Essays in Applied Microeconometrics

Hans-Martin von Gaudecker

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der
Wirtschaftswissenschaften der Universität Mannheim

Abteilungssprecher:	Prof. Dr. Enno Mammen
Referent:	Prof. Axel Börsch-Supan, Ph.D.
Korreferent:	Prof. Dr. Arthur van Soest
Tag der mündlichen Prüfung:	15. Juni 2007

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Axel Börsch-Supan for guiding me towards empirical work, creating a truly excellent research environment at MEA, and providing me with an exceptionally large number of opportunities to make contacts with outstanding researchers all over the world. The most salient reflection of my extensive use of these possibilities is the fact that this thesis is based on joint work with coauthors from London, Mannheim, Rostock, Tilburg, and Lund. I am deeply indebted to each and every one of them: Jérôme Adda, James Banks, Axel Börsch-Supan, Rob Euwals, the late Angelika Eymann, Rembrandt Scholz, Arthur van Soest, and Erik Wengström. Elaborating on each coauthor's contributions would require another thesis chapter. Arthur van Soest also agreed to be my second advisor. I am especially grateful for this.

I appreciate the valuable comments received on versions of my work at workshops and seminars at the University of Mannheim, University College London, the German Federal Pension Institute, Tilburg University, the Tinbergen Institute, as well as at conferences in Venice, Paris, Rome, Vienna, Mannheim, and Tilburg. I should like to thank a number of people who gave their time to discuss my work with me and provided detailed comments that helped to improve upon it: Syngjoo Choi, Ralf Himmelreicher, Michael Hurd, Morten Lau, Alexander Ludwig, Jürgen Maurer (who deserves a special mention for the countless fruitful discussion we had over many coffees, dinners, and shared office days), Matthias Parey, Jonas Radl, Daniel Schunk, Vladimir Shkolnikov, James Smith, Erik Sørensen, Michael Stegmann, Andreas Uthemann, and Joachim Winter. My colleagues and fellow students at MEA and CDSE were a continuous source of advice and it has been a pleasure to work there, not least because Maria Dauer, Brunhild Griesbach, Helga Gebauer, Isabella Nohe, and Petra Worms-Lickteig never interpreted their work as being limited to administrative matters which they handled superbly.

During my graduate studies, I had the chance to spend prolonged periods of time at other institutions and I thank them for their hospitality: University College London, the German Federal Pension Institute in Berlin (FDZ-RV), and Tilburg University. Financial support from the German Research Foundation (DFG), the European Commission Marie Curie program, the Land Baden-Württemberg, and the German Insurance Association (GDV) is gratefully acknowledged.

Last but certainly not least, I would like to thank my family and friends for their support during the years of my dissertation.

TABLE OF CONTENTS

1. INTRODUCTION

The field of econometrics seeks to give empirical content to models from economic theory. Arguably the most frequently encountered stumbling block on this quest is the availability of suitable data. Variables that are central for the theories under scrutiny are often unobserved, measured with error, or simultaneously determined with other relevant characteristics. A large part of econometric theory can be characterised by searching for ways to identify and estimate the parameters of interest for typical data configurations. Starting from this basis, the applied researcher needs to match these methods with a dataset in which the corresponding assumptions – some of which will be fundamentally untestable – can be made with reasonable confidence. This data-driven view of empirical work forms the thread that holds the five self-contained chapters of this dissertation together.

In Chapter 2, Rembrandt Scholz and I exploit a novel dataset in order to study life expectancy differentials among elderly German men. Unlike in many other countries, publicly available German census records do not include socio-economic variables. It was only with the inception of the Research Data Centre of the German Federal Pension Institute in 2004 that it became possible to analyse data that satisfy the heavy requirements of nonparametric life table estimators. Utilising a measure of lifetime labour income for a sample of 3.5 million pensioners in 2002, we can put a lower bound of six years on the life expectancy difference at age 65 between the groups at the extremes of the lifetime earnings distribution. Furthermore, our analysis shows that the lower overall life expectancy in the eastern part of Germany appears to be driven by a composition effect: Conditional upon socio-economic status, we cannot detect any mortality differences. It is rather the distributions of earnings that are quite distinct from each other with relatively more persons populating the high categories in the West than in the East.

Moving beyond a mere data description, the analysis performed in the third chapter aims at estimating a causal effect of income on health based on a synthetic cohort approach. In joint work with Jérôme Adda and James Banks, we model both variables as stochastic processes for the population between 30 and 60 years of age. The aggregation to the cohort level is motivated by two observations. First, there are no panel data currently available that contain both health and socio-economic variables in long enough time series as to facilitate a truly dynamic analysis at the individual level. Second, we argue that permanent shocks to cohort income are driven by other factors than health innovations, allowing us to disregard reverse-causality issues. We also present some robustness checks with respect to deviations from this assumption. Other factors that have been shown to influence both income and health trajectories – such as education and childhood circumstance – will be captured by the initial conditions. Hence we are confident that we can identify the causal effect that we aim for.

Interestingly, we find a positive and statistically significant effect on mortality, i.e. positive income innovations lead to higher subsequent mortality. There are no effects of income shocks on health as measured by a host of health indicators. The reason for the latter finding could be that causation may take more time than three years, which is what our estimators allow for. There are few reasons for risk behaviours to exhibit a very long causation lag. Accordingly, we consider them as well and find that income innovations lead to more smoking and probably also more drinking. The effects on mortality and risk behaviours are consistent with those estimated by Ruhm (2000), who shows that macroeconomic conditions are positively related to both outcomes. However, our results hint at a different interpretation. In the light of no effects on health measures, it appears unlikely that more risk-taking behaviour leads to instantaneous deaths. It is more probable that work-related accidents are the driving force. When it comes to informing policy, our results suggest that simple redistributionary policies would not help to ameliorate health inequalities.

In the last three chapters of this thesis, the focus shifts from health economics to the measurement of individual-level preferences towards risk-bearing and the timing of events. Since such preferences are intimately linked with other variables relevant for market behaviour, they are best inferred from situations that are largely controlled by the researcher. This is an active field of research and no consensus has been reached on what is a good way to elicit this type of preferences. Typical approaches include experiments or hypothetical choice tasks. However, these are far from being included in household surveys as a default and questionnaires are often tailor-made for specific research questions. I have been fortunate to be able to collaborate on these kind of data collection efforts in two specific projects.

The analysis reported in Chapter 4 is joint work with Arthur van Soest and Erik Wengström. It investigates the role of selection bias in experiments. The data stem from a preference elicitation experiment which we ran on an Internet survey that is representative of the Dutch population. Upon learning about the nature of the experiment, subjects could choose a non-participation option. The advantage of this dataset over previously existing ones is that we have a lot of background information even for subjects who did not take part in the experiment. We find that selection issues matter most for choices that cannot be rationalised by utility-maximising behaviour, namely opting for dominated alternatives and exhibiting non-monotonic choice patterns in similar decision tasks. First, the number of such inconsistencies is drastically higher for the general population than for the young and educated. Second, we find evidence for selection based on observable characteristics within the representative sample. Respondents who are more likely to participate in the experiment show more consistent choice patterns than those with a lower propensity to take part in it. Both selection effects appear to be less important for average levels of risk tolerance than for inconsistent choices.

In Chapter 5 – again with Arthur van Soest and Erik Wengström – we attempt to rationalise some empirical regularities in the same data by means of a structural model. The starting point of our analysis is a canonical expected utility of income model that we extend in two directions that have been discussed over the last thirty years. The first of these is the incorporation of loss aversion. Since the seminal work of Kahneman and Tversky (1979), it has become a widely established finding that agents are more sensitive to losses than to gains. Second, we introduce scope for preferences towards the timing of uncertainty resolution along the lines of the model of Kreps and Porteus (1978). In our two-period model, uncertainty may be resolved either in the first or in the second period. Our results can be summarised as follows. As the previous literature, we find strong effects of risk aversion and loss aversion in homogeneous specifications of the model. Augmenting it by the uncertainty resolution parameter produces mixed results. However, we can say with some confidence that a negative prospect in the outcome set significantly reduces the attractiveness of late uncertainty resolution, something that is well in line with earlier psychological findings. As one would expect, we find heterogeneity in risk aversion to be very important. Interestingly, only a very small part of the overall between-subject variation is explained by the most important demographic variables. Among these, age is associated with higher risk aversion. Men show less risk-averse behaviour than women. Finally, education and household income are linked with lower levels of risk aversion.

In the final chapter of this dissertation – joint with the late Angelika Eymann, Axel Börsch-Supan, and Rob Euwals – we employ more traditional survey instruments in order to measure risk attitudes and impatience. The resulting individual-level parameters are embedded in a model of households' choice among different asset classes. For this purpose, we included a host of tailor-made questions in the 2005 wave of the German SAVE survey. The motivation behind using many different measures of the same outcome is based on the observation that each of these can provide only an inaccurate indicator of the underlying preference parameter. We exploit factor-analytic techniques from the psychometric literature to recover the parameters of interest. We then use these preference parameters as latent traits in a structural model of household portfolio choice. Our results are very encouraging: Using the comprehensive measurement model, we find significant effects of individual preferences on households' choices of portfolio allocations. We show that such results are unlikely to emerge from analyses that employ a single proxy variable per preference parameter. It also turns out to be important to explicitly model dependencies among wealth, risk attitude, and impatience. We thus conclude that there are important merits to our structural approach as it allows to capture effects that would be hidden in reduced form strategies.

BIBLIOGRAPHY

KAHNEMAN, D. V., AND A. V. TVERSKY (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–291.

KREPS, D. M., AND E. L. PORTEUS (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

RUHM, C. J. (2000): "Are Recessions Good For Your Health?," *Quarterly Journal of Economics*, 115(2), 617–650.

2. LIFETIME EARNINGS AND LIFE EXPECTANCY

JOINT WITH REMBRANDT D. SCHOLZ

2.1 Introduction

The international literature on socio-economic status and mortality is marked by a persisting absence of Germany.[1] This scarcity of studies is probably owed to the lack of large high quality datasets. Luckily, the situation has changed since the inception of the Research Data Centre of the German Pension Insurance (Forschungsdatenzentrum der Rentenversicherung). Using these data enables us to make a number of, albeit small, contributions to the existing literature.

Most obviously, we document mortality inequalities among elderly men in Germany. The data permits us to compare the regions of the former GDR with the rest of Germany. Due to the very different institutions for forty years, differences may well be expected. In addition, the German pension system enables us to measure socio-economic status by means of a variable that we term lifetime earnings. It is a discounted sum of pensionable earnings over the life-cycle. We argue that this is a very broad measure of socio-economic status that is also readily usable for the retired population. Finally, due to the large size of our dataset, we do not need to recur to any parametric assumptions on the structure of the relationship between lifetime earnings and mortality. Hence we are able to provide credible estimates of life expectancies which provide a summary measure of mortality that is readily understood in terms of meaning and magnitude, which is not the case for many other typically used measures.

For the remaining life expectancy at age 65, our results indicate a lower bound of six years on the difference between the lowest and the highest earnings group considered in our study. That is a difference of almost fifty percent measured from the lowest group. Over a sizable part of the lifetime earnings distribution, life expectancy rises almost linearly. Despite the fact that we do find lower overall mortality in the West than in the East, our results show similar life expectancies within income groups. The unconditional difference hence comes from composition effects.

Our results document a pure correlation between lifetime earnings and mortality. From our estimates, nothing can be said about the underlying pathways that lead to these figures. In general, three broad channels of causality can be imagined. Epidemiologists stress the importance of causality from income to health (Marmot 1999). Economists are often preoccupied with quantifying the reverse direction by which it is a low health status that impairs current and future earnings capacity (Smith 2004). A third explanation is that there are one or more underlying factors determining both income and health. Among many

[1] See for example Mackenbach, Bos, Andersen, Cardano, Costa, Harding, Reid, Hemström, Valkonen, and Kunst (2003) or Huisman, Kunst, Andersen, Bopp, Borgan, Borrell, Costa, Deboosere, Desplanques, Donkin, Gadeyne, Minder, Regidor, Spadea, Valkonen, and Mackenbach (2004) and the references cited therein.

potential candidates are genetics, ability, intelligence social skills, networks, and other background or early life factors such as parental income (Case, Lubotsky, and Paxson 2002), or education (Lleras Muney 2005). We see our contribution in documenting a strong relationship between lifetime earnings and life expectancy in Germany that calls for more research on its origins.

The structure of this paper is the following. We first describe the data that we use in Section 2.2, with particular emphasis on the German pension system and the calculation of the central explanatory variable. Section 2.3 contains the presentation of our results in three stages before turning to some international comparisons. Finally, Section 2.4 concludes.

2.2 Data

We use a very large dataset of administrative records from the German Public Pension System. For reasons explained in Section 2.2.2, we consider only male individuals. Our data cover more than eighty percent of the whole male population born 1936 and earlier. Consider the first and fifth column of Table 2.1 in the Appendix on page 26. These show our sample sizes and what part of the general population is covered by our data broken down by age group and place of residence. Coverage is about three quarters in the West, but it drops substantially at older ages to less than two thirds. Part of this may be explained by overestimation of the general population in the official public records (Jdanov, Scholz, and Shkolnikov 2005), although we use numbers from the Human Mortality Database that already contain a correction for this issue. Another part is likely to be differential mortality – people outside the system tend to have a higher socio-economic status. In terms of selection bias, we do not see any reason why results on differential mortality among the pensioners covered by our data should not extend qualitatively to the rest of the population. Numbers in the East are much nicer with coverages of up to 98 percent. This is because in the former German Democratic Republic (GDR), virtually everybody was insured within the state pension system. At the same time there were exceptions for the self-employed and civil servants (see Section 2.2.1) in the Federal Republic of Germany (FRG) that explain the baseline difference.

In the next section we turn to the derivation of the central variable of our analysis, an internal measure of the Public Pension System that is used to calculate pension benefits. It is called personal earnings points and serves reasonably well as an indicator of total lifetime earnings. Section 2.2.2 presents descriptive statistics on the variables used in our analysis and deals with the creation of our dataset from several sources of the administrative records.

2.2.1 The German Public Pension System

The German Public Pension System in the form that is relevant for the cohorts studied in this analysis is a pay-as-you-go system based on a single tier. Benefits are directly related to personal earnings over the life-cycle.[2] This section provides a very brief introduction to those parts of the system that are relevant for the purposes of the paper. Our description is based on Börsch-Supan and Wilke (2004) and VDR (2004). The system covers all private and public sector employees, excluding only civil servants (about 7 percent of the workforce) and most self-employed (about 9 percent). The latter can self-insure in the system, we will get back to this in Section 2.2.2. Our focus is on old-age pensions which are paid to all retirees age 65 and above. By the end of the calender year in the course of which age 65 is completed, virtually everybody is retired.[3]

Key to the system are the so-called earnings points, which essentially are a measure of the relative annual earnings position. In any given year t, the earnings points for contribution periods (EP^{CP}) of an individual i are calculated as:

$$\mathrm{EP}_{it}^{CP} = \frac{\text{pensionable earnings}_{it}}{\text{average pensionable earnings}_{t}} \tag{2.1}$$

In 2002, pensionable earnings were the first 4,500 Euro of gross monthly wages if the individual's earnings were above the minimum earnings threshold of 325 Euro. A subset of our data contains the sum of EP_{it}^{CP} over all t with relevant contributions for each individual (we call this variable EP_{i}^{CP}). Note that this variable is subject to a difficult form of right censoring because of the annual upper limit to pensionable income. Hence we know only a lower bound for the earnings of people with high EP_{i}^{CP}. This has to be kept in mind when interpreting the results. On the contrary, the left censoring at 325 Euro is negligible. We note that because of the division by average pensionable earnings in (2.1), the discount rate inherent to EP_{i}^{CP} is the annual wage growth rate.

For administrative reasons, EP_{i}^{CP} is available only for individuals who retired after 1992, the year of a major reform to the system. The measure present for all persons in our dataset is called personal earnings points (EP^{pers}). These are calculated as follows:

$$\mathrm{EP}_{i}^{pers} = (\mathrm{EP}_{i}^{CP} + \mathrm{EP}_{i}^{NCP}) \cdot \mathrm{AF}_{i} \tag{2.2}$$

EP^{NCP} stands for earnings points from non-contributory periods. These stem from

[2] For a simple taxonomy of pension systems and an international comparison cf. OECD (2005)

[3] For the 1936 cohort (the youngest included in our analysis), internal statistics of the Deutsche Rentenversicherung show that only 0.56 percent were not retired on 1st January 2002, the starting point of our analysis.

spells with no contributions at all which are nonetheless relevant for some pension benefits. These include, for example, long-term sickness or unemployment spells, the months when disability pensions were received,[4] some allowance for advanced education, and the like. The adjustment factor AF_i scales down benefits in the case of early retirement after the 1992 reform. For our purposes it also serves to capture a type of minimum pension benefit for low earnings (Mindestentgeltpunkte bei geringem Arbeitsentgelt) that took place before 1992.

Individual pension payments are obtained directly from EP_i^{pers} by multiplication with the current pension value that is common for all pensioners. In 2002 it was 25.86 Euro for EP_i^{pers} earned in the FRG and 22.97 Euro for those EP_i^{pers} earned in the GDR. Hence some 50 EP^{pers} translate into a monthly gross pension payment of 1293.00 Euro (1148.50 Euro) in the West (East). The current pension value is adjusted annually according to complex procedures, this does not impact upon our analysis. We only need the fact that EP^{CP} and EP^{pers} remain constant once an individual receives an old-age pension. For EP^{pers} there are some minor qualifications to this, for example due to divorce or moving abroad. Since we only include pensioners living in Germany (see Section 2.2.2), the latter does not impact upon our analysis and we treat the former as negligible.[5]

We prefer EP^{CP} as a measure of lifetime earnings because EP^{pers} contains too many items that have nothing to do with lifetime earnings but rather reflect social policy measures. For cohorts born after 1928 we can compare both measures. Correlations are very high with $\rho \simeq 0.95$. We present the results of a comparative analysis of mortality experiences based on both different measures in Section 2.3.2. Our results show that the distinction is not all too important in terms of describing the mortality experience by earnings group for ages 65 to 73. For the calculation of life expectancies we also need the mortality experience of older cohorts. Hence we extrapolate the similarity result and interpret EP^{pers} as lifetime earnings although there is larger error inherent to it. The nice feature of these variable is that they give us a measure of long-run earnings. This is a much nicer measure for socio-economic status than current income typically recorded in surveys. The latter is often blurred by transitory fluctuations which are surprisingly high at first glance. These may lead to serious biases as documented in Haider and Solon (2006). Our discounted sum of lifetime earnings misses out on some things typically included in the income definition (for example bequests, capital income or transfers). Bearing the incompleteness in mind, we use lifetime earnings and lifetime income as synonyms in the remainder of the paper. Another

[4] Legislature on disability pensions has been subject to several changes over the years. They are paid until age 65 and their recipients continue to accumulate earnings points. Upon passing age 65, the disabled are treated like everybody else.

[5] Calculations based on the “Versorgungsausgleichstatistik" show that changes in EP^{pers} due to divorce affect only 2.6% of the cases in our sample.

large advantage of the measure is that it remains valid for retired persons – using broader measures may lead to biases due to differences in savings behaviour at earlier stages of the life-cycle.

This description of the pension system has focused on the FRG until 1990 and the reunified Germany thereafter. In the GDR there was a somewhat different system at work. It is beyond the scope of this paper to provide a detailed description but we note that accumulated earnings points are comparable in the sense that the ones from the GDR are also based on the length of the work life and the position in the annual earnings distribution. A detailed description of how pension entitlements were transferred is contained in Stephan (1999). It is safe to say that the amount of income needed to gain one earnings point in the GDR had much less buying power than in the FRG. On the other hand, the pension income following from GDR earnings points is only 13% less than the one stemming from FRG earnings points. Hence the relative position in the earnings distribution and pension income streams are comparable while economic status during the working life is not.

Let us close this section with a brief illustration of the monthly pensionable income necessary to accumulate a certain amount of earnings points. In 2002, the monthly gross wage that yielded one EP^{pers} was about 2,400 Euro. For simplicity, assume that this number remains constant over an individual's working life. Hence, to accumulate 50 EP^{pers}, a person with this wage would have to work 50 years. If average earnings over the life-cycle were 3,000 Euro, 40 full years of contribution would be sufficient to accumulate the same amount of EP^{pers}.

2.2.2 Description of the Dataset

The administration of the German public pension system is marked by a variety of statutory bodies. Traditionally, there have been regional pension insurance institutes for workers, a federal institute for salaried employees, and three profession-specific institutes. Except for miners, legal regulations have been the same since 1949, however. All pension insurance institutes are required by law to report statistics of all pensioners as of the end of each year as well as statistics of those pensioners who died during that year to their umbrella association.[6] We have access to this data.[7] Because only selective characteristics of the original

[6] This used to be the Verband Deutscher Rentenversicherungsträger (VDR). After a major organisational reform that took effect on 1st October 2005, its duties are fulfilled by the Deutsche Rentenversicherung Bund, the federal pensions institute.

[7] Traditionally, only aggregate statistics were published. This has changed since the beginning of 2004 with the creation of the Research Data Centre of the Public Pension Insurance. Information about datasets and access procedures can be found at http://www.fdz-rv.de. The process of data collection is described in Rehfeld (2001)

administrative records enter our dataset, some important remarks about structure and peculiarities of the data are in order.

For one thing, there is no way to link members of the same household. Ideally, we would want to use lifetime household income as a relevant measure to correlate with mortality. Because of the low female labour force participation in the cohorts relevant to our analysis, we exclude them from the analysis. It is simply unclear what the household income position of women with low EP^{CP} is because of the dominance of male earnings in total household income.

For all individuals in our dataset, we have the year and month of their birth available. We only use individuals born in 1936 and earlier because of possible health and income differentials in early retirement. Put differently, if we were to use younger pensioners, most likely we would not have a random sample available. A further demographic variable is the place of residence in three categories (eastern Germany, western Germany, foreign). We exclude people with foreign domicile (2.3%) in order to work with a subset of those recorded in official population statistics. Including them did not cause any visible differences in the mortality estimates. The data contain all deaths in 2002, these are recorded on a monthly level. The Appendix contains descriptive statistics for the entire sample in Table 2.2. Those restricted to pensioners born after 1928 are listed in Table 2.3.

In terms of variables related to pension payments, the two most important ones are those described in the last section. Note that EP^{pers} is on average about 3.1 points (6.6 percent) higher than EP^{CP} in columns 1 and 5 of Table 2.3. We also have the corresponding variables with respect to the length of being insured. These are pension-relevant insurance periods (IP^{PR}) and contribution periods (CP). The former are comprised of the latter and non-contributory periods eligible for pension benefits (for examples see Section 2.2.1). Last, we use information on the type of health insurance coverage. Employees are mandatorily covered in the public mutual funds system up to an earnings threshold that was 75% of the maximum pensionable earnings until 2003. Individuals above that threshold, the self-employed and civil servants can either insure voluntarily in the system or opt out to join a private insurance company. A small subgroup of pensioners in our sample is insured under foreign law, these persons usually worked in Germany only for short periods of time. The arrangement of the last employment spell usually carries over to retirement. We can identify the three groups (mandatorily insured / voluntarily or privately insured / insured under foreign law) in our data.

The reason why this becomes important lies in work biographies that are not confined to a single system of pension insurance. As an example, take somebody who is employed for ten years and then becomes a civil servant for the rest of his working life. If we used EP^{CP} as a measure of his lifetime earnings, we would make a huge error because he had large earnings outside the system. The health insurance variable as well as the length of pension insurance periods enable us to (partially) control for such cases. This is why we estimate life expectancies for four different subgroups of the population: All pensioners (“All"), individuals mandatorily insured within the public health insurance scheme (“HI"), those with more than 25 years of pension-relevant insurance periods (“25Y"), both the “HI" and “25Y" restrictions imposed (“HI25Y"). These make up the remaining columns of Tables 2.1 to 2.3. Looking at the population coverage in Table 2.1, it becomes clear that it drops to two thirds in the West if both restrictions are active. The decline is particularly sharp at older ages because the differential mortality effect is reinforced (see below in Section 2.3.1) and because there are more missing values at old ages. The only variable that is completely available is EP^{pers} because it is central for the pension payment. In conversations with statisticians at the Deutsche Rentenversicherung Bund, we tried to evaluate the influence of systematic effects on missing values. Except for the cases that we mention, there is no reason to expect a missing at random assumption to be violated. In the East the picture is much nicer again with coverage rates above ninety percent except for the very old ages.

There are three more variables available that we do not consider in the presentation of differential mortality by lifetime earnings for clarity reasons. These are citizenship in two categories (German/non-German), whether a pension entitlement for repatriates forms part of the total pension, and whether EP^{pers} includes a scaling-up of raw earnings points because of low earnings before 1992 (cf. Section 2.2.1). Confining our analysis to Germans that do not fall into either of the last two categories did not substantially alter any results, tables and graphs are available from the authors upon request.

2.3 Results

Our analysis is motivated by time series evidence of rising per-capita pension payments over time within cohorts. Net of changes to the current pension value (see Section 2.2.1 for details), this can only be due to changes in cohort composition. Once everybody is retired, differential mortality is the sole reason for this phenomenon to occur, i.e. persons with less than average EP^{pers} are dying relatively more frequently than those with higher earnings points. Shedding more light on this relationship by means of period life tables is the purpose of this chapter. Ideally, we would prefer a cohort analysis in the spirit of this motivation. Given the structure of the administrative records, however, there is no straightforward way to follow cohorts over time and we confine our analysis to all cohorts in 2002. We present our results in three parts. First, we look at the mortality experience of all German pensioners by means of remaining life expectancy at age 65 (e_{65}) and annual mortality rates. We compare the two earnings points measures in the second part of the chapter. Because of data availability, the best we can do is to consider the probability of reaching age 74, conditional on reaching age 65. Finally, we contrast the mortality experience of persons living in the former GDR with those residing in the western part of Germany.

Throughout the analysis, we divide the sample into eleven equally spaced groups of earnings points. We then present statistics for each of these. Due to the extremely large sample size, we can afford the luxury of not recurring to any parametric assumptions in calculating mortality rates. Life tables are based on the classic Chiang (1984) formulas, confidence intervals are calculated via a bootstrap procedure with 1000 replications.[8]

2.3.1 Remaining Life Expectancy at Age 65 by EP^{pers}

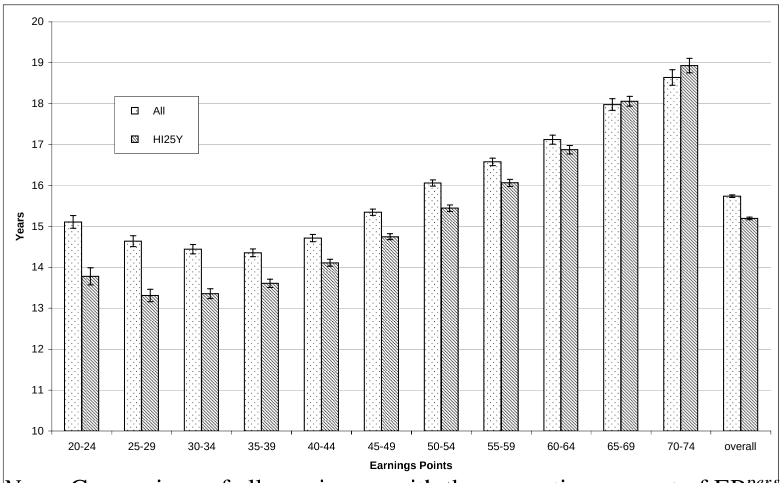
First consider the light bars in Figure 2.1. These depict e_{65} for the full sample of pensioners that we use. Overall remaining life expectancy is at 15.74 years, we postpone a comparison to estimates for the general population until the end of this section. The mortality estimates by earnings points group range from 14.35 years (35-39 EP^{pers}) to 18.65 years (70-74 EP^{pers}). Between these two extremes, life expectancy appears to rise roughly linearly over the groups. All differences among them are statistically significant at any typical confidence level.

The most striking finding at this very first glance is that the minimum life expectancy is reached close to the middle of the table and not in the lowest income group. At the bottom of the distribution, e_{65} is up to more than fifteen years again. This is quite contrary to overwhelming international evidence which indicates a monotone and positive relationship between income and life expectancy. Typically, the gradient found to be steepest in the lower tail (see for example Attanasio and Hoynes (2000)). However, there is a plausible reason at hand. As explained in the last section, we expect a very heterogeneous group at the lower end of the distribution because of persons who were covered by the pension system only during parts of their working life. These are typically well-earning academics who would be at the right tail of the distribution if we were to observe their full earnings history. To take a colourful example, we would expect to find production line workers next to their company's CEO in these groups.

A way to shed light on this issue is to exclude those for whom lifetime earnings are not observed very well. We try to do so by selecting only those who are mandatorily enrolled in the public mutual funds health insurance system or those who spent at least 25 years in the system. The dark bars in Figure 2.1 indicate the

[8] We are very grateful to Evgueni Andreev for sharing his VBA code for bootstrapping lifetables.

Fig. 2.1: Remaining Life Expectancy at Age 65 in years by EP^{pers}



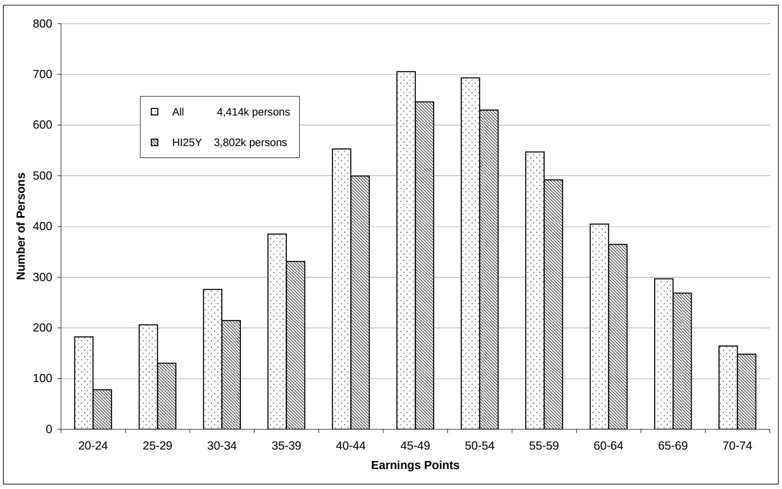
Note: Comparison of all pensioners with the respective amount of EP^{pers} and those who are mandatorily insured in the public health insurance scheme with at least 25 years of pension-relevant insurance periods (HI25Y). The vertical bars indicate 95% confidence intervals

results if we impose both restrictions. For brevity reasons we do not present the results if selection is based upon one criterion only. Again the tables are available upon request.

To begin with, note that overall remaining life expectancy drops substantially by more than half a year. Hence the fourteen percent that we excluded from the original sample must have a much higher life expectancy than the remaining selection. While it rises slightly in the top two earnings points categories, it drops significantly in all other subgroups. The decline is particularly pronounced in the lowest income categories. Minimum e_{65} is now 13.31 years for those with 25-29 EP^{pers}, the slight rise for the lowest income category is only borderline significant. We suspect this differential drop to be a combination of two effects. On the one hand, the relative size of the sample that is excluded is much higher at lower earnings points levels (57% in the lowest category as compared to less than 10% in the top seven classes, see Figure 2.2). On the other hand, if the excluded group is relatively homogeneous, the differential in e_{65} is largest in the lower categories under the hypothesis of a monotonic relationship between income and mortality.

We presume that this control is far from being perfect, hence we only have a lower bound for the mortality differential. This argument is reinforced through the right censoring of annual EP_{it}^{pers} – neither can we clearly identify the top nor the bottom earners. Note that somebody in the lowest mortality group of 30-34 EP^{pers} still would have worked 40 years at 3/4 of average wage. If somebody

Fig. 2.2: The distribution of EP^{pers} by Sample Selection



Note: “All" includes all available observations, “HI25Y" incorporates only those mandatorily enrolled in the public health insurance scheme with at least 25 years of pension-relevant insurance periods.

was living on social assistance most of the time, he would not even enter our dataset. However, even this lower bound on differences in life expectancy is quite substantial. Taking the results from the “HI25Y" selection and the unconditional e_{65} as a starting point, persons in the highest income group can expect to live 25 percent longer. On the other hand, if only 25-29 EP^{pers} were accumulated, it is 12 percent less.

Looking back at the precise relationship between earnings points and life expectancy, we find that it rises almost linearly from the group with 35-39 EP^{pers} to the one with 60-64 EP^{pers}. This is the region where neither top-coding nor earnings outside the pension system should be a major cause of measurement error. It is difficult to compare this linearity finding to other studies (see also Section 2.3.4) since all those that we are aware of either use quantiles of the income/earnings distribution or impose some functional form on the data. While the linear relationship certainly does not extrapolate to larger incomes [9] this finding shows the need to allow for flexible functional forms which accommodate (near) linearity on parts of the distribution.

Our results compare quite well with all-population mortality. Official statistics indicate a remaining life expectancy at age 65 for German males in 2002 of 16.08

[9] To see this note that if we extrapolated our results one would need to observe e_{65} = 54 years for persons who enjoyed a monthly income of 28,000 Euro over a period of 30 years. Clearly, the relationship has to level off at some point.

years (HMD 2006) This is about 4 months higher than our estimates for the full sample indicate. In terms of socio-economic status and mortality experience, most of the persons not covered in our data should be roughly comparable to those in that were excluded when we imposed the “HI25Y" restriction. Simplifying the matter a bit, the main difference between them is that one group worked for a few years in a job that covered them in the public pension system and then changed to another one; the other group started in such another job already. Following this argument we did expect a qualitatively similar rise in life expectancy if we move from “All" to the full population as the one that we see when moving from “HI25Y" to “All".

Finally, Figure 2.3 shows mortality rates for selected income groups under the “HI25Y" restriction. They were calculated separately for each one-year age band. The most salient feature of the graph is that their shape is very similar, they seem to differ by little more than a parallel shift.[10] Confidence bands for each age are not shown in order to keep the graphs readable. Until age 74, all three are statistically different from each other at the 95%-level. Mortality rates in the highest income group are significantly lower than the other two even until age 88.

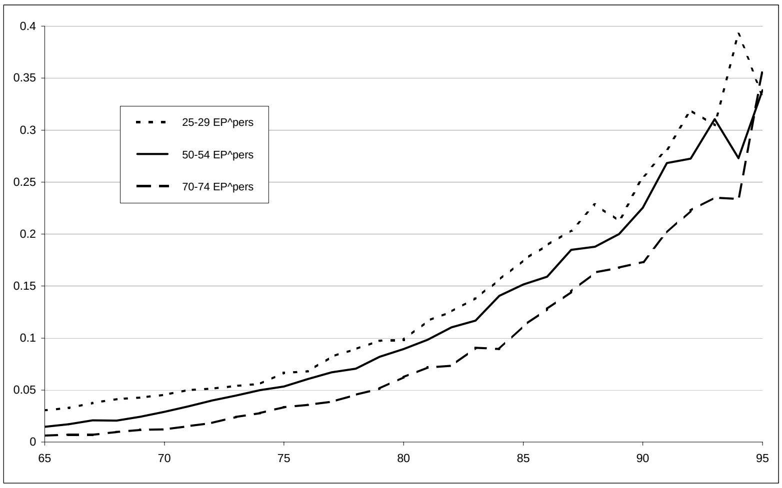
2.3.2 Mortality by EP^{pers} and EP^{CP}

The purpose of this section is to contrast the two different measures of lifetime earnings for those persons where both are available. These are the cohorts born 1929 and later because most of them did not retire before 1992. Remaining life expectancy is not a suitable summary statistic anymore because we do not have any information on old-age mortality conditional on EP^{CP}. As an alternative, we chose the probability of reaching age 74 (this is the highest we can do) conditional on reaching age 65 ($\mathbb{P}_{65}\{74\}$). Results are shown in Figure 2.4. The light bars depict $\mathbb{P}_{65}\{74\}$ conditional on EP^{pers}, the dark ones show the corresponding values based on EP^{CP}.

Overall probabilities are identical at 76.4 percent because we use the same sample in both cases. In the case of EP^{pers}, we find very much the same pattern as for life expectancy in the last section. There is a linear decline from the highest income class to those persons with 35-39 EP^{pers} which then levels out and rises again at the very bottom. The dark bars look quite similar, but there are some important differences. On the one hand, probabilities are slightly higher for all groups with more than 35 EP^{CP}. However, the decline does not level out at this point but continues linearly to the very lowest group. Based on this measure, 65-

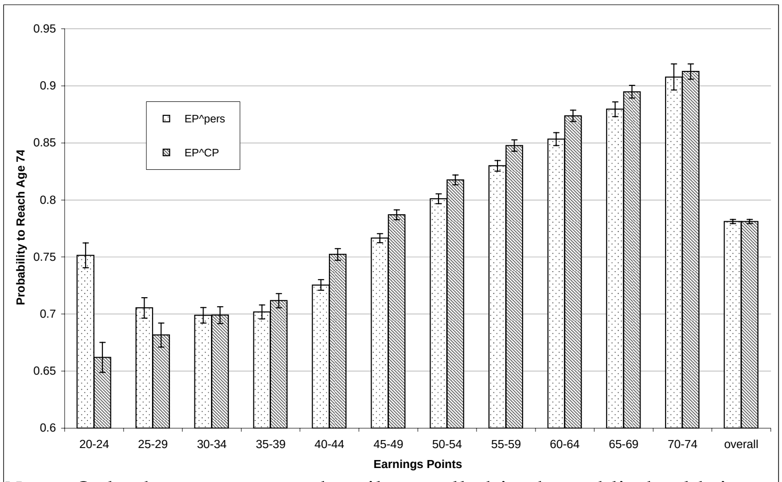
[10] It is certainly much closer to a parallel shift in mortality than to a parallel shift in log mortality (figures are available from the authors upon request), especially in the younger age groups where the data is most reliable. The latter is the parametric assumption inherent in, for example, proportional hazard models. We would like to thank one of our referees for pointing this out to us.

Fig. 2.3: Mortality Rate by EP^{pers}



Note: Only the persons mandatorily enrolled in the public health insurance scheme with at least 25 years of pension-relevant insurance periods are included in the analysis (selection HI25Y).

Fig. 2.4: Probability of Reaching Age 74 at Age 65 by EP^{pers}



Note: Only the persons mandatorily enrolled in the public health insurance scheme with at least 25 years of pension-relevant insurance periods are included in the analysis (selection HI25Y). The vertical bars indicate 95% confidence intervals.

year-old individuals in the highest earnings class have a ninety percent probability of reaching age 74. Less than two thirds in the bottom category survive to this age.

Our conclusions from this exercise are twofold. On the one hand, it does not matter much whether one uses EP^{pers} or EP^{CP} if one is interested in the mortality experience for individuals with at least 35 EP^{pers}. Our summary statistics differ by not much more than a constant that is mostly explained by the average differential of 3 points between the two measures (see the first two rows of Table 2.3). On the other hand, the choice of variable does matter in lower categories. By using EP^{CP} as a measure of lifetime earnings, we can reproduce the monotonic relationship to mortality that is documented in international studies.

In the course of the analysis of this chapter we also checked whether differences exist if we condition on actual contribution periods (CP) rather than pension-relevant insurance periods (IP^{PR}). One might have expected analogous differences as in the case of the two earnings measures. However, none such differences became apparent which is why we do not include a graph on that point.

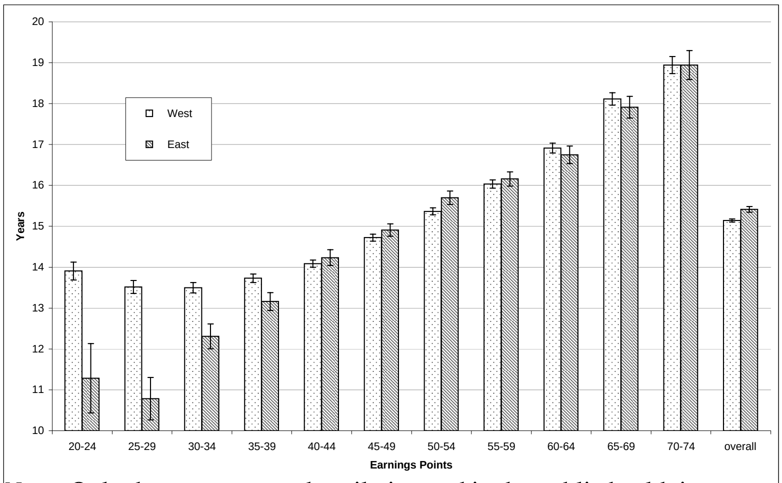
2.3.3 Life Expectancy at Age 65 by EP^{pers} and Place of Residence

In this section, we compare the mortality experience of people living in the former socialist part of Germany with those living in western Germany. This is particularly interesting to look at because of the very different biographies of people living in either part.[11] One could expect several things to happen in the eastern part. There could be a long-run effect of the more equal distribution of socio-economic status during the socialist era, resulting in smaller mortality differentials. The opposite story might read that the sharp transformation in the early 1990s led to higher inequality than in the West. Finally, one could imagine relatively quick adaption to the new institutional arrangement, hence a picture that parallels that in the West.

Naturally, the first thing to evaluate is e_{65} not stratified by income. We find it to be 15.83 years in the West and 15.41 years in the East. The difference is statistically significant at any common confidence level. This compares to full-population estimates from official statistics of 16.19 years and 15.39 years, respectively. In the West we have the effect described above: The 23% of the population not included in our sample tend to have a lower mortality than the pensioners. On the other hand, the 94% coverage in the East leads to almost identical estimates of all-population mortality, so we are very confident to have a complete depiction of the full population there. Our findings are consistent with the converging mortality experiences in the East and in the West that have been documented by several

[11] In our interpretations of this section, we neglect migration between the two regions. People in our sample were at least 53 years old at the time of reunification, so we do not expect any substantial migration movements that would blur our results

Fig. 2.5: Remaining Life Expectancy at Age 65 in years by Place of Residence and EP^{pers}



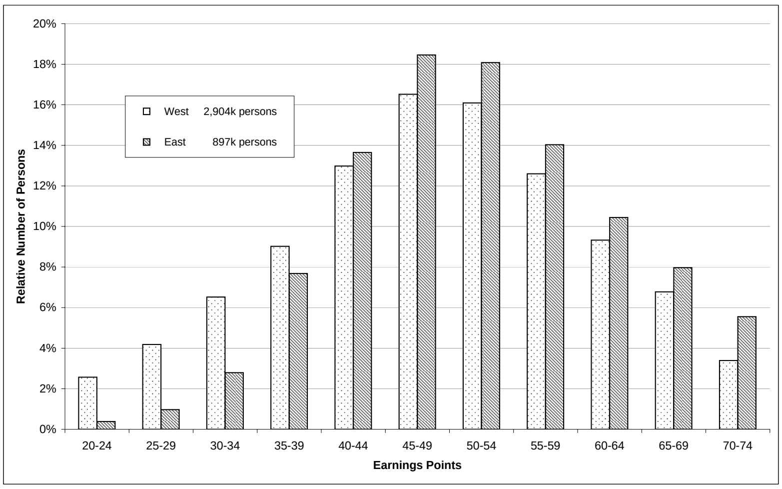
Note: Only the persons mandatorily insured in the public health insurance scheme with at least 25 years of pension-relevant insurance periods are included in the analysis (selection HI25Y). The vertical bars indicate 95% confidence intervals.

authors (cf. for example Nolte, Shkolnikov, and McKee (2000)).

Again, the main analysis concerns the comparison of life expectancy by income group. As in the previous sections, we select upon mandatory enrollment in the public mutual funds health care system and 25 years of coverage in the pension system. The reason not to consider all-pensioner mortality here is that groups are more comparable in the restricted sample. As discussed in Section 2.2.2, we expect much more heterogeneity in terms of earnings outside the system in the West than in the East, particularly in the lowest categories. This is because in the former GDR, virtually everybody was insured in the system. Hence we would obtain a stronger bias within income classes if we did not impose the restrictions. The reason is the larger heterogeneity with respect to socio-economic status in the West. The restriction of the sample makes the analysis within categories more meaningful, but a comparison of the unconditional life expectancies does not make much sense. This becomes clear from the rightmost bars in Figure 2.5. While e_{65} does not change for pensioners in the east of Germany, imposing the restriction leads to a drop in remaining life expectancy to 15.14 years in the West. This reversal is due to the fact that we excluded a large group of men with high socio-economic status in the West and only few persons in the East.

Comparing the graphs for both regions in Figure 2.5, it becomes apparent that no difference between them can be asserted for groups with more than 40 EP^{pers}. In all groups below that level, life expectancy in the West emerges larger than in

Fig. 2.6: The distribution of EP^{pers} by Place of Residence



Note: Only the persons mandatorily insured in the public health insurance scheme with at least 25 years of pension-relevant insurance periods are included in the analysis (selection HI25Y).

the East. Most notably, we find a close to strongly monotone relationship between income and life expectancy in the East (the rise in the bottom group is not detectable in a statistical sense for either region). In the lowest two groups in the East, e_{65} is only around eleven years, eight years less than at the top of the distribution. We do not take this as evidence for more inequality in the East. It is more likely that our selection procedure still does not enable us to pinpoint people with low socio-economic status in the West, just as the analysis from Section 2.3.2 suggests. The higher life expectancy in the East for this particular selection arises because the distribution of earnings points is shifted to the right as compared to western Germany under the HI25Y selection, see Figure 2.6. In other words, there are relatively more people in the higher income classes with a longer life expectancy.

This analysis sheds a new light on previous findings from the epidemiological literature. Several authors reported a larger predictive power of income for health measures in the West than in the East (Mielck, Cavelaars, Helmert, Martin, Winkelhake, and Kunst (2000), Nolte and McKee (2004)). Our results suggest this to be a consequence of measurement error, unless the link between morbidity and mortality works differently between both parts of Germany. The relevant income concept as a marker for socio-economic status is a long-run measure, the surveys employed in the cited studies contain a current income variable. Transitory fluctuations, for example due to high unemployment in the East, may well impact strongly upon the analysis. Here we conclude that inequalities with respect to pension income have the same magnitude in the East as they do in the West.

A second thing to note is the striking similarity of the results in the income classes from 40 EP^{pers} on upwards. This is somewhat surprising given that biographies have been quite different. One part lived for forty years in a socialist country with severely restricted civil rights, comparatively small wage differentials, high job security, an economy with a large amount of goods rationing, etc. People on the other side of the wall experienced much higher wage differentials, somewhat lower job security, a market economy, and quite well protected civil rights. Only from 1990 on – the cohorts in our analysis were already at least 53 years old – institutional arrangements started to converge, although large differences remain, for example on labour markets. In the light of the observed convergence in total mortality (Nolte, Shkolnikov, and McKee 2000), we note that this convergence appears to be nearly perfect conditional on our measure of socio-economic status. A natural interpretation of this would be an income effect: Reunification brought about much higher real incomes for pensioners and twelve years were enough to wipe out any lagged effects. Differences in total mortality continue to exist only because of the composition effects – in the West there are relatively more persons with a high socio-economic status. Note that it does not follow from this that redistribution would lead to equal mortality experiences – higher pension income in our case is also indicative of a higher ranking in the relative income distribution in the GDR/FRG. The parameters that give rise to this ranking are very likely to be related to mortality, both directly and through interaction effects with income. Examples include education, intelligence, social skills and networks, or genetics. However – if interpreted in this fashion – our results suggest that redistribution would lead to a convergence in mortality experiences among socio-economic groups, the extent of this remains unclear.

2.3.4 International Comparisons

In this section, we place our results in the context of the literature on differential mortality. Closest in terms of regional proximity and research question is certainly Reil-Held (2000). She uses survey panel data and finds a life expectancy (e_{37}) differential of 10 years between the top and the bottom quartile of the income distribution, employing a somewhat broader household income variable that is averaged across time. This is in the bivariate analysis that is comparable to our approach. Confidence intervals are not reported due to small sample sizes. We obtained her raw data and compared the estimates of e_{65}. Her findings suggest some 17.8 years for the top income quartile and 10.1 years for the bottom one. The slightly lower value for the top quartile may be explained by the earlier time period (she considered deaths in the 1984-1997 period). The much lower value for the bottom quartile is most likely due to our not very meaningful income data for that population segment. Taken together, the numbers compare very well and our claim to have identified a lower bound of 6 years on the life expectancy difference from the bottom to the top is reinforced by her findings of a 7.7 year differential.

Expanding the region under consideration and the variables measuring socio-economic status, the study by Huisman, Kunst, Andersen, Bopp, Borgan, Borrell, Costa, Deboosere, Desplanques, Donkin, Gadeyne, Minder, Regidor, Spadea, Valkonen, and Mackenbach (2004) serves well to place our results in a European context. They analyse differential mortality by education and housing tenure and find results that are quite similar to ours in qualitative terms. Throughout the eleven countries included in their study, relative mortality risks declined in age, but absolute mortality differences remained constant or increased until old age. This is exactly what we find from Figure 2.3. It is impossible to compare numbers because of the differences in covariates, furthermore the authors do not report life expectancies. Doblhammer, Rau, and Kytir (2005) compare mortality by occupational and educational group in Austria and also find large differences. Again, they report relative mortality risks and quantities are not easily compared.

Further corroborative evidence for our findings comes from the US – Deaton and Paxson (2004) report e_{25} to be about 10 years lower for members of families with an annual income of less than \$5,000 as compared to those earning more than \$50,000. Findings from Attanasio and Hoynes (2000) suggest that the gradient linking income and mortality is steepest at the bottom of the distribution. Since there is no reason to expect this finding to reverse in Germany, this is supportive of our view that we cannot identify people at the lower end of the distribution very well and mortality conditional on correctly measured income would be much higher. Finally, turning back to Europe, Attanasio and Emmerson (2003) show that the position in the wealth ranking has a large impact on survival probabilities in the UK. Palme and Sandgren (2004) are able to construct a measure of lifetime income for a cohort born in 1928 in the city of Malmö, Sweden. They find this variable to be only marginally significant in the case of prime-age mortality. However, their sample size is rather small and they conduct their analyses conditional on parental income. Overall, we conclude that our findings for Germany fit in very well with the existing literature on socio-economic status and mortality.

2.4 Conclusions

We found large differentials in remaining life expectancy at age 65 across classes of lifetime earnings in Germany. In our case, lifetime earnings are directly tied to pension income flows. Due to the nature of the data, we were only able to put a lower bound of six years on the difference from the lowest to the highest income group. Over the parts of the income distribution that we measure well, life expectancy rises linearly in earnings. Since we employed entirely nonparametric methods, this is not an artefact of any assumed functional form. However, the finding certainly does not extrapolate to the tails of the earnings distribution where our observations lack precision.

Comparing the former GDR with the rest of Germany, we found virtually no differences in life expectancy within income groups, except for the lowest groups. However, the latter seems to be due to peculiarities of our data. Hence, any aggregate difference in life expectancy is likely owed to composition effects. This finding is quite remarkable because of the very different institutional design people in either part were faced with during the prime ages of their working life. Inequality in life expectancy appears to be equal in both parts.

What do we learn from these results? For one thing, there is a very sizeable effect, even among the elderly, waiting to be explained. Which of the three channels mentioned in the introduction is responsible for how much of the differential? Economic analyses suggest that in the socio-economic status to health causation, income is not likely to play a large role (compare, for example, Adams, Hurd, McFadden, Merrill, and Ribeiro (2003), Meer, Miller, and Rosen (2003), or Smith (2004)). Another finding that deserves further illumination is the similarity of life expectancy in the East and in the West conditional on lifetime earnings. Finally, an interesting research question lies in the temporal changes of mortality. How was the gain in life expectancy over the last ten years distributed over different subgroups of the population? Did the poor or the rich gain more or was it evenly distributed? Is this trend likely to continue? Answers to these questions have a large impact on pension finance, the organisation of nursing home care, and many other important policy questions.

2.5 Tables

Tab. 2.1: Sample Size and Coverage of the General Population by Cohort

Birthyear	West				East			
	All	HI	25y	HI25y	All	HI	25y	HI25y
before 1902	494	396	355	283	154	130	151	127
	(64%)	(51%)	(46%)	(37%)	(96%)	(81%)	(94%)	(79%)
1902 - 1906	7,425	6,649	5,691	5,202	2,307	2,115	2,267	2,082
	(60%)	(54%)	(46%)	(42%)	(98%)	(90%)	(96%)	(88%)
1907 - 1911	52,680	48,448	43,557	41,152	13,874	13,339	13,795	13,269
	(61%)	(56%)	(50%)	(48%)	(95%)	(91%)	(94%)	(91%)
1912 - 1916	152,514	135,688	126,505	116,739	35,139	34,281	34,987	34,158
	(71%)	(63%)	(59%)	(54%)	(96%)	(93%)	(95%)	(93%)
1917 - 1921	321,267	288,243	268,837	247,470	73,661	72,450	73,381	72,216
	(76%)	(68%)	(64%)	(59%)	(92%)	(90%)	(92%)	(90%)
1922 - 1926	629,208	579,373	532,667	498,207	141,318	139,044	140,761	138,571
	(76%)	(70%)	(64%)	(60%)	(91%)	(89%)	(90%)	(89%)
1927 - 1931	1,015,922	931,499	927,315	855,694	272,274	265,174	271,711	264,689
	(79%)	(72%)	(72%)	(66%)	(93%)	(91%)	(93%)	(91%)
1932 - 1936	1,301,622	1,155,290	1,273,053	1,140,150	393,716	372,431	393,438	372,273
	(78%)	(69%)	(76%)	(68%)	(97%)	(92%)	(97%)	(92%)
Total	3,481,132	3,145,586	3,177,980	2,904,897	932,443	898,964	930,491	897,385
	(77%)	(70%)	(70%)	(64%)	(94%)	(91%)	(94%)	(91%)

Note: Sample Size as number of pensioners, coverage of the general population in parentheses. “All" includes all pensioners in the respective age range with more than 20 EP^{pers}. “HI" restricts the sample to individuals who are furthermore mandatorily insured within the public health insurance scheme. “25Y" considers only pensioners with more than 20 EP^{pers} and more than 25 years of pension-relevant insurance periods. “HI25Y" imposes both the “HI" and “25Y" restrictions. Source: DRV (2005), HMD (2006), own calculations.

Tab. 2.2: Descriptive Statistics for all Individuals

Variable	West				East			
	All	HI	25y	HI25y	All	HI	25y	HI25y
EP^{pers}	47.9	48.7	48.7	49.2	52.0	52.2	52.0	52.2
	(12.8)	(12.2)	(12.4)	(11.9)	(10.6)	(10.5)	(10.6)	(10.4)
IP^{PR}	39.0	39.5	42.3	42.5	44.1	44.2	44.2	44.3
	(11.4)	(11.3)	(4.4)	(4.1)	(2.7)	(2.6)	(2.0)	(1.9)
HI=1	0.904	1.000	0.914	1.000	0.964	1.000	0.964	1.000
HI=2	0.092	0.000	0.082	0.000	0.034	0.000	0.034	0.000
HI=3	0.004	0.000	0.004	0.000	0.002	0.000	0.002	0.000
GERMAN	0.968	0.969	0.972	0.973	0.998	0.998	0.998	0.998
MIN PENS	0.043	0.041	0.046	0.045	0.016	0.015	0.016	0.015
DEAD	0.051	0.053	0.053	0.055	0.051	0.051	0.051	0.051

Note: Mean of variables, standard errors in parentheses where appropriate. “HI=1,2,3:" Health insurance coverage by: mandatory public mutual funds, voluntary mutual funds or private insurance company, insurance under foreign law. “GERMAN" shows the fraction of pensioners with German citizenship, “MIN PENS" the fraction entitled for a scaling-up of EP^{CP} due to low earnings before 1992, and “DEAD" the fraction that died during 2002.

“All" includes all pensioners in the respective age range with more than 20 EP^{pers}. “HI" restricts the sample to individuals who are furthermore mandatorily insured within the public health insurance scheme. “25Y" considers only pensioners with more than 20 EP^{pers} and more than 25 years of pension-relevant insurance periods. “HI25Y" imposes both the “HI" and “25Y" restrictions. Source: DRV (2005), own calculations.

Tab. 2.3: Descriptive Statistics for Individuals born after 1928

Variable	West				East			
	All	HI	25y	HI25y	All	HI	25y	HI25y
EP^{pers}	47.6	48.3	48.1	48.6	50.9	51.1	50.9	51.1
	(12.6)	(11.9)	(12.2)	(11.7)	(10.3)	(10.1)	(10.3)	(10.1)
EP^{CP}	44.9	45.6	45.6	46.0	47.4	47.4	47.4	47.5
	(12.9)	(12.3)	(12.6)	(12.1)	(9.8)	(9.6)	(9.8)	(9.6)
IP^{PR}	41.0	41.4	42.4	42.6	44.0	44.1	44.1	44.1
	(8.0)	(7.8)	(4.3)	(4.0)	(2.4)	(2.2)	(2.1)	(2.0)
CP	37.8	38.3	38.4	38.7	5.9	5.9	5.9	5.9
	(7.6)	(7.2)	(6.9)	(6.8)	(6.9)	(6.8)	(6.8)	(6.8)
HI=1	0.896	1.000	0.903	1.000	0.953	1.000	0.954	1.000
HI=2	0.099	0.000	0.092	0.000	0.044	0.000	0.044	0.000
HI=3	0.005	0.000	0.004	0.000	0.002	0.000	0.002	0.000
GERMAN	0.956	0.957	0.962	0.963	0.998	0.998	0.998	0.998
MIN PENS	0.038	0.036	0.039	0.037	0.020	0.018	0.020	0.018
DEAD	0.028	0.028	0.028	0.029	0.029	0.029	0.029	0.029

Note: Mean of variables, standard errors in parentheses where appropriate. “HI=1,2,3:" Health insurance coverage by: mandatory public mutual funds, voluntary mutual funds or private insurance company, insurance under foreign law. “GERMAN" shows the fraction of pensioners with German citizenship, “MIN PENS" the fraction entitled for a scaling-up of EP^{CP} due to low earnings before 1992, and “DEAD" the fraction that died during 2002.
“All" includes all pensioners in the respective age range with more than 20 EP^{pers}. “HI" restricts the sample to individuals who are furthermore mandatorily insured within the public health insurance scheme. “25Y" considers only pensioners with more than 20 EP^{pers} and more than 25 years of pension-relevant insurance periods. “HI25Y" imposes both the “HI" and “25Y" restrictions. Source: DRV (2005), own calculations.

BIBLIOGRAPHY

ADAMS, H. P., M. D. HURD, D. MCFADDEN, A. MERRILL, AND T. RIBEIRO (2003): "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status," *Journal of Econometrics*, 112(1), 3–56.

ATTANASIO, O. P., AND C. EMMERSON (2003): "Mortality, Health Status and Wealth," *Journal of the European Economic Association*, 1(4), 821–850.

ATTANASIO, O. P., AND H. W. HOYNES (2000): "Differential Mortality and Wealth Accumulation," *Journal of Human Resources*, 35(1), 1–29.

BÖRSCH-SUPAN, A., AND C. B. WILKE (2004): "The German Public Pension System: How It Was, How It Will Be," NBER Working Paper No. 10525.

CASE, A., D. LUBOTSKY, AND C. PAXSON (2002): "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92(5), 1308–1334.

CHIANG, C. L. (1984): *The Life Table and its Applications*. Krieger Pub. Co., Malabar, Fla.

DEATON, A., AND C. PAXSON (2004): "Mortality, income, and income inequality over time in Britain and the United States," in *Perspectives on the Economics of Aging*, ed. by D. A. Wise, pp. 247–280. University of Chicago Press, Chicago, IL.

DOBLHAMMER, G., R. RAU, AND J. KYTIR (2005): "Trends in educational and occupational differentials in all-cause mortality in Austria between 1981/82 and 1991/92," *Wiener Klinische Wochenschrift*, 117(13), 468–479.

DRV (2005): "Sonderauswertung Fernrechenfile Differentielle Mortalität 2003," Restricted Use Dataset, Deutsche Rentenversicherung Bund.

HAIDER, S., AND G. SOLON (2006): "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *American Economic Review*, 96(4), 1308–1320.

HMD (2006): "Human Mortality Database," http://www.mortality.org.

Huisman, M., A. E. Kunst, O. Andersen, M. Bopp, J.-K. Borgan, C. Borrell, G. Costa, P. Deboosere, G. Desplanques, A. Donkin, S. Gadeyne, C. Minder, E. Regidor, T. Spadea, T. Valkonen, and J. P. Mackenbach (2004): "Socioeconomic Inequalities in Mortality Among Elderly People in 11 European Populations," *Journal of Epidemiology and Community Health*, 58, 468–475.

Jdanov, D. A., R. D. Scholz, and V. M. Shkolnikov (2005): "Official Population Statistics and the Human Mortality Database Estimates of Populations Aged 80+ in Germany and Nine Other European Countries," *Demographic Research*, 13(14), 335–362.

Lleras Muney, A. (2005): "The Relationship Between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72(1), 189–221.

Mackenbach, J. P., V. Bos, O. Andersen, M. Cardano, G. Costa, S. Harding, A. Reid, Ö. Hemström, T. Valkonen, and A. E. Kunst (2003): "Widening Socioeoconomic Inequalities in Mortality in Six Western European Countries," *International Journal of Epidemiology*, 32, 830–837.

Marmot, M. (1999): "Multi-Level Approaches to Understanding Social Determinants," in *Social Epidemiology*, ed. by L. Berkmann, and I. Kawachi. Oxford University Press, Oxford.

Meer, J., D. L. Miller, and H. S. Rosen (2003): "Exploring the health-wealth nexus," *Journal of Health Economics*, 22(5), 713–730.

Mielck, A., A. Cavelaars, U. Helmert, K. Martin, O. Winkelhake, and A. E. Kunst (2000): "Comparison of Health Inequalities between East and West Germany," *European Journal of Public Health*, 10(4), 262–267.

Nolte, E., and M. McKee (2004): "Changing Health Inequalities in East and West Germany since Unification," *Social Science and Medicine*, 58, 119–136.

Nolte, E., V. Shkolnikov, and M. McKee (2000): "Changing Mortality Patterns in East and West Germany and Poland. II: Short-Term Trends during Transition and in the 1990s," *Journal of Epidemiology and Community Health*, 54, 899–906.

OECD (2005): *Pensions at a Glance. Public Policies Across OECD Countries*. OECD Publishing, Paris, France.

Palme, M., and S. Sandgren (2004): "Parental Income, Lifetime Income and Mortality," Mimeo, Department of Economics, Stockholm University.

Rehfeld, U. (2001): "Die Statistiken der Gesetzlichen Rentenversicherung. Zu Stand und Perspektiven des leistungsfähigen, vielgenutzten Berichtswesens.," *Deutsche Rentenversicherung*, 3-4/2001, 160–186.

Reil-Held, A. (2000): "Einkommen und Sterblichkeit in Deutschland: Leben Reiche länger?," Sonderforschungsbereich 504 Discussion Paper No. 00-14.

Smith, J. P. (2004): "Unraveling the SES-Health Connection," IFS Working Paper 04/02.

Stephan, R.-P. (1999): "Das Zusammenwachsen der Rentenversicherung in West und Ost," *Deutsche Angestelltenversicherung*, 46(12), 546–556.

VDR (2004): *Statistik Rentenbestand am 31. Dezember 2003*. Verband Deutscher Rentenversicherungsträger (VDR), Frankfurt am Main, Germany.

3. THE IMPACT OF INCOME SHOCKS ON HEALTH: EVIDENCE FROM COHORT DATA

JOINT WITH JÉRÔME ADDA AND JAMES BANKS

3.1 Introduction

One of the most striking and frequently studied correlations observed in individual level data, is that between measures of income or socio-economic status and measures of health. Data from all over the world, collected in many different years, consistently show that those with greater levels of economic resources have better health. Consequently, an extensive and continually evolving literature has concerned itself with the precise nature of the relationship between income and health and in particular the direction and magnitude of causal links between the two.

Interpretations of the origin of the health-income gradient require researchers to deal with a classic simultaneity issue. Three broad channels of causality can be imagined. Epidemiologists stress the importance of causality from income to health (Marmot 1999). Economists have often tried to quantify the reverse direction going from poor health to impaired current and future earnings capacity (Smith 2004). A third explanation is that there are one or more underlying factors determining both income and health. Among many potential candidates are genetics, ability/intelligence and other background or early life factors such as parental income (Case, Lubotsky, and Paxson (2002), Dehejia and Lleras Muney (2004) or van den Berg, Lindeboom, and Portrait (2006)) or education (Lleras Muney 2005).

In this paper, we focus on the effect of income shocks on health over the life-cycle. First, we specify an individual dynamic model of both income and health which allows us to decompose health and income shocks into transitory and permanent ones. We also allow for a non-linear relationship between income and health at individual level. Second, we aggregate and estimate this model, using synthetic cohorts followed for up to twenty-five years, which covers a substantial part of the life-cycle. Third, we estimate the effect of *permanent* shocks to income on a number of health outcomes including subjective and objective measures, as well as health behaviour and mortality.

We use a synthetic cohort methodology for two reasons. First, by using it we aggregate out individual level variation, allowing us to focus on exogenous changes in income at the cohort level. In many countries, including the UK, there have been important changes in the income structure over time (Buchinsky (1994) or Gosling, Machin, and Meghir (2000)). The causes appear to range from changes in the return to education and experience due to skill-biased technology changes, declines in unionization and increased competition. Importantly, however, these changes are not thought to be caused by changes in health and we will use this as an exogeneity assumption to identify a causal effect of income on health. In an individual level model it would be much harder to disentangle the effects in question. [1] In the spirit of robustness analysis we also investigate briefly the results we would obtain under alternative identifying assumptions on the strength of the effect of cohort health shocks on cohort incomes.

A second advantage of cohort data is that we can exploit a wealth of data with detailed information on both income and many health outcomes because it allows the combination of various datasets. This is particularly important in our analysis because data on health and socio-economic variables typically have not been recorded jointly. Longitudinal data sets have either good information on income dynamics (e.g. the PSID in the US, or the BHPS in the UK) but no or very limited information on health, or detailed health information but very limited information on income or other socioeconomic variables. This has changed recently, but only over the last decade, and mostly for data covering the elderly and near-elderly population (e.g. the Health and Retirement Survey). Hence there are no long-T panels currently available that facilitate the study of the joint dynamics of income and health for the prime-aged population.

In our analysis we use techniques recently developed in the microeconometric literature for modelling income processes and income uncertainty and apply them to the question of identifying the effect of income shocks on subsequent health outcomes. More precisely, we use the methods of Meghir and Pistaferri (2004), who in turn build on the methods of Abowd and Card (1989), Gottschalk and Moffitt (1994), Banks, Blundell, and Brugiavini (2001), and Blundell, Pistaferri, and Preston (2005) to identify the covariance of permanent income shocks with changes in a series of health indicators. In this way we are able to extend the Granger-causality type analysis, previously carried out for US individuals aged over 70 (Adams, Hurd, McFadden, Merrill, and Ribeiro 2003) and aged over 50 (Smith 2004), and apply it to a working age population in England. These previous studies found no effect of income shocks on subsequent health changes, once one conditions on initial health and income. Differences in health care institutions, between the UK and the US, and particularly the relative lack of private provision in the UK, would hardly be expected to reverse such a result. The use of a different age group, however, could lead to different findings than those of the above studies, regardless of the country of study.

We show that different cohorts were affected by sizable permanent shocks to income over that period, especially those with low education. We find, as in Ruhm (2000), that these shocks have an impact on mortality. However, our results show that these shocks are not transmitted to any of our other health measures, whether we consider subjective ones (self-assessed health, longstanding illness) or objec-

[1] Some attempts have been made using exogenous changes in income such as lottery gains (Lindahl 2005), or “unexpected" inheritances (Meer, Miller, and Rosen 2003) as instruments although these instruments apply only to subgroups of the population which may not necessarily be representative.

tive ones (high blood pressure, cardiovascular diseases, or respiratory diseases). In contrast, we show that individuals change some of their behaviour such as total expenditure as well as expenditures on tobacco and alcohol. Our interpretation of these three findings is that risk behaviours do not seem to transmit directly into mortality or morbidity for the prime-aged population. Procyclical mortality is probably rather driven by work-related accidents and similar mechanisms, but our results are not more than merely suggestive on this point.

Our analysis is structured as follows. Section 3.2 describes our empirical strategy by setting out the structural framework for the dynamics of income and health processes, outlining our key identification assumptions and documenting our estimation methodology. Section 3.3 describes the various datasets that we draw upon in order to carry out our analysis, and provides some simple summary statistics. Section 3.4 presents our empirical results, first with reference to the effect of income shocks on mortality and then with reference to a host of health indicators and behaviour. The last part of that Section (3.4.6) contains a detailed comparison of our approach to related ones found in the literature. Finally Section 3.5 concludes.

3.2 Empirical Strategy

In this section we describe our approach to modelling income and health as stochastic processes that evolve over the life cycle. We then turn to our strategy to estimate the relevant parameters. In either of the two areas, there exists substantial previous work that we can build on. Starting with MaCurdy (1982), a large literature has emerged that models life-cycle income or earnings as stochastic processes. In terms of methodology, Abowd and Card (1989) and especially Meghir and Pistaferri (2004) are close to our approach. All these articles deal with individual panel data, applications to consumption with synthetic cohorts include Blundell and Preston (1998) and Banks, Blundell, and Brugiavini (2001).

Regarding the literature on the evolution of health, Smith (2004) gives an extensive overview of different specifications. Few authors have used pseudo panel data in connection with health outcomes, a notable exception being a series of papers by Deaton and Paxson (1998b, 2001, 2004). We provide a comparison of our approach and results to theirs (and those of Ruhm (2000)) in Section 3.4.6.

We adopt the following notation throughout the analysis. ε and u denote transitory shocks, ζ and v permanent shocks, x^Y income-related variables and x^H health-related variables. i/N denote an individual and the number of individuals respectively, c/C are cohorts and the number of cohorts, t/T a time period (one year) and the time dimension, a/A a particular age and the terminal age. L denotes the lag-operator.

3.2.1 Stochastic Process for Individual Income

We assume the following nonstationary process for the evolution of log household income over the life-cycle:

$$Y_{it} = Y_{i,t-1} + (1 - L)u^Y_{it} + v^Y_{it} \tag{3.1}$$

The error terms are assumed to have the following structure:

$$u^Y_{it} = \varepsilon^Y_{ct} + \xi_1 \varepsilon^H_{ct} + \varepsilon^Y_{it} + \phi_1 \varepsilon^H_{it} \tag{3.2}$$

$$v^Y_{it} = \zeta^Y_{ct} + \xi_2 \zeta^H_{ct} + \zeta^Y_{it} + \phi_2 \zeta^H_{it} \tag{3.3}$$

We deliberately distinguish between cohort-wide shocks denoted by x^Y_{ct} and individual-level shocks represented by x^Y_{it}. Examples of the former include changes in the returns to education, the latter may be promotions, bonuses or layoffs. Cohort and individual health shocks are denoted by x^H_{ct} and x^H_{it} respectively, with the impact on income being determined by the coefficients ξ_1, ξ_2, ϕ_1 and ϕ_2. All our estimation equations further include demographic terms which remain suppressed in this section for ease of notation.

This set up is effectively a version of the stochastic process in Blundell and Preston (1998) where individual health shocks are explicitly taken into account. These authors also show how equation (3.1) can be derived from a decomposition of income into a permanent and a transitory component. Essentially, permanent shocks are represented by the ζ's and transitory shocks are denoted by the ε's. The next section provides more discussion on the nature and role of these shocks.

In terms of generality, our specification fares well compared to the literature on income and consumption. The only feature that we do not include is a moving average component in the transitory errors as done in Meghir and Pistaferri (2004) or Banks, Blundell, and Brugiavini (2001). Incorporating it would be straightforward, in particular it would not substantially alter any of our identification results in Section 3.2.4. However, as indicated from our results in Section 3.4.1, our parsimonious specification seems to be sufficient in order to capture the features of the process at the cohort level for annual observations. With respect to the identification problem concerning cohort, age, and time effects, we assume the latter to be zero although a time trend is captured by our empirical methodology. Results reported in Banks, Blundell, and Brugiavini (2001) show that shocks to cohort income dominate any aggregate effects so that we do not view this assumption to be problematic.

3.2.2 Stochastic Process for Individual Health

For the purposes of this section, we treat health as a unidimensional stock variable. To model its evolution over the life course, we use the process employed in Deaton and Paxson (1998a) and augment it with an individual-specific age-trend m_{ia}^H:

$$H_{it} = H_{i,t-1} + m_{ia}^H + (1-L)u_{it}^H + v_{it}^H \tag{3.4}$$

Again, we assume that the error terms can be broken down additively into several components:

$$u_{it}^H = \varepsilon_{it}^H + \varepsilon_{ct}^H + \sum_{j=0}^{q} \gamma_{1j}\varepsilon_{c,t-j}^Y + \sum_{j=0}^{q} \vartheta_{1j}\varepsilon_{i,t-j}^Y \tag{3.5}$$

$$v_{it}^H = \zeta_{it}^H + \zeta_{ct}^H + \sum_{j=0}^{q+1} \gamma_{2j}\zeta_{c,t-j}^Y + \sum_{j=0}^{q+1} \vartheta_{2j}\zeta_{i,t-j}^Y \tag{3.6}$$

for some moving-average parameter q. The age trend m_{ia}^H is included to capture slow and steady biological processes such as cell-ageing or general wear and tear. Examples of individual transitory health shocks ε_{it}^H are broken bones or influenza, which in most cases do not have important long-term effects. On the other hand ζ_{it}^H can be thought of as a permanent (incurable) shock like many cancers or AIDS. Transitory cohort level shocks ε_{ct}^H could be highly infectious diseases whereas the most important example for permanent innovations that affect entire cohorts ζ_{ct}^H would be medical progress.

The income shocks described in the preceding section are allowed to have a direct effect on health, but we impose the restriction that permanent shocks to income influence health permanently and transitory shocks to income affect health only in a transitory fashion. In this specification, transitory income shocks may influence health in time periods t until $t+q$. Period-t permanent income shocks are assumed to impact health not beyond period $t+q+1$. In Section 3.4.3 we report results for $q = 0, 1, 2$. Note that we impose the same restriction on the effects of health on income while assuming that $q = 0$. [2]

Note that we do not give (3.4) to (3.6) an interpretation that is structural in the sense that all relevant decision parameters are incorporated in a model of health production. We rather specify a model that is rich enough to incorporate the most important conceivable causation channels and arrive at a causal interpretation through appropriate exclusion restrictions. This is different from the approach taken in the consumption literature. For example, Blundell, Pistaferri, and Preston

[2] Note that we can allow for even more generality at the individual level. As long as it aggregates out it doesn't impact upon our analysis.

(2005) derive a rather similar condition on the evolution of consumption from intertemporal utility maximisation. A natural point of departure in our case would have been a version of the classic model of health capital by Grossman (1972). We do not employ such a model because research in epidemiology is richly suggestive of mechanisms linking income changes to health outcomes over which individuals have only limited control (for example through byproducts of physiological stress markers). Put differently, we think that too little is known about the causation channels running from income to health outcomes as to isolate and parameterise the relevant ones a priori. As a consequence, we stick to the general statistical modelling approach.

In our analysis we also focus on health behaviour. For these models we reinterpret H_{it} in equation (3.4) as a measure of expenditures or quantity of consumption goods, such as fruits and vegetables, cigarettes and alcohol. In these cases, the economic model of informed decision-making has received a widespread application, as an example see the review of Chaloupka and Warner (2000) with respect to smoking behaviour. Nevertheless we stick to our approach for coherency reasons. The comments made in the last paragraph still apply: It is quite general and we do not see any interesting causation channels that we miss. There is a qualifier to this. It is well known that in our type of data it is not possible to control for age, cohort, and time effects simultaneously. In our particular case, the exclusion of time effects over and above a trend seems more justified in the case of health outcomes than consumption goods. For example, a jump in the relative price of cigarettes due to a tax increase may manifest itself as a shock that goes into the same direction for all cohorts. The assumption of iid-shocks would be violated. While we cannot make our inference strategy robust to this type of shocks, our point estimates remain valid. Having said this, we note that relative prices of cigarettes in the UK experienced a rather smooth upward trend over the period under scrutiny (with the exception of two slight declines in the late seventies and late eighties). All our estimates contain a full set of cohort dummies and a flexible age trend which adequately capture such trends and since we are not interested in interpreting either the age or cohort coefficients this seems sufficient for our purposes.

3.2.3 Aggregation

We need some assumptions on the idiosyncratic shocks and the fixed effect before we proceed to the cohort level. In particular we impose the restriction that their mean depends on age only. This is equivalent to specifying an age-dependent stochastic trend with zero mean of the errors. In addition to this, we assume all shocks to be independently and identically distributed. Accordingly, we have:

$$\begin{aligned}
\varepsilon_{it}^{Y} &\sim \mathrm{iid}\left(m_a^{Y,T}, \sigma^2_{\varepsilon^{Y,i}}\right)\\
\zeta_{it}^{Y} &\sim \mathrm{iid}\left(m_a^{Y,P}, \sigma^2_{\zeta^{Y,i}}\right)\\
\varepsilon_{it}^{Y} &\sim \mathrm{iid}\left(m_a^{H,T}, \sigma^2_{\varepsilon^{H,i}}\right)\\
\zeta_{it}^{H} &\sim \mathrm{iid}\left(m_a^{H,P}, \sigma^2_{\zeta^{H,i}}\right)\\
m_{ia}^{H} &\sim \mathrm{iid}\left(m_a^{H,FE}, \sigma^2_{m^{H,i}}\right)
\end{aligned}$$

where m^T stands for constants associated with transitory shocks, m^P for those associated with permanent ones, and $m_a^{H,FE}$ for the age trend of the individual-level fixed effect. The i-superscripts on the variance terms indicate that they refer to the respective individual-level parameters.

Aggregating (3.1) and (3.4) up for individuals within each cohort, c, leads to:

$$\begin{aligned}
\text{(3.7)}\quad Y_{ct} &= Y_{c,t-1} + (1-L)(m_a^{Y,T} + \phi_1 m_a^{H,T}) + m_a^{Y,P} + \phi_2 m_a^{H,P}\\
&\quad + \zeta_{ct}^{Y} + \xi_2 \zeta_{ct}^{H} + (1-L)\varepsilon_{ct}^{Y} + \xi_1(1-L)\varepsilon_{ct}^{H} + (1-L)\nu_{ct}^{Y}
\end{aligned}$$

$$\begin{aligned}
\text{(3.8)}\quad H_{ct} &= H_{c,t-1} + m_a^{H,FE} + (1-L)(m_a^{H,T} + \vartheta_1 m_a^{Y,T}) + m_a^{H,P} + \vartheta_2 m_a^{Y,P}\\
&\quad + \sum_{j=0}^{q+1} \gamma_{2j}\zeta_{c,t-j}^{Y} + \zeta_{ct}^{H} + \sum_{j=0}^{q} \gamma_{1j}(1-L)\varepsilon_{c,t-j}^{Y} + (1-L)\varepsilon_{ct}^{H} + (1-L)\nu_{ct}^{H}
\end{aligned}$$

where ν_{ct} represents cell variation as induced by summing up the stochastic terms. It can also serve to incorporate classical measurement error (which was omitted from equations (3.1) and (3.4) purely for ease of notation). Hence, both terms will be iid with zero mean.

$$\begin{aligned}
\nu_{ct}^{Y} &\sim \mathrm{iid}\left(0, \sigma^2_{\nu^Y}\right)\\
\nu_{ct}^{H} &\sim \mathrm{iid}\left(0, \sigma^2_{\nu^H}\right)
\end{aligned}$$

The transitory shocks to income and health at the cohort level may be correlated across cohorts within periods. We only impose the standard assumption of no correlation over time and an innocuous zero-mean condition.

$$\begin{aligned}
\varepsilon_{ct}^{Y} &\sim \left(0, \sigma^2_{\varepsilon^Y}\right)\\
\varepsilon_{ct}^{H} &\sim \left(0, \sigma^2_{\varepsilon^H}\right)
\end{aligned}$$

Turning to the permanent shocks, we assume them to be distributed according to:

$$
\begin{aligned}
\zeta_{ct}^{Y} &\sim \text{iid}\left(0, \sigma_{\zeta^Y}^2\right) \\
\zeta_{ct}^{H} &\sim \text{iid}\left(0, \sigma_{\zeta^H}^2\right)
\end{aligned}
$$

where the zero mean assumption again is without loss of generality. Obviously ζ_{ct}^{Y} and ζ_{ct}^{H} need to be contemporaneously uncorrelated which does not seem to be very restrictive with respect to health. Remember that the most important example for cohort-level health shocks is probably medical progress which usually does not affect all age groups in the same way. With respect to behaviour we refer to the discussion in the concluding paragraph of the previous subsection.

In order to simplify (3.7) and (3.8), note that all m-terms including their coefficients are not separately identified. Thus they can be included in a single intercept term that depends on age only:

$$
\begin{aligned}
m_a^Y &= (1-L)(\phi_1 m_a^{H,T} + m_a^{Y,T}) + \phi_2 m_a^{H,P} + m_a^{Y,P} \\
m_a^H &= m_a^{H,FE} + (1-L)(m_a^{H,T} + \vartheta_1 m_a^{Y,T}) + m_a^{H,P} + \vartheta_2 m_a^{Y,P}
\end{aligned}
$$

Hence we can rewrite (3.7) and (3.8) as:

$$
\begin{aligned}
Y_{ct} &= Y_{c,t-1} + m_a^Y \\
&\quad + \zeta_{ct}^Y + \xi_2 \zeta_{ct}^H + (1-L)\varepsilon_{ct}^Y + \xi_1 (1-L)\varepsilon_{ct}^H + (1-L)\nu_{ct}^Y
\end{aligned} \tag{3.9}
$$

$$
\begin{aligned}
H_{ct} &= H_{c,t-1} + m_a^H \\
&\quad + \sum_{j=0}^{q+1} \gamma_{2j} \zeta_{c,t-j}^Y + \zeta_{ct}^H + \sum_{j=0}^{q} \gamma_{1j} (1-L)\varepsilon_{c,t-j}^Y + (1-L)\varepsilon_{ct}^H + (1-L)\nu_{ct}^H
\end{aligned} \tag{3.10}
$$

This is our basic system of equations which we use to identify permanent shocks to income and to investigate the relation between them and various health variables.

3.2.4 Identification

Identification of the parameters in (3.9) and (3.10) closely traces the analysis in Meghir and Pistaferri (2004) and Blundell, Pistaferri, and Preston (2005) once we make the system triangular in terms of permanent shocks. To do this, our key identifying assumption is that $\xi_2 = 0$, i.e. that permanent cohort-level shocks to health do not affect income (or are not present at all) and it merits some discussion here. [3]

We first note that the model still allows for any type of health to income causality at the individual level. Here, both permanent and transitory shocks to health may affect income in an arbitrary fashion. This covers most typically mentioned channels ranging from work disability to transitorily impaired earnings capacities such as broken bones. The aggregation to the cohort level, combined with the time-series structure we impose, permits us to abstract from these issues. We also allow for transitory health innovations at the cohort level to affect current income which takes care of infectious diseases such as influenzas that might impact on different age-education cohorts differentially.

With respect to permanent changes in cohort health, the most salient example is medical progress that affects each cohort in a specific way. The introduction of by-pass surgery could be seen as a permanent health shock for older cohorts. Our identifying assumption claims that these changes are small enough as to not affect the incomes of entire cohorts. Rather, permanent income fluctuations are driven by the factors described above such as differential returns to education. This exclusion restriction appears to be rather plausible to us, especially at younger ages. Differences in early retirement seem to be more problematic in this respect. This is the reason why we restrict ourselves to people until sixty years of age. We did robustness checks with even younger cohorts only and results were not sensitive to the cutoff age.

Having triangularised the system, we use very similar moment conditions as Meghir and Pistaferri (2004) to identify the permanent shocks. In contrast to them, however, we do not consider bounding the transitory ones.[4] We start by defining g^Y_{ct} (g^H_{ct}) as the part in the growth rate of income (health) that is not explained by age effects.

$$\begin{aligned} g^Y_{ct} &= \zeta^Y_{ct} + (1-L)\varepsilon^Y_{ct} + \xi_1(1-L)\varepsilon^H_{ct} + (1-L)\nu^Y_{ct} \\ g^H_{ct} &= \sum_{j=0}^{q+1}\gamma_{2j}\zeta^Y_{c,t-j} + \zeta^H_{ct} + \sum_{j=0}^{q}\gamma_{1j}(1-L)\varepsilon^Y_{c,t-j} + (1-L)\varepsilon^H_{ct} + (1-L)\nu^H_{ct} \end{aligned}$$

Our identifying moment restrictions are based on the autocovariances and

[3] Whilst we can't identify ξ_2 we can examine robustness to this assumption by considering the estimates we would get with values other than zero. This is done briefly as a robustness check in section 3.4.5

[4] Meghir and Pistaferri (2004) find a MA(1)-process for the transitory shock to income which gives them a system of three parameters (the variance of the transitory shock to income, the MA(1)-parameter, and the variance of measurement error) in two equations based on the first and second order autocovariances of income. With a MA(0)-process, the two variances are completely confounded. In trying to exploit the cross-covariance restrictions (5 parameters in 3 equations for the MA(0) case in the health process), we passed the limits of our synthetic cohort data.

cross-covariances of these terms:

$$E\left[g^{Y}_{ct}\left(\sum_{j=-1}^{1} g^{Y}_{c,t+j}\right)\right] - E\left[g^{Y}_{ct}\right] E\left[\sum_{j=-1}^{1} g^{Y}_{c,t+j}\right] = \sigma^2_{\zeta^Y} \tag{3.11}$$

$$E\left[g^{H}_{ct}\left(\sum_{j=-(q+1)}^{q+1} g^{H}_{c,t+j}\right)\right] - E\left[g^{H}_{ct}\right] E\left[\sum_{j=-(q+1)}^{q+1} g^{H}_{c,t+j}\right] = \left(\sum_{j=0}^{q+1} \gamma_{2j}\right)^2 \sigma^2_{\zeta^Y} + \sigma^2_{\zeta^H} \tag{3.12}$$

$$E\left[g^{H}_{ct}\left(\sum_{j=-(q+1)}^{1} g^{Y}_{c,t+j}\right)\right] - E\left[g^{H}_{ct}\right] E\left[\sum_{j=-(q+1)}^{1} g^{Y}_{c,t+j}\right] = \sum_{j=0}^{q+1} \gamma_{2j} \quad \sigma^2_{\zeta^Y} \tag{3.13}$$

We note that the terms in the sum $\sum_{j=0}^{q+1} \gamma_{2j}$ are not separately identified. Hence we can say something about the aggregate impact over time of income shocks on health, but we cannot break it down further by year. Treating this sum as one parameter γ_2, we have three parameters in three equations that can be estimated via GMM.

3.2.5 Estimation

Our estimation strategy consists of three steps. We first sketch them very briefly and then describe each one in more detail.

1. Regress the health and income variables on a suitable set of regressors capturing cohort and age effects.

2. Use the first-differenced residuals to compute the terms inside the expectations of (3.11) to (3.13).

3. Calculate $\sigma^2_{\zeta^Y}$ and $\sum_{j=0}^{q+1} \gamma_{2j}$ for $q = 0, 1, 2$ via GMM, obtain confidence intervals by bootstrapping.

With respect to the first stage regression, we choose a full set of cohort dummies and a quadratic in age interacted with sex and education. Since our cohorts will be based upon the latter two variables in addition to date-of-birth intervals (see Section 3.3), this allows for very flexible profiles across the sex and education groups as well as an additive effect for year of birth. The results were invariant to higher order polynomials in age and the respective interaction terms. In particular, this modelling strategy allows us to control for cohort specific effects determined by macro-economic conditions at birth which may impact individuals as shown by van den Berg, Lindeboom, and Portrait (2006) or Dehejia and Lleras Muney (2004). They also capture the downward trends in overall mortality and morbidity over time. We checked for robustness towards the additivity restriction on cohort effects by permitting the income and health profiles to be cohort-specific. Point estimates did not change much but standard errors rose a lot as expected for an overfitted model.

First-differencing the residuals from these regressions immediately gives us the growth rates g_{ct}^{Y} and g_{ct}^{H}. In principle we could directly run a regression in first differences. This is inhibited by the fact that our time series dimension is not long enough to make the attenuation bias of OLS in this type of regressions negligible.

We then use GMM techniques to estimate the parameters of interest. All variance terms are allowed to vary across sex and education groups. After some experimentation, we restrict the γ_2-parameters to be the same. Due to the small-sample bias that is often observed in optimal minimum distance estimation of covariance structures, we use equally weighted minimum distance (Altonji and Segal 1996). Since it is very hard to keep track of the pre-estimation error analytically, we obtain confidence intervals by using the bootstrap (500 replications). As a consequence of the nonstationarity of the model, we sample entire time series of each cohort. Inference hence rests upon the cross-sectional dimension of the estimation problem.

3.3 Data

In order to study the dynamic effects of income shocks on health, we construct a synthetic cohort dataset, based on successive years of microdata from several cross-sectional surveys. This type of data has proven to be very useful in the consumption literature because it facilitates the use of large datasets in long time series. While restricting the range of admissible models to additive ones, it permits much more generality than aggregate data. Moreover the approach allows the combination of various datasets. This is particularly important in our analysis because data on health and socio-economic variables typically have not been recorded jointly. This has changed over the last decade, but mostly for data covering the elderly and near-elderly population. Hence there are no panels currently available that allow to study dynamics for the prime-aged population.

We use English data from three repeated cross-section surveys: The Family Expenditure Survey (FES), the General Household Survey (GHS), and the Health Survey for England (HSE). We will describe these datasets in turn after some general remarks. All of them contain basic demographic information such as individual education level as measured by the age left full time education, marital status, number of persons in the household, number of children and number of children under five years of age in the household, and labour market status.

We construct cohorts based on 3-year date-of-birth intervals, sex, and whether individuals attended school beyond the compulsory schooling age or not. We consider the working-age population of 30 to 60 years of age. The upper bound is set because of cohort differences in early retirement. In our data, early retirement (to the extent it is associated with a fall in income) would show up as a negative income shock, but there may be an independent influence of retirement on health that we cannot control for (see for example Snyder and Evans (2006)). We chose the lower bound because we model household income and many people live with their parents until their mid-twenties. One may well expect that risk sharing in these households is not perfect and that there is a large deal of financial autonomy between generations. If this is the case, then cohort differences in the changes of household income at these ages would not necessarily reflect the type of income shocks that we are interested in.

Table 3.1 lists the data availability by study and year and Table 3.2 provides summary statistics of the cell sizes within each date-of-birth, education and gender cohort in each of the surveys. As documented in Table 3.2, the cell sizes are reasonably large and throughout the analysis we do not explicitly consider aggregation error in calculating the standard errors. This is common practice in the consumption literature (e.g. Browning, Deaton, and Irish (1985)). The HSE displays very small sample sizes in its first two years of existence, as can be inferred from the last two rows of Table 3.2. We check robustness to exclusion of these two years in all specifications. All data are used on an annual basis from 1978 to 2003, although the HSE and some variables in the GHS do not cover the full period. The rest of this section describes the relevant variables used from each of the datasets in turn. The main variables are on household income and consumption as well as individual-level risk behaviours and health outcomes. Descriptive Statistics for all variables can be found in Table 3.3 which also indicates which variables are used at an individual or at the household level.

In addition to the three repeated cross sections, we also use all-population mortality from the Human Mortality Database (http://www.mortality.org) which is available for the period 1978-1998. For this dataset, no covariates beyond sex and age are available. Hence we cannot base cohorts on education groups when it comes to mortality. The Longitudinal Study would have allowed for such stratification, but its education measure is based on the highest degree obtained and proved to be not comparable to the years of schooling measure.

3.3.1 The Family Expenditure Survey (FES)

The FES contains detailed information on household income and consumption. Together with the Consumer Expenditure Survey in the US, it has emerged as one of the workhorses in the consumption literature (see for example Blundell, Pashardes, and Weber (1993)). Our data cover the period from 1978 to 2003 and our selected sample comprises 148,517 individuals. Unless indicated otherwise, we use expenditure and income in logarithms and weighted by the OECD-equivalence scale

$$\text{EQUIV} = 1 + 0.6 \cdot (\#\text{ADULTS} - 1) + 0.4 \cdot \#\text{CHILDREN}.$$

We conduct sensitivity analyses with unequivalised values and demographic variables on the right-hand side, results were qualitatively similar and are not reported. All variables are deflated with a general price index to obtain January 2000 prices. The evolution of mean income and mean expenditure on non- and semi-durable items is shown in Figures 3.1 and 3.2 for every other cohort.

The familiar hump-shape over the life-cycle becomes apparent for both variables and it seems to be more pronounced for expenditure as well as for the group with low education. Values for men are slightly higher than those for women, reflecting the higher income of male single households with respect to female single ones. As expected, we find pronounced education differences. Compared to these, cohort effects appear to be moderate with younger cohorts generating more income and consumption than older ones. These effects seem to be stronger for the higher educated, which is consistent with the findings of Gosling, Machin, and Meghir (2000) for male wages. In addition to total expenditure, we also use food expenditure in order to validate our methodology.

By virtue of the detailed consumption measures in the FES we are also able to obtain expenditure on items related to risk behaviours. Specifically, we consider expenditures on fruits and vegetables, alcohol, and cigarettes. While the first two of these items suffer from the problem that expenditure changes might reflect quantity or quality changes, this is less the case for cigarettes. Compared to other products, the price differences among brands are very small at any given point in time. By using a special price index for cigarettes and assuming between-product price variation to be negligible, we are able to back out actual cigarette quantity.

3.3.2 The General Household Survey (GHS)

The GHS is a general purpose survey that has been conducted since 1971. Questions on health measures and risk behaviours have been included since 1978, our data cover the period until 2003. It was not fielded in 1997 and 1999, we explain in Section 3.4.3 how we deal with these years. Please refer to Table 3.1 in the Appendix for the availability of specific variables. We select the following health variables: Self-reported health status on a three-dimensional scale (good-fair-poor), whether respondents have a chronic disease, and whether they are limited in their activity by this disease. As an example, consider the proportion reporting good health that is shown in Figure 3.3. Consistent with many findings from the literature (Case and Paxson 2005), women report worse health than men at younger ages. However, the decline is slower, leading to little if any gender differences at age 60. Education effects seem to be even stronger. At age 30, only small differences between education groups can be seen. Proportions remain nearly constant for people in their thirties, the subsequent decline begins earlier and is stronger for persons with low education. There are no apparent cohort differences in either direction.

Furthermore, the GHS includes some self-reported data on risk behaviours. We make use of questions on the quantity of cigarettes smoked and a six-category question regarding the amount of alcohol consumed, running from zero consumption to heavy drinking. With respect to the former, we use whether or not the respondent is currently smoking and an unconditional cohort average of the number of cigarettes smoked per day. Regarding alcohol consumption, we simply average the variable across cohort members.

3.3.3 The Health Survey for England (HSE)

Further health measures are available from the HSE, albeit only in a relatively small number of years, namely from 1991 to 2003. What makes matters worse is that cell sizes are very small in 1991 and 1992 which is made clear by the last two rows in Table 3.2. Accordingly, we checked robustness to exclusion of these two years in all specifications.

The HSE was set up to monitor population health in England and it includes a physical examination by a nurse. However, each year a different subset of health outcomes is emphasised (examples include cardiovascular disease and the health of older people). Consequently, few variables are contained every year in the nurse interview. In fact, the only one we are able to use is bloodpressure. We construct a binary variable that is one if the measured bloodpressure is above 140/90 mmHg (i.e. if the systolic pressure is above 140 or the diastolic pressure is above 90 or both) or if the respondent is currently taking bloodpressure medication. The fraction of persons with such a condition is listed in Figure 3.4. The age gradient is very steep compared to other variables. There is a large gender effect with women being much less likely to suffer from such a condition. There seems to be a slight tendency for lower prevalence among the better educated. We also use an indicator variable that takes the value one only if measured bloodpressure is above 140/90 mmHg and doesn't take account of medication. The two bloodpressure variables treat medical progress in a different fashion. In the course of the analysis, we term the two variables "Bloodpressure Condition" and "High Bloodpressure", respectively. Going in the same direction, we also use a self-reported measure of whether the respondent suffers from cardiovascular disease. A major reason to put an emphasis on diseases of the cardiovascular system stems from research in epidemiology which suggests causal mechanisms running from stress (that may be caused by sudden income drops) to cardiovascular markers which operate rather quickly as compared to many other diseases such as cancer or liver disease. See Brunner (1997) for a review and Steptoe, Willemsen, Owen, Flower, and Mohamed-Ali (2001) for a specific experiment.

Besides bloodpressure, we obtain some self-reported measures from the HSE. In particular, similar questions to those in the GHS are asked and we can use them to complement the years 1997 and 1999 although there is an issue with respect to comparability (see section 3.4.3). The longstanding illness question is further broken down by type of disease, of which we use the prevalence of of respiratory diseases and the just-described measure of cardiovascular disease. Again, epidemiological research suggests that smoking, a risk factor we consider below, has more rapid effects on respiratory diseases than lung cancer, for example see Peto, Darby, Deo, Silcocks, Whitley, and Doll (2000). The last health measure we include in the analysis is the mean score on the 12-item General Health Questionnaire, a widespread instrument to measure mental health status. The questions call for yes or no answers and the individual score is simply the number of affirmations. A higher score indicates poorer mental health.

The first thing to establish is that the ubiquitous health-SES gradient is indeed present in our data. From 1997 on, the HSE contains an income measure that we can use to estimate the conditional correlation at the micro level. Table 3.4 shows the results of regressions of the various health variables on log household income and a set of controls at the level of the individual. The general pattern that emerges is that persons with higher income have better health, whatever the health measure we consider, consistent with many findings from the literature.

3.4 Results

We present our results in four stages. First, we describe the income process and the variance of permanent income shocks. We then show the effect on mortality, health outcomes and risk behaviours. The concluding part of a section describes the relation of our approach to other works in the literature.

3.4.1 The Variance of Income Shocks

Table 3.5 lists the autocovariances of the unexplained growth rate of income up to order 5. Estimation is only based on equation (3.9) and moment condition (3.11). All numbers in this table and Table 3.6 are multiplied by 1000 for easier reading. An MA(0)-process implies that $E[g_t^Y g_{t-1}^Y] < 0$ and that all autocovariances of higher order are zero. The data seems to point to such a model as all covariances two years apart or more are both small in magnitude and not significantly different from zero. The exception is for men as the covariance for higher order lags rises somewhat again in absolute value. However, they are rather imprecisely estimated. In any case, Table 3.5 suggests that our parsimonious model without a moving average process for the transitory shock to income is sufficient to capture its dynamics at the cohort level. This is consistent with the analyses of Banks, Blundell, and Brugiavini (2001) on the same data, albeit over a slightly earlier time period. They find that an MA(1)-process fits the quarterly income data from the FES very well. Hence one would expect an MA(0)-process for the yearly data.

Table 3.6 presents the estimated variance of the permanent shock to income, applying the methodology described above. All variances are significantly larger than zero. The results correspond to annual shocks with a standard deviation between 3 to 5% of income. The magnitude of the shocks is larger for individuals with low education, although the difference is not statistically significant. Gender differences in the point estimates are only visible for the highly educated with men displaying a somewhat higher variance of income shocks. However, remember that our measure is household income and differences are caused only by single households which constitute around 25% of the total sample. To get a very rough idea about the magnitude of these shocks, it is useful to compare them to the findings reported in Meghir and Pistaferri (2004). They estimate the variance of permanent shocks to individual income in the US using the PSID. The magnitude of the shocks we find is somewhere around 5% of their estimates. Obviously one should not overstate cross-country comparisons, but this seems to be a reasonable magnitude of cohort-wide income shocks as compared to individual ones. [5]

3.4.2 Permanent Income Shocks and Mortality

Table 3.7 presents our estimates of γ_2 (the effect of permanent income shocks) on mortality outcomes in nine different specifications – allowing for three different time lags for the income effects and three different age trend specifications. The first thing to note is that a sufficiently flexible specification of the age trend is extremely important when it comes to mortality: Moving from the quadratic to the cubic specification all coefficients change their sign and confidence intervals do not overlap. On the other hand, the cubic trend seems to be sufficiently general as a comparison with the results on the quartic trend shows.

In general, we find that an increase in income leads to an increase in mortality. When we allow no lagged effect of income, a one percent increase in income leads to about 0.7 to 1 more deaths per 100,000 persons among the prime aged population in any given year. By and large, we obtain similar results when we extend the

[5] We also estimated the income process on different subset period and found no statistically significant differences.

lag with which income shocks can affect health although the null hypothesis of no causation is just not rejected at the 95% confidence level (P-value 0.058).

3.4.3 Permanent Income Shocks and Health

Table 3.8 contains our basic results for the relation between permanent income shocks and health outcomes.[6] Except for mental health, all dependent variables are measured prevalence rates in the population. These can be thought of as stemming from a linear probability model with our unidimensional health stock on the right hand side. Since we use household income in logarithms, the coefficients measure absolute changes in prevalence rates with respect to percentage changes in income. Taking the first entry of Table 3.8 as an example, a one percent shock to permanent income would decrease the fraction of people reporting good health by 0.065 percentage points. In other words a one standard deviation increase in income would reduce the fraction of individuals reporting good health from about 60% to 59.7%. Moreover, the effect is not significantly different from zero. This effect is obtained assuming that income has only an instantaneous effect on health. When we additionally allow a one year delay, we still find an insignificant negative effect on the probability of reporting good health. With a two year lag, the effect of an increase in income becomes slightly beneficial to health, but with a small magnitude: a one percent increase in income increases the probability of reporting good health by 0.02 percentage points. This effect is not significant either.

We obtain a symmetric effect for the probability of reporting poor health. At lags zero and one year, an increase in income increases the proportion of individuals reporting poor health while after two years, the proportion is reduced. Once again, all the effects are not significantly different from zero. Income does not appear to affect other measures of health such as the proportion of individuals with longstanding illness or limiting illness. If anything, a one percent increase in income is associated to a 0.02 percentage point increase in these proportions.

Rows 5 to 7 in Table 3.8 present the effect on objective health outcomes relating to cardiovascular health: The proportion of individuals with blood pressure condition, with high blood pressure or with cardiovascular diseases.[7] We do not find any significant effects of income on these health outcomes either. This is im-

[6] In Section 3.3 and Table 3.1 we noted that the GHS was not fielded in 1997 and 1999. The HSE also contains a question on self-assessed health, but the answer is recorded on a five-point scale (very good, good, fair, poor, very poor). We collapse the “good" and “poor" categories into one each and use a dummy variable in the first estimation stage for these years. With respect to the longstanding and limiting illness variables, we also use the values from the HSE and a dummy. In all four cases, results were largely insensitive to using either the values from HSE for all years beginning in 1993; or to cutting off the analysis in 1996.

[7] One year each is missing for blood pressure and mental health (see Table 3.1), we interpolate the adjacent values linearly for these years.

portant as one may think that this set of diseases could be linked to changes in socio-economic position due to stress, see the references listed in Section 3.3.3 among others. Neither can we reject the null hypothesis that income shocks have no effect on respiratory diseases: All point estimates in row 8 of Table 3.8 are negative but confidence intervals are too large to make statistically meaningful statements.

Mental health is measured as the simple sum of the scores on the 12-item general health questionnaire. A higher score indicates poorer mental health. We find that a positive shock to permanent income leads to poorer mental health. The confidence interval is rather large, but we can exclude the possibility of no effects at the 5% level. Looking at more lags, the coefficient appears ill identified and we cannot exclude the possibility that income has no effect on mental health.

Our strategy here has been to extract permanent shocks to income directly from income data. The life-cycle model predicts consumption to react stronger to permanent income innovations in income than to transitory ones. An alternative strategy to use as sensitivity analysis is to use data on expenditure to filter out permanent shocks as described in Blundell and Preston (1998) and Blundell, Pistaferri, and Preston (2005). We operationalise this idea by using total expenditure instead of income in the estimation procedure. The three moment conditions (3.11) to (3.13) remain unchanged even though we do not need to take care of transitory consumption movements if the life-cycle model is taken to be literal. This is because the presence of measurement and aggregation error introduces a transitory component in the consumption process anyhow. Results based on this strategy are largely comparable to our estimates from the income process. In particular, none of the coefficients changed significantly in magnitude. [8]

3.4.4 Permanent Income Shocks and Behaviour

We now turn to the effect of income on behaviour and particularly on Risk factors for health. We first look at the effect of permanent shocks on total (non durable) consumption. While total consumption is not thought to be directly linked to health, it is a useful benchmark to compare to more health related consumption decision.

The first row of Table 3.9 presents the elasticity of total expenditure with respect to permanent income shocks. A one percent increase in income is associated with a 0.4 to 0.6 percent increase in total expenditure, depending on the order of the moving average. All effects are significantly different from zero. Our estimates are in the same range as Blundell, Pistaferri, and Preston (2005) on

[8] The tables corresponding to Tables 3.8 and 3.9 calculated using this methodology are available from the authors upon request.

individual-level US data. In contrast with the findings in the previous section, permanent shocks lead to sizable changes in behaviour. We conclude that the income data contains pertinent variation (as Table 3.6 indicated as well) and that the lack of any significant effect of income on health is not due to lack in variation in the explanatory variable. Therefore, these results establish the ability of our empirical methodology to identify changes at the cohort level.

We now focus on a subset of total expenditure, the expenditure on food. A one percent increase in income leads to a 0.20 to 0.25 percent increase in food expenditure, although estimates are just shy of being significant. The elasticity of food is about half of the elasticity of total expenditure, suggesting that households are able to smooth food consumption. The third row presents the effect of permanent shocks to income on expenditure on fruits and vegetables. The effect varies with the lag order for the income effect, with an elasticity between 0.45 and zero but we cannot exclude the possibility that income has no effect. More importantly, point estimates have roughly the same magnitude as the ones on food consumption and we can never tell them apart. Hence small income shocks do not seem to cause individuals to switch to a healthier diet.

We now turn to risky behaviours such as smoking and alcohol consumption. We find a significant impact of permanent income shocks on cigarettes smoked. [9] We get similar results when we use self-reported quantities instead of quantities inferred from expenditures. This is reassuring as these two measures come from different data sets and reporting schemes. Overall, a one percent increase in income increases the unconditional number of cigarettes smoked per day by about 0.06 to 0.1. We also get significant results for the probability of being a smoker, although the magnitude of the point estimates seems a bit large as to have much confidence in this result.

Rows 7 and 8 of Table (3.9) present the results for alcohol consumption. Looking at the self-reported quantity in row 7, we find a significant effect if looking at contemporaneous effects only. Point estimates decrease in size and fail to be significant for the more general models. Turning to the elasticity of alcohol expenditure, we find it to be around 0.5. As we consider a delayed response to income shocks, the coefficients appear to be less stable with large standard errors.

To summarise, we find some evidence that permanent income shocks affect behaviour and particularly health behaviour. Broadly speaking, permanent increases in income appear to lead to poorer health behaviour.

[9] An exception is when we select a lag of one year which gives an instable parameter.

3.4.5 Robustness of Results

The results presented above rely on the identification assumption that variations in permanent income at cohort level are not caused by permanent shocks to cohort health, i.e. $\xi_2 = 0$. While we believe there is a plausible justification for this assumption, we can explore the robustness of our results to this central assumption, i.e. for values of $\xi_2 > 0$.

In this case the effect of income shocks on health can be expressed as:

$$\sum_{j=0}^{q+1} \gamma_{2j} = \frac{E\left[g_{ct}^{H}\left(\sum_{j=-(q+1)}^{1} g_{c,t+j}^{Y}\right)\right] - E\left[g_{ct}^{H}\right] E\left[\sum_{j=-(q+1)}^{1} g_{c,t+j}^{Y}\right] - \xi_2 \sigma_{\zeta^H}^2}{E\left[g_{ct}^{Y}\left(\sum_{j=-1}^{1} g_{c,t+j}^{Y}\right)\right] - E\left[g_{ct}^{Y}\right] E\left[\sum_{j=-1}^{1} g_{c,t+j}^{Y}\right] - \xi_2^2 \sigma_{\zeta^H}^2}$$

Note that when $\xi_2 = 0$, we get back to equation (3.13). The extent of the bias of the effect of income on health depends both on the magnitude of the reverse effect but also on the covariance between income, lagged income and health. The sign of the bias is a priori ambiguous.

Table 3.10 displays the effect of income on health for $\xi_2 = 0$ (baseline), for $\xi_2 = 0.1$ and for $\xi_2 = 1$, which means that a 10 percentage point rise in the proportion reporting good health would lead to an 1 (respectively 10) percent increase in cohort income. The rest of the formula is calculated using moments from the data and setting $\sigma_{\zeta^H}^2 = 0.1$ (estimated from the data). The baseline case $\xi_2 = 0$ simply reproduces the results from the first column of Table 3.8.

Our results are surprisingly robust to a relaxation of our main identifying assumption. For $\xi_2 = 0.1$, the coefficients are remarkably similar to the baseline estimates. For larger effects of cohort health on cohort income ($\xi_2 = 1$), the change in the coefficients are larger, especially for blood pressure conditions. However, if anything estimates reinforce our previous results in the sense that, from the point estimates at least, an increase in income appears to worsen health.

3.4.6 Relation to the Literature

We deem it important to compare our methodology and results to those previously found by authors who looked at mortality, health, and health behaviour with some sort of aggregate data. In particular, we think of the work of Deaton and Paxson (1998b, 2001, 2004), Ruhm (2000, 2003, 2005, 2006), and Ruhm and Black (2002). The main differences relate to the question of how the relationship between income (or macroeconomic conditions) and health is restricted by the various modelling approaches. Common to all studies is that causality running from health (behaviours) to income is not modelled. It is useful to reformulate our model of how health evolves over the life cycle (3.10) as:

$$(3.14)\quad H_{ct} \;=\; H_{c,0} + \sum_{s=0}^{t}\left(m^{H}_{s-c} + \gamma_2\zeta^{Y}_{c,s} + \zeta^{H}_{c,s}\right) + \gamma_1(\varepsilon^{Y}_{c,t} - \varepsilon^{Y}_{c,0}) + \varepsilon^{H}_{c,t} - \varepsilon^{H}_{c,0}$$

We suppress all lagged effects of the income shock on health and moving average parameters in the transitory terms as well as the measurement error terms for ease of notation. The formulation of the age trend makes use of the fact that $a = t - c$. This shows that current health is a function of some initial condition, the difference between the current and initial transitory effects, and the accumulation of the age trend, income and health shocks. From this equation (and the equivalent one on the income process) we move on to identify the parameters of interest from deviations in the cohort growth rates of income and health from their projections.

Deaton and Paxson construct datasets similar to ours for the United States (and also England and Wales in the 2004 paper). Their basic model [10] looks as follows in our notation:

$$H_{ct} = m_a + \beta_1 t + \beta_2 Y_{ct} + \beta_3 \cdot X_{ct} + u_{ct}$$

where X_{ct} represents a vector of covariates. H_{ct} is operationalised by using the log odds of mortality. Identification is then based on cohort differences in income and mortality conditional upon a nonparametric age profile, the linear time trend, and the other covariates. This strategy does not allow for cohort fixed effects that are correlated with the residuals. A cause for this could be long-term effects of macroeconomic conditions at birth on mortality (van den Berg, Lindeboom, and Portrait 2006). More general than such a fixed effect is the question of whether or not mortality may experience persistent shocks. If this is the case, then

$$\mathbb{E}[u_{c,t+s}|\mathbb{I}_{c,t}] \neq 0, \quad \text{some} \quad s > 0$$

where $\mathbb{I}_{c,t}$ represents the information available in period t. The relevance of this problem depends on the persistence of the process. In our nonstationary formulation it holds for all $s > 0$ and is potentially serious. If shocks exhibit only limited persistence, it is probably not too important in practice.

In order to compare the two approaches on our data, we ran a mortality regression based on the following model:

$$(3.15)\qquad H_{ct} \;=\; m_a + \beta_c + \beta_2 Y_{ct} + \beta_3 \cdot X_{ct} + u_{ct}$$

where β_c represents the coefficient on a cohort dummy, the age trend is interacted with gender and allowed to be quadratic (cubic, quartic), X_{ct} contains the educa-

[10] The authors also investigate a direct effect of income inequality on mortality. We do not consider the part of their analysis that focusses on this issue here.

tion measure. Y_{ct} is again specified as log equivalised household income. Overall, results for β_2 go into the same direction as the ones for γ_2, but there are important differences. If the age trend is specified to be quadratic, results show a significantly negative correlation of income and health. However, this again seems to be an artefact of specifying the age trend in a too restrictive fashion. Turning to higher order polynomials, we get point estimates of $\hat{\beta}_2 = 0.1$ which is about a tenth of the magnitude observed for the coefficient on the permanent shock. It is never significant. This is consistent with the results reported in Deaton and Paxson (2004) for a specification that is just slightly different. It is difficult to pin down the exact reason for this discrepancy. Either it could be the importance of the focus on permanent shocks and the filtering out of transitory movements and measurement error; or else the coefficients based on (3.15) are biased due to reasons laid out in the preceding paragraph. In any case, we conclude that the more involved strategy employed here is not a futile complication of matters.

Another strand of the literature has investigated the impact of macroeconomic fluctuations on aggregate mortality (Ruhm 2000), on mortality from heart disease (Ruhm 2006), on various health measures (Ruhm 2003), on drinking behaviour (Ruhm and Black 2002) and on several other risk behaviours and markers of unhealthy behaviour(Ruhm 2005). The identification strategy in this work has been based on deviations in state unemployment rates from the (US) national average, controlling for a time trend and population (in case of the mortality studies) or individual (health and health behaviour studies) characteristics. This research has shown that for the population above 20 years of age, most of these variables move in accordance with employment rates. The interpretation was that unhealthy behaviour is one of the factors underlying the procyclical variation in mortality rates.

While these results are broadly consistent with our findings in this paper, we want to highlight some important differences. The results in the mortality studies (Ruhm 2000, 2006), have been based on aggregate data. It is well known that this type of data does not permit models that are nonlinear in the *variables*. While this is probably not an issue if employment rates are the dependent variable, it does pose a problem if the interest is in the effect of income. The function linking income and health is typically thought to be concave. In our particular case, it turned out to be very important indeed to specify household income in logarithms as opposed to levels and we can aggregate such variables exactly from the underlying microdata. Another point concerns the use of log mortality as the outcome of interest which underlies most specifications in the two just-cited papers. While this may be important to capture the aggregate relationship well, it inhibits the interpretation of the results at the individual level. Finally, macroeconomic conditions may affect cohorts in a very different way. For example, unfavourable conditions at the beginning of the working life may impact negatively upon all future labour market outcomes (Raaum and Røed 2006), thereby constituting a permanent shock; whereas they may affect older cohorts' incomes only transitorily.

The only contrasting results and interpretations are found with respect to health outcomes as contained in Ruhm (2003). The basic identification strategy is the same as the one for the mortality studies, but individual-level data from repeated cross sections is used. Unfortunately, no results are reported for the overall prime-aged population which renders specific comparisons difficult. In the cases where analysis is restricted to a similar age range as the one considered here the sample is also restricted to the working population introducing a problem of sample selection. Furthermore, it is not quite clear what kind of effect is picked up by state-level unemployment rates if the sample is restricted to employed persons. Finally, the data analysis is cross-sectional in the sense that no individual-level fixed effects are controlled for. This relates immediately to the sample selection issue. Results based on these data in the US show a procyclical variation of health.

We cannot detect such an effect. On the other hand, we are rather confident that we would be able to pin it down if it were there. Our conclusion is that there may not be much instantaneous causation running from risk behaviours to health outcomes and on to mortality. Maybe one would expect the most direct effects with respect to smoking. However, if we look at respiratory diseases (which are expected to react much quicker than cancers), if anything, we find a countervailing effect. In our view, procyclical mortality is more likely to be driven by work-related accidents than by health and health behaviours. Given the indirect nature of this evidence, more research is needed to disentangle the precise causation channels.

3.5 Conclusions

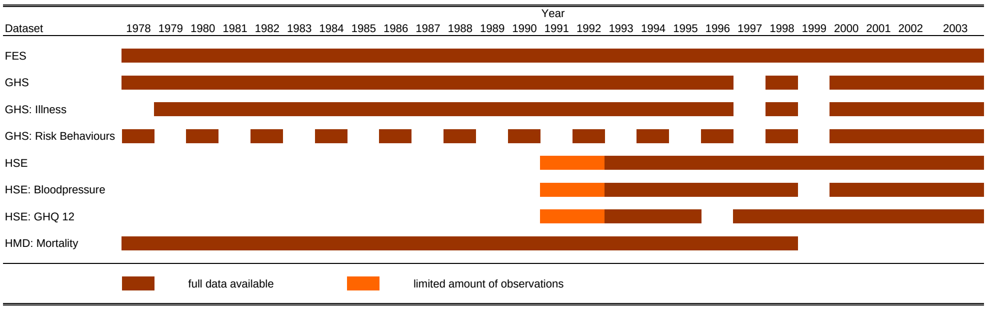
In this paper we estimate the effect of permanent shocks to income on health and health behaviour. We use a long time series of cohort data from multiple surveys which covers the life-cycle and report detailed information on income and health (both subjective and objective measures) as well as health behaviour. We exploit the fact that at the cohort level, over the eighties and nineties, there have been sizable changes in income mainly related to changes in the macro-economic environment (skill biased technological change, changes in the return to education and experience, increased competition). Our identification scheme relies on the fact that these changes are not due to changes in cohort health. While we allow for reverse causation at the individual level, as well as a rich stochastic structure, we aggregate it out at cohort level to focus only on the effect of these observed income shocks on cohort health.

As in Ruhm (2000), we find some evidence that income and mortality are related, although in our case we decompose income and filter out the transitory component. While we find some evidence that permanent income shocks lead to poorer health behaviour, we find no evidence that it affects directly any of our health measures. Despite sizable shocks to income at cohort level during our period of analysis, which can be empirically linked to changes in consumption expenditures, we fail to find any changes in (cohort) health, even if we allow those shocks to affect health with a lag. Moreover, the point estimates often hint that an increase in income leads to poorer health, except for blood pressure. This is in contrast with micro, cross-sectional evidence, where higher income is related to better health. Taken literally, our results suggest that this observed “gradient" does not stem from income innovations of the magnitude we observe at the cohort level.

Reconciliation of this lack of evidence for time series causation with the strong correlations observed in many studies, even those which control for many possible covariates, is not really necessary given the often limited extent to which the latter studies can isolate exogenous changes in income or economic resources. The results of Lindahl (2005) and Meer, Miller, and Rosen (2003), using lottery gains and unanticipated bequests respectively to exploit exogenous changes in incomes at the individual level, warrant more of a comparison. One obvious although rather uninteresting point is that different relationships may of course be obtained in different institutional settings, or for different subgroups of the population. An alternative factor may be that the relationship of health to income shocks could be non-linear or even discontinuous, with the effect that large changes in resources, of the nature used in these two studies, yield more substantial (and more identifiable) health effects than the smaller changes in cohort averages that we use here, perhaps through channels such as changes in neighbourhoods and peer groups or major changes in housing or lifestyle more generally. Finally we cannot rule out effects of income shocks that take longer to feed through into health outcomes than we have allowed for in the time-series processes estimated here (up to three years). Nevertheless, from a policy perspective our results seem particularly relevant, suggesting that the kind of changes to incomes that might be expected from redistribution towards particular age-education cohorts of working age adults are unlikely to lead to improvements in health at least in the short to medium run.

3.6 Tables

Tab. 3.1: Data Availability by Year



Tab. 3.2: Cell sizes after aggregation

Dataset	Mean	Median	First Percentile	Minimum	Total Number of Individuals
FES	137.3	134.5	65	55	148,517
GHS	257.3	221	84	58	277,084
HSE	149.6	140.5	25	18	80,541
HSE, 1993+	164.0	152	62	51	76,943

Tab. 3.3: Descriptive Statistics Averaged Across Cohorts

Men	Low Education				High Education			
	Age 30	Age 40	Age 50	Age 60	Age 30	Age 40	Age 50	Age 60
Education (FES) *)	0.555	0.489	0.476	0.589	0.445	0.511	0.524	0.411
Education (GHS) *)	0.507	0.469	0.478	0.563	0.493	0.531	0.522	0.437
Education (HSE) *)	0.521	0.440	0.430	0.538	0.479	0.560	0.570	0.462
Log Income (HH)	5.008	5.040	5.166	5.045	5.355	5.347	5.490	5.374
Log Total Expend. (HH)	4.612	4.672	4.789	4.625	4.833	4.894	5.023	4.980
Log Food Expend. (HH)	3.001	3.116	3.186	3.231	3.047	3.188	3.260	3.309
Log FruitVeg Expend. (HH)	0.741	0.904	1.031	1.085	1.088	1.175	1.357	1.382
Cigarettes (FES, HH)	2.167	2.113	2.286	2.323	2.042	1.954	1.956	2.134
Log Alcohol Expend. (HH)	1.808	1.742	2.076	1.847	1.964	1.788	2.001	1.928
Good Health	0.715	0.669	0.600	0.456	0.796	0.770	0.722	0.630
Poor Health	0.047	0.080	0.141	0.213	0.031	0.045	0.090	0.122
Longstanding Illness	0.265	0.303	0.369	0.547	0.253	0.246	0.327	0.508
Limiting Illness	0.110	0.162	0.230	0.385	0.104	0.115	0.174	0.278
Smoking Y/N	0.452	0.437	0.400	0.377	0.315	0.283	0.294	0.235
Cigarettes (GHS)	8.233	8.766	8.054	6.268	4.821	4.380	5.539	3.958
Alcohol Quantity	3.399	3.346	3.259	3.068	3.408	3.331	3.308	3.149
Bloodpressure Condition	0.243	0.278	0.488	0.654	0.243	0.220	0.423	0.624
High Bloodpressure	0.250	0.247	0.436	0.574	0.244	0.217	0.381	0.557
Cardiovascular Condition	0.017	0.041	0.100	0.218	0.007	0.037	0.080	0.176
Respiratory Condition	0.075	0.060	0.056	0.091	0.069	0.077	0.059	0.043
Mental Cond. (GHQ 12)	1.271	1.284	1.285	1.358	1.551	1.373	1.190	0.959
Mortality Rate / 1000 persons (both education groups)	0.908	1.802	5.180	15.736				

Women	Low Education				High Education			
	Age 30	Age 40	Age 50	Age 60	Age 30	Age 40	Age 50	Age 60
Education (FES) *)	0.513	0.471	0.491	0.542	0.487	0.529	0.509	0.458
Education (GHS) *)	0.532	0.456	0.534	0.591	0.468	0.544	0.466	0.409
Education (HSE) *)	0.493	0.436	0.479	0.510	0.507	0.564	0.521	0.490
Log Income (HH)	4.836	5.050	5.160	4.902	5.272	5.343	5.498	5.272
Log Total Expend. (HH)	4.495	4.682	4.788	4.543	4.786	4.912	5.038	4.903
Log Food Expend. (HH)	3.011	3.130	3.248	3.224	3.091	3.227	3.274	3.332
Log FruitVeg Expend. (HH)	0.767	0.907	1.139	1.084	1.105	1.236	1.351	1.422
Cigarettes (FES, HH)	2.134	2.143	2.270	2.271	1.937	1.864	2.029	2.187
Log Alcohol Expend. (HH)	1.521	1.739	1.963	1.671	1.697	1.692	1.993	1.714
Good Health	0.613	0.583	0.498	0.436	0.683	0.705	0.666	0.569
Poor Health	0.093	0.135	0.173	0.184	0.071	0.074	0.095	0.133
Longstanding Illness	0.280	0.314	0.408	0.545	0.242	0.260	0.321	0.510
Limiting Illness	0.124	0.190	0.250	0.341	0.098	0.124	0.175	0.308
Smoking Y/N	0.460	0.443	0.409	0.376	0.253	0.247	0.231	0.242
Cigarettes (GHS)	7.294	7.321	6.526	5.087	3.089	3.194	3.039	3.226
Alcohol Quantity	2.725	2.834	2.644	2.422	2.919	2.906	2.922	2.709
Bloodpressure Condition	0.085	0.131	0.354	0.646	0.064	0.123	0.337	0.548
High Bloodpressure	0.077	0.105	0.294	0.547	0.062	0.108	0.306	0.449
Cardiovascular Condition	0.030	0.046	0.073	0.208	0.012	0.033	0.088	0.183
Respiratory Condition	0.101	0.073	0.101	0.094	0.084	0.061	0.084	0.100
Mental Cond. (GHQ 12)	1.832	1.600	1.812	1.856	1.768	1.578	1.622	1.336
Mortality Rate / 1000 persons (both education groups)	0.475	1.189	3.305	8.865				

Note: *) Data from 1991-2000 only to ensure comparability across datasets. All other numbers are based on total number of observations for each variable, see Table 3.1. HH: Variable measured at the Household Level, equivalised by OECD scale.

Tab. 3.4: Regressions Showing the Correlations Between Health and Income.

Dependent Variable	Coef.	Std. Err.	N
Self-Reported Good Health	0.415**	0.010	36,641
Self-Reported Poor Health	-0.491**	0.014	36,641
Longstanding Illness	-0.213**	0.009	36,639
Limiting Illness	-0.332**	0.009	36,639
Bloodpressure Condition	-0.067**	0.012	23,871
High Bloodpressure	-0.053**	0.011	27,567
Cardiovascular Condition	-0.130**	0.013	36,639
Respiratory Condition	-0.092**	0.012	36,639
Mental Condition (GHQ 12)	-0.535**	0.018	35,313

Note: Coefficients of simple probit regressions of health outcomes on household income in logarithms (Mental Condition: OLS). All specifications further include year dummies as well as sex and a two-dimensional education measure, both interacted with a quadratic age trend. ** significant at the 5% level.

Tab. 3.5: The Autocovariances of the Unexplained Growth of Income

Order	Low Education Men	Low Education Women	High Education Men	High Education Women	NT
0	6.50**	7.25**	5.91**	6.12**	944
	(5.18 , 8.04)	(6.08 , 8.48)	(4.67 , 7.03)	(5.37 , 6.68)	
1	-2.38**	-2.70**	-2.22**	-2.54**	888
	(-3.18 , -1.70)	(-3.60 , -1.82)	(-2.95 , -1.50)	(-3.36 , -1.83)	
2	-0.41	0.07	-0.27	0.57	832
	(-2.13 , 1.02)	(-1.32 , 1.22)	(-1.16 , 0.80)	(-0.42 , 1.67)	
3	-0.19	-0.07	0.93	-0.22	776
	(-0.92 , 0.49)	(-0.98 , 0.89)	(-0.15 , 1.86)	(-1.39 , 0.72)	
4	0.89	-0.12	-1.30**	-0.22	720
	(-0.26 , 2.28)	(-1.22 , 0.93)	(-2.19 , -0.30)	(-1.32 , 0.94)	
5	-1.00**	-0.12	0.80	0.07	664
	(-1.75 , -0.20)	(-1.42 , 1.00)	(-0.49 , 1.90)	(-1.06 , 1.16)	

Note: Bootstrapped 95% confidence intervals in parentheses, ** significant at the 5% level. All values scaled up by 1000.

Tab. 3.6: The Variance of the Permanent Shock to Income

Time Period	Low Education Men	Low Education Women	High Education Men	High Education Women	NT
1978 – 2003	2.257**	2.027**	1.675**	0.872	832
	(0.60 , 4.09)	(0.43 , 3.72)	(0.75 , 2.57)	(-0.32 , 1.84)	

Note: Bootstrapped 95% confidence intervals in parentheses, ** significant at the 5% level. All values scaled up by 1000.

Tab. 3.7: The Effect of Permanent Income Shocks on Mortality.

Variable	Age Trend	Moving Average Parameter (q) 0	1	2	NT q = 0/1/2
Mortality	Quadratic	-0.786** (-2.118 , -0.076)	-1.108** (-3.202 , -0.136)	-1.013** (-2.204 , -0.091)	320 / 272 / 224
Mortality	Cubic	0.721** (0.329 , 1.617)	0.729 (-0.063 , 3.629)	0.527 (-0.090 , 1.561)	320 / 272 / 224
Mortality	Quartic	1.000** (0.469 , 1.953)	1.151 (-0.082 , 6.448)	0.650 (-0.231 , 2.276)	320 / 272 / 224

Note: Mortality refers to number of deaths per 1000 persons with respect to a 1% change in household income. Bootstrapped 95% confidence intervals in parentheses.

Tab. 3.8: The Effect of Permanent Income Shocks on Health Outcomes

Variable	Time Period	Moving Average Parameter (q) 0	1	2	NT q = 0/1/2
Good Health	1978-2003	-0.0647 (-0.2189 , 0.1111)	-0.1683 (-0.5471 , 0.1532)	0.0235 (-0.2835 , 0.3372)	832 / 720 / 608
Poor Health	1978-2003	0.1074 (-0.0165 , 0.2564)	0.1012 (-0.0684 , 0.3454)	-0.0781 (-0.3070 , 0.2196)	832 / 720 / 608
Longstanding Illness	1979-2003	0.0228 (-0.1252 , 0.2361)	0.0595 (-0.2589 , 0.2888)	0.0592 (-0.2525 , 0.3728)	832 / 720 / 608
Limiting Illness	1979-2003	0.0264 (-0.1083 , 0.1739)	0.0750 (-0.2272 , 0.2959)	-0.0999 (-0.3416 , 0.1470)	800 / 688 / 576
Bloodpressure Condition	1991-2003	-0.1342 (-0.9238 , 0.2461)	0.4226 (-0.3116 , 1.4468)	0.1660 (-0.3655 , 0.8034)	352 / 272 / 192
High Bloodpressure	1991-2003	-0.2106 (-1.0594 , 0.4111)	0.4100 (-0.1700 , 1.2800)	0.0361 (-0.4885 , 0.7392)	352 / 272 / 192
Cardiovascular Condition	1991-2003	0.0006 (-0.1917 , 0.3149)	-0.1112 (-0.3034 , 0.1809)	0.1978 (-0.1156 , 0.8175)	352 / 272 / 192
Respiratory Condition	1991-2003	-0.2440 (-0.6579 , 0.0592)	-0.1751 (-0.5663 , 0.2026)	-0.2114 (-0.5878 , 0.0541)	352 / 272 / 192
Mental Health	1991-2003	0.037** (0.0056 , 0.6592)	0.2244 (-0.0052 , 0.8781)	0.0081 (-0.0290 , 0.7814)	352 / 272 / 192

Note: Except for mental health, all values denote changes in percentage points of prevalence rates with respect to a 1% shock to permanent income. Mental health: change in average score on the GHQ12 with respect to a 1% shock to permanent income. Bootstrapped 95% confidence intervals in parentheses, ** significant at the 5% level.

Tab. 3.9: The Effect of Permanent Income Shocks on Behaviour

Variable	Time Period	Moving Average Parameter (q) 0	1	2	NT $q = 0/1/2$
Total Expenditure	1978-2003	0.4365** (0.2289 , 0.6381)	0.4524** (0.1547 , 0.7756)	0.6491** (0.1230 , 1.0909)	832 / 720 / 608
Food Expenditure	1978-2003	0.2135 (-0.0158 , 0.4572)	0.2494 (-0.0308 , 0.6378)	0.2435 (-0.1693 , 0.6902)	832 / 720 / 608
Fruits and Veg. Expenditure	1978-2003	0.2604 (-0.0590 , 0.8441)	0.4596 (-0.0612 , 1.2263)	-0.0085 (-0.9402 , 1.0431)	832 / 720 / 608
Fraction of Smokers	1978-2003	0.0656 (-0.1485 , 0.3393)	0.3673** (0.0521 , 0.7280)	0.3778** (0.0132 , 0.9539)	832 / 720 / 608
Cigarettes via Self-Rep. Qty.	1978-2003	0.0824** (0.0475 , 0.9420)	0.0966** (0.0589 , 0.2549)	0.1025** (0.0534 , 0.3999)	832 / 720 / 608
Cigarettes via Expenditure	1978-2003	0.0631** (0.0354 , 0.1846)	0.4539 (-0.0183 , 0.9135)	0.0592** (0.0355 , 0.5185)	832 / 720 / 608
Alcohol Self-Rep. Qty.	1978-2003	0.0059** (0.0024 , 0.0125)	0.0044 (-0.0058 , 0.0374)	0.0041 (-0.0078 , 0.2533)	832 / 720 / 608
Alcohol Expenditure	1978-2003	0.5325 (-0.3041 , 1.5006)	0.4306 (-1.0875 , 34.1617)	-0.0502 (-2.0854 , 36.3326)	832 / 720 / 608

Note: Expenditure (Total, Food, Fruit and Vegetables, and Alcohol): Coefficients shown are elasticities. Smoking Fraction: Change in percentage points of prevalence rate with respect to a 1% shock to permanent income. Cigarette variables: Change in daily number of cigarettes smoked (unconditional) with respect to a 1% shock to permanent income. Alcohol Self-Reported Quantity: Change in average reported value with respect to a 1% shock to permanent income. Bootstrapped 95% confidence intervals in parentheses, ** significant at the 5% level.

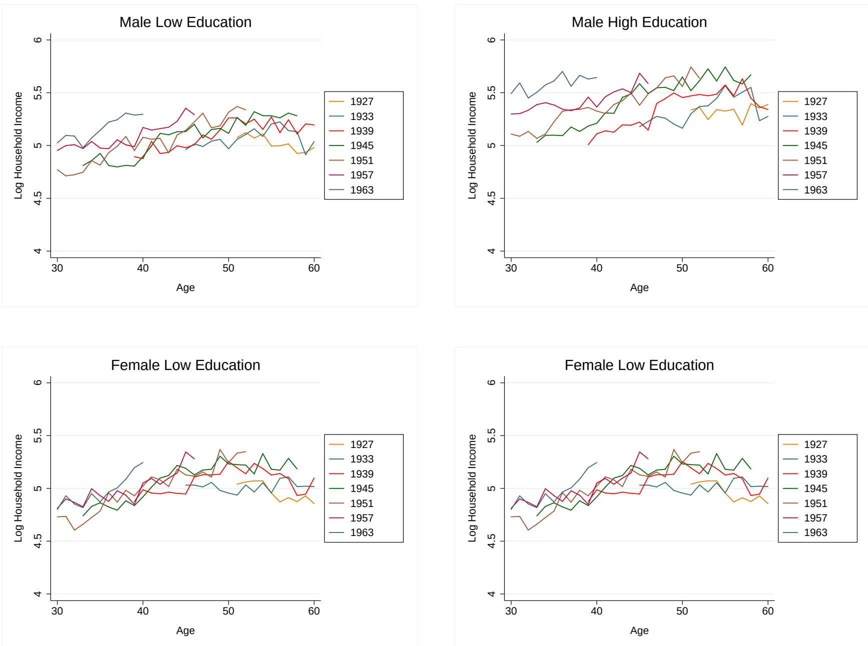
Tab. 3.10: Robustness of Results.

Variable	$\xi_2 = 0$	$\xi_2 = 0.1$	$\xi_2 = 1$
Good Health	-0.0647	-0.0717	-0.1487
Poor Health	0.1074	0.1144	0.1918
Longstanding Illness	0.0228	0.0297	0.0963
Limiting Illness	0.0264	0.0335	0.1191
Bloodpressure Condition	-0.1342	-0.1269	-0.0217
High Bloodpressure	-0.2106	-0.2033	-0.0967
Cardiovascular Condition	0.0006	0.0075	0.0716
Respiratory Condition	-0.244	-0.2372	-0.1781
Mental Health	0.037	0.0454	0.2727

Note: The calculations are described in detail in Section 3.4.5. For the health variables defined in negative terms, the sign of ξ_2 is reversed.

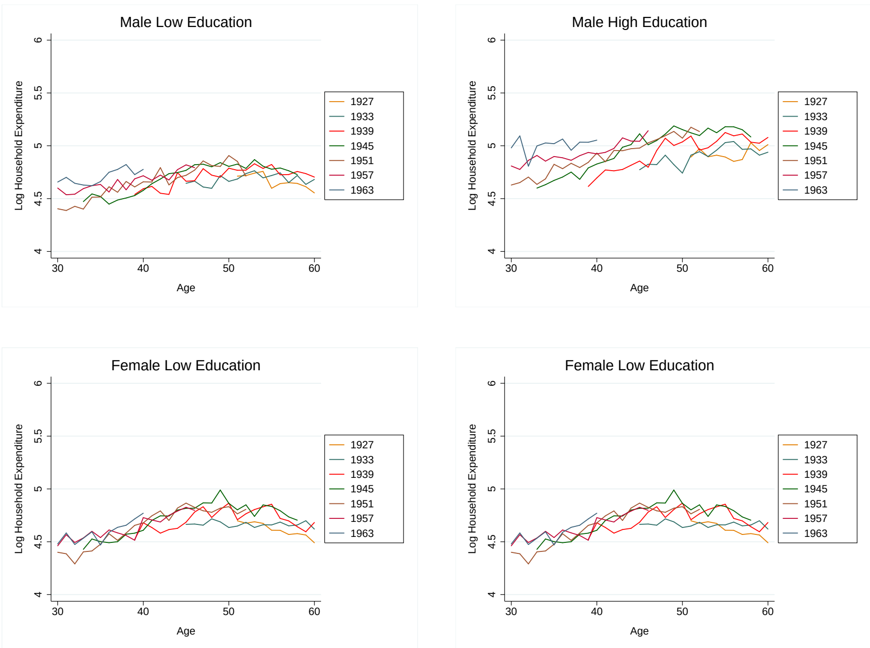
3.7 Figures

Fig. 3.1: OECD-equivalised Household Income in Logarithms by Cohort



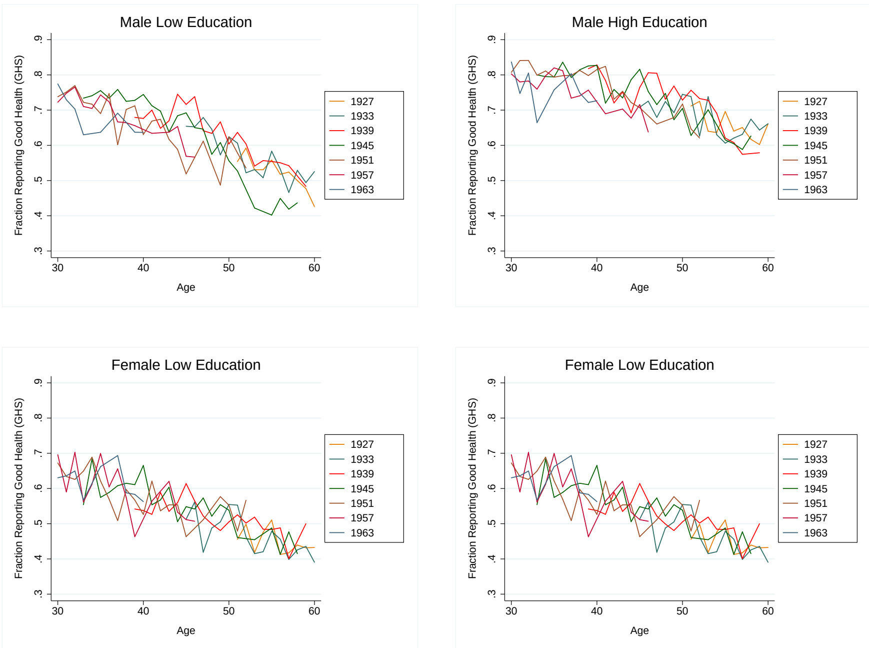
Source: FES 1978-2003, own calculations. Only every other cohort is shown for legibility reasons.

Fig. 3.2: OECD-equivalised Total Household Expenditure in Logarithms by Cohort



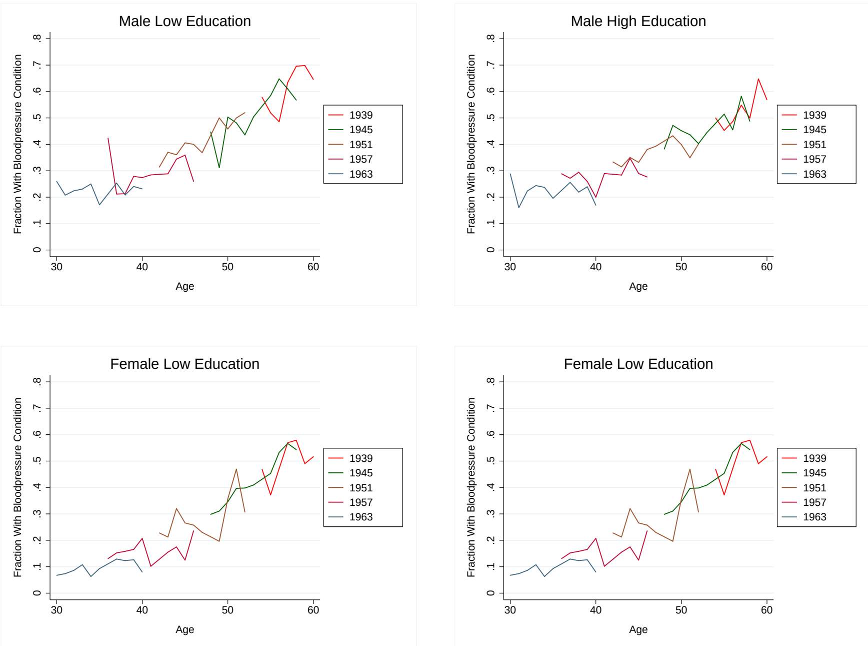
Source: FES 1978-2003, own calculations. Only every other cohort is shown for legibility reasons.

Fig. 3.3: Fraction Reporting Good Health by Cohort



Source: GHS 1978-2003, own calculations. Years 1997 and 1999 are interpolated linearly. Only every other cohort is shown for legibility reasons.

Fig. 3.4: Fraction with Bloodpressure Condition by Cohort



Source: HSE 1993-2003, own calculations. Values for 1999 are interpolated linearly. Only every other cohort is shown for legibility reasons.

BIBLIOGRAPHY

ABOWD, J. M., AND D. CARD (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–445.

ADAMS, H. P., M. D. HURD, D. MCFADDEN, A. MERRILL, AND T. RIBEIRO (2003): "Healthy, Wealthy, and Wise? Tests for Direct Causal Paths between Health and Socioeconomic Status," *Journal of Econometrics*, 112(1), 3–56.

ALTONJI, J. G., AND L. M. SEGAL (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business and Economic Statistics*, 14(3), 353–366.

BANKS, J., R. BLUNDELL, AND A. BRUGIAVINI (2001): "Risk Pooling, Precautionary Saving and Consumption Growth," *Review of Economic Studies*, 68, 757–779.

BLUNDELL, R., P. PASHARDES, AND G. WEBER (1993): "What do we Learn about Consumer Demand Patterns from Micro Data?," *American Economic Review*, 83(3), 570–597.

BLUNDELL, R., L. PISTAFERRI, AND I. PRESTON (2005): "Consumption Inequality and Partial Insurance," IFS Working Paper 04/28.

BLUNDELL, R., AND I. PRESTON (1998): "Consumption Inequality and Income Uncertainty," *Quarterly Journal of Economics*, 113(2), 603–640.

BROWNING, M., A. DEATON, AND M. IRISH (1985): "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 53(3), 503–544.

BRUNNER, E. J. (1997): "Socioeconomic Determinants of Health: Stress and the Biology of Inequality," *British Medical Journal*, 314(7092), 1472–1476.

BUCHINSKY, M. (1994): "Changes in the US Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, 62(2), 405–458.

CASE, A., D. LUBOTSKY, AND C. PAXSON (2002): "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, 92(5), 1308–1334.

Case, A., and C. Paxson (2005): "Sex Differences in Morbidity and Mortality," *Demography*, 42(2), 189–214.

Chaloupka, F. J., and K. E. Warner (2000): "The Economics of Smoking," in *Handbook of Health Economics*, ed. by A. J. Culyer, and J. P. Newhouse, pp. 1539–1627. Elsevier.

Deaton, A., and C. Paxson (1998a): "Aging and Inequality in Income and Health," *American Economic Association - Papers and Proceedings*, 88(2), 248–253.

——— (1998b): "Health, Income, and Inequality over the Life Cycle," in *Frontiers in the Economics of Aging*, ed. by D. A. Wise, pp. 431–460. University of Chicago Press, Chicago, IL.

——— (2001): "Mortality, Education, Income, and Inequality Among American Cohorts," in *Themes in the Economics of Aging*, ed. by D. A. Wise, pp. 165–170. University of Chicago Press, Chicago, IL.

——— (2004): "Mortality, income, and income inequality over time in Britain and the United States," in *Perspectives on the Economics of Aging*, ed. by D. A. Wise, pp. 247–280. University of Chicago Press, Chicago, IL.

Dehejia, R., and A. Lleras Muney (2004): "Booms, Busts, and Babies' Health," *Quarterly Journal of Economics*, 119(3), 1091–1130.

Gosling, A., S. Machin, and C. Meghir (2000): "The Changing Distribution of Male Wages in the UK," *Review of Economic Studies*, 64(4), 635–666.

Gottschalk, P., and R. Moffitt (1994): "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Paper on Economic Activity*, 2, 217–254.

Grossman, M. (1972): "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80(2), 223–255.

Lindahl, M. (2005): "Estimating the Effect of Income on Health and Mortality Using Lottery Prizes as Exogenous Source of Variation in Income," *Journal of Human Resources*, 40(1), 144–168.

Lleras Muney, A. (2005): "The Relationship Between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72(1), 189–221.

MaCurdy, T. 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.

Marmot, M. (1999): "Multi-Level Approaches to Understanding Social Determinants," in *Social Epidemiology*, ed. by L. Berkmann, and I. Kawachi. Oxford University Press, Oxford.

Meer, J., D. L. Miller, and H. S. Rosen (2003): "Exploring the health-wealth nexus," *Journal of Health Economics*, 22(5), 713–730.

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

Peto, R., S. Darby, H. Deo, P. Silcocks, E. Whitley, and R. Doll (2000): "Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies," 321, 323–329.

Raaum, O., and K. Røed (2006): "Do Business Cycle Conditions at the Time of Labor Market Entry Affect Future Employment Prospects?," *The Review of Economics and Statistics*, 88(2), 193–210.

Ruhm, C. J. (2000): "Are Recessions Good For Your Health?," *Quarterly Journal of Economics*, 115(2), 617–650.

——— (2003): "Good times make you sick," *Journal of Health Economics*, 22(4), 637–658.

——— (2005): "Healthy living in hard times," *Journal of Health Economics*, 24(2), 341–363.

——— (2006): "A Healthy Economy Can Break Your Heart," NBER Working Papers 12102, National Bureau of Economic Research, Inc.

Ruhm, C. J., and W. E. Black (2002): "Does drinking really decrease in bad times?," *Journal of Health Economics*, 21(4), 659–678.

Smith, J. P. (2004): "Unraveling the SES-Health Connection," IFS Working Paper 04/02.

Snyder, S. E., and W. N. Evans (2006): "The Effect of Income on Mortality: Evidence from the Social Security Notch," *Review of Economics and Statistics*, 88(3), 482–495.

Steptoe, A., G. Willemsen, N. Owen, L. Flower, and V. Mohamed-Ali (2001): "Acute Mental Stress Elicits Delayed Increases in Circulating Inflammatory Cytokine Levels," *Clinical Science*, 101, 185–192.

van den Berg, G. J., M. Lindeboom, and F. Portrait (2006): "Economic Conditions Early in Life and Individual Mortality," *American Economic Review*, 96(1), 290–302.

4. EXPERIMENTAL ELICITATION OF RISK PREFERENCES: SOME FURTHER STEPS TOWARDS REPRESENTATIVENESS

JOINT WITH ARTHUR VAN SOEST AND ERIK WENGSTRÖM

4.1 Introduction

Recent years have witnessed a boost in the literature on the generalisability of findings from laboratory experiments. See Harrison and List (2004) for an extensive discussion. A concern that has been recurring with a particularly high frequency is the one of subject pool bias. Extrapolation exercises are not inevitably valid and convenience samples of students do not necessarily constitute the population of interest. In some cases they may be — for example, testing predictions from industrial organisation with MBA students seems more sensible than also having assembly line workers in the laboratory — in many others they are not. The latter is for example typically the case when it comes to the elicitation of preference parameters, see for example Harrison, Lau, and Williams (2002). There are several ways of recruiting samples with more demographic variation. One rather laborious possibility is to take the laboratory from the university to the population of interest; for one of many examples taking this approach see Harrison, List, and Towe (2007). Sample sizes usually do not exceed those typically encountered in the laboratory, which may pose a problem in accounting for demographic heterogeneity. Another road along the same lines that has become available recently is to integrate experiments into existing traditional socio-economic household surveys (see Fehr, Fischbacher, von Rosenbladt, Schupp, and Wagner (2003) and, for an application in our context, Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2005)). Major advantages of this approach are that careful sampling frames are employed and that a large amount of background information on the subjects comes as a free lunch. Until now, capacity constraints in the survey instruments and the relatively high costs of personal interviews have hindered a more widespread use of this method. The two cited studies were able to use moderately-sized subsamples from the much larger German Socio-Economic Panel (N=429 and N=450, respectively). Finally, experimenters used convenience samples of Internet respondents recruited using newspaper advertising. See Lucking-Reiley (1999) and Güth, Schmidt, and Sutter (2006) for specific examples. While these approaches facilitate recruiting very large numbers of participants, there is essentially no control over the recruitment process in most cases. Selection issues may arise due to the necessity of having access to the Internet and reading a particular newspaper.

In this paper, we study a relatively standard risk preference elicitation experiment that combines the advantages of the latter two approaches. Our basic sample consists of 1,928 participants of a household survey conducted over the Internet, the CentERpanel, based on a stratified random sample of the Dutch population. Households without access to the Internet are provided access by giving them a set-top box for their TV (and a TV if necessary). For related papers using the same source of data see Donkers, Melenberg, and van Soest (2001) who analyse risk preferences using hypothetical questions, and Bellemare and Kröger (2007) who used this panel to conduct the trust game with real payoffs. In order to assess the importance of subject pool bias, we compare the results to those of parallel laboratory experiments. However, replacing the laboratory by the Internet also changes the environment, unlike the case of comparisons based on a "mobile laboratory" approach (Andersen, Harrison, Lau, and Rutström 2005). Potential differences due to demographic variation hence are confounded with environment or implementation mode effects. We address this issue from two angles. First, we introduce a treatment in the laboratory that replicates the Internet setting as closely as possible. In particular, no experimenter is present while subjects complete the experiment and there is no restriction to wait for the last person to finish before leaving the room. Second, our Internet sample is sufficiently large to restrict attention to a subsample that roughly resembles the student population in terms of age and education. If environmental factors play a role, they should become evident when comparing results from this subsample to results for the laboratory experiment.

Even in experiments designed for a broad subject pool, another selection bias may still arise from nonparticipation or incomplete participation in the experiment. This issue has not received much attention until recently, a notable exception being Harrison, Lau, and Rutström (2005). The main reason for this is that usually there is little control over the recruitment process. Experimenters typically collect some demographic information about participating subjects but the corresponding values of nonparticipants are unobserved. We have a lot of information on certain types of nonparticipants which allows us to address the selection issue in some detail.

Our results can be summarised as follows. We cannot detect any differences arising from the environmental treatment for the young and educated. When moving to the general population, the most dramatic difference is a drastic rise in the number of violations of the most basic economic principles, namely choosing dominated options and non-monotonic behaviour. Risk aversion also turns out to be higher in the overall population. Finally, we detect differential participation in the experiment and some influence on inconsistent behaviour — participants typically have observed characteristics that make them less prone to making mistakes. Risk preferences do not seem to be seriously affected by participation bias. We also find evidence of self-selection based on interest and expertise in financial matters, which calls for some caution in the interpretation of the results.

This paper is organised as follows. Section 4.2 presents the design of the experiments in the laboratory and on the Internet. Section 4.3 discusses participation and compares the composition of the Internet and the laboratory samples. Section 4.4 analyses the differences between the choices respondents have made in the two experiments, looking at errors and inconsistencies as well as implied preferences. Section 4.5 concludes.

4.2 Data and Experimental Setup

This section lays the foundations for the following chapters in providing a detailed description of our experimental design and subject pools. The next subsection illustrates the basic set-up of the experiment. Starting out from a well-established methodology for preference elicitation, the multiple price list format (MPL), we adapt it in two ways. First, we try to make it cognitively less demanding by including pie-chart representations of the probabilities in addition to the number representations. Second, the experiment is designed to elicit two preference parameters besides risk aversion, namely loss aversion (Kahneman and Tversky 1979) and uncertainty resolution preferences (Kreps and Porteus 1978). After describing the core section of the experiment, which remains constant across implementations in the laboratory and over the Internet, we point out the differences. Specifically, we highlight our design features for disentangling subject pool effects and implementation method effects. The most important of these is the introduction of two environmental treatments in the laboratory. One replicates traditional experiments, the second mimics the Internet setting as much as possible. We term them “Lab-Lab" and “Lab-Internet" to avoid confusion with the CentERpanel experiment (for which we use the term “Internet experiment" synonymously). The full experimental instructions, samples of choice problems, help screens, final questions, and the debriefing part are available upon request from the authors.

4.2.1 The Multiple Price List Format

The experiments were conducted using an adapted version of the multiple price list format pioneered in economics by Binswanger (1980) and recently employed in the context of risk preferences by Holt and Laury (2002). A general and extensive description can be found in Andersen, Harrison, Lau, and Rutström (2006). Here we stick to a brief depiction and highlight our modifications. In principle the MPL works as follows: Each subject is presented a list of four lotteries with identical payoffs but varying probabilities such as the one presented in Figure 4.1.[1] In each of the four cases, the participant can choose between Option ‘A’ and Option ‘B’. The table is designed such that the expected value of lottery ‘A’ starts out higher but moves up slower than the corresponding figure of lottery ‘B’. A participant with monotone preferences switches at some point from one to the other.

[1] We use four choices per screen to avoid the need for scrolling and to make it not too overwhelming for the respondents.

Fig. 4.1: Screenshot of Lottery 5, First Screen

Subjects who are consistent in the sense that they do not switch back and forth between ‘A’ and ‘B’ and choose the higher certain payoff in the fourth question are routed to a second screen containing lotteries with the same payoffs, but a finer probability grid. This consists of 10%-steps located roughly between their highest choice of ‘A’ and their lowest choice of ‘B’. This is essentially a slightly modified version of the iterative multiple price list (iMPL) format described in Andersen, Harrison, Lau, and Rutström (2006). In contrast to earlier studies using the MPL format we chose to supplement the verbal descriptions of the decision tasks with pie-charts describing the probabilities of the outcomes. Subjects faced seven such tasks. We believe that this design is easy to grasp and capable of generating relatively precise estimates in combination with the iterative procedure.

In order to identify loss aversion, some of the riskier lotteries involved negative outcomes, see Table 4.1 for the characteristics of each of the lotteries. Note that payments from the safer option ‘A’ were strictly positive in all cases. All payments were made three months after the experiment, which allowed us to vary the timing uncertainty resolution keeping the payment date constant. To introduce variation in the timing of uncertainty resolution we indicated on each screen whether the outcome of the lottery is revealed immediately or in three months

Tab. 4.1: Payoffs from the Seven Lotteries

Sheet	Uncertainty resolution	Payoff Low, A	Payoff High, A	Uncertainty resolution	Payoff Low, B	Payoff High, B
1	early	27	33	early	0	69
2	early	39	48	early	9	87
3	early	12	15	early	-15	48
4	early	33	36	late	6	69
5	early	18	21	late	-9	54
6	early	24	27	early	-3	60
7	late	15	18	late	-12	51

Note: These values were shown in the high incentive and hypothetical treatments. For the low incentive treatment they were divided by three. Order 0 consisted of the sequence of screens $\{2, 7, 3, 6, 1, 4, 5\}$ and order 1 of the sequence $\{1, 4, 5, 2, 6, 3, 7\}$.

time. If uncertainty was resolved late, the indication was highlighted by red letters, see again Figure 4.1 for an illustration. Pretests showed that it was important to make it very clear in the instructions that all payments would take place in three months time. Otherwise some subjects might have been confused with respect to the timing of the payment and would have confounded time preference and uncertainty resolution preferences. Consequently, we devoted the second page of the instructions entirely to the payment procedure.

To ensure incentive compatibility one row was selected at random and the subjects were paid according to the outcome of that lottery. In order to avoid negative payoffs the highest possible loss was chosen not to exceed the fixed participation fee. Respondents faced seven two-stage multiple price lists aimed at capturing the three preference parameters. Everybody who answered all first-screen questions in a consistent fashion hence faced 56 decision problems with seven different payoff configurations.

4.2.2 The CentERpanel Experiment

The subjects in the Internet experiment are respondents in the CentERpanel, aged 16 and older. Recruitment of subjects and administration of the CentERpanel is carried out by CentERdata, a survey research institute affiliated with Tilburg University. The panel contains approximately 2,000 households that are representative of the Dutch population in terms of observable characteristics. The surveys and experiments conducted on the CentERpanel are carried out using an Internet-based telepanel. To avoid selection bias, households without Internet access are provided with a set-top box for their TV. Almost every week, the panel members fill in a questionnaire on the Internet from their home.[2] Panel members are reimbursed for their costs of participation (fees of dial-up Internet connections etc.) on a regular basis. Our payments were included in one of these transactions three months after our experiment.

A major advantage of this setup is that we have access to a sample that is representative of the population. Moreover we have access to a large amount of background information on demographics and work; housing and mortgages; health and income; assets and liabilities. The latter are measured in a particularly detailed way, distinguishing over 20 different asset components and eight different forms of liabilities. We conducted the experiments in November and December of 2005.

We included a non-participation option on the welcome screen after a brief introduction to the experiment. For the two treatments with real incentives, subjects were told the amount of the participation fee and that they had the chance to win substantially more or lose (part of) this money again. It was made clear that no payment would be made upon non-participation. In the hypothetical treatment, subjects were told that the questionnaire consisted of choices under uncertainty in a hypothetical setting. They then had to tick one of two buttons, indicating whether they wanted to participate or not. See Figure 4.2 for the complete introductory screen of all treatments. In total, 2,299 persons logged into the system. Slightly less than thirteen percent opted for non-participation on the first screen, leaving us with 2,008 persons who took part in the experiment. Another eighty subjects dropped out before completing the questionnaire. Our resulting sample consists of 1928 subjects who made 98,108 choices among risky prospects. After going through two pages of online instructions these subjects faced the seven two-stage multiple price lists. The instructions and specially designed help screens could be accessed throughout the experiment. These were included to improve comparability with the laboratory experiments, compensating for the absence of an experimenter.

We employed a 3x2 treatment design with respect to payoffs and orderings of the screens. The three payoff treatments consisted of hypothetical and real lotteries with the amounts shown in Table 4.1 and another one with real payoffs, but amounts divided by three. We refer to these as hypothetical, high and low incentive treatments. All subjects in the high and low incentive groups received an upfront payment of 15 and 5 Euro, respectively. No payment at all was made in the hypothetical group. We introduced two randomly determined orderings of

[2] For more information about the CentERpanel see http://www.uvt.nl/centerdata/.

Fig. 4.2: Translations of the Welcome Screens in the CentERpanel Experiment

Hypothetical Treatment

Welcome to this economic experiment carried out by researchers of Tilburg University. The experiment is about making choices under uncertainty. Please read the instructions carefully in order to understand how the experiment works.

If you have questions after the beginning of the experiment, you can return to the instructions by clicking on a link at the top of the screen. If you have questions on the specific screen, you can click on 'Help' at the top right corner of the screen.

The questions are not designed to test you. Answers are therefore not correct or incorrect; please give the answers that reflect your own preferences.

This questionnaire is about making choices between several situations in which you can (hypothetically) gain or lose money. Your revenue depends on the choices you make and on chance. **What matters is what you would do in hypothetical situations, in reality, there is nothing at stake for you.** If you nevertheless do not want to participate out of principle, you can indicate this below. In that case you will not continue with the questionnaire.

O Yes, I proceed with the questionnaire

O No, I do **not** want to complete this questionnaire

High (Low) Incentive Treatment

Welcome to this economic experiment carried out by researchers of Tilburg University. The experiment is about making choices under uncertainty. Please read the instructions carefully in order to understand how the experiment works.

If you have questions after the beginning of the experiment, you can return to the instructions by clicking on a link at the top of the screen. If you have questions on the specific screen, you can click on 'Help' at the top right corner of the screen.

You will receive 15 (5) Euros for participating. Then you can, depending on the choices you make and on chance, earn more or lose part of the 15 (5) Euros. If completing the total experiment, you receive the reward for participating, possibly increased by your gain (or reduced by your loss) in one of the choices you have made. Whether the latter occurs and which choice then determines your payoff, will be determined by chance. **Your total reward will be added to your CentERpoints.**

The questions are not designed to test you. Answers are therefore not correct or incorrect; please give the answers that reflect your own preferences. Assume in each choice problem that this choice determines your actual payoff.

This questionnaire is about making choices, and your payoff depends on your choices and on chance. If you do not want to participate out of principle, you can indicate this below. In that case you will not continue with the questionnaire.

O Yes, I proceed with the questionnaire

O No, I do **not** want to complete this questionnaire

screens. Order 0 consisted of the sequence $\{2,7,3,6,1,4,5\}$, order 1 of the ordering $\{1,4,5,2,6,3,7\}$. In the incentives treatments, only one in ten respondents actually got paid for one of the lotteries, everybody received the participation fee (and respondents were informed about this also).

4.2.3 The Laboratory Experiment

In order to compare the answers in our Internet survey to those in the environment of a controlled laboratory experiment traditionally used in the large majority of economic experiments, we performed the same experiment in the economics laboratory at Tilburg University. In total, 203 students participated in the experiment (8 sessions in September 2005 and 8 sessions in May 2006). To enable comparisons between behaviour in the laboratory and on the Internet, the six treatments described above were carried out in the laboratory as well. Since the purpose was to distinguish effects due to different subject pools from potential effects due to moving from the controlled laboratory setting to the Internet, we also replicated this latter change in environment in the lab. This was done by performing the laboratory experiments using two different environmental structures.

The first environment treatment, labelled the “Lab-Lab” treatment was designed to replicate the traditional setup used in laboratory experiments. In particular, an experimenter was present in the room to help the subjects and answer questions. The link to the instructions and help screens was deleted from the screens; otherwise, the screens were similar to the one shown in Figure 4.1. Participants also had to wait for everyone else in the session to finish before leaving. The second environmental treatment, the “Lab-Internet" treatment, differed in that the experimenter was not present in the laboratory while participants went through the experiment. When the subjects arrived in the laboratory an experimenter announced the following before leaving the room: “1. Please do not talk to the other participants. 2. If you have any questions, please consult the online help screens and instructions. 3. If you encounter a computer break down, please contact me outside the room. 4. Make sure you complete the whole questionnaire before leaving - otherwise your answers won’t get registered and you won’t receive any payments. 5. When you have finished and have completed the whole experiment, you may leave.” The main differences compared to the “Lab-Lab” treatment were hence that there was no experimenter in the room, the availability of help screens, and the possibility to leave upon completing the experiment without having to wait for everyone else. Subjects in the laboratory experiments faced the same incentives treatments as CentERpanel respondents, but in the hypothetical treatment, in contrast to Internet participants, lab participants received a participation fee of 5 Euro. This was necessary for recruitment. The payment procedure for the incentives treatments was as in the CentERpanel experiment: The participation fee was transferred to participants’ bank accounts three months after the experiments, one in ten subjects received in addition the (possibly negative) payment from one of the lotteries he had played.

4.3 Selection Effects in the CentERpanel Experiment

A major advantage of conducting the experiment via an existing survey is that we know many characteristics of the persons eligible for participation, regardless of whether they actually take part in the experiment or not. This allows us to monitor the recruitment process much better than usual. In particular, we can relate the participation decision on the first screen (see Section 4.2.2) to observed variables. Put differently, in order to get reliable population-wide estimates, it is not enough to sample from a representative sub-population unless nonresponse is perfectly random. Since people self-select into the experiment it is unlikely that this condition holds.

The fact that selection effects may be important for the outcome of interest in our setting has been demonstrated recently by Harrison, Lau, and Rutström (2005). The advantage of our setup compared to theirs is that we have much more information about non-participants. There is a qualifier to this, however. One could think of the sampling process for our experiment on the CentERpanel in four stages. First, Dutch households are contacted at random and choose to participate in the CentERpanel or not. Second, a random subsample of CentERpanel participants is asked to take part in our experiment. Third, after getting to know the nature of the experiment, some of these people choose to not take part in it. Fourth, some of the participants drop out or click through the screens extremely rapidly. The analysis in section 4.3.1 is concerned with step three whereas section 4.3.2 takes a closer look at step four. There is a rich set of background variables available for these two analyses. For step one, CentERdata provides standard survey weights based upon a much larger Dutch household survey drawn by Statistics Netherlands, and the analysis in section 4.4 additionally takes these weights into account. The construction of the final weights which account for all four steps is described in section 4.3.3. Note that the selection at step two of the selection process is completely random by design and does not induce any selection effects.

Steps three and four are especially interesting for experimental economics because they replicate the recruitment process for laboratory experiments to some extent. Usually, some information is conveyed on payoffs and the type of experiments before subjects come to the lab. This is exactly what happens on the first screen of our experiment. We would argue that step four is a type of selection that is mixed up with step three in a laboratory setting. Fixed costs of going to the laboratory are higher than those of logging into the experiment on the Internet — subjects can be expected to have contemplated the costs and benefits of participating beforehand. This is not necessarily the case for an Internet setting, where the experiment is “just two mouse clicks away". The cost-benefit analysis may well be postponed to take place during the experiment. Indeed, people leaving the laboratory prematurely is virtually not an issue in experimental economics, whereas four percent of subjects do not finish our CentERpanel experiment. Another materialisation of this story is that some people may speed through the experiment, not caring much (or at all) about instructions or the like. Taking the shortest duration observed in the laboratory as a lower bound, some seven percent of respondents fall into this category. We now investigate in more detail whether steps three and four bear any relation to observed characteristics of participants.

4.3.1 Descriptive Statistics and Participation

Table 4.2 lists some descriptive statistics for the participants in the CentERpanel and laboratory experiments and for those who opted for non-participation in the former. The first thing to note is that there is indeed a broad variation over all characteristics in the first two columns. This is all the more true, albeit not surprising, if compared to the last column. In terms of age category and educational qualification, the laboratory population represents only five percent of that in the first two columns. The variables listed in Table 4.2 can be broadly classified into 6 groups: Incentive treatment; education; sex and age; employment status and residential area; financial literacy and experience; income and wealth. Some of the questions, particularly those on assets and financial literacy and experience, are drawn from the DNB household survey (DHS), a survey focusing on financial issues held among CentERpanel respondents once a year.

Comparing the first two columns, the results suggest some possible paths for selection effects: Less qualified persons, women, the elderly, subjects without an occupation and those with high incomes and less asset holdings seem more prone to decline participation. However, these descriptives may be blurred by correlation amongst the variables, calling for a more formal analysis. This is less of a concern for the incentive treatments since we know they have been randomised over the population. Results are striking with non-participation rates being much higher for the hypothetical treatment. This is indicative of a path through which monetary incentives may matter which has not been studied directly so far (cf. Beattie and Loomes (1997) or Gneezy and Rustichini (2000)), namely through differential participation in the experiment. Since the type of earnings is often announced in the recruitment process (“Earn 10 Euros in one hour", “Make *up to* 23 Euros in 45 minutes", etc.), different subject pools may be attracted.

We carry out a simple exercise to test for selection effects by regressing the participation decision on observed characteristics in a probit model. Results are presented in Table 4.3. The first column contains the most basic specification with the incentive treatments, education, gender, age, occupational status, and residential area as covariates. In the subsequent columns, the number of observations is somewhat lower since we have to draw upon questionnaires of the DHS that not everybody answered. Specification (2) adds the financial variables: whether the respondent manages the households finances, whether the employer offers a save-as-you-earn deduction arrangement[3], whether the respondent holds such a

[3] This is an employer provided savings plan that is heavily subsidized by the state through tax deductions. See Alessie, Hochguertel, and van Soest (2006). Up to some maximum, contributions are tax deductible and withdrawals after four years or later are not taxed, making net returns much higher than on any other safe asset for most tax payers. While it is not hard to sign up for these plans and while the employer does most of the paperwork, the default is not to participate, which

Tab. 4.2: Selected Characteristics of Participants

Variable	Participants, CentERpanel	Non-Participants, CentERpanel	Participants, Laboratory
Hypothetical Treatment	31.6	50.2	36.5
Low Incentive Treatment	36.7	23.0	27.0
High Incentive Treatment	31.7	26.8	36.5
Primary/Lower Sec.	31.0	44.1	0.0
Higher Secondary Education	14.0	13.1	0.0
Intermediate Vocational Training	19.6	16.9	0.0
Higher Vocational Training	24.0	20.0	0.0
University Degree	11.4	5.9	100.0
Female	46.5	53.6	45.5
Age 16-34	26.5	14.1	100.0
Age 35-44	19.0	13.4	0.0
Age 45-54	21.7	21.9	0.0
Age 55-64	18.2	17.0	0.0
Age 65+	15.7	32.7	0.0
Employed, Freelancer, or Self-Employed	55.8	35.7	
Unemployed, Looking for Job	2.5	2.8	
Other Occupation, Student, Pensioner, Housework	41.7	61.5	
Lives in Urban Area	60.0	62.5	
HH Financial Administrator	63.8	56.0	
Employer offers Savings Plan	43.1	25.5	
Holds Employer-Sponsored Sav. Plan	35.4	16.5	
Has Savings Account, Deposit Book, etc.	87.0	85.2	
Holds Funds, Stocks, Options, etc.	30.6	24.7	
Difficult to Manage on Budget	56.0	51.9	
HH Income <14k Euro	10.0	11.0	
HH Income 14k-22k Euro	24.2	24.1	
HH Income 22k-40k Euro	48.8	51.2	
HH Income >40k Euro	17.0	13.8	
Mean Log Total Asset Holdings excl. Accom.	7.37	2.77	
Maximum Number of Observations	2,008	291	178

Note: Except for the last two rows, numbers indicate column percentages. The next to the last row indicates mean log assets and the last row shows the number of observations, not taking into account any item non-response on some of the previous variables. "Difficult to Manage on Budget" refers to answering "very hard", "hard", or "neither hard nor easy" to the question "How well can you manage on the total income of your household?", the remaining options are "easy" and "very easy". Some households did not complete the questionnaires of the DHS from which some of the variables are drawn. Hence the number of observations is lower for some of the variables in question.

plan, whether the household's financial portfolio includes savings accounts and the like, and whether it includes mutual funds, stocks, etc.. Specifications (3) to (5) add in turn some measures of income and wealth: First, a subjective income variable; second, household income in four categories; and third, the logarithm of total asset holdings.

Results are very consistent across specifications. They indicate that persons in the top two education categories are indeed significantly more likely to participate in the experiment. The gender effect also confirms what the bivariate correlations suggested with women being more likely not to take part. Age effects start to matter at age 45, where participation rates are smaller than for younger persons. They are more pronounced for higher age categories, point estimates rise monotonously. Those beyond 65 years of age are significantly less likely to start the experiment than all other age groups. Point estimates for working (as opposed to not participating in the labour market) are negative, but small and generally insignificant. The point estimates indicate that the unemployed are more likely to participate in the experiment, but the difference with persons outside the labour force is insignificant. Residents in an urban area have somewhat lower participation rates.[4]

With respect to the incentive treatment, the results in Table 4.2 are confirmed. Being assigned to an incentive treatment significantly increases the propensity to participate in the experiment. Surprisingly, the point estimate is higher for the low incentive treatment. However, the difference to the high incentive treatment is just significant at the 5%-level in the first specification ($\chi^2 = 4.96$) and insignificant in all remaining estimations. We conclude that incentives do seem to increase overall subject turnout. Their size does not seem to have an impact, though.

Adding the financial variables mainly serves the purpose to control for individuals' preferences and their financial knowledge. Being the financial administrator of the household for example may reflect a preference for spending one's time with risky choice problems (the opposite line of argumentation is probably more intuitive: not being the financial administrator of the household is on average indicative of a general dislike to deal with such questions). It significantly increases the propensity to participate. Whether the employer offers a savings plan is just a control variable that is necessary in order to not confound the effects of holding such a plan with one's employment status (note that the latter is only imperfectly controlled for by the direct measures). The variable of interest, taking part in a save-as-you-earn savings arrangement, emerges highly significant and negative. We view this as supporting the interpretation of the financial variables as to reflect a certain preference to contemplate financial questions. This is strengthened by

may explain why employees with little financial knowledge or interest often do not sign up, cf., e.g., the work of Madrian and Shea (2001) on non-take up of 401(k) plans.

[4] We also tested for interaction effects but did not find any of them to be important.

Tab. 4.3: Nonparticipation in the CentERpanel Experiment

Specification	(1)	(2)	(3)	(4)	(5)
Low Incentive Treatment	-.475*** (.079)	-.502*** (.090)	-.541*** (.097)	-.505*** (.090)	-.501*** (.090)
High Incentive Treatment	-.289*** (.078)	-.426*** (.094)	-.419*** (.100)	-.428*** (.094)	-.425*** (.094)
Higher Secondary Education	-.187* (.104)	-.184 (.122)	-.173 (.129)	-.184 (.123)	-.187 (.122)
Intermediate Vocational Training	-.129 (.096)	-.147 (.113)	-.118 (.119)	-.151 (.113)	-.149 (.113)
Higher Vocational Training	-.202** (.090)	-.179* (.105)	-.197* (.115)	-.185* (.108)	-.181* (.105)
University Degree	-.247** (.121)	-.285* (.149)	-.258 (.159)	-.283* (.153)	-.289* (.149)
Female	.187*** (.067)	.151* (.079)	.176** (.084)	.157** (.079)	.151* (.079)
Age 35-44	.107 (.107)	.149 (.134)	.177 (.148)	.145 (.134)	.148 (.134)
Age 45-54	.152 (.101)	.236* (.124)	.264* (.136)	.237* (.125)	.233* (.124)
Age 55-64	.179* (.107)	.353*** (.127)	.423*** (.139)	.347*** (.128)	.346*** (.127)
Age 65+	.553*** (.112)	.691*** (.136)	.705*** (.148)	.683*** (.137)	.679*** (.138)
Employed, Freelancer, or Self-Employed	-.157* (.085)	-.024 (.106)	-.045 (.114)	-.032 (.107)	-.023 (.106)
Unemployed, Looking for Job	.226 (.196)	.287 (.227)	.301 (.244)	.298 (.228)	.289 (.227)
Lives in Urban Area	.106 (.067)	.151* (.080)	.208** (.086)	.150* (.080)	.151* (.080)
HH Financial Administrator		-.147* (.081)	-.187** (.087)	-.137* (.083)	-.146* (.081)
Employer offers Savings Plan		.165 (.141)	.128 (.157)	.160 (.142)	.160 (.142)
Holds Employer-Sponsored Sav. Plan		-.482*** (.150)	-.447*** (.167)	-.479*** (.151)	-.477*** (.151)
Has Savings Account, Deposit Book, etc.		.033 (.114)	-.012 (.122)	.023 (.114)	-.005 (.138)
Holds Funds, Stocks, Options, etc.		-.153* (.087)	-.242** (.096)	-.156* (.087)	-.161* (.088)
Difficult to Manage on Budget			-.059 (.087)		
HH Income 14k-22k Euro				.094 (.146)	
HH Income 22k-40k Euro				.123 (.140)	
HH Income >40k Euro				.072 (.168)	
Log Total Asset Holdings excl. Accom.					.008 (.017)
Constant	-.922*** (.119)	-.923*** (.177)	-.900*** (.207)	-1.007*** (.207)	-.945*** (.183)
No. of Observations	2296	1802	1626	1802	1802

Note: Coefficients and standard errors of simple probit regressions on opting for nonparticipation on the first screen of the CentERpanel experiment. Left-out categories of relevant variables are hypothetical treatment; primary and lower secondary education; ages 18-34; other type of occupation; and household income less than 14,000 Euro. "Difficult to Manage on Budget" refers to answering "very hard", "hard", or "neither hard nor easy" to the question "How well can you manage on the total income of your household?", the remaining options are "easy" and "very easy". All other variables are self-explanatory. Asterisks indicate significance at the 10%, 5%, and 1%-level.

the other two portfolio variables. On the one hand, having an ordinary savings account does not seem to have any predictive power for taking part in the experiment. Saving accounts are a class of products that virtually everybody knows from childhood. They have a very simple structure and do not require much expertise or effort. On the other hand, the ownership of mutual funds, stocks, and the like is significantly associated with higher participation rates. These are much more sophisticated products and investing in them requires more financial knowledge. The point made here is similar to the finding of Lazear, Malmendier, and Weber (2006) that many subjects are willing to pay a positive price in order to avoid being the proposer in a dictator game. In the context of risky choice experiments, subjects who dislike such situations just do not participate.

Finally, turning to specifications (3) to (5), it becomes evident that none of the income and wealth variables are associated with participation. Point estimates are generally small and insignificant, and a joint test confirms this.

4.3.2 Perseverance

As discussed above, Internet experiments may suffer from an additional selection issue that is not a concern with laboratory experiments – either through dropping out of the experiment before completion or through very rapid completion of the tasks without giving them serious thought. Considering the latter as a form of nonparticipation is motivated by the fact that there is a lower bound on the time needed to digest the instructions and give nonrandom answers to all the questions. It is certainly higher than the 1:43 minutes which is the minimum amount of time that one subject in the Internet experiment needed in order to click through the screens. After doing some robustness checks, we settled for the minimum duration observed in the laboratory as a cut-off point (5:20 minutes).

We attributed this fourth step in the full selection process to drastically lower fixed costs of taking part in our Internet experiment and argued that most subjects who exhibit either behaviour would not enter a laboratory in the first place. Hence selection problems are similar. There is at least one alternative interpretation, namely interaction with the experimenter and typical rules in the laboratory. The most salient mechanism of how this may matter is the possibility to ask questions. If a subject does not understand (part of) a task, he may opt for randomly ticking options or drop out entirely. Another reason is that in typical laboratory experiments everybody is expected to stay until the last subject has finished. Hence there is no point in rapid completion - people might just as well spend the time which they have to stay in the laboratory anyway on the experiment. We do not think that this explanation is very likely however, because we largely control for these effects via the “Lab-Lab" and “Lab-Internet" treatments of the laboratory experiments (see section 4.2.3). The minimum amount of time spent on the laboratory experiments is actually found in the “Lab-Lab" treatment, the smallest value in the “Lab-Internet" treatment is about 80 seconds longer. Overall, dispersion is much higher in the traditional Lab-lab treatment (standard deviation 4.71 vs. 3.92 in the “Lab-Internet" treatment), and the left hand tail of the distribution in the “Lab-lab" treatment has more mass than in the “Lab-Internet" treatment. Hence we definitely cannot find more mass in the left tail of the “Lab-Internet" treatment durations, which is what one would expect from the second explanation of the selection mechanism. Of course the comparison between these two treatment only applies to the young and highly educated lab population, and not necessarily to other population groups. Especially not understanding the tasks is a concern that we cannot fully discard. This may indeed introduce an additional problem of selection bias in Internet experiments as compared to those in the laboratory.

We investigate the two selection mechanisms from step four (not completing or completing n less than 5:20 minutes) of the process described in the introduction to this chapter by estimating a multinomial logit model. The dependent variable classifies participating subjects (N= $2,008$) into three categories. The first category is made up of 138 persons who took less than five minutes and twenty seconds to complete the entire experiment. Eighty subjects dropped out entirely after starting with the experiment and enter group two. The remaining category, our final sample used in the remainder of the paper, consists of 1,789 subjects and constitutes the baseline category of the regression. Results are presented in Table 4.4. The construction of the estimates is similar to the one in Table 4.3. We first present an estimation that maximises the number of observations. Coefficients are listed in column (1) for those who fall under the duration cut-off and in column (2) for those who drop out. We then sacrifice a sixth of our sample for a richer set of covariates. Coefficients are listed in columns (3) and (4) which correspond to columns (1) and (2), respectively.

The only case where an incentive treatment seems to matter is that for high incentives the probability of rapid completion decreases. The other estimates all point in the same direction or are very close to zero. However, standard errors are too large for inference. Columns (1) and (3) indicate that subjects with more education were less likely to rapidly click through the experimental screens. Our interpretation of this is related to the difficulty of the experiment: One reaction of subjects who do not understand the task may be to finish it as fast as possible, while not foregoing the chance to win some money in case of the real incentive treatments. Direct tests for the latter could be based on interaction effects of education and incentive treatment. They did not produce any significant results, but sample sizes relevant for such questions are very small despite the rather large pool of participants. Evidence with respect to dropping out is mixed. Results generally point in the same direction as for rapid completion, yet only one point estimate barely reaches significance.

Tab. 4.4: Abortion and Abbreviation of the CentERpanel Experiment

Specification / Dependent Variable	(1)	(2)	(3)	(4)
Low Incentive Treatment	-.212 (.211)	-.261 (.288)	-.283 (.235)	-.403 (.338)
High Incentive Treatment	-.519** (.235)	.026 (.279)	-.467* (.260)	-.205 (.337)
Higher Secondary Education	-.070 (.263)	-.110 (.376)	-.204 (.296)	-.830 (.515)
Intermediate Vocational Training	-.181 (.245)	-.164 (.355)	-.151 (.275)	-.513 (.426)
Higher Vocational Training	-1.003*** (.300)	-.243 (.341)	-1.075*** (.347)	-.709* (.419)
University Degree	-1.259*** (.457)	.280 (.361)	-1.247** (.517)	-.175 (.453)
Female	.316* (.191)	.233 (.239)	.435** (.212)	-.143 (.291)
Age 35-44	-.533** (.232)	-.289 (.335)	-.575** (.275)	-.500 (.424)
Age 45-54	-1.127*** (.260)	-1.010** (.397)	-1.094*** (.293)	-1.311*** (.492)
Age 55-64	-1.861*** (.369)	-.566 (.379)	-1.840*** (.398)	-.731* (.440)
Age 65+	-2.817*** (.609)	.027 (.383)	-2.625*** (.617)	-.263 (.477)
Employed, Freelancer, or Self-Employed	-.073 (.214)	-.054 (.313)	-.147 (.239)	.028 (.380)
Unemployed, Looking for Job	-.577 (.751)	.966* (.540)	-.290 (.766)	1.176* (.623)
Lives in Urban Area	-.231 (.186)	.215 (.247)	-.216 (.207)	.244 (.296)
HH Financial Administrator			-.763*** (.222)	.112 (.312)
Has Savings Account, Deposit Book, etc.			.352 (.341)	.208 (.447)
Holds Funds, Stocks, Options, etc.			.014 (.238)	.224 (.312)
HH Income 14k-22k Euro			.274 (.449)	-.505 (.503)
HH Income 22k-40k Euro			.457 (.432)	-.221 (.462)
HH Income >40k Euro			.840* (.482)	-.194 (.569)
Constant	-1.211*** (.281)	-2.920*** (.423)	-1.611*** (.573)	-2.499*** (.728)
No. of Observations	2006	2006	1689	1689

Note: Coefficients and standard errors of multinomial logit regression. Columns indicate categories of the dependent variable by regression type. The reference category are those who completed the experiment in more than 5:20 minutes. Columns (1) and (3) contain coefficients for completing the experiment in less than 5:20 minutes; columns (2) and (4) those for dropping out before completing the experiment. Left-out categories of relevant variables are hypothetical treatment; primary and lower secondary education; ages 18-34; other type of occupation; and household income less than 14,000 Euro. “Difficult to Manage on Budget" refers to answering “very hard", “hard", or “neither hard nor easy" to the question “How well can you manage on the total income of your household?", the remaining options are “easy" and “very easy". All other variables are self-explanatory. Asterisks indicate significance at the 10%, 5%, and 1%-level.

Women not only have a smaller propensity to participate than men, they also have a higher propensity to quickly go through the experiments. They do not seem to be more likely to drop out, though. Older participants are more likely to make it into our final sample. This is mainly due to the fact that people are less likely to complete the experiment rapidly. Part of this result is surely due to lower computer skills among the elderly, however effects are already visible for ages around forty. They then exhibit a monotonic increase. Drop-outs are least prevalent for people in their late forties and early fifties. The unemployed exhibit the highest quitting rates relative to groups with a different occupational status. Coefficients are only borderline significant, though.

The preference argument with respect to the financial variables goes through again, albeit not over such a broad range of variables as for the participation decision. In particular, ownership of mutual funds and of an employer-sponsored savings plan (not shown for brevity reasons) does not help to predict either outcome. On the other hand, being the financial administrator of the household is associated with a lower propensity to rapidly go through the tasks, reinforcing our argument from the above discussion. Finally, being in the top income class is related to a higher probability of speeding through the questionnaire, probably reflecting a larger opportunity cost of time. Given the low significance, this interpretation should not be overly stressed, though.

4.3.3 Overall Selection and Construction of Sampling Weights

Breaking up the selection process in different components and subcomponents is important for a better understanding of the mechanisms, but what matters for the outcome of interest are overall effects. For example, while results on incentive effects point in the same direction in Tables 4.3 and 4.4, this is not the case for age effects. To look at this issue, we constructed a binary variable that included our final sample in one category and all nonparticipants, drop-outs, and rapid finishers in the other. As in section 4.3.1, we then ran simple probit regressions of participation on both the parsimonious set of covariates and richer ones. For reasons of brevity, we stick to a verbal description of the effects since they are not surprising. Tables are available from the authors upon request.

The payment of incentives leads to much higher participation rates overall. The marginal effect for either incentive treatment is a little above twelve percent (with the baseline participation rate for the hypothetical treatment and all dummies in the rich specification set to zero being at 80%). Being in the top two education classes is significantly associated with selection into the experiment while females are much less likely to enter our final sample. The financial variables are still indicative of a higher participation probability, especially being the financial administrator and taking part in a save-as-you-earn arrangement. Ownership of mutual funds is not quite significant in overall participation. Being in the top income decile is borderline significant for a lower propensity to take part in the experiment. In the case of the age variables, there are some important neutralisation effects. Recall that age was associated with a higher propensity to start the experiment, but also with higher rates of speeding and dropping out. The consolidated values show an important increase in participation rates for ages 45 to 64 in the parsimonious specification, higher perseverance rates dominate here. Both effects just cancel for the elderly and also for the other groups when we move to the richer specification.

For testing whether selection effects matter for the analysis in the next section, we invoke the assumption that observations are conditionally missing at random (CMAR; see, e.g., Little and Rubin (2002)). That is, we assume that the nonresponse decision is random after controlling for factors observable to us. We then can construct weights from the probit model estimated here by taking the inverse of the predicted probability of being in the final sample. These weights do not capture the whole selection process yet, since they only correct for steps three and four. As detailed above, CentERdata provides weights similar to ours that control for the rest. By combining the two sets of weights we can account for the full process. More specifically, the CentERdata weights are multiplied with the predicted inverse probability of participation and then normalised through dividing by the population average of this product. Due to sample size considerations we opt for the parsimonious specification in the probit regression. We then can proceed to test whether results from the weighted sample are the same as those from the raw data.

A final note at this point concerns how our approach to the selection issue compares to the selection model pioneered by Heckman (1979) and less parametric versions of that model. This is the solution that comes immediately to mind upon the cue “selection problem", due to its widespread use and generality. In particular, Heckman’s model allows for a selection mechanism based on unobservable characteristics while the CMAR assumption does not. The reason for not employing it here is that the model relies on the existence of valid instruments, one or more variables that affect the selection decision while having no impact upon the outcome of interest. The above-sketched results on the financial variables can be seen as indirect evidence for selection based on preferences towards dealing with risky choice situations, but such preferences may also be correlated with risk preferences and (or) error frequencies, and are therefore unlikely to serve as valid instruments. We cannot think of any valid instruments and certainly do not see any in our set of background variables (which is much broader than what is mentioned and relevant in this paper), since participation and risk preferences and error frequencies are too closely interwoven. Therefore we stick to concentrating on the effects of selection on observables.

4.4 Errors and Preferences

This section compares behaviour between the laboratory and the Internet treatments in order to assess the importance of subject pool composition and the Internet interface. We also consider whether the selection process discussed in the previous section has an effect on the aggregated measures of behaviour.

4.4.1 Errors and Inconsistencies in the Lab vs. the CentERpanel

As a starting point for a further analysis of behaviour we analyse to what extent the subjects' choices were consistent with the standard assumption of monotonic preferences. In our design, there are three patterns of choices that we consider as inconsistent. First, choosing option 'B' when there is a zero probability for the high outcome to occur or choosing 'A' when there is a 100 percent probability for the high outcome to occur, clearly violates dominance. A second type of inconsistency emerges when subjects switch back and forth on the same screen. Finally, a third category of violations consists of inconsistent choices between screens with the same payoffs. Recall that we use an iterative version of the MPL, where after choosing among four options on a first screen, subjects are transferred to a second screen with the same payoffs but with a finer partition of probabilities. There were some overlaps of probabilities between screens enabling subjects to make a choice on the second screen that is inconsistent with the choice made on the first screen. In Table 4.5, the averages of dominance or monotonicity violations are presented by treatment and choice task. Note also that only one violation per subject is counted for each of the screens, limiting the maximum amount of mistakes to seven.

The striking fact revealed by Table 4.5 is that the numbers of violations are much higher in the Internet experiment than in the laboratory experiment. The percentage of inconsistent answers varies between 9% and 23% in the lab, whereas it ranges between 31% and 39% among the Internet participants. Clearly, answers made by the laboratory subjects were more consistent which is confirmed using the Mann-Whitney (MW) and Kolmogorov-Smirnov (KS) nonparametric tests on the number of inconsistencies per subject aggregating over all screens. Irrespective of whether we look at each type of violation separately or whether we aggregate over all types of violations we can reject the null hypothesis that observations are drawn from the same sample using both tests (two-sided p-values<0.01).

One of the possible sources behind this difference between the laboratory and the Internet is the different environment under which these experiments were conducted. The presence of an experimenter in the laboratory experiments is for example one factor that might have an impact upon answers. To assess this issue we investigate the percentage of violations in the "Lab-Lab" treatment with the “Lab-Internet" treatment. Contrary to this hypothesis, the figures in Table 4.5 indicate that violations are slightly more common in the “Lab-Lab" than the “Lab-Internet" treatment. These differences are, however, not significant if judged by the MW and KS tests (p-values>0.1). The higher frequency of violations observed in the Internet treatment is thus most likely not driven by the presence of an experimenter or the ability to leave immediately after finishing the experiment.

Following these results, the natural explanation behind the disparity between the results from the laboratory and the Internet experiment lies in the different characteristics of the subject pools. A way to investigate this is to compare choices of the laboratory sample with a subsample of the Internet sample that share some important characteristics of the laboratory sample. On these grounds, we formed a subsample labelled “Uni", consisting of all respondents between 18 and 34 years of age with a university education. When restricting attention to this subsample, behaviour of the Internet participants resembles that of the student sample. The fraction of violations among the “Uni" subsample ranges between 0.15 and 0.22, which is covered by the 0.09 to 0.23 range observed among the laboratory subjects. Overall violations in the “Lab-Lab" treatment and from the “Uni" subsample are almost exactly the same. All these differences are also insignificant using the KS and MW tests. The higher frequency of errors in the Internet treatment seems hence to be driven by the characteristics of the different subject pools. Different implementation forms do not seem to play a role, at least not for the young and educated.

A related issue is whether subjects in the two experiments commit the same type of errors. Investigating the relative frequency of the three different types of violations described above, it turns out that this in is not the case. In both the laboratory and the Internet treatments 17.9% of the violations were due to switching back and forth within screens. In the laboratory 72.5 % of the violations were due to nonmonotonicities between screens with a corresponding figure of 43.6 % in the Internet. Finally, dominance violations made up 9.6% of the violations in the lab, but 38.4% in the Internet treatment. Nonmonotonicities between screens is, hence, the prevailing source of errors in the laboratory sample, whereas violation of dominance plays a much more pronounced role in the Internet treatment. Along these lines, looking at the average number of inconsistencies per type we detect a higher number of violations within screens in the Internet treatment. Still, the main factor behind the difference is found in the number of violations of dominance. The pattern found in the laboratory replicates results reported by Loomes, Moffatt, and Sugden (2002) on a different risky choice design, namely very few dominance violations and rather frequent inconsistencies when faced with identical decision problems twice. For the general population, this changes dramatically.

Another important question is whether providing monetary incentives makes

Tab. 4.5: Average Number of Answers that Violate Monotonicity or Dominance

Sample	All Sheets	Sheet 1	Sheet 2	Sheet 3	Sheet 4	Sheet 5	Sheet 6	Sheet 7	N
Laboratory									
All	1.21	0.17	0.20	0.09	0.23	0.13	0.18	0.22	178
Lab-Lab	1.29	0.18	0.19	0.12	0.24	0.13	0.16	0.27	90
Lab-Internet	1.14	0.16	0.20	0.06	0.22	0.13	0.20	0.17	88
Order 0	1.28	0.15	0.17	0.06	0.29	0.13	0.18	0.30	87
Order 1	1.15	0.19	0.22	0.12	0.18	0.13	0.18	0.14	91
Incentive 0	1.32	0.17	0.20	0.12	0.29	0.12	0.20	0.22	65
Incentive 1	1.23	0.19	0.21	0.06	0.19	0.17	0.21	0.21	48
Incentive 3	1.09	0.15	0.18	0.08	0.20	0.11	0.14	0.23	65
CentERpanel									
All	2.43	0.34	0.37	0.31	0.39	0.34	0.34	0.35	1789
Uni	1.28	0.22	0.16	0.18	0.22	0.15	0.18	0.19	96
Order 0	2.54	0.28	0.35	0.32	0.44	0.39	0.31	0.45	913
Order 1	2.32	0.41	0.38	0.30	0.33	0.28	0.37	0.24	876
Incentive 0	2.49	0.36	0.38	0.29	0.43	0.34	0.35	0.33	555
Incentive 1	2.41	0.35	0.37	0.31	0.36	0.32	0.34	0.35	660
Incentive 3	2.40	0.32	0.34	0.31	0.39	0.35	0.33	0.36	574
CentERpanel - Weighted									
Incentive 0	2.59	0.38	0.40	0.30	0.44	0.36	0.36	0.35	555
Incentive 1	2.70	0.38	0.42	0.35	0.41	0.37	0.38	0.39	659
Incentive 3	2.50	0.35	0.34	0.33	0.40	0.37	0.33	0.38	573

Note: A choice pattern is classified as violating monotonicity or dominance if a subject switched back and forth between options ‘A’ and ‘B’ or if he chose option ‘A’ for the sure high outcome. No more than one violation is counted for each of the seven decision tasks.

subjects take more care in answering the questions and make fewer errors. Generally, this is not the case, as can be seen by comparing the percentage of violations across the different incentives treatments in Table 4.5. Also comparing the differences between screens across questions, we cannot report significant differences using both tests. Finally, we note that the ordering of screens does seem to have an impact on the number of violations. Averaging over all sheets, we can detect a marginally lower error rate for order treatment 1 in the CentERpanel, but not in the laboratory. On the other hand, differences of several individual screens show up as significant. We do not find any clear pattern in terms of learning, fatigue, or screen sequencing, however. Hence there seems to be another direction of how framing effects may matter besides inducing different levels of risk aversion (Andersen, Harrison, Lau, and Rutström 2006) — different probabilities to exhibit inconsistent behaviour.

To assess implications of the selection process sketched in section 4.3.3 on the observed distribution of preferences, the characteristics of the non-respondents have to be taken into account. Using the weights described in section 4.3.3, we conclude from the figures in Table 4.5 that not correcting for selection would lead to a slight underestimation of the number of violations. T-tests reveal a strongly significant effect overall (two-sided p-value=0.0008). Breaking this down individually by incentive treatment, we find that it is mostly driven by the low incentives group. These results go in the direction that one would expect from the the findings reported in section 4.3.3 for the education variables, since one would expect better educated people to make fewer errors. For the other variables, we did not have clear priors.

4.4.2 Preferences in the Lab vs. the CentERpanel

To obtain a crude measure of preferences we consider at which probability of the high outcome subjects switched from (the safe) option A to (the risky) option B. However, as highlighted in the previous section, a considerable amount of participants made inconsistent choices, either by switching back and forth on the same screen, or by choosing a dominated option. This implies that not all subjects have a unique switch point, and we need to define what we label a lowest switch point and a highest switch point. The lowest possible switch point is defined by the probability corresponding to the highest 'A' choice that still is lower than the minimum 'B' choice. We define the highest switch point as the minimum probability observed for option 'B' which is still higher than the maximum probability where option 'A' was chosen. If only choice 'A' ('B') was observed, both switch points were set to one (zero). To get a single measure we also calculated the mean of these two switch points for each individual.

Table 4.6 displays the average switch points by sample and treatment. A higher number indicates that subjects, on average, switched at a higher probability for the high outcome to occur, thus indicating a more risk averse behaviour. Note, that except for the rows labelled "Incentive 1", subjects in the low incentive treatment were excluded since they faced a different payoff scale which renders comparisons of switch points misleading if pooled with other subjects. A second note of caution concerns the treatment of errors which cannot be entirely separated from preferences. On the one hand, monotonicity violations lead to a large difference between the lowest and the highest switch point. We cannot judge which is the right one in this interval. Hence, comparing two mean switch points may lead to false conclusions if the intervals they are based on have different lengths.[5]

[5] A numerical example might help to make the point clearer. Assume that person 1 exhibited consistent behaviour with a low switch point of 0.6 and the high value at 0.7. She is compared with a second person who chose a pattern 'ABAB' on the first screen, leading to a low switch point of 0.25 and a high switch point of 1. A ranking in terms of preferences based on mean switch points would be seriously misleading since the interval revealed by person 1 is completed covered by the second persons possible set of switch points.

For a valid ranking of preferences both sets of lowest and highest switch points must rank in the same way. On the other hand, this ranking may be affected by dominance violations where both switch points are the same. We checked that all results reported in this section also go through if switch points are set to missing in these cases. Having said this, we often stick to comparisons of mean switch points for expositional reasons since the above-mentioned effects did not have much empirical relevance.

Comparing the answers between the laboratory and the Internet (first rows in the respective sections of Table 4.6), it is evident that there are some noteworthy differences in the observed behaviour. The average switch points in the Internet experiment are visibly higher than the corresponding figures for the laboratory experiments. Moreover, the difference between the two samples is found across all sheets. Consequently, the degree of risk aversion appears to be lower among the students in the laboratory than among the participants of the Internet experiment. Using the MW and KS we find that the difference between the laboratory and Internet samples is highly significant irrespective of which average switch point we use. That is, no matter whether we look at the lowest or highest switch point and whether we compare averages across all questions or look at each question separately, there is a significant difference between the choices made by the two samples.

To disentangle the different explanations for these observed differences, we proceed as in section 4.4.1 by comparing the average switch points in the “Lab-Lab" treatment and the“Lab-Internet" treatment. The mean switch points are slightly higher in the “Lab-Lab" treatment suggesting that the observed difference between the laboratory and the Internet samples is most probable not due to characteristics of the laboratory setting since we would then expect the difference to be in the other direction. There is, however no clear systematic difference between the two treatments considering each screen separately. Comparing the treatments using the MW and KS tests reveals that we are not able to soundly reject the null hypothesis that there is no difference between the samples for any of the measures.

Just as in the case of inconsistencies, behaviour of the Internet participants gets close to the student sample when restricting attention to the “Uni" subsample. The mean switch point of 65.1 aggregating over all screens of the “Uni" sample presented in Table 4.6 is closer to the laboratory mean of 61.5. When looking at the separate price lists the averages are comparable in most cases. Using the MW and KS tests on individual mean switch points the differences are insignificant except for sheets 3 and 5. Controlling for subject group differences hence eliminates most of the differences between the samples. The disparity found between both errors and preferences in the Internet sample and the laboratory experiments thus seems to be driven by the fact that behaviour of the homogenous student differs from the

Tab. 4.6: Mean, lowest, and highest average switch points by subsample and treatment

Sample	All Sheets	Sheet 1	Sheet 2	Sheet 3	Sheet 4	Sheet 5	Sheet 6	Sheet 7
Laboratory								
All	**61.5**	**63.0**	**57.8**	**63.5**	**53.9**	**67.7**	**63.3**	**61.8**
	55.7 67.4	57.2 68.8	51.8 63.7	58.5 68.5	47.3 60.5	62.6 72.7	57.5 69.0	54.8 68.8
Lab-Lab	**60.3**	**60.4**	**58.8**	**61.1**	**51.7**	**66.5**	**63.0**	**60.6**
	54.4 66.2	54.2 66.5	53.3 64.3	55.9 66.3	44.4 59.0	61.8 71.3	57.7 68.2	53.5 67.7
Lab-Internet	**63.1**	**66.1**	**56.5**	**66.3**	**56.5**	**69.0**	**63.6**	**63.2**
	57.2 68.9	60.7 71.5	50.1 63.0	61.5 71.1	50.8 62.3	63.6 74.5	57.3 70.0	56.4 70.1
Order 0	**59.7**	**68.2**	**55.5**	**64.3**	**52.9**	**63.4**	**57.3**	**56.0**
	54.0 65.4	63.1 73.4	50.6 60.5	59.4 69.1	46.4 59.5	58.4 68.4	51.7 62.9	48.0 64.0
Order 1	**63.4**	**57.8**	**59.9**	**62.7**	**54.8**	**71.8**	**69.1**	**67.3**
	57.3 69.4	51.4 64.3	53.0 66.8	57.6 67.9	48.1 61.5	66.7 77.0	63.1 75.0	61.3 73.4
Incentive 0	**61.3**	**63.8**	**57.7**	**63.7**	**50.3**	**67.8**	**62.7**	**62.8**
	55.4 67.2	58.1 69.6	51.7 63.6	58.9 68.5	43.6 57.1	62.8 72.8	56.7 68.8	55.7 69.9
Incentive 1	**58.2**	**58.6**	**51.3**	**61.6**	**54.8**	**64.2**	**59.4**	**57.4**
	52.1 64.3	53.2 64.1	44.5 58.0	56.4 66.9	48.9 60.7	57.5 70.8	53.9 65.0	50.5 64.4
Incentive 3	**61.8**	**62.1**	**57.9**	**63.3**	**57.4**	**67.5**	**63.8**	**60.7**
	55.9 67.7	56.2 67.9	52.0 63.8	58.0 68.5	50.9 63.9	62.3 72.6	58.3 69.3	53.8 67.6
CentERpanel								
All	**70.4**	**70.8**	**64.8**	**75.5**	**62.2**	**78.2**	**71.1**	**69.9**
	63.7 77.0	63.9 77.7	57.8 71.8	69.9 81.1	54.7 69.8	72.3 84.0	65.1 77.1	62.3 77.5
Uni	**65.1**	**65.9**	**59.0**	**71.7**	**55.9**	**73.7**	**65.6**	**64.1**
	59.1 71.1	60.5 71.4	53.4 64.5	66.0 77.3	49.7 62.1	67.1 80.3	59.5 71.8	57.4 70.7
Order 0	**68.7**	**76.7**	**63.1**	**76.6**	**61.9**	**76.1**	**64.3**	**62.5**
	62.2 75.3	71.7 81.6	56.7 69.5	71.4 81.9	54.8 69.1	69.9 82.2	58.2 70.5	53.0 72.0
Order 1	**72.1**	**64.3**	**66.7**	**74.3**	**62.5**	**80.5**	**78.5**	**78.0**
	65.3 78.9	55.3 73.3	59.0 74.4	68.3 80.3	54.5 70.5	75.0 86.0	72.6 84.4	72.4 83.5
Incentive 0	**70.6**	**71.7**	**64.0**	**76.2**	**62.0**	**78.8**	**71.7**	**70.0**
	64.0 77.2	64.7 78.7	57.2 70.8	70.7 81.7	54.6 69.5	72.9 84.6	65.8 77.5	62.4 77.6
Incentive 1	**68.0**	**68.8**	**63.3**	**71.0**	**62.1**	**73.6**	**69.6**	**67.7**
	61.2 74.8	61.5 76.1	55.6 71.0	65.0 77.1	55.0 69.1	67.4 79.9	63.5 75.6	60.4 74.9
Incentive 3	**70.1**	**69.9**	**65.6**	**74.9**	**62.4**	**77.6**	**70.5**	**69.8**
	63.4 76.8	63.1 76.7	58.4 72.8	69.2 80.5	54.7 70.1	71.7 83.5	64.3 76.8	62.2 77.4
CentERpanel - Weighted								
Incentive 0	**71.3**	**72.5**	**64.5**	**76.7**	**63.1**	**79.5**	**72.6**	**70.0**
	64.5 78.1	65.2 79.8	57.8 71.2	70.9 82.5	55.5 70.6	73.3 85.7	66.6 78.7	62.1 77.9
Incentive 1	**69.2**	**69.9**	**64.5**	**72.5**	**63.2**	**74.6**	**70.6**	**68.8**
	62.2 76.2	62.8 77.0	56.7 72.4	66.3 78.8	55.9 70.6	68.1 81.1	64.2 76.9	61.2 76.4
Incentive 3	**71.6**	**71.6**	**67.4**	**76.1**	**63.8**	**78.8**	**72.6**	**71.1**
	64.8 78.4	64.7 78.5	60.0 74.8	70.4 81.7	55.5 72.0	73.0 84.6	66.5 78.7	63.4 78.8

Note: Bold figures indicate the mean of the lowest and the highest average switch point, which are listed below it. The lowest switch point is defined as the maximum probability where option 'A' was chosen that is lower than the minimum observed probability of option B. Accordingly, the highest switch point is defined as the minimum probability observed for option B which is higher than the maximum probability where option A was chosen. If only choice 'A' ('B') was observed, both switch points were set to one (zero). Only subjects with duration longer than 5:20 minutes are included. Except for the lines labelled "Incentive 1", individuals from the incentive 1 treatment are dropped for comparability reasons. Group "Uni" refers to the subsample of the CentERpanel sample between 18 and 34 years of age who hold a university degree or are in the process of obtaining it. Sample sizes are the same as those indicated in Table 4.5.

behaviour of the general population.

Turning to the issue of incentives, the choices in the hypothetical treatment and high incentive treatments, summarised in Table 4.6, are strikingly similar. For all questions, in both the Internet and laboratory experiments, the average switch points are quite similar. Providing monetary incentives does not seem to affect the behaviour in any systematic way. This is confirmed when running the MW and KS tests on average switch points. As in Holt and Laury (2002), the tripling of incentives (and risk for that matter, loosely speaking) does not increase switch points as strong as typical preference functionals would predict. Finally, we note a similar finding on order effects as in the case of monotonicity and dominance violations. There are differences, but we cannot relate them to any of the explanations typically brought forward. In a structural model of choice behaviour, they will have to be controlled for by statistical methods.

The comparisons above highlight the impact of demographic characteristics of the subject pool on the distribution of elicited preferences. To obtain estimates valid for the general population, it is hence important to take into account the fact — discussed earlier — that participation is voluntarily and that selection into the experiment is not a random process. The summary statistics for the weighted sample are given in Table 4.6. There seems to be a slight underestimation of the mean switch points when using the unweighted sample. Using t-tests we can, however, not reject that the mean switch points for the weighted and un-weighted sample are equal irrespective of which measure we use. Although the participation decision is correlated with demographics, selection based on observables does not fundamentally alter the results on this crude measure of risk preferences.

4.5 Conclusions

We have presented evidence on different aspects of the representativeness of preference elicitation experiments. First, we looked at selection effects that may arise from voluntary participation. We concluded that selection based on observable characteristics does matter for violations of monotonicity and dominance, but seems to be a minor issue for risk preferences. We also found evidence that participation is related to variables that proxy financial knowledge and expertise, that probably are also associated with preferences to deal with risky choice problems. This suggests that nonparticipation may also be related to unobservable respondent characteristics, and we cannot exclude the possibility that such unobservable characteristics are also associated with risk preferences. We think it is difficult to find valid instruments for a sample selection model that would control for such selection effects. As a consequence, some caution must be taken in interpreting the results from experiments with voluntary participation in terms of representativeness. Second, we tried to disentangle environmental effects (laboratory vs. Internet) from traditional subject pool bias. On the one hand, we showed that the implementation mode does not matter for either errors or preference estimates among the young and educated. On the other hand, we found dramatic differences in the number of dominance and monotonicity violations when moving to a sample drawn from the general population. Subjects in the latter also exhibited a higher degree of risk aversion.

BIBLIOGRAPHY

ALESSIE, R., S. HOCHGUERTEL, AND A. VAN SOEST (2006): "Non-Take-Up of Tax-Favored Savings Plans: Evidence from Dutch Employees," *Journal of Economic Psychology*, 27, 483–501.

ANDERSEN, S., G. W. HARRISON, M. I. LAU, AND E. E. RUTSTRÖM (2005): "Preference Heterogeneity in Experiments: Comparing the Field and Lab," Centre for Economic and Business Research Discussion Paper No. 2005-03, Copenhagen.

——— (2006): "Elicitation Using Multiple Price List Formats," *Experimental Economics*, 9(4), 383–405.

BEATTIE, J., AND G. LOOMES (1997): "The Impact of Incentives on Risky Choice Experiments," *Journal of Risk and Uncertainty*, 14, 155–168.

BELLEMARE, C., AND S. KRÖGER (2007): "On Representative Social Capital," *European Economic Review*, 51(1), 183–202.

BINSWANGER, H. P. (1980): "Attitudes Towards Risk: An Experimental Measurement in Rural India," *American Journal of Agricultural Economics*, 62, 395–407.

DOHMEN, T., A. FALK, D. HUFFMAN, U. SUNDE, J. SCHUPP, AND G. G. WAGNER (2005): "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey," IZA Discussion Paper 1730, Institute for the Study of Labor (IZA).

DONKERS, B., B. MELENBERG, AND A. VAN SOEST (2001): "Estimating Risk Attitudes Using Lotteries; A Large Sample Approach," *Journal of Risk and Uncertainty*, 22(2), 165–195.

FEHR, E., U. FISCHBACHER, B. VON ROSENBLADT, J. SCHUPP, AND G. G. WAGNER (2003): "A Nation-Wide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys," IZA Discussion Paper 715, Institute for the Study of Labor (IZA).

GNEEZY, U., AND A. RUSTICHINI (2000): "Pay Enough or Don't Pay at All," *Quarterly Journal of Economics*, 115(3), 791–810.

Güth, W., C. Schmidt, and M. Sutter (2006): "Bargaining Outside The Lab – A Newspaper Experiment Of A Three Person-Ultimatum Game," *Economic Journal*, forthcoming.

Harrison, G. W., M. I. Lau, and E. E. Rutström (2005): "Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments," University of Central Florida Economics Working Paper No. 05-01.

Harrison, G. W., M. I. Lau, and M. B. Williams (2002): "Estimating Discount Rates in Denmark: A Field Experiment," *American Economic Review*, 92, 1606–1617.

Harrison, G. W., and J. A. List (2004): "Field Experiments," *Journal of Economic Literature*, 42(4), 1009–1055.

Harrison, G. W., J. A. List, and C. Towe (2007): "Naturally Occurring Preferences and Exogenous Laboratory Experiments: A Case Study of Risk Aversion," *Econometrica*, 75(2), 433–458.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153–61.

Holt, C. A., and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Kahneman, D. V., and A. V. Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–291.

Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

Lazear, E., U. Malmendier, and R. Weber (2006): "Sorting in Experiments with Application to Social Preferences," NBER Working Paper 12041.

Little, R. J., and D. B. Rubin (2002): *Statistical Analysis with Missing Data*. John Wiley & Sons Inc., New York, 2nd edn.

Loomes, G., P. G. Moffatt, and R. Sugden (2002): "A Microeconometric Test of Alternative Stochastic Theories of Risky Choice," *Journal of Risk and Uncertainty*, 24(2), 103–130.

Lucking-Reiley, D. (1999): "Using Field Experiments to Test Equivalence between Auction Formats: Magic on the Internet," *American Economic Review*, 89(5), 1063–1080.

Madrian, B. C., and D. F. Shea (2001): "The Power of Suggestion: Inertia in 401(K) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116(4), 1149–1187.

5. RISK PREFERENCES IN THE SMALL FOR A LARGE POPULATION

JOINT WITH ARTHUR VAN SOEST AND ERIK WENGSTRÖM

5.1 Introduction

Characterising decision behaviour under risk and uncertainty is one of the most prevalent themes in economics. Its pervasiveness is driven by the forward-looking nature of economic decisions and the observation that the convenient assumption of risk neutrality often provides a poor description of actual behaviour. Under nearly all circumstances, the use of observational data for investigating risky choice behaviour is only feasible under auxiliary identifying assumptions such as rational expectations or complete markets. As such restrictions often seem difficult to justify, empirical studies have been largely confined to laboratory settings. This allows the researcher to isolate the decision problem of interest. It has the drawback that interest usually lies in economically "large" quantities, while the standard practice in experimental economics with real monetary incentives is to keep the stakes quite small. Despite the negative extrapolation results of Rabin (2000), there certainly is a lot left to learn about regularities in human decision behaviour from such studies.

The early wave of experiments following Allais' (1953) seminal study was mainly devoted to documenting violations of the von Neumann-Morgenstern axioms. These findings spurred the development of numerous new models of choice under uncertainty capable of accounting for some of the empirical departures from expected utility. Machina (1987) and Starmer (2000) survey many of these developments. Following them, a second wave of experiments have put effort into distinguishing between different models, Camerer (1989) is an early example.

Quite recently the literature has witnessed a shift from testing theories toward estimating preference parameters. Holt and Laury (2002) investigate the relation between risk aversion and incentive effects using the same experimental design that we will employ in this study (multiple price lists). They report that subjects exhibit more risk aversion as measured by the Arrow-Pratt coefficient of relative risk aversion when payoffs are scaled up and actually paid in cash. Even when faced with very low incentives, most subjects are risk averse.

The multiple price list methodology was taken to the field by Harrison, Lau, and Rutström (2007), who estimated risk attitudes on a representative sample of the Danish population. They report that the average Dane is risk averse, but that the degree of relative risk aversion is nearly constant on the range of payoffs they consider. Risk attitudes are found to vary significantly with demographic variables such as age and education. Other studies have also emphasised the importance of taking heterogeneity into account and are too numerous to review completely here. As an example, Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2005) find that risk attitudes are related to background variables such as age and sex. They reach this finding by exploiting answers to a hypothetical general risk question which they compare to a risky choice experiment for a subsample of respondents. Females and older people are less tolerant towards risk. In a sample of Chilean high school students, Benjamin, Brown, and Shapiro (2006) demonstrate that small-stakes risk aversion is less common among those with higher standardised scores on tests that measure cognitive ability.

As noted before, recent decades have witnessed a boost in extensions to the classical expected utility model. One of the most influential of them has been prospect theory (Kahneman and Tversky 1979). Arguably its most prominent characteristic is loss aversion and it has become a widely established finding that – on average – agents are more sensitive to losses than to gains. Until very recently, studies of loss aversion focused on these mean parameters as opposed to estimating degrees of heterogeneity, no matter whether it relates to observables or is of idiosyncratic nature. Harrison and Rutström (2006) include negative outcomes (relative to an initial endowment) in their design. They estimate a mixture model allowing for for agents to behave along the lines of either prospect theory or expected utility theory. They find evidence that the probability of behaving according to either one is related to background characteristics of the subjects. Johnson, Gächter, and Herrmann (2006) explore loss aversion on the individual level using both a telephone mediated market situation as well as choices between gambles. They find loss aversion to vary with demographic variables. It appears to be lower for younger and more educated individuals and stronger for individuals with high levels of incomes and wealth.

A second significant extension to the canonical model consisted of accommodating preferences towards the timing of uncertainty resolution (Kreps and Porteus 1978). Experimental studies involving this model are relatively scarce. Chew and Ho (1994) investigate such preferences using hypothetical scenarios concerning disclosure of exam results and tax refunds (liabilities). Their findings suggest that preferences towards the resolution of uncertainty depend on the possible outcomes of the situation at hand. On the one hand, subjects generally prefer early resolution if losses are possible (anxiety). On the other hand, if the outcome space features only gains, subjects tend to prefer delaying the uncertainty resolution (hope). However, the latter is only the case for disclosure of exam results and not in the tax refund case (where most subjects prefer early resolution). Moreover, they find preferences towards the timing of uncertainty resolution and risk aversion to be uncorrelated. Ahlbrecht and Weber (1996) conduct a study involving hypothetical gambles and scenario-style questions. Among other things, they report that people are not indifferent towards the timing of uncertainty resolution and that most subjects prefer early resolution.

In this paper, we try to bring these lines of the literature together. We take a classical expected utility of income model to the data and extend it to allow for loss aversion and preferences towards the timing of uncertainty resolution. An experiment is conducted on the CentERpanel, a representative sample of Dutch households. The survey is administered via the Internet. In addition to the large sample size, we also obtain a lot of background information already available from previous interviews of the CentERpanel participants. The large sample size combined with the nature of our decision task allow us to specify a rich econometric model. In principle, we can estimate the joint distributions of the three individual-level parameters conditional on a rich set of demographic characteristics. Due to numerical issues, the current version of the paper uses a somewhat less general model than the one we ultimately want to aim for.

Our results can be summarised as follows. As the previous literature, we find strong effects of risk aversion and loss aversion in homogeneous specifications of the model. Augmenting it by the uncertainty resolution parameter produces mixed results. However, we can say with some confidence that a negative prospect in the outcome set significantly reduces the attractiveness of late uncertainty resolution, something that is well in line with earlier psychological findings. As one would expect, we find heterogeneity in risk aversion to be very important. Interestingly, only a very small part of the overall between-subject variation is explained by the most important demographic variables. Among these, age is associated with higher risk aversion. Men show less risk-averse behaviour than women. Finally, education and higher household income levels are linked with lower risk aversion levels. Since our experimental setting involves many choice tasks for each subject, we can relate the propensity to commit errors (as defined relative to individual preference functionals) to demographic characteristics. The first thing to note is that overall error rates are much higher than in laboratory settings with student populations. This appears to be an effect of demographics rather than implementation mode. We find a higher rate to go hand in hand with age, especially for the oldest subjects. Errors are less frequent among the well educated and members of high-income households. Regarding the magnitude of inconsistencies, we find that in the group where errors are less frequent, the predicted rate for errors that cost the utility equivalent of 30 Euros is 5 %. Among the most error-prone subjects, predicted rates are twice as high.

The paper is structured as follows. In the next section, we present a parsimonious specification of the utility functions that form the basis of our analysis. Section 5.3.2 provides a brief introduction to our experimental design and highlights some features of the data that deserve special attention in the econometric implementation. We explain our empirical strategy in Section 5.4 before turning to the results in 5.5. Section 5.6 concludes.

5.2 Theoretical Framework

In this section, we lay out the utility specifications that form the basis of our econometric analysis. As mentioned in the introduction, we start with a simple expected utility of income specification and then augment it by a loss aversion parameter, as in prospect theory. In Section 5.2.2, we introduce a temporal component to the model, which then has two periods. While all payments are made in the second period, uncertainty may be resolved either in the first or the second period. We implement a parsimonious version of the Kreps and Porteus (1978) model in order to allow for preferences towards the timing of uncertainty resolution. All this is done for a deterministic specification of the utility function; we move to a stochastic world when we describe our econometric model in Section 5.4.

5.2.1 A Simple Model of Choice Under Risk

In modelling individual decisions, we start from a standard Expected Utility formulation and gradually build it up to encompass the core features of our study (loss aversion and timing of uncertainty resolution). The basis of our framework is an exponential utility function:

$$u(z, \gamma) = -\frac{1}{\gamma} e^{-\gamma z} \tag{5.1}$$

where $z \in \mathbb{R}$ denotes lottery outcomes and $\gamma \in \mathbb{R}$ is the coefficient of absolute risk aversion. The outcomes in our experiment are framed as losses and gains relative to an initial endowment. In order to allow agents to process losses differentially from gains we augment the formulation in (5.1) with a loss aversion parameter λ. The utility function thus becomes:

$$u(z, \gamma, \lambda) = \begin{cases} -\frac{1}{\gamma} e^{-\gamma z} & \text{for } z \geq 0 \\ \frac{\lambda - 1}{\gamma} - \frac{\lambda}{\gamma} e^{-\gamma z} & \text{for } z < 0 \end{cases} \tag{5.2}$$

With $\lambda > 1$, the utility function is steeper for losses than for gains, reflecting the original prospect theory observation (Kahneman and Tversky 1979) that agents are more sensitive to losses than to gains. Many experimental as well as observational studies have confirmed that this is an important aspect in modelling behaviour under risk (see Starmer (2000) and Camerer (2000) for reviews).

However well established, the precise definition of the notion of loss aversion is not uncontroversial. In this paper, we adopt the definition of Köbberling and Wakker (2005). They define loss aversion as the ratio of the left and the right derivatives of the utility function at zero. The authors also derive implications for parametric families of utility functions and demonstrate that some severe problems are associated with the commonly used power utility formulation. Our specification (5.2) ensures that regardless of the risk attitude parameters, λ can always be interpreted as the loss aversion parameter in their sense. This is one reason for using exponential utility. Another one is that exponential utility is also convenient when extending the model to include preferences over uncertainty resolution (see below).

Our experimental setup does not contain enough variation on the negative domain to enable separate estimates of risk aversion parameters on the negative and positive domains. In (5.2) we have therefore assumed that the risk aversion parameter is constant along the real line. This stands in contrast to the standard prospect-theory specification, where agents are assumed to be risk averse on the positive domain and risk seeking when it comes to negative gambles. The motivation behind our choice of restriction is twofold. First, adopting the same parameter for the whole real line is conservative with respect to the estimation of λ because risk aversion on the negative domain implies ceteris paribus lower values of λ. Put differently, our estimates of the loss aversion parameter represent a lower bound compared to a formulation where agents switch to risk-seeking on the negative domain. Second, the observation of risk-seeking on the negative domain has been made predominantly on the basis of purely negative gambles (cf. Tversky and Kahneman (1992) or Starmer (2000)). Contrary to this, our setup involves gambles with mixed outcomes, and it is always possible to choose a less risky lottery with weakly positive outcomes only. The use of mixed gambles is motivated by the fact that most real economic decision problems involve some possibility of a loss and some possibility of a gain. However, in the light of recent findings by Baltussen, Post, and van Vliet (2006) it seems questionable if the early findings of risk seeking on the negative domain extend to the case of mixed gambles.[1]

5.2.2 Preferences towards the Resolution of Uncertainty

The timing of uncertainty resolution is a key characteristic of many real-world decision problems, ranging from insurance decision to investment choices. Individuals may have different attitudes towards living with uncertainty and investigating the nature of such preferences hence seems highly motivated. Apart from the planning advantage that early resolution brings, several authors have recently

[1] Recognising the widespread use of the prospect theory specification, we also estimate a model with $\gamma^- = -\gamma^+$. That is, we also estimate a model where we restrict the absolute value of the parameter to be same on both the positive and negative domain but the signs to be different. See for example Tversky and Kahneman (1992) who use a similar specification in a power utility function. Note that this formulation requires some rescaling of (5.2) in order to keep $u(\cdot)$ continuous, which is most easily achieved by fixing it at the origin.

highlighted the importance of anticipatory feeling such as hope and anxiety (c.f. (Wu 1999) and (Caplin and Leahy 2001)). In order to model preferences towards the resolution of uncertainty, we adopt the general model of Kreps and Porteus (1978). See also the work by Epstein and Zin (1989). In correspondence with our experimental setup we consider a two-period setting, where all decisions are made in the first period and all payments are carried out in the second period. The outcome of a gamble is either revealed directly after the choices have been made in period one (early resolution) or at the time of the payments in period two (late resolution).

Assume that agents first calculate period two utility for all outcomes based on a function $v(\cdot)$. Thereafter, they use a continuous and strictly increasing weighting function, $h(\cdot)$, to calculate their first period utility, with period two utility as its argument. The period one utility of a degenerate lottery that gives a certain outcome in period two is then simply given by $h(v(z,\cdot))$. The evaluation of nondegenerate lotteries hinges on the timing of uncertainty resolution. Depending on the latter, the expectations operator is applied either to the weighted or unweighted period two utility. Formally, let V denote the period one utility function for gambles, π where the outcomes are paid out in period two. The first period utility evaluation of π is then given by:

$$V(\pi) = \begin{cases} \mathbb{E}[h(v(z,\cdot))] & \text{for early resolution} \\ h(\mathbb{E}[v(z,\cdot)]) & \text{for late resolution} \end{cases} \tag{5.3}$$

Note that the expectations operator is always applied to the quantity that is known at the end of period one. If uncertainty resolves early, the decision-maker may apply the weighting function to the utility of the specific outcomes of π. If the outcome of π remains uncertain until the second period, he can only calculate the moments of $v(z,\cdot)$ in period one and applies the weighting function to its expected value. Kreps and Porteus (1978) show that h is convex (concave, linear) if and only if the decision maker prefers early to late resolution (late to early, exhibits indifference). We choose the following parsimonious implementation for the weighting function:

$$h(v(z,\cdot)) \quad = \quad -S\,(-S\;v(z,\cdot))^{\rho^{-S}} \tag{5.4}$$

with $\rho \in \mathbb{R}_+$ and S denoting the following sign operator:

$$S \quad = \quad \begin{cases} \;\;1 & \text{for } \gamma \geq 0 \\ -1 & \text{for } \gamma < 0. \end{cases} \tag{5.5}$$

For $\rho > 1$, $h(\cdot)$ is convex and early resolution is preferred to late resolution. Indifference is obtained for $\rho = 1$, and late resolution is preferred for $\rho < 1$. We model the second period utility function as a slightly modified version of equation (5.2):

$$v(z, \gamma, \lambda, \rho) = \begin{cases} -\frac{1}{\gamma} e^{-\gamma \rho^S z} & \text{for } z \geq 0 \\ \frac{\lambda - 1}{\gamma} - \frac{\lambda}{\gamma} e^{-\gamma \rho^S z} & \text{for } z < 0 \end{cases} \tag{5.6}$$

Formulation (5.4) in combination with (5.6) allows individuals to be either risk averse or risk seeking. The inclusion of ρ^S in the exponent serves to retain the interpretation of γ as the coefficient of absolute risk aversion for early resolving lotteries on the positive domain — for such lotteries $V(\pi)$ collapses to $\mathbb{E}[u(\pi)]$ from equation (5.2). This makes clear why the distinction between risk aversion and uncertainty resolution preferences is identified. Unfortunately, there is no such result for gambles with negative outcomes because of the additive term $\frac{\lambda-1}{\gamma}$ in (5.6) that is needed to keep $v(z, \cdot)$ continuous in the presence of loss aversion. In the model consisting of (5.3), (5.4), and (5.6), the effect introduced by this term is negligible for the parameter values of γ, λ, and ρ that we estimate.[2]

5.3 Data and Experimental Setup

In this section we are concerned with a brief description of our data and the experimental design, focussing on the issues that are most relevant for the econometric analysis. A detailed description of both the experimental setup and the data that we use is provided elsewhere (von Gaudecker, van Soest, and Wengström 2007).

We implement our experiment on the CentERpanel, a Dutch household survey that is administered via the Internet. In order to avoid selection problems due to lack of Internet access, respondents without a computer are equipped with a set-top box for their television set. The panel consists of roughly 2,000 households who are representative of the Dutch population in terms of observable characteristics. Hence there is rich variation in and background information on important demographic and socio-economic characteristics that we describe in detail in the above-cited paper. Our experiment was administered to some 2,299 individuals in November and December of 2005. Respondents are reimbursed regularly for their expenses connected with Internet use and we could use the existing system for reimbursement to make payments to the subjects who participated in the experiment.

[2] This is not true anymore if we assumed agents to have prospect theory type preferences with switching risk attitudes between the positive and negative domains. Hence we do not estimate such a specification.

Tab. 5.1: Payoffs from the Seven Lotteries

Sheet	Uncertainty resolution	Payoff Low, A	Payoff High, A	Uncertainty resolution	Payoff Low, B	Payoff High, B
1	early	27	33	early	0	69
2	early	39	48	early	9	87
3	early	12	15	early	-15	48
4	early	33	36	late	6	69
5	early	18	21	late	-9	54
6	early	24	27	early	-3	60
7	late	15	18	late	-12	51

Note: These values were shown in the high incentive and hypothetical treatments. For the low incentive treatment they were divided by 3. Order 0 consisted of the sequence of screens $\{2,7,3,6,1,4,5\}$ and order 1 of the sequence $\{1,4,5,2,6,3,7\}$.

5.3.1 Experimental Design

In order to integrate all features of decisions under risk that are necessary for estimating the parameters of the utility function, we opted for an adapted version of the multiple price list format (see Binswanger (1980) for its introduction to economics, Holt and Laury (2002) and Harrison, Lau, and Rutström (2007) for recent applications and Andersen, Harrison, Lau, and Rutström (2006) for a detailed description). Very briefly, multiple price lists work as follows. Each subject is presented a list of four pairs of lotteries. We call these pairs option 'A' and option 'B', where the latter is always the riskier alternative. Subjects may opt for either option in each of the four choice tasks. The payoffs of both options do not change, but the probabilities for the respective high payoff vary from 25 % to 100 % as one moves down the screen. The table is designed such that the expected value of option 'A' starts out higher but moves up slower than the corresponding figure of option 'B'. In total, we confronted subjects with seven such screens which are listed in Table 5.1.

If participants behave according to any utility theory that we are aware of, they should switch at some point from option 'A' to option 'B' (or choose option 'B' on the whole screen). If such consistent behaviour is observed, the subject is routed to a screen containing lotteries with the same payoffs, but a finer probability grid. Andersen, Harrison, Lau, and Rutström (2006) recommend using this method and call it "iterated multiple price list". The grid now consists of steps of 10 percentage points located roughly between the subject's highest choice of 'A' and his lowest choice of 'B'. Given this design, we obtain an unbalanced panel of between 28 and 56 binary choices for each respondent.

We tried to limit the cognitive effort by including pie-chart representations of the probabilities and by using multiple price lists with only four items instead of the standard ten. We tested the instructions thoroughly. Unlike in typical experimental settings, there was no experimenter to answer questions. In order to compensate for this, subjects had access to the instructions and specially designed help screens throughout.

We implemented six treatments in a 3×2 design along the incentives and order dimensions. Two treatments involved real incentives. We call the first one the high incentive treatment with payoffs as in Table 5.1. For the other group with real incentives, payoffs were divided by three. We term this the low incentive treatment. After the experiment, one of the 28-56 choices was selected at random and the lottery was played out for one in ten participants. The remaining incentive treatment was entirely hypothetical. Values shown on the screens were identical to those in the high incentive group. We randomly determined the two orderings of the screens that are listed in the legend of Table 5.1.

Since our theoretical model involves a loss aversion term, we incorporated gambles with negative outcomes on some of the screens. There are two things to note in this respect. First, we wanted to avoid negative overall payoffs and paid everybody in the real incentive treatments a participation fee that was equal to the maximum loss which could be incurred subsequently.[3] The potential drawback of this approach is an asymmetry of findings with respect to the loss aversion parameter: If we did not find it to be important, this could be because people decide on the basis of final earnings from the entire experiment. Anticipating the results from Section 5.5, this does not seem to be an issue. Second, the less risky option 'A' always contained only weakly positive outcomes, so that subjects were able to avoid losses altogether. Hence they were faced with mixed gambles which raises some comparability issues with respect to the earlier experimental literature on prospect theory. As noted above in Section 5.2.1, this had been predominantly concerned with gambles having purely negative outcomes in hypothetical settings.

Finally, all payoffs were made three months after the experiment. Hence we were able to reproduce the two-period setting of Section 5.2.2 by setting the timing of uncertainty resolution to either directly after the experiment; or just before the payment was actually made. We took special care in order to make clear to subjects that this concerned only the timing of uncertainty resolution and not the timing of payment.

[3] Note that everybody in the real incentive treatments received this participation fee, not only the ones that were actually chosen for playing out the lotteries later on. The upfront payment was only mentioned in the introductory screen, but not thereafter.

5.3.2 Descriptive Evidence

At this point, we want to highlight some of the features that our experimental data exhibits, thereby setting the scene for the structural analysis in the remainder of the paper. From the 2,299 persons who logged into the system, in this paper we make use of data from 1,789 subjects who completed the entire questionnaire within a reasonable time.[4] In total, we observe 98,108 binary decisions. We merely sketch the most important points that are relevant for the subsequent analysis. See von Gaudecker, van Soest, and Wengström (2007) for a more detailed analysis.

We first turn to the issue of errors. At this point, there are two types of behaviour that we classify as being inconsistent with utility theories: Switching back and forth between options 'A' and 'B' within one of the seven screens; and choosing a dominated option, i.e. option 'A' if the probability for the high payoff is one, or option 'B' if the probability for the low payoff is one. We term these errors monotonicity and dominance violations, respectively.

We find non-monotonic behaviour in some 21 % of the seven screens, dominance violations can be detected in 15 % of these. Note that both categories are not mutually exclusive, but overlap is not all too important. While the first number is broadly consistent with earlier evidence on behaviour of students in repeated choice tasks (see Loomes (2005) and the references cited therein), the number of dominance violations is much higher than what is ordinarily observed in laboratory experiments. In order to investigate whether this is to be attributed to poor experimental design on our behalf or rather to the richer demographic variation of the CentERpanel respondents we restrict attention to a subgroup of the latter who are comparable to typical experimental subjects. In particular, we select 96 individuals who study at a university or hold a degree and are less than 35 years old. The results are encouraging – dominance violations drop by two thirds and are now in the range of what is typically observed in the laboratory. We conclude that it is yet more important to model the stochastic component of utility if one deals with a representative population. While this is certainly not all too surprising, it is important to keep in mind when interpreting the results in Section 5.5.

As a very crude way of comparing decision behaviour across the different characteristics of the lotteries we consider the mean switch points on each screen. The mean switch point is calculated as the mean of the lowest and the highest observed switch point. The lowest switch point is defined as the maximum probability where option 'A' was chosen that is lower than the minimum observed probability of option 'B'. Accordingly, the highest switch point is defined as the

[4] Some respondents chose not to participate after the introductory screen which contained an explicit non-participation option, others dropped out along the way. Finally, we excluded those who went through the whole experiment in less than 5:20 minutes. See von Gaudecker, van Soest, and Wengström (2007) for more details and an investigation of selection issues.

minimum probability observed for option B which is higher than the maximum probability where option A was chosen. If only choice 'A' ('B') was observed, both switch points were set to one (zero). Note that the two may differ quite strongly because of nonmonotonicities.

First note that in order to not confound risk preferences that vary across different ranges of payoffs (Holt and Laury 2002) with the other preference parameters, we restricted attention to a relatively narrow scale of payoffs. Very loosely speaking, if utility was based on traditional risk attitude alone, we would expect broadly similar behaviour across screens. This is certainly not the case. In terms of mean switch points, the screens rank as follows from the highest to the lowest fraction of risk-averse choices: $5, 3, 6, 1, 7, 2, 4$.

With respect to loss aversion, in general we observe a higher fraction of risk-averse choices on the screens with include options with the possibility of negative payouts (screens $5, 3, 6, 7$) than on the others. The picture is less clear-cut with respect to preferences towards the timing of uncertainty resolution – we find the two lotteries where only option 'B' resolves late (screens 5 and 4) at opposite ends of the spectrum and the one where all lotteries resolve late in the middle (screen 7). A very tentative interpretation of this may be that there exist interactions between the possibility of a negative payoff and uncertainty resolution preferences – if outcomes are weakly positive, subjects may be indifferent or even prefer late resolution. On the other hand, if there is the possibility of bad news, they may want to hear it now. Such behaviour could be rationalised, for example, by some of the arguments in Caplin and Leahy (2001) that are based on earlier evidence collected by psychologists (Cook and Barnes (1964), Loewenstein (1987)). Incentive effects do not seem to be very important for mean switch points. In particular, they are slightly lower for the low incentive treatment than for the high incentive and hypothetical treatments. This mimics findings by Holt and Laury (2002), who find that switch points become higher as stakes are increased by scalar multiplication. Finally, we do observe order effects with order 1 generating more risk averse choices than order 0. However, we cannot relate their pattern to any of the explanations typically brought forward, so we will treat them as a fixed effect in the parameter estimates.

5.4 Econometric Specification

We estimate structural econometric models by maximum likelihood methods. In principle, the model is sufficiently general to allow for individual heterogeneity in preference and error parameters that may vary with observed characteristics of the individual. As we shall see shortly, estimation of even much less general versions is a formidable task and we are not yet able to estimate the most general model.

First define a binary variable Y_{ij} which is one if individual $i \in \{1, \ldots, N\}$ chooses option B in decision task $j \in \{1, \ldots, J_i\}$ and zero otherwise. Note that the number of decision tasks varies with the individual because of the routing. Subjects answered those screens with finer probability grids only upon making a monotone and non-dominated set of choices in the first round. Now define the utility difference in any decision task as:

$$\Delta\mathrm{CE}_{ij} = \mathrm{CE}(\pi^B_{ij}, \gamma_i, \lambda_i, \rho_i) - \mathrm{CE}(\pi^A_{ij}, \gamma_i, \lambda_i, \rho_i),$$

where $\mathrm{CE}(\pi^k_{ij}, \cdot)$, $k = A, B$ is the period one certainty equivalent of lottery π^k_{ij} given the utility function in (5.3) in combination with (5.4) and (5.6). For the inversion of $v(\cdot)$ and $h(\cdot)$ that lead to it, please refer to the Appendix. One reason for choosing to model decisions in terms of certainty equivalents differences over using utility differences directly is to bound the maximum difference between options B and A. Over all our seven screens it has a minimum of -30 and a maximum of 39 for all individuals, regardless of the utility function parameters. Using utility differences directly would lead to very large differences between these values which would depend on the scaling of the utility function.

In moving to a stochastic specification of utility, we follow the approach taken in Hey and Orme (1994) and model the individual's choice as:

$$Y_{ij} \quad = \quad \mathbb{I}\left\{\Delta\mathrm{CE}_{ij} + \tau_i\, \varepsilon_{ij} > 0\right\}, \tag{5.7}$$

where $\mathbb{I}\{\cdot\}$ denotes an indicator function. Throughout the analysis, we assume ε_{ij} to follow a standard logistic distribution. In contrast to most binary choice models, the scale of ΔCE does has a meaningful interpretation in monetary terms here. Hence the parameter $\tau_i \in \mathbb{R}_+$ governs the probability for individuals to make errors relative to their utility function parameters that decrease in ΔCE (Hey and Orme 1994). These are often termed Fechner errors and we include them in all the models that we estimate. Defining them in relation to certainty equivalents gives them an intuitive interpretation in terms of the monetary costs of errors upon which we will comment below. Subjects may differ in their probability to commit Fechner errors, but we do not allow the error distribution to vary over decision tasks within subjects. For extensive discussions of various error specifications, please refer to Ballinger and Wilcox (1997), Loomes, Moffatt, and Sugden (2002), Loomes (2005), or Hey (2005).

We can now write the likelihood contribution of choice j by individual i as:

$$l_{ij} \quad = \quad \Lambda\left((2Y_{ij} - 1)\frac{1}{\tau_i}\Delta\mathrm{CE}_{ij}\right), \tag{5.8}$$

where $\Lambda(\cdot)$ stands for the cumulative standard logistic distribution. A random coefficients model is our choice to specify a distribution of preference and error parameters that may vary with observed characteristics. Define

$$\eta_i = X_{ij}^{\eta}\beta^{\eta} + \xi_i^{\eta}$$

for J_i x K^{η} regressor matrices X^{η} and $\eta = \gamma, \lambda, \rho, \tau$. For ease of notation, collect the latent terms in $\xi_i = (\xi_i^{\gamma}, \xi_i^{\lambda}, \xi_i^{\rho}, \xi_i^{\tau})'$ and define $M = \dim(\xi_i)$. We assume these errors to be jointly normal with mean zero (which is without loss of generality since X_{ij}^{η} includes a constant for all η) and variance-covariance matrix $\Sigma'\Sigma$, where Σ denotes an upper triangular matrix of Cholesky-factors. Defining $\xi^* = (\Sigma')^{-1}\xi$ we can express the individual likelihood contributions as:

$$l_i \quad = \quad \int_{\mathbb{R}^M} \left[\prod_{j=1}^{J_i} l_{ij}(\cdot, \xi^*) \right] \phi(\xi^*) d\xi^* \tag{5.9}$$

where $\phi(\cdot)$ denotes the standard normal probability density function. Once this quantity is computed, its logarithm can be summed over all individuals which is then maximised by standard methods.

The integral in equation (5.9) does not have an analytical solution and has to be approximated numerically. This is a serious challenge for numerical techniques because it is characterised by very sharp peaks and long tails with zero conditional probability over most of the support where the population density $\phi(\xi^*)$ takes on substantial positive values. One reason for this lies in the long panel dimension which varies between 28 and 56. Relative to typical binary choice models that contain a linear index as the argument of the function mapping the latent variables to the unit interval, it is strongly exacerbated by the highly nonlinear nature of $\Delta(\mathrm{CE}_{ij})/\tau_i$.

For this type of problem, global approximation techniques are almost certain to provide approximations and indeed this proved to be the case in our particular application. Heuristically, very few points will lie in the region with positive values for the integral. Such issues frequently occur in Bayesian Statistics and two types of closely related methods have been discussed. Simulation of l_i is viable by means of importance sampling Geweke (1989). The method can be sketched as follows. Instead of sampling directly from the population distribution, the mode μ_i of the integrand in (5.9) is calculated in a first step. The importance sampling density is then constructed starting from a multivariate normal density with mean μ_i and its variance determined by the negative inverse Hessian of the integrand at μ_i. Iteratively, this approximation is adapted (if warranted) in each direction along every axis of the multivariate normal by means of multivariate normals with adjusted variance parameters or student-t distributions with appropriate degrees of freedom to mimic the tail behaviour of the integrand (so-called split-t transformations). Integration is then performed by simulation methods over the importance sampling density.

Closely related to importance sampling is subregion-adaptive integration with deterministic techniques (Genz and Kass 1997). It also performs integration over the importance sampling density, but instead of simulation method it uses quadrature techniques in the last step. The approach has proven to be more powerful for the moderate dimensionality of our problem (Genz and Kass 1997). It is implemented as a Fortran routine in the BAYESPACK software Genz and Kass (1998). There is one important caveat for our application. Being developed for Bayesian analyses, the numerical computation of derivatives is not an issue. Essentially, the approximation needs to be extremely precise as to not confound changes in l_i that are due to changes in the importance sampling density with those that are due to the actual perturbation of the parameters.[5] The solution to this is to use the same evaluation points for calculating the gradient that are used to calculate the original function value, see Rabe-Hesketh, Skrondal, and Pickles (2005) for a discussion of this issue. Modification of the BAYESPACK software for this purpose emerged as a very complex task that we did not yet fully solve. Hence we only report results for the heterogeneous model based on (5.1) where the integration achieves sufficient precision. Later versions of the paper will incorporate higher-dimensional heterogeneity.

Finally, note that the split-t transformations account for asymmetries in the tails of the integrand. These proved to be very important for some individuals in our case. For example, if subjects exhibited risk aversion on some screens and risk lovingness on others, this would result in bimodal integrands in (5.9). It is the reason why the routines described by Liu and Pierce (1994), Hartzel, Agresti, and Caffo (2001), and Rabe-Hesketh, Skrondal, and Pickles (2005) which are based on symmetric normal approximations of the importance sampling density were unreliable in our application.

5.5 Results

This section is devoted to the presentation of the results, alluding to earlier findings from the literature that we reviewed briefly in the Introduction. As noted before, we are limited with respect to the dimensionality of heterogeneity that we can accommodate by the currently employed integration techniques. We begin with the discussion of a model that does not incorporate any heterogeneity. On the other hand, this allows us to consider the richest model specification from Section 5.2.2.

[5] This very same problem arises in an importance sampling context.

5.5.1 Evidence from a Model not Allowing for Individual Heterogeneity

The first column of Table 5.2 contains estimates from (5.9) with $\Sigma = 0$ based on the simple utility specification (5.1). The first four rows show that agents are clearly risk averse. Order effects are statistically significant, but small in magnitude. As in Holt and Laury (2002), we find strong differences of the risk aversion estimates if we vary the dimension of the payoffs. Interpreting the between-subject finding as a within-subject one, this is evidence for decreasing absolute risk aversion. Interestingly, it does not seem to matter whether we provide real incentives or not.[6]

Turning to the following two columns that additionally augment the model by a loss aversion parameter, we find clear evidence for subjects being more sensitive to losses than to gains. Baseline estimates for λ in the low incentive treatment and order 0 are significantly larger than one and effects are even stronger in the other treatments. In particular, loss aversion appears to become more pronounced as stakes are increased. It is larger for in the group with hypothetical payoffs than in the one with tripled incentives. This is likely to be an effect of the participation fee which only the subjects in the real incentive treatments received – and some subjects may base their behaviour on the net payoffs from the experiment while others do not, see Harrison and Rutström (2006) for evidence on this point. Risk aversion estimates drop by roughly ten percent as compared to the estimates with $\lambda = 1$. This is a reflection of the typical finding that some part of ordinarily measured risk aversion is better attributed to loss aversion. Comparing the estimates based on utility function (5.2) with those from the classical prospect theory specification ($\gamma^- = -\gamma^+$), the only differences we find are with respect to the loss aversion parameters. As an inevitable result of the functional form restrictions, they increase in magnitude as one moves to switching risk preferences. Since the models are not nested, it is difficult to discriminate between them. Far from being a valid test, the lower log likelihood for our preferred specification may be interpreted as evidence in its favour. A final note concerns the magnitude of the estimated loss aversion parameter which is considerably larger (for many subjects) than most previous estimates in the literature and varies quite strongly across treatments. A similar finding emerges in Harrison and Rutström (2006). This may suggest a peculiarity of multiple price list designs that deserves further investigation.

The last three columns contain estimates of the full model in (5.3) and (5.6). In the first of these, λ is restricted to one. Comparing the risk aversion estimates with those from column one does not reveal any differences between them. The estimates of the uncertainty resolution preference parameter are mixed. On the one hand, incentives do not seem to have an effect. On the other hand, subjects

[6] Remember that the representation of payoffs is identical in the hypothetical and the high incentive treatments.

Tab. 5.2: Results of Simple Logit Models Without Heterogeneity

Parameter	**Exp U**	**Exp U Loss Av**	**Prospect Theory**	**K-P Exp U**	**K-P Loss Av**	**K-P Loss Av**
γ_{baseline}	0.1131^{**}	0.0996^{**}	0.0973^{**}	0.1118^{**}	0.1001^{**}	0.1031^{**}
	(0.0005)	(0.0012)	(0.0012)	(0.0014)	(0.0013)	(0.0015)
γ_{order1}	0.0104^{**}	-0.003^{**}	-0.002^{**}	-0.001	-0.006^{**}	-0.005^{**}
	(0.0003)	(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0005)
$\gamma_{\text{hypothetical}}$	-0.068^{**}	-0.063^{**}	-0.061^{**}	-0.066^{**}	-0.061^{**}	-0.063^{**}
	(0.0006)	(0.0012)	(0.0011)	(0.0014)	(0.0014)	(0.0014)
$\gamma_{\text{high\_incent}}$	-0.069^{**}	-0.061^{**}	-0.060^{**}	-0.066^{**}	-0.060^{**}	-0.062^{**}
	(0.0006)	(0.0012)	(0.0011)	(0.0014)	(0.0014)	(0.0014)
$\lambda_{\text{baseline}}$		1.1847^{**}	1.9045^{**}		1.3803^{**}	1.2818^{**}
		(0.0517)	(0.0718)		(0.0706)	(0.0766)
λ_{order1}		1.9213^{**}	2.4942^{**}		1.4066^{**}	1.1099^{**}
		(0.0750)	(0.0917)		(0.0943)	(0.0952)
$\lambda_{\text{hypothetical}}$		1.2102^{**}	1.5081^{**}		1.0745^{**}	0.9069^{**}
		(0.0956)	(0.1167)		(0.1099)	(0.1064)
$\lambda_{\text{high\_incent}}$		0.5406^{**}	0.9635^{**}		0.5258^{**}	0.5115^{**}
		(0.0714)	(0.1010)		(0.0931)	(0.0995)
ρ_{baseline}				0.8865^{**}	0.8266^{**}	0.6884^{**}
				(0.0306)	(0.0366)	(0.0361)
ρ_{order1}				0.5653^{**}	0.3876^{**}	0.3604^{**}
				(0.0310)	(0.0389)	(0.0344)
$\rho_{\text{hypothetical}}$				0.0595	-0.022	0.0014
				(0.0370)	(0.0412)	(0.0342)
$\rho_{\text{high\_incent}}$				0.0305	-0.006	-0.010
				(0.0372)	(0.0408)	(0.0349)
$\rho_{\text{neg\_outcome}}$						0.2016^{**}
						(0.0320)
τ_{baseline}	5.2828^{**}	5.3208^{**}	5.3199^{**}	5.2729^{**}	5.3418^{**}	5.3312^{**}
	(0.0333)	(0.0347)	(0.0346)	(0.0331)	(0.0349)	(0.0356)
τ_{order1}	-0.443^{**}	-0.458^{**}	-0.531^{**}	-0.491^{**}	-0.525^{**}	-0.553^{**}
	(0.0341)	(0.0341)	(0.0337)	(0.0336)	(0.0341)	(0.0341)
$\tau_{\text{hypothetical}}$	11.480^{**}	11.537^{**}	11.415^{**}	11.411^{**}	11.490^{**}	11.418^{**}
	(0.1155)	(0.1125)	(0.1112)	(0.1140)	(0.1129)	(0.1126)
$\tau_{\text{high\_incent}}$	11.625^{**}	11.614^{**}	11.624^{**}	11.568^{**}	11.613^{**}	11.616^{**}
	(0.1077)	(0.1076)	(0.1073)	(0.1066)	(0.1074)	(0.1083)
Number of Individuals	1,789	1,789	1,789	1,789	1,789	1,789
Log Likelihood	-50483.2	-49806.2	-49811.0	-50262.3	-49767.4	-49750.0

Note: "Exp U" refers to a specification of ΔCE based on (5.1), "Exp U, Loss Av" takes (5.2) as its basis. "Prospect Theory" is similar but risk preferences switch between the positive and negative domains. Finally, "K-P, Exp U" uses (5.3) and (5.6) to calculate ΔCE while restricting λ to one. In the "K-P, Loss Av" specification, this last restriction is dropped. Coefficients shown are baseline estimates and deviations from these.

in order treatment 0 exhibit preferences for late resolution of uncertainty while we estimate $h(\cdot)$ to be concave in the other group. The general picture does not change much when λ is estimated along with the other parameters. In the last column we test the idea from Section 5.3.2, namely differential uncertainty resolution preferences depending on whether negative prospects are involved in the gamble. This is implemented by including a dummy for a negative prospect in option 'B' among the determinants of ρ. Subjects in group two are now estimated to be indifferent between early and late resolution for positive prospects only. If there is a negative outcome involved, the attractiveness of late resolution is significantly reduced. However, the reduction is not sufficiently large as to make subjects in order group 0 prefer early resolution over it. Hence we do not reach a definite conclusion in this respect.

Finally, note that the error estimates do not change much across the six specifications. Their magnitude can be assessed by noting that the probability for choosing the option that is contrary to what is implied by ΔCE is given by $1-\Lambda(\Delta\text{CE}/\tau)$. For example, the probability to choose a dominated option with a 30-Euros difference between the two payouts is roughly 14.5 % for subjects in the high incentive or hypothetical order 0 treatment. However, we are very reluctant to call the estimates for τ error estimates in the homogeneous setting because anybody choosing according to preferences that are different from mean preferences would be contributing to them.

5.5.2 Evidence from a Model of Risk Aversion with Individual Heterogeneity

One big advantage of our data structure is that we have sufficiently rich background information combined with a large sample size as to estimate conditional distributions of preference parameters in the population. For instance, it is very interesting to which extent preference heterogeneity can be attributed to observable characteristics and to which extent it remains idiosyncratic. In the future, we will also be able to comment on the joint distribution of the preference and error parameters, but this is not possible at present due to the difficulties with the integration procedures explained above. All results reported in this section and in Table 5.3 are based on the utility specification (5.1) only.

From a comparison of the first columns in Tables 5.2 and 5.3 we note that point estimates for γ do not change substantively when moving from the homogenous to the heterogenous specification. Heterogeneity emerges as very important with the standard deviation of the risk aversion coefficient being almost as large as the coefficient itself in the low incentive treatment. This means for instance that for the low incentive treatment, 50 % of the respondents have a risk aversion coefficient in the range of $[0.047, 0.169]$. As another example, 11.7 % of the sample would be expected to show risk-loving behaviour. Trivially, there is a significant drop in the error estimates that reduces the probability to choose the dominated option to 6.6 % (see column 4 in Table 5.3). Put differently, there is a 6.6 percent chance that somebody in the high incentive - order 0 treatment will commit an error that may cost him the utility equivalent of 30 Euros. This number increases to 30 % for a utility equivalent of 10 Euros.

We now turn to the influence of demographic variables in columns two and three of Table 5.3. The two specifications are governed by the availability of covariates – we do not have all the information on all households.[7] In general, coefficients do not change much between the two specifications and we will describe both columns together, highlighting only differences.

As in large parts of the previous literature, we find women to be significantly more risk averse than men (Croson and Gneezy (2004) provide an overview). Note that this finding is not undisputed and may be most salient in abstract choice tasks, see Schubert, Brown, Gysler, and Brachinger (1999) for a critical view. Risk aversion is monotonically increasing in age, although the difference is only significant from age 55 on (ages 18-34 are the left-out category). The more educated exhibit significantly less risk aversion (left-out category is primary/lower secondary education). The effect is insignificant only for people with intermediate secondary education. The income effect goes in the expected direction with persons from higher income households displaying significantly lower risk aversion. Once income is controlled for, the age effects become slightly more pronounced which may be expected from the generally positive correlation between the two variables. Finally, being the financial manager of the household does not appear to have a significant influence on risk-taking behaviour.

Interestingly there is no substantial drop in the estimated dispersion of the risk aversion coefficient despite the large number of significant influences on risk aversion (the point estimate drops by only 2 %, but the change is not significant). Hence almost all of the observed variation in risk averse behaviour appears to be attributable to idiosyncratic characteristics and not to observed demographic variables. Although our list of covariates is certainly not exhaustive and especially lacks some typically mentioned household composition characteristics at this point, it casts some doubt on study in the literature that attempt to control for individual or household risk aversion by including its demographic correlates in the estimation equations.

Turning to the Fechner error estimates in columns five and six of Table 5.3, we also find that there is substantial variation with respect to demographic characteristics. Men have a lower error rate than women and there is a monotonically rising

[7] Education variables are not available for two persons. The sample size is reduced by another 75 if we opt for the full specification that includes household income and information on who handles the household finances in column three.

Tab. 5.3: Results of Random Coefficients Logit Models Allowing for Individual Heterogeneity in Risk Aversion

Variable	Spec (1) γ	Spec (2) γ	Spec (3) γ	Spec (1) τ	Spec (2) τ	Spec (3) τ
baseline	0.1079^{**} (0.0037)	0.1012^{**} (0.0072)	0.1077^{**} (0.0080)	3.3430^{**} (0.0166)	3.2722^{**} (0.0313)	3.5181^{**} (0.0391)
order 1	0.0204^{**} (0.0046)	0.0201^{**} (0.0046)	0.0210^{**} (0.0047)	-0.057^{**} (0.0177)	0.0073 (0.0182)	0.0032 (0.0202)
hypothetical	-0.048^{**} (0.0061)	-0.047^{**} (0.0061)	-0.047^{**} (0.0061)	8.7854^{**} (0.0667)	8.8381^{**} (0.0981)	9.5789^{**} (0.1186)
high incentive	-0.052^{**} (0.0059)	-0.051^{**} (0.0059)	-0.051^{**} (0.0061)	7.9950^{**} (0.0593)	7.8647^{**} (0.0875)	8.7070^{**} (0.1104)
female		0.0334^{**} (0.0047)	0.0340^{**} (0.0049)		0.4528^{**} (0.0202)	0.5093^{**} (0.0226)
age 35-44		0.0053 (0.0073)	0.0058 (0.0076)		0.3115^{**} (0.0285)	0.3664^{**} (0.0318)
age 45-54		0.0079 (0.0066)	0.0086 (0.0070)		0.3998^{**} (0.0281)	0.4792^{**} (0.0318)
age 55-64		0.0153^{*} (0.0074)	0.0180^{*} (0.0077)		0.6597^{**} (0.0320)	0.8414^{**} (0.0375)
age 65+		0.0174^{*} (0.0078)	0.0191^{*} (0.0081)		1.2328^{**} (0.0389)	1.3895^{**} (0.0451)
higher sec. educ.		-0.028^{**} (0.0072)	-0.029^{**} (0.0075)		-1.131^{**} (0.0253)	-1.175^{**} (0.0287)
vocational train.		−0.005 (0.0064)	−0.005 (0.0066)		-0.348^{**} (0.0262)	-0.294^{**} (0.0298)
higher voc. train.		-0.024^{**} (0.0063)	-0.020^{**} (0.0067)		-0.850^{**} (0.0230)	-0.798^{**} (0.0258)
university educ		-0.041^{**} (0.0099)	-0.038^{**} (0.0104)		-1.024^{**} (0.0261)	-0.982^{**} (0.0294)
income22-40k			-0.016^{**} (0.0053)			-0.431^{**} (0.0232)
income40k+			-0.023^{**} (0.0077)			-0.648^{**} (0.0286)
finadmin			0.0047 (0.0049)			-0.126^{**} (0.0217)
σ_γ	0.0906^{**} (0.0014)	0.0887^{**} (0.0014)	0.0885^{**} (0.0014)			
Number of Individuals	1,789	1,787	1,712			
Log Likelihood	-43626.7	-43097.8	-41199.2			

Note: All specifications are based on (5.1) to calculate ΔCE. Coefficients shown are baseline estimates and deviations from these.

age effect. It is especially pronounced for the oldest group. While part of this may be simply related to difficulties with handling the computer resources, it is likely to reflect a decline of financial numeracy levels at these ages that has been shown elsewhere (Banks and Oldfield 2006). Note that it remains unclear to what extent this it is cohort or age effects that stand behind this finding. Not surprisingly, choices are significantly more consistent for the better educated. With the exception of people in higher secondary education, the effect gets stronger with higher education categories. Finally, people in higher income groups also commit less errors, just as those who handle a household's financial affairs. The latter could be interpreted as evidence in favour of learning and (or) efficient task allocation within households, although the effect is small. To get an idea of the magnitude of the effect, the most consistent group in the high incentive treatment (young men with higher secondary or university education who handle the financial affairs of households with an annual income above 40,000 Euros) has a predicted error rate of 5 % for errors that cost the utility equivalent of 30 Euros. On the other hand, this rate is about twice as high among the most error-prone group (elderly women with primary or lower secondary education living in households with less than 22,000 Euros annual income and do not handle financial matters).

5.6 Conclusions

We characterised decision behaviour in a risky choice experiment by means of structural econometric estimations. While we were not yet able to exploit the full generality admitted by our model, some important findings emerged. First, our results on the demographic correlates of risk attitudes as well as those on the loss aversion parameter are consistent with the previous literature. Even after conditioning on an extensive set of demographic characteristics, the remaining idiosyncratic variation in choice behaviour is very large. The nature of our choice task allowed us to estimate the monetary costs behaviour that is inconsistent with the estimated individual-level utility function. We found such errors to be much more important in our data than what is typically found in laboratory experiments. Hence it seems important to allow people to make such mistakes – in other words, conclusions based on a single measurement may be seriously misleading by hiding a large amount of imprecision that may be inherent to somebody's behaviour.

5.7 Certainty Equivalents

The certainty equivalent (CE) for a lottery π in terms of period one utility, is given by:

$$\begin{aligned} V(\mathrm{CE}) &= h(v(\mathrm{CE})) \\ &= V(\pi). \end{aligned}$$

Solving this for the certainty equivalent leads to:

$$\mathrm{CE} = v^{-1}\left(h^{-1}\left[V(\pi), \rho\right], \gamma, \lambda, \rho\right).$$

In our particular framework, we have the following:

$$\begin{aligned} h^{-1}(y, \cdot) &= -S\,(-S\;y)^{\rho^S} \\ v^{-1}(y, \cdot) &= \begin{cases} -\frac{\ln(-\gamma y)}{\gamma \rho^S} & \text{for } y \geq -1/\gamma \\ -\frac{\ln(\lambda - 1 - \gamma y) - \ln(\lambda)}{\gamma \rho^S} & \text{for } y < -1/\gamma \end{cases} \end{aligned}$$

The certainty equivalent of a gamble π is hence given by:

$$\mathrm{CE}(\pi) = \begin{cases} -\frac{\ln\left(\gamma\, S\; (-S\;\, V(\pi))^{\rho^S}\right)}{\gamma \rho^S} & \text{for } V(\pi) \geq -1/\gamma \\ -\frac{\ln\left(\lambda - 1 + \gamma\, S\; (-S\;\, V(\pi))^{\rho^S}\right) - \ln(\lambda)}{\gamma \rho^S} & \text{for } V(\pi) < -1/\gamma \end{cases}$$

BIBLIOGRAPHY

AHLBRECHT, M., AND M. WEBER (1996): "The Resolution of Uncertainty: An Experimental Study," *Journal of Institutional and Theoretical Economics*, 152, 593–607.

ALLAIS, M. (1953): "Le comportement de l'homme rationelle devant le risque: critique des postulats et axiomes de l'école américaine," *Econometrica*, 21, 593–607.

ANDERSEN, S., G. W. HARRISON, M. I. LAU, AND E. E. RUTSTRÖM (2006): "Elicitation Using Multiple Price List Formats," *Experimental Economics*, 9(4), 383–405.

BALLINGER, T., AND N. WILCOX (1997): "Decisions, Error and Heterogeneity," *The Economic Journal*, 107(443), 1090–1105.

BALTUSSEN, G., T. POST, AND P. VAN VLIET (2006): "Violations of Cumulative Prospect Theory in Mixed Gambles with Moderate Probabilities," *Management Science*, 52, 1288–1290.

BANKS, J., AND Z. OLDFIELD (2006): "Understanding Pensions: Cognitive Function, Numerical Ability and Retirement Saving," The Institute for Fiscal Studies, London, Working Paper 06/05.

BENJAMIN, D., S. A. BROWN, AND J. SHAPIRO (2006): "Who is "Behavioral"? Cognitive Ability and Anomalous Preferences," Mimeo, Harvard University.

BINSWANGER, H. P. (1980): "Attitudes Towards Risk: An Experimental Measurement in Rural India," *American Journal of Agricultural Economics*, 62, 395–407.

CAMERER, C. F. (1989): "An Experimental Test of Several Generalized Utility Theories," *Journal of Risk and Uncertainty*, 2, 61–104.

——— (2000): "Prospect Theory in the Wild: Evidence from the Field," in *Choices, Values, and Frames*, ed. by D. Kahneman, and A. Tversky. Cambridge University Press, Cambridge.

CAPLIN, A., AND J. LEAHY (2001): "Psychological expected utility theory and anticipatory feelings," *Quarterly Journal of Economics*, 116, 55–79.

CHEW, S. H., AND J. L. HO (1994): "Hope: An Empirical Study of Attitude Toward the Timing of Uncertainty Resolution," *Journal of Risk and Uncertainty*, 8, 267–288.

COOK, J., AND L. BARNES (1964): "Choice of Delay of Inevitable Shock," *Journal of Abnormal and Social Psychology*, 68, 669–72.

CROSON, R., AND U. GNEEZY (2004): "Gender Differences in Preferences," Mimeo, The Wharton School of the Universtiy of Pennsylvania.

DOHMEN, T., A. FALK, D. HUFFMAN, U. SUNDE, J. SCHUPP, AND G. G. WAGNER (2005): "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey," IZA Discussion Paper 1730, Institute for the Study of Labor (IZA).

EPSTEIN, L. G., AND S. E. ZIN (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969.

GENZ, A., AND R. E. KASS (1997): "Subregion-Adaptive Integration of Functions Having a Dominant Peak," *Journal of Computational and Graphical Statistics*, 6(1), 92–111.

——— (1998): "BAYESPACK: A Collection of Numerical Integration Software for Bayesian Analysis," Mimeo, Washington State University.

GEWEKE, J. (1989): "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57(6), 1317–1339.

HARRISON, G., AND E. RUTSTRÖM (2006): "Expected Utility Theory and Prospect Theory: One Wedding and A Decent Funeral," University of Central Florida, Discussion Paper 05-18.

HARRISON, G. W., M. I. LAU, AND E. E. RUTSTRÖM (2007): "Estimating Risk Attitudes in Denmark: A Field Experiment," *Scandinavian Journal of Economics*, forthcoming.

HARTZEL, J., A. AGRESTI, AND B. CAFFO (2001): "Multinomial Logit Random Effects Models," *Statistical Modelling*, 1, 81–102.

HEY, J. (2005): "Why We Should Not Be Silent About Noise," *Experimental Economics*, 8(4), 325–345.

Hey, J., and C. Orme (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica*, 62(6), 1291–1326.

Holt, C. A., and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Johnson, E., S. Gächter, and A. Herrmann (2006): "Exploring the Nature of Loss Aversion," CeDEx Discussion Paper 2006-02.

Kahneman, D. V., and A. V. Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 47, 263–291.

Köbberling, V., and P. P. Wakker (2005): "An Index of Loss Aversion," *Journal of Economic Theory*, 122, 119–131.

Kreps, D. M., and E. L. Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory," *Econometrica*, 46, 185–200.

Liu, Q., and D. A. Pierce (1994): "A Note on Gauss-Hermite Quadrature," *Biometrika*, 81, 624–629.

Loewenstein, G. (1987): "Anticipation and the Valuation of Delayed Consumption," *The Economic Journal*, 97(387), 666–684.

Loomes, G. (2005): "Modelling the Stochastic Component of Behaviour in Experiments: Some Issues for the Interpretation of Data," *Experimental Economics*, 8(4), 301–323.

Loomes, G., P. G. Moffatt, and R. Sugden (2002): "A Microeconometric Test of Alternative Stochastic Theories of Risky Choice," *Journal of Risk and Uncertainty*, 24(2), 103–130.

Machina, M. J. (1987): "Choice under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives*, 1(1), 121–154.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles (2005): "Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects," *Journal of Econometrics*, 128(2), 301–323.

Rabin, M. (2000): "Risk Aversion and Expected Utility Theory: A Calibration Theorem," *Econometrica*, 68(5), 1281–1292.

Schubert, R., M. Brown, M. Gysler, and H. Brachinger (1999): "Financial Decision-Making: Are Women Really More Risk-Averse?," *The American Economic Review, Papers and Proceedings*, 89(2), 381–385.

Starmer, C. (2000): "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature*, 38(2), 332–382.

Tversky, A. V., and D. V. Kahneman (1992): "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, 5(4), 297–323.

von Gaudecker, H.-M., A. van Soest, and E. Wengström (2007): "Experimental Elicitation of Risk Preferences: Some Further Steps Towards Representativeness," Chapter 4 of this Dissertation.

Wu, G. (1999): "Anxiety and Decision Making with Delayed Resolution of Uncertainty," *Theory and Decision*, 46, 159–199.

6. RISK ATTITUDE, IMPATIENCE, AND ASSET CHOICE

JOINT WITH ANGELIKA EYMANN, AXEL BÖRSCH-SUPAN, AND ROB EUWALS

6.1 Introduction

Standard models of portfolio choice assume that the assets' risk and return, risk aversion, and the timing of the households' consumption expenditure guide the households' choices among different assets. In their simplest, representative-agent formulation, these models are typically soundly rejected by the data. An enormous amount of literature has suggested various explanations for why the model appears to be at odds with the data. These include taste heterogeneity, market frictions, and failure of the life-cycle model, among other things. Campbell (2006) provides a broad take and many references on the subject. In this paper, we are predominantly concerned with the role that varying attitudes towards risk and time have to play for household portfolio allocations.

Using financial market decisions of households alone, the researcher can at best identify a distribution of tastes from revealed preference type arguments. Usually this involves rather strong auxiliary assumptions on expectations and market completeness. Even if the validity of such assumptions is accepted, a test of the underlying model will have to be confined to violations of its most basic principles – estimation of the preference parameters will generally absorb all available degrees of freedom. These difficulties have inspired attempts to measure preferences directly in more controlled settings, most notably through decisions in hypothetical tasks or experiments (e.g. Barsky, Juster, Kimball, and Shapiro (1997) and Holt and Laury (2002), respectively), and behaviour in uncertain or dynamic environments with a more simple structure than the environment of interest to the researcher (Brown, Farrell, Harris, and Sessions 2006). The preference parameters thus obtained have been shown by these and other authors to vary systematically with demographic characteristics.

Yet this is only one side of the story. Just as much as preferences impact upon economic decisions, they may be influenced by them (Becker and Mulligan 1997). This interdependence is probably most salient with respect to wealth and risk preferences. On the one hand, a long line of research has attempted to describe the dependence of risk tolerance on wealth – the literature is rich in lively debates on the use of CARA, HARA, CRRA, or other functionals for this purpose (see for example Halek and Eisenhauer (2001) and the references cited therein). On the other hand, theory predicts individuals who are more tolerant towards risk to hold a larger share of their portfolio in risky assets. The resulting risk premium will lead to higher wealth levels on average, see Carroll (2002) for a related theoretical argument and empirical evidence. However, there may also be an effect in the opposite direction. Precautionary saving leads the more prudent to have higher savings rates, leaving them with higher wealth in a dynamic setting – see Haliassos and Michaelides (2003) and Gomes and Michaelides (2005) for models that generate such a pattern.

In this paper, we attempt to weave these threads into a unified model. We use a host of indicators on wealth, risk attitude, and impatience, combining them by means of a factor-analytic measurement model inspired by psychometric methods. We treat these indicators as error-ridden measurements of true wealth and attitudes and model the underlying constructs as latent variables. They may depend on each other both directly and through unobserved factors. Furthermore, they may in turn be determined by a set of socio-demographic characteristics. We then analyse the impact of wealth, risk attitude, impatience, and observable variables on households' asset choices.

Before moving on to the analysis, let us clarify our terminologies. We denote the latent individual traits that underly the measurement model by "risk attitude" and "impatience". Our model is structural in the sense that we explicitly acknowledge their joint determination with wealth and only then analyse the impact they have on the choice of portfolio components. It is not our intention to investigate the functional form of preferences with respect to wealth or consumption or even to determine *the* correct form of the individual utility function. Instead, we use endogenous latent traits as a flexible tool to approximate the unknown functional form of the individuals' preferences and thus to better explain households' asset choice. We use the term "portfolio choice" to describe the continuous choice of the portfolio share of an asset, and "asset choice" to describe the discrete decision whether to hold an asset or not.

The data for our analysis is presented in Section 6.2 along with the bivariate relationships among the key variables. This sets the scene for a verbal description of our modelling approach in Section 6.3.1. We then turn to a formal analysis of identification in and estimation of the structural equation model. Section 6.4 contains the empirical results. We close with some concluding remarks and directions for further developments of this paper.

6.2 Data and Descriptive Evidence

Datasets which contain detailed information on respondents' portfolio compositions, a variety of measures relating to behaviour under uncertainty and in a temporal setting, and sufficient background characteristics are rare to find. We make use of the German SAVE study, which provided a unique opportunity to incorporate several items in the questionnaire with our modelling setup in mind. The SAVE study is a household panel that is designed to gain a broad understanding of savings and investment behaviour of households. It is described in detail by Schunk (2007a). In total, 2305 households participated in the 2005 wave. Item nonresponse is dealt with by a multiple imputation algorithm developed by Schunk (2007b).

Here we begin with describing the main outcomes of interest, the asset components of households' portfolios. We then turn to the three key explanatory variables, namely wealth, risk attitude, and impatience. Motivating the analysis to come, we include figures depicting the connections between them and the portfolio components. Finally, we list the observable determinants of asset choice, preferences, and wealth.

6.2.1 Definition of Assets

The SAVE questionnaire addresses the member of the household who is informed best about the household finances and asks for the household's possession of any of 18 assets and current asset holdings. We aggregate these into five categories, distinguishing them by their riskiness, liquidity, and insurance aspects. In this version of the paper, we limit our ambitions in excluding housing and mortgages from the asset classes and from wealth. Essentially, the probability to own a house is almost perfectly predicted by total wealth holdings – of 616 households who report to possess assets worth more than 200,000 Euros, only 10 do not own any real estate property. Only 53 of 518 households with reported zero or negative wealth occupy their own house. Not surprisingly, this pattern provoked the probit function linking wealth and the probability to own housing property to approximate a step-wise function, dominating all other estimates. Nevertheless, the analysis in its present form yields some insights into the relationship among wealth, preferences, and asset choice behaviour.

Tab. 6.1: Definitions and Descriptive Statistics of the Asset Variables

Name	**Definition**	**Mean**
savacc	Savings accounts, building society savings contract	0.67
retinsur	Pension / endowment insurance, employer- or government-sponsored pension plan	0.41
bonds	Government or company bonds, bond funds	0.08
stocks	Stocks, mutual funds, derivatives, business assets	0.24
loans	Consumer credit loans, family loans, other loans.	0.21

Source: SAVE 2005, own calculations

Table 6.1 gives the definition of each of the five assets and overall ownership rates. Like in other countries, the percentage of the population holding risky assets is generally low. The most prevalent savings vehicle are savings accounts and buildings society savings agreements with ownership rates of about two thirds.

Possession of risky assets[1] is confined to a little more than a fourth of the households (overlap between bond and stock holdings is quite high). We now turn to describing our measures of wealth and preferences and how these variables are related to asset holdings.

6.2.2 Definition of Wealth

As explained above, the SAVE 2005 survey contains detailed questions on 18 different asset types. Among other things, the value of the respective portfolio component as of the end of 2004 is included in the questionnaire.[2] We sum up all these values (excluding housing and related loans / mortgages) to obtain our measurement of total wealth. We trim it at |2.5| million Euros because of some outliers with up to 26 million Euros net worth which may impact all too strongly upon our analysis. This way we delete 5 households from our sample, leaving us with $N = 2,300$ for the rest of the analysis.

Figure 6.1 contains a nearest-neighbour-smoothed plot of asset holdings by wealth. Restricting attention to the area with a relevant number of observations we display only values between -25,000 Euros and one million Euros.[3] Starting with the savings accounts, there is a very pronounced rise in holdings from about 50 % at the origin to roughly 80 % at a wealth level of 200,000 Euros. The profile becomes flat at higher levels of wealth. A similar but somewhat less distinct pattern can be found for retirement insurance contracts. The rise in stocks and bonds holdings, respectively, is more evenly spread across the range under consideration. These slopes level off eventually, too, although part of this is related to rather low numbers of observations in the higher wealth regions and the nearest-neighbour nature of the estimates. Finally and not surprisingly, loans are most prevalent at very low levels of wealth. They then reach a roughly stable level of slightly below 20 % for the more affluent households.

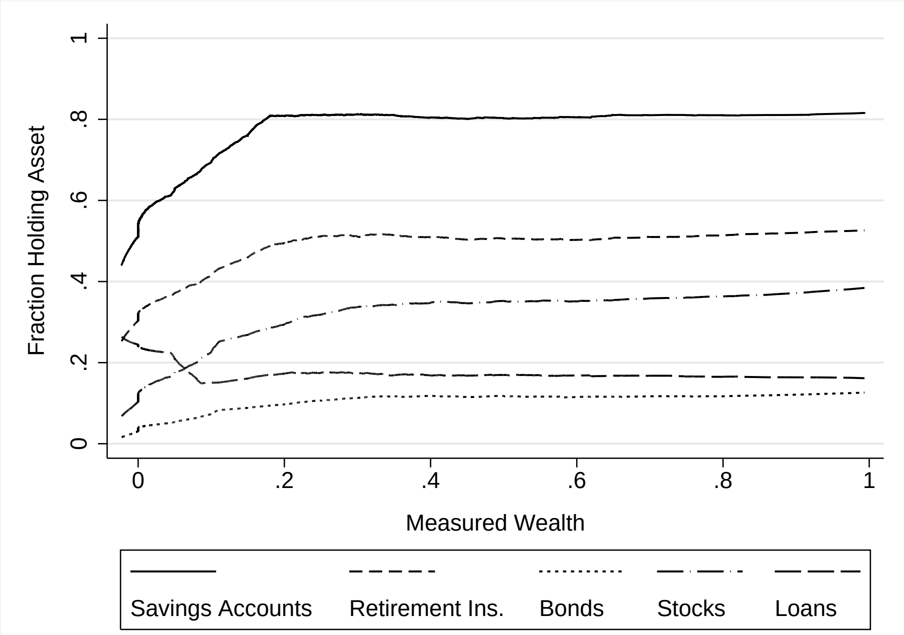
Wealth is a rather complex construct and it is notoriously difficult to measure (Juster, Smith, and Stafford 1999). Recall error, poor understanding of financial concepts, little motivation to allocate cognitive effort, among other things, may lead to misreports and item nonresponse. Valuations of illiquid assets such as real

[1] I.e. possession of risky assets which is not mediated through retirement insurance. Typical German endowment policies participate in the performance of insurers' portfolios. However, risk-return structures are quite different from direct or funds-based holdings of stocks and bonds. There are substantial minimum interest rate guarantees and any yield beyond these is smoothed over several years, leading to a relatively stable performance.

[2] Note that the SAVE survey is fielded during late spring and early summer when most people are filing their tax returns for the previous year. Accordingly, this number should be the one which is most readily available.

[3] Estimates are based on a the full 2,300 observations. The only reason to shrink the support of the graph did not want to devote two thirds of graph space to two percent of the observations.

Fig. 6.1: Mean Asset Holdings by Wealth



Source: SAVE 2005, own calculations. Wealth is measured in millions of Euros as the sum of all asset holdings except for housing and related loans and mortgages.

estate or business wealth are best characterised by an impossibility to price them precisely. We take this explicitly into account by modelling wealth as a latent variable of which the just-described measured wealth is just one indicator.

In addition to this, we make use of two further variables that are indicative of financial wealth. The first is a binary marker which is one if a household states to have any income from interest, dividends, or rents. The use of this variable is motivated by the SAVE questionnaire since the sections asking about income and wealth are separated by a substantial amount of items related to old-age provision. The income questions are yes or no questions (later on there is a question on total net income) while the wealth questions ask for the amount held immediately after a respondent indicated possession of the respective asset. Earlier findings suggest that this type of question may lead to some underreporting (Juster and Smith 1997) and we expect to ameliorate this by including the income from wealth indicator. Second, high wealth holdings lead to a considerable increase in the complexity of income tax returns. This and the fact that potential gains from exploiting details of the tax law rise with wealth increase the probability to employ a tax advisor for filing income tax returns. We include a binary variable on this issue as the third indicator of the latent variable total wealth. Descriptive statistics can be found in Table 6.7 in the Appendix.

6.2.3 Measuring Risk and Time Preferences

The canonical economic model has clear concepts of how individuals and households behave under risk and of how they allocate consumption and savings over time (Gollier 2001). It makes sharp predictions on behaviour in such circumstances. However, once this model is taken to the data and individual variation in preference parameters is acknowledged, there is a fundamental identification problem. In general, auxiliary assumptions such as rational expectations or complete markets have to be invoked in order to identify individual-level preferences based on market data alone.

The questionability of such assumptions has led researchers to make attempts at eliciting these parameters directly from behaviour in hypothetical settings and (or) experiments. The approaches differ widely in many respects, ranging from large-scale hypothetical choices (Barsky, Juster, Kimball, and Shapiro (1997), Kapteyn and Teppa (2003), Guiso and Paiella (2004)) to comparatively simple experiments with varying stakes and subject pools (Holt and Laury (2002), Harrison, Lau, and Williams (2002)). Middle grounds or combinations of these include Donkers, Melenberg, and van Soest (2001) or Dohmen et al. (2005). Finally, researches have used behaviour in other settings characterised by uncertainty to make predictions about the economic decision of interest (Brown, Farrell, Harris, and Sessions 2006). All of these approaches have their advantages and disadvantages on which we comment shortly. While hypothetical choices often ask about behaviour in a situation which is precisely the one of interest for the real-world application, the mental effort for respondents needed to put themselves in the imagined place may be substantial. Parameter findings are sometimes difficult to reconcile with real-world counterparts, see for example the concluding remarks in Kapteyn and Teppa (2003). Asking for certainty equivalents and similar rating tasks leads to well-known problems of focal point responses that may reflect other things than true valuations (Green, Jacowitz, Kahneman, and McFadden 1998). On the other hand, while preference elicitation experiments usually suffer less from these issues, they are not easily included in large household surveys (Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2005) for example consider only a small subsample of the German Socio-Economic Panel in the experimental part of their paper) and it is not perfectly clear how behaviour in small-stake gambles carries over to situations with more meaningful economic quantities. Rabin (2000) shows in a theoretical argument that the standard expected utility model cannot be expected to provide a translation mechanism. The experimental design of Holt and Laury (2002) could provide data to make such inferences, however the authors only consider aggregate behaviour. Using behaviour in other risky circumstances may be stretching the extrapolation argument rather far, as results from the experimental literature on the importance of context for a wide range of tasks suggest (Harrison and List 2004).

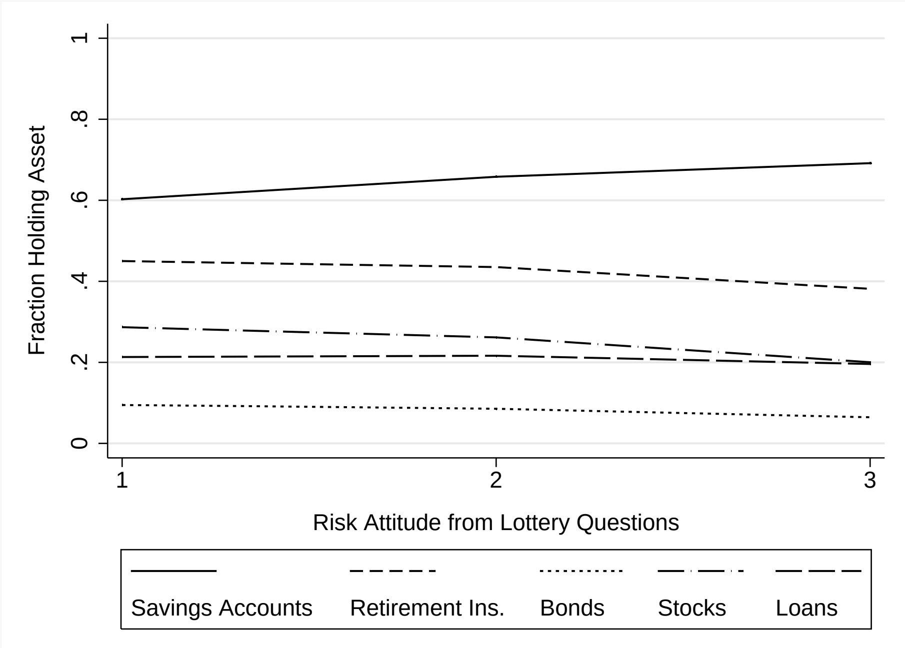
In this paper, we combine the various approaches, similar in spirit to Arrondel and Masson (2005) but quite different in the econometric implementation. We make use of five relatively simple lottery choice and temporal resource allocation tasks that we describe shortly. For the purposes of this section, they are easily aggregated to crude measures of risk attitudes and impatience. We continue the graphical analysis considering the relations to asset choice and wealth. In addition to these items, our model in Section 6.3 incorporates self-reported risk behaviour (smoking and drinking) as well as self-rated personality traits. While we discuss the measurements here, we leave their relation to wealth and asset choice to the formal model.

The SAVE 2005 questionnaire contains two sets of three hypothetical lottery tasks devoted to measuring risk preferences. Each of these six items consisted of a choice between a lottery with equal probabilities and a sure outcome. The probabilities were framed as a coin toss. In the first set of choices, the sure outcome was always 1,000 Euros, the lower lottery payoff yielded nothing. The high lottery payoff went up from 2,000 Euros in the first task to 2,500 Euros in the second and 3,000 Euros in the third task. The second set of choices had zero as its sure outcome and a loss of 100 Euros as the low outcome of the gamble. The high outcome varied from 200 to 300 and 400 Euros, respectively. A third lottery was similar but included a time preference component. The sure option would pay 500 Euros immediately, while the coin toss – leading to payoffs of nothing or 750 Euros / 1,200 Euros / 1,800 Euros – would take place six months later. In this section, we simply aggregate the number of safe choices in these nine imaginary situations. For the subsequent analysis of the full model, we generate a variable with four ordered outcomes for each of the three sets of lotteries. It takes on the value zero (three) if the risky (safe) option is chosen in all tasks. Values one and two are defined accordingly.[4] Note that this coding implies higher risk aversion among those people with higher scores on the choice tasks. Table 6.7 has the descriptive statistics.

Figure 6.2 contains the asset holding trajectories by the number of safe choices in the gambles, aggregated to three categories. Savings accounts – the most conservative asset class – are the only category that rises with risk aversion. The means of all other portfolio components are highest among those who display the

[4] Some 3.7 % of respondents exhibited non-monotonic patterns. We handled these by taking the option where the subject showed the highest risk tolerance as a reference point, ignoring all safe choices at gambles with higher expected values. A robustness check is to follow here, but we do not expect substantial sensitivity to the assumption do to the small number of relevant cases.

Fig. 6.2: Mean Asset Holdings by a Simple Risk Attitude Measure

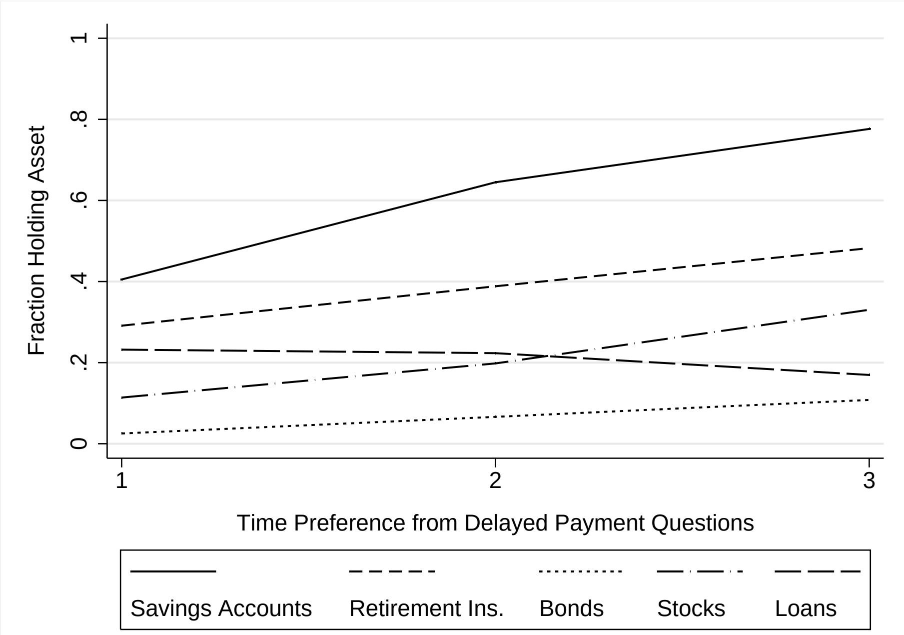


Source: SAVE 2005, own calculations. Risk Attitude is derived from the number of safe choices in the three×three lottery tasks. We group them into three categories: Up to 5 safe choices (category 1 – most risk-tolerant), between 6 and 8 safe choices (category 2), all safe choices (category 3 – most risk averse).

most risk tolerance in the lottery tasks. This is both consistent with our expectations and along the lines of findings by Barsky, Juster, Kimball, and Shapiro (1997) and Guiso and Paiella (2004). However, it is merely a motivation for the analysis to come – the graphics may largely be picking up the influence of wealth or the effects may better attributed to positive socio-economic correlates of both risk attitude and asset choice.

In order to assess individual-level time preference we make use of two key measurements. These are behaviours in hypothetical choice tasks similar in nature to the imaginary gambles that we just described. In particular, subjects were asked whether they would rather obtain 1,000 Euros right away, or 1,130 Euros (1,200, 1380 Euros) in ten months. In order to minimise a possible confounding with risk attitudes related to the fact that the delayed payment may be uncertain, we framed it as an income tax refund. We classify respondents the more patient the more often they choose the delayed payment. The second set of discounting tasks is just the converse – respondents are asked to state what they would do if they had to make a payment to the tax authority. This consisted of either 800 Euros

Fig. 6.3: Mean Asset Holdings by a Simple Measure for Impatience



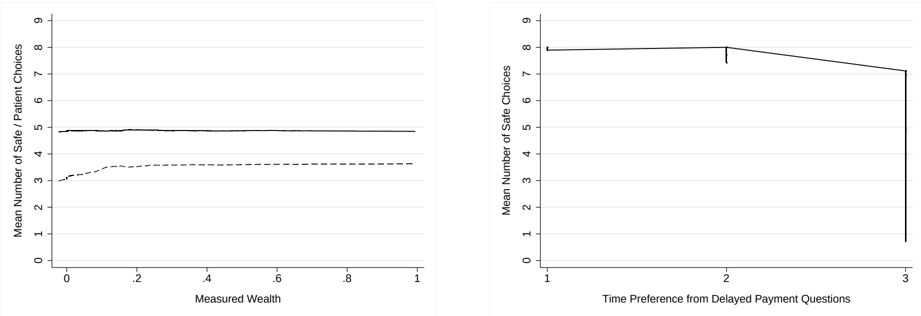
Source: SAVE 2005, own calculations. The impatience measure is derived from counting the number of "patient" choices in the two × three delayed payment tasks and the three tasks from lottery 3. We group them into three categories: Less than 2 "patient" choices (category 1 – most impatient), between 2 and 4 "patient" choices (category 2), more than 4 "patient" choices (category 3 – most patient).

now or 825 / 870 / 990 Euros. Here subjects who choose the immediate payment option are viewed as more patient than those who choose the delayed one. For later chapters, we generate four-dimensional ordered variables similar to those on risk attitudes. Here, high values signify more patient choices. They can also be found in Table 6.7. In this section, we aggregate the number of patient choices, including those in the delayed lottery task.

Figure 6.3 is the analogue to Figure 6.2, containing mean asset holdings by a three-dimensional indicator for patience. Except for loans, all portfolio components decrease with patience. The slope for loans is as one would expect a ceteris paribus effect since housing-related loans and mortgages are excluded from the analysis and consumer credit and similar loans are likely to result from a high preference for the present. Except for the most illiquid and longest-horizon component, retirement insurance contracts, the priors on the other asset classes are not so clear. We comment on these some more in Section 6.3.1.

The three bivariate relationships among the two preference parameters and

Fig. 6.4: Relations Among the Endogenous Variables



Source: SAVE 2005, own calculations. Wealth is measured in millions of Euros as the sum of all asset holdings except for housing and related loans and mortgages. The number of safe choices refers to the number of safe options chosen in the three × three lottery tasks. Accordingly, the number of "patient" choices is the number of "patient" options chosen in the two × three delayed payment tasks and the three tasks from lottery 3.

wealth are displayed in Figure 6.3. Again, there are theoretical reasons for relations to go in any direction and we find relatively flat profiles. The mean number of save choices, depicted by the blue line in the first panel of Figure 6.3, declines only very slightly in wealth. There is a notable rise in patience in the lower wealth categories (up to 200,000 Euros roughly) which levels off afterwards. The second panel containing the relation between risk attitude and impatience is shown more for completeness reasons – the decline that can be seen there is actually an artefact of the inclusion of the delayed lottery choice task in both preference equations. The profile is completely flat otherwise. However, this demonstrates the need for a joint modelling approach that is reinforced through the rest of our indicators.

The last set of indicators of risky behaviours is derived from self-reported risk behaviours and self-rated personality traits. First we use a question on alcohol consumption over the last six months. We collapse the seven categories to three, with persons who drank alcohol never or less than once a month in category zero and those who drank alcohol on at least five days per week in category two. Second, we construct a variable on smoking which is zero if the respondent smokes regularly; one if he quit doing so; and two if he never smoked more often than occasionally. The hypotheses are clear: People who smoke and (or) drink more are expected to be more risk tolerant and to discount the future more heavily than those who abstain from these stimulants.

People were asked to rate their own personality with respect to planning on a ten-point scale between living for the day (coded 0) and making precise plans for the future (coded 10). We merge the eleven outcomes to three categories (0-4=0, 5-7=1, 8-10=2). Similarly, they were asked for a rating of their own decision-making behaviour, zero meaning fast and impulsive decision-making and ten signifying long pondering over choices. Again we obtain a three-outcome variable by the same mapping as for planning. While these are certainly not perfect predictors of risk attitude or impatience, we anticipate planners and slow decision-makers to be less risk tolerant and more oriented towards present. The descriptive statistics relating to these variables complete Table 6.7.

6.2.4 Determinants of Asset Choice, Wealth and Preferences

Our data universe is completed by a set of socio-economic characteristics that may influence asset choice, wealth, and preferences. See Table 6.8 in the Appendix for a full listing of the variables, short verbal descriptions, and summary statistics. We are guided mostly by standard considerations. Among the demographic variables, we include four age categories, gender, marital status, whether any of the household members has any children and whether one or more of them live in the household, region of residence (East/West), and German citizenship of the respondent. We generate an education variable in three categories based on the highest degree obtained. The category with low education consists of respondents with primary education (Haupt- or Realschulabschluss) and no vocational training. High education individuals are those with a university or technical college degree (Fachhochschulabschluss). In addition to this, we generate a similar variable for the highest degree of the respondent and his or her partner. We only consider the high education variable of that type because there were very few households in the low education group which led to numerical difficulties.

On the economic side, we utilise labour market status and type of employment, defining a retired household as one where either the respondent or the partner is currently retired and none of them is currently working. For a household to fall into the unemployed, self-employed, or temporary job categories, it is sufficient for one member to do so. We further include monthly net household income (squared) and interact it with retirement. We do so because as labour income drops to zero we expect the relation between income and wealth to become quite different after retirement. With respect to expectations regarding income and labour market status, we include two questions asking for the probabilities for an increase in household income and for at least one household member to become unemployed, respectively, within the next year. Finally, we include a binary variable indicating credit constrained households. It is one if a household has been denied credit in the past five years; or if he did not apply for a credit because he expected the request to be turned down.

In our structural model, we will need instruments for the preference parameters. We recur to questions on childhood behaviour and ratings of the respondent's parents' behaviour. All these items are elicited on a 10-point scale. With respect to risk attitude, we include a variable on whether the respondent was ready to play risky games as a child and whether father and mother, respectively, used to be very adventurous persons. We expect risk-taking behaviour at young ages as well as adventuresome parents to be positively correlated with current risk tolerance. Turning to impatience, we employ a variable on whether the respondent used to spend his or her pocket money immediately as a child and questions on whether the parents used to plan the future in great detail. Again, we anticipate children of planners to be more patient just as those who used to spent very quickly the money they had at their disposal.

6.3 Model and Empirical Strategy

In the last section we showed that in the raw data, wealth, risk tolerance and patience are positively associated with risky asset holdings. Bivariate relations between the two preference parameters and between preferences and wealth exhibited a rather flat profile. All these findings continue to be valid if we employ partially linear models to control for the covariates described in 6.2.4, leaving the functional form of the key explanatory variable unspecified.

These results provide the motivation to investigate the relationships among wealth and preferences and their influence on asset choice in more detail. It is easy to find reasons why any of these eight variables may be endogenous to one or more of the equations determining the others. This calls for a unified model that is explicit about which effects are admitted and which are restricted. Accordingly, we now construct a structural equation model where the key variables are not directly observed. Given the complexity that this will entail, we first sketch the structure and intuition of the model informally in Section 6.3.1. Following this, Sections 6.3.2 to 6.3.4 translate it into a mathematical formulation. We first consider identification of the structural parameters under the assumption that all dependent variables were continuous. Next, we turn to identification of the covariance matrix elements before extending the identification arguments to the case where we have limited dependent variables. Finally, we describe our estimation strategy in 6.3.5.

6.3.1 Wealth, Preferences and Asset Choice: A Structural Approach

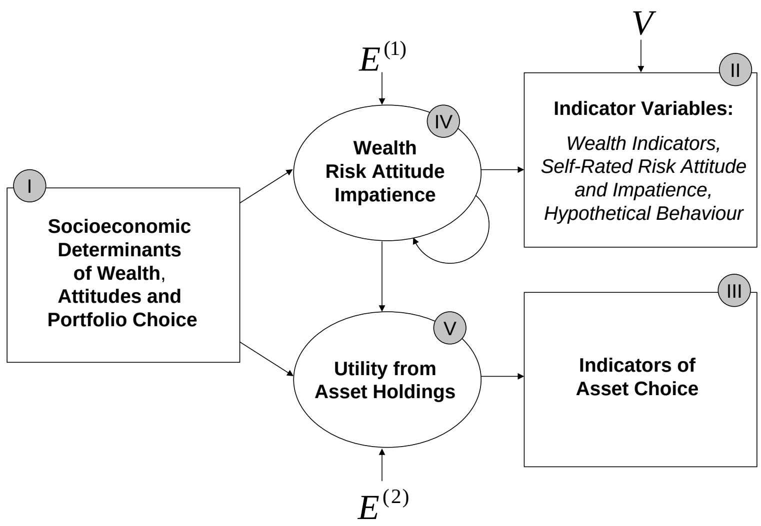
A number of recent papers have shown that subjective assessments of risk-tolerance and behaviour in hypothetical choice tasks are useful predictors of everyday financial behaviour (see Arrondel and Masson (2005), Barsky, Juster, Kimball, and

Shapiro (1997), Bertaut (1998), Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2005), Guiso and Paiella (2004), and Haliassos and Bertaut (1995) for a certainly non-exhaustive list). The approach taken has usually been to use the derived preference parameters as explanatory variables in an equation determining the outcome of interest.

Apart from a standard measurement error argument, the main shortcoming of this strategy is related to the fact that it is essentially a reduced form of a model for preferences and portfolio choice. As with any reduced form, the researcher is unable to tell – for example – direct wealth effects from those mediated through preferences. In our modelling approach, we distinguish explicitly between exogenous determinants of wealth, preferences, and asset choice on the one hand; and variables that are the outcomes of processes depending on these parameters on the other hand. This way we are very explicit about the assumed directions of causality. As usual, the exogeneity assumption is certainly debatable with respect to some of the variables we listed in Section 6.2.4. Many economic models for educational choices, labour market participation, or marriage behaviour depend on risk and (or) time preferences. However, while in principle nothing would inhibit us from extending our model by incorporating these among the endogenous variables, even the current formulation appears to be approaching the limit of our data. Hence we concentrate on the causation mechanisms that appear most important to us.

Figure 6.5 contains a graphical representation of our model. Starting from the left with Panel I, it contains the observable determinants of our key variables. The exogeneity assumption that we just stated is depicted by the fact that no arrows lead towards it. The parameter matrices associated with the connections to Panels IV and V cannot be completely unrestricted, we will comment on the necessary exclusion restrictions below. The two oval panels in the middle contain the model's core, namely the eight latent variables wealth, risk attitude, impatience, and the five utilities relating to the asset classes. They consist of an observable component (the appropriate elements of Panel I mediated through the respective parameters) and an unobservable one ($E^{(1)}$ and $E^{(2)}$ in Figure 6.5). The variables in Panel IV may depend arbitrarily upon each other, which is depicted by the arrow starting from and leading back to it. This generality is important for a variety of reasons. We expect wealth to influence risk attitude towards more risk tolerance, at least in situations with fixed stakes. The reason for this is the larger buffer leading to the decreasing absolute risk aversion typically found in between-subject studies. It is difficult to sign the impact of risk attitude on wealth a priori. On the one hand, the more risk tolerant may hold a larger share of risky assets that, on average, earns them a risk premium. On the other hand, the more prudent can be expected to save more, leading to higher wealth on average. If we observed past behaviour long enough and for a sufficient number of subjects, we could in princi-

Fig. 6.5: A Structural Model of Behavioural Traits, Wealth, and Asset Choice



ple distinguish these effects. Since we do not, they will both be picked up by our risk attitude variable. Similar arguments can be made with respect to wealth and impatience: The more patient can be expected to have a higher savings rate, leading to higher wealth. For the other direction, we would expect the more affluent to show more patient behaviour, as they will not be affected as much by market frictions such as credit constraints.

The main parameters of interest are associated with the downward arrow that leads from Panel IV to Panel V. It stands for the utility associated with holding any of the five portfolio components. We expect that, conditional on the socio-demographic variables in Panel I and the two preference parameters, the more wealthy will have a higher propensity to hold any of the assets, except for loans. The reason for this is the standard optimality of diversification result combined with fixed cost of holding one of the asset classes. Second, again conditional upon the socio-economic determinants of asset choice, wealth and impatience, the more risk-loving should hold a higher share of risky assets for obvious reasons. Finally, under the adjusted conditioning statement, we expect the more patient to possess a higher share of illiquid assets.

This generality in the model core comes at a price – we will see in the following section that we need exclusion restrictions in the equations determining wealth, attitudes, and utilities from asset holdings. As noted above in 6.2.4, we assume that behaviour at very young ages and the rating of parents' personality traits have an impact on the associated preference parameter only. While these are certainly rather soft instruments, we do not see any important direct effects that these variables should have on wealth or asset choice. The ones that we use for wealth are certainly more controversial. First, we take the dummy variable for being divorced. A marriage break-up is certainly associated with a change in wealth that is arguably not related to preferences or utilities from asset holdings. Note that the direction is not quite clear since we control for being married. Comparing the divorced with those who never married, the marriage episode may have either led to an increase or a decrease in wealth, depending on the own and the ex-spouse's endowment. The second variable that we exclude from all other equations is income in the four forms described above. Many economic models emphasise the role of idiosyncratic labour income risk on portfolio choice (see for example Heaton and Lucas (2000) or Viceira (2001)). We acknowledge this by including type of employment – i.e. whether any household member is self-employed or holds a temporary job – among the determinants as well as the self-perceived probability to become unemployed within the next year.

Next to this structural approach, the explicit consideration of measurement error is a major component of our model. While the approach of using different noisy indicators to reduce or even eliminate problems associated with measurement error is not new (see for example Bound, Brown, and Mathiowetz (2001) or Wansbeek and Meijer (2000)), we appear to be the first to apply it in this context. On each of the latent wealth and attitude parameters we observe several measurements as outlined in 6.2.2 and 6.2.3 and contained in Panel II of Figure 6.5. All of these are measures of one or more of the (appropriately scaled) latent variables together with an error, denoted by the vector V. Finally, we do not observe the utility from asset holdings directly, but only a set of binary indicators that completes Figure 6.5 through Panel III.

Finally, let us summarise the structure of our data. We observe three sets of variables: First, the determinants of asset choice, wealth, and preferences (Panel I in Figure 6.5). Second, twelve measurements on wealth and attitudes (Panel II). Third, five indicators on the presence of every single asset class in a household's portfolio (Panel III). With these observations at hand, we seek to identify five sets of parameters. First and most importantly, the ones that govern the relationships among the endogenous variables. Second, the parameters on the socio-economic determinants of asset choice and wealth. Third, the way the measurements in Panel II are generated from the underlying constructs in Panel IV. Fourth, the variances and covariances of the idiosyncratic components of wealth, attitudes, and utilities from asset holdings ($\mathbb{V}[(E^{(1)}, E^{(2)})']$). Fifth, the variances of the measurements on wealth and preferences ($\mathbb{V}[V]$). Before turning to the estimation, we first demonstrate that these parameters are indeed identified.

6.3.2 Identification in the Model with Continuous Outcomes

We cast the model depicted in Figure 6.5 into the standard LISREL form (Jöreskog 1969, 1970, 1977) . Our model also borrows from the extended heterogeneous logit models and general modelling approaches of Börsch-Supan, McFadden, and Schnabel (1994), Ben-Akiva et al. (1999), Harris and Keane (1999), and Ben-Akiva et al. (2002). The core of our model is a structural equations system with eight unobserved variables: Wealth, risk attitude, impatience, and the utility corresponding to each of the five asset types (Panels IV and V in Figure 6.5). We collect them in the $L-$vector $Z^* = \left(w^*, r^*, t^*, u_1^*, u_2^*, u_3^*, u_4^*, u_5^*\right)'$ and describe the relations among them by a parameter matrix Λ. The K determinants of Z^* from Panel I in Figure 6.5 make up the vector X which we observe directly. Their influence on Z^* is governed by the $K \times L$ parameter matrix B. Finally, we stack $E^{(1)}$ and $E^{(2)}$ in one vector E with elements $\varepsilon_l, \quad l \in \{1, 2, \ldots, L\}$ of error terms corresponding to Z^*. Hence the core of our model looks as follows:

$$Z^* \quad = \quad \Lambda \cdot Z^* + B'X + E. \tag{6.1}$$

All the variables in Z^* are not directly observed, but at least in part for different reasons. For wealth and the preference parameters, we have several error-ridden measurements which are influenced by each of these variables (Panel II in Figure 6.5). We analyse them in a factor structure. For the utility variables, we do not observe them directly but a set of binary asset choice indicators depicted in Panel III. Similar problems of limited dependent variables recur in some of the measurements. Identification arguments with respect to the latter point are straightforward in our setup (see Nelson and Olson (1978) or Wansbeek and Meijer (2000) for a textbook treatment). We separate the issues of factor analysis and limited dependent variables and treat the latter in Section 6.3.4. For now, we treat u_1^* to u_5^* and all indicators of w^*, r^*, and t^* as observed and continuous.

We collect the $M^{(1)}$ measurements on the factors described in Section 6.2 and the $M^{(2)}$ utilities of the asset types in the $M-$vector $Y^* = \left(y_1^*, y_2^*, \ldots, y_{M^{(1)}}^*, y_{M^{(1)}+1}^*, \ldots, y_M^*\right)$. The measurement model can then be written down as:

$$Y^* \quad = \quad A \cdot Z^* + V. \tag{6.2}$$

Accordingly, A is a $M \times L$-matrix with elements α_{ml} and V a vector of measurement errors with elements $\nu_m, \quad m = \{1, 2, \ldots, M\}$. In order to express the system in terms of observables and disturbance terms, we derive the reduced form of Equation (6.1) and then plug it into Equation (6.2):

$$Z^* \quad = \quad (\mathbb{I}_L - \Lambda)^{-1} \left[B'X + E\right] \tag{6.3}$$
$$Y^* \quad = \quad A \cdot (\mathbb{I}_L - \Lambda)^{-1} \left[B'X + E\right] + V \tag{6.4}$$

Tab. 6.2: Indicator Variables and Restrictions on the Loadings in the Factor Model

Measurement	**Element of Y^***	**Element of Matrix of Factor Loadings A** w^*	r^*	t^*	u_1^*	u_2^*	u_3^*	u_4^*	u_5^*	**Element of V^***
measwealth	y_1^*	1	0	0	0	0	0	0	0	v_1^*
incwealth	y_2^*	α_{2w}	0	0	0	0	0	0	0	v_2^*
taxadvisor	y_3^*	α_{3w}	0	0	0	0	0	0	0	v_3^*
lottery1	y_4^*	0	1	0	0	0	0	0	0	v_4^*
lottery2	y_5^*	0	α_{5r}	0	0	0	0	0	0	v_5^*
discount1	y_6^*	0	0	1	0	0	0	0	0	v_6^*
discount2	y_7^*	0	0	α_{7t}	0	0	0	0	0	v_7^*
lottery3	y_8^*	0	α_{8r}	α_{8t}	0	0	0	0	0	v_8^*
drink	y_9^*	0	α_{9r}	α_{9t}	0	0	0	0	0	v_9^*
smoker	y_{10}^*	0	α_{10r}	α_{10t}	0	0	0	0	0	v_{10}^*
planner	y_{11}^*	0	α_{11r}	α_{11t}	0	0	0	0	0	v_{11}^*
impulsive	y_{12}^*	0	α_{12r}	α_{12t}	0	0	0	0	0	v_{12}^*
savacc	y_{13}^*	0	0	0	1	0	0	0	0	0
retinsur	y_{14}^*	0	0	0	0	1	0	0	0	0
bonds	y_{15}^*	0	0	0	0	0	1	0	0	0
stocks	y_{16}^*	0	0	0	0	0	0	1	0	0
loans	y_{17}^*	0	0	0	0	0	0	0	1	0

$\mathbb{I}_L$ denotes the L–dimensional identity matrix. Under the additional assumptions of mean independence of the joint error terms $\mathbb{E}\left[A \cdot (\mathbb{I}_L - \Lambda)^{-1} E + V \mid X\right] = 0$ and $\mathbb{E}[X'X]$ having full rank, we can derive the following equation:

$$(\mathbb{E}[XX'])^{-1}\,\mathbb{E}[XY^{*\prime}] = B(\mathbb{I}_9 - \Lambda)^{-1\prime} A' = \Pi \tag{6.5}$$

The goal is to identify the parameters of B, Λ, and A which is clearly an impossible task without appropriate exclusion restrictions and normalisations. We start with the matrix of factor loadings A and list Equation (6.2)'s components along with their verbal descriptions in Table 6.2.

Normalising the first nonzero element of each column in A to one is without loss of generality since the scale is not identified in factor models. Taking reported wealth as the first indicator leaves the wealth factor with its natural interpretation. We do not have such straightforward scalings available for the other factors and the choice of normalisation is arbitrary.

In contrast to typical factor models, the factors in our model are not independent. Instead, the relations among them are described by the matrix Λ. These assumptions lead us to impose restrictions on A which appear to be more restrictive than the triangular structure imposed by independent factor models (see for example Carneiro, Hansen, and Heckman (2003)). However, the differences boil down to a different view of the underlying mechanisms. Take wealth and risk as an example and assume that we had a lower triangular structure of A and that $\lambda_{rw} = \lambda_{wr} = 0$. A ceteris paribus change in true wealth would not change anything about the factor risk attitude, however the measurements relating to the latter would change through, e.g., α_{4w}. In our case, this loading is restricted to zero. Still, we would expect the measurements on risk attitude to change because in our setup the risk factor itself changes through the parameter λ_{rw}. Our approach can be viewed as somewhat more modest in the sense that we do not claim to be able to identify a risk attitude parameter that is independent from wealth. We acknowledge that we always observe preferences and wealth together and that these may be influencing each other as outlined above in 6.3.1. We can then estimate the averages of these influences. In the independent factor model, one has to take up a stance on the direction in which factors influence the measurements dedicated to other factors. In other words, A has to be triangular while we can get by with a full interdependence structure in Λ.

The $K \times M$-matrix Π has at most $L < M$ independent columns. The dependence structure becomes clear from inspecting the matrix A. The second and third columns of the LHS of (6.5) are equivalent to the first column multiplied by α_{2w} and α_{3w}, respectively. Similarly, the fifth (seventh) column is equal to the fourth (sixth) scaled by α_{5r} (α_{7t}). Columns nine to thirteen are the sums of column four scaled by α_{mr} and column seven scaled by α_{mt}, $m = 9, 10, \ldots, 13$. Finally, we do not have a measurement model for the utilities and they are identical to the corresponding elements of Y^* (note that the last six elements of V are zero).

The argument in the preceding paragraph yields identification of the free parameters in A conditional upon the elements of $B(\mathbb{I}_9 - \Lambda)^{-1\prime}$ being identified. Neglecting A' and the corresponding columns on the RHS of Equation (6.5), we are left with a system of eight linear equations that has a long history in econometrics (Koopmans, Rubin, and Leipnik 1950). We first turn to the restrictions we impose upon the matrix Λ governing the relations amongst the endogenous variables:

$$\Lambda = \begin{pmatrix} 0 & \lambda_{wr} & \lambda_{wt} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{rw} & 0 & \lambda_{rt} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{tw} & \lambda_{tr} & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{u_1w} & \lambda_{u_1r} & \lambda_{u_1t} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{u_2w} & \lambda_{u_2r} & \lambda_{u_2t} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{u_3w} & \lambda_{u_3r} & \lambda_{u_3t} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{u_4w} & \lambda_{u_4r} & \lambda_{u_4t} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{u_5w} & \lambda_{u_5r} & \lambda_{u_5t} & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{6.6}$$

This translates the verbal description and intuition outlined in Section 6.3.1. There is a fully flexible interdependence structure among the wealth and preference variables. They impact upon the utilities relating to the different asset types, but there are no repercussions from or direct relations among the asset variables. As noted before, we need to impose certain exclusion restrictions on the parameters in B. While we could get by with one instrument for each of the three latent variables, we chose to use a set of three or more instruments for each of them (see the verbal outline in 6.3.1, they also become apparent from the table of results 6.11 in the Appendix). With all the parameters in A, B, and Λ at hand, we can now turn to identification of the parameters in the covariance matrices relating to E and V.

6.3.3 Identification of the Covariance Parameters

We start with some notation. Let $\Psi = \mathbb{E}[EE']$ be the variance-covariance matrix of the structural errors and $\Theta = \mathbb{E}[VV']$ be the variance-covariance matrix of the measurement errors relating to the indicators. We leave Ψ completely unrestricted at this point. Clearly we could not identify its parameters nor those of Θ if the latter was not subject to any restrictions, either. We follow the literature on factor analysis in assuming that the elements of V are pairwise uncorrelated. Hence, Θ is a diagonal matrix. Furthermore, remember that we do not have a measurement model for the utility variables, so the last six diagonal elements of Θ are zero. Last, we assume that E and V are uncorrelated, so that measurement errors are not related to the disturbances of the structural equations (6.1).

The identification arguments are based on the variance-covariance matrix of the entire disturbance terms. Under the assumptions that we just stated, they can be written as follows:

$$\begin{aligned} &(6.7) \qquad \mathbb{V}\left[A\,(\mathbb{I}_9-\Lambda)^{-1}E+V\right] &&= A\,(\mathbb{I}_9-\Lambda)^{-1}\;\Psi\,(\mathbb{I}_9-\Lambda)^{-1\prime}A' \;+\; \Theta \\ &(6.8) && = \Omega \end{aligned}$$

Because of the structure inherent to Λ, all $\mathbb{E}[\varepsilon_i\varepsilon_j], \quad i,j \in \{w^*, r^*, t^*\}$ appear in all equations. Hence we need a system of six independent equations from Ω that does not involve any unknowns except for these terms. In order to have two independent equations that involve any ε_i^2 but no v_m^2 corresponding to measurements affected directly by factor i, we need at least two indicators devoted exclusively to one factor. This criterion is just fulfilled for r^* and t^* and overfulfilled for w^*. As an example, we could then use the first two intersections by measurements on the same factor $\Omega_{2,1}$, $\Omega_{5,4}$, $\Omega_{7,6}$ together with those elements of Ω where the normalised elements of A intersect, that is $\Omega_{4,1}$, $\Omega_{6,1}$, $\Omega_{6,4}$ to identify these variables.

Now all parameters of Θ are identified from the first $M^{(1)}$ diagonal elements of Ω. Since the last $M^{(2)}$ elements of Θ are zero, the ε^2-terms corresponding to the utilities of asset types can be inferred directly from the appropriate diagonal elements of Ω. Finally, since there is no feedback from utility of assets to preferences or wealth, the corresponding covariances can be inferred one by one from the intersection of one respective measurement and the asset utility. The same argument holds for the covariances of the disturbances of the utilities themselves. This closes the discussion of identification under the presence of continuous measurements and we now turn to the additional issues raised by observing only limited dependent variables in most cases.

6.3.4 Identification in the Presence of Discrete Measurements

Of the nineteen outcomes that we measure (13 indicators on wealth and preferences; holdings of 6 types of assets), only two are continuous. Eight are of binary nature and the remaining nine outcomes are ordinal variables that take on three or four values. We follow the exposition in Wansbeek and Meijer (2000). We specify the following observation rules for the indicator variables:

$$y_m = H_m(y_m^*, \tau_m), \qquad m = \{1, 2, \dots, M\} \tag{6.9}$$

Each $H_m(\cdot)$ is a known deterministic function that maps the unobserved measurement y_m^* onto the observed variable y_m, depending additionally on a parameter vector τ_m. For the two continuous indicators, $H_m(\cdot)$ is the identity function and τ_m is empty. For the other indicators, we have the ordered response model:

$$y_m = \begin{cases} 0 & \text{if} \quad y_m^* \leq \tau_{m,1} \\ 1 & \text{if} \quad \tau_{m,1} < y_m^* \leq \tau_{m,2} \\ \vdots & \\ J & \text{if} \quad \tau_{m,J} < y_m^* \end{cases} \tag{6.10}$$

The scale of $y_m^*, \quad m = \{3, 4, \dots, M\}$ is not identified and we set the variances of the respective diagonal elements in Ω to one. This is equivalent to restricting the respective measurement error variances $\Theta_{m,m}, \quad m = \{3, 4, \dots, M^{(1)}\}$ to be functions of elements of Λ, A, and Ψ; hence they are not identified. Finally, remember that all structural equations contain an intercept, so we need to normalise one of the thresholds $\tau_{m,i}$ per latent variable which does not have a continuous indicator. We do so by setting $\tau_{m,1} = 0$ for the first element of each column in A.

Before turning to our estimation approach, we close this section with a note on nature of the discrete variables. It is crucial for our analysis that all endogenous variables are continuous and that it is *not* the discrete outcomes which carry structural meaning. Most importantly, it would not be possible to specify an interdependent system of equations for the latent variables. Instead there could only be a triangular structure. See, for example, Schmidt (1981) on this issue. On the one hand, it is not at all restrictive that it is the latent, continuous wealth and the preference variables which carry structural meaning. Indeed, this is the main point of the measurement models. On the other hand, it inhibits us from considering potential repercussions from asset holdings to wealth that we mentioned in Section 6.2.

6.3.5 Estimation of the Structural Model

Our estimation strategy is based on the assumption that all random errors follow a joint normal distribution with the above-mentioned restrictions on the variances and covariances in Ψ and Θ that translate into restrictions on the covariance matrix Ω of the joint error terms. We employ a full information maximum likelihood approach. Before moving on to the likelihood, it is useful to partition Y, Π, and Ω according to whether the corresponding outcome variable is of continuous or discrete nature:

$$\begin{aligned} Y &= (Y_c, Y_d)' \\ \Pi &= (\Pi_c, \Pi_d) \\ \Omega &= \begin{pmatrix} \Omega_{c,c} & \Omega_{c,d} \\ \Omega_{d,c} & \Omega_{d,d} \end{pmatrix}. \end{aligned}$$

We can write the likelihood for observation i as:

$$(6.11) \quad L_i = \phi\,(Y_{c,i} - \Pi_c' X_i \;;\; \Omega_{c,c}\,) \cdot \int_{\tau_{3,Y_{3,i}}}^{\tau_{3,1+Y_{3,i}}} \cdots \int_{\tau_{M,Y_{M,i}}}^{\tau_{M,1+Y_{M,i}}} \phi\,(s - \mu_{d,i|c} \;;\; \Omega_{d|c}\,)\, \mathrm{d}s_M ds_{M-1} \cdots ds_3$$

where $\phi(\mu\,;\,\Omega)$ denotes the multivariate normal density function with mean μ and variance matrix Ω and

$$\begin{aligned} s &= (s_3, s_4, \ldots, s_M)' \\ \mu_{d,i|c} &= \Pi_d' X_i + \Omega_{d,c}(\Omega_{c,c})^{-1}\,(Y_{c,i} - \Pi_c' X_i) \\ \Omega_{d|c} &= \Omega_{d,d} - \Omega_{d,c}(\Omega_{c,c})^{-1}\Omega_{c,d} \;. \end{aligned}$$

The integral to evaluate is of dimension 16 and does not have an analytical solution. Hence we resort to simulated maximum likelihood estimation, employing the GHK simulator (Börsch-Supan and Hajivassiliou 1993). For the core part of the model, we make use of the Fortran code supplied with Hajivassiliou, McFadden, and Ruud (1996),[5] taking 200 random draws per observation. To maximise the logarithm of the likelihood function (6.12), we utilise the BFGS method with numerical gradient approximations. Standard errors are computed from the inverse outer product of scores matrix. As there are many parameters in the model (Table 6.9 in the Appendix gives a concise overview), computation time is considerable even with up-to-date parallel facilities.

6.4 Results

We divide the discussion of our results in three parts. First, we present the parameter estimates from the measurement model for wealth and attitudes. Second, we report on the impact of the exogenous factors on these variables and asset choice. Finally, we come to the core of our model, i.e. the relationships among the endogenous variables, in particular the influence of attitudes and wealth on the choice of portfolio components.

6.4.1 Measuring Wealth and Attitudes

Table 6.3 contains the matrix of factor loadings A. We describe it column-wise. Looking at the first column it is evident that our binary variables "income from wealth" and "employment of a tax advisor" are indicators of a high level of wealth indeed. The magnitude of the coefficients appears to be plausible, too. To see this, we consider the marginal effects implied by the coefficients. For a baseline wealth of 140,000 Euros (the unconditional average of reported wealth), which is associated with an 18% probability to report income from wealth. A positive wealth shock of 100,000 Euros would lead to an increase of this number to 36%. The same exercise with respect to the tax advisor variable gives numbers of 24% and 27%, respectively. Hence there is more information in the "income from wealth" variable. Finally, the presence of three indicators of one unobserved variable allows us to separately identify the standard deviation of latent wealth and the standard deviation of the measurement error inherent to each of the indicators (unless they are normalised as for the binary variables), see Section 6.3.3. We estimate the idiosyncratic component of (latent) wealth in millions of Euros to have a standard error of .097 (s.e. .112), see Table 6.4. The estimate of the standard deviation of the measurement error in reported wealth is .193 (s.e. .026). According to these results, measurement error would be about twice as important as idiosyncratic variation in true wealth. This appears to be rather large, but not entirely implausible in the light of the difficulties in measuring wealth.

[5] Available at `http://econ.lse.ac.uk/staff/vassilis/pub/simulation/fortran/`.

Tab. 6.3: Results from the Measurement Model

Variable	w*	r*	t*
measwealth	1.000		
incwealth	5.693** (0.445)		
taxadvisor	1.035** (0.246)		
lottery1		1.000	
lottery2		0.567** (0.056)	
discount1			1.000
discount2			0.762** (0.156)
lottery3		1.042** (0.097)	−0.379** (0.142)
drink		0.194** (0.054)	−0.921** (0.142)
smoker		0.227** (0.056)	0.882** (0.131)
planner		0.008 (0.053)	0.828** (0.118)
impulsive		0.280** (0.039)	0.215* (0.095)

Note: Estimated elements of the matrix A as defined in Section 6.3. The first element of each column is normalised to one. Standard errors are shown in parentheses, asterisks indicate significance at the 5% and 1%-level. *Source:* SAVE 2005, own calculations.

We code all measurements of risk attitude such that higher values are associated with more risk averse behaviour. From the normalisation $\alpha_{4w} = 1$ we then expect positive coefficients throughout column 2, which is what we find. All effects except for the self-rating with respect to planning the future are highly significant. The first thing we examine are the relative magnitudes of the idiosyncratic variation in risk attitude and the measurement errors in the indicators associated with it. Remember that the latter are normalised to 1 for all 7 measurements. The standard deviation of the idiosyncratic component is estimated to be 1.002 (s.e. .252). This highlights two things: On the one hand, idiosyncratic variation in risk preferences beyond those related to demographic characteristics and economic variables is substantial. On the other hand, measurement error in typically employed direct indicators of risk attitude is very important. In regressions where such measures are directly employed, coefficients can be expected to suffer from a severe attenuation bias.

The magnitudes of the coefficients are not directly comparable because of the different thresholds, which are shown in Table 6.10 in the Appendix. We therefore compare marginal effects directly. Such comparisons reveal that lotteries 1 and 3 show the largest reaction to variation in true risk attitude by one standard deviation. For lottery 2, there is some variation in the extremes, but probabilities for the two middle categories do not change much over the bulk of the distribution of r^*. This may be related to the fact that the gambles involve negative outcomes. If loss aversion is important, explaining the behaviour by risk attitude alone would lead to a downward bias in α_{5r}. Marginal effects with respect to drinking, smoking, and impulsive decision-making are relatively small. Again, most of the action is in the extreme categories as opposed to the middle one. Finally, as could be expected from the estimated coefficient, planning the future is not indicative of risk tolerance.

The coding of the indicators for impatience is such that higher values of t^* are associated with more patient behaviour. A priori, we would expect all coefficients in the last column of Table 6.3 to be positive except for the one on lottery 3. As it happens, the risk-avoiding choices in that task are the ones with immediate payments while risky choices and delayed gratification are aligned in all other indicators. We find all variables to be significant. With the exception of drinking frequency, they all point in the expected direction. Measurement errors are about four times as high as the idiosyncratic variation in time preference, which is only imprecisely estimated. The favourable interpretation of this would be that after controlling for the covariates (which we emerges to be important as illustrated by the next section), there is little idiosyncratic variation left to explain. The less favourable view is that time preference is difficult to grasp psychologically (see Frederick, Loewenstein, and O'Donoghue (2002) who argue against the economic approach of modelling attitudes to time by means of a single parameter. Arrondel

Tab. 6.4: Correlation Matrix of Structural Errors

Variable	$\mathbf{w}^*$	$\mathbf{r}^*$	$\mathbf{t}^*$	$\mathbf{u}^*_{\text{savacc}}$	$\mathbf{u}^*_{\text{retins}}$	$\mathbf{u}^*_{\text{bonds}}$	$\mathbf{u}^*_{\text{stocks}}$	$\mathbf{u}^*_{\text{loans}}$
$\mathbf{w}^*$	0.097 (0.112)	0.172 (0.203)	0.149 (0.377)	−0.061 (0.137)	−0.096 (0.116)	−0.278 (0.195)	−0.055 (0.120)	−0.305** (0.117)
$\mathbf{r}^*$		1.002** (0.252)	−0.971 ^{x})	0.015 (0.157)	0.253* (0.124)	0.108 (0.203)	0.111 (0.125)	0.092 (0.118)
$\mathbf{t}^*$			0.199 (0.352)	0.452 (0.392)	−0.151 (0.173)	0.078 (0.305)	0.107 (0.197)	−0.262 (0.177)
$\mathbf{u}^*_{\text{savacc}}$				1.000	0.245** (0.067)	−0.024 (0.121)	0.035 (0.071)	−0.008 (0.072)
$\mathbf{u}^*_{\text{retins}}$					1.000	0.301** (0.089)	0.258** (0.053)	0.177** (0.061)
$\mathbf{u}^*_{\text{bonds}}$						1.000	0.320** (0.073)	0.003 (0.101)
$\mathbf{u}^*_{\text{stocks}}$							1.000	0.002 (0.086)
$\mathbf{u}^*_{\text{loans}}$								1.000

Note: Estimated elements of the correlation matrix associated with Ψ as defined in Section 6.3. Diagonal elements denote standard deviations, off-diagonal elements correlations. Standard errors are shown in parentheses, asterisks indicate significance at the 5% and 1%-level. ^{x}) no meaningful standard error. *Source:* SAVE 2005, own calculations.

and Masson (2005) also have more difficulties to pin down impatience than risk preferences.

Given the stronger relative importance of measurement error, marginal effects on the indicators of impatience are much smaller than those for the measurements of risk attitude. Their relative magnitudes are broadly aligned with the size of the factor loading estimates. It is noteworthy that planning the future is so strongly related to time preference in the light of the no-correlation result of Ameriks, Caplin, and Leahy (2003). This suggests the measurement error interpretation of some of their findings brought forward by Arrondel and Masson (2005): A high propensity to plan may simply be a better indicator of patience than the single hypothetical question Ameriks, Caplin, and Leahy (2003) employ to measure discounting behaviour.

We conclude from this exercise that measurement error is a potentially serious issue in attempts to elicit time preference directly. In particular, given its relative importance we would not expect many significant results to emerge from a single indicator on it, which may explain some of the negative conclusions in, for example, Kapteyn and Teppa (2003). However, even with our battery of indicators, we cannot pin down idiosyncratic preference parameters very well. In the next section, we examine whether observable variables have explanatory power for them.

6.4.2 The Determinants of Wealth and Attitudes

Before turning to the influences of observable characteristics on wealth and preferences, we begin with describing the relations among them. Table 6.5 displays the elements of the nonzero components of Λ. In this section, we are only concerned with the top three rows.

First, consider the relations among wealth and risk preference. On the one hand, we find the point estimate of risk aversion on wealth to be positive but insignificant. On the other hand, risk aversion is decreasing strongly and significantly in wealth. The marginal effect of an increase in baseline[6] wealth by 100,000 Euros would lead to a decrease in risk aversion by almost a fourth of a standard deviation ($-.23$). Patience and wealth are positively related in both directions. First, a one standard deviation difference in patience at baseline is associated with a 40,000 Euros wealth difference. This effect is likely to be mediated in large parts through higher savings rates in the past which we cannot control for. Second, 100,000 Euros more wealth than at baseline come along with a patience parameter that is about two fifth of a standard deviation higher than at baseline.

The coefficients relating the preference parameters to each other (i.e. λ_{wr} and λ_{rw}, respectively) are difficult to interpret. The reason lies in the correlation of the idiosyncratic parameters which is strongly negative and appears to be approaching the boundary of the parameter space. It is extremely imprecisely estimated. While covariance matrix estimation is a difficult task in general, we have not yet made out the precise reason for why this happens in our particular case. One may worry that the large absolute value of the correlation coefficient could drive some of the results. This does not appear to be the case: Estimates with Ψ being restricted to a diagonal matrix in many respects do not differ substantially from those reported here. Tables are available from the authors upon request.

Because of its sheer size, we move the matrix of coefficients determining the preference parameters to the Appendix in Table 6.11. Here we stick to highlighting its most salient features as self-contained as possible. Note that, except for the above-mentioned exclusion restrictions that appear in rows 2–11 of Table 6.11, none of the restrictions imposed on B is of substantive importance. We introduced them merely to limit the number of parameters.

We start with the observable determinants of wealth. We find the typical concave relationship between income and wealth. Over the bulk of the support of the income distribution, the slope is steeper for retired households. A divorce appears to affect income negatively as compared to single households, although the effect is not significant. If at least one household member holds a college or equivalent degree, household wealth is predicted to be about $27,000$ Euros higher. Since we

[6] Unless stated otherwise, in the subsequent discussion we refer to the baseline specification as one where all observed variables are set to their median. Unobservables are set to zero.

Tab. 6.5: Relations Among Wealth, Preferences, and Asset Choice

Variable	$\mathbf{w}^*$	$\mathbf{r}^*$	$\mathbf{t}^*$
$\mathbf{w}^*$		0.006 (0.015)	0.157** (0.036)
$\mathbf{r}^*$	−3.657** (1.146)		1.750** (0.537)
$\mathbf{t}^*$	1.079** (0.342)	0.131* (0.054)	
$\mathbf{u}^*_{\text{savacc}}$	2.264* (0.950)	0.026 (0.116)	1.068** (0.407)
$\mathbf{u}^*_{\text{retins}}$	2.245* (0.909)	−0.326** (0.098)	0.409 (0.319)
$\mathbf{u}^*_{\text{bonds}}$	2.031 (1.451)	−0.204 (0.162)	1.679* (0.672)
$\mathbf{u}^*_{\text{stocks}}$	2.770** (0.984)	−0.211* (0.098)	0.787* (0.356)
$\mathbf{u}^*_{\text{loans}}$	0.632 (0.992)	−0.123 (0.097)	−0.173 (0.326)

Note: Estimated elements of the matrix Λ as defined in Section 6.3. Empty elements are normalised to zero. Standard errors are shown in parentheses, asterisks indicate significance at the 5% and 1%-level. *Source:* SAVE 2005, own calculations.

control for the influence of education on wealth that is mediated through preferences, this hints at the importance of information processing or similar influences. As expected, wealth rises in age, although the difference to the left-out category of respondents between 31 and 45 years of age is only significant for those aged between 46 and 60. Finally, wealth of the self-employed is about 100,000 Euros higher on average.

Turning to risk attitude, the two instruments that emerge as being significant (playing risky games as a child, adventuresome father) point into the expected direction: For both questions, higher scores are linked with higher degrees risk tolerance. The parameter estimate on the third instrument, the respondent's rating of his or her mother being adventuresome, is small, insignificant, and points into the opposite direction than the other two. We find females to be significantly more risk averse than men, which is in line with many findings in the literature. The estimates on education and age are particularly pronounced. The risk attitude of respondents with low education is about half a standard deviation lower on average than in the middle education category. Effects are about 20% lower (higher) than this for respondents between 46 and 60 (more than 60) years of age. Wealth levels in East Germany are substantially lower than in the West.

With respect to impatience, estimates on all our instruments point in the expected direction and are statistically significant. Saving one's pocket money as a child and rating one's parents as planning the future precisely are all associated with higher levels of patience. Females are more impatient on average, while married respondents appear to be more patient. Note, however, that this effect will be counterbalanced by the effect of having children. This will be relevant for many of the married respondents and it points in the opposite direction, displaying about the same precision and magnitude. We estimate a strong and monotonous effect of the respondent's education. Higher education is associated with more patient behaviour. Finally, the self-employed appear to more impatient than tenured employees while East Germans are more patient than their Western counterparts.

6.4.3 The Determinants of Asset Choice

Before turning to the influence of the three latent variables, we describe the impacts of the exogenous variables on asset holdings. We find married households to be more likely to hold savings accounts and also to have loans in their portfolio. The young are less likely to have retirement insurance contracts or own stocks. This is in line with a preference for liquid assets at younger ages which could result from borrowing constraints. The negative association of retirement insurance contracts with the oldest age group and the retired is trivial since only contracts in the accumulation phase enter the definition. Consistent with the life-cycle model, their portfolios are also much less likely contain loans. Respondents who expect their household income to increase within the next year hold a larger share of risky assets and retirement insurance contracts. The effects of credit constraints are surprising at first glance, with households who report being credit constraint holding less savings accounts and retirement insurance contracts, but more loans. Among these effects, only the lower propensity to save in a retirement insurance contract appears sensible because of the illiquidity of such contracts. We suspect the clue to this apparent paradox to be in the definition of the credit constraints variable. At present we are using whether a credit request has not been granted at all or not been granted to its full extent. We suspect the effects to be sensitive to the inclusion of the last category and will check this in future versions.

Turning to the effects of wealth and preferences on asset choice displayed in the lower part of Table 6.5, we find that almost all of them point in the expected direction. Greater wealth leads to significantly higher probabilities to own savings accounts, retirement insurance contracts, and stocks. The effect on bonds is insignificant but positive as expected, whereas the point estimate on loans is opposite to our priors. However, note the strong negative correlation in the errors from Table 6.4 which reestablishes the observed negative correlation between wealth and loans in the population. Higher levels of risk aversion are associated with lower holdings of stocks and retirement insurance contracts. The first finding is intuitively clear – being the most risky asset category, all else being equal their possession should be positively associated with risk tolerance. The effect on retirement insurance contracts could be explained by a preference of precautionary savers for liquid assets. Finally, the last column of Table 6.5 reveals that the patient are more likely to hold savings accounts, bonds and stocks. Especially the last two effects are well in line with our priors, given the relative illiquidity of these asset classes. As a last point, the results in the right bottom corner of Table 6.4 show a strong positive correlation among the errors of retirement insurance contracts, bonds, and stocks. The unobserved factor driving this could be, for example, financial education since all three asset categories mean complicated investments relative to savings accounts. Another factor driving these results could be their relative illiquidity if our controls for credit constraints are less than perfect.

Finally we display the marginal effects of changes in wealth and preferences on asset holding probabilities in Table 6.6. Again, we use baseline values with all variables at their median. We then consider changes in asset holding percentages with respect to increases in wealth, risk *tolerance*, and impatience by one standard deviation each. We only consider direct effects since many of the indirect effects involve dynamic arguments.[7] Marginal effects for wealth are largest for savings accounts. This is is a reflection of the skewed distribution of wealth and

[7] For example, we would expect the positive effect of patience on wealth to be mediated largely through the savings rate. Looking at instantaneous marginal effects does not seem sensible.

Tab. 6.6: Marginal effects of Changes in Wealth and Preferences on Asset Holdings

	Baseline	Δ^*_w	Δ^*_r	Δ^*_t
P(savacc = 1)	0.617	0.083	-0.010	0.078
P(retins = 1)	0.171	0.063	0.095	0.022
P(bonds = 1)	0.001	0.001	0.001	0.002
P(stocks = 1)	0.070	0.046	0.033	0.024
P(loans = 1)	0.186	0.017	0.035	-0.009

Note: Marginal effects relative to the baseline specification with all variables set to their median. Only direct effects are considered for marginal effects, i.e. the first three rows in Λ are set to zero. $\Delta w^* = .1$, $\Delta r^* = -1.002$, $\Delta t^* = .199$, *Source:* SAVE 2005, own calculations.

risky asset being held predominantly at wealth levels well above the median. This becomes most evident for bonds. Here, baseline rates and marginal effects are extremely small. The population share of bond holdings (8%) appears to be almost exclusively generated by the (very) rich, risk tolerant, and patient. Wealth effects are sizeable for the retirement insurance contract and stock holdings. Influences of preferences on portfolio choice are significant in magnitude, too. For example, increasing risk tolerance by one standard deviation will lead to 50% increases in both stock and retirement insurance contract holdings. At the same time, holdings of savings accounts are predicted to decrease slightly, although this effect is insignificant. Looking at impatience, the marginal effect is once more most pronounced for savings accounts due to the reasons explained before. Effects on propensities to own stocks and retirement insurance contracts are also significant. They do not reach the same magnitude as the effects of risk tolerance though.

6.5 Conclusions

In this paper, we developed a structural model of portfolio choice behaviour. It is predominantly aimed at estimating the influences of wealth, individual risk attitudes and impatience on households' asset choices. In pursuing this aim, we were explicit about what we consider to be observable determinants of wealth, preferences, and asset choice on the one hand; and what we model as outcomes driven by these latent variables on the other hand. Accordingly, a core part of the approach is concerned with measuring the preference parameters and wealth, much in line with psychometric methods.

Our key results can be summarised as follows. We find the measurement model to be extremely important in pinning down the preference parameters. In particular, our results suggest that some of the disappointing results obtained in previous attempts to use individual preferences in portfolio choice models are likely to be due to the impact of measurement error. For example, we find idiosyncratic variation in risk attitude to be about as large as the measurement error in the best indicators that we have for it. The picture is even worse for time preference. Hence, we would not expect a single measure on either trait to produce consistent results. However, the preference parameter estimates obtained from the more comprehensive measurements turned out to have a significant impact on portfolio choice decisions in the expected directions. Allowing for the joint determination of wealth and preferences also emerged as an important ingredient to our model. This highlights the importance of employing a structural model of the type considered here; as opposed to reduced form approaches. The results for the observable determinants of wealth, preferences, and asset choice are broadly consistent with the previous literature.

The analysis in this paper has shown a fruitful approach to incorporate individual preference parameters in models of portfolio choice. Yet there are several open issues left that call for further analyses and possible extensions. Most obviously, the model should be able to accommodate housing wealth given the large share it occupies in total households' net worth. Second, we could move beyond mere probabilities to hold assets towards more comprehensive measurements of portfolio characteristics. Finally, there are some open issues with respect to estimation of the covariance parameters of the unobserved traits.

6.6 Additional Tables

Tab. 6.7: Definitions and Descriptive Statistics of the Indicator Variables

Name	**Definition**	**Min**	**Max**	**Mean**	**SD**
measwealth	Total measured household wealth, sum of all asset holdings, in mill. Euros	-.45	2.5	0.14	0.24
riskindex	Compound index from 9 different questions on risky behaviours	.61	10	6.11	1.29
incwealth	Household income from wealth.	0	1	0.28	0.45
taxadvisor	Employs tax advisor for doing tax returns	0	1	0.24	0.43
lottery1	Choice: 50-50 Lottery between 0 – 2000/2500/3000 vs. sure 1000	0	3	2.60	0.87
lottery2	Choice: 50-50 Lottery between -100 – 200/300/400 vs. sure 0	0	3	2.20	1.24
discount1	Choice: Get paid 1100 now vs. 1130/1200/1380/ in 10 months	0	3	1.04	1.11
discount2	Choice: Pay 825/870/990 in 10 months vs. 800 now	0	3	2.27	1.08
lottery3	Choice: 50-50 Lottery betw 0 – 750/1200/1800 in 6 months vs. sure 500 now	0	3	2.54	0.82
drink	Respondent drinks alcohol: very frequently / regularly / seldom-never	0	2	1.40	0.70
smoker	Respondent: smokes / quit smoking / never smoked	0	2	1.13	0.85
planner	Respondent lives for the day / plans the future precisely	0	2	1.26	0.70
impulsive	Respondent decides impulsively / thinks a lot about decisions	0	2	1.06	0.76

Source: SAVE 2005, own calculations

Tab. 6.8: Definitions and Descriptive Statistics of the Covariates

Name	Definition	Min	Max	Mean	SD
incnotret	Total net HH income in thousands of Euros, interacted with being not retired	0	40	1.92	2.57
inc2notret	Total net HH income squared / 10^8, interacted with being not retired	0	16	0.10	0.67
incret	Total net HH income in thousands of Euros, interacted with being retired	0	28	0.48	1.38
inc2ret	Total net HH income squared / 10^8, interacted with being retired	0	7.8	0.02	0.21
divorced	Divorced from spouse	0	1	0.11	0.31
riskgamkid	Was ready to play risky games as a child	0	10	2.57	3.25
advmoth	Mother was an adventurous person	0	10	1.98	2.73
advfath	Father was an adventurous person	0	10	2.71	3.09
pockmonsp	Spent pocket money immediately as a kid	0	10	3.02	3.41
planmoth	Mother used to plan the future meticulously	0	10	5.07	3.26
planfath	Father used to plan the future meticulously	0	10	5.25	3.38
female	Female	0	1	0.50	0.50
married	Married or partner living in HH	0	1	0.66	0.47
educhihh	Highest educactional degree in HH: high education	0	1	0.20	0.40
educlow	Respondent: low education	0	1	0.11	0.32
educhi	Respondent: high education	0	1	0.15	0.36
age1830	Age between 18 and 30 years	0	1	0.13	0.33
age4660	Age between 46 and 60 years	0	1	0.26	0.44
age61plus	Age above 60 years	0	1	0.31	0.46
kids	Respondent and/or spouse has kids	0	1	0.79	0.41
kidsinhh	Children living in the household	0	1	0.38	0.49
retiredhh	At least one HH member retired, none currently working	0	1	0.19	0.39
unempnowhh	At least one HH member currently unemployed	0	1	0.13	0.33
selfemphh	At least one HH member self-employed	0	1	0.09	0.29
tempjobhh	At least one member of HH: has temporary job	0	1	0.08	0.26
east	Lives in eastern germany	0	1	0.30	0.46
german	Respondent is German citizen	0	1	0.97	0.18
incincprob	Probability for an increase in HH income	0	100	19.98	31.55
unempprob	Probability for at least one HH member to become unemployed	0	100	14.59	28.69
creditconstr	Household credit constrained	0	1	0.11	0.32

Source: SAVE 2005, own calculations

Tab. 6.9: Number of Parameters in the Structural Equations Model

Symbol	**#**	**Parameter Description**
Λ	21	Slope coefficients, endogenous variables
B	138	Slope coefficients, exogenous variables
A	14	Factor loadings of measurement model
Ψ	31	Covariance parameters structural equations
Θ	1	Covariance parameters measurement equations
τ	23	Cutoff points for observation rules
	228	**Total**

Tab. 6.10: Estimated Thresholds in Ordered Probit Models

Variable	τ_0	τ_1	τ_2
incwealth	1.721** (0.110)		
taxadvisor	0.866** (0.062)		
lottery1		0.258** (0.031)	0.929** (0.053)
lottery2	0.252* (0.104)	0.373** (0.104)	0.641** (0.105)
discount1		0.672** (0.028)	1.201** (0.038)
discount2	−1.053** (0.049)	−0.614** (0.045)	−0.197** (0.043)
lottery3	−0.410** (0.148)	0.694** (0.156)	1.291** (0.166)
drink	−0.949** (0.121)	0.169 (0.116)	
smoker	0.073 (0.121)	0.782** (0.120)	
planner	−0.905** (0.115)	0.380** (0.116)	
impulsive	−0.050 (0.088)	1.071** (0.086)	

Note: Estimated threshold parameters $\tau_{m,i}$ associated with the observation rules defined in Section 6.3.4. Standard errors are shown in parentheses, asterisks indicate significance at the 5% and 1%-level. *Source:* SAVE 2005, own calculations.

Tab. 6.11: Determinants of Wealth, Preferences, and Asset Choice

Variable	$\mathbf{w}^*$	$\mathbf{r}^*$	$\mathbf{t}^*$	$\mathbf{u}^*_{\mathbf{savacc}}$	$\mathbf{u}^*_{\mathbf{retins}}$	$\mathbf{u}^*_{\mathbf{bonds}}$	$\mathbf{u}^*_{\mathbf{stocks}}$	$\mathbf{u}^*_{\mathbf{loans}}$
constant	−0.003 (0.048)	1.849** (0.270)	−0.190 (0.135)	−0.048 (0.317)	0.129 (0.272)	−2.208** (0.704)	−0.577* (0.278)	−0.817** (0.280)
incnotret	0.032** (0.004)							
inc2notret	−0.088** (0.012)							
incret	0.063** (0.009)							
inc2ret	−0.228** (0.036)							
divorced	−0.020 (0.013)							
riskgamkid		−0.062** (0.012)						
advmoth		0.009 (0.008)						
advfath		−0.032** (0.009)						
pockmonsp			−0.008** (0.002)					
planmoth			0.006* (0.002)					
planfath			0.009** (0.003)					
female		0.317** (0.083)	−0.111** (0.028)					
married	0.005 (0.016)	−0.042 (0.129)	0.117** (0.045)	0.259* (0.126)	0.161 (0.105)	−0.249 (0.193)	0.012 (0.108)	0.259* (0.117)
educhihh	0.027* (0.012)			0.186 (0.121)	−0.115 (0.100)	−0.098 (0.135)	0.085 (0.086)	0.092 (0.111)
educlow		0.513** (0.163)	−0.268** (0.049)					
educhi		−0.220 (0.114)	0.122** (0.042)					
age1830	−0.027 (0.022)	−0.177 (0.151)	0.031 (0.055)	−0.057 (0.142)	−0.493** (0.119)	−0.296 (0.270)	−0.327* (0.141)	−0.112 (0.118)
age4660	0.040** (0.014)	0.398** (0.115)	−0.061 (0.044)	−0.000 (0.123)	−0.111 (0.106)	0.049 (0.176)	−0.211 (0.108)	−0.159 (0.100)
age61plus	0.025 (0.017)	0.600** (0.130)	−0.064 (0.048)	0.230 (0.154)	−0.655** (0.139)	0.058 (0.211)	−0.193 (0.129)	−0.622** (0.141)
kids	0.000 (0.015)	0.259 (0.136)	−0.134** (0.044)	−0.152 (0.133)	0.174 (0.113)	0.054 (0.185)	0.076 (0.110)	0.128 (0.127)
kidsinhh				−0.116 (0.110)	0.005 (0.086)	−0.348* (0.150)	−0.056 (0.086)	0.121 (0.092)
retiredhh	−0.043 (0.025)	−0.153* (0.075)		−0.227 (0.144)	−0.506** (0.115)	0.049 (0.177)	−0.082 (0.114)	−0.346* (0.137)
unempnowhh	−0.017 (0.020)	0.065 (0.146)	−0.115* (0.052)	−0.295* (0.134)	−0.178 (0.114)	0.045 (0.253)	−0.097 (0.129)	−0.128 (0.113)
selfemphh	0.098** (0.015)	0.234 (0.189)	−0.137* (0.065)	−0.416* (0.170)	−0.200 (0.144)	−0.005 (0.244)	0.160 (0.143)	−0.265 (0.148)
tempjobhh	−0.011 (0.024)	−0.232 (0.173)	0.034 (0.062)	−0.228 (0.152)	−0.001 (0.133)	0.222 (0.240)	−0.043 (0.138)	0.049 (0.141)
east	−0.086** (0.012)	−0.302 (0.162)	0.161** (0.046)	−0.122 (0.141)	0.129 (0.121)	−0.078 (0.209)	0.033 (0.127)	0.282* (0.129)
german	0.058 (0.031)	0.187 (0.230)	−0.036 (0.080)	0.140 (0.223)	−0.025 (0.165)	0.606 (0.575)	−0.301 (0.183)	0.001 (0.196)
incincprob				0.000 (0.001)	0.003** (0.001)	0.003* (0.001)	0.003** (0.001)	−0.000 (0.001)
unempprob				−0.000 (0.001)	−0.000 (0.001)	−0.001 (0.001)	0.000 (0.001)	0.000 (0.001)
creditconstr				−0.426** (0.113)	−0.223* (0.097)	−0.384 (0.320)	−0.108 (0.108)	0.277** (0.092)

Note: Estimated elements of the matrix B defined in Section 6.3. Empty elements are restricted to zero. Standard errors are shown in parentheses, asterisks indicate significance at the 5% and 1%-level. *Source:* SAVE 2005, own calculations.

BIBLIOGRAPHY

Ameriks, J., A. Caplin, and J. Leahy (2003): "Wealth Accumulation and the Propensity to Plan," *Quarterly Journal of Economics*, 118(3), 1007–1047.

Arrondel, L., and A. Masson (2005): "Individual Preferences and the Distribution of Wealth," Mimeo, DELTA, Paris.

Barsky, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro (1997): "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," *Quarterly Journal of Economics*, 112(2), 537–579.

Becker, G. S., and C. B. Mulligan (1997): "The Endogenous Determination of Time Preference," *Quarterly Journal of Economics*, 112, 537–579.

Ben-Akiva, M., D. McFadden, T. Gärling, D. Gopinath, J. Walker, D. Bolduc, A. Börsch-Supan, P. Delquié, O. Larichev, T. Morikawa, A. Polydoropoulou, and V. Rao (1999): "Extended Framework for Modeling Choice Behavior," *Marketing Letters*, 10(3), 187–203.

Ben-Akiva, M., D. McFadden, K. Train, J. Walker, C. Bhat, M. Bierlaire, D. Bolduc, A. Börsch-Supan, D. Brownstone, D. Bunch, A. Daly, A. de Palma, D. Gopinath, A. Karlstrom, and M. A. Minzaga (2002): "Hybrid Choice Models: Progress and Challenges," *Marketing Letters*, 13(3), 163–175.

Bertaut, C. (1998): "Stockholding Behavior of US Households: Evidence from the 1983-1989 Survey of Consumer Finances," *The Review of Economics and Statistics*, 80(2), 263–275.

Börsch-Supan, A., and V. Hajivassiliou (1993): “Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models,” *Journal of Econometrics*, 58(3), 347–368.

Börsch-Supan, A., D. McFadden, and R. Schnabel (1994): “Living Arrangements: Health and Wealth Effects,” in *Advances in the Economics of Aging*, ed. by D. A. Wise, pp. 193–216, Chicago, IL. University of Chicago Press.

Bound, J., C. Brown, and N. Mathiowetz (2001): “Measurement Error in Survey Data,” in *Handbook of Econometrics, Vol. 5*, ed. by J. Heckman, and E. Leamer, vol. 5, pp. 3705–3843, Amsterdam. Elsevier North Holland.

Brown, S., L. Farrell, M. N. Harris, and J. Sessions (2006): “Risk preference and employment contract type,” *Journal of the Royal Statistical Society, Series A*, 169, 849–863.

Campbell, J. (2006): “Household Finance,” *Journal of Finance*, 61, 1553–1604.

Carneiro, P., K. Hansen, and J. Heckman (2003): “Estimating Distributions of Counterfactuals with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on Schooling Choice,” *International Economic Review*, 44(2), 361–422.

Carroll, C. D. (2002): “Portfolios of the Rich,” in *Household Portfolios*, ed. by L. Guiso, M. Haliassos, and T. Jappelli, Cambridge, MA. MIT Press.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005): “Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey,” IZA Discussion Paper 1730, Institute for the Study of Labor (IZA).

Donkers, B., B. Melenberg, and A. van Soest (2001): “Estimating Risk Attitudes Using Lotteries; A Large Sample Approach,” *Journal of Risk and Uncertainty*, 22(2), 165–195.

Frederick, S., G. Loewenstein, and T. O’Donoghue (2002): “Time Discounting and Time Preference: A Critical Review,” *Journal of Economic Literature*, 40(2), 351–401.

GOLLIER, C. (2001): *The Economics of Risk and Time*. MIT Press, Cambridge, MA.

GOMES, F., AND A. MICHAELIDES (2005): "Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence," *The Journal of Finance*, 60(2), 869–904.

GREEN, D., K. JACOWITZ, D. KAHNEMAN, AND D. MCFADDEN (1998): "Referendum contingent valuation, anchoring, and willingness to pay for public goods," *Resource and Energy Economics*, 20(2), 85–116.

GUISO, L., AND M. PAIELLA (2004): "The Role of Risk Aversion in Predicting Individual Behaviours," CEPR Discussion Paper 4591, Centre for Economic Policy Research, London.

HAJIVASSILIOU, V., D. MCFADDEN, AND P. RUUD (1996): "Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results," *Journal of Econometrics*, 72(1-2), 85–134.

HALEK, M., AND J. EISENHAUER (2001): "Demography of Risk Aversion," *The Journal of Risk and Insurance*, 68(1), 1–24.

HALIASSOS, M., AND C. BERTAUT (1995): "Why do so Few Hold Stocks?," *The Economic Journal*, 105(432), 1110–1129.

HALIASSOS, M., AND A. MICHAELIDES (2003): "Portfolio Choice and Liquidity Constraints," *International Economic Review*, 44(1), 143–177.

HARRIS, K., AND M. KEANE (1999): "A Model of Health Plan Choice: Inferring Preferences and Perceptions from a Combination of Revealed Preference and Attitudinal Data," *Journal of Econometrics*, 89(1-2), 131–157.

HARRISON, G. W., M. I. LAU, AND M. B. WILLIAMS (2002): "Estimating Discount Rates in Denmark: A Field Experiment," *American Economic Review*, 92, 1606–1617.

HARRISON, G. W., AND J. A. LIST (2004): "Field Experiments," *Journal of Economic Literature*, 42(4), 1009–1055.

Heaton, J., and D. Lucas (2000): "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk," *The Journal of Finance*, 55(3), 1163–1198.

Holt, C. A., and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92, 1644–1655.

Jöreskog, K. (1969): "A general approach to confirmatory maximum likelihood factor analysis," *Psychometrika*, 34(2), 183–202.

Jöreskög, K. G. (1970): "A general method for analysis of covariance structures," *Biometrika*, 57(2), 239.

Jöreskog, K. G. (1977): "Structural Equation Models in the Social Sciences: Specification, Estimation and Testing," in *Applications of Statistics*, ed. by P. Krishnaiah, pp. 265–287. North Holland, Amsterdam.

Juster, F., and J. Smith (1997): "Improving the Quality of Economic Data: Lessons from the HRS and AHEAD.," *Journal of the American Statistical Association*, 92(440).

Juster, F., J. Smith, and F. Stafford (1999): "The Measurement and Structure of Household Wealth," *Labour Economics*, 6(2), 253–275.

Kapteyn, A., and F. Teppa (2003): "Hypothetical Intertemporal Consumption Choices," *Economic Journal*, 113, C140–C150.

Koopmans, T., H. Rubin, and R. Leipnik (1950): "Measuring the equation systems of dynamic economics," *Statistical inference in dynamic economic models, Cowles Commission Monograph*, 10, 53–237.

Nelson, F., and L. Olson (1978): "Specification and Estimation of a Simultaneous-Equation Model with Limited Dependent Variables," *International Economic Review*, 19(3), 695–709.

Rabin, M. (2000): "Risk Aversion and Expected Utility Theory: A Calibration Theorem," *Econometrica*, 68(5), 1281–1292.

Schmidt, P. (1981): “Constraints on the Parameters in Simultaneous Tobit and Probit Models.,” in *Structural Analysis of Discrete Data and Econometric Applications*, ed. by C. F. Manski, and D. L. McFadden, pp. 422–434. MIT Press, Cambridge, Cambridge, MA.

Schunk, D. (2007a): “The German SAVE Survey 2001-2006. Documentation and Methodology,” Mannheim Research Institute for the Economics of Aging, MEA-Discussion Paper 109-06.

——— (2007b): “A Markov Chain Monte Carlo Algorithm for Multiple Imputation in Large Surveys,” *Advances in Statistical Analysis*, forthcoming.

Viceira, L. (2001): “Optimal Portfolio Choice for Long-Horizon Investors with Nontradable Labor Income,” *The Journal of Finance*, 56(2), 433–470.

Wansbeek, T., and E. Meijer (2000): *Measurement error and latent variables in econometrics*. Elsevier North Holland, Amsterdam.

EHRENWÖRTLICHE ERKLÄRUNG

Hiermit erkläre ich ehrenwörtlich, daß ich diese Dissertationsschrift selbständig angefertigt habe und mich anderer als der in ihr angegebenen Hilfsmittel nicht bedient habe. Entlehnungen aus anderen Schriften sind ausdrücklich als solche gekennzeichnet und mit Quellenangaben versehen.

Mannheim, den 27. Februar 2007

Hans-Martin von Gaudecker

CURRICULUM VITAE

Oct 2003 – Jun 2007	Universität Mannheim Ph.D. Studies in Economics
Sep 2004 – Aug 2005	University College London Ph.D. Studies in Economics (visiting)
Oct 2001 – Sep 2003	Universität Mannheim Graduate Studies in Economics, Diplom
Aug 2000 – Jul 2001	Pontificia Universidad Católica de Chile Graduate Studies in Economics (visiting)
Oct 1998 – Feb 2000	Ruprecht-Karls-Universität Heidelberg Undergraduate Studies in Economics, Vordiplom
Jun 1997	Leibnizschule Hannover Abitur
Aug 1978	Born in Hannover, German nationality