THE NONPARAMETRIC APPROACH TO DEMAND ANALYSIS

ESSAYS IN REVEALED PREFERENCE THEORY

A Dissertation Submitted for the Degree of
DOCTOR OF PHILOSOPHY IN ECONOMICS

Abi Adams
New College
University of Oxford

August 2013

DECLARATION

I certify that this thesis, which is here submitted for the degree of Doctor of Philosophy in Economics at the University of Oxford, is my own work except where otherwise clearly stated.

I here declare the existence of joint work in this thesis. This thesis comprises three principal essays. The first essay is soley the product of my own work, except where otherwise stated. The second and third essays incorporate joint work. The second essay is based upon my MPhil thesis, which has since been rewritten with Professor Laurens Cherchye, Professor Bram De Rock and Mr Ewout Verriest. The final chapter is my own write up of a paper that will be written in the future with Professor Richard Blundell, Professor Martin Browning and Dr Ian Crawford. I proved the sufficiency of Theorem 4.1, all consequent Theorems and Propositions and produced all of the empirical results in this essay.

Abi Adams
August 2013

ACKNOWLEDGMENTS

Every day I thank my lucky stars that I have been under the supervision of Ian Crawford for my graduate studies. Ian Crawford is one of those rare people within whom immense intelligence, incredible kindness, and generosity combine. He has been a constant source of insight, inspiration and encouragement. I have learnt so much from him and without him the production of this thesis would not have been possible. The wider Group for the Advancement of Revealed Preferences, Richard Blundell, Martin Browning and Peter Neary also deserve special mention. This thesis and my academic development have benefited hugely from my conversations and collaborations with this talented group of people. I would also like to thank Christopher Bowdler for opening so many doors for me throughout my time as a graduate student.

My family have been a never ending source of love, support and encouragement. Thank you so much for your patience and providing me with the strongest possible foundation from which to go out and explore the world. I would also like to thank my friends for keeping me (relatively) sane throughout this process and for making my life so rich and colourful — I am sceptical that there are many other social groups who can count a multivariate kernel dance move amongst their Babylove repertoire. Above all, however, I would like to thank my co-conspirator, Jeremias Prassl, for filling my life with love, joy, and possibility, and for making me happier than I could ever have imagined.

I would like to thank the Economic and Social Research Council (ESRC) for their generous financial support. My graduate studies have been funded by the ESRC grant ES/I024808/1.

Finally, I gratefully acknowledge financial support from the European Research Council (ERC) under ERC-2009-AdG grant agreement number 249529. Data supplied by TNS UK Limited. The use of TNS UK Ltd. data in this work does not imply the endorsement of TNS UK Ltd. in relation to the interpretation or analysis of the data. All errors and omissions remain my own responsibility.

ABSTRACT

This thesis comprises three principal essays, each of which provides a contribution to the literature on the nonparametric approach to demand analysis. In each essay, I develop novel techniques that follow in the revealed preference tradition, and apply them to tackle a series of questions that concern the mechanisms underlying consumer spending decisions. Each technique developed is tightly linked to a particular nonparametric theory of choice behaviour and is explicitly designed for use with a finite set of observations. My work draws heavily upon results from finite mathematics, into which I integrate insights from information theory and integer programming. The output of this endeavor is a set of methodologies that are largely free of auxiliary assumptions over the form of the unobserved structural functions of interest.

Providing greater detail on the work to come, my first essay extends and clarifies the nonparametric approach to forecasting demand behaviour at new budget regimes. Using insights from information theory and integer programming, I construct an operational nonparametric definition of global rationality and develop a methodology that facilitates the recovery of globally rational individual demand predictions. This is the first attempt in the literature to develop a systematic methodology to impose global rationality on nonparametric demand predictions. The resulting forecasts allow for unrestricted preference heterogeneity in the population and I demonstrate how these predictions can be used for coherent welfare analysis. In my second and third essays, I prove new revealed preference testability axioms for models that extend the traditional neoclassical choice framework. Specifically, in my second essay, I address the intertemporal allocation of spending by collectives, whilst my final essay integrates taste variation into the utility maximisation framework. In both of these essays, I develop my testable results into practical algorithms that allow one to recover salient features of individual preferences. In my second essay, a methodology is developed to recover the minimal intrahousehold heterogeneity in theory-consistent discount rates, whilst my final essay develops a quadratic programming procedure that facilitates the recovery of the minimal interpersonal and intertemporal heterogeneity in tastes that is required to rationalise observed choice patterns. Applying these techniques to consumption micro-data yields new empirical insights that are of relevance to the applied literatures on time discounting, family economics and the public policy debate on tobacco control.

CONTENTS

Part I

OVERVIEW

1 INTRODUCTION

Economic theory provides us with a set of models that yield hypotheses about behaviour in the presence of scarce resources. These models can be incredibly powerful; by removing irrelevant detail from consideration, they enable one to logically isolate complicated chains of cause, effect and influence to identify the salient features of a problem. Simplicity and parsimony are worthy goals of the economic theorist and Occam's razor is oft quoted within the discipline to justify certain modelling assumptions.

However, there is a trade off between accuracy and parsimony. Economic models are typically positively motivated, meaning that predictive accuracy is a prized characteristic. Yet, it is not always clear that the assumptions underlying an model are valid and thus whether a model can account for observed behaviour. In the realm of applied demand analysis, although the utility maximisation model is the paradigm of parsimony, it is not self-evidently true that individuals behave consistently with this behavioural hypotheses.

Additional issues arise from the observation that the behavioural hypotheses yielded by economic models are typically imprecise. Often, it is often not clear what consistency with an economic model implies for observed choice behaviour. For example, the utility maximisation model does not specify which of the infinitely many rational preference orderings that a consumer is endowed with. Thus, precise hypotheses concerning the responsiveness of choice to changes in prices and income are not forthcoming. This is true for economic theory more generally; the hypotheses implied by models are typically qualitative in nature as strong assumptions regarding the form of underlying structural functions are typically eschewed by theorists. Therefore, it is often not clear how economic models would manifest themselves in reality if they were "true".

These observations lead one to posit three different sorts of questions in the realm of applied demand analysis (Varian, 1982).

1. Consistency: Can a model account for observed behaviour?

2. Recoverability: Can we recover the unobserved structural functions of interest from observed behaviour?

3. Forecasting: Given a model and past behaviour, can we predict behaviour in new choice situations?

Answering these questions is a nontrivial empirical exercise. The applied demand theorist operates in a "low information environment".

We typically do not directly observe the mechanisms that economic theory is concerned with. Further, we only ever have access to a finite amount of data on choice behaviour. This implies that the underlying structural functions cannot be perfectly inferred from observables.

Traditional approaches to answering the aforementioned questions typically start by making a number of functional form assumptions regarding individual preferences and the interaction between observed and unobserved heterogeneity. These auxiliary assumptions yield well-defined hypotheses for testing, recoverability, and prediction and allow one to operate easily in a low information environment. However, Type 1 error and misspecification bias plague this approach and have led to many an erroneous conclusion in the applied demand literature.

In contrast to the traditional approach, I remain a functional form agnostic throughout this thesis. In the three essays to come, I address a series of applied demand problems using simply the functional-form-free economic theory and a finite number of consumer choice observations. To achieve this, I draw upon results from finite mathematics, into which I integrate insights from information theory and integer programming techniques. The output of this endeavour is a set of methodologies that are free of auxiliary assumptions over the form of the unobserved structural functions that are of interest to the applied demand theorist.

Specifically, in the course of this thesis I derive new testability results for a set of models of consumer choice and develop these results into practical algorithms that facilitate the recovery of salient features of individual preferences. A further original contribution of this thesis is the development of a systematic method for imposing revealed preference inequalities upon the output of alternative estimation strategies. The application of these techniques to consumption microdata yields new empirical insights that run counter to some results in the literature to date and allows me to answer questions that have thus far not been satisfactorily addressed using parametric estimation strategies. In this thesis, I am thus able to make novel empirical contributions to the applied literatures on household decision-making, time consistency and, also, the public policy debate on tobacco control.

This chapter provides a deeper discussion of the literature alluded to in this introductory statement. This will provide context for the essays to come and the apposite background information for the contribution of this thesis to be assessed. I here also provide an overview of the principle essays that comprise this thesis and discuss their contribution to the wider literature. I close with relevant comments concerning the structure of the thesis.

2 KEY THEMES IN DEMAND ANALYSIS

2.1 CONSISTENCY

Economic models comprise a set of simplifying assumptions about the mechanisms that generate observed phenomena. Models are typically positively motivated and thus, at a minimum, one would expect an endorsed model to be capable of accounting for the behaviour of interest. As Sen remarks: "The primary concern here is not with the relation of postulated models to the real economic world but with the accuracy of answers to well-defined questions posed with preselected assumptions which severely constrain the nature of the models" (1977, p. 322). Addressing this concern requires an empirical test of a model: is it consistent, or not, with data that we have on the situation observed?

However, verification of the data consistency of a model is a non-trivial empirical exercise because the information available to the econometrician is only ever incomplete. For example, how is one to determine the data consistency of the utility maximisation hypothesis given that utility is unobserved? Early economists looked forward to the day when utility could be directly measured and the validity of their theories ascertained. Edgeworth (1881) predicted that developments in physio-psychology would bring forth a contraption he referred to as the "hedonimeter" (1881, p.101). This contraption would provide an uncontroversial physiological underpinning to utility theory. However, the existence of a unit of hedonic measurement is controversial. Utility is a diffuse, multidimensional concept that does not necessarily lend itself to quantification. Thus, technological feasibility aside, it is not conceptually clear that utility measurement will ever be attainable by economists.

THE TRADITIONAL APPROACH The fundamental unobservability of utility implies that empirical tests of rational choice theory must be defined with respect to demand behaviour rather than individual preferences. The traditional approach to testability makes use of integrability theory. The key result in this literature is that a set of demand functions corresponds to the maximisation of a well-behaved utility function if and only if these demand functions embody adding up, homogeneity, Slutsky symmetry and Slutsky negative semidefiniteness (Hurwicz and Uzawa, 1971).

However, the Slutsky conditions relate to restrictions on the partial derivatives of empirical demand functions. These conditions are necessarily local and implicitly assume that we have access to an infinite amount of data. Given that one only ever has access to a finite amount of data on choice behaviour, empirical verification of these conditions requires one to assume a particular functional form for the demand function and/or the distribution of error terms. The existence of a continuous demand function then enables the Slutsky derivative restrictions to be tested.

Thus, although the Slutsky conditions are not defined with respect to preferences, operationalising these conditions still requires that one makes assumptions over the form of the utility function. For this reason, the traditional approach does not constitute an effective test of the model because falsification cannot be achieved in a finite number of steps. Using Popper's (1959) terminology, the failure of the standard test only allows one to logically derive the singular statement: "The utility maximisation model, plus the hypothesis that the demand function is of the assumed functional form, is not consistent with choice behaviour". Popper (1959) forcefully argues that one is never logically justified in inferring universal statements from singular ones. In this context, one is unable to attribute the failure of a test to a failure of utility maximisation because that failure could derive from the invalidity of one's auxiliary hypotheses. Therefore, one is not justified in rejecting the utility maximisation model on the basis of tests that follow in the Slutsky tradition. It is not enough to continue testing the Slutsky restrictions for different parametric forms of the utility function because there are an infinite number of potential parameterisations to try. Therefore, an empirical strategy based upon the Slutsky conditions precludes falsification.

THE REVEALED PREFERENCE APPROACH The axioms of revealed preference theory offer an alternative method to assess the data consistency of economic theory. Rather than start from the demand function, as one does with integrability theory, the revealed preference approach asks if there exist equivalent statements to the model of interest that are only defined in terms of the finite observations on demand behaviour that one has access to. Hurwicz (1971) describes this approach as "global" as there is no need to compare choices infinitesimally close to each other. Tests in this tradition can be classed as effective because no auxillary hypotheses regarding the underlying structure of preferences are required to operationalise these axioms for empirical work.

Samuelson (1938, 1948) was the first to consider the restrictions that utility maximisation alone imposes upon observed choice behaviour. He concluded that: "if an individual selects batch one over batch two, he does not at the same time select two over one" (1938, p.66). In this manner "the individual guinea-pig, by his market behaviour, reveals his preference pattern" (Samuelson, 1948, p.243). The following relations define how observed choice behaviour reveals an individual's preference pattern.

DEFINITION 1.1 If at prices $\mathbf{p}_t$ the agent chooses $\mathbf{q}_t$ rather than $\mathbf{q}_s$, despite $\mathbf{q}_s$ being feasible at t, i.e. $\mathbf{p}_t^{'}\mathbf{q}_t \geqslant \mathbf{p}_t^{'}\mathbf{q}_s$, then we conclude that $\mathbf{q}_t$ is "directly revealed preferred" to $\mathbf{q}_s$, or $\mathbf{q}_t\mathbb{R}^0\mathbf{q}_s$.

DEFINITION 1.2 Given a sequence $\mathbf{q}_t\mathbb{R}^0\mathbf{q}_u, \mathbf{q}_u\mathbb{R}^0\mathbf{q}_v, \ldots, \mathbf{q}_w\mathbb{R}^0\mathbf{q}_s$, we conclude that $\mathbf{q}_t$ is "revealed preferred" to $\mathbf{q}_s$, $\mathbf{q}_t\mathbb{R}\mathbf{q}_s$.The revealed preference relation is thus transitive.

Samuelson's observations were extended by Houthakker (1950) to define the Strong Axiom of Revealed Preference (SARP), which proves the equivalence between utility maximisation and demands' satisfying SARP. One cannot reject utility maximisation as a model of consumer decision making if and only if demands satisfy SARP. Additional results in the foundational revealed preference literature were largely concerned with the refinement and clarification of these equivalence results (Richter, 1966; Sen, 1971; Stigum, 1973; Uzawa, 1960). In the words of Mas-Colell, the early literature on revealed preference theory represented as "foundational and purely theoretical a subject as one can find" (1982, p.72).

It was not until the 1960's and the birth of what Pollak (1990) terms the "restricted domain" strand of revealed preference theory that tests of the revealed preference axioms became empirically implementable. The emergence of revealed preference as an empirical methodology followed from Sydney Afriat's ground breaking 1967 article and its later translation and popularisation by Diewert (1973) and Varian (1982). Afriat (1967) provided a systematic linear programming procedure that allows one to determine whether a finite set of price-quantity observations is consistent with utility maximisation. He further proved that if there is a non-empty solution set associated with this linear programme, one can always find a concave, well-behaved utility function that rationalises the observed data. These contributions are summarised by Afriat's Theorem.

AFRIAT'S THEOREM The following conditions are equivalent:

1. The data $\{\mathbf{q}_t, \mathbf{p}_t\}_{t=1,\ldots,T}$ is consistent with the utility maximisation model.

2. The data satisfy the Generalised Axiom of Revealed Preference (GARP), that is,

$$\mathbf{q}_t R \mathbf{q}_s \quad \textit{implies} \quad \mathbf{p}_s'\mathbf{q}_s \leqslant \mathbf{p}_s'\mathbf{q}_t \tag{A1}$$

   where R represents the revealed preference relation defined above.[1]

1 GARP captures the rationality requirement that the best element of one's budget set is selected; if a consumer prefers $\mathbf{q}_t$ to $\mathbf{q}_s$, then $\mathbf{q}_s$ can only be observed if $\mathbf{q}_t$ is unaf-

3. There exist numbers $u_t, \lambda_t > 0$, for $t = 1, ..., T$, such that

$$u_t - u_s \leqslant \lambda_s \mathbf{p}_s'(\mathbf{q}_t - \mathbf{q}_s) \tag{A2}$$

4. There exists a nonsatiated, continuous, concave, monotonic utility function that rationalises the data.

PROOF. See Afriat (1967), Diewert (1973), Varian (1982).

Afriat's Theorem provides the applied demand theorist with an effective test of the utility maximisation model. If one can find a set of $u_t, \lambda_t > 0$ for $t = 1, ..., T$ such that the linear programme defined by (A2) holds, then one cannot reject the hypothesis that the data were generated by a consumer acting to maximise their utility. This procedure is quite remarkable given that it makes no reference to the functional structure of the underlying utility function. The approach thus speaks well to Hicks' comment that "one has a right to an economics free of utilitarian assumptions" (1939, p.18).

RECENT WORK Although the "consistency question" has been expertly addressed for the static neoclassical consumer choice model, this by no means implies that there is no further work to be done in this area. The revealed preference characterisations of various models have been deduced in recent years (Browning, 1989; Cherchye, De Rock and Vermeulen, 2007; Crawford, 2010). Furthermore, Chambers, Echenique and Shmaya (2012) make use of results from model theory to prove that there exist necessary and sufficient revealed preference conditions, defined only in terms of observables, for any model that can be expressed as a series of universal statements.

I address the question of consistency in the second and third chapters, in which I prove new revealed preference testability axioms for models that extend the traditional neoclassical choice framework. Specifically, in my second essay, I address the intertemporal allocation of spending by collectives, whilst my final essay integrates taste variation into the utility maximisation framework. The theoretical results in the latter essay are particularly surprising. In this essay, it is proven that any data set that satisfies weak variability requirements can be rationalised once one allows for taste variation on a single good. A corollary of this result is that observed choice heterogeneity over a K-dimensional demand system can be parsimoniously summarised by a univariate parameter in the direct utility function. Further, in both of these essays, I operationalise the proven testability re-

fordable at the time s budget. Varian (1982), building upon the prior work of Afriat (1967), proves that there exists a well-behaved utility function rationalising observed behaviour if and only if demands satisfy GARP. Therefore, there is a logical equivalence equivalence between GARP and the satisfaction of the Slutsky conditions and these alternative approaches should be considered as logically equivalent ways of expressing the same conception of rationality.

sults using standard linear programming techniques and develop the conditions into practical algorithms that allow one to recover salient features of individual preferences. In my second essay, a methodology is developed to recover the minimal intrahousehold heterogeneity in theory-consistent discount rates, whilst my final essay develops a quadratic programming procedure that facilitates the recovery of the minimal interpersonal and intertemporal heterogeneity in tastes that is required to rationalise observed choice patterns.

2.2 RECOVERABILITY AND FORECASTING

In addition to assessing the data consistency of consumer theory, questions of recoverability and forecasting occupy a central place in applied demand theory. In fact, verification of the data consistency of economic models represents a relatively niche concern within the applied microeconometrics literature. Rather, most research interest centers on recovering the structural functions that underlie choice behaviour and then using economic models to predict behaviour in new circumstances. Within the realm of applied demand analysis, if behaviour is consistent with the utility maximisation model, interest quickly moves to understanding the characteristics of consumer preferences and how choices vary with changes in the budget environment.

However, these recoverability and prediction problems are typically ill-posed. The structural functions underlying observed behaviour cannot be perfectly inferred from finite choice data because these choices are consistent with an infinite number of preference specifications. Therefore, the nonparametric demand theorist can only set identify the objects of interest.

Layering additional assumptions on top of an economic model refines the set of potential answers that can be provided to one's recoverability and forecasting questions. Given a particular parametric specification for the demand function, ambiguity over the underlying structural function is assumed away. Recoverability then becomes a matter of identifying the structural parameters such that the assumed functional specification best fits the data. Although this approach yields unique preference parameters and demand predictions, it does so at the risk of misspecification; there remains a risk that the predictions implied by one's chosen functional form are not consistent with the data that we have on individual choice behaviour. This problem even afflicts researchers who use nonparametric methods to estimate statistical demand functions. In this context, functional assumptions regarding the interaction between observed and unobserved heterogeneity must be imposed for demand functions to be identified (Matzkin, 2003).

Empirical evidence suggests that the worry of misspecification bias is well-founded. Estimated demand systems are generally found to fit the data poorly (Hoderlein, 2008) and are typically unable to encompass the variety of shapes that are required to reflect the patterns in observed choice behaviour (Pendakur and Lewbel, 2009).

THE REVEALED PREFERENCE APPROACH Despite Samuelson's initial aim "to develop the theory of consumer's behaviour free from any vestigial traces of the utility concept"(1938, p.71), it is not the case that adopting a revealed preference approach commits one to an agenda that is devoid of the concept of preferences. Rather, it is simply that, in taking this approach one remains free of functional assumptions regarding the *precise structure* of the utility function. As Sen (1973, p.244) remarks: "if the theory of revealed preference makes sense it does so not because no psychological assumptions are used but because the psychological assumptions used are sensibly chosen". The revealed preference theorist does not deny the existence of preferences but rather makes no assumptions as to what form they take. In fact, preference recovery and demand prediction are central themes of the revealed preference literature. The basic approach for engaging in such exercises was covered extensively by Varian (1982, 1983) for the utility maximisation model.

RECENT WORK Within the revealed preference literature, researchers typically embrace the fact that the questions of interest are ill posed and work with the entire set of feasible solutions. Recent work by Blundell, Browning and Crawford (2003, 2008) and Blundell *et al.* (2012) has been exceptionally useful in this regard. These authors provide a methodology to refine the set of demands that the revealed preference theorist will predict at a new budget regime by "controlling" for budget variation in survey data. Alternatively, revealed preference empiricists may select "reasonable" preference constructs and demand predictions from the feasible sets delineated by theory and data. For example, the output from the linear programme defined by Afriat's inequality can be used to define a piecewise linear, concave utility function that is consistent with observed choices. However, the reader is reminded that there also exist many other candidate utility functions that are consistent with the data.

All three essays in this thesis make a contribution to the literature on nonparametric recovery and prediction. On the theme of recoverability, the techniques developed in the thesis facilitate the recovery of the minimal intrahousehold heterogeneity in theory-consistent discount rates and also the identification of the minimal interpersonal and intertemporal heterogeneity in tastes that is required to rationalise observed choice patterns. Further, on the theme of forecasting, my first essay extends and clarifies the nonparametric approach to predicting demand behaviour at new budget regimes. Using insights from information theory and integer programming, I construct an operational nonparametric definition of global rationality and develop a methodology that facilitates the forecasting of rational individual demand predictions. This is the first attempt in the literature to develop a systematic methodology to impose global rationality upon nonparametric demand predictions. The resulting forecasts allow for unrestricted preference heterogeneity in the population and I demonstrate how these predictions can be used for coherent welfare analysis.

3 THESIS STRUCTURE

The last section provided an overview of the nonparametric approach to demand analysis and alluded to the contribution that this thesis makes to the literature. This section briefly summarises the content of the essays to come and highlights how this thesis complements and extends existing results in this field.

GLOBALLY RATIONAL DEMAND FORECASTS This first essay extends, and refines, the nonparametric approach to forecasting individual demand behaviour at previously unobserved budget regimes. Nonparametric demand predictions may violate basic rationality requirements. This undermines their use for coherent welfare analysis. The primary aim of this essay is to develop a nonparametric methodology with which to recover demands that necessarily satisfy global rationality.

To achieve this aim, I first extend the traditional revealed preference approach to demand forecasting to recognise that the global rationality of a set of demand predictions requires that one's predictions are simultaneously rationalisable by a single utility function. This introduces non-trivial computational difficulties in attempts to create an operational definition of forecast rationality. I tackle these problems using mixed integer programming techniques. I then use insights from information theory to impose this working definition of global rationality upon unconstrained nonparametric demand predictions. Specifically, I recast the demand identification problem as an ill-posed inverse problem in order to select the "Minimum Information Discrimination" demand from the revealed preference support set. The resulting rationality-updated predictions can be considered those "most likely" by the statistician who is ignorant of economic theory and who predicted the original unconstrained forecasts.

I demonstrate the value of this methodology through an illustrative empirical application to individual-level panel data that is drawn from the Kantar Worldpanel. The resulting empirical strategy allows for unrestricted preference heterogeneity across individuals and returns plausible, globally rational demand predictions. I further show how these demands can be used to construct path independent welfare metrics, which can be used to evaluate consumer wellbeing at new budget regimes.

This essay makes two key contributions to the literature on nonparametric demand analysis. This essay is the first to note that demand forecasts made in the revealed preference tradition must be consistent with one another, and not simply with past choices, and provide an operational definition of the global rationality of demand forecasts. The empirical application in this essay highlights that the requirement of cross-consistency is practically relevant; the constraint is binding for a portion of the sample when I predict individual demands at a set of intersecting budgets. Second, the MDI methodology is the first attempt made in the literature to achieve globally rational nonparametric demand predictions. Past studies have only been able to impose certain elements of the set of necessary requirements for rationality upon nonparametric forecasts. This contribution is important because a violation of rationality introduces path dependency to methods that are applied to calculate measures of exact consumer surplus, which creates problems for researchers who want to engage in welfare analysis.

TIME INCONSISTENCY AND COLLECTIVE CHOICE My second essay is set in a context in which the traditional neoclassical choice model has been widely discredited. This essay provides an analysis of models that are concerned with the intertemporal allocation of spending. Choice in this setting is often labelled as "irrational" and has been rationalised using models from behavioural economics such as hyperbolic discounting. However, empirical work in this area rarely distinguishes between individual and collective choice. Insufficient regard for the separateness of persons within the household provides a route by which observed choice behaviour can appear irrational as the process of preference aggregation can generate apparent inconsistencies in aggregate choice data.

This essay develops a revealed preference methodology to test the consistency of household consumption behaviour with simple economic models of intertemporal choice. In so doing, a methodology is created to establish whether time inconsistencies in household choice can be attributed to individual heterogeneity and renegotiation within the collective unit, rather than resulting from nonstationarities at the individual level.

An empirical application to household-level microdata highlights that an explicit recognition of the collective nature of choice allows the vast majority of household behaviour to be rationalised by traditional neoclassical choice theory that assumes preference stationarity at the individual level. For the particular short panel data set that is used in this essay, simply permitting limited intrahousehold heterogeneity in time preferences allows the choices of 98.4% of the sample to be rationalised by a model that assumes exponential discounting at the individual level. It is also found that couples who are characterised by a lower divergence in spousal discount rates are older, more likely to have children and wealthier, which we take as indications of experiencing higher match quality.

This essay makes a methodological contribution to the nonparametric demand literature and an empirical contribution to the behavioural economics literature. The essay derives a revealed preference characterisation of collective models of intertemporal choice and thus provides novel results in this regard. Further, our failure to reject the hypothesis that choice is time consistent once the appropriate locus of decision making is located at the individual level is an important finding, at odds with much preceding empirical work in the literature. Our results suggest a reevaluation of the empirical evidence against the standard rational choice framework, at least in its capacity as a positivist modelling device concerning consumption choices over the short to medium run.

This essay contains joint work with Laurens Cherchye, Bram De Rock and Ewout Verriest and is an extension of my MPhil thesis. In my MPhil thesis, I independently proved all of the main theorems that are contained by this essay. I note here that the essay presented contains independent discussions and clarifications that are separate from my work with Cherchye *et al.*, on Pareto weight renegotiation as a mechanism for time inconsistent choice and also on the problems that arise in attempts to create a sufficient test of the "no-commitment" model.

RATIONALISING TOBACCO CONSUMPTION My second essay asked whether traditional neoclassical demand theory could rationalise observed choice behaviour if applied to the appropriate decision making unit. In my final essay, I consider a setting in which a consumer's demands *cannot* be rationalised by the basic utility maximisation hypothesis. This essay then addresses the questions of testability and data consistency for models that introduce taste instability into the basic rational choice framework. These results are developed into a practical algorithm that can be applied to compute the lower bound on intertemporal and interpersonal variation in marginal utility that is required to rationalise a finite set of observations on past choice behaviour.

In this essay, I apply this methodology to rationalise the patterns that have been observed in U.K. tobacco consumption since 1980. Psuedocohorts are constructed from the U.K. Family Expenditure Survey and their demands are recovered along "Sequential Maximum Power" paths using censored quantile regression techniques. It is found that taste change is a necessary component of any rationalisation of observed trends. Our analysis also uncovers evidence of significant educational differences in the "effective" and "raw" taste trajectories for tobacco. These results give strong reason to believe that less educated groups have a greater taste for tobacco relative to their more highly educated peers, for all but the heaviest smoking groups.

This essay makes a methodological contribution to the nonparametric demand literature and provides empirical results of relevance to the public policy debate on tobacco control. In addition to deriving new revealed preference constraints for models of taste change, I prove that any data set that satisfies weak variability requirements can be rationalised by allowing for taste variation on a single good. A corollary of this result is that observed choice heterogeneity over a K-dimensional demand system can be parsimoniously summarised by a univariate parameter in the direct utility function. This has interesting implications for the representation of unobserved preference heterogeneity. Furthermore, the finding of educational differences in the evolution of tobacco tastes points to a need for greater research on the mechanisms underlying this empirical phenomenon.

The material in this essay is my own write up of a paper that I will write in the future with Richard Blundell, Martin Browning and Ian Crawford. I proved the sufficiency of Theorem 4.1, all consequent Theorems and Propositions and have produced all of the empirical results in this essay.

I end this section by noting that the essays in this thesis follow in the order given above. Each essay begins with an introduction to the literature relevant to that essay only, and ends with a conclusion summarising the key results and contributions. Proofs of the Theorems that are derived in this thesis and detailed data descriptions are deferred to Appendices to improve readability.

Part II

GLOBALLY RATIONAL NONPARAMETRIC DEMAND FORECASTS

My first essay extends, and refines, the nonparametric approach to forecasting individual demand behaviour at previously unobserved budget regimes. Nonparametric demand predictions may violate basic rationality requirements. This undermines their use for coherent welfare analysis. The primary aim of this essay is to develop a nonparametric methodology with which to recover demands that necessarily satisfy global rationality.

To achieve this aim, I first extend the traditional revealed preference approach to demand forecasting to recognise that the global rationality of a set of demand predictions requires that one's predictions are simultaneously rationalisable by a single utility function. This introduces nontrivial computational difficulties in attempts to create an operational definition of forecast rationality. I tackle these problems using mixed integer programming techniques. I then use insights from information theory to impose this working definition of global rationality upon unconstrained nonparametric demand predictions. Specifically, I recast the demand identification problem as an ill-posed inverse problem in order to select the "Minimum Information Discrimination" demand from the revealed preference support set. The resulting rationality-updated predictions can be considered those "most likely" by the statistician who is ignorant of economic theory and who predicted the original unconstrained forecasts.

I demonstrate the value of this methodology through an illustrative empirical application to individual-level panel data that is drawn from the Kantar Worldpanel. The resulting empirical strategy allows for unrestricted preference heterogeneity across individuals and returns plausible, globally rational demand predictions. I further show how these demands can be used to construct path independent welfare metrics, which can be used to evaluate consumer wellbeing at new budget regimes.

4

INTRODUCTION

The aim of this essay is to develop a flexible and accurate methodology with which to recover globally rational demands. Specifically, I use insights from information theory to impose rationality upon unconstrained nonparametric demand predictions at previously unobserved budget regimes. My empirical strategy allows for unrestricted preference heterogeneity across households and returns point estimates for demands at new regimes that satisfy all of the nonparametric implications of utility maximisation. I demonstrate the value of this method via an empirical application to individual panel data, in which I explore the heterogeneity of price responses and compute welfare metrics for a hypothetical change to the price of alcohol.

IRRATIONAL FORECASTS The economic theory lying behind traditional demand analysis is extremely simple: a consumer chooses the bundle of goods that maximises their utility from the set of all bundles that they can afford. There is no uncertainty in this model; individuals are assumed to know all prices, their preferences are static and their choice behaviour is assumed to be perfectly observed. However, the theory is silent over the precise characterisation of preferences. The only assumption in this regard is that preferences are rational. This requirement of rationality is not trivial; the utility of recovered demand predictions is severely curtailed by its violation. Rationality sits at the heart of applied demand and welfare analyses. Unique consumer surplus and deadweight loss estimates are contingent upon preferences meeting this standard because irrationality generates path dependency in methods that are applied to numerically calculate exact consumer surplus (Vartia, 1983; Hausman and Newey, 1995).

Although rationality is a deeply useful property, whether it manifests itself in estimated demand systems is another matter. The predictions yielded by unconstrained forecasting exercises are not necessarily consistent. Forecast irrationality may derive from a number of sources and can arise even if utility maximisation is a valid behavioural model. For example, epistemic uncertainty regarding the structure of an individual's demand function can result in prediction errors that introduce inconsistencies into one's forecasts. The structure of the demand function is not uniquely prescribed by economic theory but this knowledge is required to perfectly predict an individual's rational demands at new budget environments. Using Jaynes' (2003) terminology, the applied demand theorist is condemned to operate at an "exploratory phase" of empirical work; the "correct" model for demand forecasting purposes cannot be directly observed, nor can it be indirectly inferred from finite choice data. The errors in demand predictions that follow from operating in this "low information environment" can lead to demand predictions failing the consistency requirements that are imposed by rationality.

A standard strategy that is often taken when operating in a low information environment is to, in effect, assume away the existence of finite-data problems by making assumptions over the functional structure of the individual demand function and/or preference heterogeneity in the population in order to estimate a statistical demand model. In this context, irrationality of recovered demand predictions could derive from functional misspecification bias rather than any inconsistency in individual decision making behaviour.

The potential for irrationality-inducing misspecification bias even afflicts nonparametric estimation techniques. In the face of unobserved preference heterogeneity, nonparametric statistical demand models must place assumptions upon the functional form of taste differences in a population so that one can make use of information across consumers. Statistical models assume that observationally equivalent consumers are identical in all dimensions except for an individual specific error term that captures the impact of unobserved heterogeneity. One would hope to be maximally flexible over the specification of this error term and proceed without recourse to assumptions that do not follow from economic theory. Sadly, Matzkin (2003) proves that this is not possible and all models in this tradition require assumptions to be made regarding the structure of unobserved heterogeneity. A nonadditive model $\mathbf{y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\epsilon})$, where $\mathbf{X}$ and $\boldsymbol{\epsilon}$ denote vectors of observable and unobservable explanatory variables respectively, is not identified without additional structure (Matzkin, 2003).

It is standard practise to specify an additive error term to reflect unobserved preference heterogeneity, e.g. $\mathbf{y} = \mathbf{f}(\mathbf{X}) + \epsilon$, with (typically) $E(\epsilon|\mathbf{X}) = 0$. Although this specification appears innocuous, it in fact imposes very strong assumptions on the structure of preferences, which may not hold in reality. Lewbel (2001) proves that recovered nonparametric demand systems of this form will violate rationality unless every individual in the population has homothetic preferences. Therefore, although nonparametric techniques assuage many concerns regarding functional form misspecification that plague parametric demands models, strong assumptions regarding the similarity of consumers still underlie the approach.[1] The failure of these un-

1 Parametric statistical demand models further assume away any uncertainty over the underlying structure of preferences via the specification of a particular functional form for the demand function. This additional structure permits point identification of demands at new budget regimes. However, any parametric specification is essentially arbitrary and almost certainly misspecified. Traditional parametric specifications place stringent assumptions upon the structure of, and heterogeneity in,

derlying assumptions can impart irrationality on estimated demand systems

IMPOSING RATIONALITY For the aforementioned reasons, (among many others,) unconstrained demand predictions may violate rationality even if utility maximisation is a valid behavioural hypothesis. A method for imposing global rationality upon a nonparametric demand system does not yet exist. To date, only certain elements of the set of necessary requirements for rationality have been imposed upon estimates. Kim and Tripathi (2003) discuss the imposition of homogeneity upon a nonparametric demand system, whilst Haag, Hoderlein and Pendakur (2009) develop a method for estimation of the full nonparametric demand system under symmetry. Recently, Blundell, Horowitz and Parey (2012) differentially weight observations to ensure the satisfaction of the Slutsky constraint in their analysis of gasoline demand. However, none of these approaches ensures the global satisfaction of the rationality of predictions. In many nonparametric regression applications, theory-consistency restrictions are imposed pointwise and thus may fail to hold *across* one's predictions. Furthermore, negative semidefiniteness of the Slutsky matrix, a necessary condition for rationality of the demand function, cannot be imposed pointwise because it is a global property that must apply across points on the considered grid.

There are methods that allow one to systematically impose rationality upon predictions at a single new budget of interest if one is willing to forego point identification of demands. Revealed preference arguments identify the "support set" (Varian, 1982) of rational demand predictions at a new budget. This is the set of demands on a new budget hyperplane that satisfy all of the nonparametric implications of utility maximisation. There exists a well-behaved utility function that can jointly rationalise demand bundles in the support set and past observations, but no such utility function exists for demands contained by the set's complement. However, application of the revealed preference methodology does not facilitate the point identification of demands at new budget regimes. There are many ways in which compensated demand curves can slope downwards and one cannot discriminate between candidate demand predictions within the support set from the perspective of the revealed preference approach alone.

ESSAY CONTRIBUTION In this essay, I develop a methodology for imposing global rationality upon sets of unconstrained nonparametric demand forecasts. The proposed methodology combines nonparametric estimation techniques with the revealed preference approach

price and income responses. However, recent empirical work (for example, Blundell, Chen and Kristensen, 2007) suggests that a variety of shapes characterise demand behaviour in a manner that cannot be easily captured by traditional parametric models.

to demand analysis using insights from information theory. Specifically, I defer to the Principle of Minimum Discrimination Information (Kullback, 1959) to update unconstrained demand predictions with rationality at previously unobserved budget regimes. Updating forecasts in accordance with this principle dictates that one select the "closest" rational demands to one's unconstrained pilot forecasts, where the distance between quantity bundles is measured by the Kullback-Leibler divergence.

The Principle of Minimum Discrimination Information is a compelling criterion to adopt in this context for three reasons. First, the Minimum Discrimination Information (MDI) result is the most likely rational demand to occur given a belief in the accuracy of the pilot prediction. Therefore, it is the least surprising result to the statistician ignorant of economic theory. Second, it corresponds to the demand that minimises the predictive loss of researchers who make their predictions without the information from economic theory. Therefore, it can be regarded as the most optimistic updating rule followed by an econometrician who is looking to make use of a statistician's estimates. Finally, the objective function that is associated with the approach has a number of attractive features that support uniqueness of the updated solution and preserve assumptions concerning the separability of preferences. The resulting MDI demand predictions can be used for coherent welfare analysis and the methodology is shown to yield plausible estimates of individual price responses.

This essay also extends the conceptual framework associated with the pure revealed preference approach to demand prediction. Until this point, the revealed preference literature has largely concerned itself with forecasting rational demands at single budgets. However, when the aim is to forecast multiple demands at a set of new budgets, global rationality requires that one's forecasts are cross-consistent such that predictions are jointly rationalisable by a single utility function. This requirement of cross-consistency can generate nonconvexities in the support set and the revealed preference bounds on demand responses become conditional upon a hypothesised preference ordering over new budgets. In high dimensions or when many new budgets are considered, this leads to complex restrictions that are difficult to interpret. I provide a mixed integer linear programming characterisation of the support set in this instance, which can be applied to impose the cross-consistency of demand predictions and compute the conditional revealed preference bounds. I find the requirement of cross-consistency to be theoretically and practically important. There are many situations in which the aim is to compare behavioural responses at sets of new budget hyperplanes that intersect. In such contexts, the cross-consistency of predictions becomes a relevant concern. An illustrative empirical investigation of the size of the support set with and without the imposition of the cross-consistency requirement is suggestive that this requirement will be a binding constraint in practise.

I demonstrate the value of the MDI methodology, and illustrate how it may be applied to recover demands and welfare metrics, via an empirical application to individual level consumption data drawn from the Kantar Worldpanel. Well-behaved demand systems are recovered at the individual level that satisfy a global rationality requirement and allow for unrestricted preference heterogeneity across households. In so doing, I find evidence of pervasive unobserved preference heterogeneity that is non-normally distributed and is independent of observable characteristics. The flexibility of the MDI methodology that I develop is thus of value because it does not necessitate the pooling of individual level information.

Irrationality of pilot predictions is not found to be a binding constraint when recovering the marginal demand curve for the particular data set used in this thesis. Only 3% of predictions fail rationality and the manipulation that these pilots require to attain global rationality is slight. However, when predicting over sets of budget hyperplanes that intersect, the attainment and imposition of global rationality is found to be of practical relevance. The size of the set of rational predictions is significantly constrained by the requirement of cross-consistency of predictions that global rationality requires. At a set of three intersecting budget hyperplanes, the pilot predictions of only 1.8% of individuals required application of the MDI updating methodology when the rationality of predictions was assessed independently. However, the pilot predictions of 10.4% of individuals required updating once the requirement of cross-consistency of forecasts was imposed even though the scale of the problem was very limited.

STRUCTURE This essay proceeds as follows. Section 5 outlines the restrictions on demand predictions that are equivalent to global rationality and extends the revealed preference methodology to address the interaction of demand predictions across new budgets of interest. Section 6 briefly considers reasons why nonparametric demand predictions may violate the revealed preference requirements even if rationality is a valid behavioural hypothesis. Sections 7 and 8 outline and justify the Minimum Discrimination Information approach to the recovery of globally rational individual demand predictions. In Sections 9 to 12, I demonstrate the value of the MDI methodology via an empirical application to individual consumption data drawn from the Kantar Worldpanel. In these sections, the heterogeneity of price and welfare effects is explored and the characteristics of the updating process is examined through the recovery of the marginal demand curve for Meat & Fish, the forecasting of demands at a set of intersecting budget hyperplanes and the evaluation of the welfare effects of a rise in the price of Alcohol. Section 13 concludes.

5 CHARACTERISING GLOBAL RATIONALITY

The aim of this essay is to recover globally rational demands at previously unobserved budgets. This section formally defines the global rationality requirement and outlines the associated restrictions on demand forecasts that this requirement imposes.

5.1 RATIONALITY

Following Lewbel (2001), I equate rational demands with those that could have been derived from the maximisation of a well behaved utility function. The behavioural hypothesis of utility maximisation specifies that individual i's choice behaviour should be modelled as the outcome to the following maximisation problem:

$$\max_{\mathbf{q}} u^i(\mathbf{q})$$

subject to

$$\mathbf{p}'\mathbf{q} = y^i$$

where $u^i(\mathbf{q})$ represents individual i's concave, locally nonsatiated utility function and y^i gives their total expenditure level. Global rationality thus requires that past behaviour and one's demand predictions can be jointly rationalised by a single, well behaved utility function. Varian (1982) proves that this is equivalent to past choices and forecasts jointly satisfying the Generalised Axiom of Revealed Preference (GARP).[1]

Not all demands that exhaust a new budget hyperplane will necessarily satisfy a global rationality requirement. The revealed preference approach to demand prediction uses this fact to set identify rational demand forecasts. Rational demand predictions are elements of the "support set" (Varian, 1982). The rest of this section defines, and extends, the concept of the revealed preference support set in order to provide a characterisation of the set of globally rational demand forecasts.

5.2 THE VARIAN SUPPORT SET

I am concerned with predicting each individual i's, $i = \{1, ..., N\}$, demands, $\mathbf{x}^i_b$, at a set of $|B|$ new budgets, $B = \{\boldsymbol{\rho}_b, m_b\}_{b=1,...,|B|}$, where

1 GARP was defined and its equivalence with the behavioural hypothesis of utility maximisation was discussed in the context of Afriat's Theorem in Part I.

$\rho \in \mathbb{R}^K_{++}$ is a strictly positive K-vector of prices and $m > 0$ is total expenditure at the new budget of interest. Let the set of unique demand predictions for individual i at the $|B|$ budgets be given by $\Omega^i = \{\mathbf{x}^i_b\}_{b=1,...,|B|}$.

I assume that we observe T past consumption choices that are made by individual i, $\{\mathbf{q}^i_t\}_{t=1,...,T}$, and the prices at which these choices were made, $\{\mathbf{p}_t\}_{t=1,...,T}$. These choices are summarised by the set $D^i = \{\mathbf{q}^i_t, \mathbf{p}_t\}_{t=1,...,T}$. I assume that D^i satisfies GARP and thus that one cannot reject the hypothesis that they were generated by a utility maximising individual.[2]

Past choices and the behavioural hypothesis of utility maximisation together constrain the location of demand responses at the budgets defined by B. To illustrate this fact, let us first consider the problem of predicting a demand response at a single new budget: $|B| = 1$ and $B = \{\boldsymbol{\rho}_b, m_b\}$. Any rational individual demand response at B is an element of the "Varian" support set (Varian, 1982), $S^{V,i}_b = S^V(\boldsymbol{\rho}_b, m_b|D^i)$. $S^{V,i}_b$ defines the set of demands on the new budget surface that are nonparametrically consistent with the existing choice data; any element of $S^{V,i}_b$ satisfies GARP for observed demands, whilst any point in the set's complement fails.

DEFINITION 2.1 Given $D^i = \{\mathbf{q}^i_t, \mathbf{p}_t\}_{t=1,...,T}$ and a single new budget $\{\boldsymbol{\rho}_b, m_b\}$, the Varian support set is defined as:

$$S^{V,i}_b = S^V(\boldsymbol{\rho}_b, m_b|D^i) = \left\{ \mathbf{x}^i_b : \begin{array}{l} \mathbf{x}^i_b \geqslant \mathbf{0} \\ \boldsymbol{\rho}_b' \mathbf{x}^i_b = m_b \\ \{D^i; \boldsymbol{\rho}_b, \mathbf{x}^i_b\} \text{ satisfy GARP} \end{array} \right.$$

PROPOSITION 2.1 $S^{V,i}_b$ is a closed, convex subset of the set of demands that exhaust a budget $\{\boldsymbol{\rho}_b, m_b\}$.

PROOF. See Appendix A.

The Varian support set is closed and convex, endowing it with a number of convenient properties for applied work. This set exhausts the empirical content of utility maximisation when forecasting demands at a single new budget and represents the set of demands that is traditionally offered by revealed preference theory in answer to the demand extrapolation problem.[3] There exists a well-behaved utility function that can jointly rationalise D^i and the demand bun-

2 Rationality of D^i is a maintained assumption throughout this essay. In our empirical application, it is upheld for 82% of individuals in my sample. In my final essay, I examine how the assumption of a time invariant utility function can be relaxed for data sets that violate GARP.

3 Until the recent work by Blundell *et al.*(2003, 2008, 2012), budget variation in survey data diluted the informativeness of the support set for demand prediction purposes. If budget hyperplanes do not cross, then utility maximisation places no restrictions

dles in the Varian support set but no such utility function exists for demands contained by the set's complement.

5.3 THE SUFFICIENT SUPPORT SET

Membership of the Varian support set is not sufficient for the global rationality of predictions once one moves to recover demands at a set of new budgets. In a world of multiple forecasts, demand predictions must be consistent with one another. This requirement of cross-consistency of predictions is not guaranteed by point wise application of Varian's methodology. I now proceed to define what I term the "sufficient support set" and develop an equivalent mixed integer linear programming definition of this set that can be more easily operationalised to calculate the conditional revealed preference bounds on globally rational demand responses.

AN EXAMPLE To illustrate the distinction between the Varian and sufficient support sets, let us consider the prediction of demands at two new budgets $\{\boldsymbol{\rho}_1, m_1\}$ and $\{\boldsymbol{\rho}_2, m_2\}$. As previously, past observations, D, and utility maximisation together constrain predictions at each budget to lie within their respective support sets.

$$\begin{aligned} \mathbf{x}_1 &\in S^V(\boldsymbol{\rho}_1, m_1|D) \\ \mathbf{x}_2 &\in S^V(\boldsymbol{\rho}_2, m_2|D) \end{aligned}$$

In Figure 1, membership of the Varian support sets defined by the past choices $\mathbf{q}_1$ and $\mathbf{q}_2$, constrains demand responses at B_1 to lie along the segment AD and constrains demand responses at B_2 to lie along BC at B_2. Further, as we are working with a 2-good example, every feasible budget share specification $\{\boldsymbol{\omega}_b\}_{b=1,2}$, where $\boldsymbol{\omega}_b = (\omega_b^1, \omega_b^2) = (\omega_b^1, 1-\omega_b^1)$, can be represented in a two-dimensional diagram as in panel (b) of Figure 1. The set of budget shares contained by their respective Varian support sets at B_1 and B_2 is given by the shaded area $AB \cup BD \cup AC \cup CD$.

on the location of demands. The revealed preference support set then coincides with the new budget hyperplane.
Blundell *et al.*(2003, 2008, 2012) develop a methodology using nonparametric Engel curves to effectively control for budget variation and provide tighter bounds on demand responses at new budget regimes. This methodology requires the pooling information across individuals to estimate the nonparametric Engel curves. In principle this extension is easily accommodated by the methodology developed in this essay; rather than define the Varian support set using D^i, one would use the set of Sequential Maximum Power demands. However, I leave the integration of this methodological extension into my own approach to a later date as I aim to allow for maximal preference heterogeneity in my later empirical application, which is curtailed by the necessary estimation of Engel curves. However, the SMP methodology is applied in Part IV to increase the power of my procedure for uncovering preference change in survey data.

However, budget-wise membership of the Varian support set is not sufficient for rationality. Demand predictions must satisfy the cross-consistency requirement. Cross-consistency of demand predictions imposes the following logical conditions upon demand responses:

$$\begin{array}{lccc} [1] & \begin{pmatrix} \mathbb{R}_{12} = \{\mathtt{True}\} \\ \boldsymbol{\rho}_1^{'}\mathbf{x}_1 \geqslant \boldsymbol{\rho}_1^{'}\mathbf{x}_2 \end{pmatrix} & \vee & \begin{pmatrix} \mathbb{R}_{12} = \{\mathtt{False}\} \\ \boldsymbol{\rho}_1^{'}\mathbf{x}_1 < \boldsymbol{\rho}_1^{'}\mathbf{x}_2 \end{pmatrix} \\ [2] & \begin{pmatrix} \mathbb{R}_{21} = \{\mathtt{True}\} \\ \boldsymbol{\rho}_2^{'}\mathbf{x}_2 \geqslant \boldsymbol{\rho}_2^{'}\mathbf{x}_1 \end{pmatrix} & \vee & \begin{pmatrix} \mathbb{R}_{21} = \{\mathtt{False}\} \\ \boldsymbol{\rho}_2^{'}\mathbf{x}_2 < \boldsymbol{\rho}_2^{'}\mathbf{x}_1 \end{pmatrix} \\ [3] & \mathbb{R}_{12} & \rightarrow & \boldsymbol{\rho}_2^{'}\mathbf{x}_2 \leqslant \boldsymbol{\rho}_2^{'}\mathbf{x}_1 \\ [4] & \mathbb{R}_{21} & \rightarrow & \boldsymbol{\rho}_1^{'}\mathbf{x}_1 \leqslant \boldsymbol{\rho}_1^{'}\mathbf{x}_2 \end{array} \tag{2.1}$$

where the Boolean variable $\mathbb{R} = \{\mathtt{True}, \mathtt{False}\}$ can be interpreted as the Revealed Preferred Relation given in Definition 1.2.

In words, condition [1] requires that either an individual prefers their chosen demand at budget 1 ($\mathbb{R}^0_{12} = \{\mathtt{True}\}$), or they prefer their demand at budget 2 ($\mathbb{R}^0_{12} = \{\mathtt{False}\}$). If an individual reveals their preference for $\mathbf{x}_1$, then it must be the case that they chose $\mathbf{x}_1$ despite $\mathbf{x}_2$ being feasible. Condition [2] is analogous. Conditions [3] and [4] impose consistency with GARP upon the preference relation.

In Figure 1, cross-consistency of demand predictions requires that if $\mathbf{x}_1 \in A$, then $\mathbf{x}_2 \notin B$ and if $\mathbf{x}_2 \in B$, then $\mathbf{x}_1 \notin A$. The characterisation of the problem in budget share space highlights the impact of the cross-consistency requirement upon the support set. Considering panel (b) of 1, we have that $\{\omega^i_b\}_{b=1,2} \in (BD \cup AC \cup CD)$ but, unlike previously, $\{\omega^i_b\}_{b=1,2} \notin AB$. Thus, in this example, cross-consistency implies a nonconvex support set.

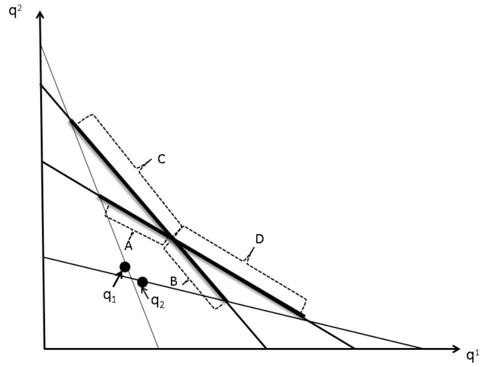
formal characterisation The "sufficient support set", $S^{S,i} = S^S(B|D^i)$, for the set of budgets $B = \{\boldsymbol{\rho}_b, m_b\}_{b=1,...,|B|}$ is the set of demand responses at each new budget $b \in B$ that satisfy GARP internally and when conjoined with existing observations.

definition 2.2 Given $D^i = \{\mathbf{q}^i_t, \mathbf{p}_t\}_{t=1,...,T}$ and the set of budgets $B = \{\boldsymbol{\rho}_b, m_b\}_{b=1,...,|B|}$, the "sufficient support set" is defined as:

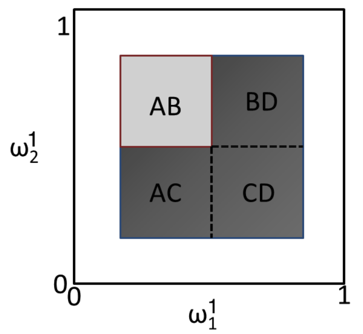
$$S^S(B|D^i) = \left\{ \Omega^i = \{\mathbf{x}^i_b\}_{b=1,...,|B|} : \begin{array}{l} \mathbf{x}^i_b \geqslant \mathbf{0} \\ \boldsymbol{\rho}^{'}_b \mathbf{x}^i_b = m_b \\ \{D^i; \boldsymbol{\rho}, \Omega^i\} \text{ satisfy GARP} \end{array} \right.$$

mixed integer programme The nonconvexity of the sufficient support set creates difficulties for optimisation over its elements. These complications have not been fully appreciated in the literature until this point. However, the constraints that define $S^{S,i}$ can be equivalently formulated as a mixed integer linear programme (MILP) that is more easily operationalised.

Figure 1: Example of Nonconvex Support Set: AB Ruled Out



(a) Good Space



(b) Budget Share Space

DEFINITION 2.3 Given $D^i = \{\mathbf{q}_t^i, \mathbf{p}_t\}_{t=1,\dots,T}$ and the set of budgets $B = \{\rho_b, m_b\}_{b=1,\dots,|B|}$, the mixed integer linear programming characterisation of the "sufficient support set" is defined as:

$$S^S(B|D^i) = \left\{ \Omega^i = \{\mathbf{x}_b^i\}_{b=1,\dots,|B|} : \begin{array}{rcll} \mathbf{x}_t^i & \geqslant & \mathbf{0} & [1] \\ \rho_t'\mathbf{x}_t^i & = & m_t & [2] \\ \rho_t'\mathbf{x}_t^i & > & 1 - R_{ts} & [3] \\ \rho_t'\mathbf{x}_s^i & \geqslant & R_{st}(\rho_t'\mathbf{x}_t^i) & [4] \\ R_{ts} + R_{su} & \leqslant & 1 + R_{tu} & [5] \\ R_{ts} & \in & \{0, 1\} & [6] \end{array} \right.$$

for all $t, s, u \in \{1, \dots, T\} \cup B$.

These constraints are linear in unknowns and provide an operational methodology with which to practically characterise the sufficient support set. Constraint [6] links the integer variable R_{ts} to the revealed preference relation. The logical occurrence $\mathbb{R}_{ts} = \{True\}$ is computationally represented by $R_{ts} = 1$ and, alternatively, $\mathbb{R}_{ts} = \{False\}$ is computationally represented by $R_{ts} = 0$.

Constraints [1] and [2] represent the standard nonnegativity and adding up requirements. Constraints [3] and [4] impose GARP upon the revealed preference restriction: for any bundle $\mathbf{x}_s$ that is revealed preferred to a bundle $\mathbf{x}_t$ (and thus $R_{st} = 1$), it must be the case that $\mathbf{x}_s$ is more expensive than $\mathbf{x}_t$ given $\boldsymbol{\rho}_t$. Constraint [4] imposes transitivity upon the revealed preferred relation: for any budget t that is preferred to budget s ($R_{ts} = 1$), that is preferred to budget u ($R_{su} = 1$), it must be the case that t is preferred to u ($R_{tu} = 1$). These constraints remain computationally demanding to implement but this can be achieved using relatively standard mixed integer linear programming techniques. This characterisation of the sufficient support set is quite different from that of the traditional Varian support set as one does not simply define a set of demand bundles but also a specification for the direct revealed preference relation.

5.4 BEYOND GARP

The concept of the support set can be applied to models beyond those considered in the current theoretical setting. This essay is concerned with predicting demands that are consistent with the behavioural hypothesis of utility maximisation. The definition of the support set that has been presented in this section reflects this fact; elements of $S^{S,i}$ jointly satisfy GARP. However, one can construct support sets for alternative models of choice. Elements of these alternative support sets would simply satisfy the consistency requirement that is imposed by the new model of interest. For example, if one was concerned with predicting household behaviour, they would define the support set as those demands consistent with their preferred model of collective choice.[4] Thus, the scope of the revealed preference approach is not limited to the traditional rational choice framework and the methodology developed in this essay can, in principle, be extended to a wide range of scenarios.

However, one should note that a general feature of the revealed preference approach to forecasting is non-unique predictions. Given finite data, the support sets associated with general models typically contain an infinite number of elements as the empirical hypotheses yielded by theory are typically qualitative in nature and are thus ill-defined.

SUMMARY This section has provided a revealed preference characterisation of the global rationality of demand predictions. In so doing, I have outlined a nonparametric answer to the question: what will

4 For example, consistency with the collective model could be imposed if they were to assume a cooperative decision making process.

demands be at a set of new budgets? I have defined the characteristics of the support set when multiple demand predictions are made and highlighted that the requirement of cross-consistency of demand predictions implies that this set is nonconvex. A MILP characterisation of the sufficient support set was then developed. This characterisation yields an operational methodology with which to practically construct $S^{S,i}$ given a finite number of observations on past choice behaviour. In so doing, this section has extended and clarified the revealed preference approach to the forecasting of demand predictions.

The next section outlines how regression techniques are traditionally used to yield demand predictions at new budgets of interest and explores why these predictions may violate global rationality to lie beyond the support set.

6

UNCONSTRAINED NONPARAMETRIC REGRESSION

In the last section, I developed a revealed preference characterisation of the set of globally rational demand predictions on a set of budgets, B. In this section, I briefly consider why unconstrained demand estimates that are yielded by traditional nonparametric regression techniques may violate global rationality and, therefore, do not necessarily constitute elements of the sufficient support set even if utility maximisation is a valid behavioural hypothesis.

FINITE DATA The behavioural hypothesis of utility maximisation, in conjunction with a finite set of observations on an individual's past demand behaviour, typically identifies a set of rational demand predictions at a new budget. However, if one assumes that the demand function is smooth, continuous and of a structure that is approximately locally constant, then observations in the close neighbourhood of a new budget of interest are informative for the location of predicted demands. Incorporating this information into one's forecasts yields more precise demand predictions. Let us specify the individual demand model as:

$$\mathbf{x}^i = \boldsymbol{\phi}^i(\boldsymbol{\rho}, m) \tag{2.2}$$

Given a long enough panel on an individual's past choice behaviour, time series variation in prices and income in the neighbourhood of new budgets of interest would facilitate nonparametric identification of $\mathbf{x}^i = \boldsymbol{\phi}^i(\boldsymbol{\rho}, m)$.[1] However, although "big T" data sets are increasingly available, the possibility of approximating asymptotia at the individual level is beyond the applied demand theorist. Thus, attempts to identify the local value of $\mathbf{x}^i = \boldsymbol{\phi}^i(\boldsymbol{\rho}, m)$ using only the finite past choice data on the individual, D^i, are generally associated with prediction error and low quality conclusions deriving from the issues that arise when one operates with a paucity of data.

The curse of dimensionality is one specific manifestation of this fact. The rate of convergence of nonparametric estimators to the truth is decreasing in the number of dimensions; intuitively, the majority of high dimensional space is "far away" from the centre. This implies that when measuring distances between coordinates in high-dimensional space, there will be little difference in the distances between different pairs of samples. This hinders the ability of nonparametric estimators

1 This is so given the consistency of kernel and alternative nonparametric estimation methods.

to capture local variation of the demand function. The only circumstance in which one can be sure of high quality estimation output with a small number of choice observations is when individual preferences are Cobb-Douglas. With Cobb-Douglas preferences there is no variation in observed budget shares and past observations are perfectly informative for demands at new budgets. However, the assumption of Cobb-Douglas preferences is decisively violated by consumer choice data.

6.1 HETEROGENEITY & POOLING

The standard approach that is taken to circumvent the problems associated with a shortage of information on an individual's past choice behaviour is to pool data across consumers' and estimate a statistical demand function. This requires that one make assumptions regarding the similarity of individuals, which, in effect, enable the applied demand theorist to assume away the fact that she operates in a low information environment. For example, if one were to assume that consumers were identical, it would be as if they had access to much more 'individual' level data with which to predict an individual's demands; data on individual i's past choice behaviour would be informative for predicting individual j's demands at new budgets.

However, the view once espoused by Becker and Stigler that preferences "are the same to all men" (1977, p.76) is undermined by patterns in consumption microdata that point to pervasive unobserved taste heterogeneity across otherwise observationally equivalent individuals. For example, least squares regressions explain a very limited amount of the variation in microdata (Hoderlein 2009; Pendakur and Lewbel 2009). Given such pervasive unobserved preference heterogeneity, assumptions must be imposed upon the functional form, and distribution, of taste differences across the population in order to make use of information across consumers.

To capture preference variation in the population, let us define the random variable $V \in \mathcal{V}$, where $\mathcal{V}$ is a Borel space that denotes preferences. Preference heterogeneity can be partly attributed to observable differences between individuals, given by the real valued J-vector O^i, and unobservable differences in tastes, given by the random variable U^i, taking values on the Borel space $\mathcal{U}$. Thus, $V = f(O, U)$, where f is a fixed $\mathcal{V}$-valued mapping defined on all sets of possible values of (O, U). This allows us to express the individual demand model as:

$$\mathbf{x}^i = \boldsymbol{\phi}^i(\boldsymbol{\rho}, m) = \Phi(\boldsymbol{\rho}, m, V^i) = \Phi(\boldsymbol{\rho}, m, O^i, U^i) \tag{2.3}$$

The population level demand model can then be specified as:

$$\mathbf{X} = \Phi(\boldsymbol{\rho}, m, O, U) \tag{2.4}$$

where $\mathbf{X}$, O and U are the stacked vectors of demand, observable and unobservable taste vectors over individuals.

Matzkin (2003) proves that such a nonadditive model is not identified with finite data on individuals' choice data in the population. Proceeding requires one to impose additional functional assumptions on the population demand model over the distribution of U and its interaction with O. A standard approach in the nonparametric literature is to capture unobserved taste heterogeneity, U, via an additive error term:

$$\mathbf{X} = \Phi(\boldsymbol{\rho}, m, O) + U \tag{2.5}$$

where U is mean independent of regressors.

This model is easily estimated via standard techniques but the returned average demands are likely to violate a global rationality requirement. As explored by Lewbel (2001), this framework, although attractive when judged according to the criteria of specification simplicity and estimation ease, is very restrictive on the form of preference heterogeneity in the population. Even when unobserved preference parameters are independent of observables, $\Phi(\boldsymbol{\rho}, m, O)$ will violate rationality restrictions unless every consumer has homothetic preferences.[2] Homotheticity is an exceptionally strong assumption that is violated in consumption data (see, for example, Cherchye *et al.* 2013) and thus irrational predictions can be yielded by such a strategy even if all individuals in the population satisfy the "homo economicus" benchmark.

6.2 IMPOSING RATIONALITY

Functional misspecification of the population demand model, and prediction errors deriving from the finite nature of past choice data, can cause the forecasts yielded by unconstrained nonparametric regressions to violate rationality. In this eventuality, one's demand predictions lie in the complement of the revealed preference support set. Violations of rationality have implications for applied researchers who want to engage in welfare analysis. No single utility function exists that can rationalise these demand predictions and standard measures of the welfare of a consumer at these demands cannot be computed. Therefore, methods applied to these predictions to estimate consumer surplus and other welfare metrics will suffer from path dependency.

In a parametric setting, rationality can be imposed upon predictions in a relatively simple and well-trodden manner using system estimation methods to restrict parameter estimates across equations. However, a method for imposing global rationality upon a full nonparametric demand system does not yet exist. Efforts to date have only imposed the local satisfaction of certain necessary requirements

2 Such preferences are also referred to as "conical" in the literature. See Afriat (1977).

for rationality. Kim and Tripathi (2003) discuss the imposition of homogeneity upon a nonparametric demand system. Haag, Hoderlein and Pendakur (2009) develop a method for estimation of the full nonparametric demand system under symmetry. Recently, Blundell, Horowitz and Parey (2012) differentially weight observations to ensure satisfaction of the Slutsky constraint in their analysis of gasoline demand.

Considering Blundell et al's (2012) method in more detail, a nonparametric kernel estimator of the gasoline demand function is presented that imposes the local satisfaction of the Slutsky constraint:

$$\frac{\partial \Phi(\rho, m, O)}{\partial \rho'} + \Phi(\rho, m, O)\frac{\partial \Phi(\rho, m, O)}{\partial m} \leqslant 0 \tag{2.6}$$

$\Phi(\rho, m, O)$ is estimated on a grid of (ρ, m) points, with observations weighted differentially at each node to ensure that the Slutsky constraint is respected. However, given the pointwise nature of this procedure, there is no guarantee that the function satisfies the constraint globally; the Slutsky condition may be violated across points on the grid and, thus, one's predictions may not be jointly rationalisable by a single utility function. Furthermore, existing methods are not sufficient given that they do not jointly impose the full set of the conditions required for rationality; negative semidefiniteness of the Slutsky matrix cannot be not imposed on nonparametric estimation techniques that proceed pointwise as it is a global property that must hold across points on the grid.

SUMMARY In summary, nonparametric regression techniques facilitate the recovery of point predictions for individual demands at new budget regimes. If infinite data on a rational individual's past choice behaviour was available, perfect prediction of rational demands would be possible as quantities at a new budget of interest could be perfectly inferred from observed behaviour in the close neighbourhood. However, there is very limited data on individual choice behaviour with which to directly recover the individual demand function. It is thus common to proceed by imposing assumptions upon the functional form of taste differences in a population such that information across consumers can be pooled and a statistical demand function estimated. Prediction errors that derive from the finite nature of panel data on individual choice, and misspecification bias in estimated demand functions, can cause one's recovered demand predictions to violate global rationality and thus lie in the complement of the support set. The next section gives my proposed method for imposing global rationality upon unconstrained nonparametric demand predictions.

7 MINIMUM DISCRIMINATION INFORMATION

In this section, I consider how to impose global rationality upon nonparamterically recovered demand predictions. I outline how the Principle of Minimum Discrimination Information (MDI) can be applied to combine the revealed preference and nonparametric estimation approaches to demand extrapolation. The Principle of Minimum Discrimination Information originates from the work of Solomon Kullback and Richard Leibler, and generalises the Principle of Maximum Entropy that was proposed by Edwin Jaynes (1957a, 1957b).

7.1 REFORMULATING THE PROBLEM

For the rest of this essay I suppress the superscript i on individual demands for notational simplicity. Please do note that all that follows is defined at the individual level.

As when motivating the concept of support set in Section 5, let us start with the problem of predicting demands at a single new budget, $\{\boldsymbol{\rho}_b, m_b\}$. Any demand bundle, $\mathbf{x}_b$, that exactly exhausts $\{\boldsymbol{\rho}_b, m_b\}$ can be expressed as a weighted combination of extreme budget simplex vertices.

$$\mathbf{x}_b = \mathbf{A}_b \boldsymbol{\omega} \tag{2.7}$$

where

$$\mathbf{A}_b = \begin{bmatrix} m_b/\rho_b^1 & 0 & \cdots & 0 \\ 0 & m_b/\rho_b^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & m_b/\rho_b^K \end{bmatrix}$$

and $\boldsymbol{\omega}$ is a $K-$vector of non-negative budget shares. The budget share simplex is then a $(K-1)$-dimensional manifold in K dimensional space. Formally, it is the set of points $\boldsymbol{\omega} \in \mathbb{R}_+^K$ such that $\omega_k \geqslant 0$ and $\sum_k \omega_k = 1$.

To incorporate the demand predictions yielded by nonparametric regression techniques, let $\widetilde{\mathbf{x}}_b$ represent an unconstrained demand response at $\{\boldsymbol{\rho}_b, m_b\}$, achieved from some unspecified estimation exercise in the spirit of the previous section. I will refer to the unconstrained estimate as a "pilot" prediction. $\widetilde{\mathbf{x}}_b$ can also be expressed in terms of budget shares:

$$\widetilde{\mathbf{x}}_b = \mathbf{A}_b \boldsymbol{\eta} \tag{2.8}$$

where A_b is defined as previously.

As was discussed in the last section, $\widetilde{\mathbf{x}}_b$ may violate rationality. Given the maintained behavioural hypothesis that individuals behave rationally, and my aim to use demand predictions for welfare analysis, $\widetilde{\mathbf{x}}_b$ must be updated to lie within the revealed preference support set, S^S. This can be achieved by modifying $\boldsymbol{\eta}$. However, economic theory does not identify a unique way of manipulating $\boldsymbol{\eta}$ to attain rationality. Even the "closest" element of the support set to the pilot estimate is not uniquely determined as there are multiple distance measures that can be applied to quantify dissimilarity. A well defined additional principle is thus required to select an element of the support set.

Taking the unconstrained pilot estimate, $\boldsymbol{\eta}$, and the information conveyed by economic theory that $\mathbf{A}_b\boldsymbol{\omega} \in S_b^V$, the updated budget share specification, $\boldsymbol{\omega}^\star$, that results from taking S_b^V into account will be chosen by minimising some functional $H(\boldsymbol{\omega}, \boldsymbol{\eta})$ in the set S_b^V.

$$H(\boldsymbol{\omega}, \boldsymbol{\eta}) = \min_{\mathbf{A}_b\boldsymbol{\omega}' \in S_b^V} H(\boldsymbol{\omega}', \boldsymbol{\eta}) \tag{2.9}$$

Given that ω^j's and η^j's have all the prerequisite characteristics of probabilities[1], they can be manipulated as such. This methodology then corresponds to the application of an abstract information operator, which takes an unrestricted pilot estimate, $\boldsymbol{\eta}$, and the information conveyed by economic theory to yield an updated theory-consistent result, $\boldsymbol{\omega}^\star$:

$$\boldsymbol{\omega}^\star = \boldsymbol{\eta} \circ S_b^V \tag{2.10}$$

7.2 The Kullback-Leibler Divergence

In this essay, I propose that this information operator should embody the Principle of Minimum Discrimination Information (MDI). This principle dictates that one selects $\boldsymbol{\omega}$ in order to minimise the functional:

$$H(\boldsymbol{\omega}, \boldsymbol{\eta}) = D(\boldsymbol{\omega} \| \boldsymbol{\eta}) = \boldsymbol{\omega}' \log_2 \frac{\boldsymbol{\omega}}{\boldsymbol{\eta}} \tag{2.11}$$

The Principle of MDI is a generalisation of the principle of maximum entropy proposed by Jaynes (1957). It corresponds to the minimisation of the Kullback-Leibler (KL) divergence between $\boldsymbol{\omega}$ and $\boldsymbol{\eta}$, $D(\boldsymbol{\omega}\|\boldsymbol{\eta})$. The technique has its historical roots in information theory and statistical physics and has been applied to answer ill-posed problems analogous to our own in a diverse range of fields (see Shore and Johnson (1981) for an early overview of the approach and Golan, Judge and Miller (1996) for more recent applications in statistics).

1 $\boldsymbol{\omega}$ are additive, non-negative, $\boldsymbol{\omega} \geq 0$, and sum to unity, $\mathbf{1}'\boldsymbol{\omega} = 1$.

In applying the principle of MDI, the "closest" rational budget share to the pilot, $\boldsymbol{\eta}$, is selected. The dissimilarity measure that is minimised during this procedure is the Kullback-Leibler (KL) divergence. The KL divergence gives a measure of how easy it is to discriminate between two distributions. It represents the amount of information, measured in bits, that is required to update an unconstrained pilot estimate with rationality. It is not a proper distance since it is not symmetric and it does not satisfy the triangular inequality.[2] However, it shares the properties of weak positivity, $D(\boldsymbol{\omega}||\boldsymbol{\eta}) \geqslant 0$, and $D(\boldsymbol{\omega}||\boldsymbol{\eta}) = 0 \longleftrightarrow \boldsymbol{\omega} = \boldsymbol{\eta}$ with a distance. Unlike other dissimilarity measures, the KL divergence has a probabilistic interpretation. As we will explore in Section 8, this makes application of the MDI extremely compelling in our context.

SINGLE PREDICTIONS For a demand prediction at $\{\boldsymbol{\rho}_b, m_b\}$, adherence to the principle of MDI dictates that one selects their rationality-updated budget share as the solution to the following maximisation problem:

$$\max_{\boldsymbol{\omega}} -D(\boldsymbol{\omega}||\boldsymbol{\eta}) = -\boldsymbol{\omega}' \log_2 \frac{\boldsymbol{\omega}}{\boldsymbol{\eta}} \tag{2.12}$$

subject to:

$$\begin{aligned} &\boldsymbol{\omega} \geq 0 \\ &\mathbf{1}'\boldsymbol{\omega} = 1 \\ &\mathbf{A}_b\boldsymbol{\omega} \in S_b^V \end{aligned}$$

There is no closed form solution expressed directly in terms of known variables for the updated MDI demand estimates. However, an indirect solution expressed as a function of the Lagrangian multipliers on the maximsation constraints is easily obtained. The Lagrangian associated with the optimisation problem above takes the form:

$$\mathcal{L} = D(\boldsymbol{\omega}||\boldsymbol{\eta}) + \beta\mathbf{1}'\boldsymbol{\omega} - \sum_{t \in RW(\boldsymbol{\rho}_b, m_b|D)} \lambda_t \mathbf{p}_t' \mathbf{A}_0 \boldsymbol{\omega} \tag{2.13}$$

where $RW(\boldsymbol{\rho}_b, m_b|D)$ denotes the set of past observations that are revealed worse to the new budget.

2 The KL divergence is not necessarily symmetric,

$$\boldsymbol{\omega}' \log_2 \frac{\boldsymbol{\omega}}{\boldsymbol{\eta}} \neq \boldsymbol{\eta}' \log_2 \frac{\boldsymbol{\eta}}{\boldsymbol{\omega}}$$

and does not satisfy the triangle inequality,

$$D(\boldsymbol{\omega}||\boldsymbol{\eta}) \nleqslant D(\boldsymbol{\omega}||\boldsymbol{\delta}) + D(\boldsymbol{\delta}||\boldsymbol{\eta})$$

Definition 2.4 The observation $\{\mathbf{p}_t, \mathbf{q}_t\}$ is directly revealed worse to a new budget $\{\boldsymbol{\rho}_b, m_b\}$, $\mathbf{p}_t \mathbb{W}^0 \boldsymbol{\rho}_b$, if $\boldsymbol{\rho}_b' \mathbf{q}_t < m_b$.

Definition 2.5 The observation $\{\mathbf{p}_t, \mathbf{q}_t\}$ is revealed worse to a new budget $\{\boldsymbol{\rho}_b, m_b\}$, $\mathbf{p}_t \mathbb{W} \boldsymbol{\rho}_b$, if $\mathbb{W}$ is the transitive closure of $\mathbb{W}^0$.

A formal expression for the updated budget share for good k at the new budget of interest is given by:

$$\omega_b^{k,\star} \leqslant \eta^k \exp\left(\sum_{t \in RW(\rho_b, m_b|D)} \lambda_t \frac{p_t^k}{\rho_b^k} m_b - \lambda_0\right) \tag{2.14}$$

where $\lambda_0 = \beta + 1$.

Multiple predictions Updating pilot predictions over the set of budgets $B = \{\boldsymbol{\rho}_b, m_b\}_{b=1,\dots,|B|}$, requires that one impose joint membership of the sufficient support set, S^S, on the full set of pilot predictions simultaneously.

Let the set of unconstrained pilot budget share predictions over B be given by $\{\boldsymbol{\eta}_b\}_{b=1,\dots,|B|}$. I seek the set of rationality-updated budget shares $\{\boldsymbol{\omega}_b^\star\}_{b=1,\dots,|B|}$ on B. With multiple predictions and the simultaneous imposition of rationality, the MDI optimisation problem is given by:

$$\max_{\{\boldsymbol{\omega}_b\}_{b=1,\dots,|B|}} -\frac{1}{|B|}\sum_{b=1}^{|B|} D(\boldsymbol{\omega}_b \| \boldsymbol{\eta}_b) = -\frac{1}{|B|}\sum_{b=1}^{|B|} \boldsymbol{\omega}_b' \log_2 \frac{\boldsymbol{\omega}_b}{\boldsymbol{\eta}_b} \tag{2.16}$$

subject to:

$$\begin{aligned} &\boldsymbol{\omega}_b \geqslant \mathbf{0} && \text{for } b = 1, \dots, |B| \\ &\mathbf{1}'\boldsymbol{\omega}_b = 1 && \text{for } b = 1, \dots, |B| \\ &\Omega = \{\mathbf{A}_b \boldsymbol{\omega}_b\}_{b=1,\dots,|B|} \in S^S && \end{aligned}$$

where $\mathbf{A}_b$ is the matrix parameterising the budget b.

Computationally, this is a much harder maximisation problem to solve because the sufficient support set, S^S, is nonconvex. We can operationalise this optimisation problem using the MILP characterisation of S^S that was developed in Section 5. Although algorithms exist to solve this optimisation problem, it remains a computationally burdensome procedure and the quality of the solution output falls quickly in $|B|$. I refer to Grossmann and Kravanja (1997) for an overview of mixed integer programming techniques, which includes a treatment of available solution algorithms and an overview of applications of the methodology.

Two-step procedure Imposing membership of S^S simultaneously on $\{\boldsymbol{\eta}_b\}_{b=1,\dots,|B|}$ typically yields a different solution to that which would be gained by imposing membership of the Varian support set budget-by-budget because predictions must satisfy the cross-consistency requirement. However, the two strategies are equivalent in two situations. Firstly, when the set of solutions yielded by imposing pointwise membership of the Varian support set constitutes an element of S^S. In this special case, the solution for rational demand predictions over B is invariant to whether one solves the single MILP optimisation problem or whether one imposes membership of the Varian support set budget-by-budget.

PROPOSITION 2.2 Let $\{\boldsymbol{\omega}_b^{V\star}\}_{b=1,...,|B|}$ represent the solution set yielded by imposing membership of the Varian support set S_b^V at each budget $b = 1, ..., |B|$. Let $\{\boldsymbol{\omega}_b^{S\star}\}_{b=1,...,|B|}$ represent the solution set when a single maximisation problem is solved to impose membership of the sufficient support set, S^S.

$$\text{If } \{\mathbf{A}_b \boldsymbol{\omega}_b^{V\star}\}_{b=1,...,|B|} \in S^S, \text{ then } \boldsymbol{\omega}_b^{V\star} = \boldsymbol{\omega}_b^{S\star} \text{ for } b = 1, ..., |B|$$

PROOF. See Appendix A.

As mentioned above, implementation of the mixed integer optimisation problem is a computationally burdensome procedure. Therefore, Proposition 2.2 is a deeply useful result that justifies the use of a two-step procedure when predicting demands on a grid of points. At the first stage, one imposes membership of the Varian support set node-by-node. If the resulting solution set satisfies cross-consistency, then implementation of the single MILP maximisation problem is not required.

The second situation in which imposing of pointwise membership of the Varian support sets and joint imposition of membership of S^S are equivalent is when interest centres on the marginal demand curve of a single good. For this set of budgets, it is impossible for $\{\mathbf{A}_b \boldsymbol{\omega}_b^{V\star}\}_{b=1,...,|B|} \notin S^S$.

PROPOSITION 2.3 Let B^{M_k} define a set of budgets that represent a portion of the marginal demand curve for good k, i.e. for all $B_b, B_{b'} \in B^{M_k}$, we have that $\rho_b^k \neq \rho_{b'}^k$, $\boldsymbol{\rho}_b^{\neg k} = \boldsymbol{\rho}_{b'}^{\neg k}$ and $m_b = m_{b'}$. The solution yielded by imposing membership of the Varian support set, S_b^V, at each budget $b = 1, ..., |B^{M_k}|$ and the solution set yielded when a single maximisation problem is solved to impose membership of the sufficient support set, $S^S(B^{M_k}|D)$, are identical.

$$\boldsymbol{\omega}_b^{V\star} = \boldsymbol{\omega}_b^{S\star} \quad \text{for any} \quad B_b \in B^{M_k}$$

PROOF. See Appendix A.

Applied interest often centres on the marginal demand function for which imposition of pointwise membership of the Varian support sets is sufficient for global rationality. In this setting, Proposition 2.3 gives us that imposing global rationality is not too onerous a procedure.

SUMMARY This section has outlined how the Principle of MDI can be applied to impose global rationality upon a set of unconstrained demand predictions. In so doing, I have provided a strategy for combining the revealed preference and nonparametric regression approaches to the demand forecasting problem. I explicitly recast the demand prediction problem as an ill-posed inverse problem in order to select the demand within the revealed preference support set that minimises the Kullback-Leibler divergence with an unconstrained pilot estimate. When predicting demands at a single budget, this optimisation problem is easily operationalised as a simple linear programming exercise. However, when predicting over a set of new budgets, the revealed preference constraints take the form of a MILP, which yields a much more burdensome computational exercise. I have shown that one is justified in following a two-step procedure in the multiple prediction case, to first impose pointwise membership of the Varian support sets before implementing the full optimisation procedure, and that the full optimisation procedure is unnecessary when interest centres on the marginal demand function of a single good.

8 JUSTIFICATION OF MDI DEMANDS

The last section outlined how the Principle of MDI can be used to update nonparametric pilot predictions with rationality. However, the application of this principle requires motivation. As I discussed previously, the problem of how to move between an unconstrained pilot and a rational prediction is underdetermined; there are an infinite number of distance measures that one could use to define the "closest" rational demand to a pilot.

This section outlines my justification for the application of the principle of MDI. I argue that the principle of MDI is a compelling criterion to adopt for three main reasons. First, the MDI result is the most likely rational demand to occur given the belief that expenditure decisions are governed by the pilot distribution. It is, therefore, the least surprising result to the statistician who is ignorant of economic theory. Second, it corresponds to the demand that minimises the predictive loss of the statistician. Therefore, it can be regarded as the most optimistic updating rule followed by an econometrician looking to make use of the statistician's estimates. Finally, the objective function associated with the approach has a number of attractive features that support uniqueness of the updated solution and preserves assumptions that a researcher may impose that concern the separability of preferences.

8.1 MAXIMUM EPISTEMIC LIKELIHOOD

The problem that I face is one of how to best update unconstrained nonparametric pilot predictions with economic theory in order to recover unique rational demand predictions at each $b \in B$. To aid discussion of my approach, I make a distinction between the information sets that statisticians and econometricians exploit when constructing their demand predictions. Statisticians are ignorant of economic theory and make their predictions only on the basis of past choice behaviour. However, econometricians are able to incorporate information from economic theory: individuals are rational and thus demand predictions must lie within the support set.[1]

Consider again the problem of predicting a single demand at the budget $\{\boldsymbol{\rho}_b, m_b\}$. Past observations lead the statistician to predict the budget share $\boldsymbol{\eta} = [\eta^1, ..., \eta^K]$. This prediction implies that she expects a proportion η^k of the new budget to be spent on good k or,

1 I remind the reader that this is an underlying assumption in this essay, which is justified if D satisfies GARP.

equivalently, that she assigns a η^k probability to the event that an infinitely small unit of expenditure drawn at random from the consumer's realised demand was spent on good k. However, the statistician's prediction does not necessarily satisfy rationality. Application of the Principle of MDI selects the rational demand that is most likely to be realised in an independent and identically distributed (iid) sample drawn from the distribution $\boldsymbol{\eta}$. Therefore, it can be considered the "least surprising" or "most epistemically likely" rational demand to the statistician who made their demand prediction without the guidance of economic theory.

What does it mean for a demand to be "least surprising"? Imagine that demand at $\{\boldsymbol{\rho}_b, m_b\}$ is realised and that the statistician's best prediction was $\mathbf{x}_b = \mathbf{A}_b\boldsymbol{\eta}$. To verify this prediction, she randomly draws a sample of n expenditure units with replacement and observes the allocation of these expenditure units across the K goods. This sample is formally represented by $\Updownarrow = \{\kappa_1, ..., \kappa_n\}$, where $\kappa_n \in \{1, ..., K\}$ denotes the good that the n^{th} sampled expenditure unit is allocated to. The budget share realised in the sample is given as $\boldsymbol{\omega} = [\omega^1, ..., \omega^K]$, where $\omega^k = \frac{1}{n}\sum_n I(\kappa_n = k)$ for $k = 1, ..., K$.

The "epistemic probability " that the statistician assigns to the realisation of a particular budget share $\boldsymbol{\omega}$ in this sample, given her belief that the underlying budget share is $\boldsymbol{\eta}$, is inversely proportional to that budget share's distance from $\boldsymbol{\eta}$:

$$\widetilde{Pr}(\boldsymbol{\omega}|\boldsymbol{\eta}, n) \propto 2^{-nD(\boldsymbol{\omega}|\boldsymbol{\eta})} \tag{2.16}$$

All possible samples that generate a particular empirical budget share $\boldsymbol{\omega}$ have the same probability of occurrence. Thus, the probability that the statistician assigns to the realisation of a particular budget share $\boldsymbol{\omega}$, given her belief that the underlying budget share is $\boldsymbol{\eta}$, is derived as follows:

$$\begin{aligned}
\widetilde{Pr}(\boldsymbol{\omega}|\boldsymbol{\eta}, n) &= \prod_{k=1}^K (\eta^k)^{\sum_n I(\kappa_n = k)} \\
&= \prod_{k=1}^K (\eta^k)^{n\omega^k} \\
&= \prod_{k=1}^K 2^{n\omega^k \log_2 \eta^k} \\
&= \prod_{k=1}^K 2^{n(\omega^k \log_2 \eta^k \pm \omega^k \log_2 \omega^k)} \\
&= 2^{n\sum_{k=1}^K (-\omega^k \log_2 \frac{\omega^k}{\eta^k} + \omega^k \log_2 \omega^k)} \\
&= 2^{n(-D(\boldsymbol{\omega}|\boldsymbol{\eta}) + \boldsymbol{\omega}' \log_2 \boldsymbol{\omega})} \\
&\propto 2^{-nD(\boldsymbol{\omega}|\boldsymbol{\eta})}
\end{aligned} \tag{2.17}$$

Therefore, she believes that sample budget shares further from $\boldsymbol{\eta}$ are exponentially less likely to occur.

RATIONAL DEMANDS Econometricians make use of the same data as statisticians but also utilise the behavioural hypothesis of utility maximisation when constructing their demand predictions. The Principle of MDI dictates that one should make use of this information in a manner that minimally prejudices the statistician's prediction. The MDI demand, $\mathbf{A}_b\boldsymbol{\omega}^\star$, is the most probable rational demand to be realised in an independent sample drawn from $\boldsymbol{\eta}$. Therefore, it can be interpreted as the least surprising rational demand to the statistician who believed that the budget share at $\{\boldsymbol{\rho}_b, m_b\}$ was described by $\boldsymbol{\eta}$.

Rational demands at the single new budget $\{\boldsymbol{\rho}_b, m_b\}$ are those contained by the support set, S_b^V. The Conditional Limit Theorem (Csiszár, 1984) gives us that the statistician must believe that $\boldsymbol{\omega}$ is close to $\boldsymbol{\omega}^\star$ with very high probability conditional on the sample demand satisfying rationality.

THEOREM 2.1 Let $\mathbf{A}_b\boldsymbol{\eta} \notin S^V(\boldsymbol{\rho}_b, m_b)$. $\Updownarrow = \{\kappa_1, ..., \kappa_n\}$ is a sample of n expenditure-to-good allocations randomly drawn with replacement by the statistician at the budget $\{\boldsymbol{\rho}_b, m_b\}$. This sample generates the sample budget share $\boldsymbol{\omega}$, with $\omega^k = \frac{1}{n}\sum_n I(\kappa_n = k)$. The statistician believes that the underlying budget share describing demand is $\boldsymbol{\eta}$. Then, the epistemic probability assigned to the realisation of any rational budget share other than the MDI solution, $\boldsymbol{\omega}^\star$, is negligible as $n \to \infty$. Formally,

$$\widetilde{Pr}(|\omega^k - \omega^{k\star}| \geqslant \epsilon) \to 0 \tag{2.T1.1}$$

as $n \to \infty$, where

$$\boldsymbol{\omega}^\star = \arg \min_{\mathbf{A}_b\boldsymbol{\omega} \in S_b^V} D(\boldsymbol{\omega}\|\boldsymbol{\eta}) \tag{2.T2.2}$$

PROOF. See Appendix A.

This result provides a natural and intuitive justification for the application of the Principle of MDI to the problem at hand. The MDI demand is the rational demand that would be considered "most likely to occur" by the statistician who gave their best prediction as $\mathbf{A}_b\boldsymbol{\eta}$ in the absence of knowledge conveyed by economic theory.

8.2 MINIMUM FORECASTING LOSS

The MDI solution also corresponds to the demand that is associated with the econometrician's most optimistic assessment of the forecast losses associated with the statistician's prediction of $\mathbf{A}_b\boldsymbol{\eta}$. Logarithmic scoring rules are commonly used to provide summary measures for the evaluation of forecasts. With a logarithmic scoring rule, for each possible outcome $k = \{1, ..., K\}$, the forecaster announces her predicted probability η^k. Then, if outcome k occurs, the forecaster is paid $\log_2(\eta^k)$ as a reward. This scoring rule is "proper"; the expected value of the score, with respect to an agent's subjective beliefs, is strictly maximised by a true report of the agent's subjective beliefs, for all possible subjective beliefs.

PROPOSITION 2.4 The expected forecasting loss that the statistician reporting budget share $\boldsymbol{\eta}$ can expect when the true budget share is given by $\boldsymbol{\omega}$ is given by:

$$E[L(\boldsymbol{\omega}, \boldsymbol{\eta})] = m_b \left\{ \boldsymbol{\omega}' \log_2 \frac{\boldsymbol{\omega}}{\boldsymbol{\eta}} \right\}$$

PROOF. See Appendix A.

With the information that demand behaviour is rational, and thus that $\mathbf{A}_b \boldsymbol{\omega} \in S_b^V$, the econometrician's most optimistic assessment of the statistician's loss is that corresponding to the MDI demand.

$$\begin{aligned} \boldsymbol{\omega}^\star &= \min\nolimits_{\mathbf{A}_b \boldsymbol{\omega} \in S_b^V} m_b \left\{ \boldsymbol{\omega}' \log_2 \tfrac{\boldsymbol{\omega}}{\boldsymbol{\eta}} \right\} \\ &= \min\nolimits_{\mathbf{A}_b \boldsymbol{\omega} \in S_b^V} D(\boldsymbol{\omega} \| \boldsymbol{\eta}) \end{aligned} \tag{2.18}$$

Therefore, the Principle of MDI can be considered the most optimistic updating rule followed by the econometrician making use of the statistician's predictions.

8.3 INSTRUMENTAL JUSTIFICATIONS

I here outline instrumental features of the updating procedure that make the Principle of MDI a compelling criterion to adopt (see Shore and Johnson (1980) and Uffink (1995) for in-depth discussions of the principle's properties from a statistical physics perspective). With the exception of Property 2, the properties discussed in this subsection hold regardless of whether membership of the Varian or the sufficient support set is imposed. For expositional ease, the properties of MDI demands are discussed with reference to the prediction of a single demand, which only requires one to impose membership of the Varian support set. Proofs of all the properties in this section are given in Appendix A.

PROPERTY 1 The updated solution satisfies $\boldsymbol{\omega}^\star = \boldsymbol{\eta}$ if and only if $\boldsymbol{\eta} \in S_b^V$. In this instance, $D(\boldsymbol{\omega}^\star \| \boldsymbol{\eta}) = 0$.

The KL divergence is minimised when the pilot and updated demand predictions coincide. Therefore, if a pilot estimate is an element of the support set, it goes unmodified by the updating procedure as it uniquely achieves the minimum value of the objective function. This is what it means to update pilot predictions in accordance with the Principle of MDI: given new information, one should select the distribution that is hardest to discriminate from the original as possible.

PROPERTY 2 The updated solution $\boldsymbol{\omega}^\star$ is unique.

The solution is unique given a set of convex, bounded constraints. Given that we only defer to a principle beyond rationality because our original problem is ill-posed, uniqueness is a key asset of the approach.[2]

PROPERTY 3 The MDI approach treats goods symmetrically.

The identity of a particular good in the objective function is irrelevant. Therefore, the solution is invariant to permutations in the ordering of commodities in the objective function.

PROPERTY 4 The MDI approach preserves weak separability of preferences in the updating procedure.

In consumer theory it is common to decompose the consumer's choice space into weakly separable subgroups that break the demand prediction problem into smaller, more manageable units. Marginal rates of substitution within each separable subgroup are independent of goods lying outside the group. Thus, given an allocation of expenditure to a particular subgroup, optimal intragroup demands are independent of demands and prices for goods outside the group under consideration.

Additivity of the objective function implies that the MDI methodology preserves the implications of separability assumptions; the solution for within-group budget share allocations is independent of whether one updates demand predictions subgroup-by-subgroup or updates the full set of K budget shares in one step. This is important because there are two alternative ways to obtain rationality updated budget shares for each commodity subgroup. One could consider each commodity group separately to obtain a conditional within-group solution. Alternatively, one could obtain a rationality-updated solution for the whole system, and then compute the conditional within-group budget share.[3]

PROPERTY 5 Without specification of a pilot budget share, the updated solution corresponds to the budget share specification that is hardest to discriminate from that dictated by symmetric Cobb-Douglas preferences.

The Principle of MDI is a generalisation of the Principle of Maximum Entropy, originally proposed by Jaynes (1957). In the informa-

2 This is the only property in this section that does not necessarily hold when multiple predictions are made. The constraints associated with the sufficient support set are not necessarily convex. Therefore, we cannot guarantee uniqueness of the full system solution when $\{\mathbf{A}_b \boldsymbol{\omega}_b^{V\star}\}_{b=1,\dots,|B|} \notin S^S$

3 However, a further point to note is that the specification of the support set is more restrictive when weak separability is imposed because this property is informative for the location of demands at new regimes.

tion theory literature, the Principle of Maximum Entropy is seen as compatible with the goal of being wholly agnostic over the specification of a probability distribution. However, in this context, leaving the pilot unspecified is not equivalent to agnosticism over the location of rational demand responses. Rather, leaving the pilot estimate unspecified is equivalent to the maximum entropy formulation of the problem. The Principle of Maximum Entropy dictates that one selects the distribution that is hardest to discriminate from a uniform prior distribution over states. Symmetric Cobb-Douglas preferences dictate a symmetric distribution of real income across goods and, therefore, a uniform prior. Thus, leaving the pilot demand prediction unspecified imposes a strong parametric prior over the structure of preferences; the MDI demand corresponds to the rational demand that is hardest to discriminate from that dictated by symmetric Cobb-Douglas preferences. Recognition of this fact provides support for our focus on the Principle of MDI over the Principle of Maximum Entropy as more satisfactory pilots can be accommodated.

8.4 INFORMATIVENESS

I now consider the properties of MDI demands as additional observations on an individual's choice behaviour become available. The support set is weakly shrinking in additional observations and strictly shrinking in "informative" observations. Thus, as we move to a more informative environment, the predictive power of nonparametric economic theory increases, causing the influence of the pilot prediction upon the eventual MDI solution to wane.

In this section, I assume that η is invariant to increases in the panel of observations on an individual's past choice behaviour. This allows me to focus on the impact of additional choice observations upon the properties of the support set and the nature of the updating procedure. The consistency properties of the particular nonparametric estimator that is applied to yield η are clearly relevant for the speed of convergence of the MDI solution but are not of direct concern in our discussion of the updating rule.

Let $D_t = \{\mathbf{p}_i, \mathbf{q}_i\}_{i=1,\dots,t}$ give the set of past t observations and $S_t^S(B)$ denote the sufficient support set consistent with D_t. We can then define $\boldsymbol{\omega}_b^{t\star}$ and $H(D_t)$ as follows:

$$\begin{aligned} \{\boldsymbol{\omega}_b^{t\star}\}_{b=1,\dots,|B|} &\in \operatorname{argmin}_{\{\boldsymbol{\omega}_b\}_{b=1,\dots,|B|}} \begin{pmatrix} \frac{1}{|B|}\sum_{b=1}^{|B|} \boldsymbol{\omega}_b' \log_2 \frac{\boldsymbol{\omega}_b}{\boldsymbol{\eta}} \\ \{\mathbf{A}_b \boldsymbol{\omega}_b\}_{b=1,\dots,|B|} \in S_t^S(B) \end{pmatrix} \\ H(D_t) &= \frac{1}{|B|}\sum_{b=1}^{|B|} \boldsymbol{\omega}_b^{t\star\prime} \log_2 \frac{\boldsymbol{\omega}_b^{t\star}}{\boldsymbol{\eta}} \end{aligned}$$

This terminology allows us to formally define what it means for an observation to be informative.

DEFINITION 2.6 An additional observation $\{\mathbf{p}_{t+1}, \mathbf{q}_{t+1}\}$ is informative if $H(D_{t+1}) > H(D_t)$.

PROPOSITION 2.5 An observation $\{\mathbf{p}_{t+1}, \mathbf{q}_{t+1}\}$ is informative iff $\{A_b \boldsymbol{\omega}_b^{t\star}\}_{b=1,\ldots,|B|} \notin S_{t+1}^S(B)$.

PROOF. See Appendix A.

In words, an observation is informative in our context if it causes a modification in how the pilot prediction is updated. The first point to note is that an additional observation can never result in a reduction in the magnitude of updating, i.e. $H(D_{t+1}) \not< H(D_t)$. The size of the support set cannot increase with the introduction of an additional observation, $S_{t+1}^S(B) \subseteq S_t^S(B)$, and thus the "closest" element of $S_{t+1}^S(B)$ can never be closer to $\boldsymbol{\eta}$ upon observation of $\{\mathbf{p}_{t+1}, \mathbf{q}_{t+1}\}$. An additional observation is always either informative or neutral.

Intuitively, informative observations refine the sufficient support set to the extent that one's previous MDI updated solution comes to violate rationality. These observations induce a larger divergence between the MDI budget share, $\boldsymbol{\omega}_b^\star$, and the pilot prediction, $\boldsymbol{\eta}$. Therefore, the influence of the pilot prediction on the eventual MDI solution wanes in informative observations. This is attractive from the perspective of the nonparametric demand analyst as, in the limit, only individual choice data and the functional form free economic theory contribute to the final MDI solution.

SUMMARY This section has outlined my justification for the application of the Principle of MDI to update unconstrained demand predictions with rationality. I argued that the principle of MDI is a compelling criterion to adopt for three main reasons. First, the MDI result is the most likely rational demand to occur given the belief that expenditure decisions are governed by the pilot distribution. Second, it corresponds to the most optimistic updating rule followed by an econometrician looking to make use of the statistician's estimates, whose forecasts are evaluated via a logarithmic scoring rule. Finally, I showed that the KL objective function that is associated with the approach has a number of attractive features that provide instrumental support for its application. I further noted that the influence of the pilot prediction is declining in the number of informative observations on an individual's demand behaviour.

This section ends our theoretical discussion of MDI demands. I now move to demonstrate the value of my approach via an application to individual panel data. The following sections examine the empirical characteristics of the methodology and illustrate how it may be applied to make rational forecasts for use in welfare analysis.

9 EMPIRICAL STRATEGY

In the remaining sections, I undertake a detailed empirical investigation of MDI demands and demonstrate how the methodology can be implemented in order to recover globally rational demand predictions. I did not motivate a particular specification for the pilot demand prediction when outlining the MDI methodology. Thus, Section 10 explores the characteristics of a number of alternative specifications for $\boldsymbol{\eta}$. This analysis leads me to endorse an "inverse distance weighted" pilot for use in the rest of this essay. This pilot requires no pooling of individual information, achieves perfect within sample fit and it is found to dominate the considered alternatives on the criterion of predictive accuracy for the data set considered in ths essay. In Section 11, I recover globally rational individual demand systems using this inverse distance weighted pilot, explore the characteristics of the updating procedure and examine the significance of the theoretical distinction between the Varian and sufficient support sets. Finally, in Section 12 I show how my rational demand predictions can be used for welfare analysis and present estimates of the equivalent variation and deadweight losses associated with hypothesised price changes.

To provide context for the empirical results to come, this section describes the data and outlines relevant details concerning the implementation of the MDI methodology.

9.1 DATA

My empirical analysis is based upon individual level consumption data drawn from the U.K. Kantar Worldpanel. The Worldpanel is one of the largest surveys of consumer behaviour in the world and contains information on domestic food and drink purchases. Participating households are issued with a barcode reader, with which they record the purchases of all barcoded products that are bought into the home. Therefore, all household scannable 'fast-moving consumer goods' are recorded.[1] Leicester (2012) estimates that approximately 20% of all household expenditures are covered by this data source.

Survey participants are recruited using quota sampling methods from a range of U.K. postcodes and the panel is maintained to be broadly representative at all times. However, there has been some worry regarding the representativeness of the demographic composition of scanner data samples and of inaccurate expenditure recording arising from scanning 'fatigue' (see, for example, Lusk and Brookes

1 Purchases from all retailers and online purchases are also covered.

(2011) and the detailed discussion in Leicester and Oldfield (2009)). Leicester and Oldfield (2009) found Worldpanel households to, on average, be poorer, more likely to be unemployed, or part-time workers than those households that are covered by the Living Costs and Food Survey (also known as the Expenditure and Food Survey).[2] However, despite these differences in demographic composition, similar patterns of spending over commodities are found in the two data sets. Furthermore, these authors did not find survey fatigue to be a relevant feature of the data. Leicester and Oldfield (2009) thus conclude that the Worldpanel may have "considerable advantages over traditional social science expenditure data" (2009, p.336), arguing that it is a high quality source of information for the applied demand theorist.

Table 1: Sample Characteristics

	Mean	St. Dev
Female	0.64	0.47
Age	53.2	14.14
Left Full Time Education:		
< 15 Years	0.25	0.44
16-18 Years	0.34	0.47
> 18 Years	0.35	0.48

Using the barcode level expenditure data of the 15,000 households contained by the primary data set in any given year, I construct a panel of food and drink expenditures for a subsample of individuals. In the Worldpanel, purchase information is recorded at the household level and all household members are required to scan their expenditures. However, as my theoretical framework concerns individual choice behaviour, I restrict my sample to single person households. This implies that I do not have to impose the assumption that a multi-person household can be accurately modelled as a representative individual.[3] In addition, my theoretical framework abstracts from the labour supply decision. Therefore, I further restrict my sample to those individuals that have stable employment status. This allows me to abstract from potential nonseparabilities between consumption and leisure that my methodology does not currently allow for. Finally, I refine my sample to those individuals whom participate in the survey for the whole period that I have data for, i.e. January 2006 to January 2010. This is to allow me to construct as lengthy a panel

2 Please note that the Living Costs and Food Survey is not necessarily representative of the population.

3 In principle, there is nothing to stop one extending my framework to impose collective rationality upon household demand predictions. As briefly discussed in Section 7, one would instead impose membership of the support set associated with the collective model.

as possible with which to recover demand predictions. The support set is weakly shrinking in additional observations, thereby increasing the probabilty that rationality will restrict the location of demand responses and that an application of the MDI methodology is called for.

These restrictions leave me with a panel of 1884 individuals observed between January 2006 and January 2010. My selection criteria result in a sample that is on average older and that disproportionately represents women and highly educated individuals. This is because of a sizable group of older, single women and the higher education credentials of younger single-person households.

Table 2: Summary Budget Share and Expenditure Statistics

	Mean	St. Dev		
		Overall	Within	Between
Expenditure	298	124	100	81
Budget Shares:				
Bread & Cereals	0.10	0.05	0.04	0.03
Meat & Fish	0.24	0.11	0.09	0.06
Dairy	0.17	0.08	0.06	0.04
Fruit & Veg	0.22	0.10	0.09	0.05
Sugar	0.13	0.08	0.07	0.05
Other	0.06	0.04	0.03	0.03
Alcohol	0.09	0.13	0.12	0.06
Prices:				
Bread & Cereals	1.86	0.16		
Meat & Fish	1.71	0.14		
Dairy	2.02	0.20		
Fruit & Veg	1.70	0.12		
Sugar	2.01	0.16		
Other	1.66	0.09		
Alcohol	2.21	0.18		

There are more than 568,000 individual product codes listed in the data set, with observations recorded at purchase-level frequency. Tractable analysis requires an aggregation of these expenditures over goods and across time. I first aggregate purchases into groups that reflect the underlying subcategories of the U.K. retail price index (RPI) at quarterly periodicity. This produces 16 observations on 31 spending categories for each individual. I further refine and aggregate these spending categories into eight final commodities: (1) Bread & Cereals; (2) Meat & Fish; (3) Dairy & Fats; (4) Fruit & Vegetables; (5) Sugar &

Confectionary; (6) Other Food; (7) Alcohol. Summary statistics are provided in Table 2.

9.2 SAMPLE RATIONALITY

Rationality has been a maintained assumption throughout this essay. The MDI methodology has been developed for forecasting the demands of individuals for whom utility maximisation cannot be rejected as a valid behavioural hypothesis. I find that 82% of the sample satisfy GARP. Thus, for 1550 individuals, the assumption of rationality cannot be rejected; i.e. one can find a well-behaved utility function for each of these individuals.

Table 3: Afriat Efficiency

			Percentiles			
	Mean	Min	5	50	95	Max
Afriat Efficiency	0.9969	0.9769	0.9901	0.9982	0.9998	1.0000

To assess the significance of the violations of the 334 individuals who fail GARP, I calculate the "Afriat efficiency index" (Varian 1990). Intuitively, this index measures the extent to which an individual's budget constraint must be relaxed for their observed choices to satisfy GARP. Table 3 suggests that the majority of these violations are marginal. However, rather than perturb the observed choices of these consumers such that their past choices satisfy the necessary rationality constraints, I exclude them from my sample because it is not clear that imposing global rationality upon their demand predictions is the appropriate empirical strategy to follow. This restriction leaves me with a sample of 1550 individuals with whom to apply the MDI methodology to.

9.3 IMPLEMENTATION

I now briefly review the techniques that I use to operationalise the MDI methodology.

In Section 11, I recover the marginal demand curve for Meat & Fish within the individual-specific convex hull of budget-normalised price space. In the context of the marginal demand curve, global rationality can be attained by imposing membership of the Varian support set on a budget-wise basis as the cross-consistency of predictions is not a binding constraint (see Proposition 2.3). The associated MDI optimisation procedure is solved using the nonlinear *fmincon* solver in MATLAB with the interior-point algorithm. This solver is also used when I explore the characteristics of alternative specifications for the pilot prediction in Section 10.

To examine the significance of the theoretical distinction between the Varian and sufficient support sets, the latter half of Section 11 moves to consider the recovery of globally rational demands at a set of three intersecting budget hyperplanes. Such a small set of new budgets is considered because the computational intensity of the mixed integer linear programme that is associated with the approach is fast increasing in the size of the set of new budgets. For those individuals for whom the set of pilot predictions fails GARP, the MILP optimisation problem is solved using the *BARON* solver in GAMS.[4] Further details of the parameterisation of new budget hyperplanes are given when we return to recover demand forecasts in Section 11.

At points throughout the following empirical sections, I use non-parametric kernel methods to explore the patterns that emerge from the output of the MDI methodology. In each case, the functions of interest are estimated using a Gaussian kernel function. In the context of univariate density estimation, the bandwidth is selected according to the Silverman (1986) "rule-of-thumb". This represents the optimal bandwidth for a normal distribution and is given as:

$$h = \left(\frac{4}{3}\right)^{0.2} \widehat{\sigma} N^{-0.2}$$

where $\widehat{\sigma}$ is the estimated standard deviation of the distribution of interest. For the estimation of regression functions by kernel methods, I use the plug-in bandwidth suggested by Azzalini and Bowman (1997, p.31). Allowing for an adaptive bandwidth had no discernible impact upon the conclusions that could be drawn from the MDI output in either case.

Finally, it is important to stress that it is not necessary for choice data to be pooled across individuals at any point when implementing the techniques developed in this essay. Globally rational MDI demand responses are recovered on an individual-by-individual basis, thereby allowing for unrestricted preference heterogeneity across individuals.

SUMMARY This section has described the data that I will use in my empirical investigation of MDI demands and has outlined relevant details concerning the implementation of the methodology. The next section uses this data set and implements the MDI methodology as here detailed in order to explore the characteristics of alternative specifications for the pilot prediction.

4 This solver was only available on a reduced capacity mode, which constrained the size of the recovery problem that could be engaged in.

10 PILOT SPECIFICATION

Recovered MDI demands are dependent upon the specification of the pilot prediction given that we operate in a low information environment. This section investigates the properties of alternative specifications for the pilot demand predictions. This pilot specification will then be used in my remaining empirical analysis of MDI demand predictions.

10.1 EVALUATION CRITERIA

In theory, any unconstrained prediction of demand at a new budget regime can be used to construct η; the properties of the MDI solution were outlined independently of the particular specification for the pilot prediction in Sections 7 and 8. Further, with infinite data, the final MDI solution is independent of η because the predictive power of the nonparametric economic theory uniquely identifies rational demand bundles at B. However, η is far from irrelevant in our own context. In the current setting, MDI demand predictions depend directly upon η and, therefore, the quality and characteristics of recovered MDI demands will depend directly upon the corresponding properties of the pilot prediction.[1]

However, the choice of pilot prediction is underdetermined; it is not uniquely prescribed by economic theory. I propose three criteria with which I will evaluate alternative specifications of pilot demand predictions that will be used throughout the remaining empirical sections of this essay. The first criterion is parsimony. Simple specifications for the pilot prediction, which yield behaviourally interpretable pilot predictions, allow one to easily intuit the characteristics of the resulting MDI demands. The second criterion is flexibility. I desire a pilot specification that is flexible enough to reflect the pervasive preference heterogeneity that is documented across otherwise observationally equivalent individuals (Heckman 2001; Lewbel and Pendakur 2009; McFadden 2001). My final criterion is accuracy. As stated in the introduction, an aim of mine in this essay is to create an "accurate methodology with which to predict individual demands." As we operate in a low information environment, the quality of MDI demands

1 This is nicely conveyed by the words of Charles Babbage, the inventor of the first programmable computing device: "On occasions I have been asked, "Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?"... I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question." (1864, p.67).

is strongly influenced by the characteristics of the pilot prediction and, thus, predictive accuracy of the pilot is a prized characteristic.

10.2 CANDIDATE PILOTS

In this section, I explore the characteristics of a set of specifications for the pilot predictions: Maximum Entropy; Cobb Douglas; Nearest Neighbour; Inverse Distance Weighted. I formally refer to this set as $\Psi = \{\boldsymbol{\eta}^{ME}, \boldsymbol{\eta}^{CD}, \boldsymbol{\eta}^{NN}, \boldsymbol{\eta}^{IDW}\}$. The simplicity of these specifications is attractive and I have sufficient information to generate these pilot predictions at the individual level.[2] Here I briefly review their construction and theoretical properties, whilst later in the section I will undertake an empirical investigation of their predictive accuracy.

MAXIMUM ENTROPY The Maximum Entropy pilot corresponds to a uniform distribution of expenditure across goods; one does not assume a greater relative preference for one good over any other. The Maximum Entropy pilot prediction for individual i at a new budget $b \in B$, then corresponds to the following specification for $\boldsymbol{\eta}$:

$$\boldsymbol{\eta}_b^{ME,i} = \boldsymbol{\eta}^{ME} = \left(\frac{1}{K}\right) \mathbf{1}_K \tag{2.19}$$

The Maximum Entropy pilot is invariant across individuals and new budgets of interest; it always dictates a symmetric distribution of total expenditure across goods.

In traditional applications of the Principle of MDI, a Maximum Entropy pilot is equivalent to maximal uncertainty regarding the specification of a probability distribution. However, as proven in Property 5 of Section 8, a Maximum Entropy pilot is equivalent to assuming symmetric Cobb-Douglas preferences. This is a strong parametric assumption, which I would be surprised to see validated in my data set. However, I investigate the properties of this pilot given its simplicity and widespread use in related literatures.

COBB DOUGLAS A Cobb Douglas pilot specification relaxes the Maximum Entropy formulation above. Rather than assuming an equal preference weighting across goods, this specification allows for deviations from symmetry to reflect differences in an individuals relative tastes for commodities.

With the individual Cobb Douglas utility function

$$u^i(\mathbf{q}) = \sum_{k=1}^{K} a^{k,i} \log q^{k,i} \quad , \tag{2.20}$$

2 This facilitates an allowance for unrestricted preference heterogeneity in the construction of the pilot prediction.

Marshallian demands take the form

$$q^{i,k} = \frac{a^{i,k}m}{p^{i,k}} \quad , \tag{2.21}$$

with the associated budget shares:

$$w^{i,k} = a^{i,k} \tag{2.22}$$

Individual preferences are highly unlikely to perfectly approximate a Cobb Douglas preference specification. Preference parameters for the pilot prediction, $\{\widehat{a}^{k,i}\}^{k=1,\dots,K-1}$, are chosen to minimise the least squares criterion function, whilst $\widehat{a}^{K,i}$ is implied by adding up.

$$\begin{aligned} \widehat{a}^{k,i} &= \arg\min_{a^k} \textstyle\sum_{t=1}^{T} (w_t^{i,k} - a^k)^2 \\ &= \tfrac{1}{T} \textstyle\sum_{t=1}^{T} w_t^{k,i} \end{aligned} \tag{2.23}$$

for $k = 1, \dots, K-1$, and

$$\widehat{a}^{K,i} = 1 - \sum_{k=1}^{K-1} \widehat{a}^{k,i} \tag{2.24}$$

The Cobb-Douglas pilot specification for individual i's predicted demand at a new budget $b \in B$ is given as:

$$\eta_b^{CD,k,i} = \eta^{CD,k,i} = \frac{1}{T} \sum_{t=1}^{T} w_t^{k,i} = \bar{w}^{k,i} \tag{2.25}$$

for $k = 1, \dots, K-1$, and

$$\eta_b^{CD,K,i} = \eta^{CD,K,i} = 1 - \sum_{k=1}^{K-1} \bar{w}^{k,i} \tag{2.26}$$

Therefore, the Cobb-Douglas pilot prediction varies across individuals to reflect the differences in their past behaviour but remains invariant across new budgets for any given individual of interest.

NEAREST NEIGHBOUR The Cobb-Douglas pilot prediction for individual i corresponds to a global average of their observed budget shares; the pilot is independent of the parameterisation of the new budget of interest. However, in the nonparametric regression literature it is common to weight local information more highly when recovering demands at a particular point. The nearest-neighbour interpolant lies at one extreme of this approach. Taking a nearest neighbour specification for the pilot, one selects the budget share associated with the past budget that is "closest" to the new budget of interest as the pilot demand prediction. Additional past observations are not considered at all. This yields a piecewise-constant pilot demand function.

Without loss of generality, let us express all individual budgets as budget-normalised price vectors.[3] Past budgets are represented by the set $\{\widetilde{\mathbf{p}}^i_t\}_{t=1,\ldots,T}$, where $\widetilde{\mathbf{p}}^i_t = \mathbf{p}_t/y^i_t$, and new budgets are given as $\{\widetilde{\boldsymbol{\rho}}_b\}_{b=1,\ldots,B}$, where $\widetilde{\boldsymbol{\rho}}_b = \boldsymbol{\rho}_b/m_b$. Let us define the distance between an observed budget, $\widetilde{\mathbf{p}}_t$, and the new budget of interest, $\widetilde{\boldsymbol{\rho}}_b$, as $d(\widetilde{\mathbf{p}}_t, \widetilde{\boldsymbol{\rho}}_b)$. Just as the distance used to measure the divergence between a pilot prediction and the support set is not uniquely determined, there exist a host of metrics by which to measure the divergence between $\widetilde{\mathbf{p}}_t$ and $\widetilde{\boldsymbol{\rho}}_b$. I discuss this in greater detail below but, for the present moment, I do not endorse one particular distance measure of interest.

With $d(\widetilde{\mathbf{p}}^i_t, \widetilde{\boldsymbol{\rho}}_b)$ characterising the distance between budgets, the nearest neighbour pilot prediction for individual i's predicted demand at a new budget $b \in B$ is given as:

$$\boldsymbol{\eta}_b^{NN,i} = \mathbf{w}^i_{t^*} \tag{2.27}$$

where

$$t^* = \arg \min_{t \in \{1,\ldots,T\}} d(\widetilde{\mathbf{p}}^i_t, \widetilde{\boldsymbol{\rho}}_b) \tag{2.28}$$

Therefore, the nearest neighbour pilot varies across individuals and budgets to reflect variation in observed choice histories and differences in the distance between past budgets and new budgets of interest.

inverse distance weighted The nearest neighbour interpolant is very simple but generates discontinuities in recovered demand functions. This is likely to generate unattractive properties in the resulting MDI predictions. Therefore, I finally consider an "Inverse Distance Weighted" (IDW) pilot specification that can be thought to lie between the extremes of the Cobb Douglas and nearest neighbour specifications. This specification adapts the standard kernel methodology to better suit our low information environment. The approach is directly analogous to the synthetic control method (Abadie and Gardeazabal, 2003; Abadie, Diamond and Hainmueller, 2010) that is applied to construct comparison units in comparative case studies when traditional regression methods are inappropriate.

I form IDW pilots as a convex combination of past budget shares, to generate a smooth pilot demand function that relies only on individual specific information. The weight on an observed budget share in the pilot is set proportional to the inverse distance of the budget that supports it to the new budget of interest. Thus, the IDW pilot associated with the budget $\widetilde{\boldsymbol{\rho}}_b$ is specified as:

$$\boldsymbol{\eta}_b^{IDW,i} = \sum_{t=1}^{T} \delta^i_t \mathbf{w}^i_t \tag{2.29}$$

3 An absence of money illusion is assumed by my theoretical framework.

where

$$\delta_t^i \propto \frac{1}{d(\widetilde{\mathbf{p}}_t^i, \widetilde{\boldsymbol{\rho}}_b)} \tag{2.30}$$

Therefore, IDW pilots vary across individuals and budgets to reflect the variation in observed choice histories and differences in the distance between past budgets and new budgets of interest. However, unlike the nearest neighbour approach, a smooth pilot function is generated by this construction.

The IDW method can be motivated by an appeal to the convexity of preferences, under which small changes in budgets should not warrant a significant demand response. I also note that an IDW pilot achieves perfect in-sample fit. Demands at observed budgets are exactly replicated given that $d(\widetilde{\mathbf{p}}_t^i, \widetilde{\boldsymbol{\rho}}_b) = 0$ iff $\widetilde{\mathbf{p}}_t^i = \widetilde{\boldsymbol{\rho}}_b$, resulting in $\boldsymbol{\eta}_t^i = \mathbf{w}_t^i$.

10.3 BUDGET DISPARITY

The IDW and nearest neighbour pilots both rest upon a characterisation of the distance between observed budgets and new budgets. Yet, just as the distance used to measure the divergence between a pilot prediction and the support set is not uniquely determined, there exist a host of metrics with which to measure the divergence between $\widetilde{\mathbf{p}}_t^i$ and $\widetilde{\boldsymbol{\rho}}_b$. Regarding the construction of $\boldsymbol{\eta}$, I do not have a strong prior in favour of any particular distance measure. This is in contrast to my stance when motivating the MDI methodology, at which point I strongly endorsed the KL divergence and the associated Principle of MDI.

It is not contradictory to endorse different distance measures at the pilot-construction stage and rationality-updating stage of the prediction procedure. The KL divergence has a probabilistic interpretation that many other distances do not. Use of the KL divergence at the updating stage thus allows one to interpret $\boldsymbol{\omega}^\star$ as the most likely rational budget share specification to occur given the belief that the true allocation of expenditure across goods is governed by $\boldsymbol{\eta}$. This interpretation cannot be applied to updated demands recovered by minimising some other distance metric. However, during the construction of $\boldsymbol{\eta}$, one is concerned with measuring the distance between budget environments rather than developing an updating procedure. The probabilistic interpretation associated with the KL divergence does not lend the generated pilot prediction attractive meaning. Furthermore, in contrast to demands, budgets cannot be naturally recast in the language of probability distributions. This leads to the application of the KL divergence being complicated and computationally intensive for pilot construction. Therefore, I investigate the properties of pilots constructed with reference to both the KL divergence and the Euclidean distance. I provide further details on how these distances are calculated in this context below.

Kullback-Leibler divergence Budgets are not easily recast in a manner analogous to probability distributions. I use a two-stage procedure to calculate the KL between a new budget $b \in B$, $\widetilde{\boldsymbol{\rho}}_b$, and $\{\widetilde{\mathbf{p}}^i_t\}_{t=1,...,T}$.

Let $B^i_V = \{\widetilde{\mathbf{p}}^i_v\}_{v=1,...,V} \subseteq \{1,...,T\}$ denote the past budgets for individual i that define the vertices of their convex hull of budget-normalised price space. Any budget b within the convex hull of observed budgets, $Conv(\{\widetilde{\mathbf{p}}^i_t\}_{t=1,...,T})$ can be expressed as a convex combination of the elements of B^i_V.

$$\widetilde{\mathbf{p}}_b = \sum_{v=1}^{V} \lambda^{v,i}_b \widetilde{\mathbf{p}}^i_v \tag{2.31}$$

where $\boldsymbol{\lambda}^i_b \geqslant \mathbf{0}$, $\mathbf{1}'\boldsymbol{\lambda}^i_b = 1$ and $\widetilde{\mathbf{p}}_b \in Conv(\{\widetilde{\mathbf{p}}^i_t\}_{t=1,...,T})$.

The distance between budgets can now be specified as the KL divergence between the vertex weight vectors:

$$\begin{aligned} d(\widetilde{\mathbf{p}}^i_t, \widetilde{\boldsymbol{\rho}}_b) &= D(\boldsymbol{\lambda}^i_t \| \boldsymbol{\lambda}_b) \\ &= \boldsymbol{\lambda}^{i\prime}_t \log \frac{\boldsymbol{\lambda}^i_t}{\boldsymbol{\lambda}_b} \end{aligned} \tag{2.32}$$

However, this procedure is more complicated than it first appears because vertex weightings are underdetermined for in-hull budgets. This is highlighted by Figure 2. In this example, the vertex weighting that defines the in-hull budget B_5 is non-unique. For example, $\boldsymbol{\lambda} = [0.33, 0, 0.67, 0]$ or, as an alternative example, $\boldsymbol{\lambda} = [0, 0.45, 0, 0.67]$. I deal with this indeterminacy by solving a first stage maximisation problem in order to select the unique Maximum Entropy vertex weighting with which to construct the distance measure. This maximisation problem takes the form:

$$\max_{\lambda^i_j} \boldsymbol{\lambda}^{i\prime}_j \log \boldsymbol{\lambda}^i_j \tag{2.33}$$

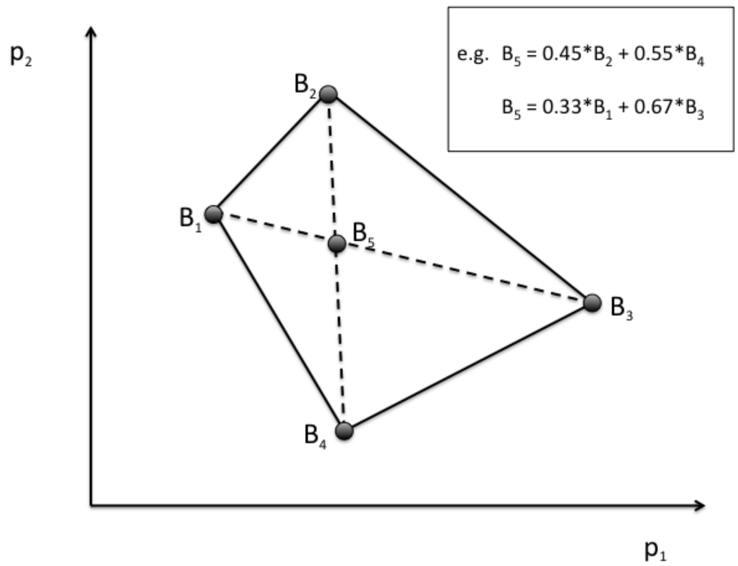
subject to

$$\begin{aligned} \boldsymbol{\lambda}^i_j &\geqslant \mathbf{0} \\ \mathbf{1}'\boldsymbol{\lambda}_j &= 1 \\ \widetilde{\mathbf{p}}^i_j &= \textstyle\sum_{v=1}^{V} \lambda^{v,i}_j \widetilde{\mathbf{p}}^i_v \end{aligned}$$

for all budgets $j \in B \cap \{1,...,T\}$ and individuals $i = 1,...,N$.

It should be noted that this procedure is extremely computationally intensive; up to $2T$ maximisation problems must be solved in order to construct a pilot prediction for demand at a new budget. Given that there is no clear justification for the use of the KL divergence in this particular setting, I also consider the Euclidean distance as a more computationally amenable alternative.

Figure 2: Vertex Weighting Underdetermined



EUCLIDEAN DISTANCE As an alternative to the KL-divergence, I consider the Euclidean distances between budgets. The Euclidean distance between budgets is defined as:

$$d(\widetilde{\mathbf{p}}_t^i, \widetilde{\boldsymbol{\rho}}_b) = \left[\left(\widetilde{\mathbf{p}}_t^i - \widetilde{\boldsymbol{\rho}}_t^i \right)' \left(\widetilde{\mathbf{p}}_t^i - \widetilde{\boldsymbol{\rho}}_t^i \right) \right]^{1/2} \tag{2.34}$$

This distance can be quickly constructed as no maximisation problems need to be solved in the process. The Euclidean distance measure is also used to construct comparison units in the synthetic control method (Abadie & Gardeazabal, 2003; Abadie, Diamond & Hainmueller, 2010), which, as noted, is an analogous procedure to the IDW method.

10.4 PREDICTIVE ACCURACY

My final criterion for evaluating competing specifications for pilot predictions was that of predictive accuracy. In order to rank alternative specifications according to this quality, I develop a cross validation algorithm that allows me to calculate the error with which different pilots, in conjunction with the MDI methodology, recover previously observed demands. As both $\boldsymbol{\eta}^{NN}$ and $\boldsymbol{\eta}^{IDW}$, achieve perfect within-sample fit, a cross validation exercise is performed in order to investigate their predictive accuracy.[4] Observations are sequentially dropped and predicted using the remaining data in conjunction with the MDI methodology. The error associated with a pilot estimate is then calculated through a comparison of predicted and realised budget shares.

PREDICTIVE ACCURACY ALGORITHM Let $\boldsymbol{\eta}_b^{n,i}(D)$ represent pilot specification $n \in \Psi$ for individual i at budget b constructed using data D. The predictive accuracy of $\boldsymbol{\eta}_b^{n,i}$ is calculated using the following algorithm:

For each $i = 1, ..., N$

For each $t = 1, ..., T$

1. Construct $\widehat{D}^i_{\neg t} = D^i \backslash \{\mathbf{p}_t, \mathbf{q}_t^i\}$

For each $n \in \Psi$

2. Construct $\boldsymbol{\eta}_b^{n,i}(D^i_{\neg t})$
3. Construct $S^V(\mathbf{p}_t, y_t | D^i_{\neg t})$
4. Calculate the MDI demand at $\{\mathbf{p}_t, \mathbf{q}_t^i\}$
$\boldsymbol{\omega}_t^{i,\star} = \arg\min_{\mathbf{A}_t \omega \in S^V(\mathbf{p}_t, y_t | D^i_{\neg t})} D(\boldsymbol{\omega} \| \boldsymbol{\eta}_b^{n,i}(D^i_{\neg t}))$
5. Calculate the absolute error
associated with the MDI prediction:
$AE_t^{n,i} = |\boldsymbol{\omega}_t^{n,i\star} - \mathbf{w}_t^i|'\mathbf{1}$

end

end

end

DISTANCE MEASURES & EFFICIENCY Before comparing the predictive accuracy of different pilots, I explore the impact of the choice of distance measure upon the nearest neighbour and IDW pilots. The first point to note is that there exist significant differences in the computational time associated with constructing the Kullback-Leibler pilots and the Euclidean pilots: $\boldsymbol{\eta}^{IDW,Euc}$ and $\boldsymbol{\eta}^{NN,Euc}$ are significantly faster to construct than their Kullback-Leibler peers as no optimisation procedures are engaged in. The greater computational efficiency with which one can construct the Euclidean pilots makes them especially attractive for applied welfare analyses that involve simulation exercises.[5]

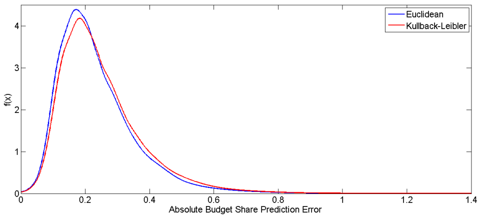
Furthermore, as highlighted by Figure 3, the Euclidean pilots are associated with marginally higher predictive accuracy than those con-

4 This follows as the distance between the previously observed budget and new budget of interest would be zero in this instance.
5 This is especially so in the light of the iterative procedure outlined in Section 12 that I use to calculate measures of compensating expenditure using MDI demands.

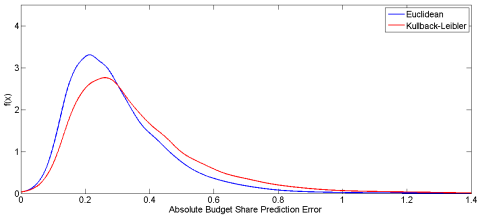
Table 4: Mean Pilot Computation Time *(secs)*

	Euclidean	KL
η^{IDW}	0.0001	2.2428
η^{NN}	0.0000	2.2427

Figure 3: Difference in Error Distribution by Distance Measure



(a) Inverse Distance Weighted



(b) Nearest Neighbour

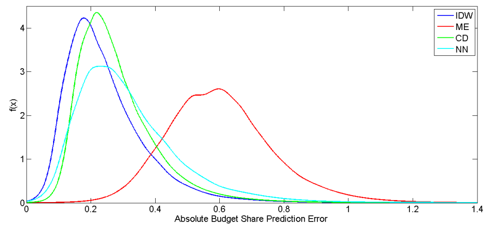
structed with reference to the KL divergence.[6] The distance between budgets is not naturally characterised by the Kullback-Leibler divergence between the vertex weight vectors and thus, the application of the measure to this setting was somewhat forced. For this reason, the remaining empirical sections will use $\eta^{IDW,Euc}$ as the specification for the pilot demand prediction at any new budget.

6 Please note that the Kullback-Leibler IDW pilot still achieves higher predictive accuracy than the Cobb-Douglas specification.

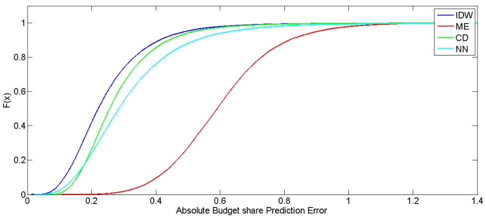
Table 5: Differences in Error by Distance Measure

	Mean Euc. Error *(St. Dev)*	Mean KL Error *(St. Dev)*	Diff.
η^{ME}	0.2377 (0.1282)	0.2518 (0.1348)	-0.0141
η^{NN}	0.3059 (0.1684)	0.3635 (0.2104)	-0.0576

Figure 4: Distribution of Prediction Errors



(a) Kernel Plot



(b) Empirical Distribution Function

Table 6: Distribution of Prediction Errors

	Mean	St. Dev	0.5	0.50	0.95
η^{ME}	0.60	0.17	0.355	0.591	0.900
η^{CD}	0.28	0.13	0.135	0.255	0.518
η^{NN}	0.31	0.17	0.119	0.285	0.630
η^{IDW}	0.24	0.13	0.095	0.220	0.491

RESULTS Figure 4 highlights that there exist clear differences in the predictive accuracy of the considered pilot specifications. Unsurprisingly, the Maximum Entropy pilot performs least favourably being independent of individual choice behaviour and the parametrisation of the new budget of interest. The IDW pilot first order stochastically dominates the alternative pilot specifications considered in this section on the criterion of predictive accuracy, whilst the Cobb-Douglas pilot is ranked second. However, as the dimensionality of the recovered demand system increases, the predictive accuracy of the IDW and Cobb-Douglas pilots will converge; when measuring distances between coordinates in high-dimensional space, there is little difference in the distances between pairwise budget comparisons and the weights forming the IDW pilot will tend towards symmetry. For some budgets, the nearest neighbour pilot achieves high predictive accuracy but this is not a global characteristic. For approximately 80% of budgets, the prediction error associated with the nearest neighbour pilot exceeds that of the Cobb-Douglas specification.

The null hypothesis of equality of distribution of predicted budget share errors is rejected at the 1% level using independent Kolmogorov-

Table 7: Kolmogorov-Smirnoff Test Statistics

	KS-Stat.	p-value
η^{ME}	0.7952	0.0000
η^{CD}	0.1568	0.0000
η^{NN}	0.2087	0.0000

Smirnov tests for any pairwise comparison of the errors associated with IDW and the alternative pilot predictions. These tests evaluate the difference in the cumulative distribution functions between data samples. The test statistics and asymptotic p-values are recorded in Table 7.

These results suggest the use of an IDW pilot in the remaining empirical sections. This pilot satisfies my criteria of flexibility and predictive accuracy: it does not require pooling of individual information, thereby allowing for maximal preference heterogeneity in the population, and I have established a first order dominance result regarding the distribution of prediction losses of IDW pilots over the alternatives considered for my particular data set. One drawback of the IDW pilot is that, although simple to construct, it does not yield a small set of behaviourally interpretable parameters with which to summarise the characteristics of the resulting MDI demands.

SUMMARY Until this section, the construction of the pilot demand prediction has gone unspecified in this essay. However, the characteristics of recovered MDI demand predictions directly depend upon those of the pilot prediction because the predictive power of nonparametric economic theory is not absolute in our finite-data context. In this section, I considered a set of alternative specifications for the pilot prediction and evaluated them against the criteria of parsimony, flexibility and predictive accuracy. This analysis has suggested the use of an inverse distance weighted pilot demand prediction, which uses a convex combination of past budget shares to form a smooth pilot demand function. The weight on an observed budget when constructing the IDW pilot is set proportional to the inverse Euclidean distance of the distance that supports it to the new distance of interest. The IDW pilot is estimated at the individual level, it is simple to construct and it dominated the considered alternatives on the criterion of predictive accuracy for my particular data set. In the following sections, I will impose global rationality upon the Euclidean IDW pilot in order to recover theory-consistent demand predictions that allow for maximal preference heterogeneity in the population.

11 MDI DEMAND SYSTEMS

In this section, I apply the MDI methodology to impose global rationality upon unconstrained pilot predictions and I explore the characteristics of this updating procedure. I begin by recovering the marginal demand curve for Meat & Fish within the convex hull of budget-normalised price space and then turn to explore the revealed heterogeneity in the recovered own-price and cross-price effects. To end this section, I proceed to investigate the characteristics of the MDI updating procedure and examine the empirical significance of the theoretical distinction that I have made between the Varian and sufficient support sets. In so doing, I find that the cross-consistency of demand predictions is not simply a theoretical curiosity but can be a binding constraint when predicting over sets of new budget hyperplanes that intersect.

11.1 MARGINAL DEMAND CURVE

I first illustrate the nature of the output yielded by application of the MDI methodology. I apply the techniques developed in Section 7 to recover the marginal demand curve for Meat & Fish. This is an interesting good to consider because censoring on this good (i.e. the existence of vegetarians) creates difficulties for traditional demand system estimation methods. Very few studies have attempted the estimation of a many-equation demand system with binding non-negativity constraints and those that do are forced to rely on restrictive functional form assumptions (see, for example, Golan, Perloff and Shen (2001)).

METHODOLOGY Demands are recovered at each individual's median level of total expenditure on a grid of points within the convex hull of budget-normalised price space for that individual. Specifically, the price of goods other than meat are fixed at the mean observed across the sample, whilst the price of meat is varied on a grid of points with 0.001 spacing. There is some variation in the price range over which demands are recovered because of differences in the size and location of the convex-hull of budget normalised price space across individuals.[1] My motivation for focusing on within-hull budgets is two-fold. First, extrapolation to beyond this space precludes construction of the IDW pilot. Second, the primary contribution of this essay is the development of the MDI updating methodol-

1 This is due to budget variation across individuals.

ogy. Past choices will not constrain the location of rational demand predictions at out-of-sample budgets and thus, only failures of the cross-consistency of demand predictions will initiate a need for pilot predictions to be manipulated in order to satisfy global rationality. However, I note that extrapolation to regions beyond the data is a risky strategy for all estimation methods. Ultimately, the extrapolation step is one of good faith and forces one to rely on unverifiable structural assumptions. Thus, my focus on within-sample prediction should not be judged in too negative a light.

Unconstrained demand predictions at each new budget of interest are first generated as the Euclidean IDW pilot that was motivated in the last section. I then update these pilot predictions using the MDI methodology such that they satisfy the global rationality requirement. Proposition 2.3 justifies a pointwise updating procedure to impose membership of the Varian support set at each new budget of interest; for budgets that define a portion of the marginal demand curve, a set of updated demand predictions is an element of the sufficient support set if and only if each prediction is an element of its own Varian support set. Therefore, solving the mixed integer optimisation problem to impose the cross-consistency of demand predictions is not required in the current setting.

One drawback of my specification for the pilot prediction is that a small set behaviourally interpretable parameters, with which to summarise individual preferences and demand behaviour, is not forthcoming. Given the complexity of individual preferences, a trade-off between parsimonious estimation output and accuracy is a feature of nonparametric methods more widely; the term "parsimonious nonparametric structural model" is something of an oxymoron. Thus, to aid interpretation of my results, I calculate the average arc elasticity of demand along the recovered portion of the marginal demand curve for each individual in addition to reporting the recovered quantity sequences.

$$\varepsilon^{k,i} = \frac{1}{|B|} \sum_{b=2}^{B} \frac{x_b^{k,i} - x_{b-1}^{k,i}}{\rho_b^k - \rho_{b-1}^k} \frac{\rho_b^k}{x_b^{k,i}}$$

Although this measure hides within-individual variation in the nature of demand responses at different regions of price space, it provides a useful parameter with which to summarise the heterogeneity in price responses across individuals.

INDIVIDUAL EXAMPLES To highlight the characteristics of the output gained from implementation of the MDI methodology, let us first consider some examples of recovered individual marginal demand curves. Figures 5 to 7 depict the recovered marginal demand curves for Meat & Fish for four individuals who are similar along dimensions of observable heterogeneity. Each figure depicts the recovered

MDI quantity trajectories as the price of Meat & Fish is varied along the x-axis. The individuals whose marginal demand curves are depicted below are female, they left school between 16-18 years old and they are aged in their late 30s.

Despite the similarity of these individuals with regard to their observable characteristics, recovered MDI demands highlight differences in the nature of price effects between them. Firstly, Individual 876, in Figure 7, is a vegetarian. Therefore, no behavioural response to changes in the price of Meat & Fish is recovered for this individual. For the remaining carnivores, the average own-price arc elasticity is in the neighbourhood of minus unity, whilst the cross-price effects are recovered to have different signs and magnitudes. For example, Fruit & Vegetables is a complement for Meat & Fish for Individual 160 in Figure 6, whilst for Individuals' 116 and 160 it is a substitute. Individual 355 does not drink alcohol in the home and thus, a null price effect on Alcohol is returned, whilst differences in the regions over which Alcohol is a complement and over which it is a substitute for Meat & Fish are revealed by comparing Individuals 116 and 160. Furthermore, the construction of the IDW pilot allows the slope of the marginal demand function and, thus, the degree of complementarity-substitutability of goods to vary across different regions of the price space. This is most clearly evident in the recovered demand responses for Individual 160. For this individual, the sign of the cross-price effects on Bread and Alcohol varies across recovered budgets; at low Meat & Fish prices, Bread is a substitute and Alcohol a complement, whilst at high Meat & Fish prices, Bread is a complement and Alcohol is a substitute.

Figure 5: Individual 116

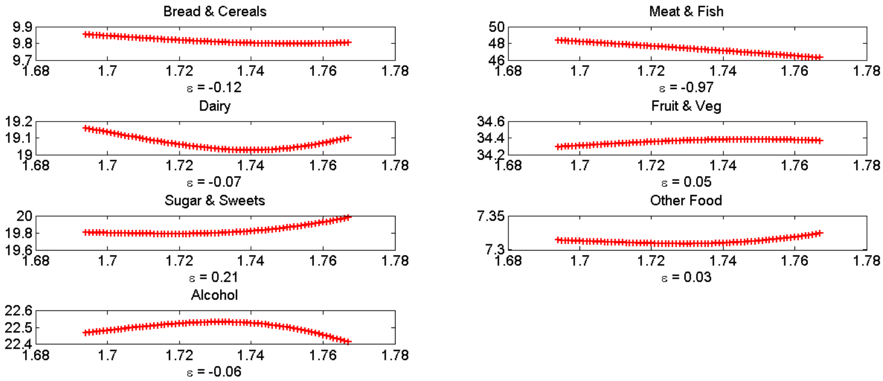


Figure 6: Individual 160

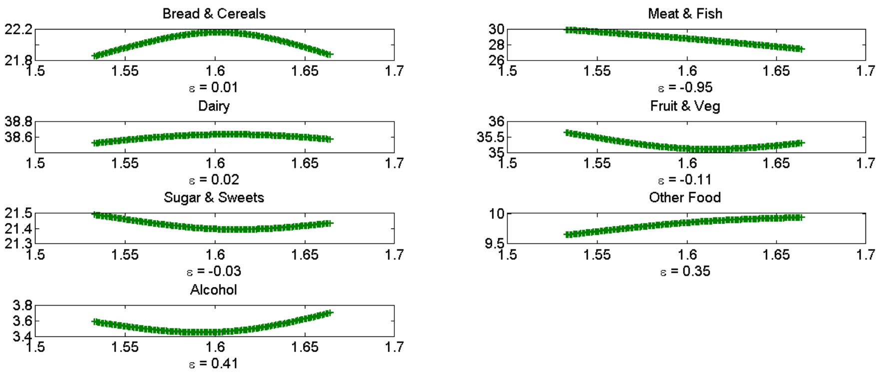


Figure 7: Individual 355

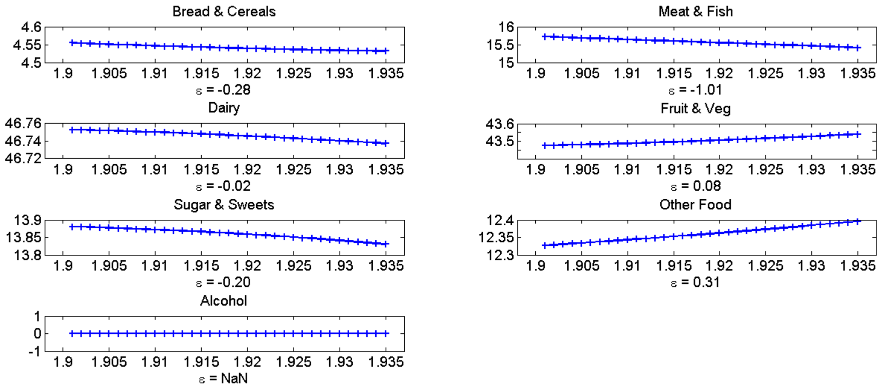


Figure 8: Individual 876

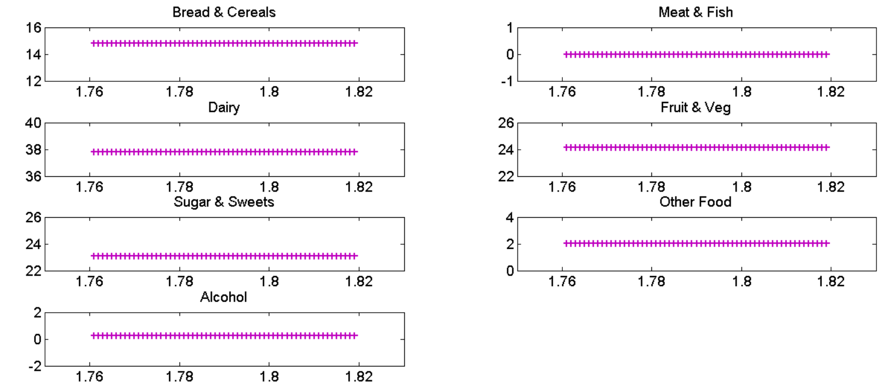


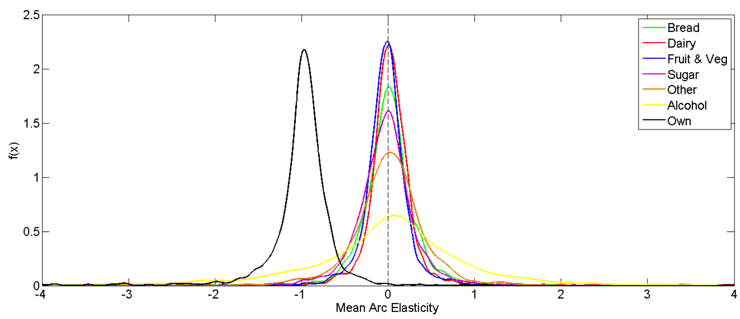
Table 8: Arc Elasticities for Marginal Demand Curve for Meat & Fish

	Mean	Percentiles				
		5	25	50	75	95
Own	-0.97	-1.59	-1.14	-0.96	-0.87	-0.62
Bread	0.02	-0.48	-0.15	0.01	0.15	0.50
Dairy	0.03	-0.36	-0.09	0.02	0.15	0.43
Fruit & Veg	-0.03	-0.40	-0.14	-0.02	0.09	0.37
Sugar	-0.03	-0.54	-0.20	-0.01	0.14	0.50
Other	0.03	-0.68	-0.19	0.03	0.23	0.69
Alcohol	-0.06	-1.97	-0.48	0.04	0.46	1.55

SAMPLE ARC ELASTICITIES Heterogeneity in recovered cross price responses is a notable feature of the MDI output. Table 10 and Figure 9 give the distribution of average arc elasticities along the marginal demand curve.

The recovered average own price effect is in the neighbourhood of minus unity for the majority of the sample; half of all recovered elasticities are in the interval $[-1.14, -0.87]$, with a median response of 0.96. There are only 11 individuals, or less than 1% of the sample, for whom a positive own price response is recovered. For this small minority, Meat & Fish must be an inferior good. Considering the distribution of average cross-price elasticities, each distribution is centred in the neighbourhood of zero but the dispersion of price responsiveness can be seen to vary across goods. There is greatest dispersion in the cross-price elasticity for Alcohol. Over 20% of the sample have an absolute average cross price elasticity exceeding unity, indicating significant price responsiveness. There is also greater heterogeneity recovered in the Other Food response. This is to be expected given the heterogenous nature of this good.

Figure 9: Kernel Density of Average Arc Elasticities



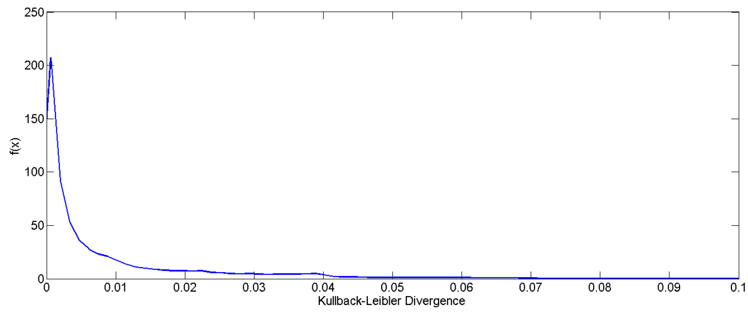
Independent nonparametric Kolmogrov-Smirnoff tests reveal that the heterogeneity in cross-price elasticities is non-normal and seemingly independent of observed characteristics. A fitted normal density function associated with elasticity estimates was computed using the unbiased sample average and variance estimates. Equality of distribution with the fitted normal density function was rejected for each of the arc-elasticity distributions. However, for the majority of sample stratifications upon observable characteristics, one cannot reject the hypothesis of equality of distribution at the 5% level between the conditional arc elasticity distributions. This is with the exception of a significantly different distribution in the cross-price elasticity of alcohol by gender. Comparison of the mean cross-price arc elasticities for Alcohol reveals that this good is more strongly complementary with Meat for men than women. This is due to a greater left skew in the distribution of the cross-price effects for males, rather than a large difference in the median response. However, with this exception, unobserved taste heterogeneity does not appear to be easily pinned down by observables nor accurately represented by a normally distributed error term.

MDI UPDATING A primary aim of this essay was to develop a methodology for imposing global rationality upon unconstrained nonparametric predictions for demands at new budget regimes. However, I did not find irrationality of the pilot predictions to be a salient feature of the recovery procedure in the context of the marginal demand curve. Pilot predictions at only 3% of budgets lay in the complement of the support set and required application of the MDI updating methodology in order to satisfy global rationality.

The value of the objective function of the MDI optimisation procedure, $D(\omega_b \| \eta_b) = \omega_b' \log_2(\omega_b/\eta_b)$, returns the minimum KL divergence between the pilot prediction and the support set. It represents the amount of information, measured in "bits", that is required to update the pilot with rationality. The magnitude of the departures of the pilot predictions from rationality appear slight. This is highlighted by reference to Figure 10 and Table 11. Three-quarters of deviations required less than 0.01 bits of information to update them with rationality.

As discussed in Section 8, the KL divergence has a probabilistic interpretation; the value of the objective can be transformed into a measure of likelihood that a sample budget share of ω would occur given a belief that η characterised the underlying distribution of expenditure to goods. To aid interpretation of the magnitude of pilot departures from rationality, I also compute the "epistemic likelihood" of the MDI solution arising in a random sample of size $n = 100$ units,

Figure 10: Kernel Density of KL Divergences



given a belief in η. Equation 2.16 gave us that the epistemic likelihood of ω given η and n can be expressed as:

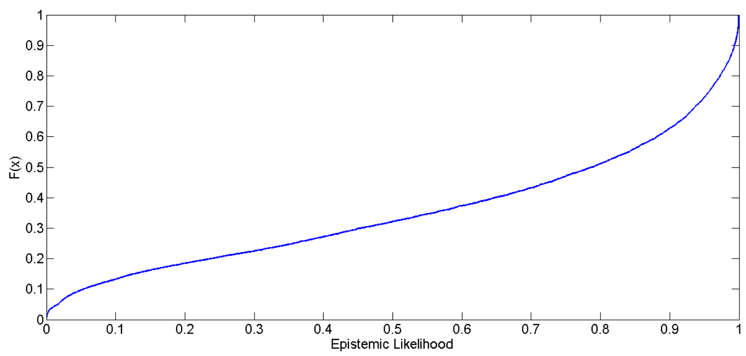
$$2^{n(-D(\omega|\eta)+\omega' \log_2 \omega)}$$

70% of deviations would be considered at least 50% likely to occur and 38% of updated budget shares would be classed as more than 90% likely to occur. This supports my interpretation of the departures of my pilot predictions' departures from rationality as "slight".

Figures 12 to 15 give examples of individuals for whom application of the MDI methodology was required. In all cases, the pilot did not generate a sufficient own-price response of Meat & Fish and a downward adjustment of the pilot predictions for the quantity of Meat & Fish purchased were required. However, there is no systematic direction to the updating of the cross-price responses and no generalisations can be made concerning the likelihood of updating and individual characteristics.

In summary, application of the MDI methodology to recover the marginal demand curve for Meat & Fish yields well-behaved, glob-

Figure 11: Empirical Distribution Function of Epistemic Likelihood



ally rational predictions and plausible estimates of the own- and cross-price elasticities of demand. Recovered demands are suggestive of heterogeneity in cross price responses that is non-normally distributed and one cannot reject the hypothesis of equality of distribution between cells stratified according to observable characteristics.

Table 9: Kolmogorov Smirnoff Test Statistic p-values

	Normal	Gender	Low/ Mod Ed.	Low/ High Ed.	Moderate/ High Ed.
Own	0.00*	0.07	0.06	0.07	0.10
Bread	0.00*	0.06	0.05	0.06	0.06
Dairy	0.00*	0.07	0.06	0.09	0.06
Fruit & Veg	0.00*	0.06	0.09	0.06	0.09
Sugar	0.00*	0.05	0.08	0.08	0.06
Other	0.00*	0.06	0.10	0.09	0.08
Alcohol	0.00*	0.01*	0.06	0.07	0.09

∗: Significant at the 5% level.

Table 10: Cross-Price Arc Elasticities for Alcohol by Sex

	Mean	St. Dev	Percentiles				
			5	25	50	75	95
Women	-0.03	1.67	-1.98	-0.50	0.05	0.54	1.60
Men	-0.10	1.45	-1.96	-0.44	0.03	0.35	1.45

Table 11: KL Divergence along Marginal Demand Curve

Mean	St. Dev	Percentiles				
		5	25	50	75	95
0.0095	0.0167	0.0000	0.0004	0.0024	0.0103	0.0407

Figure 12: Individual 341

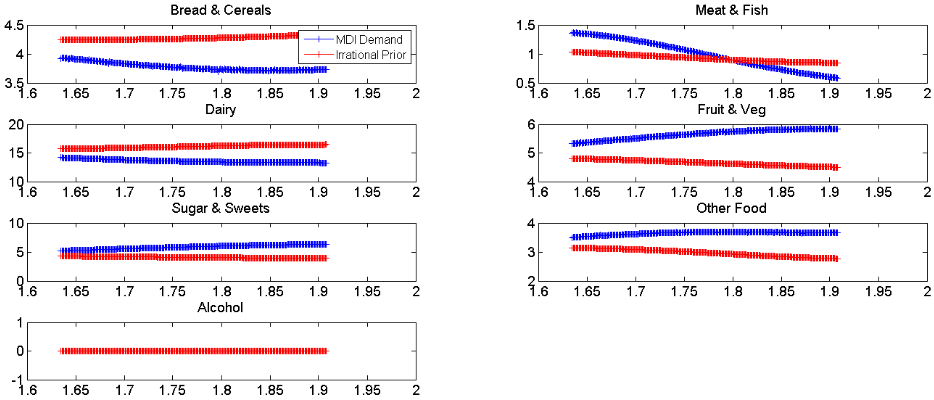


Figure 13: Individual 2

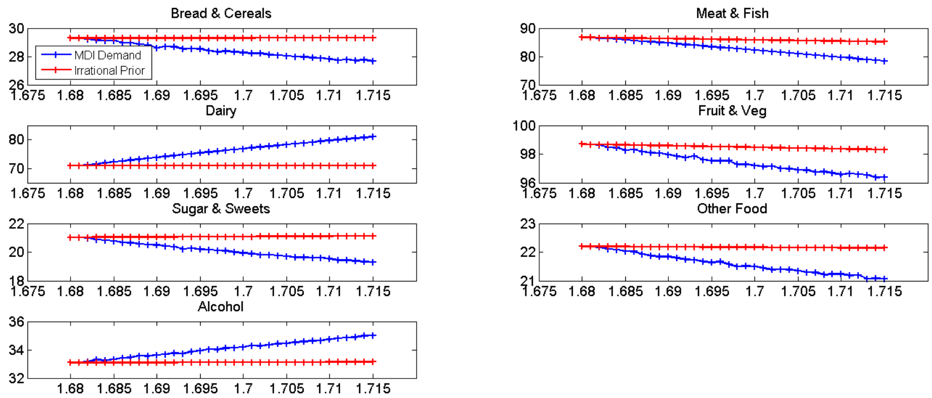


Figure 14: Individual 399

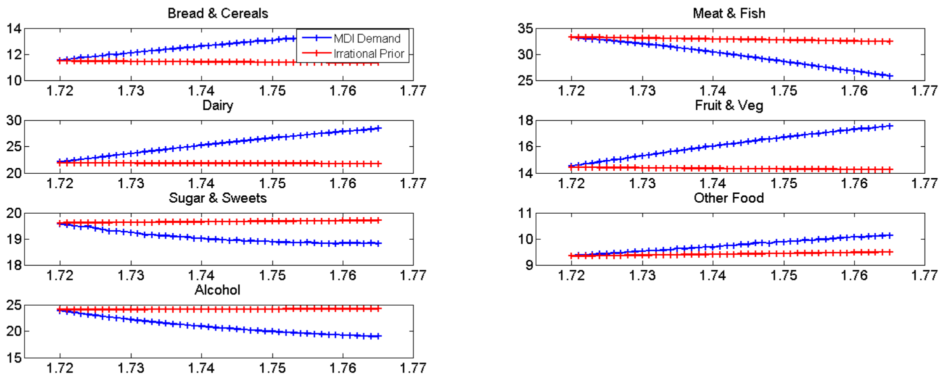
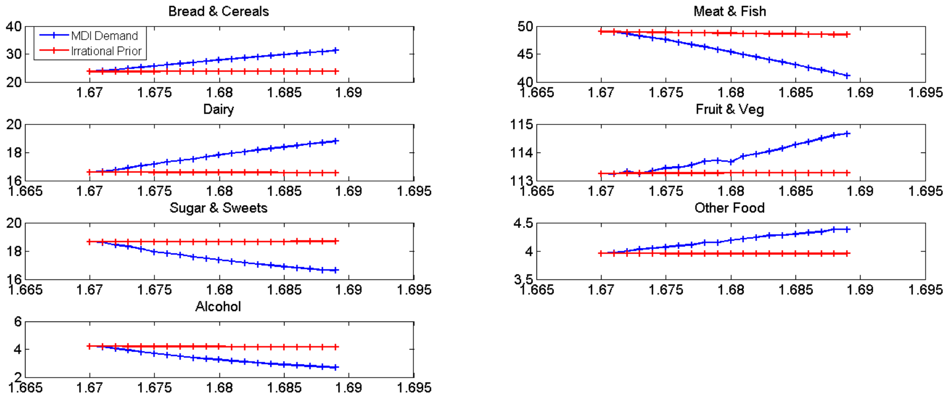


Figure 15: Individual 246



Global rationality of the IDW pilot predictions was not found to be a binding constraint in our recovery of the marginal demand curve for Meat & Fish. However, from Proposition 2.3, cross-consistency of demand predictions was necessarily achieved in our previous setting; forecasts at different points along the marginal demand curve of a single good do not interact. Yet, in many applications, the goal is to compare differences in the demand and welfare effects between regimes for which the hypothesised budget hyperplanes do intersect. In this setting, the cross-consistency of demand predictions is not necessarily attained by imposing pointwise membership of the Varian support sets. I here explore whether the requirement of cross-consistency of demand predictions is simply a theoretical curiosity or if the distinction has practical relevance by recovering demands at a small set of intersecting hyperplanes.

The updating procedure that is associated with the sufficient support set is very computationally demanding. I restrict the scale of the problem by only recovering demands at three intersecting budget hyperplanes.[2] Specifically, I independently maximise the price of Bread & Cereals, Dairy and Fruit & Vegetables within the convex-hull of budget-normalised price space, whilst all remaining good prices are kept at their mean budget-normalised level.[3] These three goods are chosen because all individuals in my sample consume strictly positive quantities of these commodities in each time period.

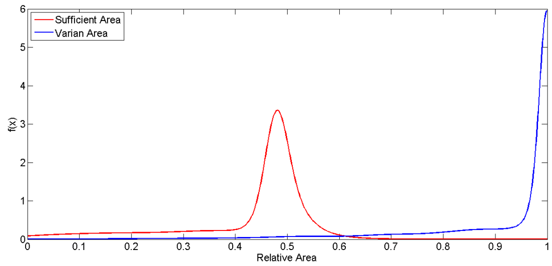
RELATIVE AREA The first question of interest is whether there is a difference in the size of the Varian and sufficient support sets. The "relative area" (Beatty and Crawford, 2010) is the size of the set of the predictions contained by the support set relative to the set of all possible outcomes. I compute the relative areas of each individual's Varian support set, $r^{V,i}$, and the sufficient support set, $r^{S,i}$, using a Monte Carlo simulation. For each individual, I randomly draw 10,000 sets of three budget shares to be associated with the new budget hyperplanes of interest.[4] To compute $r^{V,i}$, the consistency of randomly drawn demands at each budget b with past observations and rationality is verified but mutual consistency of the randomly drawn quantities is not imposed. However, to compute $r^{S,i}$, GARP is tested simultaneously on the pooled set of past choices and the full set of

2 Future work will look to improve the efficiency of the procedure such that a more ambitious empirical exercise can be engaged in.

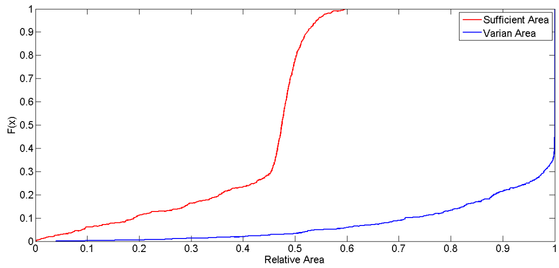
3 Due to differences in total expenditure across individuals, there is some variation in the precise budgets that demands are recovered at across individuals.

4 This is achieved by randomly drawing $|K|$ shares from the Dirichlet distribution and rescaling these shares by their sum such that they add to unity. Drawing from the Dirichlet distribution ensures a uniform distribution of budget shares across the whole surface of the new budget hyperplane.

Figure 16: Distribution of Relative Areas



(a) Kernel Density



(b) Empirical Distribution Function

random demands at the new budgets at each simulation draw. The relative area of the sufficient support set is always weakly smaller than that of the Varian support set given the additional constraint of cross-consistency of demand forecasts. The relative areas of the support sets will only be of the same size when demand forecasts do not interact, as, for example, when recovering the marginal demand curve for a single good.

I find that there is a significant difference in the relative areas associated with the Varian and sufficient support sets. When considering each new forecast independently, the revealed preference constraints only bind for 38% of the sample. For the remaining 62% of the sample, $r^{V,i} = 1$. A relative area of unity implies that rationality of the independent restrictions could never be a binding constraint on one's forecasts at the considered budgets given an individual's past choice behaviour.

However, the distribution of $r^{S,i}$ stands in stark contrast to this pattern. Once the cross-consistency of demand predictions is imposed, the relative area is reduced to lie in the neighbourhood of 0.5. Thus, the requirement of cross-consistency of demand forecasts does have quite a dramatic impact on the size of the support set.

UPDATING The differences in the relative areas associated with the Varian and sufficient support sets result, unsurprisingly, in differences of the necessity of application of the MDI updating procedure depending upon whether pointwise or global rationality is imposed.

Table 12: Support Set Relative Areas

	Mean	St. Dev	Percentiles				
			5	25	50	75	95
Varian	0.93	0.15	0.55	0.95	1.00	1.00	1.00
Sufficient	0.42	0.13	0.09	0.42	0.47	0.50	0.54

Given that the revealed preference restrictions are essentially unbinding when forecasts are treated as independent, it is of no surprise that the IDW pilot predictions failed GARP for only 33 individuals in my sample. The majority of these individuals then only required rationality to be imposed on a single pilot prediction: for 17 individuals, only one of the pilot predictions violated rationality, while a further 8 individuals required two pilots predictions to be updated and for a final 8 individuals, all pilot predictions required modification.The degree of updating required for these pilots to satisfy rationality is summarised in Table 13 and Figure 17. This represents the divergence between the pilot and MDI demand that is required to achieve rationality of a prediction.

Figure 17: Varian KL Divergences

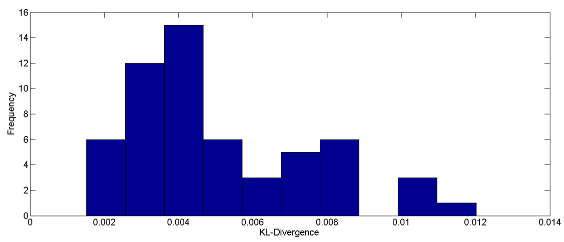


Table 13: Varian Updating

	Mean	St. Dev	Median
KL Div:	0.005	0.002	0.004

The full set of pilot predictions violated global rationality (i.e. lay in the complement of the sufficient support set) for 161 individuals, or equivalently 10.4% of the sample, despite global rationality representing a much more demanding restriction. This should be considered a positive characteristic of the specification of the pilot prediction. These results highlight that there exist 128 individuals for whom, given past choice, pilot predictions satisfied GARP when forecasts were considered independently of one another, but then violated rationality when they are considered jointly. By definition, for all those individuals whose pilots failed GARP when they were considered independently, the pilots also fail GARP when they are pooled. Table 14 and Figure 18 show the distribution of KL-divergences that are required to update the pilot predictions with rationality. I find that there is a skew in the extent to which pilot predictions have to be updated in order to satisfy global rationality. The majority of violations of cross-consistency are marginal but there is a positive skew to the distribution.

Figure 18: Sufficient KL Divergences

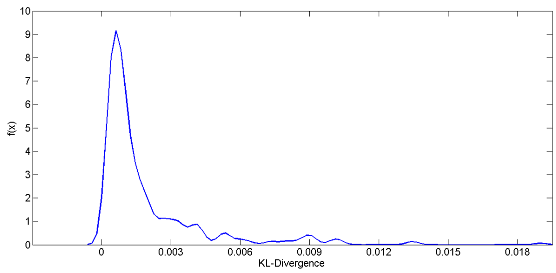


Table 14: Sufficient Updating

	Mean	St. Dev	Percentiles		
			5	50	95
KL Div:	0.0081	0.0198	0.0001	0.0022	0.0388

SUMMARY In summary, application of the MDI methodology yields well behaved, globally rational individual demand predictions, which allow for unrestricted preference heterogeneity across individuals. I found that the sample distribution of mean arc elasticities was non-normal and, with the exception of sex-differences in the cross-price response of Alcohol, I could not discern any interesting patterns relating taste differences to observed heterogeneity.

When demand forecasts are considered independently or do not interact, as is the case along the marginal demand curve, application of the MDI updating procedure is rarely required when used in conjunction with the endorsed IDW pilot. However, a Monte Carlo simulation reveals that the relative area defined by the Varian and sufficient support sets is significantly different when forecasts are jointly considered. For the particular set of budgets considered, $r^{S,i} \approx 0.5r^{V,i}$. The frequency of MDI updating is thus greater when membership of the sufficient support set is imposed.

12 WELFARE ANALYSIS

The last section showed how the MDI methodology can be usefully applied to recover globally rational demand predictions. In this section, I shift focus to demonstrate how MDI demand predictions can be used to conduct welfare analyses at the individual level.

The ability to conduct coherent welfare analyses was a primary motivation for the development of the MDI methodology in this essay. The potential applications of demand predictions are severely curtailed if they violate rationality. Rationality is necessary if one is to calculate unique welfare metrics because its violation generates path dependency in methods that are applied to numerically calculate exact consumer surplus (Vartia, 1983; Hausman and Newey, 1995). However, until this essay, a method for imposing global rationality upon nonparametric demand predictions has not existed. In this section, I apply Vartia's (1983) algorithm to recovered MDI demands in order to calculate the compensating income associated with hypothesised price changes. This enables me to calculate measures of consumer surplus and cost-of-living indices at the individual level.

12.1 COMPENSATED INCOME

I am interested in quantifying the welfare impact of price changes. This is typically advanced by calculating measures of the change in exact consumer surplus that are associated with regime changes. I continue to assume that each individual is associated with a well-behaved utility function and, thus, that the welfare metrics of interest exist. Also, please note that all of what follows is defined at the individual level. I suppress the superscript i for notational brevity.

Let us consider a change in prices from $\mathbf{p}_0$ to $\mathbf{p}_1$. The behavioural response to this price change is that dictated by the re-optimisation of the individual's rational, well behaved utility function, $u(\mathbf{q})$. When considering questions of welfare, it is useful to work with the dual formulation of the consumer's maximisation problem. The expenditure function gives the minimum cost of attaining utility u at prices $\mathbf{p}$ and is defined as:

$$e(\mathbf{p}, u) = \min \{y | y = \mathbf{p}'\mathbf{q} \quad \& \quad u(\mathbf{q}) = u\} \tag{2.37}$$

The exact consumer surplus associated with a change in prices from $\mathbf{p}_0$ to $\mathbf{p}_1$ is then defined as:

$$CS(\mathbf{p}_1, \mathbf{p}_0, u_r) = e(\mathbf{p}_1, u_r) - e(\mathbf{p}_0, u_r) \tag{2.38}$$

where u_r denotes the "reference utility level". When $r = 1$, the equivalent variation is defined. In this essay, I will focus on the case $r = 0$, which corresponds to the compensating variation.

$$\begin{aligned} CV(\mathbf{p}_1, \mathbf{p}_0, y) &= e(\mathbf{p}_1, u_0) - e(\mathbf{p}_0, u_0) \\ &= e(\mathbf{p}_1, u_0) - y \\ &= \bar{y}_1 - y \end{aligned} \tag{2.39}$$

Intuitively, the compensating variation (CV) is the change in income required to keep an individual on the same indifference curve when a change in prices occurs.

Given the assumption of maximising behaviour, the minimum expenditure required to attain the utility level in the base period is simply the level of total expenditure that we observe an individual at in this period, $e(\mathbf{p}_0, u_0) = y$. The compensated expenditure is then denoted by $\bar{y}_1 = e(\mathbf{p}_1, u_0)$. This is the expenditure that an individual requires to attain the same utility that they achieved in price regime $\mathbf{p}_0$ now that prices are $\mathbf{p}_1$. The compensated (or Hicksian) demand is the cheapest bundle of goods under $\mathbf{p}_1$ that achieves the utility level u_0. Let the compensated demand be given as: $\bar{\mathbf{q}}_1 = q(\mathbf{p}_1, \bar{y}_1)$. Then, $\bar{y}_1 = \mathbf{p}_1' \bar{\mathbf{q}}_1$.

CALCULATING $\bar{y}_1$ Calculating the compensating variation is complicated by the fact that $\bar{\mathbf{q}}_1$ and, thus, $\bar{y}_1$ are not directly observable. I apply the numerical techniques developed by Vartia (1983) to estimate these quantities from recovered ordinary MDI demands. Main Algorithm 1 (Vartia, 1983) provides an iterative method to solve the differential equation that characterises the change in expenditure that is required to move along the same indifference curve.

Following Vartia (1983), let $0 \leqslant j \leqslant 1$ represent an auxillary variable such that $\mathbf{p}(j)$ represents a differentiable curve in price space that connects $\mathbf{p}_0$ and $\mathbf{p}_1$, i.e. $\mathbf{p}(0) = \mathbf{p}_0$ and $\mathbf{p}(1) = \mathbf{p}_1$. $y(j)$ then traces out the associated expenditure evolution along this price path, starting with $y(0)$. Let us further define $\nu(j) = \nu(\mathbf{p}(j), y(j))$ as the value of the indirect utility function at the j^{th} position of the price path. Differentiation of the indirect utility function with respect to j yields:

$$\frac{dV(j)}{dt} = \sum_{k=1}^{K} \frac{\partial V(\mathbf{p}(j), y(j))}{\partial p_k(j)} \frac{dp_k(j)}{dj} + \frac{\partial V(\mathbf{p}(j), y(j))}{\partial y(j)} \frac{dy(j)}{dj} \tag{2.40}$$

This expression represents the rate of change in utility with changes in prices and expenditure along the price path. Application of Roy's identity gives us that

$$\frac{dV(j)}{dt} = \lambda(j) \left[\frac{dy(j)}{dj} - \sum_{k=1}^{K} q(\mathbf{p}(j), y(j)) \frac{dp_k(j)}{dj} \right] \tag{2.41}$$

Assuming monotonicity of the individual utility function, then $\lambda(j) > 0$. Therefore, a necessary and sufficient condition for a change in compensated income to keep an individual on the same indifference curve with changes in j is that the following differential equation is satisfied:

$$\frac{dy(j)}{dj} = \sum_{k=1}^{K} q(\mathbf{p}(j), y(j)) \frac{dp_k(j)}{dj} \tag{2.42}$$

Vartia (1983) provides a numerical solution to this differential equation, which makes use of the Collatz's and Adams interpolation of order 1 methods. Intuitively, the algorithm consists of averaging quantities at the current and previous points on the price paths, and multiplying this average by the change in price. By the envelope theorem, this returns the additional expenditure that is required to remain on the same indifference curve. The central idea motivating the procedure is to move in negligible increments from $\mathbf{q}_0$ to $\bar{\mathbf{q}}_1$. The algorithm is given formally below.

COMPENSATING INCOME ALGORITHM (VARTIA, 1983) Let $\mathbf{p}(j) = \mathbf{p}_0 + t(\mathbf{p}_1 - \mathbf{p}_0)$, $0 \leqslant j \leqslant 1$, be the linear price curve connecting $\mathbf{p}_0$ and $\mathbf{p}_1$. For a given integer S let $t_s = s/N$, $\mathbf{p}_s = \mathbf{p}(j_s)$ and generate a sequence of expenditures $y_1, \ldots, y_S$ so that

$$y_s - y_{s-1} = \frac{1}{2}(\mathbf{q}_s + \mathbf{q}_{s-1})'(\mathbf{p}_s - \mathbf{p}_{s-1}) \tag{2.A1.1}$$

where $\mathbf{q}_s = q(\mathbf{p}_s, y_s)$, $s = 1, \ldots, S$ and the starting values are $(\mathbf{p}_0, \mathbf{q}_0, y_0)$.

The solution to this series of equations is determined by iteration of the form:

$$y_k^{(m)} = y_{k-1} + \frac{1}{2}(\mathbf{q}(\mathbf{p}_s, y_s^{(m-1)}) + \mathbf{q}_{s-1})'(\mathbf{p}_s - \mathbf{p}_{s-1}) \tag{2.A1.2}$$

where $y_s^{(0)} = y_{s-1}$, $s \geqslant 1$.

When $|y_s^{(m)} - y_s^{(m-1)}|$ is negligible set $y_s = y_s^{(m)}$ and $\mathbf{q}_s = \mathbf{q}_s^{(m)}$ and start the calculation for the next s.

Assuming that the expenditure series converges, it converges to the compensated income $\bar{y}_1$ and $\mathbf{q}_S$ converges to the compensated demand $\bar{\mathbf{q}}_1$. The algorithm thus allows one to calculate estimates of exact consumer surplus and other associated welfare metrics that are associated with a hypothesised price change.

Operationalising this algorithm requires an estimate of $\mathbf{q}(\mathbf{p}_s, y_s^{(m)})$ at each step s and iteration m. I employ the MDI methodology to recover these quantities. Stepwise membership of $S^V(\mathbf{p}_s, y_s^{(m)}|D)$ is imposed and then rationality of the full quantity path verified before consumer surplus estimates are calculated. Application of the

MDI methodology has advantages over existing applications of Vartia's algorithm in the literature to date. Vartia's (1983) own application assumed a particular parametric functional form for the demand function, which risks significant misspecification bias in the resulting welfare estimates. Hausman and Newey (1995) extended Vartia's method to use nonparametric estimates of demand as inputs to the algorithm. However, although these authors provide tests of rationality, and thus of path independence of their consumer surplus estimates, they do not impose this property upon their demand estimates nor operationalise these tests in their empirical application. I impose pointwise membership of $S^V(\mathbf{p}_s, y_s^{(m)}|D)$ and also verify that the full sequence of recovered predictions satisfy rationality.

12.2 RECOVERING EXACT CONSUMER SURPLUS

In this section, I recover welfare metrics that are associated with changes in the price of the alcohol commodity aggregate; I thus recover consumer surplus measures along the marginal demand curve for alcohol. This application demonstrates how the techniques developed in this essay can be usefully applied to answer policy related questions. Specifically, I will consider the impact of a 5% rise in the price of alcohol upon consumer welfare. This is an interesting good to consider because accurate estimation of the price and welfare effects that result from taxing alcohol are of policy relevance. Furthermore, censoring of alcohol is a challenge that the applied demand theorist must tackle in this context. The methods constructed in this essay are well placed to address both of these issues; MDI demand predictions allow for unrestricted preference heterogeneity in the population and the construction of the IDW pilot implies that the recovered demands for individuals who do not drink are unaffected by changes in the price of alcohol.

IMPLEMENTATION For each individual, I specify the base price, $\mathbf{p}_0$, as the observed price vector for which the price of alcohol was at its lowest. This occurred in the last quarter of 2006. I keep total expenditure fixed at that observed in this period throughout the analysis. I then evaluate the welfare impact of a 5% rise in the price of the alcohol aggregate. Thus,

$$\begin{aligned} \mathbf{p}_0 &= [p_0^{Alc}, p_0^{Bread}, p_0^{Meat}, p_0^{Dairy}, p_0^{Veg}, p_0^{Sugar}, p_0^{Other}] \\ &= [1.90, 1.53, 1.57, 1.58, 1.52, 1.73, 1.52] \\ \mathbf{p}_1 &= [2.00, 1.53, 1.57, 1.58, 1.52, 1.73, 1.52] \\ y_0^i &= y_{t=2006,Q4}^i \end{aligned} \tag{2.43}$$

Starting from the observed situation $\{\mathbf{p}_0, \mathbf{q}_0^i, y_0^i\}$ for each individual i, I apply the Compensated Income Algorithm to recovered MDI demands in order to move along the indifference curve surface and approximate $\bar{\mathbf{q}}_1^i$. To calculate the linear price curve connecting $\mathbf{p}_0$ and $\mathbf{p}_1$, which is necessary to complete this task, I specify $S = 20$. I further set the tolerance level at $1e^{-9}$ with which to determine convergence of the main algorithm.[1]

From $\{\mathbf{p}_0, \mathbf{q}_0^i, y_0^i\}$, I first calculate an individual's demand using the MDI methodology at $(s, m) = (1, 1)$: $\mathbf{q}_1^{i,(1)} = \mathbf{q}(\mathbf{p}_1, y_0^i)$. Using this estimate, I form the average $\frac{1}{2}(\mathbf{q}_1^{i,(1)} + \mathbf{q}_0^i)$ and take the inner product $\frac{1}{2}(\mathbf{q}_1^{i,(1)} + \mathbf{q}_0^i)'(\mathbf{p}_1 - \mathbf{p}_0)$ to return $y_1^{i,(1)}$. The next step follows similarly: $\mathbf{q}_1^{i,(2)} = \mathbf{q}(\mathbf{p}_1, y_1^{i,(1)})$ and $y_1^{i,(2)} = y_0^i + \frac{1}{2}(\mathbf{q}_1^{i,(1)} + \mathbf{q}_0^i)'(\mathbf{p}_1 - \mathbf{p}_0)$. The calculations for $s > 1$ proceed in the same way. Global rationality of the full sequence of compensated demands is verified as a final step to ensure the path independence of recovered welfare estimates.

I also approximate the bounds on compensating income that are implied by revealed preference arguments alone. Varian (1982) built upon the earlier work of Afriat (1977) to illustrate how revealed preference restrictions can be used to bound the welfare effects of a price change. The upper and lower bounds on the compensated income, $\bar{y}_1^+(\mathbf{p}_1, \mathbf{q}_0)$ and $\bar{y}_1^-(\mathbf{p}_1, \mathbf{q}_0)$, are defined as:

$$\bar{y}_1^+(\mathbf{p}_1, \mathbf{q}_0) = \inf \mathbf{p}_1'\mathbf{q} \tag{2.44}$$

such that $\mathbf{q} \in RP(\mathbf{q}_0)$, and

$$\bar{y}_1^-(\mathbf{p}_1, \mathbf{q}_0) = \inf \mathbf{p}_1'\mathbf{q} \tag{2.45}$$

such that $\mathbf{q} \in NRW(\mathbf{q}_0)$, where $RP(\mathbf{q}_0)$ and $NRW(\mathbf{q}_0)$ are the sets of demands that respectively define the quantity vectors that are revealed preferred and revealed worse to $\mathbf{q}_0$.

However, computation of the theoretically-ideal bounds is difficult using standard linear programming techniques. Therefore, I follow Varian (1982) to calculate the "overcompensation" and "undercompensation" functions, which approximate the revealed preference bounds on compensating income. The methodology used for calculating these bounds is given in detail in Section 5 of Varian (1982).

RESULTS Application of these techniques returns individual estimates of compensated demands and compensated income with which I can recover individual level welfare metrics. I calculate a cost-of-living price index to reduce the comparison of individual welfare at the two budget regimes to a single scalar variable. This index gives

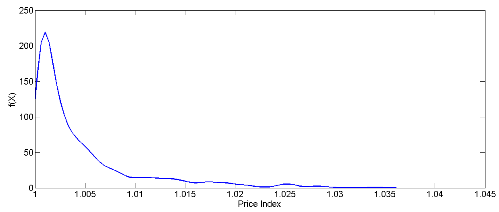
1 Final results were not found to be sensitive to the choice of S once $S > 15$. However, as convergence was typically quickly attained, I set $S = 20$ to assuage any accuracy concerns.

the ratio of the minimum expenditures necessary to achieve the base utility level at the two price regimes.

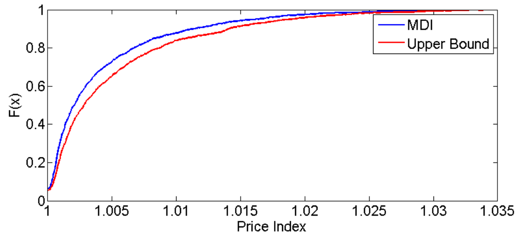
$$P(\mathbf{p}_0, \mathbf{p}_1, u_0) = \frac{\bar{y}_1}{y_0} \tag{2.46}$$

To illustrate the nature of the output yielded by application of the procedure, consider Table 15. This gives the output for three individuals whom are randomly selected from within each third of the distribution of price indices. These tables reveal heterogeneity in the budget shares for alcohol, the sign of cross-price effects and the strength of the behavioural response to the rise in price. Further, one should note that I recover no behavioural response or welfare effect for the 4% of my sample that bought no alcohol in the four years for which they were observed. This is a consequence of my choice of the IDW pilot prediction, which returns a zero budget share for alcohol at any new budget for these households.

Figure 19: Distribution of Price Index

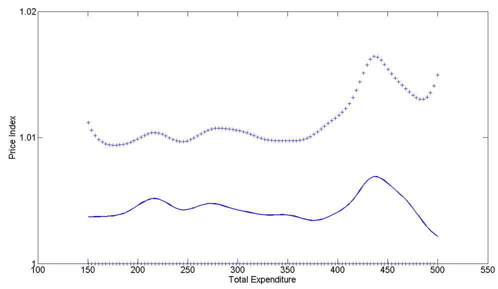


(a) Kernel Density

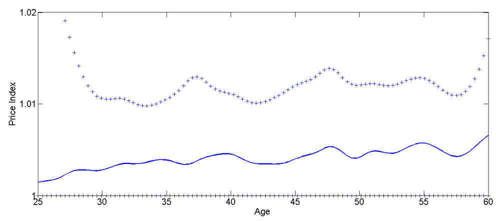


(b) Empirical Distribution Function

Figure 20: Observable Characteristics and Price Index



(a) Total Expenditure



(b) Age

Table 15: Individual Examples

(a) Individual 292

$\mathbf{q}_0$	$\mathbf{q}_1$	$\bar{\mathbf{q}}_1$
0.78	0.390	0.390
14.87	11.586	11.588
43.46	48.234	48.243
23.55	19.912	19.916
52.24	53.674	53.683
25.51	26.268	26.273
12.39	13.453	13.456

y_0	$\bar{y}_1$	Price Index
285.75	285.80	1.0002

(b) Individual 312

$\mathbf{q}_0$	$\mathbf{q}_1$	$\bar{\mathbf{q}}_1$
4.111	4.060	4.111
13.318	13.408	13.575
71.714	60.779	61.547
13.274	12.176	12.337
9.566	7.354	7.437
6.241	14.466	14.661
1.552	4.551	4.607

y_0	$\bar{y}_1$	Price Index
193.72	194.12	1.0021

(c) Individual 888

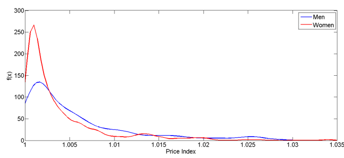
$\mathbf{q}_0$	$\mathbf{q}_1$	$\bar{\mathbf{q}}_1$
69.573	57.428	58.120
10.302	18.765	18.991
93.820	93.877	95.007
20.335	25.453	25.759
40.320	39.972	40.453
24.283	15.600	15.788
14.460	20.460	20.706

y_0	$\bar{y}_1$	Price Index
516.64	522.86	1.0120

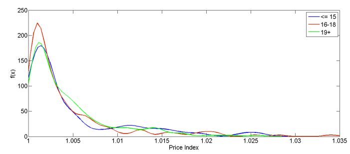
Table 16: Cost of Living Index

	Mean	St. Dev	Percentiles		
			5	50	95
All	1.0045	0.0056	1.0003	1.0023	1.0171
Upper Bound	1.0053	0.0062	1.0000	1.0028	1.0189
Lower Bound	1.0000	0.0000	1.0000	1.0000	1.0000
Women	1.0040	0.0049	1.0004	1.0022	1.0170
Men	1.0043	0.0065	1.0003	1.0023	1.0172
Age left Education:					
< 15 Years	1.0042	0.0054	1.0002	1.0022	1.0157
16-18 Years	1.0043	0.0060	1.0003	1.0018	1.0192
> 18 Years	1.0048	0.0053	1.0005	1.0027	1.0172

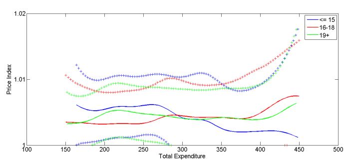
Figure 21: Observable Characteristics and Price Index



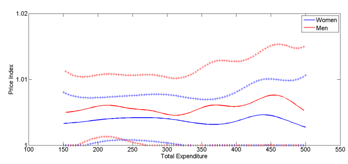
(a) Sex: Kernel Density



(b) Education: Kernel Density



(c) Sex: Kernel Regression Price Index on Total Expenditure



(d) Education: Kernel Regression Price Index on Total Expenditure

Table 17: Convergence and Updating

	Mean	St. Dev	Percentiles 5	50	95
Iterations	6.32	3.25	1	4	11
KL Divergence	0.0082	0.0181	0.0000	0.0023	0.0356

Results for the sample distribution of the price index are given in Table 16 and Figure 19. The approximation to the revealed preference lower bound to the compensated income returned the observed level of total expenditure for every individual. Although some rise in total expenditure is dictated by rationality, the Varian approximation to the lower bound does not reflect this given the paucity of data with which we work. Recovered MDI price indicies lay closer to the revealed preference upper bound. On average, individuals were found to require a 0.23% increase in their total expenditure to compensate them from a 5% rise in the price of alcohol. The distribution of welfare effects is positively skewed, which reflects the limited behavioural response of some heavy drinkers.

The heterogeneity in welfare metrics across the full sample of rational individuals does not appear to be related to observable characteristics. Nonparametric kernel regressions of recovered price indices on total expenditure and age do not show evidence of any interesting variation along this dimension. Stratifying my sample by sex and education does not reveal significant differences in the welfare consequences of the rise of a 5% rise in alcohol along these observable dimensions. Thus, as previously, taste heterogeneity and variation in welfare effects does not appear to be adequately captured by the sorts of variables commonly recorded in household surveys.

CONVERGENCE I end by briefly noting some features of the quality of the solution output. Convergence of the Vartia compensating income was typically fast, with a median of 4 iterations required at each step despite the tolerance level being set at a stringent level of $1e^{-9}$. Regarding the rationality of recovered demand predictions, 33% of my sample required pointwise modification of their pilot prediction at some point during the procedure. The extent to which these pilot predictions were modified was typically slight, as highlighted by reference to the figures in Table 17. The resulting quantity sequences returned from imposition of pointwise membership of the respective Varian supports sets all satisfied the requirement of global rationality and thus one can be confident of the path-independence of the recovered welfare metrics.

13 CONCLUSION

This essay has developed a methodology with which to recover globally rational demands at a set of previously unobserved intersecting budget hyperplanes. I have applied insights from information theory to impose membership of the "sufficient support set" upon unconstrained pilot estimates of consumer demands. My focus upon the recovery of multiple demand predictions required the extension of the revealed preference characterisation of global rationality of demand predictions to impose the cross-consistency of forecasts. I showed that forecast membership of the traditional "Varian" support set at each new budget does not guarantee the global rationality of a set of predictions because forecasts may not be jointly rationalisable by the same utility function. The requirement of cross-consistency of predictions was shown to introduce non-convexities to the revealed preference support set. This introduces nontrivial computational difficulties to attempts to practically impose a working definition of global rationality upon unconstrained estimates. In response, I developed a mixed integer linear programming characterisation of the constraints for global rationality such that this set is more easily operationalised.

I deferred to the Principle of Minimum Discrimination Information to impose global rationality on sets of unconstrained demand predictions. Updating forecasts in accordance with this principle dictates that one select the "closest" rational demand to the unconstrained pilot, where the distance between quantity bundles is measured by the Kullback-Leibler divergence. The approach has a number of compelling features and I highlighted that the resulting updated demands can be interpreted as the most likely rational demands to occur given a belief in the accuracy of the pilot specification.

An empirical application to individual consumption data drawn from the Kantar Worldpanel demonstrated the value of the MDI methodology. Well-behaved demand curves for Meat & Fish were recovered that satisfied our global rationality requirement. This application revealed heterogeneity in price effects that could not be characterised by a normally distributed error term nor could it be rationalised by an appeal to observable characteristics. Thus, the individual-specific construction of the pilot demand, which facilitates an allowance for maximal preference heterogeneity in the population, should be considered an asset of the approach. The empirical application further supported my distinction between the Varian and sufficient support sets. When demand forecasts were considered independently, or did not interact, as is the case along the marginal demand curve, application of the MDI updating procedure was rarely required. However, a Monte Carlo simulation revealed notable differences in the relative areas defined by the Varian and sufficient support sets when new budget hyperplanes intersect. As a consequence, the frequency of MDI updating was found to be greater when membership of the sufficient support set is imposed. Finally, I demonstrated how the MDI methodology can be used in conjunction with the algorithms developed by Vartia (1983) to estimate compensated income and welfare metrics at the individual level, such that path independent welfare analysis may be carried out within a nonparametric estimation setting.

My concern in this essay has been to produce an internally consistent methodology with which to recover globally rational demands and to demonstrate how this methodology may be operationalised to recover demands at new budget regimes. Future work will extend this essay in two dimensions. First, I will explore how the predictive accuracy of the MDI methodology in conjunction with the Euclidean IDW pilot compares to standard parametric and nonparametric statistical models. Second, I will research methods that increase the efficiency of the empirical strategy taken to impose membership of the sufficient support set. This will allow more ambitious recovery exercises to be engaged with. However, the work in this chapter has provided a thorough theoretical underpinning to the MDI approach and empirically demonstrated that it is operational, whilst extending the revealed preference and nonparametric demand literatures to provide a characterisation of global rationality and a systematic method to ensure its attainment.

Part III

TIME INCONSISTENCY, COLLECTIVE CHOICE & REVEALED PREFERENCE

We now move on to a setting in which the traditional neoclassical choice model has been widely discredited. This essay provides an analysis of models that are concerned with the intertemporal allocation of spending. Choice in this setting is often labelled as "irrational" and has been rationalised using models from behavioural economics such as hyperbolic discounting. However, empirical work in this area rarely distinguishes between individual and *collective* choice. Insufficient regard for the separateness of persons within the household provides provides a route through which observed choice behaviour can appear irrational because the process of preference aggregation can generate apparent inconsistencies in aggregate choice data.

This essay develops a revealed preference methodology to test the consistency of household consumption behaviour with simple economic models of intertemporal choice. In so doing, a methodology is created to establish whether time inconsistencies in household choice can be attributed to individual heterogeneity and renegotiation within the collective unit, rather than resulting from nonstationarities at the individual level.

An empirical application to household-level microdata highlights that an explicit recognition of the collective nature of choice allows the vast majority of household behaviour to be rationalised by traditional neoclassical choice theory that assumes preference stationarity at the individual level. For the particular short panel data set that is used in this essay, simply permitting limited intrahousehold heterogeneity in time preferences allows the choices of 98.4% of the sample to be rationalised by a model that assumes exponential discounting at the individual level. It is also found that couples who are characterised by a lower divergence in spousal discount rates are older, more likely to have children and wealthier, which we take as indications of experiencing higher match quality.

14

INTRODUCTION

Understanding the dynamics of intertemporal choice is a valuable, and increasingly salient, goal for modern society. Life expectancy, and life expectancy upon retirement, are on an upward trajectory, pushing issues of pension provision and retirement planning to the forefront. These trends are especially important given the rapid transition to defined-contribution pension schemes in developed countries, under which the characteristics of dynamic choice and individual self-control issues have more relevance (Thaler and Benartzi 2004).

We now move on from the setting of my previous essay to situate individual choice behaviour in an intertemporal framework. In this chapter, consumers must decide upon the allocation of spending across points in time, as well as how to split resources across goods. Furthermore, I now consider collective decision making processes and the behaviour of households as opposed to the choices of an atomistic individual. The behaviour of aggregates can appear irrational if the separateness of persons is not appropriately acknowledged. The observation of apparent irrationality in choice behaviour is another feature that distinguishes this essay from the last, in which we observed an almost universal satisfaction of the GARP constraints.

Specifically, this essay derives revealed preference tests of models of collective intertemporal choice and develops a methodology to recover the minimal heterogeneity in intrahousehold discount rates that is required to rationalise observed behaviour. My analysis reveals that textbook economic models are consistent with observed behaviour so long as they are applied to the appropriate decision making unit. Specifically, this essay provides empirical support for exponential discounting and perfect intrahousehold commitment as positivist modelling devices for short-to-medium run consumption choices. This result stands in marked contrast to much of the recent literature on time preferences and intrahousehold commitment. The following section introduces the literature relevant to this essay and places this essay's contribution in context.

THE DISCOUNTED UTILITY BENCHMARK Samuelson's (1937) canonical "discounted utility" (DU) model is the standard framework used by economists to conceptualise intertemporal choice. Consider a consumer who faces a consumption stream $\mathbf{C}_{ij} = \{\mathbf{q}_t\}_{t=i,\ldots,j}$, where $\mathbf{q}_t$ represents the quantities consumed at time t. Under the DU model, consumer preferences are represented by the following functional form:

$$U(\mathbf{C}_{ij}) = \sum_{t=i}^{j} \beta^{t-1} u(\mathbf{q}_t) \tag{3.1}$$

where u is a felicity function and $\beta = 1/(1+\sigma)$, with $\sigma \in [0, \infty)$ denoting the consumer's discount rate. These preferences are *time consistent*: the choice between alternatives does not depend upon when in time that choice occurs. Thus, if receiving X at t is preferred to receiving Y at $t+d$, the decision maker will always prefer X at τ to Y at $\tau + d$. Results from Koopmans (1960) provide an axiomatic foundation to the DU model and highlight that preferences are time consistent only if they can be expressed in the above format.

However, it is the model's virtues of simplicity and tractability that explains the widespread application of the DU framework. The predictive validity of the DU model is often thought to be on fragile footing. In experimental settings, decision makers systematically behave in a time inconsistent manner, acting impatiently today whilst planning to act patiently in the future. For example, one may prefer to receive \$100 today over receiving \$110 tomorrow, whilst they simultaneously prefer receiving \$110 in 31 days to receiving \$100 in 30 days. Such preference reversals are well documented in the psychology and economics literature (for a survey, see Frederick *et al.* 2002). Good intentions are often expressed regarding saving and responsible consumption, but the implementation of such plans tends to fall by the wayside. For example, in Choi *et al.*'s (2001) survey of 401(k) participants, 86% of the self-reported undersavers who expressed an intention to increase their savings rate, had made no changes to their behaviour four months later.

collective choice and time inconsistency Research effort has largely focused on modelling the sources of time inconsistency at the individual level. Standard methods of modelling discounting have been a prime target of criticism and the behavioural economics literature increasing favours frameworks that assume hyperbolic discount functions. Hyperbolic discount functions are characterised by a relatively high discount rate over short horizons and a relatively low rate over long horizons. This lack of constancy in the discount rate introduces a conflict between today's preferences and future preferences and a present bias to decision making. Hyperbolic discounting has been offered as an explanation for many stylised facts, from under-saving and excess co-movement of income and consumption (Laibson 1997, 1998; Angeletos *et al.* 2001) to procrastination, addiction and lack of exercise (O'Donoghue and Rabin 1999, 2001; Gruber and Koszegi 2001; DellaVigna and Malmendier

2006). Blow and Crawford (2013) have recently provided a revealed preference characterisation of this model.

However, in this essay I take a different approach and acknowledge how the *collective* nature of choice can rationalise time inconsistencies in aggregate behaviour. The preference structure associated with the DU approach is often applied to model group behaviour without modification. Under this "unitary" approach, one assumes that the collective acts as a single decision making unit, and therefore can be treated as if a rational individual. No explicit allowance is made for the separateness of persons nor the existence of preference heterogeneity within a collective.

Yet, many dimensions of intertemporal choice are better modelled as the outcome of group, rather than individual, decision making. Savings decisions are typically made at the household level, and the allocation of budgets over time within firms and committees are the product of multi-party deliberation. Even in the context of individual choice, one can consider the existence of multiple selves with distinct personalities rather than a single decision making unit. Thaler and Shefrin (1981) contrast the long-sighted "planner" within us to the short-sighted "doer", while Metcalfe and Mischel (1999) contrast our "hot" and "cool" systems. There is scientific evidence in support of the multiple selves hypothesis. Psychological studies find that different decision making systems interact within the brain during the evaluation of intertemporal prospects. Further, these systems appear to respond differently to the temporal dimensions of reward (Ainslie *et al.* 2005; McClure *et al.* 2004, 2007).

Acknowledgement of the collective nature of choice can help to rationalise the apparent time inconsistencies in household choice behaviour. In a group context, inconsistencies can arise simply from the aggregation of heterogeneous preferences. Both variation in individual discount rates and renegotiations of the household bargaining rule represent relevant considerations in this regard. First, consider the effect of discount rate heterogeneity. Deriving a time independent discount rate from the underlying preferences of a heterogeneous population has long been recognised as problematic (Marglin 1963; Feldstein 1964). When time preferences within a group differ, the collective preference is typically time inconsistent, even when the underlying population has perfectly time consistent preferences (Jackson and Yariv 2011; Zuber 2010). A present bias can be introduced to household choice through the aggregation of heterogeneous preferences. As time passes, the preferences of more impatient individuals are weighted less in the group's overall maximisation problem, causing the effective aggregate discount rate to increase over time. In fact, Jackson and Yariv (2011) show that, for a uniform distribution of discount rates in an otherwise homogeneous population, group utility maximisation generates patterns in aggregate behaviour that correspond to hyperbolic discounting.

Renegotiations of the household choice rule can also generate nonstationarities in family behaviour. Relative decision making power within the collective unit can vary with changes in the "outside options" of individuals and differences in time preferences can prompt periodic innovations in the intrahousehold preference weighting. Other things equal, it is optimal to favour impatient group members in early periods and patient members in later periods in the household bargaining process. However, as time passes and impatient individuals start to receive lower shares of the group surplus, there is an incentive for these individuals to demand a renegotiation of allocations in their favour or threaten to leave the group. These renegotiations prompt changes in the intrahousehold preference weighting, thereby generating nonstationarities in the collective preference. The degree of intrahousehold commitment is thus an important consideration in this context. Perfect commitment precludes the possibility of these renegotiations occurring, whilst deviations from the full commitment benchmark introduce instability into the intrahousehold preference weighting as an additional mechanism through which inconsistencies in observed household choice can be created.

Although it is true that nonstationarities at the individual level will translate into a failure of time consistency at the collective level, understanding whether the primary locus of inconsistent behaviour is at the individual or group level is important from both a methodological and policy perspective. The DU preference structure is tractable and parsimonious. Thus, if it cannot be rejected on the basis of choice behaviour, an appeal to Occam's razor provides a compelling analytic reason for its retention. Further, policy design should be influenced according to whether time inconsistent behaviour is the product of individual nonstationarities or collective aggregation issues.

ESSAY CONTRIBUTION This essay puts forward a nonparametric characterisation of household intertemporal choice and develops a revealed preference methodology for analysing the sources of collective time inconsistency. The approach incorporates insights gained from the extension of the revealed preference methodology to an intertemporal setting by Browning (1989) and Crawford (2010), and to the collective model by Cherchye, De Rock and Vermeulen (2007). The framework presented allows one to explore whether time inconsistencies in household choice can be rationalised by preference heterogeneity and renegotiation within the collective unit rather than individual nonstationarities. The empirical strategy further allows for the recovery of theory-consistent spousal discount rates and an assessment of the degree of intrahousehold commitment. In addition, I present further results concerning the separate nonparametric identification of intrahousehold renegotiations and binding credit constraints when one only has a finite number of observations on past collective choice behaviour available.

The methodology developed in this essay is novel and has clear advantages over existing empirical tests of the time consistency of household intertemporal behaviour. Current tests of dynamic collective choice models and time discounting are parametric and tend to reject the assumptions of constant discounting (see Frederick *et al.* 2002) and a time-independent intrahousehold preference weighting (Lich-Tyler 2004; Mazzocco 2007). However, these violations may be due to restrictive parametric assumptions, rather than a failure of the underlying functional form free theoretical framework.

Empirical studies of discounting and time consistency are sensitive to the parametric specification employed. The common assumption of linear consumption utility imparts an upward bias to discount rate estimates and is thought to contribute to the unrealistically high discount rate estimates that are observed in the literature. Although recent developments have seen the linear-utility specification somewhat relaxed (Anderson *et al.* 2008; Andreoni and Sprenger 2012), estimates are still dependent upon the set of functional form assumptions made concerning the form of the utility function. The experimental nature of existing time discounting studies can also be critiqued. Dohman *et al.* (2012) highlight that elicited preferences are not procedurally independent and discount rate estimates are hugely sensitive to the experimental design that is employed.

Turning to studies of intrahousehold commitment and renegotiation, Mazzocco's (2007) tests of intertemporal decision making and the constancy of the intrahousehold preference weighting centre on how "distribution factors" (Browning *et al.* 1994) enter log-linearised household Euler equations.[1] The risk of misspecification is high in this context. Carroll (2001), using simulated data, finds that log-linearisation introduces a significant bias to the estimation of preference parameters and concludes that the empirical estimation of approximated Euler equations should be abandoned. Furthermore, this empirical strategy requires the observation of all relevant distribution factors. This requirement is hugely demanding, especially given the limited information available in household survey data.

The methodology and empirical application presented here avoids such criticisms. The revealed preference approach is wholly nonparametric and, thus, as discussed in my "Overview" chapter, this essay's results are not contingent upon any particular specification of family member utility functions. The framework developed is also explicitly designed for use with household consumption data, although it can be profitably applied to an experimental setting. The empirical ap-

1 Distribution factors are factors that influence the relative power of family members but are independent of their preferences. See Section 15.1 for a formal discussion.

plication in this essay is one of the few in recent years to be fully grounded in "real world" household behaviour, rather than make use of choice data that has been elicited in an artificially constructed environment. This allows us to avoid many of the procedural nuances that plague experimental studies.

The primary conclusion of this essay is that accounting for the collective nature of choice allows time inconsistencies in aggregate household behaviour to be rationalised without positing nonstationarities in individual preferences. Simply allowing for some limited heterogeneity in familial discount rates allows the behaviour of 98.4% of households in our sample to be rationalised by standard models of household intertemporal behaviour. This result justifies the use of the elegant, if simplistic, assumption of exponential discounting so long as the theory is applied to the appropriate level of decision making, i.e. the individual.

OUTLINE. This essay proceeds as follows. Section 15 defines time consistency of the collective preference and outlines the revealed preference restrictions that are equivalent with the time consistency of household choice. Section 16 applies these revealed preference conditions to evaluate the empirical validity of a time consistent household consumption for a Spanish panel of household microdata. The hypothesis of time consistent household choice is heavily rejected for the data set of concern, even using nonparametric revealed preference restrictions. Given this result, Section 17 explores how a recognition of the collective nature of household choice can rationalise nonstationarities in the revealed collective preference and derives a nonparametric methodology for testing hypotheses on the sources of time inconsistent household behaviour. Section 18 continues with our empirical application and provides strong empirical support for a collective rationalisation of observed time inconsistencies. This section also considers the relationship between intrahousehold time preference heterogeneity and observable household characteristics. Section 19 concludes. A detailed data description and proofs of the Theorems in this essay can be found in Appendix B.

15 TIME CONSISTENCY AND COLLECTIVE CHOICE

The aim of this essay is to provide a framework that allows one to explore the sources of time inconsistencies in household choice. This section formally defines the concept of time consistency that is tested in this essay and derives simple revealed preference conditions that can be used to determine the time consistency of observed household choice.

15.1 THE COLLECTIVE PREFERENCE

Collective models explicitly recognise the separateness of persons within the household and allow for heterogeneity in family member felicity functions and discount rates. For notational simplicity, results are presented for a two-member household, which is constituted of members $m \in \{A, B\}$. The extension of results to an M-member ($M > 2$) household is trivial.

Individual preferences are represented by a time-additive discounted utility function, defined over private and public consumption. We assume N private goods and K public goods. At a given time t, private quantities $\mathbf{q}_t^m \in \mathbb{R}_+^N$, with associated discounted prices $\mathbf{p}_t \in \mathbb{R}_{++}^N$, are consumed non-jointly, while public quantities $\mathbf{Q}_t \in \mathbb{R}_+^K$, with associated discounted prices $\mathbf{P}_t \in \mathbb{R}_{++}^K$, are consumed jointly and non-exclusively.[1] Thus, associated with each member m is a concave and strictly increasing felicity function u^m and a discount rate $\sigma_m \in [0, \infty)$, such that a stream of public and private consumption $\mathbf{C}_{ij} = \{\mathbf{q}_t^m, \mathbf{Q}_t\}_{t=i,\dots,j}$ is evaluated as:

$$U(\mathbf{C}_{ij}^m) = \sum_{t=i}^{j} \beta_m^{t-1} u^m(\mathbf{q}_t^m, \mathbf{Q}_t) \tag{3.2}$$

where $\beta_m = 1/(1+\sigma_m)$.

Collective models do not assume *a priori* that individual preferences can be aggregated into a single time-independent household felicity function. However, the framework also does not specify a single intra-household bargaining process. Rather, the collective model simply assumes that some cooperative decision making process exists and that

1 Discounted prices are spot prices discounted by the nominal interest rate:

$$\begin{aligned} \mathbf{p}_t &= \textstyle\prod_{s=1}^{t} \rho_t/(1+i_s) \\ \mathbf{P}_t &= \textstyle\prod_{s=1}^{t} \mathbb{P}_t/(1+i_s) \end{aligned}$$

where ρ_t and $\mathbb{P}_t$ are the spot prices of private and public good at t, and i_t is the nominal interest rate in period t.

this process leads to Pareto efficient outcomes.[2] With this assumption, one can define the Pareto weight ω_t to summarise the bargaining process in period t. Consider a time horizon $|T|, T = \{1, ..., |T|\}$. For household H, the collective preference over a lifecycle consumption profile $\mathbf{C}^H = \{\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t\}_{t\in T}$ is given by:

$$U^H(\mathbf{C}^H) = \sum_{t=1}^{|T|} (\beta_A^{t-1} u^B(\mathbf{q}_t^m, \mathbf{Q}_t) + \beta_B^{t-1} \omega_t u^B(\mathbf{q}_t^m, \mathbf{Q}_t)) \qquad (3.3)$$

with $\omega_t = f(Z_t)$, where Z_t denotes the set of relevant "distribution factors" at time t. The standard theory places no restrictions on what variables count as relevant distribution factors, beyond requiring that they are independent of individual preferences (Browning *et al.* 1994). This lack of structure makes a nonparametric framework especially attractive as, unlike parametric tests of intertemporal behaviour, our methodology does not require a formal specification of the factors that jointly determine ω_t.

15.2 TIME CONSISTENCY

Given a household's time series of consumption choices, the first question of interest is whether there exists a time consistent household utility function that could have generated this choice pattern.

Following conditions given by Koopmans (1960), time consistency is seen primarily as a stationarity restriction: the passing of time should not have an effect on the preference ordering over fixed consumption streams. Formally, consider a consumption stream $\mathbf{C}_{ij}^H = \{\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t\}_{t=i,...,j}$. The collective preference over this consumption stream is defined as

$$U^H(\mathbf{C}_{ij}^H) = \sum_{t=i}^{j} (\beta_A^{t-1} u^B(\mathbf{q}_t^m, \mathbf{Q}_t) + \beta_B^{t-1} \omega_t u^B(\mathbf{q}_t^m, \mathbf{Q}_t)) \qquad (3.4)$$

Let $\mathbf{C}_{i'j'}^H$ represent an "updated" counterpart of $\mathbf{C}_{ij}^H$, which denotes the consumption streams $\mathbf{C}_{ij}^H$ shifted forward into the future by some amount $0 \leqslant \tau \leqslant |T| - j$, and let $(\mathbf{C}_{ij}^H, \mathbf{C}_{kl}^H)$ represent the combination of two (non-overlapping) consumption streams $\mathbf{C}_{ij}^H$ and $\mathbf{C}_{kl}^H$, where $i, j, k, l \in T, i < j, k < l$ and $j < k$ or $l < i$. With this notation, we can formally define time consistency of the collective preference as follows.

DEFINITION 3.1 (Jackson and Yariv, 2011) The collective preference is time consistent if the following two conditions are satisfied:

2 See, for example, Chiappori (1988, 1992) and Browning and Chiappori (1998) for detailed discussions of the Pareto efficiency assumption in collective household models. Mazzocco (2007) discusses this framework in a intertemporal context.

1. For any $i, j \in T$ with $i < j$,

$$\begin{aligned} & U^H(\mathbf{C}^H_{ij}) > U^H(\tilde{\mathbf{C}}^H_{ij}) \\ & \text{if and only if} \\ & U^H(\mathbf{C}^H_{i'j'}) > U^H(\tilde{\mathbf{C}}_{i'j'}) \end{aligned} \tag{3.D1.1}$$

2. For any $i, j, k, l \in T$ with $i < j, k < l$ and, in addition, $j < k$ or $l < i$,

$$\begin{aligned} & U^H(\mathbf{C}^H_{ij}, \mathbf{C}^H_{kl}) > U^H(\tilde{\mathbf{C}}^H_{ij}, \mathbf{C}^H_{kl}) \\ & \text{if and only if} \\ & U^H(\mathbf{C}^H_{ij}, \bar{\mathbf{C}}^H_{kl}) > U^H(\tilde{\mathbf{C}}^H_{ij}, \bar{\mathbf{C}}^H_{kl}) \end{aligned} \tag{3.D1.2}$$

Expression 3.D1.1 imposes stationarity on the collective preference; the ranking of consumption streams should not depend on when in time those streams are situated. Expression 3.D1.2 requires that the ranking of consumption streams is independent of periods with identical consumption bundles.

Conditions for time consistency Results originally given in Koopmans (1960), and applied to a collective setting by Jackson and Yariv (2012), imply that time consistency of the collective preference requires the ability to re-express the household preference in the following format:

$$\begin{aligned} U^H(\mathbf{C}^H_{ij}) &= \textstyle\sum_{t=i}^{s} (\beta_A^{t-1} u^B(\mathbf{q}^m_t, \mathbf{Q}_t) + \beta_B^{t-1} \omega_t u^B(\mathbf{q}^m_t, \mathbf{Q}_t)) \\ &= \textstyle\sum_{t=i}^{j} \beta_H^{t-1} u^H(\mathbf{q}^A_t, \mathbf{q}^B_t, \mathbf{Q}_t) \end{aligned} \tag{3.5}$$

for $\beta \in (0, 1]$.

For the collective preference to be recast in the above format, two conditions need to hold. First, household members must discount the future at the same rate, $\beta_A = \beta_B = \beta$. Second, the intrahousehold decision making mechanism must give rise to a constant Pareto weight across the lifetime of the household, $\omega_t = \omega$ for all $t \in T$. Only under these conditions can the household be modelled as a time-consistent representative agent with a latently separable, time-independent felicity function and constant discount rate.

15.3 Revealed preference conditions

The revealed preference approach to establishing the time consistency of household behaviour first derives the necessary and sufficient conditions under which observed choices can be rationalised by a stationary collective preference subject to the lifecycle budget constraint and then verifies the existence of a nonempty feasible set to these inequalities.

It is assumed that one observes $|T|$ consumption choices for household H, $T = \{1, ..., |T|\}$. For each observation t, one observes the privately consumed quantities, $\mathbf{q}_t$, and the publicly consumed quantities, $\mathbf{Q}_t$, as well as the corresponding discounted prices, $\mathbf{p}_t$ and $\mathbf{P}_t$. This defines a set of observations $S = \{\mathbf{q}_t, \mathbf{Q}_t, \mathbf{p}_t, \mathbf{P}_t\}_{t \in T}$. Note that only aggregate private quantities $\mathbf{q}_t$, but not the individual private quantities $\mathbf{q}_t^A$ and $\mathbf{q}_t^B$ (with $\mathbf{q}_t = \mathbf{q}_t^A + \mathbf{q}_t^B$), are observed. This assumption is motivated by the fact that, in most household surveys, information on who gets what is limited and the decomposition of private consumption is generally unobserved.

The rationalisation of household choice behaviour by the time consistent model is defined as follows.

DEFINITION 3.2 The set of observations S can be rationalised by the time consistent model if there exist, for all $t \in T$, private quantities $\mathbf{q}_t^A, \mathbf{q}_t^B \in \mathbb{R}_+^N$ (with $\mathbf{q}_t = \mathbf{q}_t^A + \mathbf{q}_t^B$) and a concave, strictly increasing felicity function u^H and a discount factor $\beta \in (0, 1]$ such that:

$$\sum_{t=1}^{T} \beta_H^{t-1} u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t) \geqslant \sum_{t=1}^{T} \beta_H^{t-1} u^H(\zeta_t^A, \zeta_t^B, \zeta_t^H) \tag{3.D2.1}$$

for all affordable consumption plans $\{\zeta_t^A, \zeta_t^B, \zeta_t^H\}_{t=1,...,T}$ that satisfy:

$$\sum_{t=1}^{T} \mathbf{p}_t^{'} \mathbf{q}_t + \mathbf{P}_t^{'} \mathbf{Q}_t = \sum_{t=1}^{T} \mathbf{p}_t^{'} (\zeta_t^A + \zeta_t^B) + \mathbf{P}_t^{'} \zeta_t^H \tag{3.D2.2}$$

In words, the data can be rationalised by a time consistent household preference if observed choices maximise discounted lifetime household utility out of affordable lifetime consumption plans for a coherent, stationary collective preference.[3]

Theorem 3.1 defines the revealed preference conditions that are equivalent with data rationalisation by the time consistent model.

THEOREM 3.1 The set of observations $S = \{\mathbf{q}_t, \mathbf{Q}_t, \mathbf{p}_t, \mathbf{P}_t\}_{t \in T}$ can be rationalised by the time consistent model if and only if there exist, for all $t \in T$, a utility number $u_t^H \in \mathbb{R}$ and a positive constant $\beta \in (0, 1]$ that, for any $s, t \in T$, satisfy

$$u_s^H - u_t^H \leqslant \frac{1}{\beta^{t-1}} [\mathbf{p}_t^{'} (\mathbf{q}_s - \mathbf{q}_t) + \mathbf{P}_t^{'} (\mathbf{Q}_s - \mathbf{Q}_t)]. \tag{3.T1.1}$$

PROOF See Appendix B.

Theorem 3.1 is an equivalence result. In words, if there exists a non-empty solution set to the inequalities defined by Theorem 3.1,

3 This definition of rationalisable behaviour does assume perfect foresight and perfect capital markets. However, the robustness of our empirical results to the relaxation of these assumptions is explored in Section 16.

then there exists a stationary household felicity function and discount factor that provide a perfect within-sample rationalisation of the data. The existence of a non-empty feasible set to the inequalities implies that the hypothesis of discount rate homogeneity and a stable intertemporal weighting of individual preferences within the household cannot be rejected. Verification of the existence of such a non-empty feasible set to 3.T1.1 is achieved easily. Conditioning on β, the inequalities defined by Theorem 3.1 are linear in unknowns and can be verified using standard linear programming techniques combined with a grid search for the discount factor.

SUMMARY This section has defined the concept of time consistency and noted that two conditions must be satisfied for the collective preference to meet this standard: the existence of a stationary household felicity function and the existence of a single exponential household discount rate. In this section, we have also derived the revealed preference necessary and sufficient conditions that are equivalent to rationalisation of household choices by the time consistent model. The next section applies the test defined by Theorem 3.1 to household panel data to determine the data consistency of the time consistent model.

16

AN EMPIRICAL TEST OF TIME CONSISTENCY

In this section, the conditions defined by Theorem 3.1 are tested using household panel data from the Spanish Continuous Family Expenditure Survey (the Encuesta Continua de Presupuestos Familiares, ECPF). The hypothesis of a time consistent collective preference is strongly rejected. The robustness of this result is then extensively analysed to evaluate the possibility that actual behaviour is time consistent but that the other assumptions that are embodied by the revealed preference conditions are violated.

16.1 THE DATA

The ECPF is a quarterly budget survey in which households are randomly rotated at a rate of 12.5% each quarter. Participating households are surveyed in the same week of each successive quarter, with each adult family member completing an expenditure diary in which they record their spending during the survey week. The data used are drawn from the period 1985-1997 for a sub-sample of couples in which the husband is in full time employment and the wife is out of the labour force in all periods. The requirement of stable employment status is to allow the potential nonseparability of leisure and consumption to be ignored for the time being, as our theoretical framework does not presently consider the household labour supply decision. We further restrict our sample to families with a stable number of children. This is to prevent our results being influenced by the inevitable disruption associated with the birth of a new baby, or a child leaving home, from unduly influencing our results. Finally, for matters of comparability, we only consider those households that participated in the survey for eight consecutive quarters ($|T| = 8$).[1]

We consider household choice over a commodity bundle of 15 nondurable goods. Each good is classified as either 'private' or 'public' consumption. Our bundle of private consumption consists of 11 goods: (1) Food and non-alcoholic drinks, (2) Alcohol, (3) Tobacco, (4) Clothing and footwear, (5) Nondurable medicines, (6) Medical services, (7) Transportation, (8) Petrol, (9) Leisure (cinema, theatre, clubs for sport), (10) Personal services, (11) Restaurants and bars. Our bundle of public consumption consists of 4 goods: (1) Rent, (2) Energy, (3) Home Services (heating, water and furniture repair) and (4) Nondurables at home (cleaning products). Prices are calculated from published prices aggregated to correspond to the listed expenditure cat-

1 This is the maximum number of observations on a single household in the ECPF.

egories, discounted by the average interest rate on consumer loans.[2] Summary statistics and more thorough variable definitions are given in Appendix B.

16.2 Revealed preference tests

As explained in Section 15, testing whether household consumption choices can be rationalised by a time-consistent household utility function boils down to establishing the existence of a non-empty feasible set to the linear conditions defined by Theorem 3.1, combined with a grid search for the discount factor β. Results are reported for a grid search on individual discount factors over the interval [0.9,1], with a spacing of 0.005.[3] However, our results are robust to alternative grid search specifications (including a grid search across the full interval (0,1]).

It is also worth emphasising that our tests allow for unrestricted preference heterogeneity across households. The theory-consistency of each household's behaviour is tested independently and the data is not pooled at any stage.

16.3 Results

Table 18 reports the test results obtained from applying the revealed preference conditions in Theorem 3.1 to the ECPF data. Only two households in the sample of 1585 households can be rationalised by a time consistent collective preference. Putting it differently, the pass rate is an exceptionally low 0.13%, representing a decisive failure of the time consistent model. This result complements that of Crawford (2010), who reports a pass rate of less than 5% for a similar model and his particular sample of ECPF households. Therefore, the time consistent model, as it stands, is unable to rationalise observed household intertemporal choice behaviour.

Table 18: Time Consistent Model Results

	No. Households	% Households
Pass Rate	2	0.13

The following subsections report additional results that lay claim to the robustness of this finding. In Section 18, the time consistent model is evaluated in terms of two performance measures that are frequently

2 The actual (quarterly) interest rate on savings and/or loans faced by a particular household (i_t^*) may differ from the publicly listed (quarterly) interest rate on consumer loans (i_t). However, the results prove robust if we account for some deviations, e.g. by assuming $i_t^* = i_t \pm \epsilon$, where $\epsilon = \pm 0.01$ or ± 0.02.

3 This corresponds to a search for discount rate on [0,0.11], which given the quarterly periodicity of our data is not unreasonable.

used to evaluate the empirical performance of an economic model in revealed preference analyses: discriminatory power and predictive success. These results provide further empirical evidence against the time consistent model as an explanation for observed household behaviour.

16.4 ROBUSTNESS CHECKS

The revealed preference test derived in Section 15 is based upon a number of very exacting conditions and makes some implicit assumptions that do not specifically relate to the time consistency of household behaviour. It is plausible that measurement error or a failure of perfect foresight and perfect capital market assumptions, rather than any time inconsistencies in household choice, explain the decisive rejection of the time consistent model. In the following sections, the impact of relaxing these background assumptions on the pass rate of the time consistent model is explored.

MEASUREMENT ERROR The strong rejection of the time consistent model could be the result of errors in listed expenditure data, rather than deviations of actual choice behaviour from the prescriptions of the time consistent model. Revealed preference tests are "sharp" in that household behaviour is either consistent with the model in question or it is not. Thus, small deviations in observed quantities away from the truth can have a large impact on pass rates. Varian's (1985) procedure is applied to explore the possibility that measurement error is behind the rejection of our revealed preference tests.[4]

Consider a data set S that violates the inequality constraints defined by Theorem 3.1. Our measurement error procedure first estimates the minimal adjustments of the quantity data required to obtain consistency with the revealed preference conditions, and then calculates the standard deviation on the measurement error in quantities that would be necessary for the null hypothesis of theory rationalisable behaviour to be accepted at a significance level α.

For a household observation t, let q_t^n represent the observed quantity of the n^{th} private good and Q_t^k the observed quantity of the k^{th} public good. Due to measurement error, these quantities may deviate from $q_t^{n\star}$ and $Q_t^{k\star}$, the true, but unobserved, values of the private and public quantities. The divergence between listed and true quantities can be quantified by the "relative quantity errors":

$$\epsilon_t^n = \frac{(q_t^{n\star} - q_t^n)}{q_t^n} \text{ and } \epsilon_t^k = \frac{(Q_t^{k\star} - Q_t^k)}{Q_t^k} \tag{3.6}$$

4 This procedure can also be interpreted as a mechanism through which to investigate the sensitivity of our results to seasonality trends in the considered quantity series.

As the true values, $q_t^{n\star}$ and $Q_t^{k\star}$, are not observed, the necessary measurement error required for listed expenditure data to pass the revealed preference conditions are approximated by calculating the minimal adjustments of the quantity data that is required to obtain theory consistency. Under the assumption that $\epsilon_t^i \sim N(0, \sigma_\alpha^2)$, one can calculate the critical standard deviation σ_α on measurement error that is needed such that the null hypothesis of theory-rationalisable behaviour cannot be rejected at some significance level α. The hypothesis of rationalisability at a significance level α is rejected if σ_α exceeds our prior beliefs about the true standard deviation σ. I provide more details on this procedure in Appendix B.

For each household and a significance level α, the critical standard deviation σ_α is calculated for household behaviour to be rationalisable by the time consistent model. Figure 22 shows the kernel distribution that corresponds to a significance level $\alpha = 0.05$, and Table 19 presents summary statistics on the distribution of σ_α for various significance levels.[5]

Table 19: Summary Statistics for σ_α Distribution

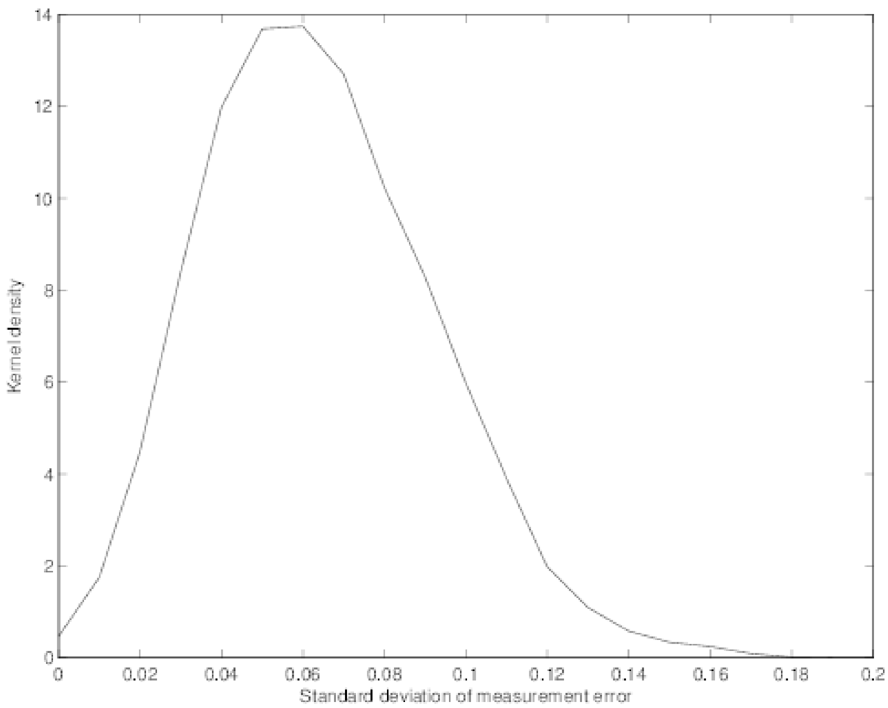
	Mean	Min	Percentiles			Max
			25	50	75	
$\alpha = 0.01$	0.061	0	0.042	0.059	0.079	0.161
$\alpha = 0.05$	0.064	0	0.043	0.061	0.082	0.168
$\alpha = 0.10$	0.065	0	0.045	0.063	0.084	0.172

Results are relatively insensitive to the chosen rejection level α. We observe considerable interhousehold variation in σ_α, ranging from zero for the two households that were consistent with the time consistent model in the absence of measurement error, to approximately 17%. The entire sample can be rationalised by the time consistent model if we believed that the standard deviation of relative quantity errors ranged to 17%. To rationalise three-quarters of the households in our sample, a standard deviation on relative quantity errors of approximately 8% is required. These results indicate that substantial measurement error in observed quantities is required to rationalise a majority of households by the time consistent model. It is unlikely that the error in the ECPF data is that large, casting doubt on the hypothesis that measurement error explains the decisive rejection of the time consistent model.

IMPERFECT CAPITAL MARKETS The theoretical framework assumes perfect capital markets and perfect foresight. These assumptions are

5 All displayed density functions in this essay are estimated using kernel methods. Specifically, a Gaussian kernel function is employed, with the bandwidth set equal to the Silverman (1986) "rule-of-thumb". See Section 9.3 for further details.

Figure 22: Kernel Plot of σ_α Distribution



clearly very strong even though we only consider household behaviour over a relatively short, two-year period.

Under the assumption of perfect capital markets, households are able to borrow and save at the same interest rate. To relax this assumption, a method analogous to Demuynck and Verriest (2012) is applied. Their method recognises that households may face an unobserved borrowing constraint that hinders their ability to smooth consumption across time periods. If this additional constraint never binds, the revealed preference conditions for the time consistent model are unchanged from those defined by Theorem 3.1. However, if the borrowing constraint binds, then household consumption will deviate from its optimal path: a binding borrowing constraint in period t implies that total household consumption in t is less than it would be under unconstrained borrowing.

Allowing for a binding borrowing constraint facilitates a rise in the pass rate associated with the time consistent model to 24%. Methodological details are given in Appendix B. Although this represents a sizeable improvement, we will see that the rise in the pass rate remains very modest compared to that achieved by an acknowledgement of the collective nature of choice.[6]

6 In the next section, additional theoretical results concerning the separate identification of binding borrowing constraints and innovations in the Pareto weight are

Table 20: Imperfect Capital Market Results

	Perfect Capital Market	Imperfect Capital Market
Pass Rate (%)	0.13	23.66

imperfect foresight Finally, the assumption that the household has perfect foresight with respect to all relevant information, such as future prices, incomes and interest rates is relaxed. In many settings, this assumption is utterly implausible. However, given the short nature of the panel used in this essay, the assumption may not be overly restrictive. In the absence of market imperfections, assuming perfect foresight is equivalent to assuming a time-constant marginal utility of wealth. The assumption of perfect foresight over the entire time horizon is relaxed in a simple way, by reducing the set of observations for which a time-constant marginal utility of wealth is imposed. In particular, the length of a time series on a household is sequentially reduced from $|T| = 8$ to $|T^{'}|$, $2 \leqslant |T^{'}| \leqslant 7$. By construction, the pass rate is weakly increasing as the set of observations shrinks.

Table 21 reports the test results. The time consistent model is still heavily rejected if perfect foresight is imposed for longer periods. This suggests that households may not be able to accurately forecast prices and incomes far into the future. However, in Section 18 it is shown that perfect foresight over the entire time horizon of eight quarters cannot be rejected for the vast majority of households as soon as we allow for intrahousehold heterogeneity in discount rates. Therefore, it is felt that the heavy rejection of the time consistent model cannot be solely attributed to the failure of the perfect foresight assumption.

Table 21: Imperfect Capital Market Results

Time Horison	2	3	4	5	6	7	8
Pass Rate (%)	90.5	51.8	20.2	4.9	1.2	0.3	0.1

summary In summary, this section has sought to establish the empirical validity of a time consistent model of household consumption behaviour. The hypothesis that this model provides an adequate framework for explaining observed patterns of family choice has been decisively rejected for our sample. This result is robust to relaxing the assumptions of an absence of measurement error, perfect capital markets and perfect foresight upon which our testing framework is prefixed. Given this negative result, in the following sections we explore the question of whether an explicit recognition of the collective

derived. As in this section, constancy of ω_t is imposed, I here abstract from such issues.

nature of household choice can rationalise observed household consumption patterns.

17 COLLECTIVE CHOICE AND TIME INCONSISTENCY

The collective preference cannot be recast in the format required for time consistency in the presence of either intrahousehold discount factor heterogeneity or innovations in the Pareto weight. The failure of these assumptions manifest themselves in different ways in observed choice behaviour. This section explores these sources of time inconsistency in greater depth and utilises Mazzocco's (2007) theoretical framework to develop an empirical strategy for distinguishing between the different sources of time inconsistent household behaviour.

17.1 INDIVIDUAL HETEROGENEITY

Within the collective unit, individual heterogeneity and innovations in the Pareto weight, ω_t, can both introduce time inconsistencies to household consumption patterns, even if individuals within the family have perfectly time consistent preferences.

With intrahousehold discount rate heterogeneity, one cannot collapse individual discount factors into a single household discount factor: if $\beta_A^t \neq \beta_B^t$, then $\beta_A^t + \beta_B^t \neq (\beta_A + \beta_B)^t$. This implies that members' preferences are weighted differently in the household allocation problem at different points in time, even if the Pareto weight remains constant. Other things equal, the preferences of the more patient member become relatively more important in future time periods. This introduces time inconsistency to the collective preference. Thus, one cannot talk of a single household discount weight with intrahousehold discounting heterogeneity.

Jackson and Yariv (2011) prove that for a group of otherwise homogeneous individuals who are selecting over a common consumption stream, any heterogeneity in time preferences imparts a present bias on the collective preference, and that with a uniform distribution of discount rates in a homogeneous population, the collective time preference is hyperbolic. Thus, time inconsistencies in group behaviour need not be derivative of nonstationarities at the individual level. Rather, present bias in household choice can arise from the aggregation of heterogenous time preferences.

THE FULL EFFICIENCY MODEL Individual heterogeneity is the only source of time inconsistency in Mazzocco's (2007) "full efficiency" model. The model assumes the existence of a perfect commitment mechanism that removes the possibility of intrahousehold renegotiation. This implies the existence of a single, fixed Pareto weight to summarise the household decision making process across the full lifetime of the household. The time consistency of allocations is not an issue in this environment and allocations lie on the *ex ante* Pareto frontier.

For a given data set S, the full efficiency model corresponds to the following rationalisation condition.[1]

DEFINITION 3.3 The set of observations S can be rationalised by the full efficiency model if there exist, for all $t \in T$, private quantities $\mathbf{q}_t^A, \mathbf{q}_t^B \in \mathbb{R}_+^N$, with $\mathbf{q}_t^A + \mathbf{q}_t^B = \mathbf{q}_t$, and concave, strictly increasing felicity functions $u^A(\mathbf{q}_t^A, \mathbf{Q}_t)$ and $u^B(\mathbf{q}_t^B, \mathbf{Q}_t)$, discount factors $\beta_A, \beta_B \in (0, 1]$ and a Pareto weight $\omega > 0$ such that:

$$\begin{aligned} &\textstyle\sum_{t=1}^{T} \beta_A^{t-1} u^A(\mathbf{q}_t^A, \mathbf{Q}_t) + \omega\beta_B^{t-1} u^B(\mathbf{q}_t^B, \mathbf{Q}_t) \\ &\geqslant \\ &\textstyle\sum_{t=1}^{T} \beta_A^{t-1} u^A(\zeta_t^A, \zeta_t^H) + \omega\beta_B^{t-1} u^B(\zeta_t^B, \zeta_t^H) \end{aligned} \tag{3.D3.1}$$

for all affordable consumption plans $\{\zeta_t^A, \zeta_t^B, \zeta_t^H\}_{t=1,\dots,T}$ that satisfy:

$$\sum_{t=1}^{T} \mathbf{p}_t'\mathbf{q}_t + \mathbf{P}_t'\mathbf{Q}_t = \sum_{t=1}^{T} \mathbf{p}_t'(\zeta_t^A + \zeta_t^B) + \mathbf{P}_t'\zeta_t^H \tag{3.D.1}$$

As for the time consistent model, the constant Pareto weight ω incorporates the combined impact of all changes in distribution factors over time; it can be considered as the average relative power of family members across the lifetime of the household. Therefore, with the existence of a perfect commitment mechanism, the only source of time inconsistency in aggregate behaviour is discount rate heterogeneity.

REVEALED PREFERENCE CONDITIONS I now consider how one can one test the importance of individual heterogeneity as a source of time inconsistency. Discount rate heterogeneity precludes the possibility of representing the collective preference in a representative-consumer format. This implies that the composition of household consumption and its distribution between family members plays a central role in revealed preference tests of the full efficiency model. This has two important implications for the revealed preference conditions associated with the full efficiency model. First, for privately consumed goods, the information on $\mathbf{q}_t^A$ and $\mathbf{q}_t^B$ is relevant. Second, for publicly consumed goods, the relevant "prices" for an individual family member are the so-called Lindahl prices, $\mathbf{P}_t^A$ and $\mathbf{P}_t^B$. These prices coincide with a family member's marginal willingness to pay and, given the maintained assumption of cooperative decision making, sum to observed prices, $\mathbf{P}_t^A + \mathbf{P}_t^B = \mathbf{P}_t$.

1 One important difference between the theoretical framework in this essay and that of Mazzocco's (2007) full efficiency model is the assumption of perfect foresight. Revealed preference tests of martingale processes lack content as, without a specification of the expectation process, one can always posit an unexpected shock to rationalise behaviour.

Theorem 3.2 gives the condition under which household choice can be rationalised by the full efficiency model. If there exists a nonempty feasible set to the inequalities defined by Theorem 3.2 but no such set exists for those defined by Theorem 3.1, then the hypothesis that time inconsistencies in aggregate behaviour are the result of discount rate heterogeneity within the family cannot be rejected.

THEOREM 3.2 The set of observations S can be rationalised by the full efficiency model if and only if there exist, for all $t \in T$, private quantities $\mathbf{q}_t^A, \mathbf{q}_t^B \in \mathbb{R}_+^N$, Lindahl prices $\mathbf{P}_t^A, \mathbf{P}_t^B \in \mathbb{R}_+^K$, utility numbers $u_t^A, u_t^B \in \mathbb{R}$ and constants $\beta_A, \beta_B \in (0, 1]$ that, for any $s, t \in T$, satisfy:

$$u_s^A - u_t^A \leqslant \frac{1}{\beta_A^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^A - \mathbf{q}_t^A) + \mathbf{P}_t^{A\prime}(\mathbf{Q}_s - \mathbf{Q}_t)]. \tag{3.T2.1}$$

$$u_s^B - u_t^B \leqslant \frac{1}{\beta_B^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^B - \mathbf{q}_t^B) + \mathbf{P}_t^{B\prime}(\mathbf{Q}_s - \mathbf{Q}_t)]. \tag{3.T2.2}$$

with

$$\begin{aligned} \mathbf{q}_t^A + \mathbf{q}_t^B &= \mathbf{q}_t \\ \mathbf{P}_t^A + \mathbf{P}_t^B &= \mathbf{P}_t \end{aligned} \tag{3.T2.3}$$

PROOF See Appendix B.

In words, if there exists a nonempty feasible set to the inequalities defined by Theorem 3.2, then there exists a pair of felicity functions and constant discount rates that provide a perfect within-sample rationalisation of the household data. Conversely, if one cannot find values of all relevant variables such that these inequalities hold, then there does not exist a theory-consistent specification of household member preferences and a constant Pareto weight that rationalise the observed consumption stream. Therefore, these inequalities allow us to test whether time inconsistencies in choice can be explained by an appeal to intrahousehold discount rate heterogeneity alone. If a nonempty feasible set is associated with the full efficiency constraints, one cannot reject the hypothesis that time inconsistency in household choice is simply the product of individual heterogeneity within the collective unit. One does not necessarily require nonstationarities in individual preferences or renegotiations of the household choice rule over time to rationalise observed behaviour.

RELATIONSHIP TO TIME CONSISTENT INEQUALITIES The inequalities that define Theorem 3.2 can be used to demonstrate the empirical equivalence of the time consistent model and the full efficiency model with homogeneous discount factors. Summing 3.T2.1 and 3.T2.2 with homogeneous household member discount rates returns:

$$\begin{aligned}(u_s^A + u_s^B) - (u_t^A + u_t^B) &\leqslant \frac{1}{\beta_H^{t-1}}\mathbf{p}_t'([\mathbf{q}_s^A + \mathbf{q}_s^B] - [\mathbf{q}_t^A + \mathbf{q}_t^B]) \\ &\quad + \frac{1}{\beta_H^{t-1}}[\mathbf{P}_t^A + \mathbf{P}_t^B](\mathbf{Q}_s - \mathbf{Q}_t) \\ \tilde{u}_s^H - \tilde{u}_t^H &\leqslant \frac{1}{\beta_H^{t-1}}\{\mathbf{p}_t'[\mathbf{q}_s - \mathbf{q}_t] + \mathbf{P}_t'[\mathbf{Q}_s - \mathbf{Q}_t]\}\end{aligned} \tag{3.7}$$

These inequalities are identical to those defined by Theorem 3.1.

With intrahousehold discount rate heterogeneity, Theorem 3.2 cannot be expressed in an equivalent format to Theorem 3.1. To provide some intuition, let us focus on a very simple case when the household only consumes a single public good, k, which both A and B derive utility from. In this instance, summing 3.T2.1 and 3.T2.2 returns:

$$\begin{aligned}\tilde{u}_s^H - \tilde{u}_t^H &\leqslant \left(\frac{1}{\beta_A^{t-1}}P_t^{k,A} + \frac{1}{\beta_B^{t-1}}P_t^{k,B}\right)(Q_s^k - Q_t^k) \\ &\leqslant \left(\phi_t^k \beta_A^{1-t} + (1-\phi_t^k)\beta_B^{1-t}\right) P_t^k (Q_t^k - Q_s^k)\end{aligned} \tag{3.8}$$

where $\phi_t^k = P_t^{k,A}/P_t^k$. This illustrates that with intrahousehold discount rate heterogeneity, the composite 'household discount factor' is a time-varying weighted average of individual discount factor. This factor is good specific and depends upon the relative preference of household members over good k at time t. This follows from the fact that $P_t^{k,A} = \beta_A^{t-1}\nabla_{Q_t^k}u^A(Q_t^k)/\eta$, where η is the Lagrange multiplier associated with the lifetime budget constraint.[2] In early periods, the composite discount factor on a good will be closely aligned to that of the family member with the greatest relative preference. However, as time passes, the composite household discount factor will converge to that of the most patient family member.

17.2 RENEGOTIATION

If household behaviour is inconsistent with the full efficiency model, an appeal to more than just discount rate heterogeneity is required. The second condition for time consistency of the collective preference is the existence of a time independent household felicity function. This rests upon the constancy of the Pareto weight across the full lifetime of the household. Whether this is necessarily attained depends upon the existence of a perfect intrahousehold commitment mechanism. Without a perfect commitment device, the Pareto weight can vary over time to reflect renegotiations of the household choice rule. These renegotiations open up an additional mechanism for time inconsistent behaviour.

2 This is shown formally in the proof of Theorem 3.2 in Appendix B.

THE NO COMMITMENT MODEL Mazzocco's (2007) "no-commitment" model weakens the assumption of perfect intrahousehold commitment that was made by the full efficiency model. Under the no-commitment model, the household solves the lifetime bargaining problem subject to additional incentive compatibility constraints. Mazzocco (2007) classes a consumption stream as incentive compatible if it does not provide an incentive for any family member to quit the household at some point to take their outside option.[3] An individual's outside option is defined as the utility that they could derive from divorcing and continuing in the world alone. Under the no-commitment model, the household solves the following maximisation problem:

$$\max_{\{\mathbf{q}_t, \mathbf{Q}_t\}_{t=1,\dots,T}} \sum_{t=1}^{T} \beta_A^{t-1} u^A(\mathbf{q}_t^A, \mathbf{Q}_t) + \omega \beta_B^{t-1} u^B(\mathbf{q}_t^B, \mathbf{Q}_t) \tag{3.9}$$

subject to

$$\begin{aligned} \textstyle\sum_{s=1}^{T-t} \beta_m^{s-1} u^m(\mathbf{q}_{t+s}^m, \mathbf{Q}_{t+s}) &\geqslant \bar{u}_t^m \\ \textstyle\sum_{s=1}^{T-t} \beta_m^{s-1} O_{t+s}^m &= \bar{u}_t^m \\ \textstyle\sum_{t=1}^{T} \mathbf{p}_t' \mathbf{q}_t + \mathbf{P}_t' \mathbf{Q}_t &= Y \end{aligned}$$

where O_t^m gives the utility of being single and Y is lifetime income. $\bar{u}_t^m$ represents the "outside option" of member m in period t. The first constraint represents the incentive compatibility condition. This constraint is absent from the full efficiency model. In words, this constraint requires that, at each point in time, each household member is guaranteed at least what they could derive if they left the household and continued as a singleton for the rest of their life. The final constraint gives the standard intertemporal budget constraint.

This maximisation problem is not recursive as the incentive compatibility constraints are forward looking; future actions limit the set of feasible allocations today. This creates difficulties when deriving the revealed preference conditions associated with this framework because whether these constraints are satisfied at any time t depends upon the future endogenous path of consumption allocations.

Using the method developed by Marcet and Marimon (1998), the model can be formulated as an equivalent recursive saddle point problem. Letting ϕ_t^m represent the discounted Lagrange multiplier on the incentive compatibility constraint for member m, the household optimisation problem can be expressed as:

$$\max \sum_{t=1}^{T} \sum_{m \in \{A,B\}} (\omega_t^m \beta_m^{t-1} u^m(\mathbf{q}_t^m, \mathbf{Q}_t) - \phi_t^m \bar{u}_t^m) \tag{3.10}$$

subject to

$$\sum_{t=1}^{T} \mathbf{p}_t^{'} \mathbf{q}_t + \mathbf{P}_t^{'} \mathbf{Q}_t = Y$$

3 This definition is akin to Shaked and Sutton's (1984) formulation.

where $\omega_1^A = 1, \omega_t^A = \omega_{t-1}^A + \phi_t^A$ and $\omega_1^B = \omega$, $\omega_t^B = \omega_{t-1}^B + \phi_t^B$.

DISCUSSION In the no-commitment model, the household choice rule, summarised by the set of intrahousehold preference weights, ω_t^m, is sequentially renegotiated to reflect changes in the slackness of the incentive compatibility constraints: $\omega_1^A = 1, \omega_t^A = \omega_{t-1}^A + \phi_t^A$ and $\omega_1^B = \omega, \omega_t^B = \omega_{t-1}^B + \phi_t^B$. Assuming positive gains to marriage continuation for at least one spouse in every time period, there will always be at least one individual who is strictly better off if the marriage continues rather than dissolving through divorce. Therefore, the incentive compatibility condition can only bind for one family member at any point in time, i.e. if $\phi_t^A \neq 0$ then $\phi_t^B = 0$, and vice versa. In periods where the incentive compatibility constraint binds for some member, the weight assigned to her preferences is increased until she is indifferent between taking her outside option and staying within the household. This new weighting of family member preferences then prevails in subsequent time periods until an incentive constraint again binds and another reweighting of preferences is implemented.

To develop some intuition for the model, consider a wife who is considering divorcing her husband to become single. First imagine that she has been promoted at work and can now achieve a higher utility by becoming single and consuming the full product of her labour rather than by staying within the family and consuming those allocations dictated by the committed model with the current sharing rule. Before signing the divorce papers, her husband can make her a new Pareto weight offer. As long as the wife's future utility implied by this new ω_t offer exceeds that implied by her outside option, the wife will not proceed with a divorce. Thus, the husband will offer a new Pareto weight regime, weighted more in his wife's favour, such that her utility from staying within the household just matches that of becoming single.

Now imagine that, although the wife's promotion raises the value of her outside option, it does not do so by enough to make divorce preferable to remaining married and continuing to consume those allocations dictated by the current Pareto weight. In this case, no renegotiation of the sharing rule will occur. Despite the wife's promotion, her husband can simply offer the allocations implied by the prior Pareto weight. This allocation will be sufficient to induce her to stay. Thus, unless an outside option constraint is binding, there is no need for a spouse to agree to any proposed renegotiation of the Pareto weight.

This discussion should have highlighted that intrahousehold allocations are sensitive to changes in outside options that cause incentive compatibility constraints to bind. This causes the no-commitment model's predictions to diverge from those of the full efficiency model. In the full efficiency model, conditions across the whole lifetime of the household influence the allocation agreed upon in any given period. However, under the no-commitment model the household only considers the future when deciding upon a period's allocation. In the honeymoon period after marriage, the full efficiency and no-commitment models yield similar consumption patterns since most of the lifetime is in the future. However, as time passes, allocations are more likely to diverge because there is less future marital surplus to lose in the event of disagreement and divorce. Allocations thereby become more responsive to temporary changes in outside options. This suggests that household behaviour evolves overtime; couples initially act with consideration and kindness towards one another but as time progresses they find themselves inconsiderately "behaving like an old married couple".

Innovations in the Pareto weight can also be driven by an interaction between discount rate heterogeneity and incentive compatibility. Consider a couple who are identical in every respect except for their patience, $\beta_A < \beta_B$. The optimal lifecycle plan would embody a more present-weighted consumption profile for A than B. Therefore, in early periods, A would receive a greater relative share of per-period expenditure, and the opposite in later periods. However, without a commitment mechanism, this plan may be infeasible. In some period, as her per-period resource share drops, A could conceivably do better by quitting the household, especially given the lower weight she attaches to future marital surpluses. The Pareto weight will then be renegotiated to re-emphasise A's preferences in the household allocation problem to prevent her from dissolving the household.

revealed preference conditions Rationalisation of a data set S by the no-commitment model is equivalent to the following condition.

definition 3.4 The set of observations S can be rationalised by the no-commitment model if there exist, for all $t \in T$, private quantities $\mathbf{q}_t^A, \mathbf{q}_t^B \in \mathbb{R}_+^N$ (with $\mathbf{q}_t^A + \mathbf{q}_t^B = \mathbf{q}_t$) and, in addition, concave, strictly increasing felicity functions $u^A(\mathbf{q}_t^A, \mathbf{Q}_t)$ and $u^B(\mathbf{q}_t^B, \mathbf{Q}_t)$, discount factors $\beta_A, \beta_B \in (0, 1]$ and Pareto weights $\omega_t^A, \omega_t^B > 0$, multipliers $\phi_t^A, \phi_t^B > 0$ and outside utilities $\bar{u}_t^A$ and $\bar{u}_t^B$ such that:

$$\begin{aligned} &\textstyle\sum_{t=1}^T \sum_{m\in\{A,B\}} (\omega_t^m \beta_m^{t-1} u^m(\mathbf{q}_t^m, \mathbf{Q}_t) - \phi_t^m \bar{u}_t^m) \\ &\geqslant \\ &\textstyle\sum_{t=1}^T \sum_{m\in\{A,B\}} (\omega_t^m \beta_m^{t-1} u^m(\zeta_t^m, \zeta_t^H) - \phi_t^m \bar{u}_t^m) \end{aligned} \tag{3.D4.1}$$

for all affordable consumption plans $\{\zeta_t^A, \zeta_t^B, \zeta_t^H\}_{t=1,\dots,T}$ that satisfy:

$$\sum_{t=1}^T \mathbf{p}_t' \mathbf{q}_t + \mathbf{P}_t' \mathbf{Q}_t = \sum_{t=1}^T \mathbf{p}_t' (\zeta_t^A + \zeta_t^B) + \mathbf{P}_t' \zeta_t^H \tag{3.D4.2}$$

The no-commitment model implies the existence of a set of mutually exclusive subsets within there is no renegotiation and the same Pareto weight is applied. Periods in which the incentive compatibility constraints bind partition these subsets.[4]

To introduce the potential for renegotiation into our revealed preference set-up, consider a partition of the set T into Υ mutually exclusive subsets T_τ of the form:

$$\mathbb{T} = \{T_1, ..., T_\Upsilon\}$$

with

$$T = \bigcup_{\tau=1}^{\Upsilon} T_\tau \text{ and } T_{\tau_s} \bigcap T_{\tau_t} = \emptyset \text{ if } \tau_t \neq \tau_s$$

such that

$$\tau_1 < \tau_2 \text{ implies } t_1 < t_2 \text{ for all } t_1 \in T_{\tau_1} \text{ and } t_2 \in T_{\tau_2}$$

Each subset, T_τ, represents a distinct Pareto weight regime within which $\omega_s^m = \omega_t^m$ for all $s, t \in T_\tau$. Let the Pareto weight in subset T_τ be denoted ω_τ^m. We then have the following testability result.

THEOREM 3.3 The set of observations S can be rationalised by the no-commitment model if and only if there exist, for all $t \in T$, private quantities $\mathbf{q}_t^A, \mathbf{q}_t^B \in \mathbb{R}_+^N$, Lindahl prices $\mathbf{P}_t^A, \mathbf{P}_t^B \in \mathbb{R}_+^K$, utility numbers $u_t^A, u_t^B \in \mathbb{R}$, and for all $\tau = \{1, ..., \Upsilon\}$, discounted marginal utilities of income $\lambda_\tau^A, \lambda_\tau^B \in \mathbf{R}_+$ and discount factors $\beta_A, \beta_B \in (0, 1]$ such that, for any $s \in T$ and $t \in T_\tau(\tau = \{1, ..., \Upsilon\})$, the following inequalities are satisfied:

$$u_s^A - u_t^A \leqslant \frac{\lambda_\tau^A}{\beta_A^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^A - \mathbf{q}_t^A) + \mathbf{P}_t^{A'}(\mathbf{Q}_s - \mathbf{Q}_t)]. \quad (3.T3.1)$$

$$u_s^B - u_t^B \leqslant \frac{\lambda_\tau^B}{\beta_B^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^B - \mathbf{q}_t^B) + \mathbf{P}_t^{B'}(\mathbf{Q}_s - \mathbf{Q}_t)]. \quad (3.T3.2)$$

and if $\lambda_{\tau_1}^m \neq \lambda_{\tau_2}^m$ for $m \in \{A, B\}$, then

$$\frac{\lambda_{\tau_1}^A}{\lambda_{\tau_1}^B} \neq \frac{\lambda_{\tau_2}^A}{\lambda_{\tau_2}^B} \quad (3.T3.3)$$

with

$$\begin{aligned} \mathbf{q}_t^A + \mathbf{q}_t^B &= \mathbf{q}_t \\ \mathbf{P}_t^A + \mathbf{P}_t^B &= \mathbf{P}_t \end{aligned} \quad (3.T3.4)$$

4 For rich enough data sets we can define these subsets by using information on outside options. In the data set used in this essay, we do not have such information and, therefore, we consider all possible partitions of T.

PROOF See Appendix B.

For sub-periods within which the same Pareto weight prevails, the full efficiency and no-commitment models imply the same pattern of behaviour.[5] Within these sub-periods individual marginal utilities of discounted income, λ_τ^A and λ_τ^B, are constant and choice behaviour will satisfy the inequalities of Theorem 3.2. The rest of this section discusses how Theorem 3.3 relates to Theorems 3.1 and 3.2 and develops an operational necessary test of the inequalities defined by Theorem 3.3.

HOMOGENEOUS DISCOUNT FACTORS With homogeneous intra-household discount factors, Pareto weight renegotiation represents the only mechanism in our theoretical framework by which time inconsistency can be generated in the revealed household preference. With $\beta_A = \beta_B = \beta$, 3.T3.1 and 3.T3.2 then reduce to the requirement that observed prices and quantities satisfy the inequalities associated with Theorem 3.1 within sub-periods with the same Pareto weight. In our empirical application, we will explore the impact of restricting discount factor heterogeneity within the no-commitment model such that the relative significance of preference heterogeneity and renegotiation in rationalising observed behaviour can be explored.

Assuming that all such sub-periods are non-degenerate (i.e. innovations are limited to occur no more frequently than every other period) then this result is identical to that of a "unitary" lifecycle model that allows for occasional revisions to the marginal utility of discounted wealth over time.[6] These revisions could occur due to failures of perfect foresight or perfect capital markets as discussed by Browning (1989). Browning found that the unitary lifecycle model was inconsistent with postwar aggregate data from the UK, US and Canada but that the rationalisation conditions did hold for surprisingly long stretches of time. This is consistent with a lifecycle model which allows for sporadic shocks that require revisions to permanent quantities. For example, Browning suggested that the lifecycle model combined with unanticipated inflation shocks in the early and late 1970s could explain the patterns uncovered in the UK data.

λ INNOVATIONS Innovations in λ_τ^A and λ_τ^B occur when a family member's incentive compatibility constraint binds. At such points, Pareto weight renegotiation results in a fall in the marginal utility of discounted income for the household member with the binding constraint.

5 However, allocations will not generally be identical as a different sharing rule may be in operation to the one which would have been employed at marriage under the assumption of perfect commitment.

6 A unitary model is one that treats a collective as a single decision making unit.

To illustrate, consider periods $t \in \tau_1$ and $t+1 \in \tau_2$ and let the incentive compatibility constraint of member A bind at $t+1$. Given the definition of ω_t^A, the Pareto weight on member A's utility, this implies that $\omega_t^A < \omega_{t+1}^A$. Defining an individual's marginal utility of discounted income as $\lambda_t^m = \eta/\omega_t^m$, where η is the Lagrange multiplier on the household lifetime budget constraint, this implies that $\lambda_t^A \neq \lambda_{t+1}^A$.[7] The greater weight assigned to A's preferences has an effect on the ratio of the marginal utility of discounted income of household members.

$$\frac{\lambda_t^A}{\lambda_t^B} = \frac{\eta/\omega_t^A}{\eta/\omega_t^B} = \frac{\omega_t^B}{\omega_t^A} \neq \frac{\omega_t^B}{\omega_t^A + \phi_{t+1}^A} = \frac{\eta/\omega_{t+1}^A}{\eta/\omega_{t+1}^B} = \frac{\lambda_{t+1}^A}{\lambda_{t+1}^B} \quad (3.11)$$

Therefore, innovations in the Pareto weight alter the ratio of individual marginal utilities of discounted income. Hence the presence of condition 3.T3.3 in Theorem 3.3.

However, the marginal utility of discounted individual income, $\lambda_t^m = \eta/\omega_t^m$, is determined by two factors: the Pareto weight ω_t^m and the marginal utility of discounted household income, η. Thus, innovations in λ_τ^m can be driven by unexpected shocks to household wealth and borrowing constraints, which result in innovations to η, as well as by renegotiation, which is reflected by changes in ω_t^m. Yet, although the levels of λ_τ^m are altered by η innovations, the ratio $\lambda_\tau^A/\lambda_\tau^B$ is not.

To illustrate this point, consider the household optimisation problem under perfect commitment and borrowing constraints:

$$\max_{\{\mathbf{q}_t\}_{1,\dots,T}} \sum_{t=1}^{\infty} \left\{ \beta_A^{t-1} u^A(\mathbf{q}_t^A, \mathbf{Q}_t) + \omega \beta_B^{t-1} u^B(\mathbf{q}_t^B, \mathbf{Q}_t) \right\} \quad (3.12)$$

such that

$$\begin{aligned} \mathbf{p}_t' \mathbf{q}_t + \mathbf{P}_t' \mathbf{Q}_t + A_t &\leqslant (x_t^A + x_t^B) + (1+i_t) A_{t-1} \\ 0 &\leqslant A_t \end{aligned}$$

where assets and income are appropriately discounted.

The associated Lagrangian is:

$$L = \sum_{t=1}^{\infty} \left\{ \begin{array}{l} \beta_A^{t-1} u^A(\mathbf{q}_t^A, \mathbf{Q}_t) + \omega \beta_B^{t-1} u^B(\mathbf{q}_t^B, \mathbf{Q}_t) \\ -\eta_t \left[\mathbf{p}_t' \mathbf{q}_t + \mathbf{P}_t' \mathbf{Q}_t + A_t - (y_t^A + y_t^B) - (1+i_t) A_{t-1} \right] \\ +\mu_t A_t \end{array} \right\} \quad (3.13)$$

7 This terminology is put to use again in the proof of Theorem 3.3

with Kuhn-Tucker conditions

$$
\begin{aligned}
\beta_A^{t-1}\nabla_{\mathbf{q}_t^A} u^A(\mathbf{q}_t^A, \mathbf{Q}_t) &= \eta_t \mathbf{p}_t \\
\omega\beta_B^{t-1}\nabla_{\mathbf{q}_t^B} u^B(\mathbf{q}_t^B, \mathbf{Q}_t) &= \eta_t \mathbf{p}_t \\
\beta_A^{t-1}\nabla_{\mathbf{Q}_t} u^A(\mathbf{q}_t^A, \mathbf{Q}_t) + \omega\beta_B^{t-1}\nabla_{\mathbf{Q}_t^c} u^B(\mathbf{q}_t^B, \mathbf{Q}_t) &= \eta_t \mathbf{P}_t \\
\eta_{t+1} + \omega_t &= \eta_t \\
\eta_t &\geqslant 0 \\
\mu_t &\geqslant 0
\end{aligned}
\tag{3.14}
$$

The assumption of strictly increasing utility functions implies that the household budget constraint always binds and thus $\eta_t > 0$ for $t = \{1, \ldots, T\}$. The multiplier μ_t is strictly positive in periods where the borrowing constraint strictly binds. The marginal utility of individual discounted income, λ_t^m, is then defined as:

$$
\begin{aligned}
\lambda_t^A &= \eta_{t+1} + \mu_t \\
\lambda_t^B &= \frac{\eta_{t+1} + \mu_t}{\omega}
\end{aligned}
\tag{3.15}
$$

Therefore, positive innovations to both λ_t^A and λ_t^B occur in periods when the household borrowing constraint binds, independent of any consideration of renegotiations in the household choice rule.

This discussion highlights that innovations in λ_τ^A and λ_τ^B can occur without renegotiation if the background assumption of perfect capital markets fail. A similar analysis can be made concerning the impact of violations of perfect foresight. Although both renegotiation and credit constraints result in λ_t^m innovations, the ratio of marginal utilities of individual discounted income is invariant to binding borrowing constraints. Let a borrowing constraint become binding at $t+1$, then:

$$
\frac{\lambda_t^A}{\lambda_t^B} = \frac{\eta_t}{\eta_t/\omega} = \omega = \frac{\eta_{t+1} + \mu_t}{(\eta_{t+1} + \mu_t)/\omega} = \frac{\lambda_{t+1}^A}{\lambda_{t+1}^B} \tag{3.16}
$$

Therefore, in theory, there is a way to nonparametrically distinguish innovations to permanent quantities arising from Pareto weight renegotiation to those arising from the failure of perfect capital markets; renegotiation alters the ratio of marginal utilities of income, whilst credit constraints leave this ratio unchanged.[8]

OPERATIONALISING Bringing Theorem 3.3 to the data is complicated as its constituent inequalities are nonlinear in unknowns; λ_τ^m and unobserved personalised prices and quantities interact. Therefore, we implement a testing strategy that is necessary, but not sufficient, for rationalisation by the no-commitment model.

8 Appendix B contains a further discussion on the incorporation of binding borrowing constraints on the revealed preference conditions for the time consistent model. These were used in order to examine the sensitivity of our empirical results for the time consistent model to the violaton of underlying theoretical assumptions.

NECESSARY NO-COMMITMENT TEST A necessary condition for the set of observations S to be rationalised by the no-commitment model is that there exist, for all $t \in T$, private quantities $\mathbf{q}_t^A, \mathbf{q}_t^B \in \mathbb{R}_+^N$, Lindahl prices $\mathbf{P}_t^A, \mathbf{P}_t^B \in \mathbb{R}_+^K$, utility numbers $u_t^A, u_t^B \in \mathbb{R}$ and discount factors $\beta_A, \beta_B \in (0, 1]$ that, for any $s, t \in T_\tau(\tau = \{1, \dots, \Upsilon\})$, satisfy:

$$u_s^A - u_t^A \leqslant \frac{1}{\beta_A^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^A - \mathbf{q}_t^A) + \mathbf{P}_t^{A\prime}(\mathbf{Q}_s - \mathbf{Q}_t)]. \quad (3.N3.1)$$

$$u_s^B - u_t^B \leqslant \frac{1}{\beta_B^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^B - \mathbf{q}_t^B) + \mathbf{P}_t^{B\prime}(\mathbf{Q}_s - \mathbf{Q}_t)]. \quad (3.N3.2)$$

with

$$\begin{aligned} \mathbf{q}_t^A + \mathbf{q}_t^B &= \mathbf{q}_t \\ \mathbf{P}_t^A + \mathbf{P}_t^B &= \mathbf{P}_t \end{aligned} \quad (3.N3.3)$$

PROOF See Appendix B.

Implementing the above test amounts to searching for a partition of the set T into subperiods within which the full efficiency conditions are satisfied. This test is not sufficient for rationalisation by the no-commitment model because it does not impose any restrictions on the ratio of marginal utilities of individual discounted income. Subset partitions could, therefore, reflect binding borrowing constraints or shocks to permanent income rather than renegotiation points.

This discussion should make it clear that the separate identification of renegotiation partitions and borrowing constraints is beyond the reach of the nonparametric demand theorist. The nonlinearity of the revealed preference constraints, and the fact that the set of rationalising λ_τ^m is nondegenerate, precludes the point identification of λ_τ^m at $\tau = 1, \dots, \Upsilon$ for $m = \{A, B\}$ and, therefore, testing for changes in the ratio of $\lambda_\tau^A / \lambda_\tau^B$. Limited progress can be made if one has access to information on an assignable good for each household member that is additively separable from non-assignable quantities in individual utility functions. In this instance, the revealed preference inequalities are linear in unknowns for the assignable quantities. However, rationalising $\{\lambda_\tau^m\}_{\tau=1,\dots,\Upsilon}$ are again only set identified and thus changes in ratios' are not uniquely defined.

SUMMARY This section has used Mazzocco's (2007) theoretical framework to develop an empirical strategy for distinguishing between the different sources of time inconsistent household choice. Theorem 3.2 is equivalent to rationalisation by the full efficiency model, under which intrahousehold discounting heterogeneity is the only source of time inconsistency in the revealed household preference. However, if there exists an empty solution set to the inequalities defined by Theorem 3.2, an appeal to more than individual heterogeneity is required and renegotiations of the Pareto weight must also be considered. The revealed preference conditions of the "no-commitment" model, which are defined by Theorem 3.3, admit the potential for renegotiation but are non-linear in unknowns. Therefore, a necessary but not sufficient test of renegotiation was developed in which one determines the existence of a partitioning of the set T into subsets, within which the inequalities defined by Theorem 3.2 hold. The next section continues our empirical analysis to apply these tests and determine whether a recognition of the collective nature of choice can rationalise the revealed time inconsistencies in household preferences.

18

RATIONALISING OBSERVED TIME INCONSISTENCY

In this section, the empirical application to the ECPF data is resumed. We test for the existence of a non-empty feasible set to the systems of inequalities defined by Theorem 3.2 and the necessary test of Theorem 3.3. Simply allowing for limited intrahousehold heterogeneity in the discount rate enables the behaviour of 98.4% of families to be rationalised without recourse to irrationality at the individual level. Given this positive result, a detailed investigation is conducted into the theory-consistent differences in spousal discount rates.

Although the vast majority of household behaviour can be explained without any mention of intrahousehold renegotiation, results are presented for a strengthened version Theorem 3.3, which imposes equal discount factors for the individual household members A and B, i.e. $\beta_A = \beta_B$. This facilitates an assessment of the relative importance of time preference heterogeneity and renegotiation in generating the observed time inconsistencies in revealed household preferences. Our results suggest that discount rate heterogeneity is the more relevant channel for explaining the patterns of household choice that are observed in our particular short panel data set.

The testing procedure is slightly modified from that of Section 16 to account for the collective nature of choice. The revealed preference tests are operationalised using a two-dimensional grid search over individual discount factors on $[0.9, 1]^2$, with a spacing of 0.005. To test the strengthened version of Theorem 3.3, we consider alternative scenarios that are defined by the maximum number of renegotiations that are permissible in the two-year period over which household is observed. This maximum can range from 0 (i.e. time consistent behaviour) to 7 (i.e. the Pareto weight changes in each different consumption quarter).

18.1 TEST RESULTS

Let us first consider the results for the full efficiency collective model, which allows for β-heterogeneity within the household but imposes a single choice rule for the period of consideration. These results are summarised in Table 22. The results associated with the full efficiency model stand in stark contrast to those reported in Section 16 for the time consistent model. The behaviour of the overwhelming majority of households can be explained using the framework of this extremely simple intertemporal collective model, under which the only source of time inconsistency in the household revealed preference is variation in the time preferences of family members. No further recourse to nonstationarities at the individual level is required. This is a surprising result given the strong assumptions of perfect foresight, perfect commitment and perfect capital markets that the theoretical framework incorporates. However, the hypothesis that these assumptions are valid in the short term, or at least for two years, cannot be rejected. For 98.4% of households a well-behaved felicity function and a constant discount rate that provide a perfect within-sample rationalisation of choice behaviour can be found for each family member.[1]

Table 22: Individual Heterogeneity

	Full Efficiency and β-heterogeneity
Pass Rate (%)	98.6

We next consider the pass rate associated with the strengthened version of Theorem 3.3 that admits renegotiation but assumes β-homogeneity. Unsurprisingly, allowing for more frequent renegotiations of the Pareto weight is associated with an increase in the pass rate.[2] For the extreme scenario that allows innovations in the Pareto weight between any two consecutive periods the pass rate amounts to 92.74%. However, as soon as a stable Pareto weight is required over two periods or more, the pass rate drops quite dramatically.

Table 23: Renegotiation Pass Rates

Breaks (max.)	All	Stable	Unstable
0	0.13	0.16	0.00
1	1.07	1.18	0.64
2	7.07	7.08	7.03
3	22.33	22.09	23.03
4	50.28	50.31	50.16
5	74.83	75.39	75.52
6	89.27	89.23	89.46
7	92.74	93.00	91.69
		χ^2	6.69

1 Here it is worth remarking that Mazzocco (2007) rejected the full efficiency model in his empirical application. The findings here suggest that this rejection could be the result of biases introduced by misspecification, omitted relevant distribution factors or the synthetic nature of the panel used, rather than a failure of commitment itself.

2 This is unsurprising since increasing the number of (possible) renegotiations obtains less stringent rationalisability conditions.

The sample of households includes a small minority whose observed characteristics (housing tenure and job classification) changed over the period considered (see Table 24). It is plausible that these changes would cause an innovation in the bargaining rule. To ensure that our conclusions are not unduly influenced by households with unstable characteristics, we compare the number of innovations in the Pareto weight that are required to rationalise behaviour of "stable" and "unstable" households (i.e. at least one characteristic changes whilst the household is under observation). Table 23 decomposes the number of Pareto weight innovations by whether a household had stable or unstable characteristics. A Likelihood Ratio Chi-Squared test for equality of categorical distribution confirms that these distributions are not statistically significantly different from one another, returning a p-value of 0.57.

Table 24: Characteristic change

	Proportion	Standard Dev.
Change in Job Title:	0.188	0.383
Change in House Tenure:	0.028	0.162
Some Change:	0.198	0.398

IMPLICATIONS Accounting for the collective nature of household choice allows the intertemporal behaviour of families in our sample to be explained using simple models that assume constant discounting at the individual level. The results in this essay provide particularly strong empirical support for a model that locates the primary source of time inconsistent family behaviour with intrahousehold β-heterogeneity. The full-efficiency model seems plausible given the short time span of our sample.

This finding is important and at odds with much of the prior empirical work in this area. Frederick, Loewenstein and O'Donoghue (2002) provide an extensive survey of empirical results in this area that suggest a systematic violation of time consistent preferences. In the collective context, Mazzocco (2007) rejected the full efficiency model using longitudinal data and estimated consumption Euler equations.

The results in this essay suggest a re-examination of this evidence. Past studies have universally employed restrictive parametric specifications and much of the work on time consistent preferences has occurred in an experimental setting. Dohman *et al.* (2012) highlight that elicited preferences are not procedurally independent and that discount rate estimates are hugely sensitive to the experimental design. In contrast, the analysis in this essay has employed a nonparametric testing methodology and utilised survey data. These results suggest that, so long as analysis is carried out at the appropriate locus of decision making, the assumption of time consistent preferences and perfect commitment over the short to medium run are appropriate positivist modelling devices for nondurable consumption choices.

18.2 DISCRIMINATORY POWER AND PREDICTIVE SUCCESS

Thus far, the pass rate has been our sole metric by which to evaluate the empirical performance of a model. However, alongside the pass rate, empirical applications of revealed preference analysis often consider two additional performance metrics: discriminatory power and predictive success. In this section, these metrics are computed for the three behavioural models under study.[3] Predictive success gives a holistic measure of the empirical performance of a behavioural model by simultaneously accounting for the pass rate and discriminatory power. As such, it is particularly interesting to compare models according to this metric. This comparison provides further empirical support for the full efficiency model as a framework for explaining the choice behaviour of households in our sample.

DISCRIMINATORY POWER. Following Bronars (1987), the discriminatory power of a revealed preference test is defined as the probability of detecting behaviour that is not rationalisable by the model. Bronars suggests an iterative procedure to compute this power metric. This procedure is applied to each household in the sample. At every iteration, the procedure simulates random behaviour by drawing $|T| * (N + K)$ uniform random budget shares. For a given household, these budget shares define a new random consumption stream $\{\mathbf{q}_t^R, \mathbf{Q}_t^R\}_{t \in T}$ that exhausts total wealth.[4] One then tests for a nonempty feasible set to the revealed preference conditions of interest on the choice set $\{\mathbf{q}_t^R, \mathbf{Q}_t^R; \mathbf{p}_t, \mathbf{P}_t\}_{t \in T}$. This procedure is iterated 1000 times per household to calculate the proportion of the randomly generated consumption streams that fail the revealed preference restrictions of the behavioural model. This proportion proxies the true probability that random household behaviour will fail the restrictions of the behavioural model for observed prices and total household expenditure. For example, if 50% of all randomly generated consumption streams fail to meet the requirements of a revealed preference test, then there is an approximately 50% chance that our tests will correctly reject random choice behaviour. Generally, high power signals a restrictive model and there is a correspondingly high probability that revealed preference tests will be violated by random behaviour.

3 For the collective model that allows for renegotiation, the following results all pertain to the specification that allows innovations in the Pareto weight between any two consecutive periods, of which the pass rate amounts to 92.75% (see Table 23).

4 This simulated random behaviour corresponds to Becker's (1962) notion of irrational behaviour as behaviour that randomly exhausts the available budget.

GRID SPECIFICATION At this point, recall that a grid search on the discount factor was conducted in order to compute pass rates for the different models. Computing power requires an analogous grid search at each iteration.[5] A household specific grid size is defined conditional on whether the observed household choices are rationalisable by the model under study. For the collective models, if observed behaviour is not rationalisable, then at each iteration the same grid size is used as before: the interval $[0.9,1]^2$ with a spacing of 0.005. However, if observed behaviour can be rationalised by the model in question, this information is used to define a finer grid. Specifically, search occurs only over (β_A, β_B) for which the difference $(\beta_A - \beta_B)$ is not greater than the minimum difference under which the observed behaviour is rationalisable. This adjusted grid search substantially limits the computational burden of our power assessment, whilst accounting for the information on individual time preferences that are revealed by observed behaviour. One proceeds similarly for the time consistent model. For nonrationalisable behaviour, β in the interval [0.9,1] are considered, with a spacing of 0.005, and, for nonrationalisable behaviour, the finer grid contains all β no lower than the maximum value under which rationalisability is obtained.

RESULTS Figure 23 shows the kernel distribution of discriminatory power for the three models under study, and Table 25 gives corresponding summary statistics. Some notable features emerge. First, the time consistent model has very high discriminatory power. For over half of the households, discriminatory power is approximately 100%, and the average power is no less than 99.97%. This is unsurprising given the extremely low pass rate, 0.13%, that we obtained before. In contrast, the collective model that allows for renegotiation but assumes homogenous discount rates has generally low power. The average power of the test in our sample is only 8.99%, and the maximal power amounts to no more than 13.30%. Finally, the full efficiency model occupies an intermediate position. The average power of the test in our sample is 47.73%. This is high compared to other revealed preference tests of the collective model (see Cherchye, De Rock and Vermuelen, 2009). There is also significant heterogeneity in discriminatory power across households: minimum power equals 0% and maximum power, 100%. Thus, the full efficiency model performs relatively well according to the metric of discriminatory power.

PREDICTIVE SUCCESS. We also compute a predictive success metric by which to compare the different models under study. This metric was recently axiomatised by Beatty and Crawford (2011) and is based upon an original proposal of Selten (1991). It combines the pass rate

5 β for the time consistent model and (β_A, β_B) for the collective models.

Table 25: Power Summary Statistics

	Mean	Min	Q1	Median	Q3	Max
Time Consistent	99.97	99.60	99.90	100.0	100.0	100.0
β-heterogeneity	47.73	0.00	22.00	48.70	72.02	100.0
β-homogeneity	8.99	3.80	7.60	9.40	10.40	13.30

and the power of a particular behavioural model into a single metric. For each household, the measure of predictive success is found by subtracting the "relative area", a, from the pass rate.[6]

$$m = r - a$$

The predictive success measure can be interpreted as a power-adjusted pass rate. The measure is always situated between -1 and 1, $m \in [1, -1]$. The higher the average predictive success measure, the better the empirical performance of the behavioural model under evaluation. A predictive success value in the neighbourhood of -1 reflects the situation in which a household fails the rationalisability conditions, even though the power of the test is low and thus relatively easy to pass. Conversely, a predictive success value in the neighbourhood of 1 reflects a household who passes the model restrictions, in a situation where the model has high power. This represents the ideal scenario. Finally, a predictive success value of zero suggests that the model is not informative for the household at hand: the model does not outperform the uninformative assumption that households exhibit random consumption behaviour, for which the power is 0 and the pass measure equals 1, by construction.

Figure 24 and Table 26 summarise the predictive success results for the three models under study. These results suggest that both the time consistent model and the no-commitment model with homogenous discount factors are uninformative for our sample of households. The empirical performance of these models is in line with a random number generator, although for different reasons. There are few choice paths that satisfy the stringent conditions that are associated with the time consistent model. The model, therefore, rejects almost any behaviour, which effectively makes it a neutral predictor. By contrast, the no-commitment model with homogenous discount factors is not very restrictive for our data set and, therefore, suffers from low power. As a result, although the model achieves a high pass rate, it barely outperforms the time consistent model according to the metric of predictive success.

Both of these models are outperformed by the full efficiency model with heterogeneous discount factors. We can even establish a (par-

6 The relative area is constructed as 1 less the power measure. For a further discussion, see Section 11.2

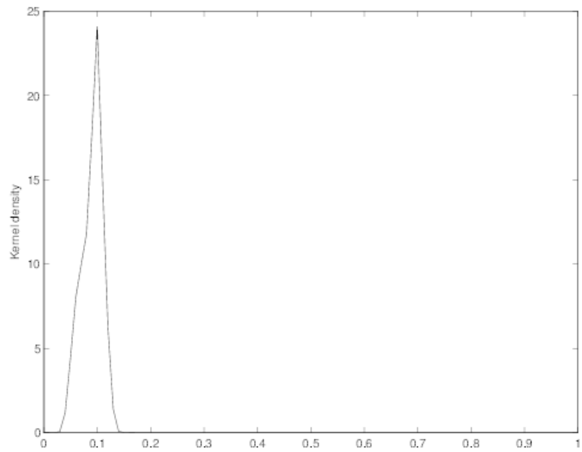
tial) stochastic dominance result in terms of predictive success. If the small set of households that are not rationalisable by the full efficiency model are disregarded, the predictive success distribution associated with the full efficiency model is situated to the right of the distributions pertaining to the other two models.[7] These results suggest that the full efficiency model is a particularly useful one for describing the short term dynamics of the consumption behaviour of the households studied in our sample.

Table 26: Predicitive Success Summary Statistics

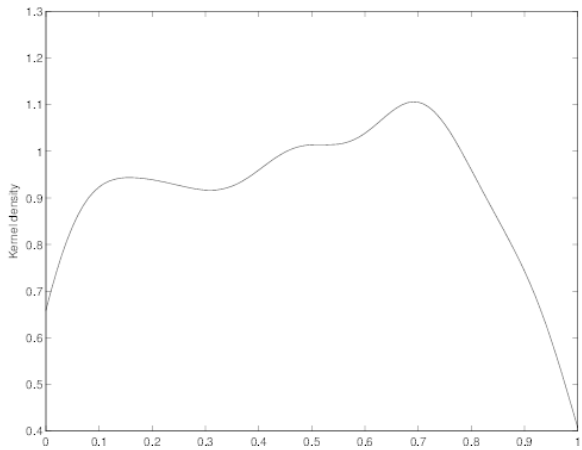
	Mean	Min	Q1	Median	Q3	Max
Time Consistent	0.09	-0.40	-0.01	00.0	00.0	00.0
β-heterogeneity	46.09	-100	22.00	48.70	72.02	100.0
β-homogeneity	8.99	3.80	7.60	9.40	10.40	13.30

7 Only 1.6% of the sample violate the conditions associated with the full efficiency model.

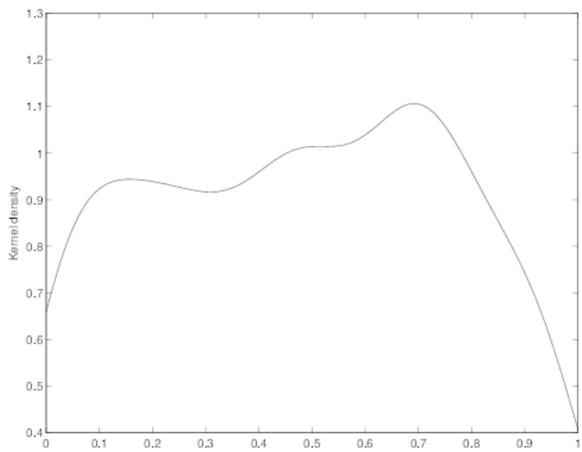
Figure 23: Kernel Plots of Power Distribution



(a) Time consistent

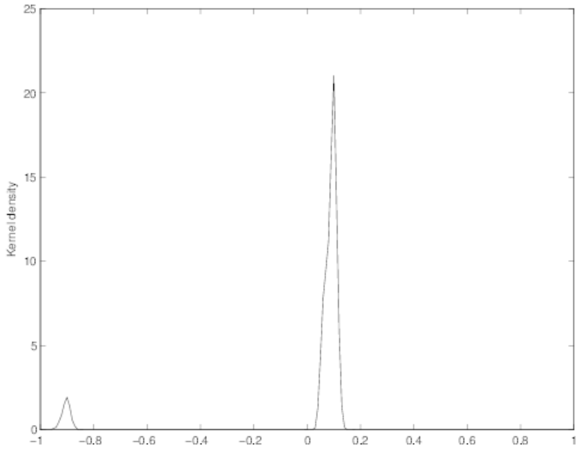


(b) β-Heterogeneity

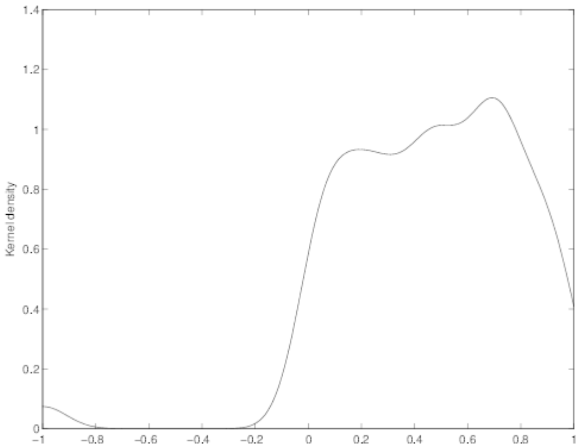


(c) Renegotiation

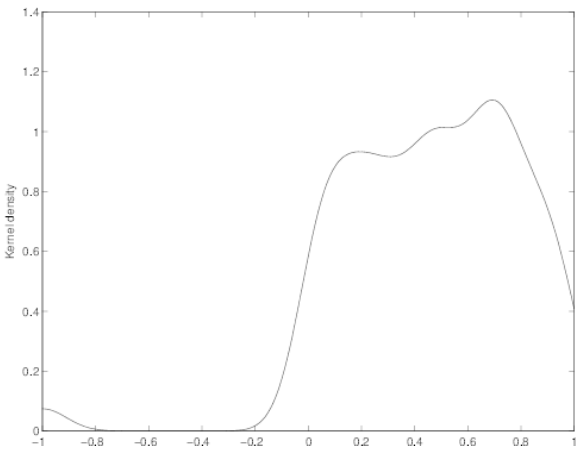
Figure 24: Kernel Plots of Predictive Success Distribution



(a) Time consistent



(b) β-Heterogeneity



(c) Renegotiation

Our analysis suggests that the full efficiency model performs well for the data at hand. The model is associated with a high pass rate and favourable predictive success measures. Discount rate heterogeneity is, therefore, important for rationalising patterns in household consumption behaviour. We now examine the characteristics of intrahousehold time preference heterogeneity in more detail and ask whether the necessary degree of unobservable preference heterogeneity correlates with observable family characteristics.

MINIMUM HETEROGENEITY For each household that can be rationalised by the full efficiency model, the discount rates that make observed consumption behaviour consistent with the rationalisability conditions in Theorem 2.2 are recovered. One drawback of using a revealed preference methodology is that the identification of individual discount rates is necessarily weakened by the lack of structure our framework imposes on individual preferences and the household choice problem. Our recovery problem is thus underdetermined and preferences are only set identified in the sense of Manski (2007). We refer to the set of potential time preferences as the set of theory consistent discount rates.

Given the multiplicity of theory-consistent discount rates, we report the *minimal* amount of discount rate heterogeneity that is necessary to rationalise a household's consumption stream. To determine the minimum difference $(\beta_A - \beta_B)$ in the discount rate grid, our testing procedure is iterated to consider nodes of the grid search in a specific order. Initially, we set $\beta_A - \beta_B = 0$, which corresponds to the time consistent model, and thus is rejected for all but two households. For the remaining households, whether behaviour can be rationalised for $\beta_A - \beta_B = 0.005$ is determined by testing for a non-empty set to the inequalities defined by Theorem 2 with $\beta_A \in [1, 0.995, ..., 0.905]$ and $\beta_B = \beta_A - 0.005$. For the remaining non-rationalisable households, the difference is then set at 0.01, and so on until the maximum difference is reached when $\beta_A = 1$ and $\beta_B = 0.90$.

It can be assumed that $\beta_A \geqslant \beta_B$ without loss of generality because our data set does not contain assignable goods and thus the Afriat inequalities for $m = \{A, B\}$ are fully symmetric.[8] Data limitations preclude a test for whether the husband or wife is the more patient household member but, if assignable information were available, our analysis could be easily extended to address gender-related questions.

8 Assignable goods are those whose consumption can be tied to a particular household member.

The distribution of the minimum theory-consistent time preference heterogeneity in our sample is summarised in Table 27, and shown with more detail in Figure 25.

Figure 25: Kernel Plot of Min. β-Heterogeneity Distribution

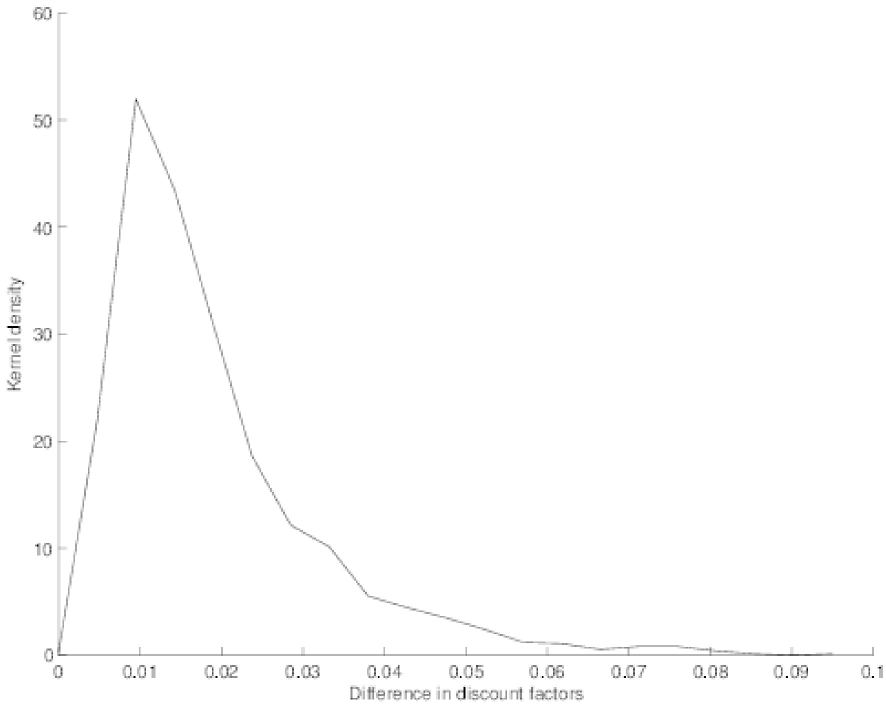


Table 27: β-Heterogeneity

	Mean	Min	Q1	Median	Q3	Max
$\min(\beta^\star_A - \beta^\star_B)$	0.020	0.000	0.010	0.015	0.025	0.095

Only limited heterogeneity is required to rationalise the behaviour of most households. Almost three quarters of the households that are consistent with Theorem 3.2 require a difference in discount factors of $\beta_A - \beta_B \leqslant 0.025$ to rationalise their behaviour. The kernel plot in Figure 25 confirms that the distribution of minimal heterogeneity shows a high and relatively narrow peak for limited amounts of within-household heterogeneity in time preferences. The maximum allowable heterogeneity level of 0.10 is not observed for any of the rationalisable households. We see this as lending plausibility to our empirical results.

appeal to observed heterogeneity. It is interesting to consider whether patterns in minimum unobservable time preference heterogeneity are related to observable household characteristics. We do this by a simple regression of the log of minimum discount rate distances on recorded household characteristics. The ECPF contains information on a number of household characteristics including: age of each spouse, schooling and occupation of the head, the number of children in the household and housing tenure.[9] Variables for total consumption expenditures and the average budget share of public goods as proxies for wealth and marital surpluses respectively are also included.

Table 28 shows the regression results. We find that the minimum intrahousehold heterogeneity in discount factors correlates with certain household characteristics.

INTERPRETATION First consider possible rationalisations of the inverse relationship between average spousal age and time preference heterogeneity. We hypothesise that a positive association between age and match quality generates this correlation. Match quality is closely aligned to similarity in time preferences in the literature. For example, Schaner (2012) classes couples as either well or badly matched on the basis of differences in their elicited discount rates. There are two alternative explanations for the positive association between the average age of a couple and match quality. First, the older people are, the more likely it is for them to have met and married someone similar to them. Second, only well-matched couples stay together in the long term. Before considering these explanations in more detail, the place of divorce in our current setting must be clarified. The empirical test in this essay assumes an invariant intrahousehold decision making rule for the two years that a family is observed. However, this is not to say that renegotiation and divorce cannot occur in the long run. Clearly, some couples do divorce in reality. However, the probability of divorce is declining in match quality (Becker *et al.*, 1977; Weiss and Willis, 1997). The higher the quality of a match, the more marital surplus is available to be shared by a couple and the less likely it is for marriage dissolution to dominate for either spouse in any time period.

Now, on the first proffered explanation for the association between match quality and average age, consider the search process leading to marriage. If this process is costly, individuals will accept imperfect matches even if it is known that differences in time preferences will create inefficiencies in the new household (see Burdett and Coles (1999) for a formal framework). Over time, if some search continues during marriage, individuals will continue to acquire new information on other potential matches. If a spouse meets a high enough

9 We created dummy variables for housing tenure and job position. The variable $high_{l}evel_{j}ob$ equals 1 if the household head is wage earner with a university degree, and 0 if he is a specialised worked without a university degree, or an unspecialised worker.

quality match, it can be worth dissolving an existing marriage to take up a new opportunity. The older someone is, the longer they will have been a participant in the marriage market and thus the more likely it is that they will have met someone similar to themselves and be classed as a member of a high quality match. In this way, costly search can create a negative association between age and discount rate heterogeneity.

Alternatively, on the second explanation, it is not unreasonable to assume that older couples have been married longer. Some model match quality as an experience good (Nelson, 1970; Jovanovic, 1979; Weiss and Willis, 1997; Chiappori and Weiss, 2006). These explanations assume that one cannot perfectly assess the suitability of a potential mate until marriage occurs and a match is experienced. Given that young couples have lower marriage experience, one can expect a greater variation in match quality (i.e. greater heterogeneity in discount rates) amongst this group, as these couples are still assessing the degree to which they are suited and not all poor matches will have been terminated. However, only well suited couples will remain married for a long period of time. This again creates a negative association between the age of a couple and discount rate heterogeneity.

The presence of children and higher total expenditure are also associated with smaller differences in time preferences. These facts can also be rationalised by appeals to match quality. Children raise the cost of divorce and so one would not want to bring children into a household unless they were sure that they are part of a high quality match. On the relationship between total expenditure and heterogeneity, more wealthy individuals are able to bear search costs for longer before entering into a match, which again raises the average quality of match that they can expect to make.

Another variable of interest is the variance in household spending. We find that higher variation is associated with more time preference heterogeneity. We do not see a direct behavioural interpretation of this phenomenon. In fact, this finding may be an artefact of our focus on the minimum β-heterogeneity that is needed for rationalisability by the full efficiency model. This model excludes the possibility of income changes causing renegotiation within the household. This may require a higher β-difference being required to obtain consistency with the conditions in Theorem 3.2 for households that face substantial income variation.

As a final note, we remark that our results suggest three further effects that are significant: a university level education, home ownership and high level job all have a positive relationship with intrahousehold β-heterogeneity. We could not lend these relationships with a collective interpretation. Follow-up research, potentially using richer data sets, could allow us to better explain these patterns.

Table 28: β-Heterogeneity and Characteristics

	Coefficient	Standard Error
Mean Age	-0.001***	0.002
Abs. Age Difference	0.006	0.006
University Degree	0.146*	0.070
High Schoo l	-0.054	0.038
Children Dummy	-0.111*	0.043
Log Expenditures	-0.322***	0.042
Var(Log Expend.)	2.010***	0.240
% Public Expend.	0.012	0.159
Home Owner	0.103*	0.042
High Level Job	0.184**	0.067
R-Squared	0.0925	
Observations	1557	
* = sig. at 5%		
** = sig. at 1%		
*** = sig. at 0.001%		

19 CONCLUSION

This essay has provided a revealed preference analysis of household intertemporal choice. We established early on that the standard DU framework could not be applied without modification to rationalise the observed patterns in family consumption; choices in our sample revealed time inconsistencies in household preferences. Adopting a collective perspective, the focus was on whether these inconsistencies could be attributed to individual heterogeneity and renegotiation within the collective unit rather than individual nonstationarities. Equivalence conditions, defined according to a finite number of observations on household choice behaviour, were derived for a set of models of intertemporal collective choice. These conditions are easily operationalisable using standard linear programming techniques and allowed us to establish in a finite number of steps whether choices were rationalisable within our theoretical framework. We further developed our testability results into a practical algorithm that facilitates the recovery of the minimal intrahousehold discount rate heterogeneity that is necessary to rationalise choice behaviour.

An empirical application to a Spanish consumption panel highlights that an explicit recognition of the collective nature of choice allows one to rationalise the vast majority of time inconsistent household behaviour. Almost all observed household behaviour can be rationalised by a simple model that assumes perfect intrahousehold commitment and exponential discounting for the two years that a household is observed. Minimum differences in familial discount rates were recovered and were found to correlate with average spousal age, the presence of children and the level and variance of household expenditures.

Perhaps most importantly, this essay has provided empirical support for exponential discounting and perfect intrahousehold commitment as valid positivist modelling devices for short-to-medium run consumption choices. This result stands in marked contrast to much of the recent literature on time preferences and intrahousehold commitment. The methodology and empirical application presented in this essay has enabled us to avoid the misspecification and procedural biases that plague past studies, and has yielded a much more favourable empirical assessment of the traditional economic models of intertemporal choice. This essay signals the need for further studies that apply our nonparametric methodology to different data sets in order to probe the external validity of our findings. In the future, I also hope to apply these tests to richer household data sets that include information on assignable goods and additional observable characteristics, in order to yield more refined insights in the context of the preference recoverability exercise begun in this essay. Furthermore, for long enough panels with more detailed household information, the framework presented in this essay will enable an investigation of how variation in individual and household characteristics relates to patterns of intrahousehold renegotiation. I look forward to pursuing such studies in the future.

Part IV

REVEALED PREFERENCE CHANGE: RATIONALISING TOBACCO CONSUMPTION

We now consider a setting in which choice behaviour cannot be rationalised by the basic utility maximisation hypothesis. This essay addresses the questions of testability and data consistency for models that introduce taste instability into the basic rational choice framework. These results are then developed into a practical algorithm that can be applied to compute the lower bound on the intertemporal and interpersonal variation in marginal utility that is required to rationalise a finite set of observations on past choice behaviour.

In this essay, I apply this methodology to rationalise the patterns that have been observed in U.K. tobacco consumption since 1980. Psuedo cohorts are constructed from the U.K. Family Expenditure Survey and their demands are recovered along "Sequential Maximum Power" paths using censored quantile regression techniques. It is found that taste change is a necessary component of any rationalisation of observed trends. Our analysis also uncovers evidence of significant educational differences in the "effective" and "raw" taste trajectories for tobacco. These results give strong reason to believe that less educated groups have a greater taste for tobacco relative to their more highly educated peers, for all but the heaviest smoking groups.

20

INTRODUCTION

In this final essay, I develop a systematic method for incorporating taste variation into the traditional revealed preference framework and I examine the implications that this extension has upon the testability of the rational choice hypothesis. These insights are used to create a methodology that enables the recovery of the minimal intertemporal and interpersonal heterogeneity in tastes that is required to rationalise the observed choice patterns in tobacco since 1980.

This introductory section motivates our focus upon tobacco consumption and provides context for the theoretical results to come.

TRENDS IN TOBACCO CONSUMPTION Since the seminal British and American reports on smoking and health (Royal College of Physicians, 1962; U.S. Department of Health, Education and Welfare, 1964), there has been great interest in how best to curb tobacco consumption. Tobacco is the only legal good that kills when used as intended and is the primary cause of preventable morbidity and premature death in virtually all Western democracies. By 2030, epidemiological analyses suggest that tobacco will be responsible for 8 million deaths annually worldwide (World Health Organisation, 2012). This figure eclipses the death toll associated with any other cause of disease (Murray and Lopez, 1996). These facts have prompted a research agenda that aims to identify the determinants of tobacco consumption and to understand how these factors can be best manipulated in order to alleviate the devastating impact that smoking has on public health.

Governments have a limited number of levers with which to influence consumption patterns. The basic utility maximisation hypothesis posits a rational consumer selecting over affordable bundles in order to best advance their preferences. Variation in observed behaviour is then attributed to variation in the constraints under which choice occurs and preference heterogeneity. Government policy can target both of these levers to influence individual consumption patterns: taxation and quantity controls can be levied in order to alter the feasible budget set, whilst information and education programmes can be designed to manipulate preferences.

Taking the first of these policy groupings, heavy taxation has long formed a central pillar of government strategies to curb tobacco consumption. In the seventeenth century, King James I of England levied the first tax for health reasons on tobacco, given his belief that smoking was a "custome loathsome to the eye, hatefull to the Nose, harmefull to the braine, dangerous to the Lungs, and in the blacke stinking fume thereof, nearest resembling the horrible Stigian smoke of the pit that is bottomelesse" (King James, 1604).[1] Tobacco remains one of the most highly taxed goods in the UK, with tax accounting for over 75% of the retail cost of a typical packet of cigarettes. Regarding government information programmes, health education and explicit anti-smoking campaigns comprise a salient feature of public health policy and aim at influencing the tastes motivating choice. The UK government's 2011 Public Health White Paper, *Healthy Lives, Healthy People*, stresses the importance of directly engaging with smokers to raise their awareness of the risks associated with tobacco consumption and modern anti-smoking adverts are often noted for their graphic content.

Understanding the relative efficacy of these levers, and how their effect varies across different socioeconomic groups, is important for the design of health policy going forward. Although smoking rates have fallen significantly since the 1960s, the rate of reduction has recently slowed. Greater knowledge of the determinants of tobacco consumption will help to inform governments on how to regain lost ground. Furthermore, declines in smoking have not occurred evenly across socioeconomic groups. Smoking is the single biggest cause of inequalities in death rates between the richest and poorest UK communities (Marmot, 2010). This is a phenomenon that has developed since the 1950s, when smoking rates were uniformly high across all social groups. Governments need to be aware if there are systematic differences in the behavioural response to policy levers as this will influence both the aggregate effectiveness of a given policy and one's assessment of its normative worth.

Answering these policy related questions requires the separate identification of price and taste effects on choice behaviour. This research question is of interest more generally. Preference change is not a phenomenon restricted to tobacco; tastes vary over one's lifecycle and with new information about a commodity. Thus, the question of how to disentangle the separate influence of preference effects and price responses has relevance in many different contexts. Furthermore, this topic is of interest to applied theorists who engage in the nonparametric testing of models. Empirical tests of economic models often utilise panel data that spans multiple years, over which taste variation is likely to occur. Therefore, it is important to know if taste change can be incorporated into economic models in a manner that allows the underlying theory to remain falsifiable.

A REVEALED PREFERENCE METHODOLOGY This essay considers the extent to which changes in tobacco consumption have been driven by price changes or by taste changes, and whether the significance of

1 His opinion changed when tobacco became a valuable cash crop in his new Virginia colony.

these two channels varies across socioeconomic groups. A revealed preference methodology is developed in order to disentangle the effects of price and preference change on tobacco consumption. The approach follows in the tradition of Afriat (1967), Diewert (1973) and Varian (1982) and builds upon the more recent work of Blundell *et al.* (2003, 2008), Browning (1989) and Crawford (2010).

Rather than calculate the correlations between the time-series variation of the variables of interest, the necessary and sufficient conditions under which observed behaviour is consistent with a particular theory of taste change are derived. In so doing, a surprising result is proven. Without strong functional assumptions on the evolution of preferences and the separability of goods in the utility function, observational data sets can always be rationalised with additively separable taste change on a *single* good. This result implies that interpersonal and intertemporal heterogeneity over a K-dimensional demand system can be parsimoniously summarised by a univariate parameter in the direct utility function. Therefore, we are able to construct a unidimensional index of unobserved multidimensional heterogeneity in our particular data context.

RATIONALISING U.K. TRENDS The empirical application in this essay focuses upon the recovery of the *minimal* amount of preference evolution that is required to rationalise choice behaviour. An algorithm is developed to set identify choice-rationalising preferences, which, conditional upon a normalisation of the utility function, identifies the unique lower bound to intertemporal and interpersonal variation in the marginal utility of tobacco. This method yields a global ranking of individuals by their taste for tobacco and enables the reconstruction of the minimum intertemporal changes in the marginal willingness to pay for tobacco that are sufficient to rationalise choice behaviour. Using our method, we are able to separately identify the minimal changes in the parameters of the direct utility function, which we label "raw" taste changes, and the minimal changes in the marginal willingness to pay for tobacco that these raw taste shifts support, which we label "effective" taste changes.

Applying this method to U.K. Family Expenditure Survey data reveals significant differences in the evolution of effective tastes for tobacco across education groups and across quantiles of the tobacco distribution. We construct psuedo-cohorts that are stratified by age and education for the period 1980 to 2000. We further track intracohort heterogeneity in tobacco tastes by recovering taste parameters at different quantiles of the budget share for tobacco distribution.

To maximise the power of our nonparametric methodology at revealing taste variation, we recover preference parameters along Blundell, Browning and Crawford's (2003) "Sequential Maximum Power" (SMP) path. Intuitively, this method enables one to control for budget variation that could otherwise preclude the identification of taste changes. Operationalising the SMP methodology requires the estimation of Engel curves for tobacco. This is complicated in our setting given significant censoring on tobacco and the endogeneity of total expenditure. We tackle these issues by applying censored quantile regression techniques. Quantile models are equivariant to monotone transformations such as censoring. This allows us to avoid the strong parametric assumptions of traditional Tobit estimators. In addition, a control function approach is employed to tackle the endogeneity of total expenditure. To implement this strategy we adapt the algorithm of Chernozhukov and Hong (2002). This yields a practical and efficient estimation method that enables us to incorporate unrestricted preference heterogeneity across psuedo cohorts.

We find that taste change is required to rationalise the choice behaviour of all cohorts. However, changes in the effective taste for tobacco have been significantly greater amongst more highly educated cohorts for all but the heaviest smokers and heavy alcohol drinkers. In other words, less educated cohorts are willing to pay more for tobacco than their more highly educated peers. Furthermore, effective tastes for tobacco amongst less educated cohorts have changed less over time. We show how one can interpret these results as changes in the virtual price of tobacco, and thereby recast the impact of government information programmes as acting as an additional tax. From this perspective, a lower taste-tax has been levied on the tobacco consumption of less educated individuals since 1980.

However, the recovered changes in the effective taste for tobacco could derive from two sources: taste change and income effects. Controlling for income effects to consider only the trends in "raw" taste changes, which are embodied by variation in the parameters of the direct utility function, weakens the conclusions that we are able to make regarding educational differences in tastes from the perspective of our nonparametric framework. The recovered parameters of the direct utility function are suggestive of a greater taste for tobacco amongst low education groups but the differences are not statistically significant. Additional structure is required on the base utility function or the marginal utility of income trajectory for one to make stronger conclusions regarding the source of the recovered differences in effective taste trajectories. In the context of our primary theoretical framework, we are unable to rule out that less educated people would demand lower quantities of tobacco if they commanded greater income.

Rather than place additional functional form assumptions upon the base utility function, we examine the impact of situating taste change in a lifecycle framework. In this setting, the marginal utility of income trajectory is restricted. This model yields stronger conclusions regarding educational differences in the parameterisation of the direct utility function. Statistically significant differences in raw taste parameters are recovered from within this framework and highlight that less educated groups enjoy a greater global taste for tobacco. From the perspective afforded by this model, more highly educated individuals would consume less tobacco than their less educated peers even if restricted to the same total expenditure path. This finding highlights that additional assumptions regarding income effects serve to strengthen our conclusions regarding educational taste asymmetries for tobacco, rather than creating additional ambiguity in the interpretation of our results.

If there is a desire for the observed socioeconomic trends in tobacco consumption to converge, then our results suggest that government education programmes must be designed to be more salient amongst less educated groups. Even if the recovered differences in effective tastes can be largely rationalised by income effects, our empirical results still imply that that the taste parameters that index the direct utility function must change by *more* for low education groups if an anti-smoking campaign is to affect a uniform "taste tax" across education groups. Our policy conclusions are stronger from the context of the lifecycle framework. From the context of this model, both effective and raw taste changes have been much more salient amongst higher education groups. This essay calls for additional research into why this is the case.

OUTLINE This essay proceeds as follows. Sections 21 and 22 outline our theoretical framework and derive the necessary and sufficient conditions under which observed behaviour and our model of taste change are consistent. Section 23 develops a quadratic programming methodology that can be applied to uncover the minimal amount of interpersonally comparable taste variation that is necessary to rationalise choice behaviour. Section 24 introduces the data that is used for our empirical investigation and discusses the construction of quantity sequences for the psuedo cohorts that we draw from the UK Family Expenditure Survey using censored quantile regression methods. Section 25 applies our method to rationalise the changes in tobacco consumption occurring in the U.K. since 1980. Our conclusions are strengthened in Section 26, in which taste parameters are recovered from the perspective of a lifecycle model. Finally, Section 27 concludes our analysis and considers the implications of our findings for government anti-smoking policy moving forward.

21

TASTE RATIONALISATION

This essay considers the extent to which variation in tobacco consumption has been driven by price changes or by taste changes, and whether the significance of these two channels varies across socioeconomic groups. This section presents the theoretical framework that is used to address this question. A simple, yet flexible, model is outlined that allows global statements to be made concerning relative tastes over particular goods independently of knowledge of the composition of the consumption bundle.

21.1 THEORETICAL FRAMEWORK

With intertemporal additive separability of preferences over lifetime consumption bundles, consumer i's maximisation problem with a time-dependent utility function can be expressed as:

$$\max_{\mathbf{q}} u^i(\mathbf{q}, \boldsymbol{\alpha}_t^i) \tag{4.1}$$

subject to

$$\mathbf{p}'\mathbf{q} = x$$

where $\mathbf{q} \in \mathbb{R}_+^K$ denotes the demanded quantity bundle, $\mathbf{p} \in \mathbb{R}_{++}^K$ denotes the exogenous price vector faced by consumer i and x gives total expenditure. $\boldsymbol{\alpha}_t^i$ is a potentially infinite-dimensional parameter that indexes consumer i's tastes at time t. We impose the common assumption that $u^i(\mathbf{q}, \boldsymbol{\alpha}_t^i)$ is locally nonsatiated[1] and concave conditional upon α_t^i.

Imagine that we observe the choice behaviour of individual i at T budget regimes: $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$ for $i = 1, ..., N$. Consistency between their observed choice behaviour and our simple model of taste change is defined as follows.

DEFINITION 4.1 Consumer i's choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$, can be "taste rationalised" by a utility function $u^i(\mathbf{q}, \boldsymbol{\alpha}^i)$ and the temporal series of taste parameters $\{\boldsymbol{\alpha}_t^i\}_{t=1,...,T}$ if the following set of inequalities is satisfied:

$$u^i(\mathbf{q}, \boldsymbol{\alpha}_t^i) \leqslant u^i(\mathbf{q}_t^i, \boldsymbol{\alpha}_t^i) \tag{4.D1.1}$$

1 Note that, in our empirical application, this assumption still permits tobacco to be a "bad" so long as utility is strictly increasing in nondurables; a very mild assumption.

for all $\mathbf{q}$ such that

$$\mathbf{p}_t'\mathbf{q} \leqslant \mathbf{p}_t'\mathbf{q}_t^i \tag{4.D1.2}$$

In words, observed behaviour can be rationalised by our theoretical framework if an individual's choice at t yields weakly higher utility than all other feasible choices at t when evaluated with respect to their time t tastes.

Marginal utility perturbations represent a simple way to incorporate taste variation into the standard utility maximisation framework. For example, in the random utility literature, it is common to represent heterogeneity across consumers as a linear perturbation to a base utility function (McFadden and Fosgerau, 2012; Brown and Matzkin, 1998). Following in this tradition to characterise intertemporal taste variation yields the temporal series of utility functions:

$$u^i(\mathbf{q}, \boldsymbol{\alpha}_t^i) = v^i(\mathbf{q}) + \boldsymbol{\alpha}_t^{i\prime}\mathbf{q} \tag{4.2}$$

where $\boldsymbol{\alpha}_t^i \in \mathbb{R}^K$.

Under this specification, $\alpha_t^{i,k}$ can be interpreted as the shift in the marginal utility of good k at time t from that dictated by the base utility function. These perturbations manifest themselves in translations of the gradients of indifference curves. The effect of a given value of a taste shifter is not invariant to transformations of $v^i(\mathbf{q})$. Thus, further analysis must be conditioned upon a particular ordinal representation of base preferences. The rationalisation conditions with additive linear taste perturbations reduce to those given by Definition 4.2.[2]

DEFINITION 4.2 Consumer i's choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,\dots,T}$, can be taste rationalised by the base utility function $v^i(\mathbf{q})$ and a temporal series of additive linear perturbations to marginal utilities $\{\boldsymbol{\alpha}_t^i\}_{t=1,\dots,T}$ where $\boldsymbol{\alpha}_t^i \in \mathbb{R}^K$ if the following set of inequalities is satisfied:

$$v^i(\mathbf{q}^i) + \boldsymbol{\alpha}_t^{i\prime}\mathbf{q}^i \leqslant v^i(\mathbf{q}_t^i) + \boldsymbol{\alpha}_t^{i\prime}\mathbf{q}_t^i \tag{4.D2.1}$$

for all $\mathbf{q}^i$ such that:

$$\mathbf{p}_t'\mathbf{q}^i \leqslant \mathbf{p}_t'\mathbf{q}_t^i \tag{4.D2.2}$$

2 When one only considers the set of observed quantities, $\{\mathbf{q}_t^i\}_{t=1,\dots,T}$, this condition can be equivalently expressed as:

$$u^i(\mathbf{q}_s^i, \boldsymbol{\alpha}_s^i) \leqslant u^i(\mathbf{q}_t^i, \boldsymbol{\alpha}_t^i) + (\boldsymbol{\alpha}_s^i - \boldsymbol{\alpha}_t^i)'\mathbf{q}_s^i$$

This condition corresponds to McFadden and Fosgerau's (2012) "Generating Function Fundamental Inequality".

There are many circumstances when one may want to restrict taste instability to a subset of goods. For example, the primary concern in this essay is to investigate taste change with respect to a single good, tobacco. More generally, it is interesting to consider the extent to which the assumption of a static utility function must be relaxed in order to rationalise observed behaviour. An appeal to Occam's razor would favour a framework in which one incorporated the minimal amount of taste variation that is necessary to capture the salient features of the data was incorporated. In the instance where taste instability is restricted to a single good we have: $\boldsymbol{\alpha}_t^i = [\alpha_t^{1i}, 0, ..., 0]$. "Good-1 Taste Rationalisability" is then defined as follows.

DEFINITION 4.3 Consumer i's observed choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$, can be "good-1 taste rationalised" by a base utility function $v^i(\mathbf{q})$ and a series of taste shifters on good 1 $\{\alpha_t^{1i}\}_{t=1,...,T}$ if the following set of inequalities is satisfied:

$$v^i(\mathbf{q}^i) + \alpha_t^{1i} q^{1i} \leqslant v^i(\mathbf{q}_t^i) + \alpha_t^{1i} q_t^{1i} \qquad (4.D3.1)$$

for all $\mathbf{q}^i$ such that:

$$\mathbf{p}_t' \mathbf{q}^i \leqslant \mathbf{p}_t' \mathbf{q}_t^i \qquad (4.D3.2)$$

Given this essay's focus on taste change over tobacco, Definition 4.3 constitutes our primary rationalisation condition. Our theoretical framework is, therefore, relatively restrictive regarding the form of taste change. Firstly, it is assumed that tastes for all goods other than good-1 are stable. Furthermore, the linear additive specification for taste change guarantees that the marginal rate of substitution between any goods $j, k \in \{2, ..., K\}$ are invariant to taste instability on good-1.

21.2 VIRTUAL BUDGETS

Taste change can be characterised as variation in "virtual budgets", a concept first suggested by Robarth (1940) and applied by Neary and Roberts (1980) to develop the theory of choice behaviour under rationing. The maximisation problem of a consumer with time-varying tastes can be expressed in a static format but with their budget determined endogeneously by tastes at t. The endogenous virtual budget at t is that which induces the individual with preferences represented by the time-invariant base utility function $v^i(\mathbf{q})$ to demand the same bundle as they would with preferences $v^i(\mathbf{q}) + \alpha_t^{1i}$ when facing market prices $\mathbf{p}_t$ and total expenditure x.

The virtual budget characterisation of taste change flows straightforwardly from the first order conditions associated with the consumer's time t maximisation problem:[3]

$$\begin{bmatrix} \frac{\partial v^i(\mathbf{q})}{\partial q^1} + \alpha_t^{1i} \\ \frac{\partial v^i(\mathbf{q})}{\partial \mathbf{q}^{-1\prime}} \end{bmatrix} = \lambda_t^i \mathbf{p}_t \tag{4.3}$$

These conditions can be rearranged as:

$$\begin{aligned} \frac{\partial v^i(\mathbf{q})}{\partial \mathbf{q}'} &= \lambda_t^i \begin{bmatrix} p_t^1 - \alpha_t^{1i}/\lambda_t^i \\ \mathbf{p}_t^{-1} \end{bmatrix} \\ &= \lambda_t^i \boldsymbol{\pi}_t^i \end{aligned} \tag{4.4}$$

The virtual budget required to support the optimal demand at t given t tastes, $\mathbf{q}_t^{i\star}$, is then:

$$\begin{aligned} \boldsymbol{\pi}_t^{i\prime} \mathbf{q}_t^{i\star} &= \mathbf{p}' \mathbf{q}_t^{i\star} - \frac{\alpha_t^{1i}}{\lambda_t^i} q_t^{1i\star} \\ &= \widetilde{x}^i \end{aligned} \tag{4.5}$$

Therefore, the maximisation problem facing the consumer can be equivalently restated as one of an individual with static tastes facing the endogenous virtual budget $(\boldsymbol{\pi}_t^i, \widetilde{x}^i)$:

$$\max_{\mathbf{q}} v^i(\mathbf{q}) \text{ subject to } \boldsymbol{\pi}_t^{i\prime} \mathbf{q}_t = \widetilde{x}_t^i \tag{4.6}$$

The properties of demand functions with unstable tastes are related to those of rationed demand functions, which are derived in Neary and Roberts (1980). The above analysis is, in effect, concerned with the behaviour of an individual with preferences $v^i(\mathbf{q})$ who is forced to purchase the consumption bundle that their time t self optimally purchases; they are 'rationed' to consume exactly $\mathbf{q}_t^{i\star}$. However, unlike results concerning behaviour under rationing, the ration levels at which virtual budgets are evaluated are not determined exogenously. Rather, ration levels are the quantities chosen by the optimising agent with time t tastes. Thus, changes in prices and income also bring about changes in ration levels due to the behavioural response of the time t agent. These effects are absent from the pure theory of rationing.

The virtual price characterisation of taste change supports the interpretation of government information programmes as supplementary tax and incomes policies. Programmes designed to cultivate a negative taste for tobacco levy a "taste-tax" on this good because they result in a rise in the virtual price for tobacco: $\pi_t^{1i} > p_t^1$ given that a negative taste shock implies $\alpha_t^{1i}/\lambda_t^i < 0$. This is an intuitive and attractive way to interpret the impact of government education programmes that we will make use of when presenting our empirical results in Section 25.

3 At time t, individual i is characterised by the utility function: $v^i(\mathbf{q}) + \alpha_t^{1i}$.

Finally, the virtual budget characterisation of our problem leads to a natural distinction between an individual's "raw" and "effective" taste for good-1. One's *raw* tastes for good-1 are reflected in the magnitude of the α parameters that index the direct utility function. However, the behavioural interpretation of these parameters is not clear. Therefore, we will also comment upon an individual's *effective* tastes, which are captured by their marginal willingness to pay for good-1, π_t^{1i}.

21.3 SINGLE CROSSING

An attractive feature of our theoretical framework is that it facilitates an unambiguous ranking of an individual's tastes for tobacco. The specification imposed upon taste instability implies that $\alpha_t^{1i} > \alpha_{t+s}^{1i}$ can be unambiguously interpreted as "individual i had a greater taste for tobacco at time t than time $t+s$". This follows because preferences satisfy the "single-crossing condition" in $(\mathbf{q};\alpha^1)$ space.

DEFINITION 4.4 (MILGROM & SHANNON, 1994) A utility function $u^i(\mathbf{q},\alpha^1)$ satisfies the single crossing property in $(\mathbf{q};\alpha^1)$ if for $\mathbf{q}' > \mathbf{q}''$ and $\alpha^{1,\prime} > \alpha^{1,\prime\prime}$

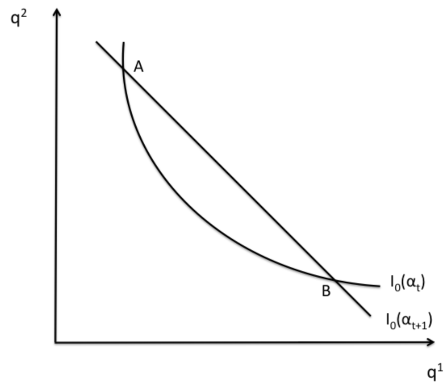
$$u^i(\mathbf{q}',\alpha^{1,\prime\prime}) \geqslant u^i(\mathbf{q}'',\alpha^{1,\prime\prime}) \text{ implies } u^i(\mathbf{q}',\alpha^{1,\prime}) \geqslant u^i(\mathbf{q}'',\alpha^{1,\prime}) \quad (4.D4.1)$$

This condition can be interpreted as stating that for $\mathbf{q}' > \mathbf{q}''$, the function $f(\alpha) = u(\mathbf{q}',\alpha) - u(\mathbf{q}'',\alpha)$ equals zero only at a single point.

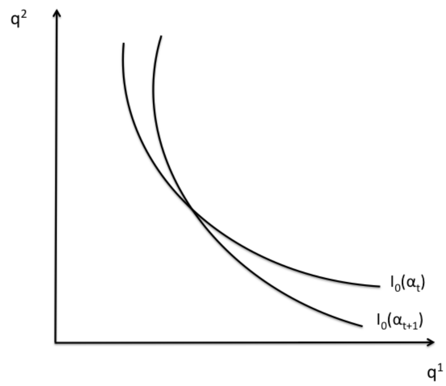
Global statements concerning relative tastes over good-1 can be made when preferences satisfy single crossing in $(\mathbf{q};\alpha^1)$ space. With single crossing, an individual's T temporal preferences can be ranked according to their taste for tobacco independently of the quantities of other goods consumed and total expenditure. To illustrate, consider Figure 26. Figure 26 depicts indifference curves associated with the different taste parameters α_t and α_{t+1}. One cannot make a global statement concerning the differences in relative taste for good-1 for the preferences depicted in panel (a). At point A, the individual with preferences indexed by α_t has a greater relative taste for good-1 than the individual with with preferences indexed by α_{t+1}. Yet, at point B, the opposite is true. However, the preferences depicted in panel (b) satisfy single-crossing and one can unambiguously conclude that the individual with tastes indexed by α_{t+1} has a greater taste for tobacco than the individual who has tastes indexed by α_t.

PROPOSITION 4.1 The individual utility function, $u^i(\mathbf{q},\alpha^{1i}) = v^i(\mathbf{q}) + \alpha^{1i}q^1$ satisfies single crossing in $(\mathbf{q},\alpha^1)$ space.

Figure 26: Single Crossing



(a) Violation of Single Crossing



(b) Single Crossing Preferences

PROOF. See Appendix C.

As our specification of the time t utility function, $u^i(\mathbf{q}, \alpha^{1i}) = v^i(\mathbf{q}) + \alpha^{1i} q^1$, satisfies single crossing, there exists a unique ranking of temporal tastes for good-1 by the magnitude of the α_t^{1i} parameters. If $\alpha_t^{1i} > \alpha_u^{1i} > \alpha_v^{1i}$, we can conclude that tastes for good-1 were greater at time t than they were at time u, and tastes at time u were greater than they were at v. This ranking is independent of the quantities of all other goods consumed and total expenditure.

21.4 INTERPERSONAL COMPARABILITY

Up until this point, our discussion has centred on formalising a theory of taste instability for a single individual i. However, in our empirical application, we want to go beyond the identification and interpretation of these minimum individual taste shifters to compare the magnitude of taste shifters across individuals. This requires the ability to meaningfully compare the values of α_t^{1i} and α_s^{1j} for any $i, j = \{1, ..., N\}$

and $s, t = \{1, ..., T\}$. This is problematic because α_t^{1i} represents a raw shift to the marginal utility of good-1. α_t^{1i} is thus measured in "utils" and its precise value does not have a behavioural interpretation in an ordinal framework.[4] Without assumptions regarding the unit comparability of individual utilities, the statement that $\alpha_t^{1i} > \alpha_t^{1j}$ does not justify the conclusion that "individual i's taste for good-1 was greater than individual j's taste for good-1". Furthermore, issues of interpersonal comparability aside, the base utility function may differ between individuals leading to a violation of single-crossing. This creates difficulties for those wanting to make global statements concerning relative tastes for good-1 across individuals.

Extending our theoretical framework to impose commonality of the base utility function provides a solution to both of these problems. $v^i(\mathbf{q}) = v^j(\mathbf{q}) = v(\mathbf{q})$ for all $i, j \in \{1, ..., N\}$ ensures that taste shifters are measured relative to a common scale and that interpersonal and intertemporal preferences satisfy the single crossing property in $(\mathbf{q}; \alpha^1)$ space. This facilitates meaningful interpersonal comparison of taste shifters and allows one to make global statements concerning relative tastes for good-1.

PROPOSITION 4.2 A sufficient condition for individual utility functions, $u^i(\mathbf{q}, \alpha^1) = v^i(\mathbf{q}) + \alpha^1 q^1$ and $u^j(\mathbf{q}, \alpha^{1j}) = v^j(\mathbf{q}) + \alpha^{1j} q^1$ to satisfy the single crossing property for any $i, j \in \{1, ..., N\}$, and thus for for $\alpha_t^{1i} > \alpha_t^{1j}$ to be interpreted as "individual i at time t has a greater taste for good-1 than individual j at time s" is $v^i(\mathbf{q}) = v^j(\mathbf{q}) = v(\mathbf{q})$.

PROOF. See Appendix C.

QUANTILE INTERPRETATION Given $v^i(\mathbf{q}) = v^j(\mathbf{q}) = v(\mathbf{q})$ for all $i, j \in \{1, ..., N\}$, the quantile rank of α_t^{1i} locates the quantile rank of individual i's demand for good-1 in a 2-good demand system. To illustrate, consider two individuals i and j for whom $\alpha^{1i} > \alpha^{1j}$, i.e. i has a stronger taste for good-1 than individual j. Then, at any budget $\{\mathbf{p}_t, x_t\}$, individual i will demand a higher quantity of good-1 than individual j at interior solutions: $q_t^{1i} > q_t^{1j}$.

This follows from the first order conditions associated with their respective maximisation problems.

$$\frac{\partial v(\mathbf{q})}{\partial q^{1n}} + \alpha_t^n = \lambda_t^n p_t^1 \qquad \frac{\partial v(\mathbf{q})}{\partial q^{2n}} = \lambda_t^n p_t^2 \tag{4.7}$$

for $n = \{i, j\}$.

4 This complicating factor does not arise in analyses of quality change, such as that of Blow and Crawford (1999), as time varying quality multipliers are measured in quantity units and are thus grounded in a common cardinal scale

At an optimal solution:

$$\frac{\partial v(\mathbf{q}^i)/\partial q^{1i\prime} + \alpha_t^{1i}}{\partial v(\mathbf{q}^i)/\partial q^{2i\prime}} = \frac{p_t^1}{p_t^2} = \frac{\partial v(\mathbf{q}^j)/\partial q^{1j\prime} + \alpha_t^{1j}}{\partial v(\mathbf{q}^i)/\partial q^{2j\prime}} \tag{4.8}$$

With $\alpha_t^{1i} \neq \alpha_t^{1j}$, this condition is not satisfied with $\mathbf{q}_t^i = \mathbf{q}_t^j$. The optimality condition above is only satisfied if $q_t^{1i} > q_t^{1j}$ with commonality of a concave, locally nonsatiated base utility function. Generalising this observation to a set of individuals $i = 1, ..., N$, we then have

$$\begin{aligned} q^{1i} &= Q_{q^1}(\tau^i) \\ \tau^i &= Q_{\alpha^1}^{-1}(\alpha^{1i}) \end{aligned} \tag{4.9}$$

Thus, the quantile rank of one's taste parameter is equivalent to their quantile rank with respect to good-1 consumption. We return to this point when discussing the assumptions that underlie our identification strategy in the empirical application presented in Section 25.

SUMMARY This section has outlined our theoretical framework and formalised the concept of taste change on a single good. Our framework imposes the assumption that individual preferences satisfy a single crossing property in $(\mathbf{q}; \alpha^1)$ space. This enables global statements to be made concerning relative tastes for good-1 over time, independently of the consumption of other goods and total expenditure. Extending our framework to impose commonality of the base utility function across individuals enables global interpersonal comparisons over tastes for good-1 to be made.

22 TESTABLE CONDITIONS

This section derives the testable conditions that are equivalent to a taste rationalisation of observed choice behaviour.

Definition 4.3 is the primary theoretical statement of interest. Intuitively, if one can find a base utility function, $v^i(\mathbf{q})$, and a series of taste shifters, $\{\alpha_t^{1\,i}\}_{t=1,...,T}$, such that the inequalities that constitute Definition 4.3 hold, then choices can be good-1 taste rationalised. This would imply that utility maximisation with time-dependent tastes provides an explanation for observed behaviour. However, the definitions presented in Section 21 do not yield an "effective" tests of the theory as they are parameterised by unobserved structural functions, i.e. the base utility function and the pattern of taste shifters. This section derives the necessary and sufficient conditions defined on observables that are equivalent to a taste rationalisation of observed choice behaviour. These conditions offer an effective test of the theory given that they facilitate falsification in a finite number of steps.

22.1 AFRIAT INEQUALITIES

We first consider the most general model of taste change that is encompassed by our theoretical framework and allow taste instability to take the form of a K-dimensional additive linear perturbation to marginal utilities. Lemma 4.1 defines the revealed preference restrictions on choice behaviour that are equivalent to this form of taste rationalisation.

LEMMA 4.1 The following statements are equivalent:

1. Individual i's observed choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$, can be taste rationalised by the set of taste shifters $\{\boldsymbol{\alpha}_t^i\}_{t=1,...,T}$, where $\boldsymbol{\alpha} \in \mathbb{R}_+^K$.
2. One can find sets $\{v_t^i\}_{t=1,...,T}$, $\{\alpha_t^{k,i}\}_{t=1,...,T}^{k=1,...,K}$ and $\{\lambda_t^i\}_{t=1,...,T}$ with $\lambda_t^i > 0$ for all $t = 1, ..., T$, such that there exists a non-empty solution set to the following inequalities:

$$\begin{aligned} v_s^i - v_t^i + \boldsymbol{\alpha}_t^i(\mathbf{q}_s^i - \mathbf{q}_t^i) &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \\ \alpha_t^{k,i} &\leqslant \lambda_t^i p_t^{k,i} \end{aligned} \qquad (4.L1.1)$$

PROOF. See Appendix C.

We are primarily interested in rationalising choice behaviour by an appeal to taste change on a single good; the empirical application that is engaged in, in this essay concerns taste change for tobacco and, further, an appeal to Occam's razor suggests that one should allow for taste instability for as few goods as possible for reasons of parsimony. Our interest therefore centres on Theorem 4.1, which defines the revealed preference restrictions on choice behaviour when taste change is restricted to a single good.

THEOREM 4.1 The following statements are equivalent:

1. Individual i's observed choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$, can be good-1 rationalised by the set of taste shifters $\{\alpha_t^{1i}\}_{t=1,...,T}$.

2. One can find sets $\{v_t^i\}_{t=1,...,T}$, $\{\alpha_t^{1i}\}_{t=1,...,T}$ and $\{\lambda_t^i\}_{t=1,...,T}$ with $\lambda_t^i > 0$ for all $t = 1, ..., T$, such that there exists a non-empty solution set to the following inequalities:

$$\begin{aligned} v_s^i - v_t^i + \alpha_t^{1i}(q_s^{1i} - q_t^{1i}) &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \\ \alpha_t^{1i} &\leqslant \lambda_t^i p_t^{1i} \end{aligned} \tag{4.T1.1}$$

PROOF. See Appendix C.

Theorem 4.1 is implied by optimising behaviour within the theoretical framework, concavity and local nonsatiation of the utility function. If the inequalities defined by Theorem 4.1 hold, then there exists a well-behaved base utility function and a series of taste shifters on good-1 that perfectly rationalise observed behaviour. These inequalities are an extension to those originally derived by Afriat (1967) for the utility maximisation model with a time-invariant utility function. Imposing $\alpha_t^{1i} = 0$ for all $t = 1, .., T$ returns the standard Afriat inequalities. Further, an absence of intertemporal variation in good-1, i.e. $q_t^{1i} = q_s^{1i}$ for all $t, s \in 1, ..., T$, returns the conditions that are implied by the standard utility maximisation model; one cannot rationalise choice behaviour by alluding to taste variation on a single good in the absence of intertemporal variation in that good.

The virtual budget characterisation of Theorem 4.1 is easily demonstrated by manipulation of its constituent inequalities.

$$\begin{aligned} v_s^i - v_t^i + \alpha_t^{1i}(q_s^{i,1} - q_t^{i,1}) &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \\ v_s^i - v_t^i &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) - \alpha_t^{1i}(q_s^{1i} - q_t^{1i}) \\ v_s^i - v_t^i &\leqslant \lambda_t^i \mathbf{p}_t^{-1\prime}(\mathbf{q}_s^{-1i} - \mathbf{q}_t^{-1i}) + \lambda_t^i(p_t^1 - \alpha_t^{1i}/\lambda_t^i)(q_s^{1i} - q_t^{1i}) \\ v_s^i - v_t^i &\leqslant \lambda_t^i \pi_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \end{aligned} \tag{4.10}$$

The above inequalities are in the same format as the standard Afriat inequalities but with virtual prices in place of those that we observe in reality.

Theorem 4.1 defines the necessary and sufficient conditions on observed behaviour that must be satisfied for choices to be good-1 rationalised by our theoretical framework. I now show that under mild assumptions on the characteristics of available choice data, one can always find a pattern of taste shifters on a single good that are sufficient to rationalise observed choices.

DEFINITION 4.5 There is "perfect intertemporal variation" in good k if $q_t^{k,i} \neq q_s^{k,i}$ for all $t, s = 1, ..., T$.

THEOREM 4.2 Given an individual i's observed choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$ where good-1 exhibits perfect intertemporal variation, one can always find sets $\{\nu_t^i\}_{t=1,...,T}$, $\{\alpha_t^{1i}\}_{t=1,...,T}$ and $\{\lambda_t^i\}_{t=1,...,T}$ with $\lambda_t^i > 0$ for all $t = 1, ..., T$, such that there exists a non-empty solution set to the following inequalities:

$$\begin{aligned} \nu_s^i - \nu_t^i + \alpha_t^{1i}(q_s^{1i} - q_t^{1i}) &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \\ \alpha_t^{1i} &\leqslant \lambda_t^i p_t^{1i} \end{aligned}$$

PROOF. See Appendix C.

This is a strong and exceptionally surprising result. Unless the data has low rounding precision, there typically exists at least one good in an observational data set that satisfies the requirement of perfect intertemporal variation. In this instance, a pattern of taste shifters on a single good can always be found that rationalise observed choice behaviour. Note that the requirement of perfect intertemporal variation is sufficient but not necessary for choices to be good-1 rationalised. For subsets S_t within which $q_t^{1i} = q$ for all $t \in S_t$, if the choice set $\{\mathbf{p}_t^{-1}, \mathbf{q}_t^{-1i}\}_{t \in S_t}$ satisfies GARP, then the inequalities defined by Theorem 4.1 will be satisfied regardless of the failure of good-1 variation.

The proof of Theorem 4.2, found in Appendix C, further gives us that the set of rationalising marginal willingness to pay for good-1 is unbounded above. Therefore, it is not possible to simultaneously identify price effects from taste change on a single good in a non-parametric framework. Choice variation can always be attributed to taste change on a single good rather than a behavioural response to prices. If $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$ passes GARP, choice variation can either be attributed to the rational response dictated by the elasticity of substitution that is embodied by a time-invariant base utility function, or taste instability on a single good, or some combination of the two. If $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$ fails GARP, then choice behaviour cannot be rationalised by an appeal to the rational responses embodied by a static utility function alone. In this instance, a non-zero lower bound on taste instability can be identified. When using this lower bound to represent taste change, choices are rationalised as far as possible by reference to the price effects embodied in the base utility function and the influence of taste instability is minimised. However, this is not to say that the identified lower bound is the "true" degree of taste change on good-1. Please note that choice behaviour could be rationalised by greater taste change. That multiple behavioural rationalisations can be applied to the same choice set is true more generally. However, what is of interest in this setting, is that one can never rule out taste change as a rationalising phenomenon.

Theorem 4.2 is implied by our theoretical framework even though we make relatively restrictive assumptions regarding the functional form of taste instability; tastes for all goods other than good-1 are time invariant, and the marginal rate of substitution between goods other than good-1 are also static. Therefore, Theorem 4.2 goes beyond "Diamond's Impossibility Theorem" (Diamond, McFadden and Rodriguez, 1978), a non-identification result in production theory. That result gives us that it is impossible to separately identify biased technical change from the elasticity of substitution for general production functions. Theorem 4.2 is stronger than this impossibility result given the restrictive assumptions that our theoretical framework embodies concerning the pattern of taste evolution.

Closer parallels can be made between Theorem 4.2 and Varian's (1988) "missing price" result. Varian (1988) establishes that the standard utility maximisation model is virtually emptied of empirical content if the econometrician does not observe the prices of all goods. One can imagine that the unobserved prices take on values so high that expenditure on these goods dominates revealed preference comparisons. The virtual budget characterisation of taste change makes clear the analogy between these results. Tastes for good-1 can always decline to the extent that the virtual prices required to support observed bundles[1] are high enough to prevent an intersection of virtual budget hyperplanes.

DEFINITION 4.6 There is "perfect variation" in good k if $q_t^{k,i} \neq q_s^{k,j}$ for all $t, s = 1, ..., T$ and for all $i, j = 1, ..., N$.

COROLLARY 4.1 Given a finite data set $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$, where $\mathbf{q}^i \in \mathbb{R}_+^K$ that satisfies "perfect variation" in some good k, then intertemporal and interpersonal preference heterogeneity on the K-dimensional demand system can be summarised by a unidimensional taste parameter in the direct utility function when following a fully nonparametric empirical strategy.

The previous section showed that globally meaningful, interpersonal comparisons of taste shifters can be achieved by imposing com-

1 At which there is perfect variation in good-1.

monality of the base utility function across individuals. A corollary of Theorem 4.1 is that, given perfect variation in good-1, there always exists a non-empty solution set to the following inequalities:

$$\begin{aligned} v_s^i - v_t^j + \alpha_t^{1j}(q_s^{1j} - q_t^{1j}) &\leqslant \lambda_t^j \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^j) \\ \alpha_t^{1i} &\leqslant \lambda_t^i p_t^{1i} \end{aligned} \tag{4.11}$$

This is a useful, if unintuitive, result. If the data set satisfies perfect variation, then a pattern of taste shifters on a single good can be found such that the finite set of population choices can be rationalised by our theoretical framework. In other words, in our context, observed heterogeneity over a K-dimensional demand system can be summarised by a univariate parameter in the direct utility function. Thus, the dimensionality of a demand system is not necessarily important for the representation of unobserved heterogeneity if separability assumptions are absent from one's theoretical framework. This is a useful result that has many applications in settings beyond those considered in this essay (for example, see Matzkin (2007)).

SUMMARY This section has proven that observed choice behaviour can always be rationalised by taste change in a single good if that good satisfies the requirement of perfect variation and one operates in a fully nonparametric setting. This has a number of important implications. First, one is unable to separately identify the elasticity of substitution from taste variation in a nonparametric framework. Second, observed heterogeneity that is defined over a K-dimensional demand system can be parsimoniously summarised by a univariate parameter in the direct utility function. The next section develops a quadratic programming procedure to recover the minimal evolution of tastes that is necessary to rationalise observed choices.

23 MINIMAL TASTE EVOLUTION

This section develops a quadratic programming method to calculate the minimal taste perturbations that are necessary to rationalise choice behaviour. If observed choices satisfy GARP, this lower bound will be zero as one does not have to appeal to taste change to rationalise the data set. However, if choices fail GARP, a non-zero lower bound on taste instability for good-1 can be identified. From this point onwards, I focus exclusively upon operationalising Definition 4.3 and Theorem 4.1. This is so because the final object of concern is the characterisation of taste evolution over a single good, tobacco, and, regardless, observational data sets will be rationalisable by an appeal to taste instability on a sole good. Thus, for notational ease, we drop the good subscript on α for what follows, i.e. $\alpha_t^i = \alpha_t^{1i}$.

basic quadratic programme The minimal squared taste perturbations on good 1, relative to tastes in period 1, that are necessary to rationalise the observed choice behaviour of individuals $i = 1, ..., N$, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$, are identified as the unique set $\{\alpha_t^i\}_{t=1,...,T}^{i=1,...,N}$ yielded by the following quadratic programming procedure.

$$\min_{\{v_t^i, \lambda_t^i, \alpha_t^i\}_{t=1,...,T}} \sum_{t=1}^{T} \alpha_t^{i\prime} \alpha_t^i$$

subject to:

1. The rationalisation constraints:

$$\begin{aligned} v_s^i - v_t^i + \alpha_t^i (q_s^{1i} - q_t^{1i}) &\leqslant \lambda_t^i \mathbf{p}_t' (\mathbf{q}_s^i - \mathbf{q}_t^i) \\ \alpha_t^i &\leqslant \lambda_t^i p_t^{1i} \end{aligned}$$

2. The normalisation conditions:

$$\begin{aligned} v_1^i &= \delta \\ \lambda_1^i &= \beta \\ \alpha_1^i &= 0 \end{aligned}$$

for $i = 1, ..., N$ and $s, t = 1, ..., T$.

The squared loss function equally weights positive and negative taste perturbations from the static base utility function and heavily penalises outliers. These are attractive properties in this context as we do not want to bias the direction of recovered taste parameters and the loss function's treatment of outliers aligns with our prior beliefs regarding the smoothness of taste evolution.

The set of rationalisation constraints ensure that the recovered pattern of minimal taste shifters are sufficient to account for observed choice behaviour, i.e. these constraints ensure that the inequalities defined by Theorem 4.1 are respected. With the normalisation conditions imposed, the Basic Quadratic Programme yields a unique solution set, $\{v_t^i, \lambda_t^i, \alpha_t^i\}_{t=1,...,T}$ for each individual $i = 1, ..., N$, where $\{\alpha_t^i\}_{t=2,...,T}$ can be interpreted as the minimal perturbations in the marginal utility of good-1 relative to tastes at $t = 1$ that are necessary to rationalise behaviour. Uniqueness is guaranteed by the linearity of the rationalisation constraints and the positive semidefiniteness of the Hessian of the objective function.

The normalisation conditions are required because, in their absence, the quadratic programming problem is ill-posed. This is so given ordinality of the utility function. Let $\{\bar{v}_t^i, \bar{\lambda}_t^i, \bar{\alpha}_t^i\}_{t=1,...,T}^{i=1,...,N}$ represent a feasible solution to the rationalisation constraints, then the following inequalities hold for all $s, t \in \{1, ..., T\}$ and $i = 1, ..., N$.

$$\bar{v}_s^i - \bar{v}_t^i + \bar{\alpha}_t^i(q_s^{1i} - q_t^{1i}) \geqslant \bar{\lambda}_s^i \mathbf{p}_s'(\mathbf{q}_s^i - \mathbf{q}_t^i) \tag{4.12}$$

However, the following set of inequalities is also feasible:

$$\beta(\bar{v}_s^i + \delta) - \beta(\bar{v}_t^i + \delta) + \beta\bar{\alpha}_t^i(q_s^{i,1} - q_t^{i,1}) \geqslant \beta\bar{\lambda}_s^i \mathbf{p}_s'(\mathbf{q}_s^i - \mathbf{q}_t^i) \tag{4.13}$$

for $\delta \in (-\infty, \infty)$ and $\beta > 0$. Therefore, without the location normalisation, $v_1^i = \delta$, the quadratic programming procedure is ill-posed as there exist an infinite number of sets of utility numbers that can be associated with a given set of feasible taste shifters. $\alpha_1^i = 0$ is imposed without loss of generality to fully specify the value of ordinal utility at $t = 1$. This allows us to interpret $\{\alpha_t^i\}_{t=2,...,T}$ as taste perturbations to good-1 relative to an individual's preferences at $t = 1$.[1] We also impose the scale normalisation $\lambda_1^i = \beta$ for computational reasons. This normalisation ensures that the output of the quadratic programming procedure is scaled sensibly and reduces the sensitivity of the solution to the termination tolerance level of the algorithm that is used to solve the programming problem.

23.1 IMPOSING COMMON BASE UTILITY

A meaningful comparison of estimated taste shifters across individuals requires that they are measured according to a common scale and that they facilitate global statements to be made concerning relative preferences for good-1. This is achieved by imposing commonality of

1 In the words of Sydney Afriat: "A change through time can be recognised only by means of some element which persists unchanged. Thus, when a material body is undergoing deformation in time, this can be known only be means of a measuring rod which is transported through time without deformation." (p.3, 1977)

the base utility function across individuals $i = \{1, ..., N\}$. We term this the "Common Base Utility" (CBU) quadratic programming procedure.

CBU QUADRATIC PROGRAMME The minimal taste perturbations on good 1, relative to individual 1's tastes in period 1, that are necessary to rationalise observed choice behaviour $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}^{i=1,...,N}$, given commonality of the base utility function across individuals $i = \{1, ..., N\}$, are identified as the unique set $\{\alpha_t^i\}_{t=1,...,T}^{i=1,...,N}$ that is yielded by the following quadratic programming procedure.

$$\min_{\{v_t^i, \lambda_t^i, \alpha_t^i\}_{t=1,...,T}^{i=1,...,N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \alpha_t^{i\prime} \alpha_t^i$$

subject to:

1. The rationalisation constraints:

$$\begin{aligned} v_s^i - v_t^j + \alpha_t^j (q_s^{1i} - q_t^{1j}) &\leqslant \lambda_t^j \mathbf{p}_t' (\mathbf{q}_s^i - \mathbf{q}_t^j) \\ \alpha_t^i &\leqslant \lambda_t^i p_t^{1i} \end{aligned}$$

   for all $i, j = 1, ..., N$ and $s, t = 1, ..., T$.

2. The normalisation conditions:

$$\begin{aligned} v_1^1 &= \delta \\ \lambda_1^1 &= \beta \\ \alpha_1^1 &= 0 \end{aligned}$$

The rationalisation conditions associated with the CBU Quadratic Programme extend those of the Basic Quadratic Programme to impose the interpersonal consistency of base utility numbers and taste shifters. The location normalisation again guarantees that the problem is well-posed and allows us to interpret the recovered α's as the minimum taste shifts relative to individual 1's tastes at $t = 1$ that are necessary to rationalise behaviour. Within this framework, the output $\alpha_t^i > \alpha_s^j$, can be interpreted as: "individual i's taste for good-1 at time t dominated that of individual j's taste for good-1 at time s along the minimal taste-evolution path".

23.2 EFFECTIVE TASTES AND VIRTUAL PRICES

The output of the quadratic programming procedures can be used to recover the virtual price of good-1 that supports observed choices. Given the output $\{\widehat{v}_t^i, \widehat{\lambda}_t^i, \widehat{\alpha}_t^i\}_{t=1,...,T}^{i=1,...,N}$ from the CBU Quadratic Programme, the lowest virtual price of good-1 that individual i could face is constructed as:

$$\pi_t^i = p_t^i - \widehat{\alpha}_t^i / \widehat{\lambda}_t^i \quad (4.14)$$

Interpersonal and intertemporal taste heterogeneity can then be represented as variation in virtual prices. Differences in virtual prices represent differences in the marginal willingness to pay for good-1, which we refer to as differences in "effective tastes". Differences in effective tastes do not only derive from α_t^i variation but also from differences in λ_t. λ_t is to be interpreted as the marginal utility of income. It varies across a population to reflect differences in budget environments across individuals. The differences in budget parameters across individuals result in choices occuring in different regions of the base preference map, over which the marginal utility of income differs.[2]

SUMMARY This section has outlined a quadratic programming procedure to recover the minimal taste perturbations that are necessary to rationalise observed choice behaviour. Interpreted from within our theoretical framework, the value of these perturbations allow one to make global statements concerning an individual's relative taste for good-1 at different time periods because preferences satisy the single-crossing property in $(\mathbf{q}, \alpha)$ space. Meaningful interpersonal comparisons of tastes for good-1 can be made if one implements the CBU Quadratic Programme. The rationalisation constraints in this problem ensure that estimated taste-shifters are measured according to a common scale and that an individual's tastes for good-1 can be ranked independently of the quantities of the other goods that they consume and their total budget. In the following sections of this essay, I apply this methodology to explore taste change over tobacco since 1980.

2 We note that it is not advisable to solve directly for theory consistent virtual prices because the set of rationalising $\{\pi_t^i\}_{t=1,\dots,T}^{i=1,\dots,N}$ is non-convex. The non-convexity of this set creates nontrivial computational difficulties when optimising over its elements.

24 CONSTRUCTING QUANTITY SEQUENCES

The preceding sections formalised the concept of taste change on a single good and developed a quadratic programming methodology that allows us to recover the minimal taste shifters that are required to rationalise observed choice behaviour. In the next two sections we apply this framework to rationalise the variation in tobacco consumption in the UK since 1980. This section introduces the data and outlines our strategy for constructing the price and quantity sequences for psuedo cohorts drawn from the Family Expenditure Survey.

24.1 THE DATA

Our empirical analysis makes use of repeated cross-section data drawn from the U.K. Family Expenditure Survey (FES) between 1980 and 2000. The FES records detailed expenditure and demographic information for 7,000 randomly selected households each year. The FES is not a panel; individuals are not followed over time. However, the FES is representative of the population in each year given random sampling. Therefore, although we cannot track individuals, we can construct a set of C psuedo cohorts that allow us to follow representative individuals over time. To this end, we construct a pseudo cohort of individuals who are aged between 25 and 35 years old in 1980 that is stratified by education level. It is a relatively mild assumption that the population of potential smokers in our constructed cohort is stable because less than 5% of smokers start smoking after they reach their 25^{th} birthday (Office for National Statistics, 2012). This allows us to abstract from issues that surround initial preference formation and reduces the likelihood of violations of our later assumption of rank invariance of the distribution of budget shares for tobacco within psuedo cohorts.

Regarding the stratification of cohorts by education level, a "low education" cohort, L, is formed from those individuals who left school no older than 16 years old, the legal minimum. Those staying in school past age 16 comprise our "high education" cohort, H. The set of psuedo cohorts that is defined upon observable characteristics is given by $C = \{L, H\}$.[1] We work with a total number of 28,252 individual observations. The number of observations that contributes to each psuedo-cohort is summarised in Table 29.

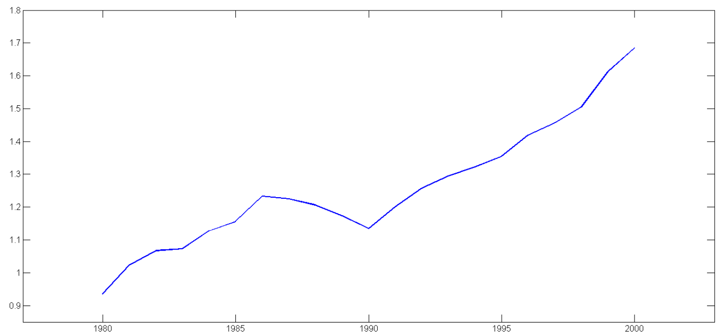
1 We proceed to track "individuals" at different quantiles of the tobacco distribution within these cohorts. This will be discussed thoroughly in due course.

Table 29: Psuedo Cohort Observations

	Min	Mean	Max
Low Education, L	446	596	713
High Education, H	665	750	828

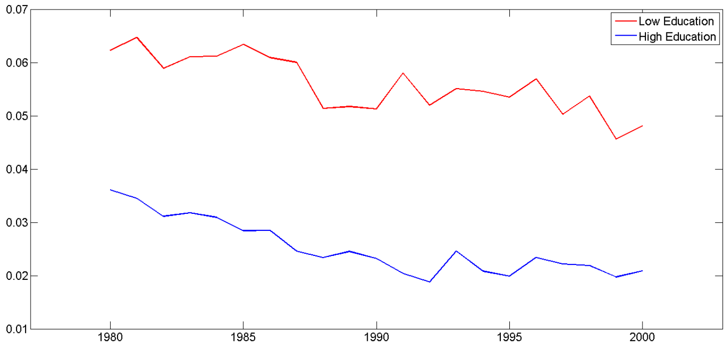
Our analysis is primarily concerned with choices over tobacco and nondurable commodity aggregates. Thus, $K = 2$. Total expenditure is defined as all spending over these goods. Appendix C provides a detailed list of goods that are classed as nondurables in our data set. Chained Laspeyres price indices were constructed for the aggregated commodity groups "tobacco" and "nondurables" from the sub-indices of the U.K. Retail Price Index. From Figure 27, the relative of price of tobacco relative to nondurables rose significantly between 1980 and 2000. This rise has been smooth and systematic apart from a brief stagnation in the late 1980s caused by a reduction in growth of the absolute price of tobacco.

Figure 27: Price of Tobacco Relative to Nondurables

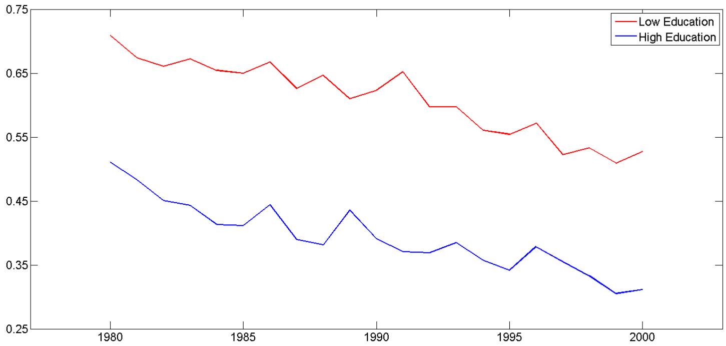


Changes in consumption behaviour have accompanied these changes in price. Figure 28 provides a basic illustration of these trends. Between 1980 and 2000, average tobacco consumption fell for both the low and high education cohorts, although there were differences in the pattern of decline across cohorts. The average proportion of income spent on tobacco was always higher for the low education cohort and declined at a slower rate over the period; the average budget share for tobacco declined by 42.1% for the high education cohort but just 22.7% for the low education cohort. A higher proportion of the low education cohort smoked in every period.

Figure 28: Raw Overtime Trends in Tobacco Consumption



(a) Mean budget share for tobacco



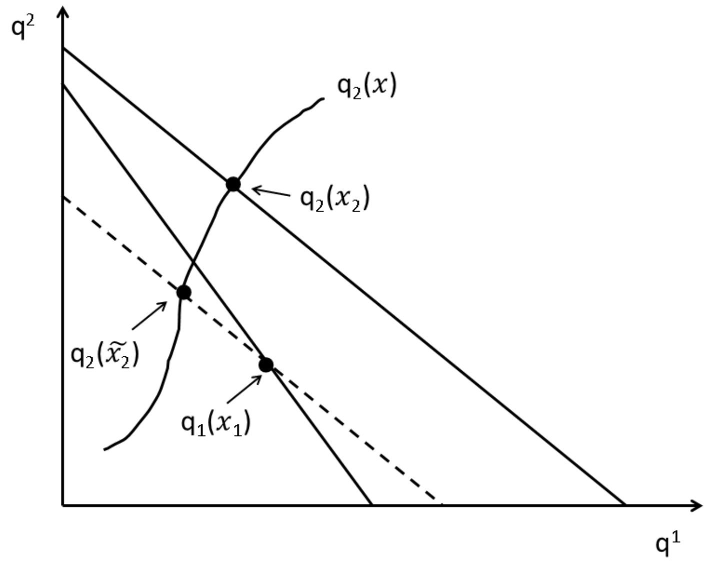
(b) Proportion of cohort with non-zero tobacco expenditure

24.2 BUDGET VARIATION

To make our methodology as powerful as possible, we evaluate each psuedo cohort's demands along the Blundell, Browning and Crawford's (2003) "Sequential Maximum Power" (SMP) paths. Income growth in survey data can prevent the intersection of budget hyperplanes. This negates the ability of our method to detect taste change. Intuitively, the SMP method controls for budget variation and provides us with a choice set that maximises the chance of detecting violations of a time invariant utility function in observational data.

To illustrate these ideas, consider Figure 29. We observe the choices $\mathbf{q}_1$ and $\mathbf{q}_2$ at the budgets $(\mathbf{p}_1, x_1)$ and $(\mathbf{p}_2, x_2)$. These budgets do not intersect. Thus, we are unable to observe that preferences cannot be rationalised by a single time-invariant utility function simply by applying the standard revealed preference methodology to the choices that these budgets support. Blundell *et al.* (2003, 2012) suggest using information on how demands vary with total expenditure to provide a powerful method for detecting violations of the standard model. In our example, using the expansion path $\mathbf{q}_2(x)$ to evaluate demands at the SMP total expenditure level, $\tilde{x}_2 = \mathbf{p}_2'\mathbf{q}_1$, facilitates a detection of

Figure 29: Sequential Maximum Power Demands



the violation of the standard model and would enable the recovery of non-zero taste change through application of our quadratic programming procedure.

In fact, in the pairwise comparison of budgets depicted in Figure 29, $\mathbf{q}_2(\tilde{x}_2)$ represents the demand that yields the most powerful test of static utility maximisation given the observation of bundle $\mathbf{q}_1$. Assuming weak normality and adding up, Proposition 1 of Blundell, Browning and Crawford (2003) implies that if there exists a demand along the expansion path $\mathbf{q}_2(x)$ that fails a GARP test, then the set $\{\mathbf{p}_1, \mathbf{p}_2; \mathbf{q}_1, \mathbf{q}_2(\tilde{x}_2)\}$ will reveal this. The concept of the SMP path of Blundell *et al.* (2003) goes beyond pairwise comparisons to provide the most powerful sequence of demands for testing revealed preference inequalities over a sequence of "preference ordered " budgets. A sequence of budgets is defined as "preference ordered" if each budget is directly revealed at least as good as the next one. If there exists a set of demands along a preference ordered sequence of budgets, $(x_s, x_t, ..., x_v)$, that reveal a violation of utility maximisation with time invariant tastes, then the SMP path for the same preference ordered sequence will reveal this.

In our empirical application, we evaluate the minimal taste evolution along the SMP path that starts at a psuedo cohort's median demand in 1980 and then continues sequentially over time to select the demand that is just weakly preferred to the SMP demand in the previous period. Specifically, we consider the set of psuedo cohort demands:

$$\begin{aligned} \mathbf{q}_1^c &= \mathbf{q}^c(\mathbf{p}_1, x_1) \\ x_1 &= Q_x(0.5) \end{aligned}$$

and

$$\begin{aligned} \mathbf{q}_t^c &= \mathbf{q}^c(\mathbf{p}_t, \widetilde{x}_t) \\ \widetilde{x}_t &= \mathbf{p}_t' \mathbf{q}_{t-1} \end{aligned}$$

for $t = 2, ..., T$ and $c = \{L, H\}$.

We are able to abstract from the complicating issues caused by transitivity in the construction of the SMP path, which are examined in Blundell *et al.* (2012), because we focus on a two-good demand system that is defined over tobacco and nondurables. As first highlighted by Rose (1958), transitivity has no empirical content in the context of a 2-good demand system.

24.3 CENSORED QUANTILE REGRESSION

Operationalising the SMP methodology requires the estimation of expansion paths for each psuedo cohort at each price regime. To estimate these expansion paths we draw upon insights from Chernozhukov and Hong (2002) and Chernozhukov, Fernandez-Val and Kowalski (2010). Specifically, we employ censored quantile regression methods and use a control function approach to correct for the endogeneity of total expenditure. We also allow for unrestricted heterogeneity across psuedo cohorts.

Censored quantile regression models are a natural choice in our context. As discussed in the motivation of our theoretical framework, the quantile rank of an individual's α identifies the quantile rank of their tobacco demand. Quantile regression models also allow covariates to shift the location, scale and shape of the distribution of tobacco budget shares at each price regime. This permits a distribution-free specification, thereby reducing the likelihood of misspecification biases being labelled as taste change. Furthermore, the latent budget share for tobacco, $w^{i\star}$ is left censored at zero: $w^i = max\{0, w^{i,\star}\}$. Quantile regression models are equivariant to monotone transformations such as censoring and allow us to avoid the strong parametric assumptions of traditional Tobit estimators.

In what follows, we express our estimation method in terms of the budget share for tobacco, w^i. Specifically, we consider the following triangular system of quantile equations at each price regime:

$$\begin{aligned} w^i &= max(0, w^{i,\star}) \\ w^{i\star} &= Q_{w^{i,\star}}(\epsilon^i | x^i, \mathbf{z}^i, \nu^i, c^i) \\ x^i &= Q_{x^i}(\nu^i | \mathbf{z}^i, m^i, c^i) \end{aligned} \tag{4.15}$$

where

$$\begin{aligned} \epsilon^i &\sim U(0,1)|x^i, \mathbf{z}^i, \nu^i, m^i, c^i \\ \nu^i &\sim U(0,1)|\mathbf{z}^i, m^i, c^i \end{aligned} \tag{4.16}$$

and x^i is total expenditure on nondurables and tobacco, $\mathbf{z}^i$ is a vector of household characteristics, ν^i is an unobserved latent variable and $c^i \in C$ denotes individual i's psuedo cohort membership.

The latent variable ν^i is included to account for the possible endogeneity of x^i. Total spending on nondurables may be jointly determined with the allocation of spending across goods. To recover the conditional quantile function of the budget share for tobacco, we must condition upon ν^i. m^i, the log of disposable income, represents our excluded instrument and allows us to recover ν^i as the conditional quantile of x given $(\mathbf{z}^i, m^i, c^i)$. ν^i plays the role of a "control variable" in our framework (see Blundell and Powell (2007) for a similar application). By conditioning on ν^i, we can identify ϵ^i as the conditional quantile of $w^\star$ given $(x^i, \mathbf{z}^i, \nu^i, c^i)$.

We use this model to identify conditional psuedo cohort demands at a set of quantiles of the tobacco distribution within each price regime along the SMP expenditure path: $\varepsilon = \{\epsilon^1, ..., \epsilon^\tau\}$. We note here that this model imposes a rank invariance restriction on the budget share for tobacco distribution within psuedo cohorts; taste evolution and price effects cannot alter the ranking of individuals within the same education cohort in the budget share for tobacco distribution. Therefore, across price regimes the following condition is imposed:

$$\text{If } w^{i,\star}_{\mathbf{p}_t} \leqslant Q_{w^\star|\mathbf{p}_t}(\epsilon|x^i, \mathbf{z}^i, \nu^i, c^i) \text{ , then } w^{i,\star}_{\mathbf{p}_s} \leqslant Q_{w^\star|\mathbf{p}_s}(\epsilon|x^i, \mathbf{z}^i, \nu^i, c^i) \tag{4.17}$$

for all $s, t = \{1, ..., T\}$. This imposes a restriction on taste evolution at different parts of the tobacco distribution. Specifically, for any uncensored quantiles of the budget share distribution, ϵ and ϵ', with $\epsilon > \epsilon'$, we have that:

$$\alpha^\epsilon_t > \alpha^{\epsilon'}_t \tag{4.18}$$

for all $t = 1, ..., T$. Intuitively, the tastes for tobacco for higher smoking cohorts must always remain greater than those for the lower smoking cohorts.

This assumption is strong but, in our opinion, not incredible. We only consider individuals who are old enough to have made an initial decision to ever smoke or not and high profile anti-smoking campaigns are often not targeted in such a manner that would lead one to expect significant re-ranking within pseudo cohorts.

ESTIMATION To estimate this model, we impose semiparametric restrictions upon the functional form of the conditional quantile functions of $w^\star$ and x. In the case of the conditional quantile function of

$w^\star$, our specification extends the standard censored quantile model of Powell (1984) to allow for unrestricted preference heterogeneity across psuedo cohorts and the endogeneity of total expenditure:

$$\begin{aligned} Q_{w^{i,\star}}(\epsilon^i|x^i, \mathbf{z}^i, \nu^i, c^i) &= \mathbf{X}^{i\prime}\boldsymbol{\beta}^{c^i}(\epsilon^i) \\ \mathbf{X}^i &= f(x^i, \mathbf{z}^i, \nu^i) \end{aligned} \tag{4.19}$$

where $c^i \in C$ and $f(x^i, \mathbf{z}^i, \nu^i)$ is a vector of transformations of x^i, $\mathbf{z}^i$ and ν^i. As outlined in Chernozhukov, Fernandez-Val and Kowalski (2010), these transformations could be polynomial, trigonometric and other basis functions that allow us to approximate the patterns in the data well.

Assuming a random sample of n observations at each price regime $\{w^i, x^i, \mathbf{z}^i, \nu^i, m^i, c^i\}^{i=1,\dots,N}$, $\boldsymbol{\beta}^c(\epsilon)$, $c = \{L, H\}$, is estimated as:

$$\boldsymbol{\beta}^c(\epsilon) = \arg\min_{\beta} \sum_{i \in S^c} \rho_\epsilon(w^i - \widehat{\mathbf{X}}^{i\prime}\boldsymbol{\beta}) \tag{4.20}$$

where $\rho_\epsilon(u) \equiv (\epsilon - 1(u \leqslant 0))u$ is the standard asymmetric absolute loss function of Koenker and Bassett (1978). $\mathbf{X}^i$ is replaced with $\widehat{\mathbf{X}}^i = f(x^i, \mathbf{z}^i, \widehat{\nu}^i)$, where $\widehat{\nu}^i$ is an estimator of ν^i, given that the true value of ν^i goes unobserved. S^c denotes the set of observations on individuals i for which $c^i = c$ and $\Pr(w^i > 0) > \delta$, i.e. the subset of individuals in psuedo cohort c for which the probability of censoring is negligible and a linear functional form for the conditional quantile is justified.

This estimator extends the estimator of Chernozhukov, Fernandez-Val and Kowalski (2010) to allow for unrestricted heterogeneity across cohorts and is implemented using an adaption of the censored quantile regression algorithm developed in Chernozhukov and Hong (2002). There are three elements to this algorithm:

1. Estimation of the control variable.
2. Estimation of the set of quantile uncensored observations.
3. Estimation of quantile coefficients.

I will briefly outline the method in full before providing greater detail on each stage. Intuitively, one first estimates the control variable ν in order to construct the set of regressors $\widehat{\mathbf{X}}$. The conditional quantile function is then estimated using an iterative procedure. Each stage takes the form of Equation 4.20, while the set S^c is sequentially updated to increase estimator efficiency. S^c is initially constructed as the set of "quantile uncensored" (Chernozhukov *et al.*, 2010) observations on a psuedo cohort, i.e. those observations for which the conditional quantile function is non-negative. This is achieved by estimating the conditional probability of censoring using some binary choice model. First stage coefficients, $\boldsymbol{\beta}^c_0(\epsilon)$, are then estimated on this initial set of observations using standard quantile regression techniques. The results of this exercise are used to update the set S^c by selecting those observations for which $\widehat{\mathbf{X}}^{i\prime}\boldsymbol{\beta}_0^c(\epsilon) > 0$. This updating procedure is repeated until an efficient estimator is obtained. These stages are now outlined in more detail.

1. CONTROL VARIABLE ESTIMATION Conditioning upon the latent control variable, ν, facilitates the recovery of the conditional quantile function of w^i when total expenditure is endogenous. However, ν is unobserved and must be estimated. ν is recovered as a residual that explains the movement in x conditional upon the excluded instrument m and demographic covariates. This constitutes the first stage of our estimation procedure.

Assuming that $Q_x(\nu|z, m, c)$ is invertible in ν, ν can be represented as:

$$\nu \equiv Q_x^{-1}(\nu|z, m, c) \tag{4.21}$$

This representation yields the estimator:

$$\widehat{\nu}_i \equiv \widehat{Q}_{x^i}^{-1}(\nu^i|z^i, m^i, c^i) \tag{4.22}$$

In our application, a linear additive model is assumed. Under this specification, ν is easily recovered from the cumulative distribution function of the least squares residuals:

$$\begin{aligned} Q_x(\nu|z, m, c) &= \mathbf{Z}'\boldsymbol{\pi}^c + Q_\nu(\nu|c) \\ \widehat{\nu}_i &= Q_\nu^{-1}(x_i - \mathbf{Z}_i'\widehat{\boldsymbol{\pi}}^c|c) \end{aligned} \tag{4.23}$$

2. QUANTILE UNCENSORED OBSERVATIONS Following the first stage estimation of the control variable, the vector of regressors $\widehat{\mathbf{X}}^i = f(x^i, \mathbf{z}^i, \widehat{\nu}^i)$ can be constructed and used to recover the conditional quantile function of $w^\star$. Our estimation procedure then continues to extend the method of Chernozhukov and Hong (2002) to allow for unrestricted heterogeneity across pseudo cohorts. Their algorithm constitutes a practical way to estimate censored quantile regression models, especially in contexts such as our own in which censoring is particularly heavy and the convergence of traditional estimators is low. Intuitively, one selects a subset of "quantile uncensored" observations within each psuedo cohort for which the probability of censoring is low enough to specify a linear functional form for the conditional quantile, $S^c = \{i : c^i = c \,\&\, \mathbf{X}^{i\prime}\boldsymbol{\beta}^c(\epsilon) > 0\}$.

Implementation proceeds by initially selecting an initial subset of observations, S_0^c, for which the quantile line exceeds zero. This is achieved by estimation of a probability model to give a likelihood of censoring for each observation, δ^i:

$$\delta^i = p^{c^i}(\mathbf{X}^{i\prime}\boldsymbol{\gamma}^{c^i}) + e^i \tag{4.24}$$

The initial set of ϵ-uncensored observations, S_0^c, is then selected as

$$S_0^c = \{i : c^i = c \,\&\, \widehat{\delta}^i > 1 - \epsilon + t_0\} \tag{4.25}$$

where $\widehat{\delta}^i = p^{c^i}(\widehat{\mathbf{X}}^{i\prime}\widehat{\boldsymbol{\gamma}}^{c^i})$ and $0 \leqslant t_0 \leqslant \epsilon$ is a cut-off constant chosen to ensure that only uncensored observations are included. These observations have conditional probability of censoring lower than ϵ given the equivalence of the events

$$\{\mathbf{X}^{i\prime}\boldsymbol{\beta}^c(\epsilon) > 0\} \equiv \{\Pr(w^{i\star} \leqslant 0|\mathbf{X}^i) < \epsilon\} \tag{4.26}$$

The specification of the probability model is not particularly important. The purpose of this step is simply to select *some* subset of quantile-uncensored observations with which to obtain an initial inefficient estimate $\widehat{\boldsymbol{\beta}}_0^c$. This is possible so long as $\mathbf{X}^i\mathbf{X}^{i\prime}1\{i \in S_0^c\}$ is invertible. We refer to Chernozhukov and Hong (2002) for further details.

3. β-ESTIMATION Given an initial set of quantile uncensored observations S_0^c, one proceeds to obtain an inefficient first stage estimator $\widehat{\boldsymbol{\beta}}_0^c(\epsilon)$ by standard quantile regression methods:

$$\boldsymbol{\beta}_0^c(\epsilon) = \arg\min_{\beta} \sum_{i \in S_0^c} \rho_\epsilon(w^i - \widehat{\mathbf{X}}^{i\prime}\boldsymbol{\beta}) \tag{4.27}$$

This estimator is used to update the set S_0^c to S_1^c so that a more efficient estimator is obtained at the next stage.

$$S_1^c = \{i : c^i = c \;\&\; \mathbf{X}^{i\prime}\widehat{\boldsymbol{\beta}}_0^c(\epsilon) > 0 + t\} \tag{4.28}$$

where t is a small positive number that tends to zero in the sample size: $t \to 0$ and $\sqrt{n_c}t \to \infty$ as $n_c \to \infty$ where $n_c = \sum_{i=1}^N 1(c^i = c)$.

This procedure is repeated over iterations $j = 2, .., J$ to asymptotically select observations that have covariate values $\mathbf{X}^{i\prime}\boldsymbol{\beta}^c(\epsilon) > 0$. Monte Carlo simulations in Chernozhukov and Hong (2002) demonstrate that only one iteration is required to obtain bias and mean squared errors that dominate, or are at least comparable, to the traditional Powell estimator.

$$\boldsymbol{\beta}_j^c(\epsilon) = \arg\min_{\beta} \sum_{i \in S_{j-1}^c} \rho_\epsilon(w^i - \widehat{\mathbf{X}}^{i\prime}\boldsymbol{\beta}) \tag{4.29}$$

where

$$S_j^c = \{i : c^i = c \;\&\; \mathbf{X}^{i\prime}\widehat{\boldsymbol{\beta}}_{j-1}^c(\epsilon) > t\}$$

24.4 EXPANSION PATHS

The method above is applied to estimate expansion paths for each psuedo cohort at each price regime. Demands are evaluated at a set of uncensored quantiles along the SMP expenditure path. These demands will be used in our quadratic programming procedure to recover taste changes for tobacco across psuedo-cohorts and quantiles of the tobacco distribution. Due to the heavy censoring of tobacco, we are unable to estimate coefficients at every quantile and thus we focus upon recovering demands for each psuedo cohort at the set of quantiles $\varepsilon = \{0.55, 0.65, 0.75\}$.

The set of transformed regressors used for the estimation procedure was specified as:

$$\mathbf{X}^i = f(x^i, \mathbf{z}^i, \nu^i) = [1, \log(x^i), (\log(x^i))^2, oecd^i, \nu^i] \quad (4.30)$$

where $oecd^i$ represents the OECD equivalence index for household i. A probit link function was used as the probability model to construct the initial set of quantile-uncensored observations. The initial cut-off constant was set as $t_0 = 0.05$ and subsequent cut-off constants were set at $t = 0.02$. Experimenting with different probability models and cut-off constants had negligible effect on estimated demands and our results are robust to sensible specification changes at this stage.

SMP DEMANDS The estimated expansion paths were used to recover a set of demands with which to investigate patterns in minimal taste evolution. We recover demands for single individuals with the median value of the control variable along the SMP expenditure path. Specifically, the following set of demands was recovered:

$$\mathbf{q}_t^{c,\epsilon} = (\mathbf{p}_t'/\tilde{x}_t^c)\mathbf{s}_t^{c,\epsilon} \quad (4.31)$$

where $\mathbf{s}_t^{c,\epsilon}$ is the vector of budget shares demanded at the ϵ-quantile of cohort c's budget share for tobacco distribution given the budget defined by price regime t and SMP total expenditure. Formally,

$$\begin{aligned} \mathbf{s}_t^{c,\epsilon} &= [\max(0, w_t^{\star c,\epsilon}), \min(1, 1 - w_t^{\star c,\epsilon})] \\ w_t^{\star c,\epsilon} &= Q_{w^{\star c,\epsilon}|\mathbf{p}_t}(\epsilon|\tilde{x}_t^c, z = 0.67, \nu = 0.5, c) \end{aligned} \quad (4.32)$$

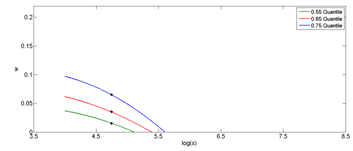
and

$$\begin{aligned} \tilde{x}_1^c &= Q_x(0.5|c) \\ \tilde{x}_t^c &= \mathbf{p}_t'\tilde{\mathbf{q}}_{t-1}^c \\ \tilde{\mathbf{q}}_t^c &= m(\tilde{x}_{t-1}^c|z = 0.67, \nu = 0.5, c, \mathbf{p}_t) \end{aligned} \quad (4.33)$$

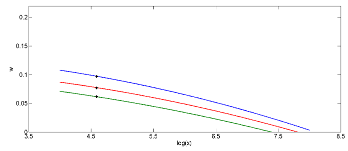
for $t = 2, ..., T$, $c = \{L, H\}$, $\epsilon = \{0.55, 0.65, 0.75\}$. $\tilde{\mathbf{q}}_t^c$ represents the "standard" SMP demand for psuedo cohort c at price regime t estimated using nonparametric kernel regression methods as in Blundell *et al.* (2003, 2008). $\tilde{\mathbf{q}}_t^c$ is used to construct the total expenditure sequence along which we recover demands to ensure comparability of our method with the prior literature on SMP paths.

ESTIMATED EXPANSION PATHS Estimated expansion paths unambiguously reveal that tobacco is an inferior good. Figure 30 depicts the estimated Engel curves for the three considered quantiles, $\varepsilon =$

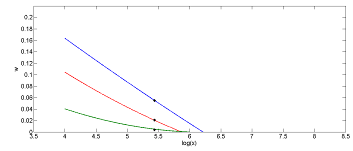
Figure 30: Engel Curves for Tobacco



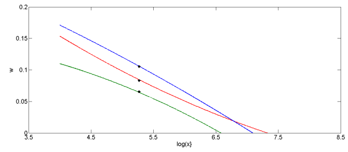
(a) High Education – 1980



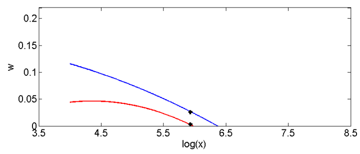
(b) Low Education – 1980



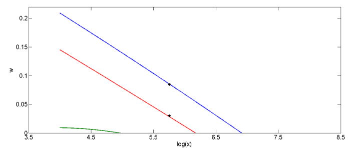
(c) High Education – 1990



(d) Low Education – 1990



(e) High Education – 2000



(f) Low Education – 2000

$\{0.55, 0.65, 0.75\}$ for each psuedo cohort at 1980, 1990 and 2000, with the SMP expenditure levels marked. Table 30 gives the set of recovered SMP demands from the estimated Engel curves with the bootstrapped 95% confidence interval. Parameter estimates with the bootstrapped 95% confidence intervals are given for all expansion paths in Appendix C.

Engel curves at each price regime are everywhere downward sloping over the observed range of total expenditure and are approximately linear in the natural logarithm of total expenditure on nondurables and tobacco. Censoring is a salient feature of the data. The 0.55-quantile for the high education group is only non-zero at low levels of total expenditure and is never above zero in the year 2000. Extreme censoring is not as prevalent for observations comprising the low education cohort. However, by the year 2000, expenditure on tobacco rarely exceeds zero at the 0.55 quantile for the less educated cohort. We also note that the "quantile crossing" problem is not a salient concern for this data set. Unconstrained quantile lines typically do not cross and if they do (as in Figure 30 (d)), it is at the boundary of the data.

SUMMARY In summary, this section has outlined a methodology that allows us to construct quantity sequences for tobacco that can be used as inputs into our quadratic programming procedure. We employ censored quantile regression methods to recover demands along the SMP expenditure path for each pseudo cohort at a set of uncensored quantiles of the tobacco distribution. Our method corrects for the endogeneity of total expenditure using a control variable approach and employs the algorithm developed by Chernozhukov and Hong (2002). Our chosen specification allows for unrestricted preference heterogeneity across cohorts and for a great deal of flexibility over the impact of regressors across the distribution of budget shares for tobacco. However, this method does impose a rank invariance assumption upon the budget share for tobacco distribution within psuedo cohorts.

Table 30: SMP Quantities with 95% Confidence Interval

Year	Ed:	Low Ed.			High Ed.		
	Quantile:	0.55	0.65	0.75	0.55	0.65	0.75
1980		4.68	5.79	7.29	1.31	3.13	5.75
		4.12	5.27	6.61	0.74	2.19	4.72
		5.41	6.42	7.90	2.28	4.35	6.60
1982		2.63	4.42	5.98	0.64	2.47	4.12
		1.82	3.71	5.25	0.00	1.45	3.55
		3.57	5.19	6.50	1.31	3.20	5.06
1984		2.88	4.42	5.75	0.08	0.58	3.41
		1.61	3.47	5.06	0.00	0.00	2.26
		3.77	5.15	7.02	0.48	2.02	4.62
1986		3.75	4.93	6.37	0.56	2.23	3.84
		2.90	4.32	5.66	0.00	1.08	2.93
		4.35	5.63	6.86	1.22	2.65	4.58
1988		3.83	4.55	5.51	0.00	0.73	3.40
		3.10	4.25	5.00	0.00	0.02	2.46
		4.35	5.05	6.03	0.30	1.65	4.20
1990		4.13	5.24	6.63	0.29	1.54	4.06
		3.15	4.64	5.95	0.00	0.86	3.08
		4.74	5.99	7.33	1.03	2.62	5.12
1992		2.42	4.18	5.60	0.40	1.44	2.98
		1.59	3.56	5.08	0.00	0.53	2.05
		3.65	5.00	6.26	0.73	1.92	3.76
1994		2.30	4.38	5.65	0.03	0.83	3.06
		1.27	3.25	4.96	0.00	0.16	2.22
		3.31	4.91	6.32	0.38	1.79	3.44
1996		2.10	4.25	5.59	0.45	1.28	3.04
		1.22	2.90	4.94	0.00	0.00	2.47
		3.03	4.87	6.20	0.75	2.19	3.45
1998		1.10	3.09	4.94	0.00	0.48	2.25
		0.59	2.07	4.05	0.00	0.00	1.36
		2.33	4.06	5.75	0.31	1.09	2.97
2000		1.28	2.67	3.67	0.00	0.16	1.31
		0.43	1.68	3.34	0.00	0.00	0.53
		1.89	3.38	4.57	0.21	0.56	2.01

25 TASTE CHANGE FOR TOBACCO

We now apply our quadratic programming procedure to recover the minimal amount of taste variation that is required to rationalise each psuedo cohorts' quantile demands along their SMP total expenditure path. We find that some degree of taste evolution is necessary to rationalise the behaviour of all psuedo cohorts. This implies that all cohorts' SMP demands fail GARP and that the virtual price trajectories of all cohorts are significantly different from the observed price path. We also find that the rise in virtual prices amongst light and moderate smokers has been significantly greater for high education cohorts. Thus, amongst all but the heaviest smokers, there are significant educational differences in the effective taste for tobacco; the marginal willingness to pay for tobacco is lower for more highly educated individuals. However, these recovered differences in effective tastes can be partially attributed to income effects. This reduces the strength of conclusions that we can make concerning educational differences in the parameters of the direct utility function.

25.1 ESTIMATION DETAILS

As outlined in the previous section, the objects of concern are psuedo cohorts formed from individuals born between 1945 and 1955, that are stratified by education level. In effect, we estimate taste change for tobacco over the period 1980 to 2000 for six psuedo cohorts because we recover quantity sequences at the set of quantiles $\varepsilon = \{0.55, 0.65, 0.75\}$ for each psuedo cohort (that is defined with respect to an education level). In what follows, we refer to the 0.55-quantile as "light smokers", we refer to the 0.65-quantile as "moderate smokers" and we refer to the 0.75-quantile as "heavy smokers".

We are interested in comparing the magnitude of taste changes across these cohorts. Thus we use estimated quantity sequences as inputs into our CBU quadratic programming procedure. Therefore, the reported minimal taste shifters are those yielded by the following quadratic programme:

$$\min_{\mathbf{v},\boldsymbol{\lambda},\boldsymbol{\alpha}} \sum_{c} \sum_{\epsilon} \sum_{t} (\alpha_t^{c,\epsilon})^2$$

subject to:

1. The rationalisation constraints:

$$v_s^{c,\epsilon} - v_t^{d,\tau} + \alpha_t^{d,\tau}(q_s^{c,\epsilon,1} - q_t^{d,\tau,1}) \geqslant \lambda_t^{d,\tau} \mathbf{p}_t'(\mathbf{q}_s^{c,\epsilon} - \mathbf{q}_t^{d,\tau})$$
$$\lambda_t^{c,\epsilon} p_t^1 - \alpha_t^{c,\epsilon} \geqslant 0$$

for all:

- $s, t \in \{1980, 1982, ..., 1998, 2000\}$
- $c, d \in \{L, H\}$
- $\epsilon, \tau \in \{0.55, 0.65, 0.75\}$

2. The normalisation conditions:

$$\begin{aligned} \nu_{1980}^{L,0.75} &= 0 \\ \lambda_{1980}^{L,0.75} &= 10 \\ \alpha_{1980}^{L,0.75} &= 0 \end{aligned}$$

Given the normalisation conditions, estimated taste shifters are to be interpreted as perturbations in the marginal utility for tobacco relative to the tastes of the "heavy smoking"-"low education" cohort in 1980. Our estimation procedure recovers a lower bound on taste shifters. We further impose the natural assumption of monotonicity by here reporting the lower envelope of minimal changes in the marginal willingness to pay. Raw results are given in Appendix C.

The implementation of our methodology is a computationally intensive procedure. There are $(T|C||\varepsilon|)^2$ constraints that are associated with the CBU quadratic programming procedure. To limit the computational burden of the empirical exercise, we restrict observations to have biennial periodicity. Therefore, there are $(11 \cdot 2 \cdot 3)^2 = 4356$ rationalisation constraints that must be satisfied by recovered taste shifters.

We bootstrap the procedure to address the issue of sampling variation in estimated quantity sequences. Specifically, we randomly draw observations with replacement within each education-time cell 1000 times and estimate quantile demands on each resampled set of observations. Minimal taste change is then estimated for these sets of perturbed quantities, allowing us to construct a simultaneous 95% confidence interval on taste perturbations for psuedo cohorts.

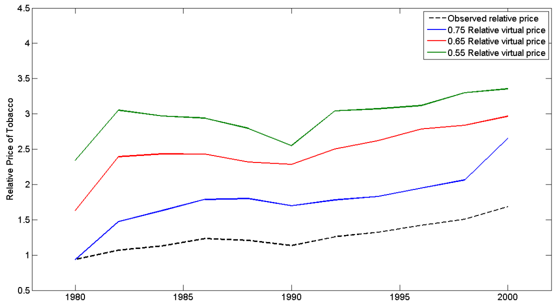
25.2 EFFECTIVE TASTES AND VIRTUAL PRICES

This procedure recovers significant differences in the virtual price trajectories along the SMP total expenditure path across psuedo cohorts. Minimal virtual prices along each psuedo cohort's SMP path are recovered as:

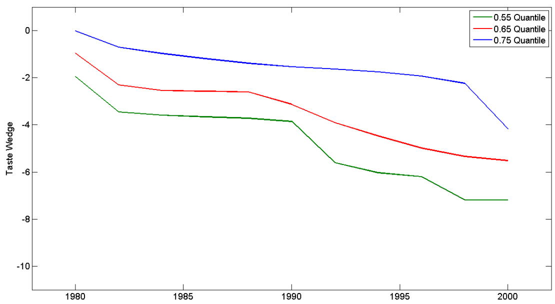
$$\widehat{\pi}_t^{c,\epsilon} = p_t^1 - \frac{\widehat{\alpha}_t^{c,\epsilon}}{\widehat{\lambda}_t^{c,\epsilon}} \tag{4.34}$$

Figures 31 and 32 depict the minimum virtual prices that are necessary to rationalise the choice behaviour of the considered psuedo cohorts, given the base normalisation relative to the highest smoking, low education cohort in 1980. Appendix C lists the full output of our

Figure 31: Low Education- Virtual Prices



(a) Virtual Prices

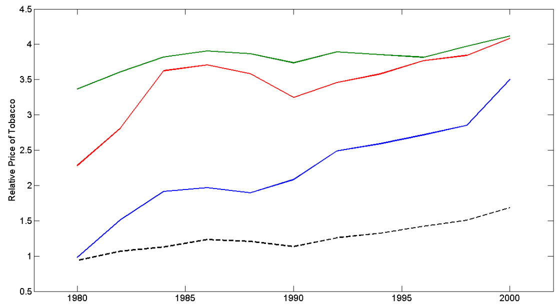


(b) Taste Wedge

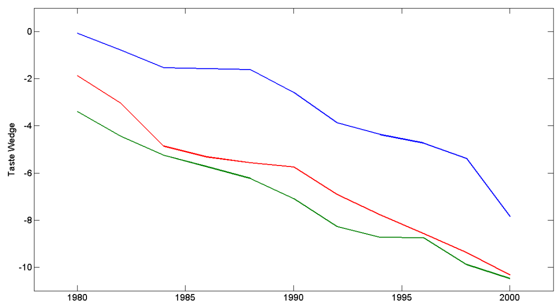
method. In addition to recovered virtual prices, the observed relative price for tobacco is depicted. This is the trajectory that each psuedo cohort's virtual prices would follow in the absence of intercohort and intertemporal taste heterogeneity. The "taste wedge", $\widehat{\alpha}_t^{c,\epsilon}/\widehat{\lambda}_t^{c,\epsilon}$ is also displayed to aid interpretation. This represents the change in the marginal willingness to pay for tobacco relative to base tastes along the SMP expenditure paths.

Some degree of taste variation is necessary to rationalise the behaviour of every cohort for the period 1980-2000; all virtual price trajectories are significantly different from the one that is observed in reality. Figures 31 and 32 display the intracohort heterogeneity in virtual prices and the taste wedge. We here suppress the 95% confidence intervals to allow for an uncluttered overview of intracohort trends but they are given in Appendix C. Unsurprisingly, effective tastes for tobacco are greater at heavier smoking quantiles, although the taste trajectories of light and moderate smokers in the high education cohort are similar. This should be expected given the heavy censoring that effects these groups.

Figure 32: High Education- Virtual Prices

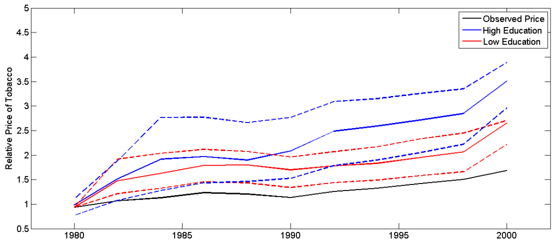


(a) Virtual Prices

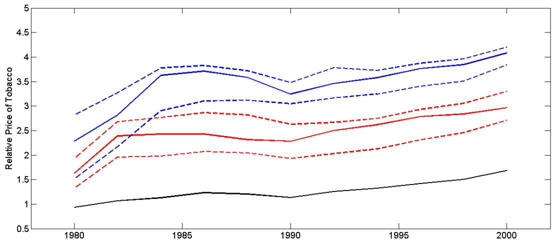


(b) Taste Wedge

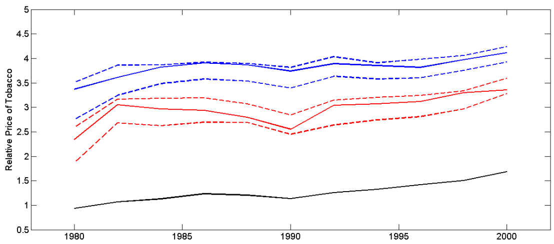
Figure 33: Virtual Prices



(a) Heavy Smokers



(b) Moderate Smokers

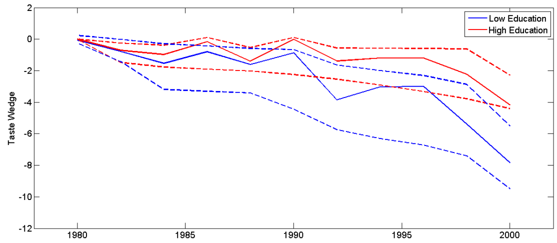


(c) Light Smokers

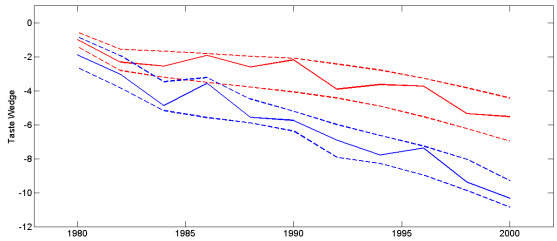
Our results uncover statistically significant heterogeneity in virtual prices across education cohorts for light and moderate smokers along their respective quantity sequences. This is highlighted by reference to Figures 33 and 34. These figures illustrate that the confidence intervals on the virtual price trajectories for low and high education cohorts are disjoint at light and moderate smoking quantiles. Therefore, there are statistically significant socioeconomic differences in virtual prices amongst light and moderate smoking groups. In fact, Figure 35 highlights that the virtual price for tobacco that is faced by "high education"-"moderate smokers" is always greater than that of "low education"-"light smokers" after 1985. Therefore, these results highlight the existence of significant educational differences in the effective taste for tobacco for these smoking groups.

However, education is irrelevant for explaining the evolution of virtual prices amongst heavy smokers. The 95% confidence intervals over the minimal virtual prices trajectories overlap for heavy smok-

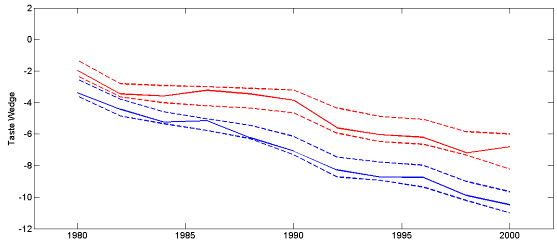
Figure 34: Taste Wedge



(a) Heavy Smokers



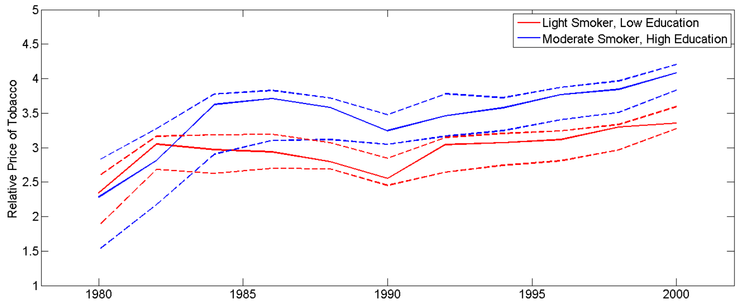
(b) Moderate Smokers



(c) Light Smokers

ers and, therefore, one cannot distinguish significant education differences in the marginal willingness to pay for tobacco amongst this group. This suggests that we should expect to observe diminished differences in smoking behaviour by education group in the future. If current trends continue, we will come to only observe positive tobacco expenditure amongst those who are currently heavy smokers and amongst whom education appears to be a less important factor in rationalising behaviour changes.

Figure 35: Disjoint Virtual Price Trajectories

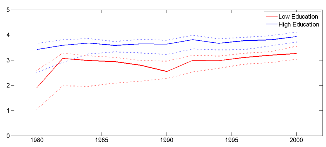


There are a number of separability assumptions that are implicit in our estimation strategy. Not including alcohol as a commodity in our estimated demand system amounts to an assumption over the separability of alcohol and tobacco in the utility function. To examine the robustness of our findings on effective tastes to this assumption, and to explore whether additional patterns emerge from the data once this condition is relaxed, we re-run our quadratic programming procedure on quantile demands that are estimated conditional on alcohol consumption.

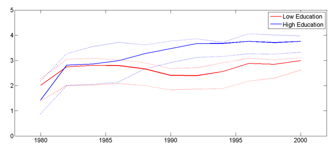
We partition the set of observations that comprise each psuedo cohort into "light" and "heavy" drinkers depending on whether an individual demands below or above the median budget share for alcohol in that time period. The quantile regression method outlined in the previous section is then applied to estimate demands within each education-alcohol-time cell. Base preferences are normalised relative to those of the heavy smoking, heavy drinking, low education cohort in 1980. This procedure allows for unrestricted heterogeneity across alcohol groups.

We recognise that the rank invariance assumption, which underlies our quantile approach to estimating SMP demands, is strong in this context. It amounts to a no re-ranking requirement on the joint distribution of tobacco-alcohol group budget shares. We cannot determine how strong this assumption is because we do not have access to panel data for the period. However, research suggests a robust, if modest, positive correlation between alcohol and smoking consumption that is replicated across many studies, which lends some support to our strategy (Bobo and Husten, 2000; Falk, Yi and Hiller-Sturmhofel, 2006).

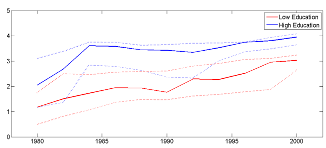
Figure 36: Virtual Prices



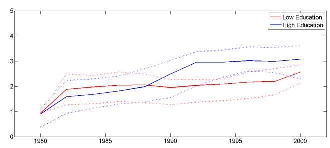
(a) Moderate Smokers - Low Alcohol



(b) Moderate Smokers - High Alcohol

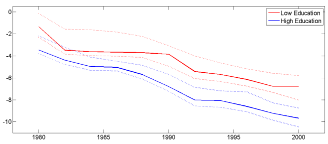


(c) Heavy Smokers - Low Alcohol

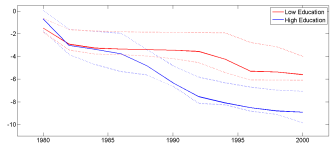


(d) Heavy Smokers - High Alcohol

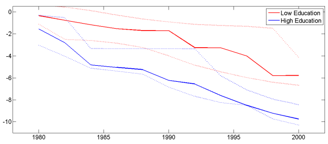
Figure 37: Taste Wedge



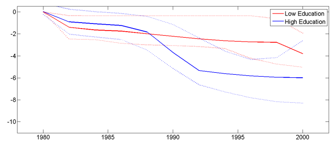
(a) Moderate Smokers - Low Alcohol



(b) Moderate Smokers - High Alcohol



(c) Heavy Smokers - Low Alcohol



(d) Heavy Smokers - High Alcohol

e

Taste shifters are only estimated for moderate and heavy smokers that are drawn from our education-alcohol cohorts. Estimating comparable taste shift parameters for all quantile, education and alcohol groups yields a quadratic programme with over 17,000 constraints. This is an exceptionally computationally demanding procedure to implement. In addition, the requirement of perfect intertemporal variability of good-1 is violated for demands on the "high education"-"light smoking"-"light drinking" cohort; for most of the period, the estimated budget share on tobacco for this cohort is zero. We deal with this by deciding not to calculate taste changes for light smokers. This ensures that there will always exist a set of comparable, choice-rationalising taste shifters that can be estimated in a reasonable amount of time. However, please note that the fact that light smokers are not included at this stage implies that the magnitude of estimated taste shifters are not comparable to the unconditional quantile results of the previous section.

RESULTS The finding of a statistically significant socioeconomic difference in the evolution of virtual prices for light and moderate smokers is robust to the conditioning of demands upon alcohol consumption. 95% confidence intervals on virtual prices and the taste wedge are disjoint across education groups for all cohorts except for the "heavy smoking"-"heavy drinking" group. Effective tastes for this group evolved very little for both education groups, as evidenced by the fact that their virtual price trajectories closely follow the observed relative price path. This finding is consistent with government and health practitioner reports that note low smoking cessation rates amongst heavy drinkers (Dollar *et al.*, 2009).

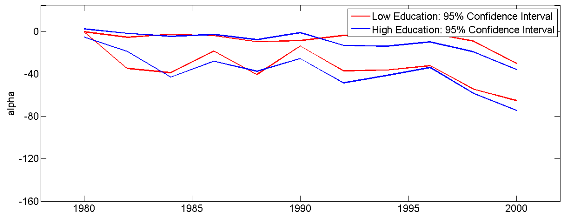
Our empirical application has revealed significant intertemporal and intercohort variation in the marginal willingness to pay for tobacco. However, heterogeneity in the marginal willingness to pay derives from two sources: taste variation, which is reflected in differences in $\alpha_t^{c,\epsilon}$, and budget variation, which is reflected in differences in the marginal utility of income across education groups, $\lambda_t^{c,\epsilon}$. Virtual prices are constructed as $\widehat{\pi}_t^{c,\epsilon,1} = p_t^1 - (\widehat{\alpha}_t^{c,\epsilon}/\widehat{\lambda}_t^{c,\epsilon})$ along each SMP expenditure path. Tracking changes in $\alpha_t^{c,\epsilon}$ enables one to make global statements concerning relative tastes for tobacco that are independent of variation in the marginal utility of income along the different SMP expenditure paths.

Variation in the marginal utility of income or income effects on tobacco cannot form part of the story constructed to rationalise differences in choice behaviour at different quantiles of the tobacco distribution because, in this context, tastes are recovered in the same region of the preference map. This is so because budget parameters are identical for each quantile of the tobacco distribution within cohorts. Therefore, the recovered patterns that concern relative tastes within education groups are unchanged; heavier smokers have a greater global taste for tobacco than lighter smokers.

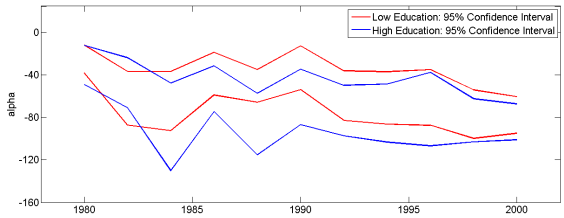
However, the divergence in recovered taste shifters across education groups is not as stark as one would imagine given the significant differences in the virtual price trajectories by education group that were explored previously. This is highlighted by Figure 38, which depicts the 95% confidence interval on the trajectory of the raw taste shifters. These intervals overlap. Total expenditure along the SMP path is greater for more highly educated groups. Therefore, tastes are recovered in different regions of the preference map for low and high education cohorts. Part of the divergence in the marginal willingness to pay across education groups can be rationalised by income effects; one cannot reject the hypothesis that less educated people would consume the same quantity of tobacco as more highly educated people if they were endowed with higher total expenditure. Thus, although educational differences in raw taste shifters exist, these are not as striking as the differences in effective tastes that we recovered previously and these differences are not statistically significant. Without additional assumptions on either the structure of the base utility function or on the movement in the marginal utility of income over time, we are unable to associate the recovered differences in effective tastes with statistically significant educational differences in the parameters of the base utility function. We explore this further in the next section.

SUMMARY This section has applied our quadratic programming methodology to repeated cross section data in the U.K. in order to

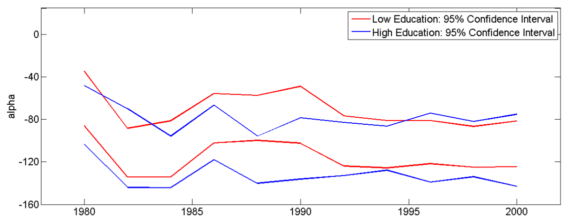
Figure 38: Raw Taste Differences



(a) Heavy Smokers



(b) Moderate Smokers



(c) Light Smokers

rationalise the changes in tobacco consumption that have occurred since 1980. We uncovered significant differences in the effective taste for tobacco across the budget share for tobacco distribution and also between education levels for light and moderate smoking cohorts. We extended our analysis to explore the impact of controlling for alcohol consumption and recovered taste parameters for below-median and above-median alcohol consumption groups. In this context, we found significant educational differences in effective tastes for all but the "heavy smoking"-"heavy drinking" cohorts. For this group, effective taste trajectories did not differ significantly between low and high education cohorts.

However, the recovered differences in effective tastes could largely be rationalised by income effects. Therefore, from within our current setting, we were unable to conclude that there exist significant educational differences in the parameters indexing the direct utility function. The next section places taste change in a framework that restricts the intertemporal evolution of λ_t. This allows us to explore what consequences restrictions on the trajectory of the marginal utility of income might have upon the strength of our conclusions over raw taste differences.

26

RESTRICTING λ

The previous section uncovered statistically significant educational differences in the effective taste for tobacco. However, despite this fact, we were unable to conclude that there exist significant educational differences in the raw parameters of the direct utility function because divergences in the marginal willingness to pay could be partially rationalised by income effects. In this section, we extend our theoretical framework to restrict the trajectory of the marginal utility of income, λ_t over time. Specifically, we place taste change within the Strong Rational Expectations lifecycle model. This allows us to explore whether additional restrictions upon the evolution of λ_t could yield stronger conclusions concerning educational differences in the parameterisation of the direct utility function.

We note here that the results in this section are underpinned by a number of strong assumptions that are required for us to easily apply a revealed preference methodology to an intertemporal setting. This section should be read as an *extension* of this essay's core theoretical framework. The aim of the following analysis is to highlight that additional restrictions on the trajectory of λ_t allow one to make stronger conclusions concerning educational differences in raw taste parameters. We do not thoroughly examine the impact of weakening the assumptions of perfect foresight and perfect capital markets on our empirical results given this exploratory motivation. This in contrast to the previous essay in this thesis, in which the core model of interest was dynamic and the robustness of our empirical results to the failure of strong background assumptions was examined.

26.1 LIFECYCLE MODEL

Extending our theoretical framework to an intertemporal setting gives rise to the following model for individual i's decision making problem:

$$\max_{\{\mathbf{q}_t^i\}_{t=1,\ldots,T}} \sum_{t=1}^{T} v^i(\mathbf{q}_t^i) + \alpha_t^i q_t^{1i} \tag{4.35}$$

subject to:

$$\sum_{t=1}^{T} \boldsymbol{\rho}_t' \mathbf{q}_t^i = A_0^i \tag{4.36}$$

where $\boldsymbol{\rho}_t = \mathbf{p}_t / \prod_{s=1}^{t}(1+i_s)$, i_t is the nominal interest rate at t and A_0^i represents discounted lifetime wealth.

This model extends the Strong Rational Expectations Lifecycle Hypothesis to incorporate taste instability on good-1. The Strong Rational Expectations Lifecycle Hypothesis was given a revealed preference treatment by Browning (1989) and is analogous to the model considered in my previous essay, in which the data consistency of intertemporal models of household choice was addressed. In contrast to the static model considered up until this point in this essay, the lifecycle framework restricts the evolution of λ_t over time. This reduces the extent to which changes in tobacco consumption can be rationalised by income effects. Specifically, optimising behaviour within this theoretical framework requires that the marginal utility of discounted income is constant over time. If this condition was violated, expenditure could be transferred from periods in which the marginal utility of income was low to those in which it was high, and overall lifetime utility would be increased.

The lifecycle model provides an alternative theoretical framework through which we can examine the impact that restrictions upon λ_t might have upon the strength of our conclusions over the significance of educational differences in minimal taste parameters. The restrictions placed on λ_t are exceptionally strong and so results in this context could be seen as sitting at the other extreme to those recovered in the previous section.

RECOVERING LIFECYCLE TASTES To assess the data consistency of the lifecycle extension of our theoretical model, we modify the CBU Quadratic Programme to impose constancy of each cohort's marginal utility of discounted income. Further, we weaken the assumptions that were made regarding the commonality of the base utility function. We experimented with imposing commonality of the base utility function across smoking groups, as in the previous section, but encountered convergence issues when doing so. Convergence was not found to cause difficulties when commonality was only imposed within smoking cohorts so we proceed using this strategy. Thus, in this section, I only require commonality of the base utility to hold across education groups at the same smoking quantiles, i.e. "low education"-"heavy smokers" and "high education"-"heavy smokers" are restricted to have the same base utility function but the base utility function of "low education"-"heavy smokers" and "low education"-"moderate smokers" can differ. Sadly, this modification implies that we cannot make conclusions about the magnitude of taste differences across quantiles of the smoking distribution.

The modified lifecycle quadratic programme enables us to recover the set of minimal taste shifters for each smoking group $\{\alpha_t^e\}_{t=1,\dots,T}$ that are necessary for observed choice behaviour to be rationalised by the lifecycle model with a common base utility function across education groups at each quantile of the smoking distribution. Appendix

C provides a further discussion, and proof, of the lifecycle rationalisation conditions with taste change over good-1. Further results are also given concerning the rejectability of these conditions. Given that these results are not central to the main argument in this section, I here only outline the quadratic programming procedure that is implemented to recover minimal taste shifters.

CBU LIFECYCLE QUADRATIC PROGRAMME The minimal taste perturbations on good 1, relative to individual 1's tastes in period 1, that are necessary to lifecycle rationalise observed choice behaviour $\{\boldsymbol{\rho}_t, \mathbf{q}_t^{c,\epsilon}\}_{t=1,\dots,T}^{i=1,\dots,N}$, given commonality of the base utility function across individuals at the same quantile of their psuedo cohort's smoking distribution, are identified as the unique set $\{\alpha_t^{c,\epsilon}\}_{t=1,\dots,T}^{i=1,\dots,N}$ that is yielded by the following quadratic programming procedure.

$$\min_{\mathbf{v},\boldsymbol{\lambda},\boldsymbol{\alpha}} \sum_{c} \sum_{\epsilon} \sum_{t} (\alpha_t^{c,\epsilon})^2$$

subject to:

1. The rationalisation constraints:

$$\nu_s^{c,\epsilon} - \nu_t^{d,\tau} + \alpha_t^{d,\epsilon}(q_s^{1\,c,\epsilon} - q_t^{1\,d,\epsilon}) \leqslant \lambda^{d,\tau} \boldsymbol{\rho}_t'(\mathbf{q}_s^{c,\epsilon} - \mathbf{q}_t^{d,\epsilon})$$
$$\lambda^{c,\epsilon}\rho_t^1 - \alpha_t^{c,\epsilon} \geqslant 0$$
$$\lambda^{c,\epsilon} \geqslant 0$$

   for all:

   - $s, t \in \{1980, 1982, ..., 1998, 2000\}$
   - $c, d \in \{L, H\}$
   - $\epsilon \in \{0.55, 0.65, 0.75\}$

2. The normalisation conditions:

$$\nu_{1980}^{L,\epsilon} = 0$$
$$\lambda_{1980}^{L,\epsilon} = 10$$
$$\alpha_{1980}^{L,\epsilon} = 0$$

   for $\epsilon \in \{0.55, 0.65, 0.75\}$.

In Appendix C it is proven that an allowance for taste change on good-1 is not necessarily sufficient to rationalise a data set within the lifecycle framework. This is an important point to note. Unlike Theorem 4.1, the rationalisation conditions that are associated with the lifecycle model are potentially rejectable, implying that the CBU Lifecycle Quadratic Programme may be associated with an empty feasible set. Innovations in α_t^i are not always sufficient to rationalise a data set and innovations in λ^i must also be allowed for in order to rationalise patterns in the allocation of total expenditure over time. This result should be regarded positively in our context because taste shifters on good-1 are the only tool that we allow to rationalise intercohort and intertemporal spending deviations from the benchmark lifecycle model. Thus, the recovered value of $\alpha_t^{c,\epsilon}$ will be set to partially offset unexpected income growth and other shocks, in addition to reflecting taste evolution. The fact that taste shocks are not sufficient for rationalisation in the lifecycle context implies that there is a limit to which α_t^i can compensate for intertemporal inconsistencies, which makes us better able to defend the claim that its value reflects taste change.

We apply the CBU Lifecycle Quadratic Programme to psuedo cohort lifecycle quantity sequences drawn from the FES. Rather than recover cohort demands along the SMP expenditure path, this section makes use of cohort demands that are estimated at the median total expenditure level in each period as the input to our programming procedure. This is because total expenditure in each period is a choice variable in the lifecycle model. Therefore, the SMP methodology cannot be applied. The same censored quantile regression methodology that was outlined in Section 24 is applied to recover this set of cohort lifecycle demands:

$$\mathbf{q}_t^{c,\epsilon} = (\mathbf{p}_t'/x_t^c)\mathbf{s}_t^{c,\epsilon} \tag{4.37}$$

where $\mathbf{s}_t^{c,\epsilon}$ is the vector of budget shares demanded at the ϵ-quantile of cohort c's budget share for tobacco distribution given the budget defined by price regime t and median total expenditure. Formally,

$$\begin{aligned} \mathbf{s}_t^{c,\epsilon} &= [max(0, w_t^{\star c,\epsilon}), min(1, 1 - w_t^{\star c,\epsilon})] \\ w_t^{\star c,\epsilon} &= Q_{w^{\star c,\epsilon}|\mathbf{p}_t}(\epsilon|\widetilde{x}_t^c, z = 0.67, v = 0.5, c) \\ \widetilde{x}_t^c &= Q_x(0.5|c) \end{aligned} \tag{4.38}$$

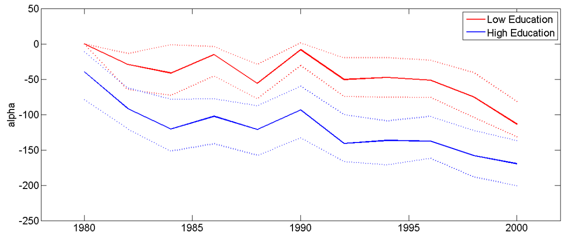
for $t = 1, ..., T$, $c = \{L, H\}$, $\epsilon = \{0.55, 0.65, 0.75\}$.

The CBU Lifecycle Quadratic Programme is parameterised by the set of discounted prices $\{\boldsymbol{\rho}_t\}_{t=1,...,T}$, as opposed to the set of observed prices that we used previously $\{\mathbf{p}_t\}$. We discount observed prices using the Bank of England base rate to construct $\boldsymbol{\rho}_t$.

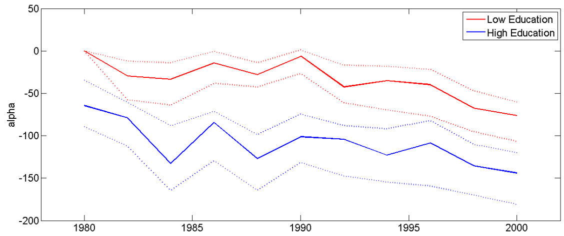
26.2 LIFECYCLE TASTE SHIFTERS

Applying the CBU Lifecycle Quadratic Programme to cohort lifecycle demands recovers statistically significant educational differences in raw tastes. When rationalisation is defined from within the lifecycle model, we observe distinct differences in the evolution of tastes across high and low education cohorts, with lower educated cohorts displaying a greater global taste for tobacco. This implies that less educated individuals would consume more tobacco than their more highly educated peers, even if restricted to the same total expenditure path.

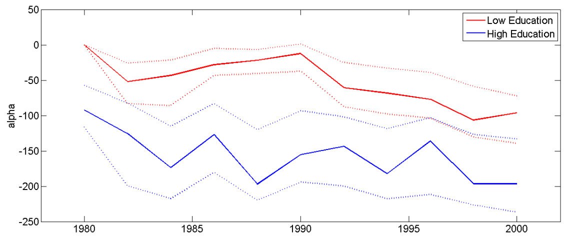
Figure 39: Raw Lifecycle Taste Differences



(a) Heavy Smokers



(b) Moderate Smokers



(c) Light Smokers

Although some of the recovered differences in taste shifters are likely to reflect differences in unexpected income growth between cohorts, α alone cannot fully compensate for shocks of this nature, making us more confident that tastes, and not simply variation in total expenditure trajectories, differ between cohorts.

SUMMARY This section extended our theoretical framework to restrict the trajectory of the marginal utility income over time. Our primary empirical results, in which taste change was situated in a static model, uncovered statistically significant differences in effective tastes but we could not reject that these taste differences arose from income effects rather than differences in the parameters that indexed the direct utility function of low and high education groups. In this section, we found that situating taste change in a model that restricts this income process enables one to make stronger conclusions regarding educational differences in raw tastes. From the perspective afforded by the lifecycle model, less educated individuals have a globally greater taste for tobacco and thus would consume more tobacco that their more highly educated peers even if they were restricted to the same total expenditure path. This model can be considered as an extreme specification over the trajectory of λ_t. These results highlight that additional restrictions on the process governing the marginal utility of income allow us to make stronger conclusions regarding taste differences across education groups rather than introducing further ambiguity to the story.

27 CONCLUSION

This essay addressed the key questions of testability and recoverability in a setting where observed behaviour could not be rationalised by the basic utility maximisation hypothesis and then developed a quadratic programming problem to compute the lower bound on the interpersonal and intertemporal taste variation that is required to rationalise a finite set of observations on past choice behaviour. We applied this methodology to explain patterns in U.K. tobacco consumption and found that taste change is a necessary component of any rationalisation of observed trends. In addition, we uncovered evidence of significant educational differences in the "effective" and "raw" taste trajectories for tobacco; our results give strong reason to believe that less educated groups have a greater global taste for tobacco relative to their more highly educated peers.

To engage in this line of research, we developed a simple theoretical framework with which to characterise taste change. Specifically, we allowed for an additive linear perturbation to the marginal utility of tobacco. Even with this relatively restrictive functional specification, we proved a surprising non-identification result: observational data sets on a K-dimensional demand system can always be rationalised by taste change on a single good in a nonparametric setting.[1] A corollary of this result is that observed choice heterogeneity can be parsimoniously summarised by a univariate parameter in the direct utility function.

Our theoretical results were used to develop a quadratic programming procedure with which to recover the minimal intertemporal and interpersonal taste heterogeneity that is required to rationalise observed choices. This method facilitated the construction of a lower bound on the changes in the marginal utility of, and the willingness to pay for, tobacco over time and across different pseudo cohorts that we constructed from the U.K. Family Expenditure Survey. The power of our methodology was maximised by integrating it with the Sequential Maximum Power technique of Blundell *et al.* (2003).

Taste variation was required to rationalise the behaviour of all cohorts. A series of strictly negative perturbations to the marginal utility of tobacco were found to be sufficient to rationalise the trends in tobacco consumption. We further recovered statistically significant educational differences in the marginal willingness to pay for tobacco; more highly educated cohorts experienced a greater shift in

1 Where one defines an "observational data set" as a data sets that includes at least one good that satisfies the requirement of perfect intertemporal variation.

their effective tastes for tobacco and thus faced a significantly higher taste-tax for tobacco than did less educated cohorts. However, in our nonparametric setting, the variation in effective tastes could be partially rationalised by income effects because total expenditure trajectories differ across education cohorts. However, a lifecycle extension to our basic model, which restricted the intertemporal evolution of the marginal utility of income, yielded stronger conclusions regarding educational differences in raw tastes. From the perspective afforded by the lifecycle model, less educated individuals have a greater global taste for tobacco than their more highly educated peers. This implies that less educated people would consume more tobacco even if they enjoyed higher total expenditure.

If there is a desire for the observed socioeconomic trends in tobacco consumption to converge, then our results suggest that government education programmes must be designed to be more salient amongst less educated groups. Even if the recovered differences in effective tastes can be largely rationalised by income effects, our empirical results still imply that that the taste parameters that index the direct utility function must change by *more* for low education groups if an anti-smoking campaign is to affect a uniform taste tax across education groups. Our policy conclusions are stronger from the context of the lifecycle framework. From the context of this model, both effective and raw taste changes have been much more salient amongst higher education groups. This essay calls for additional research into why this is the case.

Part V

CLOSING STATEMENT

28 THESIS CONCLUSION

This thesis comprises three principal essays, each of which provides an original contribution to the literature on the nonparametric approach to demand analysis. I have provided a set of methodologies that facilitate one to test a number of economic models of choice behaviour, and to recover salient features of individual preferences, in a manner that is largely free of auxiliary assumptions over the form of the unobserved structural functions of interest. These novel techniques draw upon results from finite mathematics and rely heavily upon linear programming methods for their operationalisation. Their application to consumer microdata has yielded empirical insights of relevance to the applied literatures on time-discounting, family economics and the public policy debate on tobacco control.

My first essay developed the MDI methodology, which facilitated the recovery of globally rational demand predictions at sets of budget hyperplanes. In the empirical section of this essay, I provided an internally consistent application of the MDI methodology and demonstrated how the recovered globally rational demand predictions can be used for welfare analysis. A key outstanding question of interest is how the accuracy of the proposed technique compares to alternative parametric and nonparametric methodologies. In the future, I will conduct this comparative analysis in order to assess the benefits of the MDI approach over parametric and nonparametric alternatives. This is a key outstanding question of interest. Further, I would like to explore how the KL-objective function can be applied to provide a specification test of sorts for different economic theories. The divergence has a probabilistic interpretation. Therefore, it is interesting to consider how it may be integrated with the revealed preference approach to assessing the consistency of economic theories.

My second essay proved new testability axioms for models of collective intertemporal choice and developed them into a practical algorithm that facilitates the recovery of the minimal intrahousehold heterogeneity in discount rates that is necessary to rationalise observed choice patterns. This essay provided empirical results that support the use of exponential discounting and perfect intrahousehold commitment as valid positivist modelling devices for short-to-medium run consumption choices. In the future, I would like to apply this methodology to different data sets to examine the robustness of this result and to yield more refined insights concerning the characteristics of intrahousheold discount rate heterogeneity and how patterns of renegotiation relate to variation in the outside options of family members.

My final essay incorporated taste instability into the revealed preference characterisation of the rational choice framework and proved that, in our nonparametric setting, any data set that satisfies that mild requirement of "perfect variation" can be rationalised once one allows for taste variation on a single good. A corollary of this result is that observed choice heterogeneity can be parsimoniously summarised by a univariate parameter in the direct utility function. In the future, I would like to examine, in greater detail, the implications that this result has for the representation of unobserved heterogeneity and also develop tests on the pattern of recovered taste shifters that would allow one to distinguish systematic from random deviations in an individual's behaviour from a model's prescriptions.

Part VI

APPENDIX

A

GLOBALLY RATIONAL NONPARAMETRIC DEMAND FORECASTS

This Appendix provides the proofs that are referenced in the chapter "Globally Rational Nonparametric Demand Forecasts".

PROPOSITION 2.1 S^V_b is a closed, convex subset of the set of demands that exhaust a budget $\{\boldsymbol{\rho}_b, m_b\}$.

(Closed) A set is closed if its complement is open, i.e. that the set contains its limit points. Let $Q(\boldsymbol{\rho}_b, m_b)$ represent the set of feasible demands at a new budget $\{\boldsymbol{\rho}_b, m_b\}$ and $S^c_b = Q(\boldsymbol{\rho}_b, m_b)\backslash S^V_b$. Then, for any $\widetilde{\mathbf{q}} \in S^c_b$, $\mathbf{p}'_w\widetilde{\mathbf{q}}<x_w$ for all $w \in RW(\boldsymbol{\rho}_b, m_b)$.[1] Given that this inequality is strict, S^c_b is open: for every element $\mathbf{q} \in S^c_b$, one can define a radius ε_q about it such that $B_q(\varepsilon_q) = \{\widetilde{\mathbf{q}} \in S^c_b : d(\widetilde{\mathbf{q}}, \mathbf{q}) < \varepsilon_q)$ $\subset S^c_b$. Therefore, S^V_b is closed.

(Convex) Consider any two bundles contained by S^V_b, $\mathbf{q}_1$ and $\mathbf{q}_2$. We have that,

$$\begin{array}{lll} [1] & \mathbf{q}_1 \geqslant \mathbf{0} & \mathbf{q}_2 \geqslant \mathbf{0} \\ [2] & \boldsymbol{\rho}'_b\mathbf{q}_1 = m_b & \boldsymbol{\rho}'_b\mathbf{q}_2 = m_b \\ [3] & \mathbf{p}'_w\mathbf{q}_1 \geqslant x_w & \mathbf{p}'_w\mathbf{q}_2 \geqslant x_w \end{array}$$

for $w \in RW(\mathbf{p}_0, x_0)$.

Any convex combination of $\mathbf{q}_1$ and $\mathbf{q}_2$, $\mathbf{q}_3 = t\mathbf{q}_1 + (1-t)\mathbf{q}_2$ for $0 \leqslant t \leqslant 1$, will also constitute an element of S^V_b:

$$\begin{array}{llll} [1] & \mathbf{q}_3 & = & t\mathbf{q}_1 + (1-t)\mathbf{q}_2 \geqslant 0 \\ [2] & \boldsymbol{\rho}'_b\mathbf{q}_3 & = & \boldsymbol{\rho}'_b(t\mathbf{q}_1 + (1-t)\mathbf{q}_2) \\ & & = & tm_b + (1-t)m_b \\ & & = & m_b \\ [3] & \mathbf{p}'_w\mathbf{q}_3 & = & \mathbf{p}'_w(t\mathbf{q}_1 + (1-t)\mathbf{q}_2) \\ & & \geqslant & x_w \end{array}$$

Thus, S^V_b is closed and convex.

PROPOSITION 2.2 Let $\{\boldsymbol{\omega}^{V\star}_b\}_{b=1,...,|B|}$ represent the solution set yielded by imposing membership of the Varian support set S^V_b at each budget $b = 1, ..., |B|$. Let $\{\boldsymbol{\omega}^{S\star}_b\}_{b=1,...,|B|}$ represent the solution set when a

1 The revealed worse set, $RW(\boldsymbol{\rho}_b, m_b)$, was formally defined by Definition 2.5.

single maximisation problem is solved to impose membership of the sufficient support set, S^S.

$$\text{If } \{\mathbf{A}_b \boldsymbol{\omega}_b^{V\star}\}_{b=1,\dots,|B|} \in S^S, \text{ then } \boldsymbol{\omega}_b^{V\star} = \boldsymbol{\omega}_b^{S\star} \text{ for } b = 1, \dots, |B|$$

(Proof) If the solution gained from independently imposing rationality at each new budget satisfies cross-consistency, then imposing cross-consistency explicitly has no additional impact on estimates. Let $S^V(B) = \cup_{i=1,\dots,|B|} S_b^V$ and consider a set of pilot predictions $\{\boldsymbol{\eta}_b\}_{i=1,\dots,|B|} \notin S^V(B)$.

If Proposition 2.2 did not hold, then one could find a solution $\{\boldsymbol{\omega}_b^S\}_{b=1,\dots,|B|} \neq \{\boldsymbol{\omega}_b^{V\star}\}_{b=1,\dots,|B|}$ when solving the full maximisation problem such that

$$\sum_{b=1}^{|B|} D(\boldsymbol{\omega}_b^{S\star} \| \boldsymbol{\eta}_b) \leqslant \sum_{b=1}^{|B|} D(\boldsymbol{\omega}_b^{V\star} \| \boldsymbol{\eta}_b)$$

Given that $\boldsymbol{\omega}_b^{V\star}$ uniquely minimises $D(\boldsymbol{\omega}_b \| \boldsymbol{\eta}_b)$ at each budget b, we have:

$$\sum_{b=1}^{|B|} D(\boldsymbol{\omega}_b^{V\star} \| \boldsymbol{\eta}_b) < \sum_{b=1}^{|B|} D(\boldsymbol{\omega}_b \| \boldsymbol{\eta}_b)$$

for all $\boldsymbol{\omega}_b \in S^V(B)$ and $\boldsymbol{\omega}_b \neq \boldsymbol{\omega}_b^{V\star}$ at $b = 1, \dots, |B|$.

Given that $S^S \subseteq S^V(B)$, if $\{\boldsymbol{\omega}_b\}_{b=1,\dots,|B|} \in S^S$, then $\{\boldsymbol{\omega}_b\}_{b=1,\dots,|B|} \in S^V(B)$. Thus,

$$\sum_{b=1}^{|B|} D(\boldsymbol{\omega}_b^{V\star} \| \boldsymbol{\eta}_b) < \sum_{b=1}^{|B|} D(\boldsymbol{\omega}_b^S \| \boldsymbol{\eta}_b)$$

for all $\{\boldsymbol{\omega}_b\}_{b=1,\dots,|B|} \in S^V(B)$ and $\{\boldsymbol{\omega}_b\}_{b=1,\dots,|B|} \neq \{\boldsymbol{\omega}_b^{V\star}\}_{b=1,\dots,|B|}$. Proposition 2.2 is thus proved by contradiction.

PROPOSITION 2.3 Let B^{M_k} define a set of budgets that represent a portion of the marginal demand curve for good k, i.e. for all $B_b, B_{b'} \in B^{M_k}$, we have that $\rho_b^k \neq \rho_{b'}^k$, $\boldsymbol{\rho}_b^{\neg k} = \boldsymbol{\rho}_{b'}^{\neg k}$ and $m_b = m_{b'}$. The solution yielded by imposing membership of the Varian support set S_b^V at each budget $b = 1, \dots, |B^{M_k}|$ and the solution set yielded when a single maximisation problem is solved to impose membership of the sufficient support set, $S^S(B^{M_k}|D)$ $\{\boldsymbol{\omega}_b^{S\star}\}_{b=1,\dots,|B^{M_k}|}$ are identical.

$$\boldsymbol{\omega}_b^{V\star} = \boldsymbol{\omega}_b^{S\star} \quad \text{for any} \quad B_b \in B^{M_k}$$

(Proof) Let $\{\boldsymbol{\rho}_b, m\}$ and $\{\boldsymbol{\rho}_{b'}, m\}$ represent two budgets along the marginal demand curve for good k, where $\rho_b^k > \rho_{b'}^k$. Let $\mathbf{x}_b^{V\star}$ and $\mathbf{x}_{b'}^{V\star}$ represent the MDI solutions gained by independently imposing membership of the Varian support sets. The sets $\{D; \boldsymbol{\rho}_b, \mathbf{x}_b^{V\star}\}$ and $\{D; \boldsymbol{\rho}_{b'}, \mathbf{x}_{b'}^{V\star}\}$ satisfy GARP. Let us consider interior solutions of good-k without loss of generality.

As $\boldsymbol{\rho}_b' \mathbf{x}_{b'}^{k,V\star} > m$ and $\boldsymbol{\rho}_{b'}' \mathbf{x}_b^{k,V\star} < m$, we have that:

$$\begin{array}{l} \mathbf{x}_{b'}^{k,V\star} \mathbb{R}^0 \mathbf{x}_b^{k,V\star} \\ \neg \left(\mathbf{x}_{b'}^{k,V\star} \mathbb{R}^0 \mathbf{x}_b^{k,V\star} \right) \end{array}$$

For a violation of cross-consistency to occur, there must exist an observation $\{\mathbf{p}_t, \mathbf{q}_t\} \in D$ such that:

$$\begin{array}{ll} [1] & \mathbf{q}_t \mathbb{R}^0 \mathbf{x}_{b'}^{k,V\star} \\ [2] & \mathbf{x}_b^{k,V\star} \mathbb{R}^0 \mathbf{q}_t \end{array}$$

This would induce a violation of transitivity because these relations together imply $\mathbf{x}_b^{k,V\star} \mathbb{R}^0 \mathbf{x}_{b'}^{k,V\star}$. Imagine that this were possible. Then, we would have:

$$\begin{array}{llll} [1.1] & y_t & \geqslant & \mathbf{p}_t' \mathbf{x}_{b'}^{k,V\star} \\ [1.2] & m & < & \boldsymbol{\rho}_{b'}' \mathbf{q}_t \\ [2.1] & m & \geqslant & \boldsymbol{\rho}_b' \mathbf{q}_t \\ [2.2] & y_t & < & \mathbf{p}_t' \mathbf{x}_b^{k,V\star} \end{array}$$

This is not possible. From [1.2],

$$\begin{aligned} m &< \boldsymbol{\rho}_{b'}' \mathbf{q}_t \\ &= \rho_{b'}^k q_t^k + \boldsymbol{\rho}_{b'}^{\neg k\prime} \mathbf{q}_t^{\neg k} \end{aligned}$$

From [2.1],

$$\begin{aligned} m &\geqslant \boldsymbol{\rho}_b' \mathbf{q}_t \\ &= \rho_b^k q_t^k + \boldsymbol{\rho}_b^{\neg k\prime} \mathbf{q}_t^{\neg k} \end{aligned}$$

Thus,

$$\begin{aligned} m - \boldsymbol{\rho}_b^{\neg k\prime} \mathbf{q}_t^{\neg k} &< \rho_{b'}^k q_t^k \\ m - \boldsymbol{\rho}_b^{\neg k\prime} \mathbf{q}_t^{\neg k} &\geqslant \rho_b^k q_t^k \end{aligned}$$

However, given that $\rho_b^k > \rho_{b'}^k$, $\rho_b^k q_t^k > \rho_{b'}^k q_t^k$. This contradicts [1.2] and [2.1] simultaneously holding. Proposition 2.3 is thus proved by contradiction.

THEOREM 2.1 Let $\mathbf{A}_b\boldsymbol{\eta} \notin S^V(\boldsymbol{\rho}_b, m_b)$. $\Updownarrow = \{\kappa_1, ..., \kappa_n\}$ is a sample of n expenditure-to-good allocations randomly drawn with replacement by the statistician at the budget $\{\boldsymbol{\rho}_b, m_b\}$. This sample generates the sample budget share $\boldsymbol{\omega}$, with $\omega^k = \frac{1}{n}\sum_n I(\kappa_n = k)$. The statistician believes that the underlying budget share describing demand is $\boldsymbol{\eta}$. Then, the epistemic probability assigned to the realisation of any rational budget share other than the MDI solution, $\boldsymbol{\omega}^\star$, is negligible as $n \to \infty$. Formally,

$$\widetilde{Pr}(|\omega^k - \omega^{k^\star}| \geqslant \epsilon) \to 0$$

as $n \to \infty$, where

$$\boldsymbol{\omega}^\star = \arg \min_{\mathbf{A}_b\boldsymbol{\omega} \in S_b^V} D(\boldsymbol{\omega} \| \boldsymbol{\eta})$$

(Proof) This is an application of the Conditional Limit Theorem (Csiszár, 1984), a strengthening of Sanov's Theorem (Sanov, 1957). See Cover and Thomas (1991) for a thorough treatment.

PROPOSITION 2.4 The expected forecasting loss that the statistician reporting budget share $\boldsymbol{\eta}$ can expect when the true budget share is given by $\boldsymbol{\omega}$ is given by:

$$E[L(\boldsymbol{\omega}, \boldsymbol{\eta})] = m_b \left\{ \boldsymbol{\omega}' \log_2 \frac{\boldsymbol{\omega}}{\boldsymbol{\eta}} \right\}$$

(Proof) The statistician gives $\boldsymbol{\eta}$ as her prediction for how an individual's budget will be split between commodities, or alternatively, as her prediction for which good a marginal spending decision will be devoted to. Imagine that the true distribution is given by $\boldsymbol{\omega}$. Then, the expected loss, given a logarithmic scoring rule, on any randomly drawn spending decision n that is drawn from a consumer's realised demand is represented by:

$$E[L(n|\boldsymbol{\omega}, \boldsymbol{\eta})] = \sum_{k=1}^{K} \omega_k \left(\log_2 \omega_k - \log_2 \eta_k \right)$$

Integrating over m_b randomly drawn spending decisions from a consumer's realised demand returns the expected loss for the statistician.

$$\begin{aligned} E[L(\boldsymbol{\omega}, \boldsymbol{\eta})] &= m_b \textstyle\sum_{k=1}^{K} \omega_k \left(\log_2 \omega_k - \log_2 \eta_k \right) \\ &= m_b \left\{ \boldsymbol{\omega}' \log_2 \frac{\boldsymbol{\omega}}{\boldsymbol{\eta}} \right\} \end{aligned}$$

PROPERTY 1 The updated solution satisfies $\boldsymbol{\omega}^\star = \boldsymbol{\eta}$ if and only if $\boldsymbol{\eta} \in S_b^V$. In this instance, $D(\boldsymbol{\omega}^\star||\boldsymbol{\eta}) = 0$.

(Proof) This follows from the fact that $D(\boldsymbol{\omega}||\boldsymbol{\eta}) \geqslant 0$ and $D(\boldsymbol{\omega}||\boldsymbol{\eta}) = 0$ if and only if $\boldsymbol{\omega}$=$\boldsymbol{\eta}$.

PROPERTY 2 The updated solution $\boldsymbol{\omega}^\star$ is unique.

(Proof) Let $\boldsymbol{\alpha}, \boldsymbol{\beta} \in S_b^V$ represent budget shares with the same KL-divergence from the pilot $\boldsymbol{\eta}$: $D(\boldsymbol{\alpha}||\boldsymbol{\eta}) = D(\boldsymbol{\beta}||\boldsymbol{\eta})$. Given that $D(\boldsymbol{\omega}||\boldsymbol{\eta})$ is convex, then

$$\begin{aligned} D(\boldsymbol{\alpha}||\boldsymbol{\eta}) &= D(\boldsymbol{\beta}||\boldsymbol{\eta}) \\ &= tD(\boldsymbol{\alpha}||\boldsymbol{\eta}) + (1-t)D(\boldsymbol{\beta}||\boldsymbol{\eta}) \\ &\geqslant D(t\boldsymbol{\alpha} + (1-t)\boldsymbol{\beta}||\boldsymbol{\eta}) \end{aligned}$$

This inequality is strict unless $\boldsymbol{\alpha}$=$\boldsymbol{\beta}$. Therefore, if $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$ and $D(\boldsymbol{\alpha}||\boldsymbol{\eta}) = D(\boldsymbol{\beta}||\boldsymbol{\eta})$, then there exists a budget share $t\boldsymbol{\alpha} + (1-t)\boldsymbol{\beta}$ that belongs to S_b^V (given that S_b^V is convex) and has a smaller KL-divergence than $D(\boldsymbol{\alpha}||\boldsymbol{\eta})$.

PROPERTY 3 The MDI approach treats goods symmetrically.

(Proof) The solution is invariant to permutations of the ordering of the goods. Imagine that this was not the case. We would then have that

$$D(\omega_1, .., \omega_n, ..., \omega_K||\eta_1, ..., \eta_n, ..., \eta_K) \neq D(\omega_n, .., \omega_1, ..., \omega_K||\eta_n, ..., \eta_1, ..., \eta_K)$$

However, this is contradictory because the structure of the objective function is preserved by permutations of the ordering of goods.

$$\begin{aligned} D(\boldsymbol{\omega}, \boldsymbol{\eta}) &= D(\omega_1, .., \omega_n, ..., \omega_K||\eta_1, ..., \eta_n, ..., \eta_K) \\ &= \omega_1 \log_2 \tfrac{\omega_1}{\eta_1} + ... + \omega_n \log_2 \tfrac{\omega_n}{\eta_n} + ... + \omega_K \log_2 \tfrac{\omega_K}{\eta_K} \\ &= \omega_n \log_2 \tfrac{\omega_n}{\eta_n} + ... + \omega_1 \log_2 \tfrac{\omega_1}{\eta_1} + ... + \omega_K \log_2 \tfrac{\omega_K}{\eta_K} \\ &= D(\omega_n, .., \omega_1, ..., \omega_K||\eta_n, ..., \eta_1, ..., \eta_K) \end{aligned}$$

PROPERTY 4 The MDI approach preserves weak separability of preferences in the updating procedure.

(Proof) This is an application of the "Subset Independence" property of cross-entropy minimisation. See Property 7 of Shore and Johnson (1981).

Consider a decomposition of the good space into $|G|$ disjoint subsets that correspond to weakly separable commodity subgroups, $\cup_{g=1,..,|G|} G_g =$

$\{1,...K\}$. For some subgroup G_g and $k \in G_g$, let the conditional budget share be represented by:

$$\widetilde{\omega}^k = \omega^k / \sum_{g \in G_g} \omega^g$$

The conditional pilot, $\widetilde{\eta}^k = \eta^k / \sum_{g \in G_g} \eta^g$ is analogously defined.

The Kullback-Leibler divergence of $\boldsymbol{\omega}$ with respect to $\boldsymbol{\eta}$ can be expressed as:

$$\begin{aligned} D(\boldsymbol{\omega} \| \boldsymbol{\eta}) &= \sum_{k=1}^{K} \omega^k \log_2 \frac{\omega^k}{\eta^k} \\ &= \sum_{g=1}^{|G|} \sum_{k \in G_g} m_g \widetilde{\omega}^k \log_2 \frac{m_g \widetilde{\omega}^k}{n_g \widetilde{\eta}^k} \\ &= \sum_{g=1}^{|G|} m_g D(\widetilde{\boldsymbol{\omega}}^{G_g} \| \widetilde{\boldsymbol{\eta}}^{G_g}) + \sum_{g=1}^{|G|} m_g \log_2 \frac{m_g}{n_g} \end{aligned}$$

where

$$\begin{aligned} D(\widetilde{\boldsymbol{\omega}}^{G_g} \| \widetilde{\boldsymbol{\eta}}^{G_g}) &= \sum_{k \in G_g} \omega^{k\prime} \log_2 \frac{\omega^k}{\eta^k} \\ n_g &= \sum_{k \in G_g} \eta_k \end{aligned}$$

With knowledge of m_g, $\sum_{g=1}^{|G|} m_g \log_2 \frac{m_g}{n_g}$ is a constant and is irrelevant for the minimisation.[2] Given weak separability, and therefore independence of conditional demands given m_g, minimising $D(\boldsymbol{\omega} \| \boldsymbol{\eta})$ subject to S_b^V is equivalent to the independent minimisation of each term of $\sum_g D(\widetilde{\boldsymbol{\omega}}^{G_g} \| \widetilde{\boldsymbol{\eta}}^{G_g})$ subject to $\widetilde{S}_b^g$, where $\widetilde{S}_b^g$ summarises the constraints that rationality places on the within-group allocations.

Where m_g are unknown, the updated solution for the conditional within-group demands is again independent of the strategy taken, although the full system demands will typically differ according to the updating method followed. This follows from the "Weak Subset Independence" property of cross-entropy minimisation. See Property 8 of Shore and Johnson (1981).

PROPERTY 5 Without specification of a pilot budget share, the updated solution corresponds to the budget share specification that is hardest to discriminate from that dictated by symmetric Cobb-Douglas preferences.

(Proof) The principle of maximum entropy proposed by Jaynes (1957) corresponds to minimising the functional:

$$H(\boldsymbol{\omega}) = \boldsymbol{\omega}' \log_2 \boldsymbol{\omega}$$

The demand at a new budget $\{\boldsymbol{\rho}_b, m_b\}$ that is dictated by symmetric Cobb-Douglas preferences is given by:

$$q_k = \frac{1}{K} \frac{m_b}{\rho_b^k}$$

2 Knowing m_g in advance is plausible for many applications of two-stage budgeting.

The assumption of symmetric Cobb-Douglas preferences thereby entails the pilot prediction $\eta_k = \frac{1}{K}$ for $k = 1, ..., K$. Let this pilot be represented by $\boldsymbol{\eta}^{CD}$.

Given $\boldsymbol{\eta}^{CD}$, the objective function in the MDI minimisation problem takes the form:

$$\begin{aligned} D(\boldsymbol{\omega}\|\boldsymbol{\eta}^{CD}) &= \boldsymbol{\omega}' \log_2 \tfrac{\boldsymbol{\omega}}{\boldsymbol{\eta}^{CD}} \\ &= \boldsymbol{\omega}' \log_2 \boldsymbol{\omega} + \textstyle\sum_k \omega_k \log_2 K \\ &= \boldsymbol{\omega}' \log_2 \boldsymbol{\omega} + \log_2 K \end{aligned}$$

The last term is a constant and irrelevant for minimisation. Thus,

$$H(\boldsymbol{\omega}) \equiv D(\boldsymbol{\omega}\|\boldsymbol{\eta}^{CD})$$

PROPOSITION 2.5 An observation $\{\mathbf{p}_{t+1}, \mathbf{q}_{t+1}\}$ is informative iff

$$\{A_b \boldsymbol{\omega}_b^{t\star}\}_{b=1,...,|B|} \notin S_{t+1}^S(B)$$

(Proof) By definition, $\mathbf{A}_b \boldsymbol{\omega}^t \in S_t^S$ and $\mathbf{A}_b \boldsymbol{\omega}^{t+1} \in S_{t+1}^S$. The support set is weakly shrinking in additional observations, $S_{t+1}^S \subseteq S_t^S$. Therefore, any feasible solution given an additional observation $\{\mathbf{p}_{t+1}, \mathbf{q}_{t+1}\}$ must have been feasible previously. Given that the inequality that defines when an observation is informative is strict, we must have that $\{\boldsymbol{\omega}_b^{t\star}\}_{b=1,...,|B|} \neq \{\boldsymbol{\omega}_b^{t+1\star}\}_{b=1,...,|B|}$.

Now, imagine that an observation $\{\mathbf{p}_{t+1}, \mathbf{q}_{t+1}\}$ is informative, $H(D_{t+1}) > H(D_t)$ but $\{\mathbf{A}_b \boldsymbol{\omega}_b^{t\star}\}_{b=1,...,|B|} \in S_{t+1}^S$. However, $\{\mathbf{A}_b \boldsymbol{\omega}_b^{t+1\star}\}_{b=1,...,|B|} \in S_t^S$. The fact that $\{\boldsymbol{\omega}_b^{t\star}\}_{b=1,...,|B|}$ is selected over $\{\boldsymbol{\omega}_b^{t+1\star}\}_{b=1,...,|B|}$ when one observes the data set D_t implies that:

$$\frac{1}{|B|} \sum_{b=1}^{|B|} \boldsymbol{\omega}_b^{t+1\star\prime} \log_2 \frac{\boldsymbol{\omega}_b^{t+1\star}}{\boldsymbol{\eta}_b} \geqslant H(D_t)$$

Thus, for $\{\boldsymbol{\omega}_b^{t+1\star}\}_{b=1,...,|B|}$ to be selected as the MDI solution, $\{\boldsymbol{\omega}_b^{t\star}\}_{b=1,...,|B|}$ must not be feasible upon the observation of $\{\mathbf{p}_{t+1}, \mathbf{q}_{t+1}\}$,

$$\{\mathbf{A}_b \boldsymbol{\omega}_b^{t\star}\}_{b=1,...,|B|} \notin S_{t+1}^S$$

or

$$\frac{1}{|B|} \sum_{b=1}^{|B|} \boldsymbol{\omega}_b^{t+1\star\prime} \log_2 \frac{\boldsymbol{\omega}_b^{t+1\star}}{\boldsymbol{\eta}_b} = H(D_t)$$

In which case $\{\mathbf{p}_{t+1}, \mathbf{q}_{t+1}\}$ is not informative.

Now consider an observation $\{\mathbf{p}_{t+1}, \mathbf{q}_{t+1}\}$ that alters the support set in such a way that the previous estimate is infeasible, $\{\mathbf{A}_b \boldsymbol{\omega}_b^{t\star}\}_{b=1,...,|B|} \notin S_{t+1}^S$, but that the observation is not informative, $H(D_t) > H(D_{t+1})$. This again leads to a contradiction as $\{\mathbf{A}_b \boldsymbol{\omega}_b^{t+1\star}\}_{b=1,...,|B|} \notin S_t^S$ and thus $H(D_t) \leqslant \frac{1}{|B|} \sum_{b=1}^{|B|} \boldsymbol{\omega}_b^{t+1\star\prime} \log_2 \frac{\boldsymbol{\omega}_b^{t+1\star}}{\boldsymbol{\eta}_b}$ or $\{\mathbf{A}_b \boldsymbol{\omega}_b^{t+1\star}\}_{b=1,...,|B|}$ would have been selected as the MDI solution previously.

B

TIME CONSISTENCY, COLLECTIVE CHOICE & REVEALED PREFERENCE

This Appendix provides the proofs, data description and additional methodological details for the chapter "Time Consistency, Collective Choice and Revealed Preference".

B.1 PROOFS

THEOREM 3.1 The set of observations $S = \{\mathbf{q}_t, \mathbf{Q}_t, \mathbf{p}_t, \mathbf{P}_t\}_{t \in T}$ can be rationalised by the time consistency model if and only if there exist, for all $t \in T$, a utility number $u_t^H \in \mathbb{R}$ and a positive constant $\beta \in (0, 1]$ that, for any $s, t \in T$, satisfy

$$u_s^H - u_t^H \leqslant \frac{1}{\beta^{t-1}}[\mathbf{p}_t^{'}(\mathbf{q}_s - \mathbf{q}_t) + \mathbf{P}_t^{'}(\mathbf{Q}_s - \mathbf{Q}_t)]. \quad (3.T1.1)$$

(Necessity) Suppose the set of observations S can be rationalised by the time consistency model. Let η denote the Lagrange multiplier associated with the household budget constraint. We get the following first order constraints for the household optimisation problem:

$$\beta^{t-1}\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{q}_t^m} \leqslant \eta \mathbf{p}_t \ (m \in \{A, B\}) \text{ and}$$
$$\beta^{t-1}\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{Q}_t} \leqslant \eta \mathbf{P}_t.$$

where $\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{q}_t^m}$ and $\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{Q}_t}$ represent the subgradients of the household felicity function u^H at $(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)$ for $m \in \{A, B\}$.

Under concavity of the function u^H, we have

$$u^H(\mathbf{q}_s^A, \mathbf{q}_s^B, \mathbf{Q}_s) - u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t) \leqslant \left(\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{q}_t^A}\right)'(\mathbf{q}_s^A - \mathbf{q}_t^A) + \left(\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{q}_t^B}\right)'(\mathbf{q}_s^B - \mathbf{q}_t^B) + \left(\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{Q}_t}\right)'(\mathbf{Q}_s - \mathbf{Q}_t)$$

Combining these inequalities leads to

$$u^H(\mathbf{q}_s^A, \mathbf{q}_s^B, \mathbf{Q}_s) - u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t) \leqslant \frac{\eta}{\beta^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s - \mathbf{q}_t) + \mathbf{P}_t'(\mathbf{Q}_s - \mathbf{Q}_t)]$$

By using $u_s^H = \frac{u^H(\mathbf{q}_s^A, \mathbf{q}_s^B, \mathbf{Q}_s)}{\eta}$, we obtain the inequalities in Theorem 3.1.

(Sufficiency) Suppose the inequalities in Theorem 3.1 hold. One can then define the following household felicity function:

$$u^H(\mathbf{x}^A, \mathbf{x}^B, \mathbf{X}) = \min_s \left(u_s^H + \frac{1}{\beta^{s-1}} [\mathbf{p}'_s(\mathbf{x}^A - \mathbf{q}_s^A) + \mathbf{p}'_s(\mathbf{x}^B - \mathbf{q}_s^B) + \mathbf{P}'_s(\mathbf{X} - \mathbf{Q}_s)] \right)$$

Using the argument outlined by Varian (1982), we can derive $u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t) = u_t^H$.

Consider a consumption plan $\{\mathbf{x}_t^A, \mathbf{x}_t^B, \mathbf{X}_t\}_{t \in T}$ such that

$$\sum_{t \in T} \left[\mathbf{p}'_t(\mathbf{x}_t^A - \mathbf{q}_t^A) + \mathbf{p}'_t(\mathbf{x}_t^B - \mathbf{q}_t^B) + \mathbf{P}'_t(\mathbf{X}_t - \mathbf{Q}_t) \right] \leqslant 0,$$

i.e. the consumption plan $\{\mathbf{x}_t^A, \mathbf{x}_t^B, \mathbf{X}_t\}_{t \in T}$ is affordable given the outlay associated with $\{\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t\}_{t \in T}$. Then, we need to show that

$$\sum_{t \in T} \beta^{t-1} u^H(\mathbf{x}_t^A, \mathbf{x}_t^B, \mathbf{X}_t) \leqslant \sum_{t \in T} \beta^{t-1} u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t).$$

Using the utility function defined above, we obtain

$$\begin{aligned}
& \sum_{t \in T} \beta^{t-1} u^H(\mathbf{x}_t^A, \mathbf{x}_t^B, \mathbf{X}_t) \\
\leqslant \; & \sum_{t \in T} \beta^{t-1} \left(u_t^H + \frac{1}{\beta^{t-1}} [\mathbf{p}'_t(\mathbf{x}_t^A - \mathbf{q}_t^A) + \mathbf{p}'_t(\mathbf{x}_t^B - \mathbf{q}_t^B) + \mathbf{P}'_t(\mathbf{X}_t - \mathbf{Q}_t)] \right) \\
= \; & \sum_{t \in T} \beta^{t-1} u_t^H + \sum_{t \in T} [\mathbf{p}'_t(\mathbf{x}_t^A - \mathbf{q}_t^A) + \mathbf{p}'_t(\mathbf{x}_t^B - \mathbf{q}_t^B) + \mathbf{P}'_t(\mathbf{X}_t - \mathbf{Q}_t)] \\
\leqslant \; & \sum_{t \in T} \beta^{t-1} u_t^H,
\end{aligned}$$

so that $u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t) = u_t^H$ gives the wanted conclusion.

This completes the proof of Theorem 3.1.

THEOREM 3.2 The set of observations S can be rationalised by the full efficiency model if and only if there exist, for all $t \in T$, private quantities $\mathbf{q}_t^A, \mathbf{q}_t^B \in \mathbb{R}_+^N$, Lindahl prices $\mathbf{P}_t^A, \mathbf{P}_t^B \in \mathbb{R}_+^K$, utility numbers $u_t^A, u_t^B \in \mathbb{R}$ and constants $\beta_A, \beta_B \in (0, 1]$ that, for any $s, t \in T$, satisfy:

$$u_s^A - u_t^A \leqslant \frac{1}{\beta_A^{t-1}} [\mathbf{p}'_t(\mathbf{q}_s^A - \mathbf{q}_t^A) + \mathbf{P}_t^{A'}(\mathbf{Q}_s - \mathbf{Q}_t)] \tag{3.T2.1}$$

$$u_s^B - u_t^B \leqslant \frac{1}{\beta_B^{t-1}} [\mathbf{p}'_t(\mathbf{q}_s^B - \mathbf{q}_t^B) + \mathbf{P}_t^{B'}(\mathbf{Q}_s - \mathbf{Q}_t)] \tag{3.T2.2}$$

with

$$\begin{aligned} \mathbf{q}_t^A + \mathbf{q}_t^B &= \mathbf{q}_t \\ \mathbf{P}_t^A + \mathbf{P}_t^B &= \mathbf{P}_t \end{aligned} \tag{3.T2.3}$$

(Necessity) Suppose that the set of observations S can be rationalised by the full efficiency model. Let η denote the Lagrange multiplier associated with the household budget constraint. The model is associated with the following first order constraints:

$$\begin{aligned}
\beta_A^{t-1} \frac{\partial u^A(\mathbf{q}_t^A, \mathbf{Q}_t)}{\partial \mathbf{q}_t^A} &\leqslant \eta \mathbf{p}_t \\
\omega \beta_B^{t-1} \frac{\partial u^B(\mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{q}_t^B} &\leqslant \eta \mathbf{p}_t \\
\beta_A^{t-1} \frac{\partial u^A(\mathbf{q}_t^A, \mathbf{Q}_t)}{\partial \mathbf{Q}_t} + \omega \beta_B^{t-1} \frac{\partial u^B(\mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{Q}_t} &\leqslant \eta \mathbf{P}_t.
\end{aligned}$$

where $\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{q}_t^m}$ and $\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{Q}_t}$ represent the subgradients of the household felicity function u^H at $(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)$ for $m \in \{A, B\}$.

Under concavity of each felicity functions u^m, we have

$$\begin{aligned}
&u^m(\mathbf{q}_s^m, \mathbf{Q}_s) - u^m(\mathbf{q}_t^m, \mathbf{Q}_t) \\
&\leqslant \\
&\left(\frac{\partial u^m(\mathbf{q}_t^m, \mathbf{Q}_t)}{\partial \mathbf{q}_t^m}\right)' (\mathbf{q}_s^m - \mathbf{q}_t^m) + \left(\frac{\partial u^m(\mathbf{q}_t^m, \mathbf{Q}_t)}{\partial \mathbf{Q}_t}\right)' (\mathbf{Q}_s - \mathbf{Q}_t)
\end{aligned}$$

Now define, for each $t \in T$,

$$\begin{aligned}
\lambda^A &= \eta \\
\lambda^B &= \eta/\omega \\
\mathbf{P}_t^A &= \frac{\beta_A^{t-1}}{\lambda^A} \frac{\partial u^A(\mathbf{q}_t^A, \mathbf{Q}_t)}{\partial \mathbf{Q}_t} \\
\mathbf{P}_t^B &= \mathbf{P}_t - \mathbf{P}_t^A.
\end{aligned}$$

Combining these elements yields:

$$\begin{aligned}
&u^m(\mathbf{q}_s^m, \mathbf{Q}_s) - u^m(\mathbf{q}_t^m, \mathbf{Q}_t) \\
&\leqslant \\
&\frac{\lambda^m}{\beta_m^{t-1}} \mathbf{p}_t' (\mathbf{q}_s^m - \widehat{\mathbf{q}}_t^m) + \frac{\lambda^m}{\beta_m^{t-1}} (\mathbf{P}_t^m)' (\mathbf{Q}_s - \mathbf{Q}_t)
\end{aligned}$$

By using $u^m(\mathbf{q}_s^m, \mathbf{Q}_s)/\lambda^m = u_s^m$, one obtains the inequalities in Theorem 3.2.

(Sufficiency) Suppose that the inequalities in Theorem 3.2 hold. Define the following felicity function for each member $m \in \{A, B\}$ as:

$$u^m(\mathbf{x}^m, \mathbf{X}) = \min_s \left(u_s^m + \frac{\lambda^m}{\beta_m^{s-1}} [\mathbf{p}_s'(\mathbf{x}^m - \mathbf{q}_s^m) + \mathbf{P}_s'(\mathbf{X} - \mathbf{Q}_s)] \right)$$

Consider any consumption plan $\{\mathbf{x}_t^A, \mathbf{x}_t^B, \mathbf{X}_t\}_{t \in T}$ such that

$$\sum_{t \in T} \left[\mathbf{p}_t'(\mathbf{x}_t^A - \mathbf{q}_t^A) + \mathbf{p}_t'(\mathbf{x}_t^B - \mathbf{q}_t^B) + \mathbf{P}_t'(\mathbf{X}_t - \mathbf{Q}_t) \right] \leqslant 0,$$

i.e. the consumption plan $\{\mathbf{x}_t^A, \mathbf{x}_t^B, \mathbf{X}_t\}_{t\in T}$ is affordable given the outlay associated with $\{\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t\}_{t\in T}$. Then, for $\omega^A = 1$ and $\omega^B = \omega$ we must show that

$$\sum_{m\in\{A,B\}}\sum_{t\in T}\omega^m\beta_m^{t-1}u^m(\mathbf{x}_t^m, \mathbf{X}_t) \leqslant \sum_{m\in\{A,B\}}\sum_{t\in T}\omega^m\beta_m^{t-1}u^m(\mathbf{q}_t^m, \mathbf{Q}_t).$$

Without losing generality, we can assume $\eta = 1$ or, equivalently, $\omega^m = \frac{1}{\lambda^m}$. As such, we obtain

$$\begin{aligned}
& \sum_{m\in\{A,B\}}\sum_{t\in T}\omega^m\beta_m^{t-1}u^m(\mathbf{x}_t^m, \mathbf{X}_t) \\
\leqslant\ & \sum_{m\in\{A,B\}}\sum_{t\in T}\omega^m\beta_m^{t-1}\left(u_t^m + \frac{\lambda^m}{\beta_m^{t-1}}[\mathbf{p}_t'(\mathbf{x}_t^m - \mathbf{q}_t^m) + \mathbf{P}_t'(\mathbf{X}_t - \mathbf{Q}_t)]\right) \\
=\ & \sum_{m\in\{A,B\}}\sum_{t\in T}\omega^m\beta_m^{t-1}u_t^m + \sum_{m=A,B}\sum_{t\in T}[\mathbf{p}_t'(\mathbf{x}_t^m - \mathbf{q}_t^m) + \mathbf{P}_t'(\mathbf{X}_t - \mathbf{Q}_t)] \\
\leqslant\ & \sum_{m\in\{A,B\}}\sum_{t\in T}\omega^m\beta_m^{t-1}u_t^m,
\end{aligned}$$

so that $u^m(\mathbf{q}_t^m, \mathbf{Q}_t) = u_t^m$ gives the wanted conclusion.

This completes the proof of Theorem 3.2.

THEOREM 3.3 The set of observations S can be rationalised by the no-commitment model if and only if there exist, for all $t \in T$, private quantities $\mathbf{q}_t^A, \mathbf{q}_t^B \in \mathbb{R}_+^N$, Lindahl prices $\mathbf{P}_t^A, \mathbf{P}_t^B \in \mathbb{R}_+^K$, utility numbers $u_t^A, u_t^B \in \mathbb{R}$, and for all $\tau = \{1, ..., \Upsilon\}$, discounted marginal utilities of income $\lambda_\tau^A, \lambda_\tau^B \in \mathbf{R}_+$ and discount factors $\beta_A, \beta_B \in (0, 1]$ such that, for any $s \in T$ and $t \in T_\tau (\tau = \{1, ..., \Upsilon\})$, the following inequalities are satisfied:

$$u_s^A - u_t^A \leqslant \frac{\lambda_\tau^A}{\beta_A^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^A - \mathbf{q}_t^A) + \mathbf{P}_t^{A'}(\mathbf{Q}_s - \mathbf{Q}_t)]. \tag{3.T3.1}$$

$$u_s^B - u_t^B \leqslant \frac{\lambda_\tau^B}{\beta_B^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^B - \mathbf{q}_t^B) + \mathbf{P}_t^{B'}(\mathbf{Q}_s - \mathbf{Q}_t)]. \tag{3.T3.2}$$

and if $\lambda_{\tau_1}^m \neq \lambda_{\tau_2}^m$ for $m \in \{A, B\}$, then

$$\frac{\lambda_{\tau_1}^A}{\lambda_{\tau_1}^B} \neq \frac{\lambda_{\tau_2}^A}{\lambda_{\tau_2}^B} \tag{3.T3.3}$$

with

$$\begin{aligned}
\mathbf{q}_t^A + \mathbf{q}_t^B &= \mathbf{q}_t \\
\mathbf{P}_t^A + \mathbf{P}_t^B &= \mathbf{P}_t
\end{aligned}$$

(Necessity) Suppose the set of observations S can be rationalised by the no-commitment model. Let η denote the Lagrange multiplier associated with the household budget constraint. The following first order constraints are associated with the household's optimisation problem:

$$\begin{aligned} \omega_t^A \beta_A^{t-1} \frac{\partial u^A(\mathbf{q}_t^A, \mathbf{Q}_t)}{\partial \mathbf{q}_t^A} &\leqslant \eta \mathbf{p}_t \\ \omega_t^B \beta_B^{t-1} \frac{\partial u^B(\mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{q}_t^B} &\leqslant \eta \mathbf{p}_t \\ \omega_t^A \beta_A^{t-1} \frac{\partial u^A(\mathbf{q}_t^A, \mathbf{Q}_t)}{\partial \mathbf{Q}_t} + \omega_t^B \beta_B^{t-1} \frac{\partial u^B(\mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{Q}_t} &\leqslant \eta \mathbf{P}_t. \end{aligned}$$

where $\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{q}_t^m}$ and $\frac{\partial u^H(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)}{\partial \mathbf{Q}_t}$ represent the subgradients of the household felicity function u^H at $(\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t)$ for $m \in \{A, B\}$.

Under concavity of each felicity functions u^m, we have

$$\begin{aligned} u^m(\mathbf{q}_s^m, \mathbf{Q}_s) - u^m(\mathbf{q}_t^m, \mathbf{Q}_t) \\ \leqslant \\ \left(\frac{\partial u^m(\mathbf{q}_t^m, \mathbf{Q}_t)}{\partial \mathbf{q}_t^m}\right)' (\mathbf{q}_s^m - \mathbf{q}_t^m) + \left(\frac{\partial u^m(\mathbf{q}_t^m, \mathbf{Q}_t)}{\partial \mathbf{Q}_t}\right)' (\mathbf{Q}_s - \mathbf{Q}_t). \end{aligned}$$

For each $t \in T$ and $m \in \{A, B\}$, let us define

$$\begin{aligned} \lambda_t^m &= \eta / \omega_t^m \\ \mathbf{P}_t^A &= \frac{\beta_A^{t-1}}{\lambda^A} \frac{\partial u^A(\mathbf{q}_t^A, \mathbf{Q}_t)}{\partial \mathbf{Q}_t} \\ \mathbf{P}_t^B &= \mathbf{P}_t - \mathbf{P}_t^A. \end{aligned}$$

Let us partition the set T into Υ mutually exclusive subsets T_τ of the form:

$$\mathbb{T} = \{T_1, ..., T_\Upsilon\}$$

with

$$T = \bigcup_{\tau=1}^{\Upsilon} T_\tau \text{ and } T_{\tau_s} \bigcap T_{\tau_t} = \emptyset \text{ if } \tau_t \neq \tau_s$$

such that

$$\tau_1 < \tau_2 \text{ implies } t_1 < t_2 \text{ for all } t_1 \in T_{\tau_1} \text{ and } t_2 \in T_{\tau_2}$$

Each subset, T_τ, represents a distinct Pareto weight regime within which $\omega_s^m = \omega_t^m$ for all $s, t \in T_\tau$ because individual incentive compatibility constraints do not bind within T_τ. Let the Pareto weight in subset T_τ be denoted ω_τ^m. Thus, for all $t \in T_\tau$ and $m \in \{A, B\}$:

$$\lambda_t^m = \eta / \omega_\tau^m = \lambda_\tau^m$$

Combining these elements returns 3.T3.1 and 3.T3.2:

$$u_s^m - u_t^m \leqslant \frac{\lambda_\tau^m}{\beta_m^{t-1}} \mathbf{p}_t' (\mathbf{q}_s^m - \widehat{\mathbf{q}}_t^m) + \frac{\lambda_\tau^m}{\beta_m^{t-1}} \mathbf{P}_t^{m\prime} (\mathbf{Q}_s - \mathbf{Q}_t).$$

for all $s \in T$, $t \in T_\tau$ and $m \in \{A, B\}$.

To prove the necessity of condition 3.T3.3, consider periods t and $t+1$ and let the incentive compatibility constraint of member A bind at $t+1$. Given that $\omega_t^A = \omega_{t-1}^A + \phi_t^A$, this implies that $\omega_{t+1}^A > \omega_t^A$ and therefore $\lambda_t^A > \lambda_{t+1}^A$. Renegotiation thus perturbs the ratio of the marginal utility of discounted income of members:

$$\frac{\lambda_t^A}{\lambda_t^B} = \frac{\eta/\omega_t^A}{\eta/\omega_t^B} = \frac{\omega_t^B}{\omega_t^A} \neq \frac{\omega_t^B}{\omega_t^A + \phi_{t+1}^A} = \frac{\eta/\omega_{t+1}^A}{\eta/\omega_{t+1}^B} = \frac{\lambda_{t+1}^A}{\lambda_{t+1}^B}$$

(Sufficiency) Suppose that the inequalities in Theorem 3.3 hold. Define the following felicity function for each household member $m \in \{A, B\}$:

$$u^m(\mathbf{x}^m, \mathbf{X}) = \min_s \left(u_s^m + \frac{\lambda_\tau^m}{\beta_m^{s-1}} [\mathbf{p}'_s(\mathbf{x}^m - \mathbf{q}_s^m) + \mathbf{P}'_s(\mathbf{X} - \mathbf{Q}_s)] \right).$$

Using a similar argument to that outlined by Varian (1982), we can derive $u^m(\mathbf{q}_t^m, \mathbf{Q}_t) = u_t^m$.

Consider any consumption plan $\{\mathbf{x}_t^A, \mathbf{x}_t^B, \mathbf{X}_t\}_{t \in T}$ such that

$$\sum_{t \in T} \left[\mathbf{p}'_t(\mathbf{x}_t^A - \mathbf{q}_t^A) + \mathbf{p}'_t(\mathbf{x}_t^B - \mathbf{q}_t^B) + \mathbf{P}'_t(\mathbf{X}_t - \mathbf{Q}_t) \right] \leqslant 0,$$

i.e. the consumption plan $\{\mathbf{x}_t^A, \mathbf{x}_t^B, \mathbf{X}_t\}_{t \in T}$ is affordable given the outlay associated with $\{\mathbf{q}_t^A, \mathbf{q}_t^B, \mathbf{Q}_t\}_{t \in T}$. Then, we need to show that

$$\sum_{m \in \{A,B\}} \sum_{t \in T} \omega_t^m \beta_m^{t-1} u^m(\mathbf{x}_t^m, \mathbf{X}_t) \leqslant \sum_{m \in \{A,B\}} \sum_{t \in T} \omega_t^m \beta_m^{t-1} u^m(\mathbf{q}_t^m, \mathbf{Q}_t).$$

Without losing generality, we can assume $\omega_t^m = \frac{1}{\lambda_\tau^m}$ for $t \in T_\tau$. As such, we obtain

$$\begin{aligned}
& \sum_{m \in \{A,B\}} \sum_{t \in T} \omega_t^m \beta_m^{t-1} u^m(\mathbf{x}_t^m, \mathbf{X}_t) \\
\leqslant \; & \sum_{m \in \{A,B\}} \sum_{t \in T} \omega_t^m \beta_m^{t-1} \left(u_t^m + \frac{\lambda_\tau^m}{\beta_m^{t-1}} [\mathbf{p}'_t(\mathbf{x}_t^m - \mathbf{q}_t^m) + \mathbf{P}'_t(\mathbf{X}_t - \mathbf{Q}_t)] \right) \\
= \; & \sum_{m \in \{A,B\}} \sum_{t \in T} \omega_t^m \beta_m^{t-1} u_t^m + \sum_{m = A,B} \sum_{t \in T} [\mathbf{p}'_t(\mathbf{x}_t^m - \mathbf{q}_t^m) + \mathbf{P}'_t(\mathbf{X}_t - \mathbf{Q}_t)] \\
\leqslant \; & \sum_{m \in \{A,B\}} \sum_{t \in T} \omega_t^m \beta_m^{t-1} u_t^m,
\end{aligned}$$

so that $u^m(\mathbf{q}_t^m, \mathbf{Q}_t) = u_t^m$ gives the wanted conclusion.

This completes the proof of Theorem 3.3.

---

necessary no-commitment test A necessary condition for the set of observations S to be rationalised by the no-commitment model is that there exist, for all $t \in T$, private quantities $\mathbf{q}_t^A, \mathbf{q}_t^B \in \mathbb{R}_+^N$, Lindahl prices $\mathbf{P}_t^A, \mathbf{P}_t^B \in \mathbb{R}_+^K$, utility numbers $u_t^A, u_t^B \in \mathbb{R}$ and discount factors $\beta_A, \beta_B \in (0,1]$ that, for any $s, t \in T_\tau(\tau = \{1, ..., \Upsilon\})$, satisfy:

$$u_s^A - u_t^A \leqslant \frac{1}{\beta_A^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^A - \mathbf{q}_t^A) + \mathbf{P}_t^{A'}(\mathbf{Q}_s - \mathbf{Q}_t)]. \quad (3.N3.1)$$

$$u_s^B - u_t^B \leqslant \frac{1}{\beta_B^{t-1}}[\mathbf{p}_t'(\mathbf{q}_s^B - \mathbf{q}_t^B) + \mathbf{P}_t^{B'}(\mathbf{Q}_s - \mathbf{Q}_t)]. \quad (3.N3.2)$$

with

$$\begin{aligned} \mathbf{q}_t^A + \mathbf{q}_t^B &= \mathbf{q}_t \\ \mathbf{P}_t^A + \mathbf{P}_t^B &= \mathbf{P}_t \end{aligned} \quad (3.N3.3)$$

This result follows from the fact that, given a partition of the set T into mutually exclusive subsets $\mathbb{T}$ into distinct Pareto weight regimes, consistency with the no-commitment model requires that within each subset T_τ ($\tau \in \{1, ..., \Upsilon\}$), the corresponding subset of observations $\{\mathbf{q}_t, \mathbf{Q}_t; \mathbf{p}_t, \mathbf{P}_t\}_{t \in T_\tau}$ can be rationalised by the full efficiency model.

B.2 Data description

The following table provides summary statistics of the data used in the empirical application of Chapter 3. The average expenditure shares and discounted prices over all observations and their standard deviations are listed for the good considered.

	Private goods	
	Mean share	***Mean price***
	(st. dev.)	*(st. dev.)*
Food and non-alcoholic drinks	0.3411 (0.1284)	0.9777 (0.1022)
Alcohol	0.0113 (0.0190)	0.9465 (0.1595)
Tobacco	0.0200 (0.0265)	0.9802 (0.2480)
Clothing and footwear	0.0105 (0.0225)	0.9775 (0.0467)
Nondurable medicines	0.0181 (0.0523)	0.8884 (0.2229)
Medical services	0.0774 (0.0819)	0.9099 (0.2126)
Transportation	0.0501 (0.0507)	0.9587 (0.2126)
Petrol	0.0284 (0.0364)	0.9482 (0.1409)
Leisure (cinema, theatre, clubs for sport)	0.0140 (0.0254)	0.9250 (0.1888)
Personal services	0.0127 (0.0229)	0.9702 (0.1124)
Restaurants and bars	0.1058 (0.0949)	0.9304 (0.1765)
	Public goods	
	Mean share	***Mean price***
	(st. dev.)	*(st. dev.)*
Rent	0.2406 (0.1275)	0.9610 (0.1630)
Energy	0.0435 (0.0299)	0.9556 (0.1428)
Home Services (heating, water and furniture repair)	0.0063 (0.0240)	0.9237 (0.1961)
Nondurables at home (cleaning products)	0.0204 (0.0217)	0.9682 (0.0656)

The following table records summary statistics on household characteristics that were considered in the course of our empirical exploration of time preference heterogeneity.

Characteristic	**Mean**	**St. Dev.**
Average age of couple	42.99	9.27
Age difference	1.94	3.19
University degree	0.08	0.27
High school degree dummy	0.29	0.46
Children dummy	0.70	0.45
Log total expenditure	13.04	0.48
Percentage public consumption	0.31	0.13
Home owner dummy	0.78	0.40
Skilled job dummy	0.09	0.28

In our assessment of the time consistent model, a number of robustness checks were performed to probe the sensitivity of our results. We here provide a formal presentation of the methodology that was applied in the course of these robustness checks.

MEASUREMENT ERROR Our measurement error procedure assumes the following multiplicative error structure:

$$q_t^n = (1 + \epsilon_t^n)\, q_t^{n\star} \text{ and } Q_t^k = \left(1 + \varepsilon_t^k\right) Q_t^{k\star}$$

where $q_t^{n\star}$ and $Q_t^{n\star}$ represent the true, but unobserved, values of the private and public quantities.

Assume that the true data set $S^\star = \{\mathbf{q}_t^\star, \mathbf{Q}_t^\star; \mathbf{p}_t, \mathbf{P}_t\}_{t\in T}$ is rationalisable by the time consistent model. However, we only observe the measured data set $S = \{\mathbf{q}_t, \mathbf{Q}_t; \mathbf{p}_t, \mathbf{P}_t\}_{t\in T}$, which may not be theory-rationalisable.

Define the following perturbation to observed quantities:

$$\widetilde{q}_t^n = (1 + \widetilde{\epsilon}_{t,n})\, q_t^n \text{ and } \widetilde{Q}_t^k = \left(1 + \widetilde{\varepsilon}_t^k\right) Q_t^n$$

While we cannot observe the actual errors ϵ_t^n and ε_t^k, we can calculate the smallest perturbations $\widetilde{\epsilon}_t^n$ and $\widetilde{\varepsilon}_t^k$ necessary such that the perturbed data set $\widetilde{S} = \{\widetilde{\mathbf{q}}_t, \widetilde{\mathbf{Q}}_t; \mathbf{p}_t, \mathbf{P}_t\}_{t\in T}$ is consistent with the time consistent model. These perturbations are found by minimising the sum of squared error terms,

$$\min \widetilde{\mathbf{V}} = \sum_{t\in T} \left(\sum_{n=1}^{N} (\widetilde{\epsilon}_t^n)^2 + \sum_{k=1}^{K} (\widetilde{\varepsilon}_t^k)^2 \right),$$

subject to the constraint that $\widetilde{S}$ satisfies the model. These perturbations define a lower bound on the true magnitude of measurement error:

$$\mathbf{V} = \sum_{t\in T} \left(\sum_{n=1}^{N} (\epsilon_t^n)^2 + \sum_{k=1}^{K} (\varepsilon_t^k)^2 \right)$$

$\widetilde{\mathbf{V}}$ can be used to assess the null hypothesis that $S^\star$ can be rationalised by the model once measurement error in observed quantities is accounted for. If one assumes that the true errors ϵ_t^n and ε_t^k are independently normally distributed with zero mean and constant variance σ^2, then we know that

$$\frac{\mathbf{V}}{\sigma^2} \sim \chi^2_{120},$$

where the degrees of freedom are equal to the number of quantity errors, $|T| \star N = 8 \star 15$.[1] If $\mathbf{V}$ and σ^2 were observable, the null hypoth-

1 Note that since prices are strictly positive, the relative errors are bounded from below; $\epsilon_{t,n}, \varepsilon_{t,k} \in]-1, +\infty[$. However, imposing normality should not be a concern, since these errors are typically not too far removed from zero.

esis would be rejected if $\mathbf{V}/\sigma^2 \geqslant C_\alpha$, where C_α represents the critical value from the chi-squared distribution at significance level α.

As neither $\mathbf{V}$ nor σ^2 are observed in practice, $\widetilde{\mathbf{V}}$ is used to approximate the sum of squared measurement errors. As $\widetilde{\mathbf{V}}$ corresponds to the minimal perturbation of the quantities such that the data are rationalisable, we must have $\widetilde{\mathbf{V}} \leqslant \mathbf{V}$. Therefore, our test statistic $\widetilde{\mathbf{V}}/\sigma^2$ is conservative.

Finally, as the true extent of the measurement error, σ^2, is unobserved, the procedure outlined by Varian (1985) is applied. The critical standard deviation σ_α that is required for the null hypothesis of theory-rationalisable behaviour to not be rejected at a significance level α is calculated as:

$$\sigma_\alpha = \sqrt{\frac{\widetilde{\mathbf{V}}}{C_\alpha}}.$$

One then rejects rationalisability at a significance level α if and only if σ_α exceeds their prior beliefs about the likely magnitude of σ.

IMPERFECT CAPITAL MARKETS Under the assumption of perfect capital markets, households are able to borrow and save at the same interest rate, and there are no limits regarding the maximum amount they can borrow. If we denote savings in period t by D_t, then this simply assumes that $D_t \in \mathbb{R}$, where borrowing corresponds to negative saving. When combined with the assumption of perfect foresight, this ideal setting implies perfect *ex ante* consumption smoothing across time periods or, equivalently, that the marginal utility of wealth in each period t (denoted by η_t) is kept constant over time, i.e. $\eta_t = \eta$ for all $t \in T$.

To account for possibility of binding borrowing constraints, we adapt the procedure given by Demuynck and Verriest (2012). Specifically, we allow for a household's ability to achieve a perfectly smooth consumption path to be hindered by the fact that D_t cannot fall below some (unobserved) maximum borrowing amount, d_t. If this constraint is binding in some period (i.e. $D_t = -d_t$ for some $t \in T$), then the household is forced to consume less then what would be optimal in the absence of borrowing constraints.

The possibility of borrowing constraints can be introduced into the revealed preference characterisation of the time consistent model by adding the condition $\eta_t \geqslant \eta_{t+1}$. Then, when the borrowing constraint is not binding, i.e. if $D_t > -d_t$, then $\eta_t = \eta_{t+1}$. Conversely, in periods that a borrowing constraint binds, i.e. if $D_t = -d_t$, then $\eta_t > \eta_{t+1}$ to reflect the impact of the imperfect smoothing that results.

To summarise, by allowing the marginal utility of wealth to increase over time, one can effectively account for potentially binding borrowing constraints. Including this argument in the proof of Theorem 3.1 yields the following revealed preference conditions for the time consistent model under imperfect capital markets.

COROLLARY The set of observations $S = \{\mathbf{q}_t, \mathbf{Q}_t; \mathbf{p}_t, \mathbf{P}_t\}_{t \in T}$ can be rationalised by the time consistency model with an unobserved borrowing constraint if and only if there exist, for all $t \in T$, a utility number $u_t^H \in \mathbb{R}$, a marginal utility of wealth $\eta_t > 0$ and a discount factor $\beta \in (0, 1]$ that, for any s,t $\in T$, satisfy:

$$\begin{aligned} u_s^H - u_t^H &\leqslant \frac{\eta_t}{\beta^t}[\mathbf{p}_t'(\mathbf{q}_s - \mathbf{q}_t) + \mathbf{P}_t'(\mathbf{Q}_s - \mathbf{Q}_t)], \\ \eta_{t+1} &\leqslant \eta_t. \end{aligned}$$

These revealed preference conditions are weaker than those defined by Theorem 3.1. The two sets of conditions coincide in the limiting case when borrowing constraints never bind. This corresponds to $\eta_t = \eta$ for all $t \in T$, which was used in the proof of Theorem 3.1. Interestingly, the conditions remain linear in unknowns for a given value of β.

C

REVEALED PREFERENCE CHANGE

This Appendix provides the proofs, data description and estimation results referenced in the chapter "Revealed Preference Change: Rationalising Tobacco Consumption".

C.1 PROOFS

PROPOSITION 4.1 The individual utility function, $u^i(\mathbf{q}, \alpha^{1i}) = v^i(\mathbf{q}) + \alpha^{1i} q^1$ satisfies single crossing in $(\mathbf{q}, \alpha^1)$ space.

A utility function $u^i(\mathbf{q}, \alpha^1)$ satisfies the single crossing property in $(\mathbf{q}; \alpha^1)$ if for $\mathbf{q}' > \mathbf{q}''$ and $\alpha^{1\prime} > \alpha^{1\prime\prime}$

$$u^i(\mathbf{q}', \alpha^{1\prime\prime}) \geqslant u^i(\mathbf{q}'', \alpha^{1\prime\prime}) \text{ implies } u^i(\mathbf{q}', \alpha^{1\prime}) \geqslant u^i(\mathbf{q}'', \alpha^{1\prime})$$

Without loss of generality, assume that $\alpha^1 \in \mathbb{R}_{++}$.[1] Consider $\mathbf{q}' > \mathbf{q}''$ and $\alpha^{1\prime} > \alpha^{1\prime\prime}$ such that

$$u^i(\mathbf{q}', \alpha^{1\prime\prime}) \geqslant u^i(\mathbf{q}'', \alpha^{1\prime\prime})$$

This implies

$$\begin{aligned} v^i(\mathbf{q}') + \alpha^{1\prime\prime} q^{1\prime} &\geqslant v^i(\mathbf{q}'') + \alpha^{1\prime\prime} q^{1\prime\prime} \\ v^i(\mathbf{q}') - v^i(\mathbf{q}'') &\geqslant \alpha^{1\prime\prime}(q^{1\prime\prime} - q^{1\prime}) \end{aligned}$$

Given that $\mathbf{q}'' < \mathbf{q}'$, we have

$$q^{1\prime\prime} - q^{1\prime} < 0$$

Thus, for $\alpha^{1\prime} > \alpha^{1\prime\prime}$,

$$\alpha^{1\prime\prime}(q^{1\prime\prime} - q^{1\prime}) > \alpha^{1\prime}(q^{1\prime\prime} - q^{1\prime})$$

Therefore,

$$\begin{aligned} v^i(\mathbf{q}') - v^i(\mathbf{q}'') &\geqslant \alpha^{1\prime\prime}(q^{1\prime\prime} - q^{1\prime}) \\ &> \alpha^{1\prime}(q^{1\prime\prime} - q^{1\prime}) \end{aligned}$$

This implies that

$$\begin{aligned} v^i(\mathbf{q}') + \alpha^{1\prime} q^{1\prime} &> v^i(\mathbf{q}'') + \alpha^{1\prime} q^{1\prime\prime} \\ u^i(\mathbf{q}', \alpha^{1\prime}) &> u^i(\mathbf{q}'', \alpha^{1\prime}) \end{aligned}$$

This completes the proof of Proposition 4.1.

1 This result is without loss of generality given the ordinality of utility.

PROPOSITION 4.2 A sufficient condition for individual utility functions, $u^i(\mathbf{q}, \alpha^1) = v^i(\mathbf{q}) + \alpha^1 q^1$ and $u^j(\mathbf{q}, \alpha^{1j}) = v^j(\mathbf{q}) + \alpha^{1j} q^1$ to satisfy the single crossing property for any $i, j \in \{1, ..., N\}$ is $v^i(\mathbf{q}) = v^j(\mathbf{q}) = v(\mathbf{q})$.

PROOF. Define

$$u(\mathbf{q}, \alpha^1) = v(\mathbf{q}) + \alpha^1 q^1$$

With $v^i(\mathbf{q}) = v^j(\mathbf{q}) = v(\mathbf{q})$, individual utility functions can then be expressed as:

$$\begin{aligned} u^i(\mathbf{q}, \alpha^1) = v(\mathbf{q}) + \alpha^{1i} q^1 &= u(\mathbf{q}, \alpha^{1i}) \\ u^j(\mathbf{q}, \alpha^1) = v(\mathbf{q}) + \alpha^{1j} q^1 &= u(\mathbf{q}, \alpha^{1j}) \end{aligned}$$

Single crossing of $u(\mathbf{q}, \alpha^1) = v(\mathbf{q}) + \alpha^1 q^1$ in $(\mathbf{q}; \alpha^1)$ follows from above, completing the proof of Proposition 4.2.

LEMMA 4.1 The following statements are equivalent:

1. Individual i's observed choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$, can be taste rationalised by the set of taste shifters $\{\boldsymbol{\alpha}_t^i\}_{t=1,...,T}$, where $\boldsymbol{\alpha} \in \mathbb{R}_+^K$.
2. One can find sets $\{v_t^i\}_{t=1,...,T}$, $\{\alpha_t^{k,i}\}_{t=1,...,T}^{k=1,...,K}$ and $\{\lambda_t^i\}_{t=1,...,T}$ with $\lambda_t^i > 0$ for all $t = 1, ..., T$, such that there exists a non-empty solution set to the following inequalities:

$$\begin{aligned} v_s^i - v_t^i + \boldsymbol{\alpha}_t^i(\mathbf{q}_s^i - \mathbf{q}_t^i) &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \\ \alpha_t^{k,i} &\leqslant \lambda_t^i p_t^{k,i} \end{aligned}$$

PROOF: NECCESSITY Let us consider the case where our data set has been generated by the model in question. Observed choices are then the solution to the following optimisation problem:

$$\max_{\{\mathbf{q}_t\}_{t=1,...,T}} v^i(\mathbf{q}_t) + \boldsymbol{\alpha}_t^{i\prime} \mathbf{q}_t$$

subject to

$$\mathbf{p}_t' \mathbf{q}_t \leqslant x_t$$

An optimal interior solution to the problem must satisfy:

$$\nabla_{\mathbf{q}_t} v^i(\mathbf{q}_t) + \boldsymbol{\alpha}_t^i =_t^i \mathbf{p}_t$$

Given a particular level of the taste shifter, $\boldsymbol{\alpha}^i$, concavity of the utility function implies:

$$\begin{aligned} u^i(\mathbf{q}_t, \boldsymbol{\alpha}_t^i) + \nabla_{\mathbf{q}_t} u^i(\mathbf{q}_t, \boldsymbol{\alpha}_t^i)'(\mathbf{q}_s - \mathbf{q}_t) &\geqslant v^i(\mathbf{q}_s) + \boldsymbol{\alpha}_t^{i\prime} \mathbf{q}_s \\ &\geqslant u^i(\mathbf{q}_s) + \boldsymbol{\alpha}_t^{i\prime} \mathbf{q}_s - \boldsymbol{\alpha}_s^{i\prime} \mathbf{q}_s \end{aligned}$$

Substituting the first order conditions into the concavity condition and rearranging gives:

$$v^i(\mathbf{q}_s) - v^i(\mathbf{q}_t) + \boldsymbol{\alpha}_t^{i\prime}(\mathbf{q}_s - \mathbf{q}_t) \leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i)$$

Letting $v_t^i = v^i(\mathbf{q}_t)$, returns the first set of inequalities that constitute Lemma 3.1.

The second set of inequalities are required for the base utility function to be strictly increasing in $\mathbf{q}$. From the first order conditions,

$$\alpha_t^{k,i} > \lambda_t^i p_t^{k,i}$$

would imply

$$\nabla_{\mathbf{q}_t^k} v^i(\mathbf{q}_t) < 0$$

PROOF: SUFFICIENCY Given that we can find scalars $\{v_t^i\}_{t=1,\dots,T}$, $\{\boldsymbol{\alpha}_t^{k,i}\}_{t=1,\dots,T}^{k=1,\dots,K}$ and strictly positive scalars $\{\lambda_t^i\}_{t=1,\dots,T}$ to satisfy the inequalities that define Lemma 3.1, can we find a utility function that rationalises the data set?

We remind the reader that "taste rationalisation" is defined as:

DEFINITION 4.2 Consumer i's choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,\dots,T}$, can be "taste rationalised" by the base utility function $v^i(\mathbf{q})$ and a temporal series of additive linear perturbations to marginal utilities $\{\boldsymbol{\alpha}_t^i\}_{t=1,\dots,T}$ where $\boldsymbol{\alpha}_t^i \in \mathbb{R}^K$ if the following set of inequalities is satisfied:

$$v^i(\mathbf{q}^i) + \boldsymbol{\alpha}_t^{i\prime}\mathbf{q}^i \leqslant v^i(\mathbf{q}_t^i) + \boldsymbol{\alpha}_t^{i\prime}\mathbf{q}_t^i$$

for all $\mathbf{q}^i$ such that:

$$\mathbf{p}_t'\mathbf{q}^i \leqslant \mathbf{p}_t'\mathbf{q}_t^i$$

The concavity condition associated with the taste-varying utility function, $u^i(\mathbf{q}, \boldsymbol{\alpha}_t)$ implies the existence of T overestimates of the utility of some bundle $\mathbf{q}$:

$$\begin{aligned} v^i(\mathbf{q}) &\leqslant v_t^i + \lambda_t^i \mathbf{p}_t'(\mathbf{q} - \mathbf{q}_t) - \boldsymbol{\alpha}_t^{i\prime}(\mathbf{q} - \mathbf{q}_t) \\ v^i(\mathbf{q}) &\leqslant v_t^i + \lambda_t^i \widetilde{\mathbf{p}}_t^{i\prime}(\mathbf{q} - \mathbf{q}_t) \end{aligned}$$

where $\widetilde{\mathbf{p}}_t^{i\prime} = \mathbf{p}_t - \boldsymbol{\alpha}_t^i/\lambda_t^i$, a substitution we make for notational simplicity in what follows.

A piecewise linear utility function can be derived from the lower envelope of the hyperplanes given by these T overestimates:

$$v^i(\mathbf{q}) = \min_t \left\{ v_t^i + \lambda_t^i \widetilde{\mathbf{p}}_t^{i\prime}(\mathbf{q} - \mathbf{q}_t) \right\}$$

The utility of any feasible consumption bundle cannot be strictly greater than that conferred by observed choices with the utility function defined as above. Consider an arbitrary feasible bundle, $\widehat{\mathbf{q}}$:

$$\mathbf{p}_t'\widehat{\mathbf{q}} \leqslant \mathbf{p}_t'\mathbf{q}_t$$

Given our definition of the base individual utility function:

$$v^i(\widehat{\mathbf{q}}) + \boldsymbol{\alpha}_t^{i\prime}\widehat{\mathbf{q}} \leqslant v_t^i + \lambda_t^i \widetilde{\mathbf{p}}_t^{i\prime}(\widehat{\mathbf{q}} - \mathbf{q}_t) + \boldsymbol{\alpha}_t^{i\prime}\widehat{\mathbf{q}}$$

Noting that

$$\lambda_t^i \widetilde{\mathbf{p}}_t^{i\prime}(\widehat{\mathbf{q}} - \mathbf{q}_t) = \lambda_t^i \mathbf{p}_t(\widehat{\mathbf{q}} - \mathbf{q}_t) - \boldsymbol{\alpha}_t^{i\prime}(\widehat{\mathbf{q}} - \mathbf{q}_t)$$

returns

$$\begin{aligned} v^i(\widehat{\mathbf{q}}) + \boldsymbol{\alpha}_t^{i\prime}\widehat{\mathbf{q}} &\leqslant v_t^i + \lambda_t^i \mathbf{p}_t'(\widehat{\mathbf{q}} - \mathbf{q}) + \boldsymbol{\alpha}_t^{i\prime}\mathbf{q}_t \\ &\leqslant u^i(\mathbf{q}_t, \boldsymbol{\alpha}_t^i) + \lambda_t^i \mathbf{p}_t'(\widehat{\mathbf{q}} - \mathbf{q}_t) \end{aligned}$$

Further noting that

$$\begin{aligned} \mathbf{p}_t'\widehat{\mathbf{q}} &\leqslant \mathbf{p}_t'\mathbf{q}_t \\ \mathbf{p}_t'(\widehat{\mathbf{q}} - \mathbf{q}_t) &\leqslant 0 \end{aligned}$$

Implies that

$$v^i(\widehat{\mathbf{q}}) + \boldsymbol{\alpha}_t^i\widehat{\mathbf{q}} \leqslant v^i(\mathbf{q}_t) + \boldsymbol{\alpha}_t^i\mathbf{q}_t$$

In words, any other feasible bundle yields weakly lower utility than $\mathbf{q}_t$. Therefore, we can always construct a utility function which taste rationalises the data set given that a non-empty solution set is associated with the inequalities that define Lemma 4.1. This completes our proof.

THEOREM 4.1 The following statements are equivalent:

1. Individual i's observed choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,\dots,T}$, can be good-1 rationalised by the set of taste shifters $\{\alpha_t^{1i}\}_{t=1,\dots,T}$.

2. One can find sets $\{v_t^i\}_{t=1,\dots,T}$, $\{\alpha_t^{1i}\}_{t=1,\dots,T}$ and $\{\lambda_t^i\}_{t=1,\dots,T}$ with $\lambda_t^i > 0$ for all $t = 1, \dots, T$, such that there exists a non-empty solution set to the following inequalities:

$$\begin{aligned} v_s^i - v_t^i + \alpha_t^{1i}(q_s^{i,1} - q_t^{i,1}) &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \\ \alpha_t^{1i} &\leqslant \lambda_t^i p_t^{i,1} \end{aligned}$$

PROOF. Theorem 3.1 follows from Lemma 3.1 given the restriction:

$$\boldsymbol{\alpha}_t^i = [\alpha_t^{1i}, 0, ..., 0]$$

THEOREM 4.2 Given an individual i's observed choice behaviour, $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$ where good-1 exhibits perfect intertemporal variation, one can *always* find sets $\{v_t^i\}_{t=1,...,T}$, $\{\alpha_t^{1i}\}_{t=1,...,T}$ and $\{\lambda_t^i\}_{t=1,...,T}$ with $\lambda_t^i > 0$ for all $t = 1, ..., T$, such that there exists a non-empty solution set to the following inequalities:

$$\begin{aligned} v_s^i - v_t^i + \alpha_t^{1i}(q_s^{1i} - q_t^{1i}) &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \\ \alpha_t^{1i} &\leqslant \lambda_t^i p_t^{i,1} \end{aligned}$$

PROOF. We prove the above in two stages.

1. GARP AND AFRIAT EQUIVALENCE The virtual price characterisation of taste instability, lends an alternative characterisation of the rationalisation inequalities.

$$\begin{aligned} v_s^i - v_t^i + \alpha^{1i}(q_s^{1i} - q_t^{1i}) &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s - \mathbf{q}_t) \\ v_s^i - v_t^i &\leqslant \lambda_t^i \boldsymbol{\pi}_t^{i\prime}(\mathbf{q}_s - \mathbf{q}_t) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\pi}_t^i &= \begin{bmatrix} p_t^1 - \alpha_t^{1i}/\lambda_t^i \\ \mathbf{p}_t^{-1} \end{bmatrix} \\ \pi^{1i} &\geqslant 0 \end{aligned}$$

Varian (1982) proves that the following conditions are equivalent.

1. A data set $\{\boldsymbol{\pi}_t^i, \mathbf{q}_t^i\}_{t=1,...,T}$ satisfies GARP.
2. There exist numbers $\{v_t^i\}_{t=1,...,T}$ and $\{\lambda_t^i\}_{t=1,...,T}$ with $\lambda_t^i > 0$ such that the following "Afriat" inequalities hold.

$$v_s^i - v_t^i \leqslant \lambda_t^i \boldsymbol{\pi}_t^{i\prime}(\mathbf{q}_s - \mathbf{q}_t)$$

Therefore, if a choice set $\{\boldsymbol{\pi}_t^i, \mathbf{q}_t^i\}_{t=1,...,T}$ satisfies GARP, one is always able to find numbers $\{v_t^i\}_{t=1,...,T}$ and $\{\lambda_t^i\}_{t=1,...,T}$ with $\lambda_t^i > 0$ that satisfy the Afriat inequalities.

2. EXISTENCE OF RATIONALISING $\boldsymbol{\pi}$ We observe the data set $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,...,T}$. Assume that good-1 exhibits perfect intertemporal variation, i.e. $q_t^{1i} \neq q_s^{1i}$ for all $t \neq s$. We proceed by extending Theorem 1 from Varian (1988) to the current setting.

If $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,\ldots,T}$ satisfies GARP, then, following from Varian (1982), the choice set satisfies the inequalities that define Theorem 3.1 with:

$$\alpha_t^i = \mathbf{0}$$

If $\{\mathbf{p}_t, \mathbf{q}_t^i\}_{t=1,\ldots,T}$ fails GARP, then there exist periods s and t such that

$$\begin{aligned} \mathbf{p}_s'\mathbf{q}_s^i &\leqslant \mathbf{p}_s'\mathbf{q}_t^i \\ \mathbf{p}_t'\mathbf{q}_t^i &\leqslant \mathbf{p}_t'\mathbf{q}_s^i \end{aligned}$$

or alternatively,

$$\begin{aligned} \mathbf{p}_s^{-1\prime}\mathbf{q}_s^{-1i} + p_s^1 q_s^{1i} &\leqslant \mathbf{p}_s^{-1\prime}\mathbf{q}_t^{-1i} + p_s^1 q_t^{1i} \\ \mathbf{p}_t^{-1\prime}\mathbf{q}_t^{-1i} + p_t^1 q_t^1 &\leqslant \mathbf{p}_t^{-1\prime}\mathbf{q}_s^{-1i} + p_t^1 q_s^{1i} \end{aligned}$$

Given the assumption of perfect intertemporal variation in good-1, there always exists a set of modifications to p_t^1 such that the GARP inequalities are satisfied. This result follows from Theorem 1 in Varian (1988), in which it is proved that, given perfect intertemporal variation for a good with a missing price, one can always find a price trajectory for this good such that the entire data set satisfies GARP. The new choice rationalising price trajectories in our setting, which are formed from the modifications to p_t^1, have a natural interpretation as taste-change reflecting virtual prices; the consumer is not acting as if the price of good 1 is p_t^1 when making their choices from the perspective of their base preference. Rather they act as if the price is $\pi_t^{1i} = p_t^1 - \alpha_t^{1i}/\lambda_t^i$.

To demonstrate the relevance of the Varian (1988) result in this context, let us consider what value p_t^1 would have to take on, if we were to conjecture that, once taste change is taken into account, the consumer prefers the bundle $\mathbf{q}_t^i$ to $\mathbf{q}_s^i$. This conjecture implies the need to prove the existence of a price π_t^{1i} such that:

$$\begin{aligned} \mathbf{p}_t^{-1\prime}\mathbf{q}_t^{-1i} + \pi_t^{1i} q_t^{1i} &\geqslant \mathbf{p}_t^{-1\prime}\mathbf{q}_s^{-1i} + \pi_t^{1i} q_s^{1i} \\ \pi_t^{1i} &\geqslant \frac{\mathbf{p}_t^{-1\prime}(\mathbf{q}_s^{-1i} - \mathbf{q}_t^{-1i})}{q_t^{1i} - q_s^{1i}} \end{aligned}$$

For each period t, define the lower bound on the virtual price of good 1 such that:

$$\pi_t^{1i} > \max_{s \neq t} \left\{ \frac{\mathbf{p}_t^{-1\prime}(\mathbf{q}_s^{-1i} - \mathbf{q}_t^{-1i})}{q_t^{1i} - q_s^{1i}}, 1 \right\}$$

Further define the "taste adjusted direct revealed preferred relation", $\widetilde{\mathbb{R}}^0$. If $\boldsymbol{\pi}_t^{i\prime}\mathbf{q}_t \geqslant \boldsymbol{\pi}_t^{i\prime}\mathbf{q}_s$, then we conclude that $\mathbf{q}_t$ is directly revealed taste preferred to $\mathbf{q}_s$, or $\mathbf{q}_t \widetilde{\mathbb{R}}^0 \mathbf{q}_s$.

There are then two cases to consider:

1. $q_t^{1i} > q_s^{1i}$: In this case, we must have that

$$\begin{aligned} \pi_t^{1i}(q_t^{1i} - q_s^{1i}) &> \mathbf{p}_t^{\neg 1\prime}(\mathbf{q}_s^{\neg 1i} - \mathbf{q}_t^{\neg 1i}) \\ \pi_t^{1i} q_t^{1i} + \mathbf{p}_t^{\neg 1\prime} \mathbf{q}_t^{\neg 1i} &> \pi_t^{1i} q_s^{1i} + \mathbf{p}_t^{\neg 1\prime} \mathbf{q}_s^{\neg 1i} \\ \boldsymbol{\pi}_t^{i\prime} \mathbf{q}_t^i &> \boldsymbol{\pi}_t^{i\prime} \mathbf{q}_s^i \end{aligned}$$

and set $\mathbf{q}_t^i \widetilde{\mathbb{R}}^0 \mathbf{q}_s^i$.

2. $q_t^{1i} < q_s^{1i}$. In this case, we must have that

$$\begin{aligned} \pi_t^{1i}(q_t^{1i} - q_s^{1i}) &< \mathbf{p}_t^{\neg 1\prime}(\mathbf{q}_s^{\neg 1i} - \mathbf{q}_t^{\neg 1i}) \\ \pi_t^{1i} q_t^{1i} + \mathbf{p}_t^{\neg 1i\prime} \mathbf{q}_t^{\neg 1i} &< \pi_t^{1i} q_s^{1i} + \mathbf{p}_t^{\neg 1\prime} \mathbf{q}_s^{\neg 1i} \\ \boldsymbol{\pi}_t^{i\prime} \mathbf{q}_t^i &< \boldsymbol{\pi}_t^{i\prime} \mathbf{q}_s^i \end{aligned}$$

and thus it is not the case that $\mathbf{q}_t^i \widetilde{\mathbb{R}}^0 \mathbf{q}_s^i$.

Therefore, one can determine the preference ordering of consumption bundles solely by reference to the quantity of good-1 consumed and set the taste adjusted price of good 1 to dominate the impact of revealed preference violations in the unadjusted data set. The choice set $\{\boldsymbol{\pi}_t^i, \mathbf{q}_t^i\}_{t=1,...,T}$ then passes GARP.

INTEGRATING THE RESULTS Varian's (1988) result, applied above to the current setting, gives us that we can identify a lower bound on a set of choice-rationalising virtual prices for good-1, i.e. given the requirement of perfect intertemporal variation in good-1, there always exists a set of virtual prices, $\boldsymbol{\pi}_t^i$, that is unbounded above such that the choice set $\{\boldsymbol{\pi}_t^i, \mathbf{q}_t^i\}_{t=1,...,T}$ then passes GARP.

Given the equivalence of GARP and a non-empty feasible set to the standard Afriat inequalities, this result then implies that for for any element of the rationalising price set, $\{\widetilde{\boldsymbol{\pi}}_t^i\}_{t=1,...,T}$, there exist numbers $\{\widetilde{v}_t^i\}_{t=1,...,T}$ and $\{\widetilde{\lambda}_t^i\}_{t=1,...,T}$ with $\lambda_t^i > 0$ such that the following inequalities hold.

$$\widetilde{v}_s^i - \widetilde{v}_t^i \leqslant \widetilde{\lambda}_t^i \widetilde{\boldsymbol{\pi}}_t^{i\prime}(\mathbf{q}_s - \mathbf{q}_t)$$

An element of the set of choice rationalising taste shifters associated with $\{\widetilde{\boldsymbol{\pi}}_t^i\}_{t=1,...,T}$ can then be constructed as:

$$\widetilde{\alpha}_t^{i1} = \widetilde{\lambda}_t^i(p_t^1 - \widetilde{\pi}_t^{1i})$$

for $t = 1, ..., T$.

The fact that the set of rationalising π_t^{1i} is unbounded above implies that the taste modification to virtual prices, or equivalently, the change in the marginal willingness to pay for good-1, $\alpha_t^{1i}/\lambda_t^i$, is unbounded below. Only a lower bound is identified upon the choice rationalising value of π_t^{1i},

$$\pi_t^{1i} > \max_{s \neq t} \left\{ \frac{\mathbf{p}_t^{\neg 1\prime}(\mathbf{q}_s^{\neg 1i} - \mathbf{q}_t^{\neg 1i})}{q_t^{1i} - q_s^{1i}}, 1 \right\}$$

Thus,

$$\begin{aligned}\frac{\alpha_t^{1i}}{\lambda_t^i} &= p_t^1 - \pi_t^{1i} \\ &\to -\infty\end{aligned}$$

as $\pi_t^{1i} \to \infty$.

This completes the proof of Theorem 4.2 and the note that the set of rationalising taste wedges, $\{\alpha_t^{1i}/\lambda_t^i\}_{t=1,\ldots,T}$ is unbounded below.

C.2 LIFECYCLE RESULTS

This section outlines the theoretical results that underlie the rationalisability inequalities imposed by the CBU Lifecycle Quadratic Programme. These results are found only in the Appendix given that they constitute a marginal extension to our primary theoretical framework and are not of central importance to our argument.

RATIONALISATION CONDITIONS Extending our theoretical framework to explicitly model the allocation of lifetime income across periods, gives rise to the following model for individual i's decision making problem:

$$\max_{\{\mathbf{q}_t^i\}_{t=1,\ldots,T}} \sum_{t=1}^{T} \nu^i(\mathbf{q}_t^i) + \alpha_t^i q_t^{1i}$$

subject to:

$$\sum_{t=1}^{T} \boldsymbol{\rho}_t' \mathbf{q}_t^i = A_0^i$$

where $\boldsymbol{\rho}_t = \mathbf{p}_t / \prod_{s=1}^{t}(1+i_s) = \mathbf{p}_t \delta_t$ and A_0^i is lifetime wealth.

Given the assumption of local nonsatiation, an optimal interior solution to the problem must satisfy:

$$\begin{aligned}\nabla_{q_t^1} \nu^i(\mathbf{q}_t) + \alpha_t^i &= \lambda^i \rho_t^1 \\ \nabla_{\mathbf{q}_t^{-1}} \nu^i(\mathbf{q}_t) &= \lambda^i \boldsymbol{\rho}_t^{-1}\end{aligned}$$

Substituting the first order conditions into those conditions implied by concavity of the utility function (outlined in the proof of Lemma 3.1) and rearranging then gives:

$$\nu^i(\mathbf{q}_s) - \nu^i(\mathbf{q}_t) + \alpha_t^{1i}(q_s^{1i} - q_t^{1i}) \leqslant \lambda^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i)$$

These conditions are closely related to those of the static model we consider. However, there are three central differences. Firstly, these constraints are parameterised by discounted prices. Secondly, the marginal utility of discounted income is restricted to a constant in this setting,

$\lambda_t^i = \lambda^i$. If λ^i varied across time periods then an individual would regret certain consumption decisions ex post. Finally, and most importantly, in contrast to the rationalisation conditions associated with our primary theoretical framework, these lifecycle conditions are rejectable in survey data. Deviations of intertemporal spending allocations from those prescribed by the lifecycle model cannot always be rationalised by reference to taste change on a single good; innovations in λ^i are also required.

This is most easily demonstrated by proving that one cannot reduce the inequalities that define a lifecycle rationalisation into an analogous format to those which define Theorem 3.1.

$$\begin{aligned} \nu^i(\mathbf{q}_s) - \nu^i(\mathbf{q}_t) + \alpha_t^{1i}(q_s^{1i} - q_t^{1i}) &\leqslant \lambda_t^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \\ &\leqslant \lambda^i \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) + \epsilon_t \mathbf{p}_t'(\mathbf{q}_s^i - \mathbf{q}_t^i) \end{aligned}$$

One is unable to eliminate the final term from the above expression and thus, Theorem 4.2 cannot be applied. Simulation evidence also provides a counter example to the assertion that a choice set can always be lifecycle rationalised.

C.3 DATA DESCRIPTION

This section provides details of the goods that comprise our nondurables and tobacco good aggregates and provides descriptive statistics on nominal total nondurable expenditures. The SMP total expenditure levels are also listed.

COMMODITY AGGREGATES

- **Tobacco**: the "tobacco" aggregate is comprised of the following sub-groups: Cigarettes; Other Tobacco.
- **Nondurables**: the "nondurable" aggregate is comprised of: Bread; Cereals; Biscuits; Beef; Lamb; Pork; Bacon; Poultry; Fish; Butter; Oils; Cheese; Eggs; Fresh Milk;Processed Milk; Tea; Coffee; Soft Drinks; Sugar; Sweets; Potatoes; Other Vegetables; Fruit; Other Food; Canteen; Other Snacks; Coal; Electric; Gas; Oil; Household Consumables; Pet Care; Postage; Telephone; Domestic Services; Chemical products; Petrol; Rail Fares; Bus Fares; Other Travel; Toys; New Books; Entertainment.
- **Alcohol**: the "alcohol" aggregate is comprised of: Beer; Wine and Spirits.

TOTAL NONDURABLE EXPENDITURE

Table 31: Total and SMP Nondurable Expenditures: Low Education

Year	Mean	St. Dev	SMP
1980	110.3	59.2	98.4
1982	138.7	95.6	124.1
1984	150.5	80.8	139.5
1986	192.2	123.1	154.5
1988	227.6	143.9	168.2
1990	283.8	189.0	194.6
1992	307.8	187.0	226.0
1994	307.4	206.6	249.0
1996	327.5	208.2	265.0
1998	346.5	220.8	291.8
2000	380.0	246.0	315.75

Table 32: Total and SMP Nondurable Expenditures: High Education

Year	Mean	St. Dev	SMP
1980	131.2	72.5	115.0
1982	169.2	97.6	145.7
1984	192.1	115.5	162.7
1986	246.0	198.0	177.2
1988	294.8	206.2	195.9
1990	371.6	237.6	229.4
1992	432.6	340.9	278.5
1994	420.7	255.2	289.9
1996	446.8	278.4	312.7
1998	501.2	337.5	343.4
2000	524.3	354.6	376.1

C.4 QUANTILE REGRESSION COEFFICIENTS

The following table charts the results of the censored quantile regression estimation procedure. Confidence intervals for quantile coefficients were constructed using a boostrap procedure over 1000 random samples drawn with replacement within each psuedo cohort-time cell. The SMP quantities, and their bootstrapped 95% confidence intervals, were given in the main text in Table 29.

Table 33: Control Variable Estimates

Year		Low Ed.		High Ed.	
		β^L	$se(\beta^L)$	β^H	$se(\beta^H)$
1980	cons.	1.9684	0.1470	1.7180	0.1593
	log(m)	0.5258	0.0361	0.5685	0.0330
	oecd	0.1554	0.0499	0.2515	0.0548
1982	cons.	1.2220	0.1550	1.46148	0.1512
	log(m)	0.6900	0.0327	0.6628	0.0299
	oecd	0.1880	0.0472	0.7018	0.0470
1984	cons.	1.0783	0.1394	1.2454	0.1381
	log(m)	0.7199	0.0304	0.7018	0.0278
	oecd	0.2166	0.0473	0.2133	0.0439
1986	cons.	1.1512	0.1575	2.0460	0.1324
	log(m)	0.6948	0.0324	0.5434	0.0242
	oecd	0.2831	0.0525	0.2928	0.0521
1988	cons.	1.4089	0.1371	1.7769	0.1459
	log(m)	0.7007	0.0287	0.5983	0.0271
	oecd	0.1018	0.0502	0.2942	0.0485
1990	cons.	1.5798	0.1419	2.0610	0.1364
	log(m)	0.6523	0.0285	0.5663	0.0242
	oecd	0.2428	0.0450	0.3311	0.0450
1992	cons.	2.6210	0.1372	1.9392	0.1335
	log(m)	0.4203	0.0270	0.5954	0.0240
	oecd	0.4548	0.0525	0.3045	0.0445
1994	cons.	1.6444	0.1420	2.1454	0.1337
	log(m)	0.6403	0.0301	0.5534	0.0229
	oecd	0.2637	0.0572	0.3307	0.0400
1996	cons.	1.5694	0.1505	1.6257	0.1321
	log(m)	0.6256	0.0295	0.6543	0.0235
	oecd	0.3953	0.0515	0.2973	0.0430
1998	cons.	1.3023	0.1362	2.4187	0.1434
	log(m)	0.7172	0.0276	0.5127	0.0253
	oecd	0.2212	0.0501	0.3738	0.0486
2000	cons.	2.4627	0.1412	2.5863	0.1327
	log(m)	0.4532	0.0273	0.4858	0.0240
	oecd	0.5477	0.0694	0.4051	0.0493

Table 34: Quantile Coefficients & 95% Confidence Interval: 1980-82

Year		Low Education			High Education		
		0.55	0.65	0.75	0.55	0.65	0.75
1980	cons.	0.0967	0.1127	-0.1353	-0.1531	-0.1215	-0.1792
		-0.3469	-0.2723	-0.1263	-0.6733	-0.6295	-0.5971
		0.4022	0.5460	0.5690	0.0375	0.0790	0.1508
	$\log(x)$	-0.0005	0.0019	0.0026	0.0805	0.0813	0.1362
		-0.1223	-0.1703	-0.1712	-0.0024	0.0033	0.0006
		0.1887	0.1759	0.1128	0.2818	0.2988	0.2854
	$\log(x)^2$	-0.0018	-0.0021	−0.0024	-0.0125	-0.0134	-0.0205
		-0.0229	-0.0221	-0.0149	-0.0316	-0.0349	-0.0346
		0.0099	0.0136	0.0137	-0.0030	-0.0060	-0.0058
	oecd	0.0230	0.0181	0.0182	0.0349	0.0394	0.0266
		0.0024	0.0012	0.0011	0.0046	0.0041	0.0016
		0.0413	0.0393	0.0372	0.0567	0.0749	0.0638
	$\widehat{\nu}$	-0.0362	-0.0360	−0.0355	0.0658	0.0656	0.0654
		-0.0713	-0.0645	-0.0657	0.0043	0.0067	0.0005
		0.0054	0.0048	0.0047	0.0977	0.1214	0.0924
1982	cons.	0.3749	0.6335	0.4882	-0.0487	0.0758	0.1294
		0.0536	0.1296	0.1889	-0.9070	-0.5363	-0.2996
		0.7679	0.8840	0.8623	0.2744	0.2731	0.3642
	$\log(x)$	-0.1202	-0.1914	-0.1227	0.0323	0.0128	-0.0025
		-0.2661	-0.2887	-0.2724	-0.0969	-0.0689	-0.0854
		0.0166	0.0016	-0.0052	0.3587	0.2425	0.1822
	$\log(x)^2$	0.0076	0.0131	0.0061	-0.0050	-0.0066	-0.0049
		-0.0077	-0.0065	-0.0045	-0.0405	-0.0289	-0.0239
		0.0203	0.0222	0.0209	0.0076	0.0028	0.0039
	oecd	0.0477	0.0310	0.0328	0.0077	0.0186	0.0278
		0.0164	0.0137	0.0141	-0.0104	-0.0021	-0.0026
		0.0698	0.0578	0.0539	0.1284	0.0489	0.0423
	$\widehat{\nu}$	0.0386	0.0423	0.0374	0.0230	0.0731	0.0601
		0.0386	0.0106	0.0067	0.0001	0.0267	0.0145
		0.0817	0.0720	0.0680	0.1698	0.0988	0.0926

Table 35: Quantile Coefficients & 95% Confidence Interval: 1984-86

Year		Low Education			High Education		
		0.55	0.65	0.75	0.55	0.65	0.75
1984	cons.	0.2532 0.0444 0.5049	0.4036 0.0397 0.5275	0.3186 0.0960 0.5397	0.1163 -0.7207 0.5239	0.3254 -0.1314 0.6084	0.2877 -0.1235 0.6458
	log(x)	-0.0568 -0.1592 0.0366	-0.0957 -0.1531 0.0540	-0.550 -0.1431 0.0431	-0.0388 -0.1851 0.3099	-0.0905 -0.2003 0.0837	-0.0583 -0.1968 0.1041
	$\log(x)^2$	0.0009 -0.0095 0.0114	0.0037 -0.0119 0.0105	-0.0004 -0.0108 0.0098	0.0015 -0.0404 0.0127	0.0014 -0.0146 0.0135	-0.0007 -0.0165 0.0130
	oecd	0.0296 0.0057 0.0672	0.0214 0.0023 0.0488	0.0278 -0.0061 0.0445	0.0257 0.0023 0.1228	0.0625 0.0092 0.1040	0.0376 -0.0011 0.0694
	$\widehat{v}$	0.0397 0.0041 0.0851	0.0563 0.0093 0.0864	0.0532 0.0083 0.0828	0.0347 0.0050 0.1169	0.0905 0.0196 0.1380	0.0739 0.0158 0.1128
1986	cons.	0.1862 -0.0310 0.3369	0.2269 -0.0171 0.4169	0.2804 -0.0151 0.4541	0.0499 -0.2015 0.3252	0.1284 -0.0840 0.4103	0.2347 -0.0207 0.4704
	log(x)	-0.0196 -0.0813 0.0712	-0.0263 -0.0921 0.0764	-0.0342 -0.0981 0.0889	-0.0046 -0.1025 0.1389	-0.0008 -0.1094 0.0953	-0.0288 -0.1118 0.0734
	$\log(x)^2$	-0.0011 -0.0104 0.0053	-0.0002 -0.0110 0.0057	0.0001 -0.0128 0.0061	-0.0014 -0.0204 0.0079	-0.0040 -0.0170 0.0058	-0.0021 -0.0127 0.0052
	oecd	0.0074 -0.0098 0.0316	0.0039 -0.0177 0.0225	-0.0019 -0.0199 0.0235	0.0060 -0.0069 0.0553	0.0044 -0.0019 0.0627	0.0153 -0.0023 0.0488
	$\widehat{v}$	-0.0068 -0.0551 0.0317	-0.0193 -0.0509 0.0251	0.0002 -0.0330 0.0319	0.0274 -0.0044 0.0918	0.0231 -0.0052 0.0944	0.0241 -0.0110 0.0843

Table 36: Quantile Coefficients & 95% Confidence Interval: 1988-90

Year		Low Education			High Education		
		0.55	0.65	0.75	0.55	0.65	0.75
1988	cons.	0.2820 -0.0826 0.5967	0.2855 -0.0598 0.6701	0.3291 -0.0212 0.6865	0.0000 -0.1446 0.3219	0.1196 -0.0370 0.6218	0.2580 0.0064 0.7060
	log(x)	-0.0432 -0.1633 0.0927	-0.0423 -0.1834 0.0940	-0.0461 -0.1836 0.0833	0.0000 -0.1142 0.1050	-0.0341 -0.1985 0.0244	-0.0492 -0.2040 0.0386
	$\log(x)^2$	-0.0004 -0.0135 0.0110	-0.0002 -0.0136 0.0130	-0.0004 -0.0126 0.0127	0.0000 -0.0113 0.0092	-0.0019 -0.0051 0.0151	-0.0011 -0.0072 0.0141
	oecd	0.0033 -0.0099 0.0242	0.0082 -0.0089 0.0185	0.0053 -0.0082 0.0215	0.0000 -0.0007 0.0407	0.0159 -0.0009 0.0532	0.0186 -0.0015 0.0451
	$\widehat{v}$	0.0210 -0.0047 0.0572	0.0108 -0.0066 0.0342	0.0114 -0.0093 0.0412	0.0000 -0.0059 0.0273	0.0029 -0.0083 0.0614	0.0005 -0.0140 0.0729
1990	cons.	0.1184 -0.1417 0.6238	0.4394 -0.0799 0.7434	0.3045 -0.0382 0.6668	0.3008 -0.1249 0.5594	0.3984 -0.0056 0.8296	0.4295 0.1024 0.8057
	log(x)	0.0124 -0.1677 0.1050	-0.0915 -0.2045 0.1053	-0.0310 -0.1618 0.1005	-0.0992 -0.1796 0.0547	-0.1159 -0.2485 0.0189	-0.0954 -0.2295 0.0034
	$\log(x)^2$	-0.0052 -0.0138 0.0110	0.0040 -0.0144 0.0139	-0.0022 -0.0147 0.0097	0.0079 -0.0100 0.0127	0.0061 -0.0060 0.0173	0.0021 -0.0059 0.0146
	oecd	-0.0002 -0.0177 0.0232	-0.0038 -0.0155 0.0190	-0.0026 -0.0218 0.0152	0.0032 -0.0007 0.0509	0.0304 -0.0000 0.0633	0.0312 -0.0021 0.0554
	$\widehat{v}$	0.0504 0.0082 0.0896	0.0397 0.0139 0.0800	0.0566 0.0048 0.0815	0.0142 -0.0013 0.1458	0.0825 0.0302 0.1460	0.1014 -0.0329 0.1538

Table 37: Quantile Coefficients & 95% Confidence Interval: 1992-94

Year		Low Education			High Education		
		0.55	0.65	0.75	0.55	0.65	0.75
1992	cons.	0.0471 -0.4000 0.3492	0.2120 -0.2615 0.4146	0.3259 -0.0918 0.4834	-0.1912 -0.7613 0.1975	-0.0374 -0.4987 0.2067	-0.0313 -0.3819 0.2329
	log(x)	0.0425 -0.0823 0.2018	-0.0018 -0.0816 0.1721	-0.0361 -0.0903 0.1190	0.0840 -0.0607 0.3052	0.0490 -0.0367 0.2201	0.0856 -0.0225 0.1966
	$\log(x)^2$	-0.0123 -0.0265 0.0027	-0.0076 -0.0238 0.0023	-0.0031 -0.0193 0.0028	-0.0102 -0.0323 0.0140	-0.0092 -0.0260 -0.0044	-0.0142 -0.0241 0.0146
	oecd	0.0601 0.0012 0.0969	0.0420 -0.0044 0.0789	0.0273 -0.0060 0.0551	0.0308 0.0010 0.0887	0.0444 0.0264 0.0867	0.0555 0.0204 0.0785
	$\widehat{v}$	0.1300 0.0063 0.2168	0.1048 0.0043 0.1759	0.0750 -0.0004 0.1253	0.0307 -0.0000 0.0779	0.0570 0.0068 0.0925	0.0521 0.0014 0.0835
1994	cons.	-0.1912 -0.7613 0.1975	-0.0374 -0.4987 0.2067	-0.0313 -0.3819 0.2329	0.5142 -0.000 1.0000	0.8856 0.1685 0.9674	0.4854 0.2938 0.7798
	log(x)	-0.1912 -0.7613 0.1975	-0.0374 -0.4987 0.2067	-0.0313 -0.3819 0.2329	-0.1657 -0.3189 0.0000	-0.2683 -0.3008 -0.0310	-0.1056 -0.2154 -0.0493
	$\log(x)^2$	-0.1912 -0.7613 0.1975	-0.0374 -0.4987 0.2067	-0.0313 -0.3819 0.2329	0.0125 -0.0001 0.0233	0.0182 -0.0013 0.0223	0.0034 -0.0016 0.0132
	oecd	-0.1912 -0.7613 0.1975	-0.0374 -0.4987 0.2067	-0.0313 -0.3819 0.2329	0.0124 0.0000 0.0414	0.0263 -0.0001 0.0395	0.0087 0.0008 0.0385
	$\widehat{v}$	-0.1912 -0.7613 0.1975	-0.0374 -0.4987 0.2067	-0.0313 -0.3819 0.2329	0.0246 -0.0000 0.0956	0.0770 0.0150 0.1110	0.0839 0.0301 0.1168

Table 38: Quantile Coefficients & 95% Confidence Interval: 1996-98

Year		Low Education			High Education		
		0.55	0.65	0.75	0.55	0.65	0.75
1996	cons.	-0.0939 -0.8316 0.1709	-0.0539 -0.5181 0.2551	-0.0394 -0.2834 0.2883	0.0548 -0.6701 0.8361	0.1063 -0.2597 1.0000	0.4854 -0.0192 1.0412
	log(x)	0.0640 -0.0187 0.3459	0.0848 -0.0346 0.2432	0.0870 -0.0345 0.1892	0.0063 -0.2624 0.3557	0.0507 -0.3067 0.1752	-0.1056 -0.2894 0.0868
	$\log(x)^2$	-0.0081 -0.0358 -0.0010	-0.0117 -0.0264 -0.0005	-0.0122 -0.0218 -0.0007	-0.0037 -0.0467 0.0190	-0.0135 -0.0254 0.0222	0.0034 -0.0254 0.0187
	oecd	0.0062 -0.0064 0.0537	0.0081 -0.0102 0.0503	0.0121 -0.0103 0.0365	0.0250 -0.0146 0.1240	0.0515 0.0119 0.0941	0.0087 -0.0090 0.0663
	$\widehat{\nu}$	0.0445 0.0012 0.1186	0.0279 -0.0019 0.0918	0.0460 0.0020 0.0721	0.0257 -0.0210 0.0472	0.0357 -0.0090 0.0788	0.0839 -0.0033 0.0705
1998	cons.	0.4040 -0.2775 0.4586	0.1319 -0.1659 0.5747	0.2495 -0.1475 0.5496	0.3933 -1.7032 0.4980	0.2558 -0.6889 0.5615	0.3649 -0.0596 0.7166
	log(x)	-0.1026 -1.1268 0.1329	0.0092 -0.1363 0.1208	0.0048 -0.1005 0.1383	-0.1263 -0.1584 0.5668	-0.0734 -0.1623 0.2405	-0.0589 -0.1838 0.0701
	$\log(x)^2$	0.0031 -0.0165 0.0064	-0.0065 -0.0174 0.0058	-0.0075 -0.0185 0.0024	0.0096 -0.0490 0.0116	0.0039 -0.0222 0.0109	-0.0028 -0.0132 0.0084
	oecd	0.0511 0.0005 0.0895	0.0486 0.0016 0.0793	0.0331 -0.0010 0.0559	0.0101 -0.0000 0.0288	0.0287 0.0035 0.0502	0.0665 0.0105 0.0816
	$\widehat{\nu}$	0.0970 0.0072 0.1474	0.0810 0.0176 0.1432	0.0694 0.0057 0.1056	0.0134 -0.0000 0.0674	0.0310 0.0050 0.1257	0.1032 0.0372 0.1834

Table 39: Quantile Coefficients & 95% Confidence Interval: 2000

Year		Low Education			High Education		
		0.55	0.65	0.75	0.55	0.65	0.75
2000	cons.	-0.1190	-0.3589	0.4057	-1.2069	-0.3924	0.0437
		-0.3305	-0.2240	-0.1073	-3.2298	-1.7068	-0.2504
		0.7665	0.8019	0.8057	0.2417	0.2789	0.4474
	$\log(x)$	0.0655	-0.0618	-0.0599	-0.0012	-0.0541	-0.0185
		-0.2286	-0.2159	-0.1979	-0.0784	-0.0821	-0.0958
		0.1481	0.1284	0.1231	1.0600	0.6111	0.1283
	$\log(x)^2$	-0.0084	0.0005	-0.0011	-0.0472	-0.0177	-0.0065
		-0.0158	-0.0155	-0.0159	-0.1061	-0.0641	-0.0192
		0.0174	0.0140	0.0122	0.0052	0.0041	0.0017
	oecd	0.0339	0.0165	0.0286	0.1489	0.0605	0.0492
		-0.0037	-0.0051	-0.0098	-0.0000	0.0032	0.0231
		0.0662	0.0705	0.0519	0.2744	0.1928	0.1156
	$\widehat{v}$	0.0332	0.0497	0.0640	0.2048	0.0856	0.1054
		-0.0143	-0.0099	-0.0083	-0.0000	0.0047	0.0505
		0.0912	0.1142	0.1062	0.4453	0.2847	0.2234

C.5 TASTE CHANGE PARAMETERS

The following tables give the recovered taste parameters that underlie the figures in the main text.

C.5.1 *Unconditional estimates*

The following tables give recovered taste shifters for the 2-good quantile case (i.e. without conditioning upon alcohol consumption).

Table 40: Minimal Smoothed Taste Wedge with 95% Confidence Interval

Year	π^L			π^H		
	Light	Mod.	Heavy	Light	Mod.	Heavy
1980	-1.96	-0.97	0.00	-3.39	-1.87	-0.06
	-1.29	-0.54	0.00	-2.52	-0.81	0.23
	-2.30	-1.39	0.00	-3.59	-2.62	-0.24
1982	-3.45	-2.31	-0.71	-4.42	-3.04	-0.78
	-2.81	-1.55	-0.26	-3.80	-1.92	-0.01
	-3.65	-2.81	-1.48	-4.87	-3.84	-1.42
1984	-3.59	-2.54	-0.98	-5.25	-4.87	-1.54
	-2.92	-1.66	-0.39	-3.20	-3.47	-0.29
	-4.01	l	-1.77	-5.34	-5.17	-3.19
1986	-3.66	-2.57	-1.19	-5.73	-5.32	-1.57
	-3.00	-1.80	-0.45	-5.24	-3.22	-0.43
	-4.21	-3.51	-1.90	-5.77	-5.57	-3.03
1988	-3.72	-2.60	-1.39	-6.23	-5.56	-1.62
	-3.10	-1.97	-0.53	-5.73	-4.31	-0.60
	-4.36	-3.77	-2.03	-6.30	-5.88	-3.40
1990	-3.85	-3.12	-1.53	-7.08	-5.74	-2.58
	-3.20	-2.07	-0.54	-6.58	-5.20	-0.66
	-4.65	-4.06	-2.25	-7.30	-6.37	-4.44
1992	-5.61	-3.91	-1.64	-8.27	-6.91	-3.86
	-4.34	-2.42	-0.57	-7.46	-5.99	-1.65
	-5.93	-4.42	-2.55	-8.71	-7.92	-5.75
1994	-6.03	-4.47	-1.75	-8.73	-7.78	-4.37
	-4.90	-2.78	-0.57	-7.78	-6.62	-1.99
	-6.49	-4.91	-2.91	-8.93	-8.27	-6.30
1996	-6.20	-4.98	-1.93	-8.74	-8.57	-4.73
	-5.07	-3.26	-0.59	-7.97	-7.24	-2.30
	-6.45	-5.52	-3.32	-9.36	-8.96	-6.71
1998	-7.19	-5.34	-2.23	-9.89	-9.37	-5.39
	-5.86	-3.82	-0.61	-9.00	-8.03	-2.87
	-7.33	-6.21	-3.78	-10.22	-9.86	-7.40
2000	-7.20	-5.52	-4.17	-10.45	-10.33	-7.84
	-6.00	-4.42	-2.29	-9.67	-9.27	-5.50
	-8.22	-6.95	-4.40	-11.00	-10.85	-9.50

Table 41: Minimal Raw Taste Wedge with 95% Confidence Interval

Year	π^L			π^H		
	Light	Mod.	Heavy	Light	Mod.	Heavy
1980	-1.96	-0.97	0.00	-3.39	-1.87	-0.06
	-1.29	-0.54	0.00	-2.52	-0.81	0.23
	-2.30	-1.39	0.00	-3.59	-2.62	-0.24
1982	-3.45	-2.31	-0.71	-4.42	-3.04	-0.78
	-2.81	-1.55	-0.26	-3.80	-1.92	-0.01
	-3.65	-2.81	-1.48	-4.87	-3.84	-1.42
1984	-3.59	-2.54	-0.98	-5.25	-4.87	-1.54
	-2.92	-1.66	-0.39	-3.20	-3.47	-0.29
	-4.01	-3.20	-1.77	-5.34	-5.17	-3.19
1986	-3.22	-1.92	-0.16	-5.16	-3.54	-0.79
	-2.39	-1.66	-0.05	-4.60	-3.22	-0.29
	-4.01	-3.20	-1.77	-5.34	-5.17	-3.03
1988	-3.46	-2.60	-1.39	-6.23	-5.56	-1.62
	-2.39	-0.98	-0.30	-4.42	-2.75	-0.06
	-3.63	-2.92	-2.03	-6.30	-5.89	-3.40
1990	-3.85	-2.18	-0.70	-7.08	-5.74	-0.87
	-2.68	-1.84	-0.05	-5.45	-4.74	-0.60
	-4.65	-3.01	-0.86	-7.30	-6.37	-2.86
1992	-5.61	-3.91	-1.38	-8.27	-6.91	-3.86
	-2.80	-0.89	0.78	-6.14	-3.19	0.36
	-5.93	-4.42	-2.37	-8.71	-7.92	-5.75
1994	-6.03	-3.63	-1.19	-8.73	-7.78	-3.02
	-4.34	-2.42	-0.13	-7.45	-5.98	-1.64
	-6.49	-4.91	-2.43	-8.93	-8.27	-5.02
1996	-6.20	-3.72	-1.19	-8.74	-7.37	-3.00
	-5.07	-2.49	-0.01	-7.97	-5.15	-1.27
	-6.45	-5.19	-2.20	-9.36	-8.96	-4.37
1998	-7.19	-5.34	-2.23	-9.89	-9.37	-5.39
	-5.86	-3.82	-0.61	-9.00	-8.03	-2.87
	-7.33	-6.21	-3.78	-10.22	-9.86	-7.40
2000	-6.80	-5.52	-4.17	-10.45	-10.33	-7.84
	-6.02	-4.42	-2.29	-9.67	-9.27	-5.50
	-7.22	-6.16	-4.40	-11.00	-10.85	-9.49

Table 42: Minimal α with 95% Confidence Interval

Year	π^L			π^H		
	Light	Mod.	Heavy	Light	Mod.	Heavy
1980	-59.62	-22.51	0.00	-80.33	-29.06	-2.97
	-34.44	-11.82	0.00	-47.48	-11.87	0.23
	-85.91	-38.05	0.00	-103.40	-49.00	4.84
1982	-117.08	-60.10	-14.08	-100.36	-43.09	-11.46
	-88.32	-36.54	-5.31	-69.79	-23.64	-0.01
	-134.13	-87.17	-34.70	-144.02	-70.74	-18.53
1984	-106.39	-61.91	-18.90	-126.90	-98.90	-18.64
	-81.20	-36.65	-2.35	-95.67	-47.62	-4.41
	-134.14	-92.39	-38.67	-144.38	-130.20	-42.90
1986	-80.73	-39.77	-3.99	-91.89	-44.38	-11.21
	-55.57	-18.57	-18.19	-66.33	-31.23	-2.52
	-102.17	-58.77	-18.	-117.85	-74.59	-27.72
1988	-79.38	-53.27	-24.60	-126.84	-87.00	-17.71
	-57.18	-34.92	-9.39	-95.63	-57.15	-7.38
	-99.63	-65.68	-40.39	-140.07	-115.37	-37.23
1990	-73.64	-34.63	-2.81	-115.87	-62.95	-10.49
	-48.81	-12.43	-8.23	-78.35	-34.45	-0.77
	-102.32	-53.65	-13.46	-136.20	-86.70	-25.32
1992	-109.43	-66.10	-20.11	-106.50	-64.56	-29.10
	-76.72	-36.01	-3.54	-82.76	-49.67	-12.86
	-123.87	-82.84	-36.79	-132.76	-97.23	-48.27
1994	-108.06	-55.52	-16.98	-115.83	-78.09	-23.78
	-81.01	-36.81	-1.61	-86.36	-48.53	-13.67
	-125.57	-86.28	-36.08	-137.82	-103.28	-41.33
1996	-105.21	-56.07	-17.00	-109.58	-62.73	-23.08
	-80.95	-34.97	-1.43	-74.00	-37.42	-9.69
	-121.65	-87.28	-31.91	-139.16	-106.63	-33.82
1998	-112.72	-79.16	-30.02	-115.41	-82.27	-38.41
	-86.60	-53.89	-8.79	-81.87	-62.29	-18.74
	-124.89	-99.70	-54.18	-144.67	-103.08	-58.27
2000	-98.27	-77.75	-58.29	-95.52	-88.42	-53.95
	-81.55	-60.31	-30.09	-75.00	-67.09	-35.71
	-114.61	-94.79	-65.05	-143.64	-100.98	-74.53

c.5.2 *Conditional estimates*

The following tables document the results on the taste wedge once we stratify by alcohol group.

Table 43: Low Alcohol: Smoothed Taste Wedge & 95% Confidence Interval

Year	π^L		π^H	
	Mod.	Heavy	Mod.	Heavy
1980	-1.35	-0.33	-3.46	-1.54
	-0.15	0.02	-2.19	-0.32
	-2.30	-1.11	-3.80	-3.02
1982	-3.48	-0.76	-4.49	-2.81
	-1.58	-0.06	-3.20	-0.51
	-3.85	-2.49	-4.78	-4.02
1984	-3.62	-1.16	-4.97	-4.83
	-1.62	-0.12	-4.12	-3.33
	-3.98	-2.60	-5.32	-5.11
1986	-3.66	-1.53	-5.04	-5.04
	-1.84	-0.39	-4.50	-3.34
	-4.02	-2.84	-5.38	-5.38
1988	-3.71	-1.70	-5.71	-5.24
	-2.24	-0.64	-4.86	-3.35
	-4.14	-3.24	-6.13	-5.67
1990	-3.85	-1.71	-6.79	-6.23
	-3.07	-0.91	-5.68	-3.35
	-4.95	-3.99	-7.21	-6.84
1992	-5.42	-3.25	-8.01	-6.54
	-4.00	-1.12	-6.86	-3.35
	-6.06	-4.84	-8.55	-7.67
1994	-5.69	-3.26	-8.07	-7.60
	-4.64	-1.23	-7.18	–5.82
	-6.33	-5.45	-8.68	-8.24
1996	-6.14	-4.00	-8.59	-8.49
	-5.18	-1.31	-7.28	-7.09
	-6.75	-5.98	-9.08	-8.50
1998	-6.77	-5.79	-9.21	-9.20
	-5.58	-1.49	-8.25	-7.92
	-7.31	-6.40	-9.85	-9.73
2000	-6.77	-5.79	-9.66	-9.64
	-5.80	-4.15	-8.75	-8.44
	-8.02	-6.67	-10.45	-10.29

Table 44: High Alcohol: Smoothed Taste Wedge & 95% Confidence Interval

Year	π^L		π^H	
	Mod.	Heavy	Mod.	Heavy
1980	-1.50	0.00	-0.69	0.05
	-0.63	0.00	-0.07	0.78
	-1.85	0.00	-1.74	-0.23
1982	-2.92	-1.40	-3.03	-0.90
	-1.63	-0.34	-1.62	0.25
	-3.44	-2.47	-3.83	-2.02
1984	-3.26	-1.64	-3.37	-1.07
	-1.76	-0.34	-1.78	0.02
	-3.79	-2.52	-4.71	-2.27
1986	-3.35	-1.74	-3.77	-1.23
	-1.82	-0.34	-1.95	-0.14
	-3.84	-2.86	-5.33	-2.50
1988	-3.40	-1.98	-4.82	-1.81
	-1.86	-0.35	-3.39	-0.40
	-3.96	-2.99	-5.63	-3.47
1990	-3.45	-2.21	-6.33	-3.69
	-1.87	-0.35	-4.81	-1.15
	-4.19	-3.06	-6.63	-5.16
1992	-3.56	-2.44	-7.55	-5.33
	-1.88	-0.35	-5.83	-2.36
	-4.54	-3.15	-8.14	-6.62
1994	-4.24	-2.62	-8.09	-5.62
	-1.90	-0.35	-6.32	-3.57
	-5.42	-3.32	-8.25	-7.28
1996	-5.31	-2.73	-8.52	-5.82
	-2.76	-0.35	-6.71	-4.32
	-6.07	-4.21	-8.84	-7.81
1998	-5.36	-2.76	-8.81	-5.94
	-3.15	-0.62	-6.97	-4.16
	-6.08	-4.75	-9.32	-8.16
2000	-5.60	-3.80	-8.91	-5.99
	-3.99	-1.93	-7.06	-2.61
	-6.09	-5.02	-9.85	-8.29

Table 45: Low Alcohol: Raw Taste Wedge with 95% Confidence Interval

Year	π^L		π^H	
	Mod.	Heavy	Mod.	Heavy
1980	-1.35	-0.33	-3.46	-1.54
	-0.15	0.61	-2.19	-0.32
	-2.30	-1.39	0.00	-3.59
1982	-3.48	-0.76	-4.49	-2.81
	-1.58	0.44	-3.20	-0.52
	-3.85	-2.50	-4.78	-4.02
1984	-3.10	-0.43	-4.97	-4.83
	-1.62	0.65	-4.12	-3.33
	-3.79	-2.60	-5.32	-5.11
1986	-3.39	-1.53	-5.04	-3.03
	-0.07	-1.36	-3.90	-1.37
	-2.92	-1.22	-5.38	-4.62
1988	-3.71	-1.70	-5.71	-5.24
	-2.24	-0.64	-4.86	-2.01
	-4.14	-3.24	-6.13	-5.67
1990	-3.85	-1.71	-6.79	-6.23
	-2.15	-0.16	-5.68	-3.23
	-4.95	-3.19	-7.21	-6.84
1992	-5.42	-3.25	-8.01	-6.54
	-4.00	-1.12	-6.86	-3.35
	-6.06	-4.84	-8.55	-7.67
1994	-5.69	-3.26	-8.07	-7.60
	-3.59	-0.18	-7.18	-5.82
	-6.33	-4.58	-8.68	-8.24
1996	-6.14	-4.00	-8.59	-8.49
	-2.93	-0.09	-7.28	-3.83
	-6.10	-3.96	-9.08	-8.50
1998	-6.77	-5.79	-9.21	-9.20
	-5.58	-1.49	-8.25	-7.92
	-7.31	-6.40	-9.85	-9.73
2000	-6.77	-5.79	-9.66	-9.64
	-5.55	-4.15	-8.75	-8.44
	-7.25	-6.31	-10.45	-10.29

Table 46: High Alcohol: Raw Taste Wedge with 95% Confidence Interval

Year	π^L		π^H	
	Mod.	Heavy	Mod.	Heavy
1980	-1.50	0.00	-0.69	0.05
	-0.63	0.00	0.07	0.78
	-1.86	0.00	-1.74	-0.23
1982	-2.92	-1.40	-3.03	-0.90
	-1.63	-0.34	-1.62	-0.25
	-3.43	-2.47	-3.83	-2.02
1984	-3.26	-1.64	-3.37	-0.57
	-1.60	0.40	-1.58	0.48
	-3.79	-2.52	-4.71	-1.96
1986	-3.12	-1.74	-3.77	-1.23
	-1.60	-0.17	-1.95	0.50
	-3.84	-2.86	-4.61	-2.50
1988	-3.01	-1.98	-4.82	-1.81
	-1.86	-0.13	-3.39	-0.40
	-3.71	-2.74	-5.63	-3.46
1990	-0.87	-0.97	-3.21	-0.14
	-0.24	0.61	-0.42	-1.22
	-3.15	-0.53	-5.63	-3.48
1992	-3.56	-0.88	-7.55	-5.33
	-1.68	0.40	-5.82	−2.36
	-4.54	-2.59	-8.14	-6.62
1994	-4.24	-1.26	-7.23	-2.78
	-1.90	0.16	-5.18	-0.71
	-5.42	-3.32	-7.83	-5.26
1996	-5.31	-2.73	-7.14	-3.27
	-2.76	-0.35	-5.50	-0.74
	-6.07	-4.22	-8.34	-5.30
1998	-5.36	-2.76	-7.11	-3.42
	-3.15	-0.62	-6.83	-0.67
	-6.08	-4.08	-8.62	-5.69
2000	-5.60	-3.80	-8.91	-5.99
	-3.99	-1.93	-7.06	-1.86
	-6.09	-5.03	-9.85	-8.29

Table 47: Minimal Lifecycle α with 95% Confidence Interval

Year	Light		Mod.		Heavy	
	Low	High	Low	High	Low	High
1980	0.00	-92.19	0.00	-64.52	-0.00	-39.80
	0.00	-116.22	0.00	-89.03	0.00	-78.87
	0.00	-57.11	0.00	-34.62	0.00	-11.28
1982	-52.05	-125.50	-29.70	-78.99	-29.27	-91.29
	-82.99	-199.66	-57.49	-112.47	-64.64	-120.10
	-25.89	-81.96	-11.95	-60.92	-13.54	-62.43
1984	-43.19	-173.63	-33.39	-132.72	-41.25	-120.67
	-85.90	-217.53	-63.83	-164.55	-72.74	-151.81
	-21.61	-115.17	-13.88	-88.15	-1.38	-78.59
1986	-28.09	-126.96	-14.31	-84.37	-15.40	-102.43
	-43.41	-180.28	-38.07	-129.27	-45.62	-141.40
	-5.02	-83.18	-0.55	-71.18	-3.79	-77.69
1988	-21.74	-197.25	-27.94	-126.94	-55.85	-121.22
	-40.51	-219.90	-42.40	-164.32	-77.75	-157.81
	-6.63	-119.70	-13.99	-98.38	-29.04	-87.77
1990	-12.47	-155.35	-6.24	-101.10	-8.41	-93.64
	-37.23	-193.94	-26.72	-131.60	-30.75	-132.78
	-1.13	-93.29	-1.38	-174.62	-0.99	-60.13
1992	-60.74	-143.44	-42.51	-104.31	-50.47	-140.94
	-87.53	-199.83	-61.28	-147.50	-74.45	-166.72
	-24.96	-101.87	-16.73	-87.96	-19.53	-100.13
1994	-68.10	-182.55	-35.13	-122.97	-47.40	-136.64
	-97.92	-217.47	-69.78	-154.71	-75.55	-171.35
	-33.12	-118.15	-18.17	-91.81	-19.63	-108.90
1996	-77.27	-136.01	-39.80	-108.57	-51.35	-137.57
	-103.66	-211.77	-77.07	-159.27	-75.94	-161.94
	-38.90	-103.14	-21.91	-82.11	-23.05	-102.62
1998	-106.32	-196.64	-67.41	-135.41	-75.30	-158.00
	-130.58	-226.55	-95.19	-169.93	-103.94	-187.91
	-58.81	-126.70	-46.82	-110.23	-40.67	-122.67
2000	-96.29	-196.64	-76.36	-143.98	-113.57	-169.61
	-139.26	-236.82	-106.67	-180.86	-131.75	-200.76
	-72.32	-132.92	-60.42	-120.20	-81.88	-137.05

BIBLIOGRAPHY

[1] Alberto Abadie and Javier Gardeazabal (2003), "The Economic Costs of Conflict: A Case-Control Study for the Basque Country", *American Economic Review*, 93(1), 113-132.

[2] Alberto Abadie, Alexis Diamond and Jens Hainmueller (2010), "Synthetic Control Methods for Comparative Case Studies; Estimation the Effect of California's Tobacco Control Programme", *Journal of the American Statistical Association*, 105(490), 493-505.

[3] Sydney Afriat (1967), "The Construction of Utility Functions from Expenditure Data", *International Economic Review*, 8(1), 67-77.

[4] Sydney Afriat (1977), *The Price Index*, London: Cambridge University Press.

[5] George Ainslie, Shane Frederick, Xiupeng Li, John Lynch, Page Moreau, Andrew Mitchell, Daniel Read, Alan Sawyer, Yaacov Trope, Klaus Wertenbroch, Gal Zauberman (2005), "The Psychology of Intertemporal Discounting: Why are Distant Events Valued Differently from Proximal Ones?", *Marketing Letters*, 16(3.4), 347-360.

[6] George-Mario Angeletos, David Laibson, Andrea Repetto, Jeremy Tobacman and Stephen Weinberg (2001), "The Hyperbolic Consumption Model: Calibration, Simulation and Empirical Evaluation", *Journal of Economic Perspectives*, 15(3), 47-68.

[7] James Andreoni and Charles Sprenger (2012), "Estimating Time Preferences from Convex Budget Sets", *American Economic Review*, 101(7), 3333-3356.

[8] Adelchi Azzalini and Adrian Bowman (1997), *Applied Smoothing Techniques for Data Analysis*, New York: Oxford University Press.

[9] Charles Babbage (1864), *Passages from the Life of a Philosopher*, London: Longman and Co.

[10] Timothy Beatty and Ian Crawford (2011), "How Demanding is the Revealed Preference Approach to Demand?", *American Economic Review*, 101(6), 2782-2795.

[11] Gary Becker (1962), "Irrational Behavior and Economic Theory", *Journal of Political Economy*, 70(1), 1-13.

[12] Gary Becker, Elizabeth Landes, Robert Michael (1977), "An Economic Analysis of Marital Instability", *Journal of Political Economy*, 85(6), 1141-1187.

[13] Gary Becker and George Stigler (1977), "De Gustibus Non Est Disputandum", *American Economic Review*, 67(2), 76-90.

[14] Laura Blow and Ian Crawford (1999), "Valuing Quality", *IFS Working Paper Series W99/21*, The Institute for Fiscal Studies.

[15] Richard Blundell, Martin Browning and Ian Crawford (2003), "Nonparametric Engel Curves and Revealed Preference", *Econometrica*, 71(1), 205-240.

[16] Richard Blundell, Martin Browning and Ian Crawford (2008), "Best Nonparametric Bounds on Demand Responses", *Econometrica*, 76(6), 1227-1262.

[17] Richard Blundell, Martin Browning, Laurens Cherchye, Ian Crawford, Bram De Rock and Frederic Vermeulen (2012), "Sharp for SARP: Nonparametric Bounds on the Behavioural and Welfare Effects of Price Changes", *IFS Working Paper W12/14*, The Institute for Fiscal Studies.

[18] Richard Blundell, Xiaohong Chen and Dennis Kristensen (2007), "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves", *Econometrica*, 75(6), 1613-1669.

[19] Richard Blundell, Joel Horowitz and Matthias Parey (2012), "Measuring the Price Responsiveness of Gasoline Demand: Economic Shape Restrictions and Nonparametric Demand Estimation", *Quantitative Economics*, 3(1), 29-51.

[20] Richard Blundell and James Powell (2007), "Censored Regression Quantiles with Endogenous Regressors", *Journal of Econometrics*, 141(1), 65-83.

[21] Janet Bobo and Corinne Husten (2000), "Sociocultural Infuences on Smoking and Drinking", *Alcohol Research and Health*, 24(4), 225-232.

[22] Stephen G. Bronars (1987), "The Power of Nonparametric Tests of Preference Maximization", *Econometrica*, 55(3), 693-698.

[23] Donald Brown and Rosa Matzkin (1998), "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand", *Cowles Foundation Discussion Papers 1175*, Cowles Foundation, Yale University.

[24] Martin Browning (1989), "A Nonparametric Test of the Life-Cycle Rational Expectations Hypothesis", *International Economic Review*, 30(4) 979-992.

[25] Martin Browning, Francois Bourguignon, Pierre-Andre Chiappori, Valerie Lechene (1994), "Income and Outcomes: A Structural Model of Intrahousehold Allocation", *Journal of Political Economy*, 102(6), 1067-1096.

[26] Martin Browning and Pierre Andre Chiappori (1998), "Efficienct Intra-household Allocations: A General Characterisation and Empirical Tests", *Econometrica*, 66(6), 1241-1278.

[27] Kenneth Burdett and Melvyn Coles (1999), "Long Term Partnership Formation: Marriage and Employment", *The Economic Journal*, 109(6), 307-334.

[28] Christopher Carroll (2001), "Death to the Log-Linearized Consumption Euler Equation (And Very Poor Health to the Second-Order Approximation)", *The B.E. Journal of Macroeconomics*, 1(1), 1-38.

[29] Christopher Chambers, Federico Echenique and Eran Shmaya (2010), "General Revealed Preference Theory", *California Institute of Technology Social Science Working Paper 1332*, California Institute of Technology.

[30] Laurens Cherchye, Bram De Rock and Frederic Vermeulen (2007), "The Collective Model of Household Consumption: A Nonparametric Characterization", *Econometrica*, 75(2), 553-574.

[31] Laurens Cherchye, Bram De Rock and Frederic Vermeulen (2009), "Opening the Black Box of Intrahousehold Decision Making", *Journal of Political Economy*, 117(6), 1074-1104.

[32] Laurens Cherchye, Ian Crawford, Bram De Rock and Frederic Vermeulen (2013), "Gorman Revisited: Nonparametric Conditions for Exact Linear Aggregation", *KU Leuven Discussion Paper Series DPS13.05*, Katholieke Universiteit Leuven.

[33] Victor Chernozhukov and Han Hong (2002), "Three-Step Censored Quantile Regression and Extramarital Affairs", *Journal of the American Statistical Association*, 40(459), 872-882.

[34] Victor Chernozhukov, Iván Fernández-Val and Amanda Kowalski (2010), "Quantile Regression with Censoring and Endogeneity", *Mimeo*, Massachusetts Institute for Technology.

[35] Pierre-André Chiappori (1988), "Rational Household Labor Supply", *Econometrica*, 56(1), 63-90.

[36] Pierre-André Chiappori (1992), "Collective Labor Supply and Welfare", *Journal of Political Economy*, 100(3), 437-467.

[37] Pierre-Andre Chiappori amd Yoram Weiss (2006), "Divorce, Remarriage and Welfare: A General Equilibrium Approach", *Journal of the European Economic Association 4.2/3*, Papers and Proceedings of the Twentieth Annual Congress of the European Economic Association, 415-426.

[38] James Choi, David Laibson, Brigitte Madrian and Andrew Metrick (2004), "For Better or for Worse: Default Effects and 401(k) Savings Behavior", in *Perspectives on the Economics of Ageing* edited by David Wise, University of Chicago Press, NBER: University of Chicago Press.

[39] Thomas Cover and Joy Thomas (1991), *Elements of Information Theory*, New York: John Wiley & Sons.

[40] Ian Crawford (2010), "Habits Revealed", *Review of Economic Studies*, 77(4), 1382-1402.

[41] Imre Csiszár (1998), "The Method of Types", *IEEE Transactions on Information Theory*, 44(6), 2505-2523.

[42] Stefano DellaVigna and Ulrike Malmendier (2006), "Paying Not To Go To The Gym", *The American Economic Review*, 96(3), 694-719.

[43] Thomas Demuynk and Ewout Verriest (2012), "I'll Never Forget My First Cigarette: A Revealed Preference Analysis of the 'Habits as Durables' Model", *International Economic Review*, 54(2), 717-738.

[44] Department of Health (2011), *White Paper: Healthy Lives, Healthy People: Our Strategy for Public Health in England*, London: TSO.

[45] Peter Diamond, Daniel McFadden and Miguel Rodriguez (1978), "Measurement of the Elasticity of Factor Substitution and Bias of Technical Change," in *Production Economics: A Dual Approach to Theory and Applications; Volume II, Applications of the Theory of Production* edited by Melvyn Fuss and Daniel McFadden, Amsterdam: North-Holland.

[46] W. Erwin Diewert (1973), "Afriat and Revealed Preference Theory", *The Review of Economic Studies*, 40(3), 419-425.

[47] Thomas Dohman, Armin Falk, David Huffman, Uwe Sunde (2012), "Interpreting Time Horizon Effects in Inter-Temporal Choice", *CESifo Working Paper No. 3750*, CESifo Group Munich.

[48] Francis Edgeworth (1881), *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*, New York: Augustus M. Kelly.

[49] Daniel Falk, Hsiao-Ye Yi and Susanne Hiller-Sturmhofel (2006), "An Epidemiologic Analysis of Co-Occurring Alcohol and Tobacco Use and Disorders: Findings from the National Epidemiologic Survey on Alcohol and Related Conditions", *Alcohol Research and Health*, 29(3), 162-171.

[50] Martin S. Feldstein (1964), "The Social Time Preference Discount Rate in Cost Benefitt Analysis", *The Economic Journal*, 74(294), 360-379.

[51] Shane Frederick, George Loewenstein, Ted O'Donoghue (2002), "Time Discounting and Time Preference: A Critical Review", *Journal of Economic Literature*, 40(2), 351-401.

[52] Amos Golan, George Judge and Douglas Miller (1996), *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York: John Wiley & Sons.

[53] Amos Golan, Jeffrey Perloff and Edward Shen (2001), "Estimating a Demand System with Nonnegativity Constraints: Mexican Meat Demand", *The Review of Economics and Statistics*, 83(3), 541-550.

[54] Leonard Green, Nathanael Fristoe and Joel Myerson (1994), "Temporal Discounting and Preference Reversals in Choice Between Delayed Outcomes", *Psychonomic Bulletin Review*, 1(3), 383-389.

[55] Grossmann and Kravanja (1997), "Mixed Integer Nonlinear Programming: A Survey of Algorithms and Applications", in *Large Scale Optimisation with Applications, Part II: Optimal Design and Control* edited by Lorenz Biegler, Thomas Coleman, Andrew Conn and Fadil Santosa, 93, 73-100.

[56] Jonathan Gruber and Botond Koszegi (2000), "Is Addiction 'Rational'? Theory and Evidence", *Quarterly Journal of Economics*, 116, 1261-1305.

[57] Berthold Haag, Stefan Hoderlein and Krishna Pendakur (2009), "Testing and Imposing Slutsky Symmetry in Nonparametric Demand Systems", *Journal of Econometrics*, 153(1), 33-50.

[58] Jerry Hausman and Whitney Newey (1995), "Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss", *Econometrica*, 63(6), 1445-1476.

[59] James Heckman (2001), "Micro Data, Heterogeneity and the Evaluation of Public Policy", *Journal of Political Economy*, 109(4), 673-748.

[60] John Hicks (1939), *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory*, Oxford: Clarendon Press.

[61] Stefan Hoderlein (2008), "Chapter 10 Developments in Nonparametric Demand Analysis: Heterogeneity and Nonparametrics", in *Contributions to Economic Analysis* edited by Badi Baltagi and Efraim Sadka, 288, Bingley: Emerald Group Publishing Limited.

[62] Hendrik Houthakker (1950), "Revealed Preferences and the Utility Function", *Economica*, 17(66), 59-74.

[63] Leonid Hurwicz (1971), "On the Problem of Integrability of Demand Functions", in *Preferences, Utility and Demand: A Minnesota Symposium* edited by John Chipman, Leonid Hurwicz, Gerhard Richter and Hugo Sonnenschein, 114-148.

[64] Leonid Hurwicz and Hirofumi Uzawa (1971), "On the Integrability of Demand Functions", in *Preferences, Utility and Demand: A Minnesota Symposium* edited by John Chipman, Leonid Hurwicz, Gerhard Richter and Hugo Sonnenschein, 174-214.

[65] Matthew Jackson and Leeat Yariv (2011), "Collective Dynamic Choice: The Necessity of Time Inconsistency", *Mimeo*, Stanford University.

[66] King James I of England (1604), *A Counterblaste To Tobacco*, London: R. Barker.

[67] Edwin Jaynes (1957a), "Information Theory and Statistical Mechanics I", *Physical Review*, 106(4), 620-630.

[68] Edwin Jaynes (1957b), "Information Theory and Statistical Mechanics II", *Physical Review*, 108(2), 171-190.

[69] Edwin Jaynes (2003), *Probability Theory: The Logic of Science*, Cambridge: Cambridge University Press.

[70] Boyan Jovanovic (1979), "Job Matching and the Theory of Turnover", *Journal of Political Economy*, 87(5.1), 972-990.

[71] Woocheoi Kim and Gautam Tripathi (2003), "Nonparametric Estimation of Homogeneous Functions', *Econometric Theory*, 19(4), 640-663.

[72] Tjalling Koopmans (1960), "Stationary Ordinal Utility and Impatience", *Econometrica*, 28(2), 287-309.

[73] Kris Kirby and Richard Hernstein (1995), "Preference Reversals due to Myopic Discounting of Delayed Reward", *Psychological Science*, 6(2), 83-89.

[74] Solomon Kullback (1959), *Information Theory and Statistics*, New York: John Wiley & Sons.

[75] Solomon Kullback and Richard Leibler (1951), "On Information and Sufficiency", *Annals of Mathematical Statistics*, 22(1), 79-86.

[76] David Laibson (1997), "Golden Eggs and Hyperbolic Discounting", *Quarterly Journal of Economics*, 112(2), 443-478.

[77] David Laibson (1998), "Life-Cycle Consumption and Hyperbolic Discount Functions", *European Economic Review*, 42, 861-871.

[78] Andrew Leicester (2012), "How Might In-Home Scanner Technology be Used in Budget Surveys?", *IFS Working Papers W12/01*, The Institute for Fiscal Studies.

[79] Andrew Leicester and Zoe Oldfield (2009), "Using Scanner Technology to Collect Expenditure Data", *Fiscal Studies*, 30(3-4), 309-337.

[80] Arthur Lewbel (2001), "Demand Systems With and Without Errors", *American Economic Review*, 91(3), 611-618.

[81] Arthur Lewbel and Krishna Pendakur (2009), "Tricks with Hicks: The EASI Demand System", *American Economic Review*, 99(3), 827-863.

[82] Stephen Lich-Tyler (2004), "Negotiations and Love Songs: The Dynamics of Bargained Household Decisions", *Mimeo*, University of Michigan.

[83] Jayson Lusk and Kathleen Brooks (2011), "Who Participates in Household Scanning Panels?", *American Journal of Agricultural Economics*, 93(1), 226-240.

[84] Charles Manski (2007), *Identification for Prediction and Decision*, Cambridge, MA: Harvard University Press.

[85] Albert Marcet and Ramon Marimon (1998), "Recursive Contracts", *Mimeo*, Universitat Pompeu Fabra.

[86] Stephen Marglin (1963), "The Social Rate of Discount and The Optimal Rate of Investment", *The Quarterly Journal of Economics*, 77(1), 95-111.

[87] Michael Marmot (2010), *Fair Society, Healthy Lives: Strategic Review of Health Inequalities in England Post 2010*, London: The Marmot Review.

[88] Andreu Mas-Colell (1982), "Revealed Preference after Samuelson", *Samuelson and Neoclassical Economics*, Amsterdam: Kluwer.

[89] Rosa Matzkin (2003), "Nonparametric Estimation of Nonadditive Random Functions", *Econometrica*, 71(5), 1339-1375.

[90] Rosa Matzkin (2007), "Heterogeneous Choice", in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society* edited by Richard Blundell, Whitney Newey and Torsten Persson, Cambridge: Cambridge University Press.

[91] Maurizio Mazzocco (2007), "Household Intertemporal Behaviour: A Collective Characterization and a Test of Commitment", *Review of Economic Studies*, 74(3), 857-895.

[92] Samuel McClure, David Laibson, George Lowenstein and Jonathan Cohen (2004), "Separate Neural Systems Value Immediate and Delayed Monetary Rewards", *Science*, 306(5695), 503-507.

[93] Samuel McClure, Ericson, David Laibson, George Lowenstein and Jonathan Cohen (2007), "Time Discounting for Primary Rewards", *The Journal of Neuroscience*, 27(21), 5796-5804.

[94] Daniel McFadden (2001), "Economic Choices", *American Economic Review*, 91(3), 351-378.

[95] Daniel McFadden and Mogens Fosgerau (2012), "A Theory of the Perturbed Consumer with General Budgets", *NBER Working Papers 17953*, National Bureau of Economic Research, Inc.

[96] Janet Metcalfe and Walter Mischel (1999), "A Hot/Cool System Analysis of Delay of Gratification: Dynamics of Willpower", *Psychological Review*, 106(1), 3-19.

[97] Paul Milgrom and Chris Shannon (1994), "Monotone Comparative Statics", *Econometrica*, 62(1), 157-180.

[98] Christopher Murray and Alan Lopez (1997), "Global Mortality, Disability, and the Contribution of Risk Factors: Global Burden of Disease Study", *The Lancet*, 349(9063), 1436-1442.

[99] Peter Neary and Kevin Roberts (1980), "The Theory of Household Behaviour Under Rationing", *European Economic Review*, 13(1), 25-42.

[100] Phillip Nelson (1970), "Information and Consumer Behavior", *Journal of Political Economy*, 78(2), 311-329.

[101] Ted O'Donoghue and Matthew Rabin (1999), "Doing It Now or Later", *American Economic Review*, 89(1), 103-124.

[102] Ted O'Donoghue and Matthew Rabin (2001), "Choice and Procrastiation", *Quarterly Journal of Economics*, 116(1), 121-160.

[103] Robert Pollak (1990), "Distinguished Fellow: Houthakker's Contribution to Economics", *Journal of Economic Perspectives*, 4(2), 141-156.

[104] Karl Popper (1959), *The Logic of Scientific Discovery*, London: Hutchinson and Company.

[105] James Powell (1984), "Least Absolute Deviations Estimation for the Censored Regression Model", *Journal of Econometrics*, 25(3), 303-325.

[106] Marcel Richter (1966), "Revealed Preference Theory", *Econometrica*, 34(4), 635-645.

[107] Hugh Rose (1958), "Consistency of Preference: The Two-Commodity Case", *Review of Economic Studies*, 25(2), 124-125.

[108] Royal College of Physicians (1962), *Smoking and Health*, London: Pitman.

[109] Paul Samuelson (1937), "A Note on the Measurement of Utility", *Review of Economic Studies*, 4(2), 155-161.

[110] Paul Samuelson (1938), "A Note on the Pure Theory of Consumer Behavior", *Economica*, 5(17), 61-71.

[111] Paul Samuelson (1948), "Consumption Theory in Terms of Revealed Preference", *Economica*, 15(60), 243-253.

[112] I. N. Sanov (1957), "On the Probability of Large Deviations of Random Variables", *Mat Sbornik*, 42(84.1), 11-44. (English translation (1960) in Select Translations of Mathematics, Statistics and Probability 1, p213-244)

[113] Simone G. Schaner (2012), "Do Opposites Detract? Intrahousehold Preference Heterogeneity and Inefficient Strategic Savings", *Mimeo*, Dartmouth College.

[114] Reinhard Selten (1991), "Properties of a Measure of Predictive Success", *Mathematical Social Sciences*, 21(2), 153-167.

[115] Amartya Sen (1971), "Choice Functions and Revealed Preference", *The Review of Economic Studies*, 38(3), 307-317.

[116] Amartya Sen (1973), "Behaviour and the Concept of Preference", *Economica*, 40(159), 241-259.

[117] Amartya Sen (1977), "Rational fools: A Critique of the Behavioral Foundations of Economic Theory", *Philosophy and Public Affairs*, 6(4), 317-344.

[118] Avner Shaked and John Sutton (1984), "Involuntary Unemployment as a Perfect Equilibrium in a Bargaining Model", *Econometrica*, 52(6), 1351-1364.

[119] John Shore and Rodney Johnson (1980), "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross Entropy", *IEEE Transactions on Information Theory*, 26(1), 26-37.

[120] John Shore and Rodney Johnson (1981), "Properties of Cross Entropy Minimization", *IEEE Transactions on Information Theory*, 27(4), 472-482.

[121] Bernard Silverman (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

[122] Bernt Stigum (1973), "Revealed Preference- A Proof of Houthakker's Theorem", *Econometrica*, 41(3), 411-423.

[123] Surgeon General's Advisory Committee on Smoking and Health (1964), "Report of the Advisory Committee to the Surgeon General of the United States", *Department of Health, Education and Welfare, Public Health Service Publication No. 1103*.

[124] Richard Thaler and Shlomo Benartzi (2004), "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving", *Journal of Political Economy*, 112(1), 164-187.

[125] Richard Thaler and Hersh Shefrin (1981), "An Economic Theory of Self Control", *Journal of Political Economy*, 89(2), 392-410.

[126] Jos Uffink (1995), "Can the Maximum Entropy Principle be Explained as a Consistency Requirement?", *Studies in the History and Philosophy of Science Part B*, 26(3), 223-261.

[127] Hirofumi Uzawa (1960), "Preference and Rational Choice in the Theory of Consumption", in *Mathematical Models in Social Science* edited by Kenneth Arrow, Daniel Lerner and Harold Lasswell, Stanford: Stanford University Press.

[128] Yrjo Vartia (1983), "Efficient Methods of Measuring Welfare Change and Compensated Income In Terms Of Ordinary Demand Functions", *Econometrica*, 51(1), 79-98.

[129] Hal Varian (1982), "The Nonparametric Approach to Demand Analysis", *Econometrica*, 50(4), 945-973.

[130] Hal Varian (1983), "Non-Parametric Tests of Consumer Behaviour", *Review of Economic Studies*, 50(1), 99-110.

[131] Hal Varian (1985), "Nonparametric Analysis of Optimizing Behavior with Measurement Error", *Journal of Econometrics*, 30, 445-458.

[132] Hal Varian (1988), "Revealed Preference with a Subset of Goods", *Journal of Economic Theory*, 46(1), 179-185.

[133] Hal Varian (1990), "Goodness-of-Fit in Optimising Models", *Journal of Econometrics*, 46(1-2), 125-140.

[134] Yoram Weiss and Robert Willis (1997), "Match Quality, New Information and Martial Dissolution", *Journal of Labor Economics*, 15.1 Part 2: Essays in Honor of Yoram Ben-Porath, 293-329.

[135] World Health Organisation (2012), "Mortality Attributable to Tobacco", *WHO Global Report Series*.

[136] Stephane Zuber (2011), "The Aggregation of Preferences: Can We Ignore the Past?", *Theory and Decision*, 70(3), 367-384.