Four essays in unemployment, wage dynamics and subjective expectations

by

Péter Hudomiet

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Economics) in the University of Michigan 2015

Doctoral Committee:

Professor Robert J. Willis, Chair
Professor John Bound
Professor Matthew D. Shapiro
Professor Jeffrey M. Smith
Professor Yu Xie

ACKNOWLEDGMENT

Over the past seven years I have received support and encouragement from a great number of individuals. Robert J. Willis and Gábor Kézdi have been mentors, colleagues, and friends. Our marathon discussions have been tremendously fruitful to identify intriguing research questions and bridge theory and practice. I am grateful to the rest of my committee members, Matthew Shapiro, John Bound, Jeffrey Smith and Yu Xie for all of the time and effort through this process. Their impact as motivators and resources has been immeasurable. I am indebted to Hedvig Horváth, Cynthia Doniger, Pawel Krolikowski, Lasse Brune, Breno Braga, Reid Dorsey-Palmateer and many other friends in Michigan and in Hungary for the stimulating discussions, for the feedback, and for being sources of amusement. I am also grateful for financial support from the ISR-George Katona Research Award Fund in Economic Behavior and the Rackham International Student Fellowship. Lastly, I am exceedingly thankful to my parents for supporting me unequivocally and unconditionally in every aspect of my life, and to my brother and sister and their families (especially my baby niece, Anna) for always being there for me when I needed them. Köszi!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Four essays in unemployment, wage dynamics and subjective expectations

by

Péter Hudomiet

Chair: Robert J. Willis

This dissertation contains four essays on unemployment differences between skill groups, on the effect of non-employment on wages and measurement error, and on subjective expectations of Americans about mortality and the stock market.

Chapter 1 tests how much of the unemployment rate differences between education groups can be explained by occupational differences in labor adjustment costs. The educational gap in unemployment is substantial. Recent empirical studies found that the largest component of labor adjustment costs are adaptation costs: newly hired workers need a few month get up to speed and reach full productivity. The chapter evaluates the effect of adaptation costs on unemployment using a calibrated search and matching model.

Chapter 2 tests how short periods of non-employment affect survey reports of annual earnings. Non-employment has strong and non-standard effects on response error in earnings.

Persons tend to report the permanent component of their earnings accurately, but transitory shocks are underreported. Transitory shocks due to career interruptions are very large, taking up several month of lost earnings, on average, and people only report 60-85% percent of these earnings losses. The resulting measurement error is non-standard: it has a positive mean, it is right-skewed, and the bias correlates with predictors of turnover.

Chapter 3 proposes and tests a model, the modal response hypothesis, to explain patterns in mortality expectations of Americans. The model is a mathematical expression of the idea that survey responses of 0%, 50%, or 100% to probability questions indicate a high level of uncertainty about the relevant probability. The chapter shows that subjective survival expectations in 2002 line up very well with realized mortality of the HRS respondents between 2002 and 2010 and our model performs better than typically used models in the literature of subjective probabilities.

Chapter 4 analyzes the impact of the stock market crash of 2008 on households' expectations about the returns on the stock market index: the population average of expectations, the average uncertainty, and the cross-sectional heterogeneity in expectations from March 2008 to February 2009.

CHAPTER I

The Role of Occupation-specific Adaptation Costs in Explaining the Educational Gap in Unemployment

1.1 Introduction

Unemployment rates vary in the population both in the observed cross section and over time.[1] Unskilled workers have higher levels of unemployment, and they are more exposed to aggregate labor market conditions. For example, between 2005 and 2008, the average unemployment rate was 3.8 percent among 25 – 55 year old males in the US. This average, however, masked a large cross-sectional heterogeneity. The unemployment rate was around 2 percent among college graduates, 5 percent among high school graduates, and more than 7 percent among high school dropouts. In the great recession, the unemployment rates approximately doubled in all groups.[2] Perhaps less well known is the fact that the duration of unemployment, however, is surprisingly similar across workers with different levels of

[1] See, for example, Clark and Summers (1981), Kydland (1984), Keane and Prasad (1993), Hoynes (1999), Jaimovich and Siu (2009) and Hoynes et al. (2012).

[2] These numbers are my own estimates from the Monthly Current Population Survey. See Section 1.2 for details.

schooling, which was first shown by Mincer (1991) using the PSID. This symmetry in the unemployment to employment hazards has been confirmed in more recent studies (Elsby et al., 2010; Dickens and Triest, 2012; Foote and Ryan, 2014; Hudomiet, 2014; Cairo and Cajner, 2014) using larger samples.

This paper quantifies the role of one key economic mechanism that predicts the approximate symmetry in hiring rates and large differences in layoff and unemployment rates: adjustment costs due to the relatively low productivity of newly hired workers who need time to adapt to their new workplaces. I focus on this explanation because recent studies have found these costs to be relatively large, and to correlate with the skill level of jobs. Using a Swiss employer survey, Blatter et al. (2012) report that adaptation takes approximately 80 workdays, on average, during which workers are about 30 percent less productive than experienced workers. They find that these adaptation costs are $2-3$ times larger, on average, than search costs, typically used in macro-labor models. Moreover, large differences in adaptation costs by major occupational groups were also noted. They found that it takes considerably longer to become fully productive in occupations that are considered "higher skilled."

By combining data about adaptation costs and occupational characteristics from the O*NET project, I show that adaptation costs have strong positive correlations with three occupational work activities: 1) interpreting the meaning of information for others; 2) organizing, planning, and prioritizing work; and 3) developing objectives and strategies. What is common in these work activities is that they all have many firm-specific elements. It makes sense that adapting to a new workplace is harder when workers need to acquire many firm-specific skills.

Another motivation for analyzing occupational adaptation costs is that, as I shall show, occupations, in a regression sense, explain at least two-thirds of the unemployment gap between workers with different levels of education.[3] The occupation of workers is a primary determinant of employment chances, and mechanisms that operate through occupations are reasonable candidates for explaining the educational unemployment gap.

The primary question of this paper is what fraction of the educational gap in unemployment can be explained by adaptation cost differences. Given that credible exogenous variation in adaptation costs is hard to find, and simple occupational averages are likely endogenous in a simple unemployment regression, I use a calibrated search and matching model, instead. In the model all parameters except adaptation costs are kept constant across workers. For the calibration I use empirical estimates of occupation- and education-specific adaptation costs from a US employer survey, the Multi-City Study of Urban Inequality.

In the baseline specification of the model, adaptation costs explain 68 percent of the unemployment gap between college and high school graduates.[4] Adaptation costs also account for a larger fraction of the unemployment gap between the medium and highly educated. They explain 65 percent of the unemployment gap between college graduates and college dropouts; 70 percent of the gap between college dropouts and high school graduates; and only 11 percent of the gap between high school dropouts and high school graduates. The model predicts robustly small differences in hiring rates across occupations/education groups that are evident in the data. The model also predicts an 18 percent wage differential between newly hired and experienced workers, which is in between estimated causal effects

[3] Estimates from the Current Population Survey. See Section 1.2.3 for details.

[4] In an earlier version of this paper that circulated on the web my preferred specification explained considerably less (36%) of this gap. That version only analyzed adaptation cost differences between occupations, while this version looks at costs by occupations and education jointly. It turns out that adaptations costs vary by education even within occupations: higher educated workers take more time to adopt even within detailed census occupations. I interpret this as evidence that the census occupations are not detailed enough to capture the full distribution of adaptation costs. An alternative explanation could be that the high educated learn firm-specific skills slower, but that seems implausible.

of tenure on wages reported in Topel (1991) and Altonji and Williams (2005). Moreover, this literature has noted that seniority appears to affect wages at the beginning of the employment spells more strongly, which is exactly what a model with firm-specific adaptation costs predicts. My model also predicts higher separation rates among recently hired workers, although the predicted differences fall short of empirical estimates. The usual interpretation of the separation-tenure profile is that match quality is higher among high-tenured workers (Jovanovic, 1979). My model suggests another contributing factor. Newly hired workers are less productive on average and idiosyncratic productivity shocks are more likely to make the match surplus negative.

Even though there are numerous theoretical arguments in the economic literature for why unemployment rates differ by education, we know very little about the relative importance of the factors highlighted by the theories. Understanding the mechanism is important for designing appropriate, welfare maximizing policies. Should the government run large scale training programs to provide employable skills to the unemployed? Should policymakers make labor market institutions (minimum wages, UI benefits) education-specific in order to better balance the costs and benefits of these policies? Should the government increase (or decrease) employment protection in certain jobs where adjustment costs are too low (or too high)? To answer these questions, we need to understand first what causes the large cross-sectional differences in unemployment rates. My paper endeavors to achieve this goal. There might be unemployment rate differences because: 1) Labor adjustment costs differ according to the skill level of the jobs[5]; 2) Productivity fluctuate more in low skilled jobs; 3) Low skilled wages are above market-clearing levels (e.g., due to the minimum wage[6]);

[5]This idea was suggested first by Oi (1962). Empirical papers in the last 30 years confirmed that recruitment costs are substantial and they are likely to be larger in skilled jobs. (Manning, 2011; Lerman et al., 2004; Leuven, 2005; Blatter et al., 2012).

[6]See Neumark and Wascher (2006) for a review.

4) The value of being unemployed is comparatively high for low skilled workers (e.g., the concave UI benefit formula[7] or skill mismatch[8]); 5) Some selection mechanisms favor high skilled workers compared to low skilled ones (e.g. low skilled employees might work in volatile industries). The results of the literature on these subjects are controversial.[9] For the majority of the paper I focus only on adaptation cost differences across occupations and all other mechanisms are shut off. I come back to the discussion of other mechanisms in the final sections. Economic mechanisms that predict large effects on layoff rates and small effects on hiring rates should be preferred, given the strong symmetry in hiring rates across education and occupation groups.

The elasticity of hiring rates with respect to adaptation costs is robustly small. This quasi-symmetry is remarkable given that most other mechanisms, including other adjustment costs, have large effects on hiring as discussed in Section 1.5. Adaptation costs have two opposing effects on the hiring rate that almost cancel out in equilibrium. The direct *price effect* is that longer adaptation increases the cost of hiring, which decreases the value of matches, leading to fewer jobs and longer unemployment duration. The indirect *labor hoarding effect* is that adaptation costs decrease layoff rates in experienced matches, leading to longer lasting employment spells, which then feeds back positively into job creation. In equilibrium, the direct effect is slightly larger than the indirect effect leading to similar, but moderately lower hiring rates in occupations with longer adaptation.

Differences in adaptation costs predict robustly small differences in hiring rates by ed-

[7]See Meyer (1990), Roed et al. (2002), Wolff and Launov (2004) and Bolvig et al. (2007)

[8]See Lazear and Spletzer (2012) and Sahin et al. (2012).

[9]Another approach used in the literature is to use cross-country variation in labor market institutions to identify their employment effects on workers with different levels of skill. Nickell (1997), Siebert (1997), Iversen and Wren (1998), Esping-Andersen (2000) and Saint-Paul (2004) argue that labor market rigidities in many European countries led to a dual labor market featuring inefficiently high unskilled unemployment. Card et al. (1996) and Oesch (2010), however, using more comprehensive analysis, found little evidence for this labor market rigidity hypothesis.

ucation and occupations. The role of adaptation costs in explaining differences in layoff rates, and consequently, unemployment rates, however, is sensitive to the choice of certain parameters of the model. The paper, thus, carries out a detailed sensitivity analysis. The persistence of match-specific productivity plays a very important role in determining the importance of adaptation, because persistence determines the extent of labor hoarding in the model. When shocks are permanent, adjustment costs matter less for layoff decisions; firms only care about the instantaneous values of matches, and care less about future adjustment costs. When negative shocks are relatively short lived, however, firms are willing to maintain unproductive matches to save on the cost of adjusting labor. Thus, when persistence is low, adaptation costs are more predictive of turnover and unemployment. In my baseline specification the monthly autocorrelation in match-specific productivity is 0.92, based on empirical estimates of autocorrelation in plant level productivity reported in Abraham and White (2006). This value implies that, on average, negative productivity shocks disappear in 12 months. There are reasons to believe that plant level autocorrelation is not a perfect substitute for the autocorrelation in match-specific productivity. I discuss alternative ways of proxying autocorrelation in productivity, but I conclude that it is hard to precisely pin down its value. As Pissarides (2009) also observed, we know very little about the properties of idiosyncratic productivity shocks.[10] I show that by choosing other reasonable values for the autocorrelation, the explained unemployment gap between college and high school graduates can be as low as 35 percent or as high as 122 percent.

In this paper I interpret the cost of adaptation as a special type of labor adjustment cost à la Oi (1962). One can also think of adaptation as a measure of firm-specific skills, since adaptation increases the productivity of the workers only at a particular firm. The idea

[10] Page 1345 in Pissarides (2009).

that firm-specific skills matter for equilibrium turnover and unemployment rates goes back to the early work of Becker (1962), but little empirical work has been done to evaluate the explanatory power of firm-specific skills due to data limitations. One exception is Cairo and Cajner (2014) who developed their paper independently. They use a very similar model, and a measure of on-the-job training from the Employment Opportunity Pilot Project (EOPP) that is similar to my adaptation cost measures from the MCSUI. They find that the entire unemployment gap across education groups can be explained by differences in on-the-job training, together with the volatility of unemployment over the business cycle. The main difference between their approach and mine is that they use a much lower autocorrelation in productivity and they only consider educational differences in on-the-job training, while I discuss occupational differences, too.

The paper is organized as follows. Section 2 introduces basic patterns in unemployment and turnover by education and occupations. The model with occupation-specific adaptation costs is derived and calibrated in Section 3. Section 4 carries out the imputation of explained unemployment and turnover rates by education. Section 6 discusses other economic mechanisms that might explain the rest of the unemployment rate differences across education groups, and capitulates the findings.

1.2 Data and descriptive analysis

1.2.1 The Current Population Survey

The primary dataset used in this paper is the 1978-2013 waves of the Current Population Survey (CPS). The main advantage of the CPS is its large size, roughly $60,000$ households in each month, which makes it possible to include detailed occupations in empirical models of unemployment and turnover. The CPS uses a rotating panel survey design which allows researchers to link around three-fourths of the sample between consecutive months. These so-called semi-panels (or short panels) have been used by many researchers in the past to analyze patterns in worker turnover (see e.g. Blanchard and Diamond, 1990 and Shimer, 2012a). The CPS represents the non-institutionalized adult US population cross-sectionally.

The second advantage of the CPS is that it asks detailed questions about the last jobs of the unemployed. There is information about the last occupations and last industries of the unemployed, which allows estimating, for example, unemployment to employment transition probabilities by the last occupation of the unemployed.

I restrict the sample to prime age, 25 to 55 year old, males. The age restriction is applied to minimize the effect of transitions between employment, schooling and retirement, which might systematically differ by skills. Females are excluded from the analysis, because they move considerably more in and out of the labor force and understanding patterns in the female unemployment rate needs a careful analysis of those movements.

The CPS uses the 3-digit census occupation classifications. In the most recent years it covered more than 500 different occupations. There have been two major changes in the classifications in 1983 and 2003 and two minor ones in 1991 and 2011. Available occupational crosswalks are not suitable for the purposes of this paper. The widely used crosswalks created by IPUMS,[11] for example, are not entirely consistent over time as some occupations are only available in certain years. Moreover, some of the IPUMS occupations contain very few workers, and particularly few unemployed persons, which makes the occupation-specific unemployment rate and transition estimates noisy. To address these issues I created more aggregated, but consistent occupational crosswalks over time. I use three levels of

[11] https://usa.ipums.org/usa/volii/occ_ind.shtml

aggregation: one with 10, one with 48 and one with 191 occupations.[12] The main purpose was to follow the IPUMS crosswalk as much as possible, while ensuring at least around 100 workers in each occupation and month in the sample. The secondary purpose was to have a good match between the 1980 census definitions (used between 1983 and 2002) and the 2000 census definitions (used after 2002), as the pre-1982 definitions are significantly less detailed for skilled occupations. Consequently, the quality of the proposed crosswalk in the most detailed categorization is slightly lower before 1982.

1.2.2 Basic patterns in unemployment and turnover by education

Panels A and B of Figure 1.1 show the level and the logit transformation of the unemployment rates by workers' education: (1) high school dropouts (HSD); (2) high school graduates (HS); (3) some college (SC) and (4) college graduates (BA) among $25-55$ year old males. The unemployment rates are computed from the representative, cross-sectional CPS. There is large heterogeneity in unemployment rates both in the cross-section and over time. Less educated workers are more likely to be jobless, and the increase in their unemployment rates in recessions is larger. The unemployment rate fluctuates between 6 and 16 percent among HSD workers; between 4 and 12 percent among HS workers; between 3 and 9 percent among SC workers; and between 1.5 and 4.5 percent among BA workers. Interestingly, the percentage changes over the business cycle appear to be very similar across education groups. This is even more apparent on Panel B, which shows fluctuations in the logit of the unemployment rates. The gap between education groups is roughly constant in this logit metric.

Panels C-F of Figure 1.1 show two turnover probabilities that are the most important determinants of the unemployment rate. The first is the employment to unemployment

[12] The crosswalk can be downloaded from https://sites.google.com/site/phudomiet/research.

transition probability, which is defined as the fraction of workers who are unemployed in month $t+1$ among those who were employed in month t,

$$p_{gt}^{EU} = \Pr\left(U_{ig,t+1} | E_{igt}\right). \tag{1.1}$$

$E_{ig\tau}$ and $U_{ig\tau}$ indicate that individual i in skill group g at month τ is employed or unemployed. The second probability considered is the unemployment to employment transition probability, which is defined as the fraction of workers who are employed in month $t+1$ among those who were unemployed in month t ,

$$p_{gt}^{UE} = \Pr\left(E_{ig,t+1} | U_{igt}\right). \tag{1.2}$$

These two transition probabilities are not adjusted for time aggregation bias (Shimer, 2012a). When I calibrate the model, however, I will take time aggregation bias seriously.

Panels C and E show the levels, while D and F show the logarithm of the probabilities. The log transformation is motivated by the following steady state approximation of the unemployment rate in group g at time t:

$$u_{gt} = \frac{p_{gt}^{EU}}{p_{gt}^{EU} + p_{gt}^{UE}}, \tag{1.3}$$

$$logit\left(u_{gt}\right) \equiv \ln u_{gt}^{odds} = \ln p_{gt}^{EU} - \ln p_{gt}^{UE}. \tag{1.4}$$

where u_{gt} is the level, and $u_{gt}^{odds} = u_{gt}/(1-u_{gt})$ is the odds ratio of the unemployment rate. The two approximations are equivalent and they are based on the following assumptions: (1) movements in and out of the labor force do not matter for the unemployment rate; (2) the system is in steady state; and (3) one can use the turnover estimates from the linked

CPS to approximate the unemployment rate estimates from the representative cross-sectional sample. Panels A and B of Figure 1.1 show that these simple approximations work very well. The imputed and the true unemployment rates are practically the same.[13]

The advantage of specification (1.4) is its linear representation which allows a simple decomposition of the unemployment rate differences between group A and B into contributions of the job loss and job finding probabilities:

$$\ln u_{At}^{odds} - \ln u_{Bt}^{odds} = \left(\ln p_{At}^{EU} - \ln p_{Bt}^{EU}\right) - \left(\ln p_{At}^{UE} - \ln p_{Bt}^{UE}\right). \tag{1.5}$$

Percentage differences in turnover probabilities translate one-to-one into percentage differences in the odds ratio of unemployment. By looking at the transformed differences across groups, we learn whether the unemployment gap is due to differences in hiring or layoffs.

The results of Panels C-F show that the entire educational gap in unemployment is due to differences in layoffs rather than hiring. Highly educated workers face considerably smaller chances of job loss, but conditional on being unemployed they face very similar chances of finding a job compared to less educated workers. In fact, the job finding probability appears to be slightly higher among less educated workers, particularly after 2000. This is true even though the job finding probability appears to be strongly cyclical. Recent literature argued that the job finding probability is the primary source of the business cycle variation in the aggregate unemployment rate (Davis et al., 2006; Elsby et al., 2009; Fujita and Ramey, 2009; Shimer, 2012a). My results indicate that, instead, the job loss probability is the only important determinant of the educational gap in unemployment. A model that aims at explaining cross-skill differences in the unemployment rate should have this property.

[13] This approximation turns out to work quite poorly for females, which I take as evidence that the female unemployment rate heavily depends on movements in and out of the labor force. In this paper, however, I only focus on the unemployment rate among prime age males.

1.2.3 Basic patterns in unemployment and turnover by education and occupation

This paper argues that occupational adaptation costs are important determinants of the educational gap in unemployment rates and turnover. A necessary condition for this argument to hold is that occupations are strong predictors of unemployment and turnover. To test this condition I ran the following regressions:

$$\ln(y_i) = \alpha_0^y + \alpha_1^y HSD_i + \alpha_2^y SC_i + \alpha_3^y BA_i + \sum_x \beta_x^y I_{xi}^{occ} + z_i' \gamma^y, \tag{1.6}$$

where y indicates unemployment or a turnover probability, $y_i \in \{u_i^{odds}, p_i^{EU}, p_i^{UE}\}$; HSD_i, SC_i and BA_i indicate the highest education of the worker (high school dropout, some college or at least BA) with high school graduates being the reference group; I_x^{occ} indicates the worker's occupation (current if employed or the last occupation if unemployed); and z_i is a vector of other control variables. The interesting question is by how much coefficients α_1^y, α_2^y and α_3^y shrink after I control for occupations. The more they shrink, the more occupations explain, in a regression sense, of the education gap in unemployment/turnover.

Table 1.1 shows the regressions for unemployment. Among the control variables I always include time dummies for each quarter and I sometimes include age, race and marital status dummies. Column 1 shows that unemployment rates are considerably smaller among the educated: the difference is 0.48 log points between HSD and HS workers, 0.31 between HS and SC workers and 0.96 between HS and BA workers. When one controls for basic demographic variables, the educational gap in unemployment shrinks only slightly. When I control for 1 digit occupations using 10 dummies, however, the coefficients shrink substantially. The HS-SC and HS-BA gaps shrink by roughly one half, and the HSD-HS gap shrinks by 20 percent. A large fraction of the educational gap in unemployment can be explained by the one digit occupation of the workers. As I include more and more disaggregated occupation dummies, the differences shrink even further. When I include 191 occupation dummies in columns 5, I can explain around one third of the HSD-HS gap and around two thirds of the educational gap in unemployment among workers with at least a high school degree. By further including sector dummies, the educational gap slightly shrinks further.

Tables 1.2 shows that one gets the same patterns qualitatively using linear probability models. In this table I restrict the sample to low unemployment periods, i.e. “booms”. Table 1.17 in the appendix shows the same OLS regressions in other periods, labeled as “recessions”. Booms indicate years when the aggregate unemployment rate in the 25-55 year old male sample was below 5 percent and recessions indicate that it was above 5 percent. The high unemployment periods using this definition are 1980-1986, 1991-1993 and 2009-2013. These periods are longer and somewhat lag NBER recessions, while the dot-com recession in 2001 is entirely missing as it was too mild. There are large differences in unemployment rates across education groups both in booms and in recessions; one digit occupations explain around half of the gaps among workers with at least a high school degree; and detailed occupations explain around two thirds of it. The explained HS-HSD gap is somewhat less.

Overall, occupation choice appears to be a primary reason why educated workers are less likely to be unemployed. The highly educated work in occupations where unemployment is a rare phenomenon.

Table 1.3 shows logit regressions in the monthly transition probabilities. The patterns in the employment to unemployment transition probabilities are similar to the unemployment regressions. We can see large differences between education groups, and occupations explain a large fraction of the gap between education groups. The coefficients in the job finding regressions (unemployment to employment transitions) reveal a distinct pattern. First, the job finding probability is smaller among higher educated workers. The differences are small, however, at least compared to the differential in job loss probabilities. The 0.047 differential between HS and BA workers, for example, means that BA workers should expect 4.7 percent longer unemployment duration than HS workers, on average, which is less than one week. However, after controlling for occupations, the education coefficients turn positive and they are usually significant. This means that the educated are working in occupations where finding jobs is harder, but conditional on their occupations, education is slightly positively associated with job finding chances. This finding, at least qualitatively, is in line with the adjustment cost mechanism. Appendix Tables 1.18-1.23 shows more detailed logit specifications and linear probability models. Those results are qualitatively the same.

1.2.4 Adaptation costs by occupations

Using a Swiss employer survey, Blatter et al. (2012) estimate that during the adaptation period rookies are approximately 30 percent less productive than experienced workers, and the length of the adaptation period is roughly 80 workdays on average. They also report that the 30 percent productivity gap appears to be constant across jobs and employers, while the length of the adaptation period varies greatly.

I use a US employer survey, the Multi-City Study of Urban Inequalities (MCSUI)[14]. These data were collected in 1992-1994, in order to understand why high rates of joblessness have persisted among minorities living in America's central cities. One important aspect of the study was the contacting of more than 3000 employers in four large US cities (Los Angeles, Boston, Detroit and Atlanta) to ask detailed questions about their hiring practices.

[14] The study was funded by the Russel Sage Foundation and the Ford Foundation. See Holzer (1996) for details about the survey.

Even though the intent of the study was to understand racial discrimination in hiring, the exhaustive information about the recruitment process makes this study valuable for broader purposes. The sampling procedure and the provided weights intend to represent employees who worked in Los Angeles, Boston, Detroit or Atlanta in 1992. They used the 3-digit 1980 SOC occupation classification to code the position of the workers.

The survey contains a measure of occupational adaptation costs that I use in this paper. The question reads "How many weeks or months does it take the typical employee in this position to become fully competent in it?" Unfortunately the survey did not ask questions about the average productivity difference between rookies and experienced workers, so I rely on the Blatter et al. (2012) study and assume it is 30 percent in each job.

Table 1.4 provides descriptive statistics about the adaptation period measure. The median is around 3 months and the mean is around 7 months. The large difference between the mean and the median is due to quite a few large outliers in the sample. One way of dealing with these outliers is to apply the log transformation on the reported adaptation periods to downweight the influence of these outliers. This is the method I applied in this paper.

Table 1.6 shows the coefficients of regressions of log adaptation on various occupational work activity measures. The coefficients are from simple regressions with only one measure included at a time. The work activity measures are created from the O*NET data with a procedure described in Firpo et al. (2011).[15] One might expect that adaptation takes longer when people have to work with others, and when they have to analyze information and make plans, since these activities are mostly firm-specific. The table supports these hypotheses. Work activities related to information processing and dealing with people have

[15] O*NET provides two variables: the "importance" and the "level" of work activities in occupations. Based on Firpo et al. (2011), I transform the two variables to be between 0 and 1, and then I assign a Cobb-Douglas weight of 2/3 to importance and 1/3 to level to create my final measures.

strong positive correlations with adaptation costs, while physical activities and operating machines are negative predictors of adaptation costs. The three strongest predictors (based on the t-statistics) are: 1) interpreting the meaning of information for others; 2) organizing, planning, and prioritizing work; and 3) developing objectives and strategies. Table 1.5 shows regressions of adaptation on education and various occupational work activity measures. The three main occupational activities are positive and significant even jointly and the addition of other work activity measures do not make the fit better. It appears interpreting information for others, planning and developing strategies are the primary determinants of how long it takes to get fully productive in occupations. Table 1.5 also shows that adaptation is significantly and substantially longer in jobs where the highly educated work. Compared to high school graduates, the average adaptation period is 35 percent shorter for high school dropouts, 34 percent longer for college dropouts and 73 percent longer for BA workers. When I include both education and occupational activity measures, the coefficients on education drop by about half, but remain statistically significant. In some ways it is puzzling that education remained significant. MCSUI asks employers about the average adaptation in the given position. If the positions employers have in mind are the same as the occupations I use, no variable other than occupations should matter. It appears that employers use a more detailed occupation-specification when answering these questions, and the education dummies pick up these differences.

Table 1.25 shows the average imputed length of adaptation in 46 detailed occupations: 23 with the shortest and 23 with the longest adaption out of the 191 occupations.[16] The table shows the fitted values of Models 2 and 3 from Table 1.5 and the raw averages by occupations. My preferred measure is in the first column, which is based on only the three

[16]The entire table is available on my website at https://sites.google.com/site/phudomiet/research.

main work activity measures. The addition of the other activity measures hardly changes the imputed values, and the raw averages in column 3 are very noisy, since they are based on very few observations. My preferred measure shows that the distribution of the adaptation period runs from 1.2 month (waiter's assistants) to 8.6 month (clergy and religious workers). This is a very wide distribution. Adaptation is shortest in occupations such as waiters, laundry workers, food preparation workers, janitors, machine operators, and interviewers. Adaptation is longest for managers, management support occupations, teachers, scientists, financial workers, etc. These are in line with expectations.

Table 1.7 shows logit regressions of unemployment and turnover probabilities on imputed adaptation cost measures. The imputed values are the predicted values from a regression of log adaptation costs on the three occupational work activity measures, education dummies (less than high school, high school, some college, and college) and their interactions. Model 1 of Table 1.7 includes no control variables other than year dummies, and model 2 includes education, age, race, marital status and industry dummies. In model 1 adaptation costs are significantly and strongly negatively related to unemployment and layoff rates, while the effect on hiring rates is smaller, but still negative and significant. These are qualitatively in line with the predictions of a labor hoarding model: turnover costs negatively affect both job creation and job destruction. In model 2, after controlling for education, age, race and industries, the elasticities of unemployment and job loss with respect to adaptation decrease by about 20%, but remain very large and significant; and the elasticity of job finding remains negative, but turns insignificant. These estimates are suggestive that adaptations costs are important determinants of turnover and unemployment. They are, however, likely suffer from omitted variable bias. Occupations differ in many dimensions beyond the cost of adaptation. To overcome this endogeneity problem the next section turns to a model of turnover and unemployment. In the model I allow occupations to differ only in the average length of adaptation, and all other differences are shut off.

1.3 The model

The purpose of this model is to analyze how adaptation costs affect turnover and unemployment rates in equilibrium. I use a Mortensen and Pissarides (1994) (MP) type search and matching model with endogenous job creation and job destruction as the modeling framework. In this model, unemployment is purely frictional in the sense that all unemployed persons are searching for jobs. Workers are job-less because it takes time to find new jobs after separations, not because their reservations wages are too high. In fact, in the standard setup of the model, used in this paper, workers do not even have a reservation wage: they accept any job offered to them and wages are determined by a bargaining process in which firms and workers make an agreement about how to split up the surplus of the match. The rate at which separations occur and the duration of unemployment are endogenously determined in the model. Separations occur when idiosyncratic, match-specific shocks make the value of the match (the surplus) negative. The rate at which the unemployed find jobs depends on how many vacant jobs are posted by firms and how many individuals are competing for these positions. In the standard MP model, the probability of finding a job is assumed to be the same for all workers.

The traditional adjustment cost literature (see e.g. Oi, 1962; Fair, 1985; Shapiro, 1986; Lazear, 1990) uses a neoclassical framework to analyze the effect of adjustment costs on unemployment rates. In that framework, job loss and job finding probabilities cannot be analyzed and job-less workers do not want to work. The MP modeling framework is more suitable to analyze the effect of various parameters on turnover rates, since unemployment is frictional and turnover is endogenously determined in the model.

My model closely follows the original Mortensen and Pissarides (1994) model, with a few exceptions. First, as I aim at explaining cross-sectional variation in turnover and unemployment, I only consider the steady state equilibrium of the model. Second, the labor force in my model is heterogeneous: workers can work in one of many occupations. Within occupations, however, workers are homogeneous apart from random shocks to their match-specific productivity. To keep the model simple and tractable I assume that workers are permanently assigned to their occupation and they never change their careers. When they lose their jobs, they only search in their original occupation. This assumption allows for segmenting the economy into submarkets and applying the standard solution methods on each of them. It would be interesting to consider a more general model with the possibility of switching occupations. For at least two reasons, however, the general model should behave similarly to the restricted model that is considered in this paper. First, the size of each occupation should be stable in steady state. Second, the job finding probabilities are very similar across education groups and occupations, and thus, cross-occupation flows are unlikely to contribute much to equilibrium unemployment rates.

The main deviation I introduce to the MP model is the presence of an adaptation period during which newly hired rookies are less productive than experienced workers. Occupations only differ in the expected length of the adaptation period, and all other parameters of the model are kept constant across occupations. I assume that the adaptation period in occupations is exogenous. It reflects how complex the jobs are, and how difficult it is to gain full competency in them. In reality, the adaptation period might be partly endogenous: Employers might choose to assign harder and harder tasks for newly hired workers gradually. There might be a production function of experience and employers might be able to manipulate the speed of gaining experience. In such cases one might think of my simplistic model as the second step of a two step model. In the first step firms optimize over all possible values of adaptation. Once they find the optimal production function of experience in each occupation, they can take the optimal values of adaption from the first step as a given to discover the optimal turnover decisions in the second step. In this case, my model does not reveal why adaptation costs are larger in certain occupations, but it still discerns the equilibrium turnover and unemployment rates given the observed equilibrium values of adaptation.

This is not the first paper to add an adaptation period to a search and matching model. Mortensen and Nagypal (2007) and Silva and Toledo (2009) use similar models to analyze the effect of adaptation costs on the business cycle variation of the aggregate unemployment rate. Some papers discussed the implications of training costs for unemployment rates (Pissarides, 2009; Cheremukhin, 2010; Miyamoto, 2011 and Cheron and Rouland, 2011). These papers model the cost of training as a onetime cost at the time of hiring, and they do not use empirical measures of training costs. The recent working paper of Cairo and Cajner (2014) also uses a similar model to mine.

1.3.1 The setup of the model

In what follows, subscript x refers to occupation $x \in \{1, ..., N_X\}$, and I ignore time subscripts to simplify notation. I use a continuous time version of the model. Firms can employ one worker, either a rookie or an experienced one, or zero workers. When they are matched with an experienced worker in occupation x the match produces flow profit

$$\pi_x(\varepsilon) = s_x \varepsilon, \tag{1.7}$$

where s_x is average productivity in occupation x and ε is the idiosyncratic, match-specific productivity term. The c.d.f of ε is denoted by $F(\varepsilon)$ which is assumed to follow a log-normal distribution,

$$\ln \varepsilon \sim N\left(0, \sigma^2\right). \tag{1.8}$$

The idiosyncratic shock is temporary and its value is redrawn from this distribution in an i.i.d. fashion when a new shock arrives with Poisson arrival rate λ. One can think of λ as the inverse autocorrelation in productivity: when λ is large, new shocks occur more frequently and present productivity has a smaller predictive power for future productivity. Formally, if λ is calibrated to be a monthly rate, then a t month autocorrelation in productivity is $\rho(t) = \exp(-\lambda t)$. If the productivity shock is smaller than an endogenously determined productivity threshold $\underline{\varepsilon}$, so that the surplus falls below zero, matches get destroyed endogenously. The endogenous separation rate, $f_x^{L,en}$ in the model, is

$$f_x^{L,en} = \lambda F\left(\underline{\varepsilon}_x\right). \tag{1.9}$$

Matches can also be destroyed with an additional exogenous rate, $f^{L,ex}$. Hence, the total job loss rate is

$$f_x^{L} = \lambda F\left(\underline{\varepsilon}_x\right) + f^{L,ex}. \tag{1.10}$$

Rookies' productivity is lower than experienced workers' productivity by a factor $1-\delta$,

$$\pi_x^{R}\left(\varepsilon\right) = \delta s_x \varepsilon. \tag{1.11}$$

Rookies endogenously separate from their employers if the match-specific productivity shock falls below an endogenously determined productivity threshold $\underline{\varepsilon}^R$. Furthermore, I assume that the additional exogenous separation rate is the same for rookies and experienced workers. The total job loss rate of rookies is

$$f_x^{L,R} = \lambda F\left(\underline{\varepsilon}_x^R\right) + f^{L,ex}. \tag{1.12}$$

We should expect that, in equilibrium, $\underline{\varepsilon} < \underline{\varepsilon}^R$ so that rookies are more likely to lose their jobs than experienced workers. Rookies become experienced randomly with a Poisson arrival rate φ_x. This arrival rate is assumed to be occupation-specific, but the productivity loss, $1-\delta$, is not. This is in line with the findings of Blatter et al. (2012).

Firms without workers can post a vacancy if they are willing to pay the instantaneous search cost $c_x = cs_x$. The search cost is a constant fraction of occupational productivity, s_x. This is a neutrality condition which assures that occupational productivity does not have a direct effect on turnover. One justification for this assumption is that higher skilled workers have to be interviewed (the largest component of search costs, see Blatter et al., 2012) by higher skilled workers.

If there are v_x posted vacancies (unfilled jobs) in the economy and n_x persons are searching for them, then matches are formed at rate

$$m_x = \eta v_x^{\alpha} n_x^{1-\alpha}. \tag{1.13}$$

This function is called the matching function which determines how fast workers can find jobs and firms can fill their vacancies. The matching function has constant return to scale (CRS) and it is increasing in both arguments because $0 < \alpha < 1$. The job finding rate, f_x^F, can be computed by dividing the match rate by the number of searching workers,

$$f_x^F = \frac{m_x}{n_x} = \eta \left(\frac{v_x}{n_x}\right)^{\alpha} \equiv \eta\theta_x^{\alpha}, \tag{1.14}$$

where $\theta_x = v_x/n_x$ is called labor market tightness. In a tight labor market with a lot of unfilled jobs it is easier to find jobs and f^F is relatively large. The rate at which vacancies are filled, q_x, can be computed similarly:

$$q_x = \frac{m_x}{v_x} = \eta\theta_x^{\alpha-1}. \tag{1.15}$$

The vacancy filling rate is decreasing in labor market tightness, because $\alpha < 1$. In a tight labor market with a lot of unfilled jobs it is harder to find workers and q is relatively small. Free entry of firms assures that the value of posting vacancies is zero in equilibrium. The value of being an unemployed worker, however, is positive. Unemployed workers enjoy a flow value of $b_x = bs_x$ and they expect a positive continuation value as well. Similarly to search costs, the flow value of unemployment is also normalized with occupational productivity, s_x.

Wages are determined through Nash wage bargaining, and workers' bargaining power is γ. At the time of hiring and at any time when the productivity of the match changes, wages get renegotiated. It is possible to allow for wage rigidity without changing any predictions of the model. Wage rigidity does not affect turnover under the following two conditions: 1) Newly hired workers' wages are flexible and set by Nash wage bargaining; 2) Workers and firms never leave surplus on the table. This second condition implies that if, for example, a negative shock occurs and the firm is not interested in the match at the given level of rigid wage, workers are willing to renegotiate the wage to save their jobs. Both of these conditions are reasonable.

Finally, the interest rate is r and the initial value of the idiosyncratic shock is fixed at its mean, $\varepsilon^{initial} = \varepsilon^0 = \exp\left(\frac{\sigma^2}{2}\right)$.

1.3.2 The solution

The model can be solved analytically. The Appendix shows how to write the value functions of the firm and the workers, and then how to solve for all the endogenous parameters.

The equilibrium wages of rookies and experienced workers are

$$w_x^R(\varepsilon) = (1-\gamma)\, b + \gamma\left(\delta\varepsilon + \theta_x c\right), \tag{1.16}$$

$$w_x(\varepsilon) = (1-\gamma)\, b + \gamma\left(\varepsilon + \theta_x c\right). \tag{1.17}$$

Rookies make less money, because they are less productive than experienced workers. Thus, the model predicts a wage-tenure profile. Wages increase in the bargaining power of workers (γ), in productivity (δ and ε), in the flow value of unemployment (b), in search costs (c) and in labor market tightness (θ_x). Apart from search costs, these results are intuitive. When search costs (paid by the firms) are large, the initial surplus needs to be larger so that firms are compensated for them. A larger initial surplus increases the workers' share of the initial surplus, and, consequently, workers demand higher wages. This is a classical hold-up problem[17]. Investments made before the first wage negotiation, in this case paying search costs, have to be shared with the other party. Using similar arguments, one can show that search costs that are paid by workers, as opposed to firms, decrease, rather than increase, wages. The effect of search costs on wages is similar to the wage effects of firm-specific skills described in Becker (1994). In his model, the costs and benefits of acquiring firm-specific

[17] See Acemoglu and Shimer (1999) for a more detailed discussion of the hold-up problem in search models.

skills have to be shared by workers and firms so that neither has an incentive to end the relationship too early.

Now consider equilibrium turnover. The job loss and job finding rates depend on the three endogenous parameters (θ, $\underline{\varepsilon}_x$ and $\underline{\varepsilon}_x^R$) as described in (1.10), (1.12) and (1.14). To characterize the equilibrium we need three conditions. The first one is called the job creation condition:

$$\frac{c}{\eta}\theta_x^{1-\alpha} = (1-\gamma)\frac{\delta + \frac{\varphi_x}{r+\lambda+f^{L,ex}}}{r+\lambda+f^{L,ex}+\varphi_x}\left(\varepsilon^0 - \underline{\varepsilon}_x^R\right). \tag{1.18}$$

The left-hand side of (1.18) is the expected total search costs firms pay to fill a vacancy and the right-hand side is the firms' share of the initial surplus. The job creation condition says that new vacancies are posted until labor market tightness reaches a point at which the expected value of an additional vacancy is zero.

The second equilibrium condition of the model is the job destruction condition of experienced workers:

$$b + \frac{\gamma}{1-\gamma}\theta_x c = \underline{\varepsilon}_x + \frac{\lambda}{r+\lambda+f^{L,ex}}\int_{\underline{\varepsilon}_x}^{\infty}\left(\varepsilon' - \underline{\varepsilon}_x\right) dF\left(\varepsilon'\right). \tag{1.19}$$

The left-hand side of (1.19) is the outside option of workers and the right-hand side is the total value of the match. The job destruction condition says that matches are destroyed when the surplus is below zero, or in other words, when the workers' outside option exceeds the total value of the match. The right-hand side has two terms. The first ($\underline{\varepsilon}_x$) is the flow productivity and the second, involving the integral, is the expected continuation value of the match. One can think of the second term as a labor hoarding term: matches can survive large negative shocks as long as the surplus is expected to increase soon. In fact, the only incentive to hoard labor is that the cost of adjustment is smaller than the expected cost associated with a temporary negative shock. When λ is small, and shocks are very permanent, hoarding labor is not rational, and flow productivity should be the only factor determining turnover decisions. When λ is large, and productivity shocks are short lived, hoarding labor becomes more appealing.

The last equilibrium condition is the job destruction condition of rookies, which is

$$b + \frac{\gamma}{1-\gamma}\theta_x c \quad = \quad \delta \underline{\varepsilon}_x^R + \lambda \frac{\delta + \frac{\varphi_x}{r+\lambda+f^{L,ex}}}{r+\lambda+f^{L,ex}+\varphi_x} \int_{\underline{\varepsilon}_x^R}^{\infty} \left(\varepsilon' - \underline{\varepsilon}_x^R\right) dF\left(\varepsilon'\right) + \frac{\varphi_x}{r+\lambda+f^{L,ex}} \left(\underline{\varepsilon}_x^R - \underline{\varepsilon}_x\right) \tag{1.20}$$

The left-hand side of (1.20) is the outside option of workers, which is the same for rookies and experienced workers. The right-hand side consists of the flow productivity of the match and two labor hoarding terms. The first one, involving the integral, is the expected continuation value of the match if the worker remains a rookie, and the second one is the expected gain if he becomes experienced. Rookies might be hoarded not only because a temporary negative shock is expected to disappear, but also because they might become experienced soon.

Equations (1.18)-(1.20) implicitly determine the three endogenous parameters and, consequently, the job finding and job loss rates. The equilibrium conditions in the standard MP model are similar, and they can be obtained as limiting cases of my model. For example, if rookies are as productive as experienced workers ($\delta \to 1$) or if rookies become experienced immediately ($\varphi_x \to \infty$), the equilibrium conditions collapse to the ones in the standard model. If rookies never become experienced ($\varphi_x \to 0$), the equilibrium conditions collapse to the standard model with average productivity normalized to δ, rather than 1.

1.3.3 Calibration of the model

In this section I calibrate the model to fit aggregate turnover and unemployment. The calibrated parameters might be thought of as estimates for an average or "representative" occupation. The next sections discuss how turnover and unemployment vary with adaptation costs.

The model aims at matching empirical average turnover in the CPS between 1994 and 2008. In 1994, the CPS introduced a question about employer change: workers who are employed both in month t and $t+1$ are asked in $t+1$ whether they still work for the same employer as in t. Years before 1994 are dropped, because employer switching, an important separation margin, cannot be identified in those years. Years after 2008 are dropped because of the great recession. In each year between 1994 and 2008, the aggregate unemployment rate in the prime age male sample was below 5 percent.[18]

Some parameters of the model are calibrated externally, using regular values from the literature. The first eight rows of Table 1.9 shows the choice of these parameters. The Cobb-Douglas share of vacancies in the matching function is 0.4, based on a large number of empirical estimates summarized in Pissarides and Petrongolo (2001). Workers' bargaining power is set to 0.6 so that the Hosios condition is satisfied and the market equilibrium is socially efficient. Aggregate labor market tightness is set to 0.72. This value is based on empirical estimates from the Job Openings and Labor Turnover Survey and used in most macro labor papers. One might argue that in my more restricted sample (prime age males) a different value should be used. It, however, plays no role for the estimates; the model is entirely indifferent to the choice of aggregate θ. For a formal proof, see Appendix 1.5.2.

[18] The calibration results are basically equivalent if the years of the mild dot-com recession are dropped, too.

Intuitively, the other internally calibrated parameters absorb any mismeasurement in θ. The interest rate is set to 0.004 which corresponds to ~ 5 percent annually. The average total search cost firms pay, which is c/q in the model, is set to 50 percent of the average monthly wage of an experienced worker. The average wage is computed endogenously from the model. Blatter et al. (2012) found that total search costs in Switzerland take up roughly 60 percent of the average monthly wage, while other papers (such as Pissarides, 2009; Ratner, 2013) use values around 40 percent. I use the middle point between 40 and 60 but I shall investigate the sensitivity of the model's predictions to the choice of this parameter.

The productivity of rookies is set to 30 percent lower than that of experienced workers based on Blatter et al. (2012). The rate of turning experienced is set to $\varphi = 0.31$, which corresponds to becoming experienced in about three months, on average. The value of φ was computed from the MCSUI with the following procedure. I assume that true adaptation costs only depend on the occupation of workers:

$$\ln t^a_{ix} = \ln t^a_x + m_i. \tag{1.21}$$

The log reported adaptation period in occupation x by employer i, $\ln t^a_{ix}$, is the sum of an occupation-specific term, $\ln t^a_x$, and a mean zero measurement error term, m_i. Equation (1.21) is estimated by OLS using 48 occupation dummies. I use the provided sample weights that intend to make the MCSUI sample represent employees who worked in Los Angeles, Boston, Detroit or Atlanta in 1992. After that I take the sample average of the occupational log adaptation periods using the occupational distribution of workers in the CPS sample. The average rate of turning experienced, then, is the inverse of the exponent of this number, $\varphi = \exp(-\ln t^a) = 1/t^a$. I run a sensitivity analysis on both δ and φ in the next section.

The final parameter that is externally calibrated is the autocorrelation in match-specific productivity. Empirical estimates of this parameter are hard to obtain, because we do not observe productivity in individual matches. In the baseline calibration I, instead, use measures of autocorrelation in plant level productivity, but in Section 1.4.1 I shall discuss alternative strategies. Abraham and White (2006) provides high quality estimates of plant-level productivity in the manufacturing sector. They find yearly autocorrelation to be relatively low, between 0.37 and 0.41. The corresponding monthly values are 0.92 and 0.93. I choose the value of $\rho = 0.92$. This value implies that the rate at which new i.i.d. shocks arrive is $\lambda = -\ln(0.92) \approx 0.083$, and productivity shocks get redrawn every $1/0.083 \approx 12$ months, on average. There are reasons to believe that plant level autocorrelation is not a perfect substitute for the autocorrelation in match-specific productivity. For example, as aforementioned, workers productivity might fluctuate for personal reasons such as sickness, a loss in the family, interpersonal issues with coworkers, etc. One might think that such negative shocks disappear quicker than demand shocks. Given that the plant level estimates in Abraham and White (2006) are already relatively low, I use those numbers in my baseline specification. I run a sensitivity analysis on this parameter as well.

The model has three equilibrium conditions: the job loss condition of rookies, the job loss condition of experienced workers and the job finding condition. The three conditions, together with the job loss and job finding probabilities from the CPS enable calibrating five parameters internally. These are the productivity thresholds of experienced workers ($\underline{\varepsilon}$) and rookies ($\underline{\varepsilon}^R$); the dispersion of productivity shocks (σ); the efficiency of the matching function (η); and the flow value of unemployment (b).

The next issue is the choice of turnover probabilities to match from the CPS. Table 1.8 shows different types of transition probabilities in the CPS. The average $E \rightarrow U$ transition probability is a little over 1 percent, similar to the employment to non-participation ($E \to N$) probability. The fraction of employer switchers is even larger, above 2 percent monthly. Altogether 4.4 percent of matches are destroyed in an average month in the US. It has to be determined which separations to model as endogenous and as exogenous. In the model, endogenous separations occur when the surplus falls below zero after a large negative shock to productivity. In reality, separations can occur even without productivity shocks and even in cases when the match is still productive. For example, when a worker decides to go back to school, he decides to retire, or he accepts a better job offer, arguably he would still be productive in his old match. The total monthly separations (4.4 percent) is large, and it seems unlikely that all of them involved the disappearance of the surplus. For that, one would need implausibly large and frequent productivity shocks. I, instead, assume that $E \to U$ transitions are endogenous separations while all other separations are exogenous. As a robustness check, I shall consider cases where a fraction k^{EN} of $E \to N$ transitions are endogenous, too. $E \to E$ transitions, however, are always assumed to be exogenous.

The next question is how to determine the unemployment rate in the model. The model has predictions for the different types of job loss rates and the labor market tightness, θ. Labor market tightness determines the job finding rate among searching workers, who can come either from unemployment, non-participation or employment. Because unemployed persons are all searching for jobs, the $U \to E$ transition rate equals the job finding rate: $f^{UE} = f^F$. The transition rates from other labor market statuses can be different, because, for example, not all persons in non-participation are searching. I propose to employ a simple shortcut here. In Section 1.2.2 we already saw that the unemployment rate can be very well approximated by the steady state formula, $u = p^{EU} / \left(p^{EU} + p^{UE}\right)$. The two panels of Figure 1.2 show that this approximation works well even in detailed occupations in the CPS. I, thus, assume that this formula works in the model, too. Because the $U \to E$ transition rate equals the job finding rate, the unemployment rate can be written as a function of the labor market tightness and the $E \to U$ transition rate. As described in the previous paragraph, in the baseline specification I assume that the $E \to U$ transition rate equals the endogenous separations in the model. Thus, the unemployment rate can be written as

$$u_x = \frac{\lambda F\left(\underline{\varepsilon}_x\right)}{\lambda F\left(\underline{\varepsilon}_x\right) + \eta\theta_x^{\alpha}}. \tag{1.22}$$

The formula ignores the effect of rookies on the unemployment rate, but given that they take up a tiny fraction of the employed sample, the bias is negligible.

In the more general case, when some of the endogenous separations are $E \to N$ transitions, the unemployment rate is

$$u_x = \frac{k\lambda F\left(\underline{\varepsilon}_x\right)}{k\lambda F\left(\underline{\varepsilon}_x\right) + \eta\theta_x^{\alpha}}, \tag{1.23}$$

where k is assumed to be the same in each occupation and is a function of k^{EN}.

Rows $9-11$ of Table 1.9 show the most interesting internally calibrated parameters and rows $12+$ show a few other interesting implied values for the representative occupation. The flow value of unemployment is around 66 percent of the average wage of experienced workers. There is a large disagreement about the true value of this parameter in the literature. There are papers using values as low as 40 percent (Shimer, 2005) and as high as 95 percent (Hagedorn and Manovskii, 2008) of the average wage. My estimated value is roughly in the middle of the used interval. The standard deviation of match-specific productivity shocks is 0.26, which is reasonable, but is somewhat on the high side of calibrations in other papers. The scalar of the matching function, sometimes called matching efficiency, is 0.4.

The model predicts that rookies are more likely to be laid off than experienced workers.

While the continuous employment to unemployment transition rate is only 0.014 for experienced workers, it is 0.020 for rookies. The last two rows of the table show the average wages of rookies and experienced workers using formulas (1.16) and (1.17):

$$E\left(w^R\right) = (1-\gamma)\,b + \gamma\left(\delta E\left(\varepsilon|\varepsilon > \underline{\varepsilon}^R\right) + \theta c\right), \tag{1.24}$$

$$E\left(w\right) = (1-\gamma)\,b + \gamma\left(E\left(\varepsilon|\varepsilon > \underline{\varepsilon}\right) + \theta c\right). \tag{1.25}$$

The model predicts that rookies earn roughly 18 percent less than experienced workers. This estimate is a pure effect of seniority and it should be comparable to causal estimates of tenure on wages. Simple OLS estimates of wage differential between newly hired and experienced workers is around 30 percent, but causal estimates are anywhere between 5 and 25 percent (Topel, 1991; Altonji and Williams, 2005). Taking this into account, my 18 percent is a reasonable number.

1.3.4 Turnover elasticities with respect to adaptation costs

The only occupation-specific exogenous parameter in the model is φ_x, the rate at which rookies turn experienced. To get a sense of the effect of adaptation, I calibrate the elasticities of turnover and unemployment rates with respect to the length of the adaptation period. As φ_x is a Poisson process, the expected length of adaptation is

$$t_x^a = \frac{1}{\varphi_x}. \tag{1.26}$$

Table 1.10 shows the elasticities of various endogenous variables with respect to the adaptation period evaluated at the aggregate solution. I use the following numerical approximations to compute the elasticity of any variable y

$$\varepsilon_{t^a}^{y} = \frac{\partial \ln y}{\partial \ln t^a} \approx \frac{y\left(\bar{t^a}+h\right)-y\left(\bar{t^a}-h\right)}{2h}\frac{\bar{t^a}}{y\left(\bar{t^a}\right)}, \tag{1.27}$$

where $\bar{t^a}$ is the average length of the adaptation period (~ 3 months in the baseline calibration of the model); $y(t)$ is the equilibrium value of the variable of interest when the adaptation period is t, but all other parameters are kept constant at the aggregate value; and h is a small number. In practice I use $h = 10^{-4}$.

The model predicts that longer adaptation leads to both less layoff and less hiring, but the response on the layoff margin is more than ten times larger. The steady state assumption implies that

$$\varepsilon_{t^a}^{u^{odds}} = \varepsilon_{t^a}^{f^{EU}} - \varepsilon_{t^a}^{f^{UE}}. \tag{1.28}$$

As the layoff elasticity is much larger than the hiring elasticity, the model predicts that the equilibrium unemployment rate is decreasing with the length of the adaption period. The estimated elasticity of unemployment is -0.541. To get a sense of the magnitude of this coefficient, let us consider the following back-of-the-envelope calculation. In Section 1.2.2 we saw that the unemployment gap between BA and HS workers in the logit metric was 0.962. In order to explain the entire BA-HS gap, thus, we need $\ln t_{BA}^a - \ln t_{HS}^a = 0.962/0.541 \approx 1.78$, the log difference in the adaptation periods between the two groups should be around 1.78. Alternatively, the ratio t_{BA}^a/t_{HS}^a should be around $\exp(1.78) \approx 5.9$, adaptation in jobs in which college graduates work should be 6 times longer than in jobs in which high school graduates work. This would happen, for example, if the average adaptation costs for high school and college graduates were 4 and 24 weeks, respectively. This differential is not implausible, but it seems somewhat large. In Section 1.2.4 we saw that the log difference in adaptation between the two groups was around 0.73 rather than 1.78. Based on this simple calculation, we can expect to explain around 40 percent of the educational gap in unemployment between BA and HS workers with adaptation costs. This back-of-the envelope calculation, however, might not be precise due to non-linearities in the model.

The implied close to zero hiring elasticity of adaptation costs is remarkable, given that most parameters of the model, including other adjustment costs, have large effects on hiring as shown in Hudomiet (2014). Adaptation costs have two opposing effects on the hiring rate. The direct effect is that longer adaptation makes turnover more costly, decreasing the value of the initial surplus, which lowers vacancy creation and, consequently, makes it harder to find jobs. I call this the *price effect*: when turnover is costly, there is less of it in equilibrium. The indirect effect, which I call *labor hoarding effect*, is that adaptation makes matches more stable through a decrease in layoffs. Longer lasting matches increase the value of the initial surplus, since both the firm and the worker expect to enjoy their share of the match rent for a longer time. Both the price and the labor hoarding effects are large, but they appear to almost cancel out in equilibrium. Mathematically, one can show that any reasonably calibrated job destruction condition is steep, and as adaptation costs only shift the job creation condition, labor market tightness and the job finding probabilities are not affected strongly by adaptation costs. It is worth noting that the symmetry in the hiring rates only holds if the model features endogenous job destruction. If job destruction is exogenous, that is, the labor hoarding effect is shut down, adaptation costs do not have an indirect positive effect on hiring, and there is nothing to balance out the large negative price effect of adaptation.

Adaptation costs are also different from other adjustment costs. Beyond the already mentioned price and labor hoarding effects, most adjustment costs affect turnover through a third channel as well, which I call the *bargaining effect*. Adjustment costs can change the bargaining position of workers or firms depending on who pays these costs and when. Costs that are sunk at the time of hiring, such as search costs, for example, decrease the bargaining position of the party that paid them. Adaptation costs, however, do not change the bargaining position of the parties. These costs are not sunk at hiring, and they are also firm-specific, implying that they cannot be used as a threat in wage negotiations. In equilibrium, workers and firms split up the cost of adaptation based on their fixed bargaining power. Rookies are less productive, but they also make less money.

The model predicts that $E \to N$ movements are not responsive to adaptation, because in the baseline specification they are exogenous. The job loss probabilities of rookies are not changing with the adaptation period either; the elasticity is close to zero, because of the assumption that becoming experienced is a memoryless Poisson process.

Equation (1.28) can be rewritten as

$$1 = \frac{\varepsilon_{t^a}^{f^{EU}}}{\varepsilon_{t^a}^{u^{odds}}} - \frac{\varepsilon_{t^a}^{f^{UE}}}{\varepsilon_{t^a}^{u^{odds}}}. \tag{1.29}$$

The first term shows the ratio of the job loss and the unemployment elasticities. The term can be interpreted as the ratio of the unemployment elasticity that is due to changes in layoffs. Similarly, the negative of the second term can be interpreted as the ratio of the unemployment elasticity due to changes in hiring. The last two lines of Table 1.10 show the two ratios. Layoffs explain around 110 percent of the unemployment elasticity, while hiring explains -10 percent of it. Hiring explains a negative fraction of the unemployment elasticity, because the model predicts that in occupations where adaptation costs are large, both equilibrium unemployment and hiring will be below average, that is, hiring goes in the wrong direction to explain differences in unemployment.

In order to see how robust my predictions are, I run sensitivity analyses on five externally calibrated parameters. In each dimension I use relatively wide bounds. Tables 1.11 and 1.12 show the results. Throughout all specifications, the hiring elasticity remains close to zero. It's largest absolute value is only -0.079. The layoff and unemployment elasticities are always substantially larger than the hiring elasticity. Interestingly, the ratios in (1.29) are quite robustly estimated: Layoffs explain around 110 percent of the unemployment elasticity, and hiring explains -10 percent of it. The elasticities vary with parameters in the expected way. When more $E \to N$ transitions are endogenous (k^{NE} is larger), the model predicts a lower elasticity in $E \to U$, but the sum of $\varepsilon_{t^a}^{f^{EU}}$ and $\varepsilon_{t^a}^{f^{EN}}$ is roughly a constant. Because I assume $E \to N$ transitions do not contribute to the unemployment rate, larger k^{NE} translates into a somewhat smaller unemployment elasticity. When adaptation is quicker (φ is larger), not surprisingly, the layoff and unemployment elasticities are less responsive to adaptation costs. Similarly, when rookies' productivity (δ) is closer to that of experienced workers, adaptation costs are less predictive of layoffs and unemployment. Search costs also affect the layoff elasticity, but the effect is not large.

The parameter with the largest effect on my estimates is the autocorrelation in match-specific productivity. The more persistent the productivity shocks are, the less predictive adaptation costs are for layoffs and unemployment. Why is that? The autocorrelation in productivity is the primary determinant of the extent of labor hoarding in the economy. When a bad shock hits the match, the parties can choose from two strategies: 1) They either decide to split up and search for a better match; or 2) they decide to stay together and persevere. This second strategy is called labor hoarding. Hoarding labor is more profitable when the cost of turnover is high (in order to save on these costs) and when shocks are relatively short lived (because bad times do not last long and it is worth waiting). When the autocorrelation is low, negative productivity shocks are expected to go away soon, and thus, firms are more likely to hoard their workers. And when the autocorrelation is low and labor hoarding is widespread, as Table 1.12 shows, adaptation costs are more important determinants of equilibrium layoff and unemployment rates. Conversely, when the autocorrelation is high, there is less incentive to hoard labor and adaptation costs become less predictive of turnover. At the end of the next section I return to the discussion of which values of autocorrelation might be reasonable.

1.4 Calibration of the explained educational gap in unemployment

I use a three step procedure to quantitatively evaluate how much of the unemployment gap between education groups can be explained by adaptation costs:

1. I estimate average adaptation costs by occupations and education groups from the MCSUI data.

2. I plug the estimates from step 1 into my model to impute explained turnover and unemployment rates for each occupation-education cell.

3. I aggregate estimates from step 2 to the four education groups using the observed distribution of workers in the CPS.

In the first step I estimate adaptation costs by occupational work activities and education using the empirical model in (1.21). On the right-hand side, I use the three occupational activity measures identified as the best predictors in Section 1.2.4 (interpreting the meaning of information for others; organizing, planning, and prioritizing work; and developing objectives and strategies) and their full interaction with the four education dummies.

In the second step, I use my model to compute implied unemployment and turnover rates for a detailed grid of adaptation costs. The grid is wide enough to cover the entire range of possible cost values from step 1. Once I have a grid of implied turnover and unemployment rates, I approximate them with fourth order polynomials:

$$y\left(t^{a}\right) \approx \alpha_{0}^{y}+\alpha_{1}^{y} t^{a}+\alpha_{2}^{y}\left(t^{a}\right)^{2}+\alpha_{3}^{y}\left(t^{a}\right)^{3}+\alpha_{4}^{y}\left(t^{a}\right)^{4} . \tag{1.30}$$

This approximation, in practice, fits the unemployment and turnover estimates perfectly. Figure 1.3 shows the model's predictions and the approximations using the baseline calibration of the model. The model predicts a smooth effect of adaptation costs, and thus, the approximation works very well. It works similarly well for other calibrations of the model.

In step 3, I compute average imputed unemployment rates and turnover probabilities for each education group, where I average the estimates from step 2. The standard errors in step 3 take into account the imprecision of the estimates in step 1 using the delta method. The entire procedure can be implemented using the *margins* post-estimation command in Stata after fitting model (1.21) and using out-of-sample predictions of functions (1.30) by education groups in the CPS.

The results using the baseline specification of the model are in Figure 1.4 and appendix Table 1.24. Adaptation costs, using this specification, explain 11 percent of the unemployment gap between high school dropouts and high school graduates; 71 percent of gap between high school graduates and college dropouts; and 65 percent of the gap between college dropouts and college graduates. It appears that adaptation costs are better at explaining the differences between high and medium skilled workers' unemployment rates. It is also possible, however, that this differential reflects measurement error in the adaptation cost estimates. Employers, when asked about the length of adaptation in a given occupation, might use a positive anchor to answer the question. In this case the responses would have a mean-reverting property, and the adaptation cost estimates would be upward biased at low values. There is some evidence that employers use a positive anchor. Out of the 48 occupations considered, the lowest median adaptation period is still one month, which seems a little high.

The explained unemployment gaps depend quite substantially on the assumed autocorrelation in match-specific productivity. This is not surprising, as I showed in the previous section the turnover elasticities varied with the assumed autocorrelation, too. Figure 1.5 and appendix Tables 1.26-1.27 show the results using lower and higher autocorrelation. If I use a value of 0.85 for the autocorrelation, which implies that negative shocks disappear in about half a year on average, I can explain 25 percent of the HSD-HS gap, 141 percent of the HS-SC and 106 percent of the SC-BA differential. When I use a value of 0.95, implying that shocks disappear in about 1.6 years, I can only explain 5 percent of the HSD-HS gap, 35 percent of the HS-SC and 35 percent of the SC-BA differential.

The explained gaps in employment to unemployment transition probabilities are very similar to the ones in unemployment. The model, similar to the data, consistently predicts a small negative effect of adaptation costs on job finding probabilities, independent of the choice of autocorrelation. The monthly job finding probabilities are predicted to be around 29.5 percent in all education groups.

1.4.1 Autocorrelation in match-specific productivity

The explained unemployment gap varies quite substantially with the assumed autocorrelation in match-specific productivity. Figure 1.6 visualizes this sensitivity in another way. The figure shows the model's predicted equilibrium unemployment rates as a function of adaptation costs using different values of the autocorrelation. The four lines intersect at the average adaptation period, because the model is calibrated to match the average unemployment rate for the average value of adaptation. The three larger values (0.85, 0.92 and 0.95) are the same as used earlier, with 0.92 being my preferred value. The smallest value (0.71) is the one that would approximately explain the entire educational gap even between high school dropouts and high school graduates. A value of 0.71 for the monthly autocorrelation corresponds to yearly autocorrelation below 0.02 (!), which seems inconceivably small. Another reason why the 0.71 is unlikely to be appropriate is that, using this value, the model predicts zero unemployment rates in every occupation where adaptation takes longer than 5.5 months. This happens because the autocorrelation is so low, and negative shocks disappear so quickly, that it is even worth keeping matches that produce zero value contemporaneously. In the MCSUI data there are many occupations with adaptation longer than 5.5 months. The value of 0.85 is approximately the smallest value that predicts positive unemployment rates in the entire 0-12 month range. Hence, I consider 0.85 as the lower bar for a reasonable value of autocorrelation.

What empirical evidence do we have about the autocorrelation in match-specific productivity? My baseline specification is based on Abraham and White (2006) who found yearly autocorrelation in plant level TFP to be around 0.37-0.41 using yearly observations on a large sample of establishments in the manufacturing sector. My preferred 0.92 value is the monthly equivalent of this number. Foster et al. (2008) estimates plant level productivity dynamics in traditional TFP, as well as revenue and physical TFP. The data they use is available once every five years. They find the 5-year autocorrelation to be around 0.25-0.3. The monthly equivalent of this number is around 0.98, considerably higher than in the Abraham and White (2006) study. Arguably the autocorrelation in match-specific productivity is lower than in plant level productivity because individuals might also experience shorter lived shocks, such as individual emotional issues, sickness, etc., which are missing from plant level estimates, as other co-workers can substitute for temporarily unproductive individuals. For this reason I prefer using the estimates in Abraham and White (2006) that are on the low end of estimates in this literature.

An alternative strategy to proxy autocorrelation is to look at wage dynamics. The main advantage of wage data is that it is based on individual or "match" level information. One disadvantage, however, is that wages might be more rigid than productivity. Another disadvantage is that estimates of autocorrelation are quite sensitive to the wage model used. For example, as Lillard and Willis (1978) showed, by adding a time-invariant person-specific random effect to the wage model, the autocorrelation in wages substantially falls. Their preferred yearly autocorrelation in the residual wages is around 0.35-0.4, which is close to value used in my baseline specification. Other papers, such as MaCurdy (1982) and Abowd and Card (1989), however, prefer models with high autocorrelation, close to a unit-root. It seems intuitive that there is a component of earnings that follows a random walk, for example, due to promotions and wage renegotiations after workers receive outside wage offers. Such wage changes, however, are not due to productivity shocks and are not relevant to this paper. Overall, there is a wide range of estimates in the literature for the autocorrelation depending on how the wage equation is specified. There are estimates as low as 0.10 (Hospido, 2014; monthly value is ~ 0.82), as large as a unit root, and many estimates in between (Lillard, 1999; Geweke and Keane, 2000; Alvarez and Arellano, 2004; Hospido, 2012; Altonji et al., 2013; Lemieux et al., 2014; Guvenen et al., 2015).

Yet another strategy to proxy autocorrelation is to use calibrated values from the macro-labor literature that fit certain empirical moments. Depending on the model used and the particular moment that is fit, these calibrated values, however, are different. Fujita and Ramey (2012) use a value of 0.72 for the monthly autocorrelation to fit the serial correlation in the aggregate job-loss probabilities over the business cycle. Ratner (2013) uses a value of 0.95 to fit the size of layoffs in firms. Krolikowski (2014) uses a value of 0.88 to fit the rate at which job loss probabilities fall by tenure.

The recent working paper of Cairo and Cajner (2014) has been developed in parallel with my paper. They provided estimates of the educational gap in unemployment that can be explained by educational differences in on-the-job training. The main difference between their approach and mine is that they use the Fujita and Ramey (2012) calibration of 0.72 for the monthly autocorrelation, while my preferred number is 0.92. There are other differences such as: 1) the adaptation period measure is asked in a more consistent way in the MCSUI and it is available for a markedly larger sample than in the EOPP[19]; 2) Cairo and Cajner

[19]Adaptation costs were asked in the EOPP in the following way. First, employers were asked about the current productivity of the last hired worker on a scale of 0 to 100 with 50 being the perceived average of an experienced worker in the position. Second, they asked about the productivity of the last hired worker at the time of hiring on the same scale. Third, if the initial productivity of the workers was below 50, they asked the employers about the time it took for the last hire to reach average productivity. No question about adaptation was asked if the initial productivity of the last hire was at least 50, not even if his current productivity was above his initial one. For example, if a worker's productivity was 50 to begin with and it went up to 70 over time, the survey did not ask about the length of adaptation. This approach might be particularly problematic if employers are overly optimistic about the productivity of their workers, for which there is evidence in the EOPP. The sample average of the current productivity of workers was 71 (s.e. 0.38) instead of 50 and only 14 percent of employers reported a value below 50. The average employer in the EOPP thought he employed above average workers. Moreover, two third of the employers reported that the initial productivity of their last hire was at least 50. Adaptation period estimates are only available for the rest of the sample, around 650 observations. The MCSUI data does not suffer from this inconsistency, and adaptation cost information is available for more than 3200 workers.

(2014) focus on educational differences in on-the-job training directly, without calibrating turnover for individual occupations; 3) they analyze educational unemployment rates in the entire labor force, while I use only prime age males.

Overall, the evidence is quite contradictory about which is the true value of autocorrelation in productivity.

1.5 Discussion and conclusion

Unemployment rates are systematically and substantially larger among less educated versus highly educated workers. This paper tested how much of this differential can be explained by a particular form of labor adjustment cost: occupational differences in the average time it takes for newly hired, inexperienced workers to become fully productive at their firms. The main message is that occupational differences in adaptation costs are important determinants of turnover and unemployment rate differences. Altogether, according to my estimates, adaptation costs explain around two thirds of the unemployment differences between college and high school graduates. The role of adaptation costs, however, is closely linked to productivity dynamics. In order to pin down the effect of adaptation, we need to get a better understanding of how productivity shocks hit matches. In the meantime, it would be advantageous to collect more detailed and more systematic data on adjustment costs by occupations in general, and adaptation costs, in particular. The type of questions about adaptation costs analyzed in Blatter et al. (2012) and in this paper are very predictive of turnover. Collecting such data in more detailed occupations and larger samples could improve our understanding of labor turnover and unemployment.

A natural question arising from my results is which economic mechanism is possibly responsible for the rest of the gap not explained by adaptation cost differences. A good explanation should have the property of predicting relatively small differences in hiring rates.

1.5.1 Economic mechanisms with large effects on hiring

One mechanism to explain skill differences in unemployment is that low skilled wages are above market clearing levels leading to an over-supply of unskilled workers. This could happen due to the minimum wage, unions' bargaining practices, mandatory health insurance policies, or any other labor market institutions that compress wage dispersion in the workforce. Minimum wages, at least in the US, are not indexed to occupational productivity, and thus, they are more likely to bind for low skilled workers. The minimum wage literature is rather inconclusive about whether the minimum wage affects employment, but most economists think the minimum wage does have a small negative effect on employment, particularly among the lowest skilled workers, such as teens (Neumark and Wascher, 2006). Collective bargaining might also lead to less dispersed wages[20], which might push low skilled wages above equilibrium levels. It can be shown, however, that this mechanism predicts comparatively low hiring rates among low skilled workers.

Another candidate mechanism that has been suggested in the literature is that the value of being unemployed is relatively high for low skilled workers. This could happen because of the unemployment insurance (UI) system, lack of adequate skills in some worker groups, or any other factors that influence the utility of working and the utility of not working. The value of unemployment is a primary factor determining the search effort of the unemployed. The generosity of the UI system, for example, has been found to be an important determinant of the unemployment rate (Roed and Zhang, 2003; Schmieder et al., 2010; Hagedorn et al.,

[20] See, for example, Freeman (1980), Card (2001), Card et al. (2004) and Frandsen (2012).

2013). The strong concavity of the UI benefit formula might create a disincentive for low skilled workers to search for, and firms to create, low skilled jobs. The most credible evidence we have about the negative employment effects of the UI system, however, is about the effect of the duration of the UI benefit as opposed to the replacement rate. Benefit duration, however, does not vary across occupations/education groups. Skill shortage (or skill mismatch) is another factor that might influence the ratio of workers' productivity to the value of not working. Technological change in the last thirty years might have made the skill set of some workers obsolete (Autor et al., 2003; Acemoglu and Autor, 2011; Jaimovich and Siu, 2012; Lazear and Spletzer, 2012). If the productivity of these workers fell faster than their outside option, skill mismatch could have led to excess unemployment. This mechanism, however, also predicts large differences in hiring rates by skill groups, which is inconsistent with the empirical evidence. In recent papers, Lazear and Spletzer (2012) and Sahin et al. (2012) analyzed occupational skill mismatch using occupation-specific online job-vacancy data, the Help Wanted OnLine Index. They found that occupation-specific labor market tightness (the ratio of job vacancies and unemployment) shows a large cross-sectional heterogeneity. As Lazear and Spletzer (2012) observes, this either indicates a shortage of high skilled and an abundance of low skilled workers, or measurement error. Many low skilled jobs might be advertised offline, and thus, the lack of enough low skilled vacancies in the online dataset might be irrelevant for understanding unemployment rate differences by skill. The results of my paper support the latter measurement error argument, as high skilled workers do not seem to find jobs quicker than low skilled ones.

1.5.2 Economic mechanisms with small effects on hiring

Perhaps the simplest idea to fit the entire unemployment gap is to use different values of adaptation costs. It is possible that the survey measures of occupational adaptation costs in the MCSUI data are biased. How large should adaptation cost differences be to explain the entire unemployment gap by education groups? In Table 1.13 I assume that the average length of adaptation (~ 3 months) in the MCSUI is unbiased, and I use my model to find the average length of adaptation that is needed to fit unemployment rates in the four education groups. First of all, I cannot fit the unemployment rate among high school dropouts using the baseline calibration of the model, because the implied unemployment rate is too small even when adaptation costs are zero. The imputed adaptation costs for the other three education groups are approximately 2, 4 and 7 months for high school graduates, college dropouts and college graduates respectively. These numbers are not unrealistic, but they are much more dispersed than in the MCSUI data. In Table 1.14 I repeat the same exercise using the lower value of autocorrelation (0.85). In this specification I can fit the unemployment rates in all four groups. The imputed value for high school dropouts is substantially lower than in the MCSUI data (0.8 compared to 2.5 months), but it is plausible. Moreover, the imputed values for the other three groups are actually quite similar to the data.

The variance of idiosyncratic productivity shocks is another parameter that, similarly to adaptation costs, also has a small hiring and a large layoff elasticity (Hudomiet, 2014). There are good reasons to believe that the variance of productivity shocks is larger in low skilled jobs, an example of which being capital-skill complementarity. As Griliches (1969) and other recent studies (Krusell et al., 2000; Duffy et al., 2004; Parro, 2013) argued, the elasticity of substitution is larger between capital and unskilled labor than between capital and high skilled labor. In this case the marginal product of unskilled workers is expected to fluctuate more with demand conditions, since unskilled work is more substitutable with a relatively fixed production factor, capital (Oi, 1962). Another economic mechanism that predicts larger productivity fluctuations in low skilled jobs is differences in inventory costs. One way firms can deal with temporary demand shocks is to increase inventories in bad times, and sell out these inventories in later periods when the demand conditions improve. This strategy is more profitable if storing the goods is relatively cheap. One can argue that high skilled work, such as idea generation, research, innovation, etc., is cheaper to store than low skilled work, which takes up physical space or is more linked to dealing with customers (Wen, 2005; Platt and Platt, 2011). It is, however, not trivial to quantify inventory costs and capital-skill complementarity by occupations.

Tables 1.15 and 1.16 show the values of standard deviation of the shocks that would be needed to explain the rest of the unemployment gaps, not explained by adaptation costs. Using the baseline calibration of the model, the standard deviation of shocks in the four education groups should be 0.78, 0.29, 0.25 and 0.22. The value for high school dropouts is a little large, but not implausible. The imputed values for the three more educated groups are actually quite close to each other and they are entirely plausible. The table also shows the equilibrium values of monthly turnover probabilities. The job finding probability of high school dropouts is upward biased, but the other probabilities are close to empirical estimates. When I use lower autocorrelation in Table 1.16, the imputed values are even closer to each other and all unemployment rates and turnover probabilities are close to empirical estimates.

Tables and figures

Table 1.1: Logit regressions of unemployment

	Logit of the unemployment rate						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High school dropout	0.481**	0.496**	0.397**	0.344**	0.290**	0.269**	0.267**
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
High school	reference category						
Some college	-0.307**	-0.305**	-0.147**	-0.106**	-0.088**	-0.079**	-0.078**
	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
BA or more	-0.962**	-0.901**	-0.445**	-0.400**	-0.333**	-0.320**	-0.313**
	(0.010)	(0.010)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
Age 25-29	reference category						
Age 30-34		-0.109**	-0.100**	-0.086**	-0.074**	-0.069**	-0.069**
		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Age 35-39		-0.170**	-0.159**	-0.134**	-0.111**	-0.100**	-0.100**
		(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Age 40-44		-0.233**	-0.214**	-0.181**	-0.150**	-0.131**	-0.131**
		(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Age 45-49		-0.264**	-0.240**	-0.201**	-0.160**	-0.137**	-0.137**
		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Age 50-55		-0.265**	-0.238**	-0.195**	-0.145**	-0.115**	-0.115**
		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Whites	reference category						
Non whites		0.525**	0.512**	0.516**	0.501**	0.517**	0.518**
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Not married	reference category						
Married		-0.749**	-0.735**	-0.704**	-0.675**	-0.668**	-0.666**
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Quarter dummies	Y	Y	Y	Y	Y	Y	Y
Occupations, 10 categ/s			Y				
Occupations, 48 categ/s				Y			
Occupations, 191 categ/s					Y	Y	Y
Industries, 12 categ/s						Y	
Industries, 38 categ/s							Y
Log likelihood	-1,679,795	-1,634,500	-1,618,626	-1,599,715	-1,586,857	-1,579,408	-1,577,514
Observations	8,899,239	8,899,239	8,899,239	8,899,239	8,899,239	8,899,239	8,899,239

The table shows the raw coefficients of the logistic model. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05. Sample: 25-55 year old males, CPS 1978-2013

Table 1.2: OLS regressions of unemployment in booms

	Unemployment rate						
$\bar{U} = 0.038$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High school dropout	0.028**	0.028**	0.024**	0.021**	0.017**	0.017**	0.016**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High school	reference category						
Some college	-0.011**	-0.011**	-0.005**	-0.004**	-0.003**	-0.003**	-0.003**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BA or more	-0.026**	-0.023**	-0.010**	-0.009**	-0.008**	-0.007**	-0.007**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 25-29	reference category						
Age 30-34		-0.005**	-0.005**	-0.004**	-0.004**	-0.004**	-0.004**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 35-39		-0.007**	-0.006**	-0.005**	-0.004**	-0.004**	-0.004**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 40-44		-0.008**	-0.007**	-0.006**	-0.005**	-0.004**	-0.004**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 45-49		-0.009**	-0.008**	-0.006**	-0.005**	-0.004**	-0.004**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 50-55		-0.008**	-0.007**	-0.006**	-0.004**	-0.003**	-0.003**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Whites	reference category						
Non whites		0.025**	0.024**	0.025**	0.024**	0.025**	0.025**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Not married	reference category						
Married		-0.033**	-0.032**	-0.030**	-0.029**	-0.029**	-0.029**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Quarter dummies	Y	Y	Y	Y	Y	Y	Y
Occupations, 10 categ/s			Y				
Occupations, 48 categ/s				Y			
Occupations, 191 categ/s					Y	Y	Y
Industries, 12 categ/s						Y	
Industries, 38 categ/s							Y
Observations	5,000,633	5,000,633	5,000,633	5,000,633	5,000,633	5,000,633	5,000,633

Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05. Sample: 25-55 year old males, CPS 1978-2013. Boom indicates years when the aggregate unemployment rate was below 5 percent in the sample: 1978-1979, 1987-1990 and 1994-2008.

Table 1.3: Logit regressions of monthly transition probabilities

	$E \to U$ probability			$U \to E$ probability		
	(1)	(2)	(3)	(4)	(5)	(6)
High school dropout	0.537**	0.268**	0.236**	0.034**	0.005	0.000
	(0.012)	(0.012)	(0.012)	(0.010)	(0.010)	(0.010)
High school			reference category			
Some college	-0.319**	-0.062**	-0.050**	0.004	0.046**	0.049**
	(0.012)	(0.012)	(0.012)	(0.010)	(0.011)	(0.011)
BA or more	-0.967**	-0.332**	-0.311**	-0.047**	0.054**	0.064**
	(0.013)	(0.017)	(0.017)	(0.012)	(0.014)	(0.014)
Age 25-29			reference category			
Age 30-34	-0.129**	-0.095**	-0.090**	-0.039**	-0.049**	-0.051**
	(0.014)	(0.014)	(0.014)	(0.012)	(0.012)	(0.012)
Age 35-39	-0.228**	-0.165**	-0.153**	-0.088**	-0.101**	-0.102**
	(0.014)	(0.014)	(0.014)	(0.012)	(0.013)	(0.013)
Age 40-44	-0.311**	-0.217**	-0.194**	-0.119**	-0.128**	-0.129**
	(0.015)	(0.015)	(0.015)	(0.013)	(0.013)	(0.013)
Age 45-49	-0.385**	-0.267**	-0.238**	-0.169**	-0.175**	-0.175**
	(0.016)	(0.016)	(0.016)	(0.014)	(0.014)	(0.014)
Age 50-55	-0.465**	-0.326**	-0.287**	-0.255**	-0.256**	-0.256**
	(0.016)	(0.016)	(0.016)	(0.014)	(0.014)	(0.014)
Whites			reference category			
Non whites	0.325**	0.331**	0.355**	-0.245**	-0.220**	-0.218**
	(0.012)	(0.012)	(0.012)	(0.010)	(0.010)	(0.010)
Not married			reference category			
Married	-0.545**	-0.466**	-0.453**	0.197**	0.193**	0.195**
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)
Quarter dummies	Y	Y	Y	Y	Y	Y
Occupations, 191 categ/s		Y	Y		Y	Y
Industries, 38 categ/s			Y			Y
Log likelihood	-411,862	-397,659	-393,962	-173,168	-171,819	-171,459
Observations	6,064,820	6,064,820	6,064,820	295,045	295,045	295,045

Sample: 25-55 year old males, CPS 1978-2013. The table shows average log differences in the implied probabilities relative to the baseline. For example, the coefficient on BA in the $E \to U$ regressions is $\beta_{BA}^{implied} = \frac{1}{N}\sum_i \left[\ln \Pr\left(P_i^{EU}=1|educ=BA,X_i\right) - \ln \Pr\left(P_i^{EU}=1|educ=HS,X_i\right)\right]$. More generally, in dimension j and category k the coefficient is $\beta_{jk}^{implied} = \frac{1}{N}\sum_i \left[\ln \Pr\left(P_i^{EU}=1|I_j=k,X_i\right) - \ln \Pr\left(P_i^{EU}=1|I_j=0,X_i\right)\right]$. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05.

Table 1.4: Length of adaptation periods, MCSUI

	Adaptation period in weeks	Adaptation period in month
Mean(t^a)	29.32	6.90
Median(t^a)	12.00	2.82
Mean($\ln t^a$)	2.55	1.10
exp(mean($\ln t^a$))	12.80	3.01
Observations	3243	3243

Adaptation period is the employers' estimate of the time it takes to become fully competent for an average employee in the position. The second column is created by dividing the first by 4.25.

Table 1.5: OLS regression of log adaptation period by education and occupational work activities, MCSUI

	[1]	[2]	[3]	[4]	[5]	[6]
Education						
Less than high school	-0.336**			-0.242*	-0.240*	-0.252*
	(0.108)			(0.105)	(0.105)	(0.106)
High school			reference			
Some college	0.353**			0.230**	0.218**	0.222**
	(0.057)			(0.056)	(0.056)	(0.057)
At least BA	0.744**			0.349**	0.341**	0.301**
	(0.053)			(0.060)	(0.060)	(0.065)
Occupational work activities						
Developing Objectives and Strategies		0.970**	1.346**	0.749**	1.001**	
		(0.280)	(0.323)	[0.281]	(0.326)	
Organizing, Planning, and Prioritizing Work		1.505**	1.680**	1.306**	1.621**	
		(0.310)	(0.403)	[0.310]	(0.400)	
Interpreting the Meaning of Information for Others		1.200**	1.316**	0.906**	1.144**	
		(0.325)	(0.398)	[0.327]	(0.399)	
Getting Information			-1.258*		-1.216*	
			(0.522)		(0.519)	
Interacting With Computers			0.439*		0.251	
			(0.189)		(0.190)	
Documenting/Recording Information			0.469		0.438	
			(0.324)		(0.323)	
Communicating with Supervisors, Peers, Subordinates			-1.021*		-0.883	
			(0.495)		(0.493)	
Constant	2.246**	0.634**	1.381**	0.835**	1.492**	3.552**
	(0.035)	(0.126)	(0.243)	(0.130)	(0.243)	(0.972)
Occupation dummies, 191 categ/s						YES
Observations	3,243	3,243	3,243	3,243	3,243	3,243
R-squared	0.069	0.104	0.108	0.118	0.121	0.223

The regression coefficients are from regressions of log adaptation period on occupational work activity measures. Adaptation period is the employer's estimate of the time it takes to become fully competent for an average employee in the position. The occupational skill measures are continuous variables between 0-1, and they are created from the generalized work activities measures in the O*NET. ** p<0.01, * p<0.05.

Table 1.6: Predictive power of occupation work activities on log adaptation costs, MCSUI and O*NET

#	Generalized work activities	Coefficients*	t-statistics
	Information input		
1	Getting Information	3.409**	15.130
2	Monitor Processes, Materials, or Surroundings	1.664**	7.361
3	Identifying Objects, Actions, and Events	2.976**	11.821
4	Inspecting Equipment, Structures, or Material	-0.439**	3.052
5	Estimating the Quantifiable Characteristics of Products, Events, or Information	1.140**	5.096
	Mental processes		
6	Judging the Qualities of Things, Services, or People	2.088**	9.967
7	Processing Information	2.470**	14.278
8	Evaluating Information to Determine Compliance with Standards	2.445**	12.618
9	Analyzing Data or Information	2.536**	16.988
10	Making Decisions and Solving Problems	2.608***	14.653
11	Thinking Creatively	2.486**	16.009
12	Updating and Using Relevant Knowledge	3.031**	16.888
13	Developing Objectives and Strategies	2.755**	17.488
14	Scheduling Work and Activities	2.953**	16.917
15	Organizing, Planning, and Prioritizing Work	3.402**	17.610
	Work output		
16	Performing General Physical Activities	-1.198**	10.035
17	Handling and Moving Objects	-1.366**	11.703
18	Controlling Machines and Processes	-0.834**	6.773
19	Operating Vehicles, Mechanized Devices, or Equipment	-0.614**	5.073
20	Interacting With Computers	1.496**	12.485
21	Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment	0.599***	3.978
22	Repairing and Maintaining Mechanical Equipment	-0.541**	4.102
23	Repairing and Maintaining Electronic Equipment	-0.216	1.107
24	Documenting/Recording Information	2.265**	13.660
	Interacting with others		
25	Interpreting the Meaning of Information for Others	3.089**	17.971
26	Communicating with Supervisors, Peers, or Subordinates	3.719**	14.216
27	Communicating with Persons Outside Organization	1.396**	9.208
28	Establishing and Maintaining Interpersonal Relationships	2.912**	12.200
29	Assisting and Caring for Others	0.304	1.936
30	Selling or Influencing Others	0.155	1.359
31	Resolving Conflicts and Negotiating with Others	1.541**	8.834
32	Performing for or Working Directly with the Public	-0.263*	2.008
33	Coordinating the Work and Activities of Others	2.501**	12.487
34	Developing and Building Teams	2.208**	11.642
35	Training and Teaching Others	2.164**	12.370
36	Guiding, Directing, and Motivating Subordinates	1.867**	11.637
37	Coaching and Developing Others	2.184**	13.009
38	Provide Consultation and Advice to Others	2.181**	13.819
39	Performing Administrative Activities	1.681**	11.153
40	Staffing Organizational Units	1.793**	10.361
41	Monitoring and Controlling Resources	1.635**	8.909

The regression coefficients are from simple regressions of log adaptation period on occupational work activity measures. Adaptation period is the employer's estimate of the time it takes to become fully competent for an average employee in the position. The occupational skill measures are continuous variables between 0-1, and they are created from the generalized work activities measures in the O*NET. ** $p<0.01$, * $p<0.05$.

Table 1.7: Logit regressions of unemployment and turnover on imputed adaptation costs

	(1)			(2)		
	$logit(u)$	$logit\left(p^{EU}\right)$	$logit\left(p^{UE}\right)$	$logit(u)$	$logit\left(p^{EU}\right)$	$logit\left(p^{UE}\right)$
Imputed log adaptation period	-1.051**	-1.147**	-0.119**	-0.825**	-0.967**	-0.063
	(0.007)	(0.016)	(0.019)	(0.011)	(0.025)	(0.032)
High school dropouts				0.238**	0.223**	-0.004
				(0.010)	(0.021)	(0.024)
High school			reference category			
Some college				0.131**	0.224**	0.049
				(0.010)	(0.020)	(0.025)
BA or more				-0.059**	-0.003	0.030
				(0.012)	(0.026)	(0.035)
Age 25-29			reference category			
Age 30-34				-0.108**	-0.087**	-0.073**
				(0.011)	(0.024)	(0.028)
Age 35-39				-0.110**	-0.154**	-0.151**
				(0.011)	(0.024)	(0.028)
Age 40-44				-0.128**	-0.172**	-0.185**
				(0.011)	(0.024)	(0.028)
Age 45-49				-0.126**	-0.229**	-0.257**
				(0.012)	(0.025)	(0.029)
Age 50-55				-0.080**	-0.257**	-0.420**
				(0.012)	(0.025)	(0.030)
Whites			reference category			
Non whites				0.566**	0.392**	-0.323**
				(0.008)	(0.019)	(0.022)
Not married			reference category			
Married				-0.751**	-0.525**	0.282**
				(0.007)	(0.015)	(0.017)
Year dummies	Y	Y	Y	Y	Y	Y
Industries, 38 categ/s				Y	Y	Y
Observations	3,503,304	2,439,158	91,414	3,503,304	2,439,158	91,414

Sample: CPS 1994-2008, 25-55 year old males. The imputed adaptation cost values are the predicted values from a regression of log adaptation on the three occupational work activity measures (interpreting the meaning of information for others; organizing, planning, and prioritizing work; and developing objectives and strategies), education dummies (less than high school, high school, scome college, and college) and their interactions. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05.

Table 1.8: Average monthly transition probabilities

Transition type	Description	Values
$E \rightarrow U$ transition	Separating into unemployment	0.0114
$E \rightarrow N$ transition	Separating into non-participation	0.0105
$E \rightarrow E$ transition with separation	Employer switching	0.0216
All separations		0.0435
$U \rightarrow E$ transitions	Job finding from unemployment	0.2935
$N \rightarrow E$ transitions	Job finding from non-participation	0.0907

Sample: 25-55 year old males, CPS 1994-2008

Table 1.9: Calibrated parameters of the model and other implied values

Calibrated parameters	Symbol	Value	Source
Cobb-Douglas share of vacancies	α	0.4	Pissarides and Petrongolo (2001)
Workers' bargaining power	γ	0.6	Hosios condition
Labor market tightness	θ	0.72	Everyone
Interest rate	r	0.004	5% annually
Total search costs*	c/q	$0.5 \times E(w)$	Blatter et al. (2012), Ratner (2013)
Productivity of rookies*	δ	0.7	Blatter et al. (2012)
Autocorrelation in ε*	$\exp(-\lambda)$	0.92	Abraham and White (2006)
Rate of turning experienced*	φ	0.31	own estimates, MCSUI
Flow value of unemployment in avg(wage)	$b/E(w)$	0.659	job creation condition
Sd of match-specific productivity shocks	σ	0.261	job loss probability
Scalar in the matching function	η	0.399	job finding probability
Implied other values			
$E \rightarrow U$ rate, experienced	f^{EU}	0.014	
$E \rightarrow N$ rate, experienced	f^{EN}	0.011	
$E \rightarrow U$ rate, rookies	$f^{EU,R}$	0.020	
Vacancy filling rate	q	0.486	
Job finding rate	p^F	0.350	
Mean wage, experienced workers	$E(w)$	1.049	
Mean wage, rookies	$E(w^R)$	0.862	

Sensitivity analyses are carried out on parameters with an asterisk

Table 1.10: Elasticities of turnover rates with respect to the length of the adaptation period, baseline calibration

Variables	Description	Values
$\varepsilon_{t^a}^{u^{odds}}$	Unemployment, odds ratio	-0.541
$\varepsilon_{t^a}^{f^{EU}}$	$E \rightarrow U$ rate, experienced	-0.588
$\varepsilon_{t^a}^{f^{EN}}$	$E \rightarrow N$ rate, experienced	0.000
$\varepsilon_{t^a}^{f^{EU,R}}$	$E \rightarrow U$ rate, rookies	0.006
$\varepsilon_{t^a}^{q}$	Vacancy filling rate	0.071
$\varepsilon_{t^a}^{f^F}$	Job finding rate	-0.048
$\varepsilon_{t^a}^{f^{EU}} / \varepsilon_{t^a}^{u^{odds}}$	Loss ratio	1.088
$\varepsilon_{t^a}^{f^F} / \varepsilon_{t^a}^{u^{odds}}$	Finding ratio	0.088

Table 1.11: Elasticities of turnover rates with respect to the length of the adaptation period, sensitivity analysis

		Effect of k^{EN}		Effect of φ		Effect of δ	
Elasticities of var's	Description	0.2	0.4	0.25	0.5	0.5	0.9
$\varepsilon_a^{u^{odds}}$	Unemployment, odds ratio	-0.463	-0.398	-0.593	-0.419	-0.763	-0.231
$\varepsilon_a^{f^{EU}}$	$E \rightarrow U$ rate, experienced	-0.511	-0.445	-0.649	-0.452	-0.843	-0.247
$\varepsilon_a^{f^{EN}}$	$E \rightarrow N$ rate, experienced	-0.102	-0.178	0.000	0.000	0.000	0.000
$\varepsilon_a^{f^{EU,R}}$	$E \rightarrow U$ rate, rookies	0.005	0.005	0.010	0.000	0.002	0.005
ε_a^{q}	Vacancy filling rate	0.071	0.070	0.083	0.049	0.119	0.024
$\varepsilon_a^{f^F}$	Job finding rate	-0.047	-0.047	-0.055	-0.033	-0.079	-0.016
$\varepsilon_a^{f^{EU}} / \varepsilon_a^{u^{odds}}$	Loss ratio	1.102	1.118	1.093	1.079	1.104	1.069
$\varepsilon_a^{f^F} / \varepsilon_a^{u^{odds}}$	Finding ratio	0.102	0.118	0.093	0.079	0.104	0.069
Calibrated values for agg'							
$b/E(w)$	Flow value of unemployment in avg(wage)	0.655	0.651	0.653	0.669	0.639	0.679
σ	Sd of match-specific productivity shocks	0.280	0.300	0.282	0.224	0.334	0.190
$f^{EU,R}$	$E \rightarrow U$ rate, rookies	0.018	0.017	0.021	0.018	0.022	0.015
$E(w)$	Mean wage, experienced workers	1.066	1.086	1.056	1.038	1.074	1.030
$E(w^R)$	Mean wage, rookies	0.875	0.889	0.870	0.850	0.746	0.968

Table 1.12: Elasticities of turnover rates with respect to the length of the adaptation period, sensitivity analysis

		Effect of c/q		Effect of $\exp(-\lambda)$	
Elasticities of var's	Description	0.4	0.6	0.85	0.95
$\varepsilon_a^{u^{odds}}$	Unemployment, odds ratio	-0.584	-0.506	-1.098	-0.266
$\varepsilon_a^{f^{EU}}$	$E \rightarrow U$ rate, experienced	-0.644	-0.546	-1.146	-0.312
$\varepsilon_a^{f^{EN}}$	$E \rightarrow N$ rate, experienced	0.000	0.000	0.000	0.000
$\varepsilon_a^{f^{EU,R}}$	$E \rightarrow U$ rate, rookies	0.020	-0.007	-0.017	0.009
ε_a^{q}	Vacancy filling rate	0.089	0.060	0.072	0.070
$\varepsilon_a^{f^F}$	Job finding rate	-0.059	-0.040	-0.048	-0.047
$\varepsilon_a^{f^{EU}} / \varepsilon_a^{u^{odds}}$	Loss ratio	1.102	1.079	1.044	1.175
$\varepsilon_a^{f^F} / \varepsilon_a^{u^{odds}}$	Finding ratio	0.102	0.079	0.044	0.175
Calibrated values for agg'					
$b/E(w)$	Flow value of unemployment in avg(wage)	0.721	0.597	0.668	0.640
σ	Sd of match-specific productivity shocks	0.228	0.294	0.343	0.272
$f^{EU,R}$	$E \rightarrow U$ rate, rookies	0.022	0.018	0.025	0.016
$E(w)$	Mean wage, experienced workers	1.039	1.059	1.060	1.082
$E(w^R)$	Mean wage, rookies	0.858	0.867	0.875	0.882

Table 1.13: Adaptation cost values needed to fit unemployment rates by education groups, baseline calibration

	Adaptation period, months		Unemployment rate		$E \to U$ probability		$U \to E$ probability	
	Empirical	Imputed	Empirical	Model	Empirical	Model	Empirical	Model
High school dropout	2.5160	0.0000	0.0749	0.0605	0.0244	0.0197	0.3054	0.3064
High school	2.9481	2.0640	0.0449	0.0449	0.0139	0.0140	0.3037	0.2975
Some college	3.5531	3.9160	0.0334	0.0334	0.0102	0.0101	0.2963	0.2913
BA or more	4.6848	6.9719	0.0196	0.0196	0.0053	0.0057	0.2732	0.2828

The table shows imputed adaptation cost values that would be needed to fit the unemployment rates in the education groups using the baseline calibration of the model. The first column shows the average length of the adaptation period by education; the second shows the imputed values; and the other columns show the success of the model to fit the empirical unemployment rates and turnover probabilities using these imputed values.

Table 1.14: Adaptation cost values needed to fit unemployment rates by education groups, calibration with lower autocorrelation

	Adaptation period, months		Unemployment rate		$E \to U$ probability		$U \to E$ probability	
	Empirical	Imputed	Empirical	Model	Empirical	Model	Empirical	Model
High school dropout	2.5160	0.7794	0.0749	0.0749	0.0244	0.0244	0.3054	0.3020
High school	2.9481	2.6480	0.0449	0.0449	0.0139	0.0139	0.3037	0.2953
Some college	3.5531	3.5634	0.0334	0.0334	0.0102	0.0101	0.2963	0.2925
BA or more	4.6848	5.0044	0.0196	0.0196	0.0053	0.0058	0.2732	0.2885

The table shows imputed adaptation cost values that would be needed to fit the unemployment rates in the education groups using the calibration with lower monthly autocorrelation in productivity, $\exp(-\lambda) = 0.85$. The first column shows the average length of the adaptation period by education; the second shows the imputed values; and the other columns show the success of the model to fit the empirical unemployment rates and turnover probabilities using these imputed values.

Table 1.15: Standard deviation of productivity shock needed to fit unemployment rates by education groups, baseline calibration

	Standard deviation	Unemployment rate		$E \to U$ probability		$U \to E$ probability	
	Imputed	Empirical	Model	Empirical	Model	Empirical	Model
High school dropout	0.7777	0.0749	0.0748	0.0244	0.0314	0.3054	0.3881
High school	0.2890	0.0449	0.0449	0.0139	0.0140	0.3037	0.2979
Some college	0.2517	0.0334	0.0334	0.0102	0.0101	0.2963	0.2914
BA or more	0.2201	0.0196	0.0196	0.0053	0.0057	0.2732	0.2848

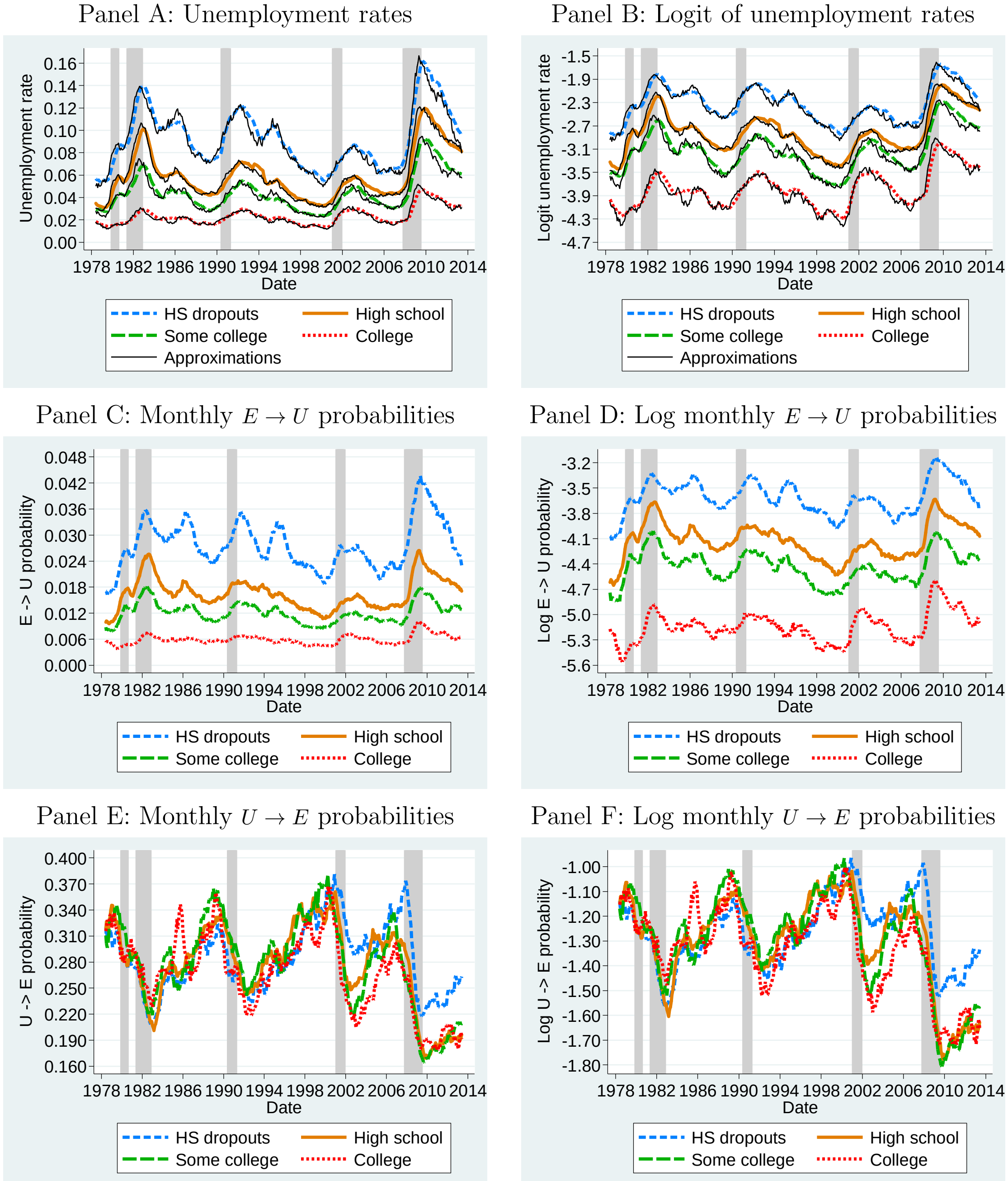
The table shows the imputed standard deviation of match-specific productivity shocks that would be needed, beyond the observed adaptation cost differences, to fit the unemployment rates in the education groups using the baseline calibration of the model. The first column shows the imputed values; and the other columns show the success of the model to fit the empirical unemployment rates and turnover probabilities using these imputed values.

Table 1.16: Standard deviation of productivity shock needed to fit unemployment rates by education groups, calibration with lower autocorrelation

	Standard deviation	Unemployment rate		$E \to U$ probability		$U \to E$ probability	
	Imputed	Empirical	Model	Empirical	Model	Empirical	Model
High school dropout	0.4597	0.0749	0.0749	0.0244	0.0253	0.3054	0.3122
High school	0.3575	0.0449	0.0449	0.0139	0.0139	0.3037	0.2961
Some college	0.3429	0.0334	0.0334	0.0102	0.0101	0.2963	0.2925
BA or more	0.3322	0.0196	0.0196	0.0053	0.0058	0.2732	0.2881

The table shows the imputed standard deviation of match-specific productivity shocks that would be needed, beyond the observed adaptation cost differences, to fit the unemployment rates in the education groups using the calibration with lower monthly autocorrelation in productivity, $\exp(-\lambda) = 0.85$. The first column shows the imputed values; and the other columns show the success of the model to fit the empirical unemployment rates and turnover probabilities using these imputed values.

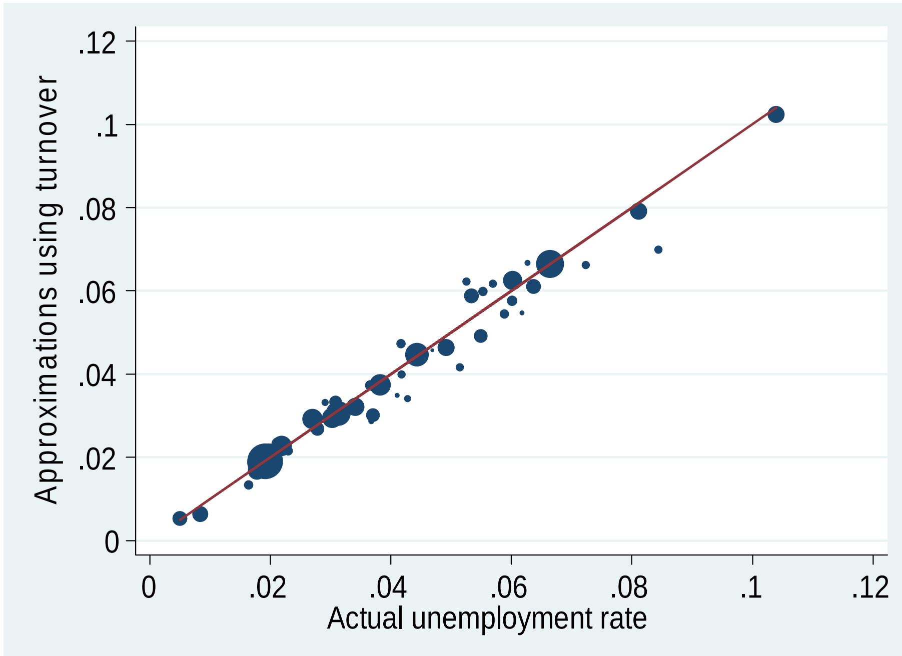
Figure 1.1: Unemployment rates and basic turnover probabilities by education



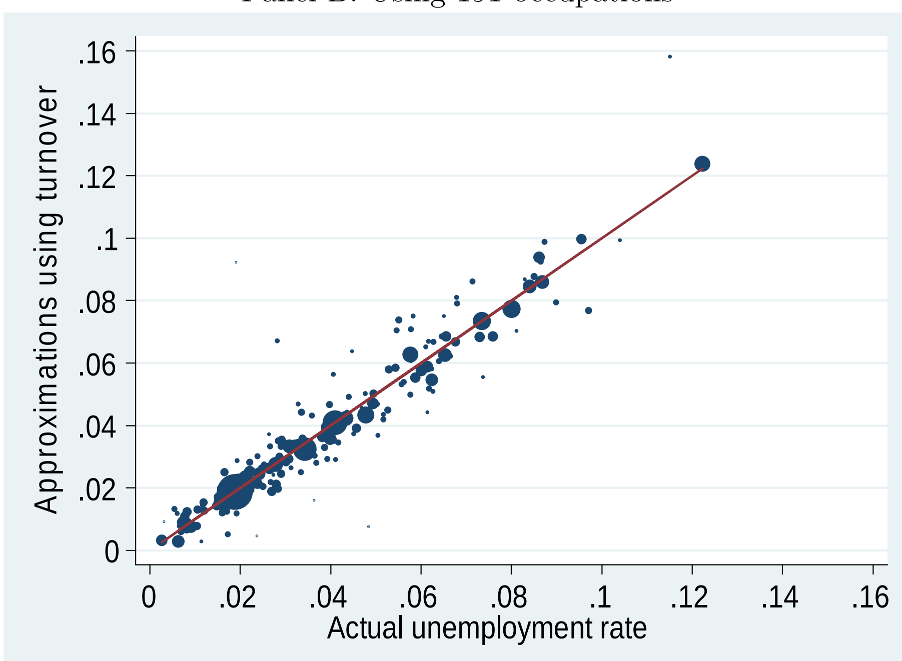
Sample: 25-55 year old males, CPS 1978-2013. All series are yearly moving averages of the monthly values. The logit of the unemployment rate is the log odds ratio, $logit\,(u) = \ln\left(\frac{u}{1-u}\right)$. The $E \to U$ transition probability is the probability that one is unemployed at month $t+1$ conditional on being employed at month t. The $U \to E$ transition probability is the probability that one is employed at month $t+1$ conditional on being unemployed at month t. The approximations in Panel A and B are $u_{gt}^{approx} = p^{EU}/(p^{EU} + p^{UE})$. The gray shaded areas indicate NBER recessions.

Figure 1.2: Occupational unemployment and its approximation with turnover rates

Panel A: Using 48 occupations



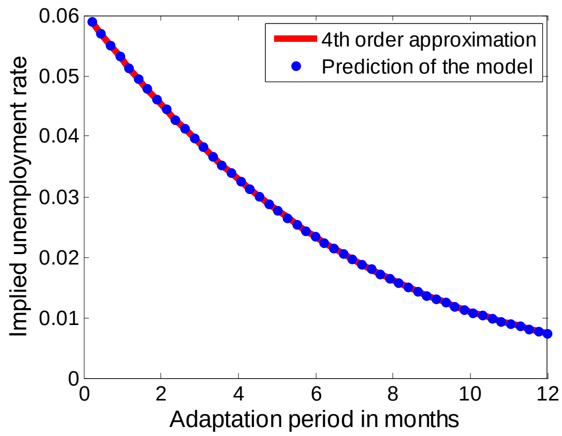
Panel B: Using 191 occupations



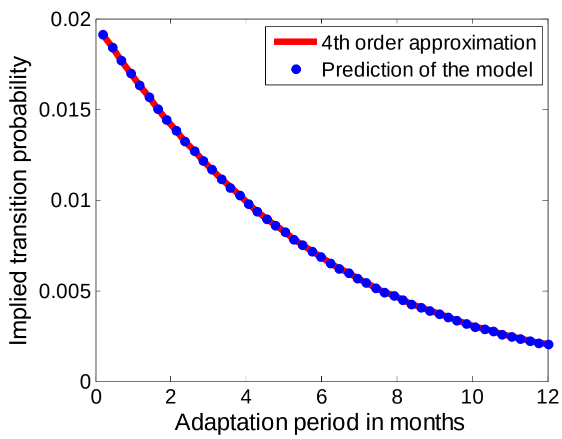
Sample: 25-55 year old males, CPS 1994-2008. Occupational unemployment is defined as $u_x = U_x/(E_x + U_x)$, where E_x is the number of employed persons with a primary job in occupation x; and U_x is the number of unemployed whose last occupation was in x. The approximations use the steady state formula $u_x^{approx} = p_x^{EU}/(p_x^{EU} + p_x^{UE})$, where p_x^{EU} is the monthly employment to unemployment transition probability in occupation x and p_x^{UE} is the monthly unemployment to employment transition probability for persons with the last occupation in x.

Figure 1.3: The effect of adaptation costs on unemployment rates and turnover probabilities, baseline specification

Panel A: Unemployment rates



Panel B: Monthly $E \rightarrow U$ probabilities



Panel C: Monthly $U \rightarrow E$ probabilities

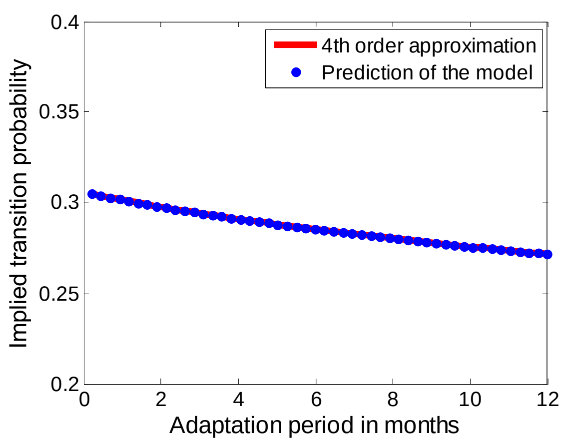
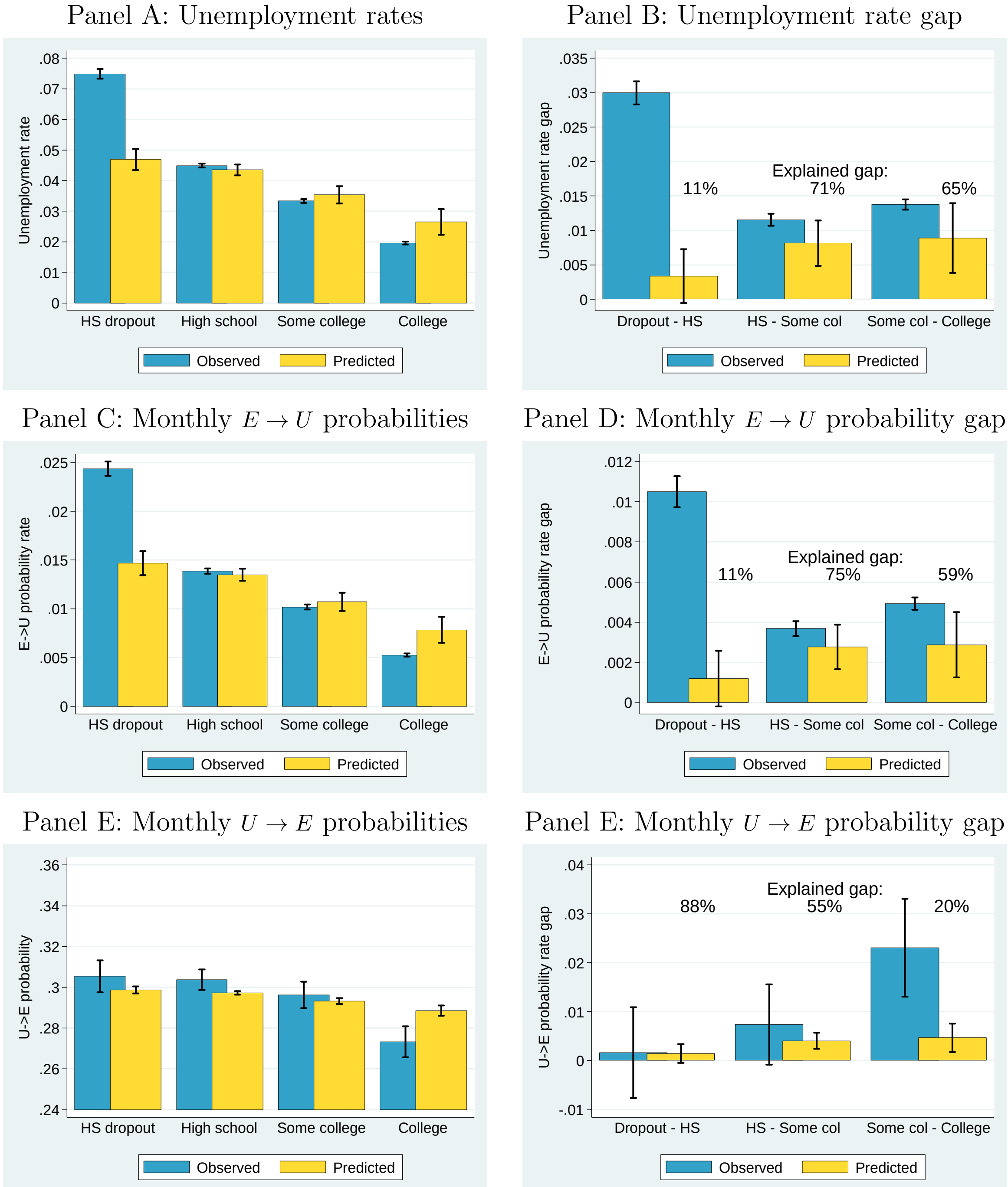


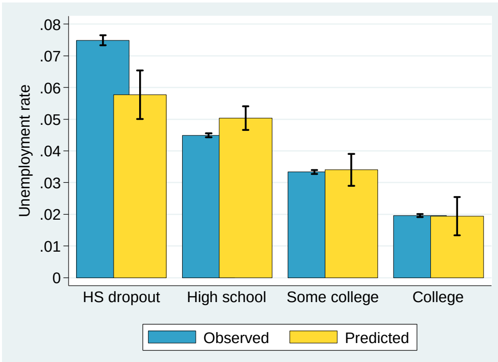
Figure 1.4: Observed and predicted unemployment rates and monthly transition probabilities by education, baseline specification



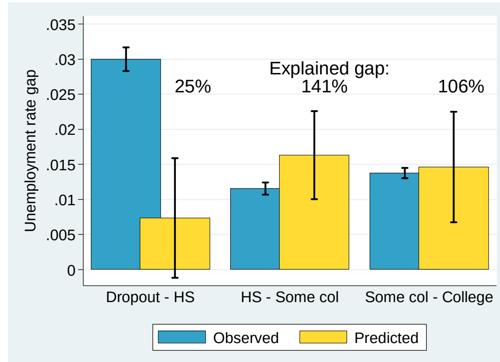
Sample: 25-55 year old males, CPS 1994-2008. The graph shows unemployment rates and monthly turnover probabilities by education from the CPS; and the model's predicted values. The model's predicted continuous transition rates are converted into monthly transition probabilities by assuming transitions follow a Markov process. The error bars indicate the 95% confidence intervals.

Figure 1.5: Observed and predicted unemployment rates by education, alternative values of autocorrelation

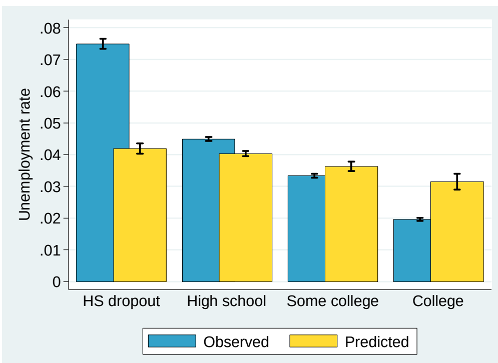
Panel A: Unemployment rates, $\exp(-\lambda) = 0.85$



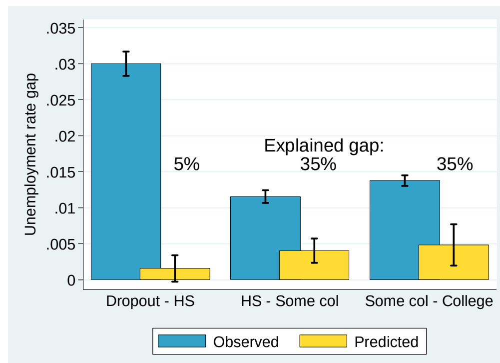
Panel B: Unemployment rate gap, $\exp(-\lambda) = 0.85$



Panel C: Unemployment rates, $\exp(-\lambda) = 0.95$

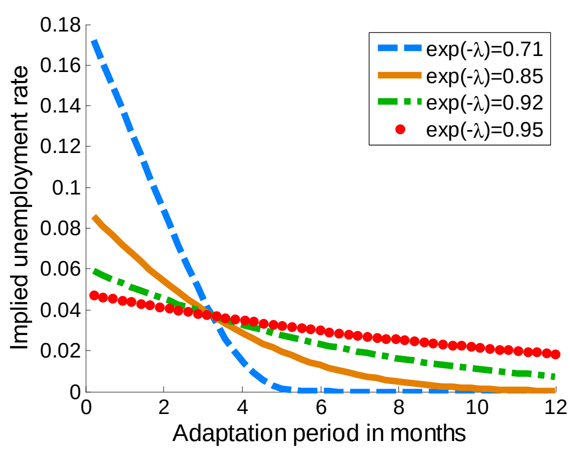


Panel D: Unemployment rate gap, $\exp(-\lambda) = 0.95$



Sample: 25-55 year old males, CPS 1994-2008. The graph shows unemployment rates and monthly turnover probabilities by education from the CPS; and the model's predicted values. The model's predicted continuous transition rates are converted into monthly transition probabilities by assuming transitions follow a Markov process. The error bars indicate the 95% confidence intervals.

Figure 1.6: The effect of adaptation costs on unemployment rates for various values of monthly autocorrelation in match-specific productivity



APPENDIX A
The value functions in the basic model

Let us consider the value functions of firms. I normalize all values with occupational productivity, s_x. Let $s_x V_x$, $s_x J_x^R(\varepsilon)$ and $s_x J_x(\varepsilon)$ denote the value of an unfilled job, being matched with a rookie and with an experienced worker with productivity ε, respectively. The Bellman equation for the value of vacancies (V) is

$$r s_x V_x = -s_x c + q_x s_x \left(J_x^R\left(\varepsilon^0\right) - V_x\right) = 0. \tag{1.31}$$

The flow value of a vacant job is the instantaneous search cost with a negative sign plus the expected capital gain, which is the value difference between employing a rookie with initial productivity ε^0 and having no worker, multiplied by the vacancy filling rate. Free entry of firms assures that the asset value of a vacant job is zero. The Bellman equation for the value of employing an experiences worker (J) is

$$r s_x J_x(\varepsilon) = s_x \varepsilon - s_x w_x(\varepsilon) + \lambda s_x \int_{\underline{\varepsilon}_x}^{\infty} J_x(\varepsilon')\, dF(\varepsilon') - \left(\lambda + f^{L,ex}\right) s_x J_x(\varepsilon). \tag{1.32}$$

The employer gets the flow profit minus the wage, plus the expected capital gains. Whenever a productivity shock or exogenous separation shock occurs, the firm loses the value of the current state and gains the value of the new state. If the productivity shock is larger than the productivity threshold $\underline{\varepsilon}_x$, the firm receives the value evaluated at the new productivity. If the shock is smaller than the reservation productivity or an exogenous separation occurs the firms' new value is $s_x V_x = 0$. The Bellman equation for employing a rookie $\left(J^R\right)$ is

$$\begin{aligned} r s_x J_x^R(\varepsilon) &= s_x \delta \varepsilon - s_x w_x^R(\varepsilon) + \lambda s_x \int_{\underline{\varepsilon}_x^R}^{\infty} J_x^R(\varepsilon') \, dF(\varepsilon') \\ &\quad - \left(\lambda + f^{L,ex}\right) s_x J_x^R(\varepsilon) + \varphi_x s_x \left(J_x(\varepsilon) - J_x^R(\varepsilon)\right). \end{aligned} \tag{1.33}$$

Apart from the last term, the Bellman equation for rookies is similar to that of experienced workers, but the flow productivity is smaller ($\delta < 1$ fraction of experienced workers' productivity) and the endogenous parameters (the wage, the value of the match and the productivity threshold) might differ, too. The last term describes the capital gain from experience as the difference in the values of being matched with an experienced worker and a rookie, multiplied by the rate at which rookies become experienced (φ_x).

Now let's consider the value functions of workers. Let $s_x U_x$, $s_x W_x^R(\varepsilon)$ and $s_x W_x(\varepsilon)$ denote the value of being unemployed, being a rookie and being an experienced worker, respectively. The Bellman equation for being unemployed (U) is

$$r s_x U_x = s_x b + f_x^F s_x \left(W_x^R\left(\varepsilon^0\right) - U_x\right). \tag{1.34}$$

The flow value of being unemployed is the instantaneous value $s_x b$, plus the capital gain from finding a job and becoming a rookie. The Bellman equation for being an experienced worker (W) is

$$r s_x W_x(\varepsilon) = s_x w_x(\varepsilon) + \lambda s_x \int_{\underline{\varepsilon}_x}^{\infty} W_x(\varepsilon') \, dF(\varepsilon') + \left(\lambda F(\underline{\varepsilon}_x) + f^{L,ex}\right) s_x U_x - \left(\lambda + f^{L,ex}\right) s_x W_x(\varepsilon). \tag{1.35}$$

Workers receive the flow wage and the expected capital gain from shocks. If the productivity shock is larger than the threshold value $\underline{\varepsilon}_x$, they gain the value of the new state and lose the value of the current state. If the productivity shock is smaller than the reservation threshold or an exogenous separation occurs, workers gain the value of being unemployed and lose the value of the current state. Finally, the Bellman equation for being a rookie (W^R) is

$$\begin{aligned} rs_x W_x^R(\varepsilon) &= s_x w_x^R(\varepsilon) + \lambda s_x \int_{\underline{\varepsilon}_x^R}^{\infty} W_x^R(\varepsilon')\,dF(\varepsilon') + \left(\lambda F\left(\underline{\varepsilon}_x^R\right) + f^{L,ex}\right) s_x U_x \\ &\quad - \left(\lambda + f^{L,ex}\right) s_x W_x^R(\varepsilon) + \varphi_x s_x \left(W_x(\varepsilon) - W_x^R(\varepsilon)\right). \end{aligned} \tag{1.36}$$

The value of being a rookie is similar to the value of being an experienced worker, but there is an extra term at the end that describes the capital gain from gaining experience.

Next, I normalize all equations by occupational productivity, s_x, and rearrange the value functions (1.31)-(1.36):

$$0 = -c + q_x J_x^R\left(\varepsilon^0\right), \tag{1.37}$$

$$\left(r + \lambda + f^{L,ex} + \varphi_x\right) J_x^R(\varepsilon) = \delta\varepsilon - w_x^R(\varepsilon) + \lambda \int_{\underline{\varepsilon}_x^R}^{\infty} J_x^R(\varepsilon')\,dF(\varepsilon') + \varphi_x J_x(\varepsilon), \tag{1.38}$$

$$\left(r + \lambda + f^{L,ex}\right) J_x(\varepsilon) = \varepsilon - w_x(\varepsilon) + \lambda \int_{\underline{\varepsilon}_x}^{\infty} J_x(\varepsilon')\,dF(\varepsilon'), \tag{1.39}$$

$$rU_x = b + f_x^F\left(W_x^R\left(\varepsilon^0\right) - U_x\right), \tag{1.40}$$

$$\begin{aligned} \left(r + \lambda + f^{L,ex} + \varphi_x\right) W_x^R(\varepsilon) &= w_x^R(\varepsilon) + \lambda \int_{\underline{\varepsilon}_x^R}^{\infty} W_x^R(\varepsilon')\,dF(\varepsilon') + \left(\lambda F\left(\underline{\varepsilon}_x^R\right) + f^{L,ex}\right) U_x \\ &\quad + \varphi_x W_x(\varepsilon), \end{aligned} \tag{1.41}$$

$$\left(r + \lambda + f^{L,ex}\right) W_x(\varepsilon) = w_x(\varepsilon) + \lambda \int_{\underline{\varepsilon}_x}^{\infty} W_x(\varepsilon')\,dF(\varepsilon') + \left(\lambda F\left(\underline{\varepsilon}_x\right) + f^{L,ex}\right) U_x. \tag{1.42}$$

Finally, to close the model, wages are continuously renegotiated and are set by generalized

Nash bargaining,

$$w_x^R(\varepsilon) = \arg\max\left\{\left(W_x^R(\varepsilon) - U_x^R(\varepsilon)\right)^\gamma J_x^R(\varepsilon)^{1-\gamma}\right\}, \tag{1.43}$$

$$w_x(\varepsilon) = \arg\max\left\{(W_x(\varepsilon) - U_x(\varepsilon))^\gamma J_x(\varepsilon)^{1-\gamma}\right\}. \tag{1.44}$$

The normalized value functions (1.37)-(1.42) together with the wage equations (1.43)-(1.44) and the matching function (1.13) fully characterize the model.

APPENDIX B
Derivation of the equilibrium conditions in the basic model

The surpluses from the matches can be written as

$$\begin{aligned}\left(r+\lambda+f^{L,ex}+\varphi_x\right)S_x^R(\varepsilon) &= \left(r+\lambda+f^{L,ex}+\varphi_x\right)\left(J_x^R(\varepsilon)+W_x^R(\varepsilon)-U_x\right)\\ &= \delta\varepsilon+\lambda\int_{\underline{\varepsilon}_x^R}^{\infty}S_x^N(\varepsilon')\,dF(\varepsilon')+\varphi_x S_x(\varepsilon)-rU_x, \qquad (1.45)\\ \left(r+\lambda+f^{L,ex}\right)S_x(\varepsilon) &= \left(r+\lambda+f^{L,ex}\right)\left(J_x(\varepsilon)+W_x(\varepsilon)-U_x\right)\\ &= \varepsilon+\lambda\int_{\underline{\varepsilon}_x}^{\infty}S_x(\varepsilon')\,dF(\varepsilon')-rU_x. \qquad (1.46)\end{aligned}$$

By taking the derivatives of both sides with respect to the idiosyncratic productivity ε we get

$$\frac{\partial S_x(\varepsilon)}{\partial\varepsilon} = \frac{1}{r+\lambda+f^{L,ex}}, \qquad (1.47)$$

$$\frac{\partial S_x^R(\varepsilon)}{\partial\varepsilon} = \frac{\delta+\varphi_x\frac{\partial S_x(\varepsilon)}{\partial\varepsilon}}{r+\lambda+f^{L,ex}+\varphi_x} = \frac{\delta+\frac{\varphi_x}{r+\lambda+f^{L,ex}}}{r+\lambda+f^{L,ex}+\varphi_x}. \qquad (1.48)$$

The surpluses are linear functions of the the match-specific productivity. Moreover, the surpluses equal to zero at the reservation productivity values $\left(\underline{\varepsilon}_x \text{ and } \underline{\varepsilon}_x^R\right)$. Thus, the surpluses can be written as

$$
\begin{aligned}
S_x(\varepsilon) &= \frac{1}{r+\lambda+f^{L,ex}}\left(\varepsilon-\underline{\varepsilon}_x\right), && (1.49)\\
S_x^R(\varepsilon) &= \frac{\delta+\frac{\varphi_x}{r+\lambda+f^{L,ex}}}{r+\lambda+f^{L,ex}+\varphi_x}\left(\varepsilon-\underline{\varepsilon}_x^R\right). && (1.50)
\end{aligned}
$$

By combining equation (1.15), (1.37) and (1.50) we get the job creation condition:

$$
\begin{aligned}
\frac{c}{\eta}\theta_x^{1-\alpha} &= J_x^R\left(\varepsilon^0\right)=(1-\gamma)\,S_x^R\left(\varepsilon^0\right)\\
&= (1-\gamma)\frac{\delta+\frac{\varphi_x}{r+\lambda+f^{L,ex}}}{r+\lambda+f^{L,ex}+\varphi_x}\left(\varepsilon^0-\underline{\varepsilon}_x^R\right). && (1.51)
\end{aligned}
$$

In the first line I used the Nash formula to rewrite the firms' value in terms of the surplus: workers and firms split the surplus according to their bargaining power.

In order to derive the job-destruction condition we need to write rU as a function of the surplus. I use equation (1.40), (1.14) and (1.15) to get:

$$
\begin{aligned}
rU_x &= b+f_x^F\left(W_x^R\left(\varepsilon^0\right)-U_x\right)=b+f_x^F\gamma S_x^R\left(\varepsilon^0\right)\\
&= b+\frac{\gamma}{1-\gamma}\frac{f_x^F}{q_x}c=b+\frac{\gamma}{1-\gamma}\theta_x c. && (1.52)
\end{aligned}
$$

To find the reservation productivity below which matches are endogenously destroyed, we need to find the productivity values at which the surplus becomes zero. By plugging (1.52) into (1.45) and (1.46) we get the job destruction conditions:

$$
\begin{aligned}
b+\frac{\gamma}{1-\gamma}\theta_x c &= \underline{\varepsilon}_x+\lambda\int_{\underline{\varepsilon}_x}^{\infty}S_x\left(\varepsilon'\right)dF\left(\varepsilon'\right), && (1.53)\\
b+\frac{\gamma}{1-\gamma}\theta_x c &= \delta\underline{\varepsilon}_x^R+\lambda\int_{\underline{\varepsilon}_x^R}^{\infty}S_x^R\left(\varepsilon'\right)dF\left(\varepsilon'\right)+\varphi_x S_x\left(\underline{\varepsilon}_x^R\right). && (1.54)
\end{aligned}
$$

By substituting in for the surplus formulas in (1.49) and (1.50) we get

$$
\begin{aligned}
b+\frac{\gamma}{1-\gamma}\theta_x c &= \underline{\varepsilon}_x+\frac{\lambda}{r+\lambda+f^{L,ex}}\int_{\underline{\varepsilon}_x}^{\infty}\left(\varepsilon'-\underline{\varepsilon}_x\right)dF\left(\varepsilon'\right), && (1.55)\\
b+\frac{\gamma}{1-\gamma}\theta_x c &= \delta\underline{\varepsilon}_x^R+\lambda\frac{\delta+\frac{\varphi_x}{r+\lambda+f^{L,ex}}}{r+\lambda+f^{L,ex}+\varphi_x}\int_{\underline{\varepsilon}_x^R}^{\infty}\left(\varepsilon'-\underline{\varepsilon}_x^R\right)dF\left(\varepsilon'\right)+\frac{\varphi_x}{r+\lambda+f^{L,ex}}\left(\underline{\varepsilon}_x^R-\underline{\varepsilon}_x\right) && (1.56)
\end{aligned}
$$

Now let us see the wages. I start with the wages of experienced workers. From (1.39) and the Nash rule we get

$$
\begin{aligned}
\left(r+\lambda+f^{L,ex}\right)J_x(\varepsilon) &= \varepsilon-w_x(\varepsilon)+\lambda\int_{\underline{\varepsilon}_x}^{\infty}J_x\left(\varepsilon'\right)dF\left(\varepsilon'\right), && (1.57)\\
\left(r+\lambda+f^{L,ex}\right)(1-\gamma)S_x(\varepsilon) &= \varepsilon-w_x(\varepsilon)+\lambda\int_{\underline{\varepsilon}_x}^{\infty}(1-\gamma)S_x\left(\varepsilon'\right)dF\left(\varepsilon'\right), && (1.58)\\
\left(r+\lambda+f^{L,ex}\right)S_x(\varepsilon) &= \frac{\varepsilon-w_x(\varepsilon)}{1-\gamma}+\lambda\int_{\underline{\varepsilon}_x}^{\infty}S_x\left(\varepsilon'\right)dF\left(\varepsilon'\right). && (1.59)
\end{aligned}
$$

Similarly from (1.35), (1.52) and the Nash rule we get

$$
\begin{aligned}
\left(r+\lambda+f^{L,ex}\right)W_x(\varepsilon) &= w_x(\varepsilon)+\lambda\int_{\underline{\varepsilon}_x}^{\infty}W_x\left(\varepsilon'\right)dF\left(\varepsilon'\right)+\left(\lambda F\left(\underline{\varepsilon}_x\right)+f^{L,ex}\right)U_x, && (1.60)\\
\left(r+\lambda+f^{L,ex}\right)\left(W_x(\varepsilon)-U_x\right) &= w_x(\varepsilon)+\lambda\int_{\underline{\varepsilon}_x}^{\infty}\left(W_x\left(\varepsilon'\right)-U_x\right)dF\left(\varepsilon'\right)-rU_x, && (1.61)\\
\left(r+\lambda+f^{L,ex}\right)\gamma S_x(\varepsilon) &= w_x(\varepsilon)+\lambda\int_{\underline{\varepsilon}_x}^{\infty}\gamma S_x(\varepsilon)\,dF\left(\varepsilon'\right)-b-\frac{\gamma}{1-\gamma}\theta_x c, && (1.62)\\
\left(r+\lambda+f^{L,ex}\right)S_x(\varepsilon) &= \frac{w_x(\varepsilon)-b}{\gamma}-\frac{1}{1-\gamma}\theta_x c+\lambda\int_{\underline{\varepsilon}_x}^{\infty}S_x\left(\varepsilon'\right)dF\left(\varepsilon'\right). && (1.63)
\end{aligned}
$$

By combining (1.59) and (1.63) we get

$$
\begin{aligned}
\frac{\varepsilon-w_x(\varepsilon)}{1-\gamma} &= \frac{w_x(\varepsilon)-b}{\gamma}-\frac{1}{1-\gamma}\theta_x c, && (1.64)\\
\gamma\left(\varepsilon-w_x(\varepsilon)\right) &= (1-\gamma)\left(w_x(\varepsilon)-b\right)-\gamma\theta_x c, && (1.65)\\
w_x(\varepsilon) &= (1-\gamma)b+\gamma\left(\varepsilon+\theta_x c\right). && (1.66)
\end{aligned}
$$

The last equation shows the equilibrium wage of experienced workers. The wages of rookies can be derived similarly. From (1.38) and the Nash rule

$$\begin{aligned}
\left(r+\lambda+f^{L,ex}+\varphi_x\right)J_x^R(\varepsilon) &= \delta\varepsilon-w_x^R(\varepsilon)+\lambda\int_{\underline{\varepsilon}_x^R}^{\infty}J_x^R(\varepsilon')\,dF(\varepsilon')+\varphi_x J_x(\varepsilon), && (1.67)\\
\left(r+\lambda+f^{L,ex}+\varphi_x\right)(1-\gamma)S_x^R(\varepsilon) &= \delta\varepsilon-w_x^R(\varepsilon)+\lambda\int_{\underline{\varepsilon}_x^R}^{\infty}(1-\gamma)S_x^R(\varepsilon')\,dF(\varepsilon')\\
&\quad+\varphi_x(1-\gamma)S_x(\varepsilon), && (1.68)\\
\left(r+\lambda+f^{L,ex}+\varphi_x\right)S_x^R(\varepsilon) &= \frac{\delta\varepsilon-w_x^R(\varepsilon)}{1-\gamma}+\lambda\int_{\underline{\varepsilon}_x^R}^{\infty}S_x^R(\varepsilon')\,dF(\varepsilon')+\varphi_x S_x(\varepsilon). && (1.69)
\end{aligned}$$

From (1.38), (1.52) and the Nash rule we get

$$\begin{aligned}
\left(r+\lambda+f^{L,ex}+\varphi_x\right)W_x^R(\varepsilon) &= w_x^R(\varepsilon)+\lambda\int_{\underline{\varepsilon}_x^R}^{\infty}W_x^R(\varepsilon')\,dF(\varepsilon')+\left(\lambda F\left(\underline{\varepsilon}_x^R\right)+f^{L,ex}\right)U_x\\
&\quad+\varphi_x W_x(\varepsilon), && (1.70)\\
\left(r+\lambda+f^{L,ex}+\varphi_x\right)\left(W_x^R(\varepsilon)-U_x\right) &= w_x^R(\varepsilon)+\lambda\int_{\underline{\varepsilon}_x^R}^{\infty}\left(W_x^R(\varepsilon')-U_x\right)dF(\varepsilon')\\
&\quad+\varphi_x\left(W_x(\varepsilon)-U_x\right)-rU_x, && (1.71)\\
\left(r+\lambda+f^{L,ex}+\varphi_x\right)\gamma S_x^R(\varepsilon) &= w_x^R(\varepsilon)-b-\frac{\gamma}{1-\gamma}\theta_x c+\lambda\int_{\underline{\varepsilon}_x^R}^{\infty}\gamma S_x^R(\varepsilon')\,dF(\varepsilon') && (1.72)\\
&\quad+\varphi_x\gamma S_x(\varepsilon), && (1.73)\\
\left(r+\lambda+f^{L,ex}+\varphi_x\right)S_x^R(\varepsilon) &= \frac{w_x^R(\varepsilon)-b}{\gamma}-\frac{1}{1-\gamma}\theta_x c+\lambda\int_{\underline{\varepsilon}_x^R}^{\infty}S_x^R(\varepsilon')\,dF(\varepsilon') && (1.74)\\
&\quad+\varphi_x S_x(\varepsilon). && (1.75)
\end{aligned}$$

By combining (1.69) and (1.74) we get

$$\begin{aligned}
\frac{\delta\varepsilon - w_x^R(\varepsilon)}{1-\gamma} &= \frac{w_x^R(\varepsilon) - b}{\gamma} - \frac{1}{1-\gamma}\theta_x c, && (1.76)\\
\gamma\left(\delta\varepsilon - w_x^R(\varepsilon)\right) &= (1-\gamma)\left(w_x^R(\varepsilon) - b\right) - \gamma\theta_x c, && (1.77)\\
w_x^R(\varepsilon) &= (1-\gamma)\, b + \gamma\left(\delta\varepsilon + \theta_x c\right). && (1.78)
\end{aligned}$$

APPENDIX C
Proof that the model is invariant to the choice of aggregate θ in Chapter I

θ appears at three places in the equilibrium conditions. The first is the job finding rate $f^F = \eta\theta^\alpha$. As the constant term of the matching function, η, is calibrated to fit the job finding rate, any change in θ is sucked up by η. The second place where θ appears is the job creation condition, where the left-hand side is the total expected cost of filling a vacancy:

$$C = \frac{c}{\eta}\theta^{1-\alpha}. \tag{1.79}$$

The total cost of filling a vacancy, however, is one of the moments I am matching. Therefore, c sucks up any variation in $\theta^{1-\alpha}/\eta$ to fit the total cost C. The third place where θ appears is the outside option of workers in the job loss conditions:

$$o = \frac{\gamma}{1-\gamma}\theta c = \frac{\gamma}{1-\gamma}\frac{C}{f^F}. \tag{1.80}$$

The outside option of workers can be rewritten in terms of only the total search costs (C) and the job finding rate (f^F), which are targeted moments. Therefore the outside option of workers is also indifferent to the choice of θ.

Overall, any mismeasurement of θ is sucked up by η and c, and the equilibrium turnover probabilities are unaffected by it.

APPENDIX D
Additional tables

Table 1.17: OLS regressions of unemployment in recessions

	Unemployment rate						
$\bar{U} = 0.065$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High school dropout	0.038**	0.038**	0.031**	0.026**	0.022**	0.021**	0.020**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High school	reference category						
Some college	-0.020**	-0.021**	-0.010**	-0.007**	-0.006**	-0.005**	-0.005**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
BA or more	-0.051**	-0.048**	-0.023**	-0.020**	-0.017**	-0.016**	-0.016**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 25-29	reference category						
Age 30-34		-0.008**	-0.007**	-0.006**	-0.005**	-0.005**	-0.005**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 35-39		-0.012**	-0.011**	-0.009**	-0.008**	-0.007**	-0.007**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 40-44		-0.017**	-0.015**	-0.013**	-0.011**	-0.010**	-0.010**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 45-49		-0.019**	-0.017**	-0.014**	-0.012**	-0.010**	-0.011**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Age 50-55		-0.020**	-0.017**	-0.014**	-0.011**	-0.010**	-0.010**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Whites	reference category						
Non whites		0.037**	0.036**	0.036**	0.035**	0.036**	0.036**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Not married	reference category						
Married		-0.047**	-0.047**	-0.045**	-0.043**	-0.043**	-0.042**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Quarter dummies	Y	Y	Y	Y	Y	Y	Y
Occupations, 10 categ/s			Y				
Occupations, 48 categ/s				Y			
Occupations, 191 categ/s					Y	Y	Y
Industries, 12 categ/s						Y	
Industries, 38 categ/s							Y
Observations	3,898,606	3,898,606	3,898,606	3,898,606	3,898,606	3,898,606	3,898,606

Sample: 25-55 year old males, CPS. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05. Recession indicates years when the aggregate unemployment rate was above 5 percent in the sample: 1980-1986, 1991-1993 and 2009-2013.

Table 1.18: Logit regressions of monthly employment to unemployment transition probability

	Log monthly $E \to U$ transition probability						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High school dropout	0.516**	0.537**	0.421**	0.338**	0.268**	0.240**	0.236**
	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
High school	reference category						
Some college	-0.316**	-0.319**	-0.129**	-0.081**	-0.062**	-0.054**	-0.050**
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
BA or more	-1.011**	-0.967**	-0.458**	-0.420**	-0.332**	-0.321**	-0.311**
	(0.013)	(0.013)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Age 25-29	reference category						
Age 30-34		-0.129**	-0.123**	-0.108**	-0.095**	-0.088**	-0.090**
		(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Age 35-39		-0.228**	-0.218**	-0.192**	-0.165**	-0.151**	-0.153**
		(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Age 40-44		-0.311**	-0.292**	-0.255**	-0.217**	-0.193**	-0.194**
		(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Age 45-49		-0.385**	-0.360**	-0.316**	-0.267**	-0.235**	-0.238**
		(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Age 50-55		-0.465**	-0.436**	-0.385**	-0.326**	-0.284**	-0.287**
		(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Whites	reference category						
Non whites		0.325**	0.325**	0.350**	0.331**	0.351**	0.355**
		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Not married	reference category						
Married		-0.545**	-0.540**	-0.504**	-0.466**	-0.454**	-0.453**
		(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Quarter dummies	Y	Y	Y	Y	Y	Y	Y
Occupations, 10 categ/s			Y				
Occupations, 48 categ/s				Y			
Occupations, 191 categ/s					Y	Y	Y
Industries, 12 categ/s						Y	
Industries, 38 categ/s							Y
Log likelihood	-417,334	-411,862	-407,189	-401,047	-397,659	-394,664	-393,962
Observations	6,064,820	6,064,820	6,064,820	6,064,820	6,064,820	6,064,820	6,064,820

Sample: 25-55 year old males, CPS 1978-2013. The table shows average log differences in the implied probabilities from baseline. For example, the coefficient on BA is $\beta_{BA}^{implied} = \frac{1}{N}\sum_i \left[\ln \Pr\left(P_i^{EU}=1|educ=BA,X_i\right) - \ln \Pr\left(P_i^{EU}=1|educ=HS,X_i\right)\right]$. More generally, in dimension j and category k the coefficient is $\beta_{jk}^{implied} = \frac{1}{N}\sum_i \left[\ln \Pr\left(P_i^{EU}=1|I_j=k,X_i\right) - \ln \Pr\left(P_i^{EU}=1|I_j=0,X_i\right)\right]$. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05.

Table 1.19: OLS regressions of monthly employment to unemployment transition probability in booms

	Monthly $E \rightarrow U$ transition probability						
$p^{\bar{E}U} = 0.013$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High school dropout	0.010**	0.010**	0.009**	0.007**	0.006**	0.005**	0.005**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
High school	reference category						
Some college	-0.004**	-0.004**	-0.001**	-0.001**	-0.001**	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BA or more	-0.009**	-0.008**	-0.003**	-0.003**	-0.002**	-0.002**	-0.002**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 25-29	reference category						
Age 30-34		-0.002**	-0.002**	-0.002**	-0.001**	-0.001**	-0.001**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 35-39		-0.003**	-0.003**	-0.003**	-0.002**	-0.002**	-0.002**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 40-44		-0.004**	-0.004**	-0.003**	-0.003**	-0.002**	-0.002**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 45-49		-0.005**	-0.004**	-0.004**	-0.003**	-0.003**	-0.003**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 50-55		-0.005**	-0.005**	-0.004**	-0.003**	-0.003**	-0.003**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Whites	reference category						
Non whites		0.005**	0.005**	0.005**	0.005**	0.005**	0.005**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Not married	reference category						
Married		-0.008**	-0.007**	-0.007**	-0.006**	-0.006**	-0.006**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Quarter dummies	Y	Y	Y	Y	Y	Y	Y
Occupations, 10 categ/s			Y				
Occupations, 48 categ/s				Y			
Occupations, 191 categ/s					Y	Y	Y
Industries, 12 categ/s						Y	
Industries, 38 categ/s							Y
Observations	3,469,345	3,469,345	3,469,345	3,469,345	3,469,345	3,469,345	3,469,345

Sample: 25-55 year old males, CPS. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05. Boom indicates years when the aggregate unemployment rate was below 5 percent in the sample: 1978-1979, 1987-1990 and 1994-2008.

Table 1.20: OLS regressions of monthly employment to unemployment transition probability in recessions

	Monthly $E \rightarrow U$ transition probability						
$p^{\bar{E}U} = 0.018$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High school dropout	0.012**	0.012**	0.010**	0.008**	0.006**	0.006**	0.006**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
High school	reference category						
Some college	-0.006**	-0.006**	-0.002**	-0.002**	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BA or more	-0.013**	-0.012**	-0.005**	-0.005**	-0.004**	-0.003**	-0.003**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 25-29	reference category						
Age 30-34		-0.003**	-0.003**	-0.003**	-0.003**	-0.003**	-0.003**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 35-39		-0.005**	-0.005**	-0.004**	-0.004**	-0.004**	-0.004**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 40-44		-0.006**	-0.006**	-0.005**	-0.005**	-0.004**	-0.004**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 45-49		-0.007**	-0.007**	-0.006**	-0.005**	-0.005**	-0.005**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age 50-55		-0.009**	-0.008**	-0.007**	-0.006**	-0.006**	-0.006**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Whites	reference category						
Non whites		0.006**	0.006**	0.006**	0.006**	0.006**	0.006**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Not married	reference category						
Married		-0.009**	-0.009**	-0.008**	-0.008**	-0.007**	-0.007**
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Quarter dummies	Y	Y	Y	Y	Y	Y	Y
Occupations, 10 categ/s			Y				
Occupations, 48 categ/s				Y			
Occupations, 191 categ/s					Y	Y	Y
Industries, 12 categ/s						Y	
Industries, 38 categ/s							Y
Observations	2,595,475	2,595,475	2,595,475	2,595,475	2,595,475	2,595,475	2,595,475

Sample: 25-55 year old males, CPS. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05. Recession indicates years when the aggregate unemployment rate was above 5 percent in the sample: 1980-1986, 1991-1993 and 2009-2013.

Table 1.21: Logit regressions of monthly unemployment to employment transition probability

	Log monthly $U \rightarrow E$ transition probability						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High school dropout	0.038**	0.034**	0.024*	0.008	0.005	0.002	0.000
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
High school	reference category						
Some college	-0.001	0.004	0.037**	0.045**	0.046**	0.048**	0.049**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)
BA or more	-0.047**	-0.047**	0.036**	0.044**	0.054**	0.059**	0.064**
	(0.012)	(0.012)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Age 25-29	reference category						
Age 30-34		-0.039**	-0.045**	-0.046**	-0.049**	-0.050**	-0.051**
		(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Age 35-39		-0.088**	-0.096**	-0.098**	-0.101**	-0.101**	-0.102**
		(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Age 40-44		-0.119**	-0.126**	-0.125**	-0.128**	-0.128**	-0.129**
		(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Age 45-49		-0.169**	-0.174**	-0.171**	-0.175**	-0.174**	-0.175**
		(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Age 50-55		-0.255**	-0.258**	-0.254**	-0.256**	-0.255**	-0.256**
		(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Whites	reference category						
Non whites		-0.245**	-0.232**	-0.223**	-0.220**	-0.218**	-0.218**
		(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Not married	reference category						
Married		0.197**	0.191**	0.195**	0.193**	0.195**	0.195**
		(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Quarter dummies	Y	Y	Y	Y	Y	Y	Y
Occupations, 10 categ/s			Y				
Occupations, 48 categ/s				Y			
Occupations, 191 categ/s					Y	Y	Y
Industries, 12 categ/s						Y	
Industries, 38 categ/s							Y
Log likelihood	-174,504	-173,168	-172,751	-172,100	-171,819	-171,628	-171,459
Observations	295,045	295,045	295,045	295,045	295,045	295,045	295,045

Sample: 25-55 year old males, CPS 1978-2013. The table shows average log differences in the implied probabilities from baseline. For example, the coefficient on BA is $\beta_{BA}^{implied} = \frac{1}{N}\sum_i \left[\ln \Pr\left(P_i^{UE} = 1 | educ = BA, X_i\right) - \ln \Pr\left(P_i^{UE} = 1 | educ = HS, X_i\right)\right]$. More generally, in dimension j and category k the coefficient is $\beta_{jk}^{implied} = \frac{1}{N}\sum_i \left[\ln \Pr\left(P_i^{EU} = 1 | I_j = k, X_i\right) - \ln \Pr\left(P_i^{EU} = 1 | I_j = 0, X_i\right)\right]$. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05.

Table 1.22: OLS regressions of monthly unemployment to employment transition probability in booms

	Monthly $U \to E$ transition probability						
$p^{\bar{U}E} = 0.33$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High school dropout	0.004	0.002	-0.002	-0.005	-0.005	-0.006	-0.006
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
High school	reference category						
Some college	-0.005	-0.003	0.007	0.010*	0.010*	0.010*	0.011**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
BA or more	-0.027**	-0.027**	-0.000	0.003	0.005	0.007	0.009
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Age 25-29	reference category						
Age 30-34		-0.015**	-0.017**	-0.017**	-0.018**	-0.018**	-0.018**
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Age 35-39		-0.030**	-0.033**	-0.033**	-0.033**	-0.033**	-0.033**
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Age 40-44		-0.040**	-0.042**	-0.041**	-0.042**	-0.042**	-0.042**
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Age 45-49		-0.056**	-0.058**	-0.056**	-0.056**	-0.055**	-0.055**
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Age 50-55		-0.086**	-0.087**	-0.084**	-0.084**	-0.083**	-0.083**
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Whites	reference category						
Non whites		-0.070**	-0.067**	-0.064**	-0.062**	-0.062**	-0.062**
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Not married	reference category						
Married		0.058**	0.057**	0.058**	0.057**	0.057**	0.057**
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Quarter dummies	Y	Y	Y	Y	Y	Y	Y
Occupations, 10 categ/s			Y				
Occupations, 48 categ/s				Y			
Occupations, 191 categ/s					Y	Y	Y
Industries, 12 categ/s						Y	
Industries, 38 categ/s							Y
Observations	129,968	129,968	129,968	129,968	129,968	129,968	129,968

Sample: 25-55 year old males, CPS. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05. Boom indicates years when the aggregate unemployment rate was below 5 percent in the sample: 1978-1979, 1987-1990 and 1994-2008.

Table 1.23: OLS regressions of monthly unemployment to employment transition probability in recessions

	Monthly $U \to E$ transition probability						
$p^{\bar{U}E} = 0.26$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
High school dropout	0.015**	0.015**	0.013**	0.008*	0.006	0.005	0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
High school	reference category						
Some college	0.003	0.005	0.012**	0.013**	0.013**	0.013**	0.013**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
BA or more	0.002	0.002	0.018**	0.019**	0.021**	0.022**	0.022**
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Age 25-29	reference category						
Age 30-34		-0.007	-0.008	-0.008	-0.008*	-0.008*	-0.009*
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Age 35-39		-0.018**	-0.019**	-0.020**	-0.021**	-0.021**	-0.021**
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Age 40-44		-0.025**	-0.026**	-0.025**	-0.026**	-0.026**	-0.026**
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Age 45-49		-0.034**	-0.035**	-0.035**	-0.036**	-0.036**	-0.036**
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Age 50-55		-0.048**	-0.049**	-0.047**	-0.048**	-0.048**	-0.048**
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Whites	reference category						
Non whites		-0.053**	-0.049**	-0.046**	-0.046**	-0.045**	-0.044**
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Not married	reference category						
Married		0.046**	0.043**	0.044**	0.044**	0.044**	0.044**
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Quarter dummies	Y	Y	Y	Y	Y	Y	Y
Occupations, 10 categ/s			Y				
Occupations, 48 categ/s				Y			
Occupations, 191 categ/s					Y	Y	Y
Industries, 12 categ/s						Y	
Industries, 38 categ/s							Y
Observations	165,077	165,077	165,077	165,077	165,077	165,077	165,077

Sample: 25-55 year old males, CPS. Robust standard errors clustered on the individual level are in parentheses. ** p<0.01, * p<0.05. Recession indicates years when the aggregate unemployment rate was above 5 percent in the sample: 1980-1986, 1991-1993 and 2009-2013.

Table 1.24: Observed and predicted unemployment rates and turnover probabilities by education, baseline specification

	Unemployment rate		$E \rightarrow U$ probability		$U \rightarrow E$ probability	
	Observed	Predicted	Observed	Predicted	Observed	Predicted
High school dropout, HSD	0.0749**	0.0469**	0.0244**	0.0147**	0.3054**	0.2987**
	(0.0008)	(0.0018)	(0.0004)	(0.0006)	(0.0040)	(0.0009)
High school, HS	0.0449**	0.0435**	0.0139**	0.0135**	0.3037**	0.2973**
	(0.0003)	(0.0009)	(0.0001)	(0.0003)	(0.0026)	(0.0004)
Some college, SC	0.0334**	0.0354**	0.0102**	0.0107**	0.2963**	0.2932**
	(0.0003)	(0.0014)	(0.0001)	(0.0005)	(0.0033)	(0.0007)
BA or more, BA	0.0196**	0.0265**	0.0053**	0.0078**	0.2732**	0.2885**
	(0.0002)	(0.0021)	(0.0001)	(0.0007)	(0.0039)	(0.0013)
Differences						
HSD-HS	0.0300**	0.0034	0.0105**	0.0012	0.0017	0.0015
	(0.0009)	(0.002)	(0.0004)	(0.0007)	(0.0047)	(0.001)
HS-SC	0.0115**	0.0082**	0.0037**	0.0028**	0.0074	0.0041**
	(0.0004)	(0.0017)	(0.0002)	(0.0006)	(0.0042)	(0.0008)
HS-BA	0.0253**	0.017**	0.0086**	0.0056**	0.0305**	0.0087**
	(0.0004)	(0.0023)	(0.0002)	(0.0008)	(0.0047)	(0.0014)
SC-BA	0.0138**	0.0089**	0.0049**	0.0029**	0.0231**	0.0047**
	(0.0004)	(0.0026)	(0.0002)	(0.0008)	(0.0051)	(0.0015)
Fraction of explained gap						
HSD-HS		0.1126		0.1136		0.8774
HS-SC		0.7067		0.7522		0.5475
HS-BA		0.6733		0.6559		0.2868
SC-BA		0.6454		0.5840		0.2032

** $p<0.01$, * $p<0.05$. Sample: 1994-2008, 25-55 year old males. The odd columns show unemployment rates and monthly turnover probabilities by education from the CPS with robust standard errors clustered on the individual level. The even columns show the model's predicted values and the fraction of the gap between education groups explained by adaptation costs. The model's predicted continuous transition rates are converted into monthly transition probabilities by assuming transitions follow a Markov process.

Table 1.25: Imputed adaptation period by detailed occupations, in months

#	Occupation	Imputation 1	Imputation 2	MCSUI avg	# in MCSUI
1	Waiter's assistant	1.154	1.086	1.654	8
2	Laundry workers	1.177	0.937	1.482	12
3	Metal and plastic processing machine operators	1.275	1.255	2.811	3
4	Crane, derrick, winch, and hoist operators	1.284	1.125	-	0
5	Waiter/waitress, food counter and fountain workers	1.348	1.339	1.342	64
6	Interviewers, enumerators, and surveyors	1.361	1.398	0.941	1
7	Recreation facility attendants	1.362	1.139	2.409	3
8	Vehicle washers and equipment cleaners	1.376	1.352	0.670	2
9	Misc food prep workers	1.380	1.645	1.197	33
10	Textile sewing machine operators	1.410	1.331	1.664	2
11	Janitors	1.428	1.225	1.713	52
12	Mail carriers for postal service	1.448	1.375	1.258	12
13	Cashiers	1.479	1.317	0.863	72
14	Stock and inventory clerks	1.518	1.717	1.544	48
15	Precision woodworking occupations	1.588	1.659	8.327	4
16	Woodworking machine operators	1.620	1.496	10.784	4
17	Auto body repairers	1.627	1.773	6.041	2
18	Other textile, apparel, and furnishings operators	1.653	1.656	3.796	4
19	Postal clerks, excluding mail carriers	1.663	1.649	0.471	1
20	Photographers	1.690	2.309	12.405	2
21	Kitchen workers	1.699	1.480	1.514	5
22	Butchers and meat cutters	1.705	1.719	15.278	2
23	Fishers, hunters, and kindred	1.718	2.156	1.548	3
	⋮				
169	Farm operators and managers	5.388	5.197	5.647	2
170	Registered nurses	5.393	5.038	4.874	29
171	Physical scientists	5.416	6.011	12.235	1
172	Vocational and educational counselors	5.538	5.965	4.893	10
173	Human resource, marketing, advertising managers	5.558	4.841	4.593	16
174	Managers of properties and real estate	5.646	6.170	2.612	3
175	Other managers	5.679	5.468	3.668	68
176	Secondary school teachers	5.750	6.632	12.144	2
177	Social workers	5.815	5.647	6.943	16
178	Managers in education and related fields	6.040	5.997	15.436	7
179	Financial services sales occupations	6.044	6.435	3.747	3
180	Special education teachers	6.120	6.911	4.440	3
181	Management support occupations	6.140	6.023	2.779	33
182	Postsecondary teachers	6.190	6.441	10.948	26
183	Primary school teachers	6.205	7.832	14.146	17
184	Life scientists	6.378	6.475	-	0
185	Psychologists	6.381	5.937	-	0
186	Financial managers	6.566	6.457	10.261	3
187	Other social scientists and urban planners	6.566	6.500	2.847	2
188	Lawyers and Judges	6.697	5.669	10.786	3
189	Managers of medicine and health occupations	6.858	6.468	6.340	2
190	Management analysts	7.287	6.642	2.824	1
191	Clergy and religious workers	8.683	9.189	14.454	8

All models are based on a fitted regression of log adaptation period on occupational skills or occupation dummies. The table shows the exponent of the predicted values of those regressions. Imputation 1 and 2 are based occupational work activity measures reported as Model 2 and 3 of Table 1.5 respectively. The MCSUI average uses occupation dummies. The table only shows occupations with the 23 shortest and 23 longest adaptation periods based on Imputation 1 out of 191 occupations. The rest of the table is available on my wedsite at https://sites.google.com/site/phudomiet/research.

Table 1.26: Observed and predicted unemployment rates and turnover probabilities by education, specification with lower autocorrelation

	Unemployment rate		$E \to U$ probability		$U \to E$ probability	
	Observed	Predicted	Observed	Predicted	Observed	Predicted
High school dropout, HSD	0.0749**	0.0577**	0.0244**	0.0183**	0.3054**	0.2984**
	(0.0008)	(0.0039)	(0.0004)	(0.0014)	(0.0040)	(0.0008)
High school, HS	0.0449**	0.0504**	0.0139**	0.0157**	0.3037**	0.297**
	(0.0003)	(0.0019)	(0.0001)	(0.0007)	(0.0026)	(0.0004)
Some college, SC	0.0334**	0.034**	0.0102**	0.0103**	0.2963**	0.2933**
	(0.0003)	(0.0026)	(0.0001)	(0.0008)	(0.0033)	(0.0007)
BA or more, BA	0.0196**	0.0194**	0.0053**	0.0058**	0.2732**	0.289**
	(0.0002)	(0.0031)	(0.0001)	(0.001)	(0.0039)	(0.0012)
Differences						
HSD-HS	0.0300**	0.0074	0.0105**	0.0026	0.0017	0.0014
	(0.0009)	(0.0043)	(0.0004)	(0.0015)	(0.0047)	(0.0009)
HS-SC	0.0115**	0.0163**	0.0037**	0.0054**	0.0074	0.0037**
	(0.0004)	(0.0032)	(0.0002)	(0.001)	(0.0042)	(0.0008)
HS-BA	0.0253**	0.031**	0.0086**	0.0099**	0.0305**	0.008**
	(0.0004)	(0.0036)	(0.0002)	(0.0012)	(0.0047)	(0.0013)
SC-BA	0.0138**	0.0146**	0.0049**	0.0046**	0.0231**	0.0043**
	(0.0004)	(0.004)	(0.0002)	(0.0013)	(0.0051)	(0.0014)
Fraction of explained gap						
HSD-HS		0.2459		0.2436		0.8233
HS-SC		1.4137		1.4628		0.5055
HS-BA		1.2228		1.1547		0.2639
SC-BA		1.0628		0.9243		0.1865

** $p<0.01$, * $p<0.05$. Sample: 1994-2008, 25-55 year old males. The odd columns show unemployment rates and monthly turnover probabilities by education from the CPS with robust standard errors clustered on the individual level. The even columns show the model's predicted values and the fraction of the gap between education groups explained by adaptation costs. The model's predicted continuous transition rates are converted into monthly transition probabilities by assuming transitions follow a Markov process. In these specifications the monthly autocorrelation in match-specific productivity shocks is assumed to be $\exp(-\lambda) = 0.85$.

Table 1.27: Observed and predicted unemployment rates and turnover probabilities by education, specification with higher autocorrelation

	Unemployment rate		$E \to U$ probability		$U \to E$ probability	
	Observed	Predicted	Observed	Predicted	Observed	Predicted
High school dropout, HSD	0.0749**	0.0419**	0.0244**	0.0131**	0.3054**	0.2988**
	(0.0008)	(0.0008)	(0.0004)	(0.0003)	(0.0040)	(0.0009)
High school, HS	0.0449**	0.0403**	0.0139**	0.0125**	0.3037**	0.2973**
	(0.0003)	(0.0004)	(0.0001)	(0.0002)	(0.0026)	(0.0004)
Some college, SC	0.0334**	0.0363**	0.0102**	0.011**	0.2963**	0.2932**
	(0.0003)	(0.0007)	(0.0001)	(0.0003)	(0.0033)	(0.0008)
BA or more, BA	0.0196**	0.0315**	0.0053**	0.0093**	0.2732**	0.2883**
	(0.0002)	(0.0013)	(0.0001)	(0.0004)	(0.0039)	(0.0013)
Differences						
HSD-HS	0.0300**	0.0016	0.0105**	0.0006	0.0017	0.0015
	(0.0009)	(0.0009)	(0.0004)	(0.0004)	(0.0047)	(0.001)
HS-SC	0.0115**	0.004**	0.0037**	0.0015**	0.0074	0.0041**
	(0.0004)	(0.0009)	(0.0002)	(0.0003)	(0.0042)	(0.0009)
HS-BA	0.0253**	0.0089**	0.0086**	0.0031**	0.0305**	0.009**
	(0.0004)	(0.0013)	(0.0002)	(0.0005)	(0.0047)	(0.0014)
SC-BA	0.0138**	0.0048**	0.0049**	0.0017**	0.0231**	0.0049**
	(0.0004)	(0.0015)	(0.0002)	(0.0005)	(0.0051)	(0.0015)
Fraction of explained gap						
HSD-HS		0.0528		0.0565		0.8822
HS-SC		0.3487		0.3955		0.5575
HS-BA		0.3498		0.3638		0.2947
SC-BA		0.3507		0.3401		0.2104

** $p<0.01$, * $p<0.05$. Sample: 1994-2008, 25-55 year old males. The odd columns show unemployment rates and monthly turnover probabilities by education from the CPS with robust standard errors clustered on the individual level. The even columns show the model's predicted values and the fraction of the gap between education groups explained by adaptation costs. The model's predicted continuous transition rates are converted into monthly transition probabilities by assuming transitions follow a Markov process. In these specifications the monthly autocorrelation in match-specific productivity shocks is assumed to be $\exp(-\lambda) = 0.95$.

CHAPTER II

Career Interruptions and Measurement Error in Annual Earnings

2.1 Introduction

Even though it is long known that survey data are prone to measurement error, there is an equally long tradition in applied economics to use survey data as if it were exactly measured. With better models of survey response error, researchers should be able to use survey data in ways that take into account potential biases in applied work. Validation studies comparing survey reports to highly accurate records are helpful in achieving this goal. In the classical measurement error (CME) model, for example, only an estimate of the noise to total variance ratio is required to adjust for the attenuation bias. Validation studies can provide estimates of this ratio, and applied researchers can use them to correct bias even when highly accurate records are not readily available. Validation studies can also be used to test the assumptions of the CME model, and if needed, suggest more appropriate models.

This paper aims at characterizing different sources of measurement error in survey reports of annual earnings of employees using a long biannual panel dataset, the Health and Retirement Study, which is linked to the Master Earnings File of the Social Security Administration. This administrative data are based on information from the W2 forms of individuals. The main novelty of this paper is to scrutinize the reporting of partial year earnings in the survey. Partial year earnings are identified from the administrative data by making use of employer identifiers. I create a complete labor history of each employee by tracking newly appearing and disappearing employers. I then compare recorded and reported earnings in years people experienced career interruptions.

I show that partial year earnings have a strong and non-standard effect on survey response error. When people report their annual earnings in surveys, they tend to report the permanent component of their earnings relatively accurately, but systematically under-report transitory shocks (Pischke, 1995). People might under-report transitory shocks because it is cognitively demanding to recall their exact values, or some might prefer reporting their usual earnings in surveys rather than their actual earnings. Transitory earnings shocks, however, are typically small. The fact that people do not report their unusual tips, bonuses or compensations for overtime work accurately makes little difference. Transitory earnings shocks due to career interruptions, however, are large, equivalent to several months of lost earnings, on average. In my preferred 50-60-year-old sample around 12.5 percent of person-years involve either a job loss or a job finding, and respondents only report $60-85\%$ of the earnings losses, on average. In years with job loss, many people report zero earnings in HRS. Those who report a positive value report around 85% of the earnings losses. In years with a new job, persons report even less accurately: around 40% of these earnings shocks remain unreported. It appears that many individuals report their annualized earnings after finding a new job and forget to report the earnings loss due to not working the entire year. Career interruptions have a strong effect on the variance of measurement error, too. Interviewees have less problem recalling their annual earnings when nothing interesting happens to their employment relationship, but the task becomes substantially harder after such large shocks.

Measurement error due to non-employment is not only substantial, but it has non-standard properties. Researchers typically think of measurement error as pure noise: supposedly it has zero mean, perhaps it is symmetric, perhaps even normally distributed, and it does not correlate with other variables. None of these properties hold for measurement error due to career interruptions. First, the error has a small positive mean, because persons under-report shocks that are negative. Second, the error is asymmetric and skewed to the right, because persons under-report shocks with negative skewness. Third and most importantly, the error correlates with predictors of turnover. Higher turnover decreases annual earnings, because individuals spend less time working. The negative effect of turnover on earnings, however, is muted in the survey data, as it is under-reported on average.

To identify under-reporting of transitory shocks I use two alternative estimation strategies. My preferred method is a fully specified dynamic earnings and measurement error model that is estimated by Markov Chain Monte Carlo (MCMC). This model incorporates permanent and transitory earnings, permanent and transitory measurement error, labor histories, and it deals with missing values and interval earnings responses in a multiple imputation fashion. Permanent earnings follow a random walk (similarly to MaCurdy, 1982), but the mean and the variance of the random walk innovation are allowed to depend on labor histories. Transitory earnings on-the-job follow a low order moving average process. The transitory earnings loss due to career interruptions follows an exponential distribution, which will be shown to approximate these shocks well. The data includes many missing values that exist due to non-response in the HRS, non-consent in the administrative data, or because earnings self-reports are not available in the off years of the survey. Moreover, a large fraction of the earnings data is only available in intervals. Dealing with missing and interval data in panel models is difficult using standard techniques, such as maximum likelihood or GMM. As a methodological contribution, I show that it is quite straightforward to model and estimate them with MCMC. However, MCMC not yet being a standard tool in economics, I shall provide simple descriptive evidence to support the main claims of the paper.

This paper is most closely related to the literature on measurement error in earnings. A handbook chapter on the subject by John Bound, Charles Brown and Nancy Mathiowetz (2001) provides a thorough overview of early developments in this literature. Researchers consistently found that the mean of the error is small, its variance is large, and the estimated reliability ratio in log earnings, within the framework of the CME model, is between 0.6 and 0.9.[21] The second robust finding was that there is a substantial negative correlation between measurement error and true earnings, and thus the CME model does not appear to be a suitable model of measurement error in earnings.[22] This property is known as the "mean reverting property of the error," as the negative correlation implies that low income people tend to report too high and high income people tend to report too low earnings in surveys. The drive behind this mean reverting property has yet to be identified. It might be that persons under-report transitory shocks in earnings (Pischke, 1995); it is possible that there are false matches between the survey and administrative data (Kapteyn and Ypma, 2007); or there might be error in the administrative data, which can create a spurious mean reversion (Abowd and Stinson, 2013). My paper is a simple extension of the Pischke (1995) mechanism with career interruptions. I shall briefly discuss other mechanisms in Section 2.2.

There is another property of measurement error that has received much less attention in the

[21] Attenuation bias is even larger in log changes in earnings, as earnings are more persistent than the error.

[22] The classical measurement error model assumes that the error is independent of the true value.

literature. Using a non-parametric estimation strategy Bollinger (1998) found that mean-reversion is mostly driven by low earning values in the Current Population Survey: mean measurement error is significantly and substantially positive at the low end of the earnings distribution, but at medium and high earnings measurement error becomes mean-zero. In other words, it becomes more classical. Kristensen and Westergaard-Nielsen (2006) found the same result using Danish data. In my sample of 50-60-year-old employees, I confirm the Bollinger result: below $\$10,000$ per year measurement error is strongly upward biased, but above this threshold the error is basically mean zero, and there is practically no mean reversion.[23] I also show that these problematic low earnings values are usually associated with career interruptions. My model, thus, is able to explain all the documented properties of measurement error: under-reporting transitory earnings shocks leads to mean reversion, and mean-reversion is strongest at the low end of the earnings distribution because those values are typically associated with large negative shocks due to non-employment.

My paper is also related to the literature on earnings dynamics. Guvenen et al. (2015), using earnings data on millions of US workers, found that earnings growth has substantial excess kurtosis compared to a simple log-normal model. Career interruptions naturally predict excess kurtosis, as persons receive large negative earnings shocks with a small probability when they lose their jobs, and they receive a large positive shock after they find a new one. This paper also has implications for the literature on displaced workers' earnings. Papers using survey based measures of earnings, such as the Displaced Worker Supplement of the Current Population Survey or the Panel Study of Income Dynamics, typically find that the initial earnings loss after displacement averages $10-20\%$ (Ruhm, 1991; Carrington and Zaman, 1994; Fallick, 1996; Stevens, 1997; Kletzer, 1998). Papers using administrative

[23]Earnings are adjusted by the CPI to 2000 dollars.

earnings records, however, usually find much larger negative effects (Jacobson et al., 1993; Couch and Placzek, 2010). One explanation for these differences is that displacement is defined differently in surveys and in administrative data (Flaaen et al., 2013). My paper suggests an alternative explanation: people might under-report earnings losses in surveys.

The paper is organized as follows. Section 2 introduces a simple model that features career interruptions and measurement error. Section 3 describes the data and basic properties of measurement error. Section 4 carries out the estimation of the dynamic earnings and error model by MCMC. Section 5 discusses implications for bias correction and Section 6 concludes.

2.2 Earnings and measurement error models

2.2.1 A simple model with labor histories

Imagine that earnings have a permanent and a transitory component.

$$Y_i^* = Y_i^{*,perm} \times Y_i^{*,trans}. \tag{2.1}$$

After taking the natural logarithm of both sides, log earnings are

$$y_i^* = y_i^{*,perm} + y_i^{*,trans}. \tag{2.2}$$

Transitory earnings are the sum of a pure noise and a shock term due to non-employment:

$$y_i^{*,trans} = \varepsilon_i - I_i k_i, \tag{2.3}$$

where I_i indicates if person i does not work the entire year, and $k_i > 0$ is earnings loss due to non-employment. Let us assume that all random variables are uncorrelated and their first and second moments are

$$\begin{aligned}
\mathbb{E}\left(y_i^{*,perm}\right) &= \mu_y, && (2.4)\\
Var\left(y_i^{*,perm}\right) &= V_p, && (2.5)\\
\mathbb{E}\left(\varepsilon_i\right) &= 0, && (2.6)\\
Var\left(\varepsilon_i\right) &= V_\varepsilon, && (2.7)\\
\mathbb{E}\left(I_i\right) &= p_I, && (2.8)\\
\mathbb{E}\left(k_i\right) &= \mu_k > 0, && (2.9)\\
Var\left(k_i\right) &= V_k. && (2.10)
\end{aligned}$$

Let us assume that the administrative data measures earnings accurately, but persons under-report transitory earnings shocks in the survey.

$$\begin{aligned}
y_i^a &= y_i^*, && (2.11)\\
y_i^s &= y_i^{*,perm} + \tau y_i^{*,trans} + \xi_i, && (2.12)
\end{aligned}$$

where y_i^a is earnings in the administrative data; y_i^s is earnings in the survey; τ is a measure of under-reporting $(0 \leq \tau \leq 1)$; and ξ_i is a classical measurement error component independent of everything with first and second moments

$$\begin{aligned}
\mathbb{E}\left(\xi_i\right) &= 0, && (2.13)\\
Var\left(\xi_i\right) &= V_\xi. && (2.14)
\end{aligned}$$

Measurement error is simply the difference between the survey and the administrative measures

$$m_i = -(1-\tau)\, y_i^{*,trans} + \xi_i. \tag{2.15}$$

The first and second moments of measurement error are

$$\mathbb{E}(m_i) = (1-\tau)\, p_I \mu_k \geq 0, \tag{2.16}$$

$$Var(m_i) = (1-\tau)^2 (V_\varepsilon + p_I V_k) + V_\xi. \tag{2.17}$$

Unlike in the classical measurement error (CME) model, the mean of the measurement error is weakly positive as $\tau \leq 1$ and $\mu_k > 0$. The mean error is expected to be close to zero when career interruptions are rare (p_I is small), when persons recall transitory shocks accurately ($\tau \approx 1$), or when the average earnings loss from career interruptions is small ($\mu_k \approx 0$).

The covariance between the error and the administrative data is

$$Cov(m_i, y_i^a) = -(1-\tau)(V_\varepsilon + p_I V_k) \leq 0. \tag{2.18}$$

In the CME model this covariance is assumed to be zero as the error is independent of true earnings. The empirical literature, however, has found that there is a strong negative correlation between the error and administrative earnings. As (2.18) shows, this simple model is able to predict a negative dependence between the error and the true value. The correlation decreases in under-reporting $(1-\tau)$ and the variance of transitory earnings $(V_\varepsilon + p_I V_k)$.

Bound et al. (2001) discuss error correction formulas when there is a negative relationship between the error and the truth. Bollinger (1998) and Kristensen and Westergaard-Nielsen (2006) found that this relationship, however, is not linear, but it is mostly driven by low earnings values. The mean of the error is substantially positive at the low end of the earnings distribution, while it is close to zero at medium and high levels of earnings. In order to derive the conditional distribution of the error given true earnings ($\mathbb{E}(m_i|y_i^a)$) I make distributional assumptions: $y_i^{*,perm}$ and ε_i are distributed normally, and k_i is a constant, $k_i = k$.[24] Using straightforward algebra (see Appendix B.1) the conditional expectation is

$$\mathbb{E}(m_i|y_i^a) = (1-\tau)\left[-\frac{V_\varepsilon}{V_\varepsilon + V_p}(y_i^a - \mu) + \frac{V_p}{V_\varepsilon + V_p} k \Pr(I_i = 1|y_i^a)\right], \tag{2.19}$$

$$\Pr(I_i = 1|y_i^a) = \frac{p \exp\left(\frac{k}{V_\varepsilon + V_p}\left(\mu - y_i - \frac{k}{2}\right)\right)}{(1-p) + p \exp\left(\frac{k}{V_\varepsilon + V_p}\left(\mu - y_i - \frac{k}{2}\right)\right)}, \tag{2.20}$$

where $\Pr(I_i = 1|y_i^a)$ is the probability that person i experienced career interruption conditional on earning y_i^a.

The first term in the square bracket of (2.19) is linear in earnings (y_i^a), but the second is not. In fact it is easy to show that:

$$\lim_{y_i^a \to -\infty} \Pr(I_i = 1|y_i^a) = 1, \tag{2.21}$$

$$\lim_{y_i^a \to \infty} \Pr(I_i = 1|y_i^a) = 0. \tag{2.22}$$

At the low end of the earnings distribution everyone is a partial year earner, while at the high end everyone works the entire year. By further assuming that there are no transitory earnings shocks on the job, $V_\varepsilon = 0$, the error mean converges to

$$\lim_{y_i^a \to -\infty} \mathbb{E}(m_i|y_i^a) = (1-\tau) k, \tag{2.23}$$

$$\lim_{y_i^a \to \infty} \mathbb{E}(m_i|y_i^a) = 0. \tag{2.24}$$

[24] Assuming stochastic k_i makes the derivation substantially more complicated. One can show that the qualitative results are very similar to the simple model discussed here when one uses stochastic k and simulates the conditional error distribution.

The model predicts the Bollinger result. At the low end of the earnings distribution the mean error is large and positive (equation 2.23), because everyone is a job loser (equation 2.21), and persons only report τ fraction of their earnings losses, k. At the high end of the earnings distribution, however, the mean of the error is zero (equation 2.24), because no one is a job loser, and the only source of error is the classical component, ξ_i, which has a zero mean.

To sum up, career interruptions can compellingly explain the documented properties of measurement error. The model predicts mean reversion, as persons under-report transitory earnings shocks, and predicts that mean reversion is driven by low earnings values where most of the sample is a partial year earner.

2.2.2 Alternative models to explain mean reversion

There have been extensive discussions in the literature about what might explain the mean-reverting property of measurement error. The first idea is described in Pischke (1995), who hypothesized that people under-report transitory shocks in earnings. My model in 2.2.1 is an extension of the Pischke (1995) mechanism with career interruptions.

Kapteyn and Ypma (2007) proposed another mechanism to explain mean reversion: false matches between the administrative and survey data. False matches lead to an apparent mean reversion in survey error, because every mismatched person is linked to an average earner, on average. False matches might be important in many contexts, especially when probabilistic matching is used instead of exact identifiers. In the context of HRS, however, mismatch is unlikely to play an important role. HRS and the Social Security Administration actively try to avoid false matching by using several variables (Social Security number, name and birth date) for matching. Moreover, false matches would lead to mean reversion at all earnings levels, not just at the low end. It is hard to argue that only the lowest earners are mismatched.

Abowd and Stinson (2013) proposed yet another mechanism to explain mean-reversion in the error: measurement error in the administrative data. The literature almost always assumes that the administrative earnings records (of employees) are accurate and therefore we can directly observe the measurement error by subtracting the records from the survey reports. Imagine, instead, that the administrative and the survey earnings both contain classical measurement error:

$$
\begin{aligned}
y_i^a &= y_i^* + \xi_i^a, && (2.25)\\
y_i^s &= y_i^* + \xi_i^s, && (2.26)\\
\mathbb{E}\left(y_i^*\right) &= \mu_y, && (2.27)\\
V\left(y_i^*\right) &= V_y, && (2.28)\\
\mathbb{E}\left(\xi_i^j\right) &= 0, && (2.29)\\
V\left(\xi_i^j\right) &= V_\xi^j, && (2.30)
\end{aligned}
$$

with $j \in \{a, s\}$. Within the Abowd and Stinson framework the difference between the survey and administrative measures is the difference between the two measurement error components:

$$m_i \equiv y_i^s - y_i^a = \xi_i^s - \xi_i^a. \tag{2.31}$$

The covariance between m_i and the administrative data is negative as

$$Cov\left(m_i, y_i^a\right) = -V_\xi^a. \tag{2.32}$$

There are many reasons to believe that the administrative data contains at least some error: unofficial income might be missing from it[25], there might be coding errors, and the legal definitions of annual earnings might be different from actual earnings.[26] Bound et al. (1994) report that some outliers in their data seem to indicate problems with the administrative data.

Measurement error in the administrative data can explain the mean reverting property of the error, but it is less obvious why only low earnings values are biased. In this paper I shall maintain the standard assumption that the administrative data are accurate. I believe, however, it would be an interesting extension to develop a technique to identify measurement error in the administrative data, which has proven elusive. See Abowd and Stinson (2013) for more discussion about identification problems.

Lastly, a model of selective memory can also explain the Bollinger finding. This model assumes that people are more likely to under-report negative than positive earnings shocks. In the psychology literature Wagenaar (1986), Thompson et al. (1996) and Tourangeau et al. (2000) discuss that people tend to remember more emotionally intense events better, but pleasant events are better stored in memory. Moreover, psychologists observed that people, when faced with a survey question about sensitive topics, tend to bias their answers toward more favorable outcomes. Some people might think of their earnings as a gauge of

[25]One might expect that survey reports for this population are more accurate as surveys are collected anonymously. Unfortunately we know very little on how people with unofficial income respond to questions about their earnings in surveys. Hurst et al. (2014) finds strong evidence that the self-employed, who are expected to have the most unofficial income due to tax incentives, seriously under-report their income in surveys as well.

[26]The regulations concerning earnings on the W2 forms are quite complex and are not solely based on economic considerations. By looking at the instructions we can find, for example, the following list of earnings components: "*The cost of accident and health insurance premiums for 2%-or-more shareholder-employees paid by an S corporation*"; "*Taxable cost of group-term life insurance in excess of $50,000*", "*Unless excludable under Educational assistance programs (see page 5), payments for non-job-related education expenses or for payments under a nonaccountable plan*", etc. It is hard to argue that there are no arbitrary elements in these instructions.

personal success and they might find it embarrassing to report too low values. It is possible, for example, that people are more likely to report their usual earnings when they received a big negative transitory income shock, but they report their total earnings after a positive shock. Appendix B.2 shows that this mechanism is able to predict the asymmetric behavior of the error discussed above. I shall show, however, that transitory earnings shocks on the job are very small. Thus, the selective memory model is unlikely to explain much of the mean reversion. Consequently, I shall not build this mechanism into my model.

2.3 Data and descriptive results

2.3.1 The Health and Retirement Study (HRS)

The survey data used in this project is the HRS, which is a biannual panel survey launched in 1992. The original sample represented the 51-62-year-old US population, but in 1998 older cohorts were added to cover everyone 51 and older. Every six years after 1998 the current 51-57-year-old cohort is added to maintain the 51+ age coverage. In this paper I use the $1992-2012$ waves of the HRS that cover annual earnings between 1991 and 2011. The HRS is a moderately large panel dataset ($\sim 18,000$ observations per wave), allowing me to look at the earnings history of persons for a relatively long time. The SIPP panels that were used in recent validation studies cover only 2.5 to 4 years.

The HRS developed a so-called unfolding bracket question sequence to reduce item non-response. People who answer "Don't know" to any of the earnings questions get a series of follow-up questions about whether their earnings are smaller or higher than some pre-determined threshold values. Table 2.1 shows that the non-response rate about wage and salary income fell below 5 percent using this method. The HRS asks about annual gross earnings in the last calendar year in five categories in the following order: 1. income from self-employment; 2 wage and salary income; 3 other income from professional practice or trade; 4 other income from tips, bonuses, commissions; 5 other income e.g. from second jobs or military reserves. After each question the interviewer says "*Other than what you have already told me about, did you have any income from [Component X]*". If the answer is yes, persons are asked about its value. In this paper I concentrate on earnings from employment, so I disregard self-employment income and the sum of components 2-5 is my measure of survey reported earnings. Table 2.2 shows that wage and salary income is by far the largest component of earnings.

For most of the paper I restrict the sample to the 50-60-year-old grouping. I shall show in Section 2.3.5 that above age 60, as people approach retirement, their earnings paths become quite complex: average earnings start falling sharply, and cross-personal variation starts increasing heavily. My preferred sample uses the relatively stable age range. I also drop the self-employed from the analysis: All person-year observations are dropped in which the person reported to be self-employed, or reported any self-employment income or in which there is any self-employment income in the administrative data.

2.3.2 Earnings in the administrative data

The source of the administrative data is the Detailed and Summary Earnings Records derived from the Master Earnings File of the Social Security Administration (SSA). For details about this data, see Olsen and Hudson (2009) and the documents on the SSA[27] and HRS websites.[28] The data are derived from the W2 forms filed by employers to the Internal Revenue Service each year. The higher quality Detailed Earnings Records (DER) is available from 1978 onward, and the less detailed Summary Earnings Records (SER) is available from 1951. My preferred sample uses years between 1991 and 2011.

The data contains employer-employee-year level information and several earnings variables. To define administrative earnings I consider total compensation and Medicare earnings. The maximum of total compensation and Medicare earnings is my measure of W2 earnings at a particular employer, and the sum over all employers is my measure of total annual W2 earnings for an individual in a given year. Total compensation amounts to the Box 1 values

[27] See http://www.ssa.gov/policy/docs/ssb/v69n3/v69n3p29.html for details.

[28] There are two relatively detailed documents under the data section at hrsonline.isr.umich.edu that can be accessed after free registration. Note, however, that the Social Security data are restricted and cannot be accessed online. The website provides information about how to get permission to use the restricted data.

of the W2s and it includes wages, bonuses, non-cash payments and tips[29]. It typically does not include deferred payments such as contributions to 401k plans, but certain plans are included. Medicare earnings are based on the Box 5 values of the W2 forms. As opposed to total compensation, Medicare earnings include most deferred payments. Before 1994, however, Medicare earnings were capped at the taxable maximum. Before 1991, the Medicare caps were equal to the Social Security caps (around $\$50,000$). Since 1994, there has been no limit on the taxable earnings for Medicare, and between 1991 and 1993 the Medicare and the Social Security taxable maximums began rapidly diverging.

Zero earnings are typically not recorded in the HRS version of the DER, but they are available in the SER. I identify zero earnings as either zero records in DER or a missing information in DER and zero records in SER.

The administrative data also contains self-employment income derived from the Form 1040 Schedule SE reported by the self-employed to the IRS. As discussed earlier, I drop all person-year observations with any self-employment income.

The administrative data are only available in intervals, although most intervals are narrow. In the descriptive part of the paper I impute values from these intervals with a simple procedure outlined in Section 2.3.3, but in the structural estimation I model interval responses in an interval regression fashion. The first type of interval response is top-coded Medicare earnings, where Medicare earnings exceed total compensation, but are at the taxable maximum. For confidentiality reasons the HRS also top-codes earnings above $\$250,000$ in a given year: When earnings are above this threshold, researchers only see if they are in one of the following intervals: $[\$250,000, \$299,999]$; $[\$300,000, \$499,999]$; $[\$500,000, \infty]$. This top-coding only affects a handful of cases. The HRS also rounded earnings below $\$250,000$ to the closest multiple of

[29]Only tips that the employee reported to the employer. Allocated tips are not part of Box 1.

\$100, with the exception of \$0 – \$49, where we can differentiate between a true \$0 and a \$1 – \$49 value.

2.3.3 Linking the HRS and administrative data

In order to match the HRS to the SSA data consent is required from the interviewees. Prior to 2006, consent only covered current and past years, but since 2006, consent covers all future years as well. For example, if the last consent from individual i was in wave 2000, we can see his SSA earnings from 1951 to 2000, but if the last consent was in wave 2008, we see his earnings from 1951 until today. Earnings histories, thus, are sometimes right-censored, but never left-censored. The coverage, however, is expected to improve over time, as more and more forward looking consent will be available.

The HRS made a lot of effort to increase the participation rate, but it remained below 100 percent. Generally the HRS has a relatively good coverage rate for earnings before 1992 and a moderately good coverage rate for earnings afterwards. Table 2.3 shows the fraction successfully matched in a few subgroups. The average match rate among 50-60-year-old employees between 1991 and 2011 is a little over 60%. The match rate is above 66% among those who reported positive earnings in the HRS. It appears that persons who do not work are less likely to consent. The match rate is almost 90% in the first 1992 wave of the survey, it is between 60% and 70% between 1994 and 2008, and it is around 50% in 2010 and 2012. Table 2.4 estimates linear probability models of successful matches on basic covariates. The strongest predictors are interview years, race and marital status. Minorities and singles are less likely to consent, but education and earnings do not predict the match rate. The somewhat selective nature of the matched sample might bias some of the estimates. In the descriptive part of the paper I shall ignore this problem, but in the structural part missing values are imputed with multiple imputation. These imputations assume that the missing values are conditionally ignorable (Little and Rubin, 2002).

Both HRS and W2 earnings contain quite a few interval responses. In the structural estimation I model them in an interval regression fashion. For the descriptive part of the paper I use a simple imputation that minimizes the error. When the reported interval (or value) and the recorded interval overlap, I define measurement error as zero and I assign the middle point of the overlapping interval to both earnings variables. If the two intervals do not overlap, I assign the values on the edge of the intervals that minimize the measurement error. The final (rare) case is when both the HRS report and the W2 record are right-open intervals. In such cases, I assign 110 percent of the maximum of the two left points of the intervals to both earnings variables. Finally, I deflated all earnings values with the CPI to 2000 dollars.

2.3.4 Labor histories

The administrative data provides employer-employee-year level information, and it contains scrambled employer IDs that identify changes of employers for the same person over time, but do not identify the same employer across individuals. I define job loss in year t as the disappearance of an employer ID from year t to year $t+1$, and job finding as the appearance of a new employer ID from year $t-1$ to year t. Employer IDs might change due to administrative reasons, which biases turnover upward, but the extent of it should be negligible compared to real turnover. Moreover, I am interested in how persons report earnings losses in years when there is evidence for a career interruption and their administrative earnings fall substantially.

Table 2.7 shows the distribution of detailed labor histories between 1991 and 2011. Column 1 and 2 use only the matched HRS-SSA sample, and column 3 and 4 add the off years when HRS responses are not available. In the latter sample, 49% of person-year observations correspond to full year earnings in a single job, when the person held one job, and he also worked for that employer in year $t-1$ and year $t+1$. The second largest category is zero earnings (24%), when the person did not work for any employer in a particular year. Around 6% of person-years involved a job loss without job finding. The majority (4.3%) correspond to the loss of a single job, and 1.8% involved the loss of at least one of multiple jobs. The table differentiates between major and minor job losses. Major job loss is the loss of a job that was the highest paying in year t or $t-1$, and minor job loss is the loss of a secondary or supplemental job that is not the highest paying neither in year t nor in $t-1$. The table shows that minor job losses are far more common than major, but not single job losses. Around 4.5% of person-years involved a job finding without a job loss. The majority (2.9%) found and held a single job. Around 10% of person-year observations involved both a job loss and a job finding. The majority (5.9%) corresponds to losing and then finding a major job; 1.9% correspond to losing and finding minor jobs; and around 1 – 1% correspond to years when a major job loss and a minor job finding, or vice versa, occurred.

Overall, there is evidence for quite high turnover in the HRS. Moreover, my procedure might actually miss some job losses. Persons who lose their jobs but return to the same employer within a year are mistakenly identified as full year earners. The administrative data, under current limitations, cannot identify such cases, and there is some evidence that the extent of this problem is non-negligible. My structural earnings model, thus, shall incorporate latent job losses as well.

Tables 2.8 and 2.9 show the distribution of earnings by labor histories. As expected, career interruptions substantially increase the likelihood of making relatively little. Around 44% of annual earnings are below \$10,000 (in 2000 dollars) when persons lose their single jobs. The number is 24% after losing a major job and only 10% after losing a minor one.

The fraction of earnings below $\$10,000$ is very similar in years with a single, major, or minor job finding (47%, 27% and 10% respectively). When persons work for the entire year, their earnings are larger, but still around 9% of the observations are below the $\$10,000$ threshold.

The distribution of earnings after a minor job loss and a minor job finding is similar to the distribution of full year earnings. Thus, I aggregate labor histories in the following way:

1. *Stable years* are observations with positive earnings without any indication of a major or single job loss or job finding. Stable years include full year earnings, minor job losses and minor job findings.

2. *Major job losses* are observations with major or single job losses and either no or only minor job findings.

3. *Major job findings* are observations with major or single job findings and either no or only minor job losses.

4. Major job loss and job findings are observations with both a major job loss and a major job finding. This category, among others, contains employer-to-employer switches.

5. *Zero earnings* are observations without any earnings.

Finally, I drop W2 earnings corresponding to the last consented year, since it is not possible to determine if persons managed to keep their jobs the entire year or not.

2.3.5 Descriptive properties of measurement error

From now on Y_{it}^a and y_{it}^a denote the level and log of annual W2 earnings; Y_{it}^s and y_{it}^s denote the level and log annual HRS earnings; and m_{it} denotes the measurement error: $m_{it} = y_{it}^s - y_{it}^a$.

Table 2.6 shows basic moments of earnings and measurement error. The first row is the classical reliability ratio defined as:

$$r^{classical} = \frac{Var\left(y_{it}^{a}\right)}{Var\left(y_{it}^{a}\right) + Var\left(m_{it}\right)}. \tag{2.33}$$

Under the assumptions of the classical measurement error model, $r^{classical}$ characterizes the attenuation bias in simple regressions with the error ridden variable on the right hand side:

$$\hat{\beta} = r^{classical} \times \beta + o_p\left(1\right). \tag{2.34}$$

According to Table 2.6 the reliability ratio is 0.82, which means that, if the CME model holds, in regressions with earnings on the right hand side, one should divide the coefficient on earnings by 0.82 to correct for the attenuation bias. Column 5 shows that when earnings growth (log changes in earnings) is on the right hand side, one should divide the coefficient by 0.50. Earnings growth is less reliable as earnings are more persistent than measurement error.

The second row of Table 2.6 shows a more general reliability ratio, $r^{correlated}$, which, as opposed to the CME model, allows for a non-negative correlation between measurement error and earnings (Bound et al., 2001):

$$r^{correlated} = \frac{Var\left(y_{it}^{a}\right)}{Var\left(y_{it}^{a}\right) + Var\left(m_{it}\right) + 2Cov\left(y_{it}^{a}, m_{it}\right)}. \tag{2.35}$$

The literature consistently finds a strong negative correlation between error and earnings. The HRS is no exception. The negative correlation decreases the denominator of (2.35), or in other words, it makes the attenuation bias less strong. As Table 2.6 shows, the correlated reliability ratio is close to one, implying that under the assumptions of the Correlated Measurement Error (CorrME) model, there is no attenuation bias in regressions with earnings on the right hand side. The negative correlation also increases the reliability ratio in earnings growth, but it is still well short of one (0.68).

The third row of Table 2.6 shows the regression coefficient of error on earnings:

$$m_{it} = \gamma_{m,y^a} y_{it}^a + v_{it}. \tag{2.36}$$

As Bound et al. (2001) discuss, regression coefficients are biased even when earnings appear on the left hand side of regression in the CorrME model, and γ_{m,y^*} is a measure of the bias:

$$\hat{\alpha} = \gamma_{m,y^a} \times \alpha + o_p(1). \tag{2.37}$$

Similarly to the literature, the γ_{m,y^a} coefficient is estimated to be around -0.1, meaning that predictors of earnings are downward biased by about 10% according to the CorrME model.

Columns 2-4 of Table 2.6 show a potential problem with the CorrME model. The basic model assumes that the negative relationship between the error and earnings is linear as in (2.36). As the table shows, the relationship is far from linear as seen when the lowest earnings values are discarded and the negative relationship approaches zero. When I discard earnings below $\$10,000$ (in 2000 dollars), γ_{m,y^a} shrinks from -0.099 to only -0.025, although it is still statistically significant. Measurement error above $\$10,000$ appears to be less correlated and is almost classical. In the case, when the relationship between the error and true earnings is not linear, $r^{correlated}$ and γ_{m,y^a} still correctly accounts for the bias, as long as total earnings (as opposed to permanent earnings) are in the regression.

Figure 2.2 visualizes the nonparametric relationship between the error and earnings. The top left panel shows the predicted value of the nonparametric regression between the 1^{st} and 99^{th} percentile of earnings together with the predicted values of the CME and the basic CorrME models. The top right panel shows average measurement error in 20 equal sized quantiles of earnings together with 95% confidence intervals. The lowest earnings are severely upward biased, but the error appears almost classical above the 15^{th} or 20^{th} percentile of earnings. Neither the CME nor the CorrME model accurately characterizes measurement error in earnings. Very similar results were found by Bollinger (1998) using the Current Population Survey and Kristensen and Westergaard-Nielsen (2006) using a Danish household survey. The bottom panels of Figure 2.2 show that the CorrME model works much better for earnings growth. It seems the relationship between error and earnings growth is close to linear, with perhaps some acceleration at the extremes.

The main claim of my paper is that transitory earnings shocks due to career interruptions are under-reported in surveys. This mechanism exactly predicts the type of non-parametric relationship between the error and earnings as in Figure 2.2. Career interruptions lead to large earnings losses and large positive measurement error at the low end of the earnings distribution if they are systematically under-reported. Figure 2.1 shows histograms of measurement error together with fitted normal densities. All histograms show excess kurtosis and right skew compared to the normal plots. The right skew appears strongest in the sample with partial year earnings. This is additional evidence that negative earnings shocks due to partial year earnings are under-reported in surveys.

Table 2.11 shows the means of measurement error by labor histories and earnings. Average error is large and positive below $10,000, and sharply decreases with earnings. The averages are very similar across the three types of career interruption (job loss, job finding or both). Average measurement error is substantially smaller in stable years, but still positive at the low end of the earnings distribution. Table 2.12 further investigates measurement error in stable years. I restrict the sample to stable years in both year t and $t-1$, and I divide the sample based on whether earnings fell or increased between $t-1$ and t. Average measurement error is close to zero when earnings increased, except in the rare cases they increased but remained under $\$1,000$. Average error is positive and substantial, however, when earnings fall. This might indicate that some observations identified as stable actually involve large negative transitory shocks as well. It is possible that these observations correspond to person-years with a job loss and a subsequent return to the same employer.

Finally, I fit a simple descriptive earnings and measurement error model à la Pischke (1995). The model assumes earnings are the sum of a permanent and a transitory component:

$$y_{it} = \mu + y_i^{perm} + y_{it}^{tran}, \tag{2.38}$$

where y_i^{perm} and y_{it}^{trans} are mean zero. Permanent earnings are fixed within person and transitory earnings follow an MA(1) process:

$$y_{it}^{tran} \quad = \quad \varepsilon_{it} - \theta\varepsilon_{i,t-1}. \tag{2.39}$$

The survey reports are

$$y_{it}^{s} = \mu + \delta + \tau^p y_i^{perm} + \tau^t y_{it}^{tran} + m_i + n_{it}, \tag{2.40}$$

where δ indicates mean error, τ^p and τ^t are the average fraction of permanent and transitory earnings reported, m_i is the permanent measurement error and n_{it} is the transitory measurement error. All error terms are assumed to be uncorrelated with each other and with the other variables of the model.

I estimate this system by GMM using all the first and second moments of earnings in year t and $t+2$. Apart from the MA(1) coefficient, θ, all parameters are identified, including the total variance of transitory earnings.

Table 2.21 in Appendix C estimates the model by 2-year age groups between the ages of 50 and 70. It shows that the mean and the variance of earnings is relatively stable between the ages of 50 and 60, the mean rapidly decreases after age 60, while the variance increases sharply. From now on I only use the relatively stable 50-60-year-old sample.

The main results are in Table 2.13. Column 1 uses all observations and column 2 restricts the sample to cases with no indication of career interruption in years t, $t+1$ or $t+2$. Note that the model in column 1 is biased because career interruptions lead to non-zero (negative) mean transitory earnings. Nevertheless, I estimate this model due to its similarities with the model Pischke (1995) used, and because I believe comparing it to column 2 is worthwhile. The average measurement error is positive in both cases, but it is close to zero in stable years. The (inter-personal) variance of permanent earnings is substantial in both models. The variance of transitory earnings is large in model 1, but it is close to zero when only stable years are used. Transitory earnings shocks on the job, thus, are small, but transitory earnings shocks due to career interruptions are large. The main parameters of interest are in the last two rows. Persons report close to 100% of their permanent earnings, but substantially less of their transitory earnings: 68% for all observations and 76% when only stable years are considered. The standard error in the latter case, though, is quite large, perhaps because it is hard to identify the under-reported component of earnings that are so small.

Overall, the descriptive evidence favors the main claims of the paper that under-reporting of earnings losses due to career interruptions has a substantial and non-standard effect on measurement error.

2.4 Structural estimation

This section introduces and estimates a dynamic earnings and measurement error model that features career interruptions and deals with missing values and interval responses.

2.4.1 The earnings and measurement error model

The W2 and HRS earnings are provided in intervals. I assume that there are latent earnings y_{it}^a and y_{it}^s that are somewhere in the provided intervals:

$$y_{it}^j \in \left[\underline{y}_{it}^j, \bar{y}_{it}^j\right], \tag{2.41}$$

where $j \in \{a, s\}$, and only the lower and upper bar earnings values are observed in the HRS and SSA data. When the HRS earnings reports are continuous values, I use the trivial intervals. When the HRS or the SSA earnings are missing, I use the $[-\infty, \infty]$ intervals.

I assume that the W2 earnings are accurate. Earnings of person i in year t depend on some potentially time-varying covariates, and they have a permanent and a transitory component:

$$\begin{aligned}
y_{it}^a &= \beta' x_{it} + y_{it}^{perm} + y_{it}^{tran}, && (2.42)\\
y_{it}^{perm} &= y_{i,t-1}^{perm} + \nu_{it}, && (2.43)\\
y_{it}^{tran} &= y_{it}^{MA(q)} - \sum_{l=1}^{L} I_{itl} k_{itl}, && (2.44)\\
y_{it}^{MA(k)} &= \varepsilon_{it} - \sum_{j=1}^{k} \theta_j \varepsilon_{i,t-j} && (2.45)
\end{aligned}$$

Permanent earnings follow a random walk with innovation term ν_{it}. Transitory earnings on-the-job follow a moving average process with q lags. My preferred model features $q = 0$ lags.[30] I_{itl} indicates a career interruption of type $l \in \{1, ...; L\}$, such as a major job loss or a job finding. Career interruptions affect both permanent and transitory earnings. Transitory earnings loss due to an l type interruption is denoted by k_{itl}. Career interruptions also affect the mean and the variance of the random walk innovation term of permanent earnings:

$$\mathbb{E}\left(\upsilon_{it}\right) = \sum_{l=1}^{L} \lambda_l^m I_{itl}, \tag{2.46}$$

$$Var\left(\upsilon_{it}\right) = \sum_{l=1}^{L} \lambda_l^V I_{itl}. \tag{2.47}$$

The mean of the innovation term is normalized to zero on-the-job, but it is free in other cases. Among other things, this specification allows for the human capital of individuals to depreciate when they are not working. The variance of the innovation term can be non-zero even on-the-job to capture promotions and demotions.

Even though my notations involve summations over labor histories in (2.44)-(2.47),a maximum of one indicator is one in a particular year for a given individual. In the previous section, we saw that there are large negative earnings shocks even in stable years, so I allow for latent job losses on-the-job with probability p. Latent job loss behaves similarly to other career interruptions: It can change the mean and the variance of permanent earnings and it can lead to a transitory earnings loss, too. The probability p shall be estimated together with the other parameters.

[30]The variance of on-the-job transitory earnings is very small, so it makes little difference how much structure one puts on these shocks. The model with no MA lags converges much faster to a stable distribution, so I use this assumption in the baseline case.

The HRS earnings reports are

$$y_{it}^{s} = (\beta + \delta)' x_{it} + \tau^{p} y_{it}^{perm} + \tau^{t} y_{it}^{MA(k)} - \sum_{l=1}^{L} \tau_{l}^{k} I_{itl} k_{itl} + m_{i} + n_{it}. \tag{2.48}$$

δ is a vector of bias terms in the regression coefficients; τ^p, τ^t and the τ_l^k terms are the average fractions of the earnings components reported (permanent, on-the-job transitory and transitory due to career interruption); m_i is permanent and n_{it} is transitory measurement error with mean zero and positive variance. The variance of transitory measurement error can also depend on labor histories:

$$Var\left(n_{it}\right) = \sum_{l=1}^{L} \lambda_{l}^{n} I_{itl}. \tag{2.49}$$

I use Markov Chain Monte Carlo (MCMC) for estimation that requires distributional assumptions. I assume that all random variables apart from the k_{itl} and I_{itl} are distributed normally and are uncorrelated with each other. The uncorrelatedness assumption is not particularly restrictive as considerable dependence is already built in the model. I_{itl} is assumed to follow a Bernoulli distribution with a constant parameter, p. I assume that the transitory earnings loss terms, the k_{itl}-s, follow an exponential distribution with p.d.f.

$$f\left(k_{itl}\right) = \gamma_{l} \exp\left(-\gamma_{l} k_{itl}\right), \tag{2.50}$$

$$k_{itl} \in [0, \infty]. \tag{2.51}$$

The motivation for this p.d.f. is the following: Imagine that an l-type career interruption implies that the person works for T_{itl} and does not work for $12 - T_{itl}$ months. His earnings loss is $K_{itl} = 1 - \frac{T_{itl}}{12}$ fraction of his earnings potential. Assuming that the fraction K_{itl} is uniformly distributed between zero and one, and taking log leads to the p.d.f. in (2.50) with parameter $\gamma_l = 1$, and $k_{itl} \equiv 1 - \ln(1 - K_{itl})$. I allow γ_l to differ from one, and I will estimate separate γ_l terms for each type of career interruption. The average log and level earnings losses decrease with γ_l:

$$\mathbb{E}(k_{itl}) = \frac{1}{\gamma_l}, \tag{2.52}$$

$$\mathbb{E}(K_{itl}) = \frac{1}{\gamma_l + 1}, \tag{2.53}$$

Figure 2.3 shows that this distributional assumption works well. It shows the histogram of earnings losses from year $t-1$ to year t for observations when $t-1$ was a stable year and year t involved a major job loss. The figure also shows a simulated density that is the sum of two terms: one with a p.d.f. from (2.50) with $\gamma_l = 1$, and one with the empirical distribution of earnings changes when both $t-1$ and t were stable years. The shape of the two distributions are similar, but the empirical histogram falls a little short of the simulated one. It appears persons lose less than half of their earnings after a typical job loss. Consequently we can expect the corresponding γ term to be above one.

2.4.2 Estimation methodology

I use a Bayesian estimation method, Markov Chain Monte Carlo (MCMC), for estimation. Gelman et al. (2013) provides a thorough overview of MCMC and other similar techniques. In some ways, MCMC is a more general version of Maximum Simulated Likelihood. It requires a fully specified model, but it can be used in very complex cases when it is too tedious or too slow to simulate the likelihood function.

The unusual feature of MCMC is that it assumes that the parameters of the model being estimated are random. The data generating process is described by the c.d.f. $F(y, \alpha)$, where y represents the data and α represents the unknown parameters. The goal of the procedure is to derive the posterior distribution of the parameters, $F(\alpha|y)$. Once the posterior is estimated, one can use its mean as the estimate, and its standard deviation as the standard error:

$$\hat{\alpha} = \mathbb{E}_F(\alpha|y), \tag{2.54}$$

$$se(\hat{\alpha}) = Std_F(\alpha|y). \tag{2.55}$$

Alternatively, one can directly define a 95% confidence interval by taking the 2.5^{th} and 97.5^{th} percentile of $F(\alpha|y)$. In this paper I shall only provide the means and standard errors from (2.54) and (2.55).

The posterior distributions of the parameters are simulated with the following procedure:

1. Take initial guesses for all α_j-s, $j \in \{1, 2, ..., J\}$. The set of α-s includes the parameters and potentially other latent or auxiliary random effects, such as the innovation terms of the random walk, the I_{itl} and k_{itl} terms, etc.

2. Define a prior distribution for all the parameters $f_0(\alpha_j)$. The priors can be defined individually, jointly or in blocks.

3. Draw new values of the parameters from their posterior distribution conditional on their prior, the data and other parameters, $F(\alpha_j|y, \alpha_{-j}, f_0(\alpha_j))$. One can update the parameters individually, jointly or in blocks.

4. Repeat step 3 N^S times, where N^S is the number of simulation draws.

5. Drop the first N^B burn-in draws and use the rest $N^S - N^B$ for inference.

The prior distributions are auxiliary objects and researchers almost always choose vague (or flat or non-informative) priors so that they do not influence the results. One can, for example, use a prior of $N(0, 100)$ for the distribution of mean log earnings. I also use vague priors in this project.

The number of necessary simulation draws, N^S, sometimes has to be large, because the consecutive simulation draws can have, and often do have, a large positive autocorrelation, which decreases the effective sample size of the simulated data. In this project, I use $100,000$ simulation draws for each parameter, and I discard the first $10,000$ burn-in draws.

The most challenging part of the estimation is to derive the posterior distributions, the $F(\alpha_j | y, \alpha_{-j}, f_0(\alpha_j))$-s, of all parameters. In standard cases, there are readily available formulas. For example, when the data are distributed normally, means, variances and regression coefficients can be updated easily. In other cases, the posterior distributions are unknown or they do not have analytical forms. In such cases the Metropolis-Hastings algorithm can be used to simulate the posterior distribution. In this project I use the Metropolis-Hastings algorithm to update the MA coefficients and the earnings loss variables due to career interruptions: I_{itl}, k_{itl} and latent transitory measurement error terms due to career interruptions. Other parameters are updated using the standard Gibbs sampler.

MCMC is particularly useful for models that feature a large number of random effects or missing data. In such cases the MCMC procedure is substantially simpler than alternatives. Other features are harder to implement with MCMC. Properly modeling moving average processes and time-varying variances, such as variances that depend on labor histories, are harder. The simulation part of the Matlab code of my preferred model has almost 500 lines. Details can be found on my website.[31]

[31] https://sites.google.com/site/phudomiet/research.

2.4.3 Results

Tables 2.14-2.18 show the main results of the paper. All of these specifications are estimated using years between 1991 and 2011, with age ranging between 50 and 60, and no lags in the moving average process of on-the-job transitory earnings. Appendix C shows results using alternative modeling assumptions.

The models are run separately for males and females. Tables 2.14 contains the regression coefficients of mean earnings; Table 2.15 shows the covariates in the means and variances of the random walk innovation terms; Table 2.16 contains other parameters of earnings; and Tables 2.17 and 2.18 show parameters of measurement error.

All Tables have six columns corresponding three alternative sets of covariates in earnings and gender. My preferred model is in Column 2. Column 1 only includes a constant term. Column 2 adds basic covariates such as race, education, age, GDP gap[32]. Column 3 adds year dummies instead of the other macro variables.

The estimated coefficients in mean earnings (Table 2.14) are standard: Minorities earn less than whites; education increases earnings; the age-earnings profile is an increasing and concave function; earnings are procyclical (and the cyclicality is stronger for males); male real earnings were slightly falling between 1991 and 2011 while female earnings stagnated or slightly increased. Cognition is a strong positive predictor of earnings, and the predictive power of education substantially decreases with the inclusion of cognition.

The mean and variance of the random walk innovation term (of permanent earnings) are allowed to depend on labor histories. I include dummies for 1. major job finding; 2. major job loss and major job finding; 3. zero earnings; and 4. latent job loss in stable years. I do not include a dummy for major job loss, as it is not separately identified from major job

[32] I use the IMF GDP gap measures divided by 100.

finding. When one loses his job, his permanent earnings are not observed until he finds a job again. Hence, it is not possible to identify the effect of a job loss and a consequent job finding separately. My models attribute all changes to job finding.

Moving from non-employment to employment has a small negative effect on permanent earnings for both males and females. In my preferred Model 2, the permanent earnings loss is 7.4% for males and 7.7% for females. This mean effect, however, masks a large cross-sectional heterogeneity. The standard deviation of the random walk innovation is close to zero on-the-job (0.06 for both males and females), but it jumps up to around 0.4 when persons find a new job. Not working for an entire year has an even more detrimental effect on earnings, on average: permanent earnings fall by 24% per year for males and 21% per year for females. Experiencing both a major job loss and a major job finding, the case that includes employment-to-employment switches, has a zero effect on permanent earnings, on average, for males, and a small positive effect for females. Finally, latent job loss in stable years is associated with a very large negative effect on permanent earnings, too: 21% earnings loss for males and 17% for females. Some of these earnings losses might be due to demotions; a reduction in work hours; or losing a job and being rehired in a lower position. The cross-sectional variation in earnings growth is very large in all cases when career interruptions occur.

Table 2.16 shows the rest of the parameters in earnings. The cross-sectional variation in earnings is quite large at age 50 already. Transitory earnings shocks on-the-job, however, are small: The standard deviation is around 0.04 for males and 0.03 for females. The corresponding variances are practically zero. Transitory earnings shocks due to career interruptions, however, are very large, on average. The estimated γ_l coefficients are larger than 1 in all cases, which means that earnings loss is less than half a year on average. Average earnings losses are between 23% and 45% based on the formula in (2.53). The largest transitory earnings loss is due to major job losses (41% for males and 45% for females) and the smallest is in latent job losses in stable years (24% for males and 25% for females).

The most important parameters, describing the properties of measurement error, are in Tables 2.17 and 2.18. The basic covariates have little effect on the mean of the error for both males and females. The standard deviation of permanent measurement error is small for both groups (0.16 for males and 0.15 for females). The standard deviation of transitory measurement error in stable years is also small (0.1 for males and 0.14 for females), but it is considerably larger when career interruptions occur. Depending on the type of shock, the standard deviation of the error is between 0.5 and 0.68. Persons report close to 100% of their permanent earnings, and substantially less of their transitory earnings. Identifying the fraction of on-the-job transitory earnings reported was an obstacle, likely because these shocks are very small, on average. In my baseline specification I simply calibrated the reported fraction to zero.[33] The fraction of transitory shocks due to career interruptions reported are between 50% and 87%. In my preferred model 2, male (female) interviewees report 84% ($\sim$ 77%) of the transitory shocks due to a major job loss, 62% (69%) of shocks due to a major job finding, 54% (65%) of shocks associated with both a job loss and a job finding, and 50% (73%) of shocks due to a latent job loss in stable years. Most parameters are very similar in the different specifications. The only exception is the fraction of reported earnings due to a latent job loss. This parameter varies between 50% and 100% in the alternative models.

Tables 2.24-2.48 in Appendix C carry out several robustness checks. Tables 2.24-2.33

[33] Table 2.38 in Appendix C show that the main results are robust to alternative assumptions in regard to this parameter: assuming that people report 50% or 100% of these shocks or that the fraction is estimated. In the case of the fraction being estimated, its standard error is quite large, and the distribution of the parameter does not appear to converge to a stable distribution.

estimates separate models by education groups; Tables 2.34-2.38 allow for higher order MA terms in transitory earnings, Tables 2.39-2.43 make alternative assumptions about the fraction of on-the-job transitory shocks reported, and Tables 2.44-2.48 uses different samples. The main results of the paper are robust to these changes: transitory earnings shocks due to career interruptions are large and are severely under-reported in the survey.

2.5 Discussion of bias correction

According to the descriptive results summarized in Table 2.6, respondents report around $1 - 0.26 = 74\%$ of their earnings growth, on average. According to Figure 2.2, under-reporting is an approximately linear function of earnings growth itself. The GMM estimates in Table 2.13 suggest that respondents report around 68% of their transitory earnings shocks. These results indicate that a simple model, in which measurement error has a classical component together with a mean-reverting one due to reporting only $\sim 70\%$ of transitory earnings fluctuations, might be a good approximation of the way people report their earnings in surveys. The MCMC estimates show a somewhat more nuanced picture: earnings losses due to job finding appears to be less salient than earnings losses due to job losses. The 70% number is approximately in the middle of the different MCMC estimates, though.

A natural question is: what do these results imply for biases in applied work and what can researchers do to correct them? The appropriate bias correction formulas depend on many factors such as:

1. Are earnings on the left- or right-hand side of regressions?

2. Does the researcher aim at including a measure of permanent earnings (or perhaps human capital or earnings potential) of individuals or their total earnings including

short-term transitory fluctuations?

3. Does the researcher model earnings levels or earnings growth?

4. Does the data have reliable measures of career interruptions or not?

5. Do particular variables in the regression correlate with turnover or not?

Discussing all of these is beyond the scope of this paper. I only consider a few interesting cases.

First, because measurement error in earnings growth appears to be an approximately linear function of earnings growth, the CorrME error model of Bound et al. (2001) might work well in these cases. According to Table 2.6, when earnings growth is on the left-hand side of regressions, the coefficients should be divided by 0.74; and when earnings growth is on the right-hand side regressions, the coefficients should be divided by 0.68. This simple procedure only works, however, if the researcher aims at including growth in total earnings, as opposed growth in permanent earnings. This is the case, for example, if one is interested in the transitory earnings loss of displaced workers either on the left or right hand side of regressions.

Second, when one is interested, instead, in permanent earnings growth, the correction formulas should be adjusted. This is the case if one wants to characterize differential earnings trends between groups over time, or one is interested in the effect of promotions/demotions/retirement on consumption growth. As people report the permanent component of their earnings accurately, no correction is needed when earnings growth is on the left hand side of regressions. Using observed earnings growth, as a proxy for permanent earnings growth on the right-hand side of regressions, however, should be done with caution. Not only is earnings growth measured with error, but it is dominated by transitory fluctuations as opposed to permanent changes.

Third, consider the case of using annual earnings (in levels) on the left hand side of regressions as a proxy for permanent earnings, such as in a Mincer earnings function. Interestingly, my results suggest that using survey-based measures is (slightly) superior to administrative records. Imagine that administrative and survey earnings take the following form:

$$y_i^a = y_i^{perm} + y_i^{trans}, \tag{2.56}$$

$$y_i^{perm} = \beta_0 + \beta_1 x_i + \eta_i, \tag{2.57}$$

$$y_i^{trans} = \varepsilon_i - I_i k_i, \tag{2.58}$$

$$y_i^s = y_i^{perm} + \tau y_i^{trans} + m_i, \tag{2.59}$$

where I_i indicates career interruptions, k_i indicates annual earnings losses due to career interruptions with mean $1/\gamma$, and the coefficient of interest is β_1. Under these assumption, the probability limit of regression coefficients with either y_i^a or y_i^s on the left hand side of the regression are

$$\text{plim}\hat{\beta}_1^a = \beta_1 - \frac{1}{\gamma}\beta_{I,x}, \tag{2.60}$$

$$\text{plim}\hat{\beta}_1^s = \beta_1 - \frac{\tau}{\gamma}\beta_{I,x}, \tag{2.61}$$

where $\beta_{I,x}$ is the coefficient from a regression of career interruption on x. The regression coefficient is biased if variable x_{it} is correlated with career interruptions. The sign of the bias depends on the sign of $\beta_{I,x}$. Variables with a positive effect on turnover have a negative bias, while negative predictors of turnover are biased upward. The survey-based regression is less biased than the administrative whenever $\tau < 1$, as persons under-report transitory shocks, and hence, under-report the bias. In the extreme case of $\tau = 0$, the survey based coefficient is actually unbiased. In most applications the regression coefficient $\beta_{I,x}$ is likely to be close to zero, since the unconditional probability of turnover is relatively low to begin with. Overall, unless variable x is expected to be a very strong predictor of turnover, one might expect negligible bias in these regressions.

Fourth, consider the case when annual earnings (in levels) appear on the right-hand side of a regression as proxies of permanent earnings.

$$\begin{aligned} z_i &= \beta_0 + \beta_1 y_i^{perm} + \eta_i, && (2.62) \\ y_i^a &= y_i^{perm} + \varepsilon_i - I_i k_i, && (2.63) \\ y_i^s &= y_i^{perm} + \tau\left(\varepsilon_i - I_i k_i\right) + m_i. && (2.64) \end{aligned}$$

Regression coefficients are inconsistent in this case. Under the assumption that all error terms are uncorrelated, $\Pr\left(I_i = 1\right) = p$ and $Var\left(k_i\right) = \frac{1}{\gamma^2}$, the probability limit of regression coefficients with either y_i^a or y_i^s on the right-hand side are

$$\begin{aligned} \text{plim}\hat{\beta}_1^a &= \beta_1 \frac{Var\left(y_i^{perm}\right)}{Var\left(y_i^{perm}\right) + Var\left(\varepsilon_i\right) + \frac{p}{\gamma^2}}, && (2.65) \\ \text{plim}\hat{\beta}_1^s &= \beta_1 \frac{Var\left(y_i^{perm}\right)}{Var\left(y_i^{perm}\right) + \tau\left(Var\left(\varepsilon_i\right) + \frac{p}{\gamma^2}\right) + Var\left(m_i\right)}. && (2.66) \end{aligned}$$

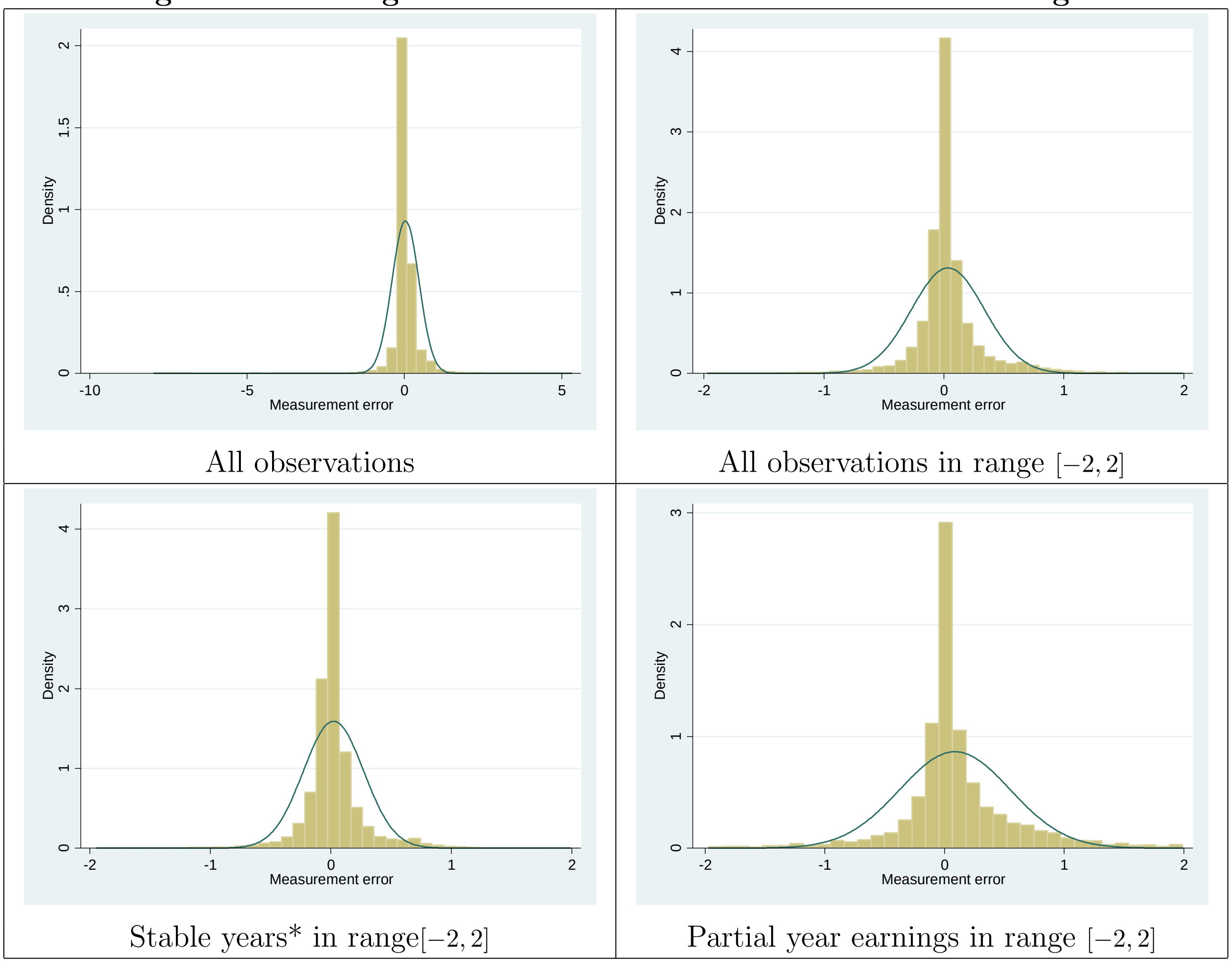
Under-reporting transitory shocks helps on the right hand side as it decreases the second term in the denominator of (2.66). The third (classical error) term in (2.66), however, increases the denominator and increases the bias. The measurement error is large enough that the bias is stronger when the survey measure is used. The bias is non-negligible in both cases, though, and it is dominated by terms related to career interruptions. The variance of transitory earnings shocks on-the-job ($Var(\varepsilon_i)$) is negligible compared to $Var(y_i^{perm})$, but $\frac{p}{\gamma^2}$ is not. Moreover, as we saw in Tables 2.17 and 2.27, the variance of measurement error is small in stable years, and much larger in years when career interruptions occur. If researchers use either y_i^a or y_i^s as right-hand side variables as proxies of y_i^{perm} it is crucial to account for partial year earnings, by either dropping those values from the regression, or better: modeling turnover appropriately.

2.6 Conclusion

This paper analyzed the properties of measurement error in annual earnings with a particular focus on how partial year earnings are reported in a household survey. I find strong evidence that transitory earnings fluctuations are severely under-reported. Transitory earnings losses due to career interruptions are very large, on average, and people systematically under-report these shocks, too. Earnings losses due to recent job losses are reported better, perhaps because these shocks are more salient. Earnings losses due to recently starting a new job are reported less accurately, perhaps because people prefer reporting their annualized earnings after starting new jobs. Overall, career interruptions have large and non-standard effects on the quality of earnings reports in surveys. The implied error has a small positive mean, it is right-skewed and it correlates with predictors of turnover. The implied biases are likely to be relatively small in most cases, as career interruptions are rare. When researchers use earnings growth in their models, or when they analyze questions that are strongly related to turnover, more caution is warranted.

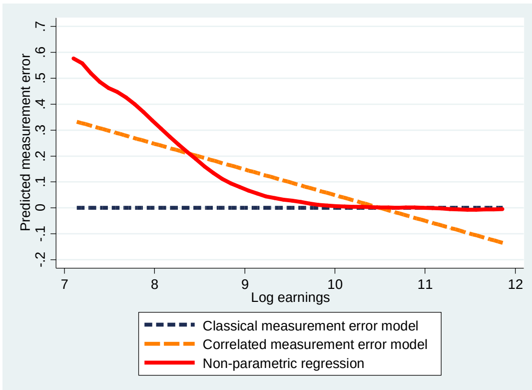
Tables anmd figures

Figure 2.1: Histogram of measurement error in annual earnings

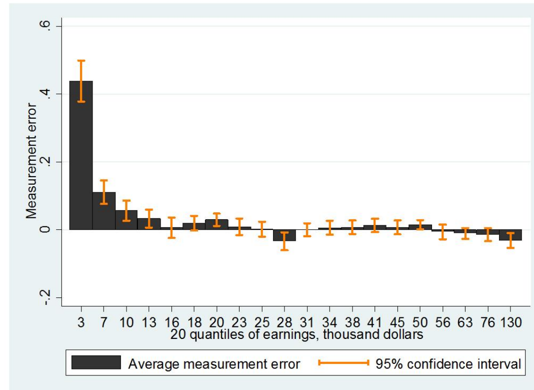


Stable years: Positive earnings and no evidence of major job loss or job finding. Major job loss: loss of a job that was the highest paid one in year t or $t-1$. Major job finding: finding a new job that is the highest paid one in year t or $t+1$.
Sample: HRS, 1991-2011, 50-60-year-old employees

Figure 2.2: Nonparametric relationship between measurement error and earnings

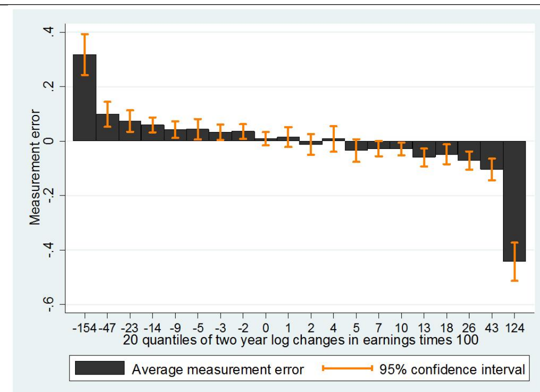


Nonparametric regression: error on earnings



Mean error by 20 quantiles of earnings

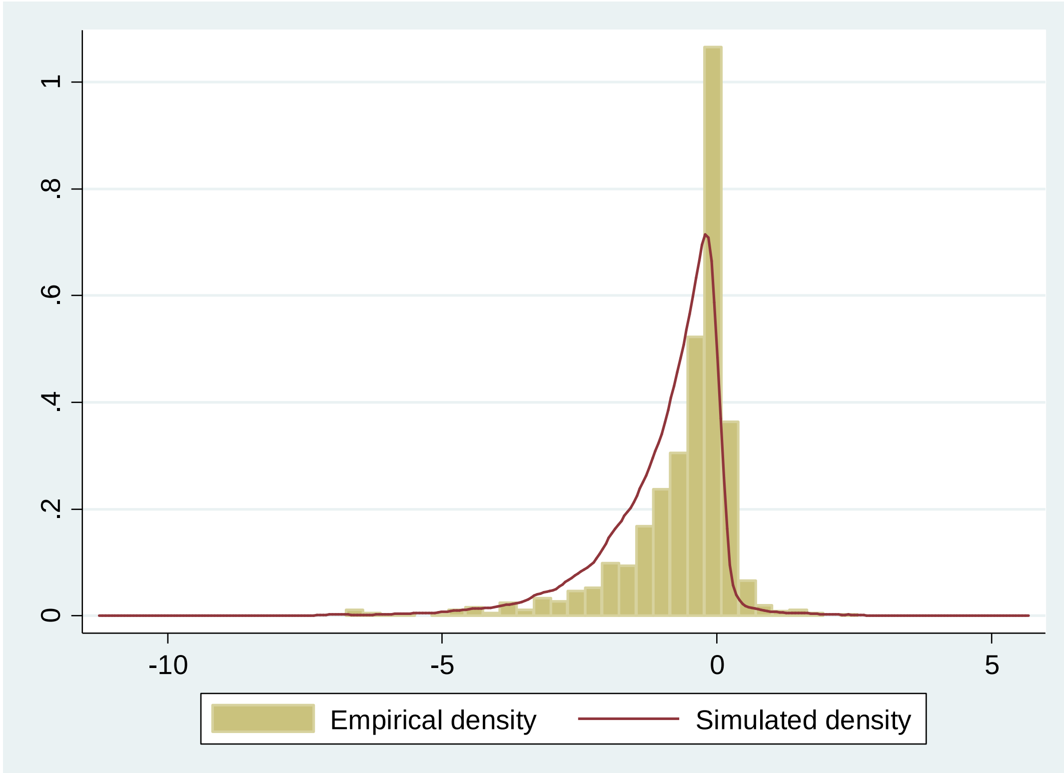
Nonparametric regression: error on earnings growth



Mean error by 20 quantiles of earnings growth

Sample: HRS, 1991-2011, 50-60-year-old employees

Figure 2.3: Histogram of earnings growth after major job loss



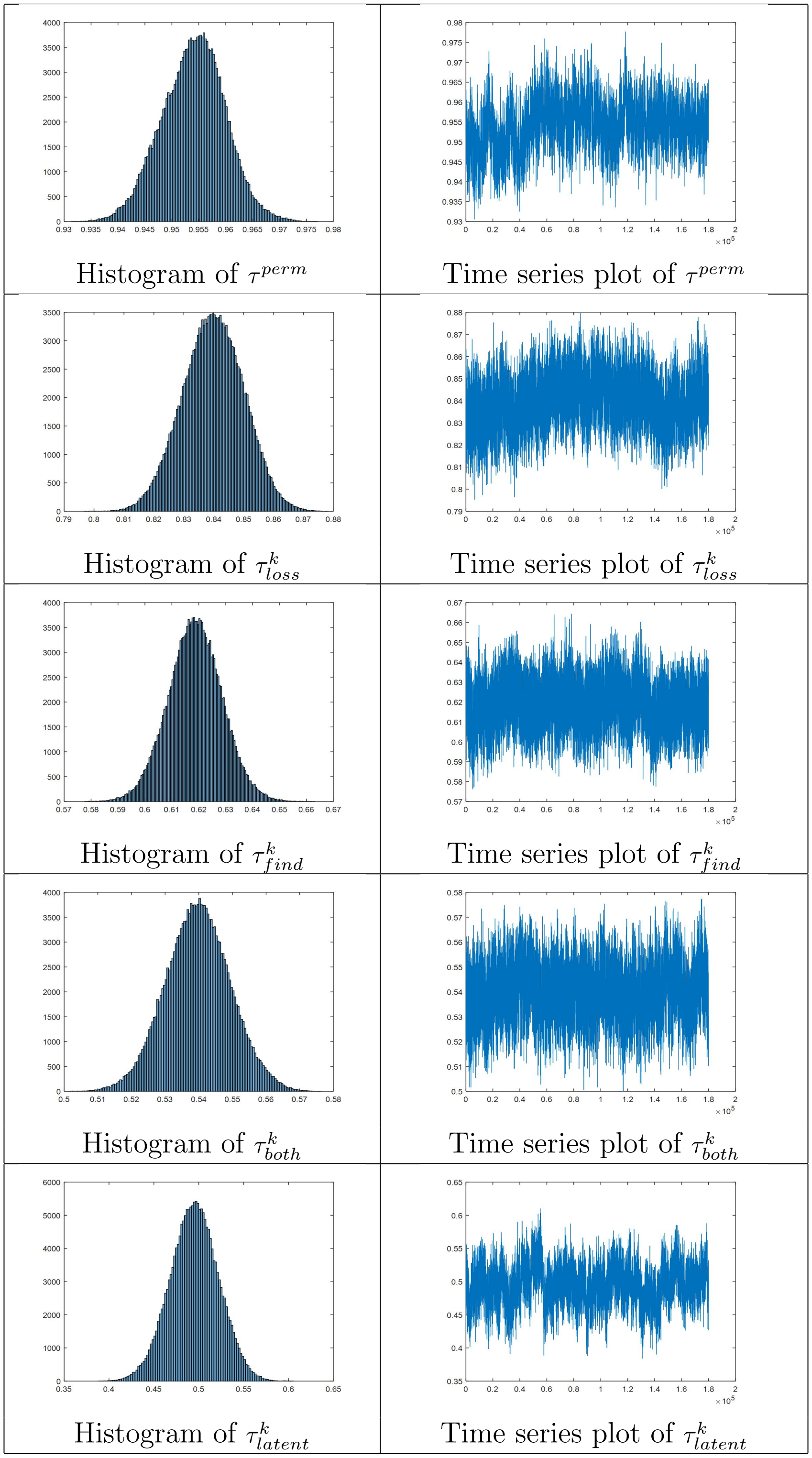
Earnings growth in a year with major job loss and predicted value assuming random job loss

Major job loss: Loss of a job that was the highest paid one in year t or $t-1$.

The random job loss model assumes that persons had one job and the date of job loss is uniformly distributed between January and December

Sample: HRS, 1991-2011, 50-60-year-old employees

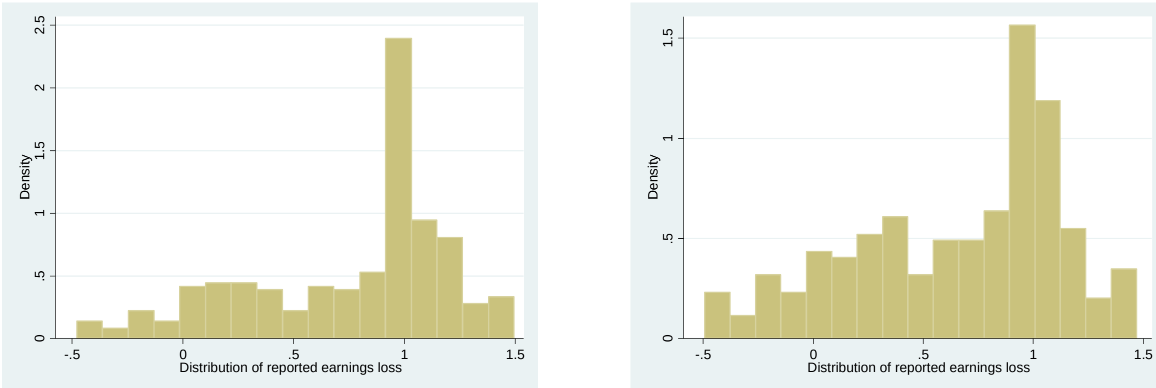
Figure 2.4: Histogram and time series plots of selected coefficients in the baseline specification of the structural model



The coefficients are from Model 2 of the baseline specification reported in Table 2.17.

Sample: HRS, 1991-2011, 50-60-year-old employees

Figure 2.5: Fraction of reported earnings loss after job loss or job finding

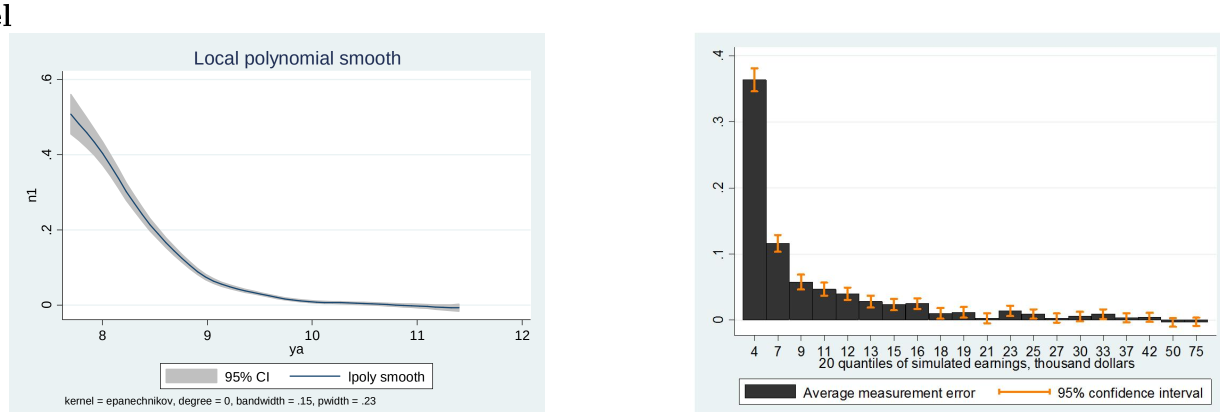


Job loss Job finding

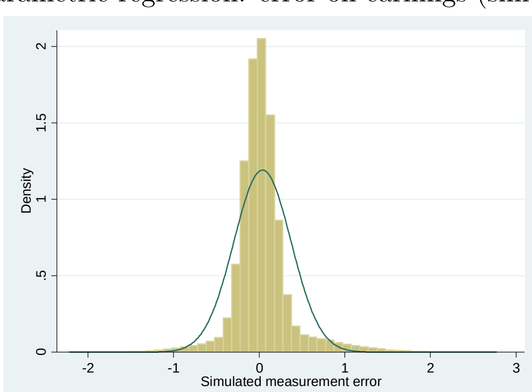
Histograms (between -0.5 and 1.5) of $\left(y^s_{it} - y^a_{i,t-1}\right) / \left(y^a_{it} - y^a_{i,t-1}\right)$ when year $t-1$ was stable, but year t involved a job loss or a job finding, and yearly earnings loss exceeded 20%. y^s_{it} and y^a_{it} are log earnings in the survey and in the administrative data.

Sample: HRS, 1991-2011, 50-60-year-old employees

Figure 2.6: Distribution of measurement error in simulated data based on the model



Nonparametric regression: error on earnings (simulated) Mean error by 20 quantiles of earnings (simulated)



Histogram of measurement error (simulated)

Parameters: $N = 100,000$; Earnings: $\mu = 10$; $\sigma^{perm} = 0.6$; $\sigma^{trans} = 0.05$; Pr(career interruption)= 0.2; $\gamma = 1.5$;

Error:$\sigma^{error}_{stable} = 0.15$; $\sigma^{error}_{interrupt}0.6$; Reported earnings:$\tau^{perm} = 1$; $\tau^{trans} = 0$; $\tau^{interruption} = 0.7$.

Table 2.1: Type of wage and salary reports

	All observations		50-60-year-old, employees	
	N	Fraction of total	N	Fraction of total
Continuous value	58,075	32.8	29,858	62.3
Complete bracket	5,181	2.9	2,476	5.2
Half open interval	1,576	0.9	501	1.0
No income	105,919	59.8	12,801	26.7
Missing	6,244	3.5	2,309	4.8
Total	176,995	100.0	47,945	100.0

Sample: HRS, 1991-2011

Table 2.2: Components of annual earnings, HRS 1991-2011, 50-60-year-old employees

	Fraction positive	Mean*	Fraction of total	N
Wage and salary income	99.2	36489	93.1	33,090
Tips, bonuses, commissions	10.1	1146	2.9	33,090
Professional practice and trade	3.3	1259	3.2	33,090
2nd job or military reserves	4.3	314	0.8	33,090
Total	100	39209	100.0	33,090

Average earnings adjusted to 2000 dollars with the CPI.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.3: Fraction of successful matches between the HRS and SSA

	N	Fraction matched
1. All observations	176,995	57.7
2. 50-60-year-old employees	47,945	61.2
3. (2) + non-missing HRS response	45,639	64.2
4. (3) + positive HRS response	33,090	66.2
5. (4) + 1992 wave	4673	88.2
6. (4) + 1994 wave	3681	64.1
7. (4) + 1996 wave	2959	65.4
8. (4) + 1998 wave	3403	73.1
9. (4) + 2000 wave	2590	70.6
10. (4) + 2002 wave	1854	71.2
11. (4) + 2004 wave	2703	73.1
12. (4) + 2006 wave	2314	60.2
13. (4) + 2008 wave	1916	60.1
14. (4) + 2010 wave	3798	48.7
15. (4) + 2012 wave	3199	46.4

Sample: HRS, 1991-2011

Table 2.4: OLS of successful matches between the HRS and SSA

	Matched
Female	0.008
	[0.008]
Less than high school	-0.014
	[0.013]
Some college	-0.018
	[0.014]
BA	-0.009
	[0.017]
Postgraduate	0.015
	[0.018]
Age - 50	-0.007**
	[0.003]
Age - 50, sq	0.001***
	[0.000]
Partnered	0.050***
	[0.019]
Single	-0.023**
	[0.010]
Black	-0.082***
	[0.012]
Hispanic	-0.077***
	[0.016]
Salaried worker	-0.006
	[0.009]
log HRS earnings	-0.003
	[0.004]
1994 wave	-0.244***
	[0.008]
1996 wave	-0.234***
	[0.010]
1998 wave	-0.156***
	[0.009]
2000 wave	-0.184***
	[0.011]
2002 wave	-0.179***
	[0.012]
2004 wave	-0.152***
	[0.010]
2006 wave	-0.282***
	[0.012]
2008 wave	-0.285***
	[0.013]
2010 wave	-0.386***
	[0.010]
2012 wave	-0.408***
	[0.011]
Constant	0.950***
	[0.043]
Observations	33,090
R-squared	0.078

Robust standard errors clustered on the household level in brackets.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.5: **HRS earnings reports and W2 earnings records**

	HRS earnings reports						
W2 earnings records	Zero	\$1-\$1,000	\$1,001-\$5,000	\$5,001-\$10,000	\$10,001+	Missing	Total
Zero	6,118	53	130	67	242	10	6620
\$1-\$1,000	367	91	52	11	21	<3	543
\$1,001-\$5,000	341	37	485	158	174	7	1202
\$5,001-\$10,000	261	15	157	797	593	16	1839
\$10,001+	921	45	118	317	18373	166	19940
Missing	6,344	103	556	640	9,855	303	17801
Total	14,352	344	1,498	1,990	29,258	503	47,945

Earnings are deflated to 2000 dollars with the CPI.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.6: **Reliability ratios and other moments of earnings and error**

	Log earnings				Earnings growth
	All observations	$Y > \$1k$	$Y > \$5k$	$Y > \$10k$	All observations
Reliability ratio, classical	0.822***	0.812***	0.785***	0.748***	0.502***
	[0.008]	[0.009]	[0.012]	[0.016]	[0.024]
Reliability ratio, correlated	0.982***	0.920***	0.834***	0.777***	0.682***
	[0.015]	[0.013]	[0.015]	[0.018]	[0.044]
γ_{m,y^a}	-0.099***	-0.073***	-0.038***	-0.025***	-0.263***
	[0.007]	[0.005]	[0.005]	[0.005]	[0.021]
Raw moments					
Mean earnings	10.167***	10.202***	10.296***	10.411***	-0.003
	[0.010]	[0.009]	[0.008]	[0.008]	[0.004]
Mean error	0.033***	0.026***	0.012***	0.004	-0.005
	[0.003]	[0.003]	[0.003]	[0.003]	[0.003]
Variance of earnings	0.849***	0.701***	0.502***	0.369***	0.255***
	[0.018]	[0.012]	[0.008]	[0.006]	[0.014]
Variance of error	0.183***	0.163***	0.138***	0.125***	0.253***
	[0.010]	[0.009]	[0.009]	[0.010]	[0.022]
Covariance of earnings and error	-0.084***	-0.051***	-0.019***	-0.009***	-0.067***
	[0.006]	[0.004]	[0.002]	[0.002]	[0.006]
Observations	21,394	21,219	20,365	18,805	11,698

Earnings growth is the log difference of earnings in year t and $t-2$.

Robust standard errors clustered on the household level in brackets; *** p<0.01, ** p<0.05, * p<0.1.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.7: Detailed labor histories created from the SSA data

	Matched sample		Including off years	
Detailed labor history	N	Fraction	N	Fraction
Single job, full year earnings	14,570	48.3	28,923	48.8
Multiple job, full year earnings	922	3.1	1850	3.1
Single job loss	1,183	3.9	2,521	4.3
Major job loss	187	0.6	363	0.6
Minor job loss	336	1.1	683	1.2
Single job finding	852	2.8	1734	2.9
Major job finding occurs	174	0.6	336	0.6
Minor job finding occurs	275	0.9	554	0.9
Minor job loss, minor job finding	577	1.9	1111	1.9
Major job loss, minor job finding	274	0.9	574	1.0
Minor job loss, major job finding	310	1.0	634	1.1
Major job loss, major job finding	1,711	5.7	3,513	5.9
Zero earnings	6,603	21.9	14,332	24.2
Last observed year	2,170	7.2	2,176	3.7
Total	30,144	100.0	59,304	100.0

The second sample includes the off years of the survey, when HRS reports are not available.

Major job loss: highest paid job at t or $t-1$

Major job finding: highest paid job at t or $t+1$

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.8: Detailed labor histories and distribution of earnings

	W2 earnings records, row percentages					
Detailed labor history	Zero	$\$1-\$1k$	$\$1k-\$5k$	$\$5k-\$10k$	$\$10k+$	Total
Single job, full year earnings	0	1.1	2.3	5.6	91.0	100.0
Multiple job, full year earnings	0	0.3	1.1	5.1	93.5	100.0
Single job loss	0	10.7	18.4	14.6	56.2	100.0
Major job loss	0	2.1	9.1	12.8	75.9	100.0
Minor job loss	0	0.3	3.9	5.7	90.2	100.0
Single job finding	0	9.2	21.4	16.4	53.1	100.0
Major job finding occurs	0	1.1	10.9	14.9	73.0	100.0
Minor job finding occurs	0	0.0	2.5	8.0	89.5	100.0
Minor job loss, minor job finding	0	0.3	3.6	8.7	87.3	100.0
Major job loss, minor job finding	0	2.2	13.5	16.1	68.2	100.0
Minor job loss, major job finding	0	1.0	16.1	19.4	63.5	100.0
Major job loss, major job finding	0	7.5	11.3	15.0	66.1	100.0
Zero earnings	100	0.0	0.0	0.0	0.0	100.0
Last observed year	1	1.3	4.4	7.7	85.8	100.0
Total	22	1.8	4.0	6.1	66.1	100.0

Earnings are deflated to 2000 dollars with the CPI.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.9: Detailed labor histories and distribution of earnings

	W2 earnings records, number of cases					
Detailed labor history	Zero	$\$1-\$1k$	$\$1k-\$5k$	$\$5k-\$10k$	$\$10k+$	Total
Single job, full year earnings	0	160	339	810	13,261	14,570
Multiple job, full year earnings	0	3	10	47	862	922
Single job loss	0	127	218	173	665	1,183
Major job loss	0	4	17	24	142	187
Minor job loss	0	1	13	19	303	336
Single job finding	0	78	182	140	452	852
Major job finding occurs	0	2	19	26	127	174
Minor job finding occurs	0	0	7	22	246	275
Minor job loss, minor job finding	0	2	21	50	504	577
Major job loss, minor job finding	0	6	37	44	187	274
Minor job loss, major job finding	0	3	50	60	197	310
Major job loss, major job finding	0	129	194	256	1,131	1,710
Zero earnings	6,603	0	0	0	0	6,603
Last observed year	17	28	95	168	1,863	2,171
Total	6,620	543	1,202	1,839	19,940	30,144

Earnings are deflated to 2000 dollars with the CPI.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.10: Detailed labor histories and distribution of earnings

	HRS earnings reports, row percentages						
Aggregate labor history	Zero	$\$1-\$1k$	$\$1k-\$5k$	$\$5k-\$10k$	$\$10k+$	Missing	Total
Stable year	5.7	0.3	2.1	4.2	86.9	0.8	100.0
Major job loss	25.8	2.1	7.0	8.4	55.8	0.9	100.0
Major job finding	10.3	3.6	11.4	11.0	62.9	0.8	100.0
Major job loss and job finding	14.3	2.1	7.7	9.0	66.3	0.6	100.0
Zero earning	92.6	0.8	2.0	1.0	3.5	0.2	100.0
Total	28.2	0.8	3.1	4.3	63.0	0.6	100.0
	HRS earnings reports, number of cases						
	Zero	$\$1-\$1k$	$\$1k-\$5k$	$\$5k-\$10k$	$\$10k+$	Missing	Total
Stable year	954	55	344	698	14,500	129	16,680
Major job loss	424	35	115	138	918	14	1,644
Major job finding	138	48	152	147	840	11	1,336
Major job loss and job finding	244	36	131	154	1,134	11	1,710
Zero earning	6,117	53	129	64	230	10	6,603
Total	7,877	227	871	1,201	17,622	175	27,973

Earnings are deflated to 2000 dollars with the CPI.

Stable years: Positive earnings and no evidence of major job loss or job finding.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.11: Measurement error by labor histories and earnings

Mean measurement error in earnings	W2 earnings records				
Aggregate labor history	$\$1-\$1k$	$\$1k-\$5k$	$\$5k-\$10k$	$\$10k+$	Total
Stable year	0.554	0.239	0.073	0.004	0.013
Major job loss	0.849	0.383	0.113	-0.011	0.078
Major job finding	0.942	0.454	0.173	0.053	0.184
Major job loss and job finding	0.972	0.419	0.120	0.001	0.089
Total	0.867	0.354	0.099	0.005	0.033

Earnings are deflated to 2000 dollars with the CPI.

Stable years: Positive earnings and no evidence of major job loss or job finding.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.12: Measurement error by earnings in stable years

	W2 earnings records, mean measurement error			
Sign of change in earnings since last year	$\$1-\$5k$	$\$5k-\$10k$	$\$10k+$	Total
Negative	0.365	0.128	0.034	0.053
Non-negative	0.153	0.037	-0.008	-0.003
Total	0.265	0.073	0.004	0.013
	W2 earnings records, number of cases			
Sign of change in earnings since last year	$\$1-\$5k$	$\$5k-\$10k$	$\$10k+$	Total
Negative	168	333	4,018	4,519
Non-negative	150	499	10,399	11,048
Total	318	832	14,417	15,567

Earnings are deflated to 2000 dollars with the CPI.

Stable years: Positive earnings and no evidence of major job loss or job finding in year t or $t-1$.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.13: GMM estimates of basic moments of earnings and error

	All observations	Stable years
Mean earnings	10.230***	10.371***
	[0.012]	[0.013]
Mean error	0.031***	0.012***
	[0.004]	[0.004]
Permanent earnings variance	0.609***	0.479***
	[0.015]	[0.014]
transitory earnings variance	0.124***	0.029***
	[0.009]	[0.004]
Permanent error variance	0.021***	0.014***
	[0.002]	[0.002]
Transitory error variance	0.114***	0.076***
	[0.010]	[0.010]
Fraction of permanent earnings reported	0.955***	0.978***
	[0.006]	[0.006]
Fraction of transitory earnings reported	0.679***	0.764***
	[0.027]	[0.051]
N	13,950	7,956

Stable years: Positive earnings and no evidence of major job loss or job finding in year t or $t+1$ or $t+2$.

Robust standard errors clustered on the household level in brackets; *** p<0.01, ** p<0.05, * p<0.1.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.14: MCMC estimates in the baseline specification, covariates in earnings

		Males			Females	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	10.655***	10.688***	10.613***	10.150***	9.911***	9.927***
	[0.011]	[0.027]	[0.020]	[0.012]	[0.030]	[0.020]
Black		-0.277***	-0.280***		-0.020	-0.025
		[0.029]	[0.029]		[0.028]	[0.028]
Hispanic		-0.316***	-0.293***		-0.102**	-0.076*
		[0.033]	[0.033]		[0.041]	[0.040]
Less than high school		-0.286***	-0.286***		-0.240***	-0.259***
		[0.033]	[0.032]		[0.037]	[0.036]
Some college		0.105***	0.111***		0.214***	0.218***
		[0.027]	[0.027]		[0.029]	[0.028]
BA		0.407***	0.406***		0.494***	0.486***
		[0.032]	[0.032]		[0.036]	[0.036]
More than BA		0.490***	0.489***		0.769***	0.754***
		[0.032]	[0.032]		[0.035]	[0.035]
Age - 50		0.011***	0.012***		0.011***	0.013***
		[0.002]	[0.002]		[0.002]	[0.002]
(Age - 50) squared		-0.000**	-0.000**		-0.000***	-0.000***
		[0.000]	[0.000]		[0.000]	[0.000]
Single		-0.015**	-0.017**		-0.004	-0.003
		[0.007]	[0.007]		[0.006]	[0.006]
Salaried worker		0.014***	0.012**		0.020***	0.020***
		[0.005]	[0.005]		[0.004]	[0.004]
GDP gap		0.400***			0.150***	
		[0.057]			[0.055]	
Year-1980		-0.005***			0.001	
		[0.001]			[0.002]	
Year dummies			YES			YES

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.15: MCMC estimates in the baseline specification, covariates in the innovation term of permanent earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Means						
Major job finding	-0.065***	-0.074***	-0.092***	-0.065***	-0.077***	-0.089***
	[0.015]	[0.013]	[0.014]	[0.014]	[0.014]	[0.013]
Major job loss and job finding	0.024**	0.007	0.015	0.073***	0.059***	0.070***
	[0.011]	[0.011]	[0.012]	[0.011]	[0.010]	[0.010]
Zero earnings	-0.275***	-0.237***	-0.240***	-0.226***	-0.205***	-0.220***
	[0.021]	[0.030]	[0.022]	[0.019]	[0.015]	[0.015]
Latent job loss in stable years	-0.212***	-0.214***	-0.207***	-0.185***	-0.171***	-0.191***
	[0.019]	[0.020]	[0.019]	[0.017]	[0.016]	[0.016]
Standard deviation						
Stable years	0.063***	0.063***	0.063***	0.069***	0.066***	0.064***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.001]	[0.001]
Major job finding	0.478***	0.433***	0.456***	0.457***	0.458***	0.418***
	[0.013]	[0.012]	[0.014]	[0.012]	[0.012]	[0.011]
Major job loss and job finding	0.413***	0.411***	0.416***	0.419***	0.419***	0.414***
	[0.010]	[0.010]	[0.010]	[0.011]	[0.010]	[0.011]
Zero earnings	0.785***	0.810***	0.797***	0.662***	0.631***	0.669***
	[0.035]	[0.034]	[0.038]	[0.025]	[0.024]	[0.024]
Latent job loss in stable years	0.834***	0.840***	0.823***	0.756***	0.719***	0.747***
	[0.022]	[0.025]	[0.026]	[0.018]	[0.019]	[0.017]
Standard deviation, raw coefficients						
Constant	-2.763***	-2.762***	-2.773***	-2.679***	-2.727***	-2.746***
	[0.021]	[0.021]	[0.022]	[0.020]	[0.018]	[0.019]
Major job finding	2.027***	1.924***	1.991***	1.894***	1.948***	1.875***
	[0.031]	[0.032]	[0.035]	[0.031]	[0.029]	[0.032]
Major job loss and job finding	1.879***	1.871***	1.894***	1.807***	1.859***	1.865***
	[0.029]	[0.030]	[0.029]	[0.029]	[0.027]	[0.029]
Zero earnings	2.525***	2.555***	2.552***	2.270***	2.270***	2.349***
	[0.039]	[0.039]	[0.044]	[0.040]	[0.037]	[0.031]
Latent job loss in stable years	2.581***	2.584***	2.573***	2.397***	2.394***	2.453***
	[0.026]	[0.025]	[0.026]	[0.023]	[0.024]	[0.025]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.16: MCMC estimates in the baseline specification, other parameters in earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Permanent std, age 50	0.608***	0.520***	0.520***	0.652***	0.582***	0.576***
	[0.009]	[0.007]	[0.008]	[0.009]	[0.008]	[0.008]
Transitory std	0.041***	0.042***	0.040***	0.027***	0.030***	0.030***
	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]
Probability of latent job loss in stable years	0.107***	0.106***	0.108***	0.117***	0.119***	0.115***
	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.003]
Avg transitory earnings loss by labor history						
Major job loss	0.409***	0.407***	0.414***	0.454***	0.454***	0.446***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Major job finding	0.372***	0.377***	0.375***	0.428***	0.426***	0.429***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Major job loss and job finding	0.367***	0.371***	0.372***	0.416***	0.410***	0.420***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Latent job loss in stable years	0.231***	0.236***	0.237***	0.250***	0.245***	0.256***
	[0.007]	[0.007]	[0.007]	[0.005]	[0.005]	[0.005]
Raw coefficients of transitory earnings loss, γ_j						
Major job-loss	1.448***	1.457***	1.413***	1.204***	1.206***	1.242***
	[0.030]	[0.031]	[0.029]	[0.023]	[0.024]	[0.024]
Major job finding	1.688***	1.657***	1.663***	1.338***	1.351***	1.332***
	[0.037]	[0.037]	[0.037]	[0.027]	[0.028]	[0.027]
Major job loss and job finding	1.726***	1.698***	1.691***	1.403***	1.442***	1.382***
	[0.034]	[0.033]	[0.033]	[0.027]	[0.027]	[0.026]
Latent job loss in stable years	3.349***	3.258***	3.241***	3.013***	3.091***	2.919***
	[0.081]	[0.077]	[0.073]	[0.064]	[0.070]	[0.067]

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.17: MCMC estimates in the baseline specification, covariates in measurement error

	Males				Females	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	0.007*	-0.024*	-0.017	-0.005	-0.038**	-0.009
	[0.004]	[0.013]	[0.012]	[0.004]	[0.016]	[0.015]
Black		0.01	0.004		-0.007	-0.002
		[0.011]	[0.011]		[0.010]	[0.010]
Hispanic		-0.026**	-0.022*		-0.033**	-0.041***
		[0.013]	[0.013]		[0.014]	[0.014]
Less than high school		-0.016	-0.023**		-0.031***	-0.026**
		[0.011]	[0.011]		[0.012]	[0.012]
Some college		0.032***	0.027***		0.014	0.011
		[0.010]	[0.010]		[0.009]	[0.009]
BA		0.030***	0.028**		-0.01	-0.012
		[0.012]	[0.012]		[0.012]	[0.012]
More than BA		0.029**	0.023**		0.006	0.009
		[0.012]	[0.011]		[0.012]	[0.013]
Age - 50		0.006*	0.005		0.001	-0.001
		[0.003]	[0.003]		[0.004]	[0.004]
(Age - 50) squared		-0.000**	-0.000*		0	0
		[0.000]	[0.000]		[0.000]	[0.000]
Single		0.011	0.01		0.006	0
		[0.008]	[0.009]		[0.007]	[0.007]
Salaried worker		0.007	0.008		0.036***	0.038***
		[0.007]	[0.006]		[0.007]	[0.007]
GDP gap		0.093			0.07	
		[0.097]			[0.121]	
Year-1980		0			0.001**	
		[0.001]			[0.001]	

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.18: MCMC estimates in the baseline specification, other parameters in measurement error

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Std permanent error	0.156***	0.160***	0.154***	0.137***	0.147***	0.133***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.005]	[0.006]
Std transitory error						
Stable years	0.103***	0.099***	0.100***	0.164***	0.144***	0.172***
	[0.003]	[0.004]	[0.003]	[0.008]	[0.010]	[0.006]
Major job loss	0.549***	0.563***	0.537***	0.633***	0.649***	0.630***
	[0.015]	[0.018]	[0.016]	[0.022]	[0.021]	[0.020]
Major job finding	0.499***	0.499***	0.489***	0.579***	0.563***	0.564***
	[0.014]	[0.014]	[0.013]	[0.021]	[0.016]	[0.015]
Major job loss and job finding	0.611***	0.609***	0.573***	0.657***	0.649***	0.645***
	[0.020]	[0.020]	[0.016]	[0.022]	[0.021]	[0.019]
Latent job loss in stable years	0.658***	0.654***	0.654***	0.681***	0.682***	0.686***
	[0.026]	[0.023]	[0.025]	[0.027]	[0.033]	[0.035]
Fraction of earnings component reported in the survey						
Permanent earnings	0.971***	0.954***	0.949***	0.966***	0.949***	0.932***
	[0.005]	[0.006]	[0.006]	[0.004]	[0.005]	[0.005]
Transitory on-the-job	0	0	0	0	0	0
Transitory due to job-loss	0.867***	0.840***	0.866***	0.753***	0.771***	0.792***
	[0.010]	[0.010]	[0.009]	[0.011]	[0.010]	[0.012]
Transitory due to job finding	0.634***	0.619***	0.564***	0.724***	0.687***	0.689***
	[0.010]	[0.010]	[0.010]	[0.009]	[0.009]	[0.010]
Transitory due to job loss and job finding	0.527***	0.540***	0.609***	0.667***	0.652***	0.711***
	[0.012]	[0.010]	[0.009]	[0.010]	[0.009]	[0.010]
Transitory due to latent job loss in stable years	0.867***	0.495***	0.869***	0.810***	0.733***	1.086***
	[0.022]	[0.027]	[0.023]	[0.023]	[0.025]	[0.022]

The fraction of transitory reports are calibrated to zero in the baseline specification.

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

APPENDIX E
Derivation of the conditional mean of the error in Section 2.2.1

The earnings model can be summarized by the following equations:

$$y_i^a = y_i^{perm} + \varepsilon_i - I_i k, \tag{2.67}$$

$$\varepsilon_i \sim N(0, V_\varepsilon), \tag{2.68}$$

$$y_i^{perm} \sim N(\mu, V_p), \tag{2.69}$$

$$\mathbb{E}(I_i) = p, \tag{2.70}$$

$$m_i = (\tau - 1)(\varepsilon_i - I_i k) + \xi_i, \tag{2.71}$$

$$\sigma_y \equiv \sqrt{V_\varepsilon + V_p} \tag{2.72}$$

The last equation (2.72) is just a definition to simplify notation. The conditional mean of the error is

$$\mathbb{E}(m_i|y_i^a) = \mathbb{E}(m_i|y_i^a, I_i = 0)\Pr(I_i = 0|y_i^a) + \mathbb{E}(m_i|y_i^a, I_i = 1)\Pr(I_i = 1|y_i^a). \tag{2.73}$$

I shall derive each of the four terms on the right hand side of (2.73). The conditional distribution of earnings given I_i is

$$f(y_i|I_i = 1) = \frac{1}{\sigma_y}\phi\left(\frac{y_i^a + k - \mu}{\sigma_y}\right), \tag{2.74}$$

$$f(y_i|I_i = 0) = \frac{1}{\sigma_y}\phi\left(\frac{y_i^a - \mu}{\sigma_y}\right), \tag{2.75}$$

where $\phi(\cdot)$ denotes the standard normal p.d.f. The unconditional distribution of earnings is

$$f\left(y_i\right)=\frac{1-p}{\sigma_y}\phi\left(\frac{y_i^a-\mu}{\sigma_y}\right)+\frac{p}{\sigma_y}\phi\left(\frac{y_i^a+k-\mu}{\sigma_y}\right). \tag{2.76}$$

Now I use the Bayes Theorem to derive the distribution of I_i given y_i^a.

$$\begin{aligned}
\Pr\left(I_i=0|y_i^a\right) &= \frac{f\left(y_{it}^a|I_i=0\right)\Pr\left(I_i=0\right)}{f\left(y_{it}^a\right)}=\frac{\frac{1-p}{\sigma_y}\phi\left(\frac{y_i^a-\mu}{\sigma_y}\right)}{\frac{1-p}{\sigma_y}\phi\left(\frac{y_i^a-\mu}{\sigma_y}\right)+\frac{p}{\sigma_y}\phi\left(\frac{y_i^a+k-\mu}{\sigma_y}\right)} && (2.77)\\
&= \frac{\left(1-p\right)\phi\left(\frac{y_i^a-\mu}{\sigma_y}\right)}{\left(1-p\right)\phi\left(\frac{y_i^a-\mu}{\sigma_y}\right)+p\phi\left(\frac{y_i^a+k-\mu}{\sigma_y}\right)}, && (2.78)\\
&= \frac{\left(1-p\right)\exp\left(-\frac{\left(y_i^a-\mu\right)^2}{2\sigma_y^2}\right)}{\left(1-p\right)\exp\left(-\frac{\left(y_i^a-\mu\right)^2}{2\sigma_y^2}\right)+p\exp\left(-\frac{\left(y_i^a+k-\mu\right)^2}{2\sigma_y^2}\right)} && (2.79)\\
&= \frac{\left(1-p\right)\exp\left(-\frac{\left(y_i^a-\mu\right)^2}{2\sigma_y^2}\right)}{\left(1-p\right)\exp\left(-\frac{\left(y_i^a-\mu\right)^2}{2\sigma_y^2}\right)+p\exp\left(-\frac{\left(y_i^a-\mu\right)^2}{2\sigma_y^2}\right)\exp\left(-\frac{k\left(2\left(y_i^a-\mu\right)+k\right)}{2\sigma_y^2}\right)} && (2.80)\\
&= \frac{\left(1-p\right)}{\left(1-p\right)+p\exp\left(\frac{k}{\sigma_y^2}\left(\mu-y_i^a-\frac{k}{2}\right)\right)} && (2.81)\\
\Pr\left(I_i=1|y_i\right) &= \frac{p\exp\left(\frac{k}{\sigma_y^2}\left(\mu-y_i^a-\frac{k}{2}\right)\right)}{\left(1-p\right)+p\exp\left(\frac{k}{\sigma_y^2}\left(\mu-y_i^a-\frac{k}{2}\right)\right)}. && (2.82)
\end{aligned}$$

The expected value of the error conditional on y_i^a and I_i is

$$\begin{aligned}
\mathbb{E}\left(m_i|y_i^a,I_i=0\right) &= \mathbb{E}\left(\left(\tau-1\right)\varepsilon_i|y_i^{perm}+\varepsilon_i\right)=-\left(1-\tau\right)\frac{V_\varepsilon}{V_\varepsilon+V_p}\left(y_i^a-\mu\right), && (2.83)\\
\mathbb{E}\left(m_i|y_i^a,I_i=1\right) &= \mathbb{E}\left(\left(\tau-1\right)\left(\varepsilon_i-k\right)|y_i^{perm}+\varepsilon_i+k\right)=\left(1-\tau\right)\left(-\frac{V_\varepsilon}{V_\varepsilon+V_p}\left(y_i^a-\mu+k\right)+k\right) && (2.84)\\
&= \left(1-\tau\right)\left(-\frac{V_\varepsilon}{V_\varepsilon+V_p}\left(y_i^a-\mu\right)+\frac{V_p}{V_\varepsilon+V_p}k\right). && (2.85)
\end{aligned}$$

(2.83)-(2.85) are standard applications of the formulas for the multivariate normal distribution. Now we have all the pieces of (2.73).

$$\mathbb{E}\left(m_i|y_i^a\right) = (1-\tau)\left[-\frac{V_\varepsilon}{V_\varepsilon+V_p}\left(y_i^a-\mu\right)+k\frac{V_p}{V_\varepsilon+V_p}\frac{p\exp\left(\frac{k}{V_\varepsilon+V_p}\left(\mu-y_i^a-\frac{k}{2}\right)\right)}{(1-p)+p\exp\left(\frac{k}{V_\varepsilon+V_p}\left(\mu-y_i^a-\frac{k}{2}\right)\right)}\right] \quad (2.86)$$

APPENDIX F
Measurement error in the Selective Memory Model

Let us assume that log earnings is the sum of a permanent and a transitory component, each distributed normally.

$$\begin{aligned} y_i^a &= y_i^{perm} + \varepsilon_i, \qquad (2.87) \\ \begin{pmatrix} y_i^{perm} \\ \varepsilon_i \end{pmatrix} &\sim N\left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} V_p & 0 \\ 0 & V_\varepsilon \end{pmatrix}\right). \qquad (2.88) \end{aligned}$$

Persons report their permanent earnings accurately, but they report only a fraction of their transitory earnings, depending on the sign of shock.

$$\begin{aligned} y_i^s &= \begin{cases} y_i^{perm} + \tau_p \varepsilon_i + \xi_i & \text{if} \quad \varepsilon_i > 0 \\ y_i^{perm} + \tau_n \varepsilon_i + \xi_i & \text{if} \quad \varepsilon_i \leq 0 \end{cases}, \qquad (2.89) \\ \tau_p &> \tau_n \qquad (2.90) \end{aligned}$$

where ξ_i is a classical error component. The total measurement error is

$$m_i = \begin{cases} -(1-\tau_p)\,\varepsilon_i + \xi_i & \text{if} \quad \varepsilon_i > 0 \\ -(1-\tau_n)\,\varepsilon_i + \xi_i & \text{if} \quad \varepsilon_i \leq 0 \end{cases} \qquad (2.91)$$

The following theorems summarize the properties of the measurement error: It has a positive mean, a negative correlation with true earnings (y_i^a) and it is convex, implying that the error is larger at low earnings values.

Theorem 1. *Under the assumptions of the model average measurement error equals to*

$$\mathbb{E}(m_i) = \frac{\tau_p - \tau_n}{4}\sqrt{V_\varepsilon}\phi(0) \approx 0.1\sqrt{V_\varepsilon}(\tau_p - \tau_n) > 0. \tag{2.92}$$

Proof.

$$\begin{aligned}
\mathbb{E}(m_i) &= \mathbb{E}((\tau_p - 1)\varepsilon_i + \xi_i|\varepsilon_i > 0)\Pr(\varepsilon_i > 0) + \mathbb{E}((\tau_n - 1)\varepsilon_i + \xi_i|\varepsilon_i \leq 0)\Pr(\varepsilon_i \leq 0) & (2.93)\\
&= (\tau_p - 1)\mathbb{E}(\varepsilon_i|\varepsilon_i > 0)\Pr(\varepsilon_i > 0) + (\tau_n - 1)\mathbb{E}(\varepsilon_i|\varepsilon_i \leq 0)\Pr(\varepsilon_i \leq 0) & (2.94)\\
&= \frac{\tau_p - 1}{2}\mathbb{E}(\varepsilon_i|\varepsilon_i > 0) + \frac{\tau_n - 1}{2}\mathbb{E}(\varepsilon_i|\varepsilon_i \leq 0) & (2.95)\\
&= \frac{\tau_p - 1}{2}\sqrt{V_\varepsilon}\frac{\phi(0)}{1 - \Phi(0)} + \frac{\tau_n - 1}{2}\sqrt{V_\varepsilon}\frac{-\phi(0)}{\Phi(0)} & (2.96)\\
&= \frac{\tau_p - \tau_n}{4}\sqrt{V_\varepsilon}\phi(0) \approx 0.1\sqrt{V_\varepsilon}(\tau_p - \tau_n) & (2.97)
\end{aligned}$$

□

Theorem 2. *Under the assumptions of the model the conditional expectation of measurement error is*

$$\mathbb{E}(m_i|y_i^a) = (\tau_n - 1)\frac{V_\varepsilon}{V_\varepsilon + V_p}(y_i^a - \mu) + (\tau_p - \tau_n)[A + B(y_i^a - \mu)] \tag{2.98}$$

where

$$\begin{aligned}
A &= \sqrt{\frac{V_\varepsilon V_p}{V_\varepsilon + V_p}}\phi\left(\sqrt{\frac{V_\varepsilon}{V_p(V_\varepsilon + V_p)}}(y_i^a - \mu)\right) & (2.99)\\
B &= \frac{V_\varepsilon}{V_\varepsilon + V_p}\Phi\left(\sqrt{\frac{V_\varepsilon}{V_p(V_\varepsilon + V_p)}}(y_i^a - \mu)\right) & (2.100)
\end{aligned}$$

Proof.

$$\begin{aligned}
\mathbb{E}\left(m_i|y_i^a\right) &= \mathbb{E}\left(\left(\tau_p-1\right)\varepsilon_i+\xi_i|\varepsilon_i>0,y_i^a\right)\Pr\left(\varepsilon_i>0|y_i^a\right) \\
&\quad +\mathbb{E}\left(\left(\tau_n-1\right)\varepsilon_i+\xi_i|\varepsilon_i\leq 0,y^a\right)\Pr\left(\varepsilon_i\leq 0|y^a\right) \qquad (2.101) \\
&= \left(\tau_p-1\right)\mathbb{E}\left(\varepsilon_i|\varepsilon_i>0,y_i^a\right)\Pr\left(\varepsilon_i>0|y_i^a\right) \\
&\quad +\left(\tau_n-1\right)\mathbb{E}\left(\varepsilon_i|\varepsilon_i\leq 0,y_i^a\right)\Pr\left(\varepsilon_i\leq 0|y_i^a\right) \qquad (2.102)
\end{aligned}$$

I compute the different terms of (2.102) separately. Using the formulas for the multivariate normal distribution and the truncated normal distribution,

$$\begin{aligned}
\mathbb{E}\left(\varepsilon_i|\varepsilon_i>0,y_i^a\right) &= \left(\frac{V_\varepsilon}{V_\varepsilon+V_p}\left(y_i^a-\mu\right)+\frac{A}{\Phi\left(\sqrt{\frac{V_\varepsilon}{V_p(V_\varepsilon+V_p)}}\left(y_i^a-\mu\right)\right)}\right), \qquad (2.103) \\
A &\equiv \sqrt{\frac{V_\varepsilon V_p}{V_\varepsilon+V_p}}\phi\left(\sqrt{\frac{V_\varepsilon}{V_p\left(V_\varepsilon+V_p\right)}}\left(y_i^a-\mu\right)\right), \qquad (2.104) \\
\mathbb{E}\left(\varepsilon_i|\varepsilon_i\leq 0,y_i^a\right) &= \left(\frac{V_\varepsilon}{V_\varepsilon+V_p}\left(y_i^a-\mu\right)-\frac{A}{\Phi\left(-\sqrt{\frac{V_\varepsilon}{V_p(V_\varepsilon+V_p)}}\left(y_i^a-\mu\right)\right)}\right), \qquad (2.105) \\
\Pr\left(\varepsilon_i>0|y_i^a\right) &= \Phi\left(\sqrt{\frac{V_\varepsilon}{V_p\left(V_\varepsilon+V_p\right)}}\left(y_i^a-\mu\right)\right), \qquad (2.106) \\
\Pr\left(\varepsilon_i\leq 0|y_i^a\right) &= \Phi\left(-\sqrt{\frac{V_\varepsilon}{V_p\left(V_\varepsilon+V_p\right)}}\left(y_i^a-\mu\right)\right). \qquad (2.107)
\end{aligned}$$

By plugging (2.103)-(2.107) into (2.102),

$$\begin{aligned}
\mathbb{E}\left(m_i|y_i^a\right) &= \left(\tau_p-1\right)\left(\frac{V_\varepsilon}{V_\varepsilon+V_p}\left(y_i^a-\mu\right)\Phi\left(\sqrt{\frac{V_\varepsilon}{V_p\left(V_\varepsilon+V_p\right)}}\left(y_i^a-\mu\right)\right)+A\right) && (2.108)\\
&\quad+\left(\tau_n-1\right)\left(\frac{V_\varepsilon}{V_\varepsilon+V_p}\left(y_i^a-\mu\right)\left(1-\Phi\left(\sqrt{\frac{V_\varepsilon}{V_p\left(V_\varepsilon+V_p\right)}}\left(y_i^a-\mu\right)\right)\right)-A\right) && (2.109)\\
&= \left(\tau_n-1\right)\frac{V_\varepsilon}{V_\varepsilon+V_p}\left(y_i^a-\mu\right)+\left(\tau_p-\tau_n\right)\left[A+B\left(y_i^a-\mu\right)\right] && (2.110)\\
B &\equiv \frac{V_\varepsilon}{V_\varepsilon+V_p}\Phi\left(\sqrt{\frac{V_\varepsilon}{V_p\left(V_\varepsilon+V_p\right)}}\left(y_i^a-\mu\right)\right) && (2.111)
\end{aligned}$$

□

Theorem 3. *Under the assumptions of the model, the derivative of the conditional expectation is*

$$\frac{\partial}{\partial y^a}\mathbb{E}\left(m_i|y_i^a\right)=\left(\tau_n-1\right)\frac{V_\varepsilon}{V_\varepsilon+V_p}+\left(\tau_p-\tau_n\right)\frac{V_\varepsilon}{V_\varepsilon+V_p}\Phi\left(\sqrt{\frac{V_\varepsilon}{V_p\left(V_\varepsilon+V_p\right)}}\left(y_i^a-\mu\right)\right)$$

Proof. Let C denote $C \equiv \frac{V_\varepsilon}{V_\varepsilon + V_p}$.

$$\begin{aligned}
\frac{\partial}{\partial y_i^a}\mathbb{E}\left(m_i|y_i^a\right) &= \left(\tau_n - 1\right)C \\
&\quad + \left(\tau_p - \tau_n\right)\sqrt{CV_p}\phi'\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right)\sqrt{\frac{C}{V_p}} \\
&\quad + \left(\tau_p - \tau_n\right)\left[C\Phi\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right)\right] \\
&\quad + \left(\tau_p - \tau_n\right)\left[C\left(y_i^a - \mu\right)\phi\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right)\sqrt{\frac{C}{V_p}}\right] \\
&= \left(\tau_n - 1\right)C \\
&\quad + \left(\tau_p - \tau_n\right)C\left[\phi'\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right) + \Phi\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right)\right] \\
&\quad + \left(\tau_p - \tau_n\right)\left[C\left(y_i^a - \mu\right)\phi\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right)\sqrt{\frac{C}{V_p}}\right] \\
&= \left(\tau_n - 1\right)C \\
&\quad + \left(\tau_p - \tau_n\right)C\left[\left(-\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right)\phi\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right) + \Phi\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right)\right] \\
&\quad + \left(\tau_p - \tau_n\right)\left[C\left(y_i^a - \mu\right)\phi\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right)\sqrt{\frac{C}{V_p}}\right] \\
&= \left(\tau_n - 1\right)C + \left(\tau_p - \tau_n\right)C\Phi\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right)
\end{aligned}$$

□

If $\tau_p > \tau_n$, the derivative reaches its maximum as $y_i^a \to \infty$ and its value is

$$\max \frac{\partial}{\partial y_i^a}\mathbb{E}\left(m_i|y_i^a\right) = \left(\tau_p - 1\right)C < 0$$

Thus the function is always downward sloping and the slope is closest to zero at the high end of the earnings distribution.

Theorem 4. *Under the assumptions of the model, the second derivative of the conditional expectations is*

$$\frac{\partial^2}{\partial \left(y_i^a\right)^2}\mathbb{E}\left(m_i|y_i^a\right) = \left(\tau_p - \tau_n\right)\frac{C^{3/2}}{\sqrt{V_p}}\phi\left(\sqrt{\frac{C}{V_p}}\left(y_i^a - \mu\right)\right) > 0$$

Proof. It is straightforward after theorem (3) □

Therefore the function is negative sloped and convex. We can also see that the function becomes linear only if $\tau_p = \tau_n$, that is, the rate of forgetting does not depend on the sign of the transitory income shock.

APPENDIX G
Additional tables

Table 2.19: **Reliability ratios and other moments of earnings and error**

Log earnings	Males	Females	HS dropouts	HS	Some college	College
Reliability ratio, classical	0.792***	0.819***	0.747***	0.811***	0.812***	0.823***
	[0.014]	[0.010]	[0.021]	[0.015]	[0.017]	[0.017]
Reliability ratio, correlated	0.983***	0.987***	0.981***	0.982***	1.019***	0.995***
	[0.030]	[0.018]	[0.043]	[0.027]	[0.035]	[0.030]
γ_{m,y^a}	-0.122***	-0.104***	-0.160***	-0.108***	-0.125***	-0.105***
	[0.013]	[0.009]	[0.023]	[0.011]	[0.015]	[0.014]
Raw moments						
Mean earnings	10.485***	9.934***	9.703***	9.997***	10.198***	10.647***
	[0.013]	[0.013]	[0.021]	[0.015]	[0.019]	[0.018]
Mean error	0.041***	0.027***	-0.001	0.022***	0.052***	0.049***
	[0.005]	[0.004]	[0.010]	[0.005]	[0.007]	[0.006]
Variance of earnings	0.675***	0.848***	0.767***	0.716***	0.757***	0.761***
	[0.024]	[0.023]	[0.042]	[0.026]	[0.036]	[0.036]
Variance of error	0.177***	0.188***	0.260***	0.167***	0.175***	0.164***
	[0.016]	[0.012]	[0.029]	[0.016]	[0.019]	[0.019]
Covariance of earnings and error	-0.083***	-0.088***	-0.122***	-0.077***	-0.095***	-0.080***
	[0.010]	[0.008]	[0.021]	[0.009]	[0.013]	[0.011]
Observations	9,053	12,341	3,336	7,384	5,151	5,523

Robust standard errors clustered on the household level in brackets; *** p<0.01, ** p<0.05, * p<0.1.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.20: **Reliability ratios and other moments of earnings and error**

Earnings growth	Males	Females	HS dropouts	HS	Some college	College
Reliability ratio, classical	0.489***	0.508***	0.509***	0.504***	0.526***	0.469***
	[0.032]	[0.031]	[0.048]	[0.043]	[0.030]	[0.055]
Reliability ratio, correlated	0.691***	0.677***	0.706***	0.672***	0.830***	0.571***
	[0.067]	[0.055]	[0.095]	[0.076]	[0.080]	[0.080]
γ_{m,y^a}	-0.298***	-0.245***	-0.273***	-0.248***	-0.348***	-0.191***
	[0.044]	[0.023]	[0.061]	[0.031]	[0.039]	[0.045]
Raw moments						
Mean earnings growth	-0.013**	0.004	0.022*	-0.001	0.003	-0.023***
	[0.006]	[0.005]	[0.013]	[0.007]	[0.008]	[0.008]
Mean error growth	-0.001	-0.008*	-0.008	-0.010*	-0.001	0
	[0.006]	[0.004]	[0.012]	[0.006]	[0.006]	[0.006]
Variance of earnings growth	0.224***	0.274***	0.348***	0.254***	0.245***	0.216***
	[0.020]	[0.018]	[0.050]	[0.023]	[0.024]	[0.023]
Variance of error growth	0.234***	0.265***	0.336***	0.250***	0.221***	0.245***
	[0.027]	[0.030]	[0.054]	[0.039]	[0.024]	[0.050]
Covariance of earnings and error growth	-0.067***	-0.067***	-0.095***	-0.063***	-0.085***	-0.041***
	[0.011]	[0.008]	[0.024]	[0.009]	[0.014]	[0.010]
Observations	4,564	7,134	1,592	4,112	2,873	3,121

Robust standard errors clustered on the household level in brackets; *** p<0.01, ** p<0.05, * p<0.1.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.21: GMM estimates of basic moments of earnings and error by age

	Means		Earnings variance		Error variance		Fraction reported		N
Age	Earnings	Error	Permanent	Transitory	Permanent	Transitory	Permanent	Transitory	
50-51	10.225***	0.022**	0.581***	0.083***	0.016**	0.089***	0.961***	0.592***	1,485
	[0.022]	[0.009]	[0.028]	[0.018]	[0.006]	[0.018]	[0.013]	[0.097]	
52-53	10.238***	0.045***	0.612***	0.160***	0.021***	0.093***	0.965***	0.695***	2,339
	[0.019]	[0.007]	[0.025]	[0.026]	[0.004]	[0.011]	[0.010]	[0.059]	
54-55	10.266***	0.036***	0.584***	0.129***	0.022***	0.123***	0.941***	0.605***	2,865
	[0.016]	[0.007]	[0.023]	[0.021]	[0.005]	[0.029]	[0.013]	[0.061]	
56-57	10.247***	0.032***	0.607***	0.103***	0.026***	0.116***	0.951***	0.677***	3,091
	[0.016]	[0.007]	[0.024]	[0.017]	[0.005]	[0.020]	[0.012]	[0.064]	
58-59	10.187***	0.026***	0.614***	0.154***	0.01	0.107***	0.963***	0.734***	2,881
	[0.017]	[0.006]	[0.026]	[0.016]	[0.006]	[0.011]	[0.012]	[0.051]	
60-61	10.172***	0.024***	0.665***	0.093***	0.020***	0.154***	0.952***	0.756***	2,440
	[0.018]	[0.008]	[0.028]	[0.022]	[0.007]	[0.038]	[0.013]	[0.094]	
62-63	10.013***	0.059***	0.784***	0.178***	0.022**	0.122***	0.951***	0.729***	1,637
	[0.024]	[0.009]	[0.042]	[0.026]	[0.009]	[0.014]	[0.015]	[0.074]	
64-65	9.765***	0.073***	0.915***	0.252***	0.044***	0.197***	0.906***	0.787***	1,081
	[0.033]	[0.015]	[0.051]	[0.037]	[0.012]	[0.052]	[0.018]	[0.060]	
66-67	9.560***	0.087***	0.955***	0.310***	0.027**	0.194***	0.950***	0.727***	723
	[0.042]	[0.017]	[0.067]	[0.054]	[0.011]	[0.028]	[0.021]	[0.085]	
68-69	9.367***	0.088***	0.801***	0.307***	0.066*	0.156***	0.927***	0.805***	528
	[0.046]	[0.020]	[0.075]	[0.081]	[0.037]	[0.054]	[0.044]	[0.115]	

Robust standard errors clustered on the household level in brackets; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Sample: HRS, 1991-2011, 50-70-year-old employees

Table 2.22: GMM estimates of basic moments of earnings and error by gender and education

	Means		Earnings variance		Error variance		Fraction reported		N
Age	Earnings	Error	Permanent	Transitory	Permanent	Transitory	Permanent	Transitory	
All	10.230***	0.031***	0.609***	0.124***	0.021***	0.114***	0.955***	0.679***	13,950
	[0.012]	[0.004]	[0.015]	[0.009]	[0.002]	[0.010]	[0.006]	[0.027]	
Males	10.563***	0.043***	0.476***	0.107***	0.027***	0.091***	0.926***	0.627***	5,780
	[0.016]	[0.005]	[0.021]	[0.013]	[0.004]	[0.011]	[0.012]	[0.051]	
Females	9.994***	0.023***	0.573***	0.134***	0.017***	0.128***	0.954***	0.688***	8,170
	[0.015]	[0.005]	[0.019]	[0.011]	[0.003]	[0.015]	[0.008]	[0.032]	
HS dropouts	9.774***	-0.003	0.483***	0.186***	0.026***	0.133***	0.931***	0.743***	1,991
	[0.027]	[0.010]	[0.030]	[0.031]	[0.008]	[0.017]	[0.020]	[0.068]	
HS	10.044***	0.020***	0.484***	0.137***	0.011***	0.112***	0.961***	0.717***	4,886
	[0.018]	[0.006]	[0.020]	[0.014]	[0.004]	[0.018]	[0.008]	[0.037]	
Some college	10.256***	0.053***	0.509***	0.127***	0.023***	0.125***	0.924***	0.628***	3,343
	[0.022]	[0.008]	[0.029]	[0.016]	[0.005]	[0.019]	[0.015]	[0.046]	
BA or more	10.695***	0.044***	0.525***	0.083***	0.035***	0.101***	0.910***	0.443***	3,718
	[0.021]	[0.007]	[0.037]	[0.013]	[0.011]	[0.021]	[0.016]	[0.091]	

Robust standard errors clustered on the household level in brackets; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.23: GMM estimates of basic moments of earnings and error by gender and education in stable years

	Means		Earnings variance		Error variance		Fraction reported		N
Age	Earnings	Error	Permanent	Transitory	Permanent	Transitory	Permanent	Transitory	
All	10.371***	0.012***	0.479***	0.029***	0.014***	0.076***	0.978***	0.764***	7,956
	[0.013]	[0.004]	[0.014]	[0.004]	[0.002]	[0.010]	[0.006]	[0.051]	
Males	10.680***	0.022***	0.374***	0.025***	0.012*	0.056***	0.973***	0.636***	3,346
	[0.017]	[0.005]	[0.020]	[0.006]	[0.007]	[0.008]	[0.014]	[0.119]	
Females	10.145***	0.006	0.440***	0.032***	0.012***	0.089***	0.973***	0.774***	4,610
	[0.016]	[0.005]	[0.017]	[0.005]	[0.004]	[0.017]	[0.009]	[0.054]	
HS dropouts	9.943***	-0.017*	0.347***	0.039***	0.029***	0.084***	0.955***	0.754***	1,054
	[0.029]	[0.010]	[0.024]	[0.011]	[0.008]	[0.016]	[0.016]	[0.178]	
HS	10.173***	-0.001	0.385***	0.026***	0.014***	0.091***	0.969***	0.741***	2,715
	[0.019]	[0.007]	[0.017]	[0.004]	[0.005]	[0.021]	[0.010]	[0.085]	
Some college	10.393***	0.027***	0.385***	0.042***	0.009*	0.057***	0.972***	0.788***	1,877
	[0.023]	[0.007]	[0.028]	[0.010]	[0.005]	[0.012]	[0.013]	[0.062]	
BA or more	10.788***	0.029***	0.409***	0.028***	0.024***	0.077***	0.928***	0.469***	2,302
	[0.021]	[0.007]	[0.034]	[0.008]	[0.009]	[0.024]	[0.022]	[0.147]	

Stable years: Positive earnings and no evidence of major job loss or job finding in year t or $t+1$ or $t+2$.

Robust standard errors clustered on the household level in brackets; *** p<0.01, ** p<0.05, * p<0.1.

Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.24: MCMC estimates by education, covariates in earnings

	Less than HS	HS	Some college	BA or more
Constant	10.410***	10.748***	10.883***	11.013***
	[0.051]	[0.040]	[0.051]	[0.048]
Black	-0.250***	-0.300***	-0.205***	-0.433***
	[0.060]	[0.045]	[0.060]	[0.072]
Hispanic	-0.273***	-0.344***	-0.285***	-0.342***
	[0.056]	[0.062]	[0.070]	[0.089]
Age - 50	0.013**	0.010***	0.013***	0.007*
	[0.005]	[0.004]	[0.004]	[0.004]
(Age - 50) squared	0.000	0.000	0.000	0.000
	[0.000]	[0.000]	[0.000]	[0.000]
Single	-0.005	-0.026*	-0.016	-0.017
	[0.015]	[0.015]	[0.016]	[0.013]
Salaried worker	0.007	0.006	0.014	0.024***
	[0.014]	[0.010]	[0.010]	[0.009]
GDP gap	0.643***	0.556***	0.394***	0.183**
	[0.161]	[0.111]	[0.131]	[0.083]
Year-1980	-0.007**	-0.007***	-0.011***	0.002
	[0.003]	[0.002]	[0.003]	[0.003]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old male employees

Table 2.25: MCMC estimates by education, covariates in the innovation term of permanent earnings

	Less than HS	HS	Some college	BA or more
Means				
Major job finding	-0.072***	-0.101***	-0.086***	-0.064**
	[0.019]	[0.031]	[0.028]	[0.029]
Major job loss and job finding	0.005	0.009	0.018	0
	[0.018]	[0.021]	[0.018]	[0.024]
Zero earnings	-0.235***	-0.243***	-0.264***	-0.266***
	[0.036]	[0.049]	[0.052]	[0.043]
Latent job loss in stable years	-0.209***	-0.207***	-0.217***	-0.186***
	[0.055]	[0.036]	[0.040]	[0.029]
Standard deviation				
Stable years	0.062***	0.067***	0.068***	0.054***
	[0.004]	[0.003]	[0.004]	[0.002]
Major job finding	0.211***	0.575***	0.430***	0.507***
	[0.020]	[0.025]	[0.022]	[0.027]
Major job loss and job finding	0.298***	0.457***	0.336***	0.481***
	[0.019]	[0.019]	[0.016]	[0.020]
Zero earnings	0.725***	0.892***	0.829***	0.750***
	[0.052]	[0.057]	[0.067]	[0.061]
Latent job loss in stable years	0.879***	0.843***	0.809***	0.779***
	[0.045]	[0.036]	[0.037]	[0.033]
Standard deviation, raw coefficients				
Constant	-2.782***	-2.713***	-2.702***	-2.930***
	[0.057]	[0.036]	[0.046]	[0.038]
Major job finding	1.218***	2.160***	1.857***	2.250***
	[0.106]	[0.052]	[0.071]	[0.067]
Major job loss and job finding	1.566***	1.926***	1.610***	2.197***
	[0.084]	[0.054]	[0.063]	[0.053]
Zero earnings	2.461***	2.601***	2.521***	2.647***
	[0.093]	[0.066]	[0.085]	[0.088]
Latent job loss in stable years	2.652***	2.539***	2.486***	2.672***
	[0.072]	[0.050]	[0.055]	[0.046]

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample: HRS, 1991-2011, 50-60-year-old male employees

Table 2.26: MCMC estimates by education, other parameters in earnings

	Less than HS	HS	Some college	BA or more
Permanent std, age 50	0.508***	0.463***	0.523***	0.582***
	[0.017]	[0.012]	[0.015]	[0.015]
Transitory std	0.054***	0.045***	0.040***	0.029***
	[0.003]	[0.002]	[0.003]	[0.002]
Probability of latent job loss in stable years	0.087***	0.102***	0.115***	0.120***
	[0.007]	[0.005]	[0.007]	[0.006]
Avg transitory earnings loss by labor history				
Major job loss	0.408***	0.413***	0.436***	0.384***
	[0.011]	[0.009]	[0.011]	[0.009]
Major job finding	0.395***	0.369***	0.370***	0.371***
	[0.012]	[0.010]	[0.011]	[0.010]
Major job loss and job finding	0.422***	0.383***	0.362***	0.309***
	[0.011]	[0.009]	[0.010]	[0.009]
Latent job loss in stable years	0.236***	0.214***	0.263***	0.261***
	[0.012]	[0.010]	[0.011]	[0.010]
Raw coefficients of transitory earnings loss, γ_j				
Major job-loss	1.460***	1.422***	1.296***	1.609***
	[0.067]	[0.052]	[0.057]	[0.063]
Major job finding	1.533***	1.714***	1.705***	1.702***
	[0.075]	[0.069]	[0.080]	[0.071]
Major job loss and job finding	1.376***	1.607***	1.764***	2.242***
	[0.057]	[0.057]	[0.073]	[0.087]
Latent job loss in stable years	3.239***	3.698***	2.823***	2.855***
	[0.199]	[0.164]	[0.130]	[0.110]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old male employees

Table 2.27: MCMC estimates by education, covariates in measurement error

	Less than HS	HS	Some college	BA or more
Constant	-0.065	-0.015	0.006	0.012
	[0.041]	[0.020]	[0.026]	[0.030]
Black	-0.057***	-0.011	0.035	0.077***
	[0.021]	[0.017]	[0.023]	[0.027]
Hispanic	-0.066***	-0.049**	0.017	0.102***
	[0.019]	[0.023]	[0.026]	[0.039]
Age - 50	0.022**	-0.003	0.003	0.013*
	[0.011]	[0.005]	[0.006]	[0.007]
(Age - 50) squared	-0.002**	0	0	-0.001**
	[0.001]	[0.000]	[0.000]	[0.001]
Single	0.014	0.002	0.028*	-0.001
	[0.021]	[0.013]	[0.016]	[0.019]
Salaried worker	0.01	0.009	0.025**	-0.002
	[0.020]	[0.010]	[0.011]	[0.014]
GDP gap	-0.043	0.221	0.357*	-0.196
	[0.325]	[0.169]	[0.188]	[0.207]
Year-1980	0	0.001*	-0.001	-0.001
	[0.001]	[0.001]	[0.001]	[0.001]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old male employees

Table 2.28: MCMC estimates in the baseline specification, other parameters in measurement error

	Less than HS	HS	Some college	BA or more
Std permanent error	0.112***	0.129***	0.149***	0.167***
	[0.012]	[0.007]	[0.007]	[0.008]
Std transitory error				
Stable years	0.150***	0.095***	0.096***	0.128***
	[0.009]	[0.005]	[0.005]	[0.010]
Major job loss	0.585***	0.566***	0.532***	0.497***
	[0.037]	[0.022]	[0.030]	[0.023]
Major job finding	0.634***	0.507***	0.430***	0.366***
	[0.043]	[0.027]	[0.023]	[0.017]
Major job loss and job finding	0.650***	0.558***	0.587***	0.573***
	[0.035]	[0.023]	[0.030]	[0.026]
Latent job loss in stable years	0.753***	0.681***	0.564***	0.629***
	[0.044]	[0.026]	[0.027]	[0.026]
Fraction of earnings component reported in the survey				
Permanent earnings	1.008***	0.952***	0.934***	0.913***
	[0.014]	[0.009]	[0.011]	[0.010]
Transitory on-the-job*	0	0	0	0
Transitory due to job-loss	0.667***	0.844***	0.857***	0.942***
	[0.028]	[0.017]	[0.022]	[0.022]
Transitory due to job finding	0.283***	0.793***	0.502***	0.629***
	[0.034]	[0.017]	[0.017]	[0.024]
Transitory due to job loss and job finding	0.417***	0.596***	0.459***	0.887***
	[0.019]	[0.018]	[0.022]	[0.025]
Transitory due to latent job loss in stable years	1.058***	0.760***	0.715***	1.099***
	[0.124]	[0.060]	[0.050]	[0.116]

*The fraction of transitory reports are calibrated to zero in the baseline specification.

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old male employees

Table 2.29: MCMC estimates by education, covariates in earnings

	Less than HS	HS	Some college	BA or more
Constant	9.697***	9.938***	10.155***	10.449***
	[0.063]	[0.048]	[0.054]	[0.050]
Black	0.028	-0.084*	0.056	-0.065
	[0.065]	[0.050]	[0.054]	[0.057]
Hispanic	-0.072	-0.093	-0.081	-0.099
	[0.066]	[0.087]	[0.087]	[0.101]
Age - 50	0.015**	0.011***	0.012***	0.010***
	[0.006]	[0.004]	[0.004]	[0.004]
(Age - 50) squared	0.000	0.000	0.000	-0.000**
	[0.000]	[0.000]	[0.000]	[0.000]
Single	0.023	0.003	-0.026**	0.004
	[0.020]	[0.012]	[0.013]	[0.011]
Salaried worker	-0.003	0.008	0.015**	0.042***
	[0.017]	[0.008]	[0.007]	[0.008]
GDP gap	0.276	0.172*	0.232**	0.000
	[0.237]	[0.102]	[0.096]	[0.092]
Year-1980	-0.006	0.000	0.000	0.007**
	[0.004]	[0.003]	[0.003]	[0.003]

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample: HRS, 1991-2011, 50-60-year-old female employees

Table 2.30: MCMC estimates by education, covariates in the innovation term of permanent earnings

	Less than HS	HS	Some college	BA or more
Means				
Major job finding	-0.128***	-0.053**	-0.092***	-0.108***
	[0.027]	[0.021]	[0.030]	[0.029]
Major job loss and job finding	0.105***	0.065***	0.056***	0.090***
	[0.019]	[0.021]	[0.021]	[0.020]
Zero earnings	-0.184***	-0.215***	-0.249***	-0.200***
	[0.027]	[0.027]	[0.045]	[0.037]
Latent job loss in stable years	-0.113**	-0.178***	-0.138***	-0.302***
	[0.047]	[0.030]	[0.028]	[0.043]
Standard deviation				
Stable years	0.078***	0.069***	0.061***	0.063***
	[0.005]	[0.002]	[0.002]	[0.002]
Major job finding	0.286***	0.404***	0.508***	0.473***
	[0.024]	[0.018]	[0.026]	[0.023]
Major job loss and job finding	0.251***	0.500***	0.438***	0.352***
	[0.019]	[0.018]	[0.018]	[0.017]
Zero earnings	0.609***	0.624***	0.691***	0.640***
	[0.047]	[0.039]	[0.048]	[0.063]
Latent job loss in stable years	0.744***	0.771***	0.630***	0.853***
	[0.044]	[0.028]	[0.027]	[0.033]
Standard deviation, raw coefficients				
Constant	-2.545***	-2.682***	-2.800***	-2.766***
	[0.065]	[0.032]	[0.035]	[0.028]
Major job finding	1.294***	1.780***	2.126***	2.018***
	[0.107]	[0.052]	[0.060]	[0.055]
Major job loss and job finding	1.157***	1.988***	1.974***	1.722***
	[0.095]	[0.046]	[0.052]	[0.056]
Zero earnings	2.052***	2.216***	2.429***	2.320***
	[0.091]	[0.069]	[0.071]	[0.104]
Latent job loss in stable years	2.251***	2.419***	2.336***	2.606***
	[0.078]	[0.042]	[0.050]	[0.042]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old female employees

Table 2.31: MCMC estimates by education, other parameters in earnings

	Less than HS	HS	Some college	BA or more
Permanent std, age 50	0.538***	0.619***	0.599***	0.569***
	[0.021]	[0.015]	[0.017]	[0.015]
Transitory std	0.055***	0.031***	0.030***	0.018***
	[0.004]	[0.003]	[0.003]	[0.003]
Probability of latent job loss in stable years	0.106***	0.120***	0.129***	0.097***
	[0.009]	[0.005]	[0.007]	[0.005]
Avg transitory earnings loss by labor history				
Major job loss	0.474***	0.452***	0.429***	0.404***
	[0.012]	[0.008]	[0.010]	[0.010]
Major job finding	0.407***	0.423***	0.416***	0.422***
	[0.012]	[0.009]	[0.010]	[0.011]
Major job loss and job finding	0.450***	0.407***	0.412***	0.373***
	[0.011]	[0.008]	[0.009]	[0.010]
Latent job loss in stable years	0.233***	0.254***	0.220***	0.287***
	[0.015]	[0.009]	[0.008]	[0.010]
Raw coefficients of transitory earnings loss, γ_j				
Major job-loss	1.113***	1.213***	1.335***	1.474***
	[0.051]	[0.039]	[0.052]	[0.064]
Major job finding	1.461***	1.371***	1.407***	1.375***
	[0.073]	[0.046]	[0.059]	[0.059]
Major job loss and job finding	1.224***	1.461***	1.428***	1.681***
	[0.055]	[0.046]	[0.052]	[0.071]
Latent job loss in stable years	3.338***	2.963***	3.571***	2.495***
	[0.252]	[0.125]	[0.143]	[0.111]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old female employees

Table 2.32: MCMC estimates by education, covariates in measurement error

	Less than HS	HS	Some college	BA or more
Constant	0.028	-0.005	-0.053**	-0.061
	[0.051]	[0.024]	[0.024]	[0.038]
Black	-0.063**	0.003	-0.023	0.02
	[0.025]	[0.014]	[0.018]	[0.028]
Hispanic	-0.057**	-0.038*	0.005	-0.024
	[0.025]	[0.022]	[0.027]	[0.048]
Age - 50	-0.012	-0.002	0.002	-0.001
	[0.014]	[0.006]	[0.006]	[0.010]
(Age - 50) squared	0.001	0	0	0
	[0.001]	[0.000]	[0.000]	[0.001]
Single	-0.011	-0.026**	0.025*	0.022
	[0.022]	[0.011]	[0.013]	[0.018]
Salaried worker	0.105***	0.022**	0.035***	0.063***
	[0.031]	[0.010]	[0.011]	[0.017]
GDP gap	0.417	0.312	-0.086	0.096
	[0.483]	[0.193]	[0.180]	[0.283]
Year-1980	0	0.001	0.002*	0.002
	[0.002]	[0.001]	[0.001]	[0.001]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old female employees

Table 2.33: MCMC estimates in the baseline specification, other parameters in measurement errors

	Less than HS	HS	Some college	BA or more
Std permanent error	0.128***	0.089***	0.148***	0.193***
	[0.018]	[0.007]	[0.013]	[0.010]
Std transitory error				
Stable years	0.206***	0.150***	0.109***	0.197***
	[0.013]	[0.011]	[0.006]	[0.009]
Major job loss	0.617***	0.596***	0.598***	0.613***
	[0.035]	[0.023]	[0.028]	[0.029]
Major job finding	0.642***	0.581***	0.597***	0.479***
	[0.038]	[0.025]	[0.035]	[0.026]
Major job loss and job finding	0.772***	0.669***	0.668***	0.407***
	[0.043]	[0.030]	[0.034]	[0.018]
Latent job loss in stable years	0.832***	0.657***	0.619***	0.657***
	[0.069]	[0.033]	[0.025]	[0.046]
Fraction of earnings component reported in the survey				
Permanent earnings	0.975***	0.963***	0.917***	0.928***
	[0.017]	[0.007]	[0.009]	[0.013]
Transitory on-the-job*	0	0	0	0
Transitory due to job-loss	0.440***	0.782***	0.870***	0.738***
	[0.042]	[0.014]	[0.017]	[0.032]
Transitory due to job finding	0.539***	0.659***	0.677***	0.867***
	[0.026]	[0.015]	[0.017]	[0.024]
Transitory due to job loss and job finding	0.450***	0.582***	0.796***	0.878***
	[0.028]	[0.016]	[0.014]	[0.024]
Transitory due to latent job loss in stable years	0.625***	0.971***	0.471***	0.927***
	[0.106]	[0.026]	[0.048]	[0.074]

*The fraction of transitory reports are calibrated to zero in the baseline specification.

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old female employees

Table 2.34: MCMC estimates with alternative modeling assumptions, covariates in earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	MA(1)	MA(2)	No latent	MA(1)	MA(2)	No latent
Constant	10.688***	10.679***	10.606***	9.927***	9.894***	9.729***
	[0.028]	[0.029]	[0.030]	[0.029]	[0.029]	[0.030]
Black	-0.313***	-0.298***	-0.262***	-0.021	-0.019	-0.026
	[0.030]	[0.030]	[0.031]	[0.029]	[0.029]	[0.030]
Hispanic	-0.323***	-0.308***	-0.280***	-0.097**	-0.088**	-0.034
	[0.034]	[0.034]	[0.036]	[0.043]	[0.042]	[0.043]
Less than high school	-0.285***	-0.279***	-0.265***	-0.249***	-0.259***	-0.189***
	[0.033]	[0.033]	[0.033]	[0.038]	[0.038]	[0.037]
Some college	0.088***	0.093***	0.086***	0.203***	0.224***	0.193***
	[0.028]	[0.028]	[0.029]	[0.029]	[0.029]	[0.030]
BA	0.395***	0.411***	0.361***	0.485***	0.500***	0.437***
	[0.033]	[0.033]	[0.034]	[0.037]	[0.037]	[0.039]
More than BA	0.483***	0.480***	0.442***	0.754***	0.767***	0.721***
	[0.032]	[0.032]	[0.034]	[0.037]	[0.037]	[0.039]
Age - 50	0.012***	0.012***	0.021***	0.015***	0.012***	0.018***
	[0.002]	[0.002]	[0.004]	[0.002]	[0.002]	[0.004]
(Age - 50) squared	-0.000**	-0.000**	-0.003***	-0.000***	-0.000**	-0.002***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Single	-0.006	-0.020**	-0.154***	-0.001	0.004	0.045***
	[0.008]	[0.009]	[0.017]	[0.007]	[0.008]	[0.015]
Salaried worker	0.017***	0.021***	0.165***	0.029***	0.027***	0.192***
	[0.006]	[0.006]	[0.013]	[0.005]	[0.005]	[0.012]
GDP gap	0.483***	0.522***	0.937***	0.254***	0.223***	0.400***
	[0.063]	[0.067]	[0.152]	[0.067]	[0.066]	[0.145]
Year-1980	-0.005***	-0.005***	-0.005***	0.000	0.001	0.004**
	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.35: MCMC estimates with alternative modeling assumptions, covariates in the innovation term of permanent earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	MA(1)	MA(2)	No latent	MA(1)	MA(2)	No latent
Means						
Major job finding	-0.088***	-0.077***	-0.081***	-0.089***	-0.097***	-0.083***
	[0.015]	[0.016]	[0.015]	[0.016]	[0.016]	[0.014]
Major job loss and job finding	0.003	0.007	-0.021**	0.018	0.041***	0.021**
	[0.013]	[0.013]	[0.009]	[0.013]	[0.013]	[0.008]
Zero earnings	-0.277***	-0.282***	-0.247***	-0.246***	-0.233***	-0.214***
	[0.024]	[0.023]	[0.011]	[0.017]	[0.018]	[0.009]
Latent job loss in stable years	-0.311***	-0.357***		-0.264***	-0.271***	
	[0.027]	[0.029]		[0.023]	[0.024]	
Standard deviation						
Stable years	0.063***	0.067***	0.085***	0.066***	0.064***	0.084***
	[0.002]	[0.002]	[0.000]	[0.002]	[0.003]	[0.000]
Major job finding	0.527***	0.532***	0.418***	0.560***	0.543***	0.412***
	[0.014]	[0.014]	[0.009]	[0.015]	[0.015]	[0.008]
Major job loss and job finding	0.541***	0.514***	0.184***	0.517***	0.542***	0.186***
	[0.013]	[0.012]	[0.004]	[0.013]	[0.014]	[0.004]
Zero earnings	0.839***	0.840***	0.140***	0.703***	0.698***	0.139***
	[0.036]	[0.038]	[0.003]	[0.028]	[0.027]	[0.002]
Latent job loss in stable years	0.981***	0.989***		0.829***	0.860***	
	[0.026]	[0.028]		[0.023]	[0.025]	
Standard deviation, raw coefficients						
Constant	-2.767***	-2.697***	-2.461***	-2.721***	-2.751***	-2.473***
	[0.030]	[0.031]	[0.006]	[0.034]	[0.041]	[0.006]
Major job finding	2.128***	2.066***	1.589***	2.143***	2.141***	1.586***
	[0.037]	[0.038]	[0.021]	[0.041]	[0.048]	[0.020]
Major job loss and job finding	2.151***	2.031***	0.766***	2.062***	2.140***	0.792***
	[0.034]	[0.036]	[0.020]	[0.039]	[0.044]	[0.020]
Zero earnings	2.594***	2.526***	0.496***	2.370***	2.393***	0.499***
	[0.043]	[0.047]	[0.019]	[0.046]	[0.054]	[0.016]
Latent job loss in stable years	2.747***	2.683***		2.530***	2.597***	
	[0.034]	[0.037]		[0.040]	[0.041]	

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.36: MCMC estimates with alternative modeling assumptions, other parameters in earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	MA(1)	MA(2)	No latent	MA(1)	MA(2)	No latent
Permanent std, age 50	0.541***	0.537***	0.576***	0.601***	0.601***	0.654***
	[0.008]	[0.008]	[0.009]	[0.009]	[0.008]	[0.009]
Transitory std	0.061***	0.068***	0.313***	0.063***	0.064***	0.282***
	[0.002]	[0.002]	[0.002]	[0.003]	[0.002]	[0.002]
MA(1) term	-0.283***	-0.298***		-0.523***	-0.565***	
	[0.034]	[0.045]		[0.034]	[0.045]	
MA(2) term		-0.089**			-0.191***	
		[0.036]			[0.044]	
Total transitory	0.064***	0.072***		0.073***	0.077***	
	[0.010]	[0.010]		[0.012]	[0.014]	
Probability of latent job loss in stable years	0.072***	0.067***		0.076***	0.077***	
	[0.002]	[0.002]		[0.002]	[0.002]	
Avg transitory earnings loss by labor history						
Major job loss	0.410***	0.405***	0.400***	0.435***	0.433***	0.421***
	[0.005]	[0.005]	[0.006]	[0.005]	[0.005]	[0.006]
Major job finding	0.354***	0.347***	0.349***	0.397***	0.398***	0.385***
	[0.005]	[0.005]	[0.007]	[0.005]	[0.005]	[0.007]
Major job loss and job finding	0.356***	0.345***	0.364***	0.362***	0.381***	0.387***
	[0.004]	[0.005]	[0.006]	[0.005]	[0.005]	[0.006]
Latent job loss in stable years	0.283***	0.286***		0.270***	0.264***	
	[0.007]	[0.007]		[0.006]	[0.008]	
Raw coefficients of transitory earnings loss, γ_j						
Major job-loss	1.441***	1.467***	1.503***	1.301***	1.307***	1.374***
	[0.029]	[0.030]	[0.040]	[0.025]	[0.025]	[0.034]
Major job finding	1.827***	1.884***	1.870***	1.525***	1.516***	1.600***
	[0.040]	[0.042]	[0.060]	[0.031]	[0.033]	[0.045]
Major job loss and job finding	1.809***	1.899***	1.750***	1.764***	1.628***	1.585***
	[0.035]	[0.037]	[0.047]	[0.033]	[0.031]	[0.040]
Latent job loss in stable years	2.545***	2.507***		2.717***	2.817***	
	[0.071]	[0.074]		[0.069]	[0.084]	

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.37: MCMC estimates with alternative modeling assumptions, covariates in measurement error

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	MA(1)	MA(2)	No latent	MA(1)	MA(2)	No latent
Constant	-0.039*	-0.03	-0.02	-0.058***	-0.051***	-0.024
	[0.020]	[0.020]	[0.030]	[0.019]	[0.019]	[0.029]
Black	0.017	0.02	0.025	-0.01	-0.003	0.001
	[0.013]	[0.013]	[0.017]	[0.011]	[0.011]	[0.015]
Hispanic	-0.02	-0.027*	-0.024	-0.037**	-0.036**	-0.031
	[0.015]	[0.014]	[0.020]	[0.015]	[0.015]	[0.020]
Less than high school	-0.024*	-0.025*	-0.034**	-0.029**	-0.023*	-0.063***
	[0.013]	[0.013]	[0.017]	[0.013]	[0.013]	[0.017]
Some college	0.033***	0.034***	0.052***	0.021**	0.017	0.031**
	[0.011]	[0.011]	[0.015]	[0.011]	[0.011]	[0.014]
BA	0.02	0.017	0.054***	-0.004	0	0.031*
	[0.013]	[0.013]	[0.018]	[0.014]	[0.014]	[0.018]
More than BA	0.034**	0.027**	0.074***	0.013	0.019	0.048***
	[0.014]	[0.014]	[0.018]	[0.014]	[0.014]	[0.018]
Age - 50	0.012**	0.010*	0.005	0.003	0.004	0.003
	[0.005]	[0.005]	[0.008]	[0.005]	[0.005]	[0.008]
(Age - 50) squared	-0.001***	-0.001**	0	0	-0.001	0
	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]	[0.001]
Single	0.019*	0.012	0.043***	0.002	0.004	-0.027**
	[0.011]	[0.010]	[0.015]	[0.008]	[0.008]	[0.011]
Salaried worker	0.015*	0.016*	-0.02	0.045***	0.045***	0.01
	[0.009]	[0.009]	[0.013]	[0.009]	[0.009]	[0.012]
GDP gap	0.205	0.183	0.473**	-0.011	-0.05	-0.18
	[0.148]	[0.149]	[0.234]	[0.149]	[0.149]	[0.232]
Year-1980	0	0	0	0.002***	0.002***	0
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.38: MCMC estimates with alternative modeling assumptions, other parameters in measurement error

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	MA(1)	MA(2)	No latent	MA(1)	MA(2)	No latent
Std permanent error	0.147***	0.147***	0.120***	0.149***	0.152***	0.108***
	[0.005]	[0.006]	[0.012]	[0.005]	[0.006]	[0.012]
Std transitory error						
Stable years	0.207***	0.200***	0.339***	0.216***	0.211***	0.363***
	[0.005]	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]
Major job loss	0.463***	0.479***	0.483***	0.540***	0.536***	0.615***
	[0.011]	[0.013]	[0.019]	[0.016]	[0.015]	[0.023]
Major job finding	0.428***	0.425***	0.525***	0.502***	0.514***	0.595***
	[0.011]	[0.010]	[0.023]	[0.016]	[0.015]	[0.023]
Major job loss and job finding	0.344***	0.398***	0.554***	0.404***	0.430***	0.637***
	[0.007]	[0.011]	[0.020]	[0.009]	[0.011]	[0.021]
Latent job loss in stable years	0.595***	0.597***		0.657***	0.648***	
	[0.040]	[0.040]		[0.034]	[0.035]	
Fraction of earnings component reported in the survey						
Permanent earnings	0.921***	0.907***	0.954***	0.935***	0.935***	1.001***
	[0.007]	[0.006]	[0.009]	[0.006]	[0.006]	[0.008]
Transitory on-the-job*	0	0	0	0	0	0
Transitory due to job-loss	0.849***	0.846***	0.865***	0.811***	0.794***	0.726***
	[0.015]	[0.015]	[0.034]	[0.013]	[0.014]	[0.034]
Transitory due to job finding	0.632***	0.677***	0.707***	0.701***	0.752***	0.692***
	[0.018]	[0.019]	[0.044]	[0.013]	[0.014]	[0.033]
Transitory due to job loss and job finding	0.587***	0.664***	0.606***	0.621***	0.739***	0.615***
	[0.015]	[0.015]	[0.036]	[0.013]	[0.012]	[0.035]
Transitory due to latent job loss in stable years	0.760***	0.601***		0.699***	0.726***	
	[0.035]	[0.059]		[0.028]	[0.027]	

*The fraction of transitory reports are calibrated to zero in this specification.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.39: MCMC estimates with alternative modeling of reporting transitory shocks, covariates in earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim
Constant	10.694***	10.696***	10.691***	9.920***	9.921***	9.917***
	[0.028]	[0.027]	[0.027]	[0.030]	[0.028]	[0.027]
Black	-0.295***	-0.305***	-0.299***	-0.016	-0.019	-0.015
	[0.029]	[0.029]	[0.029]	[0.028]	[0.028]	[0.027]
Hispanic	-0.306***	-0.315***	-0.306***	-0.088**	-0.069*	-0.117***
	[0.033]	[0.033]	[0.033]	[0.041]	[0.042]	[0.040]
Less than high school	-0.287***	-0.267***	-0.279***	-0.236***	-0.245***	-0.217***
	[0.032]	[0.031]	[0.032]	[0.035]	[0.034]	[0.039]
Some college	0.112***	0.108***	0.097***	0.221***	0.218***	0.228***
	[0.027]	[0.027]	[0.028]	[0.028]	[0.028]	[0.028]
BA	0.406***	0.411***	0.392***	0.492***	0.498***	0.501***
	[0.031]	[0.032]	[0.032]	[0.036]	[0.036]	[0.036]
More than BA	0.483***	0.482***	0.476***	0.756***	0.763***	0.768***
	[0.031]	[0.031]	[0.032]	[0.036]	[0.035]	[0.036]
Age - 50	0.011***	0.011***	0.011***	0.011***	0.012***	0.011***
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
(Age - 50) squared	-0.000**	-0.000**	-0.000**	-0.000***	-0.000***	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Single	-0.014*	-0.016**	-0.013*	-0.004	-0.006	-0.005
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
Salaried worker	0.014***	0.015***	0.014***	0.021***	0.020***	0.020***
	[0.005]	[0.005]	[0.005]	[0.004]	[0.004]	[0.004]
GDP gap	0.414***	0.445***	0.393***	0.160***	0.148***	0.161***
	[0.057]	[0.060]	[0.057]	[0.053]	[0.054]	[0.055]
Year-1980	-0.006***	-0.005***	-0.005***	0.001	0.001	0.001
	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	[0.001]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.40: MCMC estimates with alternative modeling of reporting transitory shocks, covariates in the innovation term of permanent earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim
Means						
Major job finding	-0.095***	-0.085***	-0.074***	-0.085***	-0.075***	-0.079***
	[0.014]	[0.015]	[0.014]	[0.014]	[0.013]	[0.014]
Major job loss and job finding	0.024**	0.016	0.020*	0.068***	0.065***	0.063***
	[0.011]	[0.010]	[0.010]	[0.011]	[0.010]	[0.010]
Zero earnings	-0.250***	-0.254***	-0.249***	-0.207***	-0.195***	-0.213***
	[0.023]	[0.028]	[0.022]	[0.015]	[0.017]	[0.021]
Latent job loss in stable years	-0.210***	-0.225***	-0.214***	-0.184***	-0.173***	-0.195***
	[0.019]	[0.020]	[0.019]	[0.018]	[0.018]	[0.018]
Standard deviation						
Stable years	0.064***	0.067***	0.065***	0.068***	0.070***	0.070***
	[0.002]	[0.001]	[0.002]	[0.001]	[0.001]	[0.001]
Major job finding	0.457***	0.470***	0.431***	0.449***	0.445***	0.467***
	[0.014]	[0.013]	[0.013]	[0.013]	[0.012]	[0.013]
Major job loss and job finding	0.408***	0.403***	0.413***	0.424***	0.414***	0.435***
	[0.010]	[0.010]	[0.010]	[0.011]	[0.010]	[0.012]
Zero earnings	0.795***	0.791***	0.800***	0.651***	0.630***	0.640***
	[0.035]	[0.040]	[0.044]	[0.025]	[0.025]	[0.023]
Latent job loss in stable years	0.806***	0.820***	0.817***	0.750***	0.731***	0.734***
	[0.024]	[0.024]	[0.025]	[0.021]	[0.020]	[0.019]
Standard deviation, raw coefficients						
Constant	-2.757***	-2.698***	-2.741***	-2.697***	-2.657***	-2.653***
	[0.022]	[0.021]	[0.023]	[0.019]	[0.017]	[0.020]
Major job finding	1.978***	1.945***	1.901***	1.896***	1.848***	1.893***
	[0.034]	[0.032]	[0.034]	[0.032]	[0.029]	[0.032]
Major job loss and job finding	1.859***	1.787***	1.856***	1.839***	1.776***	1.823***
	[0.030]	[0.029]	[0.030]	[0.027]	[0.025]	[0.033]
Zero earnings	2.529***	2.472***	2.526***	2.268***	2.200***	2.207***
	[0.042]	[0.040]	[0.050]	[0.042]	[0.035]	[0.037]
Latent job loss in stable years	2.536***	2.495***	2.534***	2.406***	2.340***	2.341***
	[0.027]	[0.026]	[0.028]	[0.025]	[0.025]	[0.025]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.41: MCMC estimates with alternative modeling of reporting transitory shocks, other parameters in earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim
Permanent std, age 50	0.520***	0.518***	0.522***	0.585***	0.581***	0.584***
	[0.007]	[0.007]	[0.007]	[0.008]	[0.008]	[0.008]
Transitory std	0.040***	0.037***	0.039***	0.028***	0.023***	0.026***
	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]
Probability of latent job loss	0.106***	0.105***	0.108***	0.115***	0.118***	0.114***
	[0.003]	[0.003]	[0.003]	[0.004]	[0.003]	[0.004]
Avg transitory earnings loss by labor history						
Major job loss	0.410***	0.414***	0.411***	0.452***	0.458***	0.452***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Major job finding	0.357***	0.381***	0.375***	0.425***	0.440***	0.427***
	[0.006]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Major job loss and job finding	0.375***	0.364***	0.375***	0.404***	0.408***	0.407***
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
Latent job loss in stable years	0.246***	0.235***	0.233***	0.245***	0.250***	0.236***
	[0.007]	[0.007]	[0.007]	[0.005]	[0.006]	[0.006]
Raw coefficients of transitory earnings loss, γ_j						
Major job-loss	1.442***	1.414***	1.432***	1.213***	1.184***	1.213***
	[0.029]	[0.029]	[0.029]	[0.023]	[0.023]	[0.024]
Major job finding	1.807***	1.622***	1.667***	1.353***	1.272***	1.344***
	[0.040]	[0.036]	[0.036]	[0.028]	[0.025]	[0.027]
Major job loss and job finding	1.665***	1.749***	1.668***	1.477***	1.449***	1.460***
	[0.032]	[0.034]	[0.032]	[0.028]	[0.027]	[0.027]
Latent job loss in stable years	3.084***	3.267***	3.308***	3.101***	3.023***	3.264***
	[0.069]	[0.080]	[0.076]	[0.069]	[0.065]	[0.076]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.42: MCMC estimates with alternative modeling of reporting transitory shocks, covariates in measurement error

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim
Constant	-0.027**	-0.030**	-0.017	-0.034**	-0.032**	-0.030**
	[0.014]	[0.014]	[0.014]	[0.016]	[0.015]	[0.015]
Black	0.003	0.002	0.003	0	-0.005	-0.002
	[0.012]	[0.011]	[0.011]	[0.010]	[0.009]	[0.009]
Hispanic	-0.021	-0.019	-0.025**	-0.038***	-0.038***	-0.043***
	[0.014]	[0.013]	[0.013]	[0.013]	[0.013]	[0.013]
Less than high school	-0.025**	-0.026**	-0.021*	-0.032***	-0.028**	-0.025**
	[0.012]	[0.011]	[0.011]	[0.012]	[0.011]	[0.011]
Some college	0.024**	0.026***	0.026**	0.011	0.017**	0.017*
	[0.010]	[0.010]	[0.010]	[0.009]	[0.009]	[0.009]
BA	0.028**	0.023**	0.029**	-0.012	-0.008	-0.006
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]
More than BA	0.029**	0.022*	0.031***	0.011	0.015	0.01
	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]	[0.012]
Age - 50	0.007**	0.008**	0.005*	0	0.001	0
	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.004]
(Age - 50) squared	-0.001**	-0.001***	-0.000*	0	0	0
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Single	0.019**	0.011	0.014*	-0.001	0.003	-0.002
	[0.009]	[0.008]	[0.008]	[0.007]	[0.007]	[0.007]
Salaried worker	0.009	0.012*	0.005	0.037***	0.035***	0.033***
	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]	[0.007]
GDP gap	0.059	0.085	0.039	0.041	0.045	-0.087
	[0.101]	[0.103]	[0.097]	[0.125]	[0.123]	[0.120]
Year-1980	0	0	0	0.001**	0.001*	0.001
	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.000]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.43: MCMC estimates with alternative modeling of reporting transitory shocks, other parameters in measurement error

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim	$\tau^{trans} = 0.5$	$\tau^{trans} = 1$	τ^{trans} estim
Std permanent error	0.162***	0.152***	0.160***	0.132***	0.131***	0.127***
	[0.004]	[0.004]	[0.004]	[0.005]	[0.005]	[0.005]
Std transitory error						
Stable years	0.108***	0.114***	0.091***	0.166***	0.158***	0.121***
	[0.003]	[0.004]	[0.004]	[0.008]	[0.008]	[0.016]
Major job loss	0.561***	0.536***	0.519***	0.627***	0.621***	0.623***
	[0.016]	[0.014]	[0.014]	[0.017]	[0.019]	[0.020]
Major job finding	0.498***	0.502***	0.468***	0.566***	0.572***	0.535***
	[0.014]	[0.015]	[0.012]	[0.014]	[0.018]	[0.015]
Major job loss and job finding	0.591***	0.562***	0.574***	0.653***	0.614***	0.620***
	[0.017]	[0.018]	[0.017]	[0.020]	[0.017]	[0.019]
Latent job loss in stable years	0.645***	0.671***	0.643***	0.689***	0.673***	0.675***
	[0.018]	[0.023]	[0.019]	[0.032]	[0.023]	[0.022]
Fraction of earnings component reported in the survey						
Permanent earnings	0.948***	0.949***	0.961***	0.945***	0.948***	0.966***
	[0.005]	[0.006]	[0.005]	[0.005]	[0.005]	[0.005]
Transitory on-the-job	0.5	1	-0.514***	0.5	1	-3.277***
			[0.140]			[0.681]
Transitory due to job-loss	0.885***	0.827***	0.864***	0.745***	0.730***	0.712***
	[0.010]	[0.010]	[0.011]	[0.010]	[0.010]	[0.011]
Transitory due to job finding	0.597***	0.614***	0.593***	0.757***	0.717***	0.645***
	[0.011]	[0.011]	[0.009]	[0.009]	[0.009]	[0.008]
Transitory due to job loss and find	0.536***	0.519***	0.570***	0.633***	0.594***	0.532***
	[0.010]	[0.010]	[0.010]	[0.009]	[0.009]	[0.010]
Transitory due to latent job loss	0.519***	0.770***	0.757***	0.834***	0.856***	0.800***
	[0.031]	[0.029]	[0.029]	[0.022]	[0.020]	[0.023]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.44: MCMC estimates using alternative samples, covariates in earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	1980-2011	matched sample		1980-2011	matched sample	
		with off years	no off years		with off years	no off years
Constant	10.712***	10.701***	10.700***	9.785***	9.943***	9.971***
	[0.019]	[0.024]	[0.027]	[0.021]	[0.025]	[0.026]
Black	-0.270***	-0.237***	-0.219***	0.012	-0.025	-0.030
	[0.024]	[0.029]	[0.029]	[0.025]	[0.027]	[0.025]
Hispanic	-0.285***	-0.277***	-0.269***	-0.117***	-0.093**	-0.095***
	[0.029]	[0.034]	[0.033]	[0.037]	[0.039]	[0.037]
Less than high school	-0.279***	-0.233***	-0.231***	-0.250***	-0.205***	-0.235***
	[0.025]	[0.030]	[0.029]	[0.030]	[0.033]	[0.032]
Some college	0.100***	0.128***	0.160***	0.184***	0.225***	0.219***
	[0.023]	[0.027]	[0.026]	[0.026]	[0.027]	[0.026]
BA	0.399***	0.419***	0.440***	0.446***	0.503***	0.501***
	[0.028]	[0.031]	[0.031]	[0.034]	[0.035]	[0.034]
More than BA	0.481***	0.496***	0.506***	0.751***	0.735***	0.706***
	[0.027]	[0.030]	[0.030]	[0.033]	[0.035]	[0.034]
Age - 50	0.008***	0.010***	0.010***	0.005***	0.008***	0.011***
	[0.002]	[0.002]	[0.003]	[0.002]	[0.002]	[0.003]
(Age - 50) squared	-0.000**	0.000	0.000	-0.000***	-0.000*	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Single	-0.025***	-0.011	-0.045***	-0.003	-0.001	0.029***
	[0.008]	[0.008]	[0.010]	[0.007]	[0.007]	[0.009]
Salaried worker	0.016***	0.016***	0.025***	0.021***	0.023***	0.038***
	[0.005]	[0.005]	[0.007]	[0.004]	[0.004]	[0.006]
GDP gap	0.366***	0.397***	0.174*	0.167***	0.158***	-0.153
	[0.041]	[0.062]	[0.101]	[0.040]	[0.055]	[0.099]
Year-1980	-0.006***	-0.005***	-0.006***	0.008***	0.003**	0.003**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.45: MCMC estimates using alternative samples, covariates in the innovation term of permanent earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	1980-2011	matched sample		1980-2011	matched sample	
		with off years	no off years		with off years	no off years
Means						
Major job finding	-0.082***	-0.107***	-0.091***	-0.057***	-0.124***	-0.113***
	[0.011]	[0.012]	[0.014]	[0.011]	[0.014]	[0.015]
Major job loss and job finding	0.026***	-0.01	-0.044***	0.089***	0.049***	0
	[0.009]	[0.010]	[0.010]	[0.009]	[0.010]	[0.010]
Zero earnings	-0.233***	-0.179***	-0.145***	-0.174***	-0.141***	-0.197***
	[0.016]	[0.016]	[0.019]	[0.019]	[0.010]	[0.013]
Latent job loss in stable years	-0.201***	-0.169***	-0.041***	-0.158***	-0.169***	-0.132***
	[0.015]	[0.018]	[0.012]	[0.015]	[0.017]	[0.016]
Standard deviation						
Stable years	0.069***	0.060***	0.035***	0.073***	0.065***	0.042***
	[0.001]	[0.002]	[0.002]	[0.001]	[0.001]	[0.004]
Major job finding	0.431***	0.299***	0.252***	0.449***	0.426***	0.325***
	[0.011]	[0.015]	[0.015]	[0.010]	[0.013]	[0.019]
Major job loss and job finding	0.404***	0.342***	0.240***	0.420***	0.356***	0.214***
	[0.009]	[0.011]	[0.013]	[0.009]	[0.010]	[0.016]
Zero earnings	0.745***	0.331***	0.317***	0.698***	0.165***	0.302***
	[0.033]	[0.035]	[0.034]	[0.030]	[0.035]	[0.030]
Latent job loss in stable years	0.776***	0.722***	0.377***	0.717***	0.684***	0.502***
	[0.018]	[0.023]	[0.015]	[0.017]	[0.017]	[0.021]
Standard deviation, raw coefficients						
Constant	-2.681***	-2.821***	-3.348***	-2.622***	-2.743***	-3.165***
	[0.015]	[0.026]	[0.066]	[0.017]	[0.021]	[0.048]
Major job finding	1.838***	1.606***	1.962***	1.822***	1.892***	2.048***
	[0.026]	[0.045]	[0.095]	[0.025]	[0.035]	[0.069]
Major job loss and job finding	1.774***	1.744***	1.918***	1.755***	1.708***	1.628***
	[0.022]	[0.035]	[0.088]	[0.024]	[0.031]	[0.082]
Zero earnings	2.393***	1.720***	2.194***	2.269***	0.899***	1.965***
	[0.032]	[0.102]	[0.115]	[0.032]	[0.235]	[0.109]
Latent job loss in stable years	2.427***	2.492***	2.372***	2.288***	2.362***	2.481***
	[0.022]	[0.030]	[0.066]	[0.022]	[0.027]	[0.051]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.46: MCMC estimates using alternative samples, other parameters in earnings

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	1980-2011	matched sample		1980-2011	matched sample	
		with off years	no off years		with off years	no off years
Permanent std, age 50	0.558***	0.425***	0.417***	0.642***	0.488***	0.437***
	[0.006]	[0.008]	[0.009]	[0.007]	[0.008]	[0.009]
Transitory std	0.037***	0.043***	0.053***	0.025***	0.029***	0.050***
	[0.001]	[0.001]	[0.002]	[0.002]	[0.002]	[0.002]
Probability of latent job loss	0.106***	0.107***	0.129***	0.115***	0.121***	0.121***
	[0.002]	[0.004]	[0.004]	[0.003]	[0.003]	[0.004]
Avg transitory earnings loss by labor history						
Major job loss	0.392***	0.394***	0.381***	0.421***	0.433***	0.430***
	[0.004]	[0.005]	[0.006]	[0.004]	[0.005]	[0.006]
Major job finding	0.334***	0.373***	0.418***	0.409***	0.437***	0.454***
	[0.005]	[0.005]	[0.006]	[0.005]	[0.005]	[0.006]
Major job loss and job finding	0.372***	0.352***	0.363***	0.425***	0.400***	0.431***
	[0.004]	[0.005]	[0.006]	[0.004]	[0.005]	[0.006]
Latent job loss in stable years	0.237***	0.232***	0.325***	0.239***	0.252***	0.307***
	[0.006]	[0.006]	[0.006]	[0.005]	[0.006]	[0.006]
Raw coefficients of transitory earnings loss, γ_j						
Major job-loss	1.549***	1.539***	1.627***	1.376***	1.311***	1.327***
	[0.029]	[0.033]	[0.038]	[0.025]	[0.026]	[0.028]
Major job finding	1.994***	1.679***	1.396***	1.448***	1.291***	1.203***
	[0.039]	[0.038]	[0.034]	[0.027]	[0.027]	[0.027]
Major job loss and job finding	1.688***	1.844***	1.761***	1.356***	1.499***	1.325***
	[0.030]	[0.039]	[0.039]	[0.023]	[0.030]	[0.027]
Latent job loss in stable years	3.242***	3.326***	2.082***	3.195***	2.982***	2.264***
	[0.070]	[0.084]	[0.054]	[0.063]	[0.081]	[0.066]

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.47: MCMC estimates using alternative samples, covariates in measurement error

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	1980-2011	matched sample		1980-2011	matched sample	
		with off years	no off years		with off years	no off years
Constant	-0.018	-0.021	-0.021	-0.018	-0.027*	-0.027*
	[0.014]	[0.014]	[0.014]	[0.016]	[0.016]	[0.014]
Black	0.002	0.004	0.012	-0.007	-0.007	-0.012
	[0.012]	[0.011]	[0.012]	[0.010]	[0.010]	[0.010]
Hispanic	-0.015	-0.026**	-0.031**	-0.039***	-0.043***	-0.043***
	[0.013]	[0.013]	[0.013]	[0.014]	[0.014]	[0.015]
Less than high school	-0.032***	-0.026**	-0.021*	-0.034***	-0.027**	-0.028**
	[0.012]	[0.011]	[0.011]	[0.012]	[0.012]	[0.012]
Some college	0.027***	0.027***	0.028***	0.022**	0.01	0.020**
	[0.010]	[0.010]	[0.010]	[0.009]	[0.009]	[0.009]
BA	0.027**	0.021*	0.023*	-0.005	-0.011	-0.01
	[0.012]	[0.011]	[0.012]	[0.012]	[0.012]	[0.012]
More than BA	0.027**	0.022*	0.029**	0.013	0.003	0.001
	[0.012]	[0.011]	[0.012]	[0.012]	[0.013]	[0.012]
Age - 50	0.006*	0.007**	0.004	-0.001	0	-0.001
	[0.003]	[0.003]	[0.003]	[0.004]	[0.004]	[0.003]
(Age - 50) squared	-0.001**	-0.001**	-0.000*	0	0	0
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Single	0.012	0.015*	0.006	0	0.008	-0.002
	[0.009]	[0.008]	[0.008]	[0.007]	[0.007]	[0.007]
Salaried worker	0.006	0.011*	0.006	0.035***	0.045***	0.030***
	[0.007]	[0.006]	[0.007]	[0.007]	[0.007]	[0.007]
GDP gap	0.101	0.077	0.149	0.051	-0.016	0.089
	[0.104]	[0.098]	[0.098]	[0.126]	[0.125]	[0.106]
Year-1980	0	0	0	0.001	0.001	0.001
	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]	[0.000]

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

Table 2.48: MCMC estimates using alternative samples, other parameters in measurement error

	Males			Females		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	1980-2011	matched sample		1980-2011	matched sample	
		with off years	no off years		with off years	no off years
Std permanent error	0.161***	0.149***	0.158***	0.141***	0.140***	0.149***
	[0.004]	[0.005]	[0.007]	[0.005]	[0.008]	[0.005]
Std transitory error						
Stable years	0.109***	0.102***	0.090***	0.164***	0.161***	0.115***
	[0.005]	[0.004]	[0.003]	[0.008]	[0.011]	[0.006]
Major job loss	0.527***	0.540***	0.521***	0.634***	0.632***	0.623***
	[0.014]	[0.015]	[0.017]	[0.020]	[0.021]	[0.023]
Major job finding	0.495***	0.527***	0.509***	0.573***	0.591***	0.585***
	[0.013]	[0.017]	[0.015]	[0.020]	[0.017]	[0.025]
Major job loss and job finding	0.592***	0.594***	0.581***	0.623***	0.669***	0.674***
	[0.017]	[0.018]	[0.019]	[0.018]	[0.019]	[0.023]
Latent job loss in stable years	0.638***	0.653***	0.615***	0.670***	0.668***	0.688***
	[0.034]	[0.027]	[0.034]	[0.026]	[0.034]	[0.050]
Fraction of earnings component reported in the survey						
Permanent earnings	0.954***	0.942***	0.937***	0.949***	0.910***	0.936***
	[0.006]	[0.006]	[0.010]	[0.005]	[0.006]	[0.007]
Transitory on-the-job*	0	0	0	0	0	0
Transitory due to job-loss	0.891***	0.839***	0.840***	0.772***	0.781***	0.711***
	[0.010]	[0.010]	[0.009]	[0.011]	[0.010]	[0.008]
Transitory due to job finding	0.611***	0.670***	0.704***	0.723***	0.748***	0.732***
	[0.011]	[0.010]	[0.008]	[0.010]	[0.009]	[0.007]
Transitory due to job loss and find	0.540***	0.566***	0.656***	0.632***	0.712***	0.706***
	[0.010]	[0.009]	[0.008]	[0.010]	[0.009]	[0.008]
Transitory due to latent job loss	0.986***	0.587***	0.889***	0.827***	1.013***	0.859***
	[0.042]	[0.046]	[0.020]	[0.023]	[0.039]	[0.014]

*The fraction of transitory reports are calibrated to zero in this specification.

*** p<0.01, ** p<0.05, * p<0.1. Sample: HRS, 1991-2011, 50-60-year-old employees

CHAPTER III

Estimating Second Order Probability Beliefs from Subjective Survival Data

3.1 Introduction

Textbook models of uncertainty usually treat probabilities as known or knowable objects. This is the case for outcomes of symmetric devices with known properties that are subject to random forces such as the toss of a coin or die, the shuffling of a deck of cards or the shaking of an urn containing balls of different colors. In contrast, the probabilities that confront economic decision makers are usually less precisely known.[34, 35]

In this paper, we investigate the probability beliefs held by a given individual about personal mortality risks elicited from survey questions on the Health and Retirement Study (HRS) that ask respondents about the numerical probability that he or she will survive to

[34] Frank Knight, as early as in 1921, introduced the distinction between, as he called, "risk" for known and "uncertainty" for unknown probabilities.

[35] Paté-Cornell (1996) observes that "... uncertainties in decision and risk analyses can be divided into two categories: those that stem from variability in known (or observable) populations and, therefore, represent randomness in samples (aleatory uncertainties), and those that come from basic lack of knowledge about fundamental phenomena (epistemic uncertainties also known in the literature as ambiguity)."

a given age that is 10-20 years in the future. In forming his belief about mortality risk, a person might consult a life table based on the experience of millions of individuals. The life table provides an estimate of the mean probability of survival for persons of given age and sex with essentially no sampling error. However, the individual may be ambiguous about whether this probability represents the risk he himself faces. He may have personal information that makes him think that he has a higher or lower risk than the average person of his age and sex. Moreover, he may be unsure about the influence that personal information such as the ages at death of parents and relatives, health history and current symptoms, or exercise and dietary habits will have on his likely longevity. Finally, in answering the survey question, a person must construct a probability judgment "on the fly," accessing whatever frequentist data or epistemic beliefs he may have stored in his brain and manipulating this information through reasoning or gut reaction to produce an answer to the specific question about, say, the probability he will survive to age 80, all within less than a minute.[36]

In this paper, we assume that probability beliefs about survival are ambiguous in the sense that an individual has in mind a range of possible values of the probability that can be described by a density function, $g(p)$, that can take on a variety of shapes depending on both its mean and the degree of ambiguity. The theory literature calls this density function second order probability beliefs. (See, for example, Gilboa and Marinacci, 2011).

We compare the predicted survival rates of our sample members to the actual survival of sample members eight years later, as reported in the 2010 wave of HRS. The predictions from our models track actual mortality fairly closely for sample members who are below age 80, but begin to diverge substantially at the oldest ages, with older respondents being overly

[36]The attempt to determine the probabilistic beliefs of lay people using direct questions on surveys has only become commonplace in the past two decades (Manski, 2004). There is a related but somewhat separate tradition of eliciting probability beliefs of experts as part of risk assessments in engineering and operations research applications. See Paté-Cornell (1996) for a summary of this tradition.

optimistic. The predicted survival rates co-vary with demographic characteristics, health status, parental mortality, smoking behavior and cognitive status in largely the same way that they do in regressions that explain actual mortality. Thus, it appears that the subjective survival probability answers are good candidates for modeling individual level heterogeneity in survival chances.[37]

The major contribution of this paper is to provide an estimate of ambiguity about survival probabilities that is embodied in the spread of $g(p)$. We find that survival expectations are very uncertain in the Health and Retirement Study (HRS) and that uncertainty varies in the population: more educated people have more certain beliefs, women have less certain beliefs, and deterioration of health, especially from previously excellent levels, leads to more uncertainty about survival chances.

Identifying second order probability beliefs is possible under the assumptions of our survey response model, the Modal Response Hypothesis (MRH) which is a mapping from probability beliefs $g(p)$ to survey responses. It assumes that people report the mode of $g(p)$ whenever it exists and they report 50%, whenever $g(p)$ is so ambiguous that it does not have a unique mode. The motivation for the MRH is twofold. First, imagine that people estimate the probabilities of certain events after observing some successes and failures of these events. Under some conditions a naïve estimator, which is the ratio of "good cases", is exactly equal to the mode of the Bayesian updated posterior probability distribution. Moreover, this estimator is biased in finite samples. The MRH assumes that people report this simple,

[37]Our findings about the external validity of the HRS subjective probability questions are consistent with those in earlier papers that have studied these questions in the HRS (Hurd and McGarry, 2002; Smith et al., 2001). The overly optimistic expectations of people over 75 in HRS has also been noted by Hurd et al. (2005) who find the same pattern across eleven European countries in the SHARE (Survey of Health, Ageing and Retirement in Europe). There are also a few papers that use these questions in models of behavior under uncertainty such as, for example, Picone et al. (2004) who find that people who expect to live longer are more likely to choose medical screening tests.

naïve estimator in surveys. The second motivation comes from the literature using numerical subjective probabilities on surveys. When people are asked to report a numerical probability using any digit between 0 and 100, an unusually large fraction of answers are heaped on 50. The excessive use of 50 has been interpreted as occurring because many people treat "50-50" as a synonym for "I don't know" or even for "God only knows," a sentiment that suggests that the true probability is unknowable (Fischhoff and Bruine De Bruin, 1999; Bruine de Bruin and Carman, 2012; Lillard and Willis, 2001).[38,39] Researchers have realized that these focal answers might introduce bias into estimates of subjective probabilities. As a remedy,Manski and Molinari (2010) think of focal responses as extreme versions of rounding, where some people always round to 0, 50 or 100 percent. Lumsdaine and van Loon (2013) model the probability of providing a focal answer as a separate equation in their econometric model. These papers found that accounting for focal responses is important for valid inference. However, they are agnostic about the reasons so many people provide such answers. Our approach, instead, is to model focal responses with an economic model and relate it to the precision of beliefs of individuals. In our model people answer with an epistemic 50 percent as a way of saying "I am very unsure", when their beliefs are too ambiguous. As far as we know the MRH model is the first in the literature that makes use of the focal responses to learn something about the beliefs of individuals.[40]

[38] In recent waves of the Health and Retirement Study, respondents have been asked a follow-up question if they answered 50 to survival probability question: Do you think that it is about equally likely that you will die before 75 as it is that you will live to 75 or beyond, or are you just unsure about the chances? About two-thirds say that they are "just unsure." In this paper, we use data from the 2002 wave of HRS which did not have a follow-up question.

[39] In addition, the answers to the survival questions also exhibit some heaping at 0 and 100. Taken literally, of course, these answers cannot represent rational probability beliefs except, perhaps, in the case of zero for persons who know themselves to be terminally ill or planning suicide.

[40] Manski and Molinari (2010) discuss an alternative, more direct approach to learn about the imprecision of belief. They make use of a follow-up question in a different survey (the ALP) asking about the range of probabilities responders had in mind when they provided their answers. Roughly half of the sample reported that their answers were "exact", and the rest reported relatively wide ranges of probabilities with the average

For comparison, we present an alternative model of survey response which assumes that the person reports the mean of $g(p)$. We show that the MRH model can account quite well for heaping at 0, 50 and 100 while the mean response model cannot. The overuse of focal responses can be especially problematic when very low or very high probabilities are modeled since the high ratio of 50 percent responses can arbitrarily push the estimated mean probabilities away from their true, extreme values. We show in this paper that our modal response model which explicitly models focal answers works better than the simple mean model typically used in research on subjective probabilities. When the modal response model is used for extreme probabilities both the estimated unconditional means and the average partial effects are closer to ones estimated from realized mortality data. Alternative models of focal responses, such as Manski and Molinari (2010) and Lumsdaine and van Loon (2013) might work equally well to reduce this bias and to predict enough focal answers. The main advantage of the MRH is that it is based on an economic model of limited information and uncertain beliefs.

Learning about the degree of uncertainty in individuals' beliefs can be important for several reasons. First, the value of information about mortality risk, for example, should be a function of this uncertainty. Uncertain people might value information more, and certain people might rationally ignore any new information since they have already established a good understanding of the risks they face. Even in a fully Bayesian Subjective Expected Utility (SEU) framework the degree of uncertainty might play an important role if the utility function is not linear in probabilities. This is the case, for example, if people can invest in

width being 18 percent. We believe that this approach is very informative about general uncertainty in the population, but it is less obvious how to make use of the provided ranges. First, the meaning of an "exact" answer is ambiguous. Second, one would expect that 50 percent responders provide wider ranges of possible probabilities than people with other answers ending with 0 or 5 (like 10 or 15 percent). However, they did not find a strong pattern like that in their data.

learning about their own mortality risks.[41] Second, learning about the degree of uncertainty in individuals' beliefs is useful from a survey methodological point of view, too. People whose probability beliefs are more certain may answer probabilistic survey questions with greater precision. However, if beliefs are uncertain, as is the case with mortality expectations, we expect large measurement error in survey responses, that might even account for the discrepancy between subjective and objective survival probabilities at later ages. Future research should investigate the potential role of measurement error in subjective expectation data and its link to uncertainty in individuals' beliefs. Third, learning about the degree of uncertainty in individuals' beliefs might be very important in Non-Bayesian/Non-Expected Utility models. While we don't work in these frameworks, empirical evidence on ambiguous beliefs may be useful to those who do.

The paper is organized as follows. In Section 1 we provide a quick overview of the literature on ambiguity and its role in economic decisions. Section 2 describes the subjective survival data in HRS that we use in this paper. Section 3 introduces the MRH model that can be used to model any subjective expectation data that uses the HRS framework. Section 4 describes a simple and tractable model of individual survival curves. Section 5 discusses the estimation method and identification and Section 6 shows the results.

3.2 Ambiguity in economics

Knightian uncertainty, epistemic uncertainty and ambiguity are roughly synonymous terms that figure prominently in a longstanding and ongoing debate about the link between rationality and probability beliefs, on the one hand, and the relationship between beliefs and

[41] Another example is Bommier and Villeneuve (2012), who discuss a model of mortality risk aversion, where the utility function is not additively separable over time, and thus, mortality probabilities enter the model in a non-linear fashion as well.

decisions, on the other hand. In an excellent and authoritative review of this debate since Pascal's famous bet on the existence of God in 1670, Gilboa and Marinacci (2011) discuss the different models of expectation formation, including Bayesian and Non-Bayesian models. They call the model we use in this paper "the smooth model of ambiguity" or "second order probability beliefs." In typical models of ambiguity, agents have a set of possible probability distributions in mind but they are not able to compound this information into a single probability distribution. The smooth model, however, makes compounding possible. The real question in the smooth model is whether agents have a preference for known probabilities (in other words they are ambiguity-averse) or not (in which case they are simple Bayesians). As Gilboa and Marinacci (2011) write "*... beyond the above mentioned separation [between beliefs and utilities], the smooth preferences model enjoys an additional advantage of tractability. Especially if one specifies a simple functional form for [preferences for known probabilities], one gets a simple model in which uncertainty/ambiguity attitudes can be analyzed in a way that parallels the treatment of risk attitudes in the classical literature.*", pp. 45.

The major contribution of this paper is to provide an estimate of second order probability beliefs about survival probabilities that are embodied in the spread of $g(p)$. It is important to clarify the role of this object in alternative views of uncertainty. Conventional economic theory of behavior under uncertainty, rooted in subjective expected utility (SEU) theory (Savage, 1954), is often interpreted as having erased the distinction between known and unknown probabilities because expected utility is a linear function of probabilities. That is, assume that the person's ambiguous probability beliefs are described by $g(p)$. His expected utility is

$$EU = \int_0^1 \left[pU_{live} + (1-p)U_{die}\right] g(p)dp = \bar{p}U_{live} + (1-\bar{p})\,U_{die}, \tag{3.1}$$

where is the mean of this ambiguous distribution. Clearly, expected utility is invariant to a mean preserving spread of $g(p)$; hence, decisions based on expected utility are unaffected by ambiguity.

In the famous Ellsberg experiments (Ellsberg, 1961) subjects were presented with choices of drawing balls from different urns whose compositions were either known or ambiguous. Most people revealed distaste for ambiguity which was at odds with the SEU theory. The survey by Gilboa and Marinacci (2011) provides a comprehensive and insightful discussion of this.

We separate the issues concerning ambiguity of probability beliefs from those concerning the effects of epistemic uncertainty and ambiguity aversion on decisions. We do so by utilizing survey data that asks directly about people's probability beliefs. As Manski (2004) emphasizes, this approach differs from the practice in much of applied economics of assuming that individuals have exogenously given probabilities. It also differs from Savage's theory which infers probability beliefs from choice situations. This means that we can explore empirically how beliefs about risk and uncertainty vary in the population without being required to take a stand on how decisions are affected by probability beliefs.

3.3 Subjective survival probability questions on the HRS

The Health and Retirement Study has asked probabilistic expectation questions on various topics since its beginning in 1992. The survival question that we use in this paper comes from the 2002 wave and it reads as follows: "What is the percent chance that you will live to be [TARGET AGE] or more?" The target age exceeds the individual's age by at least 10 years: it is 80 years for people below 70, and 85, 90, 95 and 100 for individuals in successive five year age intervals.

Although the subjective probability responses in HRS seem reasonable when averaged across respondents, individual responses appear to contain considerable noise and are often heaped on values of “0”, “50” and “100” (See for example Manski, 2004, for a discussion). Considering the whole group of probability questions in HRS-1998, for example, while only 5% of respondents refused to answer the probability questions, 52% of questions were heaped on either “0” or “100” and an additional 15% were heaped on “50”.

These patterns are illustrated in Figure 3.1 by histograms of responses to the HRS-2002 survival probability question. We have included separate histograms by the target age used in the survival question and a total histogram in the six panels of Figure 3.1. Each histogram shows a high frequency of focal answers, especially at 50. The ratio of 50 responses is somewhat smaller at old ages where the actual survival probabilities are low, but it still accounts for 16 percent of the responses for people above 84.

Some psychologists, especially Fischhoff, Bruine de Bruin and their colleagues (Fischhoff and Bruine De Bruin, 1999; Bruine de Bruin et al., 2000; Bruine de Bruin and Carman, 2012) have argued that answers of “50” may reflect “epistemic uncertainty;” that is, a failure to have any probability belief at all about the event in question or, at least, to have no clear idea of what the probability could be. Alternatively, of course, an answer of “50” might reflect a very precise belief about the probability that a fair coin will come up heads or perhaps a somewhat less precise belief that a given event is about equally likely to occur or not occur. Indeed, while HRS probability questions offer participants a scale of integers from 0 to 100, the large majority of “non-focal” answers are integers ending in “5” or “0”, suggesting that responses from most people involve rounding or approximation. See Manski and Molinari

(2010) for a discussion of the different rounding practices of survey respondents in the HRS and of potential remedies.

There has been much less emphasis in the psychological literature on focal answers at “0” or “100”. When a probability question concerns an event such as the chance of being alive ten or fifteen years from now, it does not seem credible to assume that a respondent who gives such an answer of “100” is completely certain he will be alive then and, apart from a person diagnosed with a terminal illness, an answer of “0” should not be taken at face value, either.[42]

Previous researchers have found that the tendency to give focal answers is associated with low education, lower cognitive ability and it varies with other demographic variables, too. (Lillard and Willis, 2001; Hurd and McGarry, 1995; Lumsdaine and van Loon, 2013). These covariates are known to correlate with mortality. Focal responses, thus, might bias estimated average survival probabilities if we take them at face value and the bias might be stronger in situations where the underlying probability is far from 50 percent. In order to test whether this is the case, in the empirical part of the paper we will compare estimated individual survival curves to actual survival of the respondents eight years later in 2010, which is the last available wave of the HRS.

[42] It is possible to regard answers of “0” or “100” as approximations which are no different in kind than rounded answers of “5”, “40” or “95”. However, in a discussion of Gan et al. (2005), Willis (2005) provides evidence against this interpretation.

3.4 Probability beliefs and the Modal Response Hypothesis

In this section, we describe a theoretical model which attempts to relate answers that an individual gives to a survey question about the subjective probability of a given event and his underlying probability beliefs. In our model we distinguish between ambiguity and epistemic uncertainty, with the latter corresponding to cases when information is limited.

Let us assume that person is faced with the problem of estimating the probability p_i of an event A. Initially he has no information about the probability of this event, he has an uninformed prior $p_i^{prior} \sim U(0,1)$, where U denotes the uniform distribution. The person observes event A happening $\alpha_i - 1$ times and not happening $\beta_i - 1$ times. In the survival context, for example, this means that a person is aware of $\alpha_i - 1$ people similar to himself who survived to the given target age and $\beta_i - 1$ similar people who died before reaching that age. It is well known that if this new information is used to Bayesian update one's beliefs about p_i, the posterior distribution has a Beta distribution with parameters α_i and β_i, $p_i \sim Beta(\alpha_i, \beta_i)$.

When faced with a survey question about the probability of event A, the person might respond with the mean or the mode of this distribution. When α_i and β_i are larger than one[43], the mean and the mode of the Beta distribution are

$$p_i^{mean} = \frac{\alpha_i}{\alpha_i + \beta_i}, \tag{3.2}$$

$$p_i^{mode} = \frac{\alpha_i - 1}{\alpha_i + \beta_i - 2}. \tag{3.3}$$

[43] Other cases will be discussed below.

A Bayesian agent would report p_i^{mean}, which is the expected value of the Bayesian updated posterior distribution. Note that p_i^{mode} is exactly equal to the naïve estimator of the probability that can be computed by the number of "good cases" which is $\alpha_i - 1$ over the number of all cases which is $\alpha_i + \beta_i - 2$. A frequentist agent, thus, would not report the mean but, rather, the mode of the distribution $g_i(p)$. The modal response hypothesis assumes that people report p_i^{mode} rather than p_i^{mean} to probabilistic survey questions for at least two reasons. First, as Lillard and Willis (2001) argue, it is cognitively less burdensome for a respondent to answer a survey probability question quickly by reporting the most likely value of p, given by the mode of $g(p)$, than it is to report the expected value given by $p_i^{mean} = \int p g_i(p) dp$. Second, p_i^{mode} is equal to a very simple rule-of-thumb estimator for the probability in question: the frequentist response. In a survey situation where people have to answer many questions in a very short timeframe, it seems a reasonable assumption that they give frequentist approximations to probability questions instead of Bayesian updating their priors. Moreover, in this model the mode is often a good approximation of the mean.

The formula in (3.3) does not give the mode of the distribution when either α_i or β_i is smaller than 1. Whenever $\alpha_i < 1, \beta_i \geq 1$ the distribution is always decreasing and has a unique mode at zero. Whenever $\alpha_i \geq 1, \beta < 1$ the distribution is always increasing and its unique mode is at one. Finally, if $\alpha_i < 1, \beta_i < 1$ the distribution has a U-shape and it has two maxima at zero and one. In this case one finds it more probable that the probability of event A is zero or one than that it is 50 percent.

We have motivated the use of the Beta distribution with a Bayesian framework where agents observe certain numbers of successes and failures. As we shall show, however, the distributions that occur when either α_i or β_i is smaller than 1 cannot be derived from Bayesian updating based on evidence. Indeed, these distributions correspond to situations in which, in effect, the agent has very little objective evidence on which to base his beliefs and, lacking evidence, tends to give conventional "epistemic " responses to survey questions about his probabilistic beliefs. In particular, we hypothesize that the person will respond with an extreme value of either zero or one when $g(p)$ is monotonically decreasing or increasing. When the distribution is U-shaped, we hypothesize that the person will answer "50" as a synonym for "God only knows" rather than as necessarily a belief that the outcome in question is equally likely to occur or not.[44]

To show how the shape of is related to the amount of evidence on which an individual bases his beliefs, we introduce two more parameters that are functions of and

$$\mu_i = \frac{\alpha_i}{\alpha_i + \beta_i}, \tag{3.4}$$

$$n_i = \alpha_i + \beta_i. \tag{3.5}$$

μ_i is the expected value of the distribution of the probability in question, and n_i is a measure of the precision of beliefs. Higher n_i means more precise beliefs: that is, a tighter $g_i(p)$ density function. Earlier, we argued that an uninformed agent with a uniform prior over the unit interval would update his prior after observing $\alpha - 1$ successes and $\beta - 1$ failures in a sample of $N = \alpha + \beta - 2$. Note that $B(1, 1)$ is a uniform distribution so that $n_i = \alpha + \beta = 2$ for an uninformed agent. Equivalently, such an agent observes no data since $N = 0$. Thus, a necessary condition for Bayesian updating is that the agent observes a positive amount of data, which implies that $\alpha > 1, \beta > 1$ and $N > 0$. As we have seen, any Beta function satisfying

[44] In the survival context, a U-shaped distribution could represent the beliefs of someone who is unsure whether he had inherited a genetically transmitted disease. In case he did, he might face a low survival probability to the target age, but if he did not he has a high probability of surviving. The posterior distribution of the survival probability in this case can have a U shape, where the extreme probabilities are more likely than any middle values. However, it is not plausible that such situations are common enough to account for the large number of "50" responses that we see in survey responses.

this condition is unimodal where the mode falls in the interval, $0 < p < 1$. Conversely, when $\alpha < 1$ or $\beta < 1$ $g(p)$ may be monotonically increasing, decreasing or U-shaped depending on the value of μ and if $\alpha + \beta < 1$, $g(p)$ is always U-shaped. Obviously, Bayesian updating cannot be the source of such beliefs since one cannot observe a negative number of signals! That is why we label such beliefs as "epistemic" and distinguish them from "ambiguous" ones.

In our development of the Beta model, the precision parameter, $n_i = \alpha_i + \beta_i$, is assumed to be an integer equal to two less than the size of the sample that is observed by agent i. A broader and more useful interpretation of precision is that it measures the confidence that an individual has in his judgment of the risk of a given event. For instance, educated individuals can utilize, in addition to their personal experience, a broader knowledge of evidence about mortality and its causes from past coursework, wider reading and better informed family and social networks. Thus, we may interpret precision as a measure of a person's capacity to assess his survival risks based on his knowledge of mortality risks and his ability to translate personal information about his own health, health behavior and family history into its implications for survival chances.

The relationship between n_i, μ_i and $g(p)$ is depicted in Figure 3.2. The figure presents a matrix of 81 probability density functions—$g(p|\mu, n)$ —corresponding to nine different values of the mean of $g(p)$, given by μ_i on the horizontal axis, and nine different degrees of precision, measured by n_i, on the vertical axis. The figure also illustrates the boundary between ambiguous beliefs that can be represented by a second order probability distribution based on Bayesian principles and epistemic uncertainty in which the individual has too little knowledge about the risk in question to be able to form an evidence-based probability judgment. Possible ambiguous densities appear in the darkly shaded, inverted U-shaped region in Figure 3.2 for which $n_i > 2$. Note that throughout this region, reports of p_i^{mean} and p_i^{mode} tend to be very close to one another. The lightly shaded area at the top of the figure corresponds to epistemic 50 percent responses. As we can see, when uncertainty is high (n_i is low) for any values of μ_i the model predicts a 50 percent response, including cases when the mean probability is very low or very high. The MRH, thus, predicts that the large fraction of 50 percent answers, typical in subjective probabilistic expectation data, can in fact correspond to mean probabilities that are far from 50 percent. The two unshaded triangular regions in Figure 3.2 correspond to epistemic 0 and 100 responses. Such responses are typical in cases when the mean probability, μ_i is also close to 0 or 100. However, when n_i is close to 2, that is beliefs are almost uniform, 0 and 100 percent answers can occur when is close to 50 percent. Thus, the MRH can predict a large fraction of focal 0, 50 and 100 answers and the corresponding bias can be either positive or negative.

To summarize, the modal response hypothesis claims that survey respondents report a potentially rounded version of the mode of $g_i(p)$ whenever it exists and they report 50 percent whenever has a U shape,

$$p_i^{mrh} = \begin{cases} round\left(\frac{\alpha_i - 1}{\alpha_i + \beta_i - 2}\right) & \text{if } \alpha_i > 1, \beta_i > 1 \\ 1 & \text{if } \alpha_i > 1, \beta_i \leq 1 \\ 0 & \text{if } \alpha_i \leq 1, \beta_i > 1 \\ 0.5 & \text{if } \alpha_i \leq 1, \beta_i \leq 1. \end{cases} \tag{3.6}$$

The round function can be rounding to the closest 1 percent, 5 percent, 10 percent or anything that seems appropriate in the context of the survey. Manski and Molinari (2010), for example, use a framework where there are individual differences in rounding practices. Their approach can also be modeled in our framework by letting the round function vary across individuals.

The hypothesis of this paper is that people answer subjective probabilistic expectation questions according to the MRH. It would be desirable, however, to test the MRH against other survey response models. A natural candidate for comparison is the mean model where people respond a potentially rounded version of the mean of $g_i(p)$

$$p_i^{mm} = round(\mu_i). \tag{3.7}$$

In the mean model the precision of beliefs (n_i) is not identified, only the mean (μ_i) is. The mean and the mode models, however, converge to each other as n_i goes to infinity. That is, the mean model is embedded in the MRH, and thus, a relatively small estimated belief precision (n) would be evidence in favor of the MRH.

3.5 Individual subjective survival curves

In the previous section, we introduced two survey response models that transform second order probability distributions $g(p|\mu_i, n_i)$ into the survey response p_i^{mrh} or p_i^{mm}. These models can be applied to any subjective probabilities in the HRS format and not just to survival data. To close the model, however, we need to specify the mean (μ_i) and the precision (n_i) of beliefs. There is no unique way of modeling these two variables; it is the task of the researcher to find the appropriate model in the context of the particular project. In this section, we show how μ_i and n_i can be modeled in the context of survival probabilities.

The so-called Gompertz model of longevity has been widely used in both demography and biology because its increasing mortality hazard assumption lines up with mortality data of humans and other species very well (Vaupel, 1997). The Gompertz model assumes that the hazard of death is exponentially increasing with age:

$$h(a) = \gamma_0 \gamma_1 \exp(\gamma_1 a), \tag{3.8}$$

where γ_0 is a positive scale and γ_1 is a positive shape parameter. By simple calculation, (3.8) leads to the following survival probability from age a to age t:

$$S(a,t) = \exp(-\gamma_0 (\exp(\gamma_1 t) - \exp(\gamma_1 a))). \tag{3.9}$$

The main advantage of subjective survival data is that we can estimate individual heterogeneity in survival chances. With objective survival data we can only identify group-specific survival probabilities by computing the ratio of survivors in a particular group. Unobserved heterogeneity within groups, however, is not identified. In contrast, subjective survival data enables us to estimate individual heterogeneity in survival chances as we collect probability data on the individual level. We follow Vaupel et al. (1979) by allowing the scale parameter (γ_0) to have a gamma distribution with shape parameter k and scale parameter θ and we assume that the shape parameter (γ_1) is fixed in the population:

$$\gamma_{0i} \sim \Upsilon(k, \theta). \tag{3.10}$$

The expected value of the gamma distribution is $k\theta$ and thus both parameters increase mortality chances and decrease the probability of survival (see equation (3.9)). The main advantage of using the gamma distribution, as we shall argue, is analytic tractability. Note, however, that this is a flexible 2 parameter distribution with both parameters being estimated, and hence this assumption is not very restrictive. The first advantage of the gamma distribution is that the average survival probabilities can be derived analytically. As we show in Appendix C2 in the e-companion online supplement, the average survival probabilities from age a to age t is

$$E\left(S_i\left(a,t|k,\theta,\gamma_1\right)\right)=\left(1+\theta\left(\exp\left(\gamma_1 t\right)-\exp\left(\gamma_1 a\right)\right)\right)^{-k}. \tag{3.11}$$

The second advantage of the gamma-Gompertz framework is that we can analytically derive the effect of individual heterogeneity on sample selection. Different survival chances are modeled by letting γ_0 have a distribution in the population. In this paper we refer to γ_0 as "frailty" (Vaupel et al., 1979), which includes genetic, environmental and behavioral factors that affect the underlying mortality of individuals other than age. As long as survival chances are heterogeneous in a population, fit individuals will be overrepresented in the sample over time as frail individuals are more likely to die and not participate in the HRS. By applying the formula from Vaupel et al. (1979) we can analytically characterize this sample selection. See the appendix for details. To sum up, we use the following structural equations for individual survival chances.

$$\begin{aligned}
\mu_i &= S_i\left(a,t\right)=\exp\left(-\gamma_{0i}\left(\exp\left(\gamma_1 t\right)-\exp\left(\gamma_1 a\right)\right)\right), && (3.12)\\
\gamma_{0i} &\sim \Upsilon\left(k,\theta\right), && (3.13)\\
\theta_i &= \frac{1}{\frac{1}{\theta^r}+\left(\exp\left(\gamma_1 a\right)-\exp\left(\gamma_1 r\right)\right)}, && (3.14)\\
\theta^r &= x_i'\beta_\theta. && (3.15)
\end{aligned}$$

θ^r represents the scale parameter of the gamma distribution in the reference cohort which is the 50 year old ($r=0.5$). Equation (3.14) shows the effect of sample selection. The scale parameter is decreasing with age, as fit individuals are increasingly overrepresented in the sample. Equation (3.15) shows that we add covariates to the scale parameter of the 50 year old cohort. Finally the Appendix shows how average partial effects of the different covariates can be derived after fitting this model.

So far we have only talked about how to model the mean survival probability, μ_i. For modeling the precision of beliefs (n_i), we use a very simple log-normal framework.

$$\ln n_i = z_i'\beta_n + u_{ni}, \tag{3.16}$$

$$u_{ni} \sim N(0, \sigma_n^2). \tag{3.17}$$

Equations (3.12)-(3.17) together with the survey response models of the previous section fully specify our model.

3.6 Estimation and identification

Our structural model has two unobservables: γ_{0i} which is a function of the mean survival probability and u_{ni} which is the unobserved heterogeneity in the precision of beliefs. Based on the distributional assumptions from the previous section, the model is fully specified and it can be estimated with maximum likelihood.

We only observe one survival probability answer in HRS. In Section 3, we proposed two survey response models. The mean model assumed that people report a rounded version of their true survival chances, while the MRH model assumed that people report a rounded version of the mode of the distribution of probability beliefs $g(p|\mu_i, n_i)$ or 50 percent when the mode does not exist.

The joint distribution of the two random variables γ_{0i} and u_{ni} is complicated because one is gamma, the other is normal and they enter the model in a non-linear fashion. The estimation of the MRH, thus, can be carried out by maximum simulated likelihood. The estimation of the mean model is more straightforward as the precision of beliefs plays no role in the model. The appendix shows how the likelihood function of these two models can be constructed.

It is worth discussing how our main parameters are identified. We seek to estimate the following set of parameters: $\gamma_1, k, \beta_\theta, \beta_n, \sigma_n$. In the case of the mean model, parameters of the belief precision, β_n and σ_n are not identified. Parameter γ_1 is the shape parameter of the individual survival function and it is identified from how fast the probability responses $\left(p_i^{hrs}\right)$ change with age $\left(\partial E\left(p_i^{hrs}|a\right)/\partial a\right)$. Parameters k and β_θ determine the scale parameter of the individual subjective survival curves (γ_{0i}) and they are identified from the location and dispersion of the individual responses $\left(E\left(p_i^{hrs}|x_i\right) \text{ and } V\left(p_i^{hrs}|a\right)\right)$.

The identification of the belief precision parameters β_n and σ_n primarily comes from the fraction of different focal answers in different demographic groups. If we have many focal answers, we expect n_i to be small. If we have many different types of focal answers (0, 50 and 100), we expect a high σ_n, indicating a high dispersion of belief precision in the population.

3.7 Empirical analysis

Beliefs about subjective survival probabilities presumably depend on an individual's knowledge of different risk factors and demographic differences in the society; his personal information about his own health, habits and family members' longevity; his ability to translate this information into a probability and on his level of optimism or pessimism. In the empirical model estimated in this section, we use the information available in the HRS to try to capture several of the major determinants of beliefs in a parsimonious fashion.

3.7.1 Sample and measures employed

The sample used in the empirical analysis consists of 13,038 age eligible respondents to the 2002 Health and Retirement Study, over age 54 in 2002,[45] who provided responses to the subjective survival probability question. Excluded from the sample are proxy respondents and non-respondents in the 2000 wave of data collection. Also excluded are persons over 90 who were not asked the survival probability questions and people for whom we did not have realized survival information in 2010.[46] Table 3.1 presents descriptive statistics for the variables used in our analyses. The average age of our sample members is just over 68 years (ranging from 54 to 89 years) and on average the target age was 16 years from their current age. The modal sample member is a white female with a high school education although there is substantial variance in each of these dimensions.

Our theory suggests that people may utilize personal information in forming their subjective survival beliefs. Parents' age at death, one's own current health status and health behavior are such variables. As Table 3.1 shows, 16 percent of our sample still has a living mother but only 5 percent has a living father. The average age at death of the mothers is 76 years while the corresponding number for the fathers is 72 years. We construct six variables about parental mortality. Three of them correspond to the mortality of the same sex parent (father-son, mother-daughter pairs) and three correspond to the opposite sex parent. Within each pair we first take the age at death of the parent if he/she is dead. If (s)he is alive we impute the expected age at death of the parent based on his/her gender and age. In all mod-

[45] The 2002 HRS is a representative panel sample of the 54+ population and their spouses.

[46] Actual mortality of HRS respondent is very precisely measured from administrative data (the National Health Index) and it is even available for people who dropped out of the survey in a later wave.

els we include dummy variables about whether the parents are alive to control for potential imputation bias. Finally we create a linear spline from the imputed parental mortality with a single cut-off point at the age of the interviewee. This approach is motivated by the idea that individuals might consider parental mortality less informative about their own survival chances if they have already lived longer than their parents.[47] These four splines (two for each parent) together with the two dummies constitute our six parental mortality variables.

We also include in our analysis three sets of variables on health related behavior. As Table 3.1 shows, 43 percent of our sample reports regular exercise at least three times a week. While only 14 percent of the sample smoked in 2002, almost 60 percent reported having smoked in the past. There is a big variation in the sample in drinking behavior. Roughly half of our sample (48 percent) reports that they drink alcohol sometimes, but the majority are not regular alcohol consumers. Among those who are, the average number of days when they drink is 3.4 days a week, and the average number of alcoholic beverages consumed is 1.9. Self-rated health in the HRS is measured on a five-point scale–1) Excellent, 2) Very Good, 3) Good, 4) Fair and 5) Poor.[48] We translate these into three categories: 1) excellent/very good; 2) good; 3) fair/poor . We then construct dummy variables representing the combination of self-rated health in 2000 and 2002 for each respondent with excellent/very good in both years as the baseline case. In Table 3.1 we only show the marginals of this joint distribution. As we can see the fraction of people in "excellent/very good" health decreased from 47 percent in 2000 to 43 percent in 2002. The fraction of people in "good" health did not change much (31 and 32 percent) and the fraction in "fair/poor" health increased from

[47]Simple exploratory work suggested that parental mortality has a stronger effect on expectations when the parent died at an old age. The specification we use in this paper is not the only way to allow for such non-linearity and future work should explore alternative models.

[48]These three categories represent relatively good, average and relatively bad health. The reason for the aggregation is that we interact subjective health in 2000 and 2002 and adding 5 x 5 = 25 interactions would make the interpretation of these variables difficult.

22 to 25 percent.

We also include two cognitive measures ("Vocabulary" and the "27-point cognitive capacity scale"[49]), and the CESD depression score which measures depressive symptoms.[50] Table 3.1 shows that the average member of our sample has higher cognitive and lower depression scores than the average HRS respondent.

3.7.2 Maximum likelihood model estimates

The objective of this section is to test the modal response hypothesis (MRH) through a series of performance tests. All the results we present in this section are based on the twelve estimated models shown in Table 3.2, 3.7, 3.11 and 3.15. Table 3.2 reports models without covariates, Table 3.7 adds basic demographic information, Table 3.11 expands the model with cognition, personality and parental mortality and finally 3.15 contains all variables including subjective health. All tables present three models. The first columns show actual eight year survival of the HRS respondents between 2002 and 2010 estimated with nonlinear least squares. The second columns shows results using subjective expectations based on the mean model and the last two columns show results from the MRH "mode" model. All the parameters of our model $(\gamma_0, \theta^{0.5}, n, \gamma_1, k, \sigma_n)$ are assumed to be positive and thus their logarithms enter the likelihood function. Covariates potentially enter two equations. The first is the equation of $\theta^{0.5}$, which is the scale parameter of the gamma distribution of the mortality hazard at the age of 50. In the actual survival models we model γ_0 directly. Covariates with positive coefficients are estimated to increase the mortality hazard and decrease the survival chances. The magnitudes of these coefficients will be analyzed later when we derive average

[49]The 27-point scale Langa-Weir method is discussed in Crimmins et al. (2011). HRS cognitive measures are described in Fisher et al. (2012).

[50]See Ofstedal et al. (2005).

partial effects of them on various survival probabilities. The second equation where covariates appear is the equation of the precision of beliefs (n). Positive coefficients mean tighter, more precise probability beliefs.

Panel A of Figure 3.3 compares estimated actual 8 year survival probabilities of HRS respondents to subjective survival beliefs in 2002 computed from the models in Table 3.2 with no covariates. The horizontal axis shows the current age of respondents in 2002 and the vertical axis shows the fitted average 8 year survival chances from the three models. As we discussed in Section 3, heterogeneity in survival chances leads to sample selection as people with better fitness are more likely to survive and become respondents of the HRS survey at older ages. In the case of realized survival, only the interpretation of the estimates changes, but we do not need to make any further adjustments of the parameters. The demography literature calls survival tables of this sort "Survival probabilities of the survivors." In the case of subjective survival chances, however, we do have to properly adjust for the unmeasured genetic and environmental differences of cohorts as discussed in Section 4. As we can see, both the mean and the MRH "mode" model track the actual survival chances very well up until about age 84, when the subjective probabilities become too optimistic. While the 8 year actual survival chance of a 90 year old is roughly 20 percent, the corresponding numbers in the MRH and mean models are ˜45 and 55 percent, respectively. Thus, although the MRH model provides numbers that are closer to the true survival chances at old ages, these numbers are still too large on average. It is not obvious, however, whether these overly optimistic numbers are biases in people's heads or biases due to measurement error in the survey. In this paper we do not try to separate these two types of bias and we simply compare the mean, the mode and the actual survival models using the raw data.

Panel B of Figure 3.3 shows the estimated heterogeneity of survival chances in our sample.

The different curves correspond to different values of γ_0 with lower values meaning better fitness. As we can see, there is notable variability in survival chances. For example, the difference in median survival (i.e., half-life) between those in the 10th and 90th percentiles of the estimated frailty distribution is about 25 years. That is, comparing two groups of 50 year olds, half of those in the 90th percentile are expected to survive to age 70 while among those in the 10th percentile half are expected to survive to age 95. Using only mortality data, one cannot identify the unobserved heterogeneity in survival chances.[51]

The reason the MRH model is somewhat better than the mean model in predicting low probability events is that the high fraction of 50 percent answers are allowed to be focal answers that do not arbitrarily push the mean survival chances up. To visualize this effect we simulated survey responses based on the estimated models of subjective survival chances in Table 3.15. To get precise numbers, we used 651,900 observations for simulation which is 50 times the size of our dataset ($651,900 = 50 \times 13,038$). As we can see in Figure 3.4, the MRH model is able to predict histograms of responses that are very similar to the histogram of actual responses in the bottom panel. The ratio of 50 percent answers is around 25 percent, while the ratio of 0 and 100 percent answers are both around 10 percent. What is more important, the MRH model recognizes that the high fraction of focal answers should not be taken at face value as a large fraction of them only reflect imprecise knowledge. The mean model, however, takes all the focal answers at face value. Consequently, the mean model is not able to predict a histogram similar to actual responses, and it seriously biases the

[51]There is a long history of discussion about the difficulty in separately identifying duration effects and unobserved heterogeneity. See, for example, Vaupel et al. (1979) and Heckman and Singer (1984). Using subjective survival data, however, identifying unobserved heterogeneity in frailty is easier, because we observe probabilities of survival on the individual level. This contrasts with the use of mortality data, where it is hard to know which survivor is fit and which is simply luckier than other non-survivors. Even though we use a particular functional form for how unobserved heterogeneity enters the model (the gamma-Gompertz framework) it is important to note that these functional form assumptions are not needed for identifying unobserved heterogeneity in subjective frailty.

estimation of low or high probability events.

Much of a person's personal information about health is likely embodied in his assessment of the level and trajectory of self-rated health. To explore the effects of other covariates, we present estimates of models with and without subjective health in Tables 3.4 and 3.3 respectively. Both tables show estimated average partial effects of surviving from age 55 to 75 and from 75 to 95. Appendix B in the e-companion contains more detailed versions of these tables and further specifications, including models with only demographic variables and partial effects of surviving from 2002 to 2010.

The tables show that the majority of the coefficients are smaller in absolute value when subjective as opposed to objective information is used, but the coefficients based on the MRH are closer to objective values. This pattern is more obvious for low probability events (surviving from age 75 to 95), as the coefficients are roughly two times as big in the MRH. Thus, not only the average survival probabilities (Figure 3.3), but the average partial effects are also closer to objective values compared to the mean model. We take it as evidence that the MRH model is more successful for modeling subjective probabilistic expectations, particularly for low probability events, because the bias from focal answers is explicitly modeled.

However, the majority of the coefficients are still smaller in absolute value in the MRH than in the actual survival models. There are two potential explanations for this. Either there is measurement error in the data or people are underestimating the effect of some variables when forming expectations about their longevity. Without further assumptions, measurement error, other than focal responses, is not identified in our model. Note, however, that not all coefficients are attenuated equally. In fact two coefficients are systematically higher in absolute value in the subjective models: same sex parents' age at death and depression.

This is consistent with a model in which agents base their expectations on easily observable determinants of mortality (parent's survival) because they have limited information on demographic differences in the society and on the role of different behavioral factors, such as smoking and exercising, on survival. In this case the partial effects of personal information is expected to be higher for subjective than for objective survival; and the pattern is expected to be the reverse for other variables. This is exactly what we see in the data.

Tables 3.3 and 3.4 also show that people are aware that regular exercising is beneficial, and smoking is harmful for them, although they underestimate the role of these factors. We can also see that regular but limited alcohol consumption is not damaging, while teetotalling and heavy drinking are. Even more interesting is the comparison of the models with and without controlling for subjective health. Not surprisingly both the objective and subjective survival probabilities are less affected by behavior when subjective health is controlled. However, the subjective values shrink more strongly. For example, as Table 3.3 shows those who quit smoking expect a 2 percentage point lower chance of surviving from age 55 to 75 compared to those who never smoked. However, this effect disappears when subjective health is taken into account (Table 3.4). It means that healthy quitters falsely believe that their survival chances are the same as those who never smoked; whereas quitters who had already acquired a disease understand its consequences.

It is also worth noting the demographic differences in subjective and objective survival. As expected, females are more likely to live longer and this is reflected in their subjective expectations. Racial differences, however, are more complex. Conditional on education, personal information, health and behavior, blacks and Hispanics have a higher chance of survival than whites in this age range. African Americans' expectations reflect this difference, but Hispanics are more pessimistic than Non-Hispanics. Table 3.8-3.10 in the e-companion show roughly similar patterns when only demographic variables enter the model. Finally even though the educated are more likely to live longer (Table 3.8-3.10), when personality, personal information and behavior are controlled, the effect of education becomes negative (!). Further investigation, not shown in the paper, shows that this result is driven by the cognitive capacity variable. Education and cognition have a strong positive correlation but cognition is a better predictor of survival than education. Moreover, the educated also have better subjective health, and thus the effect of education becomes even more negative when health is controlled (Table 3.4). Moreover, the educated believe themselves to have better chances of surviving independent of the control variables used.

Finally let us take a look at the estimated distribution of probability beliefs (second order probability distribution, $g_i(p)$) of HRS respondents. Based on the MRH model without covariates (Table 3.2) we computed the 10th, 25th, 50th, 75th and 90th quantiles of belief precision (n) and the scale parameter of the survival function (γ_0) for the cohort of age 50. These numbers can be found in Table 3.6. Figure 3.5 shows the corresponding probability belief distributions of the probability of surviving from age 50 to age 80. As we can see, there is an enormous heterogeneity in probability beliefs. The median responder in HRS (3rd row and 3rd column of Figure 3.5) has a belief distribution that is single peaked but wide, having significant probability mass for any possible probability values between zero and one. It means that although the median responder's best guess for the probability of surviving from age 50 to age 80 is roughly 50 percent, she is quite unsure about this probability. People with even less precise beliefs are very unsure. For example, already at the 25th percentile of belief precision (where $n = 1.58$) everyone provides a focal response of either 0, 50 or 100. At the 10th percentile (where $n = 0.4$) everyone has U-shaped beliefs and, thus, responds with an epistemic 50%[52]. As we increase belief precision to the 75th percentile, the second order probability belief distribution becomes quite tight, having most of its mass in the neighborhood of the mean probability. It means that at least 25 percent of the respondents have very precise beliefs about their own survival chances.

Determinants of belief precision appear in Table 3.5 and the last columns of Table 3.7,3.11 and 3.15. Positive coefficients mean tighter, more precise beliefs. As we can see more educated people have more certain beliefs. This is consistent with our hypothesis, discussed in Section 3, that more educated people may have a broader knowledge of evidence about mortality and its causes. We can also see that the deterioration of health, especially from previously excellent levels, leads to more uncertainty about survival chances, perhaps because of uncertainty about the future course of a new disease. Those who were in poor health both in 2000 and in 2002, however, hold the most certain and pessimistic beliefs about their survival chances.

The effect of age and the time horizon of the survival question in HRS have complicated relations to uncertainty. For a fixed time horizon, the net effect of age on uncertainty is negative, because the interaction term dominates for any time horizon values used in HRS (from 11 years to 26 years). Thus, older people seem to have more precise beliefs about their survival chances which might reflect learning. For a fixed age, the net effect of the time horizon on uncertainty is also negative, because the interaction term dominates again. This means that people hold more precise views about their long run than their short run survival

[52]Note, however, that the particular shape of the distribution of beliefs is only identified from the lognormal functional form in the region where n is below 1. Because there are many focal responses in the HRS data, the model estimates many uncertain responses where $n < 1$. It is hard to know, however, what the distribution of looks like, conditional on being below 1. The log-normality assumption might or might not describe this conditional distribution well. It is possible, for example, that no-one has U-shaped beliefs, but all epistemic 50 responses come from a uniform distribution. If that is the case, then the log-normality assumption of n is inappropriate.

chances.

The two splines measuring same sex parental mortality in Table 3.11 and 3.15 shows a "V" shape. It means that people are the most unsure about their survival chances when they are around the age when their same sex parent died. We can also see that active smokers and those who have already quit have less precise survival expectations compared to those who never smoked; infrequent alcohol consumption leads to more precise beliefs; women are less sure than men and depressed people are relatively more certain about their otherwise poor survival expectations.

3.8 Conclusion

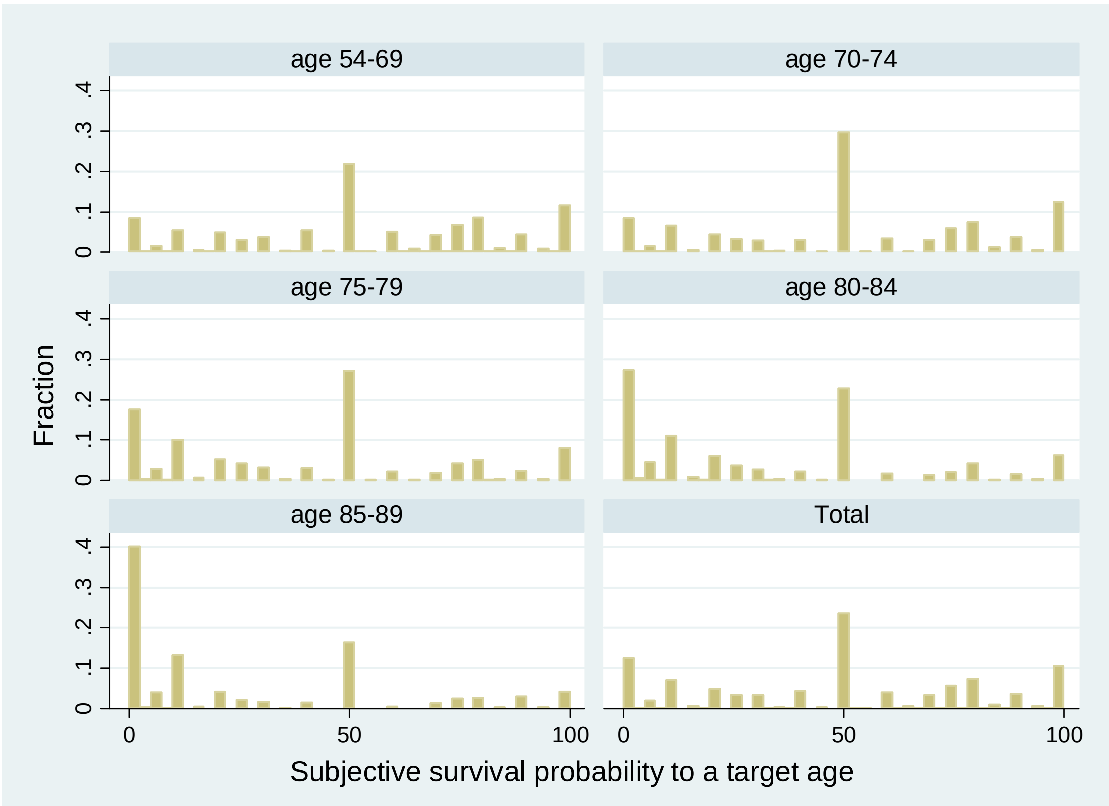
The modal response hypothesis is used in this paper as the foundation for an econometric model that is intended to provide a mapping between survey responses to probability questions and the underlying subjective probability beliefs of individuals about their chances of surviving to a target age. In this paper, we have presented the MRH as a hypothesis designed to capture the kinds of "gut response" to such questions that would be made after about 15 seconds of consideration by persons who vary in the amount of information they have about actuarial risks to health, about their own health-related circumstances and in their capacity to process such information into subjective beliefs. We argued in Section 3 that reporting the mode is relatively easier from a cognitive point of view than the mean or the median; the mode is equal to a very simple rule-of-thumb estimator for the probability in question; and that the mode often provides a good approximation to the expected probability that is called for in SEU theory.

Our empirical findings suggest that there is considerable heterogeneity in subjective survival risks, some of it associated with age, sex, race, education, health related behavioral factors, parental mortality and cognitive capacity. We have shown that subjective survival expectations line up with actual mortality very well when the objective probabilities are moderate. The subjective survival probabilities, however, become overly optimistic at old ages when the true survival probabilities are relatively low. We have shown that the MRH model does a better job compared to a standard mean model in reducing this bias as the MRH models focal answers in an explicit way. It remains for future research to learn whether the overly optimistic subjective expectations are biases in the heads of individuals, potentially having behavioral consequences, or they are a result of survey measurement error, potentially being related to uncertain beliefs.

In the empirical section of this paper, we have also found substantial uncertainty about mortality risks which is manifested by considerable spread in the estimated distribution of subjective survival probabilities for a typical respondent. In addition, we found significant variation in uncertainty, holding expected survival risk constant. It remains for future work to explore the explanation of these findings more deeply and to see whether survival risk and uncertainty about this risk play a role in decisions made by HRS respondents.

Tables and figures

Figure 3.1: Distribution of Survival Probabilities to Target Age, by Age of Respondent, HRS-2002



Note: Target age is 80 years for people below 70, and it is 85, 90, 95 and 100 for individuals in successive five year age intervals

Figure 3.2: Density of probability beliefs ($g_i(p)$) for different mean (μ_i) and precision (n_i) values

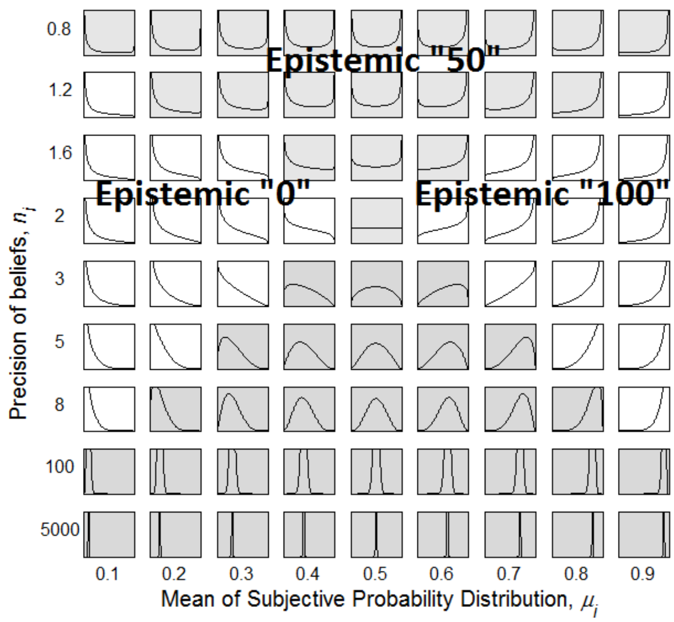
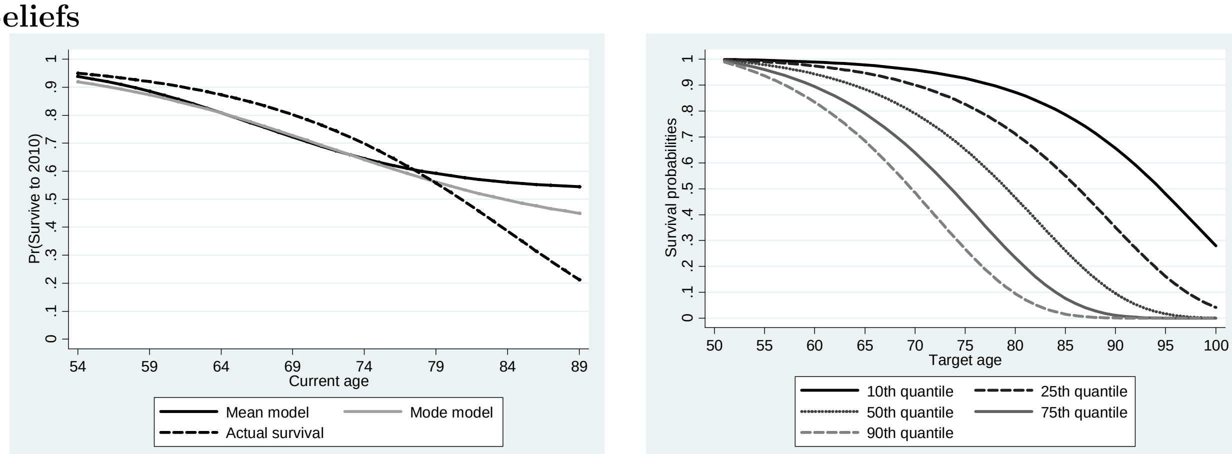


Figure 3.3: Average actual and expected survival probabilities and dispersion in beliefs



Panel A: 8 year actual and expected survival probabilities by age in 2002

Panel B: Heterogeneity in the subjective survival curves, MRH model

Figure 3.4: Simulated survey responses based on the mean and the mode models with all covariates and the empirical distribution of survey responses

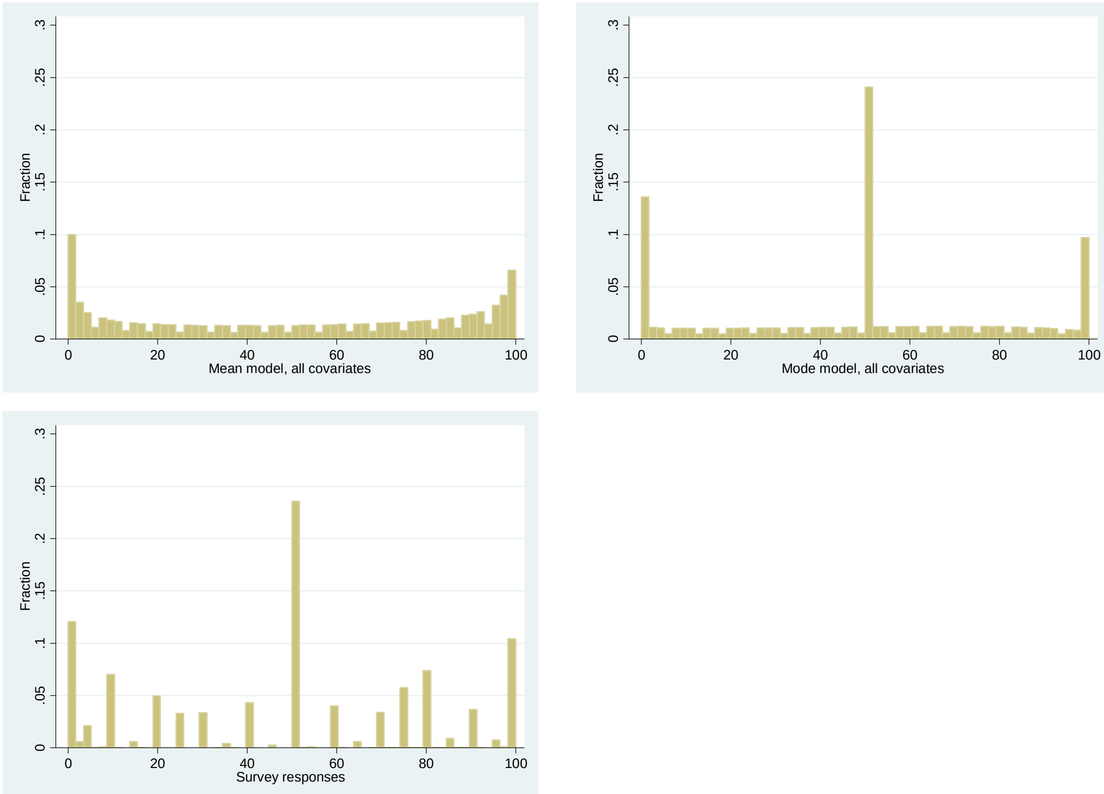


Figure 3.5: Estimated distribution of probability beliefs $g_i(p)$ of surviving from age 50 to age 80

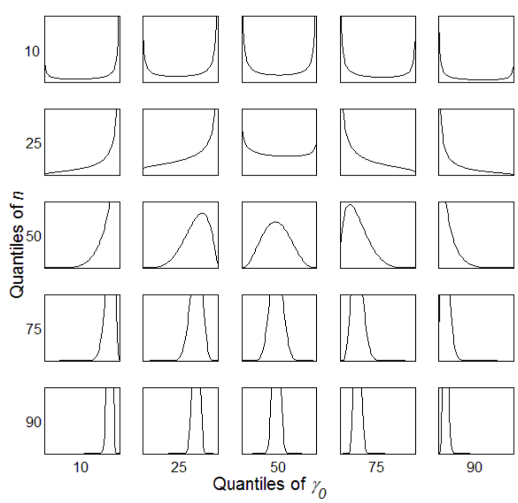


Table 3.1: Descriptive statistics, HRS-2002

	mean	sd
Alive in 2010	0.77	0.42
Subjective survival probability to target age	48.40	32.13
Age	68.15	8.69
Target age less actual age	15.97	4.17
Female	0.59	0.49
Black	0.12	0.33
Hispanic	0.06	0.24
Years of education	12.54	3.00
Mother is alive	0.16	0.37
Mother's age of death/100 or current age	0.76	0.15
Father is alive	0.05	0.23
Father's age of death/100 or current age	0.72	0.14
Exercises at least 3 times a week	0.43	0.49
Ever smoked	0.59	0.49
Smokes now	0.14	0.34
Ever drinks alcohol	0.48	0.50
# of days a week when drinks alcohol	1.10	2.08
# of days a week when drinks alcohol if positive	3.42	2.34
# of drinks when drinks alcohol	0.61	1.18
# of drinks when drinks alcohol if positive	1.92	1.36
Health excellent / very good, 2002	0.43	0.49
Health good, 2002	0.32	0.47
Health fair / poor, 2002	0.25	0.43
Health excellent / very good, 2000	0.47	0.50
Health good, 2000	0.31	0.46
Health fair / poor, 2000	0.22	0.42
Cognition score, std.	0.08	1.01
Vocabulary score, std.	0.11	0.97
CESD depression score, std.	-0.04	1.04
N	13038	

Table 3.2: Actual survival until 2010 and the mean and MRH models of subjective survival expectations, models without covariates

	Actual survival	Mean model	MRH
$\ln \gamma_0$	-8.404		
	[0.21]***		
$\ln \theta^{0.5}$		-11.174	-8.827
		[0.150]***	[0.169]***
$\ln n$			1.978
			[0.040]***
$\ln \gamma_1$	2.277	2.73	2.397
	[0.024]***	[0.012]***	[0.020]***
$\ln k$		-0.656	0.121
		[0.011]***	[0.025]***
$\ln \sigma_n$			0.814
			[0.025]***
N	13038	13038	13038
Log-likelihood		-57961.459	-47058.606

*, ** and *** denote significance at 10, 5 and 1 percent

Table 3.3: Average partial effects of surviving from age 55 to 75 and from 75 to 95 in three models with demographic, personality and personal information variables: actual survival and the mean and MRH models of subjective survival expectations

	55-75			75-95		
	Actual survival	Mean	MRH	Actual survival	Mean	MRH
Demographics						
Education	-0.003**	**0.009*****	**0.009*****	-0.004**	**0.003*****	0.005***
Female	0.073***	0.035***	**0.051*****	0.097***	0.012***	**0.029*****
Black	0.043***	**0.031*****	0.07***	0.058***	0.011***	**0.04*****
Hispanic	0.055***	**-0.052*****	-0.04**	0.073***	**-0.019*****	-0.023**
Cognition and personality, standardized scores						
Cognition	0.055***	**0.02*****	0.019***	0.073***	0.007***	**0.011*****
Vocabulary	0.003	**0.003**	-0.003	0.004	**0.001**	-0.002
Depression	-0.031***	**-0.053*****	-0.06***	-0.041***	-0.019***	**-0.034*****
Parents' age at death, linear splines with cut-off at own age						
Same sex, sp1	0.026	**0.068**	-0.048	0.035	**0.024**	-0.028
Same sex, sp2	0.159***	**0.314*****	0.415***	0.21**	0.112***	**0.239*****
Same sex lives	0.011	0.001	**0.006**	0.014	0	**0.003**
Opp. sex, sp1	0.038	0.186***	**0.156*****	0.051	**0.066*****	0.09***
Opp. sex, sp2	0.144**	0.097**	**0.116****	0.191**	0.035**	**0.067****
Opp. sex lives	0.028	**0.007**	**0.007**	0.038	0.002	**0.004**
Health related behavior						
Exercises	0.096***	0.054***	**0.061*****	0.128***	0.019***	**0.035*****
Ever smoked	-0.076***	**-0.02*****	-0.015**	-0.101***	-0.007***	**-0.008****
Smokes now	-0.097***	-0.059***	**-0.06*****	-0.128***	-0.021***	**-0.035*****
Ever drinks	0.056***	**0.044*****	0.033***	0.074***	0.016***	**0.019*****
# of days drinks	0.007***	0.004*	**0.005****	0.01***	0.001*	**0.003****
# of drinks	-0.012***	**-0.009****	-0.008**	-0.016***	-0.003**	**-0.004****

*, ** and *** denote significance at 10, 5 and 1 percent level; Bold partial effects at the Mean and the MRH models indicate that they are closer to the partial effects of actual survival.

Table 3.4: Average partial effects of surviving from age 55 to 75 and from 75 to 95 in three models with demographic, personality, personal information and subjective health variables: actual survival and the mean and MRH models of subjective survival expectations

	55-75			75-95		
	Actual survival	Mean	MRH	Actual survival	Mean	MRH
Demographics						
Education	-0.006***	0.006***	**0.004*** **	-0.007***	**0.003*** **	**0.003*** **
Female	0.067***	0.026***	**0.039*** **	0.086***	0.01***	**0.024*** **
Black	0.051***	**0.046*** **	0.086***	0.067***	0.018***	**0.053*** **
Hispanic	0.071***	-0.04***	**-0.019**	0.092***	-0.016**	**-0.012**
Cognition and personality, standardized scores						
Cognition	0.05***	**0.009** **	0.008**	0.064***	0.004**	**0.005* **
Vocabulary	0.002	**0.003**	-0.003	0.003	**0.001**	-0.002
Depression	-0.009***	**-0.026*** **	-0.031***	-0.012***	**-0.01*** **	-0.019***
Parents' age at death, linear splines with cut-off at own age						
Same sex, sp1	0.019	0.104**	**-0.022**	0.025	**0.041** **	-0.014
Same sex, sp2	0.122**	**0.268*** **	0.375***	0.157**	**0.105*** **	0.231***
Same sex lives	0.011	-0.001	**0.002**	0.014	0	**0.001**
Opp. sex, sp1	0.043	0.206**	**0.155*** **	0.056	**0.081*** **	0.095***
Opp. sex, sp2	0.093	0.061	**0.1** **	0.12	0.024	**0.061** **
Opp. sex lives	0.041	**0.006**	0.003	0.053	**0.002**	**0.002**
Health related behavior						
Exercises	0.066***	0.025***	**0.03*** **	0.086***	0.01***	**0.019*** **
Ever smoked	-0.067***	**-0.005**	-0.002	-0.086***	**-0.002**	-0.001
Smokes now	-0.093***	-0.053***	**-0.056*** **	-0.121***	-0.021***	**-0.034*** **
Ever drinks	0.045***	**0.021*** **	0.013*	0.059***	**0.008** **	**0.008* **
# of days drinks	0.005**	0.002	**0.003**	0.007**	0.001	**0.002**
# of drinks	-0.01**	**-0.009** **	-0.007**	-0.013**	-0.003**	**-0.004* **
Subjective health in 2000/2002						
Excel./excel.	ref.	ref.	ref.	ref.	ref.	ref.
Excel./good	-0.095***	-0.063***	**-0.071*** **	-0.123***	-0.025***	**-0.044*** **
Excel./poor	-0.177***	-0.118***	**-0.137*** **	-0.23***	-0.046***	**-0.084*** **
Good/excel.	-0.05***	**-0.05*** **	-0.055***	-0.064***	-0.02***	**-0.034*** **
Good/good	-0.095***	**-0.097*** **	-0.113***	-0.124***	-0.038***	**-0.069*** **
Good/poor	-0.192***	-0.17***	**-0.173*** **	-0.248***	-0.067***	**-0.106*** **
Poor/excel.	-0.124***	-0.1***	**-0.108*** **	-0.161***	-0.04***	**-0.066*** **
Poor/good	-0.142***	-0.139***	**-0.146*** **	-0.183***	-0.055***	**-0.09*** **
Poor/poor	-0.209***	**-0.237*** **	-0.263***	-0.27***	-0.093***	**-0.162*** **

*, ** and *** denote significance at 10, 5 and 1 percent level; Bold partial effects at the Mean and the MRH models indicate that they are closer to the partial effects of actual survival.

Table 3.5: Predictors of belief precision (n) in three versions of the MRH models of subjective survival expectations

	$\ln n$)		
	Model 1	Model 2	Model 3
Demographics			
Education	0.056***	0.051***	0.053***
Female	-0.139***	-0.156***	-0.157***
Black	0.053	0.041	0.02
Hispanic	0.188*	0.155	0.132
Cognition and personality, standardized scores			
Cognition		0.021	0.027
Vocabulary		0.008	0.009
Depression		0.067***	0.043*
Parents' age at death, linear splines with cut-off at own age			
Same sex, sp1		-0.493*	-0.493*
Same sex, sp2		0.467	0.5
Same sex lives		-0.064	-0.071
Opp. sex, sp1		-0.122	-0.117
Opp. sex, sp2		0.528	0.522
Opp. sex lives		0.026	0.032
Health related behavior			
Exercises		0.015	0.02
Ever smoked		-0.091*	-0.091*
Smokes now		0.007	0.006
Ever drinks		0.174***	0.177***
# of days drinks		-0.014	-0.015
# of drinks		-0.035	-0.033
Subjective health in 2000/2002			
Excel./excel.			ref.
Excel./good			-0.172**
Excel./poor			0.122
Good/excel.			-0.092
Good/good			-0.05
Good/poor			-0.059
Poor/excel.			-0.106
Poor/good			-0.042
Poor/poor			0.177**
Age and time horizon of the HRS question; variables divided by 100			
Age	-2.513*	-2.26	-2.333
Horizon	-7.503	-9.823	-10.037*
Age X Horizon	25.438**	27.698***	27.82***
Constant	1.572*	1.845*	1.882**

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.6: Quantiles of belief precision (n) and probabilities of surviving from age 50 to age 80

Quantiles	n	γ_0	$S(50, 80)$
10	0.40	0.000022	0.87
25	1.58	0.000054	0.71
50	7.23	0.000120	0.47
75	33.12	0.000229	0.23
90	130.31	0.000370	0.10

APPENDIX H
Derivations

Effect of individual survival heterogeneity on sample selection

In our model there is individual heterogeneity in survival chances. One consequence of this is that fit individuals will be overrepresented in the sample over time as frail individuals are more likely to die. This sample selection can be conveniently modeled in our framework. Let k^a and θ^a denote the shape and scale parameters in cohort a. As Vaupel (1979) shows the following is true for any cohorts

$$k^a = k^r, \tag{3.18}$$

$$\theta^a = \frac{1}{\frac{1}{\theta^r} + \left(\exp\left(\gamma_1 a\right) - \exp\left(\gamma_1 r\right)\right)}, \tag{3.19}$$

where r is a reference cohort. As we can see, older and younger cohorts share the same shape parameter k^a, but older cohorts have lower scale parameter θ^a (lower frailty) than younger cohorts.

Proof of equation (3.11)

The density function of the gamma distribution is

$$f\left(x\right) = \frac{1}{\Upsilon\left(k\right)\theta^k} x^{k-1} \exp\left(-\frac{x}{\theta}\right) = c\left(k, \theta\right) x^{k-1} \exp\left(-\frac{x}{\theta}\right). \tag{3.20}$$

It is well known that expected value of the gamma function is

$$E\left(x\right)=\int_{0}^{\infty}xc\left(k,\theta\right)x^{k-1}\exp\left(-\frac{x}{\theta}\right)dx=k\theta. \tag{3.21}$$

The expected value of a scaled gamma function is also gamma and its expected value is

$$E\left(cx\right)=ck\theta. \tag{3.22}$$

The expected value of the negative exponentiated gamma function is

$$\begin{aligned}E\left(\exp\left(-x\right)\right) &= \int_{0}^{\infty}\exp\left(-x\right)c\left(k,\theta\right)x^{k-1}\exp\left(-\frac{x}{\theta}\right)dx & (3.23)\\ &= \int_{0}^{\infty}c\left(k,\theta\right)x^{k-1}\exp\left(-\frac{x}{\theta}-x\right)dx=\int_{0}^{\infty}c\left(k,\theta\right)x^{k-1}\exp\left(-\frac{x}{\frac{\theta}{\theta+1}}\right)dx=... & (3.24)\end{aligned}$$

Note that this is very similar to the expected value formula of the gamma function, thus

$$...=\frac{\left(k-1\right)\theta}{\theta+1}\frac{c\left(k,\theta\right)}{c\left(k+1,\frac{\theta}{\theta+1}\right)}=\frac{\left(k-1\right)\theta}{\theta+1}\frac{\Upsilon\left(k-1\right)\left(\frac{\theta}{\theta+1}\right)^{k}}{\Upsilon\left(k\right)\theta^{k}}=... \tag{3.25}$$

Note that $\Gamma(k)=(k-1)\Gamma\left(k-1\right)$and thus

$$...=\frac{k-1}{k-1}\frac{\theta^{k}}{\theta^{k}}\left(1+\theta\right)^{k}=\left(1+\theta\right)^{k}. \tag{3.26}$$

Thus if $\gamma_{0i}\sim\Upsilon\left(k,\theta\right)$ then

$$E\left(S_{i}\left(a,T\right)\right)=E\left(\exp\left(-\gamma_{0i}\left(\exp\left(\gamma_{1}T\right)-\exp\left(\gamma_{1}a\right)\right)\right)\right)=\left(1+\theta\left(\exp\left(\gamma_{1}T\right)-\exp\left(\gamma_{1}a\right)\right)\right)^{-k}. \tag{3.27}$$

Details about the use of the Delta method to derive average partial effects

The goal is to derive point estimates and standard errors of the partial effects of any covariate on the survival probability from age to age

$$APE_j\left(a_1,a_2\right)=E_x\left[\frac{\partial S_i\left(a_1,a_2\right)}{\partial x_j}\right]. \tag{3.28}$$

The point estimates are

$$APE_j\left(a,T\right)=E_x\left[\frac{\partial\left(1+\theta\left(\exp\left(\gamma_1 T\right)-\exp\left(\gamma_1 a\right)\right)\right)^{-k}}{\partial x_j}\right]=... \tag{3.29}$$

Let us denote $e(a',a)=\exp\left(\gamma_1 a'\right)-\exp\left(\gamma_1 a\right)$. Then

$$\begin{aligned}
... &= E_x\left[\frac{\partial\left(1+\frac{e(T,a)}{\exp(-\beta' x_i)+e(a,r)}\right)^{-k}}{\partial x_j}\right] && (3.30)\\
&= E_x\left[-k\left(1+\frac{e(T,a)}{\exp\left(-\beta' x_i\right)+e(a,r)}\right)^{-k-1}\frac{e(T,a)\exp\left(-\beta' x_i\right)\beta_j}{\left(\exp\left(-\beta' x_i\right)+e(a,r)\right)^2}\right] && (3.31)\\
&= E_x\left[-k\left(\exp\left(-\beta' x_i\right)+e(T,r)\right)^{-k-1}\left(\exp\left(-\beta' x_i\right)+e(a,r)\right)^{k-1}e(T,a)\exp\left(-\beta' x_i\right)\beta_j\right] && (3.32)
\end{aligned}$$

By substituting the estimated coefficients into this formula we have a point estimate for the average partial effect of x_j on the survival probability from age a to age T. The standard errors can be computed with the delta method. For any differentiable transformation $g(\beta)$ and variance-covariance matrix Σ, the variance covariance matrix of the estimator of is $\left(\nabla g\right)^T\Sigma\left(\nabla g\right)$. Thus, we only need to compute the first derivatives of g. Let us see them one-by-one.

$$\begin{aligned}
\frac{\partial APE_j(a,T)}{\partial \beta_l} &= -I(j=l)\, E_x\left(k\left(\exp(-\beta' x_i)+e(T,r)\right)^{-k-1}\left(\exp(-\beta' x_i)+e(a,r)\right)^{k-1} e(T,a)\exp(-\beta' x_i)\right) \\
&-E_x\left(k(k+1)\left(\exp(-\beta' x_i)+e(T,r)\right)^{-k-2}\left(\exp(-\beta' x_i)+e(a,r)\right)^{k-1} e(T,a)\exp(-2\beta' x_i)\,\beta_j x_{il}\right) \\
&+E_x\left(k(k-1)\left(\exp(-\beta' x_i)+e(T,r)\right)^{-k-1}\left(\exp(-\beta' x_i)+e(a,r)\right)^{k-2} e(T,a)\exp(-2\beta' x_i)\,\beta_j x_{il}\right) \\
&+E_x\left(k\left(\exp(-\beta' x_i)+e(T,r)\right)^{-k-1}\left(\exp(-\beta' x_i)+e(a,r)\right)^{k-1} e(T,a)\exp(-\beta' x_i)\,\beta_j x_{il}\right). \quad (3.33)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial APE_j(a,T)}{\partial k} &= -E_x\left[\left(\exp(-\beta' x_i)+e(T,r)\right)^{-k-1}\left(\exp(-\beta' x_i)+e(a,r)\right)^{k-1} e(T,a)\exp(-\beta' x_i)\,\beta_j\right] \\
&-E_x\left[k\left(\exp(-\beta' x_i)+e(T,r)\right)^{-k-1}\left(\exp(-\beta' x_i)+e(a,r)\right)^{k-1} e(T,a)\exp(-\beta' x_i)\,\beta_j\right]\times \\
&\times\left(-\ln\left(\exp(-\beta' x_i)+e(T,r)\right)\right)+E_x\left(\ln\left(\exp(-\beta' x_i)+e(T,r)\right)\right). \qquad (3.34)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial APE_j(a,T)}{\partial \gamma_1} &= \left[T\exp(\gamma_1 T)-r\exp(\gamma_1 r)\right] e(T,a)\exp(-\beta' x_i)\,\beta_j\times \\
&\times\left\{k(k+1)\left(\exp(-\beta' x_i)+e(T,r)\right)^{-k-2}\left(\exp(-\beta' x_i)+e(a,r)\right)^{k-1}\right. \\
&\left.-k(k-1)\left(\exp(-\beta' x_i)+e(T,r)\right)^{-k-1}\left(\exp(-\beta' x_i)+e(a,r)\right)^{k-2}\right\}. \qquad (3.35)
\end{aligned}$$

By substituting the estimated coefficients into these formulas we have an estimator for the variance covariance matrix of the average partial effects.

The likelihood function of the mean model

The likelihood function can be written as

$$l_i = \Pr\left(\underline{p_i} \leq S_i(a,t) \leq \overline{p_i}\right) \qquad (3.36)$$

where $\underline{p_i}$ and $\overline{p_i}$ denote the lower and upper bound probabilities that would be rounded to the survey response. For example, if the rounding function rounds to the closest 1 percent and the survey response is 27 percent, then $\underline{p_i} = 0.265$ and $\overline{p_i} = 0.275$. If the rounding function rounds to the closest 5 percent, then the corresponding probabilities would be $\underline{p_i} = 0.225$ and $\overline{p_i} = 0.275$.[53] The likelihood function, thus, is

$$l_i = \Pr\left(\frac{-\ln \underline{p_i}}{\exp\left(\gamma_1 t\right) - \exp\left(\gamma_1 a\right)} \geq \gamma_{0i} \geq \frac{-\ln \overline{p_i}}{\exp\left(\gamma_1 t\right) - \exp\left(\gamma_1 a\right)}\right), \tag{3.37}$$

which can be easily computed from the c.d.f. of the gamma distribution in with parameters k and θ.

The likelihood function of the modal response model

The estimation of the MRH can be carried out by maximum simulated likelihood (MSL). MSL computes the likelihood function by drawing many values from the distribution of one (or several) random variables and computing the average conditional likelihood, conditioning on those values. In our case, it is worth simulating values from the distribution of u_{ni}. A standard version of the simulated likelihood would look like the following.[54]

$$l_i = \frac{1}{S}\sum_{s=1}^{S} l_i\left(u_{ni}^s\right). \tag{3.38}$$

[53] As we can see, a response that is not a multiple of 5 percent cannot be a rounded version of a true latent probability when we round to the closest 5 percent. The rounding model we have in mind is one where agents only identify bins (e.g. 0%-2.5%, 2.5%-7.5%, ..., 97.5%-100%) and any response in a particular bin only tells the econometrician that the true latent probability is also in the same bin. An alternative model would be that of Manski and Molinari (2010) where they identify individuals' rounding practices across many probability questions and use different rounding functions for each individual.

[54] Note that in case the simulated values are drawn from the distribution of u_{ni}, we do not need to weight the terms in (3.36) by the density of the draw as the simulation itself already weights the data.

where S is the number of simulation draws. The problem with this approach is that our model is full of discontinuities that make this approach infeasible. Let us rewrite the MRH formula in Section 3 with μ_i and n_i, and let us denote $\underline{n_i} = \frac{1}{n_i}$ and $\overline{n_i} = 1 - \frac{1}{n_i}$. The MRH model assumes that the survey response is

$$p_i^{mrh} = \begin{cases} round\left(\frac{\mu_i n_i - 1}{n_i - 2}\right) & \text{if } \overline{n_i} > \mu_i > \underline{n_i} \\ 1 & \text{if } \mu_i > \underline{n_i}, \overline{n_i} \\ 0 & \text{if } \mu_i < \underline{n_i}, \overline{n_i} \\ 0.5 & \text{if } \overline{n_i} < \mu_i < \underline{n_i}. \end{cases} \tag{3.39}$$

As we can see, whenever $n_i \leq 1$ an answer can only be 50 percent. Whenever $n_i \leq 2$, an answer can only be 0, 50 or 100 percent. These discontinuities make the simulation model in (3.38) hard for the following reason. Imagine that during the maximization of the likelihood function we get into a region where the precision of beliefs n_i is always below 2 for each simulation draw. This could happen if $z_i'\beta_n^k \ll 2$ and $\sigma_n^{2,k}$ is small, where k indexes the actual guesses for the parameters. In this case, the likelihood function would be undefined and the numerical maximization would fail. As a remedy, we recommend drawing separate simulation values from each region of n_i; this assures that the likelihood is well-defined in each iteration of the maximization.[55] The likelihood function, thus, is written

$$\begin{aligned} l_i &= \frac{1}{S_1}\sum_{s=1}^{S_1} l_i\left(u_{ni}^s|n_i \le 1\right)\Pr\left(n_i \le 1\right) + \frac{1}{S_1}\sum_{s=1}^{S_1} l_i\left(u_{ni}^s|1 < n_i \le 2\right)\Pr\left(1 < n_i \le 2\right) \\ &\quad + \frac{1}{S_1}\sum_{s=1}^{S_1} l_i\left(u_{ni}^s|n_i > 2\right)\Pr\left(n_i > 2\right). \end{aligned} \tag{3.40}$$

Now, whenever $z_i'\beta_n^k \ll 2$ and $\sigma_n^{2,k}$ is small in a particular iteration for a non-focal answer, the likelihood is a well-defined, extremely small number. The probabilities of the different regions of n_i are trivial since n_i is assumed to have a log-normal distribution. Whenever $n_i \le 1$ an answer can only be fifty and thus the conditional likelihood is

$$l_i\left(u_{ni}^s|n_i \le 1\right) = \begin{cases} 1 & \text{if } p_i^{hrs} = 50 \\ 0 & \text{otherwise,} \end{cases} \tag{3.41}$$

where p_i^{hrs} represents the survey response to the HRS mortality question. When $1 < n_i \le 2$, it is assured that $\underline{n_i} \ge \overline{n_i}$ and thus

$$l_i\left(u_{ni}^s|1 < n_i \le 2\right) = \begin{cases} \Pr\left(\overline{n_i} < S_i\left(a,t\right) < \underline{n_i}\right) & \text{if } p_i^{hrs} = 50 \\ \Pr\left(S_i\left(a,t\right) \le \overline{n_i}\right) & \text{if } p_i^{hrs} = 0 \\ \Pr\left(S_i\left(a,t\right) \ge \underline{n_i}\right) & \text{if } p_i^{hrs} = 100 \\ 0 & \text{otherwise.} \end{cases} \tag{3.42}$$

These probabilities can be computed analogously to the mean model derived above. The

[55] Hill et al. (2004) did not use this trick when they estimated a very similar model. The consequence was that they had to make restrictions on their model to be able to carry out the numerical estimation. It turned out that our model is identified and estimable in practice under mild conditions once these discontinuities are properly taken care of.

most complicated, although still very straightforward, case is the conditional likelihood in the region where $n_i > 2$. In this region $\underline{n_i} < \overline{n_i}$. The only complication is that a 0 and a 100 answer can now be either a focal answer or an exact rounded answer. 50 answers in this region cannot be focal answers as $n_i > 2$ and thus either α_i or β_i is larger than one. The conditional likelihood is

$$l_i\left(u_{ni}^s | n_i > 2\right) = \begin{cases} \Pr\left(S_i\left(a,t\right) \leq \underline{n_i} \text{ OR } \frac{S_i(a,t)n_i - 1}{n_i - 2} \leq \overline{p_i}\right) & \text{if } p_i^{hrs} = 0 \\ \Pr\left(S_i\left(a,t\right) \geq \overline{n_i} \text{ OR } \frac{S_i(a,t)n_i - 1}{n_i - 2} \geq \underline{p_i}\right) & \text{if } p_i^{hrs} = 100 \\ \Pr\left(\underline{p_i} \leq \frac{S_i(a,t)n_i - 1}{n_i - 2} \leq \overline{p_i}\right) & \text{otherwise.} \end{cases} \tag{3.43}$$

After straightforward algebra the conditional likelihood becomes

$$l_i\left(u_{ni}^s | n_i > 2\right) = \begin{cases} \Pr\left(S_i\left(a,t\right) \leq \max\left\{\underline{n_i}, \frac{\overline{p_i}(n_i - 2) + 1}{n_i}\right\}\right) & \text{if } p_i^{hrs} = 0 \\ \Pr\left(S_i\left(a,t\right) \geq \min\left\{\overline{n_i}, \frac{\underline{p_i}(n_i - 2) + 1}{n_i}\right\}\right) & \text{if } p_i^{hrs} = 100 \\ \Pr\left(\frac{\underline{p_i}(n_i - 2) + 1}{n_i} \leq S_i\left(a,t\right) \leq \frac{\overline{p_i}(n_i - 2) + 1}{n_i}\right) & \text{otherwise.} \end{cases} \tag{3.44}$$

and again these probabilities can be computed analogously to the mean model derived above.

APPENDIX I
Additional Tables

Table 3.7: Outputs of the ML estimation models with demographic variables; actual survival and the mean and MRH models of subjective survival expectations

	Actual survival $\ln \gamma_0$	Mean $\ln \theta^{0.5}$	MRH $\ln \theta^{0.5}$	$\ln n$
Demographics				
Education	-0.048	-0.13	-0.089	0.056
	[0.006]***	[0.007]***	[0.005]***	[0.009]***
Female	-0.343	-0.193	-0.209	-0.139
	[0.033]***	[0.037]***	[0.027]***	[0.051]***
Black	0.037	-0.007	-0.151	0.053
	[0.055]	[0.053]	[0.038]***	[0.076]
Hispanic	-0.157	0.329	0.198	0.188
	[0.079]**	[0.083]***	[0.062]***	[0.110]*
Age and time horizon of the HRS question; variables divided by 100				
Age				-2.513
				[1.478]*
Horizon				-7.503
				[6.314]
Age X Horizon				25.438
				[10.693]**
Constant	-7.538	-9.358	-7.234	1.572
	[0.223]***	[0.190]***	[0.181]***	[0.946]*
Other parameters				
$\ln \gamma_1$	2.268	2.719	2.358	
	[0.024]***	[0.012]***	[0.020]***	
$\ln k$		-0.629	0.158	
		[0.011]***	[0.025]***	
$\ln \sigma_n$			0.808	
			[0.025]***	
N	13038	13038	13038	
Log-likelihood		-57714.3	-46739.4	

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.8: Average partial effects of surviving 8 more years in three models with demographic variables: actual survival and the mean and MRH models of subjective survival expectations

	Actual survival $\ln \gamma_0$	Mean $\ln \theta^{0.5}$	MRH $\ln \theta^{0.5}$	$\ln n$
Demographics				
Education	0.008	0.011	0.012	0.056
	[0.001]***	[0.002]***	[0.002]***	[0.009]***
Female	0.061	0.017	0.027	-0.139
	[0.006]***	[0.004]***	[0.006]***	[0.051]***
Black	-0.007	0.001	0.02	0.053
	[0.010]	[0.005]	[0.006]***	[0.076]
Hispanic	0.028	-0.028	-0.026	0.188
	[0.014]**	[0.008]***	[0.009]***	[0.110]*

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.9: Average partial effects of surviving from age 55 to 75 in three models with demographic variables: actual survival and the mean and MRH models of subjective survival expectations

	Actual survival $\ln \gamma_0$	Mean $\ln \theta^{0.5}$	MRH $\ln \theta^{0.5}$	$\ln n$
Demographics				
Education	0.01	0.023	0.021	0.056
	[0.001]***	[0.004]***	[0.004]***	[0.009]***
Female	0.074	0.035	0.048	-0.139
	[0.008]***	[0.008]***	[0.010]***	[0.051]***
Black	-0.008	0.001	0.035	0.053
	[0.012]	[0.009]	[0.011]***	[0.076]
Hispanic	0.034	-0.059	-0.046	0.188
	[0.017]**	[0.017]***	[0.016]***	[0.110]*

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.10: Average partial effects of surviving from age 75 to 95 in three models with demographic variables: actual survival and the mean and MRH models of subjective survival expectations

	Actual survival $\ln \gamma_0$	Mean $\ln \theta^{0.5}$	MRH $\ln \theta^{0.5}$	$\ln n$
Demographics				
Education	0.013	0.007	0.011	0.056
	[0.002]***	[0.001]***	[0.002]***	[0.009]***
Female	0.093	0.011	0.025	-0.139
	[0.010]***	[0.003]***	[0.007]***	[0.051]***
Black	-0.01	0	0.018	0.053
	[0.015]	[0.003]	[0.006]***	[0.076]
Hispanic	0.042	-0.019	-0.024	0.188
	[0.022]**	[0.006]***	[0.009]***	[0.110]*

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.11: Outputs of the ML estimation models with demographic, personality and personal information variables; actual survival and the mean and MRH models of subjective survival expectations

	Actual survival $\ln \gamma_0$	Mean $\ln \theta^{0.5}$	MRH $\ln \theta^{0.5}$	$\ln n$
Demographics				
Education	0.015	-0.053	-0.037	0.051
	[0.006]**	[0.008]***	[0.005]***	[0.010]***
Female	-0.34	-0.198	-0.215	-0.156
	[0.037]***	[0.039]***	[0.029]***	[0.055]***
Black	-0.201	-0.177	-0.297	0.041
	[0.055]***	[0.053]***	[0.038]***	[0.076]
Hispanic	-0.254	0.297	0.171	0.155
	[0.078]***	[0.081]***	[0.060]***	[0.106]
Cognition and personality, standardized scores				
Cognition	-0.254	-0.115	-0.079	0.021
	[0.020]***	[0.021]***	[0.015]***	[0.029]
Vocabulary	-0.013	-0.017	0.013	0.008
	[0.019]	[0.021]	[0.015]	[0.028]
Depression	0.142	0.303	0.253	0.067
	[0.015]***	[0.020]***	[0.015]***	[0.025]***
Parents' age at death, linear splines with cut-off at own age				
Same sex, sp1	-0.121	-0.386	0.205	-0.493
	[0.153]	[0.292]	[0.184]	[0.291]*
Same sex, sp2	-0.733	-1.784	-1.766	0.467
	[0.279]***	[0.214]***	[0.161]***	[0.342]
Same sex lives	-0.049	-0.003	-0.025	-0.064
	[0.119]	[0.059]	[0.044]	[0.100]
Opp. sex, sp1	-0.177	-1.058	-0.664	-0.122
	[0.156]	[0.297]***	[0.190]***	[0.292]
Opp. sex, sp2	-0.667	-0.553	-0.494	0.528
	[0.282]**	[0.210]***	[0.155]***	[0.340]
Opp. sex lives	-0.131	-0.037	-0.031	0.026
	[0.120]	[0.061]	[0.045]	[0.107]
Health related behavior				
Exercises	-0.445	-0.309	-0.258	0.015
	[0.038]***	[0.035]***	[0.025]***	[0.051]
Ever smoked	0.352	0.113	0.063	-0.091
	[0.038]***	[0.038]***	[0.027]**	[0.054]*
Smokes now	0.447	0.337	0.257	0.007
	[0.050]***	[0.053]***	[0.039]***	[0.076]
Ever drinks	-0.259	-0.25	-0.141	0.174
	[0.044]***	[0.043]***	[0.031]***	[0.062]***
# of days drinks	-0.033	-0.021	-0.019	-0.014
	[0.011]***	[0.011]*	[0.008]**	[0.016]
# of drinks	0.056	0.049	0.033	-0.035
	[0.019]***	[0.019]**	[0.014]**	[0.028]

Table continues on the next page...

...Table 3.11 continued

	Actual survival $\ln \gamma_0$	Mean $\ln \theta^{0.5}$	MRH $\ln \theta^{0.5}$	$\ln n$
Age and time horizon of the HRS question; variables divided by 100				
Age				-2.26
				[1.471]
Horizon				-9.823
				[6.149]
Age X Horizon				27.698
				[10.367]***
Constant	-6.712	-8.595	-5.796	1.845
	[0.288]***	[0.239]***	[0.202]***	[0.946]*
Other parameters				
$\ln \gamma_1$	2.104	2.684	2.179	
	[0.041]***	[0.014]***	[0.029]***	
$\ln k$		-0.573	0.296	
		[0.011]***	[0.026]***	
$\ln \sigma_n$			0.773	
			[0.026]***	
N	13038	13038	13038	
Log-likelihood		-57265.8	-46196.4	

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.12: Average partial effects of surviving 8 more years in three models with demographic, personality and personal information variables: actual survival and the mean and MRH models of subjective survival expectations

	Actual survival	Mean	MRH
Demographics			
Education	-0.003	0.005	0.005
	[0.001]**	[0.001]***	[0.001]***
Female	0.058	0.017	0.029
	[0.006]***	[0.004]***	[0.007]***
Black	0.034	0.015	0.039
	[0.009]***	[0.005]***	[0.010]***
Hispanic	0.043	-0.025	-0.023
	[0.013]***	[0.008]***	[0.009]**
Cognition and personality, standardized scores			
Cognition	0.043	0.01	0.011
	[0.003]***	[0.002]***	[0.003]***
Vocabulary	0.002	0.001	-0.002
	[0.003]	[0.002]	[0.002]
Depression	-0.024	-0.026	-0.034
	[0.003]***	[0.004]***	[0.007]***
Parents' age at death, linear splines with cut-off at own age			
Same sex, sp1	0.02	0.033	-0.027
	[0.026]	[0.024]	[0.027]
Same sex, sp2	0.124	0.152	0.234
	[0.047]***	[0.032]***	[0.059]***
Same sex lives	0.008	0	0.003
	[0.020]	[0.005]	[0.006]
Opp. sex, sp1	0.03	0.09	0.088
	[0.026]	[0.026]***	[0.026]***
Opp. sex, sp2	0.113	0.047	0.065
	[0.048]**	[0.020]**	[0.028]**
Opp. sex lives	0.022	0.003	0.004
	[0.020]	[0.005]	[0.006]
Health related behavior			
Exercises	0.075	0.026	0.034
	[0.007]***	[0.005]***	[0.008]***
Ever smoked	-0.06	-0.01	-0.008
	[0.007]***	[0.004]***	[0.004]**
Smokes now	-0.076	-0.029	-0.034
	[0.009]***	[0.006]***	[0.009]***
Ever drinks	0.044	0.021	0.019
	[0.008]***	[0.005]***	[0.006]***
# of days drinks	0.006	0.002	0.003
	[0.002]***	[0.001]*	[0.001]**
# of drinks	-0.009	-0.004	-0.004
	[0.003]***	[0.002]**	[0.002]**

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.13: Average partial effects of surviving from age 55 to 75 in three models with demographic, personality and personal information variables: actual survival and the mean and MRH models of subjective survival expectations

	Actual survival	Mean	MRH
Demographics			
Education	-0.003	0.009	0.009
	[0.001]**	[0.002]***	[0.002]***
Female	0.073	0.035	0.051
	[0.009]***	[0.009]***	[0.012]***
Black	0.043	0.031	0.07
	[0.012]***	[0.011]***	[0.017]***
Hispanic	0.055	-0.052	-0.04
	[0.017]***	[0.016]***	[0.016]**
Cognition and personality, standardized scores			
Cognition	0.055	0.02	0.019
	[0.005]***	[0.005]***	[0.005]***
Vocabulary	0.003	0.003	-0.003
	[0.004]	[0.004]	[0.004]
Depression	-0.031	-0.053	-0.06
	[0.004]***	[0.009]***	[0.013]***
Parents' age at death, linear splines with cut-off at own age			
Same sex, sp1	0.026	0.068	-0.048
	[0.033]	[0.050]	[0.047]
Same sex, sp2	0.159	0.314	0.415
	[0.060]***	[0.067]***	[0.100]***
Same sex lives	0.011	0.001	0.006
	[0.026]	[0.010]	[0.010]
Opp. sex, sp1	0.038	0.186	0.156
	[0.034]	[0.054]***	[0.046]***
Opp. sex, sp2	0.144	0.097	0.116
	[0.060]**	[0.042]**	[0.048]**
Opp. sex lives	0.028	0.007	0.007
	[0.026]	[0.011]	[0.011]
Health related behavior			
Exercises	0.096	0.054	0.061
	[0.009]***	[0.011]***	[0.014]***
Ever smoked	-0.076	-0.02	-0.015
	[0.009]***	[0.007]***	[0.007]**
Smokes now	-0.097	-0.059	-0.06
	[0.012]***	[0.013]***	[0.015]***
Ever drinks	0.056	0.044	0.033
	[0.010]***	[0.010]***	[0.010]***
# of days drinks	0.007	0.004	0.005
	[0.002]***	[0.002]*	[0.002]**
# of drinks	-0.012	-0.009	-0.008
	[0.004]***	[0.004]**	[0.004]**

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.14: Average partial effects of surviving from age 75 to 95 in three models with demographic, personality and personal information variables: actual survival and the mean and MRH models of subjective survival expectations

	Actual survival	Mean	MRH
Demographics			
Education	-0.004	0.003	0.005
	[0.002]**	[0.001]***	[0.002]***
Female	0.097	0.012	0.029
	[0.012]***	[0.003]***	[0.009]***
Black	0.058	0.011	0.04
	[0.016]***	[0.004]***	[0.012]***
Hispanic	0.073	-0.019	-0.023
	[0.023]***	[0.006]***	[0.010]**
Cognition and personality, standardized scores			
Cognition	0.073	0.007	0.011
	[0.008]***	[0.002]***	[0.004]***
Vocabulary	0.004	0.001	-0.002
	[0.005]	[0.001]	[0.002]
Depression	-0.041	-0.019	-0.034
	[0.005]***	[0.004]***	[0.009]***
Parents' age at death, linear splines with cut-off at own age			
Same sex, sp1	0.035	0.024	-0.028
	[0.043]	[0.018]	[0.028]
Same sex, sp2	0.21	0.112	0.239
	[0.086]**	[0.027]***	[0.071]***
Same sex lives	0.014	0	0.003
	[0.034]	[0.004]	[0.006]
Opp. sex, sp1	0.051	0.066	0.09
	[0.044]	[0.020]***	[0.029]***
Opp. sex, sp2	0.191	0.035	0.067
	[0.086]**	[0.016]**	[0.030]**
Opp. sex lives	0.038	0.002	0.004
	[0.035]	[0.004]	[0.006]
Health related behavior			
Exercises	0.128	0.019	0.035
	[0.013]***	[0.004]***	[0.010]***
Ever smoked	-0.101	-0.007	-0.008
	[0.012]***	[0.003]***	[0.004]**
Smokes now	-0.128	-0.021	-0.035
	[0.015]***	[0.005]***	[0.010]***
Ever drinks	0.074	0.016	0.019
	[0.013]***	[0.004]***	[0.006]***
# of days drinks	0.01	0.001	0.003
	[0.003]***	[0.001]*	[0.001]**
# of drinks	-0.016	-0.003	-0.004
	[0.006]***	[0.001]**	[0.002]**

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.15: Outputs of the ML estimation models with demographic, personality, personal information and subjective health variables; actual survival and the mean and MRH models of subjective survival expectations

	Actual survival	Mean	MRH	
	$\ln \gamma_0$	$\ln \theta^{0.5}$	$\ln \theta^{0.5}$	$\ln n$
Demographics				
Education	0.026	-0.037	-0.019	0.053
	[0.006]***	[0.007]***	[0.005]***	[0.010]***
Female	-0.31	-0.152	-0.169	-0.157
	[0.036]***	[0.038]***	[0.027]***	[0.054]***
Black	-0.24	-0.264	-0.369	0.02
	[0.053]***	[0.052]***	[0.036]***	[0.075]
Hispanic	-0.332	0.229	0.083	0.132
	[0.077]***	[0.078]***	[0.056]	[0.105]
Cognition and personality, standardized scores				
Cognition	-0.231	-0.052	-0.035	0.027
	[0.019]***	[0.020]**	[0.015]**	[0.028]
Vocabulary	-0.011	-0.018	0.011	0.009
	[0.018]	[0.020]	[0.014]	[0.028]
Depression	0.043	0.148	0.131	0.043
	[0.016]***	[0.020]***	[0.014]***	[0.026]*
Parents' age at death, linear splines with cut-off at own age				
Same sex, sp1	-0.09	-0.6	0.094	-0.493
	[0.154]	[0.281]**	[0.172]	[0.287]*
Same sex, sp2	-0.567	-1.552	-1.608	0.5
	[0.272]**	[0.207]***	[0.153]***	[0.337]
Same sex lives	-0.051	0.006	-0.007	-0.071
	[0.114]	[0.057]	[0.042]	[0.099]
Opp. sex, sp1	-0.203	-1.193	-0.663	-0.117
	[0.156]	[0.285]***	[0.177]***	[0.287]
Opp. sex, sp2	-0.434	-0.353	-0.427	0.522
	[0.274]	[0.205]*	[0.148]***	[0.336]
Opp. sex lives	-0.191	-0.032	-0.014	0.032
	[0.116]*	[0.059]	[0.043]	[0.105]
Health related behavior				
Exercises	-0.309	-0.146	-0.13	0.02
	[0.038]***	[0.034]***	[0.024]***	[0.051]
Ever smoked	0.31	0.031	0.009	-0.091
	[0.038]***	[0.036]	[0.026]	[0.053]*
Smokes now	0.436	0.308	0.239	0.006
	[0.049]***	[0.051]***	[0.037]***	[0.075]
Ever drinks	-0.211	-0.119	-0.054	0.177
	[0.043]***	[0.042]***	[0.030]*	[0.061]***
# of days drinks	-0.024	-0.012	-0.011	-0.015
	[0.011]**	[0.011]	[0.008]	[0.016]
# of drinks	0.045	0.05	0.029	-0.033
	[0.019]**	[0.019]***	[0.014]**	[0.028]

Table continues on the next page...

...Table 3.15 continued

	Actual survival $\ln \gamma_0$	Mean $\ln \theta^{0.5}$	MRH $\ln \theta^{0.5}$	$\ln n$
Parents' age at death, linear splines with cut-off at own age				
Excel./excel.	ref.	ref.	ref.	ref.
Excel./good	0.442	0.363	0.306	-0.172
	[0.065]***	[0.055]***	[0.040]***	[0.084]**
Excel./poor	0.827	0.682	0.585	0.122
	[0.091]***	[0.121]***	[0.085]***	[0.157]
Good/excel.	0.231	0.29	0.237	-0.092
	[0.079]***	[0.061]***	[0.044]***	[0.095]
Good/good	0.445	0.562	0.483	-0.05
	[0.059]***	[0.050]***	[0.035]***	[0.074]
Good/poor	0.894	0.983	0.74	-0.059
	[0.065]***	[0.087]***	[0.061]***	[0.104]
Poor/excel.	0.579	0.582	0.461	-0.106
	[0.119]***	[0.156]***	[0.106]***	[0.196]
Poor/good	0.66	0.805	0.624	-0.042
	[0.075]***	[0.088]***	[0.063]***	[0.116]
Poor/poor	0.973	1.372	1.126	0.177
	[0.057]***	[0.070]***	[0.052]***	[0.085]**
Age and time horizon of the HRS question; variables divided by 100				
Age				-2.333
				[1.449]
Horizon				-10.037
				[6.053]*
Age X Horizon				27.82
				[10.204]***
Constant	-7.288	-9.366	-6.379	1.882
	[0.293]***	[0.233]***	[0.197]***	[0.933]**
Other parameters				
$\ln \gamma_1$	2.088	2.686	2.149	
	[0.041]***	[0.013]***	[0.029]***	
$\ln k$		-0.54	0.377	
		[0.011]***	[0.026]***	
$\ln \sigma_n$			0.758	
			[0.026]***	
N	13038	13038	13038	
Log-likelihood		-56994.3	-45850.0	

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.16: Average partial effects of surviving 8 more years in three models with demographic, personality, personal information and subjective health variables: actual survival and the mean and MRH models of subjective survival expectations

	Actual survival	Mean	MRH
Demographics			
Education	-0.004	0.003	0.003
	[0.001]***	[0.001]***	[0.001]***
Female	0.052	0.013	0.023
	[0.006]***	[0.004]***	[0.006]***
Black	0.04	0.023	0.05
	[0.009]***	[0.006]***	[0.012]***
Hispanic	0.056	-0.02	-0.011
	[0.013]***	[0.007]***	[0.008]
Cognition and personality, standardized scores			
Cognition	0.039	0.004	0.005
	[0.003]***	[0.002]**	[0.002]**
Vocabulary	0.002	0.002	-0.002
	[0.003]	[0.002]	[0.002]
Depression	-0.007	-0.013	-0.018
	[0.003]***	[0.003]***	[0.004]***
Parents' age at death, linear splines with cut-off at own age			
Same sex, sp1	0.015	0.052	-0.013
	[0.026]	[0.024]**	[0.024]
Same sex, sp2	0.095	0.134	0.217
	[0.046]**	[0.029]***	[0.055]***
Same sex lives	0.009	-0.001	0.001
	[0.019]	[0.005]	[0.006]
Opp. sex, sp1	0.034	0.103	0.09
	[0.026]	[0.027]***	[0.025]***
Opp. sex, sp2	0.073	0.03	0.058
	[0.046]	[0.019]	[0.026]**
Opp. sex lives	0.032	0.003	0.002
	[0.020]*	[0.005]	[0.006]
Health related behavior			
Exercises	0.052	0.013	0.018
	[0.006]***	[0.004]***	[0.005]***
Ever smoked	-0.052	-0.003	-0.001
	[0.006]***	[0.003]	[0.004]
Smokes now	-0.073	-0.027	-0.032
	[0.008]***	[0.006]***	[0.008]***
Ever drinks	0.035	0.01	0.007
	[0.007]***	[0.004]***	[0.004]*
# of days drinks	0.004	0.001	0.001
	[0.002]**	[0.001]	[0.001]
# of drinks	-0.008	-0.004	-0.004
	[0.003]**	[0.002]**	[0.002]*

Table continues on the next page...

...Table 3.16 continued

	Actual survival	Mean	MRH
Parents' age at death, linear splines with cut-off at own age			
Excel./excel.	ref.	ref.	ref.
Excel./good	-0.074	-0.031	-0.041
	[0.011]***	[0.007]***	[0.010]***
Excel./poor	-0.139	-0.059	-0.079
	[0.016]***	[0.014]***	[0.020]***
Good/excel.	-0.039	-0.025	-0.032
	[0.013]***	[0.007]***	[0.009]***
Good/good	-0.075	-0.048	-0.065
	[0.010]***	[0.009]***	[0.015]***
Good/poor	-0.15	-0.085	-0.1
	[0.012]***	[0.015]***	[0.023]***
Poor/excel.	-0.097	-0.05	-0.062
	[0.020]***	[0.015]***	[0.020]***
Poor/good	-0.111	-0.069	-0.084
	[0.013]***	[0.013]***	[0.020]***
Poor/poor	-0.163	-0.118	-0.152
	[0.010]***	[0.020]***	[0.033]***

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.17: Average partial effects of surviving from age 55 to 75 in three models with demographic, personality, personal information and subjective health variables: actual survival and the mean and MRH models of subjective survival expectations

	Actual survival	Mean	MRH
Demographics			
Education	-0.006	0.006	0.004
	[0.001]***	[0.002]***	[0.001]***
Female	0.067	0.026	0.039
	[0.008]***	[0.008]***	[0.010]***
Black	0.051	0.046	0.086
	[0.012]***	[0.012]***	[0.020]***
Hispanic	0.071	-0.04	-0.019
	[0.017]***	[0.015]***	[0.013]
Cognition and personality, standardized scores			
Cognition	0.05	0.009	0.008
	[0.004]***	[0.004]**	[0.004]**
Vocabulary	0.002	0.003	-0.003
	[0.004]	[0.003]	[0.003]
Depression	-0.009	-0.026	-0.031
	[0.003]***	[0.005]***	[0.007]***
Parents' age at death, linear splines with cut-off at own age			
Same sex, sp1	0.019	0.104	-0.022
	[0.033]	[0.048]**	[0.042]
Same sex, sp2	0.122	0.268	0.375
	[0.058]**	[0.059]***	[0.090]***
Same sex lives	0.011	-0.001	0.002
	[0.024]	[0.010]	[0.010]
Opp. sex, sp1	0.043	0.206	0.155
	[0.034]	[0.053]***	[0.043]***
Opp. sex, sp2	0.093	0.061	0.1
	[0.058]	[0.038]	[0.044]**
Opp. sex lives	0.041	0.006	0.003
	[0.025]	[0.010]	[0.010]
Health related behavior			
Exercises	0.066	0.025	0.03
	[0.008]***	[0.007]***	[0.008]***
Ever smoked	-0.067	-0.005	-0.002
	[0.009]***	[0.006]	[0.006]
Smokes now	-0.093	-0.053	-0.056
	[0.012]***	[0.012]***	[0.014]***
Ever drinks	0.045	0.021	0.013
	[0.010]***	[0.008]***	[0.007]*
# of days drinks	0.005	0.002	0.003
	[0.002]**	[0.002]	[0.002]
# of drinks	-0.01	-0.009	-0.007
	[0.004]**	[0.004]**	[0.003]**

Table continues on the next page...

...Table 3.17 continued

	Actual survival	Mean	MRH
Parents' age at death, linear splines with cut-off at own age			
Excel./excel.	ref.	ref.	ref.
Excel./good	-0.095	-0.063	-0.071
	[0.015]***	[0.014]***	[0.017]***
Excel./poor	-0.177	-0.118	-0.137
	[0.021]***	[0.028]***	[0.034]***
Good/excel.	-0.05	-0.05	-0.055
	[0.017]***	[0.013]***	[0.015]***
Good/good	-0.095	-0.097	-0.113
	[0.013]***	[0.017]***	[0.024]***
Good/poor	-0.192	-0.17	-0.173
	[0.016]***	[0.031]***	[0.038]***
Poor/excel.	-0.124	-0.1	-0.108
	[0.026]***	[0.031]***	[0.033]***
Poor/good	-0.142	-0.139	-0.146
	[0.017]***	[0.027]***	[0.033]***
Poor/poor	-0.209	-0.237	-0.263
	[0.015]***	[0.039]***	[0.054]***

*, ** and *** denote significance at 10, 5 and 1 percent level

Table 3.18: Average partial effects of surviving from age 75 to 95 in three models with demographic, personality, personal information and subjective health variables: actual survival and the mean and MRH models of subjective survival expectations

	Actual survival	Mean	MRH
Demographics			
Education	-0.007	0.003	0.003
	[0.002]***	[0.001]***	[0.001]***
Female	0.086	0.01	0.024
	[0.011]***	[0.003]***	[0.007]***
Black	0.067	0.018	0.053
	[0.015]***	[0.005]***	[0.015]***
Hispanic	0.092	-0.016	-0.012
	[0.022]***	[0.006]**	[0.009]
Cognition and personality, standardized scores			
Cognition	0.064	0.004	0.005
	[0.007]***	[0.002]**	[0.003]*
Vocabulary	0.003	0.001	-0.002
	[0.005]	[0.001]	[0.002]
Depression	-0.012	-0.01	-0.019
	[0.004]***	[0.002]***	[0.005]***
Parents' age at death, linear splines with cut-off at own age			
Same sex, sp1	0.025	0.041	-0.014
	[0.042]	[0.019]**	[0.026]
Same sex, sp2	0.157	0.105	0.231
	[0.080]**	[0.026]***	[0.068]***
Same sex lives	0.014	0	0.001
	[0.032]	[0.004]	[0.006]
Opp. sex, sp1	0.056	0.081	0.095
	[0.042]	[0.022]***	[0.029]***
Opp. sex, sp2	0.12	0.024	0.061
	[0.079]	[0.015]	[0.029]**
Opp. sex lives	0.053	0.002	0.002
	[0.032]	[0.004]	[0.006]
Health related behavior			
Exercises	0.086	0.01	0.019
	[0.012]***	[0.003]***	[0.006]***
Ever smoked	-0.086	-0.002	-0.001
	[0.012]***	[0.003]	[0.004]
Smokes now	-0.121	-0.021	-0.034
	[0.015]***	[0.005]***	[0.010]***
Ever drinks	0.059	0.008	0.008
	[0.012]***	[0.003]**	[0.005]*
# of days drinks	0.007	0.001	0.002
	[0.003]**	[0.001]	[0.001]
# of drinks	-0.013	-0.003	-0.004
	[0.005]**	[0.001]**	[0.002]*

Table continues on the next page...

...Table 3.18 continued

	Actual survival	Mean	MRH
Parents' age at death, linear splines with cut-off at own age			
Excel./excel.	ref.	ref.	ref.
Excel./good	-0.123	-0.025	-0.044
	[0.020]***	[0.006]***	[0.013]***
Excel./poor	-0.23	-0.046	-0.084
	[0.029]***	[0.012]***	[0.025]***
Good/excel.	-0.064	-0.02	-0.034
	[0.022]***	[0.006]***	[0.011]***
Good/good	-0.124	-0.038	-0.069
	[0.018]***	[0.008]***	[0.019]***
Good/poor	-0.248	-0.067	-0.106
	[0.024]***	[0.014]***	[0.029]***
Poor/excel.	-0.161	-0.04	-0.066
	[0.035]***	[0.013]***	[0.023]***
Poor/good	-0.183	-0.055	-0.09
	[0.024]***	[0.012]***	[0.025]***
Poor/poor	-0.27	-0.093	-0.162
	[0.023]***	[0.019]***	[0.042]***

*, ** and *** denote significance at 10, 5 and 1 percent level

CHAPTER IV

Stock Market Crash and Expectations of American Households

4.1 Introduction

The stock market crash of 2008 and the subsequent financial crisis constitute a rare episode whose scope and implications fall outside the life experience of American households. Whether and how those events affect people's expectations is an important question. To the extent that expectations guide investment behavior, substantial changes in expectations due to the financial crash can lead to substantial changes in investment. Besides average beliefs of "the representative household," the crisis may have an impact on heterogeneity of such beliefs.

This study uses data from the 2008 wave of the Health and Retirement Study (HRS) to study the impact of the crisis on people's expectations. We estimate the effect of the crash on the population average of expected returns, the population average of the uncertainty about returns (subjective standard deviation), and the cross-sectional heterogeneity in expected returns (an indicator of disagreement). We show estimates from simple reduced-form regressions on probability answers as well as from a more structural model that focuses on the parameters of interest and separates survey noise from relevant heterogeneity. The measurement strategy makes use of the fact that the respondents of HRS-2008 answered the survey during twelve months from February 2008 to February 2009, a time period that includes the time of the stock market crash in early October. We show that the date of interview is largely independent of the respondents' past expectations about the stock market, so even if the date of interview is non-random, it is unlikely to bias our results.

Our analysis looks at changes in expectations during the HRS sampling period of February 2008 through February 2009. It may be useful to recall some of the important events during this period. The subprime mortgage crisis began well before 2008, but the Dow Jones peaked in October 2007 above 14,000. By early 2008, though, the Dow was down to 12,000, and the rest of the year was characterized by a general decline until the crash of October. March 2008 saw the failed bailout of Bear Sterns and its subsequent sale to JP Morgan, but the rest of the Spring and the Summer went relatively quietly. On September 15, Lehman Brother filed for bankruptcy. The financial system was thought to be in severe danger, and it took a few weeks of uncertainty and heated debates before the U.S. Congress passed the TARP bill on October 3rd. The Fall of 2008 also witnessed the run-up to the Presidential election on November 4th, which focused many people's attention towards economic issues, but it also led to a natural uncertainty about future economic policy.

Figure 4.1 shows time series of four stock market variables over the course of the HRS sampling period. We divided the sampling period into four sub-periods on the figure: February to June, July to September, October to November, and December to February 2009. We shall use these sub-periods throughout our analysis; their definition was based on the stock market time series we discuss below.

The first panel shows the level of the Dow Jones Industrial Average and the VXD annualized volatility index[56]. After initial ups and downs, the level of the index started a substantial but gradual decline in June that stopped in August. The stock market crash hit in early October with a 3000-point drop in the Dow. The stock market experienced large swings in October and November, and the Dow reached a six-year low of 7500 in late November. After some recovery and a brief period of stability, the Dow experienced another period of steady decline in the first months of 2009. During the entire period, volatility showed the mirror image of the time series in levels, except that its increase started in September, and it reached its maximum in October and November. The second panel of Figure 4.1 shows the weekly volume of trade of the shares of the DJIA together with the trend of searches for the term 'Dow' on Google[57]. The latter variable is an indicator for the attention people give to news about the stock market. The figure shows a strong co-movement of the two time series: increased attention to stock market news coincided with increased volumes in March and July of 2008, February of 2009, and, especially, October of 2008. The Google index is normalized so that its five-year average is one. The maximal 8.8 value in the first week of October means that almost nine times as many searches were made from the US for the Dow Jones Industrial Average than in normal times. Looking at the two panels together, we can see that the volume of trade was the highest at times when the stock market index was decreasing, when uncertainty was increasing and when people paid a lot of attention to news about the market.

The main question of this paper is whether and how expectations changed during the stock market crash in early October 2008 and the following months. We compare post-crash

[56]The VXD index is derived from prices of options on the DJIA, and it measures the future (30-day) expected volatility of investors. Details of this index can be found at http://www.cboe.com/micro/vxd/.

[57]http://www.google.com/trends.

expectations to those earlier in 2008. It is important to keep in mind that the baseline period was characterized by early signs of the crisis and a depressed stock market. Nevertheless, the comparison can shed light on the effect of a large and perhaps qualitatively different event compared to the more "normal" declining market.

The crash may affect the population average of expected returns for various reasons. If people are unsure about the parameters of the returns process, they may use recent realizations to update their beliefs. In such a case, the crash would have a negative effect on everyone's expectations. If, on the other hand, people believe in mean reversion in stock market prices, the effect may be of the opposite sign. Of course, people may not want to update their beliefs if they don't learn from the returns. Besides stock prices, the political and policy news may have also affected people's expectations about the future of the economy and the financial sector in general, and the stock market in particular.

Empirical papers about stock market expectations usually find that average expectations track recent changes in the level of the stock market. When the stock market is increasing, average beliefs become more optimistic and conversely. See, for example, Kézdi and Willis (2008) about American households and Hurd et al. (2011) about Dutch households. According to Kézdi and Willis (2008), it took a five hundred point gain in the Dow Jones to generate a one percentage point gain in expected yearly returns in 2002. With such a relationship, expected returns of respondents in November 2008 should be more than five percentage points lower than expected returns of respondents two or three months earlier. On the other hand, the financial crisis of 2008 may have affected people's expectations in qualitatively different ways than the more gradual changes witnessed in 2002, especially if people had different views about the condition of the economy in 2002 and in 2008. People may expect asset prices to change in different ways after large sudden changes than gradual trends. This is the conclusion of Calvet et al. (2009b) who, using Swedish data, found that people tend to invest in well-performing mutual funds but also tend to dispose of winning individual stocks at the same time.

The effect of the crisis on average uncertainty is more predictably positive. Stock market risk increased dramatically, as indicated by the trend in volatility on Figure 4.1. Even those who do not follow the stock market could become more uncertain about the future of the economy in general and the stock market, in particular, as general uncertainty has been "in the air" throughout the crisis.

The crisis may also affect the cross-sectional heterogeneity in households' beliefs. Heterogeneity and potential subjectivity of people's beliefs about future stock market returns has been the focus of recent developments in finance theory (see Hong and Stein, 2007 for an overview about disagreement models in finance). Harris and Raviv (1993) and Kandel and Pearson (1995) show that public announcements can increase disagreement about the fundamental value of assets if people interpret the news in different ways (see also Kondor, 2009). As Hong and Stein (2007) observe, this pattern is precisely the opposite of what one would expect based on a simple rational-expectations model with heterogeneous priors, where public information should have the effect of reducing disagreement, rather than increasing it. Similar mechanisms may increase disagreement after the stock market crash as well. Dominitz and Manski (forthcoming), for example, assume that the population is a mix of people who believe in the random walk hypothesis, who believe in the mean reversion of stock-prices and who believe in the persistence of trends on the financial markets. When the crash hit the economy and stock prices fell sharply, people holding these various views should have interpreted its implications in different ways, and consequently the disagreement among them should have increased. Indeed, a potential explanation of the trading pattern shown in Figure 4.1 is that the increase of disagreement created space for trade as more optimistic traders wanted to buy and more pessimistic traders wanted to sell. Note that potential heterogeneity in the effect of the crash implies that the average effect could go either way.

Our results imply a temporary increase in the population average of expectations right after the crash. At the same time, average uncertainty increased, perhaps as the result of increased stock market volatility. Our most robust finding is that cross-sectional heterogeneity in expected returns, an indicator of the amount of disagreement, increased substantially with the stock market crash. The effects are found to be largest among stockholders, those who follow the stock market and those with higher than average cognitive capacity. The result on average expectations thus masks a wide distribution of effects of opposing signs. We also document the co-movement of stock market expectations with ex-post returns, implied volatility and volume of trade.

Our finding suggests that there is heterogeneity in the cognitive processes (or mental models) people use to convert public news into personal probability beliefs, in accordance with some of the disagreement literature we mentioned above. The results on changes in heterogeneity complement recent empirical investigations that show substantial heterogeneity in stock market expectations of individual investors (Vissing-Jorgensen, 2004) as well as households (Calvet et al., 2007, 2009b,a; Dominitz and Manski, 2005; Kézdi and Willis, 2008; Hurd et al., 2011; Gouret and Hollard, 2011). This paper adds new results to this empirical literature by showing that the stock market crash and the financial crisis had significant effects on average expectations, average uncertainty, and perhaps most importantly, the heterogeneity of expectations.

4.2 Data

We use stock market expectations data from HRS-2008. Before turning to our analysis, it is helpful to provide some background on the evolution of the HRS stock market expectation questions.

In 2002, HRS introduced probabilistic expectations questions about returns in the stock market to the battery of subjective expectation questions that have been asked in HRS since it began in 1992. One motivation for adding these questions is that expectations about stock returns are a key component in determining retirement saving and portfolio choice. In addition, stock market expectations are of methodological interest because the history of stock returns and their daily realizations are public information, enabling researchers to investigate how news affects the updating of beliefs without the need to adjust for differences in private information.

Like other HRS probability questions, stock market expectations are asked as a percent chance based on a “0” to “100” scale where the respondent is told that: ”0” means that you think there is absolutely no chance, and ”100” means that you think the event is absolutely sure to happen.

The instruction goes on to say, For example, no one can ever be sure about tomorrow’s weather, but if you think that rain is very unlikely tomorrow, you might say that there is a 10 percent chance of rain. If you think there is a very good chance that it will rain tomorrow, you might say that there is an 80 percent chance of rain.

Beginning in 2002, the HRS introduced a question about stock market expectations that has been asked in every wave of HRS since 2002. We call this the P_0 question. It reads: We are interested in how well you think the economy will do in the next year. By next year at this time, what is the percent chance that mutual fund shares invested in blue chip stocks like those in the Dow Jones Industrial Average will be worth more than they are today?

Like other HRS subjective probability questions, many answers to the HRS stock market questions are heaped on "50" (Hurd and McGarry, 2002) and, unlike most other probability questions, a substantial number of people fail to answer the stock expectation questions at all. A number of researchers have suggested that "50" is an indicator of "epistemic uncertainty" or imprecise probability beliefs (Fischhoff and Bruine De Bruin, 1999; Lillard and Willis, 2001). Of course, it is also possible that some people who answer "50" mean that the event in question has a 50 percent chance of occurring or that they think that the probability falls within some range such as 40-60 percent and give "50" as rounded approximation (Manski and Molinari, 2010).

Beginning in 2006 the HRS added an "epistemic" follow-up question to several probability questions, including the P_0 question, to help understand the meaning of "50" answers:

Do you think that it is about equally likely that these mutual fund shares will increase in worth as it is that they will decrease in worth by this time next year, or are you just unsure about the chances?

We now turn to a discussion of the 2008 data that we use in this paper. In 2008, HRS continued to ask the "epistemic" follow-up to persons who answered 50 to P_0. For those who did not respond 50 to P_0 or, if they answered "50", indicated that the shares were equally likely to increase or decrease in value, HRS added a follow-up question:

By next year at this time, what is the chance they will have grown by x percent or more? (For negative values of x: By next year at this time, what is the chance they will have declined by -x percent or more?)

where the probability of a gain of x per cent from the set of {+10, +20, +30, +40} or a loss of –x percent from the set of {-40, -30, -20, -10} is randomly assigned.[58] We denote the answer to this question as P_{x+} if the random value of x is positive and as P_{x-} if it is negative. Note that P_{x+} denotes the probability that returns would be greater than x, while P_{x-} denotes the probability that they would be less than -x.

The full sample consists of 17,217 individuals from 11,897 households. We restricted the sample to those 14,735 persons who participated in the last three waves of HRS (2004, 6, and 8). In 2004, the sample was refreshed by a new, younger cohort. Out of the 14,735 people, 2,850 (˜19 percent) did not answer the P_0 question (the majority answered "I do not know") leaving us a sample of 11,885 people. As we indicated earlier, HRS did not ask the P_{x+} or P_{x-} questions from those who told that they were unsure in response to the "epistemic" follow-up question (2,005 individuals). Answer to the P_{x+} or the P_{x-} question is missing for another 486 individuals, and education was missing for an additional 45 individuals. Putting all these restrictions together, we ended up with a sample of 9,348 individuals. The average age is 68 years, and 90 percent of the sample is 55 to 89 years old. We divide the sample into four sub-samples based on the date of the interview (see Figure 4.1). These subsamples are very unbalanced in terms of the number of observations. 6285 respondents gave interview between February and June 2008, 2286 between July and September 2008, 556 in October and November 2008, and 211 between December 2008 and February 2009.

As we see, there are many missing values in the HRS stock market expectation data. Out of the 14,735 people asked, only 9,348 (63 percent) gave adequate answers to both questions. The two main sources of missing values are the "I do not know" answers to any of the questions and being "unsure" after giving a 50 percent answer to P_0. In the analysis we

[58] Randomization of x was not complete in the survey: those who gave 0% for the P_0 question were assigned to get a random x with $x < 0$ but not $x > 0$, while those who answered $P_0 = 100\%$ were assigned to get a random x with $x > 0$ but not $x < 0$.

shall ignore the missing values. We think that their omission does not invalidate our results for two reasons. First, people who ”do not know” or are “unsure” might not have meaningful expectations about the stock market and thus they are not part of the population we would like to represent. These questions are not easy to answer, and if someone has no stocks and is sure that she will never have to deal with financial assets, she does not have to form expectations about the one-year-ahead returns asked in the survey. Second, our goal is to analyze the changes in expectations after the crash. As long as the crash itself did not result in an increase or decrease of missing answers, the sample selection problem does not influence our main results. Analysis of the time series of missing answers reveals that the stock market crash did not bring about more “don't know” or “unsure” answers. There is a small temporary decrease in the fraction of ”don't know” answers in October, but the decrease is both quantitatively small and statistically insignificant.

The distribution of P_0 answers is shown in the histograms of Figure 4.2. In the left panel we see the above mentioned heaping at 50. The right panel shows that the heaping disappears if we only leave in those 50 respondents who think that shares were equally likely to increase or decrease in value (rather than being unsure). Note that HRS did not ask the follow-up P_x questions from the “unsure” people so later in the analysis we will only use people from the right panel.

Figure 4.2 also highlights another interesting issue, excessive rounding. Table 4.1 shows that almost 99 percent of the answers are multiples of 5, and more than 80 percent are multiples of 10. The fact that people give approximate answers to these probability questions is not surprising, since it is very hard to compute these numbers more precisely. A careful analysis, thus, should incorporate this feature of the data.

Table 4.1, however, highlights an even more important problem, that of inconsistent answer-pairs to the probability questions. Strongly inconsistent answers are those that contradict the laws of probability: $P_0 < P_{x+}$ or $P_0 + P_{x-} > 100$. Zero mass answers are the ones that imply zero probability of returns between the asked probabilities: $P_0 = P_{x+}$ and $P_0 + P_{x-} = 100$. Nearly 17 percent of the answers are strongly inconsistent, and more than 21 percent imply zero mass. On top of these problems, Kézdi and Willis (2008) document that many HRS respondents do not give the same answer to the same probability question (say, P_0) when it is asked twice within the survey twenty minutes apart. Analyzing stock market expectations in another dataset, Gouret and Hollard (2011) show that few people give answers that imply the same expectations if they are asked in two slightly different ways within the same survey. Perhaps surprisingly, both Kézdi and Willis (2008) and Gouret and Hollard (2011) find no relationship between personal characteristics and the propensity to give problematic answers, with the potential exception of income and expectations themselves.

We argue that such answers are due primarily to question-specific survey noise due to inattention. Survey responses are the results of individual behavior under circumstances that differ from circumstances when making an actual investment decision. Answers are given in a matter of seconds and there are practically no incentives to get the answers right. Therefore, we would be wrong to assume that the survey answers are equivalent to the probabilities that represent people's subjective return distribution which forms the basis for their investment decisions. In section 4 we propose a method to separate survey noise from relevant heterogeneity in expectations.

4.3 Descriptive analysis

In this section, we analyze the answers to the probability questions in a direct way. This should be viewed as preliminary descriptive analysis that cannot estimate the magnitude of the effect of the stock market crash, for two reasons. First, the probabilities themselves are not the objects of interest and, second, survey noise can lead to biased estimates (especially on the heterogeneity of beliefs). At the same time, the descriptive analysis is free of additional assumptions that we need to make in order to recover more meaningful statistics.

Before the descriptive analysis, it is instructive to discuss how probabilities P_0 and P_x are related to the parameters of interest. Standard portfolio choice models include first and second moments of the (perceived) distribution of future returns as opposed to the probabilities themselves. With the help of additional distributional assumptions, answers to two probability questions can help identify the subjective mean and variance of the returns. Recall that the object of interest is the distribution of the one-year-ahead returns of the stock market as viewed by the respondent. If we assume that people believe that the distribution of percentage returns is normal, two points in the subjective distribution identify the entire distribution and thus both the mean and the variance. Figure 4.3 shows a normal c.d.f. that is identified by the two points. The figure depicts the case where the mean of returns is 0.07 and standard deviation is 0.15, numbers close to the post-war moments of nominal yearly returns on the Dow Jones (ending with year 2007). The probability of positive returns is around 68 per cent (1–0.32), while the probability of returns of at least 20 per cent (0.2) is around 20 per cent (1–0.80). A respondent with the postwar-pre-2008 distribution in mind would answer P_0 to be 68 per cent and P20 to be 20 per cent.

Using answers to the two probability questions, one can in principle derive the mean (μ) and the standard deviation (σ) of the beliefs of individual i. Intuitively, the mean is identified from the level of the answers, while the standard deviation is identified from the distance between the two answers (larger distance means smaller variance). Formally, we can take inverse of the appropriate probabilities:

$$P_{0i} = \Phi\left(\frac{\mu_i}{\sigma_i}\right), \tag{4.1}$$

$$P_{x+,i} = \Phi\left(\frac{\mu_i - x/100}{\sigma_i}\right), \tag{4.2}$$

$$P_{x-,i} = \Phi\left(\frac{x/100 - \mu_i}{\sigma_i}\right), \tag{4.3}$$

where, P_0 is the answer to the probability of positive returns, P_{x+} is the answer to the probability of returns at least x per cent, and P_{xi} is the answer to the probability of losses of at least x per cent. Note that a mean-preserving spread in uncertainty (σi) pushes the probabilities towards 0.5, because an increase in σi moves the index towards zero. This is very much in line with the casual interpretation of a "fifty-fifty" answer as reflecting ignorance. Using the example of positive x returns, inverting the probabilities would give this simple nonlinear but exactly identified system of two equations and two unknowns (μ_i and σ_i):

$$\Phi^{-1}(P_{0i}) = \frac{\mu_i}{\sigma_i}, \tag{4.4}$$

$$\Phi^{-1}(P_{x+,i}) = \frac{\mu_i - x/100}{\sigma_i}. \tag{4.5}$$

Unfortunately, survey answers to the probability questions are not suited for such a direct transformation at the individual level. The excessive rounding and the relatively high fraction of inconsistent probability answers discussed in the previous section would invalidate such an analysis. In the next section, we propose a method for modeling both rounding and survey noise within a structural model. Before that, we present some basic descriptive results in this section.

In order to see if the stock market crash brought about changes in expectations about stock market returns, we estimate simple OLS regressions with crude proxies for the subjective mean (μ_i) the subjective standard deviation (σ_i) and the heterogeneity of expectations. In each regression, the right-hand side variables include three dummies for the four periods we focus on: February through June 2008 is the reference category, the first dummy is for July through September 2008, the second dummy is for October through November 2008, and the third dummy is for December 2008 through February 2009.

We estimate regressions with the probability answers themselves on the left-hand side in order to assess the effects on the population average of the level of the return distribution. If people become more pessimistic on average, we expect their answers to both the P_0 and the P_{x+} question to drop on average. If the second probability question has a negative threshold, their answerP_{x-} would go up on average. We therefore run two regressions, one with P_0 on the left-hand side, and one with P_{x+} or $1 - P_{x-}$ on the left-hand side. In order to partial out any threshold-specific factors that may bias answers to the second question, the second regression includes dummies for the different thresholds. The reference category is $x = +10$.

In order to see the effect of the crash on the cross-sectional heterogeneity of expectations (which we call disagreement), we look at regressions in which the left-hand side variables are the absolute values of the residuals from the previous regressions. If disagreement increases, the residuals from the previous regression would become more dispersed, and their absolute value would therefore go up.

The effect of the crash on the population average of subjective uncertainty is approximated by a regression with the difference in the two probability answers on the left-hand side. Recall from Figure 4.2 that the difference between P_0 and P_{x+} is inversely related to the standard deviation of the subjective distribution. Another way to see the connection is in terms of the p.d.f.: a larger difference would imply a larger probability mass concentrated on the support between P_0 and P_{x+}, which implies a less dispersed distribution. If the threshold of the second probability question is negative, the probability mass between P_0 and P_{x-} is given by $1 - (P_0 + P_{x-})$. In order for an increase in uncertainty to show up with a positive sign in the regressions, we used the negative of the differences for left-hand side variables: $P_{x+} - P_0$ for positive thresholds and $[P_0 + P_{x-} - 1]$ for negative thresholds.

Before we turn to the results of the regressions, we address the question of whether the date of the interview is exogenous to prior stock market expectations. This is our most important identifying assumption in analyzing the effect of the stock market crash. The interview date was not randomly assigned. The HRS released the names of all sample households to its national field staff of interviewers at the beginning of the field period in February, 2008. Interviews were then completed in a sequence determined by each interviewer in consultation with regional field supervisors over the entire field period which ended in February, 2009. Sample members who are hardest to locate, most difficult to schedule and most reluctant to be interviewed tend to receive interviews relatively late in the field period. Ultimately, over 90 percent of eligible sample members were interviewed.

In the 2004 and 2006 waves of the survey, HRS collected data on P_0 from respondents in our sample (but there were no second probability questions asked on stock market expectations). Using these variables we can look at whether the date at which people were interviewed in 2008 is related to their answers to the P_0 questions in previous interviews. We estimated four regressions with stock market expectation variables from 2004 and 2006 on the left hand side and interview date in 2008 on the right hand side. The first two regressions have P_0 on their left hand side, while the third and fourth regressions have the residuals from those regressions (in each pair one is for 2004 and the other is for 2006). According to the discussion above, these regressions estimate the "effect" of interview date in 2008 on the average level of expectations prior to 2008 and heterogeneity of those expectations prior to 2008, respectively. The results from these "placebo" regressions are shown in Table 4.2. The only significant correlation with interview date in HRS 2008 and previous expectations is in column [2]: those who answered HRS 2008 between October and November gave slightly higher P_0 answers in 2004 on average. At the same time, no such relationship was found in more recent 2006. Overall, the results suggest that the date of the interview in HRS 2008 was largely exogenous to stock market expectations prior to 2008, a result that is especially robust in terms of disagreement.

We can now turn to the effects of the interview date in 2008 on expectations in 2008. Table 4.3 shows the results. The dependent variables in columns [1] and [2] are the probability answers, our proxies for the population average of the level of the expectations. The results from the two regressions are very similar. The summer of 2008 brought no changes, and the average level was similar to the reference period in December 2008 through February 2009 as well. However, October and November 2008 saw a significant, if temporary, increase in the average level of expectations.

Columns [3] and [4] report the results on the absolute value of the residual from the previous regressions, which are our proxies for disagreement. The estimates imply that disagreement stayed constant before October 2008, but it increased significantly after the crash. Contrary to the average level of expectations, the increase in disagreement lasted to the end of the sampling period. Column [5] shows the estimates on the difference between the two probability questions, which proxy the effects on the population average of uncertainty. Uncertainty seems to have increased already during the summer, and the crash brought about a substantially larger increase. Similarly to the average level of expectations, though, average uncertainty returned to its baseline level in the last period.

The results from these regressions suggest that on average, people became more optimistic but also more uncertain after the crash, but those increases were temporary. Cross-sectional heterogeneity in expectations also increased after the crash, and that remained high a few months later as well. Unfortunately, as we highlighted earlier, these results are not suited for drawing quantitative conclusions for two reasons: they use crude proxies for the left hand side variables of true interest, and they do not incorporate the complex survey response problems shown in the previous section.

4.4 Structural estimation

In the previous section we derived the relation between the probability answers and the first two central moments of the subjective return distribution under the assumption of normally distributed returns. Because of rounding and response error, as discussed earlier, these relations cannot be mechanically applied to the data. We incorporate rounding and survey noise in our model in two steps.

Assume that when making an investment decision, individual i thinks of one-year ahead returns as R_i^* with mean μ_i and standard deviation σ_i. Throughout the analysis, we assume that R_i^* is normally distributed. (Results are robust to alternative distributional assumptions of Student-t and log-normal as presented in the online Appendix B.) The survey answers of individual i are, however, based on a noisy version of R_i^* that we denote as R_{ji} (where j denotes the question so that j=0,x+ or x-). The noise is assumed to be additive: the mean of R_{ji} is $\mu_i + v_{ji}$, where v_{ji} is a mean-zero noise variable specific to question and individual.

The idea behind this assumption is that in a survey situation individuals have little time and no incentives to retrieve their subjective distribution of stock market returns. As a result, the subjective distribution they have in mind when answering the questions is likely to be different from the subjective distribution they would consider in an investment situation. We allow the noise terms to be different for the two probability questions (P_0 and P_x) but correlated across questions: $Corr(v_{0i}, v_{xi}) = \rho$. The estimation model will allow for estimating both the variance of the survey noise and the correlation. When estimating the noise variance, we assume that it is proportional to subjective uncertainty σ_i. The intuition behind this assumption is those who have more diffuse expectations are likely to have a harder time retrieving those expectations. A consequence of this assumption is that $Var[R_i^*]/Var[R_{ji}]$ is constant. That is, this assumption ensures that the signal-to-noise ratio is constant in terms of perceived stock market returns.

A second feature of our model is that we consider interval responses instead of the reported probabilities themselves. If the reported probability (P_{ji}) is in a pre-specified interval or 'bin' [b1,b2] then the "true" probability (including the noise component v_{ji}) is assumed to be in the same bin but not necessarily the reported probability itself. Because a large fraction of the answers are multiples of 10 (see section 2), we have defined 10 percentage point wide bins: [0, 5); [5, 15); ... [95, 100). One consequence of this assumption is that a round answer can represent any expectation that would lead to probabilities around the particular round number.

The two assumptions are combined to

$$P_{0i} \in [b_1, b_2) \Longleftrightarrow b_1 \leq \Phi\left(\frac{\mu_i + v_{0i}}{\sigma_i}\right) < b_2, \tag{4.6}$$

$$P_{x+,i} \in [b_1, b_2) \Longleftrightarrow b_1 \leq \Phi\left(\frac{\mu_i + v_{x+,i} - x/100}{\sigma_i}\right) < b_2, \tag{4.7}$$

$$P_{x-,i} \in [b_1, b_2) \Longleftrightarrow b_1 \leq \Phi\left(\frac{x/100 - \mu_i - v_{x+,i}}{\sigma_i}\right) < b_2, \tag{4.8}$$

where, as before, P_0 is the probability of positive returns; P_{x+} is the probability of returns of at least x per cent; and P_{x-} is the probability of losses of at least x per cent.

Using interval responses is quite common in the literature dealing with subjective probabilities, but the explicit modeling of survey noise and the maximum likelihood approach is not. For example, Manski and Molinari (2010) argue that, because of rounding, the parameters of interest are only partially identified, and they propose an alternative estimator based on the theory of partial identification and set estimation. Their conservative strategy resulted in very wide estimated parameter sets, especially on the HRS data, probably because of excessive rounding. To avoid this problem we have chosen instead to specify the model fully with distributional assumptions on all the unobserved random variables (see later).

We specify heterogeneity in the subjective mean and variance of returns by equations in two latent left-hand side variables μi, σi, of the form

$$\mu_i = \alpha_\mu w_i + \beta_\mu x_i + \gamma_\mu z_{\mu i} + u_i, \tag{4.9}$$

$$\ln \sigma_i = \alpha_\sigma w_i + \beta_\sigma x_i + \gamma_\sigma z_{\sigma i}. \tag{4.10}$$

In the equations, w is the vector of date of interview dummies; x is the vector of covariates such as race, gender, age, education, and cognitive capacity; the z vectors are equation-specific variables. We say more about them later when we discuss identification.

An important issue addressed in this paper is the possibility of increased cross-sectional heterogeneity in expectations, which may be labeled as disagreement. In order to capture disagreement, we let unobserved heterogeneity in µ vary with the date of the interview. Variance in u (unobserved heterogeneity in µ) measures the heterogeneity of expected returns among individuals who share the same x and z_μ variables. Formally, we let the standard deviation of u be related to the date of interview dummies (w) and the other covariates (x):

$$\ln Std\left(u_i\right) = \alpha_u w_i + \beta_u x_i. \tag{4.11}$$

The last equation is for the standard deviation of the noise, v, which is assumed to be proportional to σ_i:

$$\ln Std\left(v_{ji}\right) = \lambda \sigma_i. \tag{4.12}$$

Equations (4) and (5) describe the parameters of interest as effects on (or correlations with) the expected value of latent variables (μ_i, σ_i), and equation (6) captures the effects on (or correlations with) the standard deviation of the latent variable µi. These latent variables are mapped to the probability answers as specified by the interval response model in equations (1) to (3), which include additive question-specific noise components (v_0 and v_x), as well. The model is completed by distributional assumptions on unobservables u and v. We assume that u, v_0 and v_x are jointly normally distributed and that unobserved heterogeneity, u, is independent of survey noise. However, we allow for v0 and vx to be correlated, and we estimate their correlation. One can argue that the correlation can be different for positive versus negative thresholds in the second question, and thus we estimate two correlation coefficients, one for v_0 and v_{x+} and one for v_0 and v_{x-}.

With these elements the model is complete and can be estimated using Maximum Likelihood. Before we turn to the results, it is worthwhile to spend some time on identification issues. For simplicity, assume for a moment that there are no covariates on the right hand side of (4)-(6). In this unconditional model we would have six parameters to estimate: μ, σ, $Std(u)$, λ, $Corr(v_0, v_{x+})$ and $Corr(v_0, v_{x-})$. In order to estimate them we need at least six moments. Interesting moments are $E(P_{ji})$, $V(P_{ji})$, $E(P_{0i} - P_{xi})$, $V(P_{0i} - P_{xi})$ and the fraction of inconsistent answers. Intuitively $E(P_{ji})$ and $E(P_{0i} - P_{xi})$ help identify $E[\mu_i]$ and $E[\sigma_i]$, while $V(P_{ji})$, $V(P_{0i} - P_{xi})$ and the fraction of inconsistent answers help identify $Std(u)$, λ and the correlations.

The estimation models include covariates and some exclusion restrictions as well. We use two instruments for μ (z_μ in equation 4) and one for σ (z_σ in equation 5). The first instrument for μ is the average probability that respondents assigned to the possibility of an economic recession in the near future in the previous two waves, 2004 and 2006, of the survey. The second instrument is an average score on nine questions about depressive symptoms of the interviewees in 2004 and 2006, such as feeling lonely or feeling sad, etc. again from the previous two waves of the survey. The instrument for σ is the fraction of 50 probability answers in 2004 and 6. The idea behind using this variable is that people who are generally uncertain tend to give a lot of 50-50 answers to probability questions.[59]

4.5 Results of the structural model

The main question addressed by our analysis is how structural parameters of stock market expectations changed through the sample period. Using the model outlined in the previous

[59] This approach was first suggested by Lillard and Willis (2001) and has also been used by Sahm (2007) and Pounder (2007).

section, we estimate changes in the population average of μ_i (the subjective expected value of returns), the population average of σi (the subjective standard deviation of returns), and the population standard deviation of ui (unexplained heterogeneity in the subjective expected value). We capture the change of expectations in time by dummy variables for the four periods: February to June 2008 (the reference period, characterized by relatively high level of stock market indices and low volatility); July to September 2008 (gradual decline, relatively low but increasing volatility); October to November 2008 (the aftermath of the stock market crash and subsequently low levels and high volatility); and December 2008 to February 2009 (low levels with some further decline, and lower volatility). In order to help interpret the coefficients, all right-hand side variables except the interview date dummies are normalized to have zero mean. As a consequence, the regression constant shows the expected value of the left-hand side variable in the reference period (February through June 2008) for an average respondent in the sample. Note that the mean of the right-hand-side variables in the reference period is very close to the overall sample mean. As a result, the regression constant is very close to the actual average response in the reference period. The results are shown in table 4.4.

The estimates are in line with the reduced-form OLS results of Table 4.3. Average optimism about stock market returns increased temporarily in October-November: on average, people seemed to expect a recovery during this period. By December, the average expectations returned to where they were prior to the crash. Average uncertainty about stock market returns increased by 11 per cent during the summer, and it increased again in October-November, by almost an additional 20 per cent. However, average uncertainty seemed to return to its initial level afterwards. Unobserved cross-sectional heterogeneity in expectations increased by 13 per cent in late summer as well, and it increased substantially in the fall. By October and November, the cross-sectional standard deviation was more than 50 per cent larger than it was at the beginning of the year. Heterogeneity decreased somewhat after December, but it remained larger than before the crash.

The coefficients on the other right-hand side variables indicate that women are significantly more pessimistic and more uncertain about stock returns; minorities are substantially more uncertain; older people are less optimistic and less uncertain; more educated people are less uncertain; and smarter people are more optimistic and less uncertain. Stockholders and those who follow the stock market are significantly more optimistic, and the latter are also less uncertain. The results imply that heterogeneity in beliefs is also different in different groups; those that are characterized by higher uncertainty on average tend to be more diverse in their beliefs. The sign on the variables that serve for exclusion restrictions are intuitive: those who were more pessimistic about the economy in the past have lower expectations on average, and the same is true for those with more depressive symptoms. The fraction of fifty-fifty probability answers in the past is a strong predictor of uncertainty about stock market returns.

The last three lines of Table 4.4 contain estimates for the technical parameters. The ratio of true uncertainty to total variance that includes uncertainty as well as survey noise $(\sigma^2/(\sigma^2 + V(v))$ is constant by assumption and is estimated to be 0.645. This implies that the noise variance is almost as large as true uncertainty. The noise terms are allowed to be correlated across questions. The correlation is positive when both probability questions ask about returns higher than a particular threshold value. It is negative when the second question is about the probability of returns smaller than the predefined values.

At first sight, it is surprising that the population average of expected returns is negative during the baseline period. Note, however, that male stockholders who follow the stock market and have above average cognitive capacity expect substantial positive returns on average; their average μ of 0.06 is close to the pre-2007 historical mean of 0.07.

The coefficients on the interview date dummies in Table 4.4 show overall changes in the average level, average uncertainty, and heterogeneity of expectations. It is interesting to see whether those changes were different in different groups. In order to examine such possibilities, we estimated the model with full interaction using dummy variables that split the sample into two parts. The first model with interactions distinguishes between stockholders and non-stockholders. The second model looks at those who follow the stock market versus those who do not. The third model looks at those whose cognitive capacity is above the average versus those below average. The coefficient estimates of the three models are in the online Appendix A. The main results are summarized below with the help of three figures.

We first look at stockholders versus non-stockholders. Stockholders include all those who owned stocks directly, through mutual funds or in tax-sheltered accounts such as 401(k) accounts. Since asset holdings are defined at the household level, members of the same household were assigned the same stockholder status. From an asset pricing point of view, the effect of the crash on stockholders is more interesting than the effect on other households. Note that stockholding may be endogenous to the financial crisis. Therefore, we used the pre-crash stock-holding status from the 2006 wave of HRS. Figure 4.4 shows the results from the interaction model. The figure shows the 25th, 50th and 75th percentile of the population distribution of subjective expected returns (μ_i) for the four sampling periods. The distributions are recovered from the empirical distribution of the covariates and the normal assumption for the unobservables.

The results of Figure 4.4 indicate that stockholders have substantially higher and less uncertain expectations, consistently with standard portfolio choice models. Note that the differences are not captured in full by the interaction of stockholding status with interview date dummies presented in Table 4.6, as the two groups differ in terms of the covariates as well (e.g. demographics and education). The median of the expected return distribution among stockholders is positive throughout the sample period, while the median of the non-holder distribution is negative. Changes in average μ (and average σ, see the online appendix) are similar in the two groups, heterogeneity among stockholders reacted to the stock market crash more in relative terms. In October and November 2008, the estimated inter-quartile range in expected returns rose from about 35 percentage points to almost 60 percentage points among stockholders and from around 60 percentage points to slightly more than 80 percentage points among non-stockholders.

Next we look at the results for the better informed versus less informed individuals. HRS 2008 asked how closely the respondent follows the stock market. 8.5 per cent answered "very closely", and another 36 per cent answered "somewhat closely." The rest answered "not at all" or did not know or refused to answer. We merged the "very closely" and "somewhat closely" categories and called the subsample "informed respondents." The rest we call "uninformed respondents." Being informed and stockholding are of course correlated, but the correlation is far from being perfect. 70 per cent of stockholders claim to follow the stock market (and 30 per cent do not), while 30 per cent of non-holders claim to follow the stock market (and 70 per cent do not). Note that, similarly to stockholding status, whether one follows the stock market is potentially endogenous to the stock market crash. Unfortunately, only HRS 2004 contains the information for our sample, and it is missing there for quite a few individuals. We decided to use the 2008 measures for the analysis despite its potential endogeneity. The results are very similar if one uses the 2004 measures instead.

Figure 4.5 shows the results on the quartiles of expected returns in a way similar to the previous figure. The results are similar to the stockholder versus non-stockholder comparison, with some qualifications. The two groups differ less in terms of the initial level and heterogeneity of expectations than stockholders versus non-stockholders. The increase in the median of the distribution in October and November is more pronounced among the uninformed respondents, while the increase in the inter-quartile range is only marginally larger among the informed people. These results show that actual stockholding status is more strongly related to the effect of the stock market crash on expectations than whether one follows the stock market.

The third comparison is between those with above average cognitive capacity versus those with below average cognitive capacity. Cognitive capacity is measured by the principal component of various measures from HRS 2008. The measures include categories of self-rated memory, the score on immediate and delayed word recall and serial subtraction of seven from one hundred, and answers to three computing exercises, one of which is about compound interest rate. Cognition is correlated with whether one follows the stock market, but the correlation is not extremely strong (66 per cent of informed respondents are above average in terms of cognitive scores, compared to 40 per cent of uninformed respondents). Figure 4.6 shows the results again in terms of the estimated quartiles of subjective expected returns. In terms of the median, the patterns are more similar to what we found for informed versus uninformed people, while the patterns in terms of the inter-quartile range are closer to the patterns by stockholding.

The results of the estimates suggest that the effect on the stock market crash on expectations was different in different groups of the population. The most pronounced differences are found between stockholders and non-stockholders. Differences between informed and uninformed people or higher cognition versus lower cognition people are similar but weaker.

They all suggest, however, that disagreement increased more among those who initially agreed more.

The final question we investigate in this paper is whether the changes brought by the stock market crash are close to what one would predict by changes in different stock market indices. We seek to answer two questions. The first question is whether the patterns of stock market expectations found above are related to the evolution of the stock market. The second question is whether the link between the stock market indices and expectations broke after the crash.

We have created three indicators, all based on the Dow Jones index. The first is the monthly log-return, defined as the log of the average DJIA index from the five days before the interview minus the same lagged by one month. The second indicator is the average of the VXD annualized volatility measure from the five days before the interview. The third measure is the log of the average daily volume of trade of shares in the DJIA index, again from the five days before the interview. These indicators are defined from the same data as the series on Figure 4.1, but their exact definition is somewhat different. The first indicator enters the equation of expected returns (μ), the VDX indicator enters the equation of uncertainty (σ), and the trading volume indicator enters the equation of disagreement ($Std(u)$). The rationale for the last inclusion is that high trading periods might be the ones when traders disagree about the fundamental price of assets and thus volume patterns might be able to predict disagreement.

Table 4.5 contains the estimates from two different specifications. Specification [1] is identical to the specification of Table 4.4 above, except that the stock market indicators are entered instead of the date of interview dummies. Specification [2] differs from specification [1] by allowing for an interaction of the stock market indicators with a dummy variable that is one if the interview date is after the crash (October 2008 through February 2009) and zero otherwise. If the relationship between the stock market indicators and expectations are the same before and after the crash, their coefficients should be stable across the specifications, and all the interaction terms should be zero.

The results for μ and σ are rather clear. They indicate that their relation to the stock market indicator (monthly returns and volatility, respectively) changed dramatically after the crash. Coefficients in specification [2] suggest that before the crash, an increase in the DJIA of 1 per cent was followed by the population average of μ higher by 0.3 percentage points (i.e. 0.003). This magnitude is broadly in line with previous finings by, for example, Kézdi and Willis (2008). The post-crash relationship is just the opposite; there, the coefficient implies that the same one per cent increase would be followed by a drop of 0.4 percentage points (=0.335-0.721). This negative relationship is most likely identified from the fact that within a one month window from the crash, the monthly log returns indicator was large and negative, while average expectations were higher than before. Specification [1] shows no relationship between returns and average expectations, because it shows the mixed results of a positive relationship between changes in the DJIA and the average level of expectations in the pre-crash period and the temporary increase in expectations after the crash. The population average of uncertainty shows a similar pattern; it tracks the volatility index before the crash in specification [2] closely, but its increase after the crash is smaller than what the large increase of the VDX would have implied. Again, specification [1] mixes the two and produces an insignificant estimate.

The results in Table 4.5 are less clear on the association between disagreement and the volume of trade. Our interpretation of the results is that it was largely similar before and after the stock market crash. An increase in the volume by one per cent was associated with a subsequent increase in unobserved heterogeneity by somewhat over 0.3 per cent before the crash in both specifications. The post-crash coefficients suggest a reverse association but also a huge increase in the intercept. Taken literally, they would imply that disagreement increased by astronomical magnitudes right after the crash, and from there it tracked trading volume with a negative coefficient. Recall though that trading volume jumped substantially right after the crash (see Figure 4.1 above), and it decreased considerably and steadily for most of the following time in our sampling frame. At the same time, disagreement stayed substantially higher after the crash than before, and it may have even increased in the immediate aftermath of the crash when volume dropped the most (although we would not have enough power to detect that). We argue that the post-crash association between volume and disagreement was dominated by the large increase in both right after the crash (a strong positive connection), and subsequent movements are of second order importance.

All the results presented in this section are based on the assumption of normally distributed subjective yearly returns. We checked the sensitivity of our results to this functional form assumption by considering two alternatives, the Student-t distribution with various degrees of freedom and the shifted log-normal distribution. The Student-t has fatter tails than the normal. It is motivated by the model of Weitzman (2007) who showed that, if agents have imperfect knowledge about the "true parameters" governing the stochastic process of stock market returns, and the parameters are evolving over time, then the posterior distribution of subjective returns can be Student-t. The shifted log-normal form is motivated by finance theory. While the log-normal is practically identical to the normal for small values of the return, it is quite different for larger values. All of our important results are robust to these alternative assumptions. The detailed results are available in the online Appendix B.

4.6 Conclusions and directions for future research

Using survey data on households' subjective probability beliefs about the one-year-ahead return on the Dow Jones stock market index, we estimated the effect of the stock market crash on the population average of expected returns, the population average of the uncertainty about returns (subjective standard deviation) and heterogeneity in expected returns. We presented estimates both from reduced-form OLS regressions and a structural model that can estimate relevant heterogeneity in subjective expectations and incorporates survey noise at the same time.

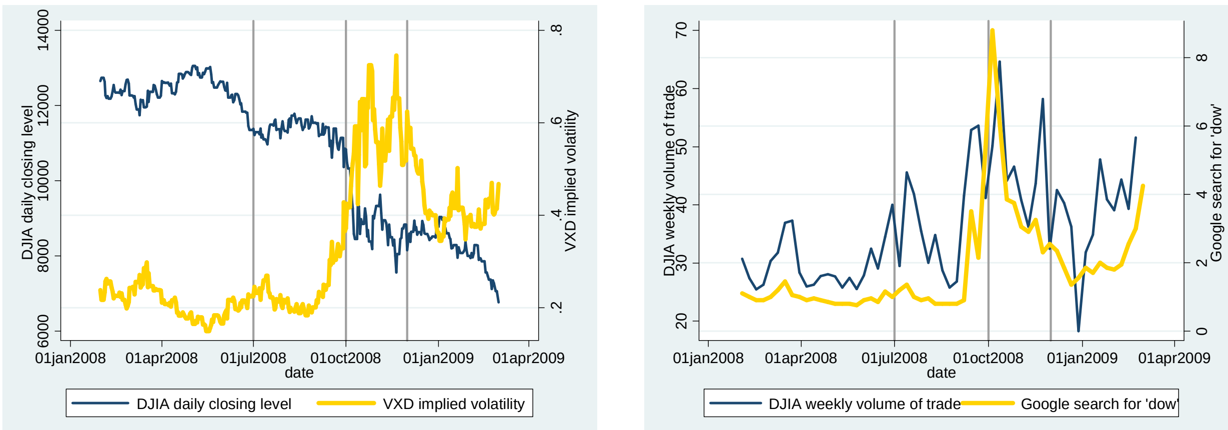
We used data from the Health and Retirement Study that was fielded in February 2008 through February 2009. We identified the effect of the crash from the date of the interview, which we showed to be exogenous to previous stock market expectations. The estimated effects are qualitatively similar in the reduced form regressions and from the structural model, and they are robust to the functional form assumption for the distribution of stock market returns. The results show a temporary increase in the population average of expectations and uncertainty right after the crash. The effect on cross-sectional heterogeneity is more significant and longer lasting, which implies substantial long-term increase in disagreement. Stockholders were found to have more positive, less uncertain and less heterogeneous expectations than non-stockholders, but the stock market crash led to a larger increase in disagreement among them than among non-stockholders. We found similar but smaller differences between those who follow the stock market and those who don't, as well as between those whose cognitive capacity is above the average and those whose cognition is below the average.

The large positive effect of the crash on disagreement suggests that there is heterogeneity in the cognitive processes (or mental models) people use to convert public news into personal probability beliefs, in line with the models of Harris and Raviv (1993) and Kandel and Pearson (1995). The differential effects on stockholders versus non-stockholders, and similar differences between informed and less informed or by cognitive capacity, may be due to the fact that those different groups receive different signals or process the signals in very different ways. These results provide empirical evidence for future research on heterogeneous beliefs in finance theory.

Another natural question for further research is whether the changes in expectations we document lead to changes in asset allocation. Data from HRS 2009 and 2010 will allow for a thorough analysis of the effect of the crisis on the reallocation of household portfolios and the role of expectations.

Tables and figures

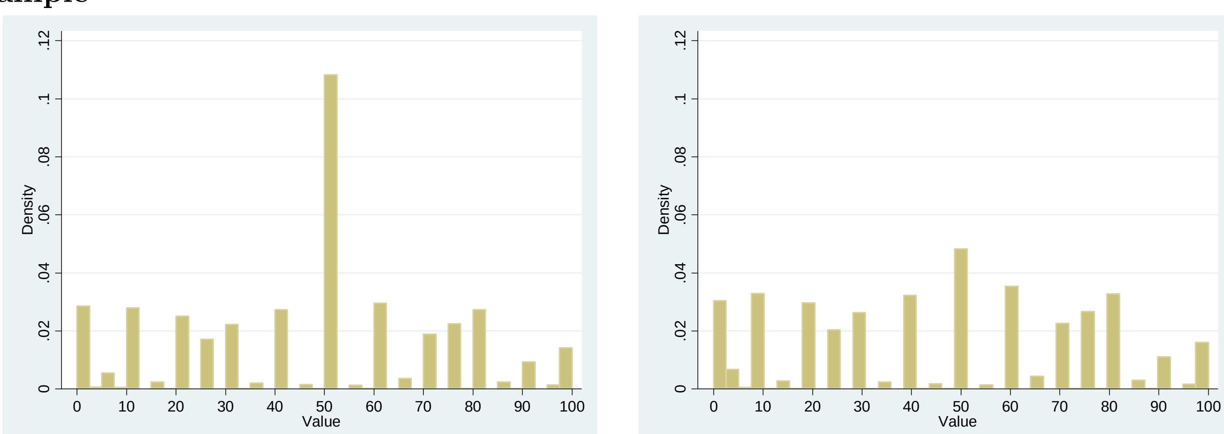
Figure 4.1: Level and volatility of the DJIA, volume of trade and Google search for ”Dow’



Level and volatility Volume of trade and Google search for ’Dow’

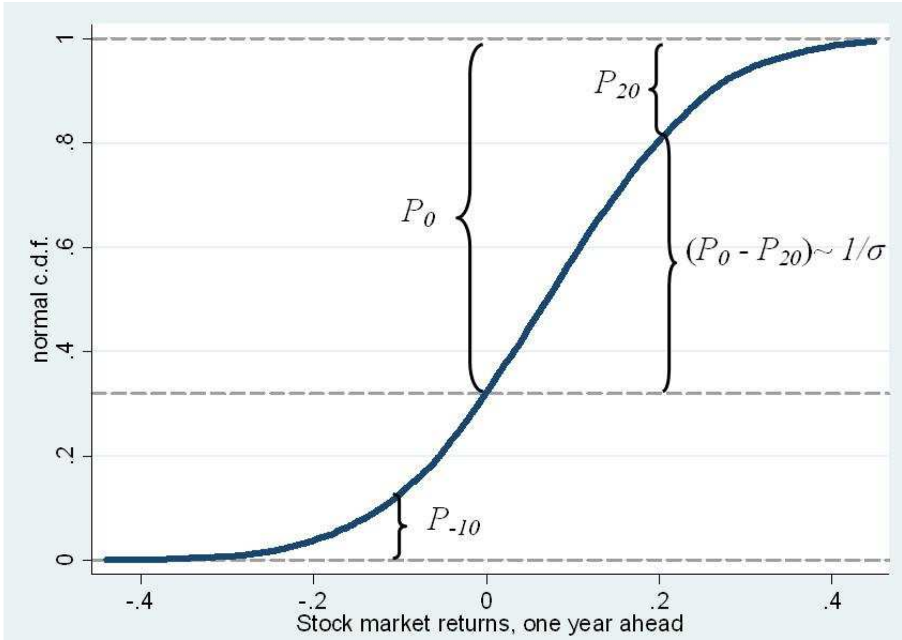
The left panel shows the level of the Dow Jones Industrial Average (daily closing), and the VXD annualized volatility index. The right panel shows the weekly volume of trade in billions of dollars and Google search for ”Dow’ from the US in the sampling period of HRS 2008 (February 2008 through February 2009).

Figure 4.2: Histogram of the P_0 answers in the total sample and in the final sample



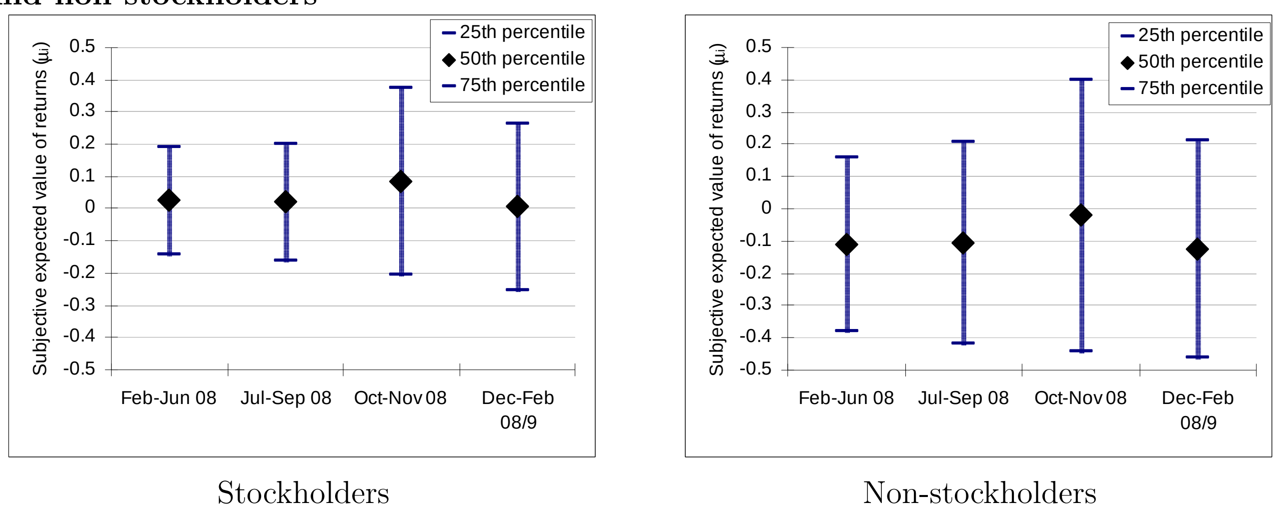
All P_0 answers, N = 11885 P_0 answers in our final sample, N = 9348

Figure 4.3: Standard normal c.d.f. with P_0 and P_{20} shown



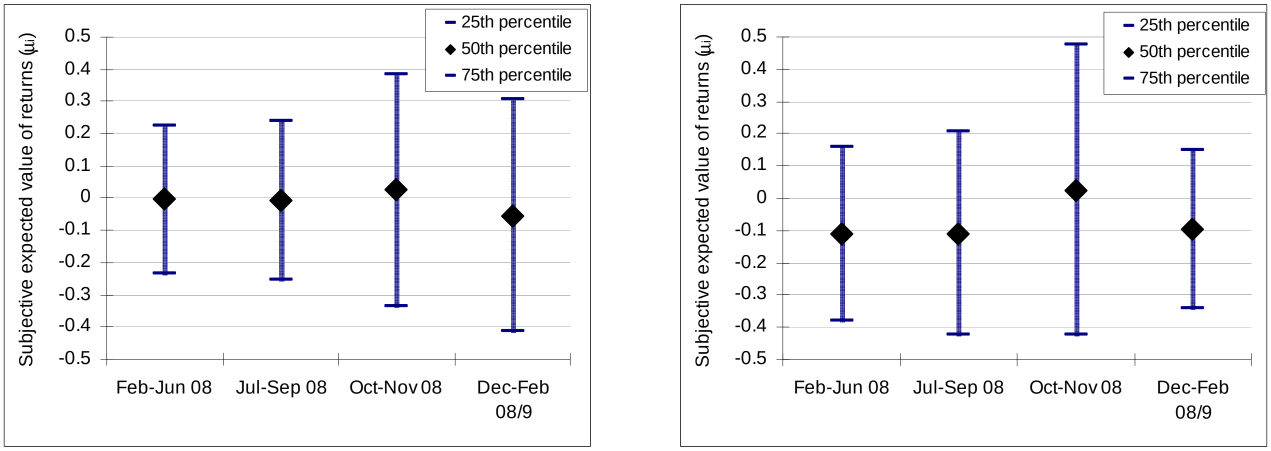
The figure shwos the standard normal c.d.f with $\mu = 0.07$ and $\sigma = 0.15$.

Figure 4.4: Cross-sectional distribution of expected returns among stockholders and non-stockholders



The figures show the quartiles of the distribution by the date of the interview, estimated from the structural model with interactions. Detailed results in Table 4.6 in the appendix.

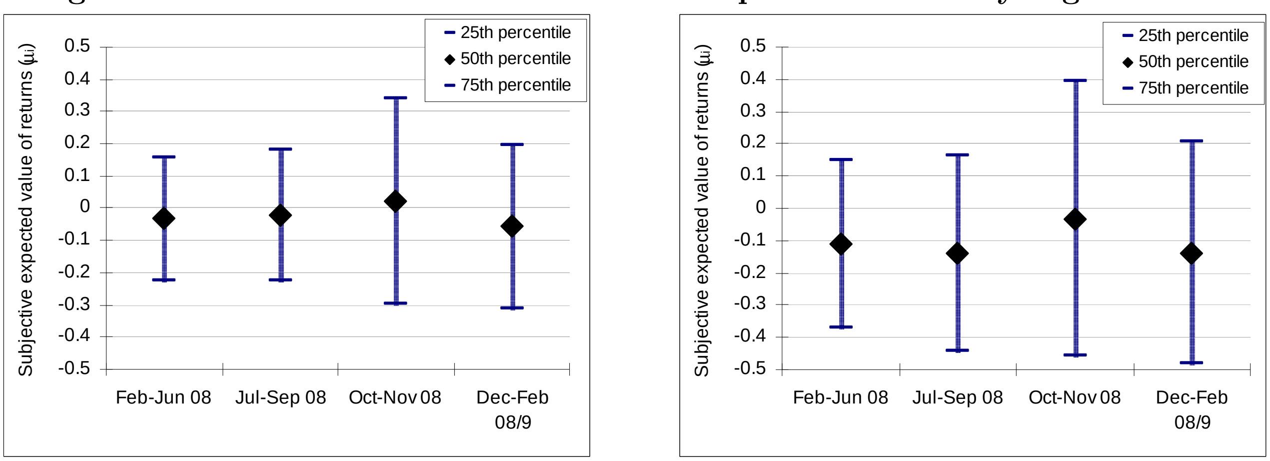
Figure 4.5: Cross-sectional distribution of expected returns among informed and uninformed individuals



Informed Uninformed

The figures show the quartiles of the distribution by the date of the interview, estimated from the structural model with interactions. Detailed results in Table 4.7 in the Appendix. Informed individuals reported to follow the stock market.

Figure 4.6: Cross-sectional distribution of expected returns by cognition



Above average cognitive capacity Below average cognitive capacity

The figures show the quartiles of the distribution by the date of the interview, estimated from the structural model with interactions. Detailed results in Table 4.8 in the Appendix.

Table 4.1: Fraction of rounded and inconsistent probability answers. HRS 2008

		fraction
Rounding	P_0 is a muliple of 10	0.806
	P_0 is 5,25,75 or 95	0.140
	P_0 is a muliple of 5	0.986
Inconsistency	Strongly inconsistent answers*	0.169
	Zero mass answers**	0.215
N		9348

* $P_0 < P_{x+}$ or $P_0 + P_{x-} > 100$

** $P_0 = P_{x+}$ or $P_0 + P_{x-} = 100$

Table 4.2: Placebo regression results: OLS estimates with proxies for the level (columns 1-2) and heterogeneity (columns 3-4) of expectations in 2004 and 2006 as dependent variables, and the time of the interview as right-hand side variables

Dependent variable	P_0 in 2006	P_0 in 2004	$\|u_{P0}\|$ in 2006	$\|u_{P0}\|$ in 2004
	[1]	[2]	[3]	[4]
Constant	49.376	50.764	20.41	20.607
	[0.358]**	[0.351]**	[0.226]**	[0.220]**
Interview in HRS 2008 is	0.728	1.312	-0.432	-0.088
July 08 to September 08	[0.695]	[0.680]	[0.439]	[0.427]
Interview in HRS 2008 is	0.878	2.489	-0.714	0.196
October 08 to November 08	[1.279]	[1.247]*	[0.809]	[0.783]
Interview in HRS 2008 is	3.903	-1.819	1.394	1.235
December 08 to February 09	[1.992]	[2.001]	[1.260]	[1.257]
N	7941	8444	7941	8444

Standard errors in brackets. * significant at 5%; ** significant at 1%. Sample: HRS 2008

"P_x" is defined as P_{x+} for positive thresholds and $(1 - P_{x-})$ for negative thresholds.

Table 4.3: OLS regressions with proxies for the level (columns 1-2), heterogeneity (columns 3-4) and uncertainty (column 5) of expectations.

Dependent Variable	P_0	P_x	$\|u_{P0}\|$	$\|u_{Px}\|$	$P_x - P_0$
	[1]	[2]	[3]	[4]	[5]
Constant	45.627	64.813	24.485	21.434	-24.532
	[0.360]**	[0.785]**	[0.184]**	[0.429]**	[0.966]**
July 08 to September 08	-0.142	0.571	-0.39	0.173	1.702
	[0.698]	[0.628]	[0.356]	[0.343]	[0.773]*
October 08 to November 08	4.922	3.645	1.714	1.762	4.343
	[1.254]**	[1.128]**	[0.640]**	[0.617]**	[1.389]**
December 08 to February 09	0.041	1.06	2.562	2.331	-0.175
	[2.000]	[1.799]	[1.021]*	[0.983]*	[2.215]
Dummies for the		YES		YES	YES
'x' categories					

Standard errors in brackets. * significant at 5%; ** significant at 1%. Sample: HRS 2008

Notes. "P_x" is defined as P_{x+} for positive thresholds and $(1 - P_{x-})$ for negative thresholds.

Table 4.4: Date of interview in 2008 and average subjective expected value of yearly stock returns (μ), average subjective standard deviation (σ) and unobserved cross-sectional heterogeneity in expectations (Std(u)). Results from structural regressions

	μ	$\ln(\sigma)$	$\ln[Std(u)]$
Constant	-0.088	-0.606	-1.23
	[0.006]**	[0.023]**	[0.079]**
July 08 to September 08	0.001	0.113	0.131
	[0.010]	[0.037]**	[0.047]**
October 08 to November 08	0.062	0.292	0.569
	[0.025]*	[0.088]**	[0.099]**
December 08 to February 09	-0.028	0.019	0.38
	[0.033]	[0.119]	[0.135]**
Female	-0.062	0.235	0.146
	[0.009]**	[0.034]**	[0.044]**
Single	0.004	0.04	0.121
	[0.010]	[0.039]	[0.048]*
Black	-0.017	0.589	0.56
	[0.025]	[0.093]**	[0.096]**
Hispanic	0.002	0.387	0.332
	[0.027]	[0.107]**	[0.114]**
Age	-0.002	-0.005	-0.004
	[0.000]**	[0.002]**	[0.002]
Years of education	0.002	-0.034	-0.031
	[0.002]	[0.007]**	[0.009]**
Above average cognition	0.031	-0.101	-0.198
	[0.010]**	[0.037]**	[0.048]**
Follow the stock market	0.049	-0.129	-0.073
	[0.010]**	[0.038]**	[0.047]
Stockholder	0.072	-0.058	-0.18
	[0.010]**	[0.038]	[0.050]**
P(economic recession)	-0.003		
2004-2006 average	[0.000]**		
Depressive symptoms	-0.017		
2004-2006 average	[0.005]**		
Ratio of fifty-fifty answers		1.512	
2004-2006 average		[0.191]**	
Log-likelihood	-42277		
N	9348		
Mean(μ)	-0.085		
Mean(σ)	0.616		
Mean(Std(u)	0.343		
σ2/(σ2+V(v))	0.645		
Rho(v0,vx-)	-0.491		
Rho(v0,vx+)	0.252		

Standard errors in brackets. * significant at 5%; ** significant at 1%

Reference categories: Interview date March to June, male,

non-Black and non-Hispanic, married

Sample: HRS 2008

Table 4.5: The effects of recent returns and volatility of the stock market index and the daily volume of trade of the shares of the DJIA, before and after the crash

	Specification [1]			Specification [2]		
	μ	$\ln(\sigma)$	$\ln[Std(u)]$	μ	$\ln(\sigma)$	$\ln[Std(u)]$
Constant	-0.085	-0.568	-8.955	-0.084	-0.785	-8.454
	[0.006]**	[0.041]**	[2.213]**	[0.005]**	[0.089]**	[2.234]**
Monthly log returns	0.048			0.335		
(log of the average of previous 5 days						
minus the same one month before)	[0.081]			[0.104]**		
VXD volatility index		0.013			0.977	
(average of previous 5 days)		[0.151]			[0.408]*	
Log of volume of trade			0.350			0.327
(average of previous 5 days)			[0.097]**			[0.099]**
Post-crash dummy				-0.007	0.210	19.295
(October 08 to February 09)				[0.032]	[0.256]	[6.101]**
Post-crash dummy interacted with				-0.721		
monthly log returns				[0.257]**		
Post-crash dummy interacted with					-0.617	
VXD volatility index					[0.601]	
Post-crash dummy interacted with						-0.839
log of volume of trade						[0.271]**
Other covariates	YES	YES	YES	YES	YES	YES
Instruments	YES	YES		YES	YES	
Log-likelihood	-42301.2			-42268.1		
N	9347			9347		

Standard errors in brackets. * significant at 5%; ** significant at 1%. Sample: HRS 2008

APPENDIX J
Detailed estimates from the models with interactions

Table 4.6: Stockholders versus non-stockholders. Date of interview in 2008 and average subjective expected value of yearly stock returns (μ), average subjective standard deviation (σ) and unobserved cross-sectional heterogeneity in expectations ($Std(u)$)

	μ		$\ln(\sigma)$		$\ln[Std(u)]$	
	coef	interaction with stockholding	coef	interaction with stockholding	coef	interaction with stockholding
Constant	-0.091	0.087	-0.597	-0.134	-1.176	-0.229
	[0.007]**	[0.013]**	[0.026]**	[0.052]**	[0.082]**	[0.063]**
July 08 to Sep08	0.008	-0.013	0.135	-0.052	0.158	-0.066
	[0.018]	[0.022]	[0.067]*	[0.082]	[0.076]*	[0.098]
Oct 08 to Nov 08	0.091	-0.031	0.29	-0.066	0.465	0.135
	[0.041]*	[0.052]	[0.155]	[0.188]	[0.161]**	[0.203]
Dec 08 to Feb 09	-0.012	-0.006	-0.113	0.194	0.231	0.251
	[0.052]	[0.068]	[0.173]	[0.246]	[0.181]	[0.268]
Female	-0.103	0.05	0.234	0.076	0.157	0.024
	[0.016]**	[0.019]**	[0.059]**	[0.073]	[0.066]*	[0.087]
Single	0.012	-0.017	0.135	-0.169	0.244	-0.232
	[0.018]	[0.022]	[0.062]*	[0.081]*	[0.070]**	[0.096]*
Black	0.015	-0.126	0.539	0.101	0.483	0.276
	[0.028]	[0.062]*	[0.104]**	[0.228]	[0.107]**	[0.241]
Hispanic	0.013	-0.04	0.408	-0.107	0.357	-0.004
	[0.035]	[0.055]	[0.134]**	[0.223]	[0.139]*	[0.242]
Age	-0.002	-0.001	-0.011	0.011	-0.012	0.018
	[0.001]**	[0.001]	[0.003]**	[0.004]**	[0.003]**	[0.004]**
Years of education	0.004	-0.001	-0.04	-0.006	-0.033	-0.028
	[0.003]	[0.004]	[0.011]**	[0.015]	[0.012]**	[0.017]
P(economic recession)	-0.004	0.001				
2004-2006 average	[0.000]**	[0.000]				
Depressive symptoms	-0.025	0.007				
2004-2006 average	[0.008]**	[0.010]				
Ratio of fifty answers			1.673	-0.42		
2004-2006 average			[0.249]**	[0.261]		
Ll	-42278.6					
N	9348					
Mean(μ)	-0.083					
Mean(σ)	0.622					
Mean($Std(u)$)	0.355					
Var[$R*$]/Var[R]	0.652					
Rho(v_0, v_{x-})	-0.535					
Rho(v_0, v_{x+})	0.225					

Standard errors in brackets. * Significant at 5%; ** significant at 1%. Sample: HRS 2008

Stockholders are those who own any stock-market based assets (stocks, mutual funds etc.) either directly or through retirement accounts.

Members of the same households are assigned the same stockholding status. For more details see main text and footnotes to Table 4

Table 4.7: Informed versus uninformed respondents. Date of interview in 2008 and average subjective expected value of yearly stock returns (μ), average subjective standard deviation (σ) and unobserved cross-sectional heterogeneity in expectations ($Std(u)$)

	μ		$\ln(\sigma)$		$\ln[Std(u)]$	
	coef	interaction with informed	coef	interaction with informed	coef	interaction with informed
Constant	-0.085	0.06	-0.588	-0.123	-1.084	-0.087
	[0.006]**	[0.012]**	[0.025]**	[0.050]*	[0.084]**	[0.054]
July 08 to Sep08	0.001	-0.004	0.106	0.005	0.165	-0.079
	[0.020]	[0.023]	[0.072]	[0.085]	[0.080]*	[0.097]
Oct 08 to Nov 08	0.136	-0.108	0.311	-0.08	0.528	-0.047
	[0.049]**	[0.056]	[0.186]	[0.210]	[0.192]**	[0.218]
Dec 08 to Feb 09	0.014	-0.064	-0.114	0.138	-0.101	0.582
	[0.054]	[0.068]	[0.211]	[0.256]	[0.225]	[0.276]*
Female	-0.096	0.058	0.166	0.128	0.107	0.081
	[0.016]**	[0.019]**	[0.063]**	[0.076]	[0.070]	[0.086]
Single	0.01	-0.029	0.111	-0.096	0.183	-0.058
	[0.018]	[0.022]	[0.069]	[0.084]	[0.075]*	[0.093]
Black	0.01	-0.077	0.458	0.221	0.461	0.244
	[0.032]	[0.048]	[0.122]**	[0.180]	[0.124]**	[0.184]
Hispanic	0.066	-0.158	0.179	0.347	0.173	0.311
	[0.032]*	[0.054]**	[0.135]	[0.211]	[0.140]	[0.219]
Age	-0.003	0.001	-0.011	0.01	-0.01	0.011
	[0.001]**	[0.001]	[0.003]**	[0.004]**	[0.003]**	[0.004]**
Years of education	0.006	0	-0.034	-0.018	-0.026	-0.046
	[0.003]	[0.004]	[0.011]**	[0.014]	[0.012]*	[0.016]**
P(economic recession)	-0.003	0				
2004-2006 average	[0.000]**	[0.000]				
Depressive symptoms	-0.022	0				
2004-2006 average	[0.008]**	[0.010]				
Ratio of fifty answers			1.634	-0.537		
2004-2006 average			[0.252]**	[0.230]*		
Ll	-42300.4					
N	9348					
Mean(μ)	-0.08					
Mean(σ)	0.608					
Mean($Std(u)$)	0.376					
Var[$R*$]/Var[R]	0.676					
Rho(v_0, v_{x-})	-0.699					
Rho(v_0, v_{x+})	0.138					

Standard errors in brackets. * Significant at 5%; ** significant at 1%. Sample: HRS 2008

Informed are those who claim to follow the stock market at least occasionally. For more details see main text and footnotes to Table 4

Table 4.8: People with above average cognitive capacity versus people with below-average cognitive capacity. Date of interview in 2008 and average subjective expected value of yearly stock returns (μ), average subjective standard deviation (σ) and unobserved cross-sectional heterogeneity in expectations ($Std(u)$)

	μ		$\ln(\sigma)$		$\ln[Std(u)]$	
	coef	interaction with cognition	coef	interaction with cognition	coef	interaction with cognition
Constant	-0.085	0.029	-0.6	-0.116	-1.175	-0.194
	[0.007]**	[0.013]*	[0.025]**	[0.051]*	[0.075]**	[0.060]**
July 08 to Sep08	-0.03	0.043	0.152	-0.056	0.169	-0.081
	[0.019]	[0.022]	[0.070]*	[0.084]	[0.079]*	[0.099]
Oct 08 to Nov 08	0.078	-0.023	0.286	0.003	0.531	0.037
	[0.049]	[0.056]	[0.185]	[0.211]	[0.186]**	[0.218]
Dec 08 to Feb 09	-0.027	0.005	-0.249	0.405	0.31	0.017
	[0.058]	[0.070]	[0.188]	[0.245]	[0.191]	[0.262]
Female	-0.097	0.04	0.2	0.129	0.1	0.155
	[0.017]**	[0.019]*	[0.063]**	[0.075]	[0.071]	[0.088]
Single	-0.016	0.013	0.128	-0.105	0.259	-0.189
	[0.019]	[0.022]	[0.069]	[0.084]	[0.077]**	[0.097]
Black	0.011	-0.113	0.485	0.285	0.457	0.336
	[0.030]	[0.053]*	[0.113]**	[0.199]	[0.115]**	[0.208]
Hispanic	0.024	-0.074	0.325	0.134	0.284	0.159
	[0.036]	[0.054]	[0.146]*	[0.214]	[0.151]	[0.228]
Age	-0.003	0.001	-0.008	0.005	-0.011	0.01
	[0.001]**	[0.001]	[0.003]**	[0.004]	[0.003]**	[0.004]*
Years of education	0.009	-0.003	-0.046	0.011	-0.037	-0.014
	[0.003]**	[0.004]	[0.011]**	[0.014]	[0.012]**	[0.017]
P(economic recession)	-0.003	0				
2004-2006 average	[0.000]**	[0.000]				
Depressive symptoms	-0.025	0.005				
2004-2006 average	[0.008]**	[0.010]				
Ratio of fifty answers			1.408	0.239		
2004-2006 average			[0.215]**	[0.268]		
Ll	-42319.7					
N	9348					
Mean(μ)	-0.083					
Mean(σ)	0.617					
Mean($Std(u)$)	0.351					
Var[$R*$]/Var[R]	0.649					
Rho(v_0, v_{x-})	-0.518					
Rho(v_0, v_{x+})	0.235					

Standard errors in brackets. * Significant at 5%; ** significant at 1%. Sample: HRS 2008

Cognitive capacity is measured by a score from memory and numeracy tasks. For more details see main text and footnotes to Table 4

APPENDIX K
Estimates based on alternative functional form assumptions

In this appendix we re-estimate our structural models with two alternative specifications for the return distribution, Student-t with various degrees of freedom and shifted log-normal. Within the Student-t framework, the relation between the probability answers and the estimated parameters is the following:

$$\begin{aligned}
P_{0i} &= T\left(\frac{\mu_i}{\sigma_i}, df\right), && (4.13)\\
P_{x+,i} &= T\left(\frac{\mu_i - x/100}{\sigma_i}, df\right), && (4.14)\\
P_{x-,i} &= T\left(\frac{x/100 - \mu_i}{\sigma_i}, df\right), && (4.15)
\end{aligned}$$

where df is the degrees of freedom of the distribution. Note that since the t distribution is symmetric, μ can still be interpreted as the mean of the subjective distribution, σ will not be the subjective standard deviation anymore. According to the properties of the Student-t distribution, the true standard deviation will be

$$\sigma_{Ri} = \sigma_i \sqrt{\frac{df}{df-2}} \tag{4.16}$$

Under the shifted log-normal assumption returns follow , it can be shown that the probability answers are:

$$
\begin{aligned}
P_{0i} &= \Phi\left(\frac{\mu_i}{\sigma_i}\right), && (4.17)\\
P_{x+,i} &= \Phi\left(\frac{\mu_i - \ln\left(x/100\right)}{\sigma_i}\right), && (4.18)\\
P_{x-,i} &= \Phi\left(\frac{\ln\left(x/100\right) - \mu_i}{\sigma_i}\right), && (4.19)
\end{aligned}
$$

For small return realizations the normal and the shifted log-normal distributions are very similar. We, however, have quite high subjective variance, and thus a very large fraction of the return distribution is in a range where the true distributions are different. An important difference between the normal and the shifted lognormal model is that they have different moments. It can be shown that the first two moments of the shifted lognormal distribution are:

$$
\begin{aligned}
\mu_{Ri} &= \exp\left(\mu_i + \frac{\sigma_i^2}{2}\right) - 1, && (4.20)\\
\sigma_{Ri}^2 &= \left(\exp\left(\sigma_i^1\right) - 1\right)\exp\left(2\mu_i + \sigma_i^2\right) && (4.21)
\end{aligned}
$$

Because it is skewed, the mean of the distribution will be larger than the estimated µ. The potential asymmetry of the subjective distribution, thus, can be a reason for having small, usually negative values for µ.

As we can see, the parameter estimates are very much the same in all specifications, and the qualitative results do not change. All specifications agree that the crash brought a moderate increase in uncertainty and a huge increase in disagreement. The normal and the Student-t models agree that the mean of the subjective distribution was weakly positively affected by the crash. The log-normal model is also equivalent to the other models if we talk about log-returns instead of actual returns. As we saw earlier, actual returns are non-linear functions of μ and σ, and the moderate increase in σ must have a positive effect on μ_R. According to these estimates, the average μ_R increased after the crash from 16.5 percent to 72.2 percent, and the overall mean is 21.2 percent. This is clearly implausible.

We can also see that the likelihood is the highest in the original normal case and thus this model fits the data the best. While asymmetric models could explain how the mean of the subjective distribution can be positive even if the average P_0 value is less than 50, we can see that the log-normal model does not fit the data well. Further investigation is needed to test for the potential non-normality of the subjective return distribution.

Table 4.9: Stock returns assumed to be distributed Student-t with 3 degrees of freedom. Date of interview in 2008 and average subjective expected value of yearly stock returns (μ), average subjective standard deviation (σ) and unobserved cross-sectional heterogeneity in expectations ($Std(u)$)

	μ	$\ln(\sigma)$	$\ln[Std(u)]$
Constant	-0.090	-0.871	-1.165
	[0.006]**	[0.024]**	[0.074]**
July 08 to September 08	0.000	0.114	0.122
	[0.010]	[0.038]**	[0.046]**
October 08 to November 08	0.059	0.277	0.552
	[0.025]*	[0.090]**	[0.099]**
December 08 to February 09	-0.027	0.003	0.364
	[0.034]	[0.122]	[0.135]**
Female	-0.063	0.248	0.157
	[0.009]**	[0.035]**	[0.043]**
Single	0.004	0.040	0.118
	[0.010]	[0.040]	[0.048]*
Black	-0.017	0.592	0.562
	[0.026]	[0.095]**	[0.098]**
Hispanic	0.000	0.383	0.327
	[0.027]	[0.109]**	[0.114]**
Age	-0.002	-0.006	-0.004
	[0.000]**	[0.002]**	[0.002]
Years of education	0.001	-0.033	-0.032
	[0.002]	[0.007]**	[0.009]**
Above average cognition	0.031	-0.094	-0.195
	[0.010]**	[0.038]*	[0.047]**
Follow the stock market	0.048	-0.125	-0.075
	[0.010]**	[0.039]**	[0.046]
Stockholder	0.071	-0.047	-0.177
	[0.010]**	[0.039]	[0.049]**
P(economic recession)	-0.003		
2004-2006 average	[0.000]**		
Depressive symptoms	-0.018		
2004-2006 average	[0.005]**		
Ratio of fifty-fifty answers		1.545	
2004-2006 average		[0.196]**	
Log-likelihood	-43091.9		
N	9348		
Mean(μ)	-0.087		
Mean(σ)	0.818		
Mean(Std(u)	0.364		
σ2/(σ2+V(v))	0.763		
Rho(v0,vx-)	-0.574		
Rho(v0,vx+)	0.193		

Standard errors in brackets. * significant at 5%; ** significant at 1%

Reference categories: Interview date March to June, male,

non-Black and non-Hispanic, married

Sample: HRS 2008

Table 4.10: Stock returns assumed to be distributed Student-t with 3 degrees of freedom. Date of interview in 2008 and average subjective expected value of yearly stock returns (μ), average subjective standard deviation (σ) and unobserved cross-sectional heterogeneity in expectations ($Std(u)$)

	μ	$\ln(\sigma)$	$\ln[Std(u)]$
Constant	-0.088	-0.678	-1.213
	[0.006]**	[0.023]**	[0.077]**
July 08 to September 08	0.001	0.113	0.128
	[0.010]	[0.038]**	[0.047]**
October 08 to November 08	0.061	0.288	0.565
	[0.025]*	[0.088]**	[0.099]**
December 08 to February 09	-0.028	0.015	0.376
	[0.033]	[0.120]	[0.135]**
Female	-0.062	0.239	0.149
	[0.009]**	[0.034]**	[0.043]**
Single	0.004	0.040	0.120
	[0.010]	[0.039]	[0.048]*
Black	-0.017	0.590	0.560
	[0.025]	[0.093]**	[0.097]**
Hispanic	0.001	0.386	0.330
	[0.027]	[0.107]**	[0.114]**
Age	-0.002	-0.005	-0.004
	[0.000]**	[0.002]**	[0.002]
Years of education	0.002	-0.034	-0.031
	[0.002]	[0.007]**	[0.009]**
Above average cognition	0.031	-0.099	-0.197
	[0.010]**	[0.038]**	[0.048]**
Follow the stock market	0.049	-0.128	-0.073
	[0.010]**	[0.038]**	[0.046]
Stockholder	0.071	-0.055	-0.180
	[0.010]**	[0.038]	[0.049]**
P(economic recession)	-0.003		
2004-2006 average	[0.000]**		
Depressive symptoms	-0.017		
2004-2006 average	[0.005]**		
Ratio of fifty-fifty answers		1.522	
2004-2006 average		[0.192]**	
Log-likelihood	-42425.5		
N	9348		
Mean(μ)	-0.085		
Mean(σ)	0.641		
Mean(Std(u)	0.348		
σ2/(σ2+V(v))	0.663		
Rho(v0,vx-)	-0.512		
Rho(v0,vx+)	0.238		

Standard errors in brackets. * significant at 5%; ** significant at 1%

Reference categories: Interview date March to June, male, non-Black and non-Hispanic, married

Sample: HRS 2008

Table 4.11: Stock returns assumed to be distributed log-normal with parameters μ_i and σ_i. Date of interview in 2008 and average subjective expected value of yearly stock returns (μ), average subjective standard deviation (σ) and unobserved cross-sectional heterogeneity in expectations ($Std(u)$)

	μ	$\ln(\sigma)$	$\ln[Std(u)]$
Constant	-0.108	-0.565	-1.185
	[0.006]**	[0.023]**	[0.078]**
July 08 to September 08	0.000	0.098	0.116
	[0.010]	[0.038]*	[0.047]*
October 08 to November 08	0.065	0.296	0.572
	[0.026]*	[0.093]**	[0.103]**
December 08 to February 09	-0.027	0.018	0.379
	[0.035]	[0.124]	[0.139]**
Female	-0.067	0.234	0.145
	[0.009]**	[0.035]**	[0.044]**
Single	0.005	0.044	0.127
	[0.011]	[0.040]	[0.049]**
Black	-0.013	0.554	0.527
	[0.025]	[0.092]**	[0.094]**
Hispanic	0.005	0.382	0.328
	[0.028]	[0.109]**	[0.115]**
Age	-0.002	-0.006	-0.005
	[0.000]**	[0.002]**	[0.002]*
Years of education	0.002	-0.031	-0.029
	[0.002]	[0.007]**	[0.009]**
Above average cognition	0.032	-0.081	-0.179
	[0.010]**	[0.038]*	[0.048]**
Follow the stock market	0.050	-0.101	-0.046
	[0.010]**	[0.038]**	[0.047]
Stockholder	0.074	-0.054	-0.177
	[0.010]**	[0.039]	[0.050]**
P(economic recession)	-0.003		
2004-2006 average	[0.000]**		
Depressive symptoms	-0.018		
2004-2006 average	[0.005]**		
Ratio of fifty-fifty answers		1.521	
2004-2006 average		[0.196]**	
Log-likelihood	-42360.2		
N	9348		
Mean(μ)	0.210		
Mean(σ)	9.017		
Mean(Std(u))	0.355		
σ2/(σ2+V(v))	0.644		
Rho(v0,vx-)	-0.484		
Rho(v0,vx+)	0.231		

Standard errors in brackets. * significant at 5%; ** significant at 1%

Reference categories: Interview date March to June, male,

non-Black and non-Hispanic, married

Sample: HRS 2008

Bibliography

Abowd, J. M. and D. Card (1989, March). On the Covariance Structure of Earnings and Hours Changes. *Econometrica, Econometric Society 57*(2), 411–45.

Abowd, J. M. and M. H. Stinson (2007, September). Estimating Measurement Error in SIPP Annual Job Earnings: A Comparison of Census Survey and SSA Administrative Data. Working paper, Federal Committee on Statistical Methodology.

Abowd, J. M. and M. H. Stinson (2013). Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data. *The Review of Economics and Statistics 95*(5), 1451–1467.

Abowd, J. M. and A. Zellner (1985). Estimating Gross Labor-Force Flows. *Journal of Business & Economic Statistics 3*(3), pp.254–283.

Abraham, A. and K. White (2006, July). The Dynamics of Plant-Level Productivity in U.S. Manufacturing. Working Papers 06-20, Center for Economic Studies, U.S. Census Bureau.

Abraham, K. G. and J. R. Spletzer (2009, May). New Evidence on the Returns to Job Skills. *American Economic Review 99*(2), 52–57.

Acemoglu, D. (1999, December). Changes in Unemployment and Wage Inequality: An Alternative Theory and Some Evidence. *American Economic Review 89*(5), 1259–1278.

Acemoglu, D. and D. Autor (2011). *Skills, Tasks and Technologies: Implications for Employment and Earnings*, Volume 4 of *Handbook of Labor Economics*, Chapter 12, pp. 1043–1171. Elsevier.

Acemoglu, D. and R. Shimer (1999, November). Holdups and Efficiency with Search Frictions. *International Economic Review 40*(4), 827–49.

Aizcorbe, A. M. (1992, July). Procyclical Labour Productivity, Increasing Returns to Labour and Labour Hoarding in Car Assembly Plant Employment. *Economic Journal 102*(413), 860–73.

Akee, R. (2011, November). Errors in Self-reported Earnings: The Role of Previous Earnings Volatility and Individual Characteristics. *Journal of Development Economics 96*(2), 409–421.

Albrecht, J. and S. Vroman (2002, February). A Matching Model with Endogenous Skill Requirements. *International Economic Review 43*(1), 283–305.

Altonji, J. G. and C. R. Pierret (2001). Employer Learning and Statistical Discrimination. *The Quarterly Journal of Economics 116*(1), pp.313–350.

Altonji, J. G., A. Smith, and I. Vidangos (2009, February). Modeling Earnings Dynamics. NBER Working Papers 14743, National Bureau of Economic Research, Inc.

Altonji, J. G., A. A. J. Smith, and I. Vidangos (2013, July). Modeling Earnings Dynamics. *Econometrica, Econometric Society 81*(4), 1395–1454.

Altonji, J. G. and N. Williams (2005, April). Do wages rise with job seniority? A reassessment. *Industrial and Labor Relations Review 58*(3), 370–397.

Alvarez, J. and M. Arellano (2004). Robust Likelihood Estimation of Dynamic Panel Data Models. Technical report, CEMFI Working Paper 0421.

Angrist, J. and A. B. Krueger (1999). Empirical Strategies in Labor Economics. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics* (1 ed.), Volume 3, Part A, Chapter 23, pp. 1277–1366. Elsevier.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*(434), 444–455.

Angrist, J. D. and A. B. Krueger (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives 15*(4), 69–85.

Arulampalam, W. and A. L. Booth (2001, August). Learning and Earning: Do Multiple Training Events Pay? A Decade of Evidence from a Cohort of Young British Men. *Economica 68*(271), 379–400.

Autor, D. H., L. F. Katz, and M. S. Kearney (2006, May). The Polarization of the U.S. Labor Market. *American Economic Review 96*(2), 189–194.

Autor, D. H., F. Levy, and R. J. Murnane (2003, November). The Skill Content Of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics 118*(4), 1279–1333.

Azariadis, C. (1975, December). Implicit Contracts and Underemployment Equilibria. *Journal of Political Economy 83*(6), 1183–1202.

Azariadis, C. (1976, February). On the Incidence of Unemployment. *Review of Economic Studies 43*(1), 115–25.

Baily, M. N. (1974, January). Wages and Employment under Uncertain Demand. *Review of Economic Studies 41*(1), 37–50.

Barnichon, R. and A. Figura (2011). Labor Market Heterogeneities, Matching Efficiency, and the Cyclical Behavior of the Job Finding Rate. Mimeo.

Bartel, A., C. Ichniowski, and K. Shaw (2007, November). How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills. *The Quarterly Journal of Economics 122*(4), 1721–1758.

Basu, S. (1996, August). Procyclical Productivity: Increasing Returns or Cyclical Utilization? *The Quarterly Journal of Economics 111*(3), 719–51.

Basu, S. and M. S. Kimball (1997, February). Cyclical Productivity with Unobserved Input Variation. NBER Working Papers 5915, National Bureau of Economic Research, Inc.

Becker, G. S. (1962). Investment in Human Capital: A Theoretical Analysis. *Journal of Political Economy 70*, 9.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal 75*(299), 493–517.

Becker, G. S. (1994, July). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education (3rd Edition)*. NBER Books. National Bureau of Economic Research, Inc.

Bernanke, B. S. and M. L. Parkinson (1991, June). Procyclical Labor Productivity and Competing Theories of the Business Cycle: Some Evidence from Interwar U.S. Manufacturing Industries. *Journal of Political Economy 99*(3), 439–59.

Bils, M. J. (1985). Real Wages over the Business Cycle: Evidence from Panel Data. *Journal of Political Economy 93*(4), pp.666–689.

Blanchard, O. and J. Galí (2010, April). Labor Markets and Monetary Policy: A New Keynesian Model with Unemployment. *American Economic Journal: Macroeconomics 2*(2), 1–30.

Blanchard, O. J. and P. Diamond (1990). The Cyclical Behavior of the Gross Flows of U.S. Workers. *Brookings Papers on Economic Activity 21*(2), 85–156.

Blatter, M., S. Muehlemann, and S. Schenker (2012). The costs of Hiring Skilled Workers. *European Economic Review 56*(1), 20–35.

Blundell, R. and T. Macurdy (1999). Labor Supply: A Review of Alternative Approaches. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics* (1 ed.), Volume 3, Part A, Chapter 27, pp. 1559–1695. Elsevier.

Bollinger, C. R. (1998, July). Measurement Error in the Current Population Survey: A Nonparametric Look. *Journal of Labor Economics 16*(3), 576–94.

Bolvig, I., I. Hardoy, M. Kauhanen, R. Lilja, M. Roed, and N. Smith (2007). *The Labour Supply of Low-Skilled-Incentives in the Unemployment Insurance Systems: A Comparative Description Based on Nordic Countries.* Nordic Council of Ministers.

Bommier, A. and B. Villeneuve (2012, 03). Risk Aversion and the Value of Risk to Life. *Journal of Risk & Insurance 79*(1), 77–104.

Borjas, G. J. (2003, November). The Labor Demand Curve Is Downward Sloping: Reexamining The Impact Of Immigration On The Labor Market. *The Quarterly Journal of Economics 118*(4), 1335–1374.

Bosworth, B. P. and R. Smart (2009, January). Evaluating Micro-Survey Estimates of Wealth and Saving. Working Papers, Center for Retirement Research at Boston College wp2009-4, Center for Retirement Research.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994). Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data. *Journal of Labor Economics 12*(3), pp.345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001). Measurement Error in Survey Data. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5 of *Handbook of Econometrics*, Chapter 59, pp. 3705–3843. Elsevier.

Bound, J. and A. B. Krueger (1991, January). The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? *Journal of Labor Economics 9*(1), 1–24.

Bound, J., T. Stinebrickner, and T. Waidmann (2010). Health, economic resources and the work decisions of older men. *Journal of Econometrics 156*(1), 106–129. Structural Models of Optimization Behavior in Labor, Aging, and Health.

Bowlus, A., H. Liu, and C. Robinson (2002, Part). Business Cycle Models, Aggregation, and Real Wage Cyclicality. *Journal of Labor Economics 20*(2), 308–335.

Brown, C. (1990). Empirical Evidence on Private Training. In *Research in Labor Economics*, Volume 11, pp. 97–113. JAI Press Inc.

Brownstone, D. and R. G. Valletta (1996). Modeling Earnings Measurement Error: A Multiple Imputation Approach. *The Review of Economics and Statistics 78*(4), pp.705–717.

Bruine de Bruin, W. and K. G. Carman (2012). Measuring Risk Perceptions: What Does the Excessive Use of 50% Mean? *32*(2), 232–236.

Bruine de Bruin, W., B. Fischhoff, S. G. Millstein, and B. L. Halpern-Felsher (2000, January). Verbal and Numerical Expressions of Probability: "It's a Fifty-Fifty Chance". *Organizational Behavior and Human Decision Processes 81*(1), 115–131.

Burdett, K. and D. T. Mortensen (1998, May). Wage Differentials, Employer Size, and Unemployment. *International Economic Review 39*(2), 257–73.

Burnside, C. and M. Eichenbaum (1996, December). Factor-Hoarding and the Propagation of Business-Cycle Shocks. *American Economic Review 86*(5), 1154–74.

Burnside, C., M. Eichenbaum, and S. Rebelo (1993, April). Labor Hoarding and the Business Cycle. *Journal of Political Economy 101*(2), 245–73.

Büttner, T., P. Jacobebbinghaus, and J. Ludsteck (2010). Occupational Upgrading and the Business Cycle in West Germany. *Economics - The Open-Access, Open-Assessment E-Journal 4*(10), 1–37.

Cahuc, P., F. Marque, and E. Wasmer (2008, August). A Theory Of Wages And Labor Demand With Intra-Firm Bargaining And Matching Frictions. *International Economic Review 49*(3), 943–972.

Cairo, I. and T. Cajner (2014, September). Human Capital and Unemployment Dynamics: Why More Educated Workers Enjoy Greater Employment Stability. Technical Report 2014-09, Federal Reserve Board, Washington, D.C.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2007, October). Down or Out: Assessing the Welfare Costs of Household Investment Mistakes. *Journal of Political Economy 115*(5), 707–747.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2009a). Fight or Flight? Portfolio Rebalancing by Individual Investors. *The Quarterly Journal of Economics 124*(1), 301–348.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2009b, May). Measuring the Financial Sophistication of Households. *American Economic Review 99*(2), 393–98.

Campbell III, C. M. (1997). The Variation in Wage Rigidity by Occupation and Union Status in the US. *Oxford Bulletin of Economics and Statistics 59*(1), 133–147.

Cannell, C. F., G. Fisher, and T. Bakker (1965). Reporting of Hospitalization in the Health Interview Survey. Technical Report 6, Vital and Health Statistics, Series 2, Public Health Service, Washington.

Card, D. (2001, January). The Effect of unions on wage inequality in the U.S. labor market. *Industrial and Labor Relations Review 54*(2), 296–315.

Card, D., F. Kramarz, and T. Lemieux (1996, March). Changes in the Relative Structure of Wages and Employment: A Comparison of the United States, Canada, and France. Working Paper 5487, National Bureau of Economic Research.

Card, D., T. Lemieux, and W. Riddell (2004). Unions and Wage Inequality. *Journal of Labor Research 25*(4), 519–559.

Carrington, W. J. and A. Zaman (1994, April). Interindustry Variation in the Costs of Job Displacement. *Journal of Labor Economics 12*(2), 243–75.

Carroll, R. J., D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu (2006, July). *Measurement Error in Nonlinear Models: A Modern Perspective* (2nd ed.). Chapman and Hall/CRC.

Chassamboulli, A. (2011). Cyclical Upgrading of Labor and Employment Differences across Skill Groups. *The B.E. Journal of Macroeconomics 11*(1).

Cheremukhin, A. (2010). Labor Matching Model: Putting the Pieces Together. 2010 Meeting Papers 260, Society for Economic Dynamics.

Cheron, A. and B. Rouland (2011, December). Inefficient Job Destructions and Training with Holdup. *LABOUR, CEIS 25*(4), 397–420.

Chodorow-Reich, G. and L. Karabarbounis (2013, November). The Cyclicality of the Opportunity Cost of Employment. Working Paper 19678, National Bureau of Economic Research.

Clark, K. B. and L. H. Summers (1981). Demographic Differences in Cyclical Employment Variation. *Journal of Human Resources 16*(1), 61–79.

Cook, K., D. Dranove, and A. Sfekas (2010). Does Major Illness Cause Financial Catastrophe? *Health Services Research 45*, 418–436.

Cortes, G. M., N. Jaimovich, C. J. Nekarda, and H. E. Siu (2013). The Micro and Macro of Disappearing Routine Jobs: A Flows Approach. Unpublished manuscript.

Couch, K. A. and D. W. Placzek (2010, March). Earnings Losses of Displaced Workers Revisited. *American Economic Review 100*(1), 572–89.

Crifo, P. (2008). Skill Supply and Biased Technical Change. Post-Print hal-00243031, HAL.

Crimmins, E. M., J. K. Kim, K. M. Langa, and D. R. Weir (2011). Assessment of Cognition Using Surveys and Neuropsychological Assessment: The Health and Retirement Study and the Aging, Demographics, and Memory Study. *Journals of Gerontology: Series B 66B*, i162–i171.

Cristia, J. and J. A. Schwabish (2009). Measurement Error in the SIPP: Evidence from Administrative Matched Records. *Journal of Economic and Social Measurement 34*(1), 1–17.

Davis, S. J., R. J. Faberman, and J. Haltiwanger (2006, Summer). The Flow Approach to Labor Markets: New Data Sources and Micro-Macro Links. *Journal of Economic Perspectives 20*(3), 3–26.

Davis, S. J., R. J. Faberman, and J. Haltiwanger (2012). Labor Market Flows in the Cross Section and Over Time. *Journal of Monetary Economics 59*(1), 1–18.

Davis, S. J. and J. Haltiwanger (1990). Gross Job Creation and Destruction: Microeconomic Evidence and Macroeconomic Implications. In *NBER Macroeconomics Annual 1990, Volume 5*, NBER Chapters, pp. 123–186. National Bureau of Economic Research, Inc.

Davis, S. J. and J. C. Haltiwanger (1992, August). Gross Job Creation, Gross Job Destruction, and Employment Reallocation. *The Quarterly Journal of Economics 107*(3), 819–63.

de Bruin, W. B., S. Potter, R. W. Rich, G. Topa, and H. W. V. der Klaauw (2010). Improving Survey Measures of Household Inflation Expectations. *SSRN eLibrary*.

Deaton, A. S. (2009, January). Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development. Working Paper 14690, National Bureau of Economic Research.

Devereux, P. J. (2000, January). Task Assignment over the Business Cycle. *Journal of Labor Economics 18*(1), 98–124.

Devereux, P. J. (2004, January). Cyclical Quality Adjustment in the Labor Market. *Southern Economic Journal 70*(3), 600–615.

Dickens, W. T. (2011). Has the Recession Increased the NAIRU? Technical report, Northeastern University and The Brookings Institution.

Dickens, W. T. and R. K. Triest (2012). Potential effects of the Great Recession on the U.S. labor market. Working Papers, Federal Reserve Bank of Boston 12-9, Federal Reserve Bank of Boston.

Dolado, J., M. Jansen, and J. Jimeno (2009, January). On-the-Job Search in a Matching Model with Heterogeneous Jobs and Workers. *Economic Journal 119*(534), 200–228.

Dominitz, J. and C. F. Manski (2005). Measuring and Interpreting Expectations of Equity Returns. NBER Working Papers 11313, National Bureau of Economic Research, Inc.

Duffy, J., C. Papageorgiou, and F. Perez-Sebastian (2004, February). Capital-Skill Complementarity? Evidence from a Panel of Countries. *The Review of Economics and Statistics 86*(1), 327–344.

Duncan, G. J. and D. H. Hill (1985). An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data. *Journal of Labor Economics 3*(4), pp.508–532.

Egger, H. and V. Grossmann (2005, December). Non-Routine Tasks, Restructuring of Firms, and Wage Inequality Within and Between Skill-Groups. *Journal of Economics 86*(3), 197–228.

Elder, T. (2013, April). The Predictive Validity of Subjective Mortality Expectations: Evidence From the Health and Retirement Study. *Demography 50*(2), 569–589.

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *75*(4), 643–669.

Elsby, M., B. Hobjin, and A. Sahin (2010). The labor market in the Great Recession. Working Paper Series 2010-07, Federal Reserve Bank of San Francisco.

Elsby, M. W. L. and R. Michaels (2008, February). Marginal Jobs, Heterogeneous Firms, & Unemployment Flows. NBER Working Papers 13777, National Bureau of Economic Research, Inc.

Elsby, M. W. L., R. Michaels, and G. Solon (2009, January). The Ins and Outs of Cyclical Unemployment. *American Economic Journal: Macroeconomics 1*(1), 84–110.

Esping-Andersen, G. (2000). *Why Deregulate Labour Markets?*, Chapter ”Who is Harmed by Labour Market Regulations? Quantitative Evidence”, pp. 66–98. Oxford: Oxford University Press.

Evans, P. (1999). Occupational Downgrading and Upgrading in Britain. *Economica 66*(261), pp.79–96.

Fair, R. C. (1985, March). Excess Labor and the Business Cycle. *American Economic Review 75*(1), 239–45.

Fallick, B. C. (1996, October). A review of the recent empirical literature on displaced workers. *Industrial and Labor Relations Review 50*(1), 5–16.

Farber, H. S. (1999). Mobility and stability: The dynamics of job change in labor markets. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 37, pp. 2439–2483. Elsevier.

Farber, H. S. (2005). What do we know about job loss in the United States? evidence from the displaced workers survey, 1984-2004. *Economic Perspectives* (Q II), 13–28.

Farber, H. S. (2011, May). Job Loss in the Great Recession: Historial Perspective from the Displaced Workers Survey, 1984-2010. Working Papers 1309, Princeton University, Department of Economics, Industrial Relations Section.

Fay, J. A. and J. L. Medoff (1985, September). Labor and Output over the Business Cycle: Some Direct Evidence. *American Economic Review 75*(4), 638–55.

Firpo, S., N. M. Fortin, and T. Lemieux (2011, February). Occupational Tasks and Changes in the Wage Structure. IZA Discussion Papers 5542, Institute for the Study of Labor (IZA).

Fischhoff, B. and W. Bruine De Bruin (1999). Fifty-Fifty=50%? *Journal of Behavioral Decision Making 12*(2), 149–163.

Fisher, G. G., H. Hassan, J. D. Faul, W. L. Rodgers, and D. R. Weir (2012). Health and Retirement Study - Imputation of Cognitive Functioning Measures: 1992-2010 Early Release.

Flaaen, A., M. D. Shapiro, and I. Sorkin (2013). Reconsidering the Consequences of Worker Displacements: Survey versus Administrative Measurements.

Foote, C. L. and R. W. Ryan (2014, January). Labor-Market Polarization Over the Business Cycle. In *NBER Macroeconomics Annual 2014, Volume 29*, NBER Chapters. National Bureau of Economic Research, Inc.

Foster, L., J. Haltiwanger, and C. Syverson (2008, March). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review 98*(1), 394–425.

Frandsen, B. R. (2012). Why Unions Still Matter: The Effects of Unionization on the Distribution of Employee Earnings. Working Paper.

Freeman, R. B. Technical report, National Bureau of Economic Research, Inc.

Freeman, R. B. (1980, October). Unionism and the dispersion of wages. *Industrial and Labor Relations Review 34*(1), 3–23.

Fujita, S. and G. Ramey (2009, May). The Cyclicality Of Separation And Job Finding Rates. *International Economic Review 50*(2), 415–430.

Fujita, S. and G. Ramey (2012). Exogenous versus Endogenous Separation. *American Economic Journal: Macroeconomics 4*(4), 68–93.

Galeotti, M., L. J. Maccini, and F. Schiantarelli (2005, April). Inventories, employment and hours. *Journal of Monetary Economics 52*(3), 575–600.

Gan, L., M. D. Hurd, and D. L. McFadden (2005, October). Individual Subjective Survival Curves. In *Analyses in the Economics of Aging*, NBER Chapters, pp. 377–412. National Bureau of Economic Research, Inc.

Gathmann, C. and U. Schönberg (2010, January). How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics 28*(1), 1–49.

Geel, R. and U. Backes-Gellner (2009, September). Occupational Mobility Within and Between Skill Clusters: An Empirical Analysis Based on the Skill-Weights Approach. Economics of Education Working Paper Series 0047, University of Zurich, Institute for Strategy and Business Economics (ISU).

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian Data Analysis* (Third ed.). Taylor and Francis Group, LLC.

Geweke, J. and M. Keane (2000). An empirical analysis of earnings dynamics among men in the PSID: 1968-1989. *Journal of Econometrics 96*, 293–356.

Gilboa, I. and M. Marinacci (2011). Ambiguity and the Bayesian Paradigm. Working Papers 379, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Goos, M. and A. Manning (2007, February). Lousy and Lovely Jobs: The Rising Polarization of Work in Britain. *The Review of Economics and Statistics 89*(1), 118–133.

Gottschalk, P. and M. Huynh (2005, May). Validation Study Of Earnings Data in the SIPP– Do Older Workers Have Larger Measurement Error? Working Papers, Center for Retirement Research at Boston College 2005-07, Center for Retirement Research.

Gouret, F. and G. Hollard (2011, April). When Kahneman meets Manski: Using dual systems of reasoning to interpret subjective expectations of equity returns. *Journal of Applied Econometrics 26*(3), 371–392.

Griliches, Z. (1969, November). Capital-Skill Complementarity. *The Review of Economics and Statistics 51*(4), 465–68.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2015, January). What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk? NBER Working Papers 20913, National Bureau of Economic Research, Inc.

Hagedorn, M., F. Karahan, I. Manovskii, and K. Mitman (2013, October). Unemployment Benefits and Unemployment in the Great Recession: The Role of Macro Effects. Working Paper 19499, National Bureau of Economic Research.

Hagedorn, M. and I. Manovskii (2008, September). The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited. *American Economic Review 98*(4), 1692–1706.

Hall, R. E. (1995). Lost Jobs. *Brookings Papers on Economic Activity 26*(1, 25th A), 221–274.

Hall, R. E. (2005, March). Employment Fluctuations with Equilibrium Wage Stickiness. *American Economic Review 95*(1), 50–65.

Haltiwanger, J. C. and L. J. Maccini (1989, January). Inventories, Orders, Temporary and Permanent Layoffs: An Econometric Analysis. *Carnegie-Rochester Conference Series on Public Policy 30*(1), 301–366.

Hamermesh, D. S. (1987, April). The Demand for Labor in the Long Run. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 1 of *Handbook of Labor Economics*, Chapter 8, pp. 429–471. Elsevier.

Hamermesh, D. S. (1995). Labour Demand and the Source of Adjustment Costs. *The Economic Journal 105*(430), pp.620–634.

Harris, M. and A. Raviv (1993). Differences of Opinion Make a Horse Race. *Review of Financial Studies 6*(3), 473–506.

Hart, R. A. and J. Malley (1996, May). Excess Labour and the Business Cycle: A Comparative Study of Japan, Germany, the United Kingdom and the United States. *Economica 63*(250), 325–42.

Hausman, J. (2001, Fall). Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left. *Journal of Economic Perspectives 15*(4), 57–67.

Heckman, J. (1974). Shadow Prices, Market Wages, and Labor Supply. *Econometrica 42*(4), 679–694.

Heckman, J., L. Lochner, and C. Taber (1998, January). Explaining Rising Wage Inequality: Explanations With A Dynamic General Equilibrium Model of Labor Earnings With Heterogeneous Agents. *Review of Economic Dynamics 1*(1), 1–58.

Heckman, J. and B. Singer (1984, March). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica 52*(2), 271–320.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Chapter 7 Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In E. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Volume 1, pp. 307–458. Elsevier.

Heckman, J. J. and S. Urzua (2009, January). Comparing IV with Structural Models: What Simple IV Can and Cannot Identify. IZA Discussion Papers 3980, Institute for the Study of Labor (IZA).

Hijzen, A., R. Upward, and P. W. Wright (2010). The Income Losses of Displaced Workers. *Journal of Human Resources 45*(1).

Hill, D., M. Perry, and R. J. Willis (2004). Estimating Knightian Uncertainty from Survival Probability Questions on the HRS.

Hobijn, B. (2012). The industry-occupation mix of U.S. job openings and hires. Working Paper Series 2012-09, Federal Reserve Bank of San Francisco.

Hoderlein, S. and J. Winter (2009, January). Structural Measurement Errors in Nonseparable Models. Boston College Working Papers in Economics 750, Boston College Department of Economics.

Holzer, H. J. (1996). *What Employers Want: Job Prospects for Less-Educated Workers*. New York: Russell Sage Foundation.

Hong, H. and J. C. Stein (2007). Disagreement and the Stock Market. *Journal of Economic Perspectives 21*(2), 109–128.

Hospido, L. (2012, April). Modelling heterogeneity and dynamics in the volatility of individual wages. *Journal of Applied Econometrics 27*(3), 386–414.

Hospido, L. (2014). Wage Dynamics in the Presence of Unobserved Individual and Job Heterogeneity. Technical report, Bank of Spain and IZA.

Hoynes, H. (1999, June). The Employment, Earnings, and Income of Less Skilled Workers Over the Business Cycle. NBER Working Papers 7188, National Bureau of Economic Research, Inc.

Hoynes, H., D. L. Miller, and J. Schaller (2012, Summer). Who Suffers during Recessions? *Journal of Economic Perspectives 26*(3), 27–48.

Hudomiet, P. (2014). The Outs and Ins of Occupational Unemployment. What Can Explain the Symmetry in Hiring Rates? University of Michigan.

Hudomiet, P., G. Kézdi, and R. J. Willis (2011, April). Stock market crash and expectations of American households. *Journal of Applied Econometrics 26*(3), 393–415.

Hudomiet, P. and R. J. Willis (2013). Estimating Second Order Probability Beliefs from Subjective Survival Data. *Decision Analysis 10*(2), 152–170.

Hurd, M., M. V. Rooij, and J. Winter (2011, April). Stock market expectations of Dutch households. *Journal of Applied Econometrics 26*(3), 416–436.

Hurd, M. D. (2009, May). Subjective Probabilities in Household Surveys. *Annual Review of Economics 1*(1), 543–564.

Hurd, M. D. and K. McGarry (1995). Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study. *The Journal of Human Resources 30*, pp. S268–S292.

Hurd, M. D. and K. McGarry (2002, October). The Predictive Validity of Subjective Probabilities of Survival. *Economic Journal 112*(482), 966–985.

Hurd, M. D., S. Rohwedder, and J. Winter (2005). Subjective Probabilities of Survival: An International Comparison. Unpublished manuscript, RAND, Santa Monica, CA.

Hurd, M. D., J. P. Smith, and J. M. Zissimopoulos (2004). The effects of subjective survival on retirement and Social Security claiming. *Journal of Applied Econometrics 19*(6), 761–775.

Hurst, E., G. Li, and B. Pugsley (2010, December). Are Household Surveys Like Tax Forms: Evidence from Income Underreporting of the Self-Employed. Working paper, Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.

Hurst, E., G. Li, and B. Pugsley (2014). Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed. *The Review of Economics and Statistics 96*(1), 19–33.

Hyslop, D. R. and G. W. Imbens (2001, October). Bias from Classical and Other Forms of Measurement Error. *Journal of Business & Economic Statistics 19*(4), 475–81.

Imbens, G. M. and J. M. Wooldridge (2008, August). Recent Developments in the Econometrics of Program Evaluation. Working Paper 14251, National Bureau of Economic Research.

Imbens, G. W. (2009, April). Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009). Working Paper 14896, National Bureau of Economic Research.

Ingram, B. F. and G. R. Neumann (2006, February). The Returns to Skill. *Labour Economics 13*(1), 35–59.

Iversen, T. and A. Wren (1998). Equality, Employment, and Budgetary Restraint: The Trilemma of the Service Economy. *World Politics 50*(4), pp.507–546.

Jäckle, A., E. Sala, S. P. Jenkins, and P. Lynn (2005). Validation of Survey Data on Income and Employment: The ISMIE Experience. Discussion Papers of DIW Berlin 488, DIW Berlin, German Institute for Economic Research.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993, September). Earnings Losses of Displaced Workers. *American Economic Review 83*(4), 685–709.

Jaimovich, N. and H. E. Siu (2009, June). The Young, the Old, and the Restless: Demographics and Business Cycle Volatility. *American Economic Review 99*(3), 804–26.

Jaimovich, N. and H. E. Siu (2012, August). The Trend is the Cycle: Job Polarization and Jobless Recoveries. NBER Working Papers 18334, National Bureau of Economic Research, Inc.

James, J. (2011). Ability Matching and Occupational Choice. Working Paper 1125, Federal Reserve Bank of Cleveland.

Jovanovic, B. (1979, October). Job Matching and the Theory of Turnover. *Journal of Political Economy 87*(5), 972–90.

Kambourov, G. and I. Manovskii (2009, February). Occupational Specificity Of Human Capital. *International Economic Review 50*(1), 63–115.

Kandel, E. and N. D. Pearson (1995). Differential Interpretation of Public Signals and Trade in Speculative Markets. *Journal of Political Economy 103*(4), 831–72.

Kapteyn, A. and J. Y. Ypma (2007). Measurement Error and Misclassification: A Comparison of Survey and Administrative Data. *Journal of Labor Economics 25*, 513–551.

Keane, M. and E. Prasad (1993, December). Skill Levels and the Cyclical Variability of Employment, Hours, and Wages. *IMF Staff Papers 40*(4), 711–743.

Khalifa, S. (2009, April). Heterogeneous Workers and Occupations: Inequality, Unemployment, and Crowding Out. *Southern Economic Journal 75*(4), 1141–1164.

Khwaja, A. (2010, May). Estimating Willingness to Pay for Medicare Using a Dynamic Life-cycle Model of Demand for Health Insurance. *Journal of Econometrics 156*(1), 130–147.

Killingsworth, M. and J. J. Heckman (1986). Female Labor Supply: A Survey. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics* (1 ed.), Volume 1, Chapter 02, pp. 103–204. Elsevier.

Kim, Y. (2005, August). Audit Misperception, Tax Compliance, and Optimal Uncertainty. *Journal of Public Economic Theory 7*(3), 521–541.

Kletzer, L. G. (1998, Winter). Job Displacement. *Journal of Economic Perspectives 12*(1), 115–136.

Kondor, P. (2009). The more we know, the less we agree: Higher-order expectations and public announcements. 2009 Meeting Papers 1018, Society for Economic Dynamics.

Kristensen, N. and N. Westergaard-Nielsen (2006, September). A Large-Scale Validation

Study of Measurement Errors in Longitudinal Survey Data. IZA Discussion Papers 2329, Institute for the Study of Labor (IZA).

Krolikowski, P. (2014, September). Job Heterogeneity and Aggregate Labor Market Fluctuations. Unpublished manuscript.

Krusell, P., L. E. Ohanian, J.-V. Rios-Rull, and G. L. Violante (2000, September). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica 68*(5), 1029–1054.

Kydland, F. E. (1984, January). Labor-force heterogeneity and the business cycle. *Carnegie-Rochester Conference Series on Public Policy 21*(1), 173–208.

Kézdi, G. and R. J. Willis (2008). Stock Market Expectations and Portfolio Choice of American Households. University of Michigan.

Lazear, E. P. (1990, August). Job Security Provisions and Employment. *The Quarterly Journal of Economics 105*(3), 699–726.

Lazear, E. P. and J. R. Spletzer (2012, September). The United States Labor Market: Status Quo or A New Normal? Working Paper 18386, National Bureau of Economic Research.

Lee, D. S. and T. Lemieux (2009, February). Regression Discontinuity Designs in Economics. Working Paper 14723, National Bureau of Economic Research.

Lemieux, T., W. B. MacLeod, and D. Parent (2014, February). Performance Pay and Earnings Dynamics. Unpublished manuscript.

Lerman, R. I., S.-M. McKernan, and S. Riegg. (2004). The Scope of Employer-Provided Training in the United States: Who, What, Where, and How Much? In R. A. S. Christopher J. O'Leary and S. A. Wandner (Eds.), *Job Training Policy in the United States*. W.E. Upjohn Institute for Employment Research.

Leuven, E. (2005, February). The Economics of Private Sector Training: A Survey of the Literature. *Journal of Economic Surveys 19*(1), 91–111.

Lilien, D. M. and R. E. Hall (1987). Cyclical fluctuations in the labor market. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 2 of *Handbook of Labor Economics*, Chapter 17, pp. 1001–1035. Elsevier.

Lillard, L. and R. J. Willis (2001, June). Cognition and Wealth: The Importance of Probabilistic Thinking. Working Papers wp007, University of Michigan, Michigan Retirement Research Center.

Lillard, L. A. (1999). Job Turnover Heterogeneity and Person-Job-Specific Time-Series Wages. *Annales d'Economie et de Statistique* (55-56), 183–210.

Lillard, L. A. and H. W. Tan (1986). Private Sector Training: Who Gets It and What Are Its Effects? Technical report, RAND, Santa Monica, CA.

Lillard, L. A. and R. J. Willis (1978, September). Dynamic Aspects of Earning Mobility. *Econometrica 46*(5), 985–1012.

Little, R. J. A. and D. Rubin (2002). *Statistical Analysis with Missing Data* (Second ed.). John Wiley and Sons, Inc.

Liu, T. and L. C. Spector (2005, January). Dynamic employment adjustments over business cycles. *Empirical Economics 30*(1), 151–169.

Ljungqvist, L. (2002, October). How Do Lay–off Costs Affect Employment? *Economic Journal 112*(482), 829–853.

Lumsdaine, R. L. and R. J. P. van Loon (2013, June). Wall Street vs. Main Street: An Evaluation of Probabilities. Working Paper 19103, National Bureau of Economic Research.

MaCurdy, T. E. (1982, January). 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.

Manning, A. (2011). *Imperfect Competition in the Labor Market*, Volume 4 of *Handbook of Labor Economics*, Chapter 11, pp. 973–1041. Elsevier.

Manski, C. F. (2004). Measuring Expectations. *Econometrica 72*(5), 1329–1376.

Manski, C. F. and F. Molinari (2010). Rounding Probabilistic Expectations in Surveys. *Journal of Business and Economic Statistics 28*(2), 219–231.

Marchetti, D. J. and F. Nucci (2001, November). Labor Effort over the Business Cycle. Temi di discussione (Economic working papers) 424, Bank of Italy, Economic Research and International Relations Area.

Mark R. Rosenzweig, K. I. W. (2000, December). Natural "Natural Experiments" in Economics. *Journal of Economic Literature 38*(4), 827–874.

Matzkin, R. L. (2007, November). Nonparametric Survey Response Errors. *International Economic Review 48*(4), 1411–1427.

McArdle, J. J., J. P. Smith, and R. Willis (2009, August). Cognition and Economic Outcomes in the Health and Retirement Survey. Working Paper 15266, National Bureau of Economic Research.

McFadden, D. (2006). How Consumers Respond to Incentives. Technical report, UC Berkeley. Working Paper, UC Berkeley.

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

Meyer, B. D. (1990, July). Unemployment Insurance and Unemployment Spells. *Econometrica 58*(4), 757–82.

Michaels, G., A. Natraj, and J. V. Reenen (2010, June). Has ICT Polarized Skill Demand? Evidence from Eleven Countries over 25 years. Working Paper 16138, National Bureau of Economic Research.

Michaillat, P. (2010, November). Do Matching Frictions Explain Unemployment? Not in Bad Times. CEP Discussion Papers dp1024, Centre for Economic Performance, LSE.

Michaillat, P. (2012, June). Do Matching Frictions Explain Unemployment? Not in Bad Times. *American Economic Review 102*(4), 1721–50.

Miller, R. L. (1971, March). The Reserve Labour Hypothesis: Some Tests of its Implications. *Economic Journal 81*(321), 17–35.

Mincer, J. (1991, September). Education and Unemployment. NBER Working Papers 3838, National Bureau of Economic Research, Inc.

Miyamoto, H. (2011, July). Efficiency in a search and matching model with training costs. *Economic Modelling 28*(4), 1838–1841.

Moffitt, R. A. (1999). Models of Treatment Effects When Responses Are Heterogeneous. *Proceedings of the National Academy of Sciences of the United States of America 96*(12), 6575–6576.

Mortensen, D. and E. Nagypal (2007). Labor-market Volatility in Matching Models with Endogenous Separations. *The Scandinavian Journal of Economics 109*(4), 645–665.

Mortensen, D. T. and C. A. Pissarides (1994, July). Job Creation and Job Destruction in the Theory of Unemployment. *Review of Economic Studies 61*(3), 397–415.

Mortensen, D. T. and C. A. Pissarides (1999a). Job Reallocation, Employment Fluctuations and Unemployment. In J. B. Taylor and M. Woodford (Eds.), *Handbook of Macroeconomics*, Volume 1 of *Handbook of Macroeconomics*, Chapter 18, pp. 1171–1228. Elsevier.

Mortensen, D. T. and C. A. Pissarides (1999b, April). Unemployment Responses to 'Skill-Biased' Technology Shocks: The Role of Labour Market Policy. *Economic Journal 109*(455), 242–65.

Moscarini, G. (2001, July). Excess Worker Reallocation. *Review of Economic Studies 68*(3), 593–612.

Moscarini, G. (2005, March). Job Matching and the Wage Distribution. *Econometrica 73*(2), 481–516.

Moscarini, G. and K. Thomsson (2008). Occupational and Job Mobility in the US. *Scandinavian Journal of Economics, Vol. 109, Issue 4, pp. 807-836, March 2008*.

Moscarini, G. and F. Vella (2008, February). Occupational Mobility and the Business Cycle. IZA Discussion Papers 3369, Institute for the Study of Labor (IZA).

Nadiri, M. I. and S. Rosen (1969, Part I Se). Interrelated Factor Demand Functions. *American Economic Review 59*(4), 457–71.

Neumark, D. and W. Wascher (2006, November). Minimum Wages and Employment: A Review of Evidence from the New Minimum Wage Research. NBER Working Papers 12663, National Bureau of Economic Research, Inc.

Nickell, S. (1997, Summer). Unemployment and Labor Market Rigidities: Europe versus North America. *Journal of Economic Perspectives 11*(3), 55–74.

Nordberg, L., I. Penttilä, and S. Sandström (2001, November). A study on the effects of using interview versus register data in income distribution analysis with an application to the Finnish ECHP-survey in 1996. Working paper 1, Chintex WP 5, Statistics Finland.

Oesch, D. (2010). What explains high unemployment among low-skilled workers? Evidence from 21 OECD countries. *European Journal of Industrial Relations 16*(1), 39–55.

Ofstedal, M. B., G. G. Fisher, and A. R. Herzog (2005). Documentation of Cognitive Functioning Measures in the Health and Retirement Study. Documentation Report DR-006, HRS.

Oi, W. Y. (1962). Labor as a Quasi-Fixed Factor. *Journal of Political Economy 70*(6), 538–555.

Olsen, A. and R. Hudson (2009). Social Security Administration's Master Earnings File: Background Information. *Social Security Bulletin 69*(3).

Oyer, P. and S. Schaefer (2011). Chapter 20 - Personnel Economics: Hiring and Incentives. Volume 4, Part B of *Handbook of Labor Economics*, pp. 1769–1823. Elsevier.

Parro, F. (2013, April). Capital-Skill Complementarity and the Skill Premium in a Quantitative Model of Trade. *American Economic Journal: Macroeconomics 5*(2), 72–117.

Paté-Cornell, M. E. (1996). Uncertainties in Risk Analysis: Six Levels of Treatment. *Reliability Engineering & System Safety 54*(2-3), 95 – 111. Treatment of Aleatory and Epistemic Uncertainty.

Pencavel, J. (1986). Labor Supply of Men: A Survey. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics* (1 ed.), Volume 1, Chapter 01, pp. 3–102. Elsevier.

Perozek, M. (2008, February). Using Subjective Expectations to Forecast Longevity: Do Durvey Respondents Know Something We Do not Know? *Demography 45*(1), 95–113.

Pfann, G. A. (2006, February). Downsizing and Heterogeneous Firing Costs. *The Review of Economics and Statistics 88*(1), 158–170.

Pfann, G. A. and F. C. Palm (1993, April). Asymmetric Adjustment Costs in Non-linear Labour Demand Models for the Netherlands and U.K. Manufacturing Sectors. *Review of Economic Studies 60*(2), 397–412.

Philipson, T. (1997, January). Data Markets and the Production of Surveys. *Review of Economic Studies 64*(1), 47–72.

Philipson, T. (2001, July). Data Markets, Missing Data, and Incentive Pay. *Econometrica 69*(4), 1099–1111.

Picone, G., F. Sloan, and D. Taylor (2004, January). Effects of Risk and Time Preference and Expected Longevity on Demand for Medical Tests. *Journal of Risk and Uncertainty 28*(1), 39–53.

Pischke, J.-S. (1995, July). Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study. *Journal of Business & Economic Statistics 13*(3), 305–14.

Pissarides, C. A. (2009, September). The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer? *Econometrica 77*(5), 1339–1369.

Pissarides, C. A. and B. Petrongolo (2001, June). Looking into the Black Box: A Survey of the Matching Function. *Journal of Economic Literature 39*(2), 390–431.

Platt, H. and M. B. P. Platt (2011). Revisiting the Labor Hoarding Employment Demand Model: An Economic Order Quantity Approach. *Journal of Financial Transformation 31*, 158–163.

Poletaev, M. and C. Robinson (2008, July). Human Capital Specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984-2000. *Journal of Labor Economics 26*(3), 387–420.

Portugal, P. and O. Blanchard (2001, March). What Hides Behind an Unemployment Rate: Comparing Portuguese and U.S. Labor Markets. *American Economic Review 91*(1), 187–207.

Poterba, J. M. and L. H. Summers (1986, November). Reporting Errors and Labor Market Dynamics. *Econometrica 54*(6), 1319–38.

Pounder, L. (2007). *Life Cycle Consumption Examined*. Ph. D. thesis.

Pries, M. and R. Rogerson (2005, August). Hiring Policies, Labor Market Institutions, and Labor Market Flows. *Journal of Political Economy 113*(4), 811–839.

Ratner, D. (2013, December). Unemployment Insurance Experience Rating and Labor Market Dynamics. Technical Report 2013-86, Finance and Economics Discussion Series, Federal Reserve Board, Washington, D.C.

Reder, M. W. (1955). The Theory of Occupational Wage Differentials. *The American Economic Review 45*(5), pp.833–852.

Reicher, C. (2011, July). The Aggregate Effects of Long Run Sectoral Reallocation. Kiel Working Papers 1720, Kiel Institute for the World Economy.

Roback, J. (1982). Wages, Rents, and the Quality of Life. *The Journal of Political Economy 90*(6), pp.1257–1278.

Robinson, C. (2011). Occupational Mobility, Occupation Distance and Specific Human Capital. University of Western Ontario, CIBC Centre for Human Capital and Productivity Working Papers 20115, University of Western Ontario, CIBC Centre for Human Capital and Productivity.

Roed, K., P. Jensen, and A. Thoursie (2002, May). Unemployment Duration, Incentives and Institutions - A Micro-Econometric Analysis Based on Scandinavian Data. Working Paper Series 3/2002, Swedish Institute for Social Research.

Roed, K. and T. Zhang (2003, January). Does Unemployment Compensation Affect Unemployment Duration? *Economic Journal 113*(484), 190–206.

Rogerson, R. and R. Shimer (2011). *Search in Macroeconomic Models of the Labor Market*, Volume 4 of *Handbook of Labor Economics*, Chapter 7, pp. Pages: 61. Elsevier.

Rogerson, R., R. Shimer, and R. Wright (2005). Search-Theoretic Models of the Labor Market: A Survey. *Journal of Economic Literature 43*(4), pp.959–988.

Rosen, S. (1987). The Theory of Equalizing Differences. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics* (1 ed.), Volume 1, Chapter 12, pp. 641–692. Elsevier.

Rossana, R. J. (1990, February). Interrelated Demands for Buffer Stocks and Productive Inputs: Estimates for Two-Digit Manufacturing Industries. *The Review of Economics and Statistics 72*(1), 19–29.

Rotemberg, J. J. and L. H. Summers (1990). Inflexible Prices and Procyclical Productivity. *The Quarterly Journal of Economics 105*(4), pp.851–874.

Ruhm, C. J. (1991, March). Are Workers Permanently Scarred by Job Displacements? *American Economic Review 81*(1), 319–24.

Sahin, A., J. Song, G. Topa, and G. L. Violante (2012, August). Mismatch Unemployment. NBER Working Papers 18265, National Bureau of Economic Research, Inc.

Sahm, C. R. (2007). Stability of Risk Preference. Finance and Economics Discussion Series 2007-66, Board of Governors of the Federal Reserve System (U.S.).

Saint-Paul, G. (2004, Fall). Why are European Countries Diverging in their Unemployment Experience? *Journal of Economic Perspectives 18*(4), 49–68.

Sakellaris, P. and A. Spilimbergo (1999, September). Business Cycles and Investment in Human Capital: International Evidence on Higher Education. Electronic Working Papers 99-009, University of Maryland, Department of Economics.

Savage, L. J. (1954). *The Foundations of Statistics.* New York, Wiley.

Sbordone, A. M. (1996, October). Cyclical Productivity in a Model of Labor Hoarding. *Journal of Monetary Economics 38*(2), 331–361.

Schmieder, J. F., T. v. Wachter, and S. Bender (2010, February). The Effects of Unemployment Insurance on Labor Supply and Search Outcomes: Regression Discontinuity Estimates from Germany. IAB Discussion Paper 201004, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].

Shapiro, M. D. (1986). The Dynamic Demand for Capital and Labor. *The Quarterly Journal of Economics 101*(3), pp.513–542.

Shimer, R. (2005, March). The Cyclical Behavior of Equilibrium Unemployment and Vacancies. *American Economic Review 95*(1), 25–49.

Shimer, R. (2007, September). Reassessing the Ins and Outs of Unemployment. NBER Working Papers 13421, National Bureau of Economic Research, Inc.

Shimer, R. (2012a, April). Reassessing the Ins and Outs of Unemployment. *Review of Economic Dynamics 15*(2), 127–148.

Shimer, R. (2012b). Wage rigidities and jobless recoveries. *Journal of Monetary Economics 59*(S), S65–S77.

Siebert, H. (1997, Summer). Labor Market Rigidities: At the Root of Unemployment in Europe. *Journal of Economic Perspectives 11*(3), 37–54.

Silva, J. I. and M. Toledo (2009, May). Labor Turnover Costs and the Cyclical Behavior of Vacancies and Unemployment. *Macroeconomic Dynamics 13*(S1), 76–96.

Sloan, F., L. Eldred, T. Guo, and Y. Xu (2013, August). Are people overoptimistic about the effects of heavy drinking? *Journal of Risk and Uncertainty 47*(1), 93–127.

Smith, C. L. (2013). The Dynamics of Labor Market Polarization. Unpublished manuscript.

Smith, V. K., D. H. Taylor, and F. A. Sloan (2001, September). Longevity Expectations and Death: Can People Predict Their Own Demise? *American Economic Review 91*(4), 1126–1134.

Solon, G., W. Whatley, and A. H. Stevens (1997, April). Wage Changes and Intrafirm Job Mobility over the Business Cycle: Two Case Studies. *Industrial and Labor Relations Review 50*(3), 402–415.

Spitz-Oener, A. (2006, April). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics 24*(2), 235–270.

Stevens, A. H. (1997, January). Persistent Effects of Job Displacement: The Importance of Multiple Job Losses. *Journal of Labor Economics 15*(1), 165–88.

Teulings, C. N. (1993). The Diverging Effects of the Business Cycle on the Expected Duration of Job-Search. *Oxford Economic Papers 45*(3), pp.482–500.

Thompson, C. P., J. J. Skowronski, S. F. Larsen, and A. L. Betz (1996). *Autobiographical Memory*. Lawrence Erlbaum Associates.

Topel, R. H. (1982, September). Inventories, Layoffs, and the Short-Run Demand for Labor. *American Economic Review 72*(4), 769–87.

Topel, R. H. (1991, February). Specific Capital, Mobility, and Wages: Wages Rise with Job Seniority. *Journal of Political Economy 99*(1), 145–76.

Tourangeau, R., L. J. Reps, and K. Rasinski (2000). *The Psychology of Survey Response*. Cambridge University Press.

Uren, L. and G. Virag (2011, May). Skill Requirements, Search Frictions, And Wage Inequality. *International Economic Review 52*(2), 379–406.

van Ours, J. C. and G. Ridder (1995). Job Matching and Job Competition: Are Lower Educated Workers at the Back of the Job Queues? *European Economic Review 39*, 1717–1731.

Van Soest, A. H. and A. Kapteyn (2009). Mode and Context Effects in Measuring Household Assets. *SSRN eLibrary*.

Vaupel, J., K. Manton, and E. Stallard (1979, August). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography 16*(3), 439–454.

Vaupel, J. W. (1997). Trajectories of Mortality at Advanced Ages. In W. KW and F. CE (Eds.), *Between Zeus and the Salmon*. National Academy Press, Washington, D.C.

Vissing-Jorgensen, A. (2004, October). Perspectives on Behavioral Finance: Does Irrationality Disappear with Wealth? Evidence from Expectations and Actions. In *NBER Macroeconomics Annual 2003, Volume 18*, NBER Chapters, pp. 139–208. National Bureau of Economic Research, Inc.

von Wachter, T., E. Handwerker, and A. Hildreth (2009, June). Estimating the T̈rueC̈ost of Job Loss: Evidence Using Matched Data from Califormia 1991-2000. Working Papers 09-14, Center for Economic Studies, U.S. Census Bureau.

Wagenaar, W. A. (1986). My Memory: A Study of Autobiographical Memory over Six Years. *Cognitive Psychology 18*(2), 225–252.

Weitzman, M. L. (2007, September). Subjective Expectations and Asset-Return Puzzles. *American Economic Review 97*(4), 1102–1130.

Wen, Y. (2005). Labor Hoarding and Inventories. Working Papers 2005-040, Federal Reserve Bank of St. Louis.

Willis, R. J. (1987, April). Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 1 of *Handbook of Labor Economics*, Chapter 10, pp. 525–602. Elsevier.

Willis, R. J. (2005). Discussion of Gan L, Hurd MD, McFadden D ”Individual Subjective Survival Curves”. In D. Wise (Ed.), *Analyses in the Economics of Aging*. University of Chicago Press.

Wolff, J. and A. Launov (2004). Unemployment Benefits and Unemployment Rates of Low-Skilled and Old Workers in West Germany: A Search Equilibrium Approach. 2004 Meeting Papers 395, Society for Economic Dynamics.

Yamaguchi, S. (2012). Tasks and Heterogeneous Human Capital. *Journal of Labor Economics 30*(1), pp.1–53.