** (0.020)	-0.159*** (0.020)
Age-squared	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
Ethnic Minority=Yes	-0.309*** (0.092)	-0.282*** (0.089)	-0.282*** (0.090)	-0.301*** (0.090)	-0.233*** (0.086)
Ethnic Minority=n.a.	-0.689*** (0.259)	-0.596** (0.256)	-0.570** (0.253)	-0.557** (0.253)	-0.484* (0.250)
Born in country=Yes	0.129** (0.064)	0.037 (0.062)	0.037 (0.062)	0.084 (0.061)	0.010 (0.061)
Born in country=n.a.	0.708 (0.669)	0.515 (0.683)	0.507 (0.704)	0.577 (0.744)	0.829 (0.755)
Years of education			0.068*** (0.016)	0.069*** (0.015)	0.049*** (0.015)
Years of education squared			-0.001*** (0.000)	-0.001*** (0.000)	-0.001** (0.000)
Religious=Yes				0.193*** (0.034)	0.190*** (0.033)
Religious=n.a.				-0.042 (0.224)	-0.009 (0.224)
Married/Cohabitating=Yes				0.614*** (0.042)	0.454*** (0.042)
Married/Cohabitating=n.a.				0.118 (0.127)	0.063 (0.103)
Domicile=Suburbs				0.046 (0.049)	0.035 (0.047)
Domicile=Town				0.095** (0.044)	0.097** (0.043)
Domicile=Village				0.219*** (0.045)	0.208*** (0.044)
Domicile=Rural				0.267*** (0.075)	0.275*** (0.076)
Domicile=n.a.				-0.222 (0.401)	-0.081 (0.331)
Household size				0.237*** (0.050)	0.099** (0.048)
Household size squared				-0.024*** (0.007)	-0.009 (0.007)
Unemployed=Yes, base					-0.979*** (0.092)
Log real household income					0.373*** (0.023)
Constant	10.173*** (0.513)	10.655*** (0.505)	9.924*** (0.517)	9.137*** (0.511)	5.648*** (0.570)
Log likelihood	-124485.221	-123899.862	-123842.369	-122815.933	-121736.466
AIC	249156.442	248001.724	247890.737	245859.866	243704.933
N	57,868	57,868	57,868	57,868	57,868

Note: * p<0.10, ** p<0.05, *** p<0.01. 75 NUTS I fixed effects additionally included.

Table B2. Mobility estimates for Western Europe (DRM), all coefficients

	(1) Linear mobility	(2) Linear mobility, controls for education	(3) Linear mobility, controls for educ. & further demographics	(4) Linear mobility, controls for educ., further demographics, UE & income	(5) Non-linear mobility, controls for educ.	(6) Non-linear mobility, controls for educ., further demographics, UE & income
Class=II	-0.160** (0.071)	-0.109 (0.073)	-0.097 (0.068)	0.004 (0.063)	-0.110 (0.073)	0.007 (0.066)
Class=III	-0.266*** (0.082)	-0.167* (0.092)	-0.131 (0.085)	0.028 (0.089)	-0.134 (0.101)	0.061 (0.094)
Class=IV	-0.661*** (0.089)	-0.510*** (0.095)	-0.518*** (0.089)	-0.286*** (0.092)	-0.493*** (0.102)	-0.261** (0.105)
Class=V	-0.663*** (0.100)	-0.506*** (0.102)	-0.480*** (0.100)	-0.203** (0.095)	-0.484*** (0.110)	-0.177* (0.101)
Class=VI	-0.833*** (0.118)	-0.628*** (0.118)	-0.607*** (0.113)	-0.245** (0.108)	-0.611*** (0.113)	-0.228** (0.108)
Class=VII	-1.096*** (0.088)	-0.879*** (0.098)	-0.812*** (0.094)	-0.420*** (0.098)	-0.877*** (0.101)	-0.420*** (0.100)
Steps mobile	0.056 (0.064)	0.059 (0.047)	0.049 (0.038)	0.031 (0.019)		
Steps down=6					-0.396 (0.310)	-0.142 (0.215)
Steps down=5					-0.306 (0.240)	-0.162 (0.150)
Steps down=4					-0.482** (0.218)	-0.323** (0.146)
Steps down=3					-0.149 (0.134)	-0.076 (0.086)
Steps down=2					-0.169 (0.106)	-0.101 (0.080)
Steps down=1					-0.082 (0.065)	-0.048 (0.050)
Steps up=1					0.044 (0.062)	0.012 (0.049)
Steps up=2					0.112 (0.098)	0.055 (0.070)
Steps up=3					0.024 (0.146)	-0.044 (0.094)
Steps up=4					0.229 (0.168)	0.132 (0.097)
Steps up=5					0.233 (0.212)	0.103 (0.114)
Steps up=6					0.449 (0.287)	0.288* (0.157)
ESS=2	-0.083 (0.053)	-0.081 (0.053)	-0.069 (0.053)	0.079 (0.054)	-0.083 (0.053)	0.078 (0.054)
ESS=3	-0.019 (0.051)	-0.037 (0.052)	-0.033 (0.051)	0.246*** (0.054)	-0.039 (0.052)	0.244*** (0.054)
ESS=4	-0.082* (0.046)	-0.102** (0.046)	-0.086* (0.047)	0.379*** (0.058)	-0.103** (0.046)	0.378*** (0.058)
ESS=5	0.006 (0.054)	-0.013 (0.054)	0.033 (0.053)	0.582*** (0.061)	-0.012 (0.054)	0.582*** (0.061)
Gender=female	0.038 (0.039)	0.046 (0.038)	0.051 (0.038)	0.076** (0.036)	0.043 (0.038)	0.073** (0.036)
Age	-0.139*** (0.024)	-0.138*** (0.024)	-0.154*** (0.023)	-0.169*** (0.022)	-0.138*** (0.024)	-0.169*** (0.022)
Age-squared	0.001*** (0.000)	0.001*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.001*** (0.000)	0.002*** (0.000)
Ethnic Minority=Yes	-0.248** (0.100)	-0.250** (0.100)	-0.257** (0.101)	-0.195* (0.101)	-0.249** (0.100)	-0.194* (0.101)
Ethnic Minority=n.a.	-0.673** (0.315)	-0.650** (0.313)	-0.660** (0.311)	-0.602* (0.311)	-0.649** (0.311)	-0.599* (0.310)
Born in country=Yes	0.078 (0.073)	0.082 (0.073)	0.130* (0.072)	0.044 (0.074)	0.084 (0.072)	0.046 (0.073)
Born in country=n.a.	0.629 (0.470)	0.600 (0.485)	0.742 (0.530)	0.778 (0.510)	0.610 (0.489)	0.785 (0.514)
Years of education		0.059*** (0.019)	0.060*** (0.019)	0.039** (0.018)	0.060*** (0.019)	0.038** (0.018)
Years of education squared		-0.001* (0.001)	-0.001* (0.001)	-0.001 (0.001)	-0.001* (0.001)	-0.001 (0.001)
Religious=Yes			0.171*** (0.038)	0.163*** (0.038)		0.162*** (0.038)
Religious=n.a.			-0.063 (0.290)	-0.043 (0.293)		-0.049 (0.296)
Married/Cohabitating=Yes			0.638*** (0.051)	0.476*** (0.051)		0.476*** (0.051)
Married/Cohabitating=n.a.			-0.077 (0.180)	-0.114 (0.144)		-0.107 (0.144)
Domicile=Suburbs			0.025 (0.056)	0.015 (0.054)		0.014 (0.053)
Domicile=Town			0.094* (0.053)	0.095* (0.053)		0.095* (0.052)
Domicile=Village			0.203*** (0.051)	0.184*** (0.051)		0.184*** (0.051)
Domicile=Rural			0.303*** (0.088)	0.311*** (0.089)		0.310*** (0.088)
Domicile=n.a.			-0.297 (0.437)	-0.128 (0.365)		-0.131 (0.366)
Household size			0.289*** (0.061)	0.133** (0.059)		0.132** (0.058)
Household size squared			-0.036*** (0.008)	-0.018** (0.008)		-0.018** (0.008)
Unemployed=Yes, base				-1.061*** (0.100)		-1.059*** (0.100)
Log real household income				0.364*** (0.025)		0.364*** (0.025)
Constant	10.678*** (0.683)	9.893*** (0.638)	9.029*** (0.606)	5.805*** (0.659)	10.323*** (0.591)	6.042*** (0.666)
P	0.601* (0.355)	0.560* (0.321)	0.561** (0.276)	0.575** (0.243)	0.557** (0.271)	0.578** (0.227)
Log likelihood	-83710.896	-83670.845	-82975.781	-82212.744	-83657.160	-82201.708
AIC	167609.793	167533.689	166165.562	164643.488	167528.320	164643.416
N	40,342	40,342	40,342	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. 75 NUTS I fixed effects additionally included.

Table B3. Reference estimates for Western Europe (DRM), all coefficients and additional specifications

	(1) Basic controls	(2) Controls for education	(3) Controls for education and further demographics	(4) Controls for educ., further demographics., UE & income	(5) Basic controls; positive. vs. negative ref. Mobility	(6) Controls for education; positive vs. negative. ref. Mobility	(7) Controls for educ. and further demographics., pos. vs. neg. ref. mobility	(8) Controls for educ., further demographics., UE & income; pos. vs. neg. ref. mobility
Class=II	-0.151** (0.076)	-0.101 (0.077)	-0.095 (0.072)	-0.021 (0.065)	-0.138* (0.071)	-0.096 (0.074)	-0.085 (0.069)	-0.006 (0.063)
Class=III	-0.269*** (0.099)	-0.168 (0.105)	-0.143 (0.097)	-0.025 (0.096)	-0.235** (0.102)	-0.149 (0.107)	-0.118 (0.101)	0.017 (0.094)
Class=IV	-0.653*** (0.121)	-0.500*** (0.123)	-0.525*** (0.116)	-0.361*** (0.117)	-0.611*** (0.118)	-0.479*** (0.121)	-0.494*** (0.117)	-0.297** (0.126)
Class=V	-0.674*** (0.146)	-0.510*** (0.145)	-0.513*** (0.143)	-0.337** (0.139)	-0.629*** (0.143)	-0.488*** (0.144)	-0.481*** (0.141)	-0.286** (0.140)
Class=VI	-0.830*** (0.183)	-0.620*** (0.186)	-0.634*** (0.179)	-0.404** (0.178)	-0.805*** (0.180)	-0.610*** (0.185)	-0.617*** (0.178)	-0.386** (0.184)
Class=VII	-1.123*** (0.196)	-0.891*** (0.196)	-0.871*** (0.190)	-0.633*** (0.185)	-1.125*** (0.194)	-0.894*** (0.196)	-0.875*** (0.189)	-0.651*** (0.184)
Steps mobile	0.082 (0.069)	0.077 (0.051)	0.071 (0.044)	0.049* (0.030)	0.076 (0.055)	0.072 (0.046)	0.068* (0.040)	0.039 (0.030)
Ref. position	0.053 (0.040)	0.045 (0.040)	-0.049 (0.039)	0.014 (0.038)	0.051 (0.040)	0.043 (0.040)	-0.047 (0.039)	0.011 (0.038)
Ref. mobility	-0.036 (0.040)	-0.027 (0.039)	-0.042 (0.039)	-0.064* (0.037)				
Ref. mob. up					-0.004 (0.043)	-0.009 (0.043)	-0.017 (0.043)	-0.022 (0.043)
Ref. mob. down					0.071 (0.056)	0.047 (0.057)	0.069 (0.054)	0.116** (0.058)
ESS=2	-0.081 (0.053)	-0.080 (0.053)	-0.068 (0.053)	0.078 (0.054)	-0.081 (0.053)	-0.079 (0.053)	-0.068 (0.053)	0.080 (0.054)
ESS=3	-0.025 (0.051)	-0.042 (0.052)	-0.039 (0.050)	0.241*** (0.054)	-0.026 (0.051)	-0.042 (0.052)	-0.039 (0.050)	0.242*** (0.054)
ESS=4	-0.091** (0.045)	-0.109** (0.046)	-0.095** (0.046)	0.372*** (0.058)	-0.088* (0.045)	-0.107** (0.046)	-0.091* (0.047)	0.378*** (0.059)
ESS=5	-0.006 (0.054)	-0.021 (0.055)	0.021 (0.054)	0.567*** (0.061)	-0.006 (0.054)	-0.021 (0.055)	0.021 (0.054)	0.567*** (0.061)
Gender=female	0.040 (0.039)	0.047 (0.038)	0.052 (0.038)	0.076** (0.036)	0.041 (0.038)	0.047 (0.038)	0.053 (0.038)	0.076** (0.036)
Age	-0.138*** (0.024)	-0.138*** (0.024)	-0.153*** (0.023)	-0.166*** (0.022)	-0.138*** (0.024)	-0.138*** (0.024)	-0.153*** (0.023)	-0.166*** (0.022)
Age-squared	0.001*** (0.000)	0.001*** (0.000)	0.002*** (0.000)	0.002*** (0.000)	0.001*** (0.000)	0.001*** (0.000)	0.002*** (0.000)	0.002*** (0.000)
Ethnic Minority=Yes	-0.246** (0.100)	-0.249** (0.100)	-0.255** (0.101)	-0.193* (0.101)	-0.245** (0.100)	-0.249** (0.100)	-0.255** (0.101)	-0.193* (0.101)
Ethnic Minority=n.a.	-0.668** (0.315)	-0.647** (0.313)	-0.656** (0.311)	-0.599* (0.312)	-0.669** (0.315)	-0.647** (0.313)	-0.657** (0.311)	-0.599* (0.311)
Born in country=Yes	0.078 (0.073)	0.082 (0.073)	0.129* (0.072)	0.044 (0.073)	0.077 (0.073)	0.081 (0.073)	0.129* (0.072)	0.043 (0.073)
Born in country=n.a.	0.599 (0.465)	0.576 (0.481)	0.713 (0.524)	0.756 (0.504)	0.615 (0.467)	0.585 (0.481)	0.726 (0.525)	0.776 (0.506)
Years of education		0.059*** (0.019)	0.059*** (0.019)	0.037** (0.018)		0.058*** (0.019)	0.058*** (0.019)	0.035* (0.018)
Years of education squared		-0.001* (0.001)	-0.001* (0.001)	-0.001 (0.001)		-0.001* (0.001)	-0.001* (0.001)	-0.001 (0.001)
Religious=Yes			0.172*** (0.038)	0.164*** (0.038)			0.173*** (0.038)	0.165*** (0.038)
Religious=n.a.			-0.061 (0.289)	-0.042 (0.292)			-0.059 (0.289)	-0.038 (0.292)
Married/Cohabitating=Yes			0.639*** (0.051)	0.477*** (0.051)			0.639*** (0.051)	0.477*** (0.051)
Married/Cohabitating=n.a.			-0.075 (0.180)	-0.115 (0.144)			-0.076 (0.181)	-0.117 (0.144)
Domicile=Suburbs			0.025 (0.056)	0.016 (0.054)			0.025 (0.056)	0.015 (0.054)
Domicile='Town			0.094* (0.053)	0.095* (0.053)			0.093* (0.053)	0.094* (0.053)
Domicile=Village			0.203*** (0.051)	0.185*** (0.051)			0.204*** (0.051)	0.185*** (0.051)
Domicile=Rural			0.304*** (0.088)	0.314*** (0.089)			0.304*** (0.088)	0.314*** (0.089)
Domicile=n.a.			-0.291 (0.434)	-0.123 (0.363)			-0.292 (0.432)	-0.123 (0.361)
Household size			0.289*** (0.061)	0.131** (0.059)			0.289*** (0.061)	0.131** (0.059)
Household size squared			-0.036*** (0.008)	-0.018** (0.008)			-0.036*** (0.008)	-0.018** (0.008)
Unemployed=Yes, base				-1.061*** (0.100)				-1.064*** (0.100)
Log real household income				0.365*** (0.026)				0.365*** (0.026)
Constant	10.660*** (0.689)	9.913*** (0.643)	9.035*** (0.616)	5.734*** (0.670)	10.639*** (0.648)	9.920*** (0.631)	9.027*** (0.611)	5.744*** (0.676)
p	0.570 (0.347)	0.546* (0.318)	0.550** (0.275	0.592** (0.233)	0.596** (0.284)	0.575** (0.293)	0.573** (0.252)	0.685*** (0.247)
Log likelihood	-83707.526	-83668.609	-82972.229	-82208.695	-83705.955	-83668.122	-82971.243	-82205.913
AIC	167607.053	167533.217	166162.458	164639.389	167605.910	167534.243	166162.486	164635.826
N	40,342	40,342	40,342	40,342	40,342	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. 75 NUTS I fixed effects additionally included.

Table B4. Destination and Origin estimates for Eastern Europe (OLS), all coefficients

	(1) Basic model	(2) Class destination added	(3) Controls for education	(4) Controls for educ. & further demographics	(5) Controls for educ., further demographics, UE & income
Orig.=II	-0.115 (0.101)	-0.028 (0.098)	-0.011 (0.103)	-0.015 (0.089)	-0.027 (0.095)
Orig.=III	-0.178 (0.141)	-0.085 (0.141)	-0.054 (0.135)	-0.046 (0.136)	-0.059 (0.127)
Orig.=IV	-0.275** (0.111)	-0.129 (0.108)	-0.088 (0.116)	-0.108 (0.099)	-0.062 (0.096)
Orig.=V	-0.269* (0.158)	-0.048 (0.150)	-0.003 (0.155)	-0.039 (0.152)	-0.046 (0.149)
Orig.=VI	-0.414*** (0.110)	-0.149 (0.101)	-0.077 (0.114)	-0.106 (0.109)	-0.082 (0.099)
Orig.=VII	-0.338*** (0.062)	-0.031 (0.055)	0.052 (0.058)	0.035 (0.053)	0.115** (0.049)
Orig.=SEI	-0.370 (0.256)	-0.179 (0.257)	-0.125 (0.256)	-0.165 (0.242)	-0.117 (0.231)
Orig.=SEII	-0.058 (0.118)	0.215* (0.117)	0.329*** (0.123)	0.288** (0.112)	0.298*** (0.102)
Dest.=II		-0.118* (0.060)	-0.045 (0.064)	-0.074 (0.078)	-0.040 (0.078)
Dest.=III		-0.519*** (0.099)	-0.364*** (0.097)	-0.407*** (0.108)	-0.340*** (0.104)
Dest.=IV		-0.590*** (0.130)	-0.382*** (0.126)	-0.425*** (0.141)	-0.325** (0.135)
Dest.=V		-0.751*** (0.182)	-0.512** (0.204)	-0.501** (0.215)	-0.385* (0.212)
Dest.=VI		-0.933*** (0.075)	-0.648*** (0.103)	-0.620*** (0.120)	-0.457*** (0.126)
Dest.=VII		-0.915*** (0.091)	-0.607*** (0.091)	-0.603*** (0.106)	-0.437*** (0.097)
Dest.=SEI		-0.132 (0.137)	0.047 (0.144)	0.023 (0.144)	0.027 (0.130)
Dest.=SEII		-0.392** (0.161)	-0.072 (0.175)	-0.100 (0.184)	-0.042 (0.180)
ESS=2	0.423** (0.169)	0.446*** (0.167)	0.442*** (0.164)	0.455*** (0.161)	0.516*** (0.172)
ESS=3	0.878*** (0.157)	0.877*** (0.155)	0.840*** (0.153)	0.865*** (0.148)	0.657*** (0.190)
ESS=4	0.916*** (0.164)	0.902*** (0.161)	0.856*** (0.157)	0.864*** (0.154)	1.246*** (0.189)
ESS=5	1.189*** (0.166)	1.193*** (0.164)	1.136*** (0.161)	1.157*** (0.157)	2.075*** (0.221)
Gender=female	0.110 (0.091)	0.020 (0.076)	0.037 (0.080)	0.125* (0.072)	0.178*** (0.064)
Age	-0.028 (0.027)	-0.027 (0.027)	-0.040 (0.026)	-0.055** (0.028)	-0.054** (0.027)
Age-squared	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Ethnic Minority=Yes	0.048 (0.141)	0.034 (0.142)	0.056 (0.142)	0.025 (0.133)	0.066 (0.123)
Ethnic Minority=n.a.	0.094 (0.289)	0.063 (0.284)	0.063 (0.277)	0.099 (0.269)	0.097 (0.276)
Born in country=Yes	0.174 (0.165)	0.168 (0.166)	0.200 (0.179)	0.211 (0.187)	0.214 (0.187)
Born in country=n.a.	-0.673 (0.562)	-0.645 (0.605)	-0.632 (0.624)	-0.482 (0.603)	-0.336 (0.587)
Years of education			0.058 (0.062)	0.042 (0.060)	0.019 (0.057)
Years of education squared			0.001 (0.002)	0.001 (0.002)	0.001 (0.002)
Religious=Yes				0.120* (0.069)	0.141* (0.072)
Religious=n.a.				-0.460 (0.332)	-0.436 (0.305)
Married/Cohabitating=Yes				0.672*** (0.068)	0.550*** (0.069)
Married/Cohabitating=n.a.				0.659*** (0.151)	0.525*** (0.136)
Domicile=Suburbs				0.090 (0.132)	0.180 (0.144)
Domicile=Town				-0.018 (0.078)	0.127* (0.071)
Domicile=Village				0.062 (0.101)	0.308*** (0.083)
Domicile=Rural				-0.797* (0.477)	-0.453 (0.452)
Domicile=n.a.				0.266 (0.298)	0.161 (0.280)
Household size				0.011 (0.057)	-0.076 (0.061)
Household size squared				-0.000 (0.005)	0.004 (0.007)
Unemployed=Yes, base					-0.839*** (0.139)
Log real household income					0.367*** (0.033)
Constant	4.398*** (0.627)	4.862*** (0.653)	4.083*** (0.718)	3.799*** (0.767)	0.639 (0.705)
Log likelihood	-51280.739	-51050.728	-50996.867	-50792.039	-50405.175
AIC	102645.478	102201.457	102097.734	101710.078	100940.350
N	22,308	22,308	22,308	22,308	22,308

Note: * p<0.10, ** p<0.05, *** p<0.01. 27 NUTS I fixed effects additionally included.

Table B5. Mobility estimates for Eastern Europe (DRM), all coefficients

	(1) Linear mobility	(2) Linear mobility, controls for education	(3) Linear mobility, controls for educ. & further demographics	(4) Linear mobility, controls for educ., further demographics, UE & income	(5) Non-linear mobility, controls for educ.	(6) Non-linear mobility, controls for educ., further demographics, UE & income
Class=II	-0.127 (0.113)	-0.030 (0.132)	-0.065 (0.133)	-0.046 (0.150)	-0.108 (0.141)	-0.118 (0.148)
Class=III	-0.661*** (0.181)	-0.462** (0.183)	-0.538*** (0.181)	-0.459*** (0.164)	-0.677*** (0.217)	-0.639*** (0.202)
Class=IV	-0.611*** (0.144)	-0.357** (0.170)	-0.426** (0.178)	-0.300* (0.168)	-0.585*** (0.177)	-0.512*** (0.181)
Class=V	-0.797*** (0.284)	-0.501 (0.326)	-0.520 (0.337)	-0.416 (0.325)	-0.674** (0.312)	-0.593* (0.311)
Class=VI	-1.113*** (0.155)	-0.753*** (0.207)	-0.755*** (0.219)	-0.588*** (0.208)	-0.824*** (0.211)	-0.658*** (0.212)
Class=VII	-0.932*** (0.123)	-0.538*** (0.117)	-0.551*** (0.128)	-0.314*** (0.092)	-0.573*** (0.107)	-0.346*** (0.089)
Steps mobile	0.060*** (0.018)	0.053*** (0.015)	0.053*** (0.015)	0.054*** (0.012)		
Steps down=6					-0.464*** (0.112)	-0.461*** (0.113)
Steps down=5					-0.527*** (0.133)	-0.448*** (0.130)
Steps down=4					-0.209 (0.220)	-0.129 (0.202)
Steps down=3					-0.107 (0.172)	-0.096 (0.174)
Steps down=2					-0.097 (0.104)	-0.103 (0.098)
Steps down=1					-0.066 (0.098)	-0.089 (0.107)
Steps up=1					-0.089 (0.086)	-0.081 (0.084)
Steps up=2					0.275** (0.135)	0.288** (0.138)
Steps up=3					0.311* (0.163)	0.286 (0.177)
Steps up=4					0.414*** (0.108)	0.311*** (0.108)
Steps up=5					0.124 (0.161)	0.130 (0.154)
Steps up=6					-0.089 (0.199)	-0.197 (0.252)
ESS=2	0.348* (0.209)	0.343* (0.204)	0.377* (0.201)	0.417* (0.217)	0.344* (0.202)	0.418* (0.215)
ESS=3	0.749*** (0.200)	0.708*** (0.195)	0.752*** (0.191)	0.495** (0.231)	0.706*** (0.195)	0.492** (0.231)
ESS=4	0.730*** (0.204)	0.682*** (0.198)	0.711*** (0.195)	1.072*** (0.231)	0.677*** (0.197)	1.069*** (0.229)
ESS=5	1.066*** (0.207)	1.007*** (0.202)	1.039*** (0.198)	1.941*** (0.259)	1.007*** (0.200)	1.943*** (0.255)
Gender=female	-0.007 (0.079)	0.012 (0.083)	0.098 (0.074)	0.148** (0.072)	-0.001 (0.084)	0.135** (0.069)
Age	-0.038 (0.028)	-0.050* (0.028)	-0.064** (0.030)	-0.062** (0.031)	-0.051* (0.028)	-0.063** (0.031)
Age-squared	0.000 (0.000)	0.000 (0.000)	0.001* (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Ethnic Minority=Yes	-0.040 (0.146)	-0.016 (0.147)	-0.050 (0.138)	-0.006 (0.127)	-0.014 (0.141)	-0.003 (0.122)
Ethnic Minority=n.a.	0.018 (0.322)	0.021 (0.311)	0.032 (0.309)	0.046 (0.317)	0.033 (0.303)	0.057 (0.307)
Born in country=Yes	0.180 (0.209)	0.209 (0.221)	0.224 (0.228)	0.225 (0.226)	0.210 (0.222)	0.228 (0.226)
Born in country=n.a.	-0.750 (0.652)	-0.750 (0.672)	-0.587 (0.657)	-0.416 (0.650)	-0.781 (0.681)	-0.438 (0.660)
Years of education		0.057 (0.060)	0.042 (0.059)	0.013 (0.056)	0.058 (0.060)	0.014 (0.056)
Years of education squared		0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	0.001 (0.002)
Religious=Yes			0.138* (0.072)	0.155** (0.076)		0.152** (0.075)
Religious=n.a.			-0.276 (0.270)	-0.280 (0.256)		-0.284 (0.257)
Married/Cohabitating=Yes			0.616*** (0.074)	0.492*** (0.074)		0.486*** (0.077)
Married/Cohabitating=n.a.			0.574*** (0.085)	0.476*** (0.081)		0.482*** (0.079)
Domicile=Suburbs			0.090 (0.159)	0.186 (0.172)		0.180 (0.172)
Domicile=Town			-0.048 (0.083)	0.100 (0.074)		0.101 (0.072)
Domicile=Village			0.068 (0.101)	0.312*** (0.091)		0.308*** (0.091)
Domicile=Rural			-0.769 (0.521)	-0.435 (0.509)		-0.432 (0.515)
Domicile=n.a.			0.354 (0.293)	0.245 (0.281)		0.279 (0.280)
Household size			0.003 (0.061)	-0.082 (0.065)		-0.080 (0.068)
Household size squared			0.000 (0.006)	0.004 (0.007)		0.003 (0.007)
Unemployed=Yes, base				-0.801*** (0.150)		-0.807*** (0.148)
Log real household income				0.360*** (0.031)		0.361*** (0.030)
Constant	4.804*** (0.680)	4.090*** (0.744)	3.820*** (0.835)	0.755 (0.714)	4.568*** (0.737)	1.227* (0.725)
p	0.627*** (0.101)	0.599*** (0.121)	0.555*** (0.113)	0.475*** (0.116)	0.515*** (0.064)	0.483*** (0.068)
Log likelihood	-52056.414	-51999.934	-51818.216	-51447.841	-51975.353	-51423.802
AIC	104198.828	104089.869	103748.433	103013.683	104062.706	102985.604
N	19,152	19,152	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. 27 NUTS I fixed effects additionally included.

Table B6. Reference estimates for Eastern Europe (DRM), all coefficients

	(1) Basic controls	(2) Controls for education	(3) Controls for education and further demographics	(4) Controls for educ., further demographics., UE & income	(5) Basic controls; positive. vs. negative ref. Mobility	(6) Controls for education; positive vs. negative. ref. Mobility	(7) Controls for educ. and further demographics., pos. vs. neg. ref. mobility	(8) Controls for educ., further demographics., UE & income; pos. vs. neg. ref. mobility
Class=II	-0.199 (0.134)	-0.114 (0.161)	-0.147 (0.151)	-0.153 (0.178)	-0.215* (0.124)	-0.127 (0.140)	-0.155 (0.140)	-0.143 (0.167)
Class=III	-0.748*** (0.180)	-0.561*** (0.182)	-0.628*** (0.167)	-0.563*** (0.155)	-0.767*** (0.172)	-0.582*** (0.165)	-0.649*** (0.169)	-0.584*** (0.164)
Class=IV	-0.917*** (0.175)	-0.684*** (0.194)	-0.747*** (0.180)	-0.622*** (0.184)	-0.980*** (0.220)	-0.753*** (0.208)	-0.801*** (0.200)	-0.658*** (0.190)
Class=V	-1.175*** (0.303)	-0.902*** (0.350)	-0.926*** (0.348)	-0.843** (0.342)	-1.254*** (0.305)	-0.993*** (0.330)	-0.990*** (0.342)	-0.872*** (0.334)
Class=VI	-1.576*** (0.203)	-1.239*** (0.244)	-1.241*** (0.252)	-1.098*** (0.269)	-1.604*** (0.201)	-1.271*** (0.230)	-1.262*** (0.240)	-1.103*** (0.259)
Class=VII	-1.545*** (0.260)	-1.177*** (0.254)	-1.189*** (0.232)	-0.973*** (0.244)	-1.500*** (0.231)	-1.120*** (0.218)	-1.139*** (0.206)	-0.921*** (0.210)
Steps mobile	0.157*** (0.040)	0.146*** (0.044)	0.156*** (0.040)	0.141*** (0.043)	0.131*** (0.042)	0.121*** (0.035)	0.130*** (0.039)	0.120*** (0.036)
Ref. position	0.110* (0.062)	0.096 (0.062)	-0.098* (0.059)	0.032 (0.060)	0.112* (0.059)	0.099* (0.060)	-0.100* (0.057)	0.034 (0.057)
Ref. mobility	-0.203*** (0.051)	-0.201*** (0.050)	-0.202*** (0.045)	-0.176*** (0.047)				
Ref. mob. up					-0.253*** (0.090)	-0.264*** (0.087)	-0.252*** (0.079)	-0.214*** (0.075)
Ref. mob. down					0.122 (0.104)	0.104 (0.100)	0.120 (0.101)	0.110 (0.098)
ESS=2	0.322 (0.208)	0.317 (0.203)	0.351* (0.200)	0.392* (0.219)	0.331 (0.207)	0.328 (0.202)	0.360* (0.199)	0.399* (0.218)
ESS=3	0.741*** (0.198)	0.701*** (0.194)	0.745*** (0.189)	0.493** (0.232)	0.748*** (0.196)	0.709*** (0.192)	0.751*** (0.187)	0.498** (0.231)
ESS=4	0.699*** (0.202)	0.652*** (0.197)	0.681*** (0.193)	1.044*** (0.232)	0.702*** (0.201)	0.655*** (0.196)	0.683*** (0.192)	1.044*** (0.231)
ESS=5	1.041*** (0.204)	0.982*** (0.200)	1.014*** (0.196)	1.909*** (0.259)	1.052*** (0.204)	0.995*** (0.199)	1.024*** (0.195)	1.915*** (0.259)
Gender=female	-0.003 (0.079)	0.016 (0.083)	0.102 (0.075)	0.152** (0.075)	-0.008 (0.077)	0.010 (0.080)	0.096 (0.072)	0.147** (0.069)
Age	-0.021 (0.029)	-0.034 (0.029)	-0.048 (0.032)	-0.050 (0.032)	-0.023 (0.029)	-0.036 (0.029)	-0.050 (0.032)	-0.052 (0.032)
Age-squared	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Ethnic Minority=Yes	-0.044 (0.146)	-0.020 (0.147)	-0.056 (0.139)	-0.010 (0.128)	-0.044 (0.147)	-0.020 (0.148)	-0.055 (0.139)	-0.010 (0.128)
Ethnic Minority=n.a.	0.004 (0.319)	0.006 (0.308)	0.016 (0.306)	0.032 (0.313)	0.004 (0.320)	0.006 (0.309)	0.016 (0.306)	0.031 (0.313)
Born in country=Yes	0.177 (0.209)	0.205 (0.221)	0.220 (0.227)	0.222 (0.224)	0.176 (0.209)	0.204 (0.220)	0.220 (0.226)	0.222 (0.224)
Born in country=n.a.	-0.763 (0.664)	-0.760 (0.684)	-0.597 (0.669)	-0.419 (0.660)	-0.773 (0.663)	-0.774 (0.683)	-0.610 (0.667)	-0.432 (0.657)
Years of education		0.057 (0.061)	0.042 (0.059)	0.014 (0.056)		0.058 (0.061)	0.043 (0.059)	0.014 (0.056)
Years of education squared		0.001 (0.002)	0.001 (0.002)	0.001 (0.002)		0.001 (0.002)	0.001 (0.002)	0.001 (0.002)
Religious=Yes			0.145** (0.069)	0.160** (0.074)			0.143** (0.069)	0.159** (0.074)
Religious=n.a.			-0.281 (0.271)	-0.283 (0.259)			-0.280 (0.269)	-0.282 (0.256)
Married/Cohabitating=Yes			0.612*** (0.073)	0.490*** (0.073)			0.611*** (0.073)	0.489*** (0.074)
Married/Cohabitating=n.a.			0.572*** (0.088)	0.477*** (0.084)			0.569*** (0.085)	0.474*** (0.085)
Domicile=Suburbs			0.092 (0.162)	0.186 (0.174)			0.088 (0.164)	0.182 (0.177)
Domicile=Town			-0.051 (0.084)	0.096 (0.073)			-0.051 (0.083)	0.095 (0.073)
Domicile=Village			0.069 (0.100)	0.313*** (0.088)			0.070 (0.099)	0.313*** (0.087)
Domicile=Rural			-0.754 (0.520)	-0.425 (0.508)			-0.756 (0.518)	-0.429 (0.508)
Domicile=n.a.			0.314 (0.285)	0.211 (0.277)			0.321 (0.285)	0.218 (0.275)
Household size			0.006 (0.060)	-0.079 (0.064)			0.005 (0.060)	-0.079 (0.063)
Household size squared			0.000 (0.006)	0.003 (0.007)			0.000 (0.006)	0.003 (0.007)
Unemployed=Yes, base				-0.802*** (0.151)				-0.802*** (0.151)
Log real household income				0.357*** (0.032)				0.356*** (0.031)
Constant	4.508*** (0.755)	3.794*** (0.842)	3.437*** (0.950)	0.343 (0.907)	4.833*** (0.861)	4.129*** (0.852)	3.757*** (0.994)	0.606 (0.785)
p	0.538*** (0.108)	0.500*** (0.149)	0.435*** (0.131)	0.358** (0.171)	0.626*** (0.126)	0.609*** (0.122)	0.549*** (0.143)	0.468*** (0.129)
Log likelihood	-52036.254	-51980.875	-51798.764	-51435.163	-52034.249	-51977.636	-51796.480	-51433.651
AIC	104162.509	104055.751	103715.528	102990.326	104160.498	104051.271	103710.960	102991.303
N	19,152	19,152	19,152	19,152	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. 27 NUTS I fixed effects additionally included.

C Robustness tests

C.1 Comparing DRMs with OLS regressions

In contrast to OLS regressions, the diagonal reference model (DRM) imposes a severe constraint on the shape of the origin and destination effects. In particular, it constrains the destination effects to be positive linear transformations of the origin effects. Doing so entails that the ratio of the effect of coming from e.g. origin II rather than III *versus* the effect of coming from e.g. origin IV rather than V is forced to be the same as the ratio of the effect of attaining destination II rather than III *versus* the effect of attaining destination IV rather than V. We may therefore be worried that these constraints cause estimates of the overall magnitude and direction of the origin, destination, and mobility effects to be misleading.

To assess this worry, we borrow from two papers by Fosse and Winship (2019b; 2019a) and an extended abstract by Fosse and Pfeffer (2019), and show that an OLS regression which omits linear terms for mobility is nevertheless informative about such linear effects. We then compare estimates from this OLS regression with our DRM estimates. Finally, we check the sensitivity of our main DRM results against relaxing constraints on the non-linear class effects in a stepwise manner.

Equation (1) of the main text can be rewritten as:

$$\begin{aligned} LS_{idom} &= \dot{\delta}d + \dot{\omega}o + \dot{\gamma}(d-o) + \tilde{\delta}_d dest_{id} + \tilde{\omega}_o orig_{io} + \tilde{\gamma}_m mob_{im} + \beta' X_i + \varepsilon_i \\ &= \left(\dot{\delta}+\dot{\gamma}\right)d + (\dot{\omega}-\dot{\gamma})o + \tilde{\delta}_d dest_{id} + \tilde{\omega}_o orig_{io} + \tilde{\gamma}_m mob_{im} + \beta' X_i + \varepsilon_i \\ &= \theta_1 d + \theta_2 o + \tilde{\delta}_d dest_{id} + \tilde{\omega}_o orig_{io} + \tilde{\gamma}_m mob_{im} + \beta' X_i + \varepsilon_i \end{aligned} \quad \text{(A1)}$$

Where we define $\dot{\delta}+\dot{\gamma}=\theta_1$ and $\dot{\omega}-\dot{\gamma}=\theta_2$. We further set the constraint $\tilde{\delta}_1 = \tilde{\delta}_7 = \tilde{\omega}_1 = \tilde{\omega}_7 = \tilde{\gamma}_{-6} = \tilde{\gamma}_6 = 0$.[15] In this rewritten form, we interpret $\dot{\delta}$, $\dot{\omega}$ and $\dot{\gamma}$ as linear trend components of the destination, origin and mobility effects, from which there may be non-linear deviations as

[15] It would have been possible to choose any other combination of levels for destination and origin to be set to zero, which would all yield different estimates of the "linear" components. Alternatively, one could also set the effect of one level to zero and let the effects of all other levels sum to zero. Fosse & Winship (2019b) use a more sophisticated "orthogonal" (see page 1979 of their paper) decomposition to separate linear and non-linear components. Our present choice appeared attractive for its ease of interpretation.

given by $\tilde{\delta}_d$, $\tilde{\omega}_o$ and $\tilde{\gamma}_m$. From the first and second line of equation (A1), it again becomes evident that an unconstrained OLS regression, which includes all origin, destination, and mobility terms, is not identified. The second and third lines additionally show that a regression in which such linear destination and origin terms are entered but no term for mobility is included, yields estimates for the linear destination and origin terms that include mobility effects (as given by θ_1 and θ_2). Indeed, in tables 2 and 5 we already make use of this point informally when interpreting the coefficients on destination and origin.

Similarly, the DRM presented in equation (2) of the main text can be rewritten as:

$$\begin{aligned} LS_{idom} = {} & p\dot{\lambda}d + (1-p)\dot{\lambda}o + \dot{\gamma}(d-o) \\ & + p\tilde{\lambda}_d dest_{id} + (1-p)\tilde{\lambda}_o orig_{io} + \tilde{\gamma}_m mob_{im} + \beta' \boldsymbol{X_i} + \varepsilon_i \end{aligned} \tag{A2}$$

With the constraint $\tilde{\lambda}_1 = \tilde{\lambda}_7 = \tilde{\gamma}_{-6} = \tilde{\gamma}_6 = 0$. Analogous to the case of equation (A1), we respectively interpret $p\dot{\lambda}$, $(1-p)\dot{\lambda}$, and $\dot{\gamma}$ as linear trend components of the destination, origin and mobility effects, from which there may be non-linear deviations as given by $p\tilde{\lambda}_d$, $(1-p)\tilde{\lambda}_o$, and $\tilde{\gamma}_m$.

It is thus possible to evaluate whether the constraints we place on the non-linear parts of the destination and origin effects in equation (A2) result in estimates of the linear trend components that are inconsistent with the linear estimates of the unconstrained model in (A1).[16] To do so, we simply compare estimates of θ_1 with estimates of $p\dot{\lambda} + \dot{\gamma}$ and estimates of θ_2 with estimates of $(1-p)\dot{\lambda} - \dot{\gamma}$.

We perform this exercise for all presented results based on the DRM. As can be discerned from appendix tables C5-C6, it turns out that our DRM estimates only deviate marginally from the implied linear estimates of the unconstrained OLS regression. The largest deviation is found for

[16] Note that $\dot{\lambda}$ and $\dot{\gamma}$ from eq.(A2) can be obtained directly from eq.(2) by respectively calculating $\dot{\lambda} = \frac{\lambda_7}{C-1}$ and $\dot{\gamma} = \frac{\gamma_{C-1} - \gamma_{-C+1}}{2C-2}$ for C social classes.

the result of table B2 column 2, where the constrained DRM estimate for origin minus mobility is 0.004 points smaller than the implied unconstrained OLS estimate.

Hence, although the DRM imposes a severe constraint on the non-linear parts of the origin and destination effects (i.e. that these are equal in shape), this constraint does not make estimates of the overall trends of the effects of destination, origin and class mobility inconsistent with those implied by an OLS regression.

Given the specification of equation (A2), we can also evaluate the degree to which our results depend on constraining the non-linear components of the origin and destination effect to be positive linear transformations. We do so by relaxing the constraints on each non-linearity in a stepwise manner. More precisely, we first allow $\tilde{\lambda}_{o=2} \neq \tilde{\lambda}_{d=2}$, we then allow $\tilde{\lambda}_{o=3} \neq \tilde{\lambda}_{d=3}$, etc. Tables C7-C8 provide such results for our preferred specifications (i.e. columns 2 and 4 of tables 3 and 5).

We find that our overall class position trend term $\dot{\lambda}$ is largely unaffected from relaxing these constraints. Moreover, in most cases our estimated mobility effects are also not strongly affected. The only noteworthy exceptions to this stems from relaxing the constraint on Class III (i.e. allowing $\tilde{\lambda}_{o=5} \neq \tilde{\lambda}_{d=5}$; note the reverse coding) for Western Europe. Doing so, yields a statistically significant coefficients for own class mobility. Hence, our results on class mobility are not fully robust. Nevertheless, since most of our robustness estimates are in line with our main results in tables 3 and 6, we maintain our conclusion to reject **H3a** and **H3b** for Western Europe, while seeing **H3b** for Eastern Europe, as well as **H5** as corroborated.

C.2 Robustness Tables

Table C1. Unconstrained mobility estimates for Western Europe (OLS model on basis of eq. A1)

	(1) Linear mobility	(2) Linear mobility, controls for education	(3) Linear mobility, controls for educ. & further demographics	(4) Linear mobility, controls for educ., further demographics, UE & income	(5) Non-linear mobility, controls for educ.	(6) Non-linear mobility, controls for educ., further demographics, UE & income
Orig. linear	0.018* (0.010)	0.007 (0.010)	0.012 (0.010)	0.002 (0.010)	-0.006 (0.020)	-0.006 (0.019)
Dest. linear	0.165*** (0.012)	0.140*** (0.013)	0.123*** (0.013)	0.069*** (0.013)	0.152*** (0.021)	0.076*** (0.020)
Dest. non-linear						
Destination=II	0.033 (0.045)	0.037 (0.045)	0.036 (0.046)	0.038 (0.045)	0.020 (0.048)	0.028 (0.047)
Destination=III	0.010 (0.061)	0.017 (0.061)	0.034 (0.060)	0.022 (0.059)	0.012 (0.066)	0.021 (0.065)
Destination=IV	-0.019 (0.058)	-0.016 (0.057)	-0.038 (0.055)	-0.037 (0.055)	-0.026 (0.059)	-0.041 (0.057)
Destination=V	0.053 (0.069)	0.051 (0.069)	0.055 (0.067)	0.049 (0.066)	0.076 (0.072)	0.074 (0.068)
Destination=VI	0.076 (0.054)	0.078 (0.054)	0.065 (0.053)	0.074 (0.054)	0.088 (0.055)	0.084 (0.054)
Orig. non-linear						
Origin=II	-0.001 (0.042)	0.007 (0.042)	0.008 (0.041)	0.040 (0.039)	0.022 (0.046)	0.054 (0.042)
Origin =III	0.072 (0.053)	0.091* (0.054)	0.093* (0.053)	0.125** (0.052)	0.127** (0.060)	0.158*** (0.057)
Origin =IV	-0.088 (0.064)	-0.054 (0.065)	-0.073 (0.064)	-0.042 (0.062)	-0.029 (0.069)	-0.017 (0.068)
Origin =V	0.026 (0.070)	0.035 (0.069)	0.017 (0.067)	0.038 (0.064)	0.035 (0.073)	0.040 (0.068)
Origin =VI	-0.006 (0.080)	0.009 (0.080)	-0.012 (0.078)	0.014 (0.075)	0.013 (0.080)	0.020 (0.076)
Steps down=5					0.024 (0.210)	-0.045 (0.207)
Steps down=4					-0.183 (0.186)	-0.208 (0.186)
Steps down=3					0.055 (0.160)	-0.018 (0.158)
Steps down=2					-0.053 (0.154)	-0.101 (0.152)
Steps down=1					-0.050 (0.126)	-0.098 (0.125)
No steps					-0.024 (0.117)	-0.071 (0.115)
Steps up=1					-0.042 (0.114)	-0.084 (0.109)
Steps up=2					-0.056 (0.117)	-0.091 (0.110)
Steps up=3					-0.241* (0.123)	-0.255** (0.118)
Steps up=4					-0.114 (0.125)	-0.123 (0.116)
Steps up=5					-0.157 (0.133)	-0.164 (0.122)
Log likelihood	-86648.287	-86607.153	-85887.887	-85097.669	-86592.379	-85085.381
AIC	173490.575	173412.306	171995.774	170419.337	173404.759	170416.761
N	40,342	40,342	40,342	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C2. Unconstrained mobility and reference estimates for Eastern Europe (OLS model on basis of eq. A1)

	(1) Basic controls	(2) Controls for education	(3) Controls for educ., further demographics., UE & income	(4) Basic controls; positive. vs. negative ref. Mobility	(5) Controls for education; positive vs. negative. ref. Mobility	(6) Controls for educ., further demographics., UE & income; pos. vs. neg. ref. mobility
Orig. linear	-0.002 (0.017)	-0.011 (0.017)	-0.005 (0.017)	-0.002 (0.017)	-0.010 (0.017)	-0.004 (0.017)
Dest. linear	0.185*** (0.028)	0.156*** (0.029)	0.110*** (0.027)	0.187*** (0.029)	0.157*** (0.029)	0.114*** (0.027)
Dest. non-lin.						
Dest.=II	0.041 (0.044)	0.044 (0.045)	0.041 (0.045)	0.040 (0.044)	0.043 (0.045)	0.040 (0.045)
Dest.=III	0.005 (0.061)	0.012 (0.061)	0.020 (0.060)	0.003 (0.061)	0.012 (0.061)	0.018 (0.059)
Dest.=IV	-0.014 (0.057)	-0.011 (0.057)	-0.034 (0.055)	-0.015 (0.058)	-0.012 (0.058)	-0.036 (0.055)
Dest.=V	0.053 (0.069)	0.051 (0.069)	0.048 (0.066)	0.051 (0.069)	0.049 (0.069)	0.045 (0.066)
Dest.=VI	0.084 (0.054)	0.085 (0.054)	0.076 (0.054)	0.083 (0.054)	0.084 (0.054)	0.075 (0.054)
Orig. non-lin.						
Origin=II	0.004 (0.043)	0.011 (0.043)	0.049 (0.039)	0.019 (0.046)	0.019 (0.046)	0.074* (0.042)
Origin =III	0.080 (0.053)	0.097* (0.054)	0.140*** (0.052)	0.109* (0.062)	0.114* (0.063)	0.189*** (0.060)
Origin =IV	-0.075 (0.066)	-0.044 (0.066)	-0.017 (0.064)	-0.035 (0.078)	-0.022 (0.079)	0.049 (0.074)
Origin =V	0.030 (0.069)	0.038 (0.069)	0.046 (0.064)	0.070 (0.083)	0.061 (0.083)	0.114 (0.077)
Origin =VI	0.006 (0.079)	0.018 (0.079)	0.035 (0.074)	0.037 (0.084)	0.036 (0.084)	0.086 (0.077)
Ref. position	0.053 (0.040)	0.045 (0.040)	0.014 (0.038)	0.051 (0.040)	0.043 (0.040)	0.011 (0.038)
Ref. mobility	-0.036 (0.040)	-0.027 (0.039)	-0.064* (0.037)			
Ref. mob. up				-0.004 (0.043)	-0.009 (0.043)	-0.022 (0.043)
Ref. mob. down				0.071 (0.056)	0.047 (0.057)	0.116** (0.058)
Log likelihood	-86644.427	-86604.473	-85093.318	-86643.245	-86604.099	-85089.807
AIC	173486.854	173410.946	170414.636	173486.490	173412.199	170409.614
N	40,342	40,342	40,342	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C3. Unconstrained mobility estimates for Eastern Europe (OLS model on basis of eq. A1)

	(1) Linear mobility	(2) Linear mobility, controls for education	(3) Linear mobility, controls for educ. & further demographics	(4) Linear mobility, controls for educ., further demographics, UE & income	(5) Non-linear mobility, controls for educ.	(6) Non-linear mobility, controls for educ., further demographics, UE & income
Orig. linear	-0.001 (0.008)	-0.015* (0.008)	-0.012 (0.008)	-0.026*** (0.008)	-0.016 (0.017)	-0.009 (0.018)
Dest. linear	0.156*** (0.016)	0.104*** (0.016)	0.104*** (0.019)	0.077*** (0.017)	0.081*** (0.018)	0.051** (0.021)
Dest. non-linear						
Destination=II	-0.019 (0.091)	-0.014 (0.095)	-0.009 (0.084)	-0.031 (0.091)	0.037 (0.091)	0.006 (0.084)
Destination=III	-0.082 (0.138)	-0.076 (0.130)	-0.061 (0.123)	-0.077 (0.126)	-0.054 (0.132)	-0.065 (0.129)
Destination=IV	-0.091 (0.121)	-0.094 (0.125)	-0.101 (0.116)	-0.087 (0.106)	-0.170 (0.143)	-0.164 (0.122)
Destination=V	-0.063 (0.126)	-0.073 (0.125)	-0.091 (0.130)	-0.145 (0.133)	-0.151 (0.146)	-0.217 (0.153)
Destination=VI	-0.140 (0.094)	-0.138 (0.096)	-0.151 (0.101)	-0.192** (0.093)	-0.207** (0.104)	-0.253** (0.105)
Orig. non-linear						
Origin=II	0.036 (0.057)	0.063 (0.058)	0.031 (0.071)	0.039 (0.071)	-0.059 (0.069)	-0.082 (0.084)
Origin =III	-0.225*** (0.085)	-0.168** (0.079)	-0.219** (0.086)	-0.200** (0.083)	-0.336*** (0.110)	-0.360*** (0.114)
Origin =IV	-0.070 (0.088)	-0.015 (0.083)	-0.058 (0.092)	-0.054 (0.097)	-0.137* (0.082)	-0.183* (0.098)
Origin =V	-0.110 (0.182)	-0.075 (0.185)	-0.070 (0.176)	-0.066 (0.178)	-0.136 (0.188)	-0.140 (0.184)
Origin =VI	-0.198* (0.102)	-0.167 (0.104)	-0.143 (0.100)	-0.124 (0.104)	-0.124 (0.116)	-0.092 (0.113)
Steps down=5					-0.139 (0.131)	-0.056 (0.120)
Steps down=4					0.112 (0.224)	0.184 (0.210)
Steps down=3					0.231 (0.154)	0.259* (0.152)
Steps down=2					0.233 (0.186)	0.266 (0.220)
Steps down=1					0.273** (0.121)	0.303** (0.135)
No steps					0.276** (0.135)	0.329* (0.169)
Steps up=1					0.126 (0.147)	0.186 (0.171)
Steps up=2					0.503*** (0.159)	0.585*** (0.192)
Steps up=3					0.540*** (0.189)	0.619*** (0.233)
Steps up=4					0.643*** (0.223)	0.651** (0.275)
Steps up=5					0.290* (0.174)	0.403* (0.236)
Log likelihood	-43752.719	-43704.935	-43552.046	-43240.000	-43682.466	-43217.266
AIC	87597.438	87505.869	87222.092	86601.999	87482.932	86578.531
N	19,152	19,152	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C4. Unconstrained mobility and reference estimates for Eastern Europe (OLS model on basis of eq. A1)

	(1) Basic controls	(2) Controls for education	(3) Controls for educ., further demographics., UE & income	(4) Basic controls; positive. vs. negative ref. Mobility	(5) Controls for education; positive vs. negative. ref. Mobility	(6) Controls for educ., further demographics., UE & income; pos. vs. neg. ref. mobility
Orig. linear	-0.035 (0.024)	-0.044* (0.023)	-0.034 (0.021)	-0.036 (0.023)	-0.046** (0.022)	-0.035* (0.021)
Dest. linear	0.294*** (0.043)	0.243*** (0.043)	0.196*** (0.041)	0.287*** (0.039)	0.235*** (0.039)	0.191*** (0.038)
Dest. non-lin.						
Dest.=II	-0.004 (0.096)	-0.001 (0.100)	-0.029 (0.096)	-0.001 (0.096)	0.002 (0.100)	-0.027 (0.096)
Dest.=III	-0.105 (0.143)	-0.099 (0.134)	-0.094 (0.128)	-0.100 (0.141)	-0.092 (0.132)	-0.090 (0.126)
Dest.=IV	-0.097 (0.117)	-0.100 (0.122)	-0.091 (0.105)	-0.094 (0.119)	-0.096 (0.124)	-0.088 (0.106)
Dest.=V	-0.079 (0.128)	-0.087 (0.127)	-0.153 (0.134)	-0.079 (0.128)	-0.088 (0.126)	-0.153 (0.134)
Dest.=VI	-0.140 (0.090)	-0.139 (0.092)	-0.195** (0.092)	-0.140 (0.090)	-0.138 (0.092)	-0.194** (0.092)
Orig. non-lin.						
Origin=II	0.060 (0.057)	0.088 (0.059)	0.063 (0.070)	0.030 (0.062)	0.054 (0.065)	0.042 (0.076)
Origin =III	-0.116 (0.071)	-0.058 (0.068)	-0.105 (0.074)	-0.166 (0.110)	-0.116 (0.105)	-0.140 (0.103)
Origin =IV	-0.052 (0.098)	0.003 (0.089)	-0.038 (0.099)	-0.140 (0.181)	-0.099 (0.172)	-0.099 (0.161)
Origin =V	-0.067 (0.188)	-0.030 (0.191)	-0.025 (0.182)	-0.165 (0.188)	-0.145 (0.190)	-0.094 (0.184)
Origin =VI	-0.146 (0.101)	-0.114 (0.102)	-0.079 (0.103)	-0.215** (0.092)	-0.195** (0.090)	-0.127 (0.090)
Ref. position	0.107* (0.063)	0.092 (0.063)	0.026 (0.059)	0.110* (0.062)	0.096 (0.062)	0.028 (0.058)
Ref. mobility	-0.204*** (0.054)	-0.205*** (0.052)	-0.175*** (0.050)			
Ref. mob. up				-0.253** (0.104)	-0.262** (0.102)	-0.209** (0.094)
Ref. mob. down				0.122 (0.107)	0.110 (0.107)	0.118 (0.098)
Log likelihood	-43735.964	-43688.831	-43229.570	-43734.520	-43686.880	-43228.836
AIC	87567.928	87477.661	86585.141	87567.040	87475.761	86585.671
N	19,152	19,152	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C5. Comparison of linear component estimates between DRM and OLS models for Western Europe

Tests for B2 *versus* C1 (mobility estimates for Western Europe)

	(1) Basic controls			(2) Controls for education,			(3) Controls for educ. & further demographics,			(4) Controls for educ. & further demographics, UE & income			(5) Controls for education			(6) Controls for educ. & further demographics, UE & income		
	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$
Orig. linear	0.073	n.a.	n.a.	0.064	n.a.	n.a.	0.059	n.a.	n.a.	0.030	n.a.	n.a.	0.065	n.a.	n.a.	0.030	n.a.	n.a.
Dest. linear	0.110	n.a.	n.a.	0.082	n.a.	n.a.	0.076	n.a.	n.a.	0.040	n.a.	n.a.	0.082	n.a.	n.a.	0.040	n.a.	n.a.
Mob. linear	0.056	n.a.	n.a.	0.059	n.a.	n.a.	0.049	n.a.	n.a.	0.031	n.a.	n.a.	0.070	n.a.	n.a.	0.036	n.a.	n.a.
Orig. – mob.	0.017	0.018	-0.002	0.006	0.007	-0.001	0.011	0.012	-0.002	-0.001	0.002	-0.003	-0.006	-0.006	0.000	-0.006	-0.006	0.000
Dest. + mob.	0.166	0.165	0.001	0.141	0.140	0.001	0.125	0.123	0.001	0.071	0.069	0.002	0.152	0.152	0.000	0.076	0.076	0.000

Tests for Tables B3 *versus* C2 (reference estimates for Western Europe)

	(1) Basic controls			(2) Controls for education			(3) Controls for educ., further demographics., UE & income			(4) Basic controls; positive. vs. negative ref. Mobility			(5) Controls for education; positive vs. negative. ref. Mobility			(6) Controls for educ., further demographics., UE & income; pos. vs. neg. ref. mobility		
	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$
Orig. linear	0.080	n.a.	n.a.	0.067	n.a.	n.a.	0.043	n.a.	n.a.	0.076	n.a.	n.a.	0.063	n.a.	n.a.	0.034	n.a.	n.a.
Dest. linear	0.107	n.a.	n.a.	0.081	n.a.	n.a.	0.062	n.a.	n.a.	0.112	n.a.	n.a.	0.086	n.a.	n.a.	0.074	n.a.	n.a.
Mob. linear	0.082	n.a.	n.a.	0.077	n.a.	n.a.	0.049	n.a.	n.a.	0.076	n.a.	n.a.	0.072	n.a.	n.a.	0.039	n.a.	n.a.
Orig. – mob.	-0.001	-0.002	0.001	-0.010	-0.011	0.001	-0.006	-0.005	-0.001	-0.001	-0.002	0.001	-0.009	-0.010	0.001	-0.005	-0.004	-0.001
Dest. + mob.	0.189	0.185	0.004	0.158	0.156	0.002	0.112	0.110	0.002	0.188	0.187	0.002	0.158	0.157	0.001	0.113	0.114	0.000

Table C6. Comparison of linear component estimates between DRM and OLS models for Eastern Europe

Tests for Tables B5 *versus* C3 (mobility estimates for Eastern Europe)

	(1) Basic controls			(2) Controls for education,			(3) Controls for educ. & further demographics,			(4) Controls for educ. & further demographics, UE & income			(5) Controls for education			(6) Controls for educ. & further demographics, UE & income		
	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$
Orig. linear	0.058	n.a.	n.a.	0.036	n.a.	n.a.	0.041	n.a.	n.a.	0.027	n.a.	n.a.	0.046	n.a.	n.a.	0.030	n.a.	n.a.
Dest. linear	0.097	n.a.	n.a.	0.054	n.a.	n.a.	0.051	n.a.	n.a.	0.025	n.a.	n.a.	0.049	n.a.	n.a.	0.028	n.a.	n.a.
Mob. linear	0.060	n.a.	n.a.	0.053	n.a.	n.a.	0.053	n.a.	n.a.	0.054	n.a.	n.a.	0.031	n.a.	n.a.	0.022	n.a.	n.a.
Orig. – mob.	-0.002	-0.001	-0.002	-0.017	-0.015	-0.002	-0.012	-0.012	0.000	-0.026	-0.026	-0.001	0.015	0.016	-0.001	0.008	0.009	-0.001
Dest. + mob.	0.157	0.156	0.001	0.107	0.104	0.002	0.104	0.104	0.000	0.078	0.077	0.001	0.080	0.081	-0.001	0.050	0.051	-0.002

Tests for Tables B6 *versus* C4 (reference estimates for Eastern Europe)

	(1) Basic controls			(2) Controls for education			(3) Controls for educ., further demographics., UE & income			(4) Basic controls; positive. vs. negative ref. Mobility			(5) Controls for education; positive vs. negative. ref. Mobility			(6) Controls for educ., further demographics., UE & income; pos. vs. neg. ref. mobility		
	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$	DRM	OLS	$\Delta_{drm-ols}$
Orig. linear	0.119	n.a.	n.a.	0.098	n.a.	n.a.	0.104	n.a.	n.a.	0.093	n.a.	n.a.	0.073	n.a.	n.a.	0.082	n.a.	n.a.
Dest. linear	0.139	n.a.	n.a.	0.098	n.a.	n.a.	0.058	n.a.	n.a.	0.157	n.a.	n.a.	0.114	n.a.	n.a.	0.072	n.a.	n.a.
Mob. linear	0.157	n.a.	n.a.	0.146	n.a.	n.a.	0.141	n.a.	n.a.	0.131	n.a.	n.a.	0.121	n.a.	n.a.	0.120	n.a.	n.a.
Orig. – mob.	-0.038	-0.035	-0.003	-0.048	-0.044	-0.004	-0.037	-0.034	-0.003	-0.038	-0.036	-0.002	-0.048	-0.046	-0.003	-0.038	-0.035	-0.003
Dest. + mob.	0.295	0.294	0.001	0.245	0.243	0.002	0.199	0.196	0.004	0.288	0.287	0.001	0.235	0.235	0.000	0.192	0.191	0.001

Table C7 Linear mobility estimates for Western Europe, relaxing constraints on origin and destination effects (DRM, on basis of Table 3, column (2))

	(1) Constraint on class VI loosened	(2) Constraint on class V loosened	(3) Constraint on class IV loosened	(4) Constraint on class III loosened	(5) Constraint on class II loosened
Linear position effect	0.148*** (0.016)	0.146*** (0.016)	0.147*** (0.017)	0.147*** (0.016)	0.146*** (0.017)
Dest. non-linearities					
Destination=II	0.025 (0.067)	0.037 (0.070)	0.040 (0.074)	0.047 (0.064)	0.009 (0.072)
Destination=III	0.122* (0.073)	0.125 (0.079)	0.121 (0.105)	0.250 (0.219)	0.123 (0.078)
Destination=IV	-0.071 (0.078)	-0.071 (0.083)	-0.089 (0.171)	-0.061 (0.092)	-0.071 (0.083)
Destination=V	0.061 (0.106)	0.074 (0.120)	0.082 (0.098)	0.089 (0.094)	0.078 (0.098)
Destination=VI	0.014 (0.110)	0.105 (0.119)	0.110 (0.119)	0.108 (0.085)	0.104 (0.118)
Orig. non-linearities					
Origin=II	0.025 (0.067)	0.037 (0.070)	0.040 (0.074)	0.047 (0.064)	0.090 (0.106)
Origin =III	0.122* (0.073)	0.125 (0.079)	0.121 (0.105)	0.017 (0.101)	0.123 (0.078)
Origin =IV	-0.071 (0.078)	-0.071 (0.083)	-0.046 (0.169)	-0.061 (0.092)	-0.071 (0.083)
Origin =V	0.061 (0.106)	0.093 (0.151)	0.082 (0.098)	0.089 (0.094)	0.078 (0.098)
Origin =VI	0.274 (0.399)	0.105 (0.119)	0.110 (0.119)	0.108 (0.085)	0.104 (0.118)
Steps mobile	0.030 (0.057)	0.058 (0.047)	0.066 (0.077)	0.080* (0.048)	0.059 (0.045)
p	0.265 (0.374)	0.435 (0.321)	0.488 (0.525)	0.597* (0.326)	0.431 (0.305)
Log likelihood	-83669.286	-83670.831	-83670.802	-83669.030	-83670.412
AIC	167532.573	167535.663	167535.604	167532.061	167534.824
N	40,342	40,342	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence, linear and squared terms for years of education.

Table C8 Non-linear mobility estimates for Western Europe, relaxing constraints on origin and destination effects (DRM, on basis of Table 3, column (5))

	(1) Constraint on class VI loosened	(2) Constraint on class V loosened	(3) Constraint on class IV loosened	(4) Constraint on class III loosened	(5) Constraint on class II loosened
Linear position effect	0.146*** (0.016)	0.146*** (0.017)	0.146*** (0.018)	0.147*** (0.016)	0.146*** (0.017)
Dest. non-linearities					
Destination=II	0.035 (0.062)	0.035 (0.068)	0.035 (0.069)	0.038 (0.065)	0.012 (0.080)
Destination=III	0.160** (0.080)	0.159* (0.086)	0.161* (0.085)	0.392 (0.404)	0.156* (0.088)
Destination=IV	-0.049 (0.090)	-0.053 (0.090)	-0.040 (0.122)	-0.046 (0.086)	-0.054 (0.090)
Destination=V	0.073 (0.128)	0.073 (0.122)	0.097 (0.118)	0.114 (0.096)	0.101 (0.110)
Destination=VI	0.022 (0.111)	0.120 (0.118)	0.115 (0.139)	0.117 (0.081)	0.120 (0.109)
Orig. non-linearities					
Origin=II	0.035 (0.062)	0.035 (0.068)	0.035 (0.069)	0.038 (0.065)	0.077 (0.104)
Origin =III	0.160** (0.080)	0.159* (0.086)	0.161* (0.085)	0.022 (0.096)	0.156* (0.088)
Origin =IV	-0.049 (0.090)	-0.053 (0.090)	-0.080 (0.202)	-0.046 (0.086)	-0.054 (0.090)
Origin =V	0.073 (0.128)	0.162 (0.173)	0.097 (0.118)	0.114 (0.096)	0.101 (0.110)
Origin =VI	0.305 (0.435)	0.120 (0.118)	0.115 (0.139)	0.117 (0.081)	0.120 (0.109)
Steps down=6	-0.249 (0.387)	-0.376 (0.315)	-0.368 (0.379)	-0.605* (0.352)	-0.398 (0.308)
Steps down=5	-0.172 (0.310)	-0.289 (0.242)	-0.282 (0.302)	-0.485* (0.277)	-0.323 (0.235)
Steps down=4	-0.356 (0.306)	-0.462** (0.221)	-0.462* (0.275)	-0.590*** (0.223)	-0.488** (0.218)
Steps down=3	-0.068 (0.187)	-0.134 (0.131)	-0.134 (0.180)	-0.242 (0.160)	-0.156 (0.135)
Steps down=2	-0.111 (0.143)	-0.166 (0.103)	-0.161 (0.127)	-0.248* (0.131)	-0.174 (0.109)
Steps down=1	-0.071 (0.073)	-0.078 (0.064)	-0.077 (0.076)	-0.127* (0.073)	-0.080 (0.065)
Steps up=1	0.023 (0.064)	0.042 (0.063)	0.040 (0.068)	0.085 (0.066)	0.043 (0.060)
Steps up=2	0.052 (0.115)	0.107 (0.102)	0.103 (0.118)	0.182 (0.111)	0.118 (0.095)
Steps up=3	-0.067 (0.196)	0.009 (0.154)	0.009 (0.189)	0.102 (0.139)	0.030 (0.139)
Steps up=4	0.100 (0.239)	0.214 (0.169)	0.210 (0.225)	0.344* (0.198)	0.240 (0.167)
Steps up=5	0.090 (0.291)	0.217 (0.212)	0.208 (0.286)	0.405 (0.254)	0.245 (0.210)
Steps up=6	0.293 (0.376)	0.429 (0.289)	0.419 (0.383)	0.659** (0.300)	0.453* (0.274)
p	0.270 (0.380)	0.419 (0.275)	0.409 (0.383)	0.681** (0.323)	0.446* (0.263)
Log likelihood	-83655.582	-83656.913	-83657.114	-83654.319	-83656.909
AIC	167527.163	167529.825	167530.229	167524.637	167529.819
N	40,342	40,342	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence, linear and squared terms for years of education.

Table C9 Linear mobility estimates for Eastern Europe, relaxing constraints on origin and destination effects (DRM, on basis of Table 6, column (2))

	(1) Constraint on class VI loosened	(2) Constraint on class V loosened	(3) Constraint on class IV loosened	(4) Constraint on class III loosened	(5) Constraint on class II loosened
Linear position effect	0.089*** (0.020)	0.090*** (0.019)	0.090*** (0.019)	0.090*** (0.020)	0.090*** (0.019)
Dest. non-linearities					
Destination=II	0.064 (0.115)	0.060 (0.122)	0.062 (0.122)	0.055 (0.131)	0.104 (0.084)
Destination=III	-0.265 (0.170)	-0.283* (0.148)	-0.280* (0.154)	-0.325* (0.193)	-0.281* (0.150)
Destination=IV	-0.076 (0.165)	-0.088 (0.153)	-0.039 (0.133)	-0.094 (0.151)	-0.094 (0.151)
Destination=V	-0.133 (0.289)	-0.140 (0.322)	-0.142 (0.288)	-0.147 (0.290)	-0.143 (0.286)
Destination=VI	-0.266 (0.209)	-0.304 (0.192)	-0.303 (0.191)	-0.306 (0.189)	-0.306 (0.190)
Orig. non-linearities					
Origin=II	0.064 (0.115)	0.060 (0.122)	0.062 (0.122)	0.055 (0.131)	-0.019 (0.204)
Origin =III	-0.265 (0.170)	-0.283* (0.148)	-0.280* (0.154)	-0.128 (0.321)	-0.281* (0.150)
Origin =IV	-0.076 (0.165)	-0.088 (0.153)	-0.203 (0.301)	-0.094 (0.151)	-0.094 (0.151)
Origin =V	-0.133 (0.289)	-0.149 (0.282)	-0.142 (0.288)	-0.147 (0.290)	-0.143 (0.286)
Origin =VI	-0.373 (0.302)	-0.304 (0.192)	-0.303 (0.191)	-0.306 (0.189)	-0.306 (0.190)
Steps mobile	0.047* (0.027)	0.053*** (0.014)	0.052*** (0.014)	0.056*** (0.016)	0.054*** (0.013)
p	0.339 (0.269)	0.400*** (0.121)	0.398*** (0.120)	0.427*** (0.132)	0.434*** (0.113)
Log likelihood	-51999.851	-51999.933	-51999.543	-51999.684	-51999.620
AIC	104091.702	104091.866	104091.087	104091.367	104091.239
N	19,152	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence, linear and squared terms for years of education.

Table C10. Non-linear mobility estimates for Eastern Europe, relaxing constraints on origin and destination effects (DRM, on basis of Table 6, column (5))

	(1) Constraint on class VI loosened	(2) Constraint on class V loosened	(3) Constraint on class IV loosened	(4) Constraint on class III loosened	(5) Constraint on class II loosened
Linear position effect	0.097*** (0.019)	0.096*** (0.018)	0.096*** (0.018)	0.095*** (0.016)	0.096*** (0.018)
Dest. non-linearities					
Destination=II	-0.026 (0.131)	-0.012 (0.137)	-0.013 (0.136)	-0.001 (0.149)	-0.056 (0.122)
Destination=III	-0.482*** (0.186)	-0.486** (0.192)	-0.482** (0.198)	-0.757** (0.363)	-0.495** (0.194)
Destination=IV	-0.289 (0.184)	-0.299 (0.190)	-0.241 (0.154)	-0.297 (0.212)	-0.297 (0.186)
Destination=V	-0.278 (0.289)	-0.292 (0.379)	-0.292 (0.291)	-0.279 (0.289)	-0.293 (0.290)
Destination=VI	-0.252 (0.209)	-0.347* (0.204)	-0.344* (0.205)	-0.333 (0.211)	-0.346* (0.208)
Orig. non-linearities					
Origin=II	-0.026 (0.131)	-0.012 (0.137)	-0.013 (0.136)	-0.001 (0.149)	0.035 (0.195)
Origin =III	-0.482*** (0.186)	-0.486** (0.192)	-0.482** (0.198)	-0.126 (0.227)	-0.495** (0.194)
Origin =IV	-0.289 (0.184)	-0.299 (0.190)	-0.373 (0.313)	-0.297 (0.212)	-0.297 (0.186)
Origin =V	-0.278 (0.289)	-0.292 (0.290)	-0.292 (0.291)	-0.279 (0.289)	-0.293 (0.290)
Origin =VI	-0.458* (0.255)	-0.347* (0.204)	-0.344* (0.205)	-0.333 (0.211)	-0.346* (0.208)
Steps down=6	-0.429*** (0.107)	-0.464*** (0.106)	-0.454*** (0.122)	-0.522*** (0.165)	-0.458*** (0.108)
Steps down=5	-0.502*** (0.146)	-0.527*** (0.122)	-0.518*** (0.139)	-0.589*** (0.167)	-0.539*** (0.134)
Steps down=4	-0.204 (0.218)	-0.209 (0.221)	-0.200 (0.232)	-0.304 (0.264)	-0.215 (0.220)
Steps down=3	-0.091 (0.164)	-0.107 (0.180)	-0.100 (0.186)	-0.160 (0.218)	-0.111 (0.172)
Steps down=2	-0.097 (0.101)	-0.097 (0.121)	-0.094 (0.100)	-0.114 (0.116)	-0.099 (0.103)
Steps down=1	-0.049 (0.100)	-0.066 (0.097)	-0.066 (0.097)	-0.059 (0.102)	-0.062 (0.094)
Steps up=1	-0.103 (0.086)	-0.089 (0.087)	-0.090 (0.087)	-0.100 (0.083)	-0.092 (0.089)
Steps up=2	0.269* (0.142)	0.275* (0.147)	0.274** (0.134)	0.290** (0.126)	0.278** (0.132)
Steps up=3	0.306* (0.168)	0.311* (0.161)	0.301** (0.152)	0.371** (0.148)	0.314* (0.163)
Steps up=4	0.413*** (0.113)	0.414*** (0.107)	0.405*** (0.115)	0.496*** (0.109)	0.423*** (0.107)
Steps up=5	0.107 (0.173)	0.124 (0.170)	0.114 (0.149)	0.190 (0.125)	0.132 (0.161)
Steps up=6	-0.125 (0.201)	-0.089 (0.218)	-0.098 (0.194)	-0.017 (0.164)	-0.096 (0.202)
p	0.424*** (0.105)	0.485*** (0.073)	0.469*** (0.074)	0.600*** (0.141)	0.474*** (0.074)
Log likelihood	-51974.643	-51975.353	-51975.095	-51972.951	-51975.192
AIC	104063.286	104064.706	104064.190	104059.903	104064.384
N	19,152	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence, linear and squared terms for years of education.

Table C11. Reference estimates for Western Europe, using non-linear mobility (DRM)

	(1) Basic controls + education	(2) Controls for educ., further demographics., UE & income	(3) Basic controls + education, pos. vs. neg. ref. mobility	(4) Controls for educ., further demographics., UE & income, pos. vs. neg. ref. mobility
Class=II	-0.102 (0.076)	-0.016 (0.069)	-0.094 (0.074)	0.001 (0.065)
Class=III	-0.132 (0.111)	0.009 (0.100)	-0.108 (0.113)	0.055 (0.097)
Class=IV	-0.481*** (0.128)	-0.333*** (0.128)	-0.451*** (0.133)	-0.253* (0.135)
Class=V	-0.485*** (0.150)	-0.309** (0.142)	-0.459*** (0.148)	-0.257* (0.142)
Class=VI	-0.600*** (0.184)	-0.384** (0.179)	-0.589*** (0.185)	-0.368** (0.187)
Class=VII	-0.884*** (0.196)	-0.628*** (0.185)	-0.889*** (0.196)	-0.650*** (0.185)
Steps down=6	-0.490 (0.318)	-0.247 (0.248)	-0.458 (0.309)	-0.193 (0.248)
Steps down=5	-0.386 (0.257)	-0.249 (0.190)	-0.359 (0.246)	-0.198 (0.193)
Steps down=4	-0.547** (0.229)	-0.395** (0.181)	-0.524** (0.223)	-0.350* (0.183)
Steps down=3	-0.197 (0.142)	-0.131 (0.106)	-0.180 (0.133)	-0.103 (0.105)
Steps down=2	-0.202* (0.109)	-0.137 (0.089)	-0.190* (0.104)	-0.118 (0.088)
Steps down=1	-0.098 (0.066)	-0.066 (0.054)	-0.092 (0.065)	-0.056 (0.055)
Steps up=1	0.060 (0.064)	0.030 (0.052)	0.054 (0.063)	0.021 (0.051)
Steps up=2	0.142 (0.104)	0.087 (0.082)	0.127 (0.102)	0.055 (0.083)
Steps up=3	0.074 (0.152)	0.009 (0.118)	0.050 (0.152)	-0.040 (0.123)
Steps up=4	0.293* (0.177)	0.200 (0.127)	0.264 (0.172)	0.144 (0.129)
Steps up=5	0.309 (0.222)	0.188 (0.153)	0.276 (0.212)	0.124 (0.157)
Steps up=6	0.545* (0.294)	0.395** (0.199)	0.501* (0.287)	0.314 (0.210)
Ref. position	0.045 (0.039)	0.015 (0.038)	0.044 (0.039)	0.011 (0.037)
Ref. mobility	-0.026 (0.039)	-0.063* (0.037)		
Ref. mob. up			-0.004 (0.043)	-0.017 (0.043)
Ref. mob. down			0.051 (0.059)	0.125** (0.060)
p	0.558** (0.268)	0.606*** (0.224)	0.602** (0.264)	0.719*** (0.243)
Log likelihood	-83654.941	-82197.734	-83654.237	-82194.309
AIC	167527.881	164639.469	167528.473	164634.618
N	40,342	40,342	40,342	40,342

* p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. Basic controls include region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C12. Reference estimates for Eastern Europe, using non-linear mobility (DRM)

	(1) Basic controls + education	(2) Controls for educ., further demographics., UE & income	(3) Basic controls + education, pos. vs. neg. ref. mobility	(4) Controls for educ., further demographics., UE & income, pos. vs. neg. ref. mobility
Class=II	-0.173 (0.159)	-0.197 (0.161)	-0.202 (0.170)	-0.212 (0.171)
Class=III	-0.754*** (0.215)	-0.724*** (0.193)	-0.801*** (0.232)	-0.764*** (0.226)
Class=IV	-0.909*** (0.182)	-0.811*** (0.193)	-0.972*** (0.209)	-0.863*** (0.213)
Class=V	-1.088*** (0.314)	-0.999*** (0.324)	-1.161*** (0.341)	-1.043*** (0.343)
Class=VI	-1.330*** (0.248)	-1.171*** (0.265)	-1.350*** (0.246)	-1.177*** (0.262)
Class=VII	-1.236*** (0.234)	-1.001*** (0.218)	-1.188*** (0.201)	-0.968*** (0.192)
Steps down=6	-1.122*** (0.215)	-0.964*** (0.216)	-0.952*** (0.199)	-0.864*** (0.161)
Steps down=5	-1.073*** (0.203)	-0.867*** (0.217)	-0.917*** (0.172)	-0.777*** (0.166)
Steps down=4	-0.645** (0.294)	-0.467* (0.272)	-0.532* (0.271)	-0.402* (0.237)
Steps down=3	-0.444* (0.238)	-0.364 (0.242)	-0.364* (0.188)	-0.314* (0.183)
Steps down=2	-0.310** (0.124)	-0.269** (0.108)	-0.239 (0.160)	-0.224 (0.149)
Steps down=1	-0.166 (0.105)	-0.166 (0.113)	-0.142 (0.095)	-0.152 (0.100)
Steps up=1	0.017 (0.086)	-0.000 (0.086)	-0.010 (0.099)	-0.016 (0.097)
Steps up=2	0.507*** (0.123)	0.461*** (0.130)	0.446*** (0.164)	0.429*** (0.160)
Steps up=3	0.649*** (0.154)	0.539*** (0.168)	0.572*** (0.201)	0.499** (0.206)
Steps up=4	0.867*** (0.144)	0.662*** (0.161)	0.756*** (0.153)	0.598*** (0.152)
Steps up=5	0.686*** (0.168)	0.564*** (0.168)	0.542** (0.253)	0.480** (0.226)
Steps up=6	0.577** (0.263)	0.316 (0.284)	0.420 (0.370)	0.226 (0.387)
Ref. position	0.098* (0.058)	0.035 (0.055)	0.099* (0.057)	0.036 (0.054)
Ref. mobility	-0.212*** (0.047)	-0.181*** (0.046)		
Ref. mob. up			-0.259*** (0.094)	-0.211** (0.085)
Ref. mob. down			0.127 (0.117)	0.124 (0.110)
p	0.398*** (0.087)	0.388*** (0.104)	0.513*** (0.134)	0.466*** (0.123)
Log likelihood	-51954.780	-51410.459	-51952.806	-51409.591
AIC	104027.560	102962.918	104025.613	102963.182
N	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. Basic controls include region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C13. Reference estimates for Western Europe, using non-linear mobility and OLS

	(1) Basic controls + education	(2) Controls for educ., further demographics., UE & income	(3) Basic controls + education, pos. vs. neg. ref. mobility	(4) Controls for educ., further demographics., UE & income, pos. vs. neg. ref. mobility
Orig.=II	0.054 (0.057)	0.056 (0.056)	0.051 (0.057)	0.050 (0.056)
Orig.=III	0.062 (0.085)	0.070 (0.080)	0.057 (0.085)	0.059 (0.080)
Orig.=IV	0.060 (0.096)	0.038 (0.089)	0.054 (0.097)	0.024 (0.089)
Orig.=V	0.184 (0.122)	0.174 (0.113)	0.176 (0.122)	0.156 (0.113)
Orig.=VI	0.230* (0.139)	0.212* (0.128)	0.221 (0.140)	0.192 (0.128)
Orig.=VII	0.163 (0.150)	0.152 (0.136)	0.154 (0.150)	0.131 (0.135)
Dest.=II	-0.146*** (0.051)	-0.067 (0.048)	-0.135** (0.052)	-0.041 (0.049)
Dest.=III	-0.211** (0.086)	-0.086 (0.083)	-0.190** (0.088)	-0.036 (0.086)
Dest.=IV	-0.536*** (0.107)	-0.381*** (0.100)	-0.507*** (0.109)	-0.313*** (0.103)
Dest.=V	-0.650*** (0.132)	-0.470*** (0.126)	-0.622*** (0.134)	-0.403*** (0.129)
Dest.=VI	-0.837*** (0.168)	-0.608*** (0.159)	-0.817*** (0.168)	-0.561*** (0.159)
Dest.=VII	-1.032*** (0.190)	-0.779*** (0.178)	-1.036*** (0.192)	-0.789*** (0.180)
Steps down=6	0.051 (0.233)	0.144 (0.230)	0.038 (0.234)	0.113 (0.231)
Steps down=5	0.067 (0.177)	0.086 (0.163)	0.058 (0.179)	0.065 (0.164)
Steps down=4	-0.144 (0.166)	-0.091 (0.147)	-0.152 (0.168)	-0.112 (0.149)
Steps down=3	0.091 (0.111)	0.088 (0.106)	0.086 (0.112)	0.078 (0.106)
Steps down=2	-0.022 (0.101)	-0.008 (0.091)	-0.025 (0.101)	-0.014 (0.091)
Steps down=1	-0.021 (0.061)	-0.014 (0.058)	-0.022 (0.061)	-0.016 (0.058)
Steps up=1	-0.021 (0.054)	-0.025 (0.050)	-0.020 (0.054)	-0.021 (0.050)
Steps up=2	-0.040 (0.071)	-0.045 (0.066)	-0.040 (0.071)	-0.044 (0.066)
Steps up=3	-0.228** (0.091)	-0.219** (0.085)	-0.227** (0.091)	-0.218** (0.085)
Steps up=4	-0.106 (0.106)	-0.102 (0.097)	-0.105 (0.106)	-0.100 (0.097)
Steps up=5	-0.155 (0.123)	-0.157 (0.112)	-0.154 (0.123)	-0.155 (0.112)
Steps up=6	0.000 (.)	0.000 (.)	0.000 (.)	0.000 (.)
Ref. position	-0.051 (0.040)	-0.019 (0.038)	-0.049 (0.039)	-0.015 (0.038)
Ref. mobility	-0.025 (0.040)	-0.065* (0.037)		
Ref. mob. up			-0.004 (0.043)	-0.014 (0.043)
Ref. mob. down			0.054 (0.063)	0.133** (0.057)
Log likelihood	-86589.718	-85081.011	-86589.004	-85076.803
AIC	173403.436	170412.022	173404.008	170405.606
N	40,342	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. Basic controls include region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C14. Reference estimates for Eastern Europe, using non-linear mobility and OLS

	(1) Basic controls + education	(2) Controls for educ., further demographics., UE & income	(3) Basic controls + education, pos. vs. neg. ref. mobility	(4) Controls for educ., further demographics., UE & income, pos. vs. neg. ref. mobility
Orig.=II	0.019 (0.119)	-0.045 (0.117)	0.025 (0.118)	-0.042 (0.115)
Orig.=III	-0.140 (0.193)	-0.188 (0.196)	-0.127 (0.187)	-0.181 (0.189)
Orig.=IV	-0.267 (0.175)	-0.327* (0.186)	-0.253 (0.175)	-0.319* (0.182)
Orig.=V	-0.288 (0.245)	-0.437 (0.280)	-0.276 (0.236)	-0.430 (0.271)
Orig.=VI	-0.356 (0.225)	-0.516* (0.273)	-0.340 (0.214)	-0.506* (0.261)
Orig.=VII	-0.178 (0.283)	-0.314 (0.326)	-0.160 (0.270)	-0.304 (0.312)
Dest.=II	-0.210** (0.083)	-0.175* (0.104)	-0.235*** (0.087)	-0.189* (0.106)
Dest.=III	-0.578*** (0.127)	-0.498*** (0.148)	-0.619*** (0.145)	-0.522*** (0.156)
Dest.=IV	-0.645*** (0.181)	-0.514** (0.223)	-0.720*** (0.202)	-0.557** (0.223)
Dest.=V	-0.794** (0.325)	-0.564 (0.369)	-0.873*** (0.305)	-0.610* (0.344)
Dest.=VI	-0.946*** (0.260)	-0.627** (0.314)	-0.989*** (0.252)	-0.651** (0.300)
Dest.=VII	-1.046*** (0.281)	-0.694** (0.329)	-1.008*** (0.294)	-0.672* (0.347)
Steps down=6	-0.550** (0.265)	-0.650* (0.335)	-0.533** (0.254)	-0.640* (0.322)
Steps down=5	-0.629** (0.268)	-0.643** (0.309)	-0.617** (0.259)	-0.636** (0.300)
Steps down=4	-0.325 (0.292)	-0.347 (0.305)	-0.318 (0.287)	-0.343 (0.300)
Steps down=3	-0.181 (0.257)	-0.232 (0.280)	-0.180 (0.257)	-0.231 (0.280)
Steps down=2	-0.121 (0.109)	-0.159 (0.097)	-0.115 (0.110)	-0.155 (0.098)
Steps down=1	-0.046 (0.115)	-0.078 (0.131)	-0.045 (0.115)	-0.078 (0.131)
Steps up=1	-0.103 (0.084)	-0.087 (0.081)	-0.104 (0.084)	-0.088 (0.081)
Steps up=2	0.329** (0.134)	0.372*** (0.140)	0.326** (0.135)	0.370** (0.140)
Steps up=3	0.406** (0.161)	0.454** (0.186)	0.404** (0.161)	0.453** (0.186)
Steps up=4	0.549*** (0.188)	0.538** (0.228)	0.547*** (0.188)	0.537** (0.227)
Steps up=5	0.244 (0.167)	0.346 (0.220)	0.242 (0.168)	0.344 (0.220)
Steps up=6	n.a.	n.a.	n.a.	n.a.
Ref. position	0.098* (0.057)	0.033 (0.054)	0.102* (0.057)	0.035 (0.054)
Ref. mobility	-0.204*** (0.050)	-0.175*** (0.048)		
Ref. mob. up			-0.256** (0.099)	-0.205** (0.091)
Ref. mob. down			0.115 (0.107)	0.124 (0.098)
Log likelihood	-43666.202	-43206.797	-43664.521	-43206.221
AIC	87454.405	86561.593	87453.041	86562.442
N	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. Basic controls include region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C15. Alternative asymmetric reference mobility estimates & reference position estimates using shares (Western Europe; DRM)

	(1) Shares in class I & class VII	(2) Shares in class I+II & class VI+VII	(3) Reference mobility, mean distances up and down
Class=II	-0.016 (0.069)	-0.016 (0.066)	-0.016 (0.065)
Class=III	-0.022 (0.098)	-0.015 (0.096)	-0.035 (0.094)
Class=IV	-0.356*** (0.116)	-0.347*** (0.116)	-0.364*** (0.116)
Class=V	-0.331** (0.145)	-0.323** (0.138)	-0.338** (0.139)
Class=VI	-0.398** (0.180)	-0.393** (0.175)	-0.382** (0.175)
Class=VII	-0.625*** (0.188)	-0.628*** (0.181)	-0.637*** (0.185)
Steps mobile	0.050* (0.029)	0.052* (0.029)	0.047* (0.026)
Ref. position (% at top)	0.085 (0.217)	0.175 (0.159)	
Ref. position (% at bottom)	-0.073 (0.212)	0.104 (0.169)	
Ref. mobility (same as table 7)	-0.064* (0.037)	-0.065* (0.037)	
Ref. position (same as table 7)			0.010 (0.038)
Ref. mob. mean upward			0.008 (0.048)
Ref. mob. mean downward			0.138** (0.058)
p	0.590** (0.233)	0.582*** (0.221)	0.601*** (0.196)
Log likelihood	-82208.483	-82207.706	-82204.159
AIC	164640.966	164639.412	164632.318
N	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence, linear and squared terms for years of education, unemployment and income.

Table C16. Alternative asymmetric reference mobility estimates & reference position estimates using shares (Eastern Europe; DRM)

	(1) Shares in class I & class VII	(2) Shares in class I+II & class VI+VII	(3) Reference mobility, mean distances up and down
Class=II	-0.148 (0.175)	-0.148 (0.175)	-0.147 (0.175)
Class=III	-0.552*** (0.164)	-0.567*** (0.168)	-0.538*** (0.129)
Class=IV	-0.617*** (0.191)	-0.630*** (0.196)	-0.609*** (0.182)
Class=V	-0.840** (0.366)	-0.853** (0.362)	-0.827** (0.337)
Class=VI	-1.099*** (0.270)	-1.091*** (0.269)	-1.095*** (0.266)
Class=VII	-0.976*** (0.229)	-0.969*** (0.256)	-0.953*** (0.220)
Steps mobile	0.141*** (0.045)	0.142*** (0.043)	0.137*** (0.038)
Ref. position (% at top)	0.227 (0.330)	-0.078 (0.343)	
Ref. position (% at bottom)	-0.067 (0.380)	-0.263 (0.387)	
Ref. mobility (same as table 7)	-0.176*** (0.047)	-0.175*** (0.048)	
Ref. position (same as table 7)			0.034 (0.060)
Ref. mob. mean upward			-0.223** (0.110)
Ref. mob. mean downward			0.110 (0.103)
p	0.361** (0.179)	0.353** (0.171)	0.380** (0.152)
Log likelihood	-51434.987	-51434.517	-51434.159
AIC	102991.973	102991.034	102990.318
N	19,152	19,152	19,152

* p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence, linear and squared terms for years of education, unemployment and income.

Table C17. Reference estimates for Western Europe, shares in reference group moving a minimal number of steps (DRM)

	(1) Minimum one step up or down	(2) Minimum two steps up or down	(3) Minimum three steps up or down	(4) Minimum four steps up or down	(5) Minimum five steps up or down	(6) Minimum six steps up or down
Class=II	0.016 (0.073)	0.007 (0.062)	0.007 (0.062)	0.025 (0.061)	-0.056 (0.064)	-0.005 (0.074)
Class=III	0.015 (0.115)	-0.027 (0.088)	0.026 (0.090)	0.006 (0.081)	-0.160 (0.102)	0.025 (0.093)
Class=IV	-0.286** (0.122)	-0.389*** (0.112)	-0.292*** (0.103)	-0.367*** (0.109)	-0.449*** (0.109)	-0.271** (0.111)
Class=V	-0.207 (0.134)	-0.312** (0.124)	-0.235 (0.146)	-0.372*** (0.117)	-0.349*** (0.113)	-0.192* (0.106)
Class=VI	-0.205 (0.157)	-0.324** (0.161)	-0.258 (0.170)	-0.479*** (0.181)	-0.498*** (0.138)	-0.258** (0.131)
Class=VII	-0.369* (0.191)	-0.511*** (0.146)	-0.470*** (0.149)	-0.686*** (0.142)	-0.838*** (0.150)	-0.565*** (0.130)
Steps mobile	0.032 (0.024)	0.036 (0.022)	0.036* (0.020)	0.041 (0.035)	0.053** (0.026)	0.033* (0.019)
Ref. position	0.013 (0.038)	0.009 (0.038)	0.010 (0.038)	0.008 (0.038)	0.007 (0.038)	0.009 (0.037)
Ref. mob. up	0.144 (0.163)	0.100 (0.150)	0.211 (0.195)	-0.484** (0.225)	-0.979*** (0.275)	-0.088 (0.436)
Ref. mob. down	0.122 (0.189)	0.395** (0.188)	0.434* (0.250)	0.850*** (0.283)	1.477*** (0.443)	3.122*** (0.894)
p	0.565** (0.250)	0.632*** (0.199)	0.598** (0.233)	0.697*** (0.269)	0.624*** (0.187)	0.674*** (0.178)
Log likelihood	-82211.497	-82206.248	-82205.886	-82198.444	-82184.218	-82193.536
AIC	164646.994	164636.496	164635.771	164620.888	164592.436	164611.071
N	40,342	40,342	40,342	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence, linear and squared terms for years of education.

Table C18. Reference estimates for Eastern Europe, shares in reference group moving a minimal number of steps (DRM)

	(1) Minimum one step up or down	(2) Minimum two steps up or down	(3) Minimum three steps up or down	(4) Minimum four steps up or down	(5) Minimum five steps up or down	(6) Minimum six steps up or down
Class=II	-0.122**	-0.004	-0.002	0.003	-0.118	-0.122**
	(0.059)	(0.187)	(0.167)	(0.161)	(0.162)	(0.059)
Class=III	-0.200**	-0.348**	-0.276**	-0.503***	-0.741***	-0.200**
	(0.099)	(0.147)	(0.138)	(0.174)	(0.227)	(0.099)
Class=IV	-0.373***	-0.308*	-0.510***	-0.570**	-0.526***	-0.373***
	(0.079)	(0.177)	(0.174)	(0.252)	(0.197)	(0.079)
Class=V	-0.532***	-0.456	-0.946**	-0.706	-0.615*	-0.532***
	(0.132)	(0.308)	(0.370)	(0.430)	(0.326)	(0.132)
Class=VI	-0.639***	-0.766***	-1.034***	-0.885***	-0.839***	-0.639***
	(0.188)	(0.208)	(0.249)	(0.310)	(0.288)	(0.188)
Class=VII	-0.623	-0.495***	-0.797***	-0.612***	-0.653***	-0.623
	(0.395)	(0.156)	(0.236)	(0.225)	(0.198)	(0.395)
Steps mobile	0.191	0.082***	0.095***	0.084***	0.086**	0.191
	(0.516)	(0.027)	(0.026)	(0.027)	(0.034)	(0.516)
Ref. position	0.033	0.038	0.037	0.042	0.034	0.033
	(0.057)	(0.060)	(0.056)	(0.056)	(0.056)	(0.057)
Ref. mob. up	-0.530	-0.483	-0.780	-0.932**	-0.973**	-0.530
	(0.416)	(0.335)	(0.485)	(0.445)	(0.432)	(0.416)
Ref. mob. down	0.013	0.111	0.857**	0.326	0.703	0.013
	(0.400)	(0.315)	(0.335)	(0.399)	(0.817)	(0.400)
p	-0.497	0.509***	0.601***	0.592***	0.520***	-0.497
	(4.161)	(0.189)	(0.131)	(0.101)	(0.169)	(4.161)
Log likelihood	-51445.477	-51441.339	-51432.105	-51438.950	-51440.416	-51445.477
AIC	103012.954	103004.678	102986.211	102999.900	103002.833	103012.954
N	19,152	19,152	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence, linear and squared terms for years of education.

Table C19. Reference estimates for Western Europe, interactions with own mobility (DRM)

	(1) Basic model	(2) Controls for education	(3) Controls for educ. & further demographics	(4) Controls for educ., further demographics., UE & income
Class=II	-0.153* (0.078)	-0.102 (0.080)	-0.096 (0.073)	-0.022 (0.066)
Class=III	-0.272*** (0.098)	-0.175* (0.106)	-0.147 (0.096)	-0.029 (0.097)
Class=IV	-0.657*** (0.122)	-0.507*** (0.126)	-0.530*** (0.115)	-0.366*** (0.116)
Class=V	-0.679*** (0.145)	-0.517*** (0.146)	-0.519*** (0.143)	-0.341** (0.139)
Class=VI	-0.833*** (0.183)	-0.624*** (0.185)	-0.637*** (0.179)	-0.406** (0.177)
Class=VII	-1.123*** (0.197)	-0.890*** (0.196)	-0.870*** (0.191)	-0.632*** (0.185)
Steps mobile	0.083 (0.073)	0.081 (0.056)	0.073 (0.047)	0.051* (0.031)
Ref. position	0.053 (0.040)	0.045 (0.040)	0.049 (0.039)	0.014 (0.038)
Ref. mobility	-0.023 (0.054)	-0.004 (0.053)	-0.024 (0.052)	-0.052 (0.051)
Steps mobile*Ref. mobility	-0.002 (0.005)	-0.003 (0.005)	-0.003 (0.005)	-0.002 (0.005)
p	0.569 (0.366)	0.530 (0.350)	0.546* (0.289)	0.582** (0.237)
Log likelihood	-83707.362	-83668.063	-82971.910	-82208.547
AIC	167608.723	167534.127	166163.820	164641.095
N	40,342	40,342	40,342	40,342

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C20. Reference estimates for Eastern Europe, interactions with own mobility (DRM)

	(1) Basic model	(2) Controls for education	(3) Controls for educ. & further demographics	(4) Controls for educ., further demographics., UE & income
Class=II	-0.198 (0.128)	-0.110 (0.151)	-0.143 (0.147)	-0.145 (0.179)
Class=III	-0.751*** (0.176)	-0.564*** (0.174)	-0.631*** (0.167)	-0.568*** (0.159)
Class=IV	-0.922*** (0.176)	-0.687*** (0.184)	-0.748*** (0.177)	-0.621*** (0.184)
Class=V	-1.182*** (0.296)	-0.908*** (0.337)	-0.927*** (0.343)	-0.837** (0.348)
Class=VI	-1.578*** (0.200)	-1.241*** (0.237)	-1.240*** (0.249)	-1.094*** (0.272)
Class=VII	-1.537*** (0.251)	-1.167*** (0.242)	-1.179*** (0.225)	-0.960*** (0.234)
Steps mobile	0.151*** (0.038)	0.140*** (0.040)	0.150*** (0.041)	0.136*** (0.039)
Ref. position	-0.109* (0.062)	-0.096 (0.062)	-0.097* (0.059)	-0.032 (0.059)
Ref. mobility	-0.170** (0.067)	-0.160** (0.068)	-0.170** (0.078)	-0.147* (0.080)
Steps mobile*Ref. mobility	-0.004 (0.008)	-0.005 (0.008)	-0.004 (0.009)	-0.004 (0.009)
p	0.562*** (0.105)	0.531*** (0.128)	0.466*** (0.138)	0.390*** (0.143)
Log likelihood	-52035.742	-51980.054	-51798.264	-51434.793
AIC	104163.485	104056.108	103714.528	102991.586
N	19,152	19,152	19,152	19,152

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C21. Origin and destination estimates for Western Europe, waves 4+5 only (OLS)

	(1) Basic model	(2) Class destination added	(3) Controls for education	(4) Controls for educ. & further demographics	(5) Controls for educ., further demographics, UE & income
Orig.=II	-0.045 (0.066)	0.035 (0.065)	0.044 (0.066)	0.044 (0.066)	0.072 (0.065)
Orig.=III	-0.086 (0.093)	0.020 (0.091)	0.037 (0.089)	0.038 (0.085)	0.058 (0.088)
Orig.=IV	-0.240** (0.092)	-0.077 (0.087)	-0.051 (0.088)	-0.095 (0.084)	-0.066 (0.083)
Orig.=V	-0.320*** (0.098)	-0.103 (0.093)	-0.074 (0.093)	-0.070 (0.091)	-0.003 (0.093)
Orig.=VI	-0.426*** (0.086)	-0.155* (0.082)	-0.110 (0.084)	-0.124 (0.077)	-0.029 (0.079)
Orig.=VII	-0.383*** (0.089)	-0.089 (0.084)	-0.041 (0.086)	-0.061 (0.085)	0.004 (0.083)
Orig.=SEI	-0.163** (0.081)	-0.010 (0.077)	0.017 (0.078)	-0.019 (0.078)	0.025 (0.076)
Orig.=SEII	-0.098 (0.097)	0.106 (0.092)	0.144 (0.093)	0.042 (0.094)	0.067 (0.090)
Dest.=II		-0.233*** (0.068)	-0.214*** (0.069)	-0.195*** (0.066)	-0.047 (0.063)
Dest.=III		-0.302*** (0.081)	-0.261*** (0.084)	-0.223*** (0.081)	-0.020 (0.077)
Dest.=IV		-0.507*** (0.092)	-0.431*** (0.100)	-0.397*** (0.098)	-0.112 (0.097)
Dest.=V		-0.743*** (0.097)	-0.665*** (0.096)	-0.627*** (0.095)	-0.303*** (0.085)
Dest.=VI		-0.955*** (0.099)	-0.850*** (0.105)	-0.795*** (0.103)	-0.399*** (0.101)
Dest.=VII		-0.988*** (0.093)	-0.871*** (0.092)	-0.781*** (0.092)	-0.378*** (0.094)
Dest.=SEI		-0.416*** (0.085)	-0.349*** (0.086)	-0.329*** (0.080)	-0.095 (0.076)
Dest.=SEII		-0.168 (0.152)	-0.075 (0.151)	-0.217 (0.149)	0.129 (0.152)
Log likelihood	-49001.546	-48749.495	-48730.695	-48317.517	-47868.460
AIC	98159.091	97670.990	97637.390	96833.034	95938.919
N	22,672	22,672	22,672	22,672	22,672

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C22. Mobility estimates for Western Europe, waves 4+5 only (DRM)

	(1) Linear mobility	(2) Linear mobility, controls for education	(3) Linear mobility, controls for educ. & further demographics	(4) Linear mobility, controls for educ., further demographics, UE & income	(5) Non-linear mobility, controls for educ.	(6) Non-linear mobility, controls for educ., further demographics, UE & income
Class=II	-0.256** (0.104)	-0.221** (0.100)	-0.193* (0.112)	-0.044 (0.138)	-0.237** (0.113)	-0.047 (0.128)
Class=III	-0.330** (0.128)	-0.260** (0.132)	-0.223* (0.130)	-0.023 (0.181)	-0.239* (0.144)	-0.011 (0.208)
Class=IV	-0.648*** (0.121)	-0.526*** (0.123)	-0.504*** (0.130)	-0.185* (0.100)	-0.523*** (0.161)	-0.176 (0.169)
Class=V	-0.842*** (0.125)	-0.709*** (0.122)	-0.687*** (0.116)	-0.301*** (0.099)	-0.694*** (0.163)	-0.273*** (0.106)
Class=VI	-1.136*** (0.145)	-0.967*** (0.153)	-0.919*** (0.152)	-0.467*** (0.141)	-0.946*** (0.151)	-0.443*** (0.146)
Class=VII	-1.161*** (0.134)	-0.977*** (0.137)	-0.885*** (0.134)	-0.433*** (0.162)	-0.967*** (0.140)	-0.434*** (0.167)
Steps mobile	0.025 (0.088)	0.020 (0.089)	0.023 (0.073)	0.001 (0.076)		
Steps down=6					-0.618 (0.652)	-0.311 (0.660)
Steps down=5					-0.113 (0.486)	0.086 (0.547)
Steps down=4					-0.689 (0.430)	-0.451 (0.445)
Steps down=3					-0.094 (0.255)	0.007 (0.270)
Steps down=2					-0.082 (0.190)	0.019 (0.185)
Steps down=1					-0.025 (0.107)	0.012 (0.099)
Steps up=1					0.031 (0.114)	-0.013 (0.102)
Steps up=2					0.037 (0.229)	-0.025 (0.257)
Steps up=3					-0.020 (0.321)	-0.108 (0.346)
Steps up=4					0.222 (0.377)	0.119 (0.408)
Steps up=5					0.156 (0.434)	0.007 (0.455)
Steps up=6					0.280 (0.516)	0.133 (0.540)
p	0.766* (0.436)	0.806 (0.521)	0.757* (0.448)	0.933 (0.885)	0.648 (0.533)	0.817 (1.141)
Log likelihood	-35274.757	-35260.005	-34962.361	-34667.877	-35240.979	-34652.251
AIC	70707.515	70682.010	70112.721	69527.754	70665.958	69514.501
N	15,941	15,941	15,941	15,941	15,941	15,941

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C23. Reference estimates for Western Europe, waves 4+5 only (DRM)

	(1) Basic controls + education	(2) Controls for educ., further demographics., UE & income	(3) Basic controls + education, pos. vs. neg. ref. mobility	(4) Controls for educ., further demographics., UE & income, pos. vs. neg. ref. mobility
Class=II	-0.217** (0.090)	-0.078 (0.071)	-0.198*** (0.075)	-0.040 (0.071)
Class=III	-0.258 (0.158)	-0.089 (0.117)	0.229** (0.124)	0.018 (0.118)
Class=IV	-0.503*** (0.157)	-0.264* (0.140)	-0.453*** (0.156)	-0.171 (0.146)
Class=V	-0.681*** (0.228)	-0.430** (0.144)	-0.632*** (0.214)	-0.345* (0.205)
Class=VI	-0.928*** (0.299)	-0.623*** (0.205)	-0.899*** (0.271)	-0.585** (0.259)
Class=VII	-0.964*** (0.301)	-0.652*** (0.258)	-1.004*** (0.302)	-0.712*** (0.278)
Steps mobile	0.019 (0.135)	0.003 (0.025)	0.005 (0.026)	0.001 (0.025)
Ref. position	0.055 (0.067)	0.024 (0.063)	0.054 (0.067)	0.018 (0.062)
Ref. mobility	-0.021 (0.066)	-0.067 (0.063)		
Ref. mob. up			0.022 (0.072)	0.015 (0.068)
Ref. mob. down			0.079 (0.101)	0.174* (0.091)
p	0.915 (0.824)	1.000 (.)	1.000 (.)	1.000 (.)
Log likelihood	-35259.028	-34666.172	-35258.031	-34662.347
AIC	70690.055	69522.344	70682.063	69516.694
N	15,941	15,941	15,941	15,941

* p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. Basic controls include region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C24. Origin and destination estimates for Eastern Europe, waves 4+5 only (OLS)

	(1) Basic model	(2) Class destination added	(3) Controls for education	(4) Controls for educ. & further demographics	(5) Controls for educ., further demographics, UE & income
Orig.=II	-0.152 (0.137)	-0.076 (0.132)	-0.061 (0.139)	-0.057 (0.119)	-0.078 (0.126)
Orig.=III	0.004 (0.145)	0.096 (0.155)	0.126 (0.144)	0.122 (0.145)	0.100 (0.155)
Orig.=IV	-0.236 (0.170)	-0.132 (0.163)	-0.103 (0.171)	-0.114 (0.142)	-0.067 (0.139)
Orig.=V	-0.138 (0.171)	0.042 (0.170)	0.080 (0.178)	0.058 (0.166)	0.054 (0.159)
Orig.=VI	-0.339** (0.142)	-0.118 (0.133)	-0.052 (0.152)	-0.081 (0.140)	-0.055 (0.128)
Orig.=VII	-0.243*** (0.068)	0.016 (0.071)	0.098 (0.069)	0.084 (0.062)	0.129** (0.063)
Orig.=SEI	-0.226 (0.277)	-0.016 (0.274)	0.048 (0.260)	-0.009 (0.261)	0.042 (0.279)
Orig.=SEII	0.043 (0.141)	0.235 (0.144)	0.355** (0.158)	0.330** (0.143)	0.365** (0.137)
Dest.=II		-0.065 (0.065)	0.016 (0.071)	0.002 (0.074)	0.025 (0.086)
Dest.=III		-0.399*** (0.072)	-0.251*** (0.077)	-0.283*** (0.087)	-0.228** (0.089)
Dest.=IV		-0.330*** (0.095)	-0.140 (0.098)	-0.178 (0.112)	-0.116 (0.121)
Dest.=V		-0.575*** (0.202)	-0.332 (0.239)	-0.307 (0.238)	-0.215 (0.244)
Dest.=VI		-0.823*** (0.108)	-0.542*** (0.153)	-0.498*** (0.169)	-0.377** (0.183)
Dest.=VII		-0.747*** (0.074)	-0.441*** (0.065)	-0.417*** (0.080)	-0.290*** (0.090)
Dest.=SEI		-0.074 (0.167)	0.098 (0.185)	0.099 (0.179)	0.065 (0.166)
Dest.=SEII		0.084 (0.204)	0.397* (0.224)	0.400* (0.230)	0.468* (0.234)
Log likelihood	-32662.936	-32542.510	-32504.358	-32394.245	-32215.624
AIC	65395.872	65171.020	65098.716	64878.490	64521.248
N	14,333	14,333	14,333	14,333	14,333

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C25. Mobility estimates for Eastern Europe, waves 4+5 only (DRM)

	(1) Linear mobility	(2) Linear mobility, controls for education	(3) Linear mobility, controls for educ. & further demographics	(4) Linear mobility, controls for educ., further demographics, UE & income	(5) Non-linear mobility, controls for educ.	(6) Non-linear mobility, controls for educ., further demographics, UE & income
Class=II	-0.058 (0.153)	0.049 (0.168)	0.021 (0.190)	0.033 (0.242)	-0.092 (0.219)	-0.102 (0.225)
Class=III	-0.472*** (0.119)	-0.280** (0.130)	-0.354** (0.140)	-0.280** (0.141)	-0.521** (0.250)	-0.487** (0.246)
Class=IV	-0.337* (0.186)	-0.096 (0.250)	-0.189 (0.306)	-0.096 (0.306)	-0.494 (0.317)	-0.440 (0.276)
Class=V	-0.499* (0.282)	-0.191 (0.337)	-0.191 (0.362)	-0.086 (0.392)	-0.416 (0.363)	-0.320 (0.330)
Class=VI	-0.948*** (0.246)	-0.584* (0.346)	-0.594* (0.343)	-0.448 (0.352)	-0.688** (0.293)	-0.536* (0.278)
Class=VII	-0.723*** (0.121)	-0.328*** (0.081)	-0.324*** (0.087)	-0.148 (0.102)	-0.378*** (0.083)	-0.194** (0.087)
Steps mobile	0.050 (0.032)	0.043 (0.026)	0.046* (0.025)	0.044** (0.017)		
Steps down=6					-0.457*** (0.117)	-0.437*** (0.110)
Steps down=5					-0.483*** (0.166)	-0.356** (0.148)
Steps down=4					-0.054 (0.246)	-0.004 (0.237)
Steps down=3					0.040 (0.194)	0.079 (0.169)
Steps down=2					-0.208 (0.131)	-0.207 (0.132)
Steps down=1					0.002 (0.144)	0.014 (0.143)
Steps up=1					-0.100 (0.100)	-0.106 (0.091)
Steps up=2					0.389** (0.166)	0.387** (0.170)
Steps up=3					0.439** (0.204)	0.427** (0.213)
Steps up=4					0.358** (0.155)	0.265* (0.143)
Steps up=5					0.249 (0.214)	0.242 (0.205)
Steps up=6					0.028 (0.265)	-0.087 (0.321)
p	0.711*** (0.212)	0.730** (0.304)	0.642** (0.283)	0.624** (0.282)	0.481*** (0.141)	0.464*** (0.110)
Log likelihood	-34080.604	-34039.053	-33938.410	-33756.783	-34016.221	-33730.564
AIC	68233.209	68154.105	67966.820	67603.566	68122.443	67553.128
N	12,726	12,726	12,726	12,726	12,726	12,726

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

Table C26. Reference estimates for Eastern Europe, waves 4+5 only (DRM)

	(1) Basic controls + education	(2) Controls for educ., further demographics., UE & income	(3) Basic controls + education, pos. vs. neg. ref. mobility	(4) Controls for educ., further demographics., UE & income, pos. vs. neg. ref. mobility
Class=II	-0.088 (0.201)	-0.127 (0.377)	-0.104 (0.274)	-0.134 (0.380)
Class=III	-0.422*** (0.129)	-0.433*** (0.160)	-0.441** (0.189)	-0.440** (0.182)
Class=IV	-0.431 (0.265)	-0.462 (0.481)	-0.469 (0.383)	-0.476 (0.472)
Class=V	-0.649** (0.293)	-0.563 (0.564)	-0.679* (0.374)	-0.580 (0.534)
Class=VI	-1.173*** (0.356)	-1.057** (0.460)	-1.199*** (0.407)	-1.067** (0.453)
Class=VII	-1.108*** (0.290)	-0.926*** (0.282)	-1.097*** (0.260)	-0.928*** (0.290)
Steps mobile	0.088 (0.119)	0.107 (0.114)	0.096 (0.099)	0.109 (0.099)
Ref. position	0.029 (0.081)	0.009 (0.078)	0.031 (0.077)	0.010 (0.074)
Ref. mobility	-0.197*** (0.056)	-0.187*** (0.053)		
Ref. mob. up			-0.212*** (0.080)	-0.194*** (0.071)
Ref. mob. down			0.176 (0.116)	0.179* (0.100)
p	0.715 (0.524)	0.544 (0.575)	0.678 (0.445)	0.535 (0.501)
Log likelihood	-34029.774	-33748.268	-34029.684	-33748.243
AIC	68139.548	67588.536	68141.368	67586.485
N	12,726	12,726	12,726	12,726

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. Basic controls include region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education include linear and squared terms for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared). 'UE' stands for unemployment. 'Income' is measured as real log household income.

D. Auxiliary regressions

Table D1. Predicting education, income and unemployment in Western Europe (OLS)

	(1) Dep. variable: education; basic controls	(2) Dep. variable: income; basic controls	(3) Dep. variable: income; controls for education	(4) Dep. variable: income; controls for educ. & further demographics	(5) Dep. variable: unemployment; controls for education	(6) Dep. variable: unemployment; controls for educ. & further demographics
Orig.=II	-0.992*** (0.071)	-0.030*** (0.011)	-0.016 (0.011)	-0.019* (0.010)	-0.001 (0.004)	0.000 (0.004)
Orig.=III	-1.485*** (0.086)	-0.073*** (0.015)	-0.051*** (0.015)	-0.046*** (0.014)	-0.004 (0.005)	-0.005 (0.005)
Orig.=IV	-2.425*** (0.086)	-0.095*** (0.015)	-0.060*** (0.015)	-0.065*** (0.014)	0.001 (0.004)	0.002 (0.004)
Orig.=V	-2.634*** (0.085)	-0.093*** (0.017)	-0.055*** (0.017)	-0.059*** (0.016)	-0.009* (0.005)	-0.008 (0.005)
Orig.=VI	-3.486*** (0.087)	-0.169*** (0.015)	-0.117*** (0.014)	-0.127*** (0.014)	0.004 (0.005)	0.006 (0.005)
Orig.=VII	-3.841*** (0.088)	-0.185*** (0.014)	-0.127*** (0.014)	-0.133*** (0.014)	0.000 (0.004)	0.002 (0.004)
Orig.=SEI	-2.230*** (0.087)	-0.090*** (0.013)	-0.058*** (0.013)	-0.061*** (0.012)	-0.005 (0.004)	-0.003 (0.004)
Orig.=SEII	-3.425*** (0.104)	-0.126*** (0.014)	-0.075*** (0.014)	-0.093*** (0.013)	-0.022*** (0.004)	-0.017*** (0.004)
Dest.=II		-0.256*** (0.013)	-0.227*** (0.014)	-0.216*** (0.013)	-0.000 (0.003)	-0.001 (0.003)
Dest.=III		-0.403*** (0.016)	-0.339*** (0.017)	-0.318*** (0.016)	0.023*** (0.004)	0.021*** (0.004)
Dest.=IV		-0.550*** (0.017)	-0.450*** (0.018)	-0.422*** (0.018)	0.018*** (0.005)	0.016*** (0.004)
Dest.=V		-0.645*** (0.017)	-0.544*** (0.018)	-0.512*** (0.018)	0.036*** (0.004)	0.033*** (0.004)
Dest.=VI		-0.782*** (0.021)	-0.649*** (0.021)	-0.598*** (0.019)	0.062*** (0.006)	0.058*** (0.006)
Dest.=VII		-0.816*** (0.018)	-0.674*** (0.019)	-0.610*** (0.019)	0.066*** (0.005)	0.061*** (0.005)
Dest.=SEI		-0.447*** (0.020)	-0.356*** (0.021)	-0.339*** (0.021)	-0.003 (0.004)	-0.004 (0.004)
Dest.=SEII		-0.825*** (0.037)	-0.690*** (0.035)	-0.667*** (0.037)	-0.017*** (0.006)	-0.017*** (0.006)
Log likelihood	-155725.460	-67064.973	-66703.276	-63045.259	3509.108	3759.207
AIC	311636.921	134331.945	133612.552	126318.517	-6812.216	-7290.414
N	57,868	57,868	57,868	57,868	57,868	57,868

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education includes linear and squared term for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared).

Table D2. Predicting education, income and unemployment in Eastern Europe (OLS)

	(1) Dep. variable: education; basic controls	(2) Dep. variable: income; basic controls	(3) Dep. variable: income; controls for education	(4) Dep. variable: income; controls for educ. & further demographics	(5) Dep. variable: unemployment; controls for education	(6) Dep. variable: unemployment; controls for educ. & further demographics
Orig.=II	-0.694*** (0.112)	-0.058** (0.026)	-0.045* (0.026)	-0.046* (0.024)	-0.005 (0.006)	-0.005 (0.006)
Orig.=III	-0.811*** (0.139)	-0.113*** (0.039)	-0.094** (0.039)	-0.082** (0.039)	0.005 (0.010)	0.005 (0.010)
Orig.=IV	-1.481*** (0.118)	-0.187*** (0.029)	-0.152*** (0.029)	-0.147*** (0.026)	-0.004 (0.008)	-0.003 (0.008)
Orig.=V	-1.826*** (0.128)	-0.123*** (0.025)	-0.082*** (0.025)	-0.069*** (0.024)	0.004 (0.010)	0.005 (0.010)
Orig.=VI	-2.366*** (0.119)	-0.161*** (0.025)	-0.105*** (0.024)	-0.088*** (0.025)	0.007 (0.007)	0.007 (0.007)
Orig.=VII	-2.714*** (0.136)	-0.251*** (0.028)	-0.184*** (0.028)	-0.155*** (0.027)	0.007 (0.007)	0.006 (0.007)
Orig.=SEI	-1.937*** (0.154)	-0.154** (0.060)	-0.107* (0.059)	-0.080 (0.060)	-0.005 (0.015)	-0.005 (0.016)
Orig.=SEII	-3.188*** (0.152)	-0.227*** (0.041)	-0.146*** (0.040)	-0.121*** (0.035)	-0.029*** (0.010)	-0.031*** (0.010)
Dest.=II		-0.233*** (0.035)	-0.185*** (0.035)	-0.188*** (0.033)	-0.006 (0.006)	-0.006 (0.006)
Dest.=III		-0.334*** (0.044)	-0.238*** (0.041)	-0.239*** (0.036)	0.012* (0.007)	0.013* (0.007)
Dest.=IV		-0.547*** (0.047)	-0.400*** (0.044)	-0.419*** (0.044)	0.013* (0.007)	0.014* (0.007)
Dest.=V		-0.616*** (0.042)	-0.454*** (0.039)	-0.441*** (0.041)	0.033*** (0.009)	0.033*** (0.009)
Dest.=VI		-0.676*** (0.047)	-0.482*** (0.042)	-0.450*** (0.041)	0.066*** (0.009)	0.063*** (0.009)
Dest.=VII		-0.748*** (0.046)	-0.528*** (0.043)	-0.492*** (0.041)	0.066*** (0.008)	0.063*** (0.008)
Dest.=SEI		-0.262*** (0.047)	-0.147*** (0.046)	-0.187*** (0.045)	0.012 (0.012)	0.013 (0.012)
Dest.=SEII		-0.725*** (0.088)	-0.505*** (0.085)	-0.451*** (0.084)	-0.048*** (0.013)	-0.061*** (0.014)
Log likelihood	-54602.453	-30373.930	-30195.800	-28955.041	-2667.915	-2614.445
AIC	109288.906	60847.859	60495.600	58036.082	5439.830	5354.890
N	22,308	22,308	22,308	22,308	22,308	22,308

Note: * p<0.10, ** p<0.05, *** p<0.01. Region- and wave-clustered standard errors in parentheses. All models include controls for: region and wave dummies, age(-squared), ethnic minority status, whether born in country of residence. Controls for education includes linear and squared term for years of education. Further demographics include religiosity, marital status, place of residence, household size(-squared).

E Formal definition of reference mobility and position measures

Our measure of reference position is defined as:

$$REFP_{ioct} = \frac{\sum_{j \in POS_{oct}} w_j d_j}{\sum_{j \in POS_{oct}} w_j} \quad (A3)$$

POS_{oct} refers to those respondents $j \neq i$ that are observed in the same wave, country, class origin, and are up to five years younger or older than respondent i. d_j indicates class destination, and w_j indicates the combined post-stratification and population weight of j. As higher levels of d_j denote more advantaged classes, the higher $REFP_{ioct}$, the higher the average class destination in the reference group.

Analogously, our measure of reference mobility is defined as the average class mobility of those in the reference group:

$$REFM_{idct} = \frac{\sum_{j \in MOB_{dct}} w_j m_j}{\sum_{j \in MOB_{dct}} w_j} \quad (A4)$$

MOB_{dct} refers to those respondents $j \neq i$ that are observed in the same wave, country, class destination, and are up to five years younger or older than respondent i.

We further define reference upward mobility as $REFM^{+}_{idct} = REFM_{idct} I(REFM_{idct} > 0)$ and reference downward mobility as $REFM^{-}_{idct} = -REFM_{idct}\, I(REFM_{idct} < 0)$, where $I(.)$ is the indicator function.

How threatening are transformations of happiness scales to subjective wellbeing research?†

Caspar Kaiser and Maarten C.M. Vendrik*

Abstract

Two recent papers argue that many results based on ordinal reports of happiness can be reversed with suitable monotonic increasing transformations of the associated happiness scale (Bond and Lang 2019; Schröder and Yitzhaki 2017). If true, empirical research utilizing such reports is in trouble. Against this background, we make four main contributions. First, we show that reversals are fundamentally made possible by explanatory variables having heterogenous effects across the distribution of happiness. We derive a simple test of whether reversals are possible by relabelling the scores of reported happiness and deduce bounds for ratios of coefficients under any labelling scheme. Second, we argue that in cases where reversals by relabelling happiness scores are impossible, reversals using an alternative method of Bond and Lang, which is based on ordered probit regressions, are highly speculative. Third, we make apparent that in order to achieve reversals, the analyst must assume that respondents use the response scale in a strongly non-linear fashion. However, drawing from the economic and psychological literature, we present arguments and evidence which suggest that respondents likely use response scales in an approximately linear manner. Fourth, using German SOEP data, we provide additional empirical evidence on whether reversals of effects of standard demographic variables are both possible and plausible. It turns out that reversals by either relabelling or by using Bond & Lang's approach are impossible or implausible for almost all variables of interest. Although our analysis uses happiness as a special case, our theoretical considerations are applicable to any type of subjective ordinal report.

Keywords: *ordinal reports, transformations of cardinal scales, happiness, subjective wellbeing, life satisfaction, Easterlin Paradox, General Social Survey, German Socio-Economic Panel*

---

† This paper is based on an alternative version of the manuscript available on the Open Science Framework (see https://osf.io/gzt7a/).

*Department of Macro, International and Labour Economics and ROA, SBE, Maastricht University; IZA, Bonn; EHERO, Erasmus University, Rotterdam. Email: m.vendrik@maastrichtuniversity.nl

1 Introduction

This paper offers an analysis of the conditions under which reports about happiness can be used to identify the effects of socio-economic variables on actual "true" happiness. Some of this analysis is a reaction to a recent paper by Bond and Lang (2019). In that paper, Bond and Lang forcefully argue that the results of most happiness research can be reversed, i.e. that the estimated sign of a variable's effect on expected happiness can be inverted by means of some suitable transformation of the cardinal happiness scale under consideration. In a similar vein, Schröder and Yitzhaki (2017) show that signs of coefficients from OLS regressions of reported happiness are potentially reversible. If Bond and Lang (2019) and Schröder and Yitzhaki (2017) are right, happiness research is in trouble. It is therefore imperative to give a comprehensive analysis of the reasons why and the conditions under which such reversals are possible.

In giving such an analysis we make four key points. First, we show that reversals are fundamentally made possible by explanatory variables having heterogenous effects across the distribution of happiness. In the context of OLS regressions, we derive a simple test of whether reversals are possible by relabelling the scores of reported happiness. Second, we argue that in cases where reversals by relabelling happiness scores are impossible, reversals based on Bond and Lang's method using ordered probit regressions are empirically unfounded and thus highly speculative. Third, we make apparent that in order to achieve reversals, the analyst must assume that respondents use the response scale in a strongly non-linear fashion. However, drawing from the economic and psychological literature, we provide arguments and discuss experimental evidence which suggest that respondents likely use response scales in an approximately linear manner. Deviations from linearity seem to be larger for response scales with as few as three response options. Fourth, using German SOEP data, we provide additional empirical evidence on whether reversals of effects of standard demographic variables are both possible and plausible. It turns out that reversals by relabelling are impossible or implausible for almost all

variables of interest. Likewise, when using Bond and Lang's approach, reversals always rely on assuming implausible ways in which respondents use the response scale.

Before making these and some additional points, it is helpful to define our terms and to frame the discussion. Happiness is typically measured via responses to questions like *"How happy are you these days?"* or *"Taking all things together, how satisfied are you with your life?"*.[1] Such responses are then recorded in their rank order, i.e. giving a "1" to the first response option, a "2" to the second response option, and so on. Call these ordered responses hr (**h**appiness **r**eports) with options $\{1,2,\ldots,R\}$. These happiness reports are assumed to be indicative of an unobservable cardinal quantity of true happiness, which is a subjective feeling whose intensity is only internally accessible to the respondent. Call this quantity ht (**h**appiness **t**rue). We take it that most research on happiness is concerned with estimating the direction and magnitude of effects of socio-economic variables on expected quantities of true happiness ht as approximated by ordered categories of reported happiness hr.

In this paper, we exclusively focus on issues that arise even if one assumes that scale use is homogenous across all respondents and times. We thus ignore the additional problems arising from inter- and intra-personal differences in scale use (see e.g. King et al. 2004). These issues are explored in paper 4 of this thesis.

The rest of the paper is structured as follows. In section 2 we give a restatement of Bond and Lang's core arguments, which we hope will clarify some of the issues at stake. In section 3, we show that reversals of the sort Bond and Lang (2019) and Schröder and Yitzhaki (2017) propose are driven by heterogeneities in the effects of explanatory variables across the distribution of happiness. We then derive a test of whether reversals of signs of OLS regression coefficients are possible by means of relabelling the underlying happiness scale. We also derive bounds on trade-

[1] Responses to the latter question are normally taken to measure life satisfaction rather than happiness. Since this distinction makes no difference to the arguments of the present paper, we will primarily use the term "happiness" in a wider sense than merely affective wellbeing.

off ratios of coefficients under any labelling scheme. In Section 4 we first note that reversals typically require that respondents use the response scale in a strongly non-linear fashion. We then present several arguments and experimental evidence showing that respondents likely use the response scale in a roughly linear fashion. Section 5 provides empirical evidence on the possibility and plausibility of reversals using German SOEP data. A final section concludes.

2 Bond and Lang's arguments

Bond and Lang (2019) make two inter-related arguments; one about non-parametric identification of rankings of group means of ht and another about using standard parametric approaches to identify such rankings. Since their arguments are opaque, sections 2.1 and 2.2 attempt to clarify them.

2.1 The non-parametric argument

As a numerical example, assume that we obtained data on hr with four response categories about two groups A and B. We want to know which group has a higher mean on true happiness ht. We obtain response frequencies across both groups as presented in table 1, columns (3) and (6).

Table 1. An example

		A			B		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
hr	τ_k	freq.	$E(ht^{(I)}\|hr=k)$	$E(ht^{(II)}\|hr=k)$	freq.	$E(ht^{(I)}\|hr=k)$	$E(ht^{(II)}\|hr=k)$
1	1.5	10	1	1	5	1	1
2	2.5	30	2	2	20	2	2
3	3.5	50	3	3	70	3	3
4	∞	10	4	7.5	5	4	7.5
Mean:		$\overline{hr}=2.60$	$\overline{ht^{(I)}}=2.60$	$\overline{ht^{(II)}}=2.95$	$\overline{hr}=2.75$	$\overline{ht^{(I)}}=2.75$	$\overline{ht^{(II)}}=2.93$

If we take the rank order of hr at face value, which is standard practice in happiness research, it seems as though group B is indeed happier on average than group A (as indicated by the means $\overline{hr}$ given in the bottom row).

However, a more prudent interpretation of the data is that we merely observe the outcome of a process where respondents apply a common reporting rule about their internal states ht. More

formally, a person i makes a response $hr_i = k$ iff $\tau_{k-1} \leq ht_i < \tau_k$, where τ_k's are thresholds common to all persons.

In this setup, the rankings of means of ht across the groups may turn out either way. Consider a world (with superscript I) in which means of ht conditional on hr are exactly those we obtain when taking hr at face value, i.e. $\mathrm{E}(ht^{(I)}|hr = k) = k$ (see columns 4 & 7). In that case the mean of hr and the mean of ht coincide. Therefore, in world I, group B is indeed happier than group A. Now consider an alternative world (with superscript II; columns 5 & 8) in which the expected value of ht in the highest response category is much higher. Consequently, because the share in the highest response category is greater for group A than B, in world II the mean of ht is higher in A than in B.[2] Since we do not observe mean ht conditional on each response category, we do not know whether we are in world I or world II. In that sense, both worlds are equally well supported by our data. We therefore do not know which of the two groups is in fact happier.

Unfortunately, even if the share of those in the highest response category was greater in group B than in group A, it could nevertheless be the case that mean ht in group A conditional on being in the highest response category is sufficiently larger than in group B (i.e. $\mathrm{E}(ht|hr = 4; A) > \mathrm{E}(ht|hr = 4; B)$) such that the overall mean is larger in group A.

The only case in which a possibility of this kind cannot obtain is when no member of group A is observed in the top response category. However, in such case, it may be that expected ht in the lowest response category is much lower for group B than for group A (i.e. $\mathrm{E}(ht|hr = 1; B) < \mathrm{E}(ht|hr = 1; A)$), thus also leading to a reversal. To prevent this possibility, we require that no members of B are observed in the lowest response category.

[2] In world I we have for A: $(1 * 10 + 2 * 30 + 3 * 50 + 4 * 10)/100 = 2.6$ and for B: $(1 * 5 + 2 * 20 + 3 * 70 + 4 * 5)/100 = 2.75$. In world II we have for A $(1 * 10 + 2 * 30 + 3 * 50 + 7.5 * 10)/100 = 2.95$ and for B $(1 * 5 + 2 * 20 + 3 * 70 + 7.5 * 5)/100 = 2.93$

Given these intuitions, and when only assuming that reporting thresholds are common across all respondents, we can then formulate Bond and Lang's condition for non-parametric identification of group means of ht. Let s_k^G denote the share in response category $k \in \{1,2,\dots,R\}$ of hr in group G. Their condition can then be stated as:

> **Bond and Lang's non-parametric condition.** In order to conclude that group B has a higher mean of ht than group A, we require (1) $s_1^B = 0$, (2) $s_R^A = 0$, and (3) $\sum_{j=1}^{k} s_j^B < \sum_{j=1}^{k-1} s_j^A \ \forall k \in \{2,\dots,R-1\}$.

Parts (1) and (2) respectively state that Group B must have no members in the lowest response category and that Group A must have no members in the highest response category. Part (3) states that for all response categories higher than 1 and lower than R, the cumulative share of members in group B up until category k is smaller than the cumulative share of members in group A up until category $k-1$.

Since this condition, especially in large samples, is rarely satisfied, Bond and Lang (2019) conclude that non-parametric identification of rankings of means of ht across two groups is almost always impossible.

2.2 The parametric argument

However, rather than aiming at non-parametric identification, happiness researchers often attempt to infer rankings of group means or effects of explanatory variables on expected ht by estimating an ordered probit regression. Bond and Lang (2019) give a related argument on how such a strategy may fail.

The underlying model of an ordered probit regression of happiness reports hr on some independent variable X and a vector of control variables $\boldsymbol{Z}$ may be written as follows:

$$hl_i = \alpha_m + \beta_m X_i + \gamma_m \boldsymbol{Z}_i + \varepsilon_i \qquad (1)$$

As in the previous section, hr takes on values $\{1, \ldots, k, \ldots, R\}$ according to the scheme $hr_i = k$ iff $\tau_{k-1} \leq hl_i < \tau_k$. For completeness we define $\tau_0 = -\infty$ and $\tau_R = \infty$. The variable hl (**h**appiness **l**atent) is an unobserved latent index of happiness. We hope that hl is a sufficiently good proxy for ht, such that when we estimate β_m to be positive, we expect ht to increase when X_i increases.

To estimate α_m, β_m as well as the remaining set of thresholds τ_k via maximum likelihood, we assume that ε_i is distributed normally with mean 0 and standard deviation σ_i. If ε_i were instead assumed to be distributed logistically, we would estimate an ordered logit regression.[3] For identification we additionally need to fix two further thresholds τ_k (e.g. $\tau_1 = 0$ and $\tau_2 = 1$).[4] In most applications we do not let σ_i vary across persons. However, forcing σ_i to be constant is not necessary and we can therefore let σ_i depend on X_i. The model $\ln(\sigma_i) = \alpha_s + \beta_s X_i + \gamma_s \boldsymbol{Z}_i$ is particularly appealing since σ_i can then only take on values larger than 0.[5]

Similar to the previous section, first consider the case in which we merely want to ascertain the ranking of some groups A and B with respect to their mean level of ht. Assume that $X_i = 0$ implies membership in A, and $X_i = 1$ implies membership in B. Following estimation, we obtain for both groups, a prediction for the distribution of hl. These predicted distributions are normal distributions with means μ_A and μ_B, and standard deviations σ_A and σ_B. Here, μ_A and μ_B are the predicted value of hl_i when respectively setting $X_i = 0$ or $X_i = 1$ (and setting controls $\boldsymbol{Z}_i$ to some set of values; see below). When $\beta_m > 0$, we find that $\mu_B > \mu_A$. From this we might conclude that mean ht is greater in group B than in group A.

[3] All the arguments to follow could be adapted to the ordered logit case.
[4] We could alternatively fix just one threshold in conjunction with fixing α_s (as defined below).
[5] Little depends on this functional form. Versions of Bond and Lang's argument could also be constructed using a linear functional form.

However, there is little in our data that informs us about the relationship between hl and ht. Crucially, we do not know whether hl is a linear transformation of ht. Instead, any positive monotone transformation $\widetilde{hl}$ of hl will reproduce the distribution of hr equally well. Since both hl and $\widetilde{hl}$ produce the distribution of hr equally well, we seem not to have a criterion by which to decide whether hl or $\widetilde{hl}$ is a better proxy of true happiness ht.

Assume that we indeed find that $\beta_m > 0$, implying that $\mu_B > \mu_A$. Further assume that we estimate $\beta_s < 0$, implying $\sigma_B < \sigma_A$. In this case, β_m and β_s have opposite signs. We may then transform hl with the function $\widetilde{hl} = e^{chl}$, where c is a positive constant to be set by the researcher. Given that hl is distributed normally, taking e^{chl} yields for any group G, a right-skewed log-normal distribution for $\widetilde{hl}$ with mean $\tilde{\mu}_G = e^{c\mu_G + 0.5c^2\sigma_G^2}$.[6] In the contrary case, where β_m and β_s have the same sign, we can alternatively apply the function $\widetilde{hl} = -e^{chl}$ (with $c < 0$) which would produce a left-skewed log-normal distribution with mean $\tilde{\mu}_G = -e^{c\mu_G + 0.5c^2\sigma_G^2}$.[7]

In either case, the mean of $\widetilde{hl}$ depends on both μ_G and σ_G. Setting $\tilde{\mu}_A = \tilde{\mu}_B$ and solving for c, we obtain the value of c at which the implied ranking of means is reversed:

$$c = \frac{2(\mu_A - \mu_B)}{\sigma_B^2 - \sigma_A^2} = \frac{2\beta_m}{e^{2(\alpha_s + \beta_s + \gamma_s \boldsymbol{z_i})} - e^{2(\alpha_s + \gamma_s \boldsymbol{z_i})}} \tag{2}$$

Thus, so long as $\sigma_A \neq \sigma_B$ there exists some level of c at which the means of $\widetilde{hl}$ of groups A and B equal. Notice that $\sigma_A = \sigma_B$ when $\beta_s = 0$. However, in large samples, this obtains almost never.[8] Thus, we are almost always able to reverse rankings of group means of ht obtained from

[6] See e.g. https://en.wikipedia.org/wiki/Log-normal_distribution.

[7] A left-skewed transform is never required for any variable in our empirical GSOEP application in section 5. This is why we focus on the right-skewed case. However, in more than half of the empirical examples in the Online Empirical Appendix of Bond and Lang, a left-skewed transform is required. All the arguments to follow could be restated for the left-skewed case.

[8] In large samples, estimates of β_s will be approximately normally distributed, and the probability of obtaining a particular value is zero for any continuous probability distribution.

an ordered probit regression. Further note that the required value of c will be positive in the right-skewed case (for equal signs of β_m and β_s) and negative in the left-skewed case (for opposing signs of β_m and β_s).

Though Bond and Lang (2019) do not derive this, a version of this result carries over to the case in which X_i is continuous. To find the required value of c when X_i is continuous, we can differentiate the expression for the expected value of $E(\widetilde{hl}_i|X_i, \boldsymbol{Z_i}) = \tilde{\mu}_i = e^{c\mu_i + 0.5c^2\sigma_i^2}$ (in the right skewed case) with respect to X_i, and set the resulting expression to 0. Doing so yields:

$$c = -\frac{2\frac{\partial E(hl_i|X_i, \boldsymbol{Z_i})}{\partial X}}{\frac{\partial E(\sigma_i|X_i, \boldsymbol{Z_i})^2}{\partial X_i}} = -\frac{2\frac{\partial(\alpha_m + \beta_m X_i + \gamma_m \boldsymbol{Z_i})}{\partial X_i}}{\frac{\partial e^{2(\alpha_s + \beta_s X_i + \gamma_s \boldsymbol{Z_i})}}{\partial X_i}} = -\frac{\beta_m}{e^{2(\alpha_s + \beta_s X_i + \gamma_s \boldsymbol{Z_i})}\beta_s} \qquad (3)$$

In the right-skewed case, we again see that c is predicted to be positive. The same expression is obtained in the left-skewed case, though c is then predicted to be negative. Again, we see that c is defined whenever $\beta_s \neq 0$.

Notice that the denominator of equation (2) depends on $\boldsymbol{Z_i}$, and the denominator of equation (3) depends on X_i and $\boldsymbol{Z_i}$. Equations (2) and (3) therefore only yield required levels of c for particular values of X_i and $\boldsymbol{Z_i}$.[9] To circumvent this issue, and in order to find a determinate value c at which the sign of the effect of an explanatory variable X_i reverses in the sample, we may use the following numerical search procedure:

> **Step 1:** Estimate equation (1) and the auxiliary equation for σ_i by ordered probit and obtain predictions of μ_i and σ_i for every observed combination of values of X_i and $\boldsymbol{Z_i}$.

[9] Notice that when $\boldsymbol{\gamma_s}$ or β_s is positive (negative), the predicted c will be smallest (largest) for the largest observed level of $\boldsymbol{Z_i}$ or X_i, and vice versa for the smallest observed value of $\boldsymbol{Z_i}$ or X_i.

Step 2: When signs of $\hat{\beta}_m$ and $\hat{\beta}_s$ differ, calculate $\tilde{\mu}_i = e^{c\mu_i + 0.5c^2\sigma_i^2}$ for some $c > 0$.

When signs of $\hat{\beta}_m$ and $\hat{\beta}_s$ equal, calculate $\tilde{\mu}_i = -e^{c\mu_i + 0.5c^2\sigma_i^2}$ for some $c < 0$.

Step 3: Numerically search for that c for which an OLS regression of $\tilde{\mu}_i$ on X_i and $\boldsymbol{Z_i}$ yields a zero coefficient on X_i.

This numerical search will always yield some reversing c that falls in the range of possible levels of c as predicted by equations (2) and (3) and observed values of X_i and $\boldsymbol{Z_i}$ in the sample. In the empirical application of section 5, we will report levels of c as found by both this numerical search and as predicted from equations (2) or (3) at the mean of all control variables.

It may not be immediately obvious why exponential transformations can be used to reverse results. Take the example of table 1 and again assume that $X_i = 0$ implies membership in A, and $X_i = 1$ implies membership in B. Running an ordered probit regression of the sort described above, we obtain $\mu_A = 1.34; \sigma_A = 1.11$ and $\mu_B = 1.53; \sigma_B = 0.86$. In the top panel of figure 1, we sketch the estimated cumulative distribution functions (CDFs) $F_A(hl)$ and $F_B(hl)$ for both groups.

Beyond the point $p = 2.19$, we find that the CDF of group A starts to be lower than that of B, i.e. for larger values of hl, we should expect relatively more members in A than in B. This implies that the effect of being in group B on hl is positive at levels of hl below p, but negative at values above p. The effect of being in a particular group is thus estimated to be heterogenous across the distribution of hl. As will be made more explicit in section 3.1, such heterogeneities are the fundamental reason for the possibility of sign reversals.

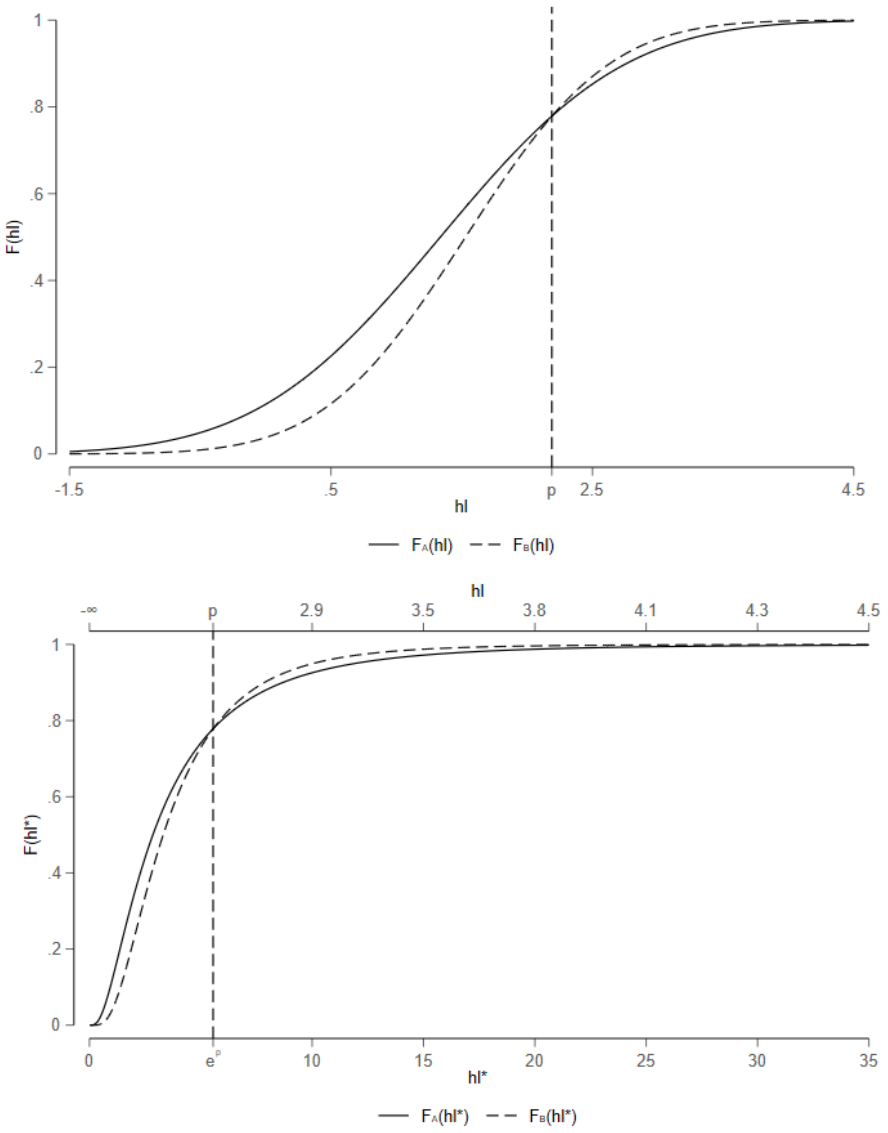


Figure 1. Cumulative distribution functions of hl (top panel) and $\widetilde{hl}$ (bottom panel) for groups A and B.

Indeed, given this heterogeneity, if values of hl larger than p are transformed to become sufficiently larger relative to values of hl smaller than p, it becomes possible to reverse the rankings of means. Functions with derivative larger for values larger than p than for values smaller than p may achieve this goal. Since $de^{chl}/dhl = e^{chl}c$, the derivative of Bond and Lang's function is increasing in hl, and is more extremely increasing in hl the larger c. Thus, for large enough c, mean reversals can be achieved. In our example, the level of c to get equal means is 0.78.

Finally, notice that when we apply a transformation function to hl, and state that such a transform is equally plausible, we also make a statement about the functional form by which ht

and independent variable(s) X may relate. Starting with our standard model $ht_i \approx hl_i = \alpha_m + \beta_m X_i + \gamma_m \boldsymbol{Z}_i + \varepsilon_i$, Bond and Lang's right-skewed transformation yields the new model: $ht_i \approx \widetilde{hl}_i = e^{chl_i} = e^{c(\alpha_m + \beta_m X_i + \gamma_m \boldsymbol{Z}_i + \varepsilon_i)} \leftrightarrow \frac{\ln(ht_i)}{c} \approx \frac{\ln(\widetilde{hl}_i)}{c} = hl_i = \alpha_m + \beta_m X_i + \gamma_m \boldsymbol{Z}_i + \varepsilon_i$. It is not clear why we should believe in this alternative log-linear functional form. Generally, researchers applying a transformation function, be it Bond and Lang's or any other function, should give reasons why the alternative functional form implied by applying a transformation function is credible.

3 Heterogeneities in effects across hr fundamentally drive reversals

Having hopefully clarified the issues raised by Bond and Lang (2019), this section will be primarily concerned with a variant of the same issue in the context of OLS regressions.

3.1 Relabeling hr

As stated in section 2.2, differences between σ_A and σ_B make reversals possible. The key reason why σ_i can depend on X is the possibility that the effect of X on individual happiness is not homogenous across the distribution of happiness.

Returning to our example of table 1, we see that there are fewer members of B than A in response categories 1, 2 and 4. However, there are more members of group B in category 3. Given that most responses for group B are in a single category, the ordered probit result of $\sigma_B < \sigma_A$, is intuitive.

We can interpret the concentration of responses in response category 3 as the effect of group B being positive at the bottom of the distribution of hr, but negative at the top of the distribution of hr. This point can be made more obvious by writing the change in mean hr in response to a change in X as a weighted sum of changes in the shares of each response category.

Let h_k denote the value with which we label the kth response category of hr. We can then write: $\frac{\partial E(hr)}{\partial X} = h_1 \frac{\partial s_1}{\partial X} + h_2 \frac{\partial s_2}{\partial X} + h_3 \frac{\partial s_3}{\partial X} + h_4 \frac{\partial s_4}{\partial X}$, where s_k denotes the share of respondents that respond with category k.[10] In our example, the first, second and fourth term are negative and the third term is positive. A 'rank-order' coding of hr implies that $h_k = k$, which is the standard choice in work on happiness and life satisfaction. If we believe that such a rank-order coding of hr is indeed a good proxy for $\mathrm{E}(ht|hr)$, i.e. that $\mathrm{E}(ht|hr = k) = k$, we would conclude that the effect of X is to increase ht overall, despite the fact that X has a negative effect for those in the highest response category.

However, there is nothing in the data that immediately precludes us from relabelling the highest response category to some larger values. Indeed, if we were assign the value $h_4 = 7.5$ to the fourth response category, and thought of this response scale as a better proxy for ht, we would have to conclude that X has a negative effect overall (cf. footnote 2).

In general, let the scale $(h_1, h_2, \ldots, h_R)$ be some labelling of the R response categories in hr such that $\mathrm{E}(ht|hr = h_k) = h_k$. Further assume that $\mathrm{E}(ht|hr = h_k; X) = \mathrm{E}(ht|hr = h_k)$. We can then write the partial effect of an independent variable X on $E(ht)$ as:

$$\frac{\partial E(ht)}{\partial X} = h_1 \frac{\partial s_1}{\partial X} + h_1 \frac{\partial s_2}{\partial X} + \cdots + h_R \frac{\partial s_R}{\partial X} \qquad (4)$$

Taking the highest response category R as the reference case, we can rewrite this to give:

$$\frac{\partial E(ht)}{\partial X} = (h_1 - h_2) \frac{\partial s_1}{\partial X} + (h_2 - h_3) \frac{\partial (s_1 + s_2)}{\partial X} + \cdots + (h_{R-1} - h_R) \frac{\partial \sum_{l=1}^{R-1} s_l}{\partial X} \qquad (5)$$

Since we should expect respondents who report a higher response category to be happier, we assume that $E(ht|hr = h_k)$ is monotonically increasing in k. Therefore, only labelling schemes

[10] In this case of a discrete dummy variable X, the partial derivatives in this formula should be interpreted as discrete difference quotients. The derivatives are partial because they hold at given values of the control variables $\boldsymbol{Z}_i$ (when such controls are included).

for hr that satisfy $h_1 < \cdots < h_R$ are permitted. Reversals are impossible when equation (5) has no roots that satisfies this constraint on the labelling scheme. Since this constraint implies that all the "coefficients" $(h_{k-1} - h_k)$ of $\partial \sum_{l=1}^{k} s_l / \partial X$ are negative, no such roots exist if and only if all the partial derivatives $\frac{\partial s_1}{\partial X}, \frac{\partial (s_1 + s_2)}{\partial X}, \ldots, \frac{\partial \sum_{l=1}^{R-1} s_l}{\partial X}$ have the same sign. This analysis can be summarised in the following proposition:

> **Proposition 1**. The effect of an explanatory variable X on mean happiness $E(ht)$ cannot change sign by relabelling the scores of reported happiness hr if and only if all effects of X on the cumulative response shares have the same sign.

Under standard OLS and ordered probit, the derivatives $\partial \sum_{l=1}^{k} s_l / \partial X$ are forced to all have the same sign. This is because β_m is assumed to be equal at each response category.[11] If this homogeneity assumption is violated, OLS and ordered probit are no longer valid in a strict standard sense and effectively yield average effect estimates across all response categories (also see Vendrik & Woltjer, pp.1428-1429 for a further discussion of this point, and see Angrist and Pischke (2009, Ch.3) for a more general "regression as a linear approximation" approach).

A recent paper by Schröder & Yitzhaki (2017) provides an equivalent condition for the special case of comparing two groups without any controls. We can then restate the above condition in terms of cumulative response shares across the two groups. For B to have a higher mean on any permitted scale of hr than A, there must be first order stochastic dominance in the cumulative response categories of group B over A, i.e. $\sum_{j=1}^{k} s_j^B < \sum_{j=1}^{k} s_j^A \; \forall k \in 1,2,\ldots,R-1$.

Whenever the conditional in Proposition 1 is violated (i.e. reversals are possible), Bond and Lang's condition as formulated in section 2.1 is also violated. However, the reverse is not true: When Bond and Lang's condition is violated, Proposition 1 may be satisfied. Moreover, Bond

[11] However, e.g. the generalized ordered probit/logit model of Williams (2006), which is not pursued here, allows parameters to differ across response categories.

and Lang's parametric condition of section 2.2, namely that $\beta_s = 0$, is almost always violated. Therefore, the conditional of Proposition 1 is weaker than Bond and Lang's conditions for non-parametric identification and parametric identification by ordered probit.

The conditional of Proposition 1 assumes that the mean of ht conditional on being in a particular response category is equal across all values of X. Formally, we assume $\mathrm{E}(ht|hr = h_k; X) = \mathrm{E}(ht|hr = h_k)$. Bond and Lang (2019) do not assume this in either of their arguments. Thus, cases in which the conditional of Proposition 1 is satisfied and Bond and Lang's condition is violated, are cases where the partial derivative $\partial E(ht|hr = h_k; X)/\partial X$, i.e. the marginal effect of X on the expectation of ht conditional on responding with category k, is non-zero for some k and has the same sign as $\partial \sum_{l=1}^{k} s_l/\partial X$.

However, nothing in the data suggests this to be the case. For example, consider a situation in which all empirically observed cumulative response shares up until category $R - 1$ decrease when an explanatory variable X rises. Substantively, this situation implies that with increases in X people tend to choose higher response categories across the entire distribution of reported happiness. The conditional of Proposition 1 is then satisfied and we expect X to be associated with an increase in $E(ht)$. For Bond and Lang's reversals to occur nevertheless, we require that in at least one response category, respondents become sufficiently less happy to counteract the overall shift towards response options with higher expected happiness. Given that all observable facts give indications to the contrary, such a large counteracting effect is speculative and implausible.[12]

[12] Moreover, note that when assuming stability of the response scale, such counteracting effects can only be arbitrarily large in the bottom and top response scale (assuming, as done in section 2, that ht is unbounded, which itself may be implausible).

3.2 A procedure to detect the possibility of reversals

To test whether reversals by relabelling are possible,[13] we need to get estimates of the partial derivatives $\frac{\partial s_1}{\partial X}, \ldots, \frac{\partial \sum_{l=1}^{R-1} s_l}{\partial X}$. Thereafter we can check whether all of them have the same sign. One procedure to get such estimates while allowing for (almost) continuous variables X (like individual income), is to run a number of separate OLS regressions with appropriately coded dummies as dependent variables.

Let $hd_{k,i}$(**h**appiness **d**ummy) be a dummy that is 1 when $hr_i \leq h_k$ and 0 otherwise. In other words, $hd_{k,i}$ equals 1 when a respondent i at most chooses response option k. We can then write:

$$hr_i = h_R + \sum_{k=1}^{R-1} (h_k - h_{k+1})\, hd_{k,i} \tag{6}$$

To see this, note that when $hr_i = h_k$, we find that $hd_{1,i} = \cdots = hd_{k-1,i} = 0$ and $hd_{k,i} = \cdots = hd_{R-1,i} = 1$. We then have $hr_i = (h_k - h_{k+1}) + (h_{k+1} - h_{k+2}) + \ldots + (h_{R-1} - h_R) + h_R = h_k$, since all other terms cancel. We may then run the following set of OLS regressions:

$$\begin{aligned} hd_{1,i} &= \alpha_{d1} + \beta_{d1} X_i + \boldsymbol{\gamma}'_{d1} \boldsymbol{Z}_i + \varepsilon_{d1,i}, \\ hd_{2,i} &= \alpha_{d2} + \beta_{d2} X_i + \boldsymbol{\gamma}'_{d2} \boldsymbol{Z}_i + \varepsilon_{d2,i}, \\ &\ldots \\ hd_{R-1,i} &= \alpha_{dR-1} + \beta_{dR-1} X_i + \boldsymbol{\gamma}'_{dR-1} \boldsymbol{Z}_i + \varepsilon_{dR-1,i}. \end{aligned} \tag{7}$$

Note that: $E(hd_{k,i} | X_i = X, \boldsymbol{Z}_i = \boldsymbol{Z}) = E(\sum_{j=1}^{N} \frac{hd_{k,j}}{N} | X_j = X, \boldsymbol{Z}_j = \boldsymbol{Z}\ \forall j) = E(\sum_{l=1}^{k} s_l | X_j = X, \boldsymbol{Z}_j = \boldsymbol{Z}\ \forall j)$. Therefore, our estimated predictions $\widehat{hd}_{k,i}$ are predictions of the cumulative response shares up to category k given X and $\boldsymbol{Z}$. Thus, $\frac{\partial \widehat{hd}_{k,i}}{\partial X_i} = \hat{\beta}_{dk} = \frac{\partial \widehat{\sum_{l=1}^{k} s_l}}{\partial X_i}$. Hence, if

[13] Schröder & Yitzhaki (2017) also propose a method to check for possible sign reversals in simple OLS regressions without controls, but this method is less general and much more complicated than ours.

$\hat{\beta}_{d1}, \hat{\beta}_{d2}, \ldots, \hat{\beta}_{dR-1}$ all have the same sign, sign reversals of $\frac{\partial E(ht)}{\partial X}$ in equation (5) are predicted to be impossible in virtue of Proposition 1.

Furthermore, equation (6) implies that predictions of hr_i as given by $\widehat{hr_i} = (h_1 - h_2)\widehat{hd}_{1,i} + (h_2 - h_3)\widehat{hd}_{2,i} + \cdots + (h_{R-1} - h_R)\widehat{hd}_{R-1,i} + h_R$ are equal to those obtained via a direct OLS regression of hr_i (using any permitted labelling scheme). In that case, estimate $\hat{\beta}$ from the regression $hr_i = \alpha + \beta X_i + \boldsymbol{\gamma}' \boldsymbol{Z}_i + \varepsilon_i$, can be expressed in terms of the coefficient estimates of $hd_{k,i}$ as

$$\hat{\beta} = (h_1 - h_2)\hat{\beta}_{d1} + (h_2 - h_3)\hat{\beta}_{d2} + \cdots + (h_{R-1} - h_R)\hat{\beta}_{dR-1}. \quad (8)$$

Therefore, the set of regressions of $hd_{k,i}$ can generate a sign reversal condition for the OLS estimate $\hat{\beta}$. We can then formulate:

> **Relabelling Condition.** Estimate $\hat{\beta}$ of the effect of an explanatory variable X_i on reported happiness hr_i does not change sign by relabelling the scores of hr_i if and only if all estimates $\hat{\beta}_{dk}$ of the effects of X_i on the happiness dummies $hd_{k,i}$ in equations (7) have the same sign.

As stated at the outset of this section, the relabelling condition implies that reversals are fundamentally made possible by heterogeneities in the signs of the effects of variables across the distribution of reported happiness. For example, when an explanatory variable makes unhappy people more happy, but also makes happy people less happy, that variable's effect on mean happiness in the entire population will depend how we weight the sizes of differences in happiness across the scale of reported happiness.[14] This point does not appear to be sufficiently recognized in the discussions of Schröder and Yitzhaki (2017) and Bond and Lang (2019). In one sense, such sign heterogeneity implies a deficiency of our happiness model, which could

[14] Such considerations of weighting across the happiness scale, which are entirely positive considerations, should be distinguished sharply from normative considerations regarding the relative weights of people with low or higher true happiness intensity ht.

potentially be repaired with the addition of further controls or higher-order and interaction terms to capture non-linearities in effects of explanatory variables.

Furthermore, Ferrer-i-Carbonell & Frijters (2004) made evident that time-invariant unobserved heterogeneity - driven by e.g. individual personality traits (Boyce 2010) - causes considerable bias in OLS estimates. As a consequence, the use of individual fixed effects has become standard in the empirical literature on happiness. Thus, obtaining reversal conditions for estimates that include such fixed effects is paramount. It is hence a major additional advantage of our approach that we can derive reversal condition for linear fixed effects models by simply demeaning every of the OLS regressions of $hd_{k,i}$.

3.3 A method for achieving reversals

Multiplicative scales of the sort proposed by Bond and Lang (2019), and discussed in section 2, can generate reversals in most instances where the relabelling condition is satisfied. In such a multiplicative scale, each subsequent difference $h_k - h_{k+1}$ in equation (6) equals a constant factor $r > 0$ times the preceding difference $h_{k-1} - h_k$, i.e. $h_k - h_{k+1} = r(h_{k-1} - h_k)$. For convenience, assume that $h_1 - h_2 = -1$. In that case, $h_2 - h_3 = -r$, $h_3 - h_4 = -r^2$, etc. We can then write equation (8) as:

$$\hat{\beta} = -\hat{\beta}_{d1} - r\hat{\beta}_{d2} - \cdots - r^{R-2}\hat{\beta}_{dR-1}. \quad (9)$$

According to Descartes' Rule of Signs, the number of positive real roots of this equation is either equal to the number of sign switches of the successive estimates of β_{dk}, or less than that by an even number. Hence, when there is only one sign switch, there exists one unique positive real root of equation (9). This appears to be the prevalent case empirically (see section 5). In general, when there is an odd number of sign switches, there exists at least one positive real root. However, when the number of sign switches of successive estimates of β_{dk} is even (e.g. two), there may not exist a positive real root. On the other hand, when there is an odd number of sign switches, there exists at least one positive real root. The number of possible sign switches can

never be larger than the order $R-2$ of the polynomial in equation (9). So, if the number of response options R is three, only one sign switch is possible, but for more than three response options there can be more than one sign switch. When a positive real root of equation (9) exists, this can be solved analytically for up to fourth-order polynomials, but in general only numerically for higher-than-fourth-order polynomials.[15]

Finally, the multiplicative scale discussed above is obtained when hr is rank-order coded initially, and the transformation $\widetilde{hr} = e^{chr}$ is applied subsequently. In that case, subsequent differences in values of hr will grow by a multiplicative ratio $r = e^{chr}$.[16] This is analogous to Bond and Lang's proposed transformation $\widetilde{hl} = e^{chl}$ of hl discussed in section 2.2.

3.4 Bounds on ratios of coefficients

When assessing the substantive implications of estimates, we may be more interested in ratios of coefficients than in their magnitude. For example, the estimation of shadow prices (Bertram and Rehdanz 2015; Levinson 2012; Luechinger 2009) or equivalence scales (Biewen and Juhasz 2017; Borah, Keldenich, and Knabe 2019; Rojas 2007) principally relies on ratios of coefficients. Unfortunately, when effects are not perfectly homogenous across the distribution of hr_i, ratios of coefficients are affected by transformations of hr_i even when no sign reversals are possible.

Fortunately, bounds on the ratio of two coefficients for any transformation of hr_i can be given. Let $\hat{\beta}$ and $\hat{\gamma}$ be coefficient estimates from a regression of a particular coding of hr_i on respectively X_i and Z_i (plus possibly other controls). Using equation (8), we can write the ratio of these coefficient estimates as the ratio of sums of corresponding coefficient estimates resulting from regressions of $hd_{k,i}$:

[15] See: https://math.stackexchange.com/questions/200617/how-to-solve-an-nth-degree-polynomial-Equation/.

[16] Since for any number n: $\frac{e^{Cn}-e^{C(n-1)}}{e^{C(n-1)}-e^{C(n-2)}} = e^{C} * \frac{e^{C(n-1)}-e^{C(n-2)}}{e^{C(n-1)}-e^{C(n-2)}} = e^{C}$

$$\frac{\hat{\beta}}{\hat{\gamma}} = \frac{\sum_{k=1}^{R-1}(h_{kj} - h_{kj+1})\hat{\beta}_{dk}}{\sum_{k=1}^{R-1}(h_{kj} - h_{kj+1})\hat{\gamma}_{dk}}. \quad (10)$$

From this we can deduce:

(1) When all ratios $\hat{\beta}_{dk}/\hat{\gamma}_{dk}$ of coefficient estimates from the regressions of $hd_{k,i}$ take the same value ρ, equation (10) reduces to $\hat{\beta}/\hat{\gamma} = \rho$. In that case, ratio $\hat{\beta}/\hat{\gamma}$ of coefficient estimates from the regression of hr_i does not depend on the particular coding scheme $(h_1, h_2, \ldots, h_R)$ of hr_i. The ratio is then invariant under any transformation of such a scheme.[17]

(2) In general, we can view ratio $\hat{\beta}/\hat{\gamma}$ as a weighted average of all ratios $\hat{\beta}_{dk}/\hat{\gamma}_{dk}$. By recoding hr_i such that $h_l - h_{l+1} = 0$ for all $l \neq k$ and $h_k - h_{k+1} < 0$, we can assign all the weight to one particular ratio $\hat{\beta}_{dk}/\hat{\gamma}_{dk}$. Doing so we obtain $\hat{\beta}/\hat{\gamma} = \hat{\beta}_{dk}/\hat{\gamma}_{dk}$. This scale is not allowed because of the zero intervals for $l \neq k$, but it acts as a bound for the allowed scales with non-zero intervals. It then follows that for all permitted transformations of the coding scheme of hr_i ratio $\hat{\beta}/\hat{\gamma}$ is larger (smaller) than the smallest (largest) ratio among all $\hat{\beta}_{dk}/\hat{\gamma}_{dk}$. We can hence give lower and upper bounds for ratio $\hat{\beta}/\hat{\gamma}$.

4 Plausibility of multiplicative happiness scales

4.1 General and theoretical arguments

When we transform a rank-order scale of reported happiness hr (or a similar scale of latent happiness hl in the ordered probit case) into a multiplicative scale $\widetilde{hr}$ (or similar scale $\widetilde{hl}$), and take such a scale as a better proxy for ht, we also change our substantive beliefs about the way in which persons reply to happiness questions. When hr is rank-order coded we believe that the difference in mean ht between each pair of adjacent response categories is constant. When we transform this scale, these differences are no longer constant. Instead, the difference in mean ht

[17] Further see Van Praag & Ferrer-i-Carbonell (2008), Ch. 2 on the invariance of trade-off ratios of coefficients when effects are homogenous across the distribution of hr_i.

between levels of $\widetilde{hr}$ grows or declines by a multiplicative factor r which is larger or smaller than one, respectively. For response scales with just three categories this may not be too problematic.

However, when the number of response categories grows large, it becomes apparent that Bond & Lang's transformation may have implausible consequences. The question on life satisfaction in the oft-used German SOEP survey has 11 response categories. With $c = 1$, which is smaller than sign-reversing c's in most of our results in section 5, and applying the right-skewed transformation, the difference on ht between $hr = 0$ (the lowest category) and $hr = 1$ is 1.72, while the difference between $hr = 9$ and $hr = 10$ is 13,923.38. This is more than 8,000 times larger.[18]

Thus, when applying such a transformation and treating our transformed variable $\widetilde{hr}$ as an equally plausible proxy of ht, we say that the following two ways of responding to a happiness question are equally plausible:

1. Respondents interpret each difference between subsequent steps[19] on their response scale as covering equal distances on their scale of experienced ht. This is illustrated in Figure 2.

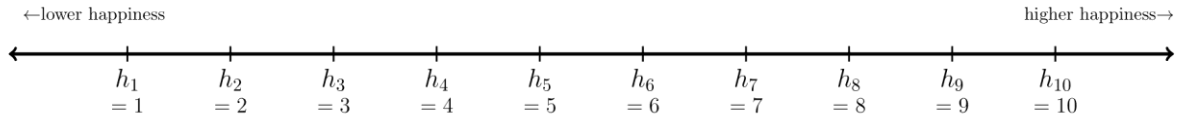


Figure 2. Linear response scale with equal intervals.

2. Respondent interpret each difference between subsequent steps on their response scale as covering a distance on their experienced scale of ht that is larger by a constant multiplicative factor $r = e^c$. This behavior is illustrated in figure 3 for $r = e^1 = 2.72$.

[18] $e^1 - e^0 = 1.72$ and $e^{10} - e^9 = 13{,}923.38$.

[19] Depending on whether one views the scale of ht as bounded or not, the top and bottom categories are exempted from that.

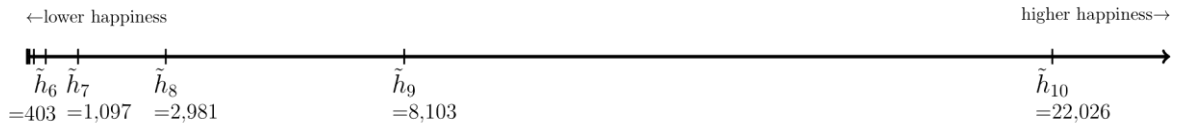


Figure 3. Multiplicative response scale with exponentially increasing intervals ($e^1 = 2.72$).

From introspection, it appears to us that the former interpretation is closer to our way of answering happiness questions. However, we may not be representative.

A theoretical argument in favour of a linear response scale as illustrated by figure 2 is provided by Van Praag (1971). Respondents may think of a finite set of response categories as discretizing an underlying continuous quantity (like ht). When reasoning about how to use such a set of categories to discretize that quantity, respondents may attempt to maximize the information that they give in the questionnaire. They can do so by minimizing the expected "inaccuracy" of their answer as modelled with a cost function like the square of the prediction error (i.e. the expectation of the square of the difference between what a respondent feels and what a researcher infers from that answer).[20] Van Praag shows that if respondents believe the underlying quantity to be uniformly distributed,[21] discretizing this quantity into equally spaced intervals minimizes this cost function. Kapteyn (1977) generalizes this result to hold for any plausible cost function.

4.2 Empirical evidence

This section presents some empirical evidence in favor of linearity of scale use.

4.2.1 Oswald (2008)

A straightforward empirical test of whether subjects treat bounded response scales as linear is given in Oswald (2008) who asked respondents (N=219) to rate their own height. To do so,

[20] See Bless, Strack, and Schwarz (1993) for arguments in favour of the assumption that respondents wish to be maximally informative with their answers.
[21] Relaxing this implausible assumption and generalizing this argument to other distributions is a subject of ongoing research. A draft of this work is available upon request. Related arguments are also put forward by Van Praag (1991), Parducci (1995) and Plant (2020).

respondents were asked to indicate their own height (relative to their gender) on a horizontal line with a label "very short" to the far left, and a label "very tall" to the far right of the line. As a visual aid, ten small vertical dashes were marked along the horizontal line. Respondents were not asked to place their height precisely on these dashes. On this data Oswald ran a regression which yielded the following result (t-statistics in parentheses) with an r-squared of 0.57:

$$subj.height = -54.28(2.8) + 0.67(2.53)\ true\ height - 0.0015(1.97)\ true\ height^2$$

Here, subjective height was scaled to range between 0 and 10. This result suggests a small amount of concavity in the response scale (as indicated by the significant squared term), but not enough to make multiplicative response scales plausible. When dropping the squared term for true height, r-squared fell only minimally to 0.56. Moreover, when running separate regressions for men and women, no evidence of concavity was found. This result suggests linearity in scale use when reporting height. However, linear scale use for height need not imply linear scale use for happiness. In particular, some experiments in psychophysics show that some objective stimuli, like light or sound, are subjectively perceived on a logarithmic scale (Shepard 1981).

4.2.2 Van Praag (1991)

Van Praag (1991) tested how persons translate ordered verbal labels (*very bad; bad; not bad, not good; good, very good*) into cardinal quantities in a context-free setting. In a first experiment he asked subjects to assign number between 1 and 1000 to each verbal label. Here, 1 was said to stand for the "very worst" and 1000 for the "very best". In a second experiment, respondents were asked to produce lines of certain length corresponding to each verbal label. Here a length of 1 unit was said to stand for the "worst", and a length of 40 to stand for the "best". No further context was given in either experiment. Scaling both sets of responses to lie on the interval [0,1], van Praag's results are shown in table 2.

Both experiments suggest roughly linear scale use. Although there is some variation in intervals between verbal labels, there is no obvious pattern to it. At most, it appears that labels "very bad"

and "bad" are less strongly distinguished than all the other labels. Van Praag explains this by noting that respondents may be trying to "leave room" for values associated with the other more positive labels.

Table 2. Van Praag's results

	Numbers experiment		Lines experiment	
Verbal label	Mean (90% CI)	Interval	Mean (90% CI)	Interval
(1) *Very bad*	0.089 (0.081-0.097)		0.073 (0.068-0.078)	
(2) *Bad*	0.201 (0.191-0.212)	0.112	0.180 (0.172-0.188)	0.107
(3) *Not bad, not good*	0.472 (0.462-0.482)	0.271	0.401 (0.392-0.410)	0.221
(4) *Good*	0.668 (0.658-0.678)	0.196	0.598 (0.588-0.608)	0.197
(5) *Very good*	0.866 (0.857-0.874)	0.198	0.823 (0.813-0.833)	0.225

Note: Confidence intervals are computed using a student's t-distribution with sample standard deviations given on p. 78 of van Praag (1991) and N = 361.

4.2.3 Studer (2012)

Studer (2012) analyses Dutch data in which respondents were asked to report their happiness with a slider on a continuous and bounded scale.[22] Using such a scale, respondents are enabled to report their ht more directly, making a cardinal interpretation of such responses more plausible. Respondents were also asked to report their happiness on a more standard Likert scale with ten response options (ranging from 0 to 9).[23] Studer can therefore directly compare response behaviour from a question that might reasonably be assumed to measure ht cardinally (up to a linear transformation), and a question that measures hr in a manner that is common in typical happiness research. In doing so, he evaluates which partitioning of the continuous scale would reproduce the observed response shares on the discrete Likert scale. More formally, he finds the set of cutoffs $\tau_k,\ k = 0, \dots, 10,$ that satisfy $\tau_0 = 0$ and $F_{cont.}(\tau_k) = \sum_{l=0}^{k-1} s_l$, where $F_{cont.}$ is the empirical CDF of responses for the question using the continuous slider and s_l denotes the share of respondents that report response category l on the question using the Likert scale. Looking at

[22] The boundedness of the scale may be justified by assuming that happiness, is bounded between states of being completely happy (or satisfied) and being completely unhappy (or dissatisfied).

[23] In both cases, the original question reads "Alles bij elkaar genomen, hoe gelukkig zou u zeggen dat u bent?" (Taking all things together, how happy would you say you are?), with extremes labelled "helemaal ongelukkig" (completely unhappy) and "helemaal gelukkig" (completely happy).

his results, we find that differences between cutoffs do not form a regular pattern: they are neither equidistant nor do they follow a multiplicatively increasing or decreasing scale. However, ratios of almost all subsequent differences of cutoffs do not exceed 1.8 or fall below 0.4.[24] The only exceptions are the ratios $(\tau_2 - \tau_1)/(\tau_1 - \tau_0)$ and $(\tau_3 - \tau_2)/(\tau_2 - \tau_1)$. These ratios are more extreme since very few people responded with the lowest two categories on the Likert scale. Studer's results therefore suggest that scales with multiplicative factors $r < 0.4$ and $r > 1.8$, corresponding to $c < -0.92$ and $c > 0.59$ are not plausible for Likert scales with relatively many response categories.

A fourth argument in favor of linearity of scale use can be found using some results in the literature on psychophysics. Since this argument is somewhat complicated, we relegate it to Appendix A. Including this additional argument, we have four different types of evidence that all point in a similar direction: People are most likely to use roughly linear reporting scales.

4.3 Scales with few reporting categories may be used in a multiplicative fashion

However, it is possible that the preceding arguments only hold in cases with at least five response categories (this is the number of categories used in van Praag's experiments). The happiness variables that are mostly analysed by B&L have just three or four response categories. We may think that respondents use these scales as collapsed versions of scales with more categories. If so, multiplicative scale use may be plausible in such cases. We therefore perform a similar exercise to that of Studer (2012). Table 3 lists cumulative shares in each response category from the happiness question of the 2006 wave of the US General Social Survey (GSS; with three categories), as well as cumulative shares for a life satisfaction question from the 5th (2006) wave of the United States sample in the World Values survey (WVS). Both samples are representative of the same population and the two questions measure strongly correlated concepts of ht.

[24] That is to say: $0.4 \leq (\tau_k - \tau_{k-1})/(\tau_{k-1} - \tau_{k-2}) < 1.8 \; \forall \; k \in 4\ldots,9$. Unfortunately, Studer (2012) only provides a graphical figure and not precise cutoff values. These values are therefore estimates.

Table 3. Cumulative response shares for happiness and life satisfaction in GSS and WVS

GSS		WVS		Mean hr
hr	Share in % (cum.)	hr	Share in % (cum.)	after collapse
1 ("Not too happy")	11.98 (11.98)	1 ("Completely dissatisfied")	0.46 (0.46)	4.14
		2	0.90 (1.36)	
		3	2.05 (3.41)	
		4	3.89 (7.30)	
		5	7.32 (14.61)	
2 ("Pretty happy")	55.80 (67.78)	6	9.71 (24.32)	7.30
		7	23.06 (47.38)	
		8	28.27 (75.65)	
3 ("Very happy")	32.22 (1.00)	9	17.65 (93.29)	9.28
		10 ("Completely dissatisfied")	6.71 (100.00)	

Note: Data from GSS and WVS wave 5 (both 2006). Design weights applied.

The observed cumulative response shares in these samples suggest that category "not too happy" in the GSS questions most closely corresponds to categories 1-5 on a 10-points scale. Likewise, category "pretty happy" seems most likely to correspond to categories 6-8 and category "very happy" corresponds to categories 9-10 on a 10-points scale. Assume now that the relative distribution of responses across the 10-points scale in the WVS sample (measuring life satisfaction) is a reasonable approximation of the distribution of responses we would observe had the GSS sample (measuring happiness) been given a 10-points scale. Assuming that the 10-points scale measures ht roughly cardinally (as argued in the previous section), we can then take mean hr across categories 1-5 of the WVS variable as indicative of mean ht in the "not too happy" response category of the GSS variable. This yields a mean of 4.14. Same arguments apply to mean hr of categories 6-8 (mean = 7.30) and 9-10 (mean = 9.28) of WVS as being indicative of mean ht in categories "pretty happy" and "very happy" of GSS. These (assumed) differences in ht between response categories of the GSS variable become smaller by a ratio of $r = (9.28 - 7.30)/(7.30 - 4.14) = 0.62$.[25]

Bond & Lang (2019), in their empirical appendices on the Easterlin Paradox and the increasing happiness gap between men and women (sections A3.1 & A3.7, respectively), also use GSS data

[25] The matching of GSS happiness shares and WVS life satisfaction shares in table 3 takes into account that people tend to be happier than they are satisfied with their life (see e.g. Knabe et al. 2010). This is confirmed by a comparison of shares of happiness and life satisfaction scores on 11-points scales in wave 3 (2006) of the ESS (see table A1 in the appendix).

with a three-points scale. For the Easterlin Paradox they find that a level of $c = -0.68$, which corresponds to a factor of $e^{-0.68} = 0.51$, is sufficient to reverse the negative trend in mean happiness in the USA. Moreover, a level of $c = -0.35$, corresponding to a factor of $e^{-0.35} = 0.70$, is found by them to be sufficient to make the growth in the happiness gap between men and women insignificant.

Though the former of these results is slightly more extreme than the level of c implied from collapsing the three-points scale, both of these results appear rather plausible in light of the preceding arguments. Thus, if mild transformations of three-points scales are sufficient to qualitatively change results, Bond and Lang's arguments have some force.[26]

5 Empirical Applications

5.1 Background and data

We now turn to assessing the wider empirical relevance of the theoretical points of the preceding sections to a range of important demographic variables using waves 1 (1984) to 32 (2015) of the German Socio-Economic Panel (GSOEP). The GSOEP is a nationally representative survey of the German population and is among the most commonly used dataset in empirical happiness research.

Our explanatory variables of interest are household income, unemployment, marriage, having children, and self-reported disability. These variables are similar to those investigated in the empirical appendix of Bond and Lang (2019).[27] Answers to the question *"How satisfied are you with*

[26] Using WVS (4-points scale for happiness) and ESS (11-points scale for happiness) data, we also applied a similar procedure to a set of 14 European countries. That exercise yielded that differences between responses on the 4-point WVS scale collapse in a roughly linear manner onto the 11-point ESS scale. See table A1 of Appendix C for results. The figures for mean hr after collapse in the fifth column of table A1 imply adjacent happiness differences from "not at all happy" to "very happy" of 2.71, 3.56, and 2.64. The subsequent ratios of these differences are given by 1.31 and 0.74.

[27] In their Appendix A.3.3, they consider the effects of the unemployment rate and inflation rate on happiness using Eurobarometer data. Unfortunately, they do not report which level of c would (just) reverse the sign of the effect of unemployment. In section A.3.6, they consider the effect of being married and of having children using BHPS data. They find that a right-skewed transformation with $c = 0.32$ reverses the effect of being married for men, while a left-skewed transformation with $c = -2.69$ reverses the effect of being married for women. Left-skewed

your life, all things considered",[28] with 11 response categories labelled from 0 to 10 are used as our dependent variable hr_i. Bond and Lang's analyses rely on questions with only three to seven response categories. Since 10 or 11 response categories are more typical for most happiness research, it will be useful to study whether plausible reversals can be obtained with the present variables. Moreover, the effects of income and unemployment have been extensively studied with these data. Consequently, there is now near universal agreement that, at least in the short run, higher incomes improve life satisfaction while unemployment reduces it. It would be a disturbing finding if these results were easily reversed.

For income we use log net post-tax household incomes, deflated to 2005 prices. We equivalize incomes using the modified OECD scale.[29] Regarding unemployment, we code a dummy that is 1 when a person reports to be unemployed, and 0 for any other possible employment status. We code similar dummies for being married, living with children in the household, and reporting a disability. Next to reporting results in which these variables are entered separately, we also report results in which all variables are entered jointly along with a set of additional control variables. The additional control variables include region ("Bundesländer") dummies, wave dummies, age, age squared, a tertiary education dummy, a home ownership dummy, log(household size) and log(1+working hours). We restrict our sample to those above the age of 18. In total, our sample includes 557,999 observations. We first present results for our method using relabelling, and then turn to Bond and Lang's method.

---

transformations with $c = -0.74$ and $c = -0.64$ reverse the effect of children in the household for men and women, respectively. Also using BHPS data, in section A.3.8 Bond and Lang show that a right-skewed transformation with $c = 1.41$ would reverse the effect of disability. None of their analyses except that of the effects of unemployment and inflation include control variables.

[28] In German: „Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?" (DIW 2016).

[29] We exclude respondents in the top and bottom percentiles of the income distribution as well as the self-employed, since there may be substantial measurement error in these observations (Berthoud and Bryan 2011; Hurst, Li, and Pugsley 2013).

5.2 Reversals using relabelling

Since Ferrer-i-Carbonell & Frijters (2004) made evident that time-invariant unobserved heterogeneity causes bias in pooled-regression estimates, applied researchers have turned to individual-fixed effects models. Unfortunately, no fixed effects estimator is thought to be readily available for the ordered probit model. This is a major drawback to Bond and Lang's method.[30] However, by demeaning each regression of $hd_{k,it}$ (index t runs over time periods) our relabelling condition can be straightforwardly applied to the fixed-effects model.[31]

Thus, table 4 shows results for pooled and fixed-effects regressions of hr_{it} on each of our explanatory variables of interest. Columns (1) and (3) show results from separate regressions in which each variable is entered individually (being married and having children are always entered jointly), while columns (2) and (4) show results from regressions in which all variables of interest, along with the additional controls discussed above, are entered jointly.[32] In all specifications, household income and being married are associated with higher life satisfaction, while unemployment and reporting a disability are associated with lower life satisfaction. Having children in the household is also associated with higher satisfaction but turns insignificant when including fixed effects as well as controls. More generally, accounting for fixed effects reduces the size of coefficient estimates for every explanatory variable.

In order to evaluate whether the sign of these coefficients can be reversed, we estimate regressions of $hd_{k,it}$ for $k = 0,1,\ldots,9$ when entering variables separately, when including controls, and when either running pooled regressions or when controlling for fixed effects. Tables A2 and A3 in Appendix B show our full results.

---

[30] A similar drawback is faced in Schröder & Yitzhaki's (2017) approach since their reversal condition does not allow for controls at all. See, however, Ferrer-i-Carbonell & Frijters (2004; section 4), for an ordered logit model with individual-fixed effects, and paper 4 of this thesis for discussion of an ordered probit estimator with individual fixed effects proposed by Van Praag (2015).

[31] In this case, with full controls, linear age drops out of the estimations since it is then perfectly collinear with the individual-fixed effects and the wave dummies.

[32] Since one may worry that marriage and having children mediate the effect of income, we also ran regressions in which these variables are excluded. This yielded very similar results for all other variables.

Table 4. OLS and fixed effects regressions

	Pooled		Fixed effects	
	(1)	(2)	(3)	(4)
	No controls	Full controls	No controls	Full controls
Log household income	0.691*** (0.011) reversing c=3.18	0.568*** (0.012) reversal impossible	0.228*** (0.011) reversing c=1.44	0.296*** (0.011) reversal impossible
Unemployed	-1.273*** (0.019) reversal impossible	-0.917*** (0.018) reversal impossible	-0.643*** (0.016) reversal impossible	-0.638*** (0.015) reversal impossible
Married	0.189*** (0.012) reversal impossible	0.290*** (0.013) reversal impossible	0.171*** (0.014) reversing c=2.17	0.168*** (0.014) reversal impossible
Children	0.175*** (0.012) reversal impossible	0.132*** (0.012) reversing c=−2.83	0.068*** (0.012) reversal impossible	0.008 (0.012) reversing c=0.13
Disability	-0.857*** (0.021) reversal impossible	-0.766*** (0.020) reversal impossible	-0.495*** (0.019) reversal impossible	-0.306*** (0.018) reversing c=2.53
Observations	557,999	557,999	557,999	557,999

Note: * p<0.10, ** p<0.05, *** p<0.01. Clustered (by respondent) standard errors in parentheses. Columns (1) and (3) result from separate models for each explanatory variable (being married and having children entered jointly). Reversing c values have been obtained numerically.

To illustrate, figure 4 plots estimated coefficients of household income for each regression of $hd_{k,it}$ (see Figures A1 to A4 in Appendix C for such plots for unemployed, married, having children and disability).

Recall that our relabelling condition states that coefficient sign reversals are impossible if and only if all coefficient estimates on $hd_{k,it}$ have the same sign for all k. For our pooled regressions, we find that the income coefficient can be reversed when no controls are included. Since the sign of the effect on $hd_{9,it}$ is positive, while all others are negative, a sufficiently convex transformation can achieve such a reversal. A numerical search reveals that a multiplicative scale in which spaces between adjacent response categories grow by a ratio of $r = 24.1$ is required to just achieve such a reversal. This corresponds to a level of $c = \ln(24.1) = 3.18$ in the transformation $\widetilde{hr}_{it} = e^{chr_{it}}$. Second, the effect of having children can be reversed when including controls. Here, since the sign of the effect on $hd_{0,it}$ is positive while the effects in all other regressions of $hd_{k,it}$ are negative, a sufficiently concave transformation can achieve such a reversal. A numerical search yields that a transformation $\widetilde{hr}_{it} = -e^{chr_{it}}$ with $c = -2.83$ or lower is sufficient. This corresponds to a multiplicative ratio of $r = 0.06$.

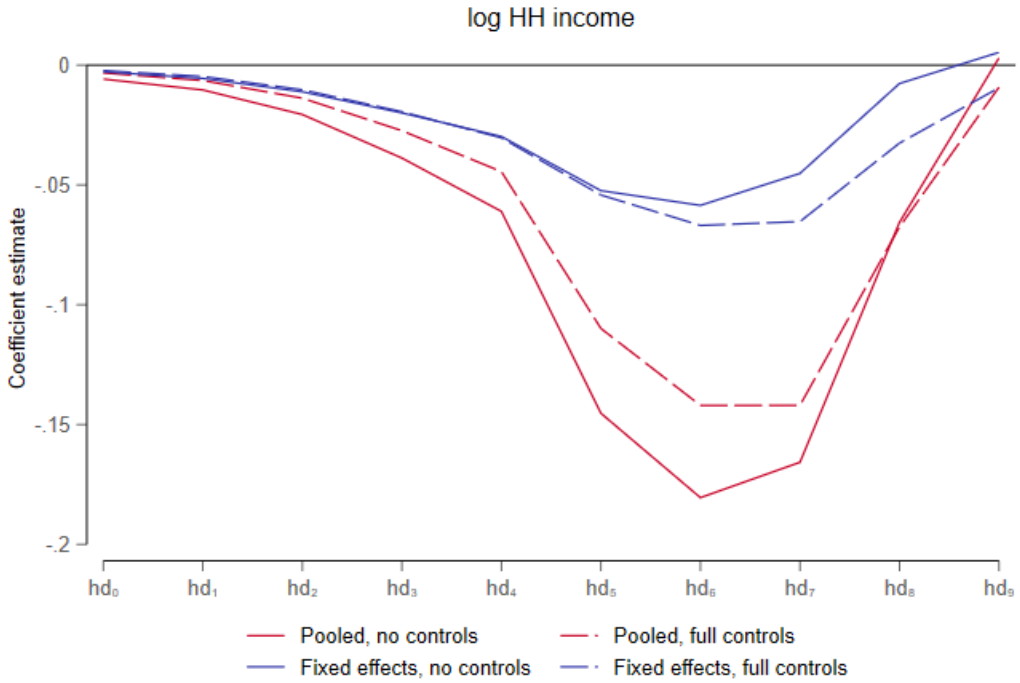


Figure 4. Coefficient estimates of income for each regression of $hd_{k,it}$, corresponding to Appendix Tables A3 and A4.

Given the results of table 4, these pooled estimates are likely biased due to the omission of fixed effects. Nevertheless, when accounting for fixed effects, reversals remain possible for both income (without controls, convex, with $c = 1.44$ or higher, corresponding to $r = 4.22$) and for children (with controls, now convex with $c = 0.13$ or higher, corresponding to $r = 1.14$).[33] In addition, the effects of being married (without controls) and disability (with controls) can now be reversed by respective convex transformations with at least $c = 2.17$ ($r = 8.76$) and $c = 2.53$ ($r = 12.56$).

The effect of unemployment cannot be reversed in any specification. We may thus conclude that the common finding that unemployment is associated with lower life satisfaction is particularly robust. Should any of the other reversals worry us? In light of the arguments given in in section 4, scales with c values in the order of at least 1 ($r \geq e^1 = 2.72$) or at most -1 ($r \leq e^{-1} = 0.37$) are implausible. Almost all just-sign-reversing scales found above fall within these ranges. The only exception is the reversal scale for $c = 0.13$, implying $r = e^{0.13} = 1.14$. This scale is

[33] Further robustness regressions not shown here indicate that the inclusion of wave dummies is sufficient to make reversals of the sign of the coefficient of log household income impossible.

reasonably close to a linear scale. We thus conclude that while reversals are *possible* for most variables in at least some specifications, the only *plausible* reversal is that of the effect of having children in a fixed-effects regression with controls. Since that result was strongly insignificant and close to zero in table 4, this is not a particularly striking result.[34]

5.3 Bounds on trade-off ratios

However, Figures 4 and A1-A4 indicate that the relative magnitudes of the effects of each explanatory variable on $hd_{k,it}$ are not the same for all k. Given the discussion in section 3.4, trade-off ratios of coefficients will consequently not be invariant under all transformations of hr_{it}. Section 3.4 also established that ratios of coefficients will be bounded from above and below by the largest and smallest corresponding ratios obtained from regressions of $hd_{k,it}$. Figure 5 plots the ratio of the coefficients of unemployment, being married, having children and disability against the coefficient for income in each of the fixed effects regressions of $hd_{k,it}$ (with full controls; corresponding to the bottom panel of table A3).

Since the magnitudes of the coefficients of being married and having children are small, their ratios with the coefficients of income vary only little across regressions of $hd_{k,it}$. Moreover, since the coefficients of having children switch sign for $k \geq 7$, the ratios of these coefficients with those of income then change sign, too. For unemployment and disability, we observe that the ratios of these coefficients to the income coefficient generally increase with higher k. Therefore, the ratios of the effects of unemployment and disability on hr_{it} to the effect of income on hr_{it} will also increase (decrease) for increasingly convex (concave) transformations of hr_{it}.[35]

[34] To achieve a significantly negative effect estimate with $p < 0.05$ of children we require $c = 0.65$, implying $r = e^{0.65} = 1.92$.

[35] Recall that convex (concave) transformations give relatively more weight to changes at higher (lower) levels of hr_{it}.

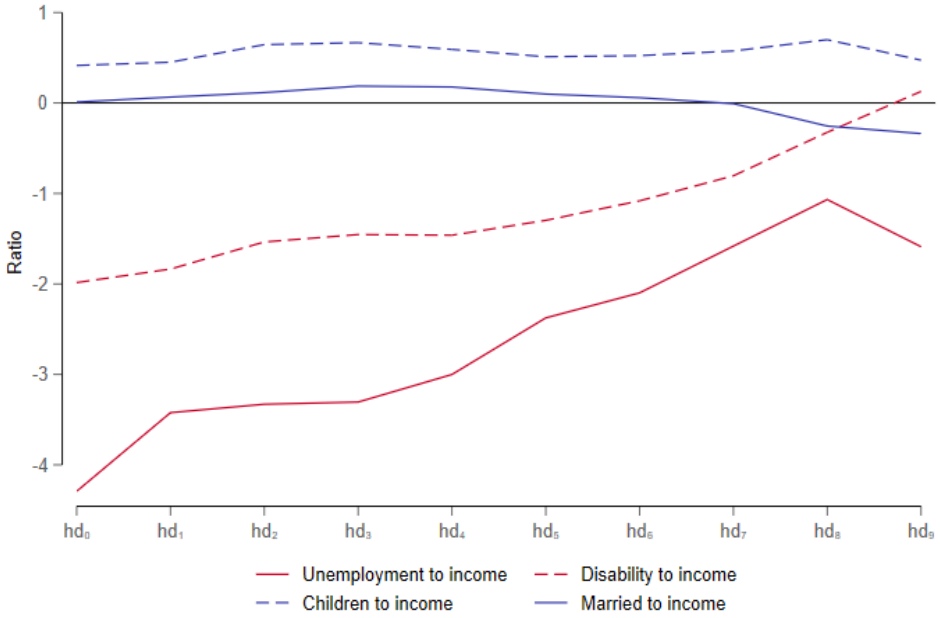


Figure 5. Ratio of coefficients of unemployment, being married, children and disability to income for each regression of $hd_{k,it}$. Regressions include the full set of controls and account for fixed effects (corresponding to the bottom panel of table A3).

Using such ratios of coefficients, we can calculate the shadow price of each explanatory variable. We define the shadow price of e.g. unemployment as the amount of additional income needed for an unemployed person with a particular income level y to be as satisfied as someone who is not unemployed. This amount is given by $\left(e^{-\beta_{ue}/\beta_{lny}} - 1\right)y$.[36] Here, β_{lny} is the coefficient of log household income, and β_{ue} is the coefficient of unemployment. Consequently, the shadow price of unemployment will fall in a range determined by the largest and smallest ratio of coefficients obtained from regressions of $hd_{k,it}$. Ranges of shadow prices for being married, having children, and disability can be found analogously. In table 5, we list ranges of shadow cost for each independent variable as evaluated at the mean of income.

[36] To see this, solve $\left[\beta_{lny}\ln(y+\Delta y)+\beta_{ue}\right]-\beta_{lny}\ln(y)=0$ for Δy.

Table 5. Shadow prices for each explanatory variable

Scale	Unemployment	Marriage	Children	Disability
Rank-order hr_{it}	€146,050 $(\frac{-0.638}{0.296})$	-€8,278 $(\frac{0.168}{0.296})$	-€500 $(\frac{0.008}{0.296})$	€34,677 $(\frac{-0.306}{0.296})$
Lower bound shadow price	€36,376 $(\frac{-0.035}{0.033})$	-€9,598 $(\frac{0.023}{0.033})$	€7,669 $(\frac{-0.003}{0.033})$	-€2,296 $(\frac{0.001}{0.009})$
Upper bound shadow price	€1,376,076 $(\frac{-0.010}{0.002})$	-€6,478 $(\frac{0.001}{0.002})$	-€3,249 $(\frac{0.004}{0.002})$	€119,769 $(\frac{-0.005}{0.002})$
$\widetilde{hr}_{it} = e^{chr_{it}}$, with $c = 0.4$	€92,782 $(\frac{-3.135}{1.772})$	-€8,370 $(\frac{1.024}{1.772})$	€1,108 $(\frac{-0.100}{1.772})$	€21,968 $(\frac{-1.358}{1.722})$
$\widetilde{hr}_{it} = e^{chr_{it}}$, with $c = -0.4$	€273,885 $(\frac{-0.037}{0.014})$	-€8,227 $(\frac{0.008}{0.014})$	-€1,590 $(\frac{0.001}{0.014})$	€53,100 $(\frac{-0.018}{0.014})$

Note: Calculations on the basis of table 4, column (4), lower panel of table A3 and fixed-effects regression of $\widetilde{hr}_{it}$ with $c = 0.4$ or $c = -0.4$ and full controls (not shown). Corresponding ratios of coefficients in parentheses. Negative shadow prices imply that a variable is estimated to benefit respondents. Thus, at the sample mean of household income and when using rank-order hr_{it}, a person who is *not* married needs to be compensated with €8,278€ of additional household income in order to be as satisfied as a person who is married.

As expected on the basis of figure 5, we find that the ranges of estimated shadow prices of unemployment and disability cover an extremely wide range. In contrast, shadow prices for marriage and children are comparatively small and vary little.[37] However, these ranges of possible shadow prices rely on rather implausible transformations of hr_{it} in which differences between response categories approach zero except for some particular chosen response category (cf. section 3.4). We therefore also evaluate how shadow prices change for a transformation $\widetilde{hr}_{it} = \pm e^{chr_{it}}$, with $c = 0.4$ and $c = -0.4$. These levels of c imply that differences in life satisfaction intensity between adjacent response categories increase or decrease by a factor of $e^{0.4} \approx 1.5$. We take it that such transformations may still be plausible.[38] This exercise shows that shadow prices for unemployment and disability still cover a rather wide range. Indeed, viewing these transformations as the most extreme plausible transformations, we do not know whether an unemployed (disabled) person can be compensated with as little as €93,000 (€22,000) or requires as much as €274,000 (€53,000).

[37] Note that since sign reversals were possible for children and disability, the signs of their shadow prices also depend on the chosen scale.

[38] The arguments of section 4 only establish that extreme departures from linearity are not plausible. They do not establish exactly when departures from linearity become implausible, which is why these choices for c are of course somewhat arbitrary.

We thus conclude that although sign reversals of the effects of explanatory variables on life satisfaction tend to be either impossible or very implausible, ratios of coefficients are heavily affected under even reasonably mild transformations. We therefore recommend that future empirical work should verify the robustness of its key results against mild convex and concave transformations of hr_{it}. We would also welcome for future work to check whether sign reversals of key coefficients are possible, and if so, to assess whether the required scale transformations seem plausible.

5.4 Reversals using Bond and Lang's method

We now turn to applying Bond and Lang's method to achieve reversals as discussed in section 2.2. To do so, in table 6 we report results from ordered probit regressions on the same data as in table 4. We again observe that higher incomes, being married, and having children are always associated with higher (mean) life satisfaction while unemployment and disability are associated with lower satisfaction. The magnitudes of these coefficients are roughly twice those obtained in the pooled OLS results of table 4. This is because differences between cutoffs are estimated to be somewhat above 2 for high levels of hr_{it} and somewhat below 2 for low levels of hr_{it}. Coefficients are therefore scaled by a factor of approximately 2 when compared to the rank-order coding used in table 4. Concerning the estimated standard deviation in latent satisfaction, σ_{it} we find the opposite result: higher incomes, being married and having children reduce σ_{it}, while unemployment and disability increase σ_{it}. Since no coefficient on $\ln(\sigma_{it})$ is estimated to be precisely zero, reversals are possible for every variable.[39] Because every explanatory variable has different coefficient signs for μ_{it} as compared to $\ln(\sigma_{it})$, convex transformations with positive c will yield reversals.

[39] However, the coefficient of children in the specification with full controls is insignificant and very close to zero. In turn, this small coefficient led to the extremely large estimate of the required value for c in Table 7.

Table 6. Heteroskedastic ordered probit regressions

	(1) HOP, variables entered separately	(2) HOP, full controls
$\boldsymbol{\mu_{it}}$		
Log HH income	1.453*** (0.043)	1.209*** (0.039)
Unemployed	-2.711*** (0.075)	-1.759*** (0.055)
Married	0.408*** (0.029)	0.577*** (0.031)
Children	0.409*** (0.030)	0.320*** (0.028)
Disability	-1.804*** (0.062)	-1.525*** (0.055)
Constant		10.336*** (0.224)
$\boldsymbol{\ln(\sigma_{it})}$		
Log HH income	-0.140*** (0.004)	-0.065*** (0.005)
Unemployed	0.066*** (0.006)	0.069*** (0.006)
Married	-0.038*** (0.004)	-0.049*** (0.005)
Children	-0.021*** (0.004)	-0.001 (0.005)
Disability	0.097*** (0.007)	0.073*** (0.007)
Constant		1.281*** (0.024)
Cutoff points		
τ_0		-∞ (assumed)
τ_1		0.000 (assumed)
τ_2		1.000 (assumed)
τ_3		2.377*** (0.036)
τ_4		3.775*** (0.068)
τ_5		4.875*** (0.094)
τ_6		7.032*** (0.145)
τ_7		8.366*** (0.177)
τ_8		10.499*** (0.228)
τ_9		13.781*** (0.306)
τ_{10}		16.257*** (0.365)
τ_{11}		∞ (assumed)
Observations	557,999	557,999

Note: * p<0.10, ** p<0.05, *** p<0.01. Clustered (by respondent) standard errors in parentheses. Column 1 results from separate models for each explanatory variable (being married & having children entered jointly). Since constants and cutoff points vary (slightly) across regressions in column 1, they are not reported here.

The numerical c values in table 7 show that reversals are not generally made harder with additional controls. Thus, adding such controls does not seem to reduce possible heterogeneities in the effects of explanatory variables across the distribution of hr_{it}. However, in order to give a correct interpretation of the degree to which happiness scales are transformed by the computed c values in table 7, we make the following observations. The large majority (about 75%) of the respondents in our sample reports happiness scores in the range of 5 to 9. In that range the differences between the estimated cutoff points in table 6 do not deviate much from their average 2.3. Therefore, for any given c, a transformation $\widetilde{hl}_{it} = e^{chl_{it}}$ is roughly 2.3 times as extreme as a similar transformation of rank-order coded hr_{it}. Hence, for a proper comparison with the latter transformations, we should multiply the c values in table 7 by 2.3. After this multiplication, it turns out that no numerical c value is below our benchmark of $c = 1$ from

Table 7. Bond and Lang's reversal conditions

		log HH income	Unemployed	Married	Children	Disability
No controls	c (via equation 3)	0.66	2.33	0.61	1.14	1.06
	c (numerical)	0.66	2.34	0.60	1.15	1.06
Full controls	c (via equation 3)	1.43	1.84	0.90	26.08	1.53
	c (numerical)	0.68	0.69	0.51	1.55	1.59

Note: Analytical reversal condition (3) for income is evaluated at the sample means of all explanatory variables.

section 4.1. The closest to this benchmark is the value required to just reverse the effect of being married with full controls, where we obtain $c = 2.3 * 0.51 = 1.17$. This implies an average ratio of $e^{1.17} = 3.22$ of subsequent differences between transformed cutoff points $\tilde{\tau}_k = e^{c\tau_k}$.

Thus, in view of the evidence in section 4.2, a sign-reversing scale with such a ratio does not even seem plausible for this least extreme value. Moreover, recall that our estimates of $hd_{k,it}$ in the pooled case yielded that reversals of the marriage effect by relabelling are impossible, since all cumulative response shares move in the same direction. Hence, as argued in section 3.2, the effect of marriage on some of the expected values $E(ht_i|hr = k)$, presumably $E(ht_i|hr = 10)$ or $E(ht_i|hr = 0)$, would have to be strongly negative in order to achieve a reversal. Such effects are neither empirically observed nor plausible in light of the observed upward shift in chosen response options.

6 Conclusions

In this paper we showed that reversals are fundamentally caused by variables having heterogenous effects across the distribution of happiness. We derived necessary and sufficient conditions under which reversals of OLS regression coefficients are made feasible. We also derived bounds on trade-off ratios of coefficients of explanatory variables under any permissible labelling scheme. Moreover, we argued that in cases where reversals by 'relabelling' are impossible, reversals using Bond and Lang's method are empirically unfounded. Furthermore, we presented arguments and evidence showing that respondents likely use the response scale in a roughly linear fashion. Finally, using GSOEP data, we empirically investigated the possibility and plausibility of reversals for a set of key variables. We found that reversals using relabelling are

largely impossible or implausible and that all reversals using Bond and Lang's method are implausible.

Thus, if the goal is to identify the direction of effects of explanatory variables, the worries flagged by Bond & Lang (2019) and Schröder & Yitzhaki (2017) do not appear to be a serious concern in practice. Nevertheless, since effects of explanatory variables are rarely homogenous across the distribution of happiness, trade-off ratios of coefficients may be severely affected by plausible transformations of reported happiness scores. We therefore urge researchers to verify the sensitivity of their results against plausible transformations of reported happiness. Moreover, we recommend that researchers ascertain whether our relabelling condition is satisfied for their application, thus verifying (or not) that their results are immune to reversals.

A limitation of our analysis is that we ignore any potential problems arising from heterogeneities in scale use. If such heterogeneities are correlated with explanatory variables, severe biases in estimated effects are possible. Hence, paper 4 of this thesis analyses these issues.

Lastly, our finding that the relative effects of explanatory variables are not homogenous across the distribution of happiness shows that estimating mean effects on happiness hides patterns in the data that are interesting and informative in their own right. As was previously done using quantile regressions (Binder and Coad 2011; 2015; Gupta et al. 2015), such patterns should be investigated more broadly.

References

Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press.

Banks, William P., and Mark J. Coleman. 1981. 'Two Subjective Scales of Number'. *Perception & Psychophysics* 29 (2): 95–105. https://doi.org/10.3758/BF03207272.

Banks, William P., and David K. Hill. 1974. 'The Apparent Magnitude of Number Scaled by Random Production.' *Journal of Experimental Psychology* 102 (2): 353.

Berthoud, Richard, and Mark Bryan. 2011. 'Income, Deprivation and Poverty: A Longitudinal Analysis'. *Journal of Social Policy* 40 (1): 135–156. https://doi.org/10.1017/S0047279410000504.

Bertram, Christine, and Katrin Rehdanz. 2015. 'The Role of Urban Green Space for Human Well-Being'. *Ecological Economics* 120: 139–52. https://doi.org/10.1016/j.ecolecon.2015.10.013.

Biewen, Martin, and Andos Juhasz. 2017. 'Direct Estimation of Equivalence Scales and More Evidence on Independence of Base'. *Oxford Bulletin of Economics and Statistics* 79 (5): 875–905. https://doi.org/10.1111/obes.12166.

Binder, Martin, and Alex Coad. 2011. 'From Average Joe's Happiness to Miserable Jane and Cheerful John: Using Quantile Regressions to Analyze the Full Subjective Well-Being Distribution'. *Journal of Economic Behavior & Organization* 79 (3): 275–90. https://doi.org/10.1016/j.jebo.2011.02.005.

———. 2015. 'Heterogeneity in the Relationship Between Unemployment and Subjective Wellbeing: A Quantile Approach'. *Economica* 82 (328): 865–91. https://doi.org/10.1111/ecca.12150.

Bless, Herbert, Fritz Strack, and Norbert Schwarz. 1993. 'The Informative Functions of Research Procedures: Bias and the Logic of Conversation'. *European Journal of Social Psychology* 23 (2): 149–165.

Bond, Timothy N., and Kevin Lang. 2019. 'The Sad Truth about Happiness Scales'. *Journal of Political Economy* 127 (4): 1629–40. https://doi.org/10.1086/701679.

Borah, Melanie, Carina Keldenich, and Andreas Knabe. 2019. 'Reference Income Effects in the Determination of Equivalence Scales Using Income Satisfaction Data'. *Review of Income and Wealth* 65 (4): 736–70. https://doi.org/10.1111/roiw.12386.

Boyce, Christopher J. 2010. 'Understanding Fixed Effects in Human Well-Being'. *Journal of Economic Psychology* 31 (1): 1–16. https://doi.org/10.1016/j.joep.2009.08.009.

DIW. 2016. 'GSOEP 2015 – Erhebungsinstrumente 2015 (Welle 32) Des Sozio-Oekonomischen Panels: Personenfragebogen'. *SOEP Survey Papers Series A - Survey Instruments (Erhebungsinstrumente)*.

Ferrer-i-Carbonell, Ada, and Paul Frijters. 2004. 'How Important Is Methodology for the Estimates of the Determinants of Happiness?' *The Economic Journal* 114 (497): 641–659.

Gupta, Prashant, Tapas Mishra, Nigel O'Leary, and Mamata Parhi. 2015. 'The Distributional Effects of Adaption and Anticipation to Ill Health on Subjective Wellbeing'. *Economics Letters* 136 (November): 99–102. https://doi.org/10.1016/j.econlet.2015.09.010.

Hurst, Erik, Geng Li, and Benjamin Pugsley. 2013. 'Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed'. *The Review of Economics and Statistics* 96 (1): 19–33. https://doi.org/10.1162/REST_a_00363.

Kapteyn, Arie. 1977. 'A Theory of Preference Formation'. Tilburg University. https://research.tilburguniversity.edu/en/publications/a-theory-of-preference-formation-2.

King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. 'Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research'. *American Political Science Review* 98 (1): 191–207. https://doi.org/10.1017/S000305540400108X.

Knabe, Andreas, Steffen Rätzel, Ronnie Schöb, and Joachim Weimann. 2010. 'Dissatisfied with Life but Having a Good Day: Time-Use and Well-Being of the Unemployed'. *The Economic Journal* 120 (547): 867–889.

Levinson, Arik. 2012. 'Valuing Public Goods Using Happiness Data: The Case of Air Quality'. *Journal of Public Economics* 96 (9): 869–80. https://doi.org/10.1016/j.jpubeco.2012.06.007.

Luechinger, Simon. 2009. 'Valuing Air Quality Using the Life Satisfaction Approach'. *The Economic Journal* 119 (536): 482–515.

Oswald, Andrew J. 2008. 'On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings'. *Economics Letters* 100 (3): 369–72. https://doi.org/10.1016/j.econlet.2008.02.032.

Parducci, Allen. 1995. *Happiness, Pleasure, and Judgment: The Contextual Theory and Its Applications*. Happiness, Pleasure, and Judgment: The Contextual Theory and Its Applications. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Plant, Michael. 2020. 'A Happy Possibility About Happiness (And Other) Scales: Understanding Why the Cardinality Assumption Is Both Defensible and Unnecessary'. *Unpublished Working Paper*.

Rojas, Mariano. 2007. 'A Subjective Well-Being Equivalence Scale for Mexico: Estimation and Poverty and Income-Distribution Implications'. *Oxford Development Studies* 35 (3): 273–293.

Schneider, Bruce, Scott Parker, Dan Ostrosky, David Stein, and Gary Kanow. 1974. 'A Scale for the Psychological Magnitude of Number'. *Perception & Psychophysics* 16 (1): 43–46. https://doi.org/10.3758/BF03203247.

Schröder, Carsten, and Shlomo Yitzhaki. 2017a. 'Revisiting the Evidence for Cardinal Treatment of Ordinal Variables'. *European Economic Review* 92 (February): 337–58. https://doi.org/10.1016/j.euroecorev.2016.12.011.

———. 2017b. 'Revisiting the Evidence for Cardinal Treatment of Ordinal Variables'. *European Economic Review* 92 (February): 337–58. https://doi.org/10.1016/j.euroecorev.2016.12.011.

Shepard, Roger N. 1981. 'Psychological Relations and Psychophysical Scales: On the Status of "Direct" Psychophysical Measurement'. *Journal of Mathematical Psychology* 24 (1): 21–57. https://doi.org/10.1016/0022-2496(81)90034-1.

Studer, Raphael. 2012. 'Does It Matter How Happiness Is Measured? Evidence from a Randomized Controlled Experiment'. *Journal of Economic and Social Measurement* 37 (4): 317–36. https://doi.org/10.3233/JEM-120364.

Van Praag, Bernard M. S. 1971. 'The Welfare Function of Income in Belgium: An Empirical Investigation'. *European Economic Review* 2 (3): 337–69.

———. 1991. 'Ordinal and Cardinal Utility: An Integration of the Two Dimensions of the Welfare Concept'. *Journal of Econometrics* 50 (1–2): 69–89.

———. 2015. 'A New View on Panel Econometrics: Is Probit Feasible After All?' *IZA Discussion Paper*, no. 9345.

Williams, Richard. 2006. 'Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables'. *Stata Journal* 6 (1): 58–82. https://doi.org/st0097.

Appendix

A An argument from psychophysics

Psychophysicists analyse how subjective intensities across different 'modalities' relate to intensities of objective stimuli. To do so, respondents are often asked to match a given stimulus s^m of modality m (e.g. a sound with a given physical intensity), which is associated with a subjective intensity ψ^m (i.e. that sound's subjective loudness), with another stimulus of a different modality s^n (e.g. the objective luminosity of a lamp) with intensity ψ^n (i.e. that lamp's subjective brightness). Such procedures are known as 'cross-modality matching' (Shepard 1981).

Answers to happiness questions can be placed in this framework. Note that a person's true happiness ht represents a subjective intensity ψ^{ht}.[40] Furthermore, the R numbered response options in any happiness question can be viewed as physical number stimuli $s_1^{num}, \ldots, s_R^{num}$, associated with subjective intensities $\psi_1^{num}, \ldots, \psi_R^{num}$. Although surveys are not explicitly framed as cross-modality matching tasks, respondents may attempt to match the subjective intensity of each numbered response option with their felt happiness intensity. When choosing a response option with number k, a respondent at time t may thus minimize the difference $\psi_k^{num} - \psi_t^{ht}$. In turn, this implies that when a respondent initially answered e.g. '6' and now answered e.g. '8', we can infer that the difference in happiness intensities is approximately equal to the difference in the subjective intensities of the number stimuli '6' and '8', i.e. that $\psi_{t=2}^{ht} - \psi_{t=1}^{ht} \approx \psi_{k=8}^{num} - \psi_{k=6}^{num}$. Therefore, if we were to identify the differences between subjective intensities of numbers, we can infer the differences in psychological magnitudes of ht.

Banks & Coleman (1981) tried to experimentally infer the function that relates objective with subjective magnitudes of a given finite set of numbers.[41] It is not feasible to directly ask how

[40] Happiness is unusual in that there is no single physical stimulus s^{ht} corresponding to a particular level of ψ^{ht}.
[41] Which is a function satisfying $f(s_k^{num}) = \psi_k^{num}$. We only report Banks & Coleman's results for cases with known bounded intervals within which numbers can occur. This is because only these results are directly relevant to typical

certain numbers "feel", since when e.g. asked whether the difference between the numbers "6" and "8" is larger than difference between the numbers "3" and "4", respondents would likely apply formal arithmetic rules. To circumvent this problem, Banks & Coleman utilized the following procedure: Respondents were given multiple series of ten different integers I on a known interval (e.g. from 1 to 1000). Each series was generated by a function of the form $I = ax^b$, with $x = 1, 2, \dots, 10$; $b = 0.1, 0.3, 0.5, 0.7, 0.9, 1.1$ (note that $b = 1$ would yield a linear function) and with a chosen such that the minimum and maximum of the series are as close as possible to the interval boundaries. Within each series, b was held constant while using every value of x. The resulting integers were then randomly shuffled. Respondents were next asked to rate how "random" these series of numbers felt. Under the assumption that respondents took "random" to mean "sampled from a uniform distribution", we should expect respondents to give series where the distribution of subjective intensities more closely resembled a uniform distribution to receive a higher subjective "randomness" rating.[42] It turns out that functions with exponents $b = 0.9$ or $b = 1.1$ received the highest mean randomness rankings across respondents. This suggests that the subjective intensity of numbers within a known bounded range is close to linear in the objective magnitude of such numbers. In a separate experimental setup, Banks & Coleman also asked respondents to rapidly think of 25 random numbers on some bounded interval (either 1-10, 10-99, or 100-999). Respondents' answers were put in rank order and means for each rank were calculated across respondents. OLS regressions of these means on their rank yielded R-squared statistics between 0.974 (for the interval 10-99) and 0.995 (for the interval 1-10). These results also suggest a linear relation between objective values of numbers and their subjective intensities.

---

response scales in happiness research. Banks & Coleman also present results for intervals without a known upper bound. In such cases, they find that power functions with exponents between 0.3 and 0.5 (i.e. concave functions) are a much better approximation of the subjective intensities of numbers. These results are in line with those of Schneider et al. (1974) as well as Banks & Hill (1974) who use similar methods.

[42] This assumes that how random a series of numbers feels only depends on the subjective intensities of the numbers in the series, and not on their objective magnitudes. Since felt randomness is a subjective perception itself, we find this assumption reasonable.

Given the aforementioned arguments, if the subjective intensity of numbers increases linearly with their objective magnitude on a bounded interval, we have further reason to believe that rank-order hr and ht also relate linearly.

Appendix B Additional tables

Table A1. Cumulative response shares for happiness and life satisfaction in ESS and WVS for European countries

WVS (Happiness) hr	ESS (Happiness) % share (cumulative)	ESS (Life Satisfaction) hr	WVS (Life Satisfaction) % share (cumulative)	$\overline{hr}$ after collapse	hr	% share (cumulative)	hr	% share (cumulative)
1	2.45 (2.45)	0	0.97 (0.97)	0.52	0	3.26 (3.26)	1	2.40 (2.40)
		1	1.04 (2.01)		1	2.15 (5.41)	2	1.99 (4.39)
2	13.06 (15.51)	2	2.11 (4.12)	3.23	2	3.37 (8.78)	3	4.16 (8.55)
		3	3.88 (8.00)		3	6.10 (14.88)	4	4.55 (13.11)
		4	4.47 (12.47)		4	5.92 (20.80)	5	11.93 (25.04)
3	58.98 (74.50)	5	14.60 (27.06)	6.79	5	14.78 (35.58)	6	10.83 (35.87)
		6	9.24 (36.30)		6	9.45 (45.03)		
		7	18.70 (55.00)		7	16.25 (61.29)	7	18.77 (54.64)
		8	24.14 (79.13)		8	21.24 (82.52)	8	25.27 (79.90)
4	25.50 (100.0)	9	11.86 (90.99)	9.43	9	9.38 (91.90)	9	11.79 (91.69)
		10	9.01 (100.0)		10	8.10 (100.0)	10	8.31 (100.0)

Note: Data from WVS wave 5 and ESS wave 3 (both 2006). Design and population weights applied. Countries included: France, Finland, Germany, Great Britain, The Netherlands, Norway, Poland, Romania, Russia, Slovenia, Spain, Sweden, Switzerland, Ukraine. WVS response options for happiness are labelled "Not at all happy" (=1), "Not very happy" (=2), "Rather happy" (=3), "Very happy" (=4). Extreme response options for happiness in ESS are labelled "Extremely unhappy" (=0) and "Extremely happy" (=10). Extreme response options for life satisfaction in ESS are labelled "Extremely dissatisfied" (=0) and "Extremely satisfied" (=10). Extreme response options for life satisfaction in WVS are labelled "Completely dissatisfied" (=0) and "Completely satisfied" (=10).

Table A2. Relabelling condition for the pooled OLS model

	(1) hr ≤ 0	(2) hr ≤ 1	(3) hr ≤ 2	(4) hr<=3	(5) hr ≤ 4	(6) hr ≤ 5	(7) hr ≤ 6	(8) hr ≤ 7	(9) hr ≤ 8	(10) hr ≤ 9
					No Controls					
Log household income	-0.006***	-0.010***	-0.021***	-0.039***	-0.061***	-0.145***	-0.180***	-0.166***	-0.066***	**0.003****
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	**(0.001**)
Unemployed	0.015***	0.025***	0.052***	0.100***	0.150***	0.258***	0.295***	0.250***	0.102***	0.028***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.001)
Married	-0.002***	-0.004***	-0.009***	-0.016***	-0.023***	-0.035***	-0.040***	-0.043***	-0.013***	-0.004***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Children	-0.001**	-0.002***	-0.004***	-0.009***	-0.014***	-0.032***	-0.042***	-0.039***	-0.029***	-0.003*
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Disability	0.010***	0.019***	0.037***	0.066***	0.096***	0.176***	0.196***	0.166***	0.077***	0.014***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)
					Full Controls					
Log household income	-0.003***	-0.006***	-0.014***	-0.027***	-0.045***	-0.110***	-0.143***	-0.143***	-0.068***	-0.009***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Unemployed	0.013***	0.022***	0.044***	0.083***	0.121***	0.190***	0.208***	0.160***	0.058***	0.017***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.001)
Married	-0.003***	-0.005***	-0.011***	-0.021***	-0.030***	-0.053***	-0.063***	-0.065***	-0.031***	-0.008***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)
Children	**0.000**	-0.001	-0.002**	-0.005***	-0.010***	-0.020***	-0.029***	-0.035***	-0.023***	-0.008***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)
Disability	0.009***	0.017***	0.034***	0.060***	0.088***	0.149***	0.164***	0.149***	0.071***	0.026***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)
Observations	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999

Note: * p<0.10, ** p<0.05, *** p<0.01. Clustered (by respondent) standard errors in parentheses. Cells in bold have opposite sign, implying possibility of reversal.

Table A3. Relabelling condition for the FE model

	(1) hr ≤ 0	(2) hr ≤ 1	(3) hr ≤ 2	(4) hr<=3	(5) hr ≤ 4	(6) hr ≤ 5	(7) hr ≤ 6	(8) hr ≤ 7	(9) hr ≤ 8	(10) hr ≤ 9
					No Controls					
Log household income	-0.003***	-0.006***	-0.011***	-0.020***	-0.030***	-0.052***	-0.059***	-0.045***	-0.008***	**0.005*** **
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	**(0.001)**
Unemployed	0.010***	0.016***	0.035***	0.065***	0.093***	0.131***	0.143***	0.104***	0.033***	0.013***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)	(0.001)
Married	-0.002***	-0.004***	-0.010***	-0.017***	-0.023***	-0.036***	-0.040***	-0.033***	-0.009***	**0.002**
	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	**(0.002)**
Children	-0.000	-0.001**	-0.003***	-0.007***	-0.011***	-0.012***	-0.013***	-0.012***	-0.003+	-0.005***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Disability	0.005***	0.011***	0.021***	0.039***	0.061***	0.095***	0.104***	0.092***	0.045***	0.020***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)
					Full Controls					
Log household income	-0.002***	-0.005***	-0.010***	-0.019***	-0.030***	-0.054***	-0.067***	-0.065***	-0.033***	-0.009***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)
Unemployed	0.010***	0.016***	0.034***	0.064***	0.091***	0.129***	0.140***	0.103***	0.035***	0.015***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)	(0.001)
Married	-0.001*	-0.002***	-0.007***	-0.013***	-0.018***	-0.028***	-0.035***	-0.038***	-0.023***	-0.004**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)
Children	-0.000	-0.000	-0.001	-0.004**	-0.005***	-0.005**	-0.004+	**0.000**	**0.008*** **	**0.003** **
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	**(0.003)**	**(0.003)**	**(0.002)**
Disability	0.005***	0.009***	0.016***	0.028***	0.044***	0.070***	0.072***	0.053***	0.011***	**-0.001**
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	**(0.002)**
Observations	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999

Note: * p<0.10, ** p<0.05, *** p<0.01. Clustered (by respondent) standard errors in parentheses. Cells in bold have opposite sign, implying possibility of reversal.

Appendix C Additional figures

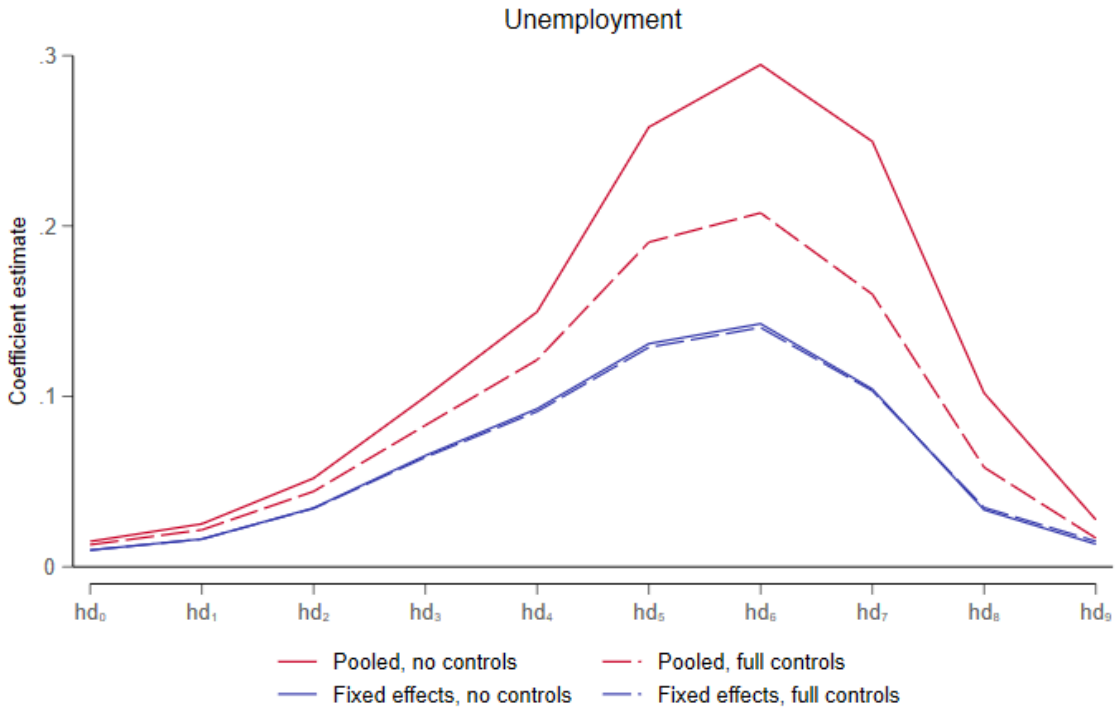


Figure A1. Coefficients estimates of unemployment for each regression of $hd_{k,it}$, corresponding to appendix tables A2 and A3.

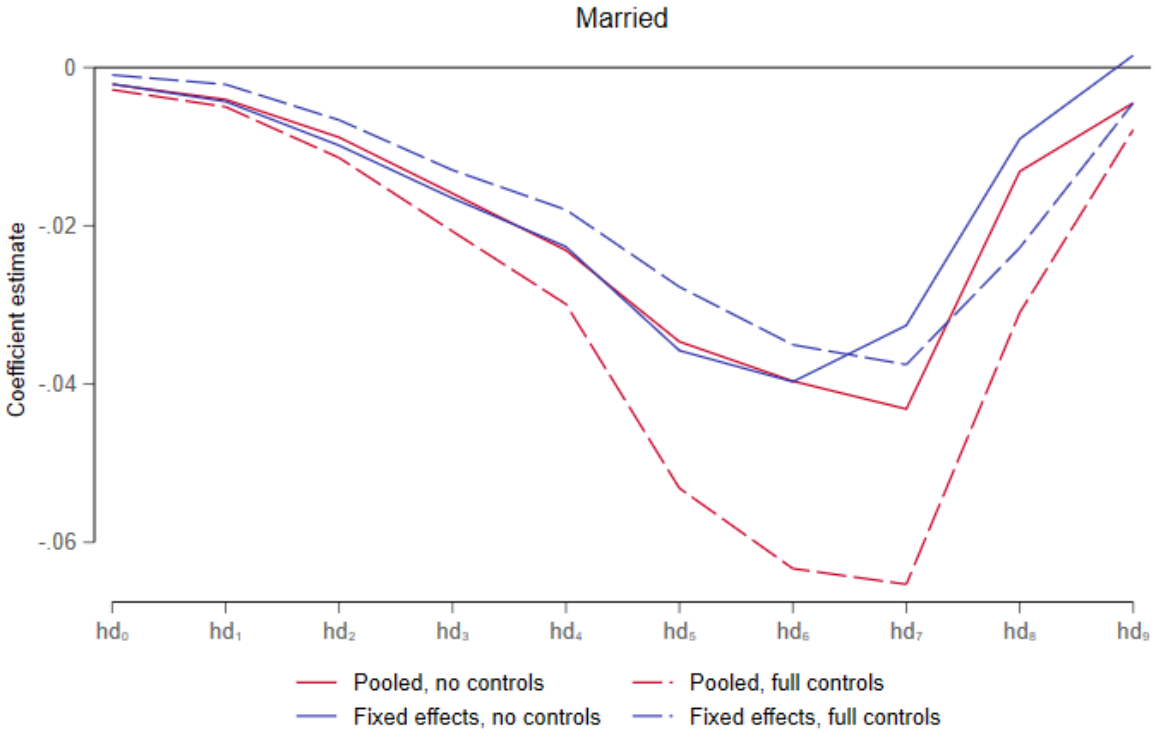


Figure A2. Coefficients estimates of marriage for each regression of $hd_{k,it}$, corresponding to appendix tables A2 and A3.

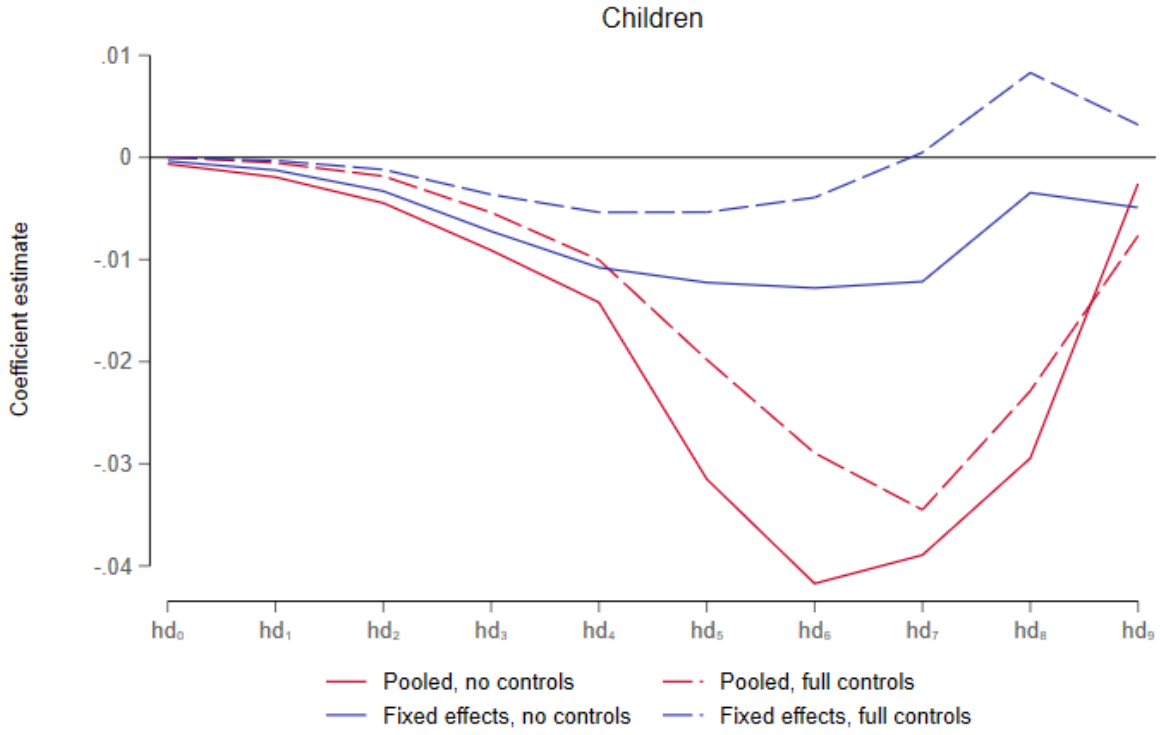


Figure A3. Coefficients estimates of children for each regression of $hd_{k,it}$, corresponding to appendix tables A2 and A3.

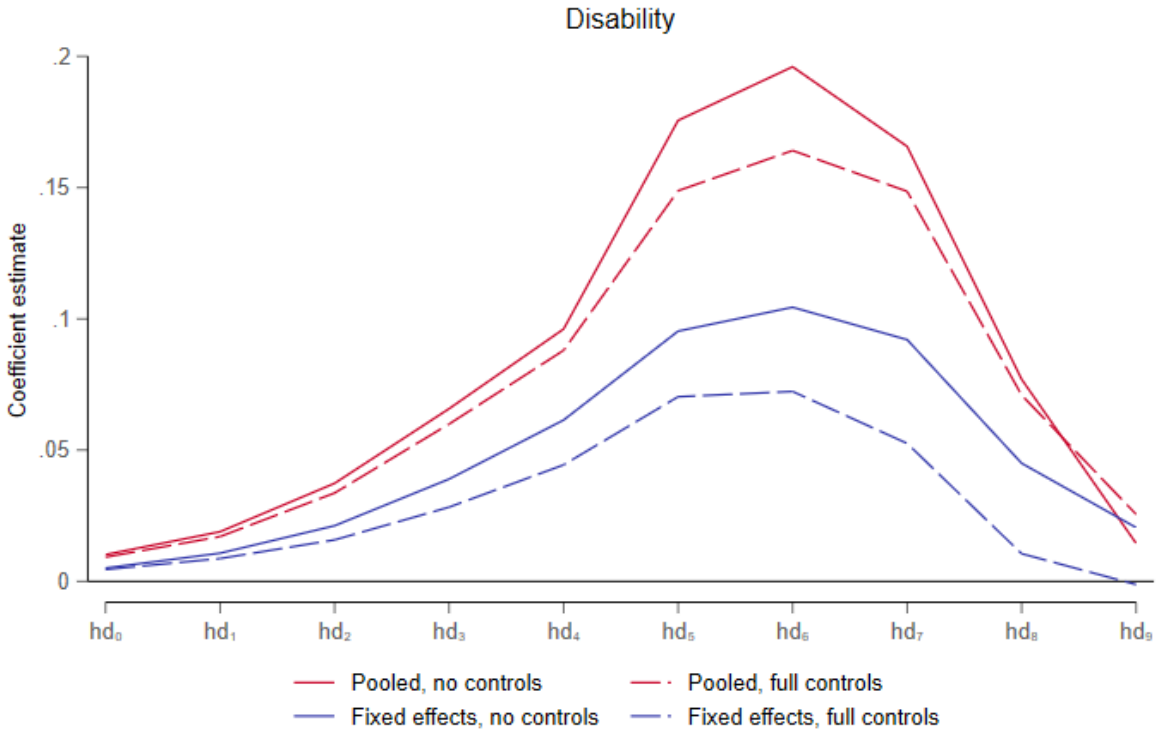


Figure A4. Coefficients estimates of disability for each regression of $hd_{k,it}$, corresponding to appendix tables A2 and A3.

Memories as anchors: Theoretical reflections and empirical analyses on the intrapersonal comparability of wellbeing reports

Caspar Kaiser

Abstract

Most empirical research on subjective wellbeing is concerned with establishing how explanatory variables are associated with the quality of unobserved mental states. Such analyses typically rely on the assumption that ratings of different respondents at different times are comparable. If this assumption of common scale use is violated, standard methods in empirical research may mislead. In attempting to explore and address this concern, I make three contributions. First, I give a theoretical analysis of the extent and direction of bias that may result from violations of the assumption of common scale use. Second, I propose to use respondents' stated memories of their past wellbeing to estimate and thereby to correct for differentials in scale use. Third, using the proposed approach, I test whether wellbeing reports are intrapersonally comparable across time. Using BHPS data, I find that the direction in which explanatory variables affect latent satisfaction is typically the same as the direction in which scale use is affected. Unemployment and bereavement appear to have particularly strong effects on scale use. Nevertheless, compared to an analysis that maintains common scale use, differences in results are never so large as to change our qualitative conclusions. This is good news for wellbeing research, as it suggests a minimum level of stability in scale use.

Keywords: *subjective wellbeing; life satisfaction; BHPS; vignettes; memories; differential item functioning*

1 Introduction

Most empirical research on subjective wellbeing (SWB) is concerned with establishing how explanatory variables are associated with the quality of unobserved or 'latent' mental states. Life satisfaction and happiness are examples of such mental states. In order to establish such associations, respondents are asked to rate their mental states by means of a numerical scale or verbal labels. Analyses based on such data typically rely on the assumption that ratings of different respondents at different times are comparable. Call this the assumption of *common scale use*. Outsiders to the field often find this assumption questionable (Adler 2013; Fleurbaey and Blanchet 2013; Viscusi 2020). Nevertheless, despite thousands of studies analysing subjective wellbeing and other subjective constructs,[1] few papers constructively engage with this issue. Are wellbeing statements interpersonally and intertemporally comparable? If such statements are not comparable, standard methods in wellbeing research may mislead. In this paper, I attempt to answer this question.

The paper proceeds as follows. In section 2, I introduce the standard set of assumptions adopted in empirical wellbeing research. I then focus on the assumption of *common scale use* and discuss several arguments as to why this assumption may be violated. In the next step, I formally explore the biases that result when *common scale use* is violated. In section 3, I first take note of vignettes (i.e. descriptions of fictional persons) as the only (to my knowledge) viable method to correct for differences in scale use. As an alternative, I then propose a new method to correct for differences in scale use. That method relies on variation in people's stated memories of past wellbeing. I discuss the plausibility of the method's identifying assumptions and derive the direction of bias in cases where my assumptions are violated. In section 4, I apply the proposed memory approach to

[1] While the ensuing discussion is stated in the context of subjective wellbeing, and life satisfaction in particular, it should be stressed that my theoretical points are in principle applicable to any type of data relating to subjective states of respondents.

investigate whether satisfaction reports are intrapersonally comparable across time. British Household Panel Study (BHPS) data are used for this purpose. Section 5 concludes.

My empirical findings are as follows. Variables that positively affect[2] latent satisfaction are found to also be associated with more 'stringent' scale use,[3] and *vice versa* for negative effects. Scale use is especially affected by unemployment and bereavement. Fortunately, compared to an analysis that maintains *common scale use*, differences in results are never so large as to change our conclusions regarding the direction and statistical significance of effects. Although this is generally good news for wellbeing research, ratios of coefficients – which may indicate the relative importance of explanatory variables - are affected by differences in scale use.

Moreover, lags and leads of explanatory variables have only limited effects on both latent satisfaction and scale use. The same holds for a set of social reference variables. Non-parallel shifts in scale use may be possible for some variables. However, such non-parallel scale shift effects turn out to be empirically indistinguishable from heterogeneous effects of covariates across the distribution of satisfaction.

2 The problem

2.1 Standard assumptions in wellbeing research

To analyse the determinants of life satisfaction, researchers typically use responses to questions of the form:

'How dissatisfied or satisfied are you with your life overall?'

When answering this question, each respondent i at time t chooses a response $r_{it} = k$ from a set of K options. In the British Household Panel Survey ('BHPS'), where this particular question is

[2] Especially when discussing my empirical estimations, I will often use phrases like "the effect of", "*x* affects *y*", etc. I am aware that these terms suggest causal relations which I cannot guarantee to hold. My estimates are only suggestive of such causal relations. However, all of my substantive points continue to hold if all estimated coefficients are interpreted as merely indicating descriptive differences in conditional means.

[3] I.e. making it harder to report a higher response category at a given level of latent satisfaction. See the next section for more details.

asked, there are seven response options with extremes labelled as *'Not at all satisfied'* (for $k = 1$) to *'Completely satisfied'* (for $k = 7$). We assume that respondents make their choice by adhering to a rule of the form:[4]

$$r_{it} = k \quad \leftrightarrow \quad \tau_{it,k-1} < s_{it} \leq \tau_{it,k}, \tag{1}$$

where s_{it} is the latent level of satisfaction that we are ultimately interested in. The τs are 'cutoffs' that need to be crossed to start reporting the next category. Call this rule the reporting function of life satisfaction (Oswald 2008).[5] When the cutoffs are shifted to take higher values, I say that scale use becomes more stringent. Vice versa, when cutoffs are shifted to take lower values, I say that scale use becomes less stringent (see figure 1).[6]

The literature standardly assumes that scale use is common across all respondents and across all times, and that cut-offs are equidistant. The former assumption can be stated as:

$$\tau_{it,k} = \tau_k, \tag{2}$$

which I call the assumption of *common scale use*. This assumption of *common scale use* will be the main subject of this paper. The assumption that cutoffs are equidistant can be written as:

$$\tau_k - \tau_{k-1} = c, \tag{3}$$

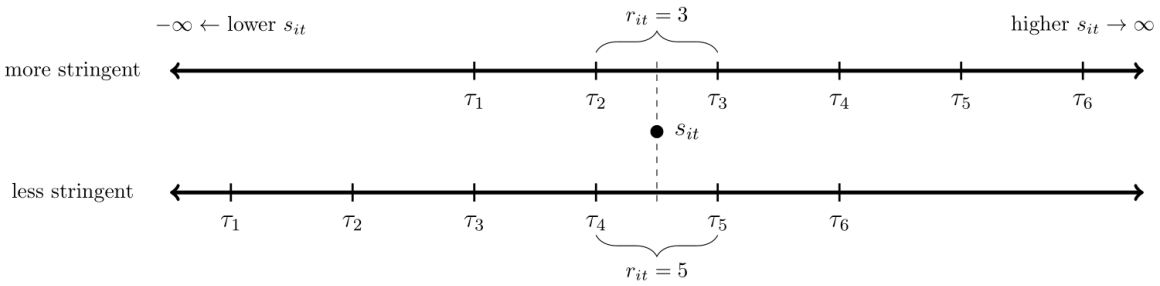
where c is some constant. We might call (3) the assumption of *equidistance*, or *linearity of scale use*. The plausibility and importance of this assumption is analysed in Schröder and Yitzhaki (2017), Bond and Lang (2019) and paper 3 of this thesis.

[4] This assumption may be questioned. We might doubt that s_{it} indeed determines the choice of r_{it}. Perhaps some other quantity, one that does not carry the sort of presumed normative content that motivates our analysis, drives the choice of r_{it} instead. Another possibility is that latent s_{it} does not really exist. Instead respondents only behave as though s_{it} exists when prompted (Bertrand and Mullainathan 2001). I leave these further doubts aside.

[5] Written differently, the reporting function is given by $r_{it}(s_{it}; \tau_{it}) = \sum_{k=1}^{K} \mathrm{I}(\tau_{it,k-1} < s_{it} \leq \tau_{it,k})k$, where $\mathrm{I}(.)$ denotes the indicator function. Thus, the value of the reporting function, i.e. r_{it}, is determined by both latent life satisfaction and the locations of thresholds.

[6] The idea of more or less stringent scale use in answers to wellbeing questions is analogous to differences in the stringency of academic grading systems: For a given level of student achievement, a less stringent grading system (e.g. that of the United States) assigns a comparatively high grade, whereas a more stringent grading system (e.g. that of the Netherlands) assigns a comparatively low grade.

Figure 1. Examples of more and less stringent scale use



Note: For a given level of s_{it}, we observe a lower level of reported satisfaction r_{it} (=3) with more stringent scale use and a higher level of reported satisfaction (=5) with less stringent scale use.

Jointly adopting assumptions (1)-(3) is taken to motivate the use of OLS regressions of r_{it} on a vector of variables $\boldsymbol{X}$ to infer the associations between s_{it} and $\boldsymbol{X}$ (Ferrer-i-Carbonell & Frijters, 2004).[7] If we are unwilling to assume *linearity of scale use*, we may infer differences $\tau_k - \tau_{k-1}$ by postulating a structural model for s_{it} and a distribution for the stochastic error of such a model. When assuming a normally distributed error, this yields the ordered probit approach.

Standard ordered probit regressions rely on *common scale use*.[8] Unfortunately however, violations of this assumption can lead to severe biases in the estimated effects of explanatory variables. These biases are formally shown in section 2.3. Before that, I will discuss some reasons for believing that response scales are not common over time and across respondents.

2.2 Are response scales common over time and across respondents?

An argument in favour of assuming *common scale use* is that respondents' own satisfaction ratings and the ratings of others about a respondent's satisfaction correlate (Diener, Inglehart, and Tay 2013). However, correlations between a respondent's own report and the ratings of others can be observed even if scale use differs across respondents and third-party raters. Such correlations only

[7] However, OLS is not necessarily sufficiently motivated even when assumptions (1)-(3) hold. Instead, in order for OLS to yield correct ratios of coefficients, we require the more stringent assumption $\mathrm{E}(s_{it}|r_{it} = k;\ \boldsymbol{X}) = \mathrm{E}(s_{it}|r_{it} = k) = a + bk$, where a and b are constants. This assumption entails that independent variables $\boldsymbol{X}$ do not affect the average level of latent satisfaction within each response category.

[8] Bond & Lang (2019) show that even if the assumption of *common scale use* is satisfied, violations of the normality assumption can lead to reversals of the signs of coefficient estimates. However, in paper 3 of this thesis Maarten Vendrik and I show that such reversals tend to be rather implausible.

establish that third-party raters are reasonably accurate at inferring interpersonal variation in s_{it}.[9] To defend *common scale use* one would have to show that there is no systematic variation across third-party raters in their ratings of subjects.[10]

Another argument notes that for any evaluative label (such as 'very satisfied' or '8 out of 10') to elicit appropriate behavioural responses, such evaluative labels must have shared meanings (Van Praag 1993; Ferrer-i-Carbonell and Frijters 2004; Plant 2020).

However, Diamond (2008), Schwarz and Strack (1999), and Fleurbaey and Blanchet (2013) assert that respondents likely report their life satisfaction in terms of their relative position in some reference distribution. Fleurbaey and Blanchet (2013) motivate this by noting that the scale with which respondents report their satisfaction is bounded, while many aspects of life are unbounded (like income). Thus, satisfaction reports may rather indicate the quantile of a respondent's life in a reference distribution of possible lives. If people's judgements are relative to a reference distribution, and if these distributions vary, then response scales cannot be common. Schwarz and Strack (1999) argue that reference distributions likely depend on intrapersonal and interpersonal comparisons. Indeed, Steffel and Oppenheimer (2009) obtain empirical evidence to suggest that people calibrate their scale against their past. In that case, better past outcomes will make todays' scale use more stringent. Similarly, if people calibrated their scale by means of fixed social comparisons, better outcomes of those in the comparison group will cause scale use to become more stringent.

However, a respondent's own contemporaneous characteristics will also influence comparisons. One such case are expectations, where respondent's contemporaneous characteristics determine her outlook of what is possible in her future.[11] This may have two effects. First, better

[9] To observe a positive correlation between third-party raters' responses and respondents' own responses, two conditions are sufficient. First, locations of thresholds and latent satisfaction of respondents must not be correlated to the extent that r_{it} is expected to decrease when s_{it} increases. Second, third-party raters' perceptions of respondents' latent satisfaction must be positively correlated with respondents' latent satisfaction.

[10] To do so, one may run rater-specific regressions of their ratings on the respondent's own ratings, and testing whether the estimated intercepts and slopes equal.

[11] See Dominitz and Manski (1997) for the case of income expectations.

contemporaneous outcomes may improve expectations about the future, thus increasing current latent life satisfaction (Clark and Georgellis 2013). Second, altered expectations may change the standard against which current satisfaction is evaluated. This may cause respondents to adjust their scale such that future events can be accommodated by the updated scale.

To illustrate, consider *Lukas*, a final year PhD student. He would rate his life a 7 out of 10. If someone were to give him a paid academic position, one of his life's major goals would be accomplished. Some months later, Lukas obtains a three-year position as a postdoc. Consequently, he feels a great sense of satisfaction and feels inclined to rate himself a 9 out of 10. However, the satisfaction of a tenure track position, an achievement he did not dare dream of when a graduate student, is now no longer impossible. Lukas wants to reserve a 10 out of 10 for a truly exceptional outcome. Thus, a tenure-track position can be no better than a 9 out of 10. However, a postdoc is worse than a tenure-track position. He must therefore score his life with an 8 out of 10. Thus, on Lukas' old scale, as a student, a postdoc position would be a 9 out of 10. However, on his new scale, as a postdoc, a postdoc position is only rated an 8 out of 10.[12]

A second reason why current circumstances may determine response scales is that contemporaneous attributes likely determine the set of members of one's comparison group (Falk and Knell 2004). One the one hand, such endogenously updated reference groups could reduce the magnitudes of the effects of explanatory variables on latent satisfaction.[13] On the other hand, if respondents indeed interpret satisfaction scales as being relative to people like them, then response scales should shift following variation in personal characteristics.

To illustrate, consider *Mary*. Last year, she rated her life a 7 out of 10. She imagined that losing her partner would be devastating, resulting in a satisfaction of 2 out of 10. This year, Mary's partner does die. Consequently, life indeed feels much less satisfying. However, when comparing herself

[12] See Sprangers and Schwartz (1999) for similar reasoning.
[13] This might help to explain why people tend to overestimate the impacts of future events on their wellbeing (Wilson and Gilbert 2005).

to others that also lost a partner, Mary thinks that she is doing relatively well. Hence, taking into account the newly acquired comparison, she rates her life a 5 out of 10. Therefore, in response to acquiring a partially new standard of comparison, her scale use has become less stringent.[14]

Quantitative evidence for this behaviour is given by Gibbons (1999). He reports on an experiment in which respondents were asked to complete a multi-item life satisfaction scale, as well as a questionnaire on self-rated functioning in various domains. After answering the first half of the life satisfaction scale questions, one group was made aware of the experimenter being confined to a wheelchair. Those respondents who were made aware of the experimenter's condition, showed a significant correlation between self-rating on physical ability and life satisfaction, which even held after controlling for life satisfaction ratings made just before revealing the experimenter's condition. Thus, respondents immediately changed their scale use once they were made aware of a new salient comparison standard.[15]

To summarise, interpersonal and intrapersonal comparisons plausibly affect scale use. When considering variation in a respondent's past characteristics, or the characteristics of a fixed social reference group, changes that are expected to decrease respondents' own satisfaction are also expected to make the response scale more stringent. However, when considering contemporaneous or future values (via anticipation) of respondents' personal characteristics, we should expect the direction in which scale use is impacted to concur with the direction in which latent satisfaction is impacted.

[14] Behaviours similar to this example, in which people 'count their blessings', are described in the qualitative interviews with cancer patients discussed by Taylor, Wood, and Lichtman (1983). Of course, in all these cases it is possible that changing one's relevant comparison group also raises latent satisfaction.

[15] Again, the contrary interpretation, namely that latent life satisfaction increased following the reveal, is also possible. However, that interpretation seems less plausible, and if true, would seem to undermine the importance of life satisfaction as something normatively valuable. Similar experimental evidence is given in Dermer et al. (1979). There respondents were asked to 'vividly imagine' positive (e.g. being a multimillionaire) or negative (e.g. suffering from a serious cancer) life events, and then to report on their current life satisfaction. Respondents' life satisfaction judgements before the experiment, and judgements made immediately after the treatment differed significantly. Respondents in the positive (negative) condition showed lowered (increased) satisfaction. This may be interpreted as evidence for a response shift if we assume that latent life satisfaction did not change following the treatment.

2.3 Biases arising from violations of the common scale use assumption

In the ordered probit approach we usually estimate a model of the following form:[16]

$$s_{it} = \boldsymbol{X}_{it*}\boldsymbol{\beta}_* + \varepsilon_{it} \tag{4}$$

Here, $\boldsymbol{X}_{it*} \equiv (\boldsymbol{X}_{it+L}, \ldots, \boldsymbol{X}_{it+l}, \ldots, \boldsymbol{X}_{it+F})$, where each element is a row vector of J independent variables, defined as $\boldsymbol{X}_{it+l} \equiv (X_{1,it+l}, X_{2,it+l}, \ldots, X_{J,it+l})$. Moreover, $\boldsymbol{\beta}_* \equiv (\boldsymbol{\beta}_L, \ldots, \boldsymbol{\beta}_l, \ldots, \boldsymbol{\beta}_F)'$, where each element is a column vector $\boldsymbol{\beta}_l \equiv (\beta_{1,l}, \ldots, \beta_{j,l}, \ldots, \beta_{J,l})'$ of parameters. Constants $L \leq 0$ and $F \geq 0$ respectively denote the numbers of included lags and leads. When no intrapersonal comparisons and anticipation effects are considered, no such lags and leads are included (i.e. $L = F = 0$). ε_{it} is assumed to be a normally distributed error with mean zero and standard deviation 1. A constant is omitted from equation (4). $\boldsymbol{X}_{it*}$ may contain personal characteristics as well as attributes of a (fixed) social comparison group.

Assume that reports r_{it} for $k = 1,2,..,K$ are generated by a rule of the form given in equation (1), imposing $\tau_0 = -\infty$, $\tau_K = \infty$. To specify all other cutoffs, write:

$$\tau_{it,k} = \gamma_k + \boldsymbol{X}_{it*}\boldsymbol{\delta}_* \tag{5}$$

For *common scale use* to hold, it is necessary that $\boldsymbol{\delta}_* = \boldsymbol{0}$, i.e. that neither contemporaneous nor lagged, nor leaded values of covariates affect the location of the thresholds. The contribution of each observation to the log-likelihood of the ordered probit model with the extension of equation (5) is then given by (see e.g. Söderbom 2011):

$$\ell_{it} = \ln \sum_{k=1}^{K} \mathrm{I}(r_{it} = k)[\Phi(\gamma_k + \boldsymbol{X}_{it*}\boldsymbol{\delta}_* - \boldsymbol{X}_{it*}\boldsymbol{\beta}_*) - \Phi(\gamma_{k-1} + \boldsymbol{X}_{it*}\boldsymbol{\delta}_* - \boldsymbol{X}_{it*}\boldsymbol{\beta}_*)]. \tag{6}$$

Here $\Phi(.)$ denotes the standard normal cumulative distribution function. $\mathrm{I}(.)$ denotes the indicator function. Throughout we maximise $\ell = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \ell_{it}$ to estimate parameters. T_i denotes

[16] Analogous analyses could also be given for an OLS model of rank-order r_{it}, but are omitted here for brevity.

the number of observations for respondent i, and N is the number of respondents. Since $\boldsymbol{X}_{it*}\boldsymbol{\delta}_* - \boldsymbol{X}_{it*}\boldsymbol{\beta}_* = \boldsymbol{X}_{it*}(\boldsymbol{\delta}_* - \boldsymbol{\beta}_*)$, equation (6) implies that $\boldsymbol{\delta}_*$ and $\boldsymbol{\beta}_*$ are not separately identified (King et al. 2004). Only the difference $\boldsymbol{\beta}_* - \boldsymbol{\delta}_*$ is identified.

A severe consequence of violating *common scale use* is that we might estimate incorrect signs for some explanatory variables. Suppose $\beta_{j,0} > 0$ and $\delta_{j,0} > 0$ for some explanatory variable $X_{j,it}$. An increase in the value of $X_{j,it}$ is here associated with a higher level of s_{it} and more stringent scale use. As discussed in the previous section, this may occur when $X_{j,it}$ records personal characteristics, variation of which could change a respondent's relevant comparison group. Now suppose that $\beta_{j,0} < \delta_{j,0}$. Hence, variation in $X_{j,it}$ impacts the location of the thresholds more strongly than it impacts latent satisfaction. In large samples, our estimate $(\widehat{\beta_{J,0} - \delta_{J,0}})$ would then be negative. If we interpreted this estimate as an estimate of $\beta_{j,0}$, we would falsely conclude that the effect of $X_{j,it}$ on s_{it} is negative, when in fact it is positive.

The opposite case is likely to occur when considering lagged values of personal covariates. If people adapt to changes in covariates (Frederick and Loewenstein 1999; Clark and Georgellis 2013; Wolbring, Keuschnigg, and Negele 2013), we should expect a variable whose contemporaneous effect is positive to have a negative lagged effect. However, the lagged effect of such a variable on scale use is likely positive. Hence, we expect that if $\beta_{j,-1} < 0$, then $\delta_{j,-1} > 0$. This means that if we are only able to estimate the difference $\beta_{j,-1} - \delta_{j,-1}$, and interpret this directly as an estimate of adaptation, we are likely to overstate the extent of adaptation observed in empirical data.[17] Conversely, in cases where we expect lags of variables to reinforce a smaller contemporaneous effect, such reinforcement will be underestimated.

For leads of explanatory variables, the previous section suggested equal signs for $\beta_{j,1}$ and $\delta_{j,1}$. If I expect a good to occur in the future, I should both be more satisfied this year and use the scale

[17] Since, if there is indeed adaptation and a positive lagged scale shift, $\beta_{j,-1} - \delta_{j,-1} < \beta_{j,-1} < 0$.

more stringently, so as to accommodate the future good. It then follows that anticipation effects on latent satisfaction will be underestimated. However, if X_j were to refer to an attribute of a fixed social comparison group (e.g. mean income in the reference group), then the previous section suggested opposite signs for $\beta_{j,*}$ and $\delta_{j,*}$.[18] Consequently, an analysis which does not account for scale shifts likely overstates the effects of fixed reference groups on latent satisfaction.

Finally, wellbeing research is increasingly interested how variables X_j and X_p can be traded off while maintaining constant levels of s_{it} (Luechinger 2009; Borah, Keldenich, and Knabe 2019; Frijters et al. 2019). We will then be interested in the ratio β_j/β_p (dropping the '0' subscript). However, we only obtain an estimate of $(\beta_j - \delta_j)/(\beta_p - \delta_p)$. Unless $\beta_j = \beta_p$ and $\delta_j = \delta_p$, biases occur in this context. Assuming that coefficients are ordered as $\beta_j \geq \beta_p > \delta_p \geq \delta_j \geq 0$ implies that our estimated trade-off ratio will be (weakly) positively biased. Likewise, assuming that $\beta_p \geq \beta_j \geq 0$, $\delta_j \geq \delta_p \geq 0$ and $\beta_p > \delta_p$ implies that our trade-off ratio will be (weakly) negatively biased.[19]

Thus, when the assumption of common response scales is violated and when response scales are correlated with explanatory variables, we might obtain incorrect signs of coefficients, false impressions of the magnitudes of effects, and biased trade-off ratios between variables. However, variation in scale use is entirely unproblematic if one is merely interested in analysing marginal effects of explanatory variables on the probability of responding with a certain category. This, together with some remarks on interpretations of absolute magnitudes of coefficients more generally, is shown in Appendix A.

[18] This should hold for lagged, contemporaneous and leaded values.

[19] The difference $(\beta_j - \delta_j)/(\beta_p - \delta_p) - \beta_j/\beta_p$ can be rearranged to give $(\beta_j\delta_p - \beta_p\delta_j)/\beta_p(\beta_p - \delta_p)$. The numerator is positive or zero if $\beta_j \geq \beta_p \geq 0$ and $\delta_p \geq \delta_j \geq 0$. The denominator is positive if $\beta_p > \delta_p \geq 0$. The whole expression is therefore positive or zero if $\beta_j \geq \beta_p > \delta_p \geq \delta_j \geq 0$. Similarly, the numerator is negative or zero if $\beta_p \geq \beta_j \geq 0$ and $\delta_j \geq \delta_p \geq 0$.

3 Anchoring response scales

In order to separately identify scale shifts from effects on latent satisfaction, previous research anchored response scales by presenting respondents with short vignettes that describe an imaginary person's life (King et al., 2004). In that approach, respondents are asked to rate the life satisfaction of the persons described in the vignettes. On the assumption that the vignettes' perceived levels of satisfaction are uncorrelated with respondents' personal characteristics, identification can be achieved. This assumption is known as the *vignette equivalence* assumption.

In the case of life satisfaction research, Kapteyn, Smith, and Van Soest (2010), Corrado and Weeks (2010), Angelini et al. (2013), Bertoni (2015) and Montgomery (2017) have adopted the vignette approach. In line with my arguments of sections 2.2 and 2.3, these studies generally find that correcting for differences in scale use increases estimates of the effects of explanatory variables on latent satisfaction. A more extensive discussion of that approach and appraisal of the associated literature is given in Appendix B.

However, the central assumption of the vignette approach, i.e. *vignette equivalence*, is untestable. As argued in Appendix B, violations of this assumption will likely lead to downwardly biased coefficients. Moreover, vignettes are expensive to administer, and the SHARE survey, which only samples individuals aged 50 or above, is the only accessible survey that contains vignettes relating to life satisfaction. There also seem to be no panel datasets that contain vignettes on life satisfaction. Thus, simultaneously accounting for time-invariant characteristics and scale use differences is currently impossible. Previous studies on life satisfaction that employ a vignette methodology are therefore hard to compare with studies employing panel data. Since biases from unobserved time-constant variables are potentially severe, previous vignette studies on life satisfaction are likely biased even if *vignette equivalence* holds. We should thus consider alternative approaches.

3.1 Using memories

This section sets out a strategy by which memories of life satisfaction can be used in a similar manner as vignettes. The general idea of using memories to identify shifts in scale use was first proposed by Howard and colleagues (Howard et al. 1979; Howard and Dailey 1979; Howard, Schmeck, and Bray 1979). These studies implemented the following experimental design. They first obtained a self-report at t_1 (the 'pre-test'). At t_2 they performed the treatment. At t_3, they asked respondents to recall their performance on the measure of interest at t_1 (the 'then-test'). They also asked respondents to evaluate their current performance at t_3 (the 'post-test'). A change in scale use stringency is interpreted to occur if the treatment is correlated with the mean difference between 'pre-test' and 'then-test' scores. Other studies employing such 'then-tests' include Hoogstraten (1985), Sprangers et al. (1999), Ahmed et al. (2004) or Townsend and Wilton (2003).

A more recent inspiration is the paper of Fabian (2019), who asked respondents about both their current life satisfaction and their life satisfaction ten years ago. Both of these questions are scored on a scale from 1 to 10. He then asked respondents to draw on a graph the development of their life satisfaction for each year between the present and 10 years ago. Inconsistencies between reports about changes in wellbeing as given in the graph, and reports about levels of life satisfaction now and ten years ago may indicate scale shifts. For example, if a respondent reports a '7' for the present year, a '6' for life satisfaction 10 years ago, and returns a drawing that indicates a flat or decreasing trajectory, respondents either have inconsistent memories about changes and levels of satisfaction, or the meaning of a '6' 10 years ago indicates a level of satisfaction similar to or higher than a level of '7' this year. In the latter case, scale use would have become less stringent. Fabian finds that between 7.8% (sample of general population) and 9.3% (sample of postgraduate students) are 'reversers' of this sort. In total he finds that between 12.6% (general population) and 16.2% (students) conform to some looser definition of likely scale adjustment.

3.2 Outline of memory approach

One possible question about memories of life satisfaction is:

'Thinking about last year rather than this year, how dissatisfied or satisfied were you with your life overall?'

Such a question would be closest to the 'then-test' methodology of Howard and colleagues. In the framework of the present paper, a natural rule by which a response $r_{it}^{(m)}$ to this question is chosen mirrors equation (1):

$$r_{it}^{(m)} = k \quad \leftrightarrow \quad \tau_{it,k-1} < s_{it}^{(m)} \leq \tau_{it,k}. \tag{7}$$

Here, $s_{it}^{(m)}$ denotes the respondent's memory of satisfaction, which is not directly observed. Respondents are assumed to use the same set of thresholds for their current satisfaction (equation (1)) and their memories of past satisfaction (equation (7)). Following the literature on vignettes, call this the assumption of *response consistency*.

The BHPS, to be used in the empirical analysis of section 4, asks a question about change in satisfaction:

Would you say that you are more satisfied with life, less satisfied or feel about the same as you did a year ago?

Here, respondents have only three response options (*more satisfied, less satisfied, about the same*). In this context, let $r_{it}^{(m)} = -1$ denote that a respondent reports to be less satisfied this year, let $r_{it}^{(m)} = 0$ denote that a respondent reports 'about the same' satisfaction, and let $r_{it}^{(m)} = 1$ denote that a respondent reports to be more satisfied this year than last year. Notice that $r_{it}^{(m)} = -1$ means $s_{it} < s_{it}^{(m)}$ and $r_{it}^{(m)} = 1$ means $s_{it} > s_{it}^{(m)}$. Furthermore, $r_{it}^{(m)} = 0$ means $s_{it} = s_{it}^{(m)} + \epsilon_{it}$, where ϵ_{it} can be positive or negative, and we do not know its size.

Recall that equation (1) states that $r_{it} = k$ implies $\tau_{it,k-1} < s_{it} \leq \tau_{it,k}$. Hence, we find the following implied reporting rule for memories:[20]

$$\begin{aligned} r_{it}^{(m)} = -1 \;\; \& \;\; r_{it} = k \;\; &\leftrightarrow \;\; s_{it}^{(m)} > \tau_{it,k-1} \\ r_{it}^{(m)} = 1 \;\; \& \;\; r_{it} = k \;\; &\leftrightarrow \;\; s_{it}^{(m)} < \tau_{it,k} \end{aligned} \tag{8}$$

This data can potentially be used to identify threshold locations. To do so, assume that one's memory at t about last period's life satisfaction is generated by:

$$s_{it}^{(m)} = \theta^{(m)} + \boldsymbol{X}_{it-1}\boldsymbol{\alpha} + \boldsymbol{X}_{it}\boldsymbol{\lambda}^{(m)} + \varepsilon_{it}^{(m)} \tag{9}$$

Here, $\varepsilon_{it}^{(m)}$ is a normally distributed error term with mean 0 and standard deviation 1. $\theta^{(m)}$ is an intercept to be estimated. For ease of exposition assume that there are no lags and leads in equation (1). This will be relaxed in section 3.4. In that case and if $\boldsymbol{\lambda}^{(m)} = \boldsymbol{0}$, the structure of the equation determining memories of satisfaction is the same as the structure of the equation determining current satisfaction. In so far as memories of satisfaction will be primarily determined by how satisfied a respondent in fact was, this is a natural choice.

Indeed, in order to point-identify the effects of contemporaneous values of covariates on scale shifts, we must assume $\boldsymbol{\lambda}^{(m)} = \boldsymbol{0}$. Call this assumption *present independence*. Substantively this assumption states that, conditional on values of last year's covariates, one's memories about last year's satisfaction do not depend on values for this year's covariates.

Present independence does not entail that reports $r_{it}^{(m)}$ are unbiased reports of past satisfaction s_{it-1}. This would imply the constraint that $\theta^{(m)} = 0$ and $\boldsymbol{\alpha} = \boldsymbol{\beta}$. I therefore allow for respondents to have biased memories of their past satisfaction and to misjudge the direction of the effects of past variables on past satisfaction (see e.g. Wilson, Meyers, and Gilbert 2003).

[20] $r_{it} = k$ & $r_{it}^{(m)} = 0$ implies $\tau_{it,k-1} < s_{it}^{(m)} + \epsilon_{it} \leq \tau_{it,k}$. Since we do not know the value of ϵ_{it}, cases where respondents answer 'about the same' are not informative about which reporting thresholds are crossed.

Using equations (1), (4), (6) and (9), together with the assumed distributions for each error term and maintaining $\boldsymbol{\lambda}^{(m)} = \mathbf{0}$, individual contributions to the log-likelihood of this model are given by:

$$\begin{aligned}\ell_{it} = \ln \sum_{k=1}^{K} \mathrm{I}(r_{\mathrm{it}} = k)[\Phi(\gamma_k + \boldsymbol{X}_{it}\boldsymbol{\delta} - \boldsymbol{X}_{it}\boldsymbol{\beta}) - \Phi(\gamma_{k-1} + \boldsymbol{X}_{it}\boldsymbol{\delta} - \boldsymbol{X}_{it}\boldsymbol{\beta})] \\ + \sum_{k=1}^{K} \mathrm{I}\left(r_{it}^{(m)} = k\right)\left[\Phi\left(\gamma_k + \boldsymbol{X}_{it}\boldsymbol{\delta} - \theta^{(m)} - \boldsymbol{X}_{it-1}\boldsymbol{\alpha}\right) - \Phi\left(\gamma_{k-1} + \boldsymbol{X}_{it}\boldsymbol{\delta} - \theta^{(m)} - \boldsymbol{X}_{it-1}\boldsymbol{\alpha}\right)\right]\end{aligned} \tag{10}$$

When using data to which equation (8) applies, as is the case for BHPS data, the contributions are defined similarly:

$$\begin{aligned}\ell_{it} = \ln \sum_{k=1}^{K} \mathrm{I}(r_{\mathrm{it}} = k)[&\Phi(\gamma_k + \boldsymbol{X}_{it}\boldsymbol{\delta} - \boldsymbol{X}_{it}\boldsymbol{\beta}) - \Phi(\gamma_{k-1} + \boldsymbol{X}_{it}\boldsymbol{\delta} - \boldsymbol{X}_{it}\boldsymbol{\beta}) \\ &+\mathrm{I}(r_{it}^{(m)} = -1)(1 - \Phi(\gamma_{k-1} + \boldsymbol{X}_{it}\boldsymbol{\delta} - \theta^{(m)} - \boldsymbol{X}_{it-1}\boldsymbol{\alpha})) \\ &+\mathrm{I}(r_{it}^{(m)} = \ \ 1)\Phi(\gamma_k + \boldsymbol{X}_{it}\boldsymbol{\delta} - \theta^{(m)} - \boldsymbol{X}_{it-1}\boldsymbol{\alpha})]\end{aligned} \tag{11}$$

Parameters $\boldsymbol{\beta}$ (effects on current satisfaction), $\boldsymbol{\alpha}$ (effects on memories of past satisfaction) and $\boldsymbol{\delta}$ (scale shift effects) are all separately identified by equations (10) or (11).

3.3 Plausibility of identifying assumptions

Since it seems implausible for respondents to vary their scale use when answering two questions about themselves, *response consistency* appears to be a weak assumption.

In contrast, *present independence* may be a strong assumption. Although *present independence* does not require that there be no memory biases, it requires that the magnitude of any memory bias does not depend on values of current covariates. There are at least four mechanisms by which memory biases could depend on current values of covariates.

First, respondents might react to differences in objective circumstances between the present and the past and infer that a concordant change in the subjective quality of their lives must have taken place. A study by Conway and Ross (1984) is suggestive of this. There, a group of participants were

treated with a program to improve their study skills. When asked to recall their abilities before the treatment (call these 'then-scores'), participants recalled their abilities as worse than what they had self-reported before the treatment ('pre-scores'). However, participants did not in fact perform better on a final exam, nor did they differ in their self-assessment after the treatment ('post-scores') suggesting that the treatment had no actual effect on study skills. Moreover, those in the control group showed no significant differences between their pre-scores and their then-scores. Conway & Ross interpret this as respondents biasing their memories to be less favourable, such that the expected gain from the treatment on objective performance is at least subjectively realised. Their results suggest opposite signs for $\boldsymbol{\lambda^{(m)}}$ and $\boldsymbol{\beta}$.[21]

A second source of potential memory bias can be found in Easterlin (2001). That argument first notes a paradox. Although people tend to report stable life satisfaction across the life course despite income growth,[22] they also typically report that the past was worse (and expect the future to be better). Easterlin explains this observation using income aspirations: When recalling their past life satisfaction, people might do so by judging their past incomes on the basis of current aspirations. In turn, aspirations might depend on current incomes. Current incomes are likely larger than past incomes, and so aspirations will have increased. Consequently, people will evaluate their past as worse than it actually was experienced in the past and the degree to which their remembered life satisfaction is biased will negatively depend on current income. Thus, *present independence* would be violated with $\boldsymbol{\lambda^{(m)} < 0}$ and $\boldsymbol{\beta = 0}$.[23]

[21] In so far as Conway & Ross (1986) ask respondents to self-report both before and after the treatment about their ability before the treatment, their design is identical to that of Howard et al. (1979). Howard et al. (1979) also observed that participants in the control group showed smaller differences between 'pre-scores' and 'then-scores', than treated participants. Next to the self-reports, Howard et al. additionally had third-party observers rate participants ('observer scores'). Howard et al. found that correlations of differences of 'post-scores' and 'then-scores' with differences of 'observer-scores' were stronger than correlations using differences of 'post-scores' and 'pre-scores', instead. Moreover, in another design, Howard et al. (1979) asked respondents to recall the actual numbers they reported prior to treatment. No significant differences were found between such 'memory-scores' and the 'pre-scores', indicating accurate memories.

[22] But see e.g. Blanchflower and Oswald (2017) on a possible U-shape in life satisfaction.

[23] However, this argument is limited to variables that generate aspirations. The argument also relies on assuming that people cannot recall their past aspirations but are able to recall their past incomes, which may be doubted. Moreover, the argument relies on assuming that aspirations depend on current incomes. A more natural view is that aspirations

A third reason why *present independence* may be violated is mood-congruency. Positive (negative) moods bias memories to be more positive (negative). For example, Urban et al. (2018) utilize a sample of US-Americans who were asked to record their emotions each day of a week. Following that week, respondents' scores for depressive symptoms positively predicted an upward bias in their recall of negative emotions and a downward bias in the recall of positive emotions. As another case, Prati (2017) shows that respondents who are subjectively more satisfied with their incomes recalled their incomes to have been larger than in reality, while those with low satisfaction had a negatively biased recall of their incomes. Experimental evidence of mood congruency is given by e.g. Lewis et al. (2005) or Klaassen et al. (2002). We should therefore expect that people in more positive moods recall their past satisfaction as more positive. Now note that explanatory variables that affect life satisfaction typically impact affect and domain-specific satisfaction in the same direction. Consequently, people's tendencies towards mood-congruent memories may cause an explanatory variable's effect on $\boldsymbol{\lambda}^{(m)}$ to have the same direction as that variable's effect on $\boldsymbol{\beta}$.

A fourth potential source of bias is anticipation, i.e. that satisfaction at t may in part be driven by expectations about explanatory variables' values at $t+1$ (Clark & Georgellis, 2013; Frijters et al., 2011; Vendrik, 2013). Assuming that memories of satisfaction depend on actual past satisfaction, equation (9) for memories of satisfaction should thus include an anticipation term.[24] However, since we model memories of satisfaction during $t-1$, that anticipation term now concerns values of covariates at t. Consequently, such anticipation terms depend on $\boldsymbol{X}_{it}$. Anticipation typically runs in the same direction as contemporaneous effects. Hence, if we observe that a variable's contemporaneous value $X_{j,it}$ has a positive effect on s_{it}, we should expect $X_{j,it+1}$ to affect s_{it} positively. To see if this is the case, we can add $X_{j,it+1}$ to the equation for s_{it} and observe the

depend on past incomes (Stutzer 2004). *Present independence* holds in that case. Moreover, the argument presumes constancy of the reporting function. If current incomes cause the reporting function to be more stringent, we may also simultaneously observe constant reported satisfaction and typical reports of the past being worse.

[24] Since expectations are unobserved, I proxy current expectations with future realizations.

estimated coefficient.[25] If this test is positive (negative), we should then expect that $X_{j,it}$ will positively (negatively) impact $s_{it}^{(m)}$, in which case *present independence* is violated. We should normally expect the bias to have the same direction as a variable's effect on contemporaneous satisfaction, i.e. that $\boldsymbol{\lambda}^{(m)}$ and $\boldsymbol{\beta}$ have the same signs.

Taking an overview, the arguments made on the basis of e.g. Conway & Ross (1986) and Easterlin (2001) predict opposing signs for $\boldsymbol{\lambda}^{(m)}$ and $\boldsymbol{\beta}$. The arguments based on mood-congruent reporting and anticipation effects predict equal signs for $\boldsymbol{\lambda}^{(m)}$ and $\boldsymbol{\beta}$.

Rewriting equation (10) as follows helps us to assess the implications of either bias:[26]

$$\begin{aligned} \ell_{it} = \ln \sum_{k=1}^{K} \mathrm{I}(r_{it} = k) \times \\ \left[\Phi\left(\gamma_k + \boldsymbol{X}_{it}(\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}) - \boldsymbol{X}_{it}(\boldsymbol{\beta} - \boldsymbol{\lambda}^{(m)})\right) - \Phi\left(\gamma_{k-1} + \boldsymbol{X}_{it}(\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}) - \boldsymbol{X}_{it}(\boldsymbol{\beta} - \boldsymbol{\lambda}^{(m)})\right)\right] \\ + \sum_{k=1}^{K} \mathrm{I}\left(r_{it}^{(m)} = k\right) \times \\ \left[\Phi\left(\gamma_k + \boldsymbol{X}_{it}(\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}) - \theta^{(m)} - \boldsymbol{X}_{it-1}\boldsymbol{\alpha}\right) - \Phi\left(\gamma_{k-1} + \boldsymbol{X}_{it}(\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}) - \theta^{(m)} - \boldsymbol{X}_{it-1}\boldsymbol{\alpha}\right)\right] \end{aligned} \quad (12)$$

We see that only the parameter differences $\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}$ and $\boldsymbol{\beta} - \boldsymbol{\lambda}^{(m)}$ are identified. If *present independence* for a particular variable X_j is satisfied, $\lambda_j^{(m)} = 0$ holds. In that case, $\delta_j - \boldsymbol{\lambda}_j^{(m)} = \delta_j$ and $\beta_j - \boldsymbol{\lambda}_j^{(m)} = \beta_j$. We can then correctly interpret these respective estimates as indicating effects on scale use and latent satisfaction. However, if $\lambda_j^{(m)}$ and β_j have opposite signs, then $|\beta_j - \boldsymbol{\lambda}_j^{(m)}| > |\beta_j|$. Estimates are then biased away from zero. In contrast, if the signs of $\lambda_j^{(m)}$ and β_j are equal and $|\lambda_j^{(m)}| < |\beta_j|$, we find that $|\beta_j - \boldsymbol{\lambda}_j^{(m)}| < |\beta_j|$. Hence, in that case, estimate $\widehat{(\beta_j - \boldsymbol{\lambda}_j^{(m)})}$, when interpreted as an estimate of β_j, is biased toward zero. Analogous analyses apply to the relationship between δ_j and $\lambda_j^{(m)}$. The results presented below can therefore be interpreted in two ways. More

[25] However, note that such coefficients may be overestimated via reverse causality of higher current satisfaction causing higher future values. See e.g. Vendrik (2013) or De Neve and Oswald (2012) for such endogeneity in the case of incomes.

[26] The argument to follow would be analogous for the case in which equation (11) applies and is therefore omitted. Likewise, the following discussion also applies to the vignettes approach.

positive (negative) estimates $(\widehat{\beta_J - \boldsymbol{\lambda}_J^{(m)}})$ are either interpreted as indicating a variable's positive (negative) effect on latent satisfaction or as indicating a variable's negative (positive) effect on memories of past satisfaction.

The possible biases described by Conway & Ross (1986) *versus* those arising from anticipation and mood-congruency may partially offset each other. Moreover, the emotional intensity of memories tends to fade with time (e.g. Levine and Bluck 2004). It therefore seems implausible that the absolute magnitude of the present-induced memory bias $\boldsymbol{\lambda}^{(m)}$ should exceed the effect $\boldsymbol{\beta}$ on latent satisfaction.

We may therefore assume $|\lambda_j^{(m)}| \leq |\beta_j|$ for any particular variables X_j. Further assuming that $\lambda_j^{(m)}$ and β_j have opposite signs (which is the more worrying bias), we find that $|\lambda_j^{(m)}| \leq 0.5|(\beta_j - \lambda_j^{(m)})|$. On the basis of this bound on $\lambda_j^{(m)}$ we can assess whether the implied scale shift has a determinate direction. We may do so by inspecting the sign of $(\widehat{\delta_J - \lambda_J^{(m)}}) - 0.5(\widehat{\beta_J - \lambda_J^{(m)}})$. We can also assess whether the effects on latent satisfaction and scale use are statistically significant. This is done by respectively testing $h_0: 0.5(\widehat{\beta_J - \boldsymbol{\lambda}_J^{(m)}}) = 0$ and $h_0: (\widehat{\delta_J - \lambda_J^{(m)}}) - 0.5(\widehat{\beta_J - \lambda_J^{(m)}}) = 0$.[27]

3.4 Interpretation of lags and leads

I now turn to the case in which latent satisfaction also depends on lagged and leaded values of $\boldsymbol{X}$. For simplicity I only consider 1-year lags and leads, which is a special case of equation (4):

$$s_{it} = \boldsymbol{X}_{it-1}\boldsymbol{\beta}_{-1} + \boldsymbol{X}_{it}\boldsymbol{\beta}_0 + \boldsymbol{X}_{it+1}\boldsymbol{\beta}_{+1} + \varepsilon_{it} \tag{13}$$

[27] Assuming that $\lambda_j^{(m)}$ and β_j have opposite signs, and assuming that $|\lambda_j^{(m)}| \leq |\beta_j|$, yields a worst case of $\beta_j = -\lambda_j^{(m)}$. In that worst case $(\beta_j - \lambda_j^{(m)}) = 2\beta_j \leftrightarrow 0.5(\beta_j - \lambda_j^{(m)}) = \beta_j$. Moreover, in that case, $(\delta_j - \lambda_j^{(m)}) = (\delta_j + \beta_j)$. Given these identities, $(\delta_j - \lambda_j^{(m)}) - 0.5(\beta_j - \lambda_j^{(m)}) = (\delta_j + \beta_j) - \beta_j = \delta_j$. Additionally note that when we assume $\boldsymbol{\delta} = \mathbf{0}$, i.e. that no scale shifts occur, estimates $(\widehat{\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}})$ are interpreted as estimates of $-\boldsymbol{\lambda}^{(m)}$, implying that the estimated difference $(\widehat{\boldsymbol{\beta} - \boldsymbol{\lambda}^{(m)}}) - (\widehat{\boldsymbol{\delta} - \boldsymbol{\lambda}^{(m)}})$ are to be interpreted as estimates of $\boldsymbol{\beta}$. Thus, the model of equation (12) contains an analysis without attempting to account for scale use as a special case.

Further lags and leads make the notation more arduous but present no further problem. Given the arguments of section 2.2, we may expect that scale use depends on the same lags and leads:

$$\tau_{it,k} = \gamma_k + \boldsymbol{X}_{it-1}\boldsymbol{\delta}_{-1} + \boldsymbol{X}_{it}\boldsymbol{\delta}_0 + \boldsymbol{X}_{it+1}\boldsymbol{\delta}_{+1} \tag{14}$$

If the equation for latent memories is to have the same structure as the equation for latent contemporaneous satisfaction, then equation (9) must be extended, too:

$$s_{it}^{(m)} = \theta^{(m)} + \boldsymbol{X}_{it-2}\boldsymbol{\alpha}_{-2} + \boldsymbol{X}_{it-1}\boldsymbol{\alpha}_{-1} + \boldsymbol{X}_{it}\boldsymbol{\alpha}_0 + \varepsilon_{it}^{(m)} \tag{15}$$

Note that $\boldsymbol{\alpha_0}$ will contain the effects of memory biases ($\boldsymbol{\lambda}^{(m)}$) discussed in the previous section.

The extended expression for the contributions to the log-likelihood can then be written as:

$$\begin{gathered}
\ell_{it} = \ln \sum_{k=1}^{K} \mathrm{I}(r_{it} = k) \times \\
[\Phi(\gamma_k + \boldsymbol{X}_{it+1}\boldsymbol{\delta}_{+1} + \boldsymbol{X}_{it}(\boldsymbol{\delta}_0 - \boldsymbol{\alpha}_0) - \boldsymbol{X}_{it+1}\boldsymbol{\beta}_{+1} - \boldsymbol{X}_{it}(\boldsymbol{\beta}_0 - \boldsymbol{\alpha}_0) - \boldsymbol{X}_{it-1}(\boldsymbol{\beta}_{-1} - \boldsymbol{\delta}_{-1})) \\
-\Phi(\gamma_{k-1} + \boldsymbol{X}_{it+1}\boldsymbol{\delta}_{+1} + \boldsymbol{X}_{it}(\boldsymbol{\delta}_0 - \boldsymbol{\alpha}_0) - \boldsymbol{X}_{it+1}\boldsymbol{\beta}_{+1} - \boldsymbol{X}_{it}(\boldsymbol{\beta}_0 - \boldsymbol{\alpha}_0) - \boldsymbol{X}_{it-1}(\boldsymbol{\beta}_{-1} - \boldsymbol{\delta}_{-1}))] \\
+ \sum_{k=1}^{K} \mathrm{I}\left(r_{it}^{(m)} = k\right) \times \\
\left[\Phi\left(\gamma_k + \boldsymbol{X}_{it+1}\boldsymbol{\delta}_{+1} + \boldsymbol{X}_{it}(\boldsymbol{\delta}_0 - \boldsymbol{\alpha}_0) - \theta^{(m)} - \boldsymbol{X}_{it-1}(\boldsymbol{\alpha}_{-1} - \boldsymbol{\delta}_{-1}) - \boldsymbol{X}_{it-2}\boldsymbol{\alpha}_{-2}\right)\right. \\
\left.-\Phi\left(\gamma_{k-1} + \boldsymbol{X}_{it+1}\boldsymbol{\delta}_{+1} + \boldsymbol{X}_{it}(\boldsymbol{\delta}_0 - \boldsymbol{\alpha}_0) - \theta^{(m)} - \boldsymbol{X}_{it-1}(\boldsymbol{\alpha}_{-1} - \boldsymbol{\delta}_{-1}) - \boldsymbol{X}_{it-2}\boldsymbol{\alpha}_{-2}\right)\right]
\end{gathered} \tag{16}$$

Only the leaded effects on latent satisfaction ($\boldsymbol{\beta_{+1}}$) and scale use ($\boldsymbol{\delta_{+1}}$), as well as the two-year lagged effect on memories of satisfaction ($\boldsymbol{\alpha_{-2}}$) are now uniquely identified. In contrast, estimates of the contemporaneous effects on latent satisfaction ($\boldsymbol{\beta_0}$) and scale use ($\boldsymbol{\delta_0}$) are contaminated by the effects on memories ($\boldsymbol{\alpha_0}$).

Fortunately, as also touched on in the previous section, since $\boldsymbol{\beta_{+1}}$ is identified (barring reverse causality), we can get some indication for the likely signs of $\boldsymbol{\alpha_0}$. This may aid us forming beliefs about the likely signs and magnitudes of $\boldsymbol{\delta_0}$. Moreover, only the difference of lagged effects on latent satisfaction and scale use ($\boldsymbol{\beta_{-1}} - \boldsymbol{\delta_{-1}}$) is identified. Given the arguments of section 2.2, $\boldsymbol{\beta_{-1}}$ and $\boldsymbol{\delta_{-1}}$ likely have opposing signs in the presence of adaptation, thus leading to an over-estimation of adaptation (or, equivalently, an underestimation of a lagged scale shift effect).

4 Empirical application

Having introduced the memory approach, I now turn to an empirical application using BHPS data. However, before doing so I will note a final difficulty for estimation.

4.1 Accounting for fixed effects

Unfortunately, failure to account for individual fixed effects severely biases coefficients in wellbeing research (Ferrer-i-Carbonell & Frijters, 2004). The standard approach to deal with this issue is to demean both sides of equation (4). However, since s_{it} is not observed in the present case, this is not an option.[28] Alternatively estimating separate parameters for each individual is known to yield inconsistent estimates for finite T (Greene, 2004). To nevertheless correct for this issue, I add individual averages of all explanatory variables to the estimation. Such a specification was proposed by Mundlak (1978).

Van Praag (2015) recently argued that a specification which includes such terms yields estimates of $\boldsymbol{\beta}_*$ that are identical to estimates that would be obtained if it were possible to demean equation (4). Hence, although this is normally thought to be infeasible (Riedl and Geishecker 2014), adequately accounting for individual fixed-effects in an ordered probit context seems possible. Appendix B reconstructs Van Praag's argument. Appendix C successfully evaluates its veracity in a Monte-Carlo study. My preferred estimates therefore include individual averages of all explanatory variables.

Finally, if latent satisfaction is partly determined by an individual fixed effect, then memories of satisfaction and scale use are also likely affected by individual fixed effects. To account for these, averages $\overline{\boldsymbol{X}}_{\boldsymbol{i}}$ are added to these equations, too. Thus, the estimates shown below control for fixed effects in scale use. They are therefore primarily concerned with intrapersonal variation in scale use. This is in contrast to vignette studies, which are always cross-sectional and therefore estimate interpersonal differences in scale use.

[28] In the OLS context, researchers demean r_{it}, which is observed.

4.2 Data

I include waves 6-10 and 12-18 of the BHPS, corresponding to the years 1996-2000 and 2002-2008. These are all the years in which data on (memories of) life satisfaction are available. Life satisfaction is measured by the question given in section 2.1. Memories of satisfaction are elicited with the second question discussed in section 3.2.

I focus on the effects of log post-tax household income[29], employment status, marital status, childbirth, and log household size. I also include a measure of reference income. Reference income is constructed as the log average income of those with similar age (i.e. respondent's age $\pm$ 5 years), and similar education (having the same education on a four category CASMIN scale). I additionally control for log(1+working hours), homeownership, having a temporary job, age(-squared),[30] and wave dummies.

4.3 Descriptive evidence

In table 1, I provide some descriptive evidence on plausible scale shifts. That descriptive analysis follows a recent study of Prati and Senik (2020). They make use of the same data but offer a different interpretation. This makes it useful to review their results. They observe that people more often state that last year was worse than that last year was better (as does Easterlin 2001). They further observe the following 'stylized facts' (quoting pages 8 and 10):

(1) *Most people report an evolution of life satisfaction that is different from the one observed over time in the panel.*

(2) *On average, people tend to overstate the improvement in their life satisfaction over time.*

(3) *More satisfied people tend to overstate the improvement in their life satisfaction; conversely, unsatisfied people tend to understate it.*

[29] This variable is provided by Levy and Jenkins (2012). I additionally apply the modified OECD equivalence scale and deflate to 2005 prices.
[30] When simultaneously accounting for fixed effects and including wave dummies, the linear term for age drops out due to collinearity.

Prati & Senik (2020) interpret these findings as driven by memory biases. Such an interpretation assumes that scale use is stable over time and that memory biases depend on the level of reported satisfaction. An alternative interpretation of (2) and (3) is to assume that scale use is not common, and that people who report relatively high life satisfaction partly do so because they use the response scale less stringently. Section 2.2 argued for the former assumption. The latter assumption may be plausible given that, *ceteris paribus*, those with less stringent scale use are more likely to report a relatively high level of satisfaction. Both interpretations may be partly true. The present paper takes the latter interpretation. Unfortunately, the data at hand does not allow us to take an ultimate stance on this.[31]

Panel A of table 1 confirms the first of Prati & Senik's (2020) observations. Although the majority state that life satisfaction did not increase (56%), more respondents state that last year was worse (28%), than that last year was better (15.3%). The higher reported satisfaction, the higher the probability of stating that last year was worse. This is intuitive: people with higher latent satisfaction are, *ceteris paribus*, more likely to state that their life satisfaction has increased. Only 0.8% of respondents state that they 'don't know' whether last year was better or worse. Thus, respondents generally feel able to recall their past wellbeing.

Panel B cross-tabulates the observed change in r_{it} against respondents' remembered change. Cases in which a decrease in reported satisfaction is observed while people state that the past was worse (4.6% of all cases), and those in which an increase in reported satisfaction is paired with stating that the past was better (2.6% of all cases), are evidently inconsistent. The former (latter) case either indicates that scale use has become more (less) stringent or that people misremember the past as worse (better) than it was.[32]

[31] Unless scale use is an entirely automatic process, reporting satisfaction on a constant scale requires respondents to remember their previous scale use. In that sense, the assumption of *common scale use* also involves an absence of memory biases. In particular, respondent's memories of last year's scale use may not depend on current values of covariates.
[32] I will now cease noting the possibility of memory biases.

Table 1. Descriptive evidence on possible scale shifts

	'Last year was:' ($r_{it}^{(m)}$)			
Panel A: *Memories of life satisfaction by level of reported life satisfaction*				
r_{it}	'Worse'	'About the same'	'Better'	'Don't know'
1	116 (5.6%)	506 (24.2%)	1,371 (65.6%)	96 (4.6%)
2	197 (6.5%)	758 (24.8%)	2,031 (66.5%)	68 (2.2%)
3	688 (8.1%)	3,058 (36.0%)	4,587 (54.1%)	153 (1.8%)
4	2,758 (14.2%)	10,682 (54.8%)	5,761 (29.6%)	293 (1.5%)
5	11,183 (27.4%)	23,940 (58.7%)	5,359 (13.1%)	297 (0.7%)
6	17,090 (37.7%)	26,315 (58.1%)	1,745 (3.85%)	171 (0.4%)
7	6,724 (33.8%)	12,666 (63.6%)	418 (2.10%)	94 (0.5%)
Total	38,756 (27.9%)	77,925 (56.0%)	21,272 (15.3%)	1,172 (0.8%)
Panel B: *Memories of life satisfaction by direction of change in reported life satisfaction*				
Δr_{it}	'Worse'	'About the same'	'Better'	'Don't know'
Decrease	**4,620 (16.3%)**	14869 (54.0%)	7792 (28.3%)	277 (1.0%)
Steady	**12,649 (27.3%)**	28571 (61.6%)	**4,924 (10.6%)**	233 (0.5%)
Increase	8,848 (34.2%)	14,296 (55.2%)	**2,613 (10.1%)**	151 (0.6%)
Panel C: *Possible shifts in scale use by levels of reported life satisfaction*				
r_{it}	Less stringent	No change	More stringent	
1	299 (21.8%)	1,007 (73.5%)	65 (4.7%)	
2	406 (19.0%)	1,626 (76.0%)	109 (5.1%)	
3	1,165 (19.3%)	4,510 (74.9%)	350 (5.8%)	
4	1,847 (13.3%)	10,680 (77.0%)	1,340 (9.7%)	
5	2,433 (8.2%)	21,608 (72.7%)	5,684 (19.1%)	
6	1,114 (3.4%)	24,476 (74.1%)	7,462 (22.6%)	
7	273 (2.0%)	11,185 (81.5%)	2,259 (16.5%)	
Total	7,537 (7.5%)	75,092 (75.2%)	17,269 (17.3%)	

Note: Bold cells indicate possible shifts in scale use. Observations are counted as 'less stringent' when respondents' reported satisfaction either increased or remained steady, but respondents reported that life last year was better. Observations are counted as 'more stringent' when respondents' reported satisfaction either decreased or remained steady, but respondents reported that life last year was worse. Case numbers for panels B and C are lower because first differences are used, thus dropping waves 6 and 12.

Panel C shows possible shifts in scale use by levels of reported satisfaction. Shifts are counted as possible when people either report a change in satisfaction, but none is observed, or when remembered and observed changes are inconsistent. These definitions follow the mechanics of the model proposed in section 3 (which ignores information on respondents stating 'about same'). Using these definitions, 24.8% of all cases record a possible change in scale use. The higher one's level of reported satisfaction, the more likely it is that scale use has become more stringent.

Although suggestive, the cross-tabulations of table 1 do not directly show that variables which are associated with higher reported satisfaction are also associated with changes towards more stringent scale use.

Therefore, table 2 shows two multinomial logistic regressions of whether scale use has become more or less stringent. Coefficients are exponentiated, thereby showing odds ratios resulting from variation in explanatory variables.

Odds ratios above (below) 1 indicate increased (decreased) odds. Results for model (1) indicate that variables known to be associated with higher reported satisfaction are also associated with higher odds of shifting to more stringent scale use. Results for model (2) indicate that even after controlling for observables, higher levels of reported satisfaction continue to be associated with more stringent scale use. Prati & Senik (2020) obtain equivalent results. While generally indicative of response shifts, these results do not show by how much scale use is impacted. They also do not show how coefficients on latent satisfaction are affected. Therefore I now turn to an analysis using the model described in section 3.2.

Table 2. Multinomial logistic regressions of changes in stringency of scale use (exponentiated coefficients)

	(1)		(2)	
	Without levels of reported satisfaction		With levels of reported satisfaction	
	Less stringent	More stringent	Less stringent	More stringent
ln(HH income)	0.945** (0.024)	1.175*** (0.026)	1.000 (0.027)	1.131*** (0.024)
Employed (base)				
Self-employed	1.044 (0.089)	1.171*** (0.070)	1.087 (0.093)	1.125** (0.067)
Unemployed	1.683*** (0.155)	0.702*** (0.052)	1.405*** (0.132)	0.790*** (0.058)
Retired	1.039 (0.085)	1.258*** (0.091)	1.101 (0.092)	1.227*** (0.088)
Family care	1.107 (0.089)	0.987 (0.057)	1.067 (0.087)	1.012 (0.058)
Lt. sick/disabled	2.477*** (0.205)	0.559*** (0.047)	1.662*** (0.138)	0.756*** (0.063)
In education	0.994 (0.094)	1.177*** (0.063)	1.062 (0.101)	1.131** (0.061)
Married (base)				
Widowed	1.413*** (0.097)	0.995 (0.070)	1.154** (0.076)	1.094 (0.076)
Divorced	1.526*** (0.081)	1.049 (0.049)	1.183*** (0.060)	1.208*** (0.056)
Never married	1.086 (0.056)	0.773*** (0.026)	0.950 (0.047)	0.830*** (0.028)
ln(# adults)	1.118*** (0.043)	0.790*** (0.021)	1.063* (0.039)	0.800*** (0.021)
Childbirth	0.858* (0.073)	1.706*** (0.071)	0.916 (0.078)	1.671*** (0.070)
Reference income	1.708*** (0.144)	2.211*** (0.144)	1.584*** (0.126)	2.160*** (0.139)
r_{it}=1 (base)				
r_{it} =2			0.878 (0.088)	0.833 (0.143)
r_{it} =3			0.941 (0.082)	0.926 (0.137)
r_{it} =4			0.652*** (0.055)	1.475*** (0.206)
r_{it} =5			0.431*** (0.036)	2.952*** (0.406)
r_{it} =6			0.173*** (0.015)	3.517*** (0.484)
r_{it} =7			0.093*** (0.010)	3.405*** (0.476)
Observations	99,260		99,260	

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01. These p values hold for t tests of the odds ratios being equal to one.

4.4 Baseline results

Table 3 presents my baseline results. These results do not consider dynamics in remembered and reported satisfaction via further lags and leads of explanatory variables. Section 4.6 considers such models. Columns (1) and (2) give results without individual averages of explanatory variables. Columns (3) and (4) include such averages and thereby account for individual fixed effects. Column (2) and (4) account for possible shifts in scale use.

Being unemployed or having a long-term sickness reduces satisfaction, compared to being employed. Being in education has a small positive effect. The self-employed, those retired and those who are in family care are not estimated to have a different level of satisfaction compared to the employed. In line with previous literature, accounting for fixed effects reduces the size of estimated coefficients (compare columns 1 and 3).[33] However, upon accounting for scale shifts (see columns 2 and 4), coefficients again become larger. Of those relating to job status, the coefficient on unemployment is increased the most. Thus, while a standard analysis would show that the effect of long-term sickness is stronger than that of unemployment, adjusting for scale use yields the opposite result. These increases in magnitudes of the positive (negative) $\boldsymbol{\beta}$ estimates[34] are driven by corresponding positive (negative) $\boldsymbol{\delta}$ estimates relating to scale use. For example, the unemployed are estimated to use the response scale much less stringently. In contrast, those in retirement use the scale much more stringently.

Higher incomes are associated with higher life satisfaction. The size of the effect is reduced when accounting for fixed effects. When additionally accounting for changes in scale use, the income effect is again increased. Hence respondents' scale use is more stringent with higher incomes. Regarding reference income, the association with scale use is positive. This is in line with the

[33] Differences between cut-off constants γ_k are almost identical across all models. Thus, the sizes of coefficients are directly comparable.

[34] To ease readability, when writing variants of 'estimates of $\boldsymbol{\beta}$', I implicitly assume that $\boldsymbol{\lambda^{(m)} = 0}$. It would be more accurate, but more cumbersome, to always write 'estimates of $\boldsymbol{\beta - \lambda^{(m)}}$'. The empirical implications if $\boldsymbol{\lambda^{(m)}}$ were not equal to $\mathbf{0}$ are discussed in section 4.5

theoretical expectation of section 2.2. In the case of column (2) this association is quite extreme, causing the coefficient on latent satisfaction to switch signs. However, this implausible result is not found in the preferred specification of column (4), where I account for fixed effects. Although the coefficient on scale use is small there, controlling for scale shifts nevertheless causes the coefficient on latent satisfaction to become insignificant.

Those who are neither married nor in a stable partnership are much less satisfied. Comparing columns (1) and (3), these estimates are generally reduced when accounting for fixed effects, though less so than what was observed for employment statuses. Upon accounting for changes in scale use, the negative effects of not being married are substantially increased. This change is particularly extreme for widowhood, where the associated coefficient is tripled in size, and becomes the largest coefficient in the model. This is also reflected by widowhood having the largest coefficient in the scale shift equation. Hence widowhood may have a particularly profound impact on what respondents mean with their answers. Such behaviour follows the example of *Mary* in section 2.2.

Finally, scale use becomes significantly more stringent following childbirth and less stringent in larger households. The respectively positive and negative effects of these variables on satisfaction are consequently magnified, too.

As noted in section 2.3, trade-offs between variables may be affected by scale shifts. Table 4 therefore records the estimated ratios $\beta_j / \beta_{ln(HH\ inc.)}$ for all explanatory variables j of interest. This exercise shows more clearly that, relative to the effect of income, the effects of unemployment, family care and widowhood are somewhat larger, while the effects of long-term sickness, education and divorce are smaller when accounting for scale shifts.

Table 3. Main Results for BHPS data

	Without averages	Without averages, with scale shift	With averages	With averages, with scale shift
	(1)	(2)	(3)	(4)
Panel A: Current life satisfaction				
ln(HH income)	0.089*** (0.010)	0.127*** (0.020)	0.037*** (0.007)	0.097*** (0.023)
Employed (base)				
Self-employed	0.043 (0.027)	0.092 (0.057)	0.022 (0.021)	0.072 (0.061)
Unemployed	-0.316*** (0.033)	-0.709*** (0.063)	-0.197*** (0.025)	-0.615*** (0.069)
Retired	0.120*** (0.030)	0.137** (0.068)	0.070*** (0.022)	0.192*** (0.073)
Family care	-0.026 (0.027)	-0.107** (0.052)	-0.002 (0.020)	-0.048 (0.055)
Lt. sick / disabled	-0.799*** (0.036)	-1.104*** (0.072)	-0.319*** (0.030)	-0.542*** (0.082)
In education	0.130*** (0.025)	0.121** (0.055)	0.095*** (0.022)	0.056 (0.063)
Married (base)				
Widowed	-0.317*** (0.032)	-1.112*** (0.113)	-0.336*** (0.044)	-1.007*** (0.137)
Divorced	-0.440*** (0.025)	-0.592*** (0.058)	-0.262*** (0.028)	-0.369*** (0.069)
Never married	-0.203*** (0.019)	-0.417*** (0.053)	-0.114*** (0.021)	-0.287*** (0.064)
ln(HH size)	-0.086*** (0.014)	-0.199*** (0.042)	-0.087*** (0.015)	-0.216*** (0.048)
Childbirth	0.143*** (0.018)	0.351*** (0.041)	0.081*** (0.015)	0.224*** (0.044)
Reference income	-0.271*** (0.035)	0.149 (0.148)	-0.254*** (0.055)	-0.220 (0.187)
Panel B: Memory of last year's life satisfaction				
ln(HH income)	0.005 (0.016)	0.034* (0.017)	0.013 (0.020)	0.027 (0.020)
Employed (base)				
Self-employed	-0.013 (0.045)	0.046 (0.051)	-0.003 (0.057)	0.010 (0.058)
Unemployed	-0.084 (0.053)	-0.198*** (0.054)	-0.132** (0.062)	-0.135** (0.063)
Retired	0.016 (0.057)	0.032 (0.065)	-0.124* (0.068)	-0.081 (0.073)
Family care	-0.060 (0.045)	-0.100** (0.047)	-0.106** (0.051)	-0.115** (0.053)
Lt. sick / disabled	-0.050 (0.058)	-0.249*** (0.065)	-0.271*** (0.076)	-0.314*** (0.078)
In education	0.008 (0.044)	-0.006 (0.048)	0.042 (0.054)	-0.008 (0.056)
Married (base)				
Widowed	-0.244*** (0.057)	-0.969*** (0.113)	-0.627*** (0.109)	-0.968*** (0.135)
Divorced	-0.313*** (0.034)	-0.466*** (0.055)	-0.328*** (0.054)	-0.419*** (0.063)
Never married	-0.105*** (0.027)	-0.309*** (0.050)	-0.180*** (0.048)	-0.285*** (0.057)
ln(HH size)	0.108*** (0.021)	0.007 (0.039)	0.081** (0.035)	-0.008 (0.046)
Childbirth	0.042 (0.036)	0.047 (0.036)	0.109*** (0.038)	0.093** (0.039)
Reference income	-0.316*** (0.048)	0.134 (0.145)	0.165 (0.144)	0.216 (0.173)
Constant	-0.884* (0.495)	-1.020** (0.510)	-1.086 (0.675)	-0.919 (0.782)
Panel C: Scale shift				
ln(HH income)		0.037** (0.018)		0.059*** (0.021)
Employed (base)				
Self-employed		0.049 (0.050)		0.051 (0.056)
Unemployed		-0.393*** (0.056)		-0.418*** (0.065)
Retired		0.017 (0.062)		0.122* (0.070)
Family care		-0.080* (0.046)		-0.047 (0.051)
Lt. sick / disabled		-0.306*** (0.064)		-0.223*** (0.077)
In education		-0.009 (0.050)		-0.039 (0.058)
Married (base)				
Widowed		-0.795*** (0.110)		-0.671*** (0.132)
Divorced		-0.152*** (0.053)		-0.107* (0.060)
Never married		-0.214*** (0.050)		-0.173*** (0.060)
ln(HH size)		-0.113*** (0.040)		-0.129*** (0.047)
Childbirth		0.208*** (0.038)		0.142*** (0.042)
Reference inc.		0.420*** (0.145)		0.034 (0.179)
Panel D: Cutoff constants				
γ_1	-4.794*** (0.310)	-4.791*** (0.310)	-4.828*** (0.350)	-4.827*** (0.350)
γ_2	-4.382*** (0.025)	-4.379 *** (0.025)	-4.413*** (0.025)	-4.412*** (0.025)
γ_3	-3.837*** (0.014)	-3.835*** (0.014)	-3.865*** (0.015)	-3.865*** (0.015)
γ_4	-3.200*** (0.010)	-3.199*** (0.010)	-3.224*** (0.010)	-3.224*** (0.010)
γ_5	-2.329*** (0.007)	-2.329*** (0.007)	-2.348*** (0.007)	-2.348*** (0.007)
γ_6	-1.283*** (0.007)	-1.283*** (0.007)	-1.297*** (0.007)	-1.297*** (0.007)
Obs. (Respondents)	118, 440 (20,572)	118, 440 (20,572)	118,440 (20,572)	118,440 (20,572)

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 4. Ratios of coefficients on each explanatory variable against coefficient on income

	Without averages		With averages	
	No shift	With shift	No shift	With shift
ln(HH income)	1.000	1.000	1.000	1.000
Employed (base)				
Self-employed	0.480	0.723	0.579	0.750
Unemployed	-3.539	-5.607	-5.291	-6.371
Retired	1.347	1.082	1.877	1.983
Family care	-0.296	-0.843	-0.048	-0.501
Lt. sick / disabled	-8.940	-8.727	-8.565	-5.613
In education	1.460	0.959	2.541	0.576
Married (base)				
Widowed	-3.549	-8.787	-9.004	-10.423
Divorced	-4.925	-4.680	-7.035	-3.826
Never married	-2.276	-3.299	-3.054	-2.970
ln(HH size)	-0.965	-1.576	-2.331	-2.234
Childbirth	1.604	2.778	2.178	2.315
Reference income	-3.038	1.177	-6.818	-2.280

Note: Ratios of coefficients computed on the basis of table 3.

Overall, my results agree with the theoretical expectations of section 2.2. Personal characteristics that are associated with higher (lower) levels of latent satisfaction are also associated with more (less) stringent scale use. Unsurprisingly, this is especially so for bereavement and unemployment, both of which have particularly profound impacts upon people's lives and their outlook upon life.

4.5 Sensitivity checks against violations of *present independence*

With the data at hand, it is impossible to test the validity of the *present independence* assumption. However, we do obtain estimates of the lagged effects of explanatory variables on memories of satisfaction (parameter α in equation (9)). These are shown in panel B of table 3. Looking at the preferred model in column (4), signs of coefficients on current *versus* remembered satisfaction typically equal.[35] For example, just as being unemployed reduces contemporaneous satisfaction, having been unemployed reduces remembered satisfaction. Thus, estimated signs of coefficients suggest that respondents are generally able to recall the direction in which past personal circumstances affected their past satisfaction. Moreover, almost all coefficients on memories have smaller magnitudes compared to those for current satisfaction.

[35] The cases for which signs differ are retirement, education, and reference income. However, in the case of education, the coefficient on latent satisfaction is close to zero, making it easy to observe sign switches.

Unfortunately, these findings are not direct evidence to support *present independence*. It is therefore important to evaluate how violations of *present independence* could affect my conclusions. If the contemporaneous effects on latent satisfaction have the same sign as potential contemporaneous effects on memory bias, then coefficients will be biased towards zero.[36] The same relation holds with respect to memory biases and effects on scale use. Note that when accounting for differences in scale use, the magnitudes of estimates on latent satisfaction almost always exceed those obtained when failing to account for scale shifts. This implies that if *present independence* is violated in a manner in which $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}^{(m)}$ have the same signs, the resulting biases (which lead to an underestimation of the true $\boldsymbol{\beta}$) are smaller than the biases resulting from failing to account for changes in scale use (which also lead to an underestimation of the true $\boldsymbol{\beta}$).

However, we do not know whether the opposite case, i.e. opposing signs for $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}^{(m)}$ occurs. Given the arguments of section 3.3, which suggested that the signs of $\boldsymbol{\lambda}^{(\boldsymbol{m})}$ are indeterminate, it seems implausible that the absolute magnitudes of $\boldsymbol{\lambda}^{(\boldsymbol{m})}$ could exceed those of $\boldsymbol{\beta}$. Thus, adopting the assumption that the magnitudes of $\boldsymbol{\lambda}^{(\boldsymbol{m})}$ are at most equal to those of $\boldsymbol{\beta}$, we can test whether effects of covariates on scale use, i.e. $\boldsymbol{\delta}$, remain significant and have a determinate sign in such a worst-case. As explained in section 3.3, this can be done by testing $h_0: (\widehat{\delta_J - \lambda_J^{(m)}}) - 0.5(\widehat{\beta_J - \lambda_J^{(m)}}) = 0$ for every variable of interest. In words, we test whether the estimated effect on the reporting function is significantly larger (or smaller) that half the estimated effect on latent satisfaction. Table 5 presents results from such an exercise on the basis of the estimates of column (4) of table 3.

[36] In an extreme case where both effects have the same sign, but the memory bias effect has a larger magnitude, it is possible to obtain inverted signs of coefficients. Since all coefficients have the expected sign, this does not appear to be the case here.

Table 5. Robustness of scale shift effects

	Worst-case estimate	Standard error	p-value
ln(HH income)	0.011	0.011	0.332
Employed (base)			
Self-employed	0.015	0.029	0.615
Unemployed	-0.111***	0.035	0.002
Retired	0.026	0.037	0.489
Family care	-0.022	0.028	0.415
Lt. sick / disabled	0.048	0.042	0.249
In education	-0.067**	0.031	0.030
Married (base)			
Widowed	-0.168**	0.070	0.017
Divorced	0.077**	0.032	0.014
Never married	-0.030	0.032	0.353
ln(HH size)	-0.021	0.025	0.391
Childbirth	0.031	0.023	0.177
Reference income	0.144	0.094	0.126

Note: Test computed under assumption that $\lambda_j^{(m)}$ is not larger than β_j and of opposite sign. * p<0.1, ** p<0.05, *** p<0.01.

The signs of most estimated scale shift effects are robust to this exercise. Own income, self-employment, being retired, and childbirth continue to be associated with more stringent scale use. Unemployment, family care, widowhood, never being married, and household size remain associated with less stringent scale use. However, only the estimates for widowhood and unemployment remain statistically significant. Moreover, contrary to the baseline assumption of $\boldsymbol{\lambda}^{(m)} = \mathbf{0}$, disability and divorce are now seen to be associated with more stringent scale use.[37] Finally note that assuming $|\boldsymbol{\lambda}^{(m)}| = |\boldsymbol{\beta}|$ (and opposing signs) causes all coefficients on latent satisfaction to be halved. Table E1 in Appendix E shows that the statistical significance of each relevant coefficient is unaffected by this reduction in estimated magnitudes.

4.6 Comparison to standard ordered probit and OLS models

The models of table 3 are unusual. It is therefore instructive to compare them to standard ordered probit and OLS models. Results from these are shown in table 6. Recall that in all columns of table 3, the effects of explanatory variables on memories of satisfaction and current satisfaction are estimated by assuming that the same set of thresholds is used for memories as for current satisfaction. If that assumption were violated, we should expect the standard ordered probit results

[37] In the cases of reference income and being in education, $\widehat{\delta_j}$ and $\widehat{\beta_j}$ have opposing signs. This leads to an exacerbation of the estimated scale use effect in this test.

given in columns (1) and (2) of table 6 to strongly differ from those in columns (1) and (3) of table 3. However, signs and magnitudes of estimated coefficients are similar across the models, and the estimated cutoff locations are in agreement, too. These results support the assumption of *response consistency*.

Furthermore, signs and significance levels of the OLS estimates shown in columns (5) and (6) - the latter of which account for fixed effects - match the corresponding results of the ordered probit models. Ratios of coefficients are also in agreement between the ordered probit and OLS models, though they do not match perfectly. One example is the ratio of the income to unemployment coefficients. This ratio is $0.036/-0.193 = -0.187$ in the ordered model with individual averages ("OP-Avg."), but $0.042/-0.278 = -0.151$ in the OLS-FE model. Part of the discrepancy is driven by the ordered probit model allowing cutoffs to not be equidistant, while the OLS model assumes constant differences in expected s_{it} across response categories. To verify this, I constrained cutoffs to be equidistant in the ordered probit models of columns (3) and (4). When doing so, the estimated coefficient ratios are even closer to the corresponding ratios estimated in the OLS models (e.g. for the income to unemployment ratio in column (4), I now obtain $0.036/-0.226 = -0.159$). Thus, in line with the theoretical discussion of Appendix C and the Monte-Carlo results of Appendix D, it indeed appears that adding individual averages to the ordered probit model accurately accounts for fixed effects.

Finally, note that the magnitudes of coefficients in the OLS model are uniformly higher than in the ordered probit models. This is because the OLS model assumes constant differences of 1 in expected s_{it} across response categories. The constrained ordered probit model with $\sigma = 1$ instead estimates differences of 0.796 between thresholds. Hence, coefficients of the OLS model can be expected to be larger by a factor of about $1/0.796 = 1.256$.

4.7 Lags and leads

We may be worried that the estimates of table 3 are affected by lags and leads of explanatory variables. Effects of past and future values of explanatory variables on latent satisfaction and scale use may also be of substantive interest. Table 7 therefore presents such results. To ease comparison, column (1) of table 7 is identical to column (4) of table 3. In column (2), I add a lag to the equation for latent satisfaction. To keep the memory equation structurally identical, a two-year lag is added there. I find that most results are not substantively changed compared to column (1).

However, for each of the variables relating to marital status, I obtain positive lagged effects on latent satisfaction. The corresponding contemporaneous coefficients are all negative (and of larger magnitude). We might interpret this result as implying that deviating from being married is subject to partial adaptation. This interpretation follows Clark and Georgellis (2013). However, the discussion of section 3.4 showed that lagged coefficients in this model (and any other model in the literature) are contaminated by lagged scale shift effects. For each variable relating to marital status, I obtain contemporaneous negative scale shift effects. There may therefore also be negative lagged scale shift effects. In that case, my estimated lagged coefficients on latent satisfaction are biased away from zero, suggesting more adaptation than really occurs. Unfortunately, the data cannot decide on this. Similar dynamics are also observed for self-employment (c.f. Hanglberger and Merz 2015) and being in education. However no lagged dynamics are observed for income, unemployment, retirement, family care, disability, and childbirth.

Column (3) differs from column (1) via the addition of leaded terms. This reduces the magnitudes of the contemporaneous coefficients on latent satisfaction relating to retirement, divorce and being single. Moreover, for all variables relating to marital status, I find large negative and statistically significant leaded coefficients. Hence, a part of the total effect relating to these variables is driven by these leads. A similar dynamic is also observed for childbirth, where the magnitudes of the leaded and contemporaneous coefficients are similar. More generally, the signs of coefficients on latent satisfaction associated with each leaded variable are always the same as the corresponding

signs of coefficients of each contemporaneous variable. As explained in section 3.4, this equality of signs reduces the likelihood that $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}^{(m)}$ have opposite signs.

In line with the theoretical considerations of section 2.2, results on scale use follow those on latent satisfaction. For example, being married in the next year is associated with both more stringent scale use and greater latent satisfaction. Finally, column (4) uses both leads and lags of explanatory variables. This joint model does not lead to substantially different results compared to columns (2) and (3).

4.8 Reference effects

Scale use may be impacted via comparisons to a fixed reference group. However, although the sign of the reference income effect on scale use was positive, it was so far insignificant. In order to investigate this further, I computed a set of additional 'reference' variables. These variables include the percentages of those in the reference group who are: employed, unemployed, retired, disabled, married, divorced, widowed, or had a birth in their household. In addition, I constructed variables measuring mean log household size and mean reported life satisfaction of those in the reference group. When jointly adding these variables into a model without lags and leads and accounting for fixed effects (i.e. akin to column (4) of table 3), almost all of these variables' coefficients on latent satisfaction, memories, and scale use are statistically insignificant and small. See table E2 in the appendix. Mean reported satisfaction in the reference group is the only major exception to this. That variable shows a large positive coefficient on latent satisfaction, which suggests that having a more satisfied reference group makes respondents themselves more satisfied. However, mean reported satisfaction neither affects memories nor scale use.[38]

[38] Mean reported satisfaction might also picks up effects from personal characteristics that are not included in the model. As can be gleaned from columns (2)-(4) of Table E2, these results are robust to the addition of lags and leads. One reason for these surprisingly insignificant results may be collinearity between the reference variables. In table E3, I therefore estimated sparser models in which only reference income and percentages of employed, disabled, widowed, having birth, and mean satisfaction were added. These variables were closest to statistical significance in the full model. However, results were largely the same in these sparser models. I also estimated further models for reference satisfaction in which reference groups where defined depending on personal characteristics. This yielded similar results. See table E4 in the appendix.

Table 6. Comparison to Ordered Probit and OLS models

	OP	OP-Avg.	OP, equidistant cutoffs imposed	OP-Avg., equidistant cutoffs imposed	OLS	OLS-FE
	(1)	(2)	(3)	(4)	(5)	(6)
ln(HH income)	0.087*** (0.010)	0.036*** (0.007)	0.088*** (0.010)	0.036*** (0.007)	0.109*** (0.011)	0.042*** (0.009)
Employed (base)						
Self-employed	0.041 (0.026)	0.021 (0.020)	0.051** (0.026)	0.025 (0.020)	0.070** (0.030)	0.032 (0.024)
Unemployed	-0.309*** (0.032)	-0.193*** (0.025)	-0.343*** (0.033)	-0.226*** (0.026)	-0.420*** (0.040)	-0.278*** (0.032)
Retired	0.119*** (0.030)	0.069*** (0.022)	0.100*** (0.029)	0.058*** (0.022)	0.108*** (0.033)	0.062** (0.026)
Family care	-0.025 (0.026)	-0.001 (0.019)	-0.040 (0.026)	-0.014 (0.020)	-0.053* (0.031)	-0.020 (0.023)
Lt. sick / disabled	-0.782*** (0.036)	-0.313*** (0.029)	-0.874*** (0.038)	-0.368*** (0.032)	-1.066*** (0.046)	-0.450*** (0.038)
In education	0.127*** (0.025)	0.093*** (0.021)	0.123*** (0.025)	0.086*** (0.022)	0.153*** (0.029)	0.101*** (0.026)
Married (base)						
Widowed	-0.310*** (0.032)	-0.330*** (0.043)	-0.308*** (0.031)	-0.323*** (0.044)	-0.362*** (0.035)	-0.368*** (0.049)
Divorced	-0.431*** (0.024)	-0.257*** (0.028)	-0.446*** (0.025)	-0.263*** (0.029)	-0.538*** (0.030)	-0.316*** (0.035)
Never married	-0.199*** (0.019)	-0.111*** (0.021)	-0.199*** (0.019)	-0.111*** (0.021)	-0.240*** (0.022)	-0.136*** (0.025)
ln(HH size)	-0.084*** (0.014)	-0.085*** (0.014)	-0.082*** (0.014)	-0.082*** (0.014)	-0.099*** (0.017)	-0.097*** (0.017)
Childbirth	0.140*** (0.018)	0.080*** (0.014)	0.132*** (0.018)	0.073*** (0.015)	0.156*** (0.021)	0.084*** (0.017)
Reference income	-0.310*** (0.032)	-0.330*** (0.043)	-0.308*** (0.031)	-0.323*** (0.044)	-0.362*** (0.035)	-0.368*** (0.049)
			Cutoffs			
τ_1	-4.745*** (0.304)	-4.780*** (0.344)	-5.077*** (0.301)	-5.108*** (0.340)		
τ_2	-4.326*** (0.024)	-4.358*** (0.024)	-4.286*** (0.005)	-4.312*** (0.005)		
τ_3	-3.789*** (0.014)	-3.817*** (0.014)	-3.496*** (0.005)	-3.516*** (0.005)		
τ_4	-3.172*** (0.010)	-3.196*** (0.010)	-2.705*** (0.005)	-2.720*** (0.005)		
τ_5	-2.334*** (0.007)	-2.352*** (0.007)	-1.915*** (0.005)	-1.924*** (0.005)		
τ_6	-1.294*** (0.007)	-1.307*** (0.007)	-1,124*** (0.005)	-1.127*** (0.005)		
Observations	118,773	118,773	118,773	118,773	118,773	118,773

Note: Standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 7. Addition of Lags and Leads

	No lags no leads (1)	One-year lag in current satisfaction, two-year lags in memories (2)	One-year lead in current satisfaction and one-year lead in scale use (3)	One-year lag and lead in current satisfaction, two-year lags in memories, one-year lead scale use (4)
Current life satisfaction				
Contemporaneous effects				
ln(HH income)	0.097*** (0.023)	0.108*** (0.024)	0.108*** (0.026)	0.113*** (0.028)
Employed (base)				
Self-employed	0.072 (0.061)	0.063 (0.065)	0.020 (0.070)	-0.008 (0.075)
Unemployed	-0.615*** (0.069)	-0.694*** (0.076)	-0.613*** (0.079)	-0.672*** (0.088)
Retired	0.192*** (0.073)	0.184** (0.078)	0.110 (0.086)	0.081 (0.092)
Family care	-0.048 (0.055)	-0.086 (0.060)	-0.075 (0.063)	-0.123* (0.068)
Lt. sick /disabled	-0.542*** (0.082)	-0.578*** (0.090)	-0.601*** (0.093)	-0.620*** (0.102)
In education	0.056 (0.063)	0.006 (0.072)	0.007 (0.075)	-0.058 (0.086)
Married (base)				
Widowed	-1.007*** (0.137)	-1.192*** (0.149)	-0.618*** (0.171)	-0.797*** (0.184)
Divorced	-0.369*** (0.069)	-0.478*** (0.074)	-0.105 (0.083)	-0.229*** (0.089)
Never married	-0.287*** (0.064)	-0.368*** (0.071)	-0.129 (0.082)	-0.227** (0.091)
ln(HH size)	-0.216*** (0.048)	-0.163*** (0.052)	-0.204*** (0.061)	-0.163** (0.066)
Childbirth	0.224*** (0.044)	0.238*** (0.049)	0.256*** (0.050)	0.282*** (0.056)
Reference inc.	-0.220 (0.187)	-0.045 (0.203)	-0.085 (0.230)	0.065 (0.250)
1-year lagged effects				
ln(HH income)		0.009 (0.008)		0.006 (0.009)
Employed (base)				
Self-employed		-0.052** (0.023)		-0.053** (0.025)
Unemployed		-0.001 (0.028)		-0.028 (0.031)
Retired		-0.009 (0.024)		-0.021 (0.027)
Family care		0.001 (0.021)		0.001 (0.023)
Lt. sick /disabled		-0.038 (0.032)		-0.042 (0.036)
In education		0.063** (0.026)		0.071** (0.029)
Married (base)				
Widowed		0.199*** (0.045)		0.247*** (0.049)
Divorced		0.181*** (0.030)		0.192*** (0.033)
Never married		0.128*** (0.025)		0.117*** (0.028)
ln(HH size)		-0.092*** (0.019)		-0.077*** (0.021)
Childbirth		0.019 (0.016)		0.021 (0.018)
Reference inc.		-0.232*** (0.067)		-0.126* (0.073)
1-year leaded effects				
ln(HH income)			0.008 (0.024)	0.001 (0.026)
Employed (base)				
Self-employed			-0.034 (0.070)	-0.016 (0.076)
Unemployed			-0.056 (0.078)	-0.111 (0.087)
Retired			0.073 (0.083)	0.064 (0.088)
Family care			-0.041 (0.064)	-0.007 (0.069)
Lt. sick /disabled			-0.290*** (0.093)	-0.302*** (0.100)
In education			0.045 (0.074)	0.087 (0.085)
Married (base)				
Widowed			-0.674*** (0.152)	-0.685*** (0.166)
Divorced			-0.456*** (0.078)	-0.504*** (0.086)
Never married			-0.281*** (0.077)	-0.249*** (0.085)
ln(HH size)			-0.042 (0.052)	-0.089 (0.057)
Childbirth			0.232*** (0.053)	0.257*** (0.058)
Reference inc.			-0.180 (0.219)	-0.339 (0.240)
Memory of last year's life satisfaction				
1-year lagged effects				
ln(HH income)	0.027 (0.020)	0.029 (0.022)	0.042* (0.023)	0.045* (0.025)
Employed (base)				
Self-employed	0.010 (0.058)	0.018 (0.065)	0.102 (0.064)	0.096 (0.070)
Unemployed	-0.135** (0.063)	-0.118* (0.070)	-0.153** (0.073)	-0.148* (0.082)
Retired	-0.081 (0.073)	-0.113 (0.084)	0.025 (0.083)	-0.014 (0.095)
Family care	-0.115** (0.053)	-0.089 (0.057)	-0.082 (0.059)	-0.086 (0.065)
Lt. sick /disabled	-0.314*** (0.078)	-0.300*** (0.087)	-0.302*** (0.087)	-0.310*** (0.096)
In education	-0.008 (0.056)	-0.025 (0.064)	-0.010 (0.065)	-0.061 (0.074)
Married (base)				
Widowed	-0.968*** (0.135)	-0.556*** (0.152)	-1.047*** (0.167)	-0.683*** (0.185)

Table 7. Addition of Lags and Leads (continued.)

Divorced	-0.419*** (0.063)	-0.352*** (0.075)	-0.401*** (0.073)	-0.362*** (0.084)
Never married	-0.285*** (0.057)	-0.238*** (0.069)	-0.273*** (0.066)	-0.283*** (0.079)
ln(HH size)	-0.008 (0.046)	-0.048 (0.058)	-0.043 (0.052)	-0.090 (0.065)
Childbirth	0.093** (0.039)	0.097** (0.044)	0.075* (0.043)	0.066 (0.050)
Reference inc.	0.216 (0.173)	0.046 (0.208)	0.173 (0.196)	-0.026 (0.232)
		2-year lagged effects		
ln(HH income)		0.052** (0.022)		0.056** (0.025)
Employed (base)				
Self-employed		0.139** (0.062)		0.116* (0.068)
Unemployed		-0.086 (0.067)		-0.074 (0.074)
Retired		0.213** (0.083)		0.179** (0.091)
Family care		-0.017 (0.056)		0.013 (0.063)
Lt. sick /disabled		-0.054 (0.087)		-0.006 (0.096)
In education		0.033 (0.063)		0.038 (0.072)
Married (base)				
Widowed		-0.712*** (0.136)		-0.678*** (0.153)
Divorced		-0.078 (0.066)		-0.101 (0.076)
Never married		-0.043 (0.060)		-0.003 (0.069)
ln(HH size)		0.078 (0.047)		0.027 (0.054)
Childbirth		-0.028 (0.041)		-0.036 (0.045)
Reference inc.		0.211 (0.191)		0.201 (0.208)
Constant	-0.920 (0.782)	-0.937 (1.035)	-1.202 (0.989)	-2.020 (1.296)
		Scale shift		
		contemporaneous effects		
ln(HH income)	0.059*** (0.021)	0.076*** (0.023)	0.071*** (0.025)	0.075*** (0.027)
Employed (base)				
Self-employed	0.051 (0.056)	0.037 (0.060)	0.027 (0.065)	-0.003 (0.070)
Unemployed	-0.418*** (0.065)	-0.471*** (0.071)	-0.399*** (0.075)	-0.445*** (0.083)
Retired	0.122* (0.070)	0.116 (0.075)	0.086 (0.083)	0.043 (0.090)
Family care	-0.047 (0.051)	-0.066 (0.056)	-0.052 (0.059)	-0.091 (0.064)
Lt. sick /disabled	-0.223*** (0.077)	-0.259*** (0.086)	-0.280*** (0.089)	-0.329*** (0.099)
In education	-0.039 (0.058)	-0.062 (0.067)	-0.092 (0.070)	-0.147* (0.080)
Married (base)				
Widowed	-0.671*** (0.132)	-0.703*** (0.143)	-0.482*** (0.167)	-0.511*** (0.177)
Divorced	-0.107* (0.060)	-0.104 (0.064)	0.034 (0.074)	0.042 (0.078)
Never married	-0.173*** (0.060)	-0.175*** (0.067)	-0.083 (0.078)	-0.112 (0.086)
ln(HH size)	-0.129*** (0.047)	-0.127** (0.050)	-0.136** (0.058)	-0.122* (0.063)
Childbirth	0.142*** (0.042)	0.145*** (0.046)	0.153*** (0.048)	0.162*** (0.053)
Reference inc.	0.034 (0.179)	0.076 (0.194)	-0.020 (0.223)	0.040 (0.241)
		1-year leaded effects		
ln(HH income)			0.008 (0.023)	0.000 (0.025)
Employed (base)				
Self-employed			-0.032 (0.067)	-0.019 (0.072)
Unemployed			0.039 (0.075)	0.007 (0.082)
Retired			0.038 (0.082)	0.051 (0.087)
Family care			-0.031 (0.061)	0.015 (0.066)
Lt. sick /disabled			-0.150* (0.088)	-0.171* (0.096)
In education			0.049 (0.071)	0.083 (0.080)
Married (base)				
Widowed			-0.386** (0.151)	-0.367** (0.165)
Divorced			-0.210*** (0.072)	-0.275*** (0.078)
Never married			-0.183** (0.073)	-0.152* (0.082)
ln(HH size)			-0.031 (0.050)	-0.084 (0.054)
Childbirth			0.127** (0.050)	0.157*** (0.055)
Reference inc.			0.085 (0.211)	-0.050 (0.231)
		Cutoff constants		
γ_1	-4.794*** (0.310)	-4.753*** (0.407)	-5.312*** (0.386)	-5.334*** (0.477)
γ_2	-4.382*** (0.025)	-4.338*** (0.027)	-4.898*** (0.027)	-4.920*** (0.030)
γ_3	-3.837*** (0.014)	-3.787*** (0.016)	-4.346*** (0.016)	-4.365*** (0.017)
γ_4	-3.200*** (0.010)	-3.145*** (0.010)	-3.702*** (0.011)	-3.718*** (0.011)
γ_5	-2.329*** (0.007)	-2.263*** (0.008)	-2.826*** (0.008)	-2.836*** (0.008)
γ_6	-1.283** (0.007)	-1.198*** (0.008)	-1.770*** (0.008)	-1.766*** (0.008)
Observations	118,440	102,924	97,939	85,472

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

4.9 Parallel and non-parallel shifts

So far I only considered shifts with equal magnitude and direction for all cutoffs. Xu & Xie (2016), call this a *parallel shift* in the cut-offs. However, it may be that independent variables impact the location of each cut-off differently. Thus, variables may cause scale use to become more stringent at high levels of reported satisfaction, but less stringent at low levels of reported satisfaction. This idea is illustrated in figure 2.

In order to allow for non-parallel shifts, it is common to postulate a nested model for the set of thresholds:[39]

$$\begin{aligned} \tau_{it,1} &= \gamma_1 + \boldsymbol{X}_{it}\boldsymbol{\delta}_1, \\ \tau_{it,2} &= \tau_{it,1} + \exp(\gamma_2 + \boldsymbol{X}_{it}\boldsymbol{\delta}_2), \\ &\dots \\ \tau_{it,K} &= \tau_{it,K-1} + \exp(\gamma_K + \boldsymbol{X}_{it}\boldsymbol{\delta}_K). \end{aligned} \tag{17}$$

This non-linear specification has the advantage that the ordering of cutoffs is always preserved, no matter the signs of $\boldsymbol{\delta}_k$ (Greene and Hensher 2010). However, this specification is inelegant since we assume a non-linear functional form for all but the first cutoff. A linear specification is therefore more natural:

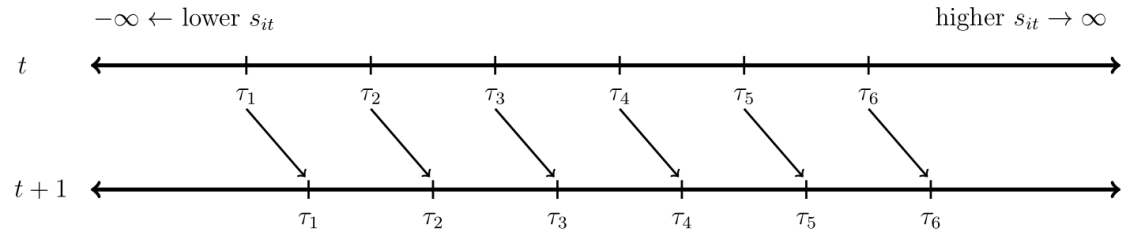
$$\begin{aligned} \tau_{it,1} &= \gamma_1 + \boldsymbol{X}_{it}\boldsymbol{\delta}_1, \\ \tau_{it,2} &= \tau_{it,1} + \gamma_2 + \boldsymbol{X}_{it}\boldsymbol{\delta}_2, \\ &\dots \\ \tau_{it,K} &= \tau_{it,K-1} + \gamma_K + \boldsymbol{X}_{it}\boldsymbol{\delta}_K. \end{aligned} \tag{18}$$

Here $\boldsymbol{\delta}_k$ for $k > 1$ may be interpreted as increasing the distance between adjacent thresholds (when $\boldsymbol{\delta}_k > 0$) or as decreasing the distance between adjacent thresholds (when $\boldsymbol{\delta}_k < 0$). In the former case, scale use is 'stretched', while in the latter case it is 'squeezed'. Note that in the specification of equation (18) it may occur that $\tau_{it,k-1} > \tau_{it,k}$ for some values of $\boldsymbol{X}_{it}$ and some k when not all corresponding elements in $\boldsymbol{\delta}_k$ and $\boldsymbol{\delta}_{k-1}$ have the same sign. Although theoretically

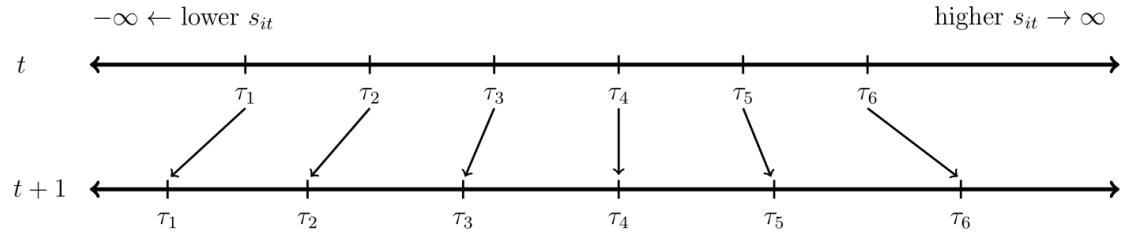
[39] In virtue of this 'nesting' property, this model is sometimes referred to as a 'Hierarchical Ordered Probit', or 'HOPIT' model.

Figure 2. Parallel and non-parallel shifts in scale use.

Panel A. Parallel Shift



Panel B. Non-parallel Shift



undesirable, empirically this does not appear to be the case. Given its greater elegance and ease of interpretation, I therefore prefer the specification of equation (18). Results for such a model are thus presented in table 8.[40] That model accounts for fixed effects. We can therefore compare these results to those of column (4) in table 3.

Results are almost unchanged, including those on scale use (compare the estimates of $\boldsymbol{\delta}$ in table 3 to the corresponding estimates of $\boldsymbol{\delta_1}$ in table 8). Consequently, I find little evidence of non-parallel shifts in scale use. The only clear such cases are observed for various employment statuses. In these, coefficients on the higher thresholds tend to be significantly negative. This may indicate that for those who are not employed, scale use is 'squeezed' at the upper end of the response scale. However, this has little impact on estimates on latent satisfaction.

In both the linear and the nonlinear case, models remain identified when omitting the information from memories, but constraining all elements in $\boldsymbol{\delta_1}$ to 0. Estimates for such models are presented in appendix tables E6 and E7.[41] With respect to latent satisfaction, the results for these models are

[40] Results using the more traditional specification of equation (17) can be found in appendix table E5. These are very similar to those given in table 8.

[41] Variants of such a test are also used by Bond & Lang (2019). See e.g. section A3.2 of their online appendix.

rather similar to those obtained in column (3) of table 3, i.e. when omitting the estimation of scale shift effects. This is because estimates for $\boldsymbol{\beta}$ capture the main scale shift effects contained in $\boldsymbol{\delta_1}$ in this specification. Therefore, these models only capture possible distortions in the 'width' of the reporting scale, but do not capture shifts in the overall location of the scale.

Table 8. Allowing for non-parallel shifts in scale use when adding individual averages (linear specification)

	Current Sat.	Memory of Sat.	Cutoff $\tau_{it,1}$	Cutoff $\tau_{it,2}$	Cutoff $\tau_{it,3}$	Cutoff $\tau_{it,4}$	Cutoff $\tau_{it,5}$	Cutoff $\tau_{it,6}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(HH income)	0.098***	0.028	0.090**	-0.013	-0.032*	0.016	-0.007	0.016
	(0.023)	(0.020)	(0.038)	(0.026)	(0.019)	(0.014)	(0.012)	(0.012)
Employed (base)								
Self-employed	0.080	0.007	0.001	-0.019	0.045	0.005	0.036	0.027
	(0.065)	(0.062)	(0.121)	(0.095)	(0.064)	(0.044)	(0.038)	(0.042)
Unemployed	-0.628***	-0.129**	-0.254**	-0.029	0.045	-0.137***	-0.155***	-0.053
	(0.069)	(0.064)	(0.111)	(0.082)	(0.058)	(0.042)	(0.040)	(0.043)
Retired	0.151**	-0.081	0.115	0.075	-0.034	-0.018	-0.077**	-0.019
	(0.072)	(0.071)	(0.121)	(0.082)	(0.059)	(0.041)	(0.036)	(0.036)
Family care	-0.073	-0.111**	-0.000	0.027	0.018	-0.045	-0.121***	-0.013
	(0.057)	(0.054)	(0.105)	(0.082)	(0.057)	(0.039)	(0.034)	(0.034)
Lt. sick / disabled	-0.549***	-0.298***	-0.131	0.062	-0.027	-0.087**	-0.149***	-0.101**
	(0.079)	(0.075)	(0.120)	(0.081)	(0.057)	(0.044)	(0.043)	(0.047)
In education	0.042	-0.013	0.117	-0.119	-0.023	0.019	-0.103***	0.058
	(0.066)	(0.058)	(0.119)	(0.090)	(0.065)	(0.046)	(0.038)	(0.044)
Married (base)								
Widowed	-0.941***	-0.937***	-0.581***	-0.065	-0.024	0.078	0.003	-0.074
	(0.125)	(0.121)	(0.164)	(0.082)	(0.065)	(0.053)	(0.046)	(0.052)
Divorced	-0.375***	-0.413***	-0.116	0.026	-0.042	0.075*	-0.030	-0.079*
	(0.070)	(0.064)	(0.100)	(0.067)	(0.050)	(0.041)	(0.039)	(0.047)
Never married	-0.296***	-0.289***	-0.100	0.002	-0.095	0.053	-0.004	-0.055
	(0.067)	(0.060)	(0.110)	(0.081)	(0.060)	(0.043)	(0.036)	(0.042)
ln(HH size)	-0.224***	-0.009	-0.106	-0.028	-0.029	0.044	-0.001	-0.033
	(0.051)	(0.047)	(0.079)	(0.054)	(0.038)	(0.028)	(0.024)	(0.027)
Childbirth	0.223***	0.087**	0.228***	-0.018	-0.017	-0.030	-0.055**	0.022
	(0.046)	(0.041)	(0.086)	(0.067)	(0.046)	(0.034)	(0.026)	(0.028)
Reference inc.	-0.200	0.241	0.188	0.021	-0.077	-0.057	-0.120	0.030
	(0.189)	(0.176)	(0.305)	(0.207)	(0.133)	(0.096)	(0.080)	(0.089)
Constant		-0.500	-2.356***	0.539***	0.532***	0.706***	0.869***	1.018***
		(0.547)	(0.099)	(0.063)	(0.042)	(0.033)	(0.030)	(0.039)
Observations	118,440							

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Moreover, the interpretation of $\boldsymbol{\delta_k}$ for $k > 1$ is generally ambiguous. Instead of interpreting these $\boldsymbol{\delta_k}$ parameters as indicating whether explanatory variables squeeze or expand the response scale, these parameters can equally well be interpreted as indicating heterogeneities in the effects of explanatory variables across the distribution of reported satisfaction. Indeed, via a change of notation, the linear specification of equation (18), together with equations (1) and (4), and when constraining $\boldsymbol{\delta_1}$ to 0, can be seen to be equivalent to the generalized ordered logit/probit model

described by Williams (2006).[42] That model explicitly allows for heterogenous effects across the distribution of the response variable. Taking unemployment as an example, the negative coefficient of -0.137 for the 4th cutoff can thus either be interpreted as unemployment bringing the 4th cutoff closer to the third (i.e. unemployment squeezing the response scale), or as indicating that the effect of unemployment is less severe for respondents whose level of life satisfaction is larger than the third response cutoff.

5 Conclusions

I hope to have made three main contributions. First, I showed how estimates of the effects of explanatory variables will be biased if the assumption of *common scale use* is violated. This analysis showed that violations of *common scale use* could in principle lead to sign reversals of the effects of explanatory variables. However, they are unlikely to do so in practice. Instead it is more likely that such violations lead to downward biases in the sizes of effect estimates. My theoretical analysis also showed that lagged scale shift effects will lead to the overestimation of adaptation and reference effects. Second, using variation in respondents' stated memories, I proposed a method by which contemporaneous scale shift effects may be separately identified from effects on latent satisfaction. Third, using the proposed memory approach, I analysed whether intrapersonal scale shift effects can plausibly be observed in British BHPS data. In my baseline analyses, I find, in line with theoretical expectation, that the direction in which explanatory variables affect scale use is typically the same as the direction in which explanatory variables affect latent satisfaction. I find particularly pronounced such effects for unemployment and bereavement. Under plausible assumptions on the

[42] In the probit version of that model, the probability of choosing response category k is given by $\Pr(r_{it} = k|\boldsymbol{X}_{it}) = \Phi(\omega_k - \boldsymbol{X}_{it}\boldsymbol{\rho}_k) - \Phi(\omega_{k-1} - \boldsymbol{X}_{it}\boldsymbol{\rho}_{k-1})$, where we impose $\omega_0 = -\infty$, $\omega_K = \infty$, and $\boldsymbol{\rho}_0 = 0$. This formulation allows the effects of explanatory variables on the latent variable to differ across response categories (hence the subscript k for $\boldsymbol{\rho}_k$). In the linear specification of the model discussed in the present paper, the same probability for $k > 1$ is given by:

$$\Pr(r_{it} = k|\boldsymbol{X}_{it}) = \Phi\left(\left[\sum_{m=1}^{k}\gamma_m + \boldsymbol{X}_{it}\boldsymbol{\delta}_m\right] - \boldsymbol{\beta}\boldsymbol{X}_{it}\right) - \Phi\left(\left[\sum_{m=1}^{k}\gamma_{m-1} + \boldsymbol{X}_{it}\boldsymbol{\delta}_{m-1}\right] - \boldsymbol{\beta}\boldsymbol{X}_{it}\right)$$

Where we impose $\gamma_0 = -\infty, \gamma_K = \infty, \gamma_K = \infty$, and $\boldsymbol{\delta}_0 = \boldsymbol{\delta}_1 = \mathbf{0}$. Since we can write $\boldsymbol{\rho}_k = \boldsymbol{\beta} - \sum_{m=1}^{k}\boldsymbol{\delta}_m$ and $\omega_k = \sum_{m=1}^{k}\gamma_m$ these formulations are equivalent. This equivalence does not hold in the case of the non-linear specification of equation (17). Nevertheless, the underlying point that we cannot separate heterogenous effects across the distribution of reported happiness from distortions of the response scale remains.

upper bound of memory biases, these scale effects turn out to be significant for the latter two variables, but not for my other explanatory variables.

In addition to these contributions, I evaluated Van Praag's (2015) approach of accounting for fixed effects in Appendices C and D. There, I showed that the method outperforms the more classical approach of using OLS to account for fixed effects. In my empirical analyses using BHPS data I further corroborated the viability of that approach. Given the ease at which Van Praag's approach can be implemented, it therefore seems useful to adopt his approach more widely.

The proposed method of using memories may be viewed as a complement to the only other viable method in the literature, i.e. the use of vignettes. Compared to the vignette approach, advantages of this approach are that the assumption of *response consistency* is more likely to be satisfied, that the approach is less costly to administer, and that it can be applied in the context of well-known panel data like BHPS or GSOEP.

However, *present independence* may be less plausible than the assumption of *vignette equivalence*. Future research may address this concern in three ways: First, it would help to obtain data on memories from multiple years in the past. With such data one could implement a version of d'Uva et al.'s (2011) test of vignette equivalence to test *present independence*.[43] Second, data in which questions on vignettes and memories are asked simultaneously would allow for a test of whether the two approaches yield equivalent results. Such cross-validation could yield strong evidence in favour of both approaches. Third, it would help to ask respondents about their current life satisfaction, about last year's (or earlier years') life satisfaction, and then – crucially – to ask respondents whether they can recall the rating they gave for their contemporaneous life satisfaction last year (following Howard and Dailey 1979).[44] Using such data, one could then perform an alternative formal test of

[43] In particular, if one had data on memories about multiple years ago, say $q = 1,2, \dots, Q$ years ago, one could extend equation (9) to read $s_{it}^{(m_q)} = \theta^{(m_q)} + \boldsymbol{X}_{it-q}\boldsymbol{\alpha}^{(m_q)} + \boldsymbol{X}_{it}\boldsymbol{\lambda}^{(m_q)} + \varepsilon_{it}^{(m_q)}$ and test if $\boldsymbol{\lambda}^{(m_q)} = \boldsymbol{0}$ for all $q > 1$.

[44] However, a concern with such a test is that biases in people's memories of their past ratings may differ from biases in peoples' memories of their latent satisfaction.

present independence by observing whether people's recollection of their past ratings is impacted by contemporaneous values of explanatory variables.

Hence, although it was not possible to validate the assumption of *present independence* with the data at hand, more detailed data would allow for such an exercise.

References

Adler, Matthew D. 2013. 'Happiness Surveys and Public Policy: What's the Use?' *Duke Law Journal* 62 (8): 1509–1601.

Ahmed, Sara, Nancy E. Mayo, Sharon Wood-Dauphinee, James A Hanley, and S Robin Cohen. 2004. 'Response Shift Influenced Estimates of Change in Health-Related Quality of Life Poststroke'. *Journal of Clinical Epidemiology* 57 (6): 561–70. https://doi.org/10.1016/j.jclinepi.2003.11.003.

Angelini, Viola, Danilo Cavapozzi, Luca Corazzini, and Omar Paccagnella. 2013. 'Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases'. *Oxford Bulletin of Economics and Statistics* 5. https://doi.org/10.1111/obes.12039.

Bago d'Uva, Teresa, Eddy Van Doorslaer, Maarten Lindeboom, and Owen O'Donnell. 2008. 'Does Reporting Heterogeneity Bias the Measurement of Health Disparities?' *Health Economics* 17 (3): 351–375.

Bertoni, Marco. 2015. 'Hungry Today, Unhappy Tomorrow? Childhood Hunger and Subjective Wellbeing Later in Life'. *Journal of Health Economics* 40 (March): 40–53. https://doi.org/10.1016/j.jhealeco.2014.12.006.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. 'Do People Mean What They Say? Implications for Subjective Survey Data'. *American Economic Review* 91 (2): 67–72.

Blanchflower, David, and Andrew Oswald. 2017. 'Do Humans Suffer a Psychological Low in Midlife? Two Approaches (With and Without Controls) in Seven Data Sets.' *NBER Working Paper*, no. 23724. https://doi.org/10.3386/w23724.

Bond, Timothy N, and Kevin Lang. 2018. 'The Sad Truth About Happiness Scales'. *Journal of Political Economy*. https://doi.org/10.1086/701679.

Bonsang, Eric, and Arthur Van Soest. 2012. 'Satisfaction with Job and Income Among Older Individuals Across European Countries'. *Social Indicators Research* 105 (2): 227–54. https://doi.org/10.1007/s11205-011-9879-5.

Borah, Melanie, Carina Keldenich, and Andreas Knabe. 2019. 'Reference Income Effects in the Determination of Equivalence Scales Using Income Satisfaction Data'. *Review of Income and Wealth* 65 (4): 736–70. https://doi.org/10.1111/roiw.12386.

Clark, Andrew E., and Yannis Georgellis. 2013. 'Back to Baseline in Britain: Adaptation in the British Household Panel Survey'. *Economica* 80 (319): 496–512. https://doi.org/10.1111/ecca.12007.

Conway, Michael, and Michael Ross. 1984. 'Getting What You Want by Revising What You Had.' *Journal of Personality and Social Psychology* 47 (4): 738.

Corrado, Luisa, and Melvyn Weeks. 2010. 'Identification Strategies in Survey Response Using Vignettes'. *CWPE Working Paper* 1031.

Crettaz, Eric, and Christian Suter. 2013. 'The Impact of Adaptive Preferences on Subjective Indicators: An Analysis of Poverty Indicators'. *Social Indicators Research* 114 (1): 139–152. https://doi.org/10.1007/s11205-013-0388-6.

Datta Gupta, Nabanita, Nicolai Kristensen, and Dario Pozzoli. 2010. 'External Validation of the Use of Vignettes in Cross-Country Health Studies'. *Economic Modelling* 27 (4): 854–65. https://doi.org/10.1016/j.econmod.2009.11.007.

De Neve, Jan-Emmanuel, and Andrew J. Oswald. 2012. 'Estimating the Influence of Life Satisfaction and Positive Affect on Later Income Using Sibling Fixed Effects'. *Proceedings of the National Academy of Sciences* 109 (49): 19953–19958.

Dermer, Marshall, Sidney J. Cohen, Elaine Jacobsen, and Erling A. Anderson. 1979. 'Evaluative Judgments of Aspects of Life as a Function of Vicarious Exposure to Hedonic Extremes.' *Journal of Personality and Social Psychology* 37 (2): 247.

Diamond, Peter. 2008. 'Behavioral Economics'. *Journal of Public Economics* 92 (8–9): 1858–62. https://doi.org/10.1016/j.jpubeco.2008.03.003.

Diener, Ed, Ronald Inglehart, and Louis Tay. 2013. 'Theory and Validity of Life Satisfaction Scales'. *Social Indicators Research* 112 (3): 497–527. https://doi.org/10.1007/s11205-012-0076.

Dominitz, Jeff, and Charles F. Manski. 1997. 'Using Expectations Data to Study Subjective Income Expectations'. *Journal of the American Statistical Association* 92 (439): 855–67. https://doi.org/10.1080/01621459.1997.10474041.

Easterlin, Richard A. 2001. 'Income and Happiness: Towards a Unified Theory'. *The Economic Journal* 111 (473): 465–84. https://doi.org/10.1111/1468-0297.00646.

Falk, Armin, and Markus Knell. 2004. 'Choosing the Joneses: Endogenous Goals and Reference Standards'. *Scandinavian Journal of Economics* 106 (3): 417–435.

Ferrer-i-Carbonell, Ada, and Paul Frijters. 2004. 'How Important Is Methodology for the Estimates of the Determinants of Happiness?' *The Economic Journal* 114 (497): 641–659.

Fleurbaey, Marc, and Didier Blanchet. 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford University Press.

Forgas, Joseph P. 1995. 'Mood and Judgment: The Affect Infusion Model (AIM).' *Psychological Bulletin* 117 (1): 39.

Frederick, Shane, and George Loewenstein. 1999. 'Hedonic Adaptation'. In *Well-Being: Foundations of Hedonic Psychology*, edited by Daniel Kahneman, Ed Diener, and Norbert Schwarz.

Frijters, Paul, Andrew Clark, Christian Krekel, and Richard Layard. 2019. 'A Happy Choice: Wellbeing as the Goal of Government'. *IZA Discussion Paper*, no. 12720.

Frijters, Paul, David W. Johnston, and Michael A. Shields. 2011. 'Happiness Dynamics with Quarterly Life Event Data'. *Scandinavian Journal of Economics* 113 (1): 190–211. https://doi.org/10.1111/j.1467-9442.2010.01638.x.

Gibbons, F. X. 1999. 'Social Comparison as a Mediator of Response Shift'. *Social Science & Medicine* 48 (11): 1517–30. https://doi.org/10.1016/S0277-9536(99)00046-5.

Greene, William. 2004. 'The Behaviour of the Maximum Likelihood Estimator of Limited Dependent Variable Models in the Presence of Fixed Effects'. *The Econometrics Journal* 7 (1): 98–119. https://doi.org/10.1111/j.1368-423X.2004.00123.x.

Greene, William H., and David A. Hensher. 2010. *Modeling Ordered Choices: A Primer and Recent Developments*. Cambridge University Press. https://papers.ssrn.com/abstract=1213093.

Hanglberger, Dominik, and Joachim Merz. 2015. 'Does Self-Employment Really Raise Job Satisfaction? Adaptation and Anticipation Effects on Self-Employment and General Job Changes'. *Journal for Labour Market Research* 48 (4): 287–303.

Hoogstraten, Joh. 1985. 'Influence of Objective Measures on Self-Reports in a Retrospective Pretest-Posttest Design'. *The Journal of Experimental Education* 53 (4): 207–10. https://doi.org/10.1080/00220973.1985.10806383.

Howard, George S., and Patrick R. Dailey. 1979. 'Response-Shift Bias: A Source of Contamination of Self-Report Measures'. *Journal of Applied Psychology* 64 (2): 144–50.

Howard, George S., Kenneth M. Ralph, Nancy A. Gulanick, Scott E. Maxwell, Don W. Nance, and Sterling K. Gerber. 1979. 'Internal Invalidity in Pretest-Posttest Self-Report Evaluations and a Re-Evaluation of Retrospective Pretests'. *Applied Psychological Measurement* 3 (1): 1–23.

Howard, George S., Ronald R. Schmeck, and James H. Bray. 1979. 'Internal Invalidity in Studies Employing Self-Report Instruments: A Suggested Remedy'. *Journal of Educational Measurement*, 129–135.

Kapteyn, Arie, James P Smith, and Arthur Van Soest. 2007. 'Vignettes and Self-Reports of Work Disability in the United States and the Netherlands'. *The American Economic Review* 97 (1): 461–73.

Kapteyn, Arie, James P. Smith, and Arthur Van Soest. 2010. 'Life Satisfaction'. In *International Differences in Wellbeing*, edited by Ed Diener, John F. Helliwell, and Daniel Kahneman. Oxford University Press.

King, Gary, Christopher J L Murray, Joshua A Salomon, Ajay Tandon, Gary King, Christopher J L Murray, World Health, and Joshua A Salomon. 2004. 'Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research'. *The American Political Science Review* 98 (1): 191–207.

Klaassen, Tineke, W. J. Riedel, N. E. P. Deutz, and H. M. Van Praag. 2002. 'Mood Congruent Memory Bias Induced by Tryptophan Depletion'. *Psychological Medicine* 32 (1): 167.

Kristensen, Nicolai, and Edvard Johansson. 2008. 'New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes'. *Labour Economics* 15 (1): 96–117. https://doi.org/10.1016/j.labeco.2006.11.001.

Layard, Richard, Andrew Clark, Christian Krekel, Daisy Fancourt, Nancy Hey, Gus O'Donell, and Jan-Emmanuel De Neve. 2020. 'When to Release the Lockdown? A Wellbeing Framework for Analysing Costs and Benefits'. *CEP Occasional Paper*, no. 49. http://cep.lse.ac.uk/pubs/download/occasional/op049.pdf.

Levine, Linda J., and Susan Bluck. 2004. 'How Emotions Fade: Valence, Appraisals, and the Emotional Impact of Remembered Events.'

Levy, Horacio, and Stephen P. Jenkins. 2012. 'Documentation for Derived Current and Annual Net Household Income Variables, BHPS Waves 1-18'. *UK Data Archive Study Number 3909*.

Lewis, P. A., H. D. Critchley, A. P. Smith, and R. J. Dolan. 2005. 'Brain Mechanisms for Mood Congruent Memory Facilitation'. *NeuroImage* 25 (4): 1214–23. https://doi.org/10.1016/j.neuroimage.2004.11.053.

Luechinger, Simon. 2009. 'Valuing Air Quality Using the Life Satisfaction Approach'. *The Economic Journal* 119 (536): 482–515.

Mayer, John D., Yvonne N. Gaschke, Debra L. Braverman, and Temperance W. Evans. 1992. 'Mood-Congruent Judgment Is a General Effect.' *Journal of Personality and Social Psychology* 63 (1): 119.

Montgomery, Mallory. 2017. 'Reversing the Gender Gap in Happiness: Validating the Use of Life Satisfaction Self-Reports Worldwide'. *Working Paper*.

Mundlak, Yair. 1978. 'On the Pooling of Time Series and Cross Section Data'. *Econometrica*, 69–85.

Murray, Christopher JL, Emre Ozaltin, Ajay Tandon, Joshua Salomon, Ritu Sadana, and Somnath Chatterji. 2003. 'Empirical Evaluation of the Anchoring Vignette Approach in Health Surveys'. *Health Systems Performance Assessment: Debates, Methods and Empiricism*, 369–99.

Oswald, Andrew J. 2008. 'On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings'. *Economics Letters* 100 (3): 369–72. https://doi.org/10.1016/j.econlet.2008.02.032.

Plant, Michael. 2020. 'A Happy Possibility About Happiness (And Other) Scales: Understanding Why the Cardinality Assumption Is Both Defensible and Unnecessary'. *Unpublished Working Paper*.

Prati, Alberto. 2017. 'Hedonic Recall Bias. Why You Should Not Ask People How Much They Earn'. *Journal of Economic Behavior & Organization* 143: 78–97.

Prati, Alberto, and Claudia Senik. 2020. 'Feeling Good or Feeling Better?' *PSE Working Papers*, no. 17. https://econpapers.repec.org/paper/halpsewpa/halshs-02545228.htm.

Ravallion, Martin, Kristen Himelein, and Kathleen Beegle. 2016. 'Can Subjective Questions on Economic Welfare Be Trusted?' *Economic Development and Cultural Change* 64 (4): 697–726.

Riedl, Maximilian, and Ingo Geishecker. 2014. 'Keep It Simple: Estimation Strategies for Ordered Response Models with Fixed Effects'. *Journal of Applied Statistics* 41 (11): 2358–74. https://doi.org/10.1080/02664763.2014.909969.

Schröder, Carsten, and Shlomo Yitzhaki. 2017. 'Revisiting the Evidence for Cardinal Treatment of Ordinal Variables'. *European Economic Review* 92 (February): 337–58. https://doi.org/10.1016/j.euroecorev.2016.12.011.

Schwarz, Norbert, and Fritz Strack. 1999. 'Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications'. In *Well-Being: The Foundations of Hedonic Psychology*, 7:61–84.

Söderbom, Måns. 2011. 'Econometrics II Lecture 2: Discrete Choice Models'. http://www.soderbom.net/metrix2/lec2.pdf.

Sprangers, Mirjam A. G, and Carolyn E Schwartz. 1999. 'Integrating Response Shift into Health-Related Quality of Life Research: A Theoretical Model'. *Social Science & Medicine* 48 (11): 1507–15. https://doi.org/10.1016/S0277-9536(99)00045-3.

Sprangers, Mirjam A. G., Frits S. A. Van Dam, Jenny Broersen, Litanja Lodder, Lidwina Wever, Mechteld R. M. Visser, Paul Oosterveld, and Ellen M.A. Smets. 1999. 'Revealing Response Shift in Longitudinal Research on Fatigue: The Use of the Thentest Approach'. *Acta Oncologica* 38 (6): 709–18. https://doi.org/10.1080/028418699432860.

Steffel, Mary, and Daniel M. Oppenheimer. 2009. 'Happy by What Standard? The Role of Interpersonal and Intrapersonal Comparisons in Ratings of Happiness'. *Social Indicators Research* 92 (1): 69–79. https://doi.org/10.1007/s11205-008-9289-5.

Stutzer, Alois. 2004. 'The Role of Income Aspirations in Individual Happiness'. *Journal of Economic Behavior and Organization* 54 (1): 89–109. https://doi.org/10.1016/j.jebo.2003.04.003.

Taylor, Shelley E., Joanne V. Wood, and Rosemary R. Lichtman. 1983. 'It Could Be Worse: Selective Evaluation as a Response to Victimization'. *Journal of Social Issues* 39 (2): 19–40. https://doi.org/10.1111/j.1540-4560.1983.tb00139.x.

Townsend, Michael, and Keri Wilton. 2003. 'Evaluating Change in Attitude towards Mathematics Using the "Then-Now" Procedure in a Cooperative Learning Programme'. *British Journal of Educational Psychology* 73 (4): 473–87. https://doi.org/10.1348/000709903322591190.

Urban, Emily J., Susan T. Charles, Linda J. Levine, and David M. Almeida. 2018. 'Depression History and Memory Bias for Specific Daily Emotions'. *PLoS ONE* 13 (9). https://doi.org/10.1371/journal.pone.0203574.

Uva, Teresa Bago d', Maarten Lindeboom, Owen O'Donnell, and Eddy Van Doorslaer. 2011. 'Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity'. *Journal of Human Resources* 46 (4): 875–906.

Van Praag, Bernard M. S. 1971. 'The Welfare Function of Income in Belgium: An Empirical Investigation'. *European Economic Review* 2 (3): 337–69.

———. 1993. 'The Relativity of the Welfare Concept'. In *The Quality of Life*, edited by Martha Nussbaum and Amartya Sen, 362–85. Oxford University Press. https://doi.org/10.1093/0198287976.003.0027.

———. 2015. 'A New View on Panel Econometrics: Is Probit Feasible After All?' *IZA Discussion Paper*, no. 9345.

Van Praag, Bernard M. S., and Ada Ferrer-i-Carbonell. 2008. *Happiness Quantified: A Satisfaction Calculus Approach*. 2nd ed. Oxford University Press.

Van Praag, Bernard M. S., and Paul Frijters. 1999. 'The Measurement of Welfare and Well-Being: The Leyden Approach'. In *Well-Being: Foundations of Hedonic Psychology*, edited by Daniel Kahneman, Ed Diener, and Norbert Schwarz, 302.

Vendrik, Maarten C. M. 2013. 'Adaptation, Anticipation and Social Interaction in Happiness: An Integrated Error-Correction Approach'. *Journal of Public Economics* 105: 131–149. https://doi.org/10.1016/j.jpubeco.2013.06.009.

Viscusi, W. Kip. 2020. 'Wellbeing Measures of Mortality Risks: Life-Cycle Contradictions and Ordinal Index Challenges'. *Behavioural Public Policy* 4 (2): 245–53. https://doi.org/10.1017/bpp.2019.47.

Williams, Richard. 2006. 'Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables'. *Stata Journal* 6 (1): 58–82. https://doi.org/st0097.

Wilson, Timothy D., and Daniel T. Gilbert. 2005. 'Affective Forecasting'. *Current Directions in Psychological Science* 14 (3): 131.

Wilson, Timothy D., Jay Meyers, and Daniel T. Gilbert. 2003. '"How Happy Was I, Anyway?" A Retrospective Impact Bias'. *Social Cognition* 21 (6): 421–46. https://doi.org/10.1521/soco.21.6.421.28688.

Wolbring, Tobias, Marc Keuschnigg, and Eva Negele. 2013. 'Needs, Comparisons, and Adaptation: The Importance of Relative Income for Life Satisfaction'. *European Sociological Review* 29 (1): 86–104. https://doi.org/10.1093/esr/jcr042.

Xu, Hongwei, and Yu Xie. 2016. 'Assessing the Effectiveness of Anchoring Vignettes in Bias Reduction for Socioeconomic Disparities in Self-Rated Health among Chinese Adults'. *Sociological Methodology* 46 (1): 84–120. https://doi.org/10.1177/0081175015599808.

Appendix

A Interpretation of absolute magnitudes of coefficients

Life satisfaction, like most other subjective constructs, has no natural cardinal scale, with no well-defined physical phenomena to ground it and no natural zero point. Despite this feature of subjective constructs, some interpretations of the absolute magnitudes of coefficients may nevertheless be both useful and meaningful.

It might be natural to interpret the magnitudes of coefficients in relation to how a (marginal) change in covariates changes the probability of exceeding some particular response level. Given the model described by equations (1), (4) and (5), the probability of responding category $r > k$ conditional on $\boldsymbol{X}_{it*}$ is given by $\Pr(r_{it} > k|\boldsymbol{X}_{it*}) = 1 - \Phi(\gamma_k + \boldsymbol{X}_{it*}\boldsymbol{\delta}_* - \boldsymbol{X}_{it*}\boldsymbol{\beta}_*) = \Phi(\boldsymbol{X}_{it*}(\boldsymbol{\beta}_* - \boldsymbol{\delta}_*) - \gamma_k)$. Taking the partial derivative with respect to some particular variable, say $X_{j,it}$, we get $\partial\Pr(r_{it} > k|\boldsymbol{X}_{it*})/\partial X_{j,it} = \varphi(\boldsymbol{X}_{it*}(\boldsymbol{\beta}_* - \boldsymbol{\delta}_*) - \gamma_k)(\beta_{j,0} - \delta_{j,0})$, where $\varphi(.)$ denotes the standard normal PDF. As explained in the main text, an analysis which does not account for shifts in scale use already estimates $\boldsymbol{\beta}_* - \boldsymbol{\delta}_*$. Therefore, if we are only interested in estimating expected effects on response probabilities, accounting for changes in scale use would serve no purpose.

One could average $\varphi\big(\boldsymbol{X}_{it*-j}(\boldsymbol{\beta}_{*-j} - \boldsymbol{\delta}_{*-j}) + X_{j,it}\beta_{j,0} - \gamma_k\big)\beta_{j,0}$ over all observed combinations of $\boldsymbol{X}_{it*}$. Here, $\boldsymbol{X}_{it*-j}$ is the same as $\boldsymbol{X}_{it*}$ but has element $X_{j,it}$ removed. Vectors $\boldsymbol{\beta}_{*-j}$ and $\boldsymbol{\delta}_{*-j}$ are defined analogously. Such an exercise can be interpreted as estimating the average marginal effect of $X_{j,it}$ on the probability of responding category $r > k$ conditional on $\boldsymbol{X}_{it*}$ under a counterfactual in which $X_{j,it}$ did not also affect scale use. However, this is cumbersome to compute and not a very intuitive interpretation of the absolute magnitude.

In the context of OLS regressions, one oft-used intuitive interpretation is to discuss by how many categories a respondent can be expected to change her response given a unit change in an

explanatory variable's value. This is very straightforward in this context and given by $\mathrm{E}(r_{it}|X_{j,it}+1;\boldsymbol{X}_{it*})-\mathrm{E}(r_{it}|X_{j,it};\boldsymbol{X}_{it*})=\beta_{j,0}$.

However, in the ordered probit context, differences between thresholds are not equal. Nevertheless, an analogue may usefully indicate how much latent satisfaction changes for a unit change in $X_{j,it}$ as a fraction of the normalized width of the satisfaction scale. Here, the natural normalization is over the number of estimated thresholds. The width of the scale can be given by the difference between the highest and lowest estimated threshold.

Formally this is given by:

$$\begin{aligned}&\left[\mathrm{E}\left(s_{it}/(\tau_{K-1}-\tau_1)\,|X_{j,it}+1;\boldsymbol{X}_{it*}\right)-\mathrm{E}\left(s_{it}/(\tau_{K-1}-\tau_1)\,|X_{j,it};\boldsymbol{X}_{it*}\right)\right](K-1)\\&=\Big[\Big(\boldsymbol{X}_{it*}\boldsymbol{\beta}_*+(X_{j,it}+1)\beta_{j,0}\Big)/\Big(\gamma_{K-1}+\boldsymbol{X}_{it*}\boldsymbol{\delta}_*+(X_{j,it}+1)\delta_{j,0}-\left(\gamma_1+\boldsymbol{X}_{it*}\boldsymbol{\delta}_*+(X_{j,it}+1)\delta_{j,0}\right)\Big)\\&\quad-\Big(\Big(\boldsymbol{X}_{it*}\boldsymbol{\beta}_*+X_{j,it}\beta_{j,0}\Big)/\Big(\gamma_{K-1}+\boldsymbol{X}_{it*}\boldsymbol{\delta}_*+(X_{j,it}+1)\delta_{j,0}-\left(\gamma_1+\boldsymbol{X}_{it*}\boldsymbol{\delta}_*+X_{j,it}\delta_{j,0}\right)\Big)\Big)\Big](K-1)\\&=\beta_{j,0}(K-1)/(\gamma_{K-1}-\gamma_1)\end{aligned}\qquad(A1)$$

Thus, the effect of $X_{j,it}$ on scale use in this expression cancels in the case of parallel shifts. Fortunately, the final line of this expression is directly obtained from our parameter estimates. We may hence use this as an intuitive interpretation of the absolute magnitudes of effects on latent satisfaction.

B Using vignettes to correct for violations of common scale use

The vignette approach tackles violations of common scale use by presenting respondents with short descriptions of imaginary persons' lives (King et al., 2004). An example from the SHARE survey for such a vignette is the following:

'John is 63 years old. His wife died 2 years ago, and he still spends a lot of time thinking about her. He has four children and ten grandchildren who visit him regularly. John can make ends meet but has no money for extras such as expensive gifts to his grandchildren. He has had to stop working recently due to heart problems. He gets tired easily. Otherwise, he has no serious health conditions. How satisfied with his life do you think John is?'.

Respondents are then asked to rate the life satisfaction of these imaginary persons. On the assumption that all respondents perceive the described person's satisfaction equally, identification of differences in scale use become possible.

No study to date used vignettes to analyse the separate effects of lags and leads of explanatory variables on both latent life satisfaction and scale use. To ease the exposition, I will therefore assume that latent life satisfaction and scale use only depend on contemporaneous values of covariates. Parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ (without subscripts) are thus only associated with contemporaneous values of covariates. If a panel dataset containing vignettes existed, an analysis of the effect of lags and leads would pose no special problem. Unfortunately, I am not aware of such data. This partly motivates the approach proposed in section 3.

B.1 Formal approach

Assume that, from a set of V vignettes, respondents give a report $r_{it}^{(v)}$ about a particular vignette v on the basis of a perceived satisfaction level $s_{it}^{(v)}$ of that vignette. Also assume that respondents use the same response scale for their own satisfaction as for rating the vignettes. This yields a reporting rule analogous to equation (1):

$$r_{it}^{(v)} = k \quad \leftrightarrow \quad \tau_{it,k-1} < s_{it}^{(v)} \leq \tau_{it,k} \tag{A2}$$

Since the same thresholds are used in equations (1) and (A2), we assume *response consistency*. Further assume that the perceived level of $s_{it}^{(v)}$ does not depend on $\boldsymbol{X}_{it}$. This assumption is known as the *vignette equivalence* assumption. Now consider:

$$s_{it}^{(v)} = \theta^{(v)} + \boldsymbol{X}_{it}\boldsymbol{\lambda}^{(v)} + \varepsilon_{it}^{(v)} \qquad (A3)$$

Here, $\theta^{(v)}$ is a vignette-specific parameter and $\varepsilon_{it}^{(v)}$ is a normally distributed error term with mean 0 and standard deviation 1. *Vignette equivalence* entails that $\boldsymbol{\lambda}^{(v)} = \mathbf{0}$. Incorporating these additional equations, the individual contributions to the likelihood then become:

$$\begin{aligned}\ell_{it} = \ln \sum_{k=1}^{K} \mathrm{I}(r_{it} = k)[\Phi(\gamma_k + \boldsymbol{X}_{it}\boldsymbol{\delta} - \boldsymbol{X}_{it}\boldsymbol{\beta}) - \Phi(\gamma_{k-1} + \boldsymbol{X}_{it}\boldsymbol{\delta} - \boldsymbol{X}_{it}\boldsymbol{\beta})] \\ + \sum_{v}\sum_{k=1}^{K} \mathrm{I}\left(r_{it}^{(v)} = k\right)\left[\Phi\left(\gamma_k + \boldsymbol{X}_{it}\boldsymbol{\delta} - \theta^{(v)} - \boldsymbol{X}_{it}\boldsymbol{\lambda}^{(v)}\right) - \Phi\left(\gamma_{k-1} + \boldsymbol{X}_{it}\boldsymbol{\delta} - \theta^{(v)} - \boldsymbol{X}_{it}\boldsymbol{\lambda}^{(v)}\right)\right]\end{aligned} \qquad (A4)$$

When assuming $\boldsymbol{\lambda}^{(v)} = \mathbf{0}$, $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are separately identified from maximising equation (A4).

B.2 Plausibility of *vignette equivalence* and *response consistency*

The vignette approach exchanges the assumption of *common scale use* with the assumptions of *response consistency* and *vignette equivalence*. Regarding *response consistency*, and extending the worries of section 2.2, respondents might rate vignettes relative to reference points that vary with characteristics given in the vignette: Persons described to be richer might be rated with a higher reference point appropriate for richer persons, and persons described as poor might be rated less stringently. Unfortunately, statistical tests of *response consistency* rely on having a further objective indicator for the dependent variable. In the case of entirely subjective variables, no objective indicators can exist, making such tests impossible. However, in cases where *response consistency* with respect to health was tested, *response consistency* was rejected (d'Uva et al., 2011; Datta Gupta et al., 2010).

Regarding *vignette equivalence*, some studies test whether respondents give a common ordering to multiple vignettes (Angelini et al., 2013; d'Uva et al., 2011; Montgomery, 2017; Murray et al., 2003).

Although this test is typically passed, failure to reject does not imply that vignette equivalence holds. d'Uva et al. (2011) propose a stronger test, noting that vignette equivalence is required for only one vignette, say for $v = 1$. For all other vignettes, *vignette equivalence* can be relaxed, thus allowing a test of $\boldsymbol{\lambda}^{(v)} = \mathbf{0}$ in $s_{it}^{(v)} = \theta^{(v)} + \boldsymbol{X}_{it}\boldsymbol{\lambda}^{(v)} + \varepsilon_{it}^{(v)}$ for $v = 2, \dots, V$. They find that vignette equivalence is rejected for almost all considered covariates.[45]

Vignettes are incomplete descriptions of a person's life. Some aspects relevant to life satisfaction must therefore be imagined by the respondent. For example, in the case of the vignette quoted above, respondents must make some assumptions about what it means to make 'ends meet'. It is plausible that a person's current circumstances (i.e. values of $\boldsymbol{X}_{it}$), will impact how such assumptions are made. In particular, for a respondent with low incomes or who is unemployed, making 'ends meet' likely means a lower level of economic welfare compared to respondents in stable employment or for those with high incomes (Van Praag and Frijters 1999; Van Praag 1971; Crettaz and Suter 2013). Thus, while income and being employed are positively associated with latent satisfaction, they will also be positively associated with respondents' perceptions of vignettes. This implies that $\lambda_j^{(v)}$ and β_j have the same sign for such variables. As shown in section 3.3, estimates of β_j are then biased towards zero. Moreover, individuals tend to make mood-congruent judgements in both self- and non-self-relevant domains (Mayer et al. 1992; Forgas 1995).[46] This means that, respondents with more positive affect can be expected to judge others more positively, too. Since affect and life satisfaction are driven by similar variables, this also suggests that $\lambda_j^{(v)}$ and

[45] Rejection of this test can be interpreted as indicating that there are differences across vignettes in how perceptions of vignettes' satisfaction depend on covariates. Passing this test does not imply that vignette equivalence holds. Instead, failure to pass indicates that vignette equivalence does not hold for at least one vignette. Rejection of this test can be interpreted as indicating that there are differences across vignettes in how perceptions of vignettes' satisfaction depend on covariates. Relatedly, Kristensen & Johansson (2008) propose to split the sample into multiple groups and to test whether estimates of $\boldsymbol{\delta}$ are the same across groups. Unfortunately, this test cannot decide between violations of *vignette equivalence* and true heterogeneity in $\boldsymbol{\delta}$.

[46] For example, in study 3 of Mayer et al. (1992), respondents in a representative sample of the population of New Hamshire (USA) were asked to rate both their mood, as well as to make a number of other-regarding judgements. Those judgements included assessments of e.g. 'the likelihood of a marriage resulting in long-term happiness for both members of the couple?', 'the likelihood that there will be an atomic war in the next five years', or 'the likelihood that the average 30-year-old in New Hampshire will be involved in a happy, loving romance?'. For all such cases, those with a better mood provided significantly higher probabilities and thus made mood-congruent judgements.

β_j have the same sign. Hence, vignette equivalence might fail. The direction in which it is likely to fail suggests that magnitudes of $\boldsymbol{\beta}$ will be biased towards zero.

B.3 Previous results using vignettes

The vignettes methodology was previously applied to investigate e.g. variation in the reporting of subjective health (Bago d'Uva et al. 2008; Xu and Xie 2016), work disability (Kapteyn, Smith, and Van Soest 2007), subjective poverty (Ravallion, Himelein, and Beegle 2016), as well as income and job satisfaction (Kristensen and Johansson 2008; Bonsang and Van Soest 2012). Five studies applied the vignette approach to studying life satisfaction. Kapteyn et al. (2010) use Dutch 'CenterPanel' data from 2006 and American 'Rand American Life' data from 2006-2007. For the USA, they find that accounting for scale use increases the estimated effects of income and household size (in the USA). In the Dutch sample, the effects of being employed, and being married are increased in size. Montgomery (2017) uses global Gallup World Poll data for the years 2011-2014. She shows that the relative gender gap in life satisfaction (with women being more satisfied) is driven by differential scale use. Furthermore, compared to being married, accounting for differential scale use marginally increases the magnitudes of the negative effects of being single, separation, divorce, and widowhood. The effects of unemployment and log income are also marginally increased. However, among two sets of vignettes available to her, one of these sets fails d'Uva et al.'s (2011) test of *vignette equivalence*. Angelini et al. (2013) use European SHARE data. They find large differences in scale use across European countries. Moreover, compared to having a partner, they also find slightly increased effects for being single and widowhood. Increased effects are also observed for not working and retirement. Finally, Bertoni (2015) and Corrado and Weeks (2010) use the same SHARE dataset. Bertoni (2015) finds that childhood deprivation is associated with less stringent scale use. Apart from the case of gender investigated by Montgomery (2017), whenever there was a change in estimated coefficients due to accounting for variation in scale use, estimated magnitudes of effects on latent satisfaction increased. This is despite the possible bias toward zero discussed above.

C Van Praag's (2015) argument

Van Praag's argument may be stated as follows. First, extend equation (4) by writing $\varepsilon_{it} = u_i + v_{it}$. This yields:

$$s_{it} = \boldsymbol{X}_{it*}\boldsymbol{\beta}_* + u_i + v_{it} \tag{A5}$$

The fixed effect u_i may be correlated with $\boldsymbol{X}_{it*}$. Assume that v_{it}, after purging it of the individual fixed effect, is normally distributed with constant variance and uncorrelated with $\boldsymbol{X}_{it*}$.

Now stack observations over the T_i periods for each individual i:

$$\boldsymbol{s_i} = \boldsymbol{X}_i\boldsymbol{\beta}_* + \boldsymbol{u_i} + \boldsymbol{v_i}. \tag{A6}$$

Here, $\boldsymbol{s_i} = (s_{i1}, \ldots, s_{iT_i})'$ and analogously for $\boldsymbol{X_i}, \boldsymbol{u_i}$, and $\boldsymbol{v_i}$. Now consider the matrix $\boldsymbol{M_i} = \frac{1}{T_i}\mathbf{1}\mathbf{1}'$, where $\mathbf{1}$ is a vector of T_i ones. For any matrix $\boldsymbol{A}_i$, define $\boldsymbol{M_iA_i} \equiv \overline{\boldsymbol{A}}_\boldsymbol{i}$. Each column of $\overline{\boldsymbol{A}}_\boldsymbol{i}$ contains the average of each column of $\boldsymbol{A}_i$. The demeaned matrix $\ddot{\boldsymbol{A}}_i$ is then obtained as $\ddot{\boldsymbol{A}}_i = \boldsymbol{A}_i - \boldsymbol{M_iA_i}$. Equation (A6) is a constrained version of:

$$\boldsymbol{s_i} = \ddot{\boldsymbol{X}}_\boldsymbol{i}\ddot{\boldsymbol{\beta}}_* + \overline{\boldsymbol{X}}_\boldsymbol{i}\overline{\boldsymbol{\beta}}_* + \boldsymbol{u_i} + \boldsymbol{v_i}. \tag{A7}$$

Equation (A7) can be recovered from (A6) when $\ddot{\boldsymbol{\beta}}_* = \overline{\boldsymbol{\beta}}_*$. Since $\boldsymbol{M_i}$ is idempotent, we find that $\ddot{\boldsymbol{X}}_\boldsymbol{i}'\boldsymbol{u_i} = \mathbf{0}$.[47] Thus, $\ddot{\boldsymbol{X}}_\boldsymbol{i}$ and $\boldsymbol{u_i}$ are uncorrelated and the exclusion or inclusion of $\boldsymbol{u_i}$ does not affect estimates of $\ddot{\boldsymbol{\beta}}_*$. Also note that $\ddot{\boldsymbol{X}}_\boldsymbol{i}'\boldsymbol{v_i} = \ddot{\boldsymbol{X}}_\boldsymbol{i}'\ddot{\boldsymbol{v}}_\boldsymbol{i}$,[48] where $\ddot{\boldsymbol{v}}_\boldsymbol{i}$ is obtained by demeaning equation (A5), as is normally done when accounting for fixed effects. Therefore, even when omitting $\boldsymbol{u_i}$, OLS estimates of $\ddot{\boldsymbol{\beta}}_*$ obtained from estimating equation (A7), would be identical to OLS estimates obtained by demeaning equation (A5).

[47] A matrix $\boldsymbol{A}$ is idempotent iff $\boldsymbol{AA} = \boldsymbol{A}$. We can therefore write: $\ddot{\boldsymbol{X}}_\boldsymbol{i}'\boldsymbol{u_i} = \ddot{\boldsymbol{X}}_\boldsymbol{i}'\boldsymbol{M_iu_i} = \big((\boldsymbol{I_i} - \boldsymbol{M_i})\boldsymbol{X_i}\big)'\boldsymbol{M_iu_i} = (\boldsymbol{X_i}'(\boldsymbol{I_i} - \boldsymbol{M_i})')\boldsymbol{M_iu_i} = (\boldsymbol{X_i}' - \boldsymbol{X_i}'\boldsymbol{M_i}')\boldsymbol{M_iu_i} = \boldsymbol{X_i}'\boldsymbol{M_iu_i} - \boldsymbol{X_i}'\boldsymbol{M_i}'\boldsymbol{M_iu_i} = \mathbf{0}$.

[48] Because: $\ddot{\boldsymbol{X}}_\boldsymbol{i}'\boldsymbol{v_i} = \ddot{\boldsymbol{X}}_\boldsymbol{i}'(\ddot{\boldsymbol{v}}_\boldsymbol{i} + \overline{\boldsymbol{v}}_\boldsymbol{i}) = \ddot{\boldsymbol{X}}_\boldsymbol{i}'\ddot{\boldsymbol{v}}_\boldsymbol{i} + \ddot{\boldsymbol{X}}_\boldsymbol{i}'\overline{\boldsymbol{v}}_\boldsymbol{i} = \ddot{\boldsymbol{X}}_\boldsymbol{i}'\ddot{\boldsymbol{v}}_\boldsymbol{i} + (\boldsymbol{X_i}' - \boldsymbol{X_i}'\boldsymbol{M_i}')\boldsymbol{M_iv_i} = \ddot{\boldsymbol{X}}_\boldsymbol{i}'\ddot{\boldsymbol{v}}_\boldsymbol{i} + \boldsymbol{X_i}'\boldsymbol{M_iv_i} - \boldsymbol{X_i}'\boldsymbol{M_i}'\boldsymbol{M_iv_i} = \ddot{\boldsymbol{X}}_\boldsymbol{i}'\ddot{\boldsymbol{v}}_\boldsymbol{i}$.

Although van Praag (2015) does not take this step, we can rewrite equation (A7) as:

$$\boldsymbol{s_i} = \boldsymbol{X_i}\ddot{\boldsymbol{\beta}}_* + \bar{\boldsymbol{X}}_{\boldsymbol{i}}\widetilde{\boldsymbol{\beta}}_* + \boldsymbol{u_i} + \boldsymbol{v_i} \tag{A8}$$

Here, $\bar{\boldsymbol{\beta}}_* - \ddot{\boldsymbol{\beta}}_* \equiv \widetilde{\boldsymbol{\beta}}_*$. This yields the specification proposed by Mundlak (1978).[49] Van Praag (2015) makes a further claim. When $\boldsymbol{s_i}$ is not observed, and only $\boldsymbol{r_i}$ is, an estimation of equation (A7) by ordered probit will also yield a consistent estimate of $\ddot{\boldsymbol{\beta}}_*$ when omitting $\boldsymbol{u_i}$.

However, estimates of $\ddot{\boldsymbol{\beta}}_*$ will only by consistent up to a proportionality factor. Since that factor is constant across all elements in $\ddot{\boldsymbol{\beta}}_*$ estimation should nevertheless yield consistent estimates of the signs, ratios of coefficients, and normalized magnitudes (c.f. appendix A).[50]

[49] Wooldridge (2002) considers Chamberlain's random effects estimator in this context. That estimator is largely identical to the approach of Van Praag (2015). It merely adds a further random effect that is assumed to be uncorrelated with $\boldsymbol{X_{i*}}$ (when conditioning on $\bar{\boldsymbol{X}}_{\boldsymbol{i}}$). Since this addition is more complicated to implement and only presents an efficiency gain, I do not pursue this approach here.

[50] The proportionality factor depends on our choice of the assumed standard deviation σ and the true standard deviation σ^* of ϵ_{it}. In ordered probit we consistently estimate $\ddot{\boldsymbol{\beta}}_*\sigma/\sigma^*$ (Wooldridge 2002). The value of σ^* positively depends on $\mathrm{Var}(v_{it})$ and $\mathrm{Var}(u_i)$, but negatively on $\mathrm{Cov}(\bar{X}_{it}, u_i)$. When $\mathrm{Cov}(\bar{X}_{it}, u_i) = 0$, then $\mathrm{Var}(\epsilon_{it}) = \mathrm{Var}(v_{it}) + \mathrm{Var}(u_i)$. Thus, $SD(v_{it}) \leq \sigma^* \leq \sqrt{\mathrm{Var}(v_{it}) + \mathrm{Var}(u_i)}$. If we choose $\sigma = 1$ and $\mathrm{SD}(v_{it}) = 1$, we will get estimates with negatively biased magnitudes by a factor $1/\sigma^*$. Estimates of the thresholds will be biased by the same factor $1/\sigma^*$ as the coefficients on latent satisfaction. Hence magnitudes of coefficients that are standardized by the width of the response scale (see appendix A) are not affected by the proportionality factor.

D Monte-Carlo study

This small Monte-Carlo study evaluates how well the proposed approach of using memories works at identifying variables' separate effects on latent satisfaction and scale use in the presence of fixed effects.

D.1 Experimental design

The set-up is as follows. I set $N = 5{,}000$ and $T = 6$. With respect to T, this is close to the BHPS data used in the main text. N is somewhat smaller than what is available in BHPS to reduce computing time. Latent satisfaction is generated by:

$$s_{it} = \beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it} + u_i + v_{it}, \tag{A9}$$

where $u_i \sim N(0,1)$ and $v_{it} \sim N(0,1)$. Furthermore, $X_{1,it} = X_{1,it-1} + e_{1,it}$ for $t > 1$ and $X_{1,it} = u_i$ for $t = 1$, with $e_{1,it} \sim N(0,1)$. $X_{2,it} \sim N(u_i, 1)$. Finally, $X_{3,it} = 1$ if $u_i + e_{2,it} > 0$ and $X_{3,it} = 0$ otherwise, with $e_{2,it} \sim N(0,1)$. Thus, each independent variable is correlated with the fixed effect u_i and is strongly autocorrelated (as would be the case for most demographic variables). $X_{1,it}$ and $X_{2,it}$ are continuous, while $X_{3,it}$ is a dummy. I further set $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5$. I evaluate two different processes by which the cutoffs $\tau_{it,k}$ are generated to determine r_{it} (which in turn is generated as in equation (1)). As a first case, I generate $\tau_{it,k} = -7 + \bar{s} + 2k$, where $\bar{s}$ is the average of s_{it} across i and t. As a second case, I generate $\tau_{it,k} = -2.75 + \delta_1 X_{1,it} + \delta_2 X_{2,it} + \delta_3 X_{3,it} + \bar{s} + 1.25k$. With $\delta_1 = 0.2$, $\delta_2 = 0.8$, $\delta_3 = 0.8$. In both cases, the number of response categories K is set to 7. Both cases yield equidistant cutoffs and roughly even shares across each of the seven response categories for r_{it}. However, in the first case all cutoffs are independent of $\boldsymbol{X}_{it}$, while in the latter case cutoffs strongly depend on $\boldsymbol{X}_{it}$. Memories of satisfaction are generated by $s_{it}^{(m)} = \alpha_1 X_{1,it-1} + \alpha_2 X_{2,it-1} + \alpha_3 X_{3,it-1} + u_i + v_{it}^{(m)}$, with $\alpha_1 = 0.8$, $\alpha_2 = 0.4$, $\alpha_3 = 0.6$ and $v_{it}^{(m)} \sim N(0,1)$.

I draw 250 independent sets of observations and estimate six different models on each set. Those models are:

- An OLS regression of r_{it} ('OLS'),
- An OLS regression of demeaned r_{it} ('OLS-FE'),
- A standard ordered probit regression ('OP'),
- An ordered probit regression with individual averages for each independent variable added (as discussed in Appendix C; 'OP-Avg.')
- An ordered probit regression with individual averages, extended by an equation for memories of past satisfaction (as discussed in section 3, 'MOP-Avg.').
- An ordered probit regression with individual averages, extended by an equation for memories of past satisfaction and using data for memories as available for the UK (as discussed in section 3; 'UKMOP-Avg.')

For both MOP-Avg. and the UKMOP-Avg. model, I add individual averages to the memory and the cutoff equations as well as the equation on latent satisfaction.

D.2 Results in case of cutoffs being independent of X_{it}

Table D1 presents mean estimates of coefficients and their ratios from 250 independent trials. Absolute magnitudes of coefficients are affected by arbitrary scaling choices for σ (also see Van Praag and Ferrer-i-Carbonell 2008). As motivated in Appendix A, I therefore present estimates of normalized coefficients on latent satisfaction. These are given by $\text{Norm.}\beta = 6\beta/(\gamma_6 - \gamma_1)$. True values are $\text{Norm.}\beta_1 = 0.6$ and $\text{Norm.}\beta_2 = \text{Norm.}\beta_3 = 0.3$. With respect to these targets, the OLS and OP models perform worst overall. However, interestingly, the OLS-FE model is furthest from the target value for the normalized coefficient for β_1. This is despite the fact cutoffs are specified to be equidistant, which is normally taken to motivate the use of OLS regressions (see section 2.1), but which is in line with footnote 6 of the main text. In contrast, the OP-Avg. and

MOP-Avg. models both perform very well. The UKMOP-Avg. also performs much better than OLS-FE, though coefficients seem somewhat biased towards zero.

Table D1. Results from Monte-Carlo simulation with cutoffs independent of X_{it}

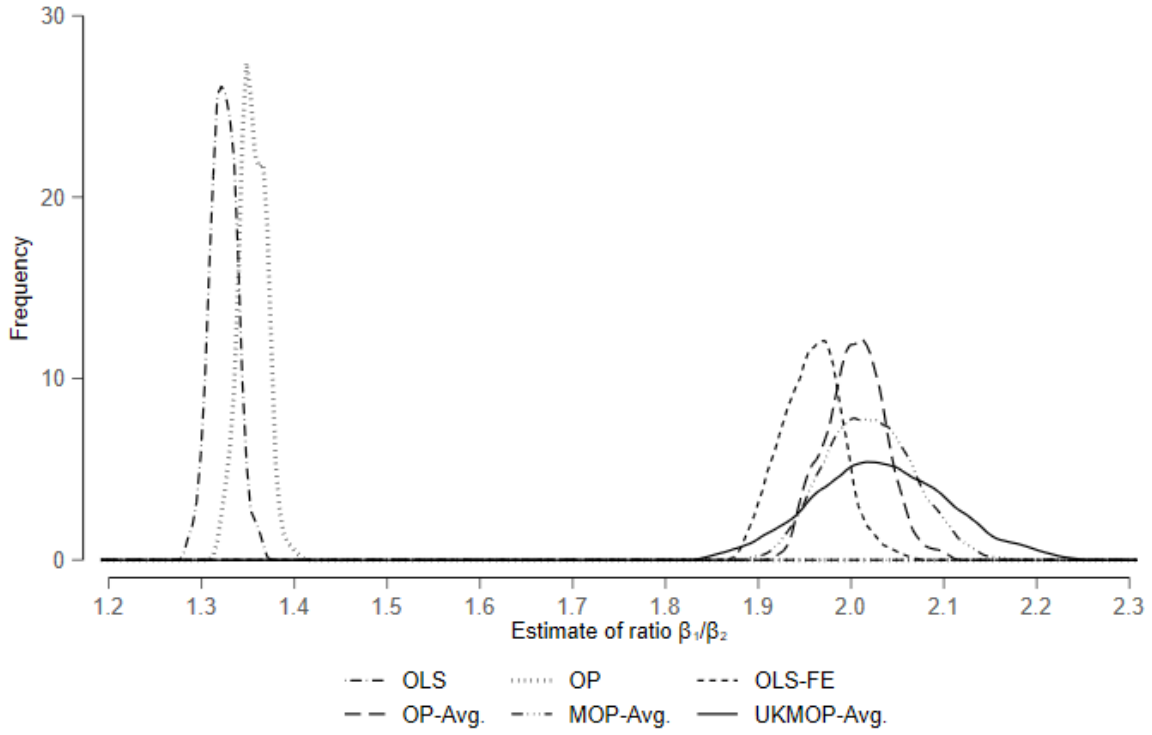
Parameter estimates	(1) OLS	(2) OP	(3) OLS-FE	(4) OP-Avg.	(5) MOP-Avg.	(6) UKMOP-Avg.
β_1	0.526	0.987	0.459	0.958	0.963	0.937
β_2	0.398	0.729	0.235	0.478	0.478	0.464
β_3	0.542	0.926	0.244	0.477	0.479	0.466
β_1/β_2	1.323 (0.677)	1.354 (0.646)	1.957 (0.054)	2.003 (0.032)	2.015 (0.054)	2.024 (0.081)
β_1/β_3	0.971 (1.029)	1.066 (0.934)	1.884 (0.136)	2.011 (0.076)	2.019 (0.118)	2.022 (0.151)
β_2/β_3	0.734 (0.266)	0.787 (0.213)	0.963 (0.054)	1.004 (0.041)	1.002 (0.060)	1.000 (0.080)
$\mathrm{Norm.}\,\beta_1$	0.526 (0.074)	0.687 (0.087)	0.459 (0.141)	0.602 (0.005)	0.605 (0.010)	0.559 (0.043)
$\mathrm{Norm.}\,\beta_2$	0.398 (0.098)	0.507 (0.207)	0.235 (0.065)	0.301 (0.005)	0.301 (0.007)	0.278 (0.026)
$\mathrm{Norm.}\,\beta_3$	0.542 (0.242)	0.645 (0.345)	0.244 (0.057)	0.300 (0.011)	0.301 (0.017)	0.278 (0.030)
α_1					0.781	0.767
α_2					0.383	0.376
α_3					0.576	0.557
α_1/α_2					2.041 (0.061)	2.045 (0.102)
α_1/α_3					1.358 (0.056)	1.383 (0.105)
α_2/α_3					0.665 (0.026)	0.678 (0.053)
δ_1					0.029	-0.045
δ_2					0.000	-0.039
δ_3					0.002	-0.036
δ_1/δ_2					-14.849	0.756
δ_1/δ_3					-35.011	1.302
δ_2/δ_3					-13.251	1.333

Note: $T = 6$ and $N = 5000$. Numbers show mean estimates of coefficients and their ratios from 250 independent trials. Root mean squared error (RMSE) shown in parentheses. True values for each parameter are: $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \alpha_1 = 0.8, \alpha_2 = 0.4, \alpha_3 = 0.6, \delta_1 = 0, \delta_2 = 0, \delta_3 = 0$. For ratios of δ the RMSE is not defined due to division by zero. For OP regressions '$\mathrm{Norm.}\,\beta$' give normalized coefficients, which are given by $(K-1)\beta/(\gamma_{K-1} - \gamma_1) = 6\beta/(\gamma_6 - \gamma_1)$. For OLS regressions, '$\mathrm{Norm.}\,\beta$' is identical to β since r is rank-order coded. True values are $\mathrm{Norm.}\,\beta_1 = 0.6, \mathrm{Norm.}\,\beta_2 = \mathrm{Norm.}\,\beta_3 = 0.3$.

Moreover, from the experimental setup, it follows that the true ratios β_1/β_2 and β_1/β_3 are equal to 2, while the ratio β_2/β_3 equals 1. With respect to these target values, the OLS regressions in column (1) perform worst in terms of both the average prediction and the root mean squared error (RMSE). Simple ordered probit regressions shown in column (2) perform almost as badly, with all ratios of coefficients being far below the target. All other models perform rather similar and quite well. Of these, the OLS-FE models shows the worst average estimate of each ratio of coefficients, with a slight negative bias compared to the true value.

Hence, and encouragingly, the ordered probit model with individual averages added (see column (4)) performs best with almost no bias in either direction and low RMSE. Thus, the argument of Appendix C, namely that accounting for fixed effects in an ordered probit setting is possible,

Figure D1. Density plots of estimates of β_1/β_2 when scale use is independent of $\boldsymbol{X_{it}}$



Note: The OP-Avg., the MOP-Avg. and UKMOP-Avg. model give estimates which are centred close to the true value (=2), while the OLS-FE model is somewhat negatively biased, but with little variance. The OLS and OP model also have little variance but yield a strongly negatively biased ratio of coefficients.

appears to be correct. Models which are extended by an equation for memories of past satisfaction ('MOP' and 'UKMOP') perform almost as well as the OP-Avg. model, with the average estimates being close to the true values. However, RMSEs are larger than for the OP-Avg. model because these models estimate the $\boldsymbol{\beta}$ coefficients using one fewer wave. On top, by explicitly modelling dependence of scale use on $\boldsymbol{X_{it}}$,[51] as well as memories of past satisfaction, these models estimate three times as many free parameters as the other models, leading to a loss in efficiency. One reason why the UKMOP model may perform worse than the MOP model is because each observation of $r_{it}^{(m)}$ in that model carries much less information than $r_{it}^{(m)}$ in the MOP model (cf. section 3.2). The location of the cutoffs is thus less precisely estimated. This can also be seen in the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\delta}$, which are somewhat less precise in the UKMOP model compared to the MOP model.

[51] Although the coefficients are explicitly modelled, in the present example all true coefficients relating to scale use are zero.

D.3 Results in case of cutoffs depending on X_{it}

I now turn to the case in which higher values of each independent variable makes scale use more stringent. The assumption of common scale use is thus violated. Table D2 presents results from the same 250 trials as in table D1, with only the data-generating process for the cutoffs altered. Because true differences between thresholds are now smaller, the true values for normalized coefficients are given by $\text{Norm.}\,\beta_1 = 0.96$, $\text{Norm.}\,\beta_2 = \text{Norm.}\,\beta_3 = 0.48$.

With respect to both target values for ratios and normalized magnitudes, the OLS and OP model continue to be strongly biased. When considering estimated ratios of coefficients, the OLS-FE and OP-Avg. model are now similarly biased.[52] When considering normalized magnitudes, these models are even more biased than the simple OLS and OP models. Importantly, since the true effects of $X_{2,it}$ and $X_{3,it}$ on scale use are stronger than on latent satisfaction (and of the same sign), these models fail to yield the correct signs for these coefficient estimates. In contrast, the MOP-Avg. and the UKMOP-Avg. models continue to yield correct coefficient ratios for each β coefficient.[53] Indeed, these estimates of ratios are almost unaffected by the change in how cutoffs are determined. Moreover, MOP and UKMOP models also yield largely correct normalized magnitudes of coefficients, though the coefficient on β_1 estimated by the UKMOP model seems to be underestimated slightly. Finally, ratios δ_1/δ_2 and δ_1/δ_3, which indicate relative shifts in scale use, appear positively biased for the MOP model, and negatively biased for the UKMOP model, though in the latter case that bias is negligible. The ratio δ_2/δ_3 is well estimated by both models. Signs of coefficients are always correctly estimated by both the MOP and the UKMOP model.

[52] However, on average the OP-Avg. provides the correct estimate for the ratio β_1/β_2. As flagged in section 2.32., this is because the special case $\beta_2 = \beta_3$ and $\delta_2 = \delta_3$ obtains here.

[53] And, on average, ratios of coefficients for memories of satisfaction are also estimated correctly. RMSEs for the UKMOP model are even improved compared to Table D1.

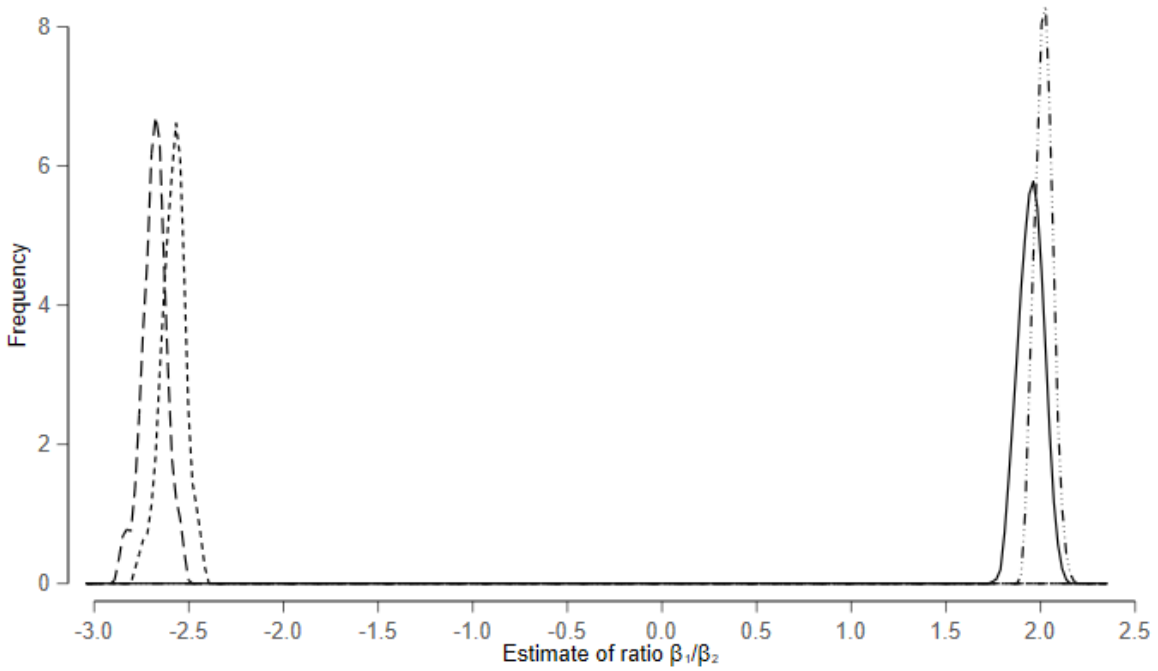
Table D2. Results from Monte-Carlo simulation with cutoffs depending on values of X_{it}

Parameter estimates	(1) OLS	(2) OP	(3) OLS-FE	(4) OP-Avg.	(5) MOP-Avg.	(6) UKMOP-Avg.
β_1	0.672	0.817	0.564	0.767	0.963	0.937
β_2	0.039	0.039	-0.219	-0.286	0.478	0.464
β_3	0.221	0.233	-0.227	-0.289	0.480	0.466
β_1/β_2	17.48 (15.66)	21.68 (19.98)	-2.582 (4.582)	-2.684 (4.685)	2.016 (0.058)	2.024 (0.081)
β_1/β_3	3.054 (1.068)	3.515 (1.529)	-2.491 (4.493)	-2.666 (4.668)	2.014 (0.112)	2.022 (0.151)
β_2/β_3	0.178 (0.823)	0.166 (0.834)	0.965 (0.069)	0.994 (0.061)	1.000 (0.057)	1.000 (0.080)
$\mathrm{Norm.}\,\beta_1$	0.672 (0.288)	0.910 (0.051)	0.564 (0.396)	0.772 (0.189)	0.967 (0.014)	0.929 (0.036)
$\mathrm{Norm.}\,\beta_2$	0.039 (0.441)	0.043 (0.437)	-0.219 (0.699)	-0.288 (0.768)	0.480 (0.012)	0.479 (0.016)
$\mathrm{Norm.}\,\beta_3$	0.221 (0.260)	0.260 (0.221)	-0.227(0.707)	-0.290 (0.770)	0.482 (0.026)	0.475 (0.035)
α_1					0.781	0.802
α_2					0.382	0.392
α_3					0.574	0.580
α_1/α_2					2.042 (0.062)	2.047 (0.092)
α_1/α_3					1.362 (0.055)	1.388 (0.099)
α_2/α_3					0.667 (0.024)	0.678 (0.044)
δ_1					0.22	0.193
δ_2					0.764	0.798
δ_3					0.769	0.798
δ_1/δ_2					0.288 (0.040)	0.242 (0.022)
δ_1/δ_3					0.287 (0.039)	0.243 (0.024)
δ_2/δ_3					0.995 (0.024)	1.002 (0.044)

Note: $T = 6$ and $N = 5000$. Numbers show mean estimates of coefficients and their ratios from 250 independent trials. Root mean squared error (RMSE) shown in parentheses. True values for each parameter are: $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \alpha_1 = 0.8, \alpha_2 = 0.4, \alpha_3 = 0.6, \delta_1 = 0.2, \delta_2 = 0.8, \delta_3 = 0.8$. For OP regressions '$\mathrm{Norm.}\,\beta$' give normalized coefficients, which are given by $(K-1)\beta/(\gamma_{K-1}-\gamma_1) = 6\beta/(\gamma_6-\gamma_1)$. For OLS regressions, '$\mathrm{Norm.}\,\beta$' is identical to β since r is rank-order coded. True values are $\mathrm{Norm.}\,\beta_1 = 0.96, \mathrm{Norm.}\,\beta_2 = \mathrm{Norm.}\,\beta_3 = 0.48$.

I draw three conclusions. First, the approach of adding individual averages to an ordered probit model in order to account for individual fixed effects works well and appears to yield consistent estimates of the normalized coefficients and trade-off ratios. Remarkably, the OLS-FE model, which is standard in empirical happiness research, performs worse than the OP-Avg. model. This is despite the fact that cutoffs are equidistant, which is an assumption normally maintained in order to motivate the OLS approach. Second, in cases where independent variables affect scale use, both the OP-Avg. and OLS-FE model yield biased estimates. Third, the MOP and UKMOP models appear to yield close to consistent estimates whether or not scale use is affected by any of the independent variables. However, the MOP and even more so the UKMOP model are less efficient than the OLS-FE and the OP-Avg. model.

Figure D2. Density plots of estimates of β_1/β_2 when scale use depends on $\boldsymbol{X_{it}}$



Note: The MOP-Avg. and UKMOP-Avg. model continue to give estimates which are centred close to the true value (=2). The OLS-FE and OP-Avg. model now yield a strongly negatively biased estimate. The OP and OLS model are omitted from this figure.

E Additional results tables

Table E1. Robustness of effect on latent satisfaction

	Estimate	Standard error	p-value
ln(HH income)	0.048***	0.011	0.000
Employed (base)			
Self-employed	0.036	0.031	0.235
Unemployed	-0.308***	0.035	0.000
Retired	0.096***	0.037	0.009
Family care	-0.024	0.028	0.380
Lt. sick / disabled	-0.271***	0.041	0.000
In education	0.028	0.031	0.376
Married (base)			
Widowed	-0.503***	0.069	0.000
Divorced	-0.185***	0.034	0.000
Never married	-0.143***	0.032	0.000
ln(HH size)	-0.108***	0.024	0.000
Childbirth	0.112***	0.022	0.000
Reference income	-0.110	0.093	0.238

Note: Test computed under assumption that $\lambda_j^{(m)}$ is not larger than β_j and of opposite sign. * p<0.1, ** p<0.05, *** p<0.01.

Table E2. Estimates for fixed reference groups (with lags and leads)

	No lags no leads (1)	One-year lag in current satisfaction, two-year lags in memories (2)	One-year lead in current satisfaction and one-year lead in scale use (3)	One-year lag and lead in current satisfaction, two-year lags in memories, one-year lead in scale use (4)
		Current life satisfaction		
		Contemporaneous effects		
ln(HH income)	0.094*** (0.023)	0.103*** (0.026)	0.107*** (0.027)	0.115*** (0.031)
Employed (base)				
Self-employed	0.102 (0.063)	0.131* (0.072)	0.055 (0.074)	0.052 (0.085)
Unemployed	-0.615*** (0.072)	-0.697*** (0.084)	-0.592*** (0.084)	-0.657*** (0.099)
Retired	0.169** (0.078)	0.177** (0.086)	0.140 (0.090)	0.138 (0.102)
Family care	-0.070 (0.057)	-0.101 (0.065)	-0.097 (0.066)	-0.160** (0.074)
Lt. sick / disabled	-0.543*** (0.085)	-0.598*** (0.099)	-0.591*** (0.097)	-0.664*** (0.112)
In education	0.012 (0.066)	-0.068 (0.079)	-0.046 (0.077)	-0.130 (0.093)
Married (base)				
Widowed	-1.038*** (0.141)	-1.219*** (0.162)	-0.711*** (0.182)	-0.908*** (0.207)
Divorced	-0.337*** (0.072)	-0.448*** (0.083)	-0.090 (0.089)	-0.252** (0.099)
Never married	-0.329*** (0.066)	-0.410*** (0.078)	-0.182** (0.083)	-0.290*** (0.097)
ln(HH size)	-0.184*** (0.051)	-0.142** (0.057)	-0.177*** (0.063)	-0.153** (0.072)
Childbirth	0.209*** (0.046)	0.229*** (0.054)	0.239*** (0.052)	0.295*** (0.064)
Reference income	0.065 (0.231)	0.246 (0.262)	0.014 (0.280)	0.141 (0.317)
Reference employed	-1.211* (0.641)	-0.995 (0.754)	-0.528 (0.783)	-0.229 (0.918)
Reference unemployed	-0.237 (0.816)	-0.853 (0.979)	-0.359 (0.997)	-0.171 (1.202)
Reference retired	0.459 (0.659)	-0.353 (0.779)	-0.064 (0.907)	-0.503 (1.050)
Reference % disabled	0.512 (0.936)	-0.204 (1.079)	2.264** (1.095)	1.407 (1.246)
Reference % married	0.290 (0.440)	0.223 (0.531)	-0.495 (0.599)	-0.855 (0.702)
Reference % divorced	0.138 (0.846)	0.408 (0.988)	-0.426 (1.029)	-0.587 (1.207)
Reference % widowed	-1.378 (1.115)	-1.418 (1.272)	-1.979 (1.534)	-2.421 (1.657)
Reference % birth	-0.249 (0.651)	0.390 (0.794)	-0.263 (0.779)	0.561 (0.956)
Reference ln(HHsize)	-0.056 (0.336)	-0.085 (0.416)	0.615 (0.459)	0.437 (0.538)
Reference ln(Job hours)	0.163 (0.204)	-0.033 (0.239)	0.099 (0.241)	0.005 (0.282)
Reference life satisfaction	0.659*** (0.125)	0.610*** (0.148)	0.632*** (0.144)	0.468*** (0.173)
		1-year lagged effects		
ln(HH income)		0.006 (0.009)		-0.000 (0.010)
Employed (base)				
Self-employed		-0.033 (0.025)		-0.033 (0.028)
Unemployed		0.009 (0.031)		-0.027 (0.035)
Retired		-0.013 (0.027)		-0.023 (0.030)
Family care		0.000 (0.023)		-0.003 (0.026)
Lt. sick / disabled		-0.041 (0.035)		-0.064 (0.039)
In education		0.058** (0.029)		0.065** (0.033)
Married (base)		0.000 (.)		0.000 (.)
Widowed		0.216*** (0.049)		0.258*** (0.055)
Divorced		0.191*** (0.033)		0.193*** (0.036)
Never married		0.106*** (0.028)		0.087*** (0.031)
ln(HH size)		-0.075*** (0.021)		-0.053** (0.024)
Childbirth		-0.001 (0.018)		0.011 (0.020)
Reference income		-0.171** (0.087)		-0.116 (0.099)
Reference employed		-0.294 (0.268)		-0.104 (0.323)
Reference unemployed		0.328 (0.327)		0.476 (0.378)
Reference retired		0.034 (0.218)		0.164 (0.245)
Reference % disabled		-0.194 (0.338)		0.090 (0.393)
Reference % married		0.192 (0.184)		0.210 (0.223)
Reference % divorced		-0.282 (0.320)		-0.316 (0.364)
Reference % widowed		-0.443 (0.324)		-0.235 (0.371)
Reference % birth		-0.094 (0.272)		-0.063 (0.316)
Reference ln(HHsize)		-0.159 (0.134)		-0.169 (0.162)
Reference ln(Job hours)		0.058 (0.081)		0.020 (0.095)

Reference life satisfaction		-0.074* (0.039)		-0.079* (0.045)
		1-year leaded effects		
ln(HH income)			0.006 (0.025)	-0.003 (0.029)
Employed (base)				
Self-employed			-0.033 (0.073)	-0.001 (0.082)
Unemployed			-0.062 (0.083)	-0.061 (0.098)
Retired			0.052 (0.089)	0.035 (0.097)
Family care			-0.081 (0.067)	0.015 (0.077)
Lt. sick /disabled			-0.309*** (0.097)	-0.331*** (0.109)
In education			-0.003 (0.079)	0.033 (0.095)
Married (base)				
Widowed			-0.543*** (0.163)	-0.524*** (0.189)
Divorced			-0.421*** (0.082)	-0.483*** (0.097)
Never married			-0.287*** (0.079)	-0.264*** (0.093)
ln(HH size)			-0.014 (0.056)	-0.063 (0.064)
Childbirth			0.234*** (0.057)	0.280*** (0.066)
Reference income			0.055 (0.269)	-0.027 (0.317)
Reference employed			-0.453 (0.765)	0.027 (0.902)
Reference unemployed			0.526 (1.003)	-1.078 (1.206)
Reference retired			0.746 (0.768)	0.822 (0.909)
Reference % disabled			-2.460** (1.089)	-2.613** (1.270)
Reference % married			0.791 (0.536)	0.779 (0.626)
Reference % divorced			0.023 (0.968)	0.890 (1.113)
Reference % widowed			0.309 (1.336)	1.228 (1.503)
Reference % birth			0.448 (0.797)	1.109 (0.951)
Reference ln(HHsize)			-0.405 (0.432)	-0.199 (0.518)
Reference ln(Job hours)			0.098 (0.239)	-0.135 (0.277)
Reference life satisfaction			-0.046 (0.132)	0.067 (0.152)
		Memory of last year's life satisfaction		
		1-year lagged effects		
ln(HH income)	0.020 (0.021)	0.027 (0.024)	0.032 (0.024)	0.041 (0.027)
Employed (base)				
Self-employed	-0.035 (0.061)	-0.021 (0.071)	0.070 (0.068)	0.094 (0.078)
Unemployed	-0.117* (0.065)	-0.080 (0.076)	-0.125 (0.077)	-0.092 (0.090)
Retired	-0.149* (0.078)	-0.157* (0.090)	-0.056 (0.089)	-0.059 (0.102)
Family care	-0.119** (0.055)	-0.096 (0.062)	-0.084 (0.062)	-0.064 (0.072)
Lt. sick /disabled	-0.324*** (0.081)	-0.257*** (0.092)	-0.303*** (0.091)	-0.226** (0.104)
In education	-0.011 (0.059)	-0.018 (0.071)	0.007 (0.070)	-0.017 (0.084)
Married (base)				
Widowed	-0.985*** (0.137)	-0.503*** (0.163)	-1.072*** (0.168)	-0.672*** (0.198)
Divorced	-0.398*** (0.065)	-0.374*** (0.081)	-0.372*** (0.075)	-0.401*** (0.090)
Never married	-0.314*** (0.059)	-0.287*** (0.075)	-0.302*** (0.069)	-0.343*** (0.086)
ln(HH size)	-0.024 (0.047)	-0.096* (0.058)	-0.045 (0.053)	-0.139** (0.063)
Childbirth	0.095** (0.041)	0.133*** (0.048)	0.066 (0.045)	0.078 (0.055)
Reference income	0.277 (0.214)	-0.104 (0.269)	0.012 (0.246)	-0.377 (0.304)
Reference employed	-0.566 (0.606)	-0.733 (0.765)	-0.375 (0.718)	-0.106 (0.904)
Reference unemployed	-0.049 (0.735)	-0.626 (0.921)	-0.097 (0.855)	-0.198 (1.046)
Reference retired	1.091* (0.638)	0.136 (0.876)	1.522** (0.714)	1.244 (0.958)
Reference % disabled	-0.141 (0.854)	-1.377 (1.054)	-0.387 (0.983)	-1.604 (1.218)
Reference % married	-0.443 (0.424)	-0.527 (0.583)	-0.424 (0.520)	-0.438 (0.701)
Reference % divorced	-0.360 (0.779)	-0.033 (0.959)	-0.635 (0.897)	-0.507 (1.103)
Reference % widowed	-0.856 (1.073)	-0.577 (1.388)	-0.909 (1.289)	-0.577 (1.596)
Reference % birth	0.242 (0.615)	0.173 (0.762)	0.419 (0.735)	0.419 (0.893)
Reference ln(HHsize)	0.204 (0.334)	-0.005 (0.431)	0.146 (0.399)	-0.060 (0.504)
Reference ln(Job hours)	0.206 (0.187)	0.247 (0.237)	0.273 (0.219)	0.281 (0.275)
Reference life satisfaction	0.073 (0.107)	0.045 (0.121)	0.087 (0.122)	-0.005 (0.138)
		2-year lagged effects		
ln(HH income)		0.044* (0.024)		0.042 (0.027)
Employed (base)				

Self-employed		0.155** (0.069)		0.120 (0.076)
Unemployed		-0.071 (0.074)		-0.064 (0.084)
Retired		0.145 (0.090)		0.120 (0.100)
Family care		-0.000 (0.061)		0.039 (0.068)
Lt. sick /disabled		-0.094 (0.095)		-0.011 (0.106)
In education		0.052 (0.071)		0.057 (0.081)
Married (base)		0.000 (.)		0.000 (.)
Widowed		-0.780*** (0.152)		-0.775*** (0.172)
Divorced		-0.091 (0.074)		-0.105 (0.087)
Never married		-0.041 (0.067)		0.018 (0.079)
ln(HH size)		0.098* (0.051)		0.051 (0.058)
Childbirth		-0.008 (0.045)		-0.036 (0.050)
Reference income		-0.020 (0.261)		-0.154 (0.290)
Reference employed		1.001 (0.739)		0.610 (0.870)
Reference unemployed		-1.195 (0.887)		-1.809* (0.999)
Reference retired		-0.171 (0.716)		-0.769 (0.822)
Reference % disabled		-0.328 (0.969)		-1.105 (1.098)
Reference % married		0.105 (0.522)		0.137 (0.604)
Reference % divorced		-1.544* (0.923)		-1.728 (1.058)
Reference % widowed		-1.798 (1.279)		-1.710 (1.490)
Reference % birth		-0.293 (0.748)		0.112 (0.866)
Reference ln(HHsize)		-0.119 (0.398)		-0.326 (0.456)
Reference ln(Job hours)		-0.275 (0.226)		-0.265 (0.267)
Reference life satisfaction		-0.163 (0.126)		-0.166 (0.140)
Constant	-0.664 (2.032)	2.485 (2.475)	0.274 (2.434)	1.139 (2.874)
		Scale shift		
		contemporaneous effects		
ln(HH income)	0.062*** (0.022)	0.079*** (0.025)	0.075*** (0.026)	0.087*** (0.029)
Employed (base)				
Self-employed	0.068 (0.058)	0.094 (0.066)	0.057 (0.069)	0.050 (0.079)
Unemployed	-0.414*** (0.067)	-0.503*** (0.078)	-0.377*** (0.078)	-0.454*** (0.092)
Retired	0.154** (0.075)	0.135 (0.084)	0.141 (0.088)	0.094 (0.100)
Family care	-0.056 (0.053)	-0.085 (0.061)	-0.066 (0.062)	-0.129* (0.069)
Lt. sick /disabled	-0.207*** (0.080)	-0.266*** (0.093)	-0.260*** (0.093)	-0.352*** (0.108)
In education	-0.052 (0.061)	-0.108 (0.073)	-0.117 (0.073)	-0.194** (0.086)
Married (base)				
Widowed	-0.727*** (0.136)	-0.702*** (0.156)	-0.599*** (0.180)	-0.564*** (0.202)
Divorced	-0.077 (0.064)	-0.063 (0.072)	0.046 (0.080)	0.049 (0.088)
Never married	-0.173*** (0.062)	-0.175** (0.072)	-0.102 (0.078)	-0.147* (0.089)
ln(HH size)	-0.114** (0.049)	-0.108** (0.054)	-0.120** (0.061)	-0.107 (0.069)
Childbirth	0.134*** (0.044)	0.137*** (0.051)	0.141*** (0.051)	0.157*** (0.061)
Reference income	0.074 (0.223)	0.205 (0.253)	-0.030 (0.271)	0.068 (0.308)
Reference employed	-0.547 (0.608)	-0.257 (0.715)	0.066 (0.746)	0.496 (0.877)
Reference unemployed	-0.338 (0.779)	-0.537 (0.937)	-0.503 (0.963)	0.094 (1.158)
Reference retired	0.526 (0.645)	0.079 (0.760)	0.039 (0.895)	0.119 (1.033)
Reference % disabled	0.544 (0.896)	0.078 (1.034)	2.196** (1.064)	1.839 (1.202)
Reference % married	0.291 (0.420)	0.266 (0.498)	-0.669 (0.572)	-0.863 (0.672)
Reference % divorced	-0.096 (0.809)	0.292 (0.942)	-0.949 (0.992)	-0.709 (1.168)
Reference % widowed	-0.830 (1.096)	-1.019 (1.248)	-1.803 (1.532)	-2.066 (1.651)
Reference % birth	-0.117 (0.613)	0.568 (0.742)	-0.220 (0.739)	0.888 (0.897)
Reference ln(HHsize)	0.078 (0.329)	0.008 (0.401)	0.636 (0.448)	0.527 (0.523)
Reference ln(Job hours)	0.089 (0.194)	-0.076 (0.227)	0.046 (0.229)	-0.003 (0.269)
Reference life satisfaction	0.023 (0.121)	0.008 (0.143)	0.012 (0.140)	-0.120 (0.167)
		1-year leaded effects		
ln(HH income)			0.010 (0.024)	-0.003 (0.027)
Employed (base)				
Self-employed			-0.052 (0.070)	-0.021 (0.079)
Unemployed			0.029 (0.079)	0.037 (0.092)
Retired			0.040 (0.087)	0.049 (0.095)
Family care			-0.063 (0.063)	0.023 (0.073)

Lt. sick / disabled			-0.154* (0.092)	-0.196* (0.104)
In education			0.006 (0.075)	0.044 (0.088)
Married (base)				
Widowed			-0.239 (0.162)	-0.243 (0.188)
Divorced			-0.174** (0.076)	-0.251*** (0.088)
Never married			-0.160** (0.075)	-0.124 (0.087)
ln(HH size)			-0.007 (0.053)	-0.050 (0.061)
Childbirth			0.132** (0.054)	0.165*** (0.062)
Reference income			0.002 (0.001)	0.000 (0.001)
Reference employed			-0.242 (0.734)	0.118 (0.864)
Reference unemployed			0.148 (0.977)	-1.342 (1.139)
Reference retired			0.831 (0.750)	0.704 (0.881)
Reference % disabled			-2.231** (1.057)	-2.685** (1.221)
Reference % married			0.816 (0.515)	0.714 (0.598)
Reference % divorced			0.045 (0.939)	0.528 (1.074)
Reference % widowed			0.592 (1.335)	1.162 (1.500)
Reference % birth			0.898 (0.767)	1.555* (0.900)
Reference ln(HHsize)			-0.291 (0.419)	-0.300 (0.496)
Reference ln(Job hours)			0.060 (0.228)	-0.144 (0.263)
Reference life satisfaction			-0.069 (0.131)	0.078 (0.150)
		Cutoff constants		
γ_1	2.212** (1.035)	3.190** (1.316)	2.451** (1.214)	3.412** (1.458)
γ_2	-0.873*** (0.025)	-0.865*** (0.028)	-0.876*** (0.028)	-0.872*** (0.031)
γ_3	-0.590*** (0.015)	-0.577*** (0.016)	-0.581*** (0.016)	-0.573*** (0.018)
γ_4	-0.439*** (0.010)	-0.435*** (0.011)	-0.434*** (0.011)	-0.426*** (0.012)
γ_5	-0.122*** (0.007)	-0.113*** (0.008)	-0.123*** (0.008)	-0.113*** (0.008)
γ_6	0.063*** (0.008)	0.080*** (0.008)	0.068*** (0.008)	0.085*** (0.009)
Observations	111,683	90,280	91,631	73,654

Note: All percentages divided by 100. Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E3. Sparse estimates for fixed reference group (with lags and leads)

	No lags no leads (1)	One-year lag in current satisfaction, two-year lags in memories (2)	One-year lead in current satisfaction and one-year lead in scale use (3)	One-year lag and lead in current satisfaction, two-year lags in memories, one-year lead in scale use (4)
		Current life satisfaction		
		Contemporaneous effects		
ln(HH income)	0.094*** (0.023)	0.103*** (0.026)	0.106*** (0.027)	0.115*** (0.031)
Employed (base)				
Self-employed	0.100 (0.063)	0.128* (0.071)	0.052 (0.074)	0.053 (0.084)
Unemployed	-0.613*** (0.072)	-0.695*** (0.084)	-0.594*** (0.084)	-0.658*** (0.099)
Retired	0.177** (0.077)	0.183** (0.086)	0.133 (0.090)	0.133 (0.101)
Family care	-0.070 (0.057)	-0.099 (0.065)	-0.093 (0.066)	-0.160** (0.074)
Lt. sick / disabled	-0.539*** (0.085)	-0.601*** (0.098)	-0.588*** (0.097)	-0.663*** (0.112)
In education	0.017 (0.065)	-0.064 (0.078)	-0.018 (0.077)	-0.111 (0.092)
Married (base)				
Widowed	-1.041*** (0.141)	-1.221*** (0.163)	-0.705*** (0.183)	-0.888*** (0.207)
Divorced	-0.338*** (0.072)	-0.442*** (0.082)	-0.088 (0.088)	-0.248** (0.099)
Never married	-0.331*** (0.066)	-0.411*** (0.077)	-0.183** (0.084)	-0.289*** (0.098)
ln(HH size)	-0.187*** (0.050)	-0.144** (0.057)	-0.181*** (0.063)	-0.155** (0.072)
Childbirth	0.211*** (0.046)	0.228*** (0.054)	0.241*** (0.052)	0.294*** (0.064)
Reference income	0.139 (0.217)	0.256 (0.246)	0.051 (0.261)	0.122 (0.292)
Reference employed	-0.599** (0.284)	-0.776** (0.344)	-0.225 (0.380)	-0.281 (0.456)
Reference % disabled	0.691 (0.792)	0.338 (0.925)	2.552** (0.992)	1.846* (1.114)
Reference % widowed	-1.573 (1.067)	-1.591 (1.206)	-1.662 (1.444)	-1.981 (1.556)
Reference % birth	0.261 (0.573)	0.500 (0.729)	0.059 (0.694)	0.478 (0.877)
Reference life satisfaction	0.690*** (0.123)	0.627*** (0.146)	0.622*** (0.142)	0.456*** (0.170)
		1-year lagged effects		
ln(HH income)		0.006 (0.009)		-0.000 (0.010)
Employed (base)				
Self-employed		-0.031 (0.025)		-0.033 (0.028)
Unemployed		0.010 (0.030)		-0.025 (0.035)
Retired		-0.014 (0.027)		-0.024 (0.030)
Family care		0.002 (0.023)		-0.002 (0.026)
Lt. sick / disabled		-0.039 (0.034)		-0.061 (0.039)
In education		0.059** (0.028)		0.061* (0.033)
Married (base)				
Widowed		0.214*** (0.049)		0.260*** (0.055)
Divorced		0.186*** (0.033)		0.190*** (0.036)
Never married		0.105*** (0.028)		0.085*** (0.031)
ln(HH size)		-0.079*** (0.021)		-0.054** (0.023)
Childbirth		0.001 (0.018)		0.012 (0.020)
Reference income		-0.127 (0.080)		-0.109 (0.092)
Reference employed		0.016 (0.121)		0.063 (0.142)
Reference % disabled		0.056 (0.273)		0.167 (0.329)
Reference % widowed		-0.490* (0.286)		-0.205 (0.315)
Reference % birth		-0.005 (0.237)		0.052 (0.282)
Reference life satisfaction		-0.057 (0.039)		-0.065 (0.044)
		1-year leaded effects		
ln(HH income)			0.007 (0.025)	-0.002 (0.029)
Employed (base)				
Self-employed			-0.030 (0.072)	0.006 (0.082)
Unemployed			-0.059 (0.083)	-0.068 (0.098)
Retired			0.050 (0.088)	0.036 (0.096)
Family care			-0.082 (0.067)	0.015 (0.077)
Lt. sick / disabled			-0.300*** (0.097)	-0.324*** (0.109)
In education			0.010 (0.078)	0.043 (0.093)

Married (base)				
Widowed			-0.549*** (0.164)	-0.529*** (0.190)
Divorced			-0.419*** (0.082)	-0.475*** (0.097)
Never married			-0.289*** (0.079)	-0.271*** (0.093)
ln(HH size)			-0.004 (0.055)	-0.052 (0.064)
Childbirth			0.229*** (0.057)	0.270*** (0.067)
Reference income			0.086 (0.255)	-0.056 (0.306)
Reference employed			-0.048 (0.346)	-0.223 (0.421)
Reference % disabled			-2.536*** (0.943)	-2.893*** (1.097)
Reference % widowed			-0.027 (1.260)	0.682 (1.431)
Reference % birth			0.197 (0.722)	0.548 (0.859)
Reference life satisfaction			-0.022 (0.130)	0.054 (0.148)
Memory of last year's life satisfaction				
		1-year lagged effects		
ln(HH income)	0.019 (0.021)	0.027 (0.024)	0.033 (0.024)	0.042 (0.027)
Employed (base)				
Self-employed	-0.034 (0.061)	-0.020 (0.071)	0.065 (0.068)	0.096 (0.078)
Unemployed	-0.120* (0.065)	-0.082 (0.076)	-0.134* (0.077)	-0.096 (0.091)
Retired	-0.132* (0.077)	-0.151* (0.089)	-0.027 (0.088)	-0.042 (0.101)
Family care	-0.119** (0.055)	-0.095 (0.062)	-0.086 (0.062)	-0.066 (0.071)
Lt. sick / disabled	-0.320*** (0.080)	-0.252*** (0.092)	-0.301*** (0.091)	-0.223** (0.104)
In education	-0.014 (0.058)	-0.022 (0.070)	-0.005 (0.068)	-0.040 (0.082)
Married (base)				
Widowed	-0.993*** (0.137)	-0.511*** (0.163)	-1.088*** (0.169)	-0.667*** (0.199)
Divorced	-0.398*** (0.065)	-0.377*** (0.081)	-0.378*** (0.075)	-0.410*** (0.090)
Never married	-0.305*** (0.059)	-0.280*** (0.075)	-0.296*** (0.069)	-0.337*** (0.086)
ln(HH size)	-0.026 (0.047)	-0.103* (0.058)	-0.052 (0.054)	-0.147** (0.063)
Childbirth	0.096** (0.041)	0.135*** (0.048)	0.069 (0.045)	0.086 (0.055)
Reference income	0.350* (0.198)	0.049 (0.250)	0.147 (0.229)	-0.186 (0.284)
Reference employed	-0.382 (0.282)	-0.305 (0.366)	0.024 (0.324)	0.302 (0.419)
Reference % disabled	-0.793 (0.747)	-1.517 (0.949)	-1.201 (0.872)	-2.113* (1.102)
Reference % widowed	-0.562 (1.023)	-0.222 (1.297)	-0.413 (1.181)	-0.041 (1.460)
Reference % birth	-0.343 (0.527)	-0.105 (0.657)	-0.108 (0.652)	0.170 (0.789)
Reference life satisfaction	0.077 (0.105)	0.057 (0.119)	0.114 (0.119)	0.068 (0.134)
		2-year lagged effects		
ln(HH income)				
Employed (base)				
Self-employed		0.153** (0.068)		0.116 (0.076)
Unemployed		-0.076 (0.074)		-0.073 (0.083)
Retired		0.139 (0.089)		0.116 (0.099)
Family care		-0.001 (0.061)		0.038 (0.068)
Lt. sick / disabled		-0.095 (0.095)		-0.013 (0.106)
In education		0.049 (0.069)		0.048 (0.080)
Married (base)				
Widowed		-0.785*** (0.152)		-0.791*** (0.173)
Divorced		-0.090 (0.074)		-0.106 (0.087)
Never married		-0.032 (0.067)		0.024 (0.078)
ln(HH size)		0.103** (0.051)		0.050 (0.058)
Childbirth		-0.008 (0.045)		-0.033 (0.050)
Reference income		0.071 (0.238)		-0.050 (0.264)
Reference employed		0.350 (0.312)		0.265 (0.359)
Reference % disabled		-0.082 (0.864)		-0.594 (0.977)
Reference % widowed		-1.747 (1.217)		-1.911 (1.399)
Reference % birth		-0.812 (0.665)		-0.096 (0.755)
Reference life satisfaction		-0.109 (0.123)		-0.077 (0.138)
Constant	0.691 (1.314)	1.134 (1.641)	0.165 (1.563)	-0.699 (1.936)

Scale shift				
contemporaneous effects				
ln(HH income)	0.061*** (0.022)	0.079*** (0.025)	0.073*** (0.026)	0.087*** (0.029)
Employed (base)				
Self-employed	0.064 (0.058)	0.090 (0.066)	0.053 (0.069)	0.050 (0.078)
Unemployed	-0.414*** (0.067)	-0.502*** (0.078)	-0.382*** (0.078)	-0.455*** (0.093)
Retired	0.162** (0.074)	0.146* (0.083)	0.136 (0.087)	0.096 (0.099)
Family care	-0.056 (0.053)	-0.084 (0.061)	-0.064 (0.062)	-0.129* (0.069)
Lt. sick / disabled	-0.203** (0.080)	-0.269*** (0.093)	-0.257*** (0.093)	-0.349*** (0.107)
In education	-0.044 (0.060)	-0.109 (0.072)	-0.090 (0.072)	-0.179** (0.085)
Married (base)				
Widowed	-0.728*** (0.137)	-0.701*** (0.157)	-0.593*** (0.181)	-0.544*** (0.202)
Divorced	-0.076 (0.063)	-0.056 (0.072)	0.048 (0.080)	0.053 (0.088)
Never married	-0.174*** (0.062)	-0.176** (0.072)	-0.102 (0.079)	-0.146 (0.089)
ln(HH size)	-0.112** (0.048)	-0.107** (0.054)	-0.123** (0.061)	-0.108 (0.069)
Childbirth	0.135*** (0.044)	0.134*** (0.051)	0.142*** (0.051)	0.154** (0.061)
Reference income	0.109 (0.209)	0.164 (0.237)	0.011 (0.253)	0.038 (0.283)
Reference employed	-0.192 (0.275)	-0.313 (0.329)	0.106 (0.366)	0.191 (0.438)
Reference % disabled	0.451 (0.776)	0.166 (0.895)	2.218** (0.980)	1.805* (1.087)
Reference % widowed	-0.974 (1.053)	-1.213 (1.191)	-1.299 (1.450)	-1.633 (1.557)
Reference % birth	0.435 (0.543)	0.620 (0.689)	0.047 (0.665)	0.570 (0.836)
Reference life satisfaction	0.051 (0.119)	0.020 (0.141)	0.006 (0.138)	-0.131 (0.165)
1-year leaded effects				
ln(HH income)			0.011 (0.024)	-0.002 (0.027)
Employed (base)				
Self-employed			-0.049 (0.069)	-0.013 (0.079)
Unemployed			0.029 (0.079)	0.031 (0.092)
Retired			0.041 (0.086)	0.051 (0.094)
Family care			-0.065 (0.063)	0.024 (0.073)
Lt. sick / disabled			-0.144 (0.092)	-0.190* (0.104)
In education			0.019 (0.073)	0.055 (0.087)
Married (base)				
Widowed			-0.246 (0.163)	-0.250 (0.189)
Divorced			-0.172** (0.076)	-0.243*** (0.088)
Never married			-0.162** (0.075)	-0.128 (0.088)
ln(HH size)			0.004 (0.053)	-0.039 (0.061)
Childbirth			0.127** (0.054)	0.156** (0.063)
Reference income			0.058 (0.246)	-0.093 (0.292)
Reference employed			-0.005 (0.333)	-0.133 (0.404)
Reference % disabled			-2.551*** (0.929)	-3.008*** (1.070)
Reference % widowed			0.210 (1.267)	0.753 (1.440)
Reference % birth			0.615 (0.694)	0.848 (0.812)
Reference life satisfaction			-0.040 (0.128)	0.071 (0.146)
Cutoff constants				
γ_1	0.262 (0.652)	0.870 (0.835)	0.484 (0.737)	1.358 (0.946)
γ_2	-0.873*** (0.025)	-0.866*** (0.028)	-0.876*** (0.028)	-0.872*** (0.031)
γ_3	-0.590*** (0.015)	-0.577*** (0.016)	-0.582*** (0.016)	-0.573*** (0.018)
γ_4	-0.439*** (0.010)	-0.435*** (0.011)	-0.434*** (0.011)	-0.426*** (0.012)
γ_5	-0.122*** (0.007)	-0.114*** (0.008)	-0.123*** (0.008)	-0.114*** (0.008)
γ_6	0.062*** (0.008)	0.079*** (0.008)	0.067*** (0.008)	0.084*** (0.009)
Observations	111,683	90,280	91,631	73,654

Note: All percentages divided by 100. Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E4. Inclusion of reference satisfaction variables

	Average	Each individually
	(1)	(2)
	Current life satisfaction	
ln(HH income)	0.055** (0.024)	0.056** (0.026)
Employed (base)		
Self-employed	0.085 (0.064)	0.076 (0.064)
Unemployed	-0.467*** (0.073)	-0.408*** (0.079)
Retired	0.175** (0.076)	0.187** (0.077)
Family care	-0.042 (0.057)	-0.030 (0.058)
Lt. sick / disabled	-0.239*** (0.088)	-0.121 (0.103)
In education	0.060 (0.065)	0.069 (0.065)
Married (base)		
Widowed	-0.930*** (0.143)	-0.923*** (0.145)
Divorced	-0.160** (0.073)	-0.126 (0.081)
Never married	-0.243*** (0.067)	-0.224*** (0.069)
ln(HH size)	-0.213*** (0.051)	-0.234*** (0.051)
Childbirth	0.179*** (0.047)	0.253*** (0.050)
Reference income	0.037 (0.196)	0.044 (0.198)
Ref. sat. income quintile		0.226*** (0.051)
Ref. sat. employment		0.325*** (0.051)
Ref. sat. marital status		0.259*** (0.058)
Ref. sat childbirth		-0.172 (0.108)
Ref. sat HH size		0.340*** (0.060)
Avg. reference satisfaction	1.127*** (0.088)	
	Memory of last year's life satisfaction	
ln(HH income)	0.010 (0.021)	0.005 (0.023)
Employed (base)		
Self-employed	-0.041 (0.061)	-0.046 (0.062)
Unemployed	-0.093 (0.066)	-0.044 (0.071)
Retired	-0.120 (0.076)	-0.118 (0.077)
Family care	-0.116** (0.055)	-0.108* (0.055)
Lt. sick / disabled	-0.294*** (0.083)	-0.203** (0.094)
In education	-0.006 (0.058)	-0.002 (0.059)
Married (base)		
Widowed	-0.985*** (0.139)	-1.012*** (0.140)
Divorced	-0.363*** (0.066)	-0.378*** (0.072)
Never married	-0.291*** (0.060)	-0.301*** (0.062)
ln(HH size)	-0.023 (0.048)	-0.020 (0.049)
Childbirth	0.086** (0.041)	0.101** (0.044)
Reference income	0.238 (0.182)	0.269 (0.183)
Ref. sat. income quintile		0.064 (0.046)
Ref. sat. employment		0.108*** (0.040)
Ref. sat. marital status		0.011 (0.051)
Ref. sat childbirth		-0.047 (0.093)
Ref. sat HH size		-0.002 (0.053)
Avg. reference satisfaction	0.169** (0.075)	
Constant	3.518*** (0.945)	5.327*** (0.974)
	Scale shift	
ln(HH income)	0.063*** (0.023)	0.065** (0.025)
Employed (base)		
Self-employed	0.061 (0.059)	0.062 (0.059)
Unemployed	-0.418*** (0.069)	-0.420*** (0.075)
Retired	0.164** (0.074)	0.162** (0.075)
Family care	-0.051 (0.054)	-0.050 (0.054)
Lt. sick / disabled	-0.207** (0.083)	-0.213** (0.098)
In education	-0.033 (0.061)	-0.032 (0.061)
Married (base)		
Widowed	-0.741*** (0.138)	-0.773*** (0.140)
Divorced	-0.088 (0.065)	-0.133* (0.073)
Never married	-0.179*** (0.063)	-0.195*** (0.065)
ln(HH size)	-0.114** (0.049)	-0.125** (0.050)

Childbirth	0.143*** (0.045)	0.148*** (0.049)
Reference income	0.064 (0.188)	0.067 (0.189)
Ref. sat. income quintile		-0.023 (0.050)
Ref. sat. employment		-0.019 (0.050)
Ref. sat. marital status		-0.112** (0.054)
Ref. sat childbirth		-0.017 (0.110)
Ref. sat HH size		0.081 (0.057)
Avg. reference satisfaction	-0.068 (0.086)	
γ_1	0.984** (0.418)	3.117*** (0.428)
γ_2	-0.851*** (0.025)	-0.829*** (0.025)
γ_3	-0.569*** (0.015)	-0.551*** (0.015)
γ_4	-0.421*** (0.010)	-0.406*** (0.010)
γ_5	-0.107*** (0.007)	-0.096*** (0.007)
γ_6	0.075*** (0.007)	0.084*** (0.007)
Observations	118, 440	118, 440

Note: 'Ref. sat. income' measures mean reported life satisfaction of those in the pre-defined reference group (see section 4.1) who are additionally in the same quintile of the income distribution. Likewise, 'Ref. sat. marital status' measures mean reported life satisfaction of those in the predefined reference group who additionally share the same marital status. Variables 'Ref. sat. employment', 'Ref. sat childbirth', and 'Ref. sat HH size' are defined analogously. Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E5. Allowing for non-parallel shifts in scale use when adding individual averages (non-linear specification)

	Current sat. (1)	Memory of sat. (2)	Cutoff $\tau_{it,1}$ (3)	Cutoff $\tau_{it,2}$ (4)	Cutoff $\tau_{it,3}$ (5)	Cutoff $\tau_{it,4}$ (6)	Cutoff $\tau_{it,5}$ (7)	Cutoff $\tau_{it,6}$ (8)
ln(HH income)	0.102***	0.027	0.090***	-0.050*	-0.025	0.018	-0.003	0.015
	(0.023)	(0.021)	(0.030)	(0.026)	(0.025)	(0.019)	(0.013)	(0.013)
Employed (base)								
Self-employed	0.082	-0.001	0.068	-0.079	0.027	-0.007	0.050	0.024
	(0.065)	(0.062)	(0.124)	(0.175)	(0.125)	(0.071)	(0.040)	(0.036)
Unemployed	-0.631***	-0.134**	-0.327***	0.098	0.125	-0.301***	-0.177***	-0.057
	(0.069)	(0.064)	(0.125)	(0.198)	(0.124)	(0.074)	(0.051)	(0.049)
Retired	0.159**	-0.080	0.134	0.136	-0.074	-0.028	-0.080**	-0.000
	(0.072)	(0.071)	(0.100)	(0.109)	(0.086)	(0.060)	(0.040)	(0.034)
Family care	-0.072	-0.113**	-0.020	0.052	0.061	-0.070	-0.137***	-0.015
	(0.056)	(0.054)	(0.081)	(0.088)	(0.081)	(0.055)	(0.037)	(0.031)
Lt. sick	-0.558***	-0.301***	-0.238*	0.381**	-0.034	-0.201***	-0.180***	-0.138**
/ disabled	(0.080)	(0.075)	(0.133)	(0.189)	(0.124)	(0.077)	(0.063)	(0.063)
In education	0.038	-0.022	0.103	-0.203	-0.069	0.038	-0.100**	0.057
	(0.065)	(0.058)	(0.117)	(0.203)	(0.142)	(0.073)	(0.042)	(0.037)
Married (base)								
Widowed	-0.910***	-0.912***	-0.659***	-0.165	0.139	0.114	0.017	-0.083
	(0.125)	(0.121)	(0.163)	(0.197)	(0.212)	(0.105)	(0.067)	(0.065)
Divorced	-0.371***	-0.427***	0.041	-0.352**	0.061	0.115	-0.010	-0.091*
	(0.070)	(0.065)	(0.124)	(0.158)	(0.146)	(0.072)	(0.048)	(0.048)
Never married	-0.303***	-0.296***	-0.164	0.104	-0.107	0.065	0.000	-0.046
	(0.067)	(0.060)	(0.117)	(0.183)	(0.128)	(0.067)	(0.038)	(0.038)
ln(HH size)	-0.219***	-0.003	-0.166**	0.073	-0.062	0.051	-0.006	-0.026
	(0.050)	(0.047)	(0.076)	(0.082)	(0.068)	(0.042)	(0.026)	(0.025)
Childbirth	0.220***	0.086**	0.199***	0.034	-0.045	-0.041	-0.054*	0.006
	(0.046)	(0.041)	(0.077)	(0.090)	(0.085)	(0.050)	(0.030)	(0.024)
Reference	-0.117	0.255	0.276	-0.288	-0.091	-0.063	-0.098	0.019
income	(0.188)	(0.175)	(0.242)	(0.227)	(0.237)	(0.152)	(0.096)	(0.088)
Constant		0.057	-1.188	-3.440***	-1.516**	1.094**	-1.968***	-6.816***
		(0.798)	(0.771)	(1.050)	(0.747)	(0.481)	(0.340)	(0.371)
Observations				118,440				

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E6. Allowing for non-parallel shifts in scale use when adding individual averages (non-linear specification, omitting first cut-off)

	Current sat. (1)	Memory of sat. (2)	Cutoff $\tau_{it,1}$ (3)	Cutoff $\tau_{it,2}$ (4)	Cutoff $\tau_{it,3}$ (5)	Cutoff $\tau_{it,4}$ (6)	Cutoff $\tau_{it,5}$ (7)	Cutoff $\tau_{it,6}$ (8)
ln(HH income)	0.009			-0.048*	-0.029	0.014	-0.001	0.019
	(0.021)			(0.025)	(0.024)	(0.018)	(0.013)	(0.013)
Employed (base)								
Self-employed	0.035			0.014	-0.003	-0.017	0.040	0.029
	(0.114)			(0.169)	(0.113)	(0.067)	(0.039)	(0.036)
Unemployed	-0.276***			0.092	0.145	-0.274***	-0.184***	-0.065
	(0.100)			(0.169)	(0.109)	(0.068)	(0.048)	(0.048)
Retired	0.013			0.101	-0.069	-0.041	-0.087**	0.002
	(0.072)			(0.100)	(0.080)	(0.056)	(0.039)	(0.034)
Family care	-0.052			0.047	0.055	-0.075	-0.139***	-0.016
	(0.064)			(0.082)	(0.075)	(0.051)	(0.035)	(0.031)
Lt. sick	-0.273***			0.448**	-0.031	-0.209***	-0.170***	-0.145**
/ disabled	(0.103)			(0.175)	(0.115)	(0.073)	(0.061)	(0.061)
In education	-0.076			-0.231	-0.065	0.033	-0.097**	0.054
	(0.099)			(0.175)	(0.125)	(0.069)	(0.040)	(0.036)
Married (base)								
Widowed	-0.231**			-0.209	0.196	0.125	0.010	-0.081
	(0.111)			(0.176)	(0.214)	(0.104)	(0.066)	(0.064)
Divorced	-0.414***			-0.335**	0.050	0.108	-0.001	-0.088*
	(0.106)			(0.155)	(0.136)	(0.066)	(0.046)	(0.047)
Never married	-0.161*			0.109	-0.138	0.048	-0.000	-0.049
	(0.096)			(0.156)	(0.116)	(0.062)	(0.037)	(0.037)
ln(HH size)	-0.029			0.101	-0.055	0.051	-0.008	-0.029
	(0.059)			(0.082)	(0.067)	(0.040)	(0.025)	(0.025)
Childbirth	-0.016			-0.014	-0.062	-0.037	-0.060**	0.004
	(0.062)			(0.083)	(0.079)	(0.048)	(0.028)	(0.024)
Reference	-0.385**			-0.276	-0.073	-0.076	-0.108	0.006
income	(0.167)			(0.203)	(0.225)	(0.144)	(0.093)	(0.087)
Constant			-2.068***	-3.608***	-1.552**	1.133**	-1.943***	-6.778***
			(0.776)	(0.991)	(0.710)	(0.471)	(0.334)	(0.370)
Observations				118,440				

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table E7. Allowing for non-parallel shifts in scale use when adding individual averages (linear specification, omitting first cut-off)

	Current sat.	Memory of sat.	Cutoff $\tau_{it,1}$	Cutoff $\tau_{it,2}$	Cutoff $\tau_{it,3}$	Cutoff $\tau_{it,4}$	Cutoff $\tau_{it,5}$	Cutoff $\tau_{it,6}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln(HH income)	0.010			-0.009	-0.032*	0.013	-0.006	0.019
	(0.031)			(0.026)	(0.018)	(0.013)	(0.011)	(0.012)
Employed (base)								
Self-employed	0.087			0.004	0.045	-0.004	0.025	0.032
	(0.109)			(0.097)	(0.061)	(0.041)	(0.036)	(0.042)
Unemployed	-0.331***			-0.003	0.056	-0.130***	-0.156***	-0.055
	(0.093)			(0.081)	(0.055)	(0.039)	(0.037)	(0.042)
Retired	0.039			0.081	-0.025	-0.031	-0.078**	-0.016
	(0.098)			(0.081)	(0.057)	(0.039)	(0.034)	(0.035)
Family care	-0.056			0.047	0.020	-0.052	-0.120***	-0.014
	(0.092)			(0.080)	(0.054)	(0.036)	(0.031)	(0.034)
Lt. sick	-0.402***			0.070	-0.021	-0.095**	-0.139***	-0.104**
/ disabled	(0.096)			(0.080)	(0.055)	(0.041)	(0.041)	(0.046)
In education	-0.069			-0.107	-0.023	0.011	-0.097***	0.053
	(0.106)			(0.088)	(0.061)	(0.042)	(0.036)	(0.044)
Married (base)								
Widowed	-0.334***			-0.055	-0.010	0.082	-0.002	-0.074
	(0.114)			(0.080)	(0.064)	(0.052)	(0.044)	(0.051)
Divorced	-0.259***			0.028	-0.047	0.066*	-0.025	-0.077
	(0.078)			(0.065)	(0.047)	(0.038)	(0.037)	(0.047)
Never married	-0.223**			0.003	-0.114**	0.039	-0.005	-0.061
	(0.093)			(0.080)	(0.056)	(0.040)	(0.034)	(0.042)
ln(HH size)	-0.123**			-0.032	-0.028	0.044*	-0.005	-0.037
	(0.062)			(0.052)	(0.036)	(0.026)	(0.022)	(0.026)
Childbirth	-0.032			-0.036	-0.027	-0.028	-0.054**	0.019
	(0.074)			(0.063)	(0.042)	(0.031)	(0.024)	(0.028)
Reference	-0.361			0.048	-0.062	-0.068	-0.126*	0.017
income	(0.251)			(0.206)	(0.130)	(0.090)	(0.076)	(0.088)
Constant			-2.324***	0.548***	0.526***	0.686***	0.839***	1.009***
			(0.097)	(0.061)	(0.039)	(0.031)	(0.028)	(0.038)
Observations				118,440				

Note: Clustered standard errors in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Conclusions

1 Main findings and contributions to the wider literature

Evidence on subjective wellbeing is increasingly entering efforts in public policy.[1] In order for these efforts to be successful, it is paramount that our evidence is sufficiently accurate. This thesis made contributions in two areas of this important endeavour. The first of these concerns the ways in which other people as well as respondents' past outcomes are crucial (or not) in determining individual life satisfaction. The second area is more theoretical. It is concerned with the ways in which people use response scales.

1.1 Interpersonal and intertemporal dynamics

Papers 1, 2, and 4 contribute to the literature on interpersonal and intertemporal dynamics in the determination of life satisfaction.

As noted in paper 1, the previous literature often assumed that incomes were subject to full adaptation, leading to near-zero effects of income in the long run. Full adaptation would imply that the effects of income are entirely relative to past internal reference standards. Paper 1 showed this to be false; incomes are not subject to full adaptation. Instead, previous evidence of adaptation was driven by omitted variable bias. Such omitted variable bias may also be relevant in other domains. This makes it plausible that adaptation is also less widespread in other domains. Paper 1 therefore showed that lags (or leads) of controls should be additionally included when studying adaptation (or anticipation) dynamics with respect to any explanatory variable of interest. Indeed, paper 4, which analyses effects of lags and leads of a broad set of covariates, only finds adaptation dynamics for self-employment and marital status. It finds no evidence of adaptation to income, unemployment, childbirth, or long-term sickness. Moreover, paper 4 theoretically argued that

[1] Prominent examples of that are New Zealand's "Wellbeing Budget" and the UK Treasury's "Green Book" (i.e. the Treasury's guide on policy evaluation) endorsing subjective wellbeing reports as an important source of evidence.

lagged effects on reported satisfaction, which are typically interpreted as indicating adaptation, may instead be driven by scale shift effects.

However, paper 2 showed that intergenerational class mobility matters to people's life satisfaction. In particular, when people move to a more advantaged social class, the mere fact of moving 'up' seems to improve life satisfaction, and vice versa when moving to a less advantaged social class. This suggests that internal reference standards do have some importance. However, paper 2 assumed that the internal reference standard is set by the attainment of the respondent's parents. In contrast, papers 1 and 4 assumed that the internal reference standard is set by respondents' attainment no more than three years in the past. Internal reference standards may therefore be set earlier in a respondent's life (i.e. in their childhood) than typically assumed. To assess this possibility, future work should carefully analyse how intergenerational mobility in income and other variables affect satisfaction.

Regarding interpersonal dynamics, paper 1 found that estimated reference effects were always sufficiently large in the long run so as to render the combined effects of own and reference income insignificant and relatively close to zero. This has the implication that a positive income shock that is common to all members of the population is expected to fail to increase average life satisfaction. Paper 2 extended this investigation of social reference effects to the case of social class. It turns out that the effects of others' intergenerational social mobility have opposite signs compared to the effects of respondents' own social mobility. Since this reference mobility effect has roughly the same magnitude as the effect of own mobility, it appears that when everybody is upwardly (downwardly) mobile, there is no direct societal benefit (harm) from mobility.[2] This is similar to

[2] Another descriptive contribution of the paper is the finding that members of the least advantaged social class in Western Europe still exhibited a higher level of mean life satisfaction than members of the most advantaged social class in Eastern Europe. In addition, being self-employed turned out to be a much more advantaged social class than was the case for Western Europe. This has the wider implication that standard collapses of the ESeC scheme, which place the self-employed in the middle of the class 'hierarchy' are likely misleading in the case of Eastern Europe.
Finally, a methodological contribution of paper 2 was to clarify the assumptions of Sobel's (1981) influential 'diagonal reference model' and to propose ways of testing the robustness of its assumptions. Making a Stata package for the estimation of diagonal models available to the research community is another methodological contribution.

the findings of paper 1. In contrast, paper 4 only finds very limited evidence in favour of social reference effects when further extending the analysis to the domains of employment, marital status, job hours, etc. The only consistently significant and sizable variable relating to reference effects was the mean reported satisfaction of those in the reference group, which entered with a positive coefficient.

1.2 Shape of the reporting function

The manner in which respondents report their satisfaction may be described by a reporting function (Oswald 2008). Papers 1 and 2 follow standard practice in the literature and assume that the reporting function is linear and common across respondents and across time. Adopting these assumptions entails that rank-order coded life satisfaction reports measure latent satisfaction in a cardinal and interpersonally comparable manner. Papers 3 and 4 contribute to the literature by exploring the implications of relaxing these assumptions.

In particular, paper 3 explores the assumption of a linear reporting scale. This assumption is required if satisfaction reports are to measure latent satisfaction cardinally. The paper was motivated by recent papers by Schröder and Yitzhaki (2017) and Bond and Lang (2019). They show that dropping the assumption of linear scale use may lead to reversals of signs of coefficients in regression models using satisfaction reports as dependent variables. In response, paper 3 contains both good and bad news for the empirical literature on life satisfaction. The good news is that the reversals of signs of coefficients as proposed by Schröder and Yitzhaki (2017) and Bond and Lang (2019) are shown to either be impossible or implausible for almost all variables of interest. The only case in which a reversal was both possible and plausible (that of additional children in a fixed effects regression), was one in which the estimated coefficient was small and insignificant in the baseline analysis and about which there is no consensus in the literature. The bad news is that relative magnitudes of coefficients are sensitive to even fairly mild transformations of the assumed underlying response scales. Although this cannot lead to reversals in the signs of estimated trade-

off ratios, it means that estimates of e.g. shadow prices are only identified within a range, and no point estimates can be given without further assumptions about scale use. Unfortunately, these ranges are large, which casts doubt on current the viability of the life satisfaction approach to valuing non-market goods (see e.g. Luechinger 2009).

The analyses of paper 3 have three immediate practical implications. First, when using OLS regressions of life satisfaction (or other subjective) data, future empirical work should assess whether reversals of signs of coefficients of explanatory variables of interest are possible. Second, if reversals are possible, future research should numerically assess whether reversals can occur within some *a priori* specified bounds for plausible non-linear scale use. In the paper we suggest respectively increasing and decreasing multiplicative ratios of 1.8 and 0.4 between response options as one such set of bounds. Third, using the same bounds, and if ratios of coefficients are of interest, future research should then also assess the robustness of reported ratios (and results that depend on these).

As a parallel and related endeavour, paper 4 explores the assumption of common scale use. This assumption too is required if satisfaction reports are to be treated as cardinal and interpersonally comparable. So far, possible scale shift effects (i.e. violations of common scale use) have either been assumed away (Ferrer-i-Carbonell and Frijters 2004), or have been used to discount the viability of life satisfaction research (Fleurbaey and Blanchet 2013; Viscusi 2020). However, the extent and probable direction of bias has not been analysed in the literature. The first task of the paper was therefore to show that violations of common scale use likely lead to an underestimation of the effects of contemporaneous values of personal characteristics. At the same time, adaptation and social reference effects are likely overestimated, while reinforcements dynamics are likely underestimated. This contribution may help future work to assess whether scale shifts are a serious concern.

As the main methodological contribution, paper 4 then proposed a method to detect and account for scale shifts. Using this method, together with a novel approach by Van Praag (2015),[3] the paper shows that scale shifts seem especially strong for unemployment and bereavement. Surprisingly, reference effects seem to play little to no role in scale shifts. Moreover, signs of coefficients are never reversed when accounting for scale shifts, and ratios of coefficients are not heavily affected. This is very good news for past and future research on subjective wellbeing.

2 Policy implications

If we believe that the goal of government is to maximize the expectation of average welfare in the population and if we can treat satisfaction reports as tracking welfare cardinally, then the findings of papers 1 and 2 in particular have some clear policy implications.

The omitted variable bias that was unearthed in paper 1, and which drove observed adaptation to income in previous studies was partly caused by reinforcement of a negative effect of household composition (i.e. that the negative effects of additional adults and children increased over time).[4] Since equivalence scales depend on welfare effects of household composition, paper 1 showed that equivalence scales computed on the basis of life satisfaction data are sensitive to such over-time dynamics. In particular, compared to the commonly used OECD equivalence scale, a scale computed on the basis of satisfaction data suggests greater economies of scale in the short-run, but, due to reinforcement of the household composition effect, much smaller economies of scale in the long-run. Therefore, and if we are convinced that equivalence scales should be computed on the basis of life satisfaction data, this finding has the policy implication that weights on adults and children in equivalence scales should be increased.

Moreover, in so far as paper 1 showed that the long run effect on life satisfaction of raising the income of all is near zero, another policy implication of the paper is that countries are not helped

[3] A further methodological contribution of the paper is the validation of Van Praag's (2015) approach to accounting for fixed effects via a simulation study and a comparison to the more standard OLS method using BHPS data.
[4] The other cause was the omission of a control variable for the employment status of a respondent's partner.

by pursuing growth *per se*. Countries should rather seek to improve the efficiency with which economic growth is converted into life satisfaction gains. Since incomes have convincingly been shown to be subject to strongly diminishing marginal wellbeing gains (Layard, Mayraz, and Nickell 2008), progressive redistribution may well lead to efficiency gains in this sense. However, in order for such welfare gains to obtain, the effects of income gains and income losses of satisfaction must be roughly symmetric in the long run. This was not shown in this thesis. However, section 3 of this conclusion makes some suggestions for future research in that direction.

Another policy implication to be taken from paper 1 is to prioritize goals that are different from economic growth. This is also suggested by the findings of paper 2. Indeed, paper 2 showed that social class is an important, and perhaps underappreciated determinant of life satisfaction. Strikingly, the association of social class with life satisfaction was estimated to be similar in strength to that of income. Moreover, no sizable reference effects relating to others' class position could be obtained. In contrast to the analogous case relating to incomes, shifting respondents to more advantaged class positions should therefore result in an increase in average wellbeing.[5]

Furthermore, paper 2 showed that respondent's social origins are a potentially important determinant of life satisfaction. Evidently, people have no control over their social origins. Thus, even if the goal of policy is not the maximization of (average) welfare, equity considerations may also motivate policy to minimize the negative life satisfaction effects emanating from social origins.

However, as stated at the beginning of this section, these policy implication rely, at least to some degree, on satisfaction reports measuring latent satisfaction in a cardinal and interpersonally comparable manner. Papers 3 and 4 showed that if these assumptions are violated, results based on satisfaction reports may not be able to generate evidence robust enough for clear policy advice.

[5] However, the conclusions of paper 2 are based on limited variation in societal class structures. *Prima facie*, it seems that much greater over-time variation in class structures (say, a complete shift towards membership in the salariat) may turn out to uncover reference effects of class position.

A negative policy implication of this thesis is therefore that governments should treat the estimates from regressions of the sort presented in papers 1 and 2 (as well as the wider literature) with some caution. Indeed, should governments wish to routinely perform cost-benefit analyses using satisfaction data, they may usefully supplement standard analyses with further bounding analyses on how extreme possible curvatures and shifts in the reporting function would have to be so as to affect policy appraisals. With the same motivation and as a more positive policy implication, governments may also have an interest to complement their standard data collection exercises with additional efforts designed to gather data that would allow us to routinely give tighter bounds on possible scale shifts and non-linear scale use. Some suggestions for questions generating such data are given in the next section.

3 Avenues for future research

The findings, policy implications, and methodological contributions of the empirical chapters suggest a number of avenues for future research. Sketches for potential research endeavours are given in this section. Suggestions for further research that directly follow from each individual paper are given first. A final section then suggests some further avenues for future research. Although these further avenues are informed by the present thesis, they do not directly follow from any particular chapter.

3.1 Further research following paper 1

One key limitation of paper 1 was the use of annual income data. If the adaptation process is completed within a year, adaptation cannot be observed with such data. One way of circumventing this issue with standard data (like GSOEP and UKHLS) is to recognize that we observe both a respondent's household income in the previous month, as well as her average monthly household income over the last 12 months. Both of these can be simultaneously entered into a regression model. If the estimated coefficient on the previous month's household income were to significantly

exceed the estimated coefficient for the average monthly household income, this may indicate a rapid adaptation process.

Another limitation of paper 1 was its failure to allow for asymmetries in the effects of income gains and losses. In the behavioural literature on choice, such asymmetries are often observed, with 'losses looming larger than gains' (Kahneman and Tversky 1979). More recently short-run asymmetries have also been observed in the domain of reported life satisfaction (Boyce, Wood, and Ferguson 2016; Carter and McBride 2013; De Neve et al. 2018; Vendrik and Woltjer 2007). This evidence is sometimes viewed as testing a hedonic version of prospect theory (Kanninen and Mahler 2017). It therefore appears pertinent to investigate whether such asymmetries persist in the long run. As pointed out in the previous section, evidence on such potential long run asymmetries could have important policy implications regarding the likely welfare effects of redistribution. One way of investigating long run asymmetries is to follow Bahmani-Oskooee et al. (2017) and Hovi & Laamanen (2017) and decompose incomes into partial sums of positive and negative changes since the beginning of the observation period, i.e. to define variable $y_{it}^{+} = \sum_{l=1}^{t} max\ (\Delta y_{il}, 0)$ and $y_{it}^{-} = \sum_{l=1}^{t} |min\ (\Delta y_{il}, 0)|$. Such variables, along with their lags, could then be entered into models that are otherwise identical to those used in paper 1.

If there are no long-run asymmetries, this would imply that choice and experience do not concur in the long run. However, if there are long run asymmetries, a respondent's level of life satisfaction is dependent on a person's entire history of income changes. This has the counterintuitive implication that two otherwise identical persons who have been on the same income for an arbitrarily long time could nevertheless feel different levels of life satisfaction. Figure 1 illustrates this point.[6]

[6] Other cases in which such "hysteresis" or "path-dependency" are invoked are over-time changes in unemployment rates (Blanchard and Summers 1986), habit formation (Vendrik 1993), and the stability of political institutions (Pierson 2000).

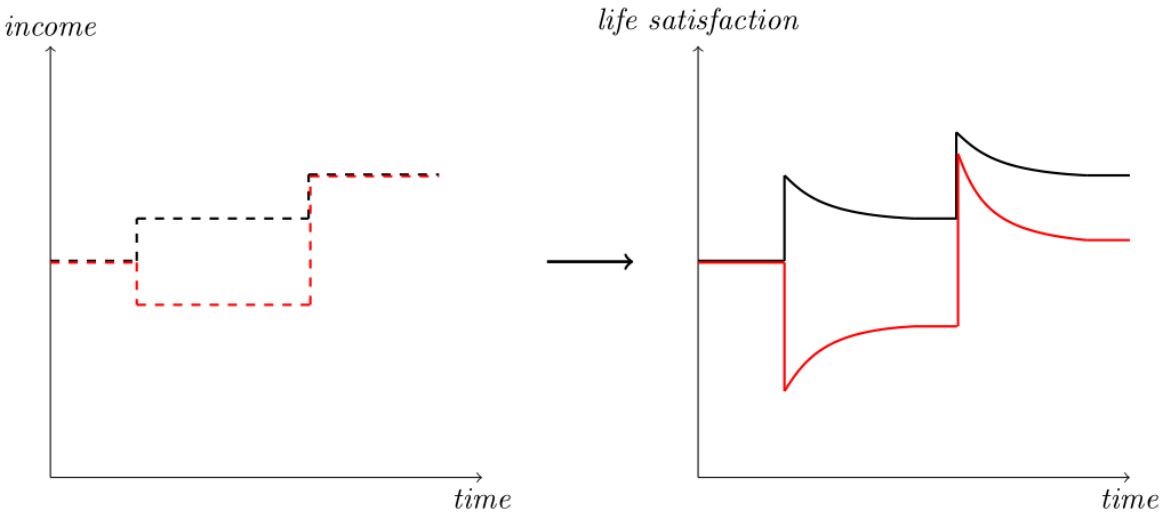


Figure 1. Illustration of long-run loss aversion. The left panel shows the income trajectory of two otherwise identical individuals. The person represented by black lines experiences two small gains, while the person represented by red lines experiences a small loss and a large gain. The right panel shows how life satisfaction reacts to income changes shown in the left panel. I assume short-run and long-run asymmetries, such that in both cases losses loom 1.5 times larger than gains. Despite equal incomes at the end of the period, the life satisfaction of the two individuals differs even in the long-run.

3.2 Further research following paper 2

Of all included chapters, paper 2 was the only one to use cross-sectional data (from the European Social Survey). Advantages of this choice of data were that a very large number of individuals were included, and that differences across Europe could be observed. However, an important disadvantage was that time-invariant heterogeneities could not be accounted for via individual fixed effects. Moreover, respondent's intrapersonal class mobility across the life course could not be observed. In order to address these concerns, future research may try to tackle similar empirical questions using panel data such as the German SOEP. In this context, one potentially fruitful idea may be to extend the diagonal reference model to three rather than two dimensions. Concretely, in the standard diagonal reference model, life satisfaction is given by the weighted sum of an origin, a destination, and an intergenerational mobility effect (see equation 2 in paper 2). Such a model

could be extended with further terms for previous social class as well as intragenerational mobility, allowing for the joint estimation of all these terms.[7]

However, one may take issue with the strong assumptions that diagonal reference models make. As noted in the paper, Fosse and Winship (2019) propose, in the context of age-period-cohort data, a method for partial identification using much weaker assumptions. Their method could be applied to social mobility data, which would allow us to arrive at more robust, albeit less precise conclusions. Moreover, one reason why paper 2 did not present disaggregated results for individual countries was due to numerical convergence issues. Since the method of Fosse and Winship (2019) relies on straightforward OLS regressions, it will not face this problem, thus allowing for such disaggregated analyses.

Another shortcoming of the paper was that its reference effects only measured variation in absolute class mobility. However, an important strand of the literature on class mobility is concerned with relative class mobility instead (Breen 2004; Bukodi and Goldthorpe 2019; Bukodi, Paskov, and Nolan 2019).[8] One problem with measures of relative class mobility, such as the UNIDIFF parameter in sociological log-linear models or global log odds ratios from mobility tables (see e.g. Bukodi, Paskov, and Nolan 2019) is that they cannot be readily computed at the individual level. However, the recently proposed entropy-based $\boldsymbol{M}$-index of Seiler and Jann (2019) allows for such decompositions, making it possible to investigate the effects of relative class mobility on individual life satisfaction. In so far as relative class mobility is a concept closely related to equality of

[7] To be more precise, one might estimate $LS_{it} = p_1\lambda_d dest_{it} + p_2\lambda_d dest_{it-1} + p_3\lambda_o orig_i + \gamma_m intermob_{it} + \gamma_n intramob_{it} + \beta' \boldsymbol{X_{it}} + \varepsilon_{it}$, where we impose the constraints that $\sum_{k=1}^{3} p_k = 1$ and $p_k \geq 0$ and where $intermob_{it} = dest_{it} - orig_i$ and $intramob_{it} = dest_{it} - dest_{it-1}$. Substantively, $p_2\lambda_d$ would give the effect of the previous social class (in that sense, an intragenerational origin effect), while $intermob_{it}$ and $intramob_{it}$ would respectively yield estimates of intergenerational and intragenerational mobility effects.

[8] The distinction between these concepts is the following. Absolute class mobility measures the proportion of people who attain a different social class than their parents. Relative class mobility measures how much more (or less) likely it is for a person with some particular class origin to attain some particular class destination than another person with a different class origin.

opportunity, which in turn has some policy relevance, understanding such effects seems to be an important research direction.

3.3 Further research following paper 3

As discussed in paper 3, the plausibility of sign reversals as well as the robustness of ratios of coefficients depend on our beliefs about respondent's scale use. Oddly, beyond the papers discussed in paper 3, little work has so far been done on this important issue. As noted in section 2, obtaining clearer and more precise evidence on non-linearities in scale use would make evidence based on satisfaction reports a significantly more powerful tool to generate policy advice.

To make progress on a theoretical level, it seems useful to generalize Van Praag's (1991) argument. His argument suggests that linear scale use is maximally informative. However, in the context of regressions of life satisfaction, the applicability of Van Praag's (1991) argument rests on assuming that respondents think of life satisfaction as being uniformly distributed. This is implausible. It should therefore be explored what ways of using response scales would be maximally informative when assuming different distributions, such as the normal or the log-normal distribution.[9]

On a more empirical level, a straightforward next step would be to directly ask respondents how they use life satisfaction scales. To do so one might show respondents a set of possible response scales, say a linear, and a multiplicative scale (e.g. like those shown in figures 3 and 4 in paper 3), and ask how they used the scale when answering the life satisfaction question. As noted in paper 3, a more indirect method is that of Oswald (2008), who asked respondents to rate their height on a bounded scale with extremes labelled 'Very short' and 'Very tall', and observed that these ratings mapped linearly onto their true cardinally measured height. A worry is that his finding is particular to height, or to variables that are distributed normally (like height). One could therefore ask respondents to rate themselves on a bounded scale against very many cardinally observable

[9] I now started to explore this issue, showing that scale use would indeed be roughly linear when assuming a normal distribution. A draft of this work is available upon request.

quantities, such age, income, weight, number of friends, etc., many of which are non-normally distributed, and see whether the linear mapping observed by Oswald (2008) continues to hold. If so, we would have strong reasons to believe that a linear response scale is also used in the case of life satisfaction in particular and subjective wellbeing more generally.[10]

Versions of the reversibility arguments discussed in paper 3 will also hold for multi-variable scales such as the 'Satisfaction with Life Scale' (Diener et al. 1985). Though less popular in sociology and economics, such scales are extremely popular in psychology, and used for nearly every construct of interest. Crucially, even if one were to only allow for common non-linearities in scale use across items in such scales, their aggregates will not necessarily yield scale with even ordinal content.[11] This is problematic. Exploring how the analyses of paper 3 extend to such multi-item scales would therefore be useful. Furthermore, a simplifying assumption in the analyses of paper 3 was to assume that covariates do not affect the expectation of latent satisfaction within response categories. In an expanded version online, we give some discussion of this issue. However, more progress is possible in this area.

On an empirical level, although the results show that reversals are impossible for most variables investigated with GSOEP, it is not clear whether this is also the case for other variables in other commonly used datasets. It thus looks useful to systematically replicate other findings in the literature and to see how robust these findings are.

[10] However, previous results in psychophysics indicate that subjective perceptions of physical stimuli like sound our light follow a logarithmic mapping (Shepard 1981).

[11] The following numerical example makes the point. Let the measure m_i be given by the sum of two sub-scales s_{1i} and s_{2i}. m_i is taken to measure the intensity of some latent construct c_i in persons i. Consider the following case. For respondent a we observe $s_{1a} = 2$ and $s_{2a} = 5$. For respondent b we observe $s_{1b} = 4$ and $s_{2b} = 4$. Thus $m_a = 7$ and $m_b = 8$. Since $m_b > m_a$, respondent b is measured to have a higher intensity on construct c than a. Now apply the non-linear transformation $\tilde{s}_{1i} = e^{s_{1i}}$ and $\tilde{s}_{2i} = e^{s_{2i}}$, with $\tilde{m}_i = \tilde{s}_{1i} + \tilde{s}_{2i}$ as a transformed measure of c. We obtain $\tilde{s}_{1a} = e^2 = 7.4$, $\tilde{s}_{2a} = e^5 = 148.4$, $\tilde{s}_{1b} = \tilde{s}_{2b} = e^4 = 54.6$. Thus, $\tilde{m}_a = 155.8$ and $\tilde{m}_b = 109.2$. Now it looks as though a scores higher than b. Hence, multi-item scales may not be ordinarily comparable if response scales are non-linear, even if the non-linearity is common across scales.

3.4 Further research following paper 4

Similar to concerns about non-linear scale use outlined in paper 3, concerns about scale shifts considered in paper 4 may undermine the potential of satisfaction reports to generate policy advice. In order to bolster the usefulness of life satisfaction research in guiding policymaking, it therefore seems important to generate more evidence on the likely extent of scale shifts in the population.

A natural next step is to apply the approach proposed in paper 4 to other panel datasets that include data on memories. GSOEP is one such dataset for life satisfaction. However, the data on memories there is in the form of pictures of trajectories in satisfaction over time. The proposed model would thus have to be adapted for such data. Another suitable dataset, albeit measuring subjective health, is the Whitehall II study, a long-running and large scale panel survey of British civil servants.[12] On a global level, the Gallup World Poll includes a question on how respondents rate themselves on the 'Cantril Ladder' five years in the past. Unfortunately, the World poll is not a panel dataset. Nevertheless, by aggregating responses within a country, one could, by appropriately adapting the memory approach, test for an average scale shift across time. Such country-wide over-time shifts in scale use may serve as another yet-overlooked explanation for the much-debated Easterlin Paradox.

Moreover, evidence generated from existing datasets may be supplemented with efforts to gather and analyse novel data. This would allow us to generate more methodologically robust findings. As explained in the main text of paper 4, there are several ways in which the method of paper 4 could be extended. First, following Howard et al. (1979), one could ask respondents to recall which response option they chose in the previous measurement occasion. If errors in the recalled response option are unrelated to contemporaneous covariates, the assumption of *present independence* is rendered more plausible. Second, one could ask about memories about multiple years in the past.

[12] See: https://www.ucl.ac.uk/epidemiology-health-care/research/epidemiology-and-public-health/research/whitehall-ii/data-collection.

One could then test the plausibility of *present independence* by adapting a version of d'Uva et al.'s (2011) test. An even more robust method is to cross-validate the proposed memory approach against the vignette approach by obtaining data on both from the same set of respondents. One could then formally test whether *present independence* and *vignette equivalence* hold jointly.

In addition, it seems that vignettes and memories could additionally be supplemented with questions about respondents' peers. For example, a question eliciting respondent's assessment of people's life satisfaction in their own country could also be used to anchor response scales.[13] Finally, the analyses of paper 3 (on linearity of scale use) and paper 4 (on scale shifts) were conducted independently. Future work might usefully combine these analyses.

3.5 Additional avenues for future research

Responses to questions relating to subjective wellbeing have the potential to become a key tool for policy analysis. However, in order to realize this potential, there are at least three further fundamental issues that ought to be more thoroughly considered. Those issues are: (1) Do questions about life satisfaction and related concepts really track welfare? (2) Can we locate the level at which welfare is zero? (3) What really explains reported satisfaction?

Regarding the first issue, recent evidence suggests people's choices and anticipated effects on life satisfaction (and SWB more widely) do not concur (Benjamin et al. 2014).[14] One interpretation of this evidence is that people do not choose what they think is best for them. This may be the case when respondents attempt to maximize others' welfare alongside their own. Another more worrying interpretation is that respondents do not make an exclusive report about their contemporaneous personal welfare when answering questions about their life satisfaction. Indeed,

[13] Such a question might go as follows: *'Using the same scale as you used to rate your own life satisfaction, what do you think is the share of people in [respondent's country] with a life satisfaction of* $[k]$*?'* Here, $[k]$ is any response category on the satisfaction scale.

[14] Findings in which choices and *experienced* subjective wellbeing do not concur could be explained by respondents mis-predicting their subjective wellbeing when making the choice. Such an explanation is not on offer in the case of *anticipated* subjective wellbeing.

another recent paper by Benjamin and colleagues (Benjamin et al. 2020) provides suggestive evidence that life satisfaction responses reflect a mixture of self-regarding and other-regarding concerns, as well as a mixture of current and life-time levels of welfare. In that case, life satisfaction responses (or other SWB responses) do not track welfare sufficiently well.

Consequently, an important avenue for future empirical research should be to find (sets of) question(s) that are more reliably interpreted by respondents as reflecting their judgement of their contemporaneous welfare level (in line with subjectivism, c.f. section 1 of this thesis' introduction). Given the worries raised in papers 3 and 4, such questions should be posed in a manner which makes scale shifts across respondents and over time unlikely, and which is explicit about the relative differences in welfare levels across response categories. *How* to generally pose such questions should also be an important area of future research. Recent work on quadratic voting applied to Likert scales may be useful in that endeavour (Cavaillé and Chen 2020). A parallel avenue for theoretical research is to give an adequate account of whether respondents' (hypothetical) choices or people's stated evaluations of their welfare (or something else entirely) is ultimately authoritative in assessing welfare levels. This seems to be both a normative and an epistemological question.

Regarding the second issue, note that even when we measure welfare cardinally and in an interpersonally comparable manner, we do not measure it on a ratio scale. To obtain such a ratio scale we additionally need to ascertain the level at which individual welfare is zero. We do not require such a ratio scale for policy analyses in which we can assume that the population is fixed.[15] However, most policy proposals will have some non-zero expected effects on who lives and how long people live. In those cases, we need a ratio scale. Consider this example: Which of the following options is better: (1) One person with a reported satisfaction of 9 out of 10 lives for one more year. (2) Two persons with reported satisfaction of 5 out of 10 living for one more year. Without a neutral level (and without cardinal measurement), the life satisfaction approach to policy

[15] And when assuming that the badness of a unit of negative welfare is equal in magnitude to the goodness of a unit of positive welfare.

analysis is silent on this type of choice.[16] To find the neutral level, Peasgood et al. (2018) propose to ask respondents which level of satisfaction they deem not worth living. Peasgood et al. (2018) find 2 out of 10 to be the average answer in their feasibility study. While a helpful start, much more work can be done on this topic. First, it would seem important to assess whether the neutral level varies with observables, particularly across age and the contemporaneous level of reported satisfaction. If this is the case, we may believe this to be evidence suggesting scale shifts and non-linear scale use. Second, it would help to clarify for how long respondents would require to be at a certain negative welfare level in order for them to deem their lives not worth living. Third, hypothetical assessments of the neutral level should best be supplemented with data on observed choice: conditional on levels of reported welfare, with what probability do people choose to end their lives?[17]

Finally, a well-known but rarely mentioned feature of research on life satisfaction (and subjective wellbeing more generally) is that objectively observable variables only explain very little of the observed variance in reported life satisfaction (Bellet and Frijters 2019). This implies that we may be systematically missing crucial determinants of satisfaction. If such determinants were manipulable, finding these determinants would have large policy implications. To make headway in this direction, it may be useful to shift focus. Instead of attempting to find causally interpretable estimates on latent satisfaction, we may rather seek to predict satisfaction reports as accurately as possible. Recent work by Bhushan (2020), who uses machine learning techniques to predict reported satisfaction, is a first step in this direction. Future work should build on this.

[16] Unfortunately (and surprisingly), there can exist no decision rule satisfying a set of intuitive desiderata in cases where population sizes vary, even if welfare is measured on a ratio scale (Arrhenius 2000). This is similar to Arrow's (1950) impossibility theorems pertaining to the case of ordinal preferences in a fixed population. Nevertheless, any reasonable decision rule will require welfare to be measured on a ratio scale (Broome 2004).

[17] For at least four reasons we should expect suicide decisions to provide a lower bound for the neutral level. First, people are likely loss-averse, particular concerning their life as a whole. Second, people may expect the future to be better than the present (as is often observed empirically, see Easterlin 2001). The present must therefore be sufficiently bad so as to offset future goods. Third, people may expect to inflict harm upon others by ending their lives, and therefore avoid suicide. Fourth, there may be societal or religious norms against suicide. However, affect-driven suicide decisions may upwardly bias estimates of the neutral level.

References

Arrhenius, Gustaf. 2000. 'An Impossibility Theorem for Welfarist Axiologies'. *Economics & Philosophy* 16.

Arrow, Kenneth J. 1950. 'A Difficulty in the Concept of Social Welfare'. *Journal of Political Economy* 58 (4): 328–346.

Bahmani-Oskooee, Mohsen, Seyed Hesam Ghodsi, and others. 2017. 'Asymmetric Causality and Asymmetric Cointegration between Income and House Prices in the United States of America'. *International Real Estate Review* 20 (2): 127–165.

Bellet, Clément, and Paul Frijters. 2019. 'Big Data and Well-Being'. *World Happiness Report 2019*.

Benjamin, Daniel J, Jakina Debnam, Marc Fleurbaey, Ori Heffetz, and Miles Kimball. 2020. 'What Do Happiness Data Mean? Evidence from a Survey of Happiness Respondents'. *Working Paper*, 1–41.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones. 2014. 'Can Marginal Rates of Substitution Be Inferred from Happiness Data? Evidence from Residency Choices'. *American Economic Review* 104 (11): 3498–3528.

Bhushan, Sidharth. 2020. 'Using Machines to Understand Human Wellbeing'. MPhil Thesis, University of Oxford.

Blanchard, Olivier J., and Lawrence H. Summers. 1986. 'Hysteresis and the European Unemployment Problem'. *NBER Macroeconomics Annual* 1: 15–78.

Bond, Timothy N., and Kevin Lang. 2019. 'The Sad Truth about Happiness Scales'. *Journal of Political Economy* 127 (4): 1629–40. https://doi.org/10.1086/701679.

Boyce, Christopher J., Alex M. Wood, and Eamonn Ferguson. 2016. 'Individual Differences in Loss Aversion: Conscientiousness Predicts How Life Satisfaction Responds to Losses versus Gains in Income'. *Personality and Social Psychology Bulletin* 42 (4): 471–484.

Breen, Richard, ed. 2004. *Social Mobility in Europe*. Oxford ; New York: Oxford University Press.

Broome, John. 2004. 'Weighing Lives'. *OUP Catalogue*.

Bukodi, Erzsébet, and John Goldthorpe. 2019. *Social Mobility and Education in Britain: Research, Politics and Policy*. Cambridge University Press.

Bukodi, Erzsébet, Marii Paskov, and Brian Nolan. 2019. 'Intergenerational Class Mobility in Europe: A New Account'. *Social Forces*, 1–32. https://doi.org/10.1093/sf/soz026.

Carter, Steven, and Michael McBride. 2013. 'Experienced Utility versus Decision Utility: Putting the "S" in Satisfaction'. *The Journal of Socio-Economics* 42: 13–23.

Cavaillé, Charlotte, and Daniel L Chen. 2020. 'A Decision Theoretic Approach to Understanding Survey Response: Likert vs. Quadratic Voting for Attitudinal Research'. *The University of Chicago Law Review*, 22.

De Neve, Jan-Emmanuel, George Ward, Femke De Keulenaer, Bert Van Landeghem, Georgios Kavetsos, and Michael I. Norton. 2018. 'The Asymmetric Experience of Positive and Negative Economic Growth: Global Evidence Using Subjective Well-Being Data'. *Review of Economics and Statistics* 100 (2): 362–375.

Diener, E. D., Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. 1985. 'The Satisfaction with Life Scale'. *Journal of Personality Assessment* 49 (1): 71–75.

Easterlin, Richard A. 2001. 'Income and Happiness: Towards a Unified Theory'. *The Economic Journal* 111 (473): 465–84. https://doi.org/10.1111/1468-0297.00646.

Ferrer-i-Carbonell, Ada, and Paul Frijters. 2004. 'How Important Is Methodology for the Estimates of the Determinants of Happiness?' *The Economic Journal* 114 (497): 641–59. https://doi.org/10.1111/j.1468-0297.2004.00235.x.

Fleurbaey, Marc, and Didier Blanchet. 2013. *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford University Press.

Fosse, Ethan, and Christopher Winship. 2019. 'Bounding Analyses of Age-Period-Cohort Effects'. *Demography* 56 (5): 1975–2004. https://doi.org/10.1007/s13524-019-00801-6.

Hovi, Matti, and Jani-Petri Laamanen. 2017. *Adaptation and Loss Aversion in the Relationship between GDP and Subjective Well-Being*. Tampereen yliopisto. http://tampub.uta.fi/handle/10024/102152.

Howard, George S., Kenneth M. Ralph, Nancy A. Gulanick, Scott E. Maxwell, Don W. Nance, and Sterling K. Gerber. 1979. 'Internal Invalidity in Pretest-Posttest Self-Report Evaluations and a Re-Evaluation of Retrospective Pretests'. *Applied Psychological Measurement* 3 (1): 1–23.

Kahneman, Daniel, and Amos Tversky. 1979. 'Prospect Theory: An Analysis of Decision under Risk'. *Econometrica: Journal of the Econometric Society*, 263–291.

Kanninen, Ohto, and Daniel H. Mahler. 2017. 'The Value Function with Life Satisfaction Data'.

Layard, Richard, Guy Mayraz, and Stephen Nickell. 2008. 'The Marginal Utility of Income'. *Journal of Public Economics* 92 (8): 1846–1857.

Luechinger, Simon. 2009. 'Valuing Air Quality Using the Life Satisfaction Approach'. *The Economic Journal* 119 (536): 482–515.

Oswald, Andrew J. 2008. 'On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings'. *Economics Letters* 100 (3): 369–72. https://doi.org/10.1016/j.econlet.2008.02.032.

Peasgood, Tessa, C. Mukuria, M. Karimi, and J. Brazier. 2018. 'Eliciting Preference Weights for Life Satisfaction: A Feasibility Study'. *Unpublished Manuscript*.

Pierson, Paul. 2000. 'Increasing Returns, Path Dependence, and the Study of Politics'. *American Political Science Review*, 251–267.

Schröder, Carsten, and Shlomo Yitzhaki. 2017. 'Revisiting the Evidence for Cardinal Treatment of Ordinal Variables'. *European Economic Review* 92 (February): 337–58. https://doi.org/10.1016/j.euroecorev.2016.12.011.

Seiler, Simon, and Ben Jann. 2019. 'International Comparison of Social Mobility Using the M-Index'. *Working Paper*, 1–31.

Shepard, Roger N. 1981. 'Psychological Relations and Psychophysical Scales: On the Status of "Direct" Psychophysical Measurement'. *Journal of Mathematical Psychology* 24 (1): 21–57. https://doi.org/10.1016/0022-2496(81)90034-1.

Sobel, Michael E. 1981. 'Diagonal Mobility Models: A Substantively Motivated Class of Designs for the Analysis of Mobility Effects'. *American Sociological Review* 46 (6): 893–906. https://doi.org/10.2307/2095086.

Uva, Teresa Bago d', Maarten Lindeboom, Owen O'Donnell, and Eddy Van Doorslaer. 2011. 'Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity'. *Journal of Human Resources* 46 (4): 875–906.

Van Praag, Bernard M. S. 1991. 'Ordinal and Cardinal Utility: An Integration of the Two Dimensions of the Welfare Concept'. *Journal of Econometrics* 50 (1–2): 69–89.

———. 2015. 'A New View on Panel Econometrics: Is Probit Feasible After All?' *IZA Discussion Paper*, no. 9345.

Vendrik, Maarten C. M. 1993. 'Collective Habits and Social Norms in Labour Supply. From Micromotives to Macrobehaviour'. Maastricht: University of Limburg.

Vendrik, Maarten C. M., and Geert B. Woltjer. 2007. 'Happiness and Loss Aversion: Is Utility Concave or Convex in Relative Income?' *Journal of Public Economics* 91 (7–8): 1423–1448. https://doi.org/10.1016/j.jpubeco.2007.02.008.

Viscusi, W. Kip. 2020. 'Wellbeing Measures of Mortality Risks: Life-Cycle Contradictions and Ordinal Index Challenges'. *Behavioural Public Policy* 4 (2): 245–53. https://doi.org/10.1017/bpp.2019.47.