working paper

2205

Income, Employment and Health Risks of Older Workers

Siqi Wei

July 2022

CENTRO DE ESTUDIOS
MONETARIOS Y FINANCIEROS
Casado del Alisal 5, 28014 Madrid, Spain
www.cemfi.es

CEMFI Working Paper No. 2205
July 2022

Income, Employment and Health Risks of Older Workers

Abstract

This paper begins with the observation that many olderworkers move to "bridge" jobs with lower wages and fewer working hours before exiting the labor force for good. To explain this gradual transition to full retirement, I propose a nonlinear agingrelated shock — mismatch shock, which mismatches workers with their existing job and triggers job leaves. I develop an empirical framework of employment and job transitions jointly with stochastic wage and hour processes to separate health risks, individual-specific productivity risks, firm-specific mismatch risks, quality of outside offers, and job destruction risks faced by older workers. The model is estimated with a sample of male individuals aged 51 to 70 in the US Health and Retirement Study applying a novel parameter-expanded stochastic EM algorithm. The paper finds that mismatch shocks play an important role in explaining the reduction in wages and hours for movers. Furthermore, I calculate the welfare cost of risks and quantify how much individuals value the possibility of a flexible transition to full retirement by constructing a utility-based structural model of consumption, employment and job movements where agents face the same risks as in the empirical model. The model is estimated using a novel simulation-based algorithm that exploits the connection to the empirical model and the estimates from the empirical model. The results show that the median cost of mismatch risks amounts to a reduction in consumption flow by 5?3%-7?1% depending on the education group. Banning job changes and re-entry causes a welfare loss equivalent to a consumption drop of 12% 4%.

JEL Codes: J26, J24, J22, C51, I14.

Keywords: Income risks, health risks, mismatch, bridge jobs, latent variables.

Siqi Wei
CEMFI
siqi.wei@cemfi.edu.es

Acknowledgement

I am deeply indebted tomy advisor Manuel Arellano for his generous guidance, support, and encouragement during my study at CEMFI. I would also like to express special thanks to my thesis committee members, Dmitry Arkhangelsky, Pedro Mira, and Josep Pijoan-Mas for their valuable help during all stages of this project. Great comments have been received from Martin Almuzara, Jesús Bueren, Micole De Vera, Yan Hu, Liyang Sun, Jiayi Wen, and seminar participants at CEMFI. Funding from Spain’s Ministerio de Economía, Industria y Competitividad (BES-2017-082506) and María de Maeztu Programme for Units of Excellence in R&D (MDM-2016-0684) is gratefully acknowledged. All errors are my sole responsibility.

1 Introduction

Around 30% to 50% of older workers in the United States experience post-retirement employment before exiting the labor force for good (Alcover et al., 2014; Wang et al., 2014; Cahill and Quinn, 2020), and more than half of them move to jobs with lower wages and fewer working hours (Ameriks et al., 2020). This type of "gradual transition" to full retirement is important to study to understand employment dynamics near retirement age. It is also policy-relevant because many countries have considered retirement policy reforms, including increasing the statutory retirement age to relieve the pressures on the pension system, as population aging is becoming an increasingly large challenge.[1] Evaluation of the effectiveness and welfare effects of these policies requires researchers to understand how individuals make retirement decisions, which further requires understanding the risks that older people face and how those risks interact with their retirement decisions.

The interest in the different types of risks that individuals face is not only limited to policy evaluation. Indeed, the study of the risks that older workers face is essential for understanding income dynamics near retirement. However, the identification of risks requires controlling for endogenous choices such as employment decisions. This is so because individuals may react to different shocks by making contingent choices. For example, individuals who receive a firm-specific productivity shock that worsens the match with their job may move to a different one. The total variation in the observed wage changes is therefore the result of both exogenous shocks and endogenous choices.[2] Consequently, taking those employment choices into account is crucial in distinguishing exogenous risks from observed earning variation.

This paper studies the risks that older people face and their gradual transition to full retirement. First, using Health and Retirement Study (HRS) data, I document the left-skewness in the wage and hour changes for job movers. To explain this type of movement to worse-paying and less-demanding jobs, I propose an aging-related shock, *mismatch shock*, that mismatches workers with their existing jobs and triggers job leave. One possible interpretation is the stamina decreases: as people age, it becomes more difficult to work and focus for a long period of time, which causes productivity drops and stress,

[1]Reported in World Population Ageing 2019 by Population Division of the United Nations Department of Economic and Social Affairs, the number of 65+ persons in the world is projected to double to 1.5 billion in 2050, and by the end of this century, there will be 155 countries with more than one fifth of the population older than 65 years old

[2]Low et al. (2010) obtain smaller variance of permanent shocks adding endogenous choices in employment and job.

and triggers movement to worse-paying but less-demanding jobs. Following the income dynamics literature, I further build a latent-variable wage and hour processes jointly with flexible employment and job transition model. This empirical model is flexible but interpretable, and allows us to disentangle health risks, individual-specific productivity risks, firm-specific mismatch risks, quality of outside offers, and job destruction risks from endogenous choices. I study the role of different shocks in explaining job movement to worse-paying and less-demanding jobs. Finally, I calculate the welfare cost of different risks and the value of the smooth transition to full retirement by building a utility-based model where agents make consumption and employment choices facing the same set of risks, identified from my flexible model.

Using a sample of male individuals aged between 50 to 70 in the HRS data, I start by documenting statistics on employment, wage, and working hours for older workers in the US. Descriptive statistics show that in each period, around 11% of individuals start working in a new job. Workers leave the previous job for different reasons, including business closure (6.9%), lay-offs (20.3%), poor health (13.5), better job (7.1%), job quits (9.6%), and retirement (39.4%), each followed by different employment status in later waves. For those who continue working with a different employer, most of them face wage and hour decreases, leading to a left-skewed distribution of wage and hour changes.

I then build a panel data-based estimation framework of wages, hours, employment, and job dynamics. The model consists of three blocks. The first block is stochastic wage and hour equations. The wage equation consists of individual's characteristics, latent health, unobserved fixed effect, general productivity, firm-specific productivity, and measurement error.[3] I specify the hour equation in a parallel way allowing for a firm-specific working-hour component.

The second block includes equations describing the uncertainty people face. Specifically, the model includes a rich set of risks associated with which are shocks to health, general productivity, firm-specific productivity, and employment (i.e., job destruction). Additionally, individuals also receive a random outside job offer, a package of firm-specific wage and hour components, in each period. In particular, I define mismatch shock as an aging-related discrete shock that is connected with changes in firm-specific productivity and job leave. Moreover, it can also have a persistent effect on future job offers.

[3]Individual characteristics include age, education and race

The last block of the model is a list of equations describing the employment and job transitions between periods. Possible transitions for workers include becoming non-employed, staying at the same job, and starting a new job. The movements are depicted by employment-to-employment and job change equations, which are affected by the quality of offers, their current wages, health, and wealth. In contrast, there are only two paths possibly observed for those who were non-employed: either remaining non-employed or moving to a new job. The probability of being employed is captured by the non-employment-to-employment equation. All these equations involve latent variables, are flexible and empirically rich. They can be seen as approximate reduced form rules of a class of more restrictive utility-based structural models.

In this empirical model, mismatch shock is expected to be a key component to explain the gradual transition to full retirement. This is because those who suffer mismatch shocks are more likely to receive worse job offers and at the same time more likely to accept a worse offer due to reservation wage changes. The 0-1 discrete format therefore creates a natural asymmetry and left-skewness in wage and hour changes. Mismatch shocks also provide a possible justification for non-linear persistence (as in Arellano et al., 2017): compared with receiving some marginal shocks to general productivity, individuals who receive mismatch shocks and start a new job lose all the accumulated tenure, which potentially corresponds to a much smaller persistence of wage history.[4] Intuitively, shocks to general productivity are identified through job stayers whereas mismatch shocks are identified through job movers especially those with non-marginal wage changes.

As an extension, I study the welfare cost of these risks and quantify how much individuals value the possibility of a flexible transition to full retirement. To do this, I construct and estimate a utility-based structural model of consumption, employment, and job movements. Agents in the model derive utility from consumption and leisure and pay extra costs for working, re-entry, and changing jobs depending on their age and health. They make decisions on consumption and saving, labor supply, and job changes by optimizing the expected discounted utility facing the same health, income, and employment risks as in the non-structural empirical model. This structural model can be seen as a more restrictive version of the empirical model, which serves as a

[4]Friedrich et al. (2019) has similar similar framework and allows for firm-level shocks identified from matched employer-employee data. The difference is that they use Swedish data, they do not specifically focus on older people, and they use earning data for the lack of hour measurement.

tool for measuring the welfare changes in different settings. Specifically, I conduct two counterfactual analyses. In the first exercise, I evaluate the welfare cost from the risks of being mismatched or losing jobs involuntarily by eliminating the risks and compare to the baseline model. In the second exercise, I evaluate the welfare loss from not being able to change jobs or re-enter after 65 years of age, which speaks of the value of bridge jobs. Two measures are constructed, where one is a lump-sum transfer of assets, and the other is consumption flows.

To estimate the empirical model, which contains various latent variables, I exploit the parameter-expanded stochastic EM algorithm (PX-SEM) studied in Wei (2021). PX-SEM is a simulation-based algorithm that combines the parameter expansion techniques in Liu et al. (1998) with the stochastic EM algorithm (Diebolt and Celeux, 1993), which is a variant of the original EM algorithm (Dempster et al., 1977). Similar to the standard SEM algorithm, PX-SEM alternates between an E-step, where we simulate draws from the posterior distribution of the latent components, and an M-step, where we update parameters. The difference is that in the M-step of PX-SEM, I exploit more robust estimators to the draws from the E-step by imposing some assumptions from the model itself, taking into account that the E-step draws might be simulated under values of the parameters that are far away from the optimum. This is achieved by applying parameter expansion techniques, which requires in the M-step: 1) expanding the original model by allowing for extra correlations between latent variables and observables and by adding extra parameters to relax the covariance matrix, 2) estimating the expanded model, and 3) reducing to the original model space while keeping the likelihood of observables unchanged. The PX-SEM method improves the algorithmic efficiency in estimating the empirical model which otherwise is intractable using standard SEM algorithm.

To estimate the structural model, I propose a new simulation-based algorithm that makes use of the estimates of the empirical model. Specifically, parameters are chosen such that the structural model best approximates the estimated empirical model. This method is different from Indirect Inference because it exploits the information on latent variables. Under the premise that the empirical model can be treated as a flexible approximation to the reduced form of the structural model, it improves both the algorithmic and statistical efficiency. Precisely, we can bring directly the risks from the empirical model to the structural one as input, and estimate the remaining parameters of the structural model more efficiently by comparing its simulations with the estimated empirical model instead of data.

This paper constructs two types of models to combine their advantages. The empirical model brings us more credible estimates of risks due to its flexibility; whereas the structural model allows us to do welfare analysis.[5] These two models are well-connected by both the nested structure and the new simulation-based algorithm.

The models are estimated using the 1996-2016 data from the Health and Retirement Study (HRS). Specifically, I focus on male individuals aged between 50 and 70. The estimation results show that our model can capture the dynamics in employment and job movements, as well as the left-skewness in the distribution of wage changes and hour changes among job movers. Mismatch shocks play an important role in explaining the job movements of older workers to jobs that pay less and/or require fewer working hours. Shutting down mismatch shocks implies that the mean wage drop of movers is reduced by around 55%, and the variance of wage changes goes down by 17% to 28%, depending on the age groups. This is mainly driven by the bottom percentiles, as $P10$ of wage changes increases by 17% and 23% for the younger and older group respectively, and the $P10$ of hours changes increases by 28% and 49%, whereas the $P90$ is almost unchanged in all cases. A similar pattern also applies to hour changes. Additionally, I look at bridge job movements and the associated wage and hour changes. The conclusion is that mismatch shocks are the main driver of bridge job movements associated with cuts in wages and hours.

Finally, the welfare calculation shows that the median cost of mismatch shocks amounts to a lump sum transfer of assets at age 55/56 of around \$62,300 for the high education group and \$26,700 for the low education group, or a reduction in consumption flow by 7.11% for the high education (HE) and 5.33% for the low education (LE). Individuals seem to value the possibility of a flexible transition at their older age. Imposing restrictions on job changes and re-entry causes a median loss of around \$107,300 for HE and \$58,400 for LE, which is equivalent to a reduction in consumption flow by 12% – 14%.

Literature and contribution. This paper is closely related to several strands of the literature. The first one is the vast literature on income dynamics. From the perspective of the identification of income risks, models can be classified into two types: univariate and multivariate (Arellano, 2014). Univariate models refer to those that use income history alone and identify the risks from income variation (Lillard and Willis, 1978; Hause, 1980;

[5] Structural models tend to have simplified risk dynamics due to dimensionality restriction. In addition, the model is more restrictive due to the parametric assumption.

MaCurdy, 1982; Chamberlain and Hirano, 1999). Compared to the canonical random-walk permanent transitory models, the recent literature has documented new features of income dynamics and accommodates them by further relaxation in different dimensions.[6] That being said, this paper belongs to the group of multivariate models that exploit not only income history but also other information, like working hours and employment history (Abowd and Card, 1989; Altonji et al., 2013; Low et al., 2010; Friedrich et al., 2019). By modelling a list of risks and their contingent decisions jointly, we can hope to distinguish exogenous risks from endogenous choices when modelling the total variation of income. Relative to other papers in this literature, a distinguishing feature of this paper is a specific focus on older workers. Related to this, another contribution is that the rich set of risks is intended to capture the risk specificities of older workers—in particular by allowing for mismatch shocks that capture firm-worker-level productivity declines and trigger the move of some workers to worse-paying jobs. Key papers in the multivariate literature, such as Altonji et al. (2013) and Low et al. (2010), do not include health shocks and the productivity of a match remains constant. Low and Pistaferri (2015) includes health shocks but no firm-level productivity. Friedrich et al. (2019) studies the pass-through of firm-level shocks to wages, and allows for variation of firm-level productivity but they do not focus on older workers.

Secondly, this paper is related to retirement studies (Gustman and Steinmeier, 1983; Rust and Phelan, 1997; Blau and Gilleskie, 2008; French, 2005; French and Jones, 2011, among others). In this literature, structural models are commonly used as tools to study labor participation changes and to conduct counterfactual policy analysis.[7] Fewer papers have focused on job dynamics and the corresponding patterns of wage change for older workers. Blau (1994) provides a descriptive analysis of employment dynamics for older workers. Berkovec and Stern (1991), to which this paper is closely related, model job movements as a result of shocks and choices. However, they do not allow for an individual persistent productivity component, ignore the effect of wealth in labor supply decisions,

[6]Guvenen et al. (2021) and the papers in the Global Income Dynamics project document the evolution of skewness and kurtosis of income and income changes using administrative data over the years in different countries. Arellano et al. (2017) shows the evidence of the nonlinear persistence of income dynamics with flexible quantile models. Browning et al. (2010) models earnings processes allowing for high-dimensional heterogeneity. Hospido (2012) allows for heterogeneity in the conditional variance of wages. Almuzara (2020) models both permanent and transitory income changes allowing for cross-sectional heterogeneity. Sanchez and Wellschmied (2020) models age-varying positive and negative earning shocks over the life-cycle and finds out the negative shocks dominates for workers between 50-55.

[7]Such as analyzing the role of social security system (French, 2005), Medicare (Blau and Gilleskie, 2008; French and Jones, 2011), and etc.

and assume perfect information of future health.[8] They do not focus on explaining the pattern of wage changes associated with job movements either. Recent work by Jacobs and Piyapromdee (2016) tries to explain the re-entry behavior by incorporating a burnout-recovery process to utility functions (Maestas and Li, 2007). This paper complements their study by trying to explain the dynamics from the risk side on the basis of a rich set of risks and paying attention to job-to-job changes in addition to revised retirement choices.[9]

A third strand to which this paper is related is the interdisciplinary literature on bridge jobs. The interest in bridge jobs draws from different social science, including economics (Ruhm, 1990; Maestas, 2010; Cahill et al., 2011, 2018; Cahill and Quinn, 2020; Ameriks et al., 2020; Brunello and Langella, 2013), psychology (Zhan et al., 2009; Wang et al., 2008), sociology (Han and Moen, 1999) and management (Kim and Feldman, 2000; Sullivan and Al Ariss, 2019). The detailed definition of bridge job varies according to disciplines, the main question under study, or the available datasets. This paper contributes to this literature by studying bridge jobs in an economic model that is not built around a narrow definition of bridge job. The bridge job movement manifests itself through the employment and job movements in the model. Furthermore, the model allows us to control for selection, explore quantitatively the reasons and outcomes of starting a bridge job, and conduct counterfactual analysis.

Additionally, this paper is also related to the literature on health dynamics and their impact on other outcomes. In terms of the modelling of health dynamics, this paper fits in the line of research that assumes health as an exogenous process (Amengual et al., 2017; Lange and McKee, 2012; Halliday, 2008; Heiss, 2011).[10] Once again, this paper specifically focuses on the impact on the labor supply of older workers (Blundell et al., 2016, 2017; French, 2005; French and Jones, 2011).[11] This paper contributes to the health literature by estimating a latent health process that interacts flexibly with other model components.

Finally, from a methodological point of view, this paper belongs to an expanding literature that considers the application of the EM algorithm (Dempster et al., 1977) and its variants as well as Indirect Inference in estimating latent variable models (Diebolt and

[8] They argue that wealth is partly captured by the non-wage utility flow, which is constant during the tenure and employment history has not effect on the random job-specific opportunities.

[9] Usually they distinguish full time and part time.

[10] Examples of endogenous health dynamics in which pecuniary and/or non-pecuniary investments are allowed include Grossman (1972); Hokayem and Ziliak (2014); Halliday et al. (2019); Margaris and Wallenius (2020).

[11] Research on saving includes Bueren (2018); De Nardi et al. (2010), and research on the effect over the life cycle include De Nardi et al. (2017); Capatina et al. (2018); Jolivet and Postel-Vinay (2020); Hosseini et al. (2021)

Celeux, 1993; Arcidiacono and Jones, 2003; Arellano and Bonhomme, 2016; Liu et al., 1998; Keane and Smith, 2003). This paper contributes to this literature in two ways. First, I apply the idea of parameter-expanded stochastic EM algorithm studied in a companion paper to the estimation of the nonlinear panel data model (Wei, 2021). Second, I propose a novel simulation-based estimation method that exploits the flexible reduced-from model in order to estimate the structural model.

Organization. The paper proceeds as follows. In section 2, I discuss the institutional background of the US labor market for workers near retirement, which also motivates the model specification. Section 3 presents the empirical model of earning dynamics. Section 4 explains the mapping between variables in the model and the data. In section 5, I discuss the estimation methods. In section 6, I present model fit diagnostics and empirical results. Next, I quantify the welfare cost of risks in section 7 by conducting counterfactual analysis in a utility-based model. Finally, Section 8 concludes.

2 Older Workers In HRS

For many older Americans, retirement does not mean a one-time and permanent withdrawal from the labor market. On the contrary, retirement is a gradual and dynamic process, accompanied by moving in and out of the labor market and job changes (Cahill et al., 2011; Cahill and Quinn, 2020; Ameriks et al., 2020). Jobs that take place between the career employment period and full exit from the labor force are often referred to as bridge jobs. In the rest of this section, I will present evidence related to employment dynamics of older workers in the US using the HRS dataset, which further inspires our modelling strategy.

HRS is a longitudinal panel study that surveys a representative sample of non-institutionalized individuals aged 50+ in the US.[12] It includes detailed information on health and economic characteristics and has been conducted every other year since 1992.[13] For consistency of the questionnaire, we use data from 1996 to 2016, a total of 11 waves. We present statistics for male individuals aged between 51 to 70. See Section 4 for further details of sample selection.

Table 1 shows that around 32% of HRS individuals between 51 to 70 make at least

[12]The Health and Retirement Study is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan.

[13]The exception is the AHEAD cohort whose second wave and third wave are in year 1993 and 1995 respectively.

one job change.[14] Two situations are further distinguished depending on whether there is a spell of non-employment between two jobs. We categorize a job change that happens between two consecutive waves as a job-to-job transition and a job movement with at least one wave of non-employment in between as a re-entry.[15] Around 21% of HRS individuals have experienced job-to-job transitions whereas fewer, around 14%, have experienced re-entry. Measures based on individual-year observation show that 4% of the non-employed transition to employment, and among those who stay employed, 11% change jobs. The number remains relatively stable for different age groups, except that the job-to-job movement decreases slightly after 65 to around 8%. In general, there is a non-negligible number of employment and job movements, and the movements do not concentrate on a narrow age group.

Table 1: Employment Dynamics

		Age group			
	All	≤ 55	$55 \sim 60$	$60 \sim 65$	> 65
Individual					
People who have started new jobs	0.32	0.16	0.21	0.17	0.13
job-to-job transition	0.21	0.12	0.15	0.10	0.05
re-entry	0.14	0.03	0.06	0.06	0.08
Individual-year					
Employment to Employment	0.60	0.87	0.78	0.54	0.30
Employment to Nonemployment	0.14	0.08	0.10	0.18	0.14
Nonemployment to Employment	0.04	0.03	0.04	0.03	0.04
Nonemployment to Nonemployment	0.23	0.02	0.08	0.24	0.52
Job-to-job transition	0.07	0.11	0.09	0.06	0.02
conditional on employment	0.11	0.12	0.11	0.11	0.08

Notes: HRS sample, male 51-70, further selection discussed in Section 4.

Among those who left their jobs, HRS shows a diversity of motivations for doing so. Table 2 presents the main reasons for leaving the employer recorded in the previous wave, which includes retirement (39.4%), laid-off (20.3%), poor health (13.5%), quit (9.6%), better job (7.1%), and business close (6.9%).[16] Moreover, different reasons for leaving are

[14]Including individuals who have not reached 70, removing individuals with missing wage and hours, or attrition can lead to an underestimation of this statistics.

[15]The job-to-job transition may also contain a non-employment period of up to two years between two surveys. To be categorized as re-entry, at least 3 waves are required: employed, non-employed, and employed. In Section 4, we discuss in detail the definition of employment variables and their limitations.

[16]Other reasons includes family care (1.5%), family moved (1.4%), ownership changed (1.3%), pension rule changed (0.1%), divorce/separation (0.0%), handed over responsibilities to other family members (0.0%), transportation (0.6%), to travel (0.1%), early retirement incentive/offer (0.5%), financial advantageous (0.2%), transferred (0.0%).

associated with different employment dynamics. For instance, people with poor health are much more likely to be non-employed in the following wave. In contrast, around half of people who left due to business close start a new job in the following wave. The results suggest that there is substantial heterogeneity in the shocks that people receive (e.g., involuntary job separation, health shocks, productivity shocks, etc.), which further leads to distinct employment dynamics in later life.

Table 2: Fractions of Reasons For Leaving The Job

		In the following wave	
	All	Employed	Non-employed
Business close	6.9	3.4	3.4
Laid off	20.3	9.0	11.3
Poor health	13.5	0.6	12.8
Better job	7.1	6.5	0.6
Quit	9.6	5.3	4.3
Retire	39.4	5.1	34.3

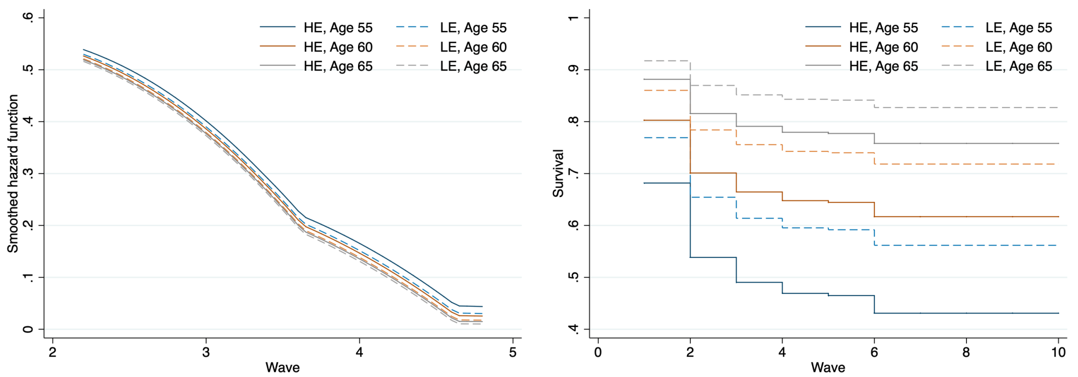
Notes: HRS sample.

Additionally, for those who left their jobs, a job-to-job transition is much more likely than re-entry. Figure 1 presents the result of fitting a Cox proportional hazards model for unemployment duration with age and education as independent variables. Figure 1(a) is the estimated hazard function, and 1(b) is the survivor function. The results show that for those who left their jobs, the probability of starting a new job decreases with the non-employment duration: most of them start working immediately in the next wave.

Figure 1: Cox Proportional Hazards Model For Unemployment Duration

(a) Hazard function

(b) Survivor function



Notes: the hazards model controls for education and age.

Finally, we focus on the group of individuals with a job-to-job transition and compare their wage and hour changes with those of the group of stayers. Figure 2 plots the density of log wages and log hours for stayers and movers. Specifically, Figure 2a compares the log wage changes. It is obvious that the distribution for stayers is more concentrated around zero, whereas the one for movers is much more widespread. Furthermore, the wage changes of movers are more left-skewed, with more negative changes than positive ones. The result indicates that stayers face much less wage volatility than movers. Indeed, most movers end up with a worse-paying job, and only a small fraction of them end up with a better-paying job. Figure 2b presents the distribution for hour changes. The comparison between stayers and movers is qualitatively the same as that for wage changes, which suggests that the switch to worse-paying jobs might be accompanied by a reduction in working hours. We further restrict the sample to workers with more than 10yrs of tenure to get closer to the notion of bridge jobs. As shown in Figures 2c and 2d, the features are more pronounced. In Appendix B, we include additional results by education, similar figures in which "business close" movers are excluded, figures in which both "business close" and "laid-off" movers are excluded, and the pattern remains the same.[17]

To sum up, there are a non-negligible amount of employment and job changes with distinct reasons. In terms of wage and hour changes, there is heterogeneity between stayers and movers. Among movers, a large fraction switch to jobs that are worse paid and/or less time demanding. Motivated by these facts, we construct a model that takes into account different types of shocks and aims to capture the heterogeneity in employment transitions, job mobility, as well as changes in wages and hours.

3 An Empirical Model of Earning Dynamics

In this section, I propose an empirical model of earning dynamics following Altonji et al. (2013) with extensions in adding health and mismatch shocks and endogenizing working hours. The model specifies a set of interpretable shocks and their transmission to wages, hours of work, and employment status. It also allows for observed and unobserved heterogeneity.

The timing of events is as follows. Consider individual i, $i \in \{1, 2, ..., N\}$ at period t, $t \in \{1, ..., T\}$ whose current or most recent job is indexed by $j(t)$.[18] Firstly, he receives a set of shocks to employment status, latent wage rate, and working hours. At the same

[17]For prime-age workers, Bowlus et al. (2021) document that non-laid-off movers on average experience positive earnings growth, Karahan et al. (2019) document the left skewness of earning change distribution for both stayers and movers and the level vary with lifetime earnings, and Halvorsen et al. (2019) document that the left skewness is mainly driven by stayers.

[18]We use job and firm interchangeably.

Figure 2: Log Wage And Hour Changes For Movers And Stayers

Notes: HRS sample.

time, he also gets an outside job offer j', which is a package of wage rate and working hours. Then the worker will face three possibilities: stay at the old job, move to the new job or become non-employed. The non-employed only face two possibilities: move to the new job or stay non-employed. Finally, the observed wage rate and working hours are determined based on the employment outcomes.

Notation. Throughout the model specification, we uniformly use γ to represent coefficients, whose superscript and subscript refer to the dependent variable and explanatory variable, respectively. We use $f_k(\cdot)$ to represent a function whose output is variable k or a component of variable k. Letter ε stands for an i.i.d. error component, and its superscript refers to the variable on which ε has a direct effect.

Log wages. We start by introducing the wage process. Denote the observed log wage rate by w_{it} for individual i at period t. We only observe w_{it} for workers ($E_{it} = 1$), and it equals the latent wage w^*_{it} plus the measurement error ε_{it}, as indicated by equation (1).[19]

$$w_{it} = E_{it} \times (w^*_{it} + \varepsilon_{it}) \tag{1}$$

$$w^*_{it} = X'_{it}\gamma^w_X + h_{it}\gamma^w_h + \mu_i + \omega_{it} + v_{ij(t)} \tag{2}$$

Equation (2) describes the composition of the latent wage. The first component contains a vector of exogenous observables X_{it} including education, race, polynomials of age, and a recession indicator. The second term contains a latent health status variable h_{it}. The third one is an unobserved individual fixed effect μ_i.[20] Finally, the last two components are individual productivity, ω_{it}, and firm-specific productivity, $v_{ij(t)}$.

Working hours. Let l_{it} denote the annual working hours. They are determined in the following way:

$$l_{it} = E_{it} \times (X'_{it}\gamma^l_X + h_{it}\gamma^l_h + \mu_i\gamma^l_\mu + \xi_{ij(t)} + \varepsilon^l_{it}) \tag{3}$$

Equation (3) says that we only observe positive working hours for workers. The amount depends on the demographic characteristics and aggregate shocks X_{it}, health h_{it}, unobserved fixed effect μ_i, a firm-specific working-hour requirement $\xi_{ij(t)}$ and a transitory shock ε^l_{it}.

Health dynamics. Latent health h_{it} follows an age-dependent Markovian exogenous process:

$$h_{it} = f_h(h_{i,t-1}, age_{i,t-1}, edu_i, \varepsilon^h_{it}) \tag{4}$$

We regard h_{it} as an underlying continuous index associated with self-reported health level in the HRS survey, srh_{it}, which is a discrete variable varying from 1 (Excellent) to 5 (Poor). The measurement equation is:

$$srh_{it} = \sum_{k=1}^{5} \mathbb{1}(h_{it} > \tau^{srh}_k) \tag{5}$$

The range of h_{it}, which is $(-\infty, \infty)$, is divided into five intervals by the thresholds τ^{srh}.

Self-reported health is a subjective measure. It is comprehensive and is considered as a powerful predictor of mortality (Idler and Benyamini, 1997; Heiss, 2011). In addition, self-reported health alone is enough for identifying the parametric model of latent health dynamics. Adding more variables might improve the efficiency, but might also change

[19] In principle, ε_{it} could also capture transitory shocks that are independent of employment outcomes.

[20] We assume μ_i is independent of the education and race in X_{it}. Therefore, μ_i represents the extra heterogeneity in addition to the linear effects of education and race.

the interpretation of the latent variable. Health investments, either pecuniary or non-pecuniary, are not considered.

Individual earning power. In our model, people differ in productivity ω_{it} and productivity risk σ_{it}. The productivity risk σ_{it} is determined by lagged risk, age, and health, and further affects the process of productivity ω_{it}. Their joint dynamics are as follows

$$\omega_{it} = \rho(\sigma_{it}, age_{i,t-1})\omega_{i,t-1} + \varepsilon_{it}^{\omega}\sigma_{it} \tag{6}$$

$$\sigma_{it} = f_{\sigma}(age_{i,t-1}, h_{i,t-1}, \sigma_{i,t-1}, edu_i, \varepsilon_{it}^{\sigma}) \tag{7}$$

As shown in equation (6), productivity risk σ_{it} affects ω_{it} through both the variance of the shocks and the persistence. The interpretation is that some people might face larger volatility and/or faster depreciation of productivity than others. Furthermore, this heterogeneity can change over time and be partly explained by their age and health history.

This specification is motivated by the recent literature that finds non-linear persistence in income dynamics (Arellano et al., 2017; Almuzara, 2020).[21] The nonlinearity means that the persistence, instead of being constant, changes with the current status and the shocks. As a result, some extreme shocks could fully erase the accumulation of human capital. One way to allow for this feature in our model is through σ_{it}: negative health history causes non-marginal drop in productivity.

Firm-specific wage component and mismatch shock. The firm-specific component $v_{ij(t)}$ consists of two parts, a deterministic polynomial trend of tenure $ten'_{ij(t)}\gamma_e$ and a stochastic term $\nu_{ij(t)}$ that captures the job fit. The value of $\nu_{ij(t)}$ changes with jobs and can also change during a job tenure when workers become mismatched. Let $\nu'_{ij(t)}$ denote the wage component attached to the job offer, and m_{it} denote a 0-1 discrete mismatch shock. We have the following equations:

$$v_{ij(t)} = ten'_{ij(t)}\gamma_{ten}^{v} + \nu_{ij(t)} \tag{8}$$

$$\nu_{ij(t)} = \begin{cases} \nu'_{ij(t)} & \text{if start a new job } (ue_{it} = 1 \text{ or } jc_{it} = 1) \\ \nu''_{ij(t)} \ll \nu_{ij(t-1)} & \text{if not start a new job and } m_{it} = 1 \\ \nu_{ij(t-1)} & \text{Otherwise} \end{cases} \tag{9}$$

Equation (9) says that variable $\nu_{ij(t)}$ is determined by both shocks and choices. When people move to a new job, either from non-employment ($ue_{it} = 1$) or their old job ($jc_{it} = 1$), $\nu_{ij(t)}$ changes to the value of the new job $\nu'_{ij(t)}$. If people maintain their employment and

[21] A special case of equation (7) is a fixed σ_i as in Almuzara (2020).

job status without suffering from mismatch shock, $\nu_{ij(t)}$ keeps the same value as in the previous period. However, if an individual receives a mismatch shock, $\nu_{ij(t)}$ is reduced to a much smaller value $\nu''_{ij(t-1)}$ such that staying at the old job implies earning a much lower wage rate in the following periods.

The probability that a mismatch shock happens depends on age, education, the stochastic firm-specific wage component, and the employment status in the previous period.

$$m_{ij(t)} = \mathbb{1}\{f_m(age_{i,t-1}, edu_i, \nu_{i,t-1}, E_{i,t-1}) + \varepsilon^m_{it} > 0\} \tag{10}$$

A mismatch shock is defined as a shock that reduces wages and also triggers a job change.[22,23] The shock is intended to capture a non-marginal drop in firm-specific productivity, which is associated with the randomness of the aging processes in life. On the one hand, deteriorating health as people age might cause a decrease in productivity. On the other hand, the impact of aging on productivity is not necessarily associated with diseases. For instance, fluid intelligence, which refers to the ability of reasoning, learning new things, and other mental activities that depend only minimally on prior learning and acculturation, tends to decline during late adulthood (McArdle et al., 2002). The decline of cognitive skills associated with fluid intelligence is heterogeneous across individuals and rather general at the within-individual level as people age (Ghisletta et al., 2012). Therefore, this may cause workers in cognitive-skill demanding occupations to mismatch with their job requirements and other similar jobs.[24] However, finding a comprehensive cognitive health measure is often difficult.[25] Relying on the latent mismatch shock, we expect to capture part of the changes in cognitive health as well.

In combination with the dynamics of $\nu_{ij(t)}$, once mismatch shock happens in the model, staying at the current job becomes too costly. The options opened to a mismatched worker will only be non-employment or to accept an outside offer, which will also be impacted by the mismatch shock. Mismatch shocks can also happen during non-employment, thus reflecting the depreciation of job-specific productivity, which further affects the

[22]Alternatively, a mismatch shock could be specified as a shock to wages only. However, such a model would not be able to capture the pattern of wage changes. This is because we observe much smaller wage volatility among stayers than movers.

[23]In principle the mismatch shock might also capture the wage changes due to the pass-through of the firm productivity. Without matched employ-employee data, we can not distinguish the two.

[24]Wen (2018) provides evidence that cognitive health affects labor supply depending on occupation for older workers.

[25]For instance, Wen (2018) uses memories only, Blundell et al. (2017) constructs an index from memories and two variables of IADL list, Heineck and Anger (2010) uses results of symbol correspondence test.

probability of transitioning to employment as well as future wages.

Firm-specific effect on working hours. The firm-specific hours component $\xi_{ij(t)}$ is included in the hours equation to reflect some potential fixed requirements of each job. It stays the same during the tenure of a worker and changes to $\xi'_{ij(t)}$ when the worker starts a new job.

$$\xi_{ij(t)} = \begin{cases} \xi'_{ij(t)} & \text{if } ue_{it} = 1 \text{ or } jc_{it} = 1 \\ \xi_{ij(t-1)} & \text{otherwise} \end{cases} \tag{11}$$

Job destruction and new offer. In each period, there are some workers involuntarily losing their jobs due to job destruction. We define jd_{it} as a 0-1 binary variable that equals 1 when job destruction happens. Its probability varies with education to allow people with different education levels to have different levels of work stability (equation **??**). s

Meanwhile, everyone receives a job offer in each period. In our model, a job offer is defined as a package of wage component $\nu'_{ij(t)}$ and hours requirement $\xi'_{ij(t)}$. The hours requirement can be treated as a sort of amenity (Lamadon et al., 2019; Card et al., 2018). People either take it or leave it. The distribution of offers is expressed in equations (12) and (13):

$$\nu'_{ij(t)} = f_\nu(\nu_{ij(t0)}, m_{it}, jd_{it}, E_{i,t-1}, edu_i) + \varepsilon^\nu_{it} \tag{12}$$

$$\xi'_{ij(t)} = f_\xi(\xi_{ij(t0)}, m_{it}, jd_{it}, E_{i,t-1}, edu_i, \nu'_{ij(t)}) + \varepsilon^\xi_{it} \tag{13}$$

Variables $\nu_{ij(t0)}$ and $\xi_{ij(t0)}$ represent the initial firm-specific wage and hours components of the current or most recent job. The dependence of new offers on $(\nu_{ij(t0)}, \xi_{ij(t0)})$ may reflect different aspects. For example, new employers may refer to workers' previous wages when they make an offer.[26] In addition, the employment status can also affect the job offer. Whether someone was working or not (E_{it-1}), or whether someone lost the job involuntarily or not (jd_{it}) might reflect their effort in searching for new jobs. Moreover, the mismatch shock m_{it} also has an influence on the job offer. This is because one of the potential causes of job mismatch is impaired firm-specific skills. Even if a person moves from job j to j', depending on the type of job, the new place may still to some extent requires the same skill, which in turn affects the wages. In other words, in general, workers can not neutralize the mismatch shock by changing firms if the new job is similar

[26] Altonji et al. (2013) give three potential reasons including 1) employers base the offer on their previous wage, 2) reflect some common demanding shocks to a certain group of firms and, 3) quality of firm may affect the workers network

to the previous one. If the person is lucky (through ε^{ν}_{it}), he might still receive a well-paying offer, for example, when the skills required in job j are not essential in job j'. Finally, we allow for correlation between $\nu'_{ij(t)}$ and $\xi'_{ij(t)}$.

As a result, mismatched workers will leave their jobs and be more likely to accept job opportunities that they would not have considered otherwise. This is so because, in addition to having an effect on the reservation wage, m_{it} will also affect the quality of new job offers as we discuss below.

Employment to employment transition. After drawing all shocks, workers who receive neither the job destruction shock nor the mismatch shock ($E_{i,t-1} = 1$, $jd_{it} = 0$, and $m_{it} = 0$) can make decisions of whether and where to work. Two equations are used to describe their decisions. We start by depicting their decisions of labor supply. Define $ee_{it} = 1$ if workers continue working , either in the same place or in a new one, and $ee_{it} = 0$ otherwise. Given $E_{i,t-1} = 1$, $jd_{it} = 0$ and $m_{it} = 0$, we specify the following equation:

$$ee_{it} = \mathbb{1}\{f_{ee}(X_{it}, A_{i,t-1}, h_{it}, ten_{i,t-1}, w^*_{i,t-1}, \omega_{it}, \nu_{ij(t-1)}, \nu'_{ij(t)}, \xi_{ij(t-1)}, \mu_i) + \varepsilon^{ee}_{it} > 0\} \quad (14)$$

where ε^{ee}_{it} is the exogenous shock to the labor supply decision and $A_{i,t-1}$ denotes assets. Demographics X_{it}, health h_{it}, and tenure $ten_{i,t-1}$ can affect employment through the wage equation, but may also reflect some systematic difference among groups in the utility cost of working.[27] Employees observe $h_{i,t}$, $\nu_{ij(t-1)}$, $\xi_{ij(t-1)}$, ω_{it} and hence know the wage if they continue working in the same place. The quality of the outside offer ν'_{ijt} also affects the employment choice.[28] In principle, we expect that a good offer increases the probability of working, potentially in a nonlinear way.[29]

Job change. If an employee decides to continue working (i.e. $E_{i,t-1} = 1$, $jd_{it} = 0$, $m_{it} = 0$, and $E_{it} = 1$), he still needs to decide whether to change job or not. The binary variable jc_{it} equals one if worker i accepts the new offer opened to him and changes jobs.

$$jc_{it} = \mathbb{1}\{f_{jc}(X_{it}, A_{i,t-1}, h_{i,t}, ten_{i,t-1}, \omega_{it}, \nu_{ij(t)}, \nu'_{ij(t)}, \xi_{ij(t)}, \xi'_{ij(t)}, \mu_i) + \varepsilon^{jc}_{it} > 0\} \quad (15)$$

where ε^{jc}_{it} is an exogenous shock to the job decision. If a worker moves to a new job, his stochastic wage and hours component will be updated to $\nu'_{ij(t)}$ and $\xi'_{ij(t)}$, but at the same time he will lose the accumulated tenure in the previous job. In addition to job-specific

[27]For instance, unhealthy people may bear more cost of working; the uncertainty in medical expenditure affects the employment choice too (French and Jones, 2017). Tenure may also capture their stickiness to labor market.

[28]We tried to add $\xi'_{ij(t)}$. However estimation becomes less stable.

[29]In a structural model in which people choose between staying and moving, what is important is the maximum of the expected discounted utility of staying and of moving.

characteristics, other components including health, individual-specific productivity, and fixed effect can also affect job decisions. Even though these components do not change with jobs, they may still matter for evaluating the two options. In a model where changing jobs induces utility loss, one way is by affecting income and thus assets, which affects the trade-off between the utility cost and the marginal change in consumption. Another way is through expectations. Potentially, healthier workers might expect a longer employment spell such that the discounted value of moving to a better-paying job is more profitable.

We have specified workers' choices in equations (14) and (15) conditioning one after the other. However, this does not necessarily mean that workers make decisions sequentially. These options can be seen as reduced form rules from a structural labor supply model, where an agent would simultaneously decide whether and where to work.[30] On the other hand, if workers make choices sequentially, this would be equivalent to adding constraints on the arguments of equations (14) and (15).[31]

Non-employment to employment transition. Individuals who were non-employed ($E_{i,t-1} = 0$), or lost their job due to job destruction shocks ($jd_{it} = 1$), or individuals who left their job due to mismatch shocks ($m_{it} = 1$), will choose to remain non-employed or accept a new job opportunity. If they accept the job offer opened to them, we set $ue_{it} = 1$, otherwise $ue_{it} = 0$. The decision rule is specified through the following equation:

$$ue_{it} = \mathbb{1}\{f_{ue}(X_{it}, A_{i,t-1}, h_{it}, ud_{i,t-1}, jd_{it}, m_{it}, \omega_{it}, \nu'_{ij(t)}, \xi'_{ij(t)}, \mu_i) + \varepsilon^{ue}_{it} > 0\} \tag{16}$$

where ε^{ue}_{it} is the exogenous shock to labor supply, and variable $ud_{i,t-1}$ is the number of years of non-employment since leaving the last job. The characteristics of the offer $\nu'_{ij(t)}$ and $\xi'_{ij(t)}$ affect the probability of transition to employment. In addition, whether someone has been non-employed for some periods or just left due to jd_{it} or m_{it} may also affect their assessment of the new job opportunity. Equations (14-16) describe the employment transitions and job mobility. They are empirically flexible and have an economic interpretation. Moreover, they control for selection and play an important role in identifying different risks.

After an individual receives exogenous shocks and makes employment and job decisions, his employment status E_{it}, non-employment duration ud_{it}, and tenure ten_{it} are determined endogenously as described in equations (17-19), and therefore the observed

[30]Equation (14) describes the probability that $\max(A, B) > C$ and equation (15) describes the probability that $B > C$ conditional on $\max(A, B) > C$.

[31]For example, if individuals make employment decisions before receiving the offer, then it is equivalent to restricting the coefficients of $\nu'_{ij(t)}$ and $\xi'_{ij(t)}$ in equation (14) to zero.

wage in equation (1) is determined.

$$E_{it} = ee_{it}E_{i,t-1} + ue_{it}(1 - E_{i,t-1}) \tag{17}$$

$$ud_{it} = (1 - E_{it}) \times (ud_{i,t-1} + 2) \tag{18}$$

$$ten_{it} = ee_{it}(1 - jc_{it})(ten_{i,t-1} + 2) + ue_{it} \times 2 + jc_{it} \times 2 \tag{19}$$

Asset accumulation. Finally, at the end of each period, assets evolve according to the following conditional probability distribution:

$$\mathbb{E}(\mathbb{1}(A_{it} < \tau_k^A)) = \Phi(f_{A,k}(X_{it}, A_{i,t-1}, E_{it}, w_{it}^*, l_{it}, h_{it}, \omega_{it}, \nu_{it}, \mu_i)) \tag{20}$$

where $\tau_k^A \in (0,1)$. We specify assets as a function of demographics, aggregate shocks, lagged assets, labor supply, wage rate, individual-specific productivity, firm-specific productivity, and unobserved heterogeneity. The standard budget constraint can be treated as a special case when consumption and other expenditures are functions of the arguments in equation (20).

Initial conditions and error components. To complete the model, we specify the distributions of the initial variables as follows:

$$h_{i1} \sim N(\mu_{h1}(X_{i1}, A_{i1}, E_{i1}), 1) \tag{21}$$

$$\nu_{i1} \sim N(\mu_{\nu 1}(age_{i1}, edu_i), \sigma_{\nu 1}^2(edu_i)) \tag{22}$$

$$\xi_{i1} \sim N(\mu_{\nu 1}(age_{i1}, edu_i, \nu_{i1}), \sigma_{\xi 1}^2(edu_i)) \tag{23}$$

$$\sigma_{i2} \sim Gamma(k_{\sigma_2}, f_{\sigma 2}(edu_i, age_{i1}, h_{i1})/k_{\sigma_2}) \tag{24}$$

$$\omega_{i1} \sim N(0, \sigma_{\omega 1}^2) \tag{25}$$

$$\mu_i \sim N(0, \sigma_\mu^2) \tag{26}$$

The conditional distribution of health on initial status is Normal with a conditional mean that depends on demographics, assets, and employment status. Initial firm-specific wage and hour components are drawn from a joint Normal distribution whose mean depends on age and education, and whose variance depends on education. Initial individual productivity risk is assumed to follow a conditional Gamma distribution with a shape parameter $k_{\sigma 2}$ and a scale parameter $f_{\sigma 2}(edu_i, age_{i1}, h_{i1})/k_{\sigma_2}$, and thus the conditional mean is $f_{\sigma 2}(edu_i, age_{i1}, h_{i1})$. Finally, we assume both initial individual productivity and unobserved heterogeneity are from Normal distributions with means normalized to be zero.

All the error components are i.i.d. over individuals and time, and independent of each other. Specifically, we assume that ε_{it}^{ee}, ε_{it}^{jc}, ε_{it}^{ue}, ε_{it}^{m}, ε_{it}^{jd} and $\varepsilon_{it}^{\omega}$ follow standard Normal distributions. In addition, the error components of the wage equation ε_{it}, the hours equation ε_{it}^{l} and the latent health equation ε_{it}^{h} are also normally distributed with zero mean and unknown variances. The offer error components ε_{it}^{ν} and ε_{it}^{ξ} are draws from Normal Mixture distributions whose CDF is $(1-\rho)\Phi(x)+\rho\Phi(x/k)$, with $\rho = 1/(k^2-1)$.[32] The dynamics of individual productivity risk σ_{it} is assumed to follow an autoregressive gamma process (Gourieroux and Jasiak, 2006), and will be discussed in Section 5.1.

Additional discussion of the mismatch shocks. The mismatch shock is an important tool in our model to explain the worker's movement to a worse-paying job. It is discrete, and thus, the change in the firm component takes the form of a step function. This is a simple way to create asymmetry in log wage changes in correspondence with its left-skewness in data. At the same time, mismatch shocks also generate nonlinearity in persistence: those who receive mismatch shocks and end up in a new job lose their accumulated tenure.

Different from shocks to individual-specific productivity, mismatch shocks force workers to leave their current job. At the same time, mismatch shocks have a persistent negative effect on productivity, whereas the job destruction shocks, which also lead to leaving the job, do not have. Intuitively, individual-specific productivity shocks are mostly identified from wage variation among the stayers, whereas the mismatch shocks are identified from the wages of job movers, especially those who switch to jobs that pay less and require fewer working hours.

4 Mapping Variables in Model to Data

In this section, we discuss the construction of the main variables of our model. The data comes from RAND HRS, which is a user-friendly version provided by the RAND Center for the Study of Aging.[33] For consistency, our sample covers 11 waves from 1996 to 2016.

Wage w_{it} is the log real hourly wage calculated from the usual hours worked per week, the usual weeks worked per year, and a pay rate using 2016 as the base year. Labor supply l_{it} is the log total working hours calculated from usual number of hours per week

[32]See Ali (1974). We estimate the model whose parameter k varies with education, mismatch shock, and employment status in period $t-1$. Further detail is in Section 5.1.

[33]This file was developed at RAND with funding from the National Institute on Aging and the Social Security Administration.

and usual weeks worked per year.

Employment status E_{it} equals one if individuals are employed at the time of the interview, whether full-time or part-time, retired or not. On the contrary, it equals zero if individuals are unemployed or not in the labor force.[34] Accordingly, employment transition variables ee_{it} and ne_{it} are constructed based on the labor force status in two consecutive waves: $ee_{it} = \mathbb{1}(E_{i,t-1} = E_{it} = 1)$ and $ne_{it} = \mathbb{1}(E_{i,t-1} = 0, E_{it} = 1)$. Furthermore, for those with $ee_{it} = 1$ but do not work for the same employer as in the last wave, job change jc_{it} equals one.

The way E_{it} is defined implies that variables ee_{it} and ne_{it} are silent about the individual's employment status between the two surveys. Similarly, jc_{it} does not inform us about potential job changes in between waves, or whether a mover has experienced unemployment. Indeed, people labeled as "employed" in both periods may have experienced unemployment, or those labeled as "job movers" may have gone through a period of unemployment in between waves, in which cases job-to-job movement would not be an accurate description. With this in mind, our employment transition equations could be interpreted as a reduced form describing employment changes at two points in time.

HRS also asks why people left the employer in the last wave. Those who mentioned "business closure" are assigned jd_{it} equal to one. Additionally, the education measure edu_i is an indicator of higher education which is defined as some college or above. The health measure srh_{it} is a self-reported general health status categorical variable. Its value ranges from "1" for Excellent to "5" for Poor. Finally, assets A_{it} are defined as the sum of all wealth components (excluding secondary home) net of all debt.

Observations for a given person-year are retained if the person is aged 51 to 70, never self-employed, and has non-missing data on education, race, self-reported health, birth, wages, working hours, and employment status.[35] We restrict the sample to male individuals who have eligible observations for at least three consecutive waves and are employed for at least one wave. See Appendix A for further details. We end up with a sample of $2,897$ individuals and $15,277$ individual-year observations.

[34] Including people who not in the labor force potentially affects the transition probability to employment. However, we expect the impact to be small, because based on the sample selection, the sample is limited to individuals who have been employed for at least one period.

[35] The estimation algorithm we use, which is a variant of EM algorithm, would allow us to handle missing data, such as missing wages and working hours. However, increasing the amount of missing information may decrease the rate of convergence and increase the computing time.

5 Estimation Strategy

5.1 Empirical Specification

We now explain the empirical specification of the model components. Health dynamics consist of a polynomial age trend, the accumulation from the previous period (h_{it} interacted with an age polynomial and health in the last period), and the random shocks whose variance varies with previous self-reported health.[36] Individual productivity is expressed as the sum of the productivity accumulation ($\omega_{i,t-1}$ interacted with a linear function of age and productivity risk σ_{it}) and shocks whose variance is σ^2_{it}. Productivity risk is assumed to follow the autoregressive gamma process (Gourieroux and Jasiak, 2006).

We assume the mean of job offers are a linear function of the current firm-specific component, mismatch shocks, employment status at the beginning of the current period, and its interaction with mismatch shocks, education and its interaction with mismatch shocks. At the same time, education, mismatch shocks, and employment status at the beginning of the current period also affect the variance of the shocks.

In the employment-to-employment equation, in addition to polynomials for each component, we also add interactions of age with $ten_{i,t-1}$, $\nu_{ij(t-1)}$ and $\nu'_{ij(t)}$, and interactions of education with $\nu_{ij(t-1)}$ and $\nu'_{ij(t)}$. This is motivated by the data pattern that the effect of tenures and wages on future employment vary with education and age groups. Similarly, in the non-employment-to-employment equation, we add interactions of age and education with $\nu'_{ij(t)}$. Additionally, we allow a flexible specification for job destruction shocks by adding the interactions of jd_{it} with other elements, to take into account that the non-employed and those who just lost their job involuntarily may have different employment dynamics. Finally, the job change equation adds the interaction of age and education with $\upsilon_{ij(t-1)}$, $\Delta\upsilon_{ij(t)}\mathbb{1}(\Delta\upsilon_{ij(t)} > 0)$, $\Delta\upsilon_{ij(t)}\mathbb{1}(\Delta\upsilon_{ij(t)} <= 0)$ allowing for different reactions to offers depending on the total firm-specific component.[37] Further details on the specification can be found in Appendix C.

[36] The coefficient of $h_{i,t-1}$ is limited between 0 and 1

[37] Variable $\Delta\upsilon_{ij(t)}$ is defined as $\nu'_{ij(t)} - \nu_{ij(t-1)} + \Delta ten'_{ij(t)}\gamma^{\upsilon}_{ten}$. Starting a new job correspondes a tenure of two years.

5.2 Estimation Algorithm

The Expectation-Maximization (EM) algorithm proposed by Dempster et al. (1977) is a useful tool for empirical models with latent variables for obtaining maximum-likelihood estimates. Starting from an initial guess of parameters, the algorithm iterates between an E-step, which computes the conditional mean of certain functions of latent variables given observables, and an M-step, which solves the optimization problem and updates parameters until the convergence to the maximum of the likelihood. The EM algorithm has also been extended to introduce GMM estimation in the M-step (Arcidiacono and Jones, 2003). However, in complicated models where computing the E-step analytically is infeasible, simulated versions of the EM algorithm are often implemented. A prominent example is the stochastic expectation-maximization (SEM) algorithm (Diebolt and Celeux, 1993). In this case, in the E-step, we draw latent variables from the posterior distribution given observables, and in the M-step, update parameters as if the draws were observables. We iterate between two steps until the convergence of the estimates to the stationary distribution. However, a drawback of the EM algorithm and its variants is the slow convergence in some situations, especially when the models contain multiple latent variables over multiple periods, such as our case.

In this paper, I estimate the model using a modified (parameter-expanded) stochastic EM algorithm (PX-SEM), which is developed in Wei (2021). The algorithm combines the parameter expansion techniques in Liu et al. (1998) with the SEM algorithm. Similar to the standard SEM algorithm, PX-SEM also consists of an E-step where we draw latent variables and an M-step where we update parameters. The E-step is the same for both algorithms. However, in the M-step, PX-SEM requires 1) expanding the original model, 2) estimating the expanded one, and 3) reducing to the original model space to obtain the estimator. The main motivation of the algorithm is to make better use of some model assumptions in each M-step to accelerate the convergence: we realize that the draws from the E-step given a bad guess will violate some model assumptions; therefore, we exploit a more robust estimator to the guesses in E-step.[38] Wei (2021) shows that PX-SEM can

[38]Consider a simple deconvolution example of $y_i = y_i^* + \epsilon_i$, $\epsilon_i \sim N(0,1)$, $y_i^* \sim N(0,\sigma^2)$, and $y_i^* \perp \epsilon_i$. If we estimate this model using SEM algorithm, in the E-step, we need to draw $y_i^* \sim f(y_i^*|y_i;\sigma^{(s)})$. It is easy to show that $cov(y_i^*,\epsilon_i) = \frac{\sigma^{(s)2}(\sigma_{true}^2-\sigma^{(s)2})}{(\sigma^{(s)2}+1)^2}$, which is only zero when the guess $\sigma^{(s)}$ is correct, $\sigma^{(s)} = \sigma_{true}$. Otherwise, we will observe correlations between sampled y^* and $y-y^*$ that violates the model assumptions. Intuitively, if we have a guess $\sigma^{(s)}$ that is significantly larger than the true value, then a negative correlation between y^* and $y-y^*$ will be observed given the variance of y_i is always preserved. PX-SEM takes into account the possibility of this type of violation.

greatly reduce the computing time.[39,40]

Let Y denote all observables $Y \equiv \{X, w, l, E, jc, jd, ten, srh, A\}$, Z denote all latent variables $Z \equiv \{h, \mu, \omega, \sigma, \nu', \nu, m, \xi', \xi\}$, and Θ denote all parameters. Then the density function of the joint distribution of Y and Z is expressed as $f(Y, Z; \Theta)$. In addition, we use $f_Y(Y; \Theta)$ for the marginal density of observables, and $f_{Z|Y}(Z|Y; \Theta)$ for the posterior density of latent variables given all observables. Starting with an initial guess $\hat{\Theta}^{(0)}$, we iterate between the following two steps until the convergence of the $\hat{\Theta}^{(s)}$ process to its stationary distribution:

1. Stochastic E-step: draw $Z^{(s)}$ from $f_{Z|Y}(Z|Y; \hat{\Theta}^{(s-1)})$
2. PX-M-step: update to $\hat{\Theta}^{(s)} = \arg\max_{\Theta} \sum g(Y, Z^{(s)}; \Theta)$

where $g(\cdot)$ is the overall estimation objective function.[41] The final estimator is $\hat{\Theta} = \frac{1}{S^0} \sum_{S-S^0+1}^{S} \hat{\Theta}^{(s)}$.

There are many options for the function $g(\cdot)$ in the M-step depending on the way the original model is expanded. So far, there is not a general rule. A simple rule is that the expansion and the reduction need to be easy to operate. Considering that our original model is complex enough, we will only expand the model by allowing linear dependence of latent variables on observables and scaling up and down the variances of each latent variable. The explicit form of $g(\cdot)$, as well as the likelihood function and detailed estimation steps, can be found in Appendix D. We use a random-walk Metropolis-Hastings sampler in the E-step. The acceptance rate is controlled to be between 20% and 40%.

In practice, we take a two-stage estimation strategy. First, we estimate the health dynamics process (equations 4 and 5). The reason is that health dynamics are exogenous to the model, and the parameters are separately identified. Secondly, we estimate the rest of the parameters taking health dynamics as given. The results are based on 500

[39] SEM algorithm starts with some initial guess of unknown parameters and iterates between E-step and M-step until the convergence of the parameter process. However, in practice, if the initial guess is far away from the true values, it may take very long time to converge. Furthermore, very often it is even difficult to know whether the initial guess is close to the true vales or not, researchers usually go through SEM algorithm many times from different initial guess and pick one based on some criteria such as largest likelihood. All these reasons prolong the computing time for the SEM algorithm.

[40] Nielsen (2000) proves the ergodicity of the process. Liu et al. (1998) proves the parameter expansion technique could accelerate EM algorithm.

[41] In the M-step of SEM, the optimization is based on the likelihood: $g(Y, Z^{(s)}; \Theta) = \ln f(Y, Z^{(s)}; \Theta)$, but $g(\cdot)$ could also combine a likelihood component for some parameters with a GMM component for others, such as in Arcidiacono and Jones (2003). As another example, in Arellano and Bonhomme (2016), the M-step is based on quantile regression.

iterations ($S = 500$), with 150 MH draws in each iteration. We take the average of the last 200 iterations as the estimates ($S^0 = 200$).

6 Results

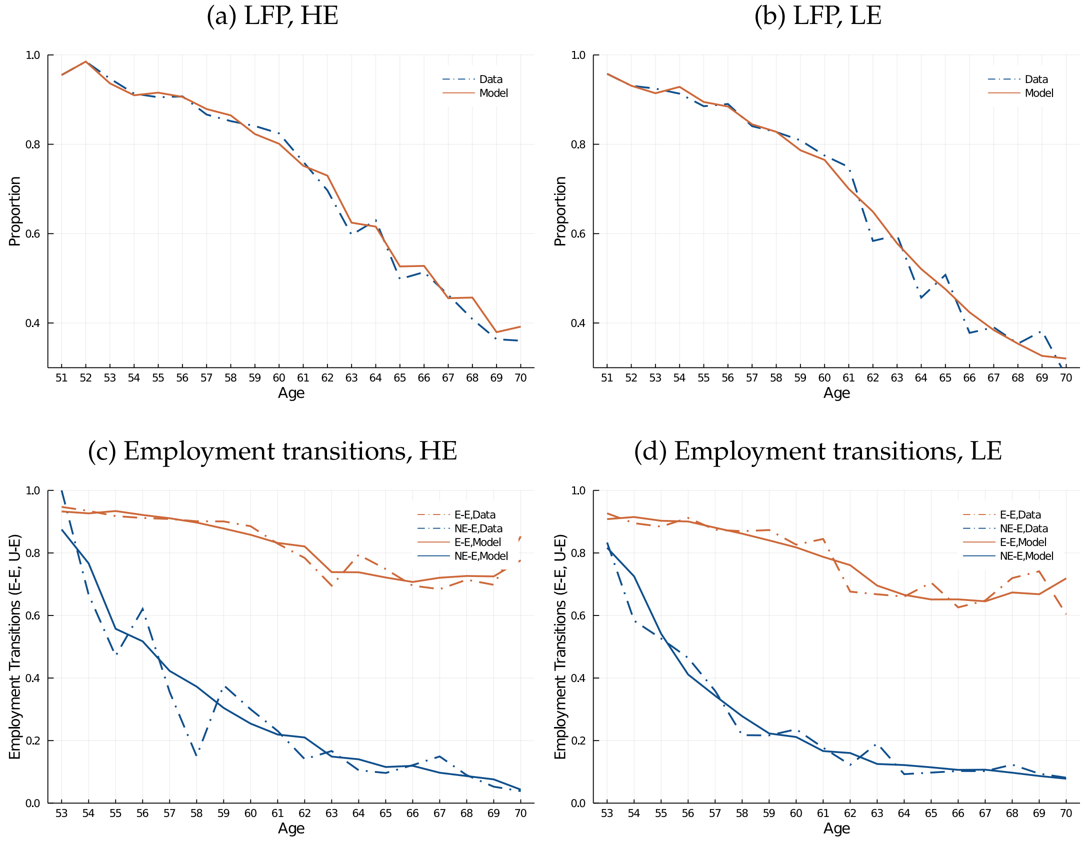
In this section, we present the estimation results from our empirical model. We start by checking model fit by comparing the model simulations with the data along relevant margins. Then we present the model implications. Specifically, we will analyze the importance of mismatch shocks in explaining job movements to worse-paying jobs and study the contribution of different risks and fixed effects to the dynamics of employment, wages and hours of work.

6.1 Model Fit

To evaluate the fit of the model, we simulate the employment and wage trajectories for $20 \times N$ individuals from the estimated model. Then we calculate and compare several statistics from simulated and actual data.

Employment. Figure 3 presents results on employment rates and employment transitions. Specifically, Figures 3a and 3b show employment profiles for HE and LE, respectively. In each figure, we plot the proportion of the employed at each age using both data and the simulations from our model. The employment rate for both education groups decreases with age. The employment rate of LE is similar to HE before 61, but decreases faster after 61. Figures 3c and 3d show the proportions of individuals transitioning to employment conditional on being employed and non-employed in the previous period at each age for HE and LE, respectively. For the employed, around 90% remain employed in the next period in their 50s. The number decreases to around 70% in their 60s. For the non-employed, the proportion of individuals being employed in the next period decreases sharply from around 80% in their early 50s to around 10% since early 60s. In general, Figure 3 shows that our model captures the trends in employment rates and transitions well.

Figure 3: Employment And Employment Transition By Age



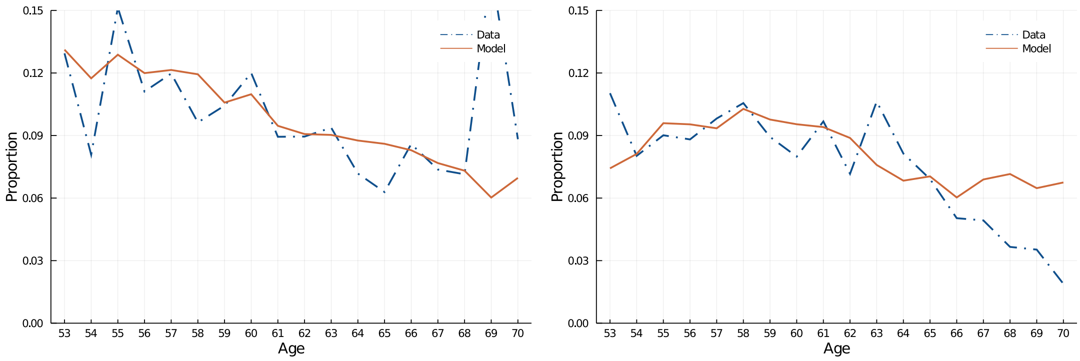
Notes: Comparison of the employment profile in the HRS sample (dashed line) and the simulations from the estimated model (solid line). Figures (a) and (b) plot the employment rate by age for the subpopulation with high education (HE) and low education (LE), respectively. Figures (c) and (d) plot the probability of being employed in the following period for the employed (E-E) and the non-employed (NE-E) by age for HE and LE, respectively.

Figure 4 displays the proportions of workers who change jobs in the following wave by age. The proportion decreases with age for HE, whereas for LE, the number is relatively stable before 61 and declines after. For both education groups, our model fits the pattern well.

Wages. In Figure 5, we show how the percentiles of log wages vary with age in the data and the simulations. Results are presented by the education group. We see that for both education groups, the median wage decreases with age. The HE group has larger dispersion in wages than LE. Furthermore, the dispersion for the HE increases with age as the top percentiles decline less than the bottom percentiles with age, as we can see in Figures 6a and 6b where we plot the $P90 - P10$, and 6c and 6d where the dispersion is further decomposed into $P90 - P50$ and $P50 - P10$. However, the dispersion for LE is relatively stable until around 65, and then it starts to decrease as the top percentiles

Figure 4: Job Change By Age

(a) Job-to-job move conditional on EE, HE (b) Job-to-job move conditional on EE, LE

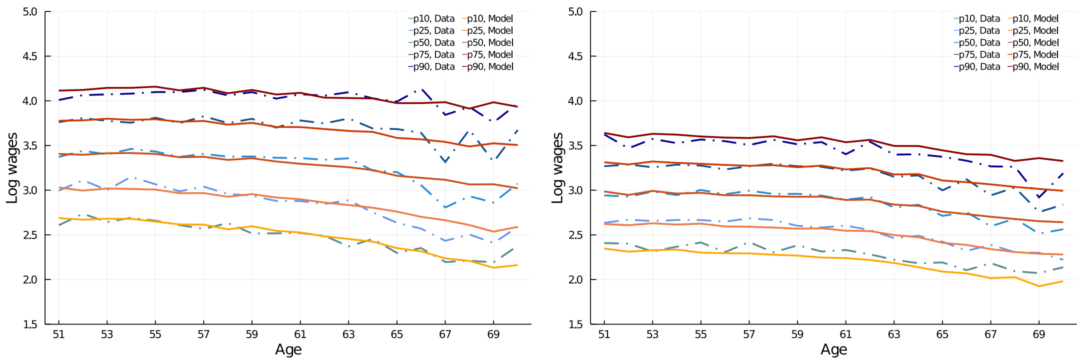


Notes: Comparison of the job change profile in the HRS sample (dashed line) and the simulations from the estimated model (solid line). Figures (a) and (b) plot the probability of changing jobs in the following period for the employed, that is, $Prob(jc_{it}|E_{i,t-1} = 1, E_{it} = 1)$. The rate is computed by age for the subpopulation with high education (HE) and low education (LE), respectively.

drop more than the bottom ones. Our simulations fit the data better for workers before age 65. The model performs less well for the 65+ in fitting the dispersion and other higher moments. The counterpart of quantile-based measures, the mean and standard deviation, are presented in Appendix E.

Figure 5: Quantiles of Log Wages

(a) Quantiles of wages (excluding zeros), HE (b) Quantiles of wages (excluding zeros), LE



Notes: Comparison of the log wage profile in the HRS sample (dashed line) and the simulations from the estimated model (solid line). Figures (a) and (b) plot the 10th, 25th, 50th, 75th, and 90th percentiles of log wages by age for the subpopulation with high education (HE) and low education (LE), respectively.

Figure 6: Dispersion of Log Wages

(a) $P90 - P10$ of log wages, HE (b) $P90 - P10$ of log wages, LE

(c) $P90 - P50$ and $P50 - P10$ of log wages, HE (d) $P90 - P50$ and $P50 - P10$ of log wages, LE

Notes: Comparison of the log wage profile in the HRS sample (dashed line) and the simulations from the estimated model (solid line). The upper panel plots the overall dispersion measured by $P90 - P10$ for the subpopulation with high education (HE) and low education (LE), respectively. The lower panel decomposes the overall dispersion into upper dispersion $P90 - P50$ and lower dispersion $P50 - P10$ for HE and LE, respectively.

Next, we check how well our model can reproduce the heterogeneity in wage change distributions for stayers and movers. Table 3 compares wage change distributions between the data and the simulations for both stayers and movers in each education group. We can see that our model fits the data both qualitatively and quantitatively. Compared with stayers, the wages of movers are more volatile. The median movers face a wage decline. Additionally, the distribution of movers is left-skewed, with the $10th$ percentile being more than 1.5 times larger than the $90th$ percentile in absolute value.

Table 3: Percentiles of log wage changes

	HE				LE			
	Stayers		Movers		Stayers		Movers	
	Data	Model	Data	Model	Data	Model	Data	Model
P10	-0.168	-0.187	-0.583	-0.622	-0.118	-0.139	-0.557	-0.606
P25	-0.055	-0.076	-0.266	-0.318	-0.044	-0.052	-0.237	-0.315
P50	0.005	0.007	-0.023	-0.042	0.001	0.022	-0.043	-0.062
P75	0.083	0.091	0.163	0.156	0.059	0.096	0.103	0.089
P90	0.203	0.201	0.373	0.377	0.15	0.181	0.254	0.234

Notes: Comparison of the distribution of log wage changes in the HRS sample (Data) and the simulations from the estimated model (Model) for job stayers and job movers. The left panel shows the distributions for the high education group (HE), and the right one shows the distributions for the low education group (LE).

A similar pattern can be found in the distribution of hour changes in Table 4. Our model reproduces the left-skewed and more dispersive hour changes for movers. Among the movers, more than 10% reduce their work hours by at least half, whereas the $10th$ percentile of stayers reduce them by around 20%.

Table 4: Percentiles of log hour changes

	HE				LE			
	Stayers		Movers		Stayers		Movers	
	Data	Model	Data	Model	Data	Model	Data	Model
P10	-0.218	-0.363	-0.601	-0.601	-0.207	-0.361	-0.634	-0.638
P25	-0.085	-0.208	-0.282	-0.302	-0.061	-0.207	-0.266	-0.339
P50	-0.012	-0.037	-0.051	-0.044	-0.008	-0.037	-0.015	-0.075
P75	0.036	0.135	0.052	0.188	0.016	0.134	0.043	0.16
P90	0.173	0.288	0.236	0.394	0.166	0.288	0.262	0.365

Notes: Comparison of the distribution of log hour changes in the HRS sample (Data) and the simulations from the estimated model (Model) for job stayers and job movers. The left panel shows the distributions for the high education group (HE), and the right one shows the distributions for the low education group (LE).

In Appendix E, we also provide figures about conditional job change rates, job destruction rates, distributions of tenures, distributions of assets by age, and health profiles by age.

6.2 Model Implications

In this subsection, we first discuss the prevalence of mismatch shocks 1) by age groups and 2) by age groups and range of log wage changes. Next, we assess the relative importance of different risks and fixed effects for the dynamics of employment, wages and hours of work. We achieve this by simulating from alternative versions of the model in which particular shocks are switched off. Those simulations are compared with the mean and variance from the baseline model. Finally, we focus on the group of individuals who move to a "bridge" job, which in this exercise we define as any job that takes place after a job with a tenure longer than ten years.

Mismatch shocks. Table 5 reports the probability of receiving mismatch shock for employed workers ($E_{i,t-1} = 1$) in each education and age group at each period. We can see that people in their 60s are more likely to receive mismatch shocks than those in their 50s. The probability of receiving mismatch shocks per period ranges from 0.4% and 0.8% for LE and HE workers aged between 51 and 55 to 3.1% and 4.2% for LE and HE workers aged between 66 and 70. For all age groups, the mismatch probability for HE workers is larger than for the LE. We also compute the probability of receiving at least one mismatch shock for employed workers by age 65 and 70, and the numbers are 10.3% and 13.5%, respectively.

Table 5: Proportion of $m_{it} = 1$ per period for employed workers by age group

	Age group			
	51 ~ 55	56 ~ 60	61 ~ 65	66 ~ 70
HE	0.008	0.025	0.044	0.042
LE	0.004	0.014	0.027	0.031

Notes: Probability of receiving mismatch shocks per period for the employed by age and education group computed based on the simulations from the estimated model.

Table 6 shows the mismatch probability for the movers ($jc_{it} = 1$) whose log wage changes are below a given threshold k. For instance, around 18% of job movers who experience a wage cut greater than 50% and who are aged between 51 and 55 have received a mismatch shock. For movers older than 60 with a similar wage cut, the proportion is as high as 75%. The results indicate that mismatch shocks are important to explain job movements toward worse-paying jobs despite their relatively low probability of occurrence.

Table 6: Proportion of $m_{it} = 1$ for movers with $\Delta \ln(wage_t) \leq k$ by age groups

	Age group			
k	$51 \sim 55$	$56 \sim 60$	$61 \sim 65$	$66 \sim 70$
-0.1	0.098	0.308	0.564	0.632
-0.3	0.135	0.402	0.648	0.699
-0.5	0.184	0.482	0.719	0.741
-0.7	0.212	0.553	0.774	0.75

Notes: Proportion of mismatched workers among job movers whose log wage changes is smaller than threshold k. Results computed based on the simulations from the estimated model.

Relative importance. We conduct simulations to measure the relative importance of various sources of risks and unobserved fixed effects for the patterns of employment, job mobility, wages, and working hours. In each simulation, we remove certain risks by setting the variance of the shock to zero or remove the unobserved fixed effects by fixing their values to the median. Then we compare the simulated employment rate, employment transition, job change rate, and both the mean and the variance of wages, wage changes, hours, and hour changes with those from the baseline model.

Table 7 presents the comparisons by age groups. The top panel is for the 51 to 60 years old, and the bottom one is for 61 to 70 years old. In each panel, the first row shows the results for the baseline model. Entries from the second row to the last display the ratios relative to the baseline model (the first row). First, we look at mismatch shocks. Table 7 says that removing mismatch shocks barely affects the overall labor force participation (LFP), and the mean and variance of wages for both age groups, but it reduces the cross-sectional dispersion in working hours for both age groups. In both groups, there is more non-employment to employment transition and less job-to-job movement. However, what is more interesting is its effect on wage changes and hour changes for the movers. As we can see, without mismatch shocks, the average wage changes and hour changes increase for both age groups. The dispersion of wage changes and hour changes are smaller too. These effects are mainly driven by the bottom percentiles, as the $P10$ of wage changes increases by 17% and 23% for the younger and older group respectively, and the $P10$ of hours changes increases by 28% and 49%, whereas the $P90$ is almost unchanged in all cases. Additionally, this pattern is not seen in other simulation exercises. The results illustrate the importance of mismatch shocks in explaining the job movements of older workers toward worse-paying jobs from a different angle.

Table 7: Relative importance of different risks and initial conditions (continues)

A. Age group 51 to 60

	Employment				w, all		h, all		Δw, movers				Δh, movers			
	LFP	E-E	NE-E	JC	*Mean*	*Var*	*Mean*	*Var*	*Mean*	*Var*	*P*10	*P*90	*Mean*	*Var*	*P*10	*P*90
Baseline	0.87	0.87	0.29	0.11	3.16	0.36	7.81	0.1	-0.09	0.15	-0.58	0.31	-0.06	0.16	-0.53	0.39
No mismatch shocks	1.0	1.0	1.32	0.87	1.0	0.98	1.0	0.84	0.55	0.83	0.83	1.03	-0.04	0.58	0.72	0.99
No jd shocks	1.01	1.01	1.2	0.89	1.0	1.0	1.0	1.0	0.82	0.94	0.93	1.01	1.06	1.02	1.04	1.0
No offer shocks	1.0	1.0	1.19	1.27	1.0	0.97	1.0	0.87	0.96	0.52	0.8	0.61	0.68	0.68	0.86	0.92
No productivity shocks	1.0	1.0	1.22	1.01	1.0	0.93	1.0	0.99	1.07	0.79	0.94	0.79	0.99	0.99	1.0	1.01
Median σ	1.0	1.0	1.22	1.0	1.0	0.95	1.0	1.0	1.03	0.86	0.95	0.87	1.0	1.0	0.99	1.01
No health shocks	1.02	1.03	1.35	1.04	1.0	1.0	1.0	1.0	1.01	0.96	0.99	0.98	0.88	0.97	0.98	1.01
No fix effect	1.0	1.0	1.23	0.99	1.0	0.85	1.0	0.99	1.01	1.18	1.06	1.2	1.01	0.98	0.99	1.0

B. Age group 61 to 70

	Employment				w, all		h, all		Δw, movers				Δh, movers			
	LFP	E-E	NE-E	JC	*Mean*	*Var*	*Mean*	*Var*	*Mean*	*Var*	*P*10	*P*90	*Mean*	*Var*	*P*10	*P*90
Baseline	0.56	0.69	0.11	0.08	3.0	0.39	7.57	0.17	-0.2	0.19	-0.73	0.27	-0.22	0.23	-0.86	0.33
No mismatch shocks	1.02	0.99	1.41	0.61	1.01	0.92	1.01	0.57	0.58	0.72	0.77	1.04	0.29	0.36	0.51	0.92
No jd shocks	1.01	1.01	1.12	0.9	1.0	1.0	1.0	0.99	0.97	0.96	0.99	1.02	1.13	1.07	1.08	0.93
No offer shocks	1.0	1.0	1.09	1.35	0.99	0.94	1.0	0.74	0.75	0.51	0.79	0.61	0.76	0.59	0.77	0.87
No productivity shocks	1.0	1.0	1.14	1.03	1.0	0.88	1.0	0.99	0.97	0.71	0.93	0.69	1.03	1.03	1.03	1.03
Median σ	1.0	1.0	1.15	1.02	1.0	0.91	1.0	1.0	0.93	0.81	0.91	0.8	1.07	1.03	1.07	0.98
No health shocks	1.1	1.06	1.32	1.08	1.0	0.99	1.0	0.98	0.96	0.94	0.99	0.95	0.98	1.01	1.0	1.01
No fix effect	1.0	1.0	1.14	1.01	1.0	0.83	1.0	0.99	0.93	1.04	1.0	1.15	1.06	1.0	1.03	0.95

Notes: In both panels, entries in the second row to the last display the ratios relative to the Baseline (first row). The variables w and h are log wages and log hours, respectively.

The third row shows the result of removing job destruction shocks. As before, there is no significant effect on the overall employment rate, the mean and variance of wages and hours of work. The job change rate is slightly lower. What is different is its effect on wage and hour changes compared with mismatch shocks. Even though both job destruction shocks and mismatch shocks force workers to leave their jobs, mismatch shocks have persistent negative effects on firm-specific productivity and affect future job offers. For the younger group, removing job destruction shocks reduces the wage cut for movers by 18%, smaller than the case without mismatch shocks. Additionally, the hours changes for both groups and the wage changes for the older group do not vary by much.

Shutting down the uncertainty in job offers increases the job change rate by 27% for the younger group and by 35% for the older group, which are the highest rates among all the exercises. The variation in wage changes of movers is much smaller, but in contrast to mismatch shocks, it is more driven by the top percentiles as the chances of moving to better-paying jobs are smaller.

Table 7: –Continued

A. Age group 51 to 60								
	Δw, stayers				Δh, stayers			
	$Mean$	Var	$P10$	$P90$	$Mean$	Var	$P10$	$P90$
Baseline	0.01	0.04	-0.16	0.19	-0.03	0.06	-0.35	0.3
No mismatch shocks	0.98	0.99	1.0	1.0	1.01	1.0	1.0	1.0
No jd shocks	0.97	1.0	1.0	1.0	1.01	1.0	1.0	1.0
No offer shocks	0.89	1.0	1.01	0.99	0.99	1.0	1.0	1.0
No productivity shocks	1.0	0.14	0.49	0.56	1.03	1.0	1.0	1.0
Median σ	1.0	0.34	0.8	0.83	1.01	1.0	1.0	1.0
No health shocks	1.0	1.01	1.0	1.0	1.01	1.0	1.0	1.0
No fix effect	1.04	1.78	1.64	1.56	1.07	1.0	1.0	1.0

B. Age group 61 to 70								
	Δw, stayers				Δh, stayers			
	$Mean$	Var	$P10$	$P90$	$Mean$	Var	$P10$	$P90$
Baseline	0.01	0.05	-0.18	0.21	-0.09	0.06	-0.42	0.23
No mismatch shocks	0.94	1.0	1.01	1.0	1.0	1.0	1.0	1.0
No jd shocks	0.97	1.0	1.0	0.99	1.0	1.0	1.0	1.0
No offer shocks	0.96	1.0	1.0	0.99	1.0	1.0	1.0	1.0
No productivity shocks	1.05	0.11	0.41	0.51	0.99	1.0	1.0	1.01
Median σ	1.07	0.26	0.69	0.75	0.97	1.0	0.99	1.01
No health shocks	0.97	0.97	0.99	0.99	0.98	1.0	0.99	1.0
No fix effect	0.99	1.28	1.24	1.22	0.99	1.0	0.99	1.0

Notes: In both panels, entries in the second row to the last display the ratios relative to the Baseline (first row). The variables w and h are log wages and log hours, respectively.

Individual-specific productivity shocks are an important source of wage volatility for job stayers, because eliminating these shocks (fifth row) reduces significantly the variance of wage changes for stayers. A related exercise is to remove the heterogeneity in individual-specific productivity risk by setting the standard deviation of the shocks to the median for all individuals over all periods ($\sigma_{it} = Med(\sigma_{it}) \approx 0.08$). The results show that this has a large effect on wage volatility too (sixth row). We would expect this heterogeneity to be essential in generating the high kurtosis in wage changes for stayers. This is because the baseline specification allows for a group of individuals with very small risks who therefore have very stable wages across periods. Looking at the kurtosis of wage changes in the baseline specification for stayers only, the quantile-based Crow-Siddiqui measure (i.e., $\frac{P972.5-P2.5}{P75-P25}$) turns out to be 4.86 for the younger group and 5.13 for the older one (not in the table), whereas removing the heterogeneity leads to a value of 2.91 which is the kurtosis for the Gaussian distribution.

Finally, the unobserved fixed effect is the largest contributor to the overall wage variation in both age groups. After removing the fixed effect (eighth row), the variance of wages is reduced by 15% for the younger group and by 17% for the older one. This is a clear indication that unobserved heterogeneity matters even after including a rich specification of shocks. On the other hand, health shocks have small effects on all dimensions of our analysis.

Bridge jobs. Next, we look at workers who move to "bridge" jobs. These are jobs that bridge career employment with the full exit from the labor force. In the calculations below, we define any job with more than ten years of tenure as career employment.[42] We conduct the same simulation exercises, that is, removing risks and the unobserved fixed effects, with a specific focus on those who move to bridge jobs. We first compute the proportion of individuals with bridge jobs (first column). Then we look at their first bridge job after leaving a career job and separate them by whether it is a job-to-job movement (JC) or there is a non-employment gap in between (NE-E). We compute their proportions (second and third columns), mean and variance of the associated changes in wages and hours (fourth column to the last).

Table 8: Job, mean and variance of wage change conditional on tenure larger than 10yrs

	Proportion of IDVI			$E(\Delta w)$			$Var(\Delta w)$			$E(\Delta h)$			$Var(\Delta h)$		
	Bridge	JC	NE-E	Bridge	JC	NE-E	Bridge	JC	NE-E	Bridge	JC	NE-E	Bridge	JC	NE-E
Model	0.13	0.08	0.06	-0.49	-0.35	-0.67	0.25	0.19	0.28	-0.31	-0.2	-0.45	0.33	0.23	0.42
No mismatch shocks	0.88	0.67	1.2	0.78	0.7	0.71	0.66	0.79	0.57	0.32	-0.01	0.38	0.35	0.41	0.28
No jd shocks	0.89	0.83	0.96	0.99	0.95	0.98	1.03	0.98	1.03	1.09	1.24	0.97	1.04	1.07	1.01
No offer shocks	0.91	0.91	0.98	1.23	1.31	1.14	0.56	0.39	0.61	1.07	1.21	0.96	0.54	0.64	0.45
No productivity shocks	1.01	1.03	1.0	1.02	1.0	1.03	0.83	0.78	0.81	0.99	1.01	0.99	0.97	0.98	0.97
Median σ	1.0	1.01	0.99	1.03	0.99	1.05	0.9	0.84	0.88	1.02	1.05	1.01	1.0	1.0	1.0
No health shocks	1.04	1.08	0.99	0.98	1.0	1.0	0.99	0.94	1.05	0.95	0.95	0.98	0.98	0.97	1.01
No fix effect	1.0	0.99	1.0	0.99	1.0	0.98	1.12	1.08	1.19	1.02	1.05	0.99	1.0	1.0	1.0

Notes: Entries in the second row to the last display the ratios relative to the Baseline (first row). The variables w and h are log wages and log hours, respectively.

Table 8 shows that mismatch shocks are the most important element in explaining bridge job movements. Shutting down mismatch shocks reduces the proportion of individuals moving to bridge jobs and increases the wage and hour changes regardless of whether there is a non-employment spell or not.[43] Consistent with our previous analysis, we also see a less cross-sectional variation in the change of wages and hours.

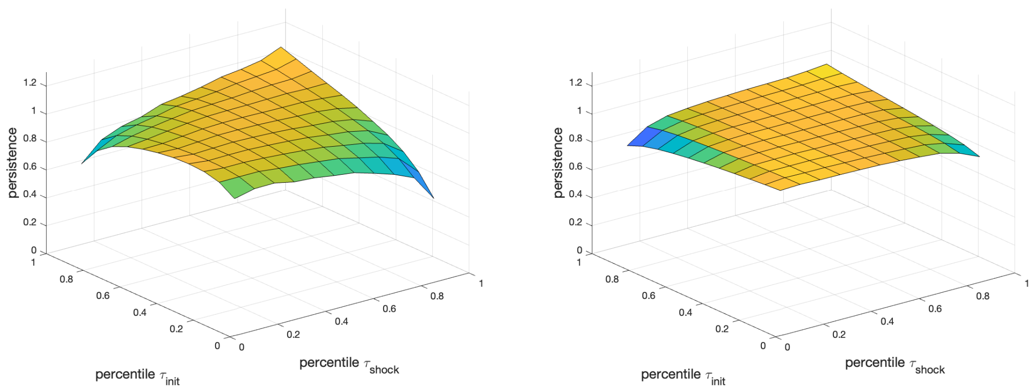
[42]A common definition of career employment is a full-time job lasting more than 10yrs (Brunello and Langella, 2013; Cahill et al., 2011). Here we do not require full-time. One reason is that the hour measure we use is the total annual working hours in all jobs. Without further information, we cannot identify whether someone is working in a full-time job or not even if the $exp(h_{it})$ is larger than 1600.

[43]Our proportion of individuals in bridge jobs is significantly lower than the literature (30% ~ 50%). This is related to the sample feature. First, we have unbalanced panel. Some individuals in our sample only have 3 waves. Second, we do not follow a specific cohort until their full exit from labor market. Some individuals are still in their career job during their last interview. Third, we include those who do not have career jobs. All these factors make it difficult to compare the number with the literature, but they do not affect the comparison between simulation exercises.

Nonlinear persistence in wages. The last feature that we check about the wage dynamics is its nonlinear persistence. Arellano et al. (2017) find that the earnings process features nonlinear persistence where the earnings are constructed from household labor earnings for male heads aged between 25 and 60. In our case, we focus on the wages of male workers at older ages from 51 to 70.[44] Therefore, we will first check whether the nonlinear persistence exists in our sample and then evaluate to what extent our model can generate the same pattern. Figure 7 plots the persistence for data and the simulations. We can see that our model qualitatively can capture some nonlinear persistence but not as much as we observe in the data.

Figure 7: Wage Persistence

(a) Data (b) Model



Notes: Comparison of the wage persistence in the HRS sample (a) and the simulations from the estimated model (b) following Arellano et al. (2017).

7 Quantifying The Welfare Costs of Risks and of Inflexibility in Transitioning to Retirement

In this section, we will quantify the welfare cost of risks and the value of a flexible transition to full retirement. To do this, we need to understand how people near retirement age make decisions about employment. Therefore, we construct and estimate a utility-based model of employment where agents face the same health, income, and employment risks as in the empirical model developed in Section 3. The structural model will serve us as a tool for analyzing counterfactuals and measuring changes in welfare. Moreover, the utility-based model can be treated as a version of the empirical model subject to structural

[44]In their appendix, they show that the empirical pattern is also present for male wages.

restrictions. This nested structure makes the empirical model a natural target on which to base the estimation of structural parameters. Accordingly, a new simulation-based estimation algorithm is discussed to implement the idea.

7.1 Model

In this model, agents start their life at age 50 (t=1) and live at most up to 90. Each period lasts for two years: t = 1,2,...,T, with T = 20. People are ex-ante heterogeneous in education level edu, health status h, individual productivity ω, volatility of individual productivity σ, firm-specific productivity υ, initial employment status E and asset A. At each period, they seek to optimize their expected discounted lifetime utility by choosing consumption level C and employment status d ($d = 0$ if not work, $d = 1$ if keep working in the same job, $d = 2$ if move to a new job), facing the uncertainty of survival, health, income, and employment.

Preferences. Individuals derive utility from consumption and leisure. The within period utility function takes the following form

$$U(C_t, d_t, d_{t-1}, jd_t, h_t, \epsilon_t^d) = \frac{1}{1-\nu} C_t^{1-\nu} + L_t \tag{27}$$

where L_t is the disutility of working. The disutility of working depends on people's age, health, and employment history. Specifically, working (E = 1) induces disutility.[45] There is also an extra cost for re-entry (RE = 1) and job changes (JC = 1).[46] All costs depend linearly on health. In addition, we use a piecewise linear function for variable age t, allowing for different preference forms before and after age 62, which is the earliest retirement age.[47] The last component of disutility is the idiosyncratic preference shocks associated with the employment choices ϵ_t^d. It is specified as i.i.d. over time and distributed Type-I extreme

[45] Note there is a direct mapping between employment status variables E and d: $E_t = \mathbb{1}\{d_t = 1 \text{ or } d_t = 2\}$

[46] We count it as a re-entry when one becomes employed from non-employment or from an involuntary job separation or from mismatch leave: $RE_t = \mathbb{1}\{(E_{t-1} = 0, E_t = 1) \text{ or } (E_{t-1} = 1, jd_t = 1, E_t = 1) \text{ or } (E_{t-1} = 1, m_t = 1, E_t = 1)\}$. Job change happens when one moves to a new job without being non-employed or being involuntarily separated from the old job or suffering mismatch shocks: $JC = \mathbb{1}\{E_{t-1} = 1, jd_t = 0, m_t = 0, E_t = 1\}$

[47] An alternative to piecewise linear function is polynomials. In practice the piecewise linear function fits better.

Value with zero mean.[48] Therefore the specification for L_t is given by:

$$\begin{aligned} L_t = &- (\theta_{e0} + \theta_{e1}t + \theta_{e2}\mathbb{1}(t > 6)(t-6) + \theta_{e3}h_t)E_t \\ &- (\theta_{r0} + \theta_{r1}t + \theta_{r2}\mathbb{1}(t > 6)(t-6) + \theta_{r3}h_t)RE_t \\ &- (\theta_{j0} + \theta_{j1}t + \theta_{j2}\mathbb{1}(t > 6)(t-6) + \theta_{j3}h_t)JC_t + \epsilon_t^d \end{aligned} \tag{28}$$

Individuals also derive utility from leaving a bequest if they die at period t. The bequest function is of the form

$$b(A_t) = \kappa A_t \mathbb{1}(A_t > 0) \tag{29}$$

Parameter κ captures the intensity of leaving bequest.

Health And Survival Rate. Health dynamics is an exogenous process that only depends on health status in the last period, age, and education. We use the same functional form as in the empirical model (equation 4). Moreover, health also affects the survival rate. The probability of being able to survive from $t-1$ to t is a function of age and health:

$$E(s_t) = f_s(h_{t-1}, t) \tag{30}$$

where $s_t \in \{0, 1\}$ is the indicator of survival.

Labor Income And Employment Shocks. Labor income is a product of the hourly wage W_t and working hours N. In the model, the amount of hour supply is fixed. For each education group, it is set to be the median working hours.[49]

$$Y_t = W_t \times N \tag{31}$$

The log hourly wage rate $\ln W_t$ takes the same form as in the empirical model, Equation (2). We treat transitory shocks as measurement errors and hence exclude them from the structural analysis. In addition, to reduce the dimensionality of state variables, we combine unobserved heterogeneity μ, tenure accumulation $ten'\gamma_e$, and the firm-specific component ν (ν') altogether, and define a composite firm component $\widetilde{v}$ ($\widetilde{v}'$).[50] More details about the construction of this composite term can be found in Appendix G. Thus,

$$\ln W_t = X_t'\gamma_X + h_t\gamma_h + \omega_t + \widetilde{v}_{jt} \tag{32}$$

[48]Note that with Type-I extreme Value assumption, the policy function of employment and job dynamics can be rewritten as logistic regressions. As long as the functional form inside the Probit regression in the non-utility-based model is flexible enough, we can still regard the non-utility-based model as an approximate reduced form of the utility-based one.

[49]For HE, N is 5154, which is equivalent to 2577 hrs per year. For LE, the number is 5142, which is 2571 hrs per year.

[50]To keep the name and notation simpler, I continue using *firm component* for $\widetilde{v}_{jt}$, even though now it contains individual unobserved heterogeneity as well.

The firm component $\widetilde{\upsilon}_{jt}$ evolves in a deterministic way during the tenure of a given job. As described in equation (33), it is a function of its value in the previous period only.[51] Once an individual takes a new offer, its value is replaced by the new one $\widetilde{\upsilon}'_{j(t)}$. Each period, individuals receive a new offer which is drawn from a distribution that depends on their previous firm component $\widetilde{\upsilon}_{j,t-1}$, mismatch shocks m_{jt}, job destruction shocks jd_{it}, and employment status in the last period E_{t-1}:

$$\widetilde{\upsilon}_{j(t)} = \begin{cases} \rho_{\upsilon 0} + \rho_{\upsilon}\widetilde{\upsilon}_{j(t-1)} & \text{if stay at the same job} \\ \widetilde{\upsilon}'_{j(t)} & \text{if move to new job} \end{cases} \tag{33}$$

$$\widetilde{\upsilon}'_{j(t)} \sim f_{\upsilon'}(\widetilde{\upsilon}_{j,t-1}, m_t, jd_{it}, E_{t-1}) \tag{34}$$

Mismatch shocks $m_{jt} \in \{0,1\}$ follow the same dynamics as in equation (10), only with $\nu_{j,t-1}$ replaced by $\widetilde{\upsilon}_{j,t-1}$. Workers who receive mismatch shocks will face a substantial step-decrease in the wage in their current job such that in practice they will either choose to take the new offer or become non-employed. However, they can not fully escape from the negative effect of mismatch shocks by jumping to a new job because the shocks also affect the average quality of new offers. People who are not working can also receive mismatch shocks. This mimics the depreciation of job-related productivity during non-employment, and similarly, it affects offers:

$$E(m_{jt}) = f_m(t, edu, \widetilde{\upsilon}_{t-1}, E_{t-1}) \tag{35}$$

Similarly, the process of job destruction jd_t, which depends on age, education, firm component $\widetilde{\upsilon}_{t-1}$ and mismatch shocks m_t, is an approximation to equation (**??**):

$$E(jd_t) = f_{jd}(t, edu, \widetilde{\upsilon}_{t-1}, m_t) \tag{36}$$

Other components, including the individual component ω_t and the productivity risks σ_t are the same as in the empirical model (equations 6 and 7).

Social Security. According to the rules of the Social Security Administration, social security benefits depend on Average Indexed Monthly Earnings (AIME), which summarizes the 35 years that represent an individual's top earnings. The age at which people retire and the employment status after receiving social security benefits could also affect ones' benefits.

That being said, we do not model the process of application for social security benefits for simplicity. We assume that everyone starts collecting social security benefits ss_t once

[51] We tried more flexible functional forms for $\widetilde{\upsilon}_{jt} = f(\widetilde{\upsilon}_{j(t-1)})$. It turns out AR(1) model can already approximate the dynamics well, with $R^2 = 0.999$.

they reach 65 years of age regardless of their employment status, that is $B_t = \mathbb{1}\{t \geq 8\}$. We also simplify the dynamics of ss_t: they will receive a constant payment, which only depends on the education group, until death.[52]

Budget Constraint. During period t, individuals potentially receive income from four different sources: asset income rA_t, labor earnings Y_t, social security benefits ss_t and government transfers tr_t. The only expenditure comes from consumption. Assets are accumulated as follows:

$$A_{t+1} = (1+r)A_t + Y_t + ss_t \times B_t + tr_t - C_t \tag{37}$$

We assume there is a borrowing constraint $A_{t+1} \geq A_{min}$. The minimum amount of assets, A_{min}, takes the lowest value observed in the data, which is around $-\$40,000$. Government transfers provide a consumption floor that guarantees a minimum amount of consumption C_{min}.

$$tr_t = \max\{0, C_{min} - ((1+r)A_t + Y_t + ss_t \times B_t - A_{min})\}. \tag{38}$$

where C_{min} is set to be $\$10,000$, which amounts to $\$5,000$ per year. The number lies in between the range suggested by the literature.[53]

Choice Set. At the beginning of each period, individuals receive shocks to survival, health, productivity, employment, and preferences for work, as well as an outside offer. Then they will make decisions about consumption and employment status. However, the options available depend on their state variables.

1. For those who are working ($E_{t-1} = 1$, $jd_t = 0$, $m_t = 0$), they can choose among three options: quit the job and become non-employed ($d_t = 0$), keep the old job ($d_t = 1$) or move to the new job ($d_t = 2$)
2. For those who are not working (maybe because they were non-employed $E_{t-1} = 0$, or their job is destroyed $jd_t = 1$, or they leave their old job as a consequence of being mismatched $m_t = 1$), they can only choose between two options: stay non-employed ($d_t = 0$), take the outside offer ($d_t = 2$)

Value Function. Let Ω_t denote the set of state variables, namely $\Omega_t = (A_{t-1}, \widetilde{v}_{t-1}, \widetilde{v}'_t, \omega_t, \sigma_t, h_t, m_{j,t}, t, d_{t-1}, jd_t, edu, \epsilon^d_t)$. Individuals make optimal decision of consumption

[52] In the later calibration part, ss_t is set to be the median social security benefits of each education group.
[53] Bueren (2018); French and Jones (2011)

C_t and employment d_t based on the following value function

$$\begin{aligned} V_t(\Omega_t) = \max_{C_t, d_t} \Big\{ & U(C_t, d_t, d_{t-1}, jd_t, h_t, \epsilon_t^d) + \beta(1 - s_{t+1}) b(A_{t+1}) \\ & + \beta s_{t+1} \mathbb{E}\Big(V_{t+1}(\Omega_{t+1}) \Big| \Omega_t, C_t, d_t \Big) \Big\} \end{aligned} \tag{39}$$

subject to equations (30) - (38), (6), (7) and (4).

7.2 Estimation Procedure

We propose a novel simulation-based estimation method that exploits the empirical model. The method is motivated by the following premise: the empirical (or non-utility-based, NU) model and the structural (or utility-based, U) model share the same wage equation and latent variable dynamics, and the NU employment and job transitions can be treated as an approximate reduced form of the U ones. The main idea is to take advantage of the estimated NU model, especially information on latent variables.

Specifically, the parameters are chosen such that the U model best approximates the estimated NU model in terms of Kullback-Leibler divergence. Denote NU model likelihood by $f_{NU}(Y, Z; \Theta)$, where Y is the set of all observables, Z is the set of all latent variables, Θ denotes all parameters. Correspondingly, $\hat{\Theta}$ is the NU model estimates we have obtained. Denote parameter set of U model by Ω, then the proposed estimator $\hat{\Omega}$ can be expressed by the following equation:

$$\hat{\Omega} = \arg\max_{\Omega} \sum \sum_{\widetilde{Z}} \ln f_{NU}(Y, \widetilde{Z}; \Theta(\Omega))$$

where $\widetilde{Z} \sim f_{NU}(Z|Y; \hat{\Theta})$, and $\Theta(\Omega) = \arg\max_{\Theta} \sum \sum_s \ln f_{NU}(Y^s(\Omega), Z^s(\Omega); \Theta)$ is a mapping from the U model parameters to NU model parameters.

It is straightforward that we make use of information on latent variables learned from NU model through the draws $\widetilde{Z}$. Additionally, under the premises, this method allows us to directly bring the NU model results of the wage equation and the latent components to U model as input.[54,55]

[54]With the separability of parameters, we can rewrite the objective as $f_{NU}(Y, \widetilde{Z}; \Theta(\Omega_2, \Omega_1)) = f(Y_{-W}|\widetilde{Z}; \Theta(\Omega_2)) f(W, \widetilde{Z}; \Theta(\Omega_1))$ where Ω_1 denotes parameters in common parts of U and NU models, that is wage equation and latent variable dynamics, Ω_2 denotes remaining parameters, W is the wages, and Y_{-W} is all observables other than wages. Due to $\Theta(\Omega_1) = \Omega_1 = \Theta_1$, maximization requires $\hat{\Omega}_1 = \hat{\Theta}_1$.

[55]Even though this method is motivated by the premises discussed before, the application can be extended to cases where nest structure is not satisfied as shown in Appendix I. However, we may not be able to directly use part of the NU model estimates when the premises are not met.

This method improves algorithmic efficiency compared with estimating the U model as a whole due to a reduced vector of unknowns. Moreover, the primitive parameters, that are estimated under the more flexible NU model, are less prone to misspecification problems. On the other hand, the flexibility of the NU model makes it suitable as a tool to measure the difference between the U model with real data — its flexibility enriches the dimensions of the comparison. Finally, the estimator is expected to be statistically more efficient, given the inclusion of extra information on latent variables learned from the NU model.[56]

In this application, the method is executed in two steps. In the first step, we estimate the survival probability jointly with health dynamics for people between 51 and 90.[57] An extra approximation is required due to the composite firm component $\widetilde{\upsilon}$. Denote all the primitive parameters by $\hat{\Omega}_1$ including those in the wage and hours equations, and the dynamics of health, productivity (both ω and $\widetilde{\upsilon}$), offer, and job separation. Further details can be found in Appendix H.

In the second step, we estimate the remaining parameters $\Omega_2 \equiv (\nu, \theta_e, \theta_r, \theta_j, \kappa, \beta)$ by solving the following equation.

$$\hat{\Omega}_2 = \arg\max_{\Omega_2} \sum \sum_{\widetilde{Z}} \ln f_{NU}(Y, \widetilde{Z}; \Theta(\Omega_2, \hat{\Omega}_1))$$

In practice, we start with some initial guess of Ω_2. Given $\hat{\Omega}_1$ and Ω_2, we simulate M statistically independent data sets from the U model: $\{Y, Z\}^m$, $m = 1, ..., M$, where each data set consists of N_M individuals and T_M periods. Then we compute $\Theta(\Omega_2, \hat{\Omega}_1) = \frac{1}{M} \sum \hat{\Theta}^m(\Omega_2, \hat{\Omega}_1)$, where $\hat{\Theta}^m$ is the estimator for each of the M simulated data sets: $\hat{\Theta}^m(\Omega_2, \hat{\Omega}_1) = \arg\max_\Theta \ln f_{NU}(Y^m, Z^m; \Theta)$. Finally, we evaluate the objective function $\sum \sum_{\widetilde{Z}} \ln f_{NU}(Y, \widetilde{Z}; \Theta(\Omega_2, \hat{\Omega}_1))$. Our estimator $\hat{\Omega}_2$ is generated by choosing the value of Ω_2 that maximizes the objective function.[58]

Note that the part of the objective function related to wages and latent variables remains unchanged between iterations.[59] It is the employment transitions E-E and NE-E,

[56]Since we ignore approximation errors in the specification of the empirical model as a proper reduced form, it is important that the empirical model fits the data well as we hope to be the case in light of our results.

[57]The estimation of health is similar to the empirical part, except that now we expand to 90 years old and add survival probability.

[58]We choose $M = 50$, $N_M = 10,000$ and $T_M = 6$.

[59]For example, we could separate latent variables and have $f_{NU}(Y, \widetilde{Z}; \Theta(\Omega_2, \hat{\Omega}_1)) = f(Y|\widetilde{Z}; \Theta(\Omega_2, \hat{\Omega}_1)) f(\widetilde{Z}; \hat{\Omega}_1)$ where the second component does not change. Similar decomposition works with wage equations too.

job changes JC as well as asset accumulation, which are playing an active role.[60]

It can be shown that in the case when the NU model increases with sample size ($N, T \to \infty$), both NU model and U model estimators are consistent and the standard errors are around the true values. If the NU model does not increase with sample size, then we have a pseudo-maximum likelihood estimator and the standard errors are around pseudo-true values. A more detailed discussion about the method can be found in Appendices H and I.

Relation to Indirect Inference. Our method is closely related to but different from the standard Indirect Inference method. The main difference is that we make use of NU model estimation results. Instead of choosing parameters by comparing simulations with data, our method compares the simulations with the NU model. On the other hand, if one chooses the NU model as the auxiliary model to conduct Indirect Inference, then the estimator which takes LR metric can be expressed as follows:

$$\hat{\Omega}_{II} = \arg\max_{\Omega} \sum \sum_{\tilde{\tilde{Z}}} \ln f_{NU}(Y, \tilde{\tilde{Z}}; \Theta(\Omega))$$

where $\tilde{\tilde{Z}}$ is drawn from posterior $f_{NU}(\tilde{\tilde{Z}}|Y; \Theta(\Omega))$. Using our method, we draw latent variable from the posterior distribution given NU model estimates of real data $\hat{\Theta}$, whereas with I-I method, it is the estimates of simulated data $\Theta(\Omega)$ that change with iterations.

7.3 Estimation Results

Table 9 lists the estimated preference parameters. The discount factor implied by our model is 0.899 for HE and 0.881 for LE: the HE is more patient than the LE. The parameter, which is identified from both consumption and labor choices, is slightly lower than other estimates.[61] The coefficient of risk aversion is 1.666 for HE and 1.891 for LE. The results are similar to the literature.[62]

The estimates of the non-pecuniary part of utility show that the cost of working increases with age. People in good health (small h) face fewer working costs than those in bad health (large h). The relatively high cost of re-entry and job change corresponds to the low RE and JC rates in the data.

[60] The asset accumulation in our structural model is very simplified. We had problems matching the original asset equation in the NU model, which depends on many observables and latent variables. In the end, we change to match only the quantiles (0.2, 0.4, 0.6, 0.8) of assets by age group (51-52, 53-54,...).

[61] French and Jones (2011) has the estimates that vary with the type from 0.86 to 1.12, and the average equals 0.91.

[62] Van der Klaauw and Wolpin (2008); Blau and Gilleskie (2008); Rust and Phelan (1997); Haan and Prowse (2014); Wen (2018)

Table 9: Preference Parameter Estimates

	Parameters	HE	LE		Parameters	HE	LE
θ_{e0} :	Cost of working	-0.347	-0.212	θ_{j0} :	Cost of job change	2.295	2.005
θ_{e1} :	Cost of working: age dependent ($\times t$)	0.106	0.04	θ_{j1} :	Cost of job change: age dependent ($\times t$)	-0.003	0.11
θ_{e2} :	Extra cost of working for 60+: age dependent ($\times t$)	0.009	0.061	θ_{j2} :	Extra cost of job change: for 60+: age dependent ($\times t$)	-0.044	0.002
θ_{e3} :	Cost of working: health dependent ($\times h$)	0.446	0.373	θ_{j3} :	Cost of job change: health dependent ($\times h$)	0.01	0.011
θ_{r0} :	Extra cost when reentering labor market	1.925	0.978				
θ_{r1} :	Reentry cost: age dependent ($\times t$)	0.147	0.328	ν :	Coef. risk aversion	1.666	1.896
θ_{r2} :	Extra Reentry cost for 60+: age dependent ($\times t$)	-0.188	-0.417	κ :	Bequest intensity	0.029	0.037
θ_{r3} :	Reentry cost: health dependent ($\times h$)	-0.055	-0.084	β :	Discount factor	0.899	0.881

Notes: Table shows the point estimation of parameters.

Besides the point estimation, we also plot the simulated age profile of employment, job change, asset, and wage against the data in Appendix J. Simulated profiles, in general, match the characteristics of the data quite well.

7.4 Welfare Calculation

In this section, we implement two exercises aimed at providing answers to two different questions. The first one is about the welfare cost of risks. We focus on mismatch shocks and job destruction shocks. Specifically, we compute the welfare changes by comparing the baseline world to the alternative where certain risk is eliminated. The second question is about the value of bridge jobs. We compute the welfare cost of living in an alternative economy where re-entry and job changes are not allowed after age 65. The exercise speaks of how much people value the probability of a flexible transition to their full retirement.

Two measures are formed to quantify the welfare gain/loss. The first one is a lump-sum transfer of the asset, ΔA, received at age 55/56 ($t = 3$). The amount is the equivalent variations in the asset such that people are indifferent between the baseline economy and the alternative one. It can be expressed as follows:

$$V_3(A_2 + \Delta A, \Omega_3 \backslash A_2) = \widetilde{V}_3(\Omega_3)$$

where V_t represents the expected discounted utility given a certain set of state variables at t in the baseline world, and $\widetilde{V}_t$ represents the alternative world. Symbol $\Omega_3 \backslash A_2$ denotes all state variables in Ω_3 but A_2. Therefore, the asset transfer, ΔA, is defined as the amount that people receive in the baseline world at $t = 3$ such that the expected discounted utility is the same as in the alternative world.

The second measure is based on consumption flows. It is defined as the proportion of optimal consumption, π, at all ages since 55/56($t \geq 3$) that people receive such that they have the same expected discounted utility as in the alternative world.

$$V_3(\Omega_3) + \sum_{t=3} \beta^{t-3} E_3\left(s(t)\frac{1}{1-\nu}(\pi C_t^*)^{1-\nu}\right) = \widetilde{V}_3(\Omega_3)$$

where $s(t) = (\prod_{k=3}^{k=t} s_k)/s_3$ is the survival rate until period t conditional on being alive at period 3, and C_t^* is the policy function for consumption in the baseline world. The LHS of the equality above has two parts: the value in the baseline world and the extra expected utility from π more consumption. The RHS is the value of living in the alternative world.

Both measures can be either positive or negative. A positive value means that moving to the alternative world improves welfare, while a negative one implies the opposite. However, the sign and magnitude can vary from person to person depending on their state variables.

7.4.1 Welfare cost of risks

Table 10 presents the welfare gains of eliminating mismatch and employment risks in the first four and last four columns, respectively. The welfare cost of risks is computed for high and low education groups separately. Additionally, we expect the welfare changes to be heterogeneous, as people with different state variables react differently to the experiments. Hence, we report the distribution of the welfare gains by displaying the $10th$, the $50th$, and the $90th$ percentiles of the welfare gains in the first three rows, and the conditional median on different assets, employment status, and health levels starting from the $4th$ row.

Mismatch Shocks. Results in Table 10 show that mismatch shocks bring welfare loss to both education groups. The median welfare cost accounts for a lump sum transfer of assets around \$62,300 for HE and \$26,700 for LE. We also observe heterogeneities within each education group: the 90th percentile is around more than two times the 10th percentile for both groups. Furthermore, we calculate the welfare loss in each subgroup with different asset levels, employment status, tenure, wage rate, and health levels. The cost measured in transfers increases with assets: the number for the bottom

Table 10: Welfare Cost Of Mismatch And Employment Risks

	No mismatch risk				No jd risk			
	$\Delta A(\times\$10,000)$		$\pi(\%)$		$\Delta A(\times\$10,000)$		$\pi(\%)$	
	HE	LE	HE	LE	HE	LE	HE	LE
P10	3.98	1.45	5.9	4.34	1.69	1.78	2.71	5.41
P50	6.23	2.67	7.11	5.33	3.35	3.63	3.53	7.3
P90	8.62	4.57	7.99	6.18	5.03	6.63	4.31	9.29
By assets level								
$A_{t-1} \leq P33$	4.8	1.81	7.43	5.14	2.24	2.31	3.39	6.77
$P33 < A_{t-1} \leq P66$	6.2	2.61	6.91	5.18	3.31	3.52	3.5	7.1
$A_{t-1} > P66$	7.97	4.05	6.88	5.68	4.38	5.74	3.69	8.21
By employment status								
Non-employed	3.02	1.4	5.31	3.73	1.39	1.77	2.4	4.71
Employed, ten$\geq$ 10 yrs	6.66	3.18	7.09	5.56	3.69	4.43	3.72	7.87
Employed, ten$<$ 10 yrs	5.94	2.39	7.28	5.2	2.99	3.16	3.42	6.86
Employed, high wage ($\geq P50$)	7.2	3.53	7.04	5.77	4.21	5.09	3.99	8.4
Employed, low wage ($< P50$)	5.55	2.18	7.35	5.1	2.6	2.81	3.23	6.65
By health level								
Good ($h_{t-1} > P75$)	7.39	3.33	7.4	5.86	4.04	4.65	3.94	8.27
Far ($P25 < h_{t-1} \leq P75$)	6.27	2.76	7.07	5.38	3.39	3.79	3.55	7.4
Bad ($h_{t-1} \leq P25$)	5.08	1.97	6.54	4.87	2.49	2.64	3.19	6.43

Notes: The columns 1-4 show the distribution of welfare cost caused by mismatch risk measured by asset transfer ΔA and consumption flow π for the high educated (HE) and low educated (LE). The columns 5-8 show the distribution of welfare cost caused by job destruction risk measured by asset transfer ΔA and consumption flow π for the high educated (HE) and low educated (LE).

one-third is $\$48,000(\$18,100)$, and it increases to $\$79,700(\$40,500)$ for the top one-third. By employment status, we find that workers with higher tenure and higher wages suffer greater losses. Lastly, people in better health require more asset transfer.[63]

Next, we switch to the measure based on the consumption flow (columns 3-4). The median cost of mismatch shocks is equivalent to 7.11% consumption for HE, and 5.33% for LE. There is relatively less heterogeneity compared to the other measure: the 90th percentile is 7.99%(6.18%), and the 10th percentile is 5.9%(4.34%). What's more, the correlation between the cost and observables for HE is different from the previous result. The equivalent proportion of consumption decreases with asset level. Workers with lower tenure and lower wages face a higher cost. But in general, the differences between subgroups are minor.

[63]However, considering the multi-dimensional nature of health, one needs to be cautious in its interpretation

JD shocks. Columns 5 to 8 in Table 10 show the welfare changes of eliminating the probability of receiving job destruction shocks. The median welfare gain accounts for $33,500 lump sum transfer or 3.53% consumption for HE and $36,300 lump sum transfer or 7.3% consumption for LE. In terms of asset transfer, wealthier and healthier people, workers with higher tenure and higher wages gain more from a smaller jd risk. The pattern of consumption flow is similar except for the correlation with assets for the low educated.

7.4.2 Value of flexible employment transitions

Finally, we assess how much people value the possibility of a flexible transition to full retirement, which indirectly speaks of the value of bridge jobs. Specifically, in the alternative economy, people older than 65 are banned from moving to a new job, either through re-entry or job-to-job movement. Indeed, workers can only stay in their job until their full exit from the labor market: they can not change jobs even if they have better options. Similarly, the non-employed who recovered from previous health or productivity shocks can neither return to the labor market.

The first four columns of Table 11 show the cost of imposing such UE and JC restrictions. The median cost is around $107,300 for HE and $58,400 for LE. For the most affected, the loss can be higher than $174,000. Even for those less affected, the rigidities of the labor market cause a loss of around $29,800. If we measure the loss using consumption changes, then the negative effect is the amount to a $10\% - 17\%$ reduction in consumption. The consumption cost is slightly higher and more dispersive for LE. Lastly, conditional on subgroups, we find that wealthier, healthier, experienced workers face higher costs.

We further analyze the effect of mismatch shocks on the cost. By comparing the baseline world without mismatch shocks and the alternative world without mismatch shocks but with UE and JC restrictions, we find that the cost of the employment restrictions is slightly higher in all cases. This implies that when people are fully insured against the mismatch risks in an alternative world, the flexibility in the transition at the end of their career becomes more valuable.

Table 11: Value Of Flexible Transition To Full-retirement

	No UE/JC after 65 (to baseline)				No UE/JC after 65 + no mismatch risk (to baseline + no mismatch risk)			
	$\Delta A(\times\$10,000)$		$\pi(\%)$		$\Delta A(\times\$10,000)$		$\pi(\%)$	
	HE	LE	HE	LE	HE	LE	HE	LE
p10	-17.4	-13.4	-13.78	-16.79	-18.0	-13.63	-14.04	-16.92
p50	-10.73	-5.84	-12.65	-13.53	-11.13	-5.99	-12.84	-13.68
p90	-5.55	-2.98	-10.88	-10.52	-5.85	-3.07	-10.99	-10.61
By asset level								
$A_{t-1} \leq p30$	-6.87	-3.78	-11.38	-11.4	-7.14	-3.88	-11.51	-11.52
$p33 < A_{t-1} \leq p66$	-10.89	-5.78	-12.61	-13.23	-11.3	-5.94	-12.8	-13.37
$A_{t-1} > p66$	-15.74	-10.91	-13.51	-15.88	-16.3	-11.17	-13.76	-16.03
By employment status								
Non-employed	-4.99	-3.5	-10.99	-11.49	-5.2	-3.59	-11.12	-11.61
Employed, ten≥ 10	-12.0	-7.4	-12.88	-14.31	-12.46	-7.6	-13.08	-14.46
Employed, ten< 10	-9.1	-4.8	-12.29	-12.68	-9.4	-4.91	-12.46	-12.8
Employed, high wage ($\geq p50$)	-13.99	-8.95	-13.25	-15.08	-14.48	-9.18	-13.48	-15.23
Employed, low wage ($< p50$)	-8.34	-4.52	-11.93	-12.26	-8.57	-4.59	-12.07	-12.38
By health level								
Good ($h_{t-1} > p75$)	-13.03	-7.43	-13.03	-14.3	-13.5	-7.63	-13.25	-14.45
Far ($p25 < h_{t-1} \leq p75$)	-10.85	-6.26	-12.64	-13.76	-11.24	-6.42	-12.83	-13.9
Bad ($h_{t-1} \leq p25$)	-8.46	-4.38	-12.1	-12.32	-8.71	-4.46	-12.27	-12.46

Notes: The columns 1-4 show the distribution of welfare cost of banning job change and re-entry after age 65 for the high educated (HE) and low educated (LE), respectively. The columns 5-8 show the distribution of welfare cost of banning job change and re-entry after age 65 without mismatch risk (both in baseline economy and alternative economy) for the high educated (HE) and low educated (LE), respectively. The cost is measured by asset transfer ΔA and consumption flow π.

8 Conclusion

In this paper, I focus on the gradual transition to full retirement of older workers in the United States: instead of having a one-time and permanent withdrawal from the labor force, many older workers experience job movements, which often means moving towards worse-paying and less-demanding jobs. To explain this pattern, I propose an aging-related shock â mismatch shock, that mismatches workers with their exiting job. I built latent-variable wage and hour processes jointly with a flexible model of employment and job transitions to disentangle health risks, individual-specific productivity risks, firm-specific mismatch risks, quality of outside offers, and job destruction risks for older workers near retirement. The model is estimated by applying a simulation-based method in a companion paper of mine and using 50+ individuals in the HRS data. The results show that the mismatch shock could capture the left-skewness in wage and hour changes observed in the data for job movers. Indeed, the mismatch shock plays an important role in explaining the non-marginal reduction in wages and hours for movers.

Furthermore, I compute the welfare cost of risks and quantify how much individuals value the possibility of retiring smoothly in a utility-based structural model. Agents in the structural model face the same risks as in the empirical model. Moreover, the empirical model is flexible enough to be treated as a good approximation of the reduced form of the structural model. Taking advantage of these connections, I propose a novel simulation-based estimation algorithm for the structural model that exploits the results of the empirical model. The results show that the mismatch risk causes a non-negligible amount of welfare loss: the loss caused by mismatch risk is heterogeneous with a median loss amount of $\$62,300$ for the highly educated (HE) and $\$26,700$ for the lowly educated (LE), equivalent to a 7.11% reduction in consumption flow for HE and 5.33% for LE. Furthermore, results suggest that people value the flexibility in transitioning to full retirement: banning job changes and re-entry causes a median loss of around $\$107,300$ for HE and $\$58,400$ for LE, which is equivalent to a $12\% - 14\%$ reduction in consumption flow.

References

Abowd, John M and David Card (1989) "On the covariance structure of earnings and hours changes," *Econometrica: Journal of the Econometric Society*, 411–445.

Alcover, Carlos-Maria, Gabriela Topa, Emma Parry, Franco Fraccaroli, and Marco Depolo (2014) *Bridge employment: A research handbook*: Routledge.

Ali, Mukhtar M (1974) "Stochastic ordering and kurtosis measure," *Journal of the American Statistical Association*, 69 (346), 543–545.

Almuzara, Martín (2020) "Heterogeneity In Transitory Income Risk.," working papers, Federal Reserve Bank of New York.

Altonji, Joseph G, Anthony A Smith, and Ivan Vidangos (2013) "modelling earnings dynamics," *Econometrica*, 81 (4), 1395–1454.

Amemiya, Takeshi (1978) "On a two-step estimation of a multivariate logit model," *Journal of Econometrics*, 8 (1), 13 – 21, https://doi.org/10.1016/0304-4076(78)90086-6.

Amengual, Dante, Jesús Bueren, and Julio Crego (2017) "Endogenous health groups and heterogeneous dynamics of the elderly,"Technical report, HEDG, c/o Department of Economics, University of York.

Ameriks, John, Joseph Briggs, Andrew Caplin, Minjoon Lee, Matthew D Shapiro, and Christopher Tonetti (2020) "Older Americans would work longer if jobs were flexible," *American Economic Journal: Macroeconomics*, 12 (1), 174–209.

Arcidiacono, Peter and John Bailey Jones (2003) "Finite mixture distributions, sequential likelihood and the EM algorithm," *Econometrica*, 71 (3), 933–946.

Arellano, Manuel (2014) "Uncertainty, persistence, and heterogeneity: A panel data perspective," *Journal of the European Economic Association*, 12 (5), 1127–1153.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme (2017) "Earnings and consumption dynamics: a nonlinear panel data framework," *Econometrica*, 85 (3), 693–734.

Arellano, Manuel and Stéphane Bonhomme (2016) "Nonlinear panel data estimation via quantile regressions."

Berkovec, James and Steven Stern (1991) "Job exit behavior of older men," *Econometrica: Journal of the Econometric Society*, 189–210.

Blau, David M. (1994) "Labor Force Dynamics of Older Men," *Econometrica*, 62 (1), 117–156, http://www.jstor.org/stable/2951478.

Blau, David M and Donna B Gilleskie (2008) "The role of retiree health insurance in the employment behavior of older men," *International Economic Review*, 49 (2), 475–514.

Blundell, Richard W, Jack Britton, Monica Costa Dias, and Eric French (2016) "The dynamic effects of health on the employment of older workers."

——— (2017) "The impact of health on labor supply near retirement," *Available at SSRN 3103887*.

Bowlus, Audra, Ãmilien Gouin-Bonenfant, Huju Liu, Lance Lochner, and Youngmin Park (2021) "Four Decades of Canadian Earnings Inequality and Dynamics across Workers and Firms," Working Paper 28757, National Bureau of Economic Research, 10.3386/w28757.

Browning, Martin, Mette Ejrnaes, and Javier Alvarez (2010) "Modelling income processes with lots of heterogeneity," *The Review of Economic Studies*, 77 (4), 1353–1381.

Brunello, Giorgio and Monica Langella (2013) "Bridge jobs in Europe," *IZA Journal of Labor Policy*, 2 (1), 1–18.

Bueren, Jesús (2018) "Long-term care needs: Implications for savings, welfare, and public policy," *Wroking Paper. Accessed online on February*, 2, 2020.

Cahill, Kevin E, Michael D Giandrea, and Joseph F Quinn (2011) "Reentering the labor force after retirement," *Monthly Labor Review*, 134 (6), 34–42.

Cahill, Kevin E., Michael D. Giandrea, and Joseph F. Quinn (2018) "Is Bridge Job Activity Overstated?" *Work, Aging and Retirement*, 4, 330–351, 10.1093/workar/way006.

Cahill, Kevin E and Joseph F Quinn (2020) "The importance of gradual retirement in America today," *Public Policy & Aging Report*, 30 (3), 107–112.

Capatina, Elena, Michael P Keane, and Shiko Maruyama (2018) "Health Shocks and the Evolution of Consumption and Income over the Life-cycle," *UNSW Business School Research Paper*, 14.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline (2018) "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36 (S1), S13–S70.

Chamberlain, Gary and Keisuke Hirano (1999) "Predictive distributions based on longitudinal earnings data," *Annales d'Economie et de Statistique*, 211–242.

Choi, S. C. and R. Wette (1969) "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and Their Bias," *Technometrics*, 11 (4), 683–690, `http://www.jstor.org/stable/1266892`.

Cox, David R (1975) "Partial likelihood," *Biometrika*, 62 (2), 269–276.

De Nardi, Mariacristina, Eric French, and John B Jones (2010) "Why do the elderly save? The role of medical expenses," *Journal of political economy*, 118 (1), 39–75.

De Nardi, Mariacristina, Svetlana Pashchenko, and Ponpoje Porapakkarm (2017) "The lifetime costs of bad health,"Technical report, National Bureau of Economic Research.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977) "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, 39 (1), 1–22.

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

French, Eric (2005) "The effects of health, wealth, and wages on labour supply and retirement behaviour," *The Review of Economic Studies*, 72 (2), 395–427.

French, Eric and John Bailey Jones (2011) "The effects of health insurance and self-insurance on retirement behavior," *Econometrica*, 79 (3), 693–732.

——— (2017) "Health, health insurance, and retirement: a survey," *Annual Review of Economics*, 9, 383–409.

Friedrich, Benjamin, Lisa Laun, Costas Meghir, and Luigi Pistaferri (2019) "Earnings dynamics and firm-level shocks,"Technical report, National Bureau of Economic Research.

Ghisletta, Paolo, Patrick Rabbitt, Mary Lunn, and Ulman Lindenberger (2012) "Two thirds of the age-based changes in fluid and crystallized intelligence, perceptual speed, and memory in adulthood are shared," *Intelligence*, 40 (3), 260–268, https://doi.org/10.1016/j.intell.2012.02.008.

Gourieroux, Christian and Joann Jasiak (2006) "Autoregressive gamma processes," *Journal of forecasting*, 25 (2), 129–152.

Grossman, Michael (1972) "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80 (2), 223–255, http://www.jstor.org/stable/1830580.

Gustman, Alan L and Thomas L Steinmeier (1983) "A structural retirement model."

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song (2021) "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?" *Econometrica*, 89 (5), 2303–2339, https://doi.org/10.3982/ECTA14603.

Haan, Peter and Victoria Prowse (2014) "Longevity, life-cycle behavior and pension reform," *Journal of Econometrics*, 178, 582–601, https://doi.org/10.1016/j.jeconom.2013.08.038.

Halliday, Timothy J (2008) "Heterogeneity, state dependence and health," *The Econometrics Journal*, 11 (3), 499–516.

Halliday, Timothy J, Hui He, Lei Ning, and Hao Zhang (2019) "Health investment over the life-cycle," *Macroeconomic Dynamics*, 23 (1), 178–215.

Halvorsen, Elin, Hans Holter, Kjetil Storesletten, and Serdar Ozkan (2019) "Dissecting Idiosyncratic Income Risk,"Technical report.

Han, Shin-Kap and Phyllis Moen (1999) "Clocking out: Temporal patterning of retirement," *American journal of sociology*, 105 (1), 191–236.

Hause, John C (1980) "The fine structure of earnings and the on-the-job training hypothesis," *Econometrica (pre-1986)*, 48 (4), 1013.

Health and Retirement Study, "(RAND HRS Longitudinal File 2016) public use dataset," *Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740). Ann Arbor, MI.*

Heineck, Guido and Silke Anger (2010) "The returns to cognitive abilities and personality traits in Germany," *Labour Economics*, 17 (3), 535–546, https://doi.org/10.1016/j.labeco.2009.06.001.

Heiss, Florian (2011) "Dynamics of self-rated health and selective mortality," *Empirical economics*, 40 (1), 119–140.

Hokayem, Charles and James P Ziliak (2014) "Health, human capital, and life cycle labor supply," *American Economic Review*, 104 (5), 127–31.

Hospido, Laura (2012) "Modelling heterogeneity and dynamics in the volatility of individual wages," *Journal of Applied Econometrics*, 27 (3), 386–414.

Hosseini, Roozbeh, Karen A Kopecky, and Kai Zhao (2021) "The evolution of health over the life cycle," *Review of Economic Dynamics*.

Idler, Ellen L and Yael Benyamini (1997) "Self-rated health and mortality: a review of twenty-seven community studies," *Journal of health and social behavior*, 21–37.

Jacobs, Lindsay and Suphanit Piyapromdee (2016) "Labor Force Transitions at Older Ages: Burnout, Recovery, and Reverse Retirement."

Jolivet, Grégory and Fabien Postel-Vinay (2020) "A structural analysis of mental health and labor market trajectories."

Karahan, Fatih, Serdar Ozkan, and Jae Song (2019) "Anatomy of lifetime earnings inequality: Heterogeneity in job ladder risk vs. human capital," *FRB of New York Staff Report* (908).

Keane, Michael and Anthony A Smith (2003) "Generalized indirect inference for discrete choice models," *Yale University*.

Kim, Seongsu and Daniel C Feldman (2000) "Working in retirement: The antecedents of bridge employment and its consequences for quality of life in retirement," *Academy of management Journal*, 43 (6), 1195–1210.

Van der Klaauw, Wilbert and Kenneth I Wolpin (2008) "Social security and the retirement and savings behavior of low-income households," *Journal of econometrics*, 145 (1-2), 21–42.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler (2019) "Imperfect Competition, Compensating Differentials and Rent Sharing in the U.S. Labor Market," Working Paper 25954, National Bureau of Economic Research, 10.3386/w25954.

Lange, Fabian and Douglas McKee (2012) "The evolution of latent health over the life course,"Technical report, Citeseer.

Lillard, Lee A and Robert J Willis (1978) "Dynamic aspects of earning mobility," *Econometrica: Journal of the Econometric Society*, 985–1012.

Liu, Chuanhai, Donald B Rubin, and Ying Nian Wu (1998) "Parameter expansion to accelerate EM: the PX-EM algorithm," *Biometrika*, 85 (4), 755–770.

Low, Hamish, Costas Meghir, and Luigi Pistaferri (2010) "Wage risk and employment risk over the life cycle," *American Economic Review*, 100 (4), 1432–67.

Low, Hamish and Luigi Pistaferri (2015) "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off," *The American Economic Review*, 105 (10), 2986–3029, `http://www.jstor.org/stable/43821365`.

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

Maestas, Nicole (2010) "Back to work expectations and realizations of work after retirement," *Journal of Human Resources*, 45 (3), 718–748.

Maestas, Nicole and Xiaoyan Li (2007) "Burnout and the retirement decision."

Margaris, Panagiotis and Johanna Wallenius (2020) "Can Wealth Buy Health? A Model of Pecuniary and Non-Pecuniary Investments in Health."

McArdle, John J, Emilio Ferrer-Caja, Fumiaki Hamagami, and Richard W Woodcock (2002) "Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span.," *Developmental psychology*, 38 (1), 115.

Nielsen (2000) "The stochastic EM algorithm: estimation and asymptotic results," *Bernoulli*, 6 (3), 457–489.

Ruhm, Christopher J (1990) "Bridge jobs and partial retirement," *Journal of labor economics*, 8 (4), 482–501.

Rust, John and Christopher Phelan (1997) "How social security and medicare affect retirement behavior in a world of incomplete markets," *Econometrica: Journal of the Econometric Society*, 781–831.

Sanchez, Manuel and Felix Wellschmied (2020) "Modeling life-cycle earnings risk with positive and negative shocks," *Review of Economic Dynamics*, 37, 103–126, https://doi.org/10.1016/j.red.2019.11.003.

Sullivan, Sherry E and Akram Al Ariss (2019) "Employment after retirement: A review and framework for future research," *Journal of Management*, 45 (1), 262–284.

Wang, Mo, Lee Thomas Penn, and BERTONE AGUSTINA (2014) "Bridge employment in the United States," in *Bridge Employment*, 221–241: Routledge.

Wang, Mo, Yujie Zhan, Songqi Liu, and Kenneth S Shultz (2008) "Antecedents of bridge employment: a longitudinal investigation.," *Journal of applied Psychology*, 93 (4), 818.

Wei, Siqi (2021) "Estimating Nonlinear Panel Data Models By Parameter-Expanded Stochastic EM Methods," Unpublished Manuscript.

Wen, Jiayi (2018) "Occupational Retirement and Social Security Reform: the Roles of Physical and Cognitive Health," Working Papers 2018-12-04, Wang Yanan Institute for Studies in Economics (WISE), Xiamen University, https://ideas.repec.org/p/wyi/wpaper/002390.html.

Zhan, Yujie, Mo Wang, Songqi Liu, and Kenneth S Shultz (2009) "Bridge employment and retirees' health: A longitudinal investigation.," *Journal of occupational health psychology*, 14 (4), 374.

APPENDIX to "Income, Employment And Health Risks Of Older Workers"

A The Data

The assets A_{it} are defined as the sum of all wealth components (excluding secondary home) net of all debt. It is the sum of the primary residence, real estate, vehicles, businesses, IRA/Keogh accounts, stocks, mutual funds, investment trusts, checking, savings, money market accounts, CD, government savings bounds, T-bills, bonds, bonds funds, and all other savings less all mortgages/land contracts, other home loans and other debt.

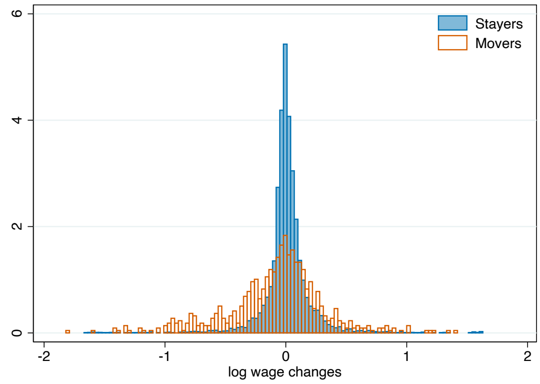
Observations for a given person-year are retained if the person is aged 51 to 70, never self-employed, and has non-missing data on education, race, self-reported health, birth information, wages, working hours and employment status. Follow Arellano et al. (2017), we do not use observed wage rates when they display extreme âjumpsâ from one year to the next and treat them as measurement error. Additionally, we also do not use assets that display extreme "jumps" but for a different reason. This could potentially be a measurement error, or this could reflect some real changes in capital income, real estate and etc. Our model do not include other sources of income but labor earnings, and it is difficult to explain the changes only through labor income. Potentially we can incorporate these possibilities by modelling some extreme shocks on top of normal shocks. But this is not what we focus on. Instead, we simply remove those observations. We restrict the sample to male individuals who have eligible observations for at least three consecutive waves and are employed for at least one wave.

Finally, following Altonji et al. (2013), we censor reported hours at 4000, add 200 to reported hours before taking logs to reduce the impact of very low values of hours on the variation in the logarithm, and censor observed hourly wage rates to increase by no more than 500% and decrease to no less than 20% of their lagged values.

B Supplement Figures

Figure B1: Log Wage Changes For Movers (excluding business close) And Stayers

(a) Log wage changes



(b) Log hour changes

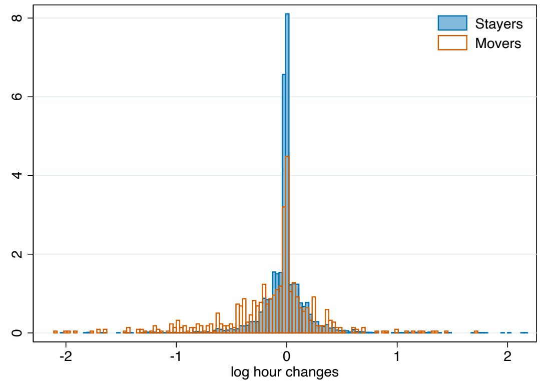
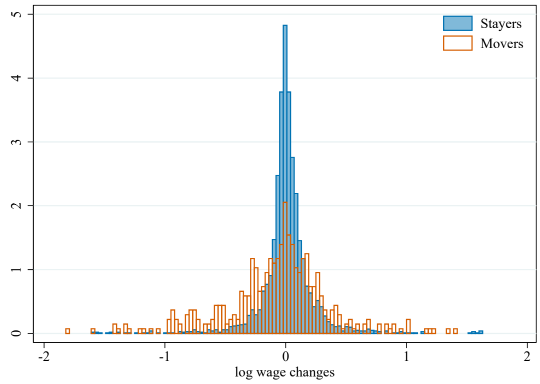


Figure B2: Log Wage Changes For Movers And Stayers, By Education

(a) HE



(b) LE

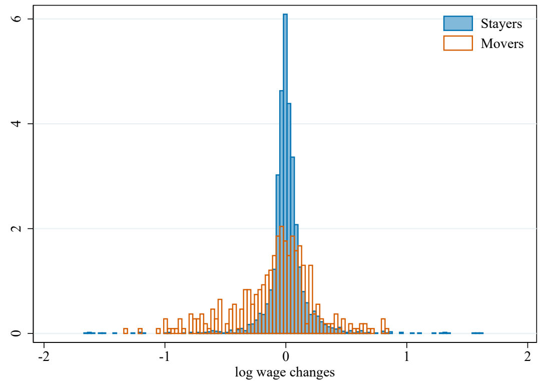


Figure B3: Log Wage Changes For Movers (excluding business close and laid-off) And Stayers

(a) HE (b) LE

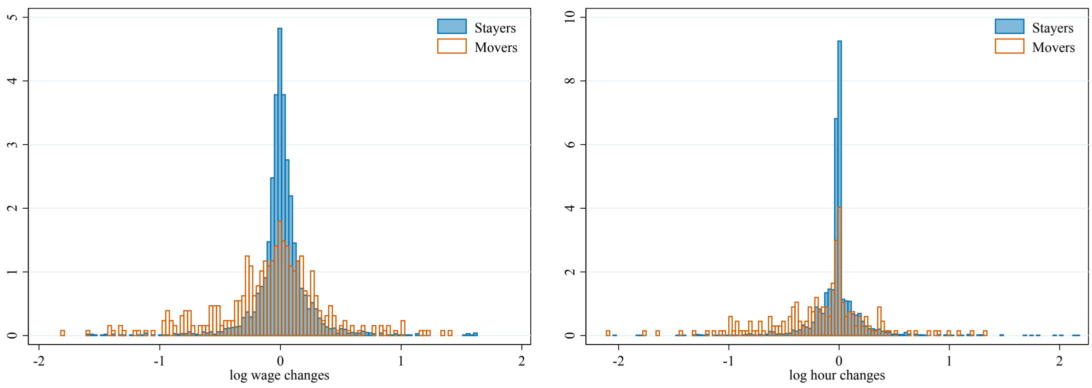


Figure B4: Log Wage And Hour Changes For Movers And Stayers With Tenure $\geq 10yrs$, By Education

(a) LE (b) HE

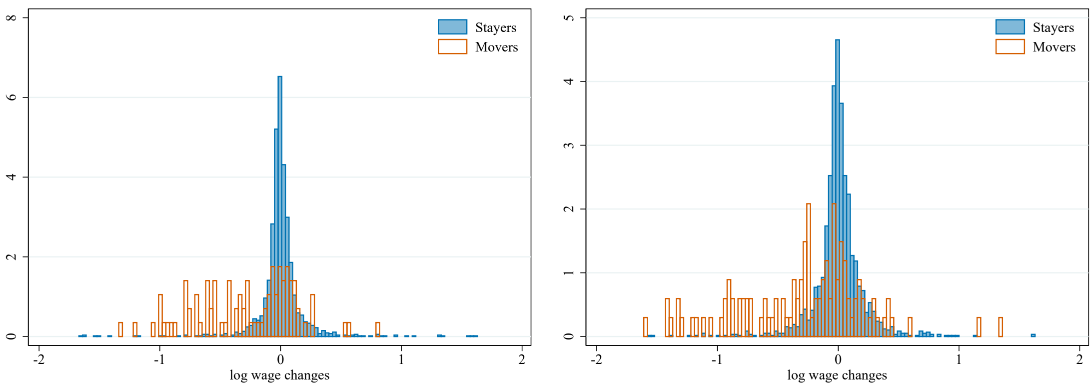


Figure B5: Log Hour Changes For Movers And Stayers, By Education

(a) LE (b) HE

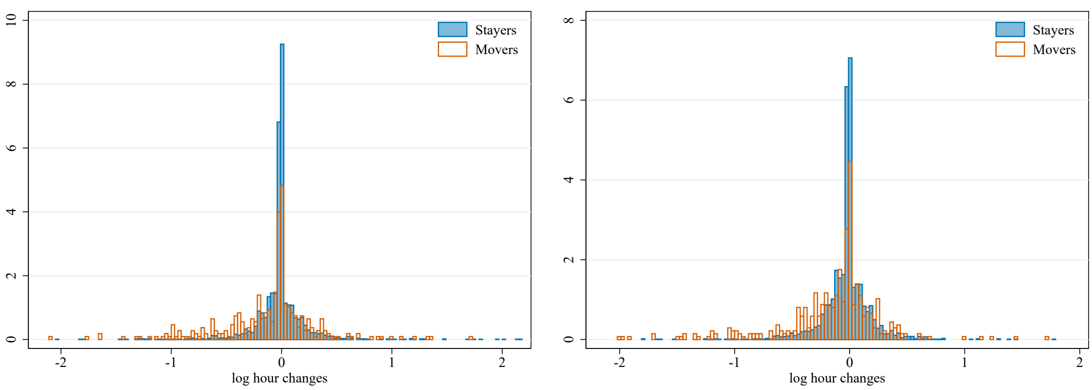
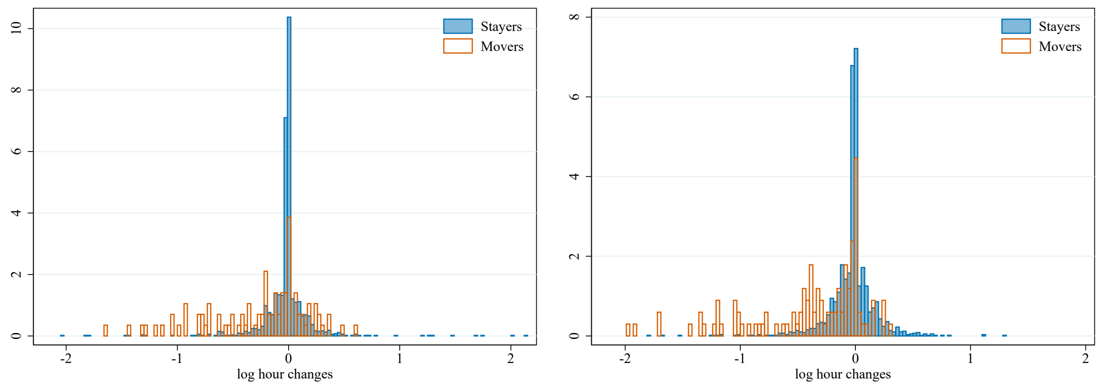


Figure B6: Log Hour Changes For Movers And Stayers With Tenure ≥ $10yrs$, By Education

(a) LE (b) HE



C Specification of Empirical Model

In this subsection, we describe in detail the specification of our empirical model.

Log Wages. The log wage $w_{it}$of individual i at period t is given by:

$$w_{it} = E_{it} \times (w^*_{it} + \varepsilon_{it}), \quad \varepsilon_{it} \sim N(0, \varsigma^2_m) \tag{C1}$$

$$w^*_{it} = X'_{it}\gamma^w_X + h_{it}\gamma^w_h + \mu_i + \omega_{it} + v_{ij(t)} \tag{C2}$$

Working hours. The log hour h_{it} of individual i at period t is specified as:

$$l_{it} = E_{it} \times (X'_{it}\gamma^l_X + h_{it}\gamma^l_h + \mu_i\gamma^l_\mu + \xi_{ij(t)} + \varepsilon^l_{it}) \tag{C3}$$

Individual earning power. The persistence of individual-specific productivity shock is assumed to be an exponential of a linear function of σ_{it} and $age_{i,t-1}$. We restrict the persistence to be non-negative.

$$\omega_{it} = exp(\gamma^\omega_1 + \gamma^\omega_\sigma \sigma_{it} + \gamma^\omega_{age} age_{i,t-1})\omega_{i,t-1} + \varepsilon^\omega_{it}\sigma_{it} \tag{C4}$$

In addition, we assume σ_{it} follows an autoregressive gamma process (Gourieroux and Jasiak, 2006). Variable σ_{it} follows an ARG(1) if and only if $\frac{\sigma_t}{c} \sim NoncentralGamma(\delta, 2\beta\sigma_{i,t-1})$, or equivalently $\frac{2\sigma_t}{c} \sim NoncentralChi-squared(2\delta, 2\beta\sigma_{t-1})$, where c, δ and β are parameters. It can be interpreted as σ_t is drawn from a Gamma distribution whose shape parameter depends on its value in previous period σ_{t-1} and parameter σ, and whose scale parameter is c.[1]

$$\frac{2\sigma_{it}}{exp(\gamma_1^\sigma + \gamma_{edu}^\sigma edu_i + \gamma_h^\sigma h_{i,t-1} + \gamma_{age}^\sigma age_{i,t-1} + \gamma_{age^2}^\sigma age_{i,t-1}^2)} \sim \chi_{2\delta}^2(2\beta\sigma_{i,t-1}) \tag{C5}$$

As described in Equation (C5), we specify parameter c as an exponential of a linear function of education, health, age, and age squared.

Health dynamics. The latent health h_{it} depends on a polynomial age trend and its value in previous period. How much the past health $h_{i,t-1}$ could affect the following period depends on it own value and age. Finally, $h_{i,t-1}$ also affects the size of health shocks.

$$\begin{aligned} h_{it} = &\gamma_1^h + \gamma_2^h age_{i,t-1} + \gamma_3^h age_{i,t-1}^2 + logistic(\gamma_4^h + \gamma_5^h h_{i,t-1} + \gamma_6^h h_{i,t-1}^2 + \\ &\gamma_7^h h_{i,t-1} age_{i,t-1} + \gamma_8^h age_{i,t-1} + \gamma_9^h age_{i,t-1}^2) h_{i,t-1} + \varepsilon_{it}^h \sigma(srh_{i,t-1}) \end{aligned} \tag{C6}$$

$$srh_{it} = \sum_{k=1}^{5} \mathbb{1}(h_{it} > \tau_k), \tau_1 = -\infty \tag{C7}$$

where $\sigma(srh_{i,t-1}) = \sum_k^5 \sigma_k \mathbb{1}(srh_{i,t-1} = k)$. We estimate the process by education group.

Firm-specific component and mismatch shocks. We assume there is a firm-specific wage component and a firm-specific hour component. The mismatch shock follows Equation (C10), a Probit model with regressors including age, age squared, education, stochastic firm component $\nu_{i,t-1}$, employment status, and its interaction with $\nu_{i,t-1}$.

$$\upsilon_{ij(t)} = \gamma_{ten1}^{\upsilon} ten_{ij(t)} + \gamma_{ten2}^{\upsilon} ten_{ij(t)}^2 + \gamma_{ten3}^{\upsilon} ten_{ij(t)}^3 + \nu_{ij(t)} \tag{C8}$$

$$\nu_{ij(t)} = \begin{cases} \nu'_{ij(t)} & \text{if start a new job } (ue_{it} = 1 \text{ or } jc_{it} = 1) \\ \nu''_{ij(t)} \ll \nu_{ij(t)} & \text{if stay at the same job and } m_{it} = 1 \\ \nu_{ij(t-1)} & \text{Otherwise} \end{cases} \tag{C9}$$

$$\begin{aligned} m_{ij(t)} = \mathbb{1}\{&\gamma_1^m + \gamma_{age}^m age_{i,t-1} + \gamma_{age^2}^m age_{i,t-1}^2 + \gamma_{edu}^m edu_i + \gamma_\nu^m \nu_{i,t-1} + \\ &\gamma_E^m (1 - E_{i,t-1}) + \gamma_{\nu E}^m \nu_{i,t-1}(1 - E_{i,t-1}) + \varepsilon_{it}^m > 0\} \end{aligned} \tag{C10}$$

$$\xi_{ij(t)} = \begin{cases} \xi'_{ij(t)} & \text{if } ue_{it} = 1 \text{ or } jc_{it} = 1 \\ \xi_{ij(t-1)} & \text{otherwise} \end{cases} \tag{C11}$$

Offers. The offer is a package of firm-specific wage and hour components. As we can see in Equation (C12), mismatch shocks affect the offer in two different ways. First, workers who receive mismatch shocks will face an average change of the offer quality by γ_m^ν. Secondly, if he does not accept the offer and becomes non-employed, then the

[1]To simulate σ_{it}, one could draw z from a Poisson distribution with parameter $\beta\sigma_{t-1}$ (i.e., $z \sim Poisson(\beta\sigma_{t-1})$), and then draw σ_t from a Gamma distribution: $\sigma_t \sim Gamma(z + \sigma, c)$.

following offers are also affected through the term $(1-E_{i,t-1})\sum_{to}^{t} m_{ij(t)}$. The accumulated mismatch shocks $\sum_{to}^{t} m_{ij(t)}$ is the key to express that mismatch shocks have persistent effect on firm-specific productivity and hence the offers. This term $(1-E_{i,t-1})\sum_{to}^{t} m_{ij(t)}$ also allows for different offer distributions for workers with job destruction shock and with mismatch shocks: those who only receive job destruction shock and become non-employed have $\sum_{t0}^{t} m_{ij(t)} = 0$.

$$\nu'_{ij(t)} = \gamma^{\nu}_{\nu}\nu_{ij(t0)} + \gamma^{\nu}_{m} m_{it}E_{i,t-1} + \gamma^{\nu}_{E}(1-E_{i,t-1}) + \gamma^{\nu}_{mE}(1-E_{i,t-1})\sum_{to}^{t} m_{ij(t)} + \gamma^{\nu}_{edu} edu_i + \gamma^{\nu}_{edumE} edu_i \times m_{it}E_{i,t-1} + \varepsilon^{\nu}_{it} \tag{C12}$$

$$\xi'_{ij(t)} = \gamma^{\xi}_{\xi}\xi_{ij(t0)} + \gamma^{\xi}_{m} m_{it}E_{i,t-1} + \gamma^{\xi}_{E}(1-E_{i,t-1}) + \gamma^{\xi}_{mE}(1-E_{i,t-1})\sum_{to}^{t} m_{ij(t)} + \gamma^{\xi}_{edu} edu_i + \gamma^{\xi}_{edumE} edu_i \times m_{it}E_{i,t-1} + \gamma^{\xi}_{\nu}\nu'_{ij(t)} + \varepsilon^{\xi}_{it} \tag{C13}$$

where $\varepsilon^{\nu}_{it} \sim m_{ij(t)} \times \sigma_{\nu 0}\sigma_{\nu 1}\varepsilon_{it} + (1-m_{ij(t)}) \times \sigma_{\nu 0}\varepsilon_{it}$, and the CDF of ε_{it} is $(1-\rho)\Phi(x) + \rho\Phi(x/k)$, with $\rho = 1/(k^2-1)$. Additionally, we allow parameters $\sigma_{\nu 0}$, $\sigma_{\nu 1}$, and k to be education and employment status $E_{i,t-1}$ specific. The distribution of $^{\xi}_{it}$ is specified in a similar way.

Employment transitions and Job dynamics. Job destruction only depends on education, which captures the systematic differences in stability of the job of workers with different education levels.

$$jd_{it} = \mathbb{1}\{\gamma^{jd}_{1} + \gamma^{jd}_{edu} edu_i + \varepsilon^{jd}_{it} > 0\} \tag{C14}$$

where $\varepsilon^{jd}_{it} \sim N(0,1)$.

Employees without job destruction shocks and mismatch shocks. The employment-to-employment transition is realized based on Equation (C15). Specifically, we allow individuals in their 70s to have different patterns from those in their 60s by adding interactions of an age dummy with polynomials of tenures and wage offers.

$$\begin{aligned} ee_{it} = \mathbb{1}\{ & \gamma^{ee}_{X} X_{it} + \gamma^{ee}_{A} A_{i,t-1} + \gamma^{ee}_{ten0}\mathbb{1}(ten_{i,t-1}=2) + \gamma^{ee}_{ten} ten_{i,t-1} + \gamma^{ee}_{ten^2} ten^2_{i,t-1} + \gamma^{ee}_{ten^3} ten^3_{i,t-1} + \\ & (\gamma^{ee}_{age,ten} ten_{i,t-1} + \gamma^{ee}_{age,ten^2} ten^2_{i,t-1} + \gamma^{ee}_{age,ten^3} ten^3_{i,t-1}) \times \mathbb{1}(age_{i,t-1} \geq 61) + \\ & \gamma^{ee}_{h1} h_{t-1} + \gamma^{ee}_{h2}\mathbb{1}(srh_{i,t-1}=4) + \gamma^{ee}_{h3}\mathbb{1}(srh_{i,t-1}=5) + \\ & \gamma^{ee}_{h4} h_{t} + \gamma^{ee}_{h5}\mathbb{1}(srh_{i,t}=4) + \gamma^{ee}_{h6}\mathbb{1}(srh_{i,t}=5) + \\ & \gamma^{ee}_{w1}\max(w^s_{it}, w^{new}_{it}) + \gamma^{ee}_{w2}\max(w^s_{it}, w^{new}_{it})\mathbb{1}(age_{i,t-1} \geq 61) + \\ & \gamma^{ee}_{w3}\max(w^s_{it}, w^{new}_{it}) edu_i + \gamma^{ee}_{l} l_{i,t-1} + \gamma^{ee}_{\mu}\mu_i + \varepsilon^{ee}_{it} > 0\} \end{aligned} \tag{C15}$$

where $\varepsilon_{it}^{ee} \sim N(0,1)$. Variable w_{it}^{s} is defined as wage the worker will earn if he stays at his job, and variable w_{it}^{new} is the wage if he moves to the new job.[2]

The job movement is specified as in Equation (C16). Workers value the difference in firm-specific component between the new job and the old job, that is, $(\xi'_{ij(t)} - \xi_{ij(t-1)})$ and $dv_{it} \equiv v'_{ij(t)} - v_{ij(t-1)}$. Additionally, the evaluation could depend on the characteristics of their current job $v_{ij(t)}^{s}$, and the sign of changes $|dv_{it}|$.[3]

$$\begin{aligned}
jc_{it} = \mathbb{1}\{&\gamma_X^{jc} X_{it} + \gamma_A^{jc} A_{i,t-1} + \gamma_{ten0}^{jc}\mathbb{1}(ten_{i,t-1} = 2) + \gamma_{ten}^{jc} ten_{i,t-1} + \gamma_{ten^2}^{jc} ten_{i,t-1}^2 + \gamma_{ten^3}^{jc} ten_{i,t-1}^3 + \\
&\gamma_{h1}^{jc} h_{t-1} + \gamma_{h2}^{jc}\mathbb{1}(srh_{i,t-1} = 4) + \gamma_{h3}^{jc}\mathbb{1}(srh_{i,t-1} = 5) + \\
&\gamma_{h4}^{jc} h_t + \gamma_{h5}^{jc}\mathbb{1}(srh_{i,t} = 4) + \gamma_{h6}^{jc}\mathbb{1}(srh_{i,t} = 5) + \\
&\gamma_{\omega}^{jc}\omega_{it} + \gamma_{\xi 1}^{jc}(\xi'_{ij(t)} - \xi_{ij(t-1)}) + \gamma_{\xi 2}^{jc}(\xi'_{ij(t)} - \xi_{ij(t-1)})^2 + \\
&\gamma_{v1}^{jc} v_{ij(t)}^{s} + \gamma_{v2}^{jc} v_{ij(t)}^{s}\mathbb{1}(age_{i,t-1} \geq 61) + \gamma_{v3}^{jc} v_{ij(t)}^{s} edu_i + \gamma_{v4}^{jc} v_{ij(t)}^{s}\mathbb{1}(age_{i,t-1} \geq 61) edu_i + \\
&\gamma_{dv1}^{jc}\log(|dv_{it}|)\mathbb{1}(dv_{it} > 0) + \gamma_{dv2}^{jc}\log(|dv_{it}|)\mathbb{1}(dv_{it} > 0)\mathbb{1}(age_{i,t-1} \geq 61) + \\
&\gamma_{dv3}^{jc}\log(|dv_{it}|)\mathbb{1}(dv_{it} > 0) edu_i + \gamma_{dv4}^{jc}\log(|dv_{it}|)\mathbb{1}(dv_{it} > 0)\mathbb{1}(age_{i,t-1} \geq 61) edu_i + \\
&\gamma_{dv5}^{jc}\log(|dv_{it}|)\mathbb{1}(dv_{it} \leq 0) + \gamma_{dv6}^{jc}\log(|dv_{it}|)\mathbb{1}(dv_{it} \leq 0)\mathbb{1}(age_{i,t-1} \geq 61) + \\
&\gamma_{dv7}^{jc}\log(|dv_{it}|)\mathbb{1}(dv_{it} \leq 0) edu_i + \gamma_{dv8}^{jc}\log(|dv_{it}|)\mathbb{1}(dv_{it} \leq 0)\mathbb{1}(age_{i,t-1} \geq 61) edu_i + \\
&\gamma_{\mu}^{jc}\mu_i + \varepsilon_{it}^{jc} > 0\}
\end{aligned} \tag{C16}$$

where $\varepsilon_{it}^{jc} \sim N(0,1)$.

Non-employed or with job destruction shocks or with mismatch shocks. The non-employment-to-employment transition is specified as follows:

$$\begin{aligned}
ue_{it} = \mathbb{1}(&\gamma_X^{ue} X_{it} + \gamma_A^{ue} A_{i,t-1} + \gamma_{h1}^{ue} h_{t-1} + \gamma_{h2}^{ue}\mathbb{1}(srh_{i,t-1} = 4) + \gamma_{h3}^{ue}\mathbb{1}(srh_{i,t-1} = 5) + \\
&\gamma_{h4}^{ue} h_t + \gamma_{h5}^{ue}\mathbb{1}(srh_{i,t} = 4) + \gamma_{h6}^{ue}\mathbb{1}(srh_{i,t} = 5) + \\
&\gamma_{w_1}^{ue} w_{it}^{new} + \gamma_{w_2}^{ue} w_{it}^{new}\mathbb{1}(age_{i,t-1} \geq 61) + \gamma_{w_3}^{ue} w_{it}^{new} edu_i + \\
&\gamma_{jd1}^{ue} jd_{it} + jd_{it} \times (\gamma_{jd2}^{ue} edu_i + \gamma_{jd3}^{ue} age_{i,t-1} + \gamma_{jd4}^{ue} age_{i,t-1}^2 + \gamma_{jd5}^{ue} age_{i,t-1}^3 + \gamma_{jd6}^{ue} h_{i,t-1}) + \\
&jd_{it} \times (\gamma_{jd7}^{ue} w_{it}^{new} + \gamma_{jd8}^{ue} w_{it}^{new}\mathbb{1}(age_{i,t-1} \geq 61) + \gamma_{jd9}^{ue} w_{it}^{new} edu_i) + \\
&\gamma_m^{ue} m_{it} E_{i,t-1} + \gamma_{ud}^{ue} ud_{i,t-1} + \gamma_{\mu}^{ue}\mu_i + \varepsilon_{it}^{ue} > 0\}
\end{aligned} \tag{C17}$$

where $\varepsilon_{it}^{ue} \sim N(0,1)$. In data, we observe that people with $jd = 1$ are significantly more likely to return to labor market than those who have to non-employed for many

[2]We have $w_{it}^{s} \equiv X'_{it}\gamma_X^w + h_{it}\gamma_h^w + \mu_i + \omega_{it} + \gamma_{ten1}^{v}(ten_{ij(t-1)} + 2) + \gamma_{ten2}^{v}(ten_{ij(t-1)} + 2)^2 + \gamma_{ten3}^{v}(ten_{ij(t-1)} + 2)^3 + v_{ij(t-1)}$, and $w_{it}^{new} \equiv X'_{it}\gamma_X^w + h_{it}\gamma_h^w + \mu_i + \omega_{it} + \gamma_{ten1}^{v} ten_{ij'(t0)} + \gamma_{ten2}^{v} ten_{ij'(t0)}^2 + \gamma_{ten3}^{v} ten_{ij'(t0)}^3 + v'_{ij(t)}$.

[3]We take log of $|dv_{ti}|$ because it tends to have a long tail in practice. We transform it to increase the stability.

periods. For this reason, we add flexible interactions of the job destruction dummy jd_{it} and demographics, health, and wages.

Assets accumulation. We specify the conditional probability distribution of assets as in Equation (C18), which is a set of Probit regressions whose parameters depend on the threshold τ_k^A.

$$\begin{aligned}\mathbb{1}(A_{it} < \tau_k^A) = \mathbb{1}\{&\gamma_X^A(\tau_k^A)X_{it} + \gamma_A^A(\tau_k^A)A_{i,t-1} + \gamma_w^A(\tau_k^A)w_{it}^* E_{it} + \gamma_l^A(\tau_k^A)l_{it}E_{it} + \\ &\gamma_E^A(\tau_k^A)(1 - E_{it}) + \gamma_h^A(\tau_k^A)h_{it} + \gamma_\omega^A(\tau_k^A)\omega_{it} + \gamma_\nu^A(\tau_k^A)\nu_{it} + \gamma_\mu^A(\tau_k^A)\mu_i + \varepsilon_{it}^A > 0\}\end{aligned} \tag{C18}$$

Initial distribution. Finally, the distributions of the initial variables are specified in Equations (C19 - C24)

$$h_{i1} = \gamma_X^{h1}X_{i1} + \gamma_A^{h1}A_{i1} + \gamma_E^{h1}E_{i1} + \varepsilon_{i1}^{h1}, \varepsilon_{i1}^{h1} \sim N(0,1) \tag{C19}$$

$$\nu_{i1} = \sum_{k=1}^{7} \gamma_k^{\nu 1}\mathbb{1}(2k + 51 \le age_{i1} \le 2k + 52) + \gamma_{edu}^{\nu 1} edu_i + \varepsilon_{i1}^{\nu 1} \tag{C20}$$

$$\xi_{i1} = \sum_{k=1}^{7} \gamma_k^{\xi 1}\mathbb{1}(2k + 51 \le age_{i1} \le 2k + 52) + \gamma_{edu}^{\xi 1} edu_i + \gamma_{\nu 1}^{\xi 1}\nu_{i1} + \varepsilon_{i1}^{\xi 1} \tag{C21}$$

$$\sigma_{i2} \sim Gamma(k_{\sigma_2}, exp(\gamma_{01}^\sigma + \gamma_{02}^\sigma edu_i + \gamma_{03}^\sigma h_{i1} + \gamma_{04}^\sigma age_{i1} + \gamma_{05}^\sigma age_{i1}^2)/k_{\sigma_2}) \tag{C22}$$

$$\omega_{i1} \sim N(0, \sigma_{\omega 1}^2) \tag{C23}$$

$$\mu_i \sim N(0, \sigma_\mu^2) \tag{C24}$$

where $\varepsilon_{i1}^{\nu 1} \sim N(0, \sigma_{\nu edu1}^2 edu_i + (1 - edu_i)\sigma_{\nu edu0}^2)$. Further extension is to allow for selection of latent variables in initial period, for example, by adding an employment equation for the first period.

D The Estimation Of The Empirical Model

Let $Y^\dagger$ denote the observable set $\{w_{it}, E_{it}, sh_{it}, jd_{it}, jc_{it}, A_{it}, l_{it}\}, i = 1, ..., N, t = 1, ..., T_i$, let $X^\dagger$ denote the observable set $\{age_{it}, edu_i, race_i, E_{i1}, ten_{i1}, A_{i1}, recess_{it}\}, i = 1, ...N, t = 1, ..., T_i$, and let Z denote the latent variable set $\{h_{it}, \mu_i, \omega_{it}, \sigma_{it}, \nu'_{it}, \nu_{i1}, m_{it}, \xi'_{it}, \xi_{i1}\}, i = 1, ..., N, t = 1, ..., T_i$, denote Θ as all parameters. In addition, I define $Y \equiv \{x | x \in Y^\dagger, x \neq sh\}$ as all elements in $Y^\dagger$ except health measures sh. Correspondingly, Θ_h represents all parameters in health dynamics (eq 4 and 5), and $\Theta_{-h} \equiv \{x | x \in \Theta, x \notin \Theta_h\}$ is all other parameters except Θ_h. For simplicity, I use $f(\cdot)$ to represent the density function implied by the empirical model throughout this section.

The log-likelihood function of the empirical model can be decomposed as follows

$$L(Y^{\dagger}|X^{\dagger};\Theta) = \sum_{i=1}^{I} \ln f(sh_i|X_i^{\dagger};\Theta_h) + \sum_{i=1}^{I} \ln f(Y_i|sh_i, X_i^{\dagger};\Theta_h,\Theta_{-h})$$

Instead of estimating Θ jointly by maximizing $L(\Theta)$, we proceed sequentially: we obtain estimates $\hat{\Theta}_h$ from maximizing $\sum_{i=1}^{I} \ln f(sh|X^{\dagger};\Theta_h)$ and estimates $\hat{\Theta}_{-h}$ from maximizing $\sum_{i=1}^{I} \ln f(Y_i|sh, X^{\dagger};\hat{\Theta}_h,\Theta_{-h})$. The two-step estimators are consistent.[4]

Health part. Estimation of Θ_h involves maximizing $\sum_{i=1}^{I} \ln f(sh_i|X_i^{\dagger};\Theta_h)$. Considering that there is only one latent variable, which usually implies relatively faster convergence, we implement the SEM algorithm. Given initial guess $\Theta_h^{(0)}$, we iterate between E-step and M-step

1. E-step: draw $h^{(s)}$ from $f(h_i|sh_i, X_i^{\dagger};\Theta_h^{(s-1)}) \propto f(h_i, sh_i|X_i^{\dagger};\Theta_h^{(s-1)})$
2. M-step: update to $\Theta_h^{(s)} = \arg\max_{\Theta_h} \sum_{i=1}^{I} \ln f(h_i^{(s)}, sh_i|X_i^{\dagger};\Theta_h)$

where the joint distribution is

$$f(h_i, sh_i|X_i^{\dagger};\Theta_h) = f(h_{i1}|X_i^{\dagger};\Theta_h)\Big(\prod_{t=2}^{T_i} f(h_{it}|h_{i,t-1}, X_i^{\dagger};\Theta_h)\Big) f(sh_i|h_i;\Theta_h)$$

The remaining. We exploit PX-SEM algorithm to estimate the remaining parameters Θ_{-h}. This is motivated by the fact that our model contains multiple latent variables across multiple periods, which is often accompanied by a relatively slow convergence in SEM iterations. In this situation, the value of improving the convergence speed is no longer negligible.

PX-SEM also requires iterating between E-step and M-step, where the E-step is the same as the one of SEM algorithm. In M-step, we need to 1) expand the original model (O model) to a larger model (L model) space that contains the original model 2) estimate the larger model 3) reduce to the original model space keeping the likelihood of observables unchanged. By expanding and reducing, we are actually proposing a more robust estimator to bad E-step guess. And the lesson is more than the parameter expansion technique itself: even without expanding the model (L model = O model), exploiting more robust estimators in M-step can help too (i.e., the estimator that uses as less as possible the latent draws from E-step).

There is no unique way to expand the original model. In principle, a larger model should always be better than or the same as SEM in terms of converging speed. However,

[4]Check Cox (1975), Amemiya (1978) and Arcidiacono and Jones (2003) for details about asymptotic properties of these sorts of two-step estimators.

PX-SEM may cost extra time in expanding and reducing models especially when the L model is too complicated, and therefore extending the total computing times. Following the strategy in Wei (2021), we will expand the model in a *linear* way.[5]

The specific steps of PX-SEM are as follows: starting with initial guess $\Theta^{(0)}_{-h}$, we iterate between the E-step and M-step:

1. E-step: draw $Z^{(s)}$ from $f(Z_i|Y_i^{\dagger}, X_i^{\dagger}; \hat{\Theta}_h, \Theta_{-h}) \propto f(Y_i^{\dagger}, Z_i|X_i^{\dagger}; \hat{\Theta}_h, \Theta_{-h})$
2. PX-M-step: update to $\hat{\Theta}^{(s)} = \arg\max_{\Theta} \sum g(Y_i^{\dagger}, Z_i^{(s)}; \Theta)$

In the following paragraphs, we discuss E-step and M-step in reverse order. We firstly present the M-step, which including our choice of L model, its estimation and reduction, and then we compute the likelihood function of the original model $f(Y_i^{\dagger}, Z_i|X_i^{\dagger}; \Theta_h, \Theta_{-h})$ that is required for MH in E-step.

L model is based on O model. Firstly we revisit the latent wage equation (2) and firm component equation (8). For simplicity, we separate the recession indicator rec_t from the variable vector X_{it} and name the remaining demographic part as D_{it}, that is $[D_{it}; rec_t] = X_{it}$. The equation (2) and (8) are replaced by equation (D1) and (D2)

$$w^*_{it} = D'_{it}\boldsymbol{\gamma_{1,t}} + rec_{it}\boldsymbol{\gamma_{2,t}} + h_{it}\gamma_h + \mu_i\boldsymbol{k_\mu} + \omega_{it}\boldsymbol{k_\omega} + \upsilon_{ij(t)} \tag{D1}$$

$$\upsilon_{ij(t)} = ten'_{ij(t)}\boldsymbol{\gamma_{3,t}} + \nu_{ij(t)}\boldsymbol{k_\nu} \tag{D2}$$

Compared with O model, L model allows for a time-variant dependence of latent wages on observables including X and ten as shown in equation (D1) and (D2). Furthermore, the coefficient of latent variable μ_i, ω_{it} and $\nu_{ij(t)}$ can be different from one. Then we move to working hours. The equation (3) is replaced by the following one

$$l_{it} = E_{it} \times (X'_{it}\gamma^l_X + h_{it}\gamma^l_h + k_\mu\mu_i\gamma^l_\mu + \xi_{ij(t)}\boldsymbol{k_\xi} + \varepsilon^l_{it}) \tag{D3}$$

where the coefficient of $\xi_{ij(t)}$ is also allowed to be other than one.

Next, define $\widetilde{\mu}_i \equiv \mu_i - D'_{i1}\boldsymbol{\alpha_1} - rec_t\boldsymbol{\alpha_2} - ten'_{ij(1)}\boldsymbol{\alpha_3}$, and $\widetilde{\omega}_{it} \equiv \omega_{it} - D'_{it}\boldsymbol{\beta_{1,t}} - rec_t\boldsymbol{\beta_{2,t}} - ten'_{ij(t)}\boldsymbol{\beta_{3,t}}$. We replace μ_i, ω_{it}, $\nu_{ij(t)}$ and $\xi_{ij(t)}$ in the rest of O model, which is all equations and assumptions in the O model except equation (2), (3) and (8), with $k_\mu\widetilde{\mu}_i$, $k_\omega\widetilde{\omega}_{it}$, $k_\nu\nu_{ij(t)}$ and $k_\xi\xi_{ij(t)}$. For example, we then have $k_\mu\widetilde{\mu}_i \sim N(0, \sigma^2_\mu)$. As a result, the fixed effect μ_i and the individual-specific productivity ω_{it} can vary with X and ten.

[5] It is linear in the sense that the L model expands the O model by affine transformation.

Finally we assume $\begin{pmatrix} \gamma_{1,t} + k_\mu g_1(\alpha_1) + k_\omega \beta_{1,t} \\ \gamma_{2,t} + k_\mu \alpha_2 \times \mathbb{1}(t=1) + k_\omega \beta_{2,t} \end{pmatrix} = \gamma_X^w$, $\alpha_2 = 0$, $\gamma_{3,t} + k_\mu \alpha_3 \times \mathbb{1}(t = 1) + k_\omega \beta_{3,t} = \gamma_{ten}^v$, and $\alpha_3 = 0$.[6]

Apart from all parameters in O model Θ, L model contains a set of auxiliary parameters including $\Lambda = (k_\mu, k_\omega, k_\nu, k_\xi, \alpha_1, ..., \alpha_3, \beta_1, ..., \beta_3)$. It is easy to check that when $k_\mu = k_\omega = k_\nu = k_\xi = 1$, $\alpha_1 = \beta_{1,t} = \vec{0}$, $\alpha_2 = \beta_2 = 0$, $\alpha_3 = \beta_{3,t} = \vec{0}$, $\forall t \le T$, L model equals O model. On the other hand, L nests O model (L model $\supseteqq$ O model) because the auxiliary parameters can aways take other values. We could further expand L model, for instance, by adding dependent variables including observables and latent variables of all periods in defining $\widetilde{\mu}_i$ and $\widetilde{\omega}_{it}$, etc. However, it will complicate the estimation and reduction.

The estimation of L model is based on the pseudo-complete data including the draws from E-step of h_{it}, μ_i, ω_{it}, σ_{it}, ν'_{it}, ν_{i1}, m_{it} and ξ_{it}. We decompose the parameters into several groups to estimate sequentially. Firstly, we estimate the auxiliary parameters: 1) regress μ_i on X_{i1} and ten_1 to obtain $\hat{\alpha}_1$ and compute the residual res_i^μ, 2) regress ω_{it} on X_{it} and ten_{it} to obtain $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ and compute residual res_{it}^ω.[7]

Secondly, we recover coefficients of wage equation, k_μ, k_ω and k_ν, standard deviation of μ and ω_{i1} through following steps: 1) regress Δw_{it} on ΔX_{it}, Δten, Δh_{it}, Δres_{it}^ω and $\Delta \nu_{ij(t)}$ for individual i who works both in period $t-1$ and t to get coefficients of variables other than education and race. The coefficients of Δres_{it}^ω and $\Delta \nu_{ij(t)}$ are the estimates of k_ω and k_ν respectively, 2) regress $w_{it} - X'_{it}\hat{\gamma}_X - h_{it}\hat{\gamma}_h^w - res_{it}^\omega \hat{k}_\omega - ten'_{ij(t)}\hat{\gamma}_{ten}^v - \nu_{ij(t)}\hat{k}_\nu$ on edu, $race$ and res_i^μ.[8][9] The coefficient of res_i^μ is the estimate of k_μ. Step 1) and 2) altogether provide $\hat{\gamma}_X^w$, $\hat{\gamma}_h^w$, $\hat{\gamma}_{ten}^v$, $\hat{k}_\mu$, $\hat{k}_\omega$ and $\hat{k}_\nu$. We can also compute latent wage for each individual w_{it}^*.

Next, we estimate the parameters in working hour equation. By regressing non-zero working hours on X_{it}, h_{it}, $\hat{k}_\mu res_i^\mu$, and $\xi_{ij(t)}$, we obtain $\hat{\gamma}_X^l$, $\hat{\gamma}_\mu^l$ and $\hat{k}_\xi$.

All parameters in m, jd, jc, ee, ue and A equations are estimated with Probit regression function in julia package GLM. Specifically, asset dynamics are estimated with a set of Probit regressions. Each regression has a different τ and a dependent variable $\mathbb{1}(A_{it} \le \tau)$.

Finally, parameters in the distribution of latent variables are estimated as follows: 1)

[6]function $g_1(\alpha_1)$ is a vector of polynomials of α_1. The function transforms the coefficients of age_1 and its polynomials into coefficient of the polynomials of age_t.

[7]In the estimation of ω, we use the assumption that ω_{i1} and ϵ_{it}^ω are independent from other components.

[8]Combining equation (1), equation (D1), equation (D2), definition of $\widetilde{\mu}$ and $\widetilde{\omega}_{it}$, and restrictions on auxiliary parameters, we can get $w_{it} = X'_{it}\gamma_X^\omega + h_{it}\gamma_h^w + k_\mu \widetilde{\mu}_i + k_\omega \widetilde{\omega}_{it} + ten'_{ij(t)}\gamma_{ten}^v + k_\nu \nu_{ij(t)} + \epsilon_{it}^m$ for workers.

[9]In Step 1), ΔX removes edu and $race$ because they do not vary with time. We use $\hat{\gamma}_X$ to represent coefficients of the rest.

distribution of μ: $\hat{\sigma}_\mu = \hat{std}(\hat{k}_\mu res_i^\mu)$. 2) initial distribution of ω: $\hat{\sigma}_{\omega 1} = \hat{std}(\hat{k}_\omega res_{i1}^\omega)$. 3) dynamics of ω: use nonlinear least squares regression by solving min $\sum(\hat{k}_\omega res_{it}^\omega/\sigma_{it} - \rho(\cdot)\hat{k}_\omega res_{i,t-1}^\omega/\sigma_{i,t-1})^2$, where NLopt package is exploited. 4) initial distribution of ν and ξ: regress $\hat{k}_\nu \nu_{ij(1)}$ on age group dummies (51-52, 53-54,...,69-70) and compute the std of the residual. Regress $\hat{k}_\xi \xi_{ij(1)}$ on age group dummies as well as $\hat{k}_\nu \nu_{ij(1)}$ and compute the std of the residual. 5) offer distribution: regress $\hat{k}_\nu \nu'_{ij(t)}$ on $\hat{k}_\nu \nu_{ij(t-1)}$, interactions of m_{it}, jd_{it}, $E_{i,t-1}$, and compute the std of the residual. Regress $\hat{k}_\xi \xi'_{ij(1)}$ on $\hat{k}_\nu \nu_{ij(t-1)}$, interactions of m_{it}, jd_{it}, $E_{i,t-1}$, as well as $\hat{k}_\nu \nu'_{ij(t)}$. Then compute the std of the residual. 6) initial distribution of σ: we estimate parameters in scale parameter by maximizing log-likelihood function of Gamma distribution (the shape parameter can be separately and do not affect the result of scale paraters). Because $\sigma_{i1}/exp(\beta x) \sim Gamma(k, 1/k)$, we solve for the shape parameter by Newton-Raphson algorithm as described in Choi and Wette (1969). 7) dynamics of σ: The autoregressive gamma process is estimated following Gourieroux and Jasiak (2006). We obtain a pseudo-maximum likelihood estimator based on a Gaussian pseudo-family.

The rule of reducing from L to O model is to keep the likelihood of observables unchanged. We specify the L model in such a way that the reduction is easy, we simply keep the estimates of all O model $\hat{\Theta}$.[10].

Finally, we present the likelihood function of the **original** model.

$$
\begin{aligned}
& f(Y_i^\dagger, Z_i | X_i^\dagger; \Theta) \\
=& f(w_i, E_i, sh_i, jd_i, jc_i, A_i, l_i, h_{it}, \mu_i, \omega_i, \sigma_i, \nu'_i, \nu_{ij(1)}, m_i, \xi'_i, \xi_{ij(1)} | X_i^\dagger; \Theta) \\
=& f(\mu_i|X_i^\dagger) f(\sigma_{i2}|X_i^\dagger) f(\nu_{ij(1)}|X_i^\dagger) f(\xi_{ij(1)}|X_i^\dagger, \nu_{i1}) f(h_{i1}|X_i^\dagger) f(\omega_{i1}|X_i^\dagger) \\
& \Big(\prod_{t=2}^{T_i} f(jd_{it}|X_i, E_{i,t-1}) f(m_{it}|X_i, h_{i,t-1}, \nu_{ij(t-1)}, E_{i,t-1}) f(h_{it}|h_{i,t-1}, X_i) \\
& f(\nu'_{ij(t)}|\nu_{ij(t-1)}, m_{it}, jd_{it}, E_{i,t-1}) f(\xi'_{ij(t)}|\xi_{ij(t-1)}, m_{it}, jd_{it}, E_{i,t-1}, \nu'_{ij(t)}) \\
& f(\sigma_{it}|\sigma_{i,t-1}, X_i, h_{i,t-1}) f(\omega_{it}|X_i, \sigma_{it}) \\
& f(E_{it}|E_{i,t-1}, X_i, h_{it}, ten_{i,t-1}, ud_{i,t-1}, jd_{it}, m_{it}, w^*_{i,t-1}, \omega_{it}, \nu_{ij(t-1)}, \nu'_{ij(t)}, \xi_{ij(t-1)}, \xi'_{ij(t)}, \mu_i) \\
& f(jc_{it}|E_{i,t-1}, E_{it}, jd_{it}, m_{it}, X_i, h_{i,t}, ten_{i,t-1}, \omega_{it}, \nu_{ij(t)}, \nu'_{ij(t)}, \xi_{ij(t)}, \xi'_{ij(t)}, \mu_i) \\
& f(A_{it}|A_{i,t-1}, X_i, E_{it}, w^*_{it}, l_{it}, h_{it}, \omega_{it}, \nu_{it}, m_{it}, \mu_i) \Big) \\
& f(w_i|w_i^*) f(l_i|X_i, h_i, \mu_i, \xi_i) f(sh_i|h_i)
\end{aligned}
$$

where w^*_{it} is determined by $X_{it}, h_{it}, \mu_i, \omega_{it}, ten_{ij(t)}$ and $\nu_{ij(t)}$. Similarly, $ten_{ij(t)}$ is determined by $ten_{ij(1)}$ and employment history E_i^t, whereas realized firm wage and hour component,

[10] Until now, the objective function for estimation in M-step, $g(\cdot)$, is implicitly decided by the steps above.

$\nu_{ij(t)}$ and $\xi_{ij(t)}$, can be expressed as a function of all offers $(\nu_i'^t, \xi_i'^t)$, initial component $(\nu_{ij(1)}, \xi_{ij(1)})$ as well as employment histories E_i^t and jc_i^t.

We have a further discussion about the asset dynamics, which are estimated flexibly using Distribution Regression. The asset data are two-side bounded. During the preliminary data cleaning, I censor reported net wealth at $1st$ and $99th$ quantiles, which are -4.134379×10^4\$ and 322.161×10^4\$ respectively. Distribution Regression adapts easily to this feature. In practice, I take 19 cutoffs. In addition to the maximum and minimum values in the data, I also add $10th$, $20th$,...,$80th$, $83th$,..., $95th$, $96th$,...,$99th$ quantiles of the empirical distribution of net wealth. I denote the cutoffs as $\tau_{A,1} < \tau_{A,2} <, \ldots, < \tau_{A,19}$. The cdf is

$$F_A(\tau_A|X_A) = \mathbb{1}(A_{it} \leq \tau_A|X_A) = \Phi(\beta(\tau_A)'X_A)$$

where $X_A = [A_{i,t-1}, X_i, E_{it}, w_{it}^*, l_{it}, h_{it}, \omega_{it}, \nu_{it}, m_{it}, \mu_i]'$. Moreover, I assume $\beta(\tau_A)$ is a linear spline. Then we can easily get the density function

$$f(A_{it}|X_A) = \phi(\beta(\tau_A)'X_A)\frac{\beta(\tau_{A,k+1})'X_A - \beta(\tau_{A,k})'X_A}{\tau_{A,k+1} - \tau_{A,k}}$$

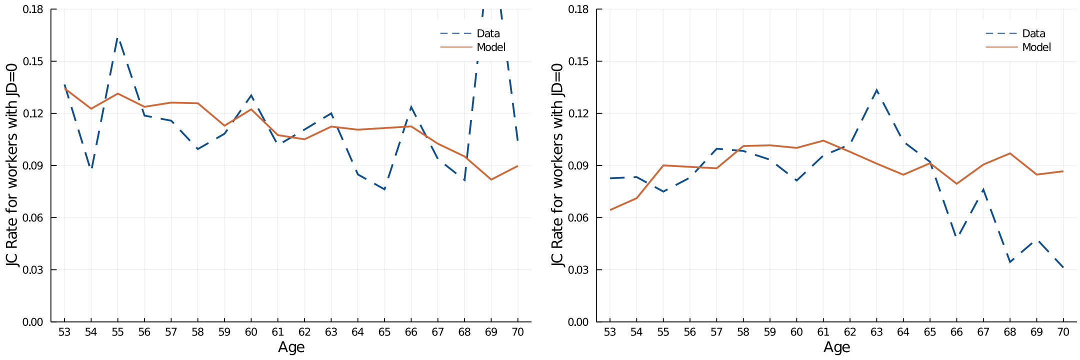
where $\tau_{A,k}$ and $\tau_{A,k+1}$ are closest cutoffs: $\tau_{A,k} \leq \tau_A < \tau_{A,k+1}$, $k \in \{1, \ldots, 18\}$. Two probability mass points, the minimum and the maximum, have the mass function $F_A(\tau_{A,1}|X_A)$ and $(1 - F_A(\tau_{A,19}|X_A))$

In practice, to keep the CDF function non-decreasing monotonically, the rearrangement operator is exploited.[11]

E Model Fit

Figure E1: Job Change Over Age

(a) Job-to-job move conditional on $ee_{it} = 1$ and $jd_t = 0$, HE (b) Job-to-job move conditional on $ee_{it} = 1$ and $jd_t = 0$, LE



[11]Details can be found in notes by Victor Chernozhukov and Ivan Fernandez-Val *Distribution Regression And Counterfactual Analysis*

Figure E2: Job Destruction over Age

(a) JD rate, HE   (b) JD rate, LE

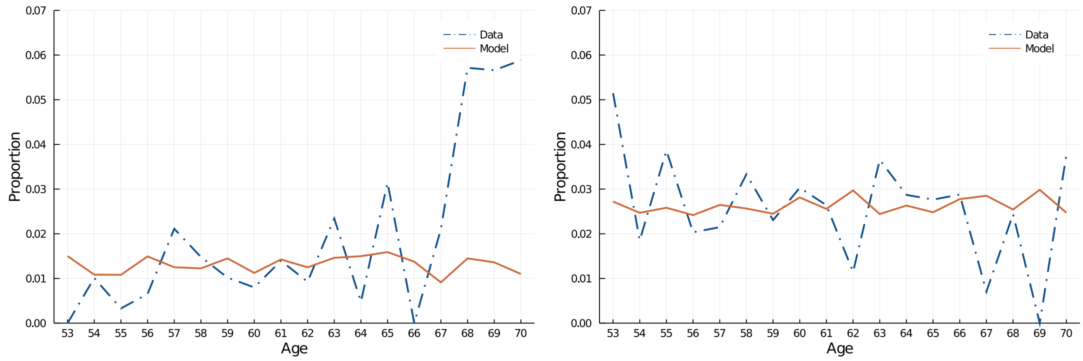


Figure E3: Percentiles of Tenures over Age

(a) Percentiles of tenures, HE   (b) Percentiles of tenures, LE

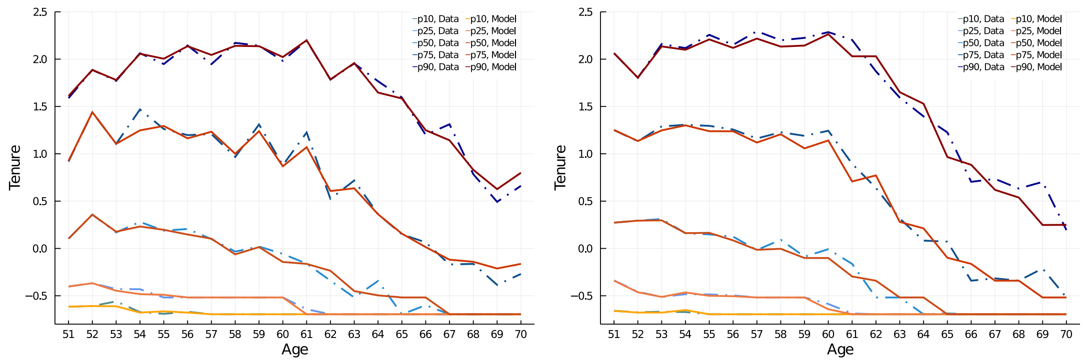
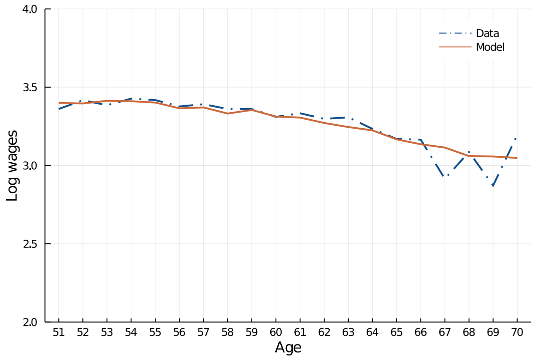
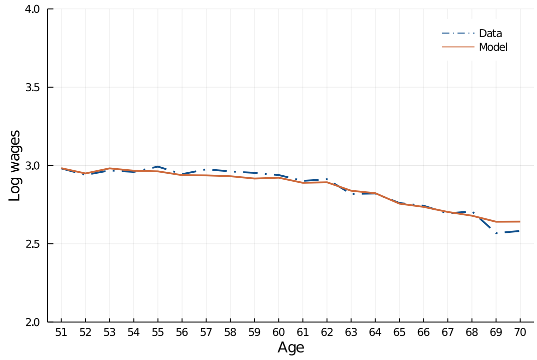


Figure E4: Moments of Wages

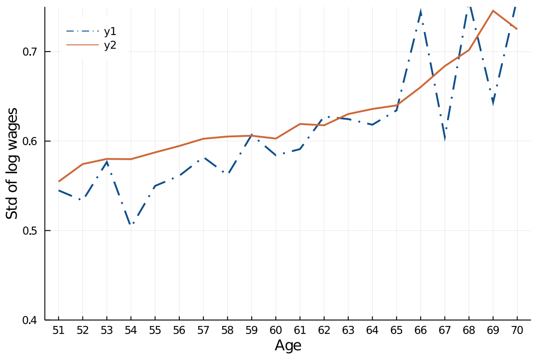
(a) Mean wages, HE



(b) Mean wages, LE



(c) Std wages, HE



(d) Std wages, LE

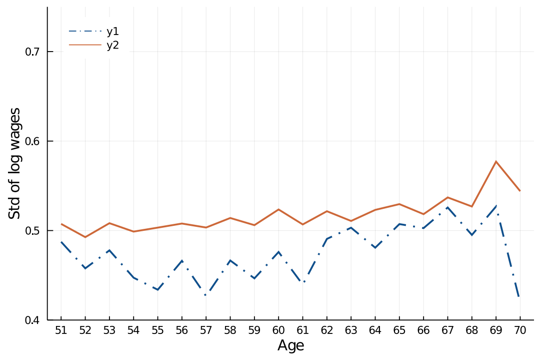
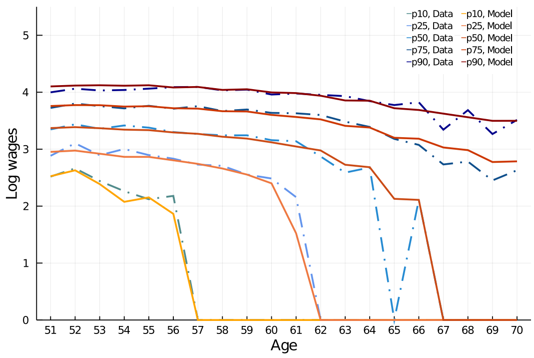


Figure E5: Percentiles of Wages

(a) Percentiles of wages (including zeros), HE



(b) Percentiles of wages (including zeros), LE

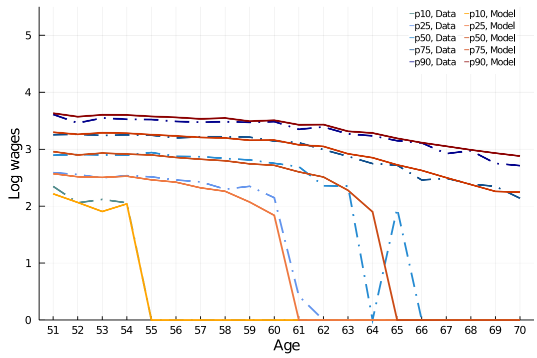
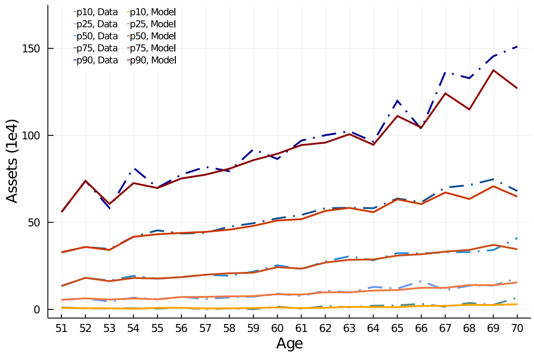
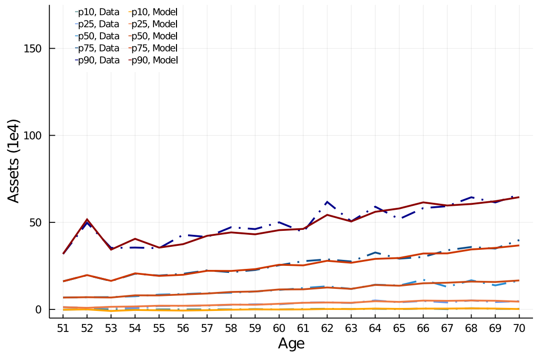


Figure E6: Percentiles of Assets

(a) Percentiles of assets, HE



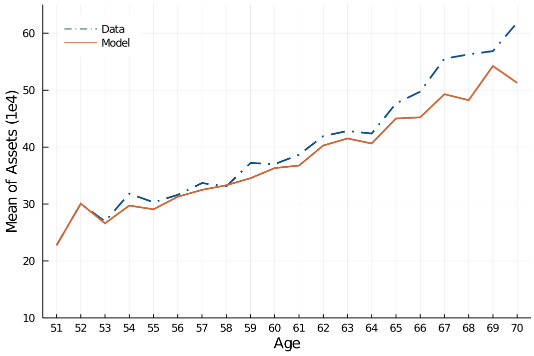
(b) Percentiles of assets, LE



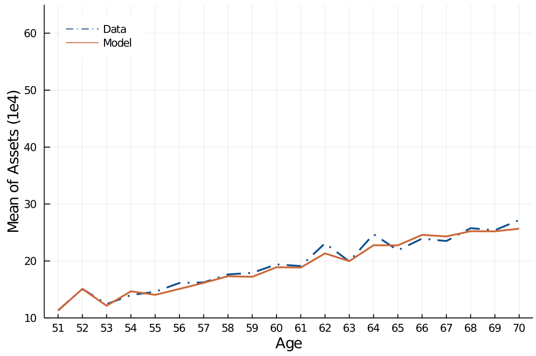
Assets

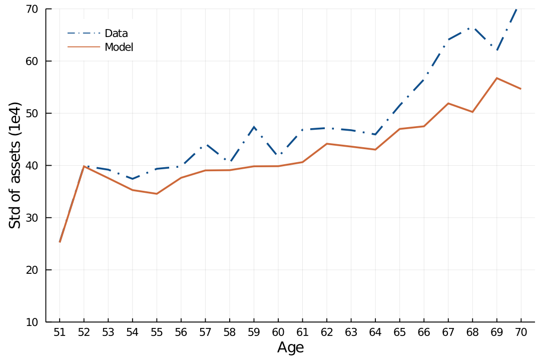
Figure E7: Moments Of Assets

(a) Mean Assets, HE

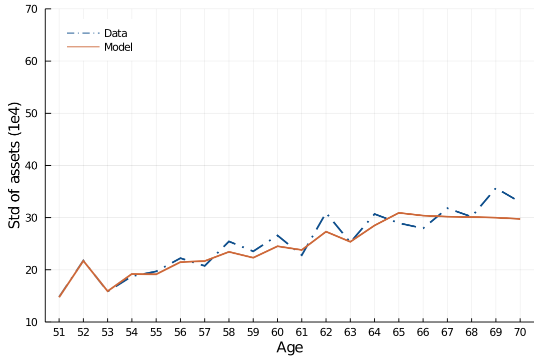


(b) Mean Assets, LE

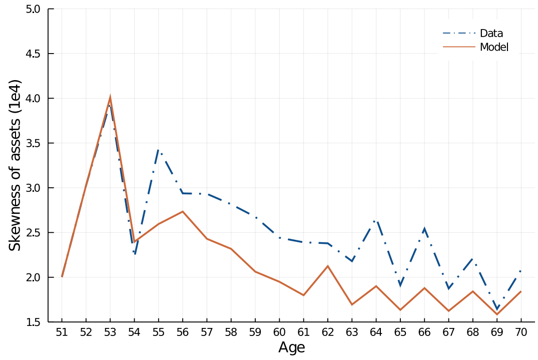




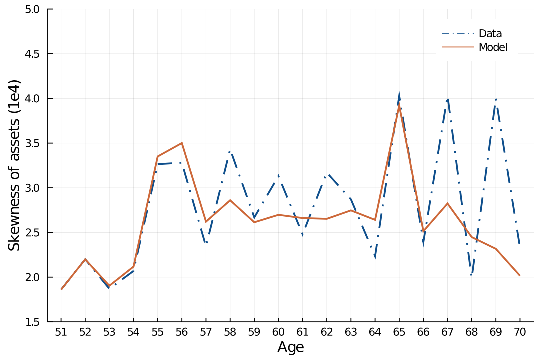
(c) Std of assets, HE



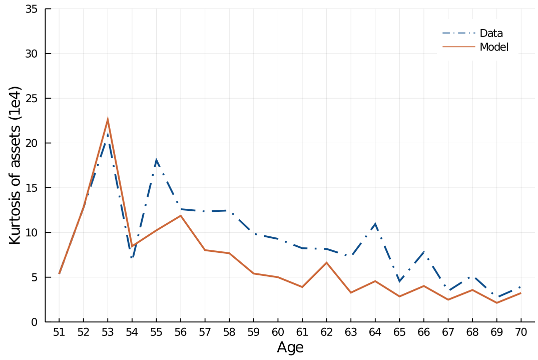
(d) Std of assets, LE



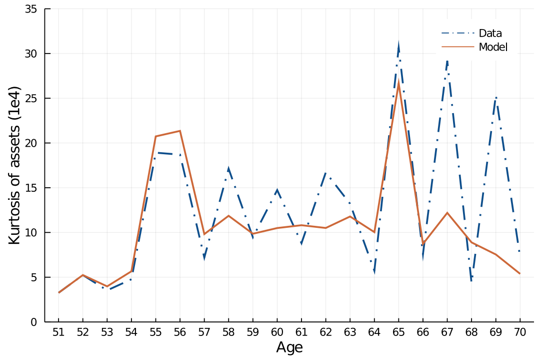
(e) Skewness of assets, HE



(f) Skewness of assets, LE



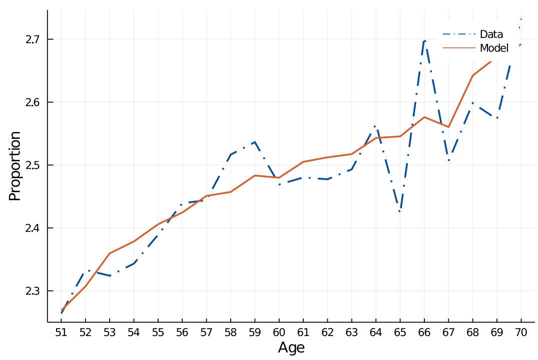
(g) Kurtosis of assets, HE



(h) Kurtosis of assets, LE

Figure E7: Health Profiles by Age

(a) Mean self-reported health, HE



(b) Mean self-reported health, LE

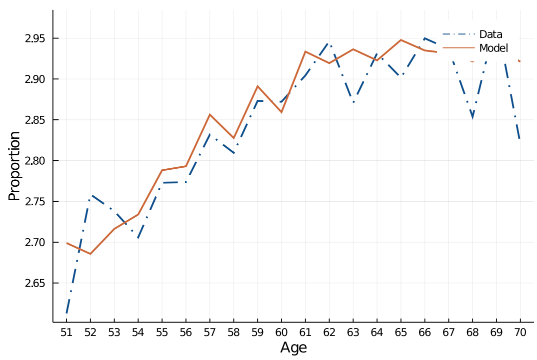
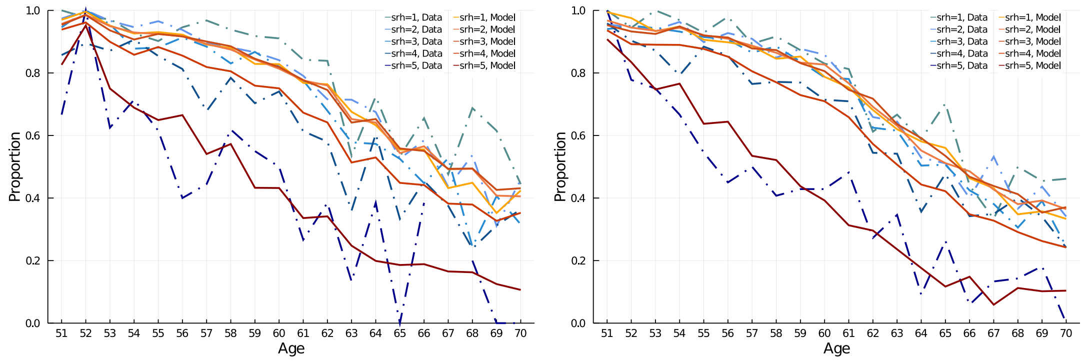


Figure E8: Employment Rate by Self-reported Health and Age

(a) Employment rate by health and age, HE (b) Employment rate by health and age, LE



F The Solution of the Structural Model

I will solve the model using backward induction. The state variables of the model are $\Omega_t^\dagger = [A_{t-1}, \widetilde{v}_{t-1}, \widetilde{v}'_t, \omega_t, \sigma_t, h_t, m_t, t, d_{t-1}, jd_t, edu, \epsilon_t^d]$. Among all the elements of $\Omega_t^\dagger$, jd_t and d_{t-1} only affects the choice set but not the dynamics of shocks. I also solve and estimate for each education group separately. For simplicity, define $\Omega_t = [A_{-1}, \widetilde{v}_{t-1}, \widetilde{v}'_t, \omega_t, \sigma_t, h_t, m_t, t]$

Value function:

$$\begin{aligned} V_t(\Omega_t^\dagger) \quad = \quad & \max_{C_t, E_t, jc_t} \Big\{ U(C_t, d_t, d_{t-1}, jd_t, h_t, \epsilon_t^d) + \beta(1 - s_{t+1}) b(A_{t+1}) + \\ & \beta s_{t+1} \mathbb{E}\Big(V_{t+1}(\Omega_{t+1}^\dagger) \Big| \Omega_t^\dagger, C_t, d_t \Big) \Big\} \\ s.t. \qquad & A_{t+1} = (1+r)A_t + Y_t + ss_t * B_t + tr_t - C_t \\ & A_{t+1} \geq A_{min}, C_t \geq C_{min} \end{aligned}$$

Discretization I discretise variables $\omega, \sigma, \nu, \nu'$ and h. For each of the variable, a set of grid-points are taken based on the distribution. The elements of transition matrix represent the conditional probability on a set of grids of having a draw in an area around the certain grid. For example, we have ω_t depending on ω_{t-1} and σ_t. Accordingly, the transitions matrix is of dimension 3. Let us use $g_1^\omega \leq g_2^\omega \leq ... \leq g_{k_\omega}^\omega$ to represent grids for ω, and $g_1^\sigma \leq g_2^\sigma \leq ... \leq g_{k_\sigma}^\sigma$ for σ. The (i, j, p) element of the transition matrix is computed as: $F(\frac{g_{p+1}^\omega + g_p^\omega}{2} | g_i^\omega, g_j^\sigma) - F(\frac{g_p^\omega + g_{p-1}^\omega}{2} | g_i^\omega, g_j^\sigma)$.

Steps Started for the last period T, I iterated among the following steps:

1. Given each possible value of state variable Ω and employment and job decision d, I compute the optimal consumption C which does not depend on the preference

shocks ϵ. With this optimal consumption, we can obtain the present value associated to the specific group of state variable and labor force participation choice.

2. We obtain the optimal choice given the state variable by choosing the one that are associated with highest present value.

In step 1, we need to compute $E_t(V_{t+1}|\Omega_t^\dagger)$. Given the transition matrix and the assumption on the preference shocks, it can be computed easily: $E_t(V_{t+1}|\Omega_t^\dagger) = E_t(E_t(V_{t+1}|\Omega_t^\dagger, \Omega_{t+1})|\Omega_t^\dagger)$. The inner part, the integral over preference shocks, has closed-form, whereas the outer part can be computed from the transition matrix.

Simulation After solving the model, we get policy functions of labor supply, job movements and consumption. Then I use these decision rules to generate simulated histories. Specifically, the initial distribution comes from the empirical distributions of observables and estimated distributions of unobservables from the non-utility-based model. In each period, we draw $\omega, \sigma, \nu, \nu'$ and h from the continuous processes. Based on the realisation of state variables, policy function and budget constraint, we get their labor force participation and the accumulated wealth.

Relation to the non-utiltiy-based employment transitions and job movements The non-utility-based model, whose employment decisions and job movements are characterized by two separate equations, can be seen as an approximation of reduced form of the utility-based model. Especially, the two-step estimation of ee and jc does not contradict the assumption of simultaneous decision making in the utility-based model.

Denote $V_t^{old*}, V_t^{new*}, V_t^{non*}$ as the present value net of preference shocks associated with three different choices: work at the old job ($E_t = 1, jc_t = 0$), work at the new job ($E_t = 1, jc_t = 1$) and Not work ($E_t = 0$). Each of the three is a function of state variables. Then according to the utility-based model, people make decisions by choosing the one that optimises the present value.

For individuals who were previously non-employed or lose the job because of the job destruction shock or mismatch shock, they choose between *new job* and *non-employed*. It can be specified as

$$E_t = \mathbb{1}\Big(V_t^{new*} - V_t^{non*} + \epsilon_t^{new} - \epsilon_t^{non} > 0\Big)$$

Combining with the assumption of ϵ, the equation above is the counterpart of ue equation. Being Logistic regression of Probit regression will not matter as long as the other elements in the ue equation is flexible enough.

For individuals who were employed and did not lose the job, we have

$$E_t = \mathbb{1}\Big(V_t^{non*} + \epsilon_t^{non} < \max(V_t^{old*} + \epsilon_t^{old}, V_t^{new*} + \epsilon_t^{new})\Big)$$

$$jc_t = \mathbb{1}\Big(V_t^{new*} + \epsilon_t^{new} > V_t^{old*} + \epsilon_t^{old} > 0, E_t = 1\Big)$$

Combining with the assumption of ϵ, it can be shown that

$$P(E_t = 1|E_{t-1} = 1, jd_t = 0) = 1 - \frac{exp(V_t^{non*})}{exp(V_t^{old*}) + exp(V_t^{new*}) + exp(V_t^{non*})}$$

$$P(jc_t = 1|E_t = 1) = \frac{exp(V_t^{new*})}{exp(V_t^{old*}) + exp(V_t^{new*})}$$

which are counterparts of ee and jc equations.

G The construction of $\widetilde{\upsilon}$

Labor earnings are the product of hourly wage, W_t, and working hours, N. In the model, the amount of hour supply is fixed. By assumption, N = 4000, which is amount to 2000 hours per year.

$$Y_t = W_t \times N \tag{G1}$$

Log hourly wage rate $\ln W_t$ takes the same structure as the one in non-utility-base model.

$$\ln W_t = X_t'\gamma_X + h_t\gamma_h + \omega_t + \widetilde{\upsilon}_{jt} \tag{G2}$$

To reduce the dimension of state variables, I merge unobserved heterogeneity μ, firm component υ_{ijt} and offer ν'_{ijt}. Specifically, I define $\widetilde{\upsilon}_{jt} \equiv \mu_i + \upsilon_{jt} = ten_t'\gamma_e + \nu_{ij(t)} + \mu_i$ and $\widetilde{\nu}'_{jt} \equiv ten_0'\gamma_e + \nu'_{ij(t)} + \mu_i$.

The new firm component $\widetilde{\upsilon}_{jt}$ contains two parts: $\mu_i + \nu_{ij(t)}$ stays constant during the tenure and drop significantly when there is mismatch shocks. $ten_t'\gamma_e$ represents the accumulation of firm-specific experience.

$$\widetilde{\upsilon}_{j(t)} = \begin{cases} \rho_{\upsilon 0} + \rho_\upsilon \widetilde{\upsilon}_{j(t-1)} & \text{if stay at the same job} \\ \widetilde{\upsilon}'_{j(t)} & \text{if move to new job} \end{cases} \tag{G3}$$

During the tenure, the dynamics of $\widetilde{\upsilon}_{jt}$ is approximated by the equation G3, that is a linear function of its value in last period.

H Structural Model Estimation Procedure

The estimation procedure is motivated by such a premise: non-utility-based (NU) and utility-based (U) model share the same latent variable dynamics, and the NU employment and job transitions can be treated as an approximate reduced form of the U ones. These two characteristics make it attractive to use the estimated latent variable dynamics from NU model as input in the U one. This is because the NU model is under less parametric assumption in the employment and job part which matters for the estimation of the latent variable dynamics. In addition, the flexibility of NU model makes it suitable as an intermediary to compare simulated data and true data.

For simplicity, I refer to the non-utility-based model as *model 1*, whereas utility-based model as *model 2*. The observed data Y consist of a set of observations on N individuals in each of $T_i \leq T$ periods: $\{y_{it}\}, i = 1, ..., N, t = 1, ..., T_i$. There are also a set of latent variables Z unobservable to researchers.

Model 1 is represented by $f_1 \equiv f_1(Y, Z^*; \theta), \theta \in \Theta$, whose marginal distribution of Y is denoted as $f_{1Y}(Y; \theta)$. The family is defined as $M_1 \equiv \{f_1 | \theta \in \Theta\}$. By exploiting stochastic EM algorithm, we have obtained MLE of model 1:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \sum \ln f_{1Y}(Y; \theta)$$

Similarly, the family of model 2 is defined as $M_2 \equiv \{f_2(Y, Z^*; \beta) | \beta \in B\}$, whose marginal distribution is $f_{2Y}(Y; \beta)$. The goal is to estimate parameter β.

The estimation procedure is described as follows:

0) we draw a large sample of $\widetilde{Z}$ from $f_1(\widetilde{Z}|Y; \hat{\theta}_{MLE})$;

1) given β, we use model 2 to generate M statistically independent simulated data set $\{Y, Z\}^m$;

2) we compute $\theta^m(\beta) = \frac{1}{M} \sum \widetilde{\theta}^m(\beta)$, where $\widetilde{\theta}^m(\beta)$ is the MLE estimator for each of the M simulated data sets: $\widetilde{\theta}^m(\beta) = \arg\max_{\theta \in \Theta} \ln f_1(Y, Z; \theta)$;

3) evaluate the objective function $\sum \sum_{\widetilde{Z}} \ln f_1(Y, \widetilde{Z}; \theta^m(\beta))$

Our estimator $\hat{\beta}_{NEW}$ of the model 2 is generated by choosing β to maximize objective function in step 3). In practice, we need to iterate from step 1) to step 3) until the convergence of the objective function. Note that step 0) is not included in the iteration. The intuition is to treat $\widetilde{Z}$ as if they are the observed data. Our estimator can be described as follows:

$$\hat{\beta}_{NEW} = \arg\max_{\beta \in B} \sum \sum_{\widetilde{Z}} \ln f_1(Y, \widetilde{Z}; \theta(\beta))$$

where $\widetilde{Z}$ is drawn from posterior $f_1(\widetilde{Z}|Y;\hat{\theta}_{MLE})$.

The non-utility-based model and the utility-based model in this chapter are special cases of model 1 and model 2 in the sense that two models have the latent variable dynamics in common. Under such preconditions, we can further simplify the estimation procedure: directly use MLE of latent variable dynamics $\hat{\beta}^Z_{MLE} = \hat{\theta}^Z_{MLE}$, and estimate the rest of the parameters $\beta \backslash \hat{\beta}^Z_{MLE}$ in iterations.

In practice, a two-step strategy is exploited. In the first step, I estimate survival probability, which depends on latent health and age, for people between 51 and 90. Estimates for latent health, wage equation, individual and firm component, unemployment risks and random offers from the NU based model are directly taken ($\hat{\beta}^Z_{MLE}$).[12] In the second step, I estimate the rest of parameters $\beta \backslash \hat{\beta}^Z_{MLE} \equiv (\nu, \theta_e, \theta_r, \theta_j, \kappa, \beta, C_{min})$ using the iteration procedure described above. Specifically, I take random draws from the empirical joint distribution of assets, health status, initial employment status, and demographics for initial wave, and simulate the life course according to model 2 given a set of parameters. To compute the objective function, in step 3, ee, ue and jc equations (eq 14-16) as well as asset accumulation (eq 20) are used.[13]

Relation to Indirect Inference Note that the procedure above is closely related to but different from the standard Indirect Inference method. Under the same notation, I-I estimator which uses Model 1 as auxiliary model and takes LR metric can be expressed as follows:

$$\hat{\beta}_{II} = \arg\max_{\beta \in B} \sum \ln f_{1Y}(Y;\theta(\beta))$$

where $\theta(\beta) = \arg\min_{\theta \in \Theta} E(\ln \frac{f_{2Y}(\beta)}{f_{1Y}(\theta)})$. However, the difficulty is that the analytical form f_{1Y} is difficult to get because of latent variables. One natural idea is to combine EM algorithm, with which we have I-I estimator expressed as:

$$\hat{\beta}_{II} = \arg\max_{\beta \in B} \sum \sum_{\widetilde{Z}} \ln f_1(Y, \widetilde{Z}; \theta(\beta))$$

where $\widetilde{Z}$ is drawn from posterior $f_1(\widetilde{Z}|Y;\theta(\beta))$. The problem of implementing $\hat{\beta}_{II}$ is that it may be very time consuming: for each $\theta(\beta)$, we need to draw $\widetilde{Z} \sim f_1(\widetilde{Z}|Y;\theta(\beta))$. When the direct sampling is not feasible, which is our case, we have to go for MCMC in each iteration.

[12] A further step is needed to approximate the dynamics of new firm component and its interactions with other elements conditional on age groups.

[13] In practice, I put extra restriction on equation 20 by assuming assets only depends on previous asset level and age

Now it is clear how $\hat{\beta}_{NEW}$ is different from $\hat{\beta}_{II}$: instead of drawing $\widetilde{Z}$ in each iteration, we use the draw from the posterior distribution given the estimates of Model 1.

I Further Discussion About Structural Model Estimation

Imagine the true model produces observables Y and latent variables Z^* from DGP:

$$g(Y, Z^*)$$

with a marginal distribution of observable Y denoted as $g_Y(Y)$

There are two models:

Model 1 is assumed to be a flexible model.

$$f_1 \equiv f_1(Y, Z^*; \theta), \theta \in \Theta$$

with a marginal distribution of Y denoted as $f_{1Y}(Y;\theta)$. The family is defined as $M_1 \equiv \{f_1|\theta \in \Theta\}$

Model 1 is estimated using MLE (maybe combine EM algorithm). The MLE estimator is defined as

$$\hat{\theta}_{MLE} = \arg\max_{\theta} \sum \ln f_{1Y}(Y;\theta)$$

Model 2 is a structural model that we are interested in. The model could potentially contain other variables. But it produces the following joint distribution of Y and Z^*

$$f_2 \equiv f_2(Y, Z^*; \beta), \beta \in B$$

with marginal distribution $f_{2Y}(Y;\beta)$. The family is defined as $M_2 \equiv \{f_2|\beta \in B\}$.

First we try to estimate Model 2 using Indirect Inference with Model 1 being auxiliary model. I choose the LR approach: the metric is the likelihood function associated with the auxiliary model. The procedure is

1) given $\beta^{(s)}$, simulate from Model 2 $\{Y^{(s)}, Z^{(s)}\}$
2) estimate Model 1 with $\{Y^{(s)}, Z^{(s)}\}$ and obtain $\theta^{(s)} = \theta(\beta^{(s)})$
3) evaluate objective function $\ln f_{1Y}(Y;\theta^{(s)})$

The mapping from Model 2 to Model 1 in step 1) and 2):

$$M : \{f_2(\beta)|\beta \in B\} \to \{f_1(\theta)|\theta \in \Theta\}$$

Specifically, I define the range as $M_{21} \equiv \{M(f_2)|\beta \in B\}$. By definition $M_{21} \subseteq M_1$.

In step 2), if we use MLE, then we know that:

$$\forall \beta^0 \in B, M(f_2(\beta^0)) = f_1(\theta^0), \text{ where } \theta^0 = \arg\min_{\theta \in \Theta} E\Big(\ln \frac{f_2(Y^{(s)}, Z^{(s)}; \beta^0)}{f_1(Y^{(s)}, Z^{(s)}; \theta)}\Big)$$

Combine three steps together with MLE in step 2), II is to choose optimal β to maximize the objective function.

$$\hat{\beta}_{II} = \arg\max_{\beta \in B} \sum \ln f_{1Y}(Y; \theta(\beta))$$

where $\theta(\beta) = \arg\min_{\theta \in \Theta} E(\ln \frac{f_2(\beta)}{f_1(\theta)})$.

Note that $\hat{\beta}_{II}$ is a MLE of a more restricted version of Model 1.

However, the analytical form f_{1Y} is difficult to get because of latent variables. One possible solution is to combine the EM algorithm:

$$\hat{\beta}_{II} = \arg\max_{\beta \in B} \sum \sum_{\widetilde{Z}} \ln f_1(Y, \widetilde{Z}; \theta(\beta))$$

where $\widetilde{Z}$ is drawn from posterior $f_1(\widetilde{Z}|Y; \theta(\beta))$. The limitation of this estimator is that it is time consuming: for each $\theta(\beta^{(s)})$, we need to draw $\widetilde{Z} \sim f_1(\widetilde{Z}|Y; \theta(\beta^{(s)}))$. If direct sampling is not feasible, it means that we may need to combine with MCMC.

So instead, here I use an estimator that makes use of the $\hat{\theta}_{MLE}$.[14] The estimator $\hat{\beta}_{NEW}$ is defined as follows:

$$\hat{\beta}_{NEW} = \arg\max_{\beta \in B} \sum \sum_{\widetilde{Z}} \ln f_1(Y, \widetilde{Z}; \theta(\beta))$$

where $\widetilde{Z}$ is drawn from posterior $f_1(\widetilde{Z}|Y; \hat{\theta}_{MLE})$. Take $\hat{\theta}_{MLE}$ as given, the II procedure is as follows:

0) draw a large sample of $\widetilde{Z}$ from $f_1(\widetilde{Z}|Y; \hat{\theta}_{MLE})$

1) given $\beta^{(s)}$, generate $\{Y^{(s)}, Z^{(s)}\}$ from Model 2

2) estimate the auxiliary model (Model 1) with $\{Y^{(s)}, Z^{(s)}\}$ and obtain $\theta^{(s)} = \theta(\beta^{(s)})$

3) evaluate the objective function $\sum \sum_{\widetilde{Z}} \ln f_1(Y, \widetilde{Z}; \theta(\beta^{(s)}))$ and return to step 1) with $\beta^{(s+1)}$

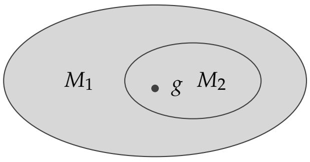
Now I will discuss about the consistency of $\hat{\beta}_{NEW}$ under different assumptions.

[14]Why we may want to estimate the potentially complicated model 1 as auxiliary model? It could be that Model 1 is interesting on its own/helps identify some parameters outside Model 2. Or it may be some flexible models in the literature that has interesting empirical facts that we want to match with Model 2.

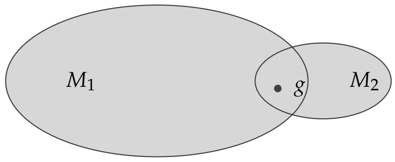
Case 1: Both Model 1 and Model 2 are correctly specified

Assumption: True model $g(Y, Z^*) \in M_1$, $g(Y, Z^*) \in M_2$

Model 1 does not necessarily nest Model 2. The possible cases include:



$M_2 \quad \subseteq \quad M_1$



$M_1 \cap M_2 \neq \emptyset, \exists f_1 \in M_1 : f_1 \notin M_2$

The estimator $\hat{\beta}_{NEW}$ converges to the true value in limit.

Sketch:

$$S_0(\beta) = \int g_Y(Y) \int \ln f_1(Y, \widetilde{Z}; \theta(\beta)) f_1(\widetilde{Z}|Y; \boldsymbol{\theta_{MLE}}) d\widetilde{Z} dY$$

where $\theta_{MLE} = \arg\max_{\theta\in\Theta} \int g_Y(Y) \ln f_1(Y, \theta) dY$

Under assumption $g \in M_1$, we know $f_1(Y, Z^*; \theta_{MLE}) = g(Y, Z^*)$.

$$S_0(\beta) = \int\int \ln f_1(Y, \widetilde{Z}; \theta(\beta)) f_1(Y, \widetilde{Z}; \boldsymbol{\theta_{MLE}}) d\widetilde{Z} dY$$

In addition, $g \in M_2$, then $\exists \beta^* \in B$ such that $f_2(Y, Z^*; \beta^*) = g(Y, Z^*) = f_1(Y, Z^*; \theta_{MLE})$. Now it can be proved that $\theta(\beta^*) = \theta_{MLE}$.

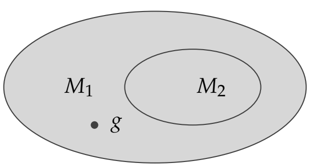
$$\theta(\beta^*) \equiv \arg\max_{\theta\in\Theta} \int \ln f_1(Y, Z; \theta) f_2(Y, Z; \beta^*) dYZ$$

$$= \arg\max_{\theta\in\Theta} \int \ln f_1(Y, Z; \theta) f_1(Y, Z; \theta_{MLE}) dYZ = \theta_{MLE}$$

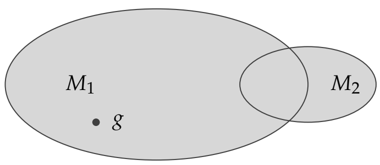
Then it is easy to show that

$$\begin{aligned} S_o(\beta) - S_0(\beta^*) =& \int\int \ln \frac{f_1(Y, \widetilde{Z}; \theta(\beta))}{f_1(Y, \widetilde{Z}; \theta_{MLE})} f_1(Y, \widetilde{Z}; \boldsymbol{\theta_{MLE}}) d\widetilde{Z} dY \\ \leq& \ln \int\int \frac{f_1(Y, \widetilde{Z}; \theta(\beta))}{f_1(Y, \widetilde{Z}; \theta_{MLE})} f_1(Y, \widetilde{Z}; \boldsymbol{\theta_{MLE}}) d\widetilde{Z} dY \\ =& 0 \end{aligned}$$

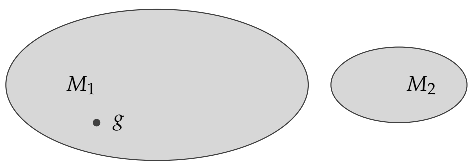
Case 2: Model 1 is correctly specified but Model 2 not



$M_2 \quad \subseteq \quad M_1$



$M_1 \cap M_2 \neq \emptyset, \exists f_1 \in M_1 : f_1 \notin M_2$



$M_1 \cap M_2 \quad = \quad \emptyset$

Assumption: True model $g(Y, Z^*) \in M_1$, $g(Y, Z^*) \notin M_2$

$$S_0(\beta) = \int g_Y(Y) \int \ln f_1(Y, \widetilde{Z}; \theta(\beta)) f_1(\widetilde{Z}|Y; \boldsymbol{\theta_{MLE}}) d\widetilde{Z} dY$$

where $\theta_{MLE} = \arg\max_{\theta \in \Theta} \int g_Y(Y) \ln f_1(Y, \theta) dY$
Under assumption $g \in M_1$, we know $f_1(Y, Z^*; \theta_{MLE}) = g(Y, Z^*)$.

$$S_0(\beta) = \int \int \ln f_1(Y, \widetilde{Z}; \theta(\beta)) f_1(Y, \widetilde{Z}; \boldsymbol{\theta_{MLE}}) d\widetilde{Z} dY$$

Maximum $S_0(\beta)$ is equivalent to minimize $\int \ln \frac{f_1(Y, Z^*; \theta_{MLE})}{f_1(Y, Z^*; \theta(\beta))} f_1(Y, Z^*; \theta_{MLE}) dZ^* dY$

$\hat{\beta}$ is a pseudo-maximum likelihood estimator that minimizes Kullback-Leibler divergence from $f_1(Y, Z^*; \theta(\beta)), \beta \in B$ to true model $f_1(Y, Z^*; \theta_{MLE})$

Case 3: Model 1 is not correctly specified

$$\begin{aligned} S_0(\beta) &= \int g_Y(Y) \int \ln f_1(Y, \widetilde{Z}; \theta(\beta)) f_1(\widetilde{Z}|Y; \boldsymbol{\theta_{MLE}}) d\widetilde{Z} dY \\ &= \int \int \ln f_1(Y, \widetilde{Z}; \theta(\beta)) \underbrace{f_1(\widetilde{Z}|Y; \boldsymbol{\theta_{MLE}}) g_Y(Y)}_{p(Y, \widetilde{Z}; \theta_{MLE})} d\widetilde{Z} dY \end{aligned}$$

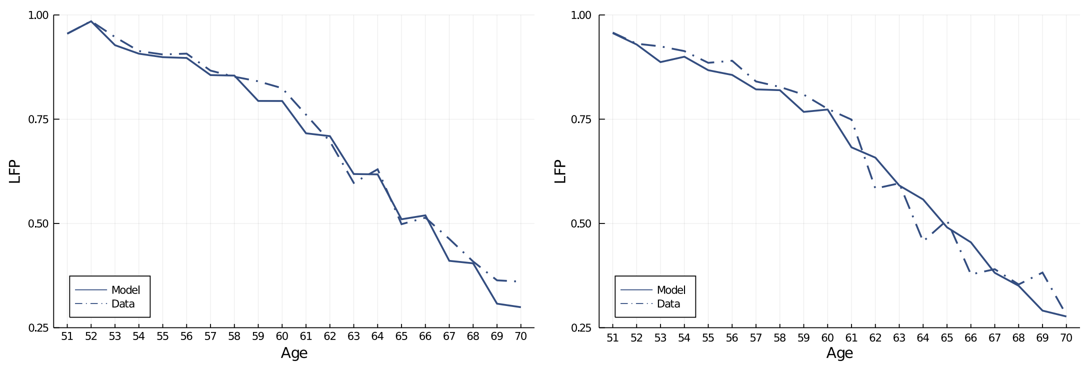
$\hat{\beta}$ is chosen in a way to best approximate a new density $p(Y, \widetilde{Z}; \theta_{MLE})$. The marginal distribution of observables $g_Y(Y)$ is from the true model, while the conditional distribution of latent variables on observables are taken from the MLE $f_1(\widetilde{Z}|Y; \theta_{MLE})$.

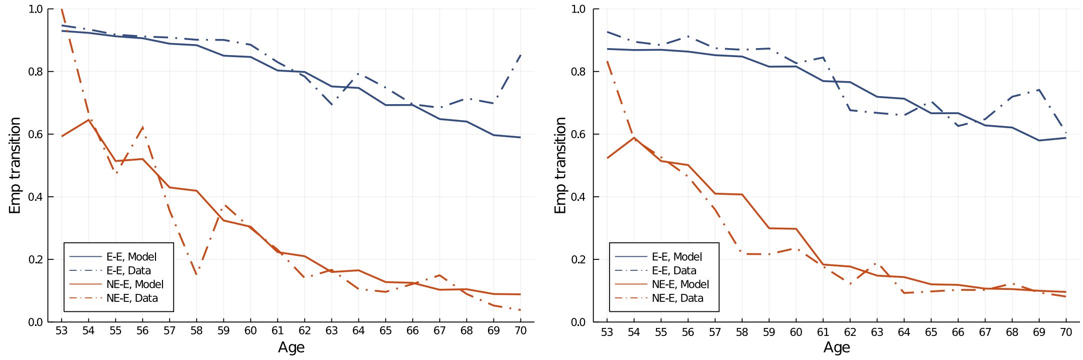
J The Fit Of Utility-based Model

Model fit

Figure J1: Employment

(a) LFP, HE (b) LFP, LE





(a) Employment transitions, HE (b) Employment transitions, LE

Figure J1: Job Destruction

(a) Proportion of $jd_{it} = 1$ by age, HE (b) Proportion of $jd_{it} = 1$ by age, LE

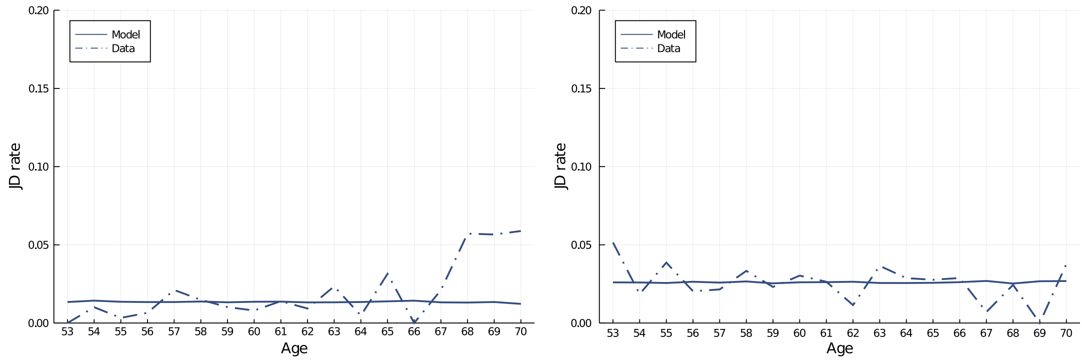
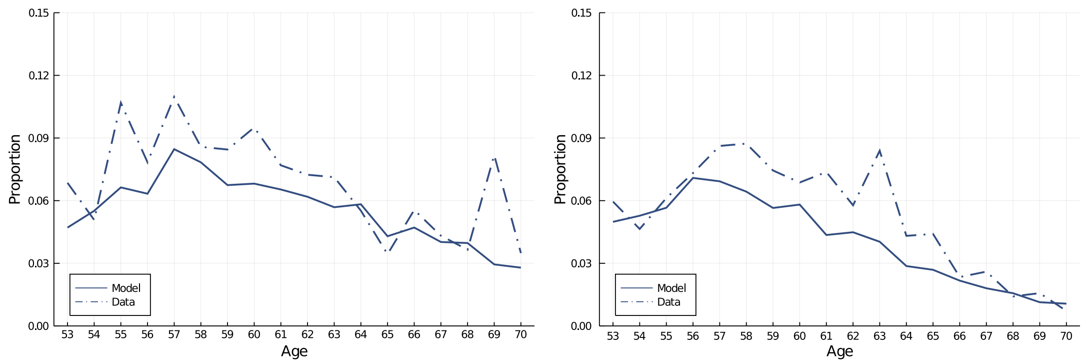
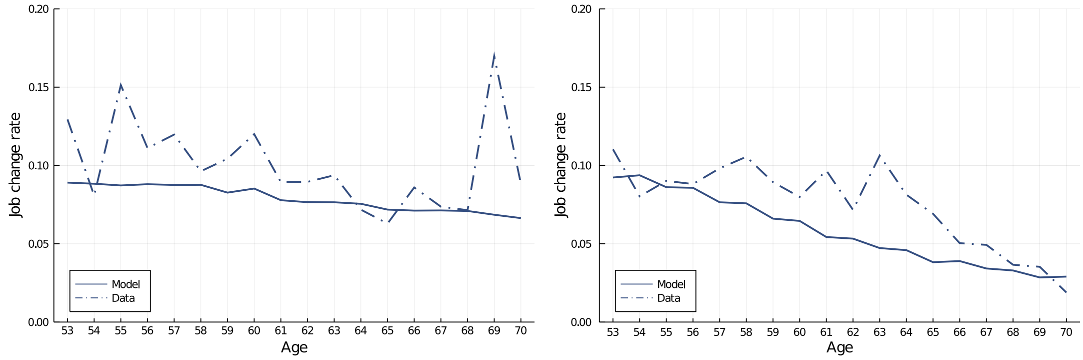


Figure J2: Job Change

(a) Proportion of $jc_{it} = 1$ by age, HE (b) Proportion of $jc_{it} = 1$ by age, LE





(a) Proportion of $jc_t = 1$ cond. on $E_{t-1} = 1$, HE (b) Proportion of $jc_t = 1$ cond. on $E_{t-1} = 1$, LE

Figure J2: Asset Accumulation

(a) Quantiles of asset by age, HE (b) Quantiles of asset by age, LE