Department of Economics Discussion Paper Series

How should we model health as a dynamic process?

Alexandra Elizabeth Brown

Number 1023
August, 2023

Department of Economics
Manor Road Building, Manor Road
Oxford, OX1 3UQ

ISSN 1471-0498

How should we model health as a dynamic process?

Alexandra Elizabeth Brown[†]

August 2023

Abstract

Health is a complex dynamic process that impacts many economic decisions in ways that remain poorly understood. This paper comprehensively reviews how health is modelled in the literature, showing that baseline models typically fail to take into account how persistence and frequency of health shocks vary by past health history and magnitude and direction of past shocks. Methods from the earnings dynamics literature are adapted to produce improved health persistence estimates. This paper also investigates how medical biomarker data can be incorporated in dynamic models of health as a proxy for underlying health. There is significant scope for further work in this area as more medical data becomes available to researchers.

JEL classifications: I10, I31, C5

[†]University of Oxford, Department of Economics, 10 Manor Rd, Oxford, OX1 3UQ, United Kingdom. Email: alexandra.brown@economics.ox.ac.uk. I would like to thank Hamish Low for his guidance and support with this paper

Contents

1 Introduction

Health is an important determinant of an individual's economic decision making, affecting labour supply, consumption, family composition and access to government-provided insurance. Accurately modelling health as a dynamic process is needed to answer several important and open questions in the literature, including what is the relationship between health and earnings inequality, how effective are the current government-provided safety nets for those who fall ill, and how best should governments respond to the increasing economic burdens of chronic disease, rising disability rates and an ageing population. There has been a significant amount of reduced-form work on these questions (Prinz et al. (2018) provides a good summary). More recently, structural approaches have been used to better understand some of the complex endogeneity between health and economic decision making. These models typically capture health dynamics in a highly simplified way to minimise computational burden. The contributions made by this paper are relevant to both these reduced form and structural approaches.

This paper makes three key contributions to the literature. Firstly, it provides a comprehensive assessment of the different ways health dynamics have been modelled in the literature. Researchers have typically borrowed from the earnings dynamics literature and modelled health as a simple linear process such as an autoregressive moving average process, often discretised as a first-order Markov process, or the sum of a permanent and a transitory shock. To the author's knowledge, there have been no prior attempts to systematically evaluate these modelling approaches and their underlying assumptions. I use Understanding Society data, a commonly-used UK household panel dataset, to replicate the most common models of health dynamics using standard panel data techniques. I then evaluate how well they capture key features of the health process, focussing on estimates of persistence and cross-sectional heterogeneity caused by different health shock realisations. I show that an ARMA(1,1) model with a large AR(1) coefficient close to one and a moderately-sized negative MA(1) coefficient fit the data reasonably well. An alternative linear model that combines a permanent process and a transitory process can be a desirable alternative as it allows for two types of shocks with different properties, but at the expense of some very strong and potentially incorrect persistence assumptions.

One of the most important components of health dynamics to accurately model is persistence. An individual is likely to respond very differently to a highly persistent health shock compared to a moderately persistent one. In the ARMA(1,1) model, the AR(1) term captures the average persistence of health from last period, which is then modified by the MA(1) term depending on the magnitude of the prior period's error term. I show that persistence heterogeneity is much greater than captured by this model, and varies systematically by past health and the features of the health shock. On average, health shocks are more persistent if they are negative (a decline in health rather than an improvement), if the individual was in poor health prior to the shock, and if the health shock is large. The standard linear models of health dynamics do not capture this heterogeneity, and therefore tend to be overly-optimistic in modelling the health dynamics of those with a history of poor health who experience additional negative health shocks.

A related limitation is that these models are not particularly effective in capturing the different distributions of health shock risks that individuals face. While an ARMA(1,1) model can be estimated using GMM techniques that are fairly robust to various error distribution assumptions, the most obvious application of the model would impose a mean-zero independent and identically distributed (i.i.d.) normal error distribution, while the model that is a sum of a random walk and a moving average transitory process has a normally-distributed error term. I document several ways that the error terms, which I interpret as health shocks, deviate from an i.i.d. normal distribution. Firstly, there is a strong relationship between past health and the expected distribution of future health shocks. Those in poor health face an increased risk of both large negative and large positive health shocks, while the variance of health shocks faced by those in good health is much lower. The variance of health shocks is higher for negative shocks than positive shocks, even when controlling for past health. Finally, the baseline models do not accurately replicate the higher order moments of the data.

The second contribution of this paper is to address many of the limitations of these standard linear models of health by adapting a recent panel data technique from the earnings dynamics literature. I use Arellano, Blundell and Bonhomme (2017)'s quantile-based method to produce non-linear persistence estimates that allow for a large amount of heterogeneity. One attraction of this framework is that it allows for

persistence to vary depending on the size and sign of the health shock that occurs in period t, which cannot be done using the standard linear models due to the endogeneity between the shock and persistence estimates that relate health in period $t-1$ to health in period t. Applying this framework to my health data produces persistence estimates that range from 0.6 to 1.2, depending on prior health and characteristics of the shock in period t. This framework is able to capture that the persistence of the health process is higher among individuals in poor prior health, and that positive health shocks are typically less persistent than negative health shocks. These improved persistence estimates better capture the health risks faced by individuals, with implications for our understanding of the impact of health on different economic decisions such as labour supply and consumption. I also estimate an extended version of this framework that is able to strip out time-invariant unobserved heterogeneity from the persistence estimates, and consider the wider applicability of the framework by applying it to produce non-linear persistence estimates of an index of mental health.

Finally, this paper investigates how best to use increasingly-available medical data to improve health modelling. These data are available for a subset of individuals in the Understanding Society dataset. Previous studies have shown that biomarker data such as inflammation markers and steroid hormones in the blood can predict future adverse health outcomes in ostensibly healthy people (Davillas and Pudney, 2020*a*). To the author's knowledge, this data has never been used to better model health dynamics. I find that incorporating the biomarker data into my models of health dynamics does not improve their persistence estimates. However, the data can be used to better model the different health risks individuals face. I show that the ARMA(1,1) model performs less well in cases where biomarker data suggests that an individual's underlying health is very poor. These are typically cases where an individual does not report any serious health conditions, but they face a significantly elevated risk of negative health shocks. This is an important source of risk to capture. Secondly, I find that variation in biomarker data are strongly correlated with the variation captured by the fixed effect component of the persistence estimates produced using the Arellano, Blundell and Bonhomme (2017) framework. This suggests that biomarker data can be used to better understand and model individual heterogeneity in health outcomes, a topic that remains poorly understood.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature and Section 3 describes the data, focussing on the construction of indicies to capture observed and underlying health. Section 4 reviews the baseline dynamic health models in the literature and section 5 identifies their limitations. Section 6 applies methods from the earnings literature to produce non-linear estimates of persistence. Section 7 concludes.

2 Literature Review

There is a body of literature that develops methods of aggregating survey health data into an index of overall health, which I summarise in the data section of this paper. However, answering questions on the relationship between health and economic outcomes often require us to take a stance on how health evolves over time. There is some reduced form work on this question (ODonnell, Van Doorslaer and Van Ourti, 2015), but the most common approach in the literature has been to apply the vast literature on modelling earnings dynamics to model health as a simple linear process, most commonly as an ARMA(p,q) process or the sum of a persistent and a transitory component. In the structural literature, a discrete version of this approach; a first-order Markov process with a small number of discrete health states, has been common. However, the implications and limitations of these models has only very recently begun to be examined in the literature. I review the modelling health as a dynamic process literature, highlighting the gaps that this paper seeks to fill.

The canonical papers that model the time series properties of the mean of earnings, such as Lillard and Willis (1978), MaCurdy (1982) and Abowd and Card (1989) use panel data to fit ARMA-type processes to earnings data. A recent example of this approach applied to health data is Blundell et al. (2020), who represent health $\widetilde{h}_t$ using the error correction model: $\widetilde{h}_t = \pi_t + \varepsilon_t$. The persistent component (π_t) evolves as a random walk: $\pi_t = \pi_{t-1} + \eta_t$, and ε_t is a MA(0) transitory component. Alternate specifications include modelling the persistent component as an AR(1) or higher order process so the effect of a shock to the persistent component decays over time, and adding more structure to the transitory component, such as by incorporating moving-average terms (Blundell et al., 2016), or by modelling health as a stock that decays (Wallenius, 2020). To reduce dimensionality, health processes that are included in structural models are often discretised into a first-order Markov process that models transitions between discrete health states. There are many examples: Palumbo (1999), French (2005), De Nardi, French and Jones (2010), Attanasio, Kitao and Violante (2010), French and Jones (2011), Capatina (2015), Jung and Tran (2016), Braun, Kopecky and Koreshkova (2017), Imrohoroglu and Zhao (2018), Jolivet and Postel-Vinay (2020), Nygaard (2021) and Amengual, Bueren and Crego (2021). Earlier structural papers typically only modelled two health states, good and

bad health, while more recent papers tend to include additional states, for example Jolivet and Postel-Vinay (2020) model four states of mental health: good, average, poor and severe. Some of these papers endogenise the health process by incorporating the impact of choices such as unhealthy consumption of cigarettes (Nygaard, 2021) or medical expediture choices (Prados, 2018). Zweifel, Breyer and Kifmann (2009) model people choosing the level of health investment to marginally alter their transition probabilities between different health states. An important distinction between these Markov models and ARMA models is that the latter imposes a symmetry between positive and negative health shocks. Markov models do not have this feature, and the data suggest that the transition probability from good to poor health tends to differ from the transition probability from poor health to good health.

Both ARMA and Markov models emphasise the state dependency of the health process. This can downplay the importance of individual heterogeneity in explaining the large cross-sectional variance in health observed in the data. Halliday (2008) finds that individual characteristics that trace back to childhood and early adulthood play an important role in determining how long health shocks persist, while the importance of state dependence varies significantly. However, he acknowledges that his first-order Markov model with only two health states limits his ability to pin down state dependence. Hauck and Rice (2004) similarly emphasise the importance of individual heterogeneity relative to state dependence in modelling mental health transitions. Pashchenko, Porapakkarm and Nardi (2017) find that variation in health transitions due to 'health types' is much larger than variation due to state-dependence in men with high-school education. They do not attempt to explain what causes these different 'health types'; they are simply used as a modelling device to allow for different health transition probabilities for different agents. However, the authors do highlight the recent empirical literature that emphasises the life-long economic consequences of genetics and early childhood experiences such as Barth, Papageorge and Thom (2020), Conti and Heckman (2010), Case, Fertig and Paxson (2004), Harris et al. (2016) and Cronqvist and Siegel (2015). Understanding the nature of this individual heterogeneity is of central importance to answering questions such as what causes the relationship between health and education, or health and earnings inequality, but it remains poorly understood.

Some of the recent structural model papers have made progress in capturing ad-

ditional complexity of health dynamics, most commonly by adding an extra variable that varies health shock risk such as 'health type' or 'underlying health'. Pashchenko, Porapakkarm and Nardi (2017) augment a standard first-order Markov model of health with transition probabilities that also depend on the duration of the current health spell and 'health type', which is a proxy for individual heterogeneity and affects transition likelihood. They find evidence of 'duration dependence' where the longer that someone has stayed in a particular state of health the less likely they are to transition states next period. This is not consistent with a low-order Markov process of health dynamics. Salvati (2021) incorporates a similar fixed-effect variable which is described as a proxy for high or low health into her model of health. She embeds a health equation into her life-cycle model that consists of an AR(1) process, a binary fixed effect term, a labour-market health interaction term, and various independent variables. Ozkan (2017) models two types of health capital: physical health capital that determines survival probability and preventative health capital that is subject to health shocks and can be modified by health investment. Keane, Capatina and Maruyama (2020) make progress in modelling individual heterogeneity by incorporating an asymptomatic health risk variable estimated with medical data. In this model, individuals have functional health that is subject to three types of shocks: predictable and persistent shocks, unpredictable and persistent shocks, and unpredictable and transitory shocks. Asymptomatic health risk captures conditions such as high cholesterol, high blood pressure and high BMI that do not directly affect daily life but increase the probability for future predictable adverse shocks to functional health.

While these models have made progress in capturing health dynamics, these equations tend to be a small component of large and complex structural models with computational demands that limit what these models can capture. Reduced form approaches for the health block of these models are common. In addition, the 'black-box' nature of these models can make it difficult to understand the mechanics of the interactions between health and other variables. This paper identifies some of limitations of modelling health in this way.

3 Data

The main dataset used in this paper is Understanding Society - the UK Household Longitudinal Study. This is a longitudinal, nationally representative dataset with good coverage of health, education, employment, family life and income variables. I build an unbalanced panel using waves 1-12 of the study, which include observations from 2009-2021. Excluding a small number of individuals with insufficient health information results in a sample of 265,830 observations from 29,886 unique individuals. Table 1 reports the summary statistics of this sample and indicates good coverage over age, education, family type and employment.

3.1 Health index construction

In many settings, the theoretically-ideal health index would be an overall stock of health measure, or a related concept such as a work-capacity index. Such an index would be continuous and bounded from below (death). Since these are unobservable concepts, we can instead construct a proxy index by aggregating various health data from household panel surveys. The available data can be grouped into three main categories. Objective health data are data on specific diagnoses and disabilities. Subjective health data are based on survey respondents' assessment of their own health. A third category of data is medical data such as pulse, blood pressure readings, blood tests or genomic data that can be used to predict health outcomes. Some of these medical data, such as genetic information, may be plausibly exogenous to any experiences or choices of the individual, which can be valuable for statistical analysis.

The limitations of each of these categories of health data as proxies for overall health has been thoroughly evaluated in several papers (Blundell et al. (2021) provides a good summary of this literature). To briefly summarise, objective measures are vulnerable to omitted variable bias, they can only capture a subset of relevant conditions, and often lack disease severity information. The rate of omission of life-changing medical diagnoses such as heart attacks and strokes reported by survey respondents has been found to be surprisingly high when compared to linked hospital admission data, suggesting measurement error could be large (Caraballo et al., 2020). Subjective health measures are fairly crude and vulnerable to reporting error and justification bias. For a given disease presentation, people will vary hugely in how

Table 1: Summary statistics

	men	women
age		
<30	17,205	24,567
30-39	17,026	24,991
40-49	21,968	30,142
50-59	21,938	28,848
60-69	19,815	24,204
70-79	12,489	13,950
80-89	3,553	4,448
90+	282	404
education		
below GSCEs	25,743	31,135
GSCEs	31,755	41,889
A-level	13,476	16,622
degree	43,302	61,908
family type		
cohabitating/married	80,910	95,630
widowed	3,315	10,427
separated/divorced	7,111	16,091
single	22,772	29,068
number of children		
0	86,026	106,463
1	11,818	19,822
2	12,271	18,718
3	3,299	5,088
4+	862	1,463
currently employed		
yes	72,589	88,493
no	41,500	62,779
occupation class		
professional	6,324	4,820
managerial & technical	28,854	34,801
skilled non-manual	9,532	24,154
skilled manual	18,719	9,105
partially skilled	7,269	14,316
unskilled	2,651	2,005
N (observations)	*114,276*	*151,554*

poorly they rate their health and to what degree they report that the disease has a negative impact on their life (French and Jones, 2017). Medical data is less commonly collected in household surveys and there is limited research on how best to use them to model health.

A challenge in the literature has been how best to use this data to construct an overall health index that minimises these biases and approximate the ideal theoretical health concept. Lack of consensus on this question has contributed to the ongoing uncertainty of the relationship between health and employment (Blundell et al., 2017). For example, large differences have been found when estimating the impact of poor health on labour supply using objective or subjective health data (Anderson and Burkhauser, 1984). To reduce these biases, a common approach has been to instrument subjective health data with objective data. This approach is still regularly used, with Blundell et al. (2020) providing a recent example, although the approach is not without criticism. Bound (1991) argues that the different types of biases affecting subjective health measures roughly offset, so that incorporating objective health data adds little value and may increase bias. Alternative approaches to aggregating health data have included taking the first principle component over a large number of objective measures (Poterba, Venti and Wise, 2017), constructing multiple indices (Blau and Gilleskie, 2001), and converting medical conditions into World Health Organisation disability weights that represent the magnitude of health loss associated with specific health outcomes, which can then be aggregated (Prados, 2018). A helpful contribution was made by Blundell et al. (2021) who comprehensively evaluated the different approaches in the literature to identify how should health data be combined to bet represent overall health. They conclude that objective measures, provided that a large enough set of them are used, subjective measures, and subjective measures instrumented with objective measures can produce similar estimates of the impact of health on employment, and any of these modelling approach can reasonably be used. This finding was broadly supported by Hosseini, Kopecky and Zhao (2022) who compares the performance of a 'frailty index' which aggregates objective indicators with a subjective health index and a principle components analysis and similarly finds that predictive power of the different approaches to be broadly comparable.

I follow the literature and use both subjective and objective health data to construct a single health index that functions as a proxy for an individual's overall stock

of health. The subjective data comes from the survey question '*In general, would you say your health is: poor, fair, good, very good or excellent?*'. The objective health data used is reported in Table 2.

Table 2: Objective health indicators

Objective measure	Data
Disabilities (specified as causing 'some difficulty' or 'much difficulty')	12 indicators: manual dexterity, mobility, lifting/moving objects, continence, hearing, sight, communication/speech, memory/ability to concentrate and learn, recognising danger, physical co-ordination, personal care, other
Mental wellbeing	General Health Questionnaire (GHQ) Caseness measure. Measures common mental health problems e.g. depression, anxiety, somatic symptoms, social withdrawal to detect those at risk of developing psychiatric disorders.
Ever diagnosed with condition	asthma, congestive heart failure, coronary heart disease, angina, heart attack, stroke, emphysema, hypothyroidism, chronic bronchitis, liver condition, epilepsy, hypertension, multiple sclerosis, COPD, osteoarthritis, rheumatoid arthritis, other arthritis, cancers: bowel/colorectal, lung, breast, prostate, liver, skin, other, diabetes: type 1, type 2, gestational and other, anxiety, depression, psychosis/schizophrenia, bipolar/manic depression, eating disorders, PTSD, other emotional/nervous/psychiatric condition, other chronic condition
Still have previously diagnosed condition	Conditions: asthma, congestive heart failure, coronary heart disease, angina, hypothyroidism, chronic bronchitis, liver condition, epilepsy, hypertension, COPD, osteoarthritis, rheumatoid arthritis, cancers: bowel/colorectal, breast, prostate and skin, type 2 diabetes, anxiety, depression, eating disorders, PTSD
Hospital out-patient	1-2 days, 3-5 days, 6-10 days, >10 days in the past year
Hospital in-patient	1-2 days, 3-5 days, 6-10 days, >10 days in the past year

To construct a single health index, I follow the approach of Blundell et al. (2020) and estimate an ordered probit of an individual's subjective reported health on a rich dataset of objective health measures, and then take the predicted values from this regression to be the individual's health index. I run the following ordered probit regression where H_{it}^* is the unobserved continuous latent general health variable and

H_{it} is the observed ordinal general health score assessed by the individual in period t. $H_{it} = \{1, 2, 3, 4, 5\}$ where 1 = poor, 2 = fair, 3 = good, 4 = very good and 5 = excellent. X_{it} is a vector of objective measures and some additional controls, and ϵ_{it} is the individual error term. The included controls are age, sex, an employment dummy, occupation class, and month and year dummies. The Pseudo-R square from these ordered probit regressions is around 0.2. Each wave is estimated seperately, and a sample of the regression output is reported in Appendix 8.2.

$$H^*_{it} = X^{'}_{it}\beta + \epsilon_{it}, \quad \epsilon_{it} \sim \mathcal{N}(0,1) \;\; \forall \; i = 1...N, t = 1...T$$

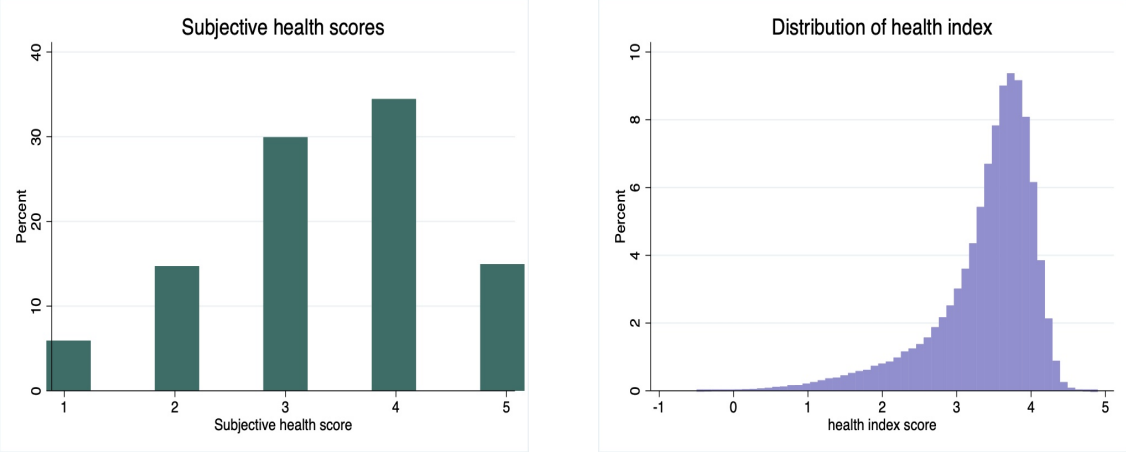
$$H_{it} = j \quad \text{if } \mu_j < H^*_{it} < \mu_{j-1}, \quad j = \{1, 2, 3, 4, 5\}$$

The probability that individual i selects general health value j in period t is:

$$Pr(H_{it} = j) = \Phi(\mu_j - X^{'}_{it}\beta) - \Phi(\mu_{j-1} - X^{'}_{it}\beta)$$

I then map the ordered probit fitted values onto the general health scores using a linear regression of the subjective scores onto the predicted values, and re-calculate the fitted values. The distribution of the original subjective health scores and constructed health index is shown in Figure 1.

Figure 1



The constructed health index can be interpreted as the average subjective health score reported by all individuals with the same medical condition diagnoses and disabilities, controlling for individual characteristics such as age and sex. The distribution of these scores is left-skewed due to a large tail of individuals in poor health, and

its kurtosis is around double that of a normal distribution, with many individuals bunching around the modal score.

Figure 2 indicates that differences in health index values between men and women are small. I include men and women in the same regression to calculate the health indices but include a gender dummy variable to allow for some variation by gender. Average health index scores gradually decline with age, although they are fairly stable between the ages of around 55-65. Variance in health scores increases with age, especially from the age of around 50. To strip out this decline in health by age, I demean the health index by regressing the health index against age, higher powers of age up to order four, sex, and month/year dummies. The residuals from this regression become the 'demeaned health index' that I use to model health as a dynamic process in subsequent chapters of this thesis

Figure 2: Distribution of health index values by age and gender

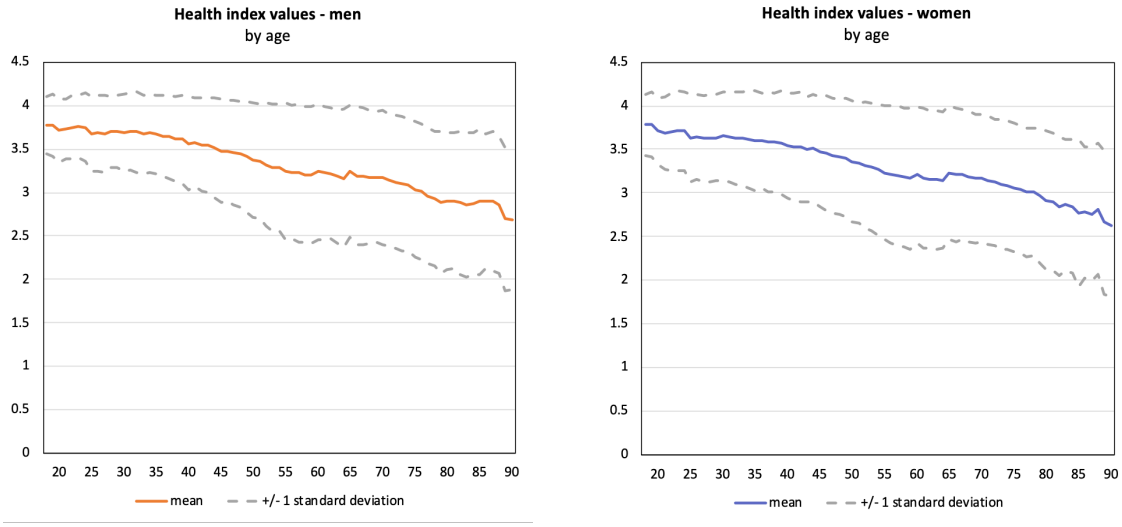
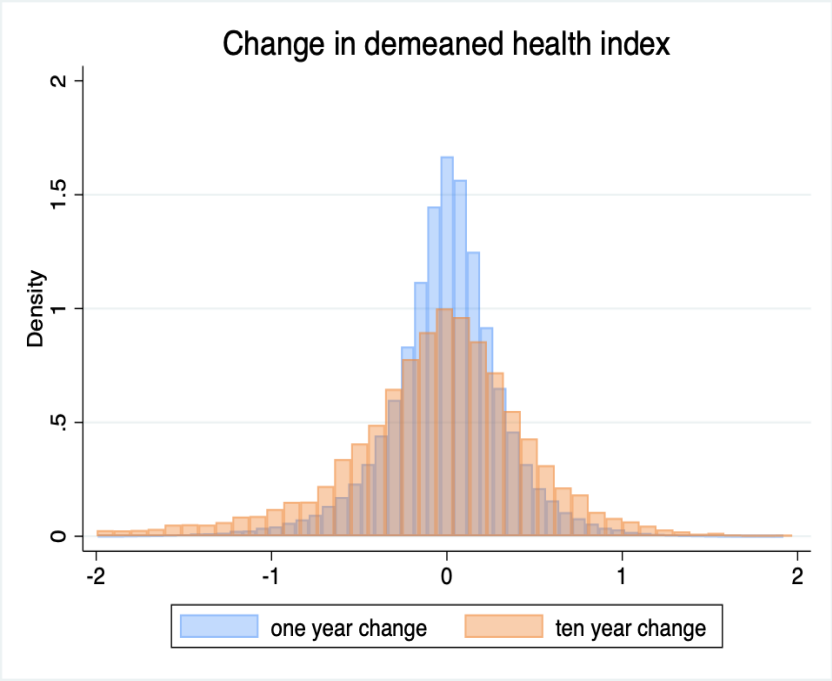


Figure 3 shows the distribution of changes to an individual's demeaned health index over one year and ten years. In both cases the distribution has a slight negative skew, of -0.5 and -0.8 respectively. The approximate symmetry of shocks supports the use of simple linear ARMA models that impose symmetry of shocks.

Figure 3



A potential concern is that attrition rates vary systematically by health. In my dataset, around 18 per cent of observations do not have an observation next period, either due to attrition or missing data. Estimating a linear probability model of attrition indicates that those in the lowest health quintile are two per cent more likely to not report health data next period relative to those in better health. However, the literature is fairly sanguine about the risks using health indices for economic research when there is differential attrition risk by health (Jones, Koolman and Rice, 2006), Pudney and Watson (2013), and I conclude that attrition is not a significant threat to my modelling of health dynamics. Further analysis of attrition in my dataset is reported in Appendix 8.1.

3.2 Allostatic scores from biomarker data

Between 2010-12, a subset of 8,465 individuals from waves 2 and 3 of the main Understanding Society survey were visited by a nurse for a physical health check and gave a blood sample. I use this biomarker (biological marker) data to construct a second index that approximates a component of underlying health called 'allostatic load'. This is a medical concept that reflects the risk from the cumulative effects of exposure to physical, psychosocial and environmental stressors that increase the risk of developing chronic diseases (Group, 2001). As a measure of cumulative wear and tear to the body, allostatic load is theoretically quite close to overall health stock or working capacity, although it cannot capture mental health or physical injury or disability.

Biomarker data can improve health dynamics modelling for several reasons. They can be measured with less error than other health data that rely on an individual accurately describing their health. The availability of biomarker data is likely to grow rapidly following the increasing popularity of health technology such as smart watches. They can help predict future health and mortality risk in ostensibly healthy individuals (Davillas and Pudney, 2020*b*,*c*). Davillas and Pudney (2020*a*) find that combining subjective health data with biomarker data significantly improves their predictions of future disability risk. This is because biomarker data incorporates health information such as kidney function and hormonal balance that may not be known by the individual, and because it offsets people's bias towards overweighting certain health information such as obesity and blood pressure and under-weighting other information such as strength and lung function. Biomarker data can also help disentangle the endogeneity between health and economic outcomes, and have been used to better understand the income-health gradient (Davillas, Jones and Benzeval, 2019), the impact of economic insecurity and childhood economic circumstances on health (Niedzwiedz et al., 2017; Davillas and Jones, 2020), and comparing the health impact of becoming re-employed in poor quality work compared to remaining unemployed (Chandola and Zhang, 2017).

To construct the allostatic score index, I normalise and then aggregate the biomarker data. I follow the approach of Davillas and Pudney (2020*a*) and take the simple average of the z-scores of 12 biomarkers and physical indicators reported in Table 4. I

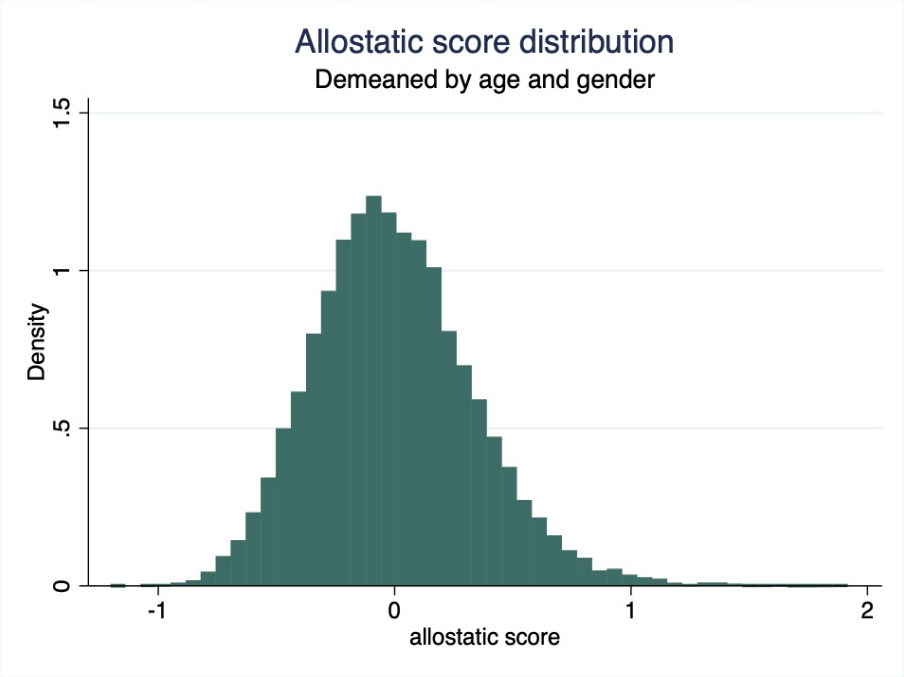
then demean the allostatic scores by age and gender to match how I constructed my health index. The subsequent distribution is approximately normal (Figure 5).

Figure 4: Biomarkers used in allostatic load index construction

Indicator	Data	Description
waist-to-height ratio	waist circumference, body mass index	obesity indicator
pulse	resting heart rate	lower heart rate associated with more efficient heart function
blood pressure	systolic, diastolic	two readings treated as separate indicators
lung function	forced vital capacity (FVC)	total amount of air forcibly blown out after a full inspiration using a spirometer
blood sugar	glycated haemoglobin levels (HbA1c)	measures glucose intolerance, a good indicator of diabetes risk
inflammation	C-reactive protein (CRP)	is a protein in the blood that rises in response to inflammation general chronic or systemic inflammation. High levels are risk factor for cardiovascular disease and mortality.
kidney function	creatinine	Creatinine is a waste product of muscle function, which is passed through the kidneys and excreted in urine. Glomerular filtration rate (eGFR) calculated using creatinine data according to calculation cited in Benzeval et al. (2014). Indicates how effectively the kidneys are 'cleaning' the blood.
liver function	albumin levels	albumin is main protein made by the liver. Low levels may be indicative of a loss of liver function
steroid hormone	ihydroepian-drosterone sulphate (DHEAS)	one of the primary mechanisms through which psychosocial stressors may affect health. Low levels associated with cardiovascular risk and all-cause mortality
cholesterol	high-density lipoprotein cholesterol (HDL)	'good' cholesterol that helps remove other forms of cholesterol from the bloodstream. High levels lower risk of cardiovascular disease.
grip strength	maximum grip strength	correlated with overall body strength, lower scores associated with decreased physical function, disability and mortality

A limitation of using this data is that I only have one set of biomarker observations per individual. However, the predictive content of allostatic scores is fairly stable over time. I show this by regressing my health index against the allostatic score index, varying the time gap between the data used for the health index and the

Figure 5



allostatic score (Table 6). An allostatic score has similar predictive power for a health index based on survey data collected one year later to a health index based on data collected ten years later. This suggests that allostatic scores capture a stable, long-term measure of health. There is a planned second round of biomarker data collection during wave 16 of Understanding Society in 2024-26, which can be used to check the stability of biomarker data over time more formally (Meena Kumari and Benzeval, 2022).

Figure 6: Health index predictive content of allostatic scores

	Number of waves between collection of allostatic score and health index data								
	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
allo.	0.527***	0.533***	0.526***	0.517***	0.522***	0.562***	0.583***	0.468***	0.439***
	(0.0208)	(0.0207)	(0.0218)	(0.0217)	(0.0219)	(0.0231)	(0.0240)	(0.0240)	(0.0237)
R-sq	0.080	0.082	0.073	0.073	0.075	0.083	0.086	0.061	0.058
Obs	7434	7456	7400	7269	7024	6519	6249	5895	5536

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Alongside biomarker data, genetic data was also collected. However, the Understanding Society genetic data is safeguarded special license data. Therefore, I perform

some preliminary analysis using an alternate dataset, The English Longitudinal Study of Ageing (ELSA) on whether genetic data can be used as an additional health risk indicator when modelling health dynamics. The ELSA dataset reports polygenic scores for a variety of behavioural, emotional and health-related phenotypes, which estimate an individual's propensity to develop many different physical and mental health conditions. However, I find that while the genetic scores do contain additional information on future health outcomes not captured by the the health or allostatic indices, the size of the effect is too small to significantly improve my modelling of the overall health process. Further details of my analysis is reported in Appendix 8.3.

4 Modelling health as a dynamic process

Health is a complex dynamic process that is subject to shocks that vary in magnitude and persistence. Heterogeneity between individuals is also large. In this section I use standard panel data techniques to identify how best to model health as a simple linear process. I estimate two baseline models that replicate the two most commonly used approaches to modelling health dynamics: an ARMA(p,q) model and a linear additive shock model that is the sum of a permanent process and a transitory MA(1) process. I find that an ARMA(1,1) model with a large AR coefficient and a moderately-sized negative MA coefficient best fits the data, although there are circumstances where the extra flexibility of the linear additive shock model to capture two different shocks may be desirable. I then evaluate how effective these models are in capturing health dynamics accurately. I show that while these models can be appealing due to their simplicity and intuitive interpretation, they have some important limitations which I discuss in detail in the next section.

4.1 ARMA(p,q) baseline model

The two data attributes that I wish to capture in any baseline model of health are the persistence of innovations and cross-sectional heterogeneity between individuals. My starting point is the simplest linear models that incorporate persistence; the autoregressive moving average (ARMA) class of models. I model health of individual i in period t, h_{it}, as an ARMA(p,q) process that includes a fixed effect μ_i:

$$h_{it} = \sum_{k=1}^{p} \rho_k h_{i,t-k} + \sum_{j=1}^{q} \theta_j \varepsilon_{i,t-j} + \mu_i + \varepsilon_{it}$$
$$i = 1...N, \quad t = 1...T$$

The p lags of the ρ term make up the autoregressive AR(p) components, and the q lags of the θ term make up the moving average MA(q) components. It is important to note that the health process I estimate is based on data that has been detrended by age and gender. This was done by regressing the raw health index against these observable variable and taking the residuals as the detrended health index. This detrending is quite common in the literature, perhaps due to familiarity with modelling the component of earnings growth that is unexplained by observables

such as experience and education. Furthermore, detrending by age removes the time trend as health declines over time, reducing the risk that the process is non-stationary. Small changes in survey design between waves is controlled for by including time dummies. Nonetheless, it may be attractive for the researcher to explicitly model the decline in health as people age. I replicate the key empirical work in this section with the original non-detrended health index, reported in Appendix 8.4. I find that my results are robust to using a non-detrended index.

I firstly test for stationarity, and find that at least a significant proportion of the series is stationary. I use the Born and Breitung (2016) test for panel series correlation, as it is designed to be robust to fixed effects and heteroskedasticity. A non-stationary pure random walk model would result in the autocorrelation of differenced health with its second (and higher) lag to be zero, which is not what we observe. Instead, this pattern of gradually decreasing autocorrelation in first differences is consistent with a persistent autoregressive process or a MA(q) process with a large q.

Table 3: Born and Breitung test for panel series correlation

	levels		first difference	
	LM(k)-stat*	p-value	LM(k)-stat	p-value
lag 1	36.30	0.000	-44.24	0.000
lag 2	24.63	0.000	8.36	0.000
lag 3	8.09	0.000	9.20	0.000
lag 4	-14.10	0.000	7.69	0.000
lag 5	-25.87	0.000	3.16	0.002
lag 6	-23.97	0.000	8.52	0.000
lag 7	-25.21	0.000	2.87	0.004
lag 8	-24.45	0.000	1.82	0.068
lag 9	-20.21	0.000	1.90	0.057
lag 10	-15.72	0.000	-1.22	0.224

*LM(k) test statistic is a modified t test of $\zeta = -1/(T-1)$. ζ from equation $h_{it} - \overline{h_i} = \zeta(h_{i,t-k} - \overline{h_i}) + \epsilon_{it}$. k is the lag order being tested

In general, the literature finds mixed evidence of health following a random walk process as opposed to a highly persistent one. Blundell et al. (2020) do find evidence of a random walk, while Blundell et al. (2016) estimate the coefficient on the first lag of health to be 0.9-1.1 depending on the sub-sample, and Heiss, Venti and Wise (2014) estimate an overall coefficient of 0.9. Whether papers model health as highly persistent or permanent processes likely reflects sample selection or health index con-

struction. For example, the use of a dataset such as ELSA or Health and Retirement Study (HRS) that only include older individuals will have a higher proportion of highly-persistent health shocks compared to a more representative sample which will contain a higher proportion of less-persistent health shocks such as changes in mental health index scores. Secondly, there are different ways to construct health indices, and some may place more weight on more permanent health indicators such as disability diagnoses compared to indicators of temporary health conditions such as infectious disease history or mental health indexes. Blundell et al. (2020) use ELSA data and a health index that emphasises disability indicators, therefore it is unsurprisingly that they find evidence of a random walk.

I begin by estimating an AR(p) model using OLS with various values of p, reported in Table 4. The OLS estimates indicate that health is highly persistent, with the sum of coefficients on the lagged health terms consistently around 0.9. A major concern of using OLS is that the coefficient estimates may be spuriously high due to the presence of fixed effects. I strip them out using first differencing and avoid the resultant Nickell bias by using GMM estimation techniques. I use the Arellano-Bond 'Difference GMM' estimator which mitigates Nickell bias by instrumenting the lagged dependent variable terms with further lagged terms in levels. I re-estimate the AR(p) model, adopting the following specifications which are selected to be conservative and robust: two-step estimator, time dummies, robust standard errors clustered at the individual level and an 'unadjusted' initial weighting matrix (Windmeijer, 2000). I include the Windmeijer correction to correct for the usually-negative bias in finite samples when the two-step estimator is used (Windmeijer, 2005). To prevent over-proliferation of instruments, I 'collapse' the instrument set and only include instruments based on the first to fifth lag of the variable being instrumented. My results are robust to various alternate specifications such as forward orthogonal deviations and different weighting matrices. I report the results of this exercise in Table 5. The MA(0) and MA(1) specifications reflect whether I allow the first lag to be used as an instrument.

Table 4: OLS estimates of the health process as AR(p) model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
L1.health	0.853***	0.603***	0.537***	0.504***	0.485***	0.470***	0.469***	0.454***
	(386.25)	(163.95)	(116.05)	(95.91)	(82.66)	(70.34)	(62.31)	(49.62)
L2.health		0.300***	0.243***	0.221***	0.213***	0.197***	0.203***	0.208***
		(80.71)	(47.12)	(38.08)	(32.27)	(27.06)	(25.85)	(22.96)
L3.health			0.134***	0.108***	0.0938***	0.0825***	0.0878***	0.0922***
			(31.61)	(18.40)	(13.57)	(10.67)	(10.23)	(9.33)
L4.health				0.0801***	0.0646***	0.0613***	0.0646***	0.0643***
				(16.12)	(9.61)	(7.75)	(7.15)	(6.23)
L5.health					0.0533***	0.0350***	0.0352***	0.0423***
					(8.93)	(4.29)	(3.79)	(3.89)
L6.health						0.0592***	0.0335***	0.0351**
						(8.45)	(3.55)	(2.99)
L7.health							0.0176*	-0.00891
							(2.17)	(-0.73)
L8.health								0.0149
								(1.48)
Observations*	222172	177129	142823	114489	89994	69512	52415	37934

Standard errors in parentheses; clustered standard errors
*Exclude individuals where n<5 (non-consecutively)
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Difference-GMM estimates of the health process as an AR(p) model

	AR(1)		AR(2)		AR(3)		AR(4)		AR(5)	
	MA(0)	MA(1)	MA(0)	MA(1)	MA(0)	MA(1)	MA(0)	MA(1)	MA(0)	MA(1)
L1.health	0.241***	0.944***	0.478***	0.971***	0.571***	1.008***	0.574***	1.184***	0.614***	0.979***
	(0.00949)	(0.0342)	(0.0152)	(0.0528)	(0.0196)	(0.0991)	(0.0235)	(0.156)	(0.0281)	(0.134)
L2.health			0.147***	-0.0106	0.203***	-0.0453	0.223***	-0.190	0.241***	-0.00671
			(0.00748)	(0.0186)	(0.0101)	(0.0560)	(0.0125)	(0.103)	(0.0154)	(0.0888)
L3.health					0.0681***	-0.0101	0.0871***	-0.0605	0.0916***	-0.00324
					(0.00678)	(0.0195)	(0.00861)	(0.0375)	(0.0106)	(0.0352)
L4.health							0.0351***	-0.0170	0.0380***	-0.00104
							(0.00688)	(0.0158)	(0.00894)	(0.0169)
L5.health									0.00398	-0.00925
									(0.00850)	(0.0106)
AB test, order 1, z score	-60.29	-31.22	-51.10	-18.73	-45.02	-9.10	-38.71	-6.38	-34.47	-6.14
AB test, order 1, p value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AB test, order 2, z score	11.19	16.04	-1.79	9.89	2.00	4.58	-1.28	3.68	-2.32	2.64
AB test, order 2, p value	0.000	0.000	0.075	0.000	0.045	0.000	0.199	0.000	0.020	0.008
Hansen J test stat	550.65	0.91	127.31	2.51	35.64	6.83	28.25	5.51	20.60	10.99
Hansen J test p value	0.000	0.823	0.000	0.473	0.000	0.078	0.000	0.138	0.000	0.012
moment conditions	16	15	16	15	16	15	16	15	16	15
Observations	222095	222095	184734	184734	151622	151622	121698	121698	94513	94513

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

These results strongly suggest that an ARMA(1,1) model best suits the data. All MA(0) specifications that include instruments based on the immediately preceding lag result in a strong rejection of the null of the Hansen J test, indicating that the model is wrongly specified. However, excluding this instrument, which the MA(1) specifications do, is typically sufficient to change the result of this test and fail to reject the null. For example, excluding $h_{i,t-2}$ as an instrument for $\Delta h_{i,t-1}$ and only use $h_{i,t-3}$ and earlier lags lead to the non-rejection of the null. This strongly suggests that the errors follow an MA(1) process. This result is also supported by the Arellano-Bond autocorrelation tests, which identify autocorrelation up to order 2. When I exclude the first lag as an instrument, the point estimate of the coefficient on the first lag of health is much higher at around unity while the coefficients on all the subsequent lags are small and not significant. This indicates that including only one lag of health is sufficient.

Incorporating some additional moment conditions by using Blundell-Bond 'System GMM' estimation leads to improved ARMA(1,1) estimates. It is well known that the Arellano-Bond estimator does not function well when persistence is high. At the limit, if health follows a random walk ($\rho_1 = 1$) then the difference GMM instruments are uninformative. Blundell and Bond (1998) suggest there is a risk of serious finite sample bias at ρ_1 values of 0.8 and higher, although they suggest the bias is smaller with very large samples. The System GMM estimator typically performs much better in these circumstances. The additional moment conditions can also contribute to more precise coefficient estimation. This is particularly helpful as having to only use further lags as instruments due to the MA(1) error structure increases the risk of weak instruments. The additional initial moment restriction of $\mathbb{E}(\varepsilon_{it} h_{i1}) = 0$ that is required for System GMM estimation is not a particularly onerous restriction for my data. Blundell and Bond (2023) state that this restriction holds automatically if the same process has generated the series for long enough before the start of the sample period. Since my first observation occurs at least 18 years after the the start of the health process at birth this undoubtedly holds. This condition is also easier to fulfil when ρ_1 is large, which is also the case in my data.

Table 6 reports the AR(p) model coefficients estimated using System-GMM and allowing for MA(1) errors. Differences between the Difference and System GMM coefficient estimates are small, although using System GMM leads to much more precisely

estimated coefficients, especially for the first lag. The coefficient estimates of the first lag are mostly not significantly different for the AR(1) AR(2) and AR(3) specification, and the coefficients on additional lags are typically not significant. Therefore, including only one lag is sufficient to capture the persistence dynamics in this model.

Table 6: System-GMM estimates of the health process as an AR(p) model

	AR(1)	AR(2)	AR(3)	(AR4)	(AR5)
			MA(1) assumption		
L1.health	0.872***	0.901***	1.032***	1.113***	1.149***
	(0.0123)	(0.0310)	(0.0790)	(0.0864)	(0.0980)
L2.health		-0.00940	-0.0775	-0.126**	-0.135**
		(0.0170)	(0.0430)	(0.0448)	(0.0481)
L3.health			-0.0194	-0.0361	-0.0432*
			(0.0166)	(0.0199)	(0.0217)
L4.health				-0.0160	-0.0115
				(0.00844)	(0.0116)
L5.health					0.00608
					(0.00720)
AB test, order 1 z score	-53.86	-26.25	-11.81	-11.8	-10.78
AB test, order 1 p value	0.000	0.000	0.000	0.000	0.000
AB test, order 2 z score	19.19	12.69	6.27	6.88	6.32
AB test, order 2 z score	0.000	0.000	0.000	0.000	0.000
AB test, order 3 z score	0.032	-0.54	-0.39	0.08	1.75
AB test, order 3 p value	0.974	0.588	0.693	0.935	0.080
AB test, order 4 z score	0.867	0.54	-0.16	-0.36	-1.32
AB test, order 4 p value	0.386	0.592	0.871	0.721	0.188
Hansen J test stat	6.42	8.259	8.109	11.27	35.24
Hansen test p value	0.170	0.143	0.23	0.127	0.000
moment conditions	16	17	18	19	20
Observations	222095	184734	151622	121698	94513

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

I conclude that the best single-equation linear specification to capture the health process is the following ARMA(1,1):

$$h_{it} = 0.87h_{i,t-1} - 0.33\varepsilon_{i,t-1} + \eta_i + \varepsilon_{it}$$

Since the AR term has already been estimated as 0.87, I estimate that the coefficient on the MA(1) term, θ, is -0.33 by re-arranging the ARMA(1,1) model as: $h_{it} - 0.87h_{i,t-1} = \tilde{h_{it}} = \eta_i + +\theta\varepsilon_{i,t-1} + \varepsilon_{it}$. This is now a simple MA(1) process

that can be estimated using GMM. I use the following three variance and covariance moments and report the coefficient estimates in Table 7:

$$\text{Var}(\tilde{h_{it}}) = \mathbb{E}\eta_i^2 + (1+\theta^2)\mathbb{E}\varepsilon^2$$

$$\text{Cov}(\tilde{h_{it}}, \tilde{h_{i,t-1}}) = \mathbb{E}\eta_i^2 + \theta\mathbb{E}\varepsilon^2 \quad \text{Cov}(\tilde{h_{it}}, \tilde{h_{i,t-2}}) = \mathbb{E}\eta_i^2$$

Table 7: GMM estimates of MA(1) process

	ρ =0.87
η_i	0.0380***
	(0.00285)
θ	-0.334***
	(0.00593)
ε_{it}	0.335***
	(0.00125)
Observations	222095

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As a robustness exercise, I check whether there are large subgroups with health dynamics that are better captured by a different linear model. If this were the case, describing the health dynamics of the entire sample using a single ARMA(1,1) model may be misleading. I use the Sarafidis and Weber (2015) K-means clustering algorithm to divide the sample into as many clusters as required for the estimated slope coefficients of an AR(1) model to be the same within each cluster, accounting for individual-specific fixed effects. The algorithm divides my sample into two groups, containing 40 and 60 per cent of the sample respectively. This suggests that only two groups are needed to capture any heterogeneity in model coefficients. I then estimate AR(p) models separately for each group using GMM, re-assessing whether including one lag is sufficient and whether the error structure follows an MA(1) process. The regression tables are reported in Appendix 8.5, as well as some summary statistics for each group. I determine that the models for the two groups that best fit the data are an AR(2) and ARMA(1,1) respectively.

$$\text{Group 1: } h_{it} = 0.79h_{i,t-1} + 0.09h_{i,t-2} + \eta_i + \varepsilon_{it}$$

$$\text{Group 2: } h_{it} = 0.83h_{i,t-1} - 0.55\varepsilon_{i,t-1} + \eta_i + \varepsilon_{it}$$

The two models are quite similar. Both capture that health is a highly persistent process, and have an additional term that helps distinguish between highly-persistent health shocks such as chronic health conditions, and transitory health shocks. I conclude that an ARMA(1,1) model is sufficient to describe the entire sample and slope heterogeneity is not a significant concern.

4.2 Linear additive shock model

I conclude this section with estimating a slightly different model that allows for more flexibility in capturing shock persistence, but at the expense of imposing other restrictions. A specification used very commonly in the earnings dynamics literature, and sometimes in the health dynamics literature, relaxes the restriction of individuals being subject to only one type of shock. Instead, the variable is modelled as the sum of two independent random processes: a permanent shock process which is typically a random walk, and a transitory process which is either an MA(0) or MA(1):

$$y_{it} = p_{it} + \upsilon_{it}$$

$$\text{permanent process: } p_{it} = p_{i,t-1} + \zeta_{it}$$

$$\text{transitory process: } \upsilon_{it} = \varepsilon_{it} - \theta\varepsilon_{it,1}$$

An additive classical measurement error $r_{it} \sim N(0, \sigma_r^2)$ can also be included. This clear distinction between permanent and transitory shocks reflects the influence of Friedman's Permanent Income Hypothesis on earnings dynamics research, but it is also conceptually attractive as researchers can cleanly classify most income shocks as either temporary, such as overtime or one-off-bonuses, or permanent, such as a job change (Meghir and Pistaferri, 2004). A similar intuition for health shocks being divided into permanent shocks such as a physical disability and temporary shocks such as some mental health episodes is compelling, and adopted in papers such as Blundell et al. (2020) and Blundell et al. (2016).

The canonical moment conditions used to estimate these models require that the permanent process is a random walk to achieve identification. This is a strong as-

sumption for my data. It is challenging to distinguish between highly persistent and random walk processes in small-T panel data with significant individual heterogeneity, and I cannot reject that the coefficient on the lagged health term is 1 in many of the ARMA(p,q) models I estimated using GMM. However, my serial correlation tests do indicate that a reasonable proportion of the data are best characterised as following a persistent process rather than a random walk. It is also unclear how robust the resulting coefficient estimates are to small violations of the random walk assumption implied by the moment conditions. Proceeding with caution, I use the following moment conditions to estimate health as the sum of a permanent walk and MA(1) transitory process. Letting g_{it} be a change in h_{it} (equivalent to $h_{i,t} - hi, t-1$) we can identify the variance of the permanent condition using the following moment condition from Meghir and Pistaferri (2004):

$$\mathbb{E}(\zeta_{it}^2) = \mathbb{E}\left[g_{it} \left(\sum_{j=-(1+q)}^{1+q} g_{i,t+j} \right) \right]$$

. Since we cannot separately identify the variance of any measurement error, the variance of the transitory shock, and θ, we can only use the moment conditions to place bounds on these coefficients with the following moment conditions:

$$\sigma_r^2 = \mathbb{E}(g_{it}, g_{i,t-1}) - \frac{(1+\theta)^2}{\theta}\mathbb{E}(g_{it}, g_{i,t-2})$$

$$\sigma_\varepsilon^2 = \frac{\mathbb{E}(g_{it}, g_{i,t-2})}{\theta}$$

By setting σ_r^2 to zero we can estimate the lower or upper bound of θ, which we assume is bounded between -1 and 1. The sign of $\mathbb{E}(g_{it}, g_{i,t-2})$ defines the sign of θ. In my case it is negative, therefore the maximum value of θ is the case where $\sigma_r^2 = 0$. I use these moment conditions to estimate the variance of the two shocks, as well as the coefficients of the MA(1) transitory process. These estimates are reported in Table 8. My estimates of the magnitude of the variances of the two shocks are quite similar to the findings of Blundell et al. (2016), although they do not find evidence of a MA(1) transitory process. Blundell et al. (2020) obtain quite different results and argue that transitory and permanent shocks contribute fairly equally to health variance. However, they use a very different estimation strategy and do not use these

canonical moments from the earning dynamics literature.

Table 8: Coefficient estimates of linear additive shock model

$\mathbb{E}(\zeta_{it}^2)$	0.155***
	(47.26)
σ_ε^2 if $\sigma_r^2 = 0$	0.050***
	(40.89)
θ if $\sigma_r^2 = 0$ (upper bound)	-0.072***
	(-6.42)

t statistics in parentheses, *p<0.05, *p<0.01, *p<0.001

If transitory shocks do not explain much variation of the overall health process, then using an ARMA(p,q) model that only allows for one type of shock is sufficient. My results suggest that the permanent process is responsible for the majority of the variance in the health process over time, although the transitory process does make some contribution. In addition, the ARMA(p,q) model does not require the persistent shock to follow a random walk, which my preferred ARMA(1,1) specification does not. I conclude that the ARMA(p,q) model is a superior fit for my data. I further consider the limitations of these two models in the next chapter, and suggest some improvements.

5 Capturing more complex dynamics

The two linear health models estimated in the previous section are simple to use and incorporate into more complex structural models. However, there is a cost to their simplicity. Since the baseline ARMA(p,q) model attempts to capture the average persistence of a health shock, it imposes uniformity of persistence on shocks of different sizes, for positive and negative health shocks, and for individuals with very different levels of health and health histories pre-shock. I find evidence of significant heterogeneity in persistence once I allow persistence to vary by these characteristics. Simple extensions of the ARMA(p,q) baseline model can capture some of this variation, however we can make further progress with more sophisticated modelling approaches, which I discuss in the subsequent chapter. In addition, I do not need to assume stationarity or that the error terms follow a white noise process for my ARMA(p,q) coefficient estimates to be valid. However, I document some features of the error distributions that are important to capture when modelling the heterogeneity in health shock risk that individuals face. I show that biomarker data can be used to capture some of the elevated negative health shock risk faced by some individuals.

This section focusses on the ARMA(1,1) model as my preferred linear model, but most of the limitations I identify can also be applied to the linear additive shock model. De Nardi, Fella and Paz-Pardo (2019) provide a good summary of the key limitations of this model when applied to earnings data which are equally valid when using health data. They key model assumptions they identify that do not match the data are: age independence of the second and higher moments of the conditional distribution of both the transitory and persistent components, normality of the shock distribution, and linearity of the process of the persistent component.

5.1 Recent health history

The average persistence of a health shock varies significantly depending on the health history of the individual prior to the shock taking place. This makes intuitive sense; someone's capacity to recover from an illness is a function of how healthy they were just before getting sick. The MA term in an ARMA(1,1) model takes into account the size of the shock last period, but there is significant additional persistence information in the level of health. To illustrate this I add an interaction term to the

baseline ARMA(1,1) model that assigns individuals to a quintile of their health just before the shock and interacts it with the lagged health term. This shows that that persistence is the highest for those with prior bad health, and lowest for those with prior average health. Simple modifications to the ARMA specification can allow for some heterogeneity in the estimates of persistence of health shocks experienced between period t and period $t-1$. However, they are unable to allow coefficients to depend on the sign or magnitude of the health shock between periods $t-1$ and t. I adopt more complex econometric techniques in a later section of this paper to capture this heterogeneity in the persistence of health shocks.

Table 9: ARMA(1,1) model with interaction dummy for lagged health level quintile

	Diff-GMM	Sys-GMM
Q1 - lagged health index	0.966***	0.923***
	(0.0697)	(0.0163)
Q2 - lagged health index	1.102***	0.910***
	(0.159)	(0.104)
Q3 - lagged health index	0.754***	0.665***
	(0.100)	(0.0719)
Q4 - lagged health index	0.901***	0.829***
	(0.0569)	(0.0433)
Q5 - lagged health index	0.871***	0.868***
	(0.0349)	(0.0260)
AB test, order 1 z score	-24.76	-56.093
AB test, order 2 z score	0.000	0.000
AB test, order 2 z score	14.3	20.835
AB test, order 2 p value	0.000	0.000
AB test, order 3 z score	0.0567	0.0915
AB test, order 3 p value	0.9548	0.927
Hansen J test stat	25.925	67.035
Hansen J test p value	0.0388	0.000
Observations	222095	222095

Standard errors in parentheses

GMM estimation specifications identical to those used in baseline model in Ch4

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

As well as the relationship between recent health history and persistence, there is also a relationship between recent health history and the expected distribution of future health shocks. This is difficult to capture in a simple linear model but is an important component of health risk to capture. To illustrate the relationship between past health and the expected distribution of health shocks, I graph the higher moments of the health index as a function of health percentile in the previous period.

Figure 7

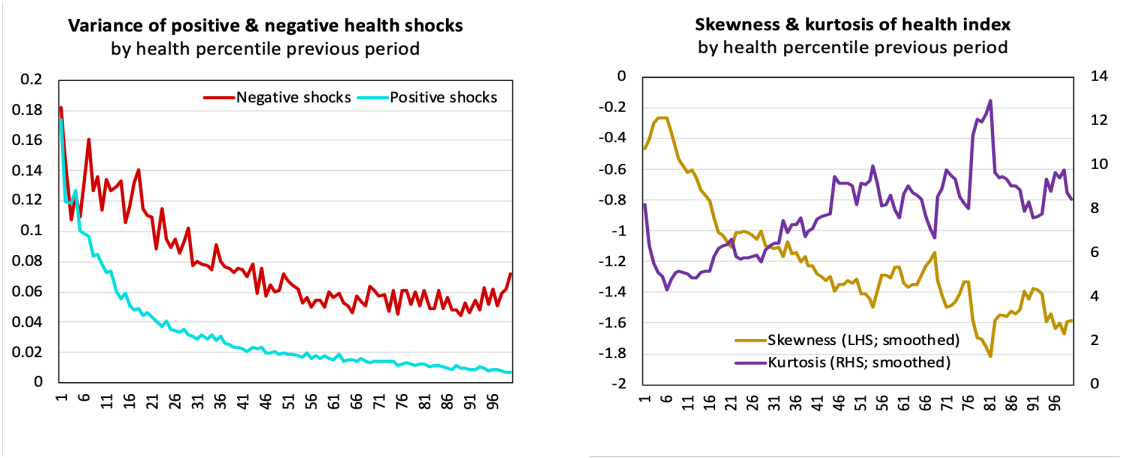
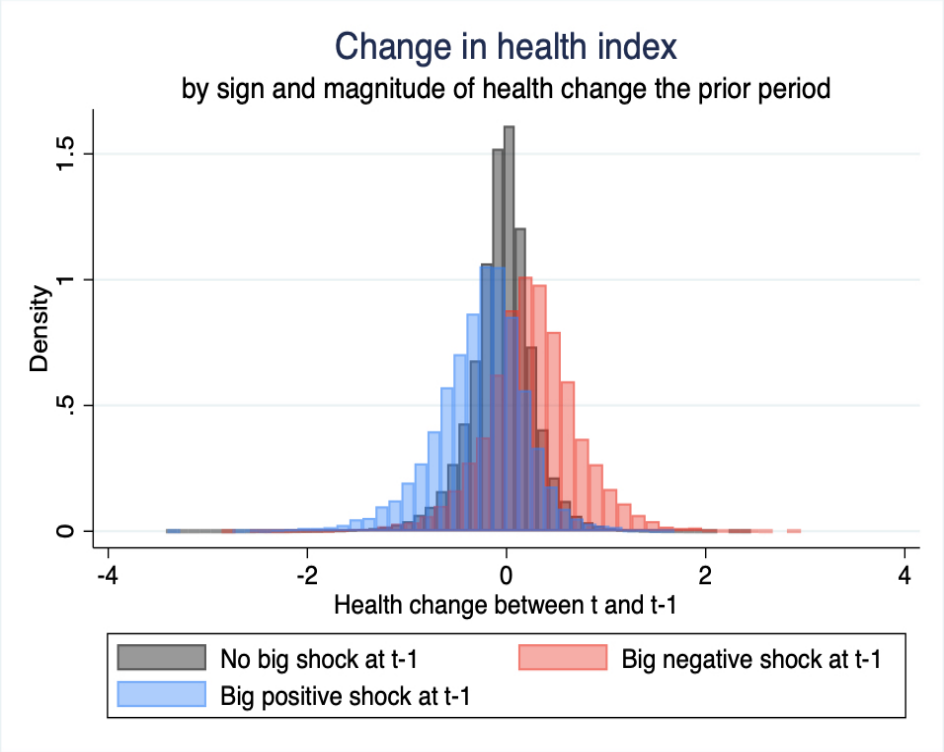


Figure 7 shows variance, skewness and kurtosis by health percentile the previous period. Variance, skewness and kurtosis all systematically vary by health the previous period. Notably, those in poor health have more volatile health in subsequent periods, with increased risk of both large negative and positive changes to health relative to those in good or average health. This elevated risk is difficult to capture in simple linear models. One plausible way of capturing this feature of the data is estimating an autoregressive conditional heteroskedasticity (ARCH) model. ARCH models are able to capture differences in variance depending on size of the error term the previous period. For example, a large shock in period $t-1$ may mean a large shock is more likely in period t. Figure 8 compares the histograms of health changes between period $t-1$ and t for those who experienced a greater than one standard deviation change in health in the prior period (between $t-2$ and $t-1$) to those who did not. It shows that large changes in health is associated with more volatile health next period, and that, on average, a large negative health shock is associated with an improvement in health next period, and the opposite for those who experience a positive health shock in the prior period. Interestingly, the distribution of health changes for those

who experience large positive or negative shocks the prior period is closer to a normal distribution than the distribution of health changes for those who experience neither. For this group, their is very little mass at the tails as stable health in the past is correlated with stable health in subsequent periods.

Figure 8



I assess this relationship between errors and variance more formally by estimating an ARCH(1) model with the following specification for health variance: $\exp(\gamma_0 + \gamma_1 \varepsilon_{i,t-1}^2 + \gamma_2 \varepsilon_{i,t-1})$. The γ_2 term accounts for possible heterogeneity between positive and negative shocks. I describe my estimation procedure in Appendix 8.6, but I do not find any evidence that γ_1 or $\gamma_2 \neq 0$, and therefore do not find evidence of ARCH effects in my data. However, this specification only models the relationship between shock magnitude in two consecutive periods. I do find evidence that individuals who experience a large negative health shock are more likely to experience another large negative health shock in subsequent years. However, a majority of these later shocks occur several years afterwards, which cannot be captured in an ARCH(1) model and requires a more complex econometric approach. Table 10 reports the number of large negative shocks, defined as at least one standard deviation fall in the detrended

health index, experienced by those of different ages in the sample. Conditional on experiencing one negative shock, individuals are more likely to experience a second. For example, those aged 20-29 at the beginning of the sample period have a 16 per cent chance of experiencing a negative shock in the next decade, but 28 per cent of those who experience one negative shock experience a second, with an average gap between shocks of four years. The average gap between negative health shocks rises with age.

Table 10: Number of negative shocks of ≥ 1 st. dev. over 10 year sample

Population share by age

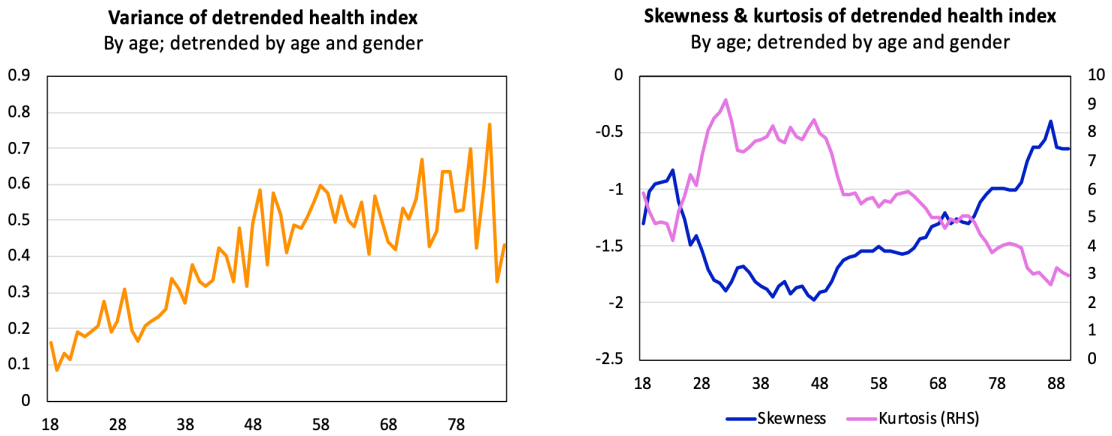
Age in first wave	0 shocks	1 shock	2+ shocks	ave yrs b/tween shocks
20-29	0.801	0.155	0.044	3.5
30-39	0.784	0.169	0.047	3.7
40-49	0.744	0.198	0.058	4.2
50-59	0.722	0.222	0.056	4.4
60-69	0.704	0.242	0.054	4.6
70-79	0.610	0.300	0.090	4.5

5.2 Age and model stationarity

A different source of heterogeneity in persistence and shock distribution is age of the individual. Age is closely related to the statistical property of stationarity. Since the time dimension of my panel data is fairly short, stationarity is difficult to assess. However, the ARMA(1,1) process I estimated is stationary in the long run, provided that $\rho + \theta \neq 0$, $\rho < |1|$, and some restrictions are imposed on the distribution of ε_{it}. Stationarity implies that the moments of the data are age independent. For the first moment, this is mechanically achieved by detrending the health index by age and age polynomials. However, higher moments of the detrended health data are not age-independent. Figure 9 graphs the second, third and fourth moments of the detrended health data by age. Older individuals are more likely to experience health shocks, and so the standard deviation of the health index increases by age. The distribution of the detrended health index of older people is less negatively skewed, reflecting

their increased propensity to experience positive health shocks. Young people are much less likely to experience positive health shocks as their health is typically good and so cannott be improved further. The health index distribution for young people is platykurtic and so extreme health changes are rarer, while the kurtosis for older people is close to a normal distribution. Therefore, higher moments do systematically vary by age, which should be captured in life-cycle models or models that consider the long-term impacts of health shocks on economic outcomes. This can be achieved by imposing a shock sequence that is a function of age rather than assuming a normal distribution for the error term. The ARMA(1,1) coefficients I estimate using GMM are robust to conditional heteroskedasticity and the patterns of kurtosis and skewness I identify; assuming mean-zero errors and no serial correlation of the errors is sufficient (Arellano and Bond, 1991). However, higher moments are an important component of capturing the health risk people face.

Figure 9



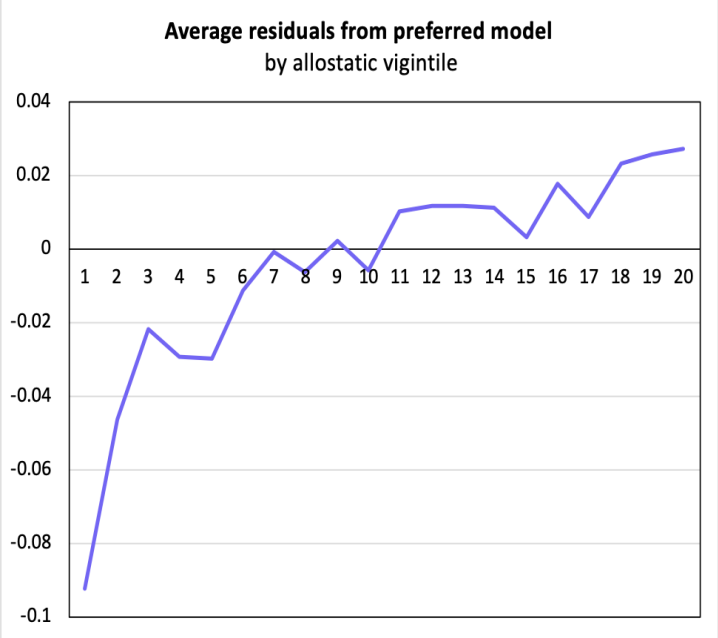
5.3 Underlying health

I conclude with considering how allostatic scores can be used as an additional data source to improve the performance of linear health dynamics models. I find that the main informational content of allostatic scores relates to the likelihood of a large negative shock realised in period t. In addition, I find that allostatic scores do not help predict the persistence of already realised health shocks (see Appendix 8.7 for further details).

The ARMA(1,1) model has the worst performance when predicting the health

of individuals with poor allostatic scores. Figure 10 shows the average difference between health predicted in period t using my preferred ARMA(1,1) model, and actual health in period t by allostatic score vigintile. The biggest forecast misses occurs for the population with the worst 10 per cent of allostatic scores, indicating very poor underlying health. If someone is in average health in period $t-1$ but has bad underlying health, they are more likely to be hit by a large negative shock in period t. In these cases, the ARMA(1,1) model performs the most poorly and significantly overestimates the level of health. This result is in line with previous research that finds that biomarker data can predict future negative health outcomes among ostensibly healthy people (Davillas and Pudney, 2020*c*). This increased propensity to experience a large negative health shock is an important source of risk to capture in models of health dynamics.

Figure 10



6 Non-linear health dynamics

The complex dynamics of health, as described in the prior section, can be better understood by adapting the latest panel data techniques. I estimate the health process using the non-linear panel data framework developed by Arellano, Blundell and Bonhomme (2017). This method is from the earnings literature, although it has been applied to a small number of non-earnings contexts, such as non-linear productivity and investment-dynamics in firms (Fella et al., 2021).* A major attraction of this method is that it allows for heterogeneity in persistence depending on the size and direction of the health shock that occurs in period t. This is not possible to do using the methods used to estimate the ARMA(p,q) models due to the fundamental endogeneity between the shock in period t and the persistence estimates that relate health in period $t-1$ to health in period t. This framework also allows persistence estimates to vary by the level of health in period $t-1$, which I previously showed can have a large impact on persistence estimates.

Adapting the Arellano, Blundell and Bonhomme (2017) framework to a health context produces persistence estimates that range from 0.6 to 1.2. While the linear methods from the prior section produce persistence estimates around the midpoint of these estimates, this range is large enough to have meaningful implications for economic decision making. People faced with a health shock with persistence at the lower end of this range are likely to behave quite differently to those facing a much more persistent health shock. Accurately capturing these differences in persistence can improve our understanding of the relationship between health and labour supply decisions, wages, productivity, consumption, and government insurance access. I also document some interesting patterns in how persistence estimates vary depending on whether the shock at period t is positive or negative, the magnitude of the shock, and the level of health immediately prior to the shock. I find that negative health shocks are more persistent than positive health shocks, and that negative health shocks are more persistent if someone was in poor health prior to the shock. I also estimate an additional model that includes fixed effects as an additional source of heterogeneity. Accounting for fixed effects does reduce the persistence estimates a little, especially

*Dal Bianco and Moro (2022) have written a working paper concurrent to this one that also applies this framework to a health context

for those in poor health who experience large negative shocks. I find some evidence that the size and sign of the fixed effect is correlated with allostatic scores which helps us understand the variation captured by the fixed effect. I conclude this section by extending this method to better capturing the complex dynamics of other health indicators by estimating the non-linear persistence of an index of mental health.

6.1 Non-linear persistence estimates of overall health

The non-linear framework of Arellano, Blundell and Bonhomme (2017) models their variable of interest as the sum of a persistent component (η_{it}) and a transitory innovation (ε_{it}). The linear model estimated in the previous section as likewise the sum of a permanent component and transitory innovation can be considered a special, highly-restrictive case of Arellano, Blundell and Bonhomme (2017)'s model. The persistent component is assumed to follow a general first-order Markov process, and so the η_{it} terms are dependent over time, although the nature of their dependence does not need to be specified, allowing for flexible temporal dynamics. The τth conditional quantile ($\tau \in (0,1)$) of this persistent component, given η_{it-1}, is $Q_t(\eta_{it-1}, \tau)$. υ_{it} is then defined as a random process such that :

$$\eta_{it} = Q_t(\eta_{i,t-1}\upsilon_{it}), \text{ where } (\upsilon_{it}|\eta_{i,t-1}, \eta_{i,t-2}\ldots) \sim \text{Uniform(0,1)}$$

. The quantile function maps draws of υ_{it} from a uniform distribution into quantile draws for the persistent component. The transitory component ε_{it} is assumed to be mean-zero, independent over time, and independent of $\eta_{i,t-s}$ for all s, and is assumed to also include any measurement error. This method allows for general forms of heteroskedasticity, conditional skewness and kurtosis in η_{it}. A caveat to this specification is that it excludes the possibility for the transitory component to follow an MA(1) process, which I do find some evidence for when estimating the baseline models. The t subscript refers to age. The permanent and transitory components are assumed to be mean-independent of age t, but the conditional quantile functions and marginal distributions of the transitory component may all depend on t. Non-linear persistence (ρ_t) of the persistent component can then be defined as:

$$\rho_t(\eta_{i,t-1}, \tau) = \frac{\partial Q_t(\eta_{it-1}, \tau)}{\partial \eta}, \quad \rho_t(\tau) = \mathbb{E}\frac{\partial Q_t(\eta_{it-1}, \tau)}{\partial \eta}$$

$\delta Q_t/\delta\eta$ is the partial derivative of Q_t with respect to its first argument, and the expectation is taken with respect to the distribution of η_{t-1}. This approach estimates persistence as the derivative effect of how much the persistent component of earnings at period t varies with the persistent component of earnings at period $t-1$ when hit with a shock at period t. I estimate $\rho_t(\eta_{i,t-1}\tau)$ of the health process, which is the persistence of $\eta_{i,t-1}$ when hit by shock with rank τ.

A major attraction of this method is that it allows for one shock to wipe out the memory of past shocks. This incorporates an important additional source of heterogeneity in health shock persistence that is unavailable in the simple linear models. This allows, for example, a big negative shock in period t, such as a sudden permanent severe disability, to wipe out the persistence of past shocks. By contrast, the ARMA(p,q) and simple linear additive shock models cannot allow ρ to vary by any features of the shock that occurs in period t. Despite its computational complexity, the method is easy to use as Arellano, Blundell and Bonhomme (2017) make available full MATLAB replication files.[†] Furthermore, De Nardi, Fella and Paz-Pardo (2019) propose a simulation-based method to discretize nonlinear and non-normal stochastic processes, so that these estimates can be incorporated into a life-cycle model with minimal state-space cost.

To estimate the model, the quantile functions for ε_{it}, η_{i1} and η_{it} are first parameterised as low order Hermite polynomials. Since the persistent and transitory components of the process are not seperately observable, the estimation algorithm begins with an initial guess for the coefficients and then iterates sequentially between draws from the posterior distribution of the latent persistent component and quantile regression estimation until convergence is achieved. The algorithm used is closely related to the stochastic EM algorithm (Diebolt and Celeux, 1993), although the quantile specification of the model avoids the need for a likelihood-based approach to estimation.

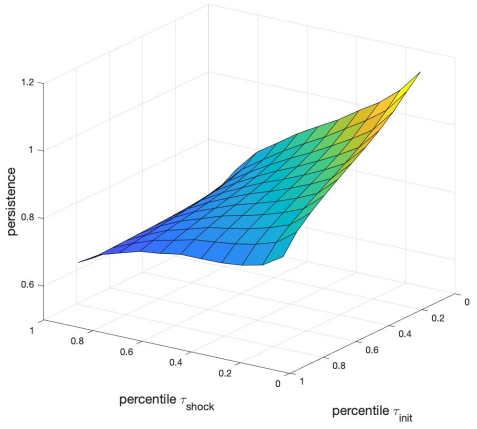
I apply this method to estimating the persistence of health, and report the results in Figure 11 and Table 11 by deciles for the magnitude of the shock at period t and health decile in period *t-1*. Since the health index has been demeaned by age, the health shocks are approximately symmetric, so the lowest decile constitutes large

[†]All replication files and supplementary material can be downloaded from: `https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13795`

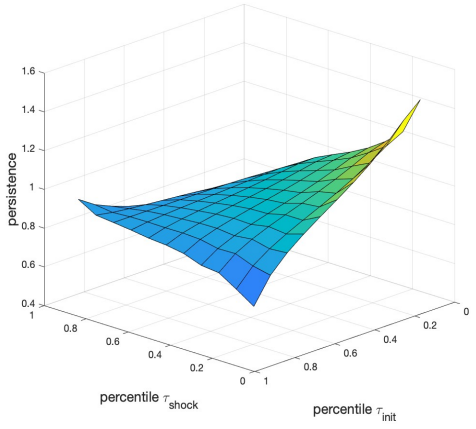
negative health shocks, the median decile consists of very small health shocks or unchanged health, and the highest decile consists of very large positive shocks. I report both the persistence estimates for the overall health process, and just the persistent component η_{it}, which strips out the transitory component from the overall estimates. The persistent-component-only estimates are on average higher, with two notable exceptions; large positive shocks experienced by those in poor prior health, and large negative shocks experienced by those in prior good health. Transitory shocks are likely to be more important in these cases.

Figure 11: Non-linear persistence estimates

(a) Health ($\eta_{it} + \varepsilon_{it}$)



(b) Persistent component only (η_{it})



I find that persistence of health shocks varies greatly, depending on past health, shock size and sign. While the average of my estimates is approximately the estimate of persistence from my baseline models, my non-linear persistence estimates range from 0.6 to 1.2. Such variation has significant implications for economic decision making. Furthermore, there is a large difference in the persistence of positive health shocks and negative health shocks. Large negative health shocks are almost twice as persistent as large positive health shocks. Another notable result is that those in poorer health pre-shock take much longer to recover from a negative shock relative to those in better health pre-shock. Individuals who are both in poor health in period $t-1$ and then experience a large negative health shock in period t have an estimated persistence coefficient of 1 or more, suggesting that a negative shock is likely to be permanent for these individuals. By comparison, an ARMA(p,q) model will underestimate the persistence of a large negative health shock and overestimate

Table 11: Health persistence estimates using Arellano, Blundell and Bonhomme (2017) framework

	Shock size percentiles*										
	1	2	3	4	5	6	7	8	9	10	11
$health_{t-1}$*											
	Overall health persistence										
1	1.16	1.10	1.05	1.01	0.97	0.94	0.90	0.86	0.82	0.76	0.67
2	1.13	1.06	1.01	0.98	0.94	0.91	0.87	0.83	0.79	0.73	0.65
3	1.08	1.02	0.98	0.94	0.91	0.88	0.85	0.81	0.78	0.72	0.64
4	1.05	0.99	0.95	0.92	0.89	0.86	0.83	0.80	0.77	0.71	0.64
5	1.02	0.96	0.93	0.90	0.87	0.85	0.82	0.79	0.76	0.71	0.64
6	0.99	0.93	0.90	0.88	0.86	0.83	0.81	0.78	0.75	0.71	0.64
7	0.96	0.91	0.88	0.86	0.84	0.82	0.80	0.78	0.74	0.71	0.65
8	0.93	0.88	0.86	0.84	0.82	0.81	0.79	0.77	0.74	0.70	0.65
9	0.90	0.85	0.83	0.82	0.80	0.79	0.78	0.76	0.73	0.70	0.65
10	0.87	0.82	0.80	0.79	0.78	0.78	0.76	0.75	0.73	0.70	0.66
11	0.81	0.77	0.76	0.75	0.75	0.75	0.74	0.74	0.72	0.70	0.66
	Persistent component of health shocks										
1	1.45	1.25	1.17	1.11	1.06	1.02	0.98	0.92	0.86	0.77	0.56
2	1.37	1.21	1.15	1.09	1.05	1.01	0.98	0.93	0.87	0.80	0.62
3	1.28	1.15	1.11	1.06	1.03	1.00	0.96	0.92	0.88	0.81	0.67
4	1.21	1.10	1.07	1.03	1.01	0.98	0.95	0.92	0.88	0.83	0.71
5	1.14	1.05	1.03	1.00	0.99	0.97	0.94	0.91	0.88	0.84	0.74
6	1.08	1.01	1.00	0.98	0.97	0.95	0.93	0.91	0.88	0.84	0.77
7	1.02	0.97	0.97	0.95	0.95	0.94	0.92	0.90	0.88	0.85	0.79
8	0.96	0.93	0.94	0.93	0.93	0.92	0.91	0.90	0.88	0.86	0.82
9	0.89	0.89	0.91	0.90	0.91	0.91	0.90	0.89	0.88	0.87	0.85
10	0.81	0.83	0.86	0.86	0.88	0.89	0.88	0.89	0.88	0.88	0.89
11	0.68	0.74	0.79	0.80	0.84	0.85	0.86	0.87	0.88	0.89	0.95

*1=most negative, 11=most positive

the pace and magnitude of recovery, especially for those in poor past health.

6.2 Fixed effects

These non-linear persistence estimates demonstrate the crucial importance of allowing for heterogeneity in health shock features and health history when estimating persistence. Time-invariant, individual fixed effects are an additional important source of heterogeneity, and not accounting for them may bias the persistence estimates upwards. The literature also emphasises the importance of individual heterogeneity, such as initial conditions from childhood, education, or generic variation, as potentially more important than state dependence in determining health outcomes for most people (Halliday, 2008). I re-estimate persistence allowing for fixed effects by using an extension to the Arellano, Blundell and Bonhomme (2017) framework included in their supplementary appendix. I find that accounting for fixed effects does reduce the persistence estimates, and the magnitude of the reduction varies by past health and shock magnitude. The reductions are largest for those in prior poor health who experience a large negative health shock. Therefore, the extremely high persistence previously observed for this group partially reflects fixed effects, although the new persistence estimates remain high. Accounting for fixed effects also removes the asymmetry between positive and negative shock persistence.

To capture time-invariant fixed effects, the persistent component η_{it} is now defined as being equal to $Q_t(\eta_{i,t-1}, \zeta_i, \upsilon_{it})$ where ζ_i is the fixed effect. I report the new persistence estimates in Figure 12 and Table 12. I report the estimates for the persistent component rather than overall health as estimates of this component are most likely to be overstated by not accounting for fixed effects.

Figure 12: Persistent component of health, accounting for fixed effects

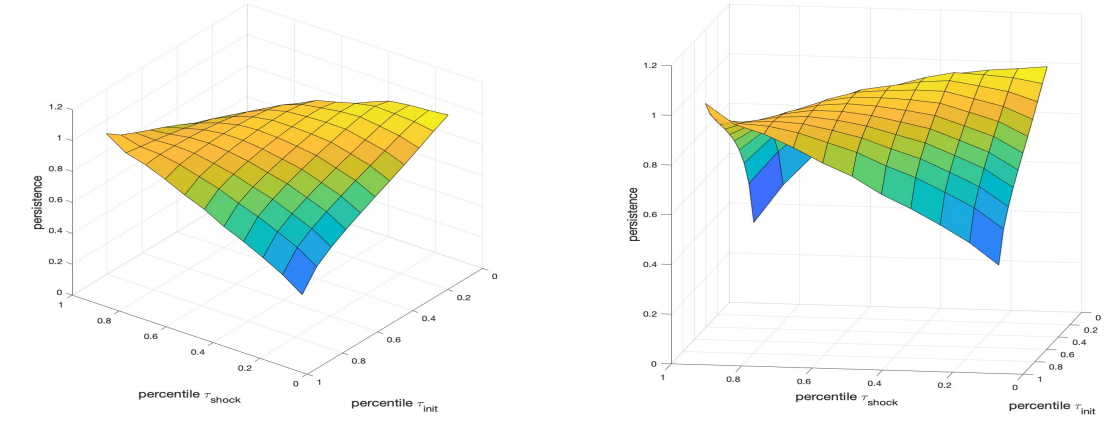


Table 12: Persistent component of health, accounting for fixed effects

	Shock size deciles*										
	1	2	3	4	5	6	7	8	9	10	11
$health_{t-1}$*											
1	1.01	0.99	0.99	0.97	0.92	0.86	0.81	0.72	0.62	0.50	0.34
2	0.96	0.98	0.99	0.99	0.97	0.93	0.90	0.84	0.75	0.66	0.52
3	0.91	0.94	0.96	0.98	0.97	0.95	0.93	0.89	0.82	0.74	0.63
4	0.86	0.90	0.93	0.96	0.96	0.95	0.94	0.91	0.86	0.79	0.70
5	0.81	0.86	0.90	0.93	0.94	0.95	0.95	0.92	0.88	0.83	0.76
6	0.76	0.82	0.86	0.90	0.92	0.93	0.94	0.93	0.90	0.86	0.80
7	0.71	0.78	0.83	0.87	0.89	0.92	0.93	0.93	0.91	0.88	0.84
8	0.66	0.74	0.79	0.83	0.86	0.89	0.91	0.93	0.92	0.90	0.88
9	0.61	0.69	0.74	0.79	0.83	0.87	0.89	0.92	0.92	0.92	0.92
10	0.54	0.62	0.68	0.74	0.78	0.83	0.87	0.90	0.92	0.93	0.97
11	0.41	0.50	0.57	0.63	0.68	0.75	0.80	0.86	0.91	0.95	1.03

*1=most negative, 11=most positive.

Several of the key results from the original non-linear persistence estimates are unaffected by accounting for fixed effects. The range of the persistence estimates, depending on past health and characteristics of the shock at period t remain large, ranging from 0.3 to 1.0 depending on the features of the shock at period t and past health. Persistence estimates in cases of negative health shocks continue to be much higher for individuals in poor health, and the opposite is true for positive health shocks.

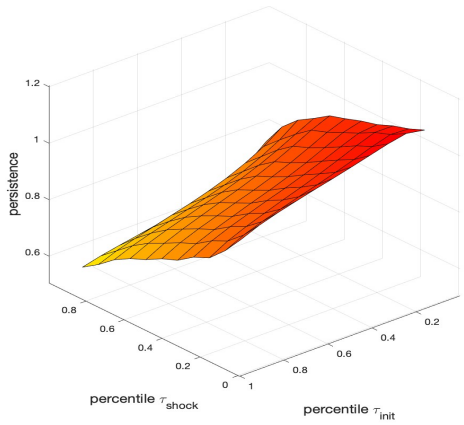
On average, the persistence estimates are smaller when fixed effects are taken into account. This is mostly driven by reductions to the persistence estimates in cases of

negative shocks at period t. The largest reductions are observed for the largest decile of negative shocks, where persistence estimates fall by 0.3-0.4. Accounting for fixed effects has little impact on the persistence estimates when there are positive shocks. As a result, while the original non-linear estimates were much higher for negative shocks than positive shocks, this difference disappears when we take fixed effects into account. Accounting for fixed effects also reduces the persistence estimates for those in very poor health in $t-1$ who experience a positive shock in period t. These results indicate that those who experience large negative shocks, or have a history of poor health, are also more likely to have some unobserved time-invariant trait that subtracts from overall health, such as poor underlying health, and this partially explains the persistently very poor health we observe after large negative shocks and among those with poor health in the prior period. This result is not symmetrical for those who experience positive shocks.

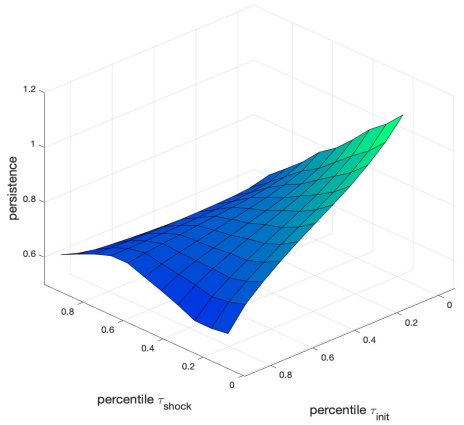
While these fixed effects I account for cannot be observed directly, I do find some evidence that they are related to allostatic scores. Allostatic scores attempt to measure underlying health, which may be associated with vulnerability to suffer negative health shocks, and propensity and speed of recovery from negative health shocks. I divide my sample into two groups based on whether allostatic scores are above or below the sample median allostatic score, and then re-estimate persistence for these two groups (Figure 13).

Figure 13: Non-linear persistence estimates; by allostatic score

(a) High allostatic scores (poor underlying health)



(b) Low allostatic scores (good underlying health)



I observe significant differences in the persistence estimates between those with better and worse allostatic scores. Those with worse underlying health experience more-persistent negative health shocks and less-persistent positive health shocks. The biggest difference between the two groups is that the persistence estimates for those who experience a large negative health shock but are in good health prior to the shock are about 0.5 lower for the group with above average (good) allostatic scores.

There are two possible reasons why the non-linear persistence estimates vary by allostatic score. Allostatic scores may be correlated with the persistence of shocks that people experience; for example those with worse underlying health may be more vulnerable to highly persistent conditions such as many chronic health conditions, or it may simply reflect a correlation between allostatic scores and fixed effects. I find some evidence that the differences mostly reflect fixed effects by repeating the non-linear persistence estimates illustrated in Figure 13, but using the estimation procedure that takes fixed effects into account. The difference between the two groups becomes much smaller when fixed effects are taken into account, suggesting that a majority of the difference was due to fixed effects. Table 13 reports the difference between the fixed-effect-adjusted non-linear persistence estimates by subtracting the estimates for the group with above average (good) allostatic scores from the group with below average (bad) allostatic scores. The difference between the two groups is now never more than 0.2.

This result strongly suggests that allostatic scores capture some aspect of fixed effects that are helpful to model when modelling health dynamics. There is significant scope for further research on modelling this fixed effect and linking it to, for example, education, early childhood experiences and genetics, to better understanding cross-sectional variation in health.

6.3 Application to other health indicators

This paper has focussed on estimating overall health as a dynamic process. However, the methods used in this paper can be easily applied to other health indices used in the economics literature. As an illustrative example, I calculate the persistence of GHQ scores, which are a subcomponent of my health index and can be considered a measure of overall mental health. GHQ scores range from 1-36 based on a

Table 13: Difference between persistent component estimates of poor and good allostatic health groups, accounting for fixed effects

	Shock size deciles*										
	1	2	3	4	5	6	7	8	9	10	11
$health_{t-1}$*											
1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	-0.1
2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1
3	0.1	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	-0.1	-0.1
4	0.1	0.0	-0.1	0.0	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
5	0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
6	0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
7	0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.1
8	0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.2	-0.1
9	0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.2	-0.1
10	0.1	-0.1	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.2	-0.2	-0.1
11	0.2	-0.1	-0.1	-0.1	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2	-0.1

*1=most negative, 11=most positive.

questionnaire designed to identify non-psychotic and minor psychiatric disorders, so it does not accurately capture conditions such as schizophrenia. I de-trend the raw GHQ scores from age, gender and time trends and estimate an ARMA(p,q) model as a linear baseline. I find that GHQ scores can be represented as an ARMA(1,1) model in a similar manner to an overall health index, although the level of persistence is much lower.

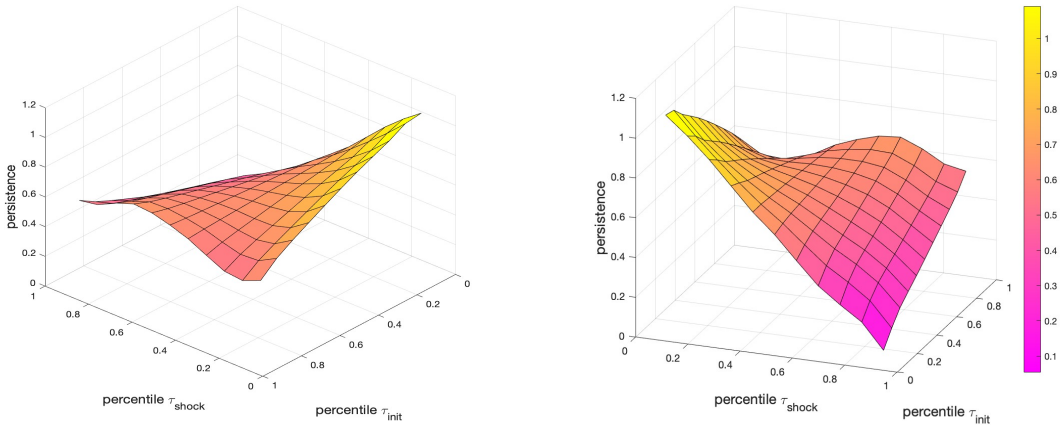
Table 14: ARMA(1,1) model of GHQ scores

	Diff-GMM	Sys-GMM
L.ghq	0.610***	0.693***
	(0.0508)	(0.0388)
AB test, order 1, z score	-17.27	-22.72
AB test, order 1, p value	0.000	0.000
AB test, order 2, z score	8.960	11.42
AB test, order 2, p value	0.000	0.000
AB test, order 3, z score	0.158	0.262
AB test, order 3, p value	0.874	0.793
Hansen J test stat	2.223	7.41
Hansen J test p value	0.527	0.115
Observations	222095	222095

Standard errors in parentheses
Incorporating additional lagged terms found to be insignificant
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Like overall health, the persistence of mental health shocks varies hugely depending

Figure 14: Persistence of GHQ index



on past mental health and the magnitude and sign of the shock. The range of persistence estimates is very large, ranging from 0.1 to 1.1, and persistence is higher when there are negative shocks relative to positive ones. Negative mental health shocks that occur among individuals with prior poor mental health have a persistence of around 1 so should be considered permanent. An interesting difference compared to the overall health index persistence estimates, is that for those with good prior mental health, persistence estimates do not vary much by shock size or sign regardless of size or sign, at around 0.5-0.6. For those with bad prior mental health, persistence estimates range from over 1 for very large negative shocks to 0.1 for very large positive shocks.

Mental health indicators are increasingly being used in the economics literature, such as Jolivet and Postel-Vinay (2020)'s investigation of the relationship between mental health and labour market outcomes. My estimates suggest that it is important to account for past mental health when incorporating mental health indices into economic models, rather than just taking the linear estimates of 0.7.

7 Conclusion

This paper investigated how best to model health as a dynamic process. It evaluated the strengths and weaknesses of the most commonly used approaches in the literature, and adapted recent techniques from the earnings dynamics literature to better capture some of the complexities around modelling shock persistence, frequency and magnitude. The paper also explored how biomarker data can be used to improve our ability to model health dynamics, although further research on this topic is recommended as increasing availability of medical data offers researchers the opportunity to model health in much more sophisticated ways.

Worsening health is an important risk that individuals face, and understanding the many ways it affects economic decision making remains an open question in the literature. An improved method of capturing health as a statistical process helps us better capture the complexities of decision making. This is especially the case as individuals typically know much more about their health than the statistician. The persistence and distribution of health shocks affects both ex-ante choices (how people prepare in advance for a potential health shock) and ex-post choices (how people respond to a realised health shock). For example, an additional channel by which past health history could impact saving behaviour is that individuals may revise their priors on their future health and modify their saving behaviour accordingly. An interesting avenue for future research is to better understand to what degree people modify their expectations of the frequency and persistence of future health shocks following a period of poor health.

8 Appendix

8.1 Attrition

I estimate a linear probability model of attrition to identify the relationship between health and attrition risk.

Table 15: Linear probability model of attrition

	Missing next period
health index quantile 1 (lowest)	0.0213***
	(8.14)
health index quantile 2	-0.00200
	(-0.80)
health index quantile 3 (baseline)	0
health index quantile 4	-0.00131
	(-0.52)
health index quantile 5 (highest)	0.00166
	(0.62)
age	-0.0205***
	(-5.88)
age squared	0.0577***
	(5.08)
age cubed	-0.0861***
	(-5.61)
age quartic	0.0498***
	(6.75)
sex	-0.00706***
	(-4.34)
Observations	228886

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

While I do find a relationship, the literature is fairly sanguine about the risks of using health indices for economic research when there is differential attrition risk by healthJones, Koolman and Rice (2006) find that response rates to the British Household Panel Survey (BHPS) vary by health, with elderly or low-income individuals who

start the survey in poor health particularly likely to attrit, but find that attempting to account for this using inverse probability weights is unnecessary for most research applications. Similarly, Pudney and Watson (2013) investigate the impact of reducing the effort made to chase up non-responders of BHPS and HILDA (an Australian panel dataset). While they find that effort exerted to chase up non-responders changes the sample prevalence of disability and ill health, their subsequent statistical modelling of the relationship between health and unemployment is unaffected. I conclude that the observed level of attrition in my dataset does not pose a significant threat to the robustness of my analysis.

8.2 Regression output from health index construction

I estimate the health index separately for each data wave using an ordered probit. Below I report the output for the the second last wave of data.

Table 16: Estimation of health index - wave 11

	subjective health	t-stat
mobility - some difficulty	-0.701***	(-18.97)
mobility - significant difficulty	-1.069***	(-13.40)
lifting, carrying, moving objects - some difficulty	-0.405***	(-10.77)
lifting, carrying, moving objects - significant difficulty	-0.489***	(-7.65)
manual dexterity - some difficulty	-0.173***	(-3.05)
manual dexterity - significant difficulty	-0.401***	(-3.90)
continence - some difficulty	-0.241***	(-4.81)
continence - significant difficulty	-0.339***	(-3.33)
hearing - some difficulty	-0.0565	(-0.76)
hearing - significant difficulty	-0.0403	(-0.34)
sight - some difficulty	-0.101	(-1.44)
sight - significant difficulty	-0.134	(-1.18)
communication, speech problems - some difficulty	-0.140	(-1.23)
communication, speech problems - significant difficulty	-0.455*	(-2.20)
memory, ability to concentrate, learn, understand - some difficulty	-0.333***	(-6.02)
memory, ability to concentrate, learn, understand - significant difficulty	-0.398***	(-4.19)
recognise danger - some difficulty	0.116	(0.63)
recognise danger - significant difficulty	0.543*	(2.25)
physical coordination - some difficulty	-0.0759	(-1.29)
physical coordination - significant difficulty	0.111	(0.98)
personal care - some difficulty	-0.144	(-1.91)
personal care - significant difficulty	-0.321**	(-2.87)
other - some difficulty	-0.581***	(-14.95)
other - significant difficulty	-0.745***	(-10.50)
age 20-24	-0.244***	(-4.32)
age 25-29	-0.415***	(-7.33)
age 30-34	-0.452***	(-7.73)
age 35-39	-0.531***	(-9.61)
age 40-44	-0.649***	(-12.00)
age 45-49	-0.653***	(-12.49)
age 50-54	-0.732***	(-14.15)
age 55-59	-0.692***	(-13.66)
age 60-64	-0.764***	(-14.60)
age 65-69	-0.625***	(-11.88)
age 70 and older	-0.616***	(-12.39)

female	0.117***	(6.43)
asthma - ever had	0.0609	(0.29)
arthritis - ever had	0.0407	(0.26)
congestive heart failure - ever had	-0.285	(-0.58)
coronary heart disease - ever had	0.855**	(2.76)
angina - ever had	-0.175	(-0.45)
heart attack - ever had	-0.358	(-1.93)
angina - ever had	-0.0386	(-0.27)
emphysema - ever had	-0.480	(-1.15)
hypothyroidism - ever had	-0.0582	(-0.14)
chronic bronchitis - ever had	-0.532	(-1.15)
chronic liver condition - ever had	0.405	(1.33)
cancer - ever had	0.324*	(2.17)
diabetes - ever had	-0.277	(-1.25)
epilepsy - ever had	0.0659	(0.16)
high blood pressure - ever had	-0.0356	(-0.32)
other chronic condition - ever had	-0.209***	(-3.78)
multiple sclerosis - ever had	0.148	(0.51)
COPD - ever had	0.158	(0.43)
emotional, nervous, psychiatric problem - ever had	-0.252	(-1.08)
other cancer - ever had	-0.590**	(-2.90)
anxiety - ever had	-0.0187	(-0.07)
depression - ever had	-0.267	(-1.41)
asthma - still have	-0.194	(-0.89)
arthritis - still have	-0.128	(-0.78)
congestive heart failure - still have	-0.327	(-0.60)
coronary heart disease - still have	-1.019**	(-3.00)
angina - still have	-0.310	(-0.73)
hypothyroidism or underactive thyroid - still have	-0.0505	(-0.12)
chronic bronchitis - still have	0.144	(0.27)
liver condition - still have	-0.606	(-1.75)
cancer - still have	-0.884***	(-4.44)
diabetes - still have	-0.115	(-0.50)
epilepsy - still have	-0.105	(-0.20)
high blood pressure - still have	-0.134	(-1.14)
COPD - still have	-0.488	(-1.28)
emotional or nervous or psychiatric condition - still have	0.521	(1.91)
anxiety - still have	-0.0641	(-0.24)
depression - still have	0.000729	(0.00)
1-2 visits to hospital outpatient in yr	-0.132***	(-6.50)
3-5 visits to hospital outpatient in yr	-0.384***	(-12.12)
6-10 visits to hospital outpatient in yr	-0.476***	(-9.93)
>10 visits to hospital outpatient in yr	-0.652***	(-9.31)
No job dummy	-0.269***	(-9.84)
professional occupation	0.0977*	(2.02)
skilled non-manual occupation	-0.154***	(-5.00)
skilled manual occupation	-0.1000**	(-2.95)
partly skilled occupation	-0.188***	(-5.45)
unskilled occupation	-0.181**	(-2.58)
GHQ score - 1	0.192	(0.66)
GHQ score - 2	-0.309	(-1.30)
GHQ score - 3	-0.261	(-1.19)
GHQ score - 4	-0.424*	(-2.12)
GHQ score - 5	-0.476**	(-2.60)
GHQ score - 6	-0.531**	(-3.04)
GHQ score - 7	-0.639***	(-3.65)
GHQ score - 8	-0.720***	(-4.11)
GHQ score - 9	-0.851***	(-4.85)
GHQ score - 10	-0.962***	(-5.49)
GHQ score - 11	-1.051***	(-6.01)
GHQ score - 12	-1.163***	(-6.66)
GHQ score - 13	-1.226***	(-6.93)
GHQ score - 14	-1.221***	(-6.82)
GHQ score - 15	-1.268***	(-7.04)
GHQ score - 16	-1.354***	(-7.48)
GHQ score - 17	-1.292***	(-7.10)

GHQ score - 18	-1.433***	(-7.76)
GHQ score - 19	-1.389***	(-7.48)
GHQ score - 20	-1.402***	(-7.47)
GHQ score - 21	-1.447***	(-7.61)
GHQ score - 22	-1.634***	(-8.59)
GHQ score - 23	-1.698***	(-8.95)
GHQ score - 24	-1.497***	(-7.89)
GHQ score - 25	-1.788***	(-8.56)
GHQ score - 26	-1.864***	(-8.35)
GHQ score - 27	-1.823***	(-6.19)
GHQ score - 28	-2.013***	(-7.98)
GHQ score - 29	-1.703***	(-6.75)
GHQ score - 30	-2.027***	(-7.47)
GHQ score - 31	-1.980***	(-7.51)
GHQ score - 32	-1.922***	(-6.90)
GHQ score - 33	-2.272***	(-6.62)
GHQ score - 34	-2.033***	(-8.03)
GHQ score - 35	-1.946***	(-4.30)
GHQ score - 36	-1.527***	(-3.66)
Education - GSCEs	0.0511*	(1.96)
Education - Yr12	0.125***	(3.66)
Education - Degree	0.275***	(10.81)
pregnant	0.174	(1.70)
N	24987	

t stats in parentheses, time dummies, *p<0.05, **p<0.01, *** p<0.001

8.3 Genetic data

PGS are scores that estimate an individuals propensity to a phenotype, which is an observable trait. They are calculated from genome-wide association studies (GWAS), which are systematic analyses of genetic variation across the entire human genomes and their association with various phenotypes. I make use of the latest (2022) version of the ELSA polygenic scores (Ajnakin and Andrew Steptoe, 2022) and select a subset of these to cover the health conditions with the biggest disease burdens, which I report in Table 17. I use these to create two health aggregates; one capturing chronic physical health conditions, the other mental health. The correlation between the two indices is small.

My method of aggregation is identical to how I aggregate the biomarker data; I normalise each polygenic score (PGS) and then add them together for each individual. I do consider alternate aggregation methods, including factor analysis and converting each PGS into a binary variable with the highest 10-20 per cent of scores coded as '1' for higher likelihood of developing condition, however the resulting indexes are all quite similar. I report in Table 18 a regression of a health index constructed using ELSA data to be comparable to my health index constructed using Understanding Society data, against the normalised polygenic scores for each genetic index. The

Table 17: Polygenic Score Aggregation

Physical index	Mental index
Coronary artery disease (2016)	Alzheimer's disease (2019)
Type II diabetes (2018)	Depressive Symptoms
Rheumatoid arthritis	Major depressive disorder (2018)
Myocardial infarction	Anxiety (case-control)
Migrane (2016)	Schizophrenia (2020)
Chronic pain	Bipolar disorders (2021)
Waist-hip-ratio	Subjective wellbeing
	Loneliness

major depressive disorder PGS has the highest predictive power although the predictive power of many of the eight mental health PGS are quite similar. The predictive power of the physical PGS are much more varied, with chronic pain being by far the most important.

Table 18: How polygenic score components predict health index

	mental index	physical index
z score: depressive symptoms	-0.0302***	
	(0.00486)	
z score: major depressive disorder	-0.0521***	
	(0.00515)	
z score: anxiety	0.00763	
	(0.00490)	
z score: schizophrenia	0.0192***	
	(0.00575)	
z score: bipolar	-0.00397	
	(0.00468)	
z score: subjective wellbeing	0.0119**	
	(0.00433)	
z score: alzheimers	-0.0170***	
	(0.00394)	
z score: loneliness	-0.0340***	
	(0.00490)	
z score: arthritis		0.00911*
		(0.00426)
z score: coronary heart disease		0.00193
		(0.00401)
z score: diabetes		-0.0342***
		(0.00409)
z score: chronic pain		-0.112***
		(0.00416)
z score: myocardial infarction		-0.0192***
		(0.00408)
z score: waist-hip ratio		-0.0121**
		(0.00394)
z score: migraines		-0.00235
		(0.00378)
N	37543	37546

Finally, I report in Table 19 how well the two genetic indices predict the level of the health index as well as its variance over time for each individual.

Table 19: Predictive content of polygenic data

	level of health index			variance of health index		
	(1)	(2)	(3)	(4)	(5)	(6)
mental index	-0.00746***	-0.00125*	-0.00768***	0.000810***	0.000392	0.000654*
	(-8.42)	(-2.10)	(-6.31)	(4.48)	(1.93)	(2.52)
physical index	-0.0229***	-0.00476***	-0.0187***	0.00165***	0.000809**	0.00136***
	(-18.28)	(-5.42)	(-10.71)	(6.24)	(2.63)	(3.56)
mental index ^{2}	-0.000496***	-0.000187*	-0.000463**	0.0000500*	0.0000226	0.0000302
	(-4.73)	(-2.45)	(-3.19)	(2.37)	(0.93)	(1.01)
physical index ^{2}	-0.000519*	0.0000618	-0.000417	0.0000270	0.0000366	0.00000785
	(-2.05)	(0.34)	(-1.18)	(0.49)	(0.57)	(0.10)
lagged health index		0.840***			-0.0485***	
		(201.75)			(-29.36)	
allostatic index			-0.385***			0.0330***
			(-18.59)			(10.38)
Observations	37543	25256	18304	37412	25256	18203

t statistics in parentheses; also control for age and gender

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

While the coefficient for both indices are significant in predicting the level of health and its variance over time, and contains additional information not captured by lagged health information or biomarkers, the size of the coefficients are very small. I conclude that the coefficient sizes are too small to be a useful tool to model the overall dynamics of the health index. However, a more granular approach may be more effective. This could include using only individual PGS with higher predictive power or for diseases with high incidence rates such as diabetes and depression rather than rarer conditions such as schizophrenia. Using PGS to model only a specific component of health is also likely to be more effective in better capturing health shocks and individual heterogeneity in health outcomes.

8.4 Replication exercise using non-detrended health data

I replicate the preferred baseline ARMA(1,1) model estimated using Difference-GMM using health data that has not been detrended by age and sex. This exercise shows that the persistence estimates are robust to the detrending; the slightly higher estimates are due to a gradual decline in health as people age being incorporated into the persistence estimates.

Table 20: Estimation of ARMA(1,1) model with non-detrended health index

	(1)
	health_index
L.health_index	1.038***
	(0.0227)
Hansen J test stat	2.756
Hansen J p value	0.431
Observations	222095

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

8.5 Additional information on groups from clustering analysis

The partition clustering algorithm makes an initial partition of individuals into clusters based on the number of desired clusters, and then reallocates individuals until the final partition minimises the residual sun of squares objective function. The optimal number of clusters for a given model is selected as the one with the lowest model information criterion (MIC).

I describe the main characteristics of the two clusters in table 21. Relative to group 2, group 1 contains individuals with, on average, worse health. The individuals in the two groups are just as likely to experience a large negative health shock of at least one standard deviation (6.6 and 6.7 per cent of observations respectively), but the second group is more likely to experience a positive shock (5.3 versus 6.4 per cent).

I estimate some OLS and within group models for the two groups and find significant differences.

Table 21: Group characteristics

	Group 1	Group 2
mean	0.024	0.134
median	0.186	0.254
st. dev	0.670	0.565
mean allostatic	0.023	-0.046
N	28644	44231

* Only include individuals where T=12

Table 22: Regressions by group

	OLS		WG	
	Group 1	Group 2	Group 1	Group 2
L.demean_health	0.775***	0.483***	0.575***	0.00817
	(0.00623)	(0.00464)	(0.00668)	(0.00543)
L2.demean_health	0.125***	0.401***	-0.0190**	-0.0381***
	(0.00630)	(0.00469)	(0.00691)	(0.00538)
Observations	26040	40210	26040	40210

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Since group 2 is best modelled using an ARMA(1,1) model, the below table reports the results of estimating the MA coefficient for this model.

Table 23: Group 2 MA estimation

	(1)
Constant	0.0851***
	(0.00280)
Constant	-0.544***
	(0.0249)
sig_e	
Constant	0.289***
	(0.00411)
Observations	44231

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

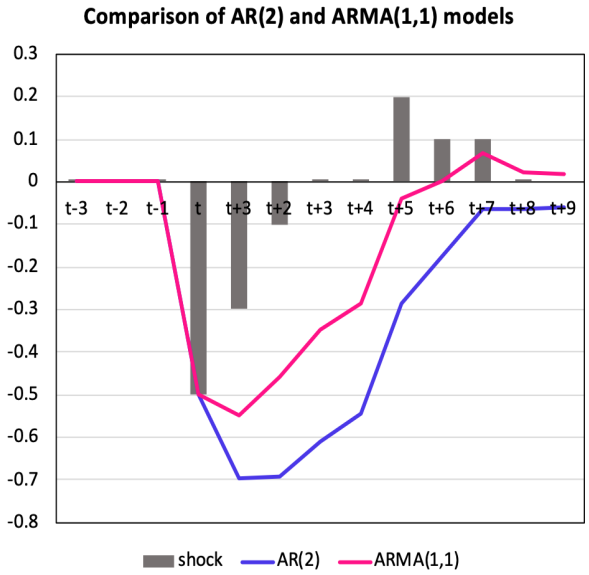
Table 24: GMM estimates of the health process, by cluster

	Difference GMM								System GMM	
	group 1				group 2				group 1	group 2
	MA(0)		MA(1)		MA(0)		MA(1)		MA(0)	MA(1)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
L.health	0.709***	0.791***	0.898***	0.922***	0.104***	0.187***	0.997***	0.893***	0.742***	0.828***
	(0.0203)	(0.0218)	(0.0311)	(0.100)	(0.0119)	(0.0172)	(0.132)	(0.164)	(0.00819)	(0.0461)
L2.health		0.0876***		-0.0207		0.0737***		-0.00265	0.0807***	
		(0.0124)		(0.0822)		(0.0127)		(0.0236)	(0.0106)	
AB test, order 1 z score	-26.87	-27.77	-21.04	-6.27	-35.4	-31.92	-9.46	-6.99	-33.00	-19.15
AB test, order 1 p value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AB test, order 2 z score	3.44	-1.21	3.59	1.27	3.36	-1.46	6.02	4.2	-1.11	9.39
AB test, order 2 p value	0.001	0.226	0.000	0.203	0.001	0.145	0.000	0.000	0.266	0.000
Hansen J test statistic	63.74	3.39	1.47	1.37	78.77	43.01	1.71	7.19	12.74	4.09
Hansen J p value	0.000	0.495	0.689	0.712	0.000	0.000	0.634	0.066	0.047	0.394
Observations	28644	26040	28644	26040	44231	40210	44231	40210	26040	44231

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The ARMA(1,1) tends to be associated with a higher degree of mean reversion, as illustrated by Figure 15, which shows a hypothetical multi-period negative shock that gradually dissipates.

Figure 15



8.6 ARCH model estimates

An autoregressive conditional heteroskedasticity (ARCH) model accounts for whether individuals who have recently experienced a health shock, captured in the model as a large ε_{it} term, are more or less likely to experience additional health shocks in the subsequent period. I estimate an ARCH(1) effects model with the following exponential variance specification:$\sigma_{it}^2 = \exp(\gamma_0 + \gamma_1 \varepsilon_{i,t-1}^2 + \gamma_2 \varepsilon_{i,t-1})$. To estimate ARCH effects using GMM I use the following moment condition derived by Arellano (1995). For robustness I also estimate the more general specification by Meghir and Windmeijer (1999) and obtain similar results.

$$\mathbb{E}\left[h_{i,t-k}\left(\varepsilon_{i,t-1}^2 - \frac{\varepsilon_{i,t}^2(1+\sigma_{it-1}^2)}{(1+\sigma_{i,t}^2)}\right)\right] = 0, \text{ k} = 1...\text{t-3}$$

An ARCH effect exists if the γ_1 and γ_2 terms on the lagged error terms in the variance specification $\sigma_{it}^2 = \exp(\gamma_0 + \gamma_1 \varepsilon_{i,t-1}^2 + \gamma_2 \varepsilon_{i,t-1})$ are significantly different from zero. While the point estimates for the γ terms are reasonable, and suggest that a

person who experiences a large health shock has more volatile health next period, and that this effect is significantly reduced if the health shock is positive, the estimates are not statistically significant. I conclude that there is no evidence for ARCH effects in the data.

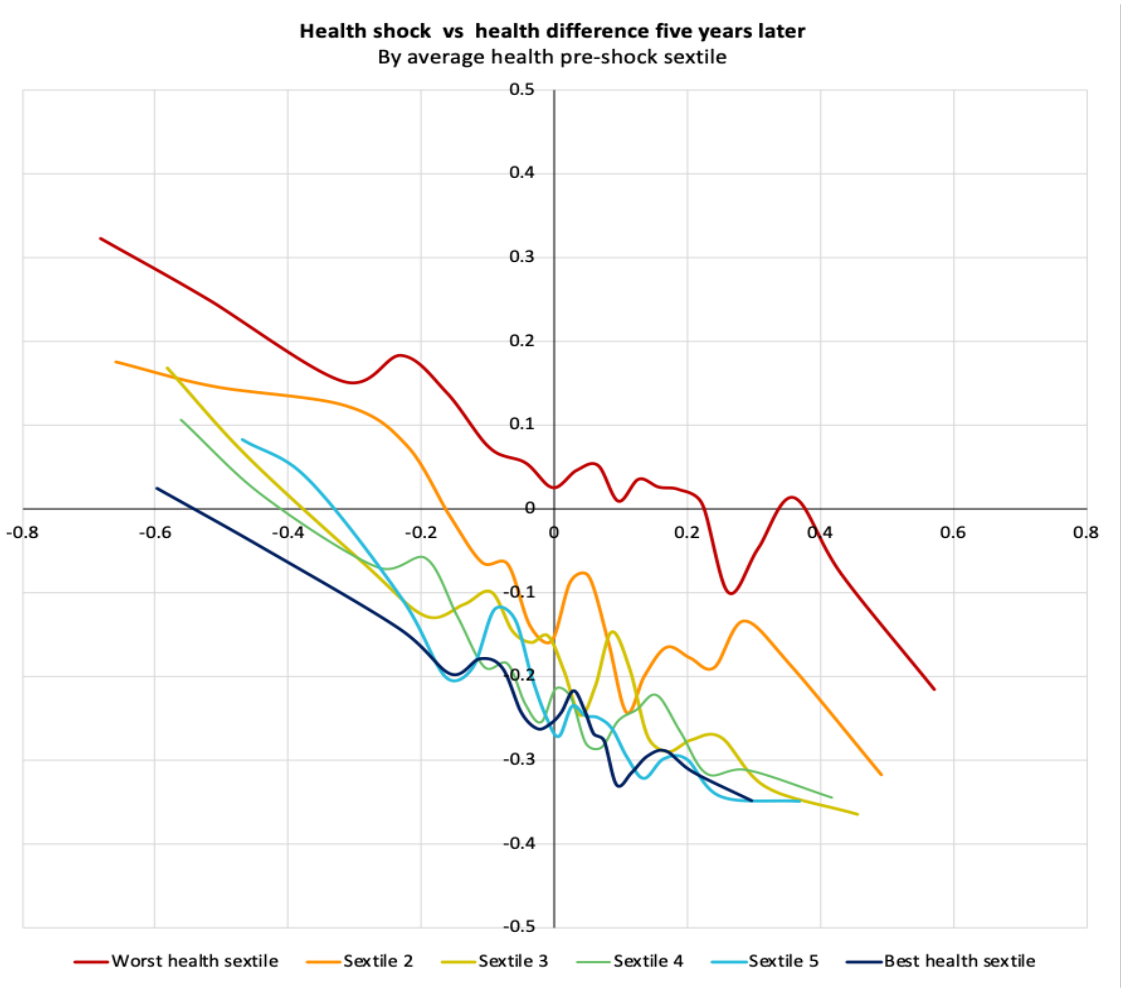
Table 25: GMM estimates of ARCH effect: $\sigma_{it}^2 = \exp(\gamma_0 + \gamma_1 \varepsilon_{i,t-1}^2 + \gamma_2 \varepsilon_{i,t-1})$

Coefficients	Estimates
γ_0	-2.5176
	(5.7445)
γ_1	0.3301
	(0.5848)
γ_2	-0.1648
	(0.8703)

standard errors in parentheses

In addition, I consider shock heterogeneity in a different way. Figure 16 graphs the relationship between the magnitude of health shock in period t and extent of recovery/mean-reversion five years later. I plot the initial health shock ($h_{i,t} - h_{i,t-1}$ on the x axis and the difference between health at period t and five years later ($h_{i,t+5} - health_{i,t}$) on the y axis. A steeper downward-sloping line suggests a faster rate of mean-reversion, while a flat line would indicate that there has been no change in health between period t and period $t + 5$. I graph this relationship for six sub-samples based on average health in the period prior to the shock ($t+1$). By contrast, the baseline ARMA(1,1) model with white noise error process would get around a 60 per cent mean reversion.

Figure 16



*Adapted from an earnings dynamics graph by Guvenen et al. (2021)

Figure 16 shows that there is significantly less recovery from negative shocks than would be predicted by an AR(2) model with normally distributed white-noise shocks, especially among those in persistently poor health prior to the negative shock. The health dynamics of those whose health is in the bottom sixth of the sample are much less well captured by a simple linear AR(2) model relative to those in the upper two-thirds. In reality, those in bad health who experience more negative health shocks remain in poor health, while AR models predict a much higher degree of mean reversion. This exercise suggests that simple AR models are adequate for approximating the health process for the healthier section of the population, but are much less accurate for those with a history of poor health.

8.7 ARMA models with additional allostatic regressors

Adding allostatic scores as an additional regressor, or as a dummy variable for whether individuals have 'bad' allostatic scores interacted with the lagged health term, or estimating the linear models on only a subset of allostatic scores all have very little impact on model estimates. For example, table 26 reports the regression output if I include them in my preferred specification.

Table 26: Preferred model estimates with additional allostatic regressors

	(1) demean_health	(2) demean_health
L.demean_health	0.890***	0.876***
	(0.0286)	(0.0218)
demean_allo	0.203	-0.348
	(0.302)	(0.178)
L.demean_health × demean_allo		-0.131
		(0.0711)
Observations	70039	70039

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

8.8 Additional non-linear persistence estimates results

I report the full results of re-estimating the non-linear persistence model when the sample is split into two sub-groups based on allostatic score (the graphs of these results are reported in Chapter 6).

Table 27: Estimates of coefficient of persistent component $\rho_t(\tau)$ - Low allostatic (healthy)

	Shock size percentiles*										
	1	2	3	4	5	6	7	8	9	10	11
$health_{t-1}$*											
1	1.15	1.09	1.04	0.98	0.92	0.88	0.82	0.76	0.71	0.64	0.55
2	1.08	1.02	0.98	0.93	0.87	0.84	0.79	0.74	0.70	0.63	0.54
3	1.02	0.97	0.93	0.89	0.84	0.81	0.78	0.73	0.69	0.63	0.55
4	0.98	0.92	0.89	0.86	0.82	0.79	0.77	0.72	0.69	0.63	0.55
5	0.93	0.89	0.85	0.83	0.80	0.78	0.76	0.72	0.69	0.63	0.56
6	0.89	0.85	0.82	0.80	0.78	0.77	0.75	0.72	0.68	0.63	0.57
7	0.85	0.81	0.79	0.78	0.76	0.75	0.74	0.72	0.68	0.64	0.58
8	0.81	0.77	0.75	0.75	0.74	0.74	0.74	0.71	0.68	0.64	0.59
9	0.76	0.73	0.72	0.72	0.72	0.73	0.73	0.71	0.69	0.65	0.60
10	0.70	0.68	0.67	0.69	0.70	0.71	0.72	0.71	0.69	0.65	0.61
11	0.62	0.61	0.61	0.64	0.67	0.69	0.71	0.71	0.69	0.66	0.63

*1=most negative, 11=most positive

Table 28: Estimates of coefficient of persistent component $\rho_t(\tau)$ - high allostatic (unhealthy)

	Shock size percentiles*										
	1	2	3	4	5	6	7	8	9	10	11
$health_{t-1}$*											
1	1.09	1.06	1.05	1.02	1.01	0.99	0.97	0.93	0.90	0.85	0.77
2	1.10	1.07	1.05	1.02	0.99	0.96	0.94	0.90	0.85	0.80	0.72
3	1.08	1.05	1.02	0.99	0.96	0.93	0.91	0.87	0.82	0.77	0.69
4	1.07	1.02	1.00	0.96	0.93	0.90	0.88	0.84	0.79	0.74	0.66
5	1.05	1.00	0.97	0.94	0.90	0.88	0.85	0.81	0.77	0.72	0.64
6	1.03	0.98	0.95	0.92	0.88	0.85	0.83	0.79	0.74	0.70	0.63
7	1.02	0.96	0.93	0.90	0.86	0.83	0.81	0.77	0.73	0.68	0.62
8	1.00	0.94	0.90	0.87	0.84	0.81	0.78	0.75	0.71	0.66	0.60
9	0.98	0.92	0.88	0.85	0.81	0.79	0.76	0.73	0.69	0.64	0.59
10	0.96	0.90	0.85	0.82	0.79	0.76	0.73	0.71	0.67	0.62	0.58
11	0.93	0.87	0.81	0.79	0.75	0.73	0.70	0.67	0.64	0.60	0.56

*1=most negative, 11=most positive

Table 29: Persistence estimates - persistent component only

	Shock size percentiles*										
	1	2	3	4	5	6	7	8	9	10	11
$health_{t-1}$*											
	Good allo										
1	1.48	1.35	1.25	1.15	1.05	0.97	0.91	0.86	0.80	0.68	0.39
2	1.35	1.21	1.13	1.04	0.96	0.90	0.85	0.81	0.76	0.66	0.46
3	1.24	1.11	1.02	0.96	0.89	0.85	0.80	0.76	0.72	0.64	0.50
4	1.15	1.02	0.95	0.89	0.84	0.80	0.76	0.73	0.69	0.63	0.53
5	1.08	0.95	0.88	0.84	0.80	0.77	0.73	0.70	0.67	0.61	0.55
6	1.02	0.90	0.83	0.79	0.76	0.74	0.71	0.68	0.65	0.60	0.56
7	0.96	0.85	0.79	0.76	0.73	0.71	0.68	0.66	0.63	0.59	0.58
8	0.91	0.80	0.74	0.72	0.70	0.69	0.66	0.64	0.61	0.58	0.59
9	0.86	0.75	0.70	0.68	0.67	0.66	0.64	0.62	0.60	0.57	0.60
10	0.80	0.69	0.65	0.64	0.64	0.63	0.61	0.59	0.58	0.56	0.61
11	0.70	0.61	0.57	0.57	0.58	0.59	0.57	0.56	0.55	0.54	0.63
	Bad allo										
1	1.35	1.25	1.16	1.03	0.96	0.92	0.89	0.83	0.78	0.68	0.52
2	1.26	1.19	1.11	1.00	0.94	0.91	0.89	0.83	0.79	0.71	0.58
3	1.17	1.13	1.05	0.96	0.91	0.89	0.87	0.83	0.79	0.73	0.61
4	1.08	1.07	1.00	0.93	0.89	0.88	0.86	0.82	0.79	0.73	0.64
5	1.01	1.01	0.96	0.90	0.87	0.86	0.85	0.82	0.79	0.74	0.66
6	0.94	0.96	0.92	0.87	0.85	0.85	0.84	0.81	0.79	0.74	0.67
7	0.88	0.92	0.88	0.85	0.84	0.84	0.83	0.81	0.79	0.75	0.69
8	0.82	0.87	0.84	0.82	0.82	0.83	0.82	0.81	0.79	0.75	0.70
9	0.75	0.82	0.80	0.80	0.80	0.81	0.81	0.80	0.79	0.75	0.72
10	0.67	0.76	0.75	0.76	0.78	0.80	0.80	0.79	0.78	0.75	0.73
11	0.53	0.66	0.67	0.71	0.74	0.77	0.77	0.78	0.78	0.76	0.76

*1=most negative, 11=most positive

Table 29 reports output from the same estimation process, but this time only reporting the persistent component. Fixed effects were not taken into account here.

Table 30 reports some additional results from the non-linear persistence estimates of mental health. The table is shown in panel (a) of the below graphs, while panel (b) shows the impact of taking fixed effects into account.

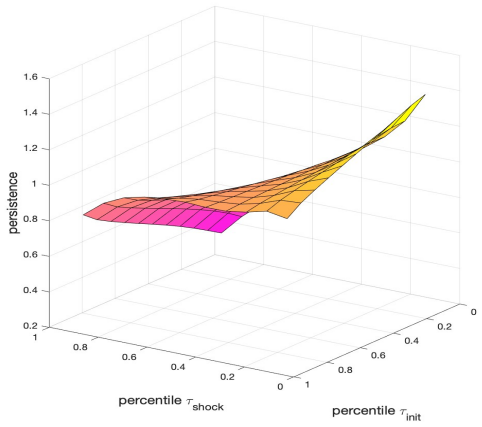
Table 30: Estimates of coefficient of persistent component $\rho_t(\tau)$ - GHQ scores (mental health)

	Shock size percentiles*										
	1	2	3	4	5	6	7	8	9	10	11
$health_{t-1}$*											
1	1.09	1.00	0.90	0.78	0.66	0.56	0.45	0.34	0.26	0.18	0.06
2	1.08	0.96	0.86	0.76	0.66	0.58	0.49	0.40	0.33	0.25	0.13
3	1.02	0.90	0.81	0.73	0.66	0.59	0.52	0.44	0.38	0.30	0.19
4	0.97	0.86	0.78	0.71	0.65	0.60	0.54	0.48	0.41	0.34	0.24
5	0.92	0.82	0.74	0.69	0.65	0.61	0.56	0.50	0.45	0.37	0.28
6	0.87	0.77	0.71	0.67	0.65	0.62	0.58	0.53	0.48	0.40	0.32
7	0.81	0.72	0.67	0.65	0.64	0.62	0.60	0.56	0.51	0.44	0.36
8	0.75	0.67	0.63	0.63	0.63	0.63	0.62	0.59	0.54	0.47	0.41
9	0.69	0.62	0.60	0.61	0.63	0.64	0.64	0.62	0.57	0.51	0.46
10	0.63	0.56	0.55	0.59	0.62	0.65	0.66	0.65	0.61	0.54	0.50
11	0.54	0.48	0.50	0.55	0.61	0.66	0.69	0.70	0.65	0.59	0.57

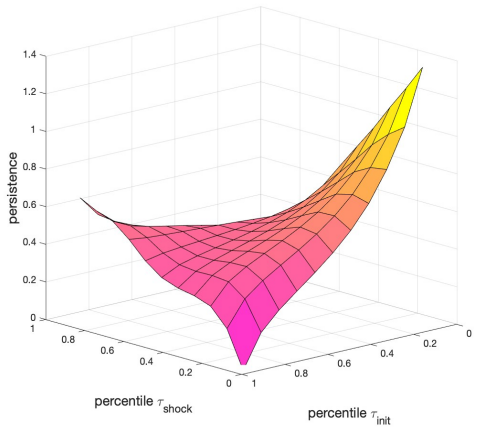
*1=most negative, 11=most positive

Figure 17: Persistence of GHQ index

(a) Persistent component only



(b) Persistent component w FE



References

Abowd, John, and David Card. 1989. "On the Covariance Structure of Earnings and Hours Changes." *Econometrica*, 57(2): 411–45.

Ajnakin, Olesya, and University College London Andrew Steptoe. 2022. "The English Longitudinal Study of Ageing (ELSA) Polygenic Scores." University College London Documentation Report.

Amengual, Dante, Jess Bueren, and Julio A. Crego. 2021. "Endogenous health groups and heterogeneous dynamics of the elderly." *Journal of Applied Econometrics*, 36(7): 878–897.

Anderson, Kathryn H., and Richard V. Burkhauser. 1984. "The importance of the measure of health in empirical estimates of the labor supply of older men." *Economics Letters*, 16(3): 375–380.

Arellano, Manuel. 1995. "Panel data models with predetermined instruments, Paper presented at Review of Economic Studies lecture, Royal Econonomic Society Conference, Kent."

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies*, 58(2): 277–297.

Arellano, Manuel, Richard Blundell, and Stphane Bonhomme. 2017. "Earnings and Consumption Dynamics: A Non-linear Panel Data Framework." *Econometrica*, 85(3): 693–734.

Attanasio, Orazio, Sagiri Kitao, and Giovanni L. Violante. 2010. "Financing Medicare: A General Equilibrium Analysis." *Demography and the Economy*, 333–366. University of Chicago Press.

Barth, Daniel, Nicholas W. Papageorge, and Kevin Thom. 2020. "Genetic Endowments and Wealth Inequality." *Journal of Political Economy*, 128(4): 1474–1522.

Benzeval, Michaela, Apostolos Davillas, Meena Kumari, and Peter Lynn. 2014. "Understanding Society: The UK Household Longitudinal Study." Institute for Social and Economic Research Biomarker User Guide and Glossary.

Blau, David M., and Donna B. Gilleskie. 2001. "Retiree Health Insurance and the Labor Force Behavior of Older Men in the 1990s." *The Review of Economics and Statistics*, 83(1): 64–80.

Blundell, Richard, and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." *Journal of Econometrics*, 87(1): 115–143.

Blundell, Richard, and Steve Bond. 2023. "Initial conditions and BlundellBond estimators." *Journal of Econometrics*, 234: 101–110. Jubilee Issue - Celebrating our Fiftieth Anniversary: 1973–2023.

Blundell, Richard, Jack Britton, Monica Costa Dias, and Eric French. 2016. "The Dynamic Effects of Health on the Employment of Older Workers." University of Michigan Retirement Research Center (MRRC) Working Paper 348.

Blundell, Richard, Jack Britton, Monica Costa Dias, and Eric French. 2017. "The impact of health on labour supply near retirement." Institute for Fiscal Studies IFS Working Papers W17/18.

Blundell, Richard, Jack Britton, Monica Costa Dias, and Eric French. 2021. "The impact of health on labour supply near retirement." *The Journal of Human Resources*, 1217.

Blundell, Richard, Margherita Borella, Jeanne Commault, and Mariacristina De Nardi. 2020. "Why Does Consumption Fluctuate in Old Age and How Should the Government Insure It?" National Bureau of Economic Research Working Paper 27348.

Born, Benjamin, and Jrg Breitung. 2016. "Testing for Serial Correlation in Fixed-Effects Panel Data Models." *Econometric Reviews*, 35(7): 1290–1316.

Bound, John. 1991. "Self-Reported Versus Objective Measures of Health in Retirement Models." *The Journal of Human Resources*, 26(1): 106–138.

Braun, R. Anton, Karen A Kopecky, and Tatyana Koreshkova. 2017. "Old, Sick, Alone, and Poor: A Welfare Analysis of Old-Age Social Insurance Programmes." *The Review of Economic Studies*, 84(2 (299)): 580–612.

Capatina, Elena. 2015. "Life-cycle effects of health risk." *Journal of Monetary Economics*, 74: 67–88.

Caraballo, Csar, Rohan Khera, Philip G. Jones, Carole Decker, Wade Schulz, John A. Spertus, and Harlan M. Krumholz. 2020. "Rates and Predictors of Patient Underreporting of Hospitalizations During Follow-Up After Acute Myocardial Infarction." *Circulation: Cardiovascular Quality and Outcomes*, 13(7): e006231.

Case, Anne, Angela Fertig, and Christina Paxson. 2004. "The Lasting Impact of Childhood Health and Circumstance." Princeton University, Woodrow Wilson School of Public and International Affairs, Center for Health and Wellbeing. Working Papers 246.

Chandola, Tarani, and Nan Zhang. 2017. "Re-employment, job quality, health and allostatic load biomarkers: prospective evidence from the UK Household Longitudinal Study." *International Journal of Epidemiology*, 47(1): 47–57.

Conti, Gabriella, and James J. Heckman. 2010. "Understanding the Early Origins of the EducationHealth Gradient: A Framework That Can Also Be Applied to Analyze GeneEnvironment Interactions." *Perspectives on Psychological Science*, 5(5): 585–605. PMID: 21738556.

Cronqvist, Henrik, and Stephan Siegel. 2015. "The Origins of Savings Behavior." *Journal of Political Economy*, 123(1): 123–169.

Dal Bianco, Chiara, and Andrea Moro. 2022. "The Welfare Effects of Nonlinear Health Dynamics." Center for Open Science SocArXiv dvhby.

Davillas, Apostolos, and Andrew M Jones. 2020. "Ex ante inequality of opportunity in health, decomposition and distributional analysis of biomarkers." *Journal of Health Economics*, 69: 102251.

Davillas, Apostolos, Andrew M. Jones, and Michaela Benzeval. 2019. "Chapter 22 - The Income-Health Gradient: Evidence From Self-Reported Health and Biomarkers in Understanding Society." In *Panel Data Econometrics.* , ed. Mike Tsionas, 709–741. Academic Press.

Davillas, Apostolos, and Stephen Pudney. 2020*a*. "Biomarkers as precursors of disability." *Economics and Human Biology*, 36: 100814.

Davillas, Apostolos, and Stephen Pudney. 2020*b*. "Biomarkers, disability and health care demand." *Economics and Human Biology*, 39: 100929.

Davillas, Apostolos, and Stephen Pudney. 2020*c*. "Using biomarkers to predict healthcare costs: Evidence from a UK household panel." *Journal of Health Economics*, 73: 102356.

De Nardi, Mariacristina, Eric French, and John B. Jones. 2010. "Why Do the Elderly Save? The Role of Medical Expenses." *Journal of Political Economy*, 118(1): 39–75.

De Nardi, Mariacristina, Giulio Fella, and Gonzalo Paz-Pardo. 2019. "Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare." *Journal of the European Economic Association*, 18(2): 890–926.

Diebolt, Jean, and Gilles Celeux. 1993. "Asymptotic properties of a stochastic EM Algorithm for estimating mixing proportions." *Communications in Statistics. Stochastic Models*, 9(4): 599–613.

Fella, Giulio, Julio Galvez, Beatriz Gonzalez, Juan Carlos, Ruiz-Garcia, and Tatsuro Senga. 2021. "Non-Linear Productivity and Investment Dynamics." Queen Mary University of London Working Paper.

French, Eric. 2005. "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour." *The Review of Economic Studies*, 72(2): 395–427.

French, Eric, and John Bailey Jones. 2011. "The Effects of Health Insurance and Self-Insurance on Retirement Behavior." *Econometrica*, 79(3): 693–732.

French, Eric, and John Bailey Jones. 2017. "Health, Health Insurance, and Retirement: A Survey." *Annual Review of Economics*, 9(1): 383–409.

Group, Biomarkers Definitions Working. 2001. "Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework." *Clinical Pharmacology & Therapeutics*, 69(3): 89–95.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2021. "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?" *Econometrica*, 89(5): 2303–2339.

Halliday, Timothy. 2008. "Heterogeneity, state dependence and health." *Econometrics Journal*, 11: 499–516.

Harris, Sarah E, Saskia P Hagenaars, Gail Davies, W David Hill, David CM Liewald, Stuart J Ritchie, Riccardo E Marioni, International Consortium for Blood Pressure Genome-Wide Association Studies METASTROKE Consortium, International Consortium for Blood Pressure Genome-Wide Association Studies, CHARGE Consortium Aging, Longevity Group, CHARGE Consortium Cognitive Group, Cathie LM Sudlow, Joanna M Wardlaw, Andrew M McIntosh, Catharine R Gale, and Ian J Deary. 2016. "Molecular genetic contributions to self-rated health." *International Journal of Epidemiology*, 46(3): 994–1009.

Hauck, Katharina, and Nigel Rice. 2004. "A longitudinal analysis of mental health mobility in Britain." *Health Economics*, 13(10): 981–1001.

Heiss, Florian, Steven F Venti, and David A Wise. 2014. "The Persistence and Heterogeneity of Health among Older Americans." National Bureau of Economic Research Working Paper 20306.

Hosseini, Roozbeh, Karen A. Kopecky, and Kai Zhao. 2022. "The evolution of health over the life cycle." *Review of Economic Dynamics*, 45: 237–263.

Imrohoroglu, Ayse, and Kai Zhao. 2018. "The chinese saving rate: Long-term care risks, family insurance, and demographics." *Journal of Monetary Economics*, 96: 33–52.

Jolivet, Grgory, and Fabien Postel-Vinay. 2020. "A Structural Analysis of Mental Health and Labor Market Trajectories." IZA Institute of Labor Economics Discussion Paper Series DP13518.

Jones, Andrew M., Xander Koolman, and Nigel Rice. 2006. "Health-Related Non-Response in the British Household Panel Survey and European Community Household Panel: Using Inverse-Probability-Weighted Estimators in Non-Linear Models." *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 169(3): 543–569.

Jung, Juergen, and Chung Tran. 2016. "Market inefficiency, insurance mandate and welfare: U.S. health care reform 2010." *Review of Economic Dynamics*, 20: 132–159.

Keane, Michael, Elena Capatina, and Shiko Maruyama. 2020. "Health Shocks and the Evolution of Earnings over the Life-Cycle." School of Economics, The University of New South Wales Discussion Papers 2018-14b.

Lillard, Lee A., and Robert J. Willis. 1978. "Dynamic Aspects of Earning Mobility." *Econometrica*, 46(5): 985–1012.

MaCurdy, Thomas E. 1982. "The use of time series processes to model the error structure of earnings in a longitudinal data analysis." *Journal of Econometrics*, 18(1): 83–114.

Meena Kumari, Tarek Al Baghal, and Michaela Benzeval. 2022. "A framework for the re-collection of biomarkers in Understanding Society at Wave 16." Understanding Society, University of Essex 2022-10.

Meghir, Costas, and Frank Windmeijer. 1999. "Moment Conditions for Dynamic Panel Data Models with Multiplicative Individual Effects in the Conditional Variance." *Annales d'conomie et de Statistique*, 55/56: 317–330.

Meghir, Costas, and Luigi Pistaferri. 2004. "Income Variance Dynamics and Heterogeneity." *Econometrica*, 72(1): 1–32.

Nelson, Daniel B. 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica*, 59(2): 347–370.

Niedzwiedz, Claire L, Srinivasa Vittal Katikireddi, Aaron Reeves, Martin McKee, and David Stuckler. 2017. "Economic insecurity during the Great Recession and metabolic, inflammatory and liver function biomarkers: analysis of the UK Household Longitudinal Study." *Journal of Epidemiology & Community Health*.

Nygaard, Vegard M. 2021. "Causes and consequences of life expectancy inequality." *Manuscript, University of Houston*.

Ozkan, Serdar. 2017. "Preventive vs. curative medicine: A macroeconomic analysis of health care over the life cycle." *Manuscript, University of Toronto*.

ODonnell, Owen, Eddy Van Doorslaer, and Tom Van Ourti. 2015. "Chapter 17 - Health and Inequality." In *Handbook of Income Distribution*. Vol. 2 of *Handbook of Income Distribution*, , ed. Anthony B. Atkinson and Franois Bourguignon, 1419–1533. Elsevier.

Palumbo, Michael G. 1999. "Uncertain medical expenses and precautionary saving near the end of the life cycle." *The Review of Economic Studies*, 66(227): 395–421.

Pashchenko, Svetlana, Ponpoje (Poe) Porapakkarm, and Mariacristina De Nardi. 2017. "The Lifetime Costs of Bad Health." National Bureau of Economic Research Working Paper 23963.

Poterba, James M., Steven F. Venti, and David A. Wise. 2017. "The asset cost of poor health." *The Journal of the Economics of Ageing*, 9(C): 172–184.

Prados, Mara Jos. 2018. "Health And Earnings Inequality Over The Life Cycle: The Redistributive Potential Of Health Policies." PhD diss. Columbia University, Department of Economics.

Prinz, Daniel, Michael Chernew, David Cutler, and Austin Frakt. 2018. "Health and Economic Activity Over the Lifecycle: Literature Review." National Bureau of Economic Research Working Paper 24865.

Pudney, Stephen, and Nicole Watson. 2013. "If at first you don't succeed? Fieldwork, panel attrition, and health-employment inferences in BHPS and HILDA." Institute for Social and Economic Research (ISER), University of Essex 2013-27.

Salvati, Francesca. 2021. "Health Inequality, Labor Supply and Retirement Policies." PhD diss. University College London, Department of Economics.

Sarafidis, Vasilis, and Neville Weber. 2015. "A Partially Heterogeneous Framework for Analyzing Panel Data." *Oxford Bulletin of Economics and Statistics*, 77(2): 274–296.

Wallenius, Panagiotis Margaris Johanna. 2020. "Can Wealth Buy Health? A Model of Pecuniary and Non-Pecuniary Investments in Health." Centre for Economic Policy Research Discussion Paper DP14796.

Windmeijer, Frank. 2005. "A finite sample correction for the variance of linear efficient two-step GMM estimators." *Journal of Econometrics*, 126(1): 25–51.

Zweifel, Peter, Friedrich Breyer, and Mathias Kifmann. 2009. *Health economics [electronic resource]*. . 2nd ed. ed., Dordrecht ; New York:Springer.