TWO APPLICATIONS OF LIFECYCLE MODELS: TEACHERS' RETIREMENT UNDER TIME-VARYING PENSION RULES AND THE INCOME-HEALTH CORRELATION IN PSID

A Dissertation presented to

the Faculty of the Graduate School

at the University of Missouri

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

WEIWEI WU

Dr. SHAWN NI, Dissertation Supervisor

May 2017

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

TWO APPLICATIONS OF LIFECYCLE MODELS: TEACHERS' RETIREMENT UNDER TIME-VARYING PENSION RULES AND THE INCOME-HEALTH CORRELATION IN PSID

presented by Weiwei Wu,

a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Shawn Ni

Dr. Michael Podgursky

Dr. Aaron Hedlund

Dr. Chong He

ACKNOWLEDGMENTS

I would like to express my special appreciation and thanks to my advisor Professor Shawn Ni: you have been a tremendous mentor to me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career has been priceless. I would also like to thank my committee members, Professor Michael Podgursky, Professor Aaron Hedlund, and Professor Chong He, for serving as my committee members even at hardship. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions. I would especially like to express gratitude for the support from the Economic & Policy Analysis Research Center (EPARC). Professor Mark Ehlert, Dr. Dongwoo Kim, Mrs. Margie Gurwit, Mrs. Valerie Kulp, and all the people in EPARC, you all make EPARC a great research place with a family-like working environment.

A special thank to my graduate study director, Dr. X.H. Wang: you have been my mentor since the beginning of my journey. Your suggestions and kindness to me are unforgettable.

Special thanks to my family. Words cannot express how grateful I am to my beloved wife, Xinying. Thank you for your love and bring me my treasure, our lovely son, Andrew. Being a father myself make me understand that there are no words to thank my parents for all of the love and support you have offered to me.

I would also like to thank Xiang Li, Han Liu, Chen Huang, and all of my friends who supported me and incented me to strive towards my goal during the years in Mizzou.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Abstract: This study uses two structural models to predict public school teachers' retirement decisions and explore the correlations among people's health, occupational choices, and labor income.

Extensive literature suggests that the early exit of the late career teachers is inefficient. The first chapter uses the option value model introduced by Stock & Wise (1990) to analyze the retirement behavior of public school teachers in Missouri (MO) and simulate the effect of several retention policies. The existing studies failed to consider the sample selection bias of the late career teachers and did not consider the time-varying pension rules. I propose two algorithms that fix the sample selection bias, and I examine the competing theories on teachers' expectation of the time-varying pension rules. My modified model produces an excellent out-of-sample fit to the observed data. Simulations of counter-factual experiments show that the MO pension enhancements lead to earlier retirements, and well-designed retention policies can delay retirements.

The second chapter builds a life-cycle model to explain the widely acknowledged income-health correlation. I explore the dynamic relationships among occupations, health, and earnings observed in Panel Study of Income Dynamics (PSID). The simulations from the calibrated model fit the data well. Several simulation-based counter-factual experiments show that the occupations are mainly decided by education. "Manual" occupations lead to faster health deterioration and lower income compared to "non-manual" occupations. Given an occupation, people's health affects the labor supplied and labor income. Meanwhile, labor income affects health level through occupational choices, which is the main source for the health-income correlation.

These two applications reveal that life-cycle models can successfully model teachers' retirement and the health-income correlation, which can then help in evaluating the potential policies on retaining experienced teachers, and also help in exploring the correlations among health, occupation, and labor income.

Key Words: Teachers' Pension, Stock-Wise Model, Time-varying Pension Rules, Retention Policies, Health, Labor Income, Occupation, PSID

Chapter 1

Analyzing the Effect of Teacher Pension Rules Using a Structural Model of Retirement

1.1 Introduction

This chapter analyzes the retirement behavior of public school teachers using a structural model. The model is the option value model introduced by Stock & Wise (1990). I take the parameters estimated in an earlier study using Missouri (MO) data and find good out-of-sample fit to other Missouri data sets. The model can predict teachers' retirement decisions in different counter-factual scenarios. Therefore, it helps to evaluate the policy changes in the past and to forecast the effects of potential policies.

There are two reasons such a study is important for policy makers.

First, the existing pension rules appear to result in inefficient allocation of human capital of teaching. Experienced teachers have significant effects on students' future achievements. Rivkin *et al.* (2005) argue that teacher effectiveness determines the quality of schools; and Chetty *et al.* (2011) conclude that good teachers create substantial economic value. Both studies note the importance of retaining experienced teachers. However, public teachers retire earlier relative to other professionals. For example, the average retirement age of MO teachers is 57.7 during 2008-2013, which is lower than the average retirement age of other professionals (64.3) in the Current Population Survey during the same period. Harris & Adams (2007) show that the workforce exit rate of late career teachers is significantly higher than that of other professionals. The early exits of experienced teachers make inefficient allocation of the human capital of teaching. This study examines the effect of the current pension rules and explores potential retention policies.

The second policy-related motivation of this study is that the teacher pension costs create fiscal stresses on many states and local governments. According to Costrell & Podgursky (2009) [1], employer costs for teacher pensions averaged at 11.9% in March 2004, and this number rose to 19.9% by September of 2015. In contrast, the cost for private sector managers and professionals have remained relatively stable at 11.0% over the same period of time. Despite the rising fiscal costs, most of the teacher pension plans report large unfunded liabilities. Hence reform of teacher pensions is on the agenda due to the necessities of state and local public finance. Teacher pension reforms need empirical and theoretical guidance. This study seeks to provide a better assessment of policies based on a theoretical model.

The first chapter uses Stock & Wise (1990) to model how observable factors such as age, experience, salary, and pension rules influence the retirement decision. The basic idea of the model is to evaluate and compare the expected utility of retiring in the current year and in future years. If retiring in any one of the future years

[1] updated at
http://www.uaedreform.org/downloads/2013/12/quarterly-employer-contribution-chart-update.pdf

generated a higher expected utility than retiring in the current year, employees would continue to work, otherwise, they would retire. Recently, this model has been used to explain teachers retirement decisions. The model uses unobservable preference errors to capture unobservable factors that affect retirement. The parameters in Stock & Wise (1990) that pertain to teachers' preference towards discounting future income, intertemporal elasticity of substitution, distaste for working (relative to retiring), etc. do not change when pension rules do. Such a structural model is suitable for policy analysis for two reasons.

One reason is that the structural model can be used to simulate the effects of counter-factual pension policies. On estimating the effects of policies, the traditional empirical methods require policy variations and data observed from both the target sample and the relevant control group, which may not be available in most cases. Furthermore, it takes money and time to conduct policy experiments for social science. However, with a reliable structural model, the counter-factual simulations from the structural model can predict and evaluate the effects of different potential policies, without subjecting the economy to the resource cost of a real experiment.

Another reason is that structural models are more reliable for predicting the long-term effects of policies. Suppose retention bonuses are given to teachers for several years or longer in a real experiment, then researchers can collect data to estimate the effects of the bonus policy. However, experiments are often limited in duration because of money and time constraints. Most research using the observed experimental data can only show the short-term effect of a temporary policy. But, temporary policies are different from permanent policies, and the short-term effect of a policy may differ from its long-term effect. The estimates based on the short-term responses may be misleading, therefore a reliable structural model is needed for evaluating the long-term effects.

The parameters used in this chapter are estimated by Ni & Podgursky (2016) from a 2002-2008 panel data of Missouri teachers aged 47-59 in 2002, while other samples are used for policy evaluations. So, the first step is to examine the out-of-sample fit.

There are two main difficulties for the model to fit the out-of-sample data. The first one is the sample selection bias. The Stock-Wise model targets late career teachers (e.g., age 48-65 for Missouri). But the observed sample of the late career teachers always includes those teachers who were eligible to retire but chose to stay, meanwhile, the observed sample does not include those teachers who retired before the first observation year. In other words, the observed sample is biased towards the teachers who preferred to continue teaching. In the model, the teaching preference is captured by an error term. In order to adjust the model to fit the observed selected sample, I adjust the error terms of different teachers. Secondly the 1994 cohort of MO teachers suffered consecutive rule changes in the late 1990s. If under fixed pension rules, teachers would use the fixed rule to evaluate the expected utility of retiring in their current year and in their future years. However, if there are some changes in the pension rules, teachers may be able to predict the future pension rules and evaluate their expected utility of retiring in future years accordingly. It is difficult to model how teachers predict the time-varying pension rules. There are several competing assumptions, such as perfect forecast, short-term forecast, and no forecast. I evaluate these assumptions and find that teachers actually do not make any forecasts of their future pension rules. They make their retirement decisions only based on their ongoing pension rules.

With appropriate adjustments for the sample selection bias and an appropriate expectation assumption, the model produces good out-of-sample fits to other observed samples. I then seek to answer policy questions such as “would teachers change their retirement decisions if there were no enhancements in the 1990s?” To answer this question I use this model to simulate different counter-factual scenarios in the past. By comparing the scenarios with pension rule changes and the scenarios without pension rule changes, I conclude that those changes in the late 1990s did not efficiently retain experienced teachers, instead, those changes actually created early retirement incentives. For example, one of the most important policy changes is the introduction of the "25 and out" policy. It allows teachers with 25 years of experience to retire at any age. My result shows this policy actually decreases the average retirement age by 0.6 years.

The model can also evaluate the effects of several potential retention policies. The first type of retention policy is retention bonus. The result shows that teachers can be effectively retained by permanent retention bonus but not by a temporary bonus. Temporary policies tend to be more costly than permanent policies for each extra year of teaching.

The second type of retention policy is Deferred Retirement Option Plans (DROP) which permit employees to retire and begin collecting all or part of their pension annuities, and to continue working full-time for a limited period of time. My result shows that DROP plans can retain experienced teachers at a lower cost compared to the retention bonuses.

Besides the above data-dependent analysis, I also conduct comparative analysis to examine how the pension rules affect the retirement incentives. Under a number of pension rules, the model can predicts the retirement probabilities at all combinations of age and experience. My simulations show that the both the contribution rate and the replacement rate in a Defined Benefit (DB) pension plan play important roles in retirement decisions. A higher contribution rate reduces the salary net of the contribution, hence increases the probability of retirement. Also, a higher replacement rate increases the benefits after retirement and increases the probability of retirement.

More flexible retirement criteria, such as "25 and out" and "rule of 80", also result in earlier retirement.

The next section introduces the structural model used in this chapter. Section three describes the data. Section four discusses the sample selection problem and the two methods for adjustment. Section five reports the history of MO pension rules and discusses the assumptions of teachers' expectations on the time-varying pension rules. Section six reports the estimated effects of the pension benefit enhancements. Section seven examines the effects of different retention bonuses to Missouri teachers. Section eight shows the effects of the DROP plans. Section nine conducts comparative analysis. The last section concludes.

1.2 Stock-Wise Model

Stock & Wise (1990) built an "option value" model on analyzing the pension plan effects on the retirement decisions of old employees in big firms. I apply the Stock-Wise (SW) model to teacher retirement. The teacher's expected utility in period t can be written as a function of the expected retirement year: m (with $m = t, \cdots, T$ and $T = 101$ is an upper bound on age). Denote the pension rule in year t as R_t, the teachers' contribution rate as c_t. Additionally, I assume teachers calculate their expected pension benefits based on the current pension rules. In year t, the expected utility of retiring in period m is the discounted sum of pre- and post-retirement expected utility

$$\mathbb{E}_t V_t(m, R_t) = \mathbb{E}_t \{ \sum_{s=t}^{m-1} \beta^{s-t} [(k_s(1-c_t)Y_s)^\gamma + w_s] + \sum_{s=m}^{T} \beta^{s-t} [(B_s(R_t))^\gamma + \xi_s] \}, \quad (1.1)$$

where $0 < k_s < 1$ captures the disutility of working, and Y is real salary. $B_s(R_t)$ is the real pension benefit collected in year s under the rules of year t, R_t. The unobserved innovations in preferences are AR(1): $w_s = \rho w_{s-1} + \epsilon_{ws}$, $\xi_s = \rho \xi_{s-1} + \epsilon_{\xi s}$. Denote the error terms $\nu_s = w_s - \xi_s$, $\epsilon_s = \epsilon_{ws} - \epsilon_{\xi s}$. Then it follows that:

$$\nu_s = \rho \nu_{s-1} + \epsilon_s. \tag{1.2}$$

I assume ϵ_s is iid $N(0, \sigma^2)$.

The dis-utility of work, k_s, is assumed to be decreasing with age: $k_s = \kappa(\frac{60}{age})^{\kappa_1}$. The retirement decision in year t is choosing $m = t, \cdots, T$ that maximizes $\mathbb{E}_t V_t(m)$.

The future is uncertain, teachers are risk averse, but the retirement decision is irreversible. So there is a value associated with continuing teaching and keeping the retirement option open. This is why this model is known as an "option value" model.

Besides the uncertainty in preference shocks there is an uncertainty of survival: for a teacher alive in year t, I denote the probability of survival to period $s > t$ as $\pi(s|t)$. To quantify the option value, write the expected gain from retirement in year m over retirement in the current period t as:

$$G_t(m, R_t) = \mathrm{E}_t V_t(m, R_t) - \mathrm{E}_t V_t(t, R_t) = g_t(m, R_t) + K_t(m)\nu_t, \tag{1.3}$$

where

$$\begin{aligned} g_t(m, R_t) = &\sum_{s=t}^{m-1} \pi(s|t)\beta^{s-t}\mathrm{E}_t(k_s(1-c_t)Y_s)^\gamma + \sum_{s=m}^{T} \pi(s|t)\beta^{s-t}\mathrm{E}_t(B_s(R_t))^\gamma \\ &- \sum_{s=t}^{T} \pi(s|t)\beta^{s-t}\mathrm{E}_t(B_s(R_t))^\gamma \end{aligned}$$

is the difference in expected utility between retiring in year $m > t$ and retiring now (in year t). Because the teacher's future salary and pension benefits are predictable, in the empirical analysis, I replace the expected salary and benefit in $g_t(m, R_t)$ with a forecast based on historical data. In the last term in (1.3), $K_t(m) = \sum_{s=t}^{m-1} \pi(s|t)(\beta\rho)^{s-t}$ depends on unknown parameters and the AR(1) error term ν_t given in (1.2). Let $m_t^\dagger(R_t) = argmax\ g_t(m, R_t)/K_t(m)$, then the probability that the teacher retires in period t ($G_t(m, R_t) \leq 0$ for all $m > t$) is

$$Prob(\frac{g_t(m^\dagger, R_t)}{K_t(m^\dagger)} \leq -\nu_t). \tag{1.4}$$

The condition (1.4) is affected by the time-varying pension rules. Suppose a teacher is observed for period $t, t+1, .., t+n$ and she retired in $t+n$, then the likelihood is the probability of the joint event $\pi(\frac{g_t(m_t^\dagger,R_t)}{K_t(m_t^\dagger)} > -\nu_t, .., \frac{g_{t+n-1}(m_{t+n-1}^\dagger,R_{t+n-1}}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}, \frac{g_{t+n}(m_{t+n}^\dagger,R_{t+n})}{K_{t+n}(m_{t+n}^\dagger)} < -\nu_{t+n})$.

Denote $\boldsymbol{\nu}_{t,t+n} = (\nu_t, \cdots, \nu_{t+n})'$.

The event $(\frac{g_t(m_t^\dagger,R_t)}{K_t(m_t^\dagger)} > -\nu_t, .., \frac{g_{t+n-1}(m_{t+n-1}^\dagger,R_{t+n-1})}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}, \frac{g_{t+n}(m_{t+n}^\dagger,R_t)}{K_{t+n}(m_{t+n}^\dagger)} < -\nu_{t+n})$ can be expressed as $\boldsymbol{\nu}_{t,t+n} \in$ a corresponding region in space R^n. The marginal distribution of $\nu_t \sim N(0, \sigma_\nu^2)$ where $\sigma_\nu^2 = \frac{\sigma^2}{1-\rho^2}$. Given $\nu_t = \rho\nu_{t-1} + \epsilon_t$, the covariance of $\boldsymbol{\nu}_{t,t+n}$ is given by

$$\boldsymbol{\Sigma} = \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & ... & \rho^{n-1} & \rho^{n-1} \\ \rho & 1 & \rho & ... & \rho^{n-2} & \rho^{n-2} \\ \rho^2 & \rho & 1 & ... & \rho^{n-3} & \rho^{n-3} \\ .. & .. & .. & .. & .. & \\ \rho^{n-1} & \rho^{n-2} & ... & ... & \rho & 1 \\ & & & & & \end{pmatrix}.$$

The covariance matrix $\mathbf{\Sigma}$ permits a Cholesky decomposition $\mathbf{\Sigma} = \mathbf{V}\mathbf{V}'$ with

$$\mathbf{V} = \begin{pmatrix} \frac{1}{\sqrt{1-\rho_i^2}} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \frac{\rho_i}{\sqrt{1-\rho_i^2}} & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \frac{\rho_i^2}{\sqrt{1-\rho_i^2}} & \rho_i & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ \frac{\rho_i^{T-1}}{\sqrt{1-\rho_i^2}} & \rho_i^{T-2} & \rho_i^{T-3} & \rho_i^{T-4} & \cdots & \rho_i^2 & \rho_i & 1 \end{pmatrix}.$$

The log likelihood is

$$logL(\gamma, \kappa, \kappa_1, \beta, \sigma, \rho \mid \mathbf{Y}, \mathbf{B}, \mathbf{R}) = \sum_{i=1}^{I} log\pi_i(\boldsymbol{\nu}_{t,t+n} \in A_i) = \sum_{i=1}^{I} log \int_{A_i} \phi(\boldsymbol{\nu}_{t,t+n}) d\boldsymbol{\nu}_{t,t+n} \quad (1.5)$$

where for teacher i retiring in period $t+n$, $\boldsymbol{\nu}_{t,t+n} \in A_i$ if $\frac{g_t(m_t^\dagger(R_t))}{K_t(m_t^\dagger)} > -\nu_t, .., \frac{g_{t+n-1}(m_{t+n-1}^\dagger(R_{t+n-1}))}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}, \frac{g_{t+n}(m_{t+n}^\dagger(R_{t+n}))}{K_{t+n}(m_{t+n}^\dagger)} < -\nu_{t+n}$, and $\phi(.)$ denotes multivariate normal density distribution of $N(0, \mathbf{\Sigma})$.

In estimating the structural parameters, an obstacle to evaluating the likelihood is the large computational time of n -dimensional integration. In this study, I use the estimates from Ni & Podgursky (2016), in which, the likelihood is numerically evaluated using the Geweke-Hajivassiliou-Keane (GHK) simulator proposed by Börsch-Supan & Hajivassiliou (1993). For a longer data panel, the GHK simulator is more efficient than the Monte Carlo approach for the frequency of retirement. Ni & Podgursky (2016)'s estimation on Missouri 2002-2008 PSRS data yields $\beta = 0.965$, $\gamma = 0.716$, $\rho = 0.643$, $\kappa_s = \kappa(60/age\ at\ year\ s)^{\kappa_1}$, $\kappa = 0.640$, $\kappa_1 = 0.976$, $\sigma = 3660.166$. These parameters are used in this study. Using the estimates from Ni & Podgursky (2016) is not merely for convenience. It is of interest to examine whether the parameters on teacher preferences are applicable to different teacher samples who are under different pension rules.

1.3 Data

This chapter uses the administrative data of public school teachers in Missouri. Public school teachers in most of the states in the U.S. are covered by a DB pension plan. It typically takes 3-5 years for teachers to become vested in the system; once vested, a teacher can collect her pension upon becoming collection eligible. The normal retirement age is one way that collection eligibility is determined. It varies across plans between the ages of 60 and 65, and can be service-based as well (e.g., 30 years of service). There are also early-retirement provisions in most systems that allow individuals to retire and begin collecting benefits prior to normal retirement. These provisions typically depend on either (1) work experience alone or (2) a combination of age and work experience.

Take Missouri as an example, Missouri public teachers are covered by a state-wide DB pension system, which is the Public School Retirement System (PSRS). Benefits at retirement in PSRS are determined by the following formula (some variant of which is nearly universal in teacher DB systems):

$$Annual\ Benefit = S \times FAS \times R$$

where S is service years (essentially years of experience in the system), FAS is final average salary calculated as the average of the highest three years of salary, and R is the replacement factor. Suppose teachers earn 2.5% for each year of teaching service up to 30 years. Thus, a teacher with 30 years experience and a final average salary of $\$60,000$ would receive $30 \times \$60,000 \times 0.025 = \$45,000$. The pension rules in PSRS are changing over time. Under the current rules, Missouri teachers become eligible for a full pension if they meet one of three conditions: a) they are sixty years of age with at least five years of teaching experience, b) thirty years of experience (and any age), or c) the sum of age and years of service equals or exceeds 80 ("rule of 80"). I follow the 1994 cohort of PSRS teachers who were in the "retirement window" (aged 48-65 with at least 5 years of experience) in 1994 for 15 years till 2008. The sample means for age and experience are listed in Table 1.9. Since the pension rules changed over time, the average retirement age and retirement experience also change over time. There are more discussions about the changing of PSRS pension rules in section five. I will also examine the counter-factual scenarios that without or with fewer pension enhancements in the late 1990s in Missouri.

Most (92.22%) of the teachers in the 1994 cohort had retired by 2008, but the retention policies should focus on more recent teachers in Missouri. Besides the 15-year panel of the 1994 cohort in Missouri, I have three shorter panels of MO teachers:

- 2008 cohort: MO teachers who were aged 50-65 and had 10 or more years of experience in the 2007-2008 school year. The summary statistics are in Table 1.10
- 2011 cohort: MO teachers who were aged 48-65 and had 5 or more years of experience in the 2010-2011 school year. The summary statistics are in Table 1.11.
- 2011 STEM cohort: MO teachers who were aged 48-65, had 5 or more years of experience, and were teaching STEM courses in the 2010-2011 school year. The summary statistics are in Table 1.12.

The 2008 cohort is used for testing the performance of different bonus policies, while the 2011 cohort is used for testing the performance of the proposed approaches for correcting the sample selection bias and the targeted DROP policies for the STEM teachers.

Since the parameters used in our model are taken from the estimates in Ni & Podgursky (2016) based on 2002-2008 Missouri data, all predictions made in this chapter are out-of-sample predictions.

Suppose we use Sample A of teachers to estimate structural parameters to answer policy questions for Sample B. The first step is to measure the out-sample fit for Sample B. For each teacher in Sample B, I will draw a pre-set large number (say 10,000) of time series of preference errors given by the AR(1) process in equation 1.2, and simulate the binary decisions (1=retire, 0=stay) based on inequality 1.4 under the pension rules for Sample B. The simulated frequency of the retirement decision for the teacher in each period is predicted retirement probability of the teacher in the period.

Because the quantities of interest are aggregate variables (e.g., the fraction of all teachers in 2011 who choose not to retire, i.e., the survival rate of 2011), for testing the goodness of fit I compare the simulated aggregate quantities such as the survival rates with the observed ones in Sample B. As described in the paragraph above, for each teacher in Sample B we can generate a probability of surviving in 2011. Averaging over all teachers yields the point prediction of the survival rate, which can be compared with the observed one. I test the out-of-sample fit of each data set before answering policy questions.

To quantify the statistical significance of the mismatch of the predicted and observed quantities we should consider two types of errors. One is sampling errors that affected the parameters estimated from Sample A. This is the base of confidence band commonly seen in other studies. The width of the confidence band is a function of the standard errors of the estimator. Because of the high cost of simulations using multiple parameters on a large number of teachers, in this study I only use one value of the parameters and ignore this type of errors.

The second type of sampling errors comes from the draws of preference errors for each teacher. The aggregate quantities such as survival rate in 2011 are also affected by this type of errors. Suppose aggregation of the retirement decision for each draw of the preference error for each teacher yields a survival rate of 2011 for Sample B. I can obtain 10,000 survival rates of 2011 from sets of preference errors, therefore I can plot the 99% simulation band of the survival rates. In the study, when I test the goodness of fit of the out-of-sample predictions, the 99% simulation bands are always plotted along with the aggregate quantities, such the survival rate, age distribution of retirees, and the experience distributions of the retirees. In most cases, the observed data are covered in these bands. This suggests that the observed data are within the range likely covered by simulated quantities; even I take the single set of parameter value estimated from another sample (Sample A) and ignore the sampling errors in Sample A.

1.4 Sample Selection Problem

Sample selection bias occurs when the observed sample is not randomly selected from the population. Due to the sample selection problem, the estimates based on the observed sample are biased, which can lead to misleading conclusions and inferences. Some researchers have developed methodologies to eliminate the sample selection bias in estimation (Heckman (1981), Hausman & Wise (1979), and Wooldridge (2005)).

My study also has a sample selection problem since the observed late-career teacher sample *includes* teachers who were eligible for retirement but who chose to wait, but *excludes* those who chose to retire prior to 1994 (where $t = 1$). However, the sample selection problem is special in several ways. First, my observed samples are panel data, so the inclusion (selection) condition doesn't only affect the decision in the first period, but also affect the decisions in all the following periods. Second, the sample selection problem in this study results in biased retirement rates and survival rates in the simulation, instead of biased estimates of parameters of the model. In this section, I explain the sample selection problem within the Stock-Wise framework. Then introduce two possible ways to adjust the sample selection problem.

In the Stock-Wise model, if a teacher is observed for period $t, t+1, .., t+n$ and she retired in $t+n$, then the likelihood is the probability of the joint event $\pi(\frac{g_t(m_t^\dagger, R_t)}{K_t(m_t^\dagger)} > -\nu_t, .., \frac{g_{t+n-1}(m_{t+n-1}^\dagger, R_{t+n-1}}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}, \frac{g_{t+n}(m_{t+n}^\dagger, R_{t+n})}{K_{t+n}(m_{t+n}^\dagger)} \leq -\nu_{t+n})$. By the definition of conditional probability, this joint probability can be viewed as products of a sequence of conditionals:

$$
\begin{aligned}
&\pi(\frac{g_t(m_t^\dagger, R_t)}{K_t(m_t^\dagger)} > -\nu_t, .., \frac{g_{t+n-1}(m_{t+n-1}^\dagger, R_{t+n-1})}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}, \frac{g_{t+n}(m_{t+n}^\dagger, R_{t+n})}{K_{t+n}(m_{t+n}^\dagger)} < -\nu_{t+n}) \\
&= \pi[(\frac{g_{t+n}(m_{t+n}^\dagger, R_{t+n})}{K_{t+n}(m_{t+n}^\dagger)} \leq -\nu_{t+n}) \mid \frac{g_t(m_t^\dagger, R_t)}{K_t(m_t^\dagger)} > -\nu_t, .., \frac{g_{t+n-1}(m_{t+n-1}^\dagger, R_{t+n-1})}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}] \\
&\times \pi[(\frac{g_{t+n-1}(m_{t+n-1}^\dagger, R_{t+n-1})}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}) \mid \frac{g_t(m_t^\dagger, R_t)}{K_t(m_t^\dagger)} > -\nu_t, .., \\
&\frac{g_{t+n-2}(m_{t+n-2}^\dagger, R_{t+n-2})}{K_{t+n-2}(m_{t+n-2}^\dagger)} > -\nu_{t+n-2}] \\
&\dots \\
&\times \pi[(\frac{g_{t+1}(m_{t+1}^\dagger, R_{t+1})}{K_{t+1}(m_{t+1}^\dagger)} > -\nu_{t+1}) \mid \frac{g_t(m_t^\dagger, R_t)}{K_t(m_t^\dagger)} > -\nu_t] \\
&\times \pi[\frac{g_t(m_t^\dagger, R_t)}{K_t(m_t^\dagger)} > -\nu_t].
\end{aligned}
$$

The last term in the above equation, $\pi[\frac{g_t(m_t^\dagger, R_t)}{K_t(m_t^\dagger)} > -\nu_t]$, is the initial condition that makes the teacher observable in the first observation year $t = 1$. If there are some retirement decisions are made before period $t = 1$, the "entering" (selection) condition contains information for the individuals before they enter the observed panel, which is known as the initial value problem (Akay (2009)). The choice of staying or retiring before $t = 1$ is correlated with the characteristics of the teachers. Ignoring or mis-specifying the initial value problem leads to serious bias in the estimates.

The observed teachers in my sample are late career teachers. All the samples contain teachers who were eligible to retire but chose to wait, but do not contain the teachers who retired before the first observation year ($t = 1$). These lead to a sample selection problem: the percentage of stayers in the observed data is higher than that in the population. The stayers are more likely to stay than other teachers, and their initial condition is

$$-\frac{g_1(m_1^\dagger, R_1)}{K_1(m_1^\dagger)} < \nu_1 \tag{1.6}$$

This implies that initial shocks (ν_1) of stayers are more restricted than those of others. So I cannot draw ν_1 from the unconditional stationary distribution for everyone. If ignore the sample selection problem, the model would over-predicts the retirement rates and under-predicts the survival rates.

There are two prevalent solution strategies for the initial value problem in estimation proposed by Heckman (1981) and Wooldridge (2005). Both realized that the initial value ν_1 varies across different individuals. Heckman (1981) uses a reduced-form to approximate the unknown conditional distribution for the initial value (conditional on other exogenous covariates, and the unobserved personal characteristics) and put this approximated distribution into the joint likelihood function. Wooldridge (2005) offers a simple alternative by specifying the distribution of the initial value condition on the fact that the individual can be observed in the first period ($t = 1$).

1.4.1 Two Approaches to Correcting the Sample Selection Bias

To correct the sample selection bias in the predicted retirement rates and survival rates, the distribution of the initial value, ν_1, should be sample dependent. There are two possible approaches to restrict the distribution of ν_1. The first method set a unique mean value for ν_1 of all teachers in the sample, while the second method customizes the distribution of ν_1 for different teachers according to their previous decisions before they were observed in 1994.

Method 1: Choose a constant initial condition ν_0 for the whole sample and draw $\nu_t(t = 1, ..., T)$ from equation (1.2).

As mentioned, the model underestimates the survival rates due to the sample selection problem. To get better fits on the survival rates, ν_1 should have a lower bar in the model. The AR(1) process in equation (1.2), $\nu_1 = \rho\nu_0 + \epsilon_1$, says a bigger ν_0 gives a bigger ν_1. A positive ν_0 creates a lower bar for ν_1 in the model, then the predicted survival rates can fit the observed survival rates better. For a different sample, the value of ν_0 can be different. ν_0 measures the researcher's belief concerning the average sample selection bias in the initial year. The larger ν_0 is, the greater the average sample selection bias.

Method 2: Simulate ν_0 based on pre-sample history.

Suppose a teacher was eligible to retire J years prior to the first observation year ($t = 1$) and chose to stay. I back-track the retirement decision in each of the J years and compute g(-J+1),..,g(0).Then compare them with preference errors ν_j prior to $t = 1$, and make sure (1.6) holds.

The steps are:

(a). draw ν_{-J} from $N(0, \frac{\sigma^2}{1-\rho^2})$, generate ν_t for $t = -J + 1, .., 0$ based on (1.2).

(b). if draws imply the teacher did not retire before year $t = 0$ (all ν_t for $t =$

$-J+1,..,0$ satisfies (1.6)), then keep the draws of ν_0, and draw $\nu_t, (t = 1, ..., t+n)$ based on (1.2).

(c). repeat steps (a) and (b) until the number of draws in step (b) reaches a pre-set number (say 10,000).

1.4.2 Performance of Methods for Adjusting the Sample Selection Bias

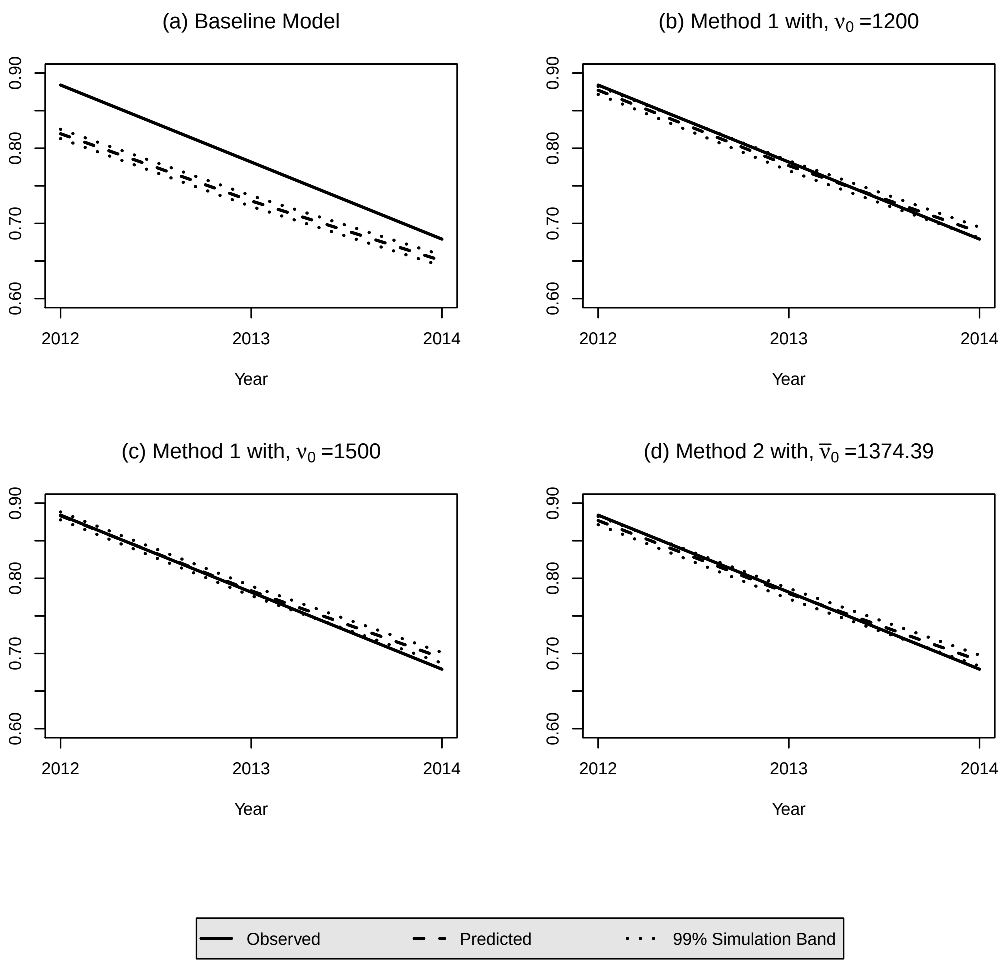
Due to the sample selection problem, our observed sample contains more teachers who are less likely to retire. Before applying any method to adjust the sample selection bias, the model over-predicts the retirement rates and under-predicts the survival rates. After applying either of the adjusting methods, predicted survival rates match the observed survival rates much better.

Both the sample selection problem and the time-varying pension rules affect the observed data during 1994-2008 simultaneously, and the time-varying rules complicate the problem of sample selection. In order to show the performance of the above two solution methods for the sample selection bias, I use the 2011 cohort of the PSRS data to get rid of the effect of time-varying pension rules. The pension rules were largely unchanged in 2011-2014. There were 19,683 MO teachers in the retirement window (aged 48-65 and having more than five years of experience) in the 2010-2011 school year. By the end of 2014, 13367 teachers were still working. As seen in Figure 1.1 (a), before considering the sample selection problem, the model under-predicts the survival rates, the observed survival rates are even not in the 99% simulation band. When applying method 1 to correct the sample selection bias, I report two cases: $\nu_0 = 1200$ and $\nu_0 = 1500$ in (b) and (c) of Figure 1.1. Both of them fit the survival rates much better due to the non-zero value for ν_0. The ν_0 measures the average sample selection bias of the initial year. The effect of ν_0 diminished as time goes on, since $\rho < 1$ in equation (1.2).

Method 2 is also good to deal with the sample selection problem. For Missouri, J in method 2 is decided by two conditions: 1). $age \geq 60$ and $exp \geq 5$ or 2). $age + exp \geq 80$. For example, if a teacher had $age + exp = 85$ in 2011, then $J = 5$. I compute his $g(-4)$ to $g(0)$ and compare them with the simulated ν_{-4} to ν_0 to make sure the teacher did not retire in these 5 years. The predicted survival rates with method 2 fit the observed survival rates fairly well in (d) of Figure 1.1. The average ν_0 simulated in method 2 is 1374.39, which is a number in between the mean values for method 1.

In short, both method 1 and method 2 can fix the sample selection bias and produce a fairly good fit to the observed data. In the following sections, I fix the sample selection bias by method 2.

Figure 1.1: **Observed and Predicted Survival Rates, Missouri**



Notes: Data are the administrative data from 2011-2014 for Missouri teachers aged 48-65 with at least 5 years of experience in 2011. Predicted survival rates are from the following model settings respectively:

(a) Baseline model: the model without any efforts on correcting the sample selection problem.

(b) Method 1 with $\nu_0 = 1200$: use method 1 to correct the sample selection problem and set $\nu_0 = 1200$.

(c) Method 1 with $\nu_0 = 1500$: use method 1 to correct the sample selection problem and set $\nu_0 = 1500$.

(d) Method 2: use method 2 to correct the sample selection problem.

The 99% simulation band describes the sampling errors of the draws of the preference errors for each teacher.

1.5 Pension Enhancements in Missouri

In the late 1990s, the National Conference of State Legislators (NCSL) reported that educator pensions were enhanced in more than half the states. In most states, teachers' benefits were automatically and retroactively adjusted to reflect the enhancements at the time of their enactment without additional required contributions. Therefore, teachers whose retirement plans happened to coincide with the timing of the benefit enhancements were able to collect the more generous pensions even though their lifetime contributions were structured to fund a much less remunerative flow of benefits.

Take Missouri as an example, Table 1.1 describes the series of enhancements that occurred to the PSRS in Missouri. In 1995 the formula factor was 0.023, the final average salary was calculated based on the highest five years of earnings, and early retirement was only possible through the "55 and 25 rule". The "55 and 25 rule" allowed a teacher to retire and collect benefits without penalty if both of two conditions were met: (1) the teacher had to be at least 55 years old, and (2) he/she had to have accrued at least 25 years of system service. By 2002 the formula factor had been raised from 0.023 to 0.025, the final-average-salary calculation changed from the highest five to highest three years of earnings, and the "25 and out" and "Rule of 80" provisions had been incorporated into the system (the "Rule of 80" is a more flexible version of the "55 and 25 rule", while "25 and out" allows teachers with 25 years of experience to retire at any age). In addition, the cap on post-retirement cost-of-living adjustments (COLAs) was raised from 65 to 80 percent of the baseline annual pension payment, and a retroactive bonus was added for teachers who reached their 31st year of system service.

In order to generate good out-of-sample predictions, the model has to involve the benefit enhancements of the teacher pension system in Missouri.

Table 1.1: **PSRS Pensions Rule Changes**

Effective Year	FAS	COLA	Retirement Age and Experience
1995	0.023	0.65	Age$\geq$ 55 and Exp $\geq$ 25, or
	FAS using average salary		Age$\geq$ 60 and Exp $>$ 5, or EXP$\geq$ 30,
	of the highest 5 years		
1996	0.023	0.65	Add "25 and out" early retirement
	district health insurance		(with EXP$\geq$25),
	added to the FAS		
1997	0.023	0.75	same
1998	0.025	0.75	"25 and out" formula factors increased
1999	0.025	0.75	same
	FAS using average salary		
	of the highest 3 years		
2000	0.025	0.75	Add the "rule of 80" Age+ Exp $\geq$ 80
2001	0.025	0.80	same
2002	0.0255 if EXP$\geq$ 31	0.80	same

Notes: The table lists important changes in pension benefit rule of the state-wide educator plan–the Public School Retirement System (PSRS) in Missouri from 1995 to 2002. The "25 and out" rule in 1996 permits retirement at a reduced benefit factor (replacement rate) R if teachers have 25 or more years of experience, with the following benefit factors: 2% for teachers with 25 years of experience, 2.05% for 26 years, 2.1% for 27 years, 2.15% for 28 years and 2.2% for teachers with 29 years of experience. The "25 and out" rule in 1998 raises the benefit factors to 2.2% for 25 years, 2.25% for 26 years, 2.3% for 27 years, 2.35% for 28 years and 2.4% for teachers with 29 years of experience. The "rule of 80" permits regular retirement when age+experience $\geq$ 80.
Notes: The teacher's contribution rate c was 10.5% in 1990's and has since increased to 14.5% in late 2000's.

1.5.1 Different Expectations

Under fixed pension rules, teachers would use the fixed rule to evaluate the expected utility of retiring in current year and in their future years. However, if there are changes in the pension rules, teachers may be able to predict the future pension rules and evaluate their expected utility of retiring in future years accordingly. Given the large and persistent benefit enhancements in the teacher pension system of Missouri, the model has to consider the teachers' expectations regarding the time-varying pension rules when teachers make their retirement decisions. There are several competing assumptions on teachers' forecast of future pension rules. I discuss and evaluate these assumptions in this section.

Myopic Expectation

The option-value model in section 2 assumes that the teachers calculate their expected pension benefit for all future years based on the current pension rule that they observed, and teachers assume that the current pension rule will persist in the future. In period t, the expected utility for any future retirement year m is a function of the current rule R_t: $\mathrm{E}_t V_t(m, R_t)$. For example, a teacher who made a decision in 1995 would use R_{1995} to evaluate the expected utilities for any future retirement year, like

$$\mathrm{E}V_{1995}(1996, R_{1995}), \mathrm{E}V_{1995}(1997, R_{1995}), \ldots, \mathrm{E}V_{1995}(2000, R_{1995}) \ldots$$

Also, a teacher who made a decision in 2000 would use R_{2000} to evaluate the expected utilities for any future retirement year:

$$\mathrm{E}V_{2000}(2001, R_{2000}), \mathrm{E}V_{2000}(2002, R_{2000}), \ldots, \mathrm{E}V_{2000}(2010, R_{2000}) \ldots.$$

Myopic Expectation is labeled as **Case (A)** in the latter section.

Perfect Foresight

The perfect foresight assumption of the future pension rules assumes that teachers can correctly predict the future pension rules and will be able to evaluate their future retirement benefits based on those future pension rules. With perfect foresight of future pension rules, the expected utility in period t of retiring in period m becomes

$$\mathrm{E}_t V_t(m, R_m) = \mathrm{E}_t\{\sum_{s=t}^{m-1} \beta^{s-t}[(k_s(1-c_s)Y_s)^\gamma + w_s] + \sum_{s=m}^{T} \beta^{s-t}[(B_s(R_m))^\gamma + \xi_s]\}, \quad (1.7)$$

And the difference in expected utility between retiring in year $m > t$ and retiring now (in year t) is

$$\begin{aligned} g_t(m, R_m) = \; & \textstyle\sum_{s=t}^{m-1} \pi(s|t)\beta^{s-t}\mathrm{E}_t(k_s(1-c_s)Y_s)^\gamma + \sum_{s=m}^{T} \pi(s|t)\beta^{s-t}\mathrm{E}_t(B_s(R_s))^\gamma \\ & - \textstyle\sum_{s=t}^{T} \pi(s|t)\beta^{s-t}\mathrm{E}_t(B_s(R_s))^\gamma \end{aligned}$$

Continuing with the previous example, under the perfect foresight assumption, a teacher who a made decision in 1995 can perfectly predict the future pension rule in year m ($m > 1995$): R_m and evaluate the expected utilities accordingly, like

$$\mathrm{E}V_{1995}(1996, R_{1996}), \mathrm{E}V_{1995}(1997, R_{1997}), \ldots, \mathrm{E}V_{1995}(2000, R_{2000}) \ldots$$

Also, a teacher who made a decision in 2000 would use R_m to evaluate the expected utilities for any future retirement year m ($m > 2000$):

$$\mathrm{E}V_{2000}(2001, R_{2001}), \mathrm{E}V_{2000}(2002, R_{2002}), \ldots, \mathrm{E}V_{2000}(2010, R_{2010}) \ldots$$

If the pension rules do not change at all or change consistently every year, the perfect foresight assumption is reasonable. Perfect foresight assumption is labeled as **Case (B)** in the latter section.

Mixed Expectations

Both the myopic and perfect foresight assumptions are extremes. Given the observed consecutive enhancements in pension rules, few would think that current rules would last forever. Also, no one can perfectly predict all the pension rules in the future. Therefore I consider another two hybrid assumptions between the extremes.

First, a more realistic expectation assumption can be a compromise of the myopic assumption and perfect foresight assumption. In this case, teachers are able to forecast the pension rules clearly for the following one or two years, but have no information beyond that. For simplicity, assume teachers can only forecast one year ahead, then in the previous example, a teacher who made a decision in 1995 can forecast the rules in 1996, R_{1996}. But the teacher cannot predict beyond 1996, and would use R_{1996} for evaluating what they would receive for all the following future years. Thus, I have

$$\mathrm{E}V_{1995}(1996, R_{1996}), \mathrm{E}V_{1995}(1997, R_{1996}), \ldots, \mathrm{E}V_{1995}(2000, R_{1996}) \ldots$$

Also, a teacher who made a decision in 2000 would use R_{2001} to evaluate the expected utilities for all future retirement years m:

$$\mathrm{E}V_{2000}(2001, R_{2001}), \mathrm{E}V_{2000}(2002, R_{2001}), \ldots, \mathrm{E}V_{2000}(2010, R_{2001}) \ldots$$

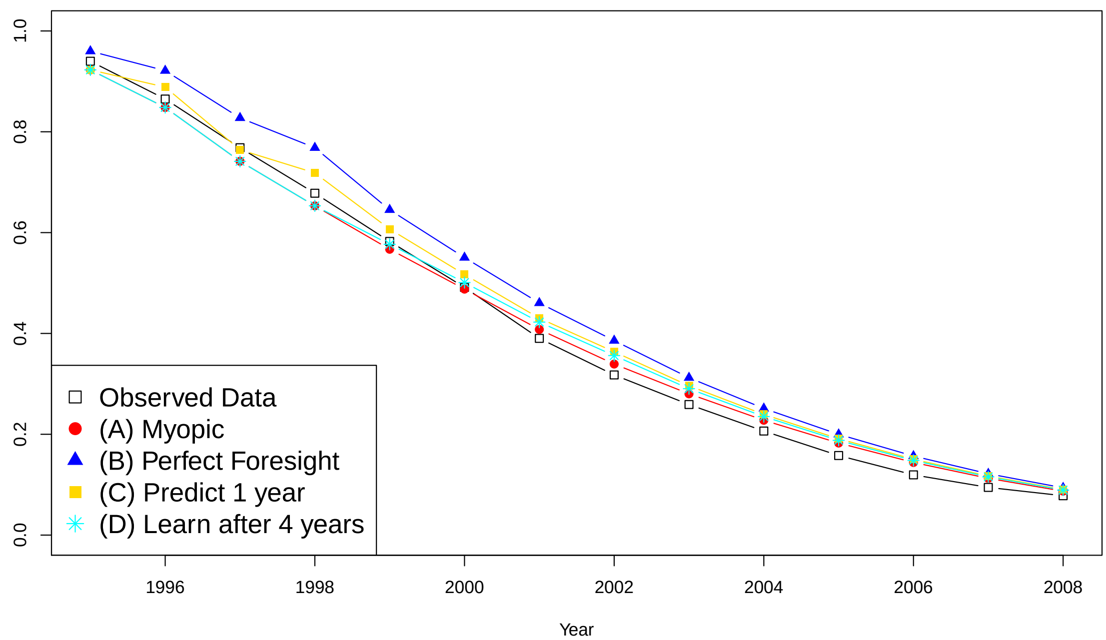
This assumption is labeled as **Case (C)** in the latter section.

Second, teachers can also have adaptive learning in forming expectations on future rule changes. Under this assumption, in the first few years, teachers will still make decisions based on the implemented rule even if the pension rules were enhanced in the past. Because they think enhancements are rare and they don't expect the future pension rules will be further enhanced at that time. However, when teachers see persistent enhancements in all of the first several years, they would start to believe the future rules will be further enhanced, and they start to make their decisions according to their expected future pension rules. With more details, in this case, I assume teachers had the myopic expectation in 1995-1998. Because 1995-1998 are the first 4 years that the pension rules started to enhance, teachers would not expect a further change in the future pension rules at that time, so they made their decisions based on the implemented pension rules in 1996-1998. But by 1998, teachers had seen several consecutive enhancements, and they realized that the rules can be more generous in the future. So, after 1998, I assume teachers can perfectly predict future pension rules. This case is labeled as **Case (D)** in the latter section. Case (D) is a combination of the case (A) and the case (B). In the setting of the previous example, in 1995-1998, teachers have : $\mathrm{E}V_t(m, R_t)$ in year $t \in [1995, 1998]$ for any $(m > t)$. After 1998, teachers have $\mathrm{E}V_t(m, R_m)$ in year $t(t > 1998)$ for any $(m > t)$.

1.5.2 Performance of Different Assumptions on Expectation

Inappropriate assumptions about teachers' expectation of future pension rules is another source of bias in the simulated retirement rate. For example, under the myopic expectation assumption, the teachers expect that the current rules will not change. However, teachers expecting enhancements in the near future may postpone retirement. If that is the case, then the myopic model would over-predict near-term retirement. Under the perfect foresight expectation assumption for future pension rules, the model may under-predict retirement. Therefore, it is necessary to examine the performance of different assumptions and find the most appropriate one. The basic idea of finding a reasonable expectation is to see which one gives a better fit to the observed data. This section compares the performance of Case (A) to Case (D) mentioned previously.

Figure 1.2: **Observed and Predicted Survival Rates** [a]

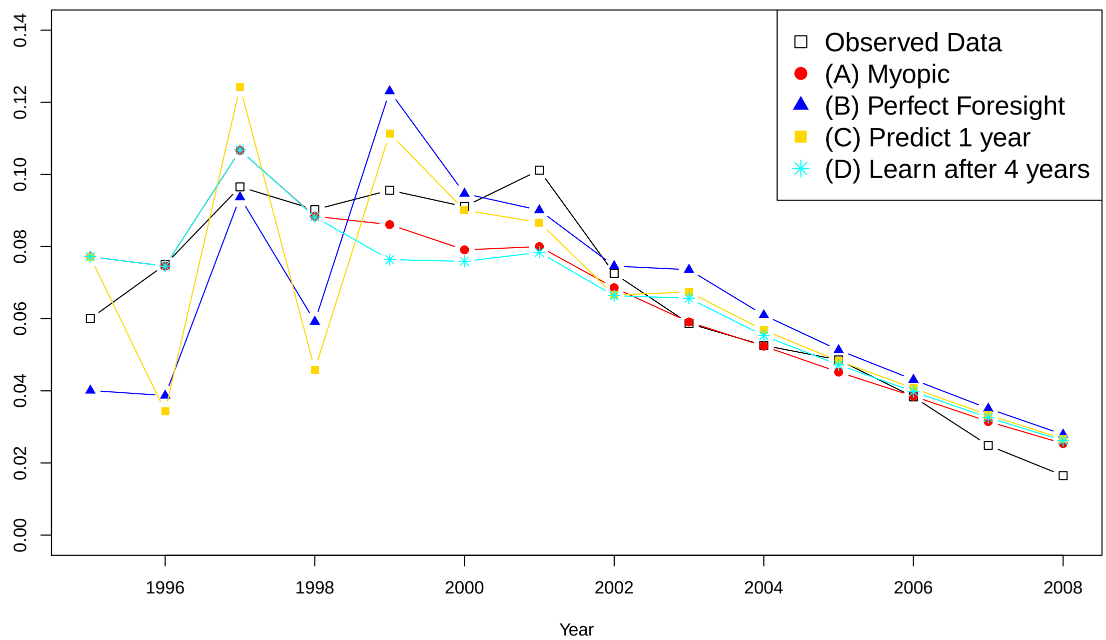


Notes: Survival rates are the percentages of the 1994 cohort teachers who survived in 1995-2008. Predicted survival rates are simulated with different assumptions on expectations as mentioned in the text respectively. Please see Table 1.9 for sample descriptions.

[a]Figures with controls are animated throughout this study.

Figure 1.2 reports the observed survival rates of the 1994 cohort in the PSRS data and the predicted survival rates under different assumptions of expectation, respectively. Overall, the predicted survival rates have the same trend as the observed one. The survival rates generated by perfect foresight (B) are always higher than the observed survival rates. It is not hard to imagine that the expected generous future pension benefits delay teachers' retirement. The survival rates generated by myopic expectation (A) are lower than the observed survival rates before 1999 but higher than the observed rates after the year 2000. The survival rate in (C) swings around the observed survival rate before 1999. Case (D) is a combination of (A) and (B), and its survival rates are identical to Case(A) before 1998, then a little bit higher than Case (A) after 1998.

Figure 1.3: **Observed and Predicted Retirement Distribution over Time**



Notes: The retirement distribution over time shows the attrition rate of the 1994 cohort PSRS teachers in 1995-2008. Predicted ones are simulated with different assumptions on expectations as mentioned in the text respectively. Please see Table 1.9 for sample descriptions.

Figure 1.3 presents the retirement distribution over time for both the observed PSRS data and the predictions in Case (A) to Case (D). The retirement distribution over time shows the attrition rate in every year while the $(1 - survivalrate)$ is the accumulated attrition for all previous years. Overall, the predicted retirement distributions successfully mimic the observed pattern. But differences exist: the simulated retirement rates under myopic expectations (A) match the observed ones quite well for the first four years, then after 1998 they tend to underestimate retirement. Under perfect foresight expectations (B) the model tends to under-predict retirement at the beginning of the sample. This is expected, because if everyone foresaw future enhancements, then fewer teachers would have retired. But in the late 1990's under perfect expectations (B) the model predicted higher retirement than under the myopic expectations(A). A plausible assumption on the formation of expectations is that in the initial years of enhancements, teachers expected the current rules to persist. But after a few enhancements the remaining teachers started to anticipate future enhancements. Case (D) is designed for testing this.

In short, different expectations for the retirement rules make different predictions. Teachers do consider the future pension rules when making their retirement decisions. To make good out-of-sample predictions, the model should consider the change of retirement rules and teachers' expectations over future pension rules.

To quantify the performance of these various expectation assumptions, I calculate three different statistics for the predicted retirement distribution (A)-(D) in respect to the observed distribution. The first statistic is the "Kolmogorov-Smirnov" statistic, which measures the maximum difference of the cumulative annual retirement rate between the model prediction and the observation. Also, I calculate the correlation coefficient and the sum of squared error between the predictions and the observations. All these statistics are reported in Table 1.2.

Among all of the above expectation assumptions, the Myopic (Case (A)), which assumes teachers make retirement decisions based on the currently observed pension rules, makes the best fit to the data, which implies that teachers mainly make their

Table 1.2: **Comparative Statistics for Different Predicted Retirement Distribution over Time**

	(A)	(B)	(C)	(D)
Kolmogorov-Smirnov	**0.0269**	0.0902	0.0459	0.0384
Correlation Coefficient	**0.9386**	0.7944	0.7496	0.9112
Sum of Squared Error	**0.0001**	0.0003	0.0004	0.0001

Notes: Several comparative Statistics measures the the differences between the simulated and the observed retirement distribution over time. The bold font indicates the best performing assumption according to the corresponding statistic.

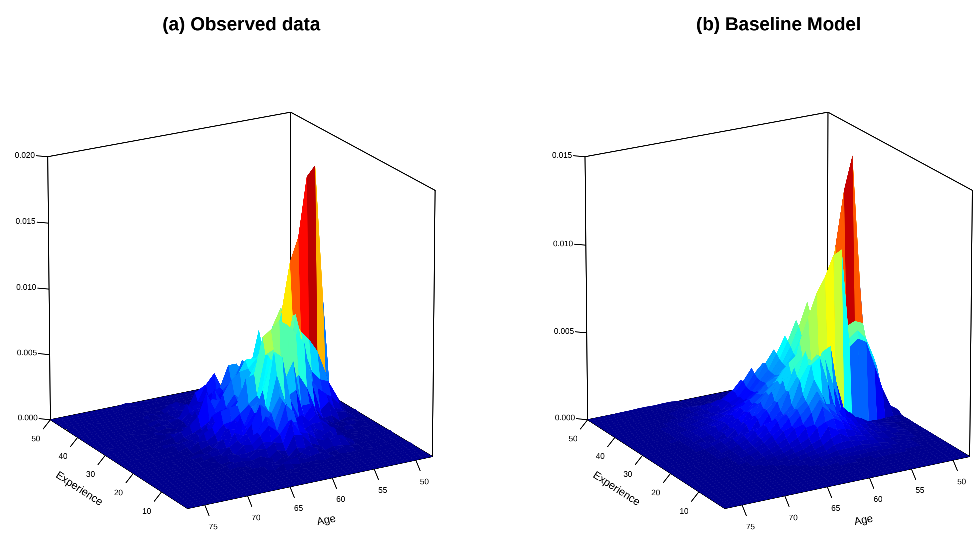
retirement decisions based on the current pension rules that they observed. Of course, there may be some other assumptions concerning expectations that can fit the data better or different teachers may have different types of expectation for the future pension rules. These are reserved for future studies. In this research, based on the statistics in Table 1.2, the myopic expectation (Case (A)), is used in the baseline model for the counter-factual experiments in the following sections.

1.6 What if There Are No Pension Enhancements in Missouri?

As noted in the introduction, a structural model is useful in analyzing the effects of major changes in retirement plans. Many enhancements in PSRS pension rules happened in the late 1990s. As a result, the current patterns of retirement reflect strong, but rather arbitrary, incentives built into pension rules. For example, the "rule of 80" provision creates a ridge of increased retirement probability along the age+experience = 80 line if one plots retirement rates against age and experience. A similar spike in retirements occurs at 25 years of experience. There is no obvious efficiency rationale for these kinks in the intertemporal budget constraint and it is likely that a more rational retirement plan would eliminate them in favor of smoother life-cycle benefit accrual. One interesting question is "What would happen if there were no such changes in the retirement rules in the late 1990s?" Did teachers change their behaviors due to these pension enhancements? Thus it is important to have a model that can yield accurate behavioral predictions in the absence of such enhancements. The option-value model in this chapter can help us to simulate people's behavior under different counter-factual scenarios and answer these questions.

Before the simulation experiments, I examine the overall out-of-sample fit for the 1994 cohort of the Missouri teachers in Figure 1.4.

Figure 1.4: **Joint Distribution of Retirement over Age and Experience**



Notes: The observed and baseline predicted joint distribution of age and experience at the time of retirement in 1994-2008 for Missouri PSRS teachers who were 48-65 and had ten or more years of experience in 1994. Baseline model is the model with Myopic (Case (A)) expectation introduced above. The parameters are MLE based on panel data of 2002-2008 of Missouri PSRS teachers aged 47-59 in 2002.

The observed joint distribution by age and experience at the time of retirement for the Missouri PSRS teachers who were 48-65 and had five or more years of experience in 1994 is plotted in (a) of Figure 1.4, while the model predicted joint distribution of retirement is on the right-hand side. Figure 1.4 shows that the model nicely mimics the joint distribution of age and experience for retirees. We can see the ridge at "rule of 80" in both the observed distribution and the prediction distribution from the model. Overall, the goodness of fit of my model to Missouri data is quite good.

According to the descriptions in Table 1.1 of the enhancements, the dramatic changes happened in 1996 and 2000 by adding the "25 and out" and "Rule of 80" respectively. Therefore, I consider the following three counter-factual scenarios:

(a) **No enhancements at all**. Assume the pension rules did not change at all since 1995, in which both the effect of "25 and out" and the effect of "rule of 80" are eliminated.

(b) **No "25 and out"**. The "25 and out" policy provides large retirement incentives for those teachers who have less than 30 years of experience. I examine the counter-factual scenario without the "25 and out" policy.

(c) **No "Rule of 80"**. The "rule of 80" also provides large incentives for teachers to retire. In the observed joint distribution of retirement over age and experience, there is a ridge at "rule of 80". In this scenario, I consider what would happen if the "Rule of 80" did not exist. Since "Rule of 80" was introduced in 2000, I apply the pension rule of 1999 rule to all the following years after 1999 in this scenario.

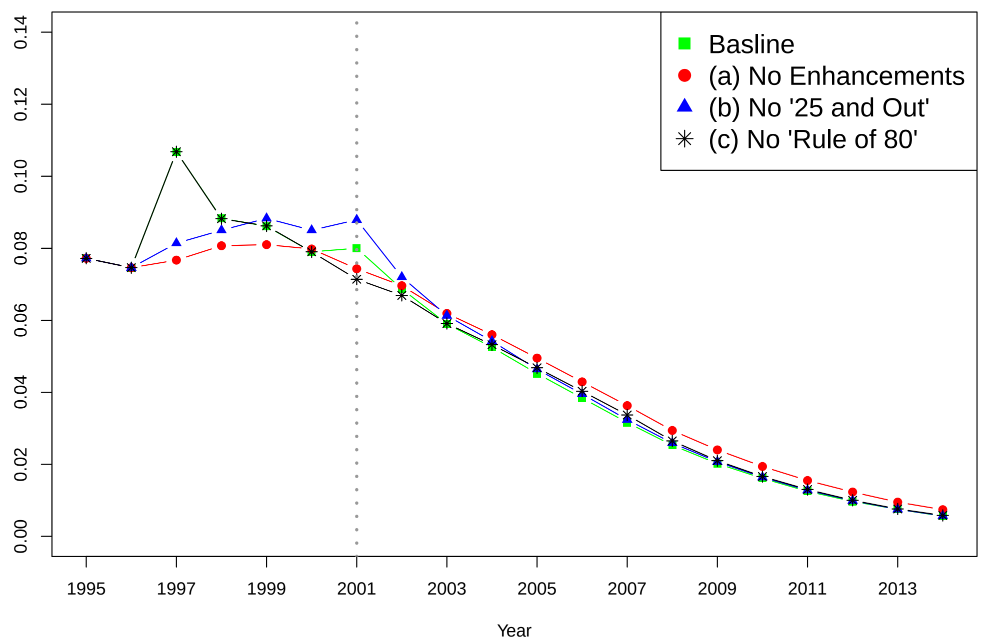
In all of the counter-factual scenarios and the baseline scenario, I run the model for the next 30 years. All teachers in the 1994 cohort would have retired by the end of the 30 years. Besides the 1994 cohort, some younger teachers would be newly included in the retirement window every year. I consider both the effect only on the 1994 cohort and the effect including the subsequent cohorts. Compared to the subsequent cohorts, the 1994 cohort has two defects. First, the 1994 cohort contains a large number of experienced teachers at old age. These teachers are strictly selected and they may have already passed their optimal retirement age under new rules. So, their retirement age and experience are less affected by the new pension rules. Second, the number of teachers in the 1994 cohort is fixed. Those retired teachers are not affected by new rules. Therefore, the time-varying pension rules will have a greater effect on all teachers compared to the 1994 cohort only.

1.6.1 Effects on the 1994 Cohort

Figure 1.5 shows the retirement distribution over time under the baseline model and the three different counter-factual scenarios. The retirement distribution of scenario (a) (without any pension enhancements) is smaller in the first several years, which means the case without enhancements delays teachers' retirements. The retirement distribution of scenario (b) (without "25 and out") deviates from the baseline model after 1996. Scenario (b) reduces the retirements in 1997 and 1998 but increases the retirements in 1999 to 2002 compared to the baseline model. Scenario (c) (without the "rule of 80") deviates from the baseline model in 2001. Since the "rule of 80" was put into effect to the PSRS system in 2000, the retirement rate in 2001 would be driven up by the "rule of 80", which is the same as in the observed data. After 2003, the difference in the retirement distribution is very small, since more than 60% of the teachers in the 1994 cohort had retired by 2003.

Figure 1.6 shows the age and experience distribution of retirees under different model settings. In panel (a) of Figure 1.6, removing the "rule of 80" results in less retirement before 25 years of experience, since some teachers would be able to retire

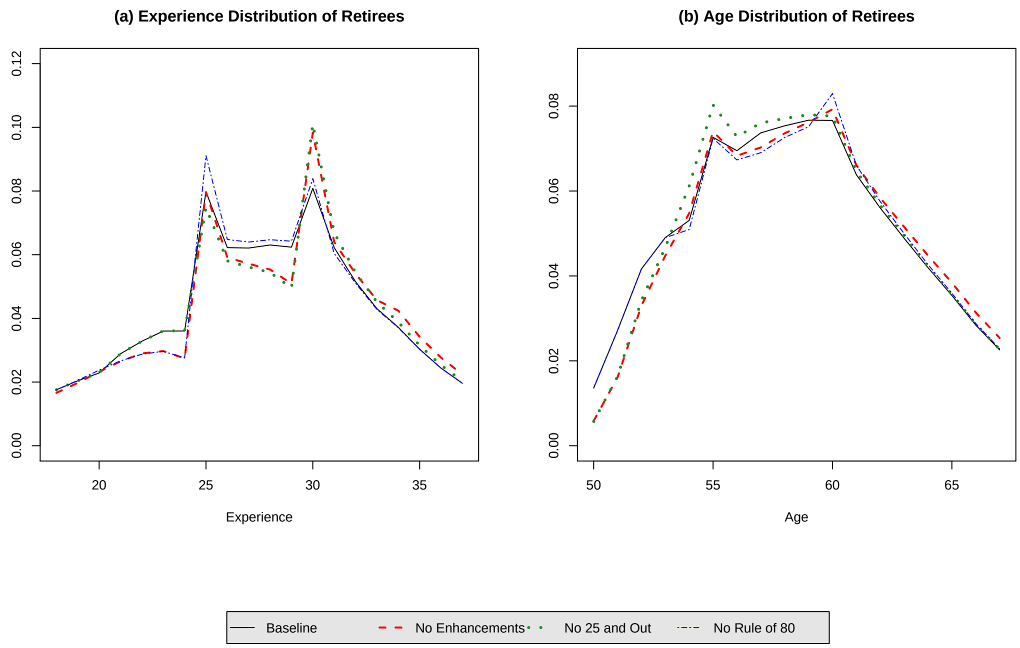
Figure 1.5: **Retirement distributions over Time in Different Models**



Notes: Predicted retirement distributions over from the baseline model and the different counter-factual scenarios as mentioned in the text respectively. Only the first 20 years are shown in the plot.

with old age and less experience previously. Without “25 and out”, fewer teachers would retire with experience fewer than 30 years. If not enhanced at all, both the “symptom” of missing “rule of 80” and missing “25 and out” are shown, the peak at normal retirement experience of 30 years stands out and also more teachers choose to retire at more experience. In panel (b), without “25 and out”, more teachers exit at age 55 since those teachers have more than 25 years of experience and can retire with full retirement benefits at age 55. Meanwhile, without the “rule of 80”, the number of teachers retiring between age 53 and 60 is lower than the baseline model. In short,

Figure 1.6: **Age and Experience distributions of Retirees in Different Models**



Notes: Predicted age and experience distribution of retirees under the baseline model and those three counter-factual scenarios as mentioned in the text respectively.

for teachers who want to retire early, if there were no "rule of 80", they would choose to retire with 25 years of experience; if there were no "25 and out", they would fully use the "rule of 80"; if both are absent, they would retire at 30 years of experience or age 60, which are the normal retirement criteria.

Table 1.3 calculates the predicted averages of age and experience of the 1994 cohort by their time of retirement under different scenarios. The baseline model has the smallest numbers compared to the other counter-factual settings, while scenario (a) gives the highest average retirement age and experience. Even though the difference is not too large, we can tell the enhancements in the late 1990s lead to early retirements. It is not hard to explain it intuitively: when teachers can get more benefits after

Table 1.3: **Average Retirement Age and Experience Predicted, 1994 cohort**

	Average Retirement Age	Average Retirement Experience
Baseline	60.27	28.89
No Enhancements (a)	60.80 (*0.53*)	29.42 (*0.53*)
NO 25 and out (b)	60.42 (*0.15*)	29.04 (*0.15*)
NO Rule of 80 (c)	60.34 (*0.07*)	28.96 (*0.07*)

Notes: The average retirement age and retirement experience of the 1994 cohort teachers under the baseline model and the counter-factual scenarios. Sample size of the 1994 cohort: 11691. The *italic numbers* in the parenthesis in each row are the differences between that scenario and the baseline model.

retirement, or if they can retire at early stages with less "penalty", they choose to retire earlier. One of the explanations of the very small difference between the case (c) and the base model is that: 50% of the 1994 cohort teachers have retired by 2000. When I remove the "rule of 80" after 2000, only the remaining teachers would respond to the counter-factual pension rules. By comparing (b) and (c), the effects of removing the "25 and out" policy is larger than the effect of removing the "rule of 80" to the 1994 cohort. There are two reasons. First, "25 and out" was introduced in 1996, so it must have a bigger effect for the 1994 cohort compared to the "rule of 80" which was introduced in 2000. Second, given "25 and out" in the pension rule, those teachers who did not use "25 and out" for early retirement are more likely to prefer to continue teaching, so the "rule of 80" has less effect on their decisions.

1.6.2 Effects on All Teachers in the Retirement Window

Besides the 1994 cohort, in the next 30 years approximately 40,000 or more teachers would enter the retirement window (age between 48 and 65, and have five or more years of experience). The newly included group in the next 30 years is about three times larger than the 1994 cohort. They will make their retirement decisions based on the enhanced pension rules as well. Therefore, the newly included teachers should not be ignored when evaluating the effects of different policies.

To simplify the calculation, I assume the newly included cohorts are the same in every year (1457 teachers[2]). However, different cohorts are exposed to different pension rules in different years. For example, the cohort that entered the retirement window in 1996 would make their decisions based on the pension rules in 1996, 1997, and so on; whereas the cohort that entered in 2000 would make their decisions according to the pension rules after 2000.

Table 1.4: **Average Retirement Age and Experience Predicted, All Teachers**

	Number of Retirees	Average Retirement Age	Average Retirement Exp
Baseline	41374	57.47	27.93
No Enhancements (a)	39351	58.73 (*1.26*)	28.98(*1.05*)
NO 25 and out (b)	40584	58.07 (*0.60*)	28.39(*0.46*)
NO Rule of 80 (c)	40792	57.72(*0.25*)	28.16 (*0.23*)

Notes: The average retirement age and retirement experience of all teachers in the next 30 years under the baseline model and the counter-factual scenarios, respectively. The *italic numbers* in the parenthesis in each row are the differences between that scenario and the baseline model.

With the newly included cohorts, the difference generated by different policies would be larger. Table 1.4 shows the average age and experience of all teachers at their retirement year. Compare Table 1.4 to Table 1.3, and several patterns appear: 1), the newly included cohorts would retire at a younger age and with less experience. For example, the average retirement age for the 1994 cohort is 60.27 under the baseline model, but the number decreases to 57.47 when the new cohorts are included. This is because most of the teachers in newly included cohorts would make their decisions based on the generous pension rules, generating early retirements. 2), The differences between the counter-factual scenarios and the baseline model are greater in Table 1.4 than the ones only for 1994 cohort in Table 1.3. For example, the average age in scenario (a) is 1.26 years greater than that in the baseline model in Table 1.4, while the corresponding difference is only 0.53 in Table 1.3. The greater gap is mainly due to more teachers being affected by different pension rules when all teachers

[2] 1457 teachers per year, thus 43710 teachers in the next 30 years.

are considered. This is also because the 1994 cohort is more strictly selected as I mentioned in Section 4.1: the 1994 cohort includes more teachers who have passed their eligible retirement age or experience. These older teachers enjoy teaching and are less likely to be affected by generous pension rules. Both phenomenon 1) and 2) conclude that the enhancements in the late 1990s were not effective in retaining experienced teachers; instead, the enhancements provided early retirement incentives.

In Table 1.4, compare scenario (a) with the baseline model: those 39351 teachers could have worked to 58.73 years old without enhancements in the late 1990s. But they retired earlier at age 57.47 in the baseline model with enhancements. The loss is $39351 \times (58.73 - 57.47) = 49582$ years of teaching given by experienced teachers. [3]

[3]This number underestimates the actual loss. Because 57.47 is the average retirement age of 41374 retirees. The 39351 retirees are the early leavers among those 41374 retirees, so the 39351 retirees should have a lower average retirement age in the baseline model, which would result in a larger gap in the average retirement age.

1.7 Retention Bonuses

1.7.1 Bonus for Missouri Teachers

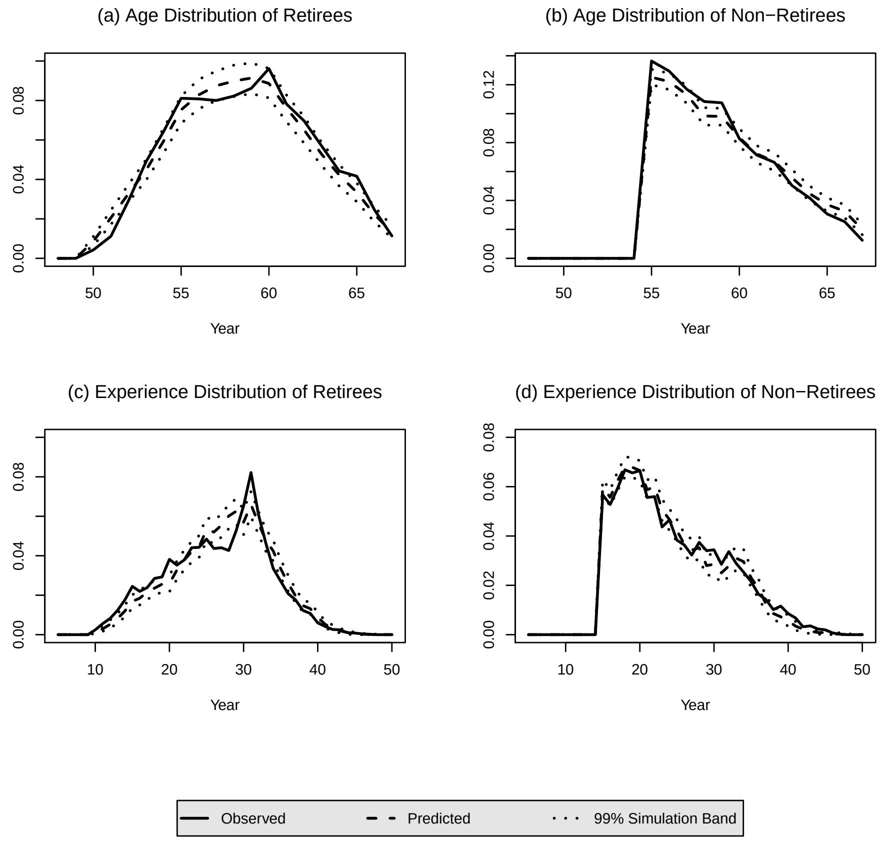
The retirements of Missouri teachers peak at age 60 and 31 years of experience. The relatively low retirement age and experience suggest it is reasonable to lengthen the typical teacher's career. This section designs some bonus policies to retain experienced teachers.

Before considering alternatives to the current plan, I first examine the ability of my model to fit the out-of-sample behavior of 15459 Missouri teachers who were aged 50-65 and had at least ten years of experience in the 2007-2008 school year. I track these teachers forward in time for the next six years until 2014. In this period, roughly 65 percent of the teachers in the sample retired.

Figures 1.7 reports the age and experience distributions of retired and non-retired teachers over this six-year interval (2009-2014). The non-retired teachers are those teachers who had not retired by 2014. The observed distribution are mostly covered by the 99% simulation band, which means the model provides an excellent fit to the age and experience distribution of both of these groups.

I will now consider a variety of bonuses designed to retain MO teachers who are likely to retire. The idea is to neutralize in part the push incentives for late career teachers. The bonuses I consider are paid to teachers who continue working. The bonuses do not enter into the calculation of the retirement annuity. In order to find the most efficient retention policy, the first question is what should be the target of the retention policies? Age, experience or the combination of age and experience? According to Figure 1.7, MO teachers get to the retirement peak at age 60 and 31 years of experience. To aim at delaying the retirement age, the first set of bonus policies give a bonus to teachers who continue teaching at age 60, 61, 62, and 63

Figure 1.7: **Age and Experience Distribution of MO Teachers 2008-2014**



Notes: The observed and predicted distribution of age and experience at the time of retirement for MO teachers aged 50-65 and had 10 or more years of experience in 2008. The parameters are MLE based on panel data of 2002-2008 of Missouri teachers aged 47-59 in 2002. The 99% simulation band describes the sampling errors of the draws of the preference errors for each teacher.

respectively. For retaining more experienced teachers, the second set of bonus policies give a bonus to teachers who continue teaching and have 31, 32, 33, and 34 years of experience respectively. The bonus policies can also target the combination of age and experience. The “rule of 80” in MO PSRS is a large “push” for teachers to retire. So the third set of bonus policies give teachers bonuses if they continue teaching and have age+experience= 80 or 81, 82 or 83, and 84 or 85 respectively. For the convenience of comparing all these policies, the bonus amount is $10,000 for all these policies. The second question is what is the optimal amount of the retention bonus? The higher the amount of the bonus, the more teachers can be retained, but at the same time, more bonus would be wasted on those teachers who would teach even without the bonus. To find the optimal bonus amount, I examine three bonus amount ($1,000, $5,000, and $10,000) that targeted those teachers who have 31 years of experience. In summary, I consider following retention policies:

(1) a $1,000 bonus given to teachers who continue teaching at experience=31.

(2) a $5,000 bonus given to teachers who continue teaching at experience=31.

(3) a $10,000 bonus given to teachers who continue teaching at experience=31.

(4) a $10,000 bonus given to teachers who continue teaching at experience=32.

(5) a $10,000 bonus given to teachers who continue teaching at experience=33.

(6) a $10,000 bonus given to teachers who continue teaching at age=60.

(7) a $10,000 bonus given to teachers who continue teaching at age=61.

(8) a $10,000 bonus given to teachers who continue teaching at age=62.

(9) a $10,000 bonus given to teachers who continue teaching at age+experience=80 or 81.

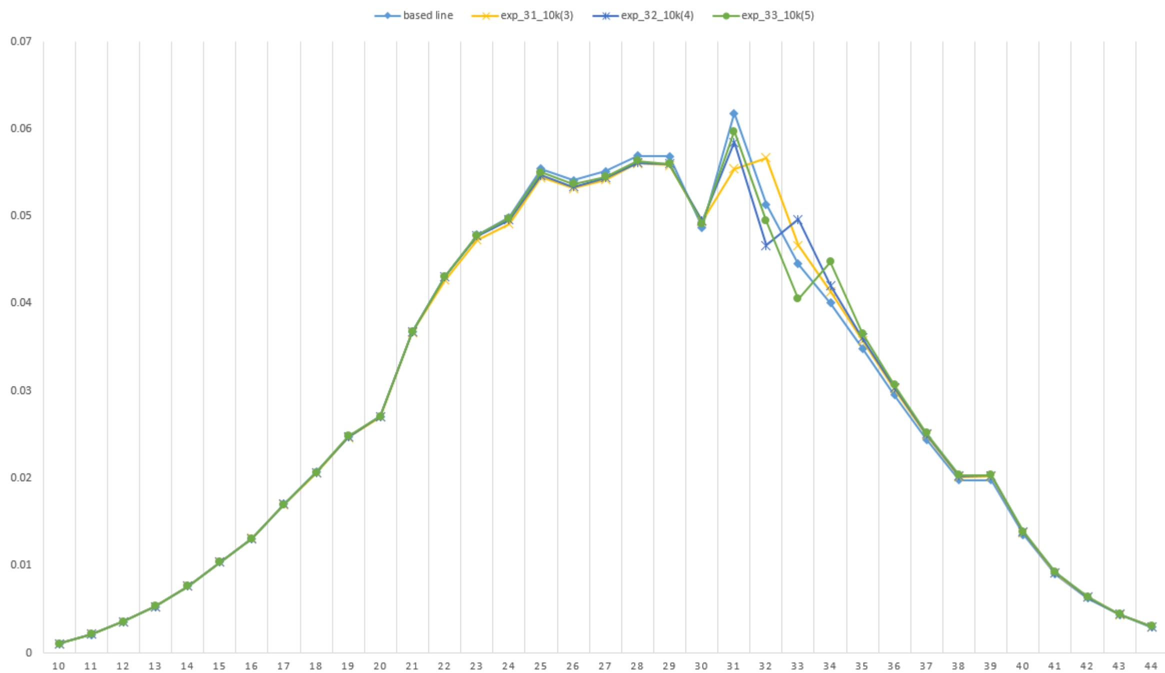
(10) a $10,000 bonus given to teachers who continue teaching at age+experience=82 or 83.

(11) a $10,000 bonus given to teachers who continue teaching at age+experience=84 or 85.

All the bonus policies above are permanent in the sense that all teachers know that they will receive the bonus once they reach the target age or experience.

The retention bonus policies change teachers' retirement age or experience. Figure 1.8 shows the effects of experience-based bonus policies: (3), (4), and (5). Take policy (3) as an example: if a \$10,000 bonus is given to teachers who have experience=31 and continue teaching, then fewer teachers would retire with 31 years of experience and more teachers would retire at the year of experience=32. The same kind of switches in age distributions can be found on age-based policies as well. Also, as the value of the bonus increases, more teachers switch.

Figure 1.8: **Experience Distribution of Retirees under Different Retention Policies**



Notes: The predicted distributions of experience at the time of retirement for Missouri teachers aged 50-65 with ten or more years of experience in 2008 under the baseline model and the model with experience based bonus policies (3), (4), and (5).

Table 1.5: **Effect of Various Retention Bonuses for MO Teachers after One, Five, and Ten Years**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Target:	Exp=31	Exp=31	Exp=31	Exp=32	Exp=33	Age=60	Age=61	Age=62	Rule of 80/81	Rule of 82/83	Rule of 84/85
Bonus Size:	$1,000	$5,000	$10,000	$10,000	$10,000	$10,000	$10,000	$10,000	$10,000	$10,000	$10,000
One Year											
Additional Years	3.17	29.96	68.85	62.24	56.63	88.08	80.71	67.85	48.67	69.39	82.25
Gross Cost/Year	$200,450	$130,811	$122,412	$104,832	$89,989	$128,455	$124,551	$106,610	$273,740	$204,309	$156,082
Elasticity	0.69	1.06	1.13	1.32	1.54	1.08	1.11	1.30	0.51	0.68	0.89
Five Year											
Additional Years	20.73	192.12	462.97	424.97	385.17	640.63	604.09	546.66	344.40	482.91	591.72
Gross Cost/Year	$143,519	$106,260	$99,488	$95,340	$87,977	$106,223	$101,053	$95,198	$189,695	$147,182	$122,685
Elasticity	0.98	1.33	1.42	1.48	1.61	1.33	1.40	1.48	0.74	0.96	1.15
Ten Year											
Additional Years	31.61	343.74	843.67	734.33	644.92	1,221.37	1,212.51	1,150.39	657.88	924.05	1,143.60
Gross Cost/Year	$134,710	$97,007	$91,796	$90,454	$86,561	$106,065	$99,738	$94,527	$167,928	$135,430	$116,425
Elasticity	1.06	1.48	1.56	1.59	1.66	1.35	1.44	1.52	0.85	1.06	1.23

Notes: Simulations from the structural model in the text. The parameters are MLE based on panel data of 2002-2008 of Missouri teachers aged 47-59 in 2002. The data used for the simulations are the 2008-2014 panel of 15459 MO teachers aged 50-65 and had 10 or more years of experience in 2008. The effect of bonus on teaching years and costs is measured by the difference between the quantities in the presence of the bonus and those in the baseline without the bonus. The columns report the effects of different bonus schemes that target different teachers. The rows report the effect of the bonus policies in the next “One Year”, “ Five Years”, and “Ten Years” respectively. In each time block, the first row reports the estimated gain in teacher years resulting from the bonus payment. The second row reports the average total cost (that includes pension, salary, and bonus) per incremental year. The third row reports the estimates of the elasticity of extra years of teaching with respect to total costs. “Rule of 80/81” means give a bonus to teachers if they continue teaching when they have age+exp=80 or 81.

Table 1.5 reports the effects over 10 years of a range of retention bonus schemes. The columns report the effects of different bonus schemes that target different teachers. The rows report the effect of the bonus policies in "One Year", "Five Years", and "Ten Years", respectively. They can be interpreted as the short-term effect and long-term effect. In each timing block, the first row reports the estimated gain in teacher years resulting from the bonus payment. The second row reports the average total cost (that includes pension, salary, and bonus) per incremental year. The third row reports the estimates of elasticity of extra years of teaching with respect to total teaching costs [4]; e.g. in column (1), a \$1,000 bonus paid to teachers who have 31 years of experience would yield 31.61 additional teacher years over ten years at an average cost per year of \$134,710.

Several patterns emerge. First, as time goes, the incremental cost per year decreases and the elasticities go up. Take column (3) as an example. After one year of the implementation of the \$10,000 bonus given to teachers who have 31 years of experience, 68.85 teachers changed their mind and continued teaching, with an average

[4] Elasticity is defined as $:\frac{(\frac{ChangeinTeachingYears}{BaselineTeachingYears})}{(\frac{ChangeinTotalCost}{BaselineTotalCost})}$, where Teaching Years is the number of years taught by this cohort of teachers, and the Total cost includes pension, salary, and bonus that paid to this cohort of teachers.

cost around \$122,000. After applying the age bonus for five years, the teachers in the sample work for 462.97 extra years. The incremental cost decreases to \$99,000. After ten years, the extra working years rise to 843.67, and incremental cost is smaller and the elasticity is greater, which are \$92,000 and 1.56, respectively. In other words, for any of the policies, the long-term effect is greater than the short-term effect. Second, for those policies that target age or experience, there are trade offs between the gain of extra years and the incremental cost. Take the ten-year effect of the age policies as an example, in Columns (6)-(8). The bonus policy targets age 60 generates 1221.37 extra teaching years in ten years, but only 1150.39 teaching years for the policy targets age 62. However, the incremental cost for the age 62 bonus policy is lower than that of the age 60 bonus policy. In terms of elasticity, the age 62 policy is more efficient. Third, for those policies targeted at the sum of age and experiences, Columns (9)-(11), there is no trade-off between the gain of extra years and the incremental cost. Column (11) gives the greatest extra teaching years and smallest incremental cost. Four, comparing Columns (1)-(3), larger bonuses produce more additional years and lower incremental cost. For example, after one year, a \$1,000 bonus at 31 years of experience produces only 3.17 additional teacher years, whereas a \$5,000 bonus yields 29.96 years and a \$10,000 bonus yields 68.85 years. \$10,000 is the most efficient size of the bonus, since Column (3) have more additional years and smaller incremental costs than Column (1) and (2) in both short-term and long-term.

1.7.2 One-Time Versus Permanent Bonuses

All the bonus policies in the previous sections are permanent bonuses in the sense that all teachers know that they will receive the bonus if they reach the targeted experience range or age range in the future. This is in contrast to a temporary bonus that would operate for one year only. Along with the permanent bonuses discussed in the previous sections, I also conducted simulations for a similar range of temporary bonus programs. The effect of temporary programs is considerably smaller than the effect of the permanent bonus policies in both short-term (one year) and long-term (5 years). This is illustrated in Figure 1.9 and Figure 1.10, where I simulate the one-year and five-year effect of a temporary versus a permanent $10,000 bonus for teachers with 31 years of experience.

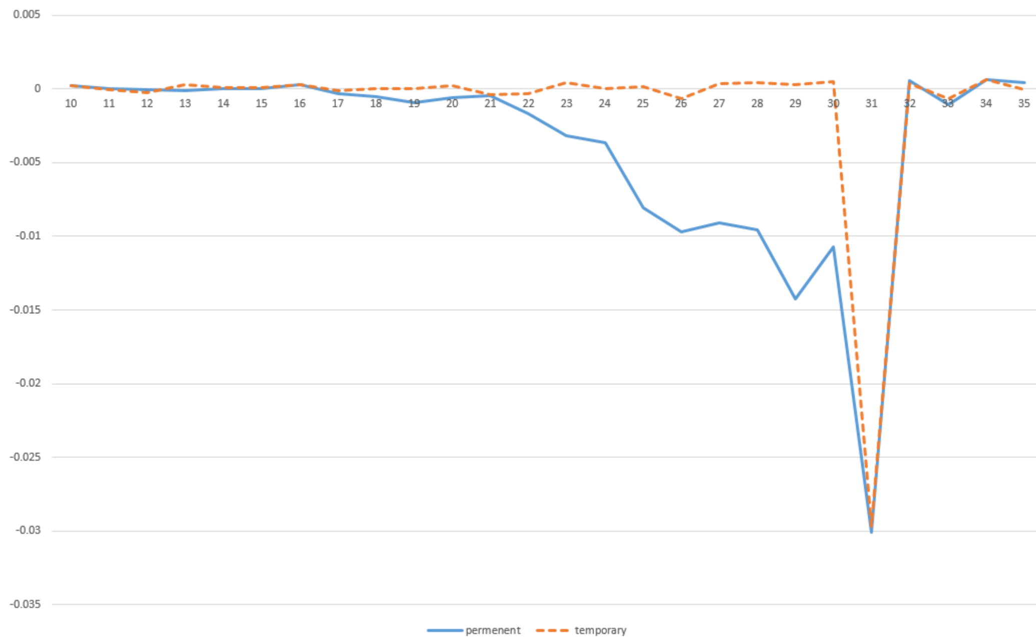
In Figure 1.9, the short-term effect for teachers in the affected cell (exp=31) is identical for both the temporary bonus and the permanent bonus. However, the permanent bonus increases retention for teachers with years of experience less than 31, who anticipate the bonus and alter retirement plans accordingly.

In Figure 1.10, after five years of the bonus in place, the difference between temporary bonus and the permanent bonus gets bigger. After 5 years, most of the teachers are affected by the permanent bonus while only teachers who had 31 years of experience in 2008 are affected by the temporary bonus. Teachers who had 27 years of experience in 2008 are most impacted by the permanent bonus given to exp=31, since they have been exposed to the permanent bonus policy for 5 years, and the fifth year is the year that they have exp=31.

These findings suggest that a temporary experiment or pilot program providing retention bonuses, however well designed from a research point of view, would significantly understate the short run and long run impact of a permanent scheme. The temporary program will also be much less effective in retention and be more costly. Simulation results not included in the tables show that it costs over $275,000 to gain an incremental year of teaching using a temporary $10,000 bonus for teachers with 31 years of experience, in contrast to a cost of about $91,000 in Table 1.5.

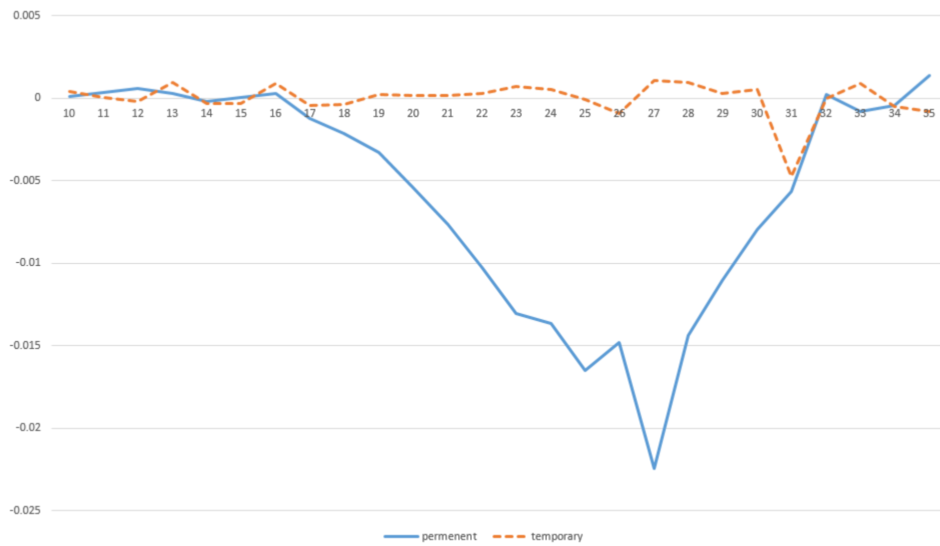
Temporary policies underestimate the effects and overestimate the costs of permanent policies. The inferences from temporary policies would mislead the decisions on permanent policies. However, most social experimental policies usually only last for a few periods because of the limits of money and time. So, those estimated effects from these temporary policy experiments cannot be used to predict the effects of permanent policies.

Figure 1.9: **Change in Retirement Rate in the First Year**



Notes: The difference of the simulated retirement rates comparing the case with permanent/temporary bonus and the baseline case in the first year. The parameters are MLE based on panel data of 2002-2008 of Missouri teachers aged 47-59 in 2002. The data used for simulations are Missouri teachers aged 50-65 with ten or more years of experience in 2008. The permanent \$10,000 bonus will be given to teachers who reach experience of 31 years in any future years. A temporary \$10,000 bonus is only given to teachers who had 31 years of experience in 2008.

Figure 1.10: **Change in Retirement Rate in the First 5 Years**



Notes: The difference of the simulated retirement rates comparing the case with permanent/temporary bonus and the baseline case in the first 5 years. The parameters are MLE based on panel data of 2002-2008 of Missouri teachers aged 47-59 in 2002. The data used for simulations are Missouri teachers aged 50-65 with ten or more years of experience in 2008. The permanent $10,000 bonus will be given to teachers who reach experience of 31 years in any future years. A temporary $10,000 bonus is only given to teachers who had 31 years of experience in 2008.

1.8 Deferred Retirement Option Plans (DROP)

Deferred Retirement Option Plans (DROP) are an alternative way to retain experienced teachers. DROP plans permit employees to retire and begin collecting all or part of their pension annuities and to continue working full-time for a limited period of time. From the date that the teacher enters the program forward she no longer contributes to the pension plan, nor does she accrue additional service. The annuity payments are usually put into an escrow account while the teacher continues to work, and become available with interest when she actually stops working. For example, teachers in Arkansas for many years have had the option of participating in a DROP plan, receiving roughly 70 % of their pension annuity, and continuing full-time teaching for up to ten years.[5] Florida teachers can retire, have their annuity deposited in an escrow account, and continue in full-time employment for up to five years, at the end of which they terminate employment and collect their accumulated annuity payments plus interest.[6] The Louisiana teacher retirement system provides a DROP option for retirement eligible teachers for up to three years. [7]

While several states offer or have offered this option for educators, the program is available to all retirement age teachers. In the absence of an actual targeted DROP, in this section, I simulate the effect of DROP plans for MO STEM teachers. Of course, no teacher is forced to take the DROP plan. She can continue working or she can terminate employment and take the conventional annuity.

Before simulating the effect of a DROP plan for MO STEM teachers, I check the out-of-sample fit of the MO STEM teachers (sample size: 2131; see Table 1.12 for

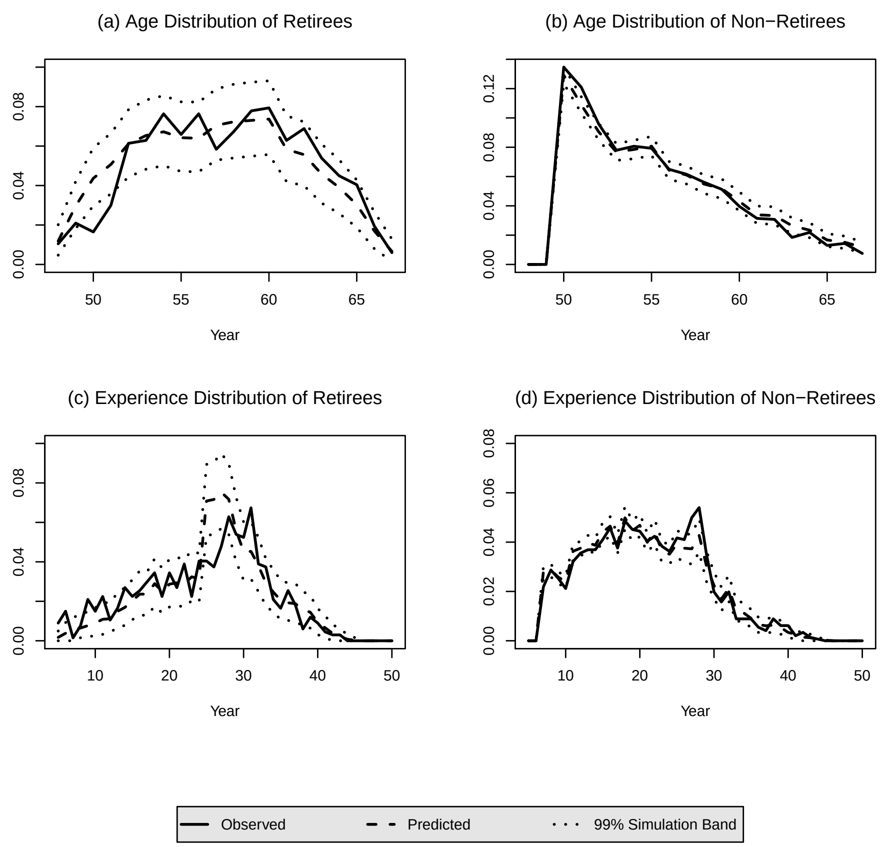
[5] https://www.artrs.gov/members/teacher-deferred-retirement-option-t-drop/how-does-t-drop-work

[6] http://www.dms.myflorida.com /workforce$_o$$perations/retirement/members/deferred_retirement_option_program_drop$

[7] https://www.trsl.org/main/optional_programs/drop

more sample information.) Figure 1.11 reports the age and experience distributions of retired and non-retired teachers during 2012-2014. The non-retired teachers are those teachers who had not retired by 2014. The observed distribution are mainly covered by the 99% simulation band, which means the model provides an excellent fit to the age and experience distribution of all these groups.

Figure 1.11: **Age and Experience Distribution of MO STEM Teachers 2011-2014**



Notes: The observed and predicted distribution of age and experience at the time of retirement for MO teachers aged 50-65 and had 10 or more years of experience in 2008. The parameters are MLE based on panel data of 2002-2008 of Missouri teachers aged 47-59 in 2002. The 99% simulation band describes the sampling errors of the draws of the preference errors for each teacher.

Since the MO STEM teachers have their retirement peak at 31 years of experience, I consider a 2-year DROP plan for MO STEM teachers if they have 31 years of experience. It requires the teachers to retire with 31 years of experience and to collect a fraction of the pension annuity for 2 years, in addition to salary, and to leave the teaching force after 2 years. The teacher needs to commit to working for 2 years regardless of the draws of the preference shocks. Different fractions of pension that they can collect influence the benefit flows in those two years and teacher's decisions; therefore I try three different fractions: 50%, 70%, and 100%.

This proposed DROP plan affects the decisions of those teachers in the 2011 cohort and also the decisions of all teachers in the subsequent cohorts. I discuss the effects respectively.

1.8.1 Effects on the 2011 Cohort

Table 1.6: **Long-Term Effects of DROP Plans, 2011 Cohort**

	Baseline	50%	70%	100%
Total DROP takers	0	94	130	170
Total Retirees at Exp=31	102	115	134	170
Total Teachers Reach Exp=31	439	451	473	515
Total Extra Teacher-years		210	352	647
Average Cost per Teacher Year ($)		39,068	49,330	59,113

Notes: The retirement decisions of MO STEM teachers who were aged 48-65 and had 5 or more years of experience in 2011.

- Total DROP takers: number of teachers who join the DROP plan.
- Total Retirees at Exp=31: number of teachers who retire with 31 years of experience, including the DROP takers and the conventional retirees.
- Total Teachers Reach Exp=31: number of teachers who reach their 31st year of experience, including retirees and non-retirees.
- Total Extra Teacher-years: number of years gained compared to the baseline model.
- Average Cost per Teacher Year ($): cost of per incremental year.

Table 1.6 shows the effect of the DROP plan to the the 2011 cohort of teachers in long term, after 30 years, during which all the 2131 STEM teachers in the 2011 cohort would have retired. Without the DROP plans, there are 439 teachers who reached the year that they have 31 years of experience and make retirement decisions in the baseline scenario, where 102 out of 439 teachers retire with 31 years of experience. Take the case of 70% as an example. When teachers have 31 years of experience, the DROP plan allows teachers to get 70% of their retirement annuity in addition to their salaries if they sign the DROP contract and promise to teach and only teach for the next two years. In this scenario, there are 473 teachers reach the year with 31 years of experience and make retirement decisions with the option of taking the DROP deal. There are 134 teachers decide to retire and 130 out of 134 (94%) choose to take the DROP deal. These 130 teachers work for an extra 260 years in the next 30 years compared to the baseline scenario. Among these 130 teachers, 98 of them ("Pure DROP Takers") originally retired with 31 years of experience using the conventional retirement track, but they join the DROP plan now, while 32 of them ("Pulled DROP Takers") would have retired before they have 31 years of experience, but they are pulled to join the DROP plan. For these 32 teachers, they contribute some more years before they join the DROP plan. These extra years before DROP are known as "Pull" effect. For example, if a teacher would originally retire with 27 years of experience, but now wait 4 years to join the DROP plan, then the "Pull" effect for this teacher is 4 years. And the total "Pull" effect is 92 years for the teachers in the 2011 cohort. If one plots the cumulative distribution of the retirement experience, the difference between the 70% case and the baseline case is 1.6% at exp=30, which means about 1.6% of the teachers are pulled to their 31st year. In total, the 70% DROP plan can get 352 extra years taught by the teachers in the 2011 cohort. Table 1.7 shows the decomposition of these 352 extra years as I described above.

The bigger the fraction of the annuity given out, the bigger the effect. The 50% DROP case, the 70% DROP case, and the 100% DROP case can get 210, 352, 547 ex-

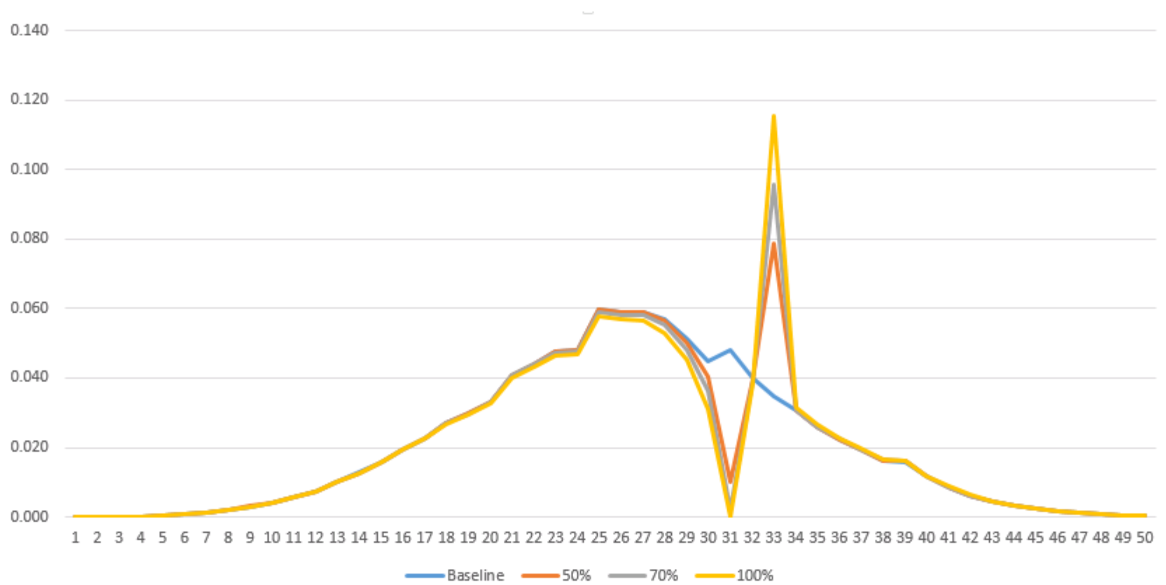
Table 1.7: **Decomposition of the Total DROP Effect, 70% Case**

<table>
<tr><td colspan="3">Total Extra Years=352</td></tr>
<tr><td rowspan="2">Pulled Effect=92</td><td colspan="2">Years from the DROP Takers=260</td></tr>
<tr><td>Years from the Pulled DROP Takers=64</td><td>Years from the Pure DROP Takers=196</td></tr>
</table>

Notes:
Pure DROP Takers: those teachers originally retire at exp=31, but join the DROP plan.
Pulled DROP Takers: those teachers retire before exp=31, but wait to join the DROP plan.
Pulled Effect: the waiting years of the Pulled DROP Takers.

tra years, respectively. The incremental cost of the DROP plans are $39,068; $49,330; $ 59,113, which are lower than the incremental costs of the retention bonus policies in the table 1.5.

Figure 1.12: **Distribution of Exit Experience, MO STEM Teachers 2011-2014**



Notes: The distribution of experience when teachers actually exit the teaching labor force. Those teachers who take the DROP plan are counted as exiting with 33 years of experience.

The effect can be shown by the exit distribution of the 2011 cohort. Figure 1.12 shows the distribution of experience when they really exit the teaching labor force. Those teachers who take the DROP plan are counted as exiting with 33 years of experience. For example, in the 70% case, there are 6.1% teachers who take the

DROP plan, and there are 3.5% teachers who do the conventional retirement at exp=33, so the total exit at 33 years of experience is 6.1%+3.5%=9.6%.

1.8.2 Effects on Multi-cohorts

Besides the 2011 cohort, more teachers would enter the retirement window (age between 48 and 65, and have five or more years of experience). They will make their retirement decisions under the baseline pension rules or with DROP plans. Hence, the newly included teachers should not be ignored when evaluating the effects of the DROP plans. To simplify the calculation, I assume the newly included cohorts are the same in every year (252 teachers). With newly included cohorts, it is impossible to define the long-term effects which requires all teachers to retire; therefore, I report the short-term effects: the 1-year effect, 5-year effect, and 10-year effect.

Multi-cohort short term effect is the summation of the short-term effects on the main 2011 cohort and a newly included cohort in each year. For example, the multi-cohort 5 year effect= the effect on the main 2011 cohort in the first 5 years+ the effect on the first newly included cohort for 4 years+ the effect on the second newly included cohort for 3 years+ the effect on the third newly included cohort for 2 years+ the effect on the fourth newly included cohort for 1 year.

The results are interesting. Take the 50% DROP case as an example. In the first year, there are 10 DROP takers but only 11 extra years, which is less than 20. It is because those 10 DROP takers only complete the first year of the DROP contract, and they will work in the second year as promised. In 5 years, there are 50 DROP takers, but only have 98 extra years. The reason is that, more than 2 of these 50 takers only complete the first year of the contract.

When the newly included teachers are counted, DROP plans have bigger effects.

Table 1.8: **Short Term Effects of DROP Plans, Multi-cohort**

	Baseline	50%	70%	100%
1 year effect:				
Total DROP takers	0	10	10	11
Total Retirees at Exp=31	9	10	10	11
Total Teachers Reach Exp=31	40	40	40	40
Total Extra Teacher-years		11	14	23
5 Year effect:				
Total DROP takers	0	50	58	70
Total Retirees at Exp=31	44	51	58	70
Total Teachers Reach Exp=31	199	205	212	224
Total Extra Teacher-years		98	137	223
10 Year effect:				
Total DROP takers	0	106	128	160
Total Retirees at Exp=31	94	110	128	160
Total Teachers Reach Exp=31	428	442	462	497
Total Extra Teacher-years		218	327	565

Notes: The retirement decisions of MO STEM teachers who were aged 48-65 and had 5 or more years of experience in 2011, in the first 1, 5, and 10 years.

- Total DROP takers: number of teachers who join the DROP plan.
- Total Retirees at Exp=31: number of teachers who retire with 31 years of experience, including the DROP takers and the conventional retirees.
- Total Teachers Reach Exp=31: number of teachers who reach their 31st year of experience, including retirees and non-retirees.
- Total Extra Teacher-years: number of years gained compared to the baseline model.

For example, the 50% plan can gain 218 extra years within 10 years when the newly included cohorts are counted, whereas it would take 30 years to get an extra 210 years if we only consider the 2011 cohort.

In short the DROP plans can retain experienced teachers and the effects size will be bigger as more teachers are involved.

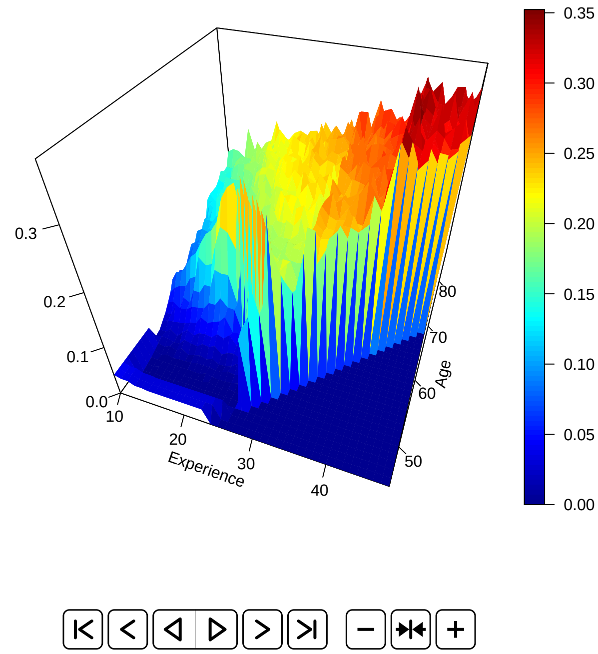
1.9 Simulations: How Pension Rules Affect the Probability of Retirement

So far, all analyses and results are based on the observed data sample and about a certain historical or a potential change in the pension rule. Other elements in the pension rule may also affect teachers' retirement decisions. In order to provide a broader idea of how teachers respond to pension rule changes, I conduct some generalized model-based simulations. With these simulations, I can examine the effect of some other elements in the pension rules, such as contribution rate and replacement rate.

For any pension rule, the retirement probability (hazard rate) at any age and experience is given by condition (1.4) in the model, which is the foundation of this section. The contribution rate, replacement rate, and retirement eligibility are three main factors in characterizing the pension rules. In this section, I keep all elements of the pension rules constant but change one of these three factors at a time, to see the effects of these three factors on the retirement incentives given experience and age. The retirement incentive is measured by hazard rate of retirement.

I use the current pension rule as the baseline model for the comparative analysis. In the baseline model, the contribution rate is 14.5%; the replacement rate (R) varies by experience: R=0.0255 if experience is greater or equal to 31 years, otherwise R=0.025; the pension system allows retirements at the "rule of 80" for full retirement benefits and also allow retirements at "25 and out" with reduced pension payments. First of all, I plot the baseline retirement probability (hazard rate of retirement) over experience and age in Figure 1.13. Overall, the retirement rate increases to 30%-35% as age or experience goes up. The sample selection problem mentioned in section 4 is also fixed by method 2 in this section, which means that for those teachers who

Figure 1.13: **Hazard Rate over Age and Experience, Baseline Model**



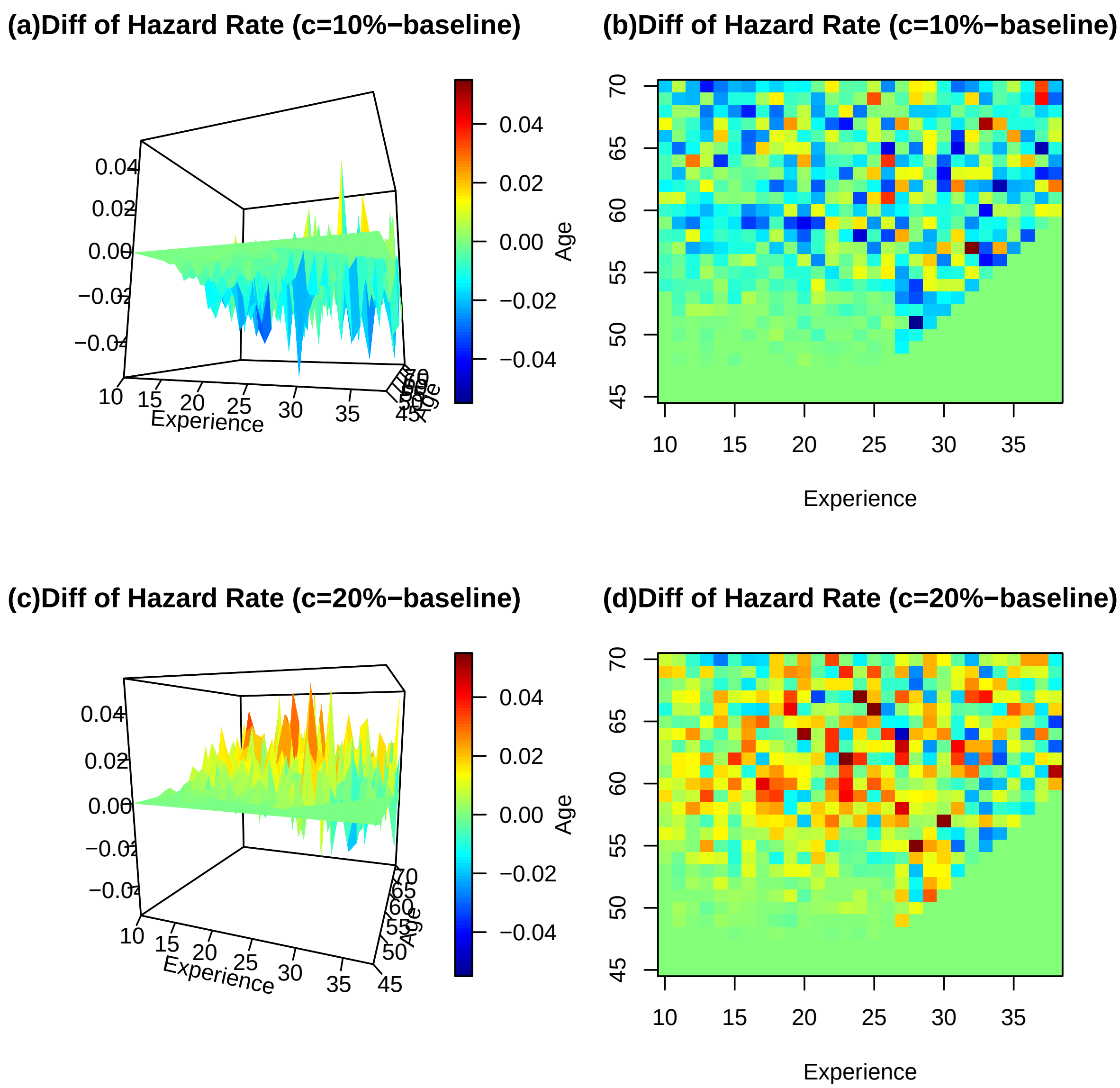
Notes: The simulated hazard rates for each experience and age combination under the current PSRS pension rule (baseline model).

survive to high age and high experience, they have larger preferences of teaching and are less likely to retire compared to other teachers. There are two ridges at "age=60" and "age+experience=80", which are the criteria for normal retirement in the baseline model (current rules).

Effects of Contribution Rate Changes

The baseline contribution rate paid by PSRS teachers is 14.5%. I compare the scenarios when the contribution rate changes to 10% or 20% with the baseline model. Panel (a) of Figure 1.14 is a 3D plot over age and experience. The graph shows the difference between the retirement probability in the scenario with a 10% contribution rate and the retirement probability in the baseline model with a 14.5% contribution rate. With a lower contribution rate, teachers have less incentive to retire, so more

Figure 1.14: **The Effects of Contribution Rate on Hazard Rates**



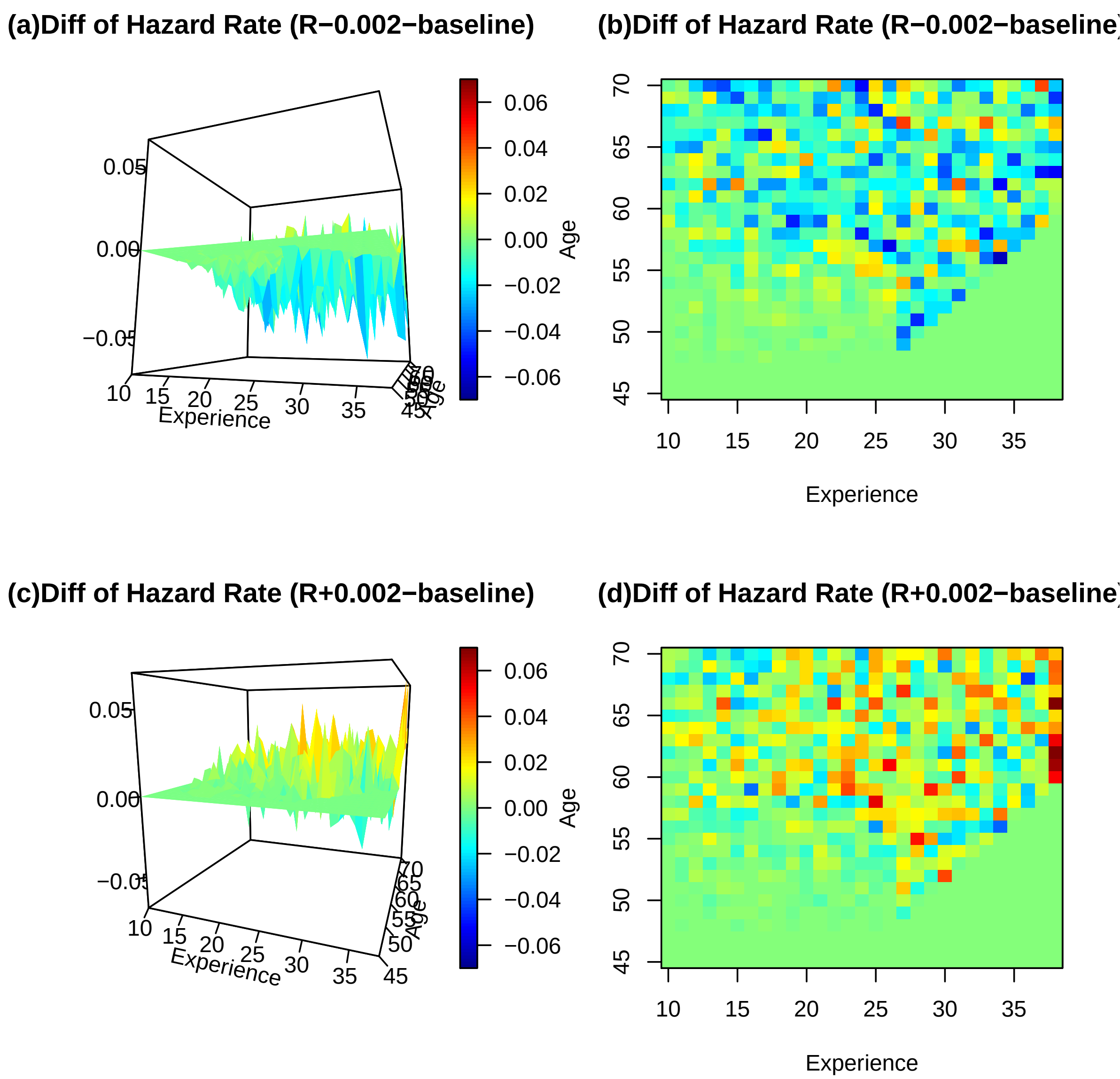
Notes: Panel (a) and (b) present the difference between the retirement probability in the scenario with 10% contribution rate and the retirement probability from the baseline model. Panel (c) and (d) present the difference between the retirement probability in the scenario with 20% contribution rate and the retirement probability from the baseline model. (b) and (d) are top views of (a) and (c) respectively.

of the cells are less than zero. The same information is shown in panel (b) with a top view picture, where colors distinguish the value of numbers. Panel (c) and panel (d) are parallel to panel (a) and panel (b), respectively, but with the contribution rate=20%. Panel (b) gives more negative numbers while panel (d) gives more positive numbers. Therefore the conclusion is that a higher contribution rate leads to higher retirement incentives. It is not hard to explain this, since higher contribution rate reduces the net benefit of working. The "zeros" at "younger" cells (low experience and low age) show that the changes of contribution rates only affect those teachers in the retirement window, but do nothing to those teachers with younger age or less experience, since their exit decision is almost not related to the pension rules.

Effects of Replacement Rate Changes

In the PSRS pension system, the Annual Benefit=S $\times$ FAS $\times$ R, where S is service years (essentially years of experience in the system), FAS is the final average salary calculated as the average of the highest three years of salary, and R is the replacement rate. The current replacement rate varies by experience: R=0.0255 if the experience is greater or equal to 31 years, otherwise R=0.025. In order to measure the effects generated by the replacement rate on the retirement incentives, I assume another two scenarios: 1) all replacement rates decrease by 0.002; 2) all replacement rates increase by 0.002. Figure 1.15 compares the hazard rates given experience and age under the above two scenarios and the current pension rule (baseline model). Panel (a) and (b) show many negative changes in retirement hazard rates when the replacement rates decrease by 0.002, while panel (c) and (d) have many positive numbers when the replacement rates increase by 0.002. Therefore, the conclusion is that higher replacement rates provide more benefits after retirement and leads to greater retirement incentives. Similar to the effects of the contribution rate, the replacement rates does not affect the exit decisions of those teachers with low age or less experience.

Figure 1.15: **The Effects of Contribution Rate on Hazard Rates**



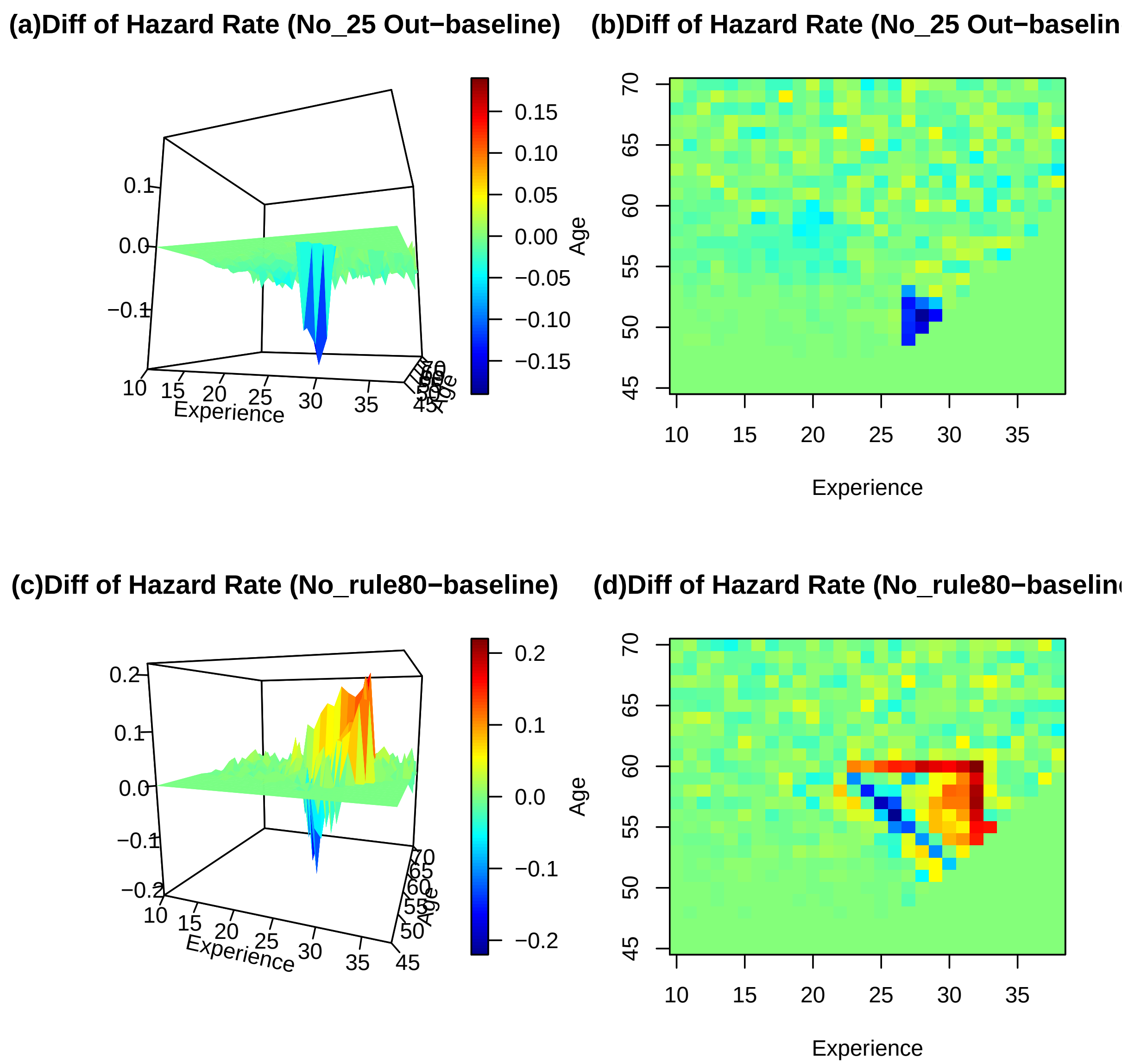
Notes: Panel (a) and (b) present the difference between the retirement probability in the scenario with all replacement rates minus 0.002 and the retirement probability from the baseline model. Panel (c) and (d) present the difference between the retirement probability in the scenario with all replacement rates plus 0.002 and the retirement probability from the baseline model. (b) and (d) are top views of (a) and (c) respectively.

Effects of Retirement Eligibility

The current pension rules allow early retirement with reduced retirement benefits when they have 25 years of experience ("25-and-out"), and also allow normal retirement without penalty if the sum of their age and experience is equal to or greater than 80 ("rule-of-80"). These two conditions provide large incentives for experienced teachers to retire. The significant effects of these two rules to the 1994 cohort have been shown in section 5. In this section, I show the effects of them to the general retirement incentives. Figure 1.16 illustrates the difference between the hazard rates in the scenarios without "25-and-out" or "rule of 80" and the hazard rates in the baseline model. In panel (a) and (b), the hazard rates decrease a lot for those cells that can use "25 and out" to retire. In panel (c) and (d), without the "rule of 80", the retirement probabilities at those cells that satisfied "age+experience=80" decrease a lot, while the retirement probabilities for cells have age=60 or experience $\geq$ 30 increase, since "age=60" and "experience $\geq$ 30" are other normal retirement criteria besides "rule of 80". Therefore, both "25 and out" and the "rule of 80" have significant effects on those teachers in the retirement window.

Overall, contribution rate, replacement rate, and retirement eligibility influence the retirement incentives of the late-career teachers significantly. The enhancements in the late 1990s that changed these factors provided more retirement incentives that would definitely result in earlier retirements.

Figure 1.16: **The Effect of Retirement Criteria on Hazard Rates**



Notes: Panel (a) and (b) present the difference between the retirement probability in the scenario without "25 and out" policy and the retirement probability from the baseline model. Panel (c) and (d) present the difference between the retirement probability in the scenario without "rule of 80" policy and the retirement probability from the baseline model. (b) and (d) are top views of (a) and (c), respectively.

1.10 Conclusion

This chapter applies the model in Stock & Wise (1990) to analyze how observable factors such as age, experience, salary, and pension rules influence the retirement decision. Even though parameters used in our model are taken from the estimates in Ni & Podgursky (2016) based on 2002-2008 Missouri data for teachers aged 47-59, the model produces excellent out-of-sample fit to other observed data in different periods.

There were two main difficulties for the model to fit the out-of-sample data: the sample selection problem and the pension enhancements in the late 1990s of Missouri. I proposed two methods to adjust for the sample selection bias, and modeled teacher's expectations over the time-varying pension rules.

With appropriate adjustments for sample selection bias and appropriate expectation assumptions in the change of pension rules, the model produces a fairly good out-of-sample fit to the observed data. I then seek to answer a number of policy questions. First, "what were the effects of those enhancements in the 1990s to MO teachers?" The simulated counter-factual predictions show that teachers would retire later without or with fewer enhancements of the pension benefits. Second, "what are efficient policies to retain late career public school teachers?" I simulate the effects of two types retention policies: retention bonus and DROP plans. Result shows that teachers can be retained by permanent retention bonuses but not so by temporary bonus. Temporary policies and permanent policies have different effects. Temporary policies tend to be more costly than permanent policies for each extra year of teaching. For permanent policies, the long-term effect is larger than the short-term effect. In terms of the size of the bonuses, $10,000 is more efficient than $1,000 or $5,000. The DROP plans can retain experienced teachers at a lower cost compared to the retention bonuses.

This study provides theoretical guidance for the pension system reforms for states that seek to retain experienced teachers in a cost-effective way. The approach of the study is applicable to the analysis of the effect of time-varying pension rules on retirement decisions and to simulate the effects of potential retention policies.

1.11 Appendix

Table 1.9: **Sample Averages by the Year of Retirement for Missouri**

Sample Year	Number of Teachers	Age	Experience	Male
Base year				
All 1994	11691	53.01	21.62	0.26
Retirement year				
1995	702	58.40	28.43	0.33
1996	877	58.63	28.36	0.32
1997	1129	57.37	28.82	0.32
1998	1055	58.14	28.84	0.32
1999	1118	58.36	28.66	0.28
2000	1065	58.76	28.42	0.28
2001	1183	59.34	28.44	0.25
2002	848	60.00	28.53	0.23
2003	686	60.64	27.82	0.21
2004	615	61.19	28.40	0.19
2005	568	62.04	27.83	0.17
2006	448	62.51	27.91	0.19
2007	291	63.29	28.90	0.19
2008	193	64.47	28.84	0.13
Not Retired by 2008	913	64.99	28.91	0.21

Notes: Missouri teachers aged 48-65 and had 5 and more experience in the 1993-1994 school year. "All 1994" is the total cohort of 11691 teachers in the base year; age and experience are the averages in the base year. The rows with retirement year labels 1995-2008 are contemporaneous averages for teachers who retired in that year. The row for "Not Retired by 2008" are the contemporaneous averages for teachers who remained employed at the end of the sample period. Male=1 for male teachers.

Table 1.10: **Sample Averages by the Year of Retirement for MO Teachers 2008-2013**

Sample Year	Number of Teachers	Age	Experience	Male
Base Year				
All 2008	15459	55.60	22.39	0.18
Retirement year				
2009	1860	57.33	26.13	0.19
2010	1498	58.02	25.31	0.20
2011	1929	58.54	26.69	0.18
2012	1728	59.16	26.75	0.17
2013	1471	59.48	26.48	0.17
2014	1363	60.48	27.08	0.18
Not Retired by 2014	5610	59.22	24.47	0.18

Notes: The Missouri teachers aged 50-65 and had 10 and more experience in the 2007-2008 school year. The rows with retirement year labels 2008-2014 are contemporaneous averages for teachers who retired in that year. The row for "Not Retired by 2014" are the contemporaneous averages for teachers who remained employed at the end of the sample period. Male=1 for male teachers.

Table 1.11: **Sample Averages by the Year of Retirement for MO Teachers 2011-2014**

Sample Year	Number of Teachers	Age	Experience	Male
Base Year				
All 2011	19683	54.74	20.17	0.18
Retirement Year				
2012	2281	58.49	25.19	0.18
2013	2017	58.83	25.28	0.18
2014	2018	59.43	26.31	0.18
Not Retire by 2014	13367	56.69	21.54	0.18

Notes: The Missouri teachers aged 48-65 and had 5 and more years of experience in the 2010-2011 school year. The rows with retirement year labels 2012-2014 are contemporaneous averages for teachers who retired in that year. The row for "Not Retired by 2014" are the contemporaneous averages for teachers who remained employed at the end of the sample period. Male=1 for male teachers.

Table 1.12: **Sample Averages by the Year of Retirement for MO STEM Teachers 2011-2014**

Sample Year	Number of Teachers	Age	Experience	Male
Base Year				
All 2011	2131	54.31	20.30	0.31
Retirement Year				
2012	238	58.27	25.21	0.31
2013	217	58.62	24.80	0.37
2014	213	59.27	26.57	0.35
Not Retire by 2014	1463	56.21	21.82	0.30

Notes: The Missouri STEM teachers aged 48-65 and had 5 and more years of experience in the 2010-2011 school year. The rows with retirement year labels 2012-2014 are contemporaneous averages for teachers who retired in that year. The row for "Not Retired by 2014" are the contemporaneous averages for teachers who remained employed at the end of the sample period. Male=1 for male teachers.

Chapter 2

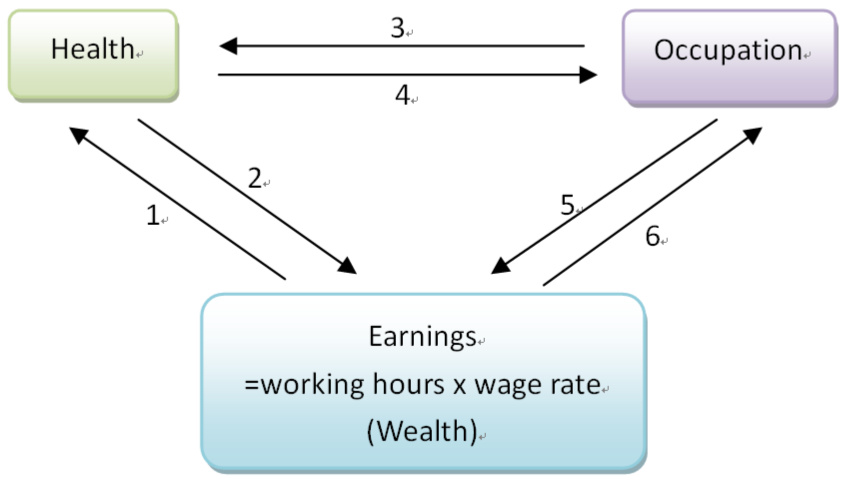
Dynamics of Occupations, Health, and Earnings

2.1 Literature Review

In the last two decades, numerous studies have sought to explain the pair-wise correlations of individuals' health status, occupation, and labor income. It has been widely recognized that health status, occupation, and labor income are related to one another. Most existing studies on micro-level data are based on reduced form regressions and cover a short period of time. There are two limitations. The regressions often suffer from the presence of endogenous regressors and lack of behavioral justification. Consequently, researchers draw different conclusions on the causality of the correlations between health and income. Furthermore in the short run health can affect income, but in the long run income can affect health as well. Thus studies based on the short run data can be misleading by omitting the long run effect. This chapter explores the dynamic relationships among health, occupational choice, and earnings using a life cycle model on a PSID data of 30 years.

The correlations among health, occupation, and labor income reported in the existing empirical studies can be roughly summarized in Figure 2.1. There are bi-directional caustics between any two of the three variables. Empirical literature can be found to support all of these channels corresponding to the numerical labels:

Figure 2.1: **Channels among Health, Occupation, Working Hours, and Earnings**



Channel 1

Frijters *et al.* (2005) use the German Socio-Economics Panel(GSOEP) to estimate the causal effect of income on health satisfaction. Their estimation relies on a large exogenous increase in the real income of East Germany from German Reunification. They find that the increased income leads to a small improvement in health satisfaction. Ettner (1996) analyzes several health indicators, including working and functional limitations, bed days, and alcoholic behaviors, and show that increases in income significantly improve mental and physical health. Adams *et al.* (2003) define causality with conditional probabilities, then use the data from Asset and Health Dynamics among the Oldest Old (AHEAD) to show that Social-economic Status (SES) does not cause a sudden change in health, but does cause a gradual change in health.

Channel 2

Rather than attributing the causality from income to health, some authors suggest the effect is mainly from health to income. Mullahy & Sindelar (1994) regress logarithm of income on health related behavior variables and suggest that for prime age (working age) males, alcohol reduces income significantly. Schultz (2002) uses heights as an instrument variable for health, and shows that there are significant health effects on wage. Bound *et al.* (1999) find the current health status and the trend of health status influence labor market decisions. A rapidly declining trend in health predicts retirements. Oster *et al.* (2013) find that individuals with more limited life expectancy will obtain less education and job training. Health status is also important for non-labor market decisions. Rosen & Wu (2004) show that poor health decreases the ownership of retirement accounts, bonds, and risky assets. Michaud & Van Soest (2008) use a vector auto-regression (VAR) model and the data from the US Health and Retirement Study (HRS) to test the causality between health and wealth. They find no evidence of the causal effect from wealth to health, but strong evidence of the causal effect from health to wealth.

Channel 3

Case & Deaton (2005) conclude from the estimates of an ordered Probit model that manual workers have worse health and rapidly declining health during working years. Carson (2014) reaches the opposite conclusion, showing that farmers and unskilled workers were in better health based on the prison data of the U.S. between the 19th and the 20th century. One possible reason for the conflicting results is that Carson (2014) measures people's health with Basal Metabolic Rate (BMR) while Case & Deaton (2005) measure health with self-assessed health status. At the same time, Ravesteijn *et al.* (2013) say that workplace condition increases the health inequality if workers with low Social-Economic Status (SES) choose a harmful occupation.

Channel 4

Health may influence people's occupational choice. Case & Paxson (2006) state that taller individuals are likely to be selected into high paying occupations and get higher returns. Pagan & Davila (1997) conclude that women with obesity have difficulties in moving across occupations. Case *et al.* (2009) estimate from the data from British Household Panel Survey that each inch of height increases wages by 1.5% via more education and higher-status occupations. Different studies use different measures for health, but the common conclusion is that poor health limits occupational choice.

Channel 5

Many researchers have shown the effects from occupation to earnings. Haller & Spenner (1977) state that income difference increases during a life time. Stevens & Featherman (1981) show that different occupations generate different incomes. Some other researchers argue that different occupations correspond to different labor incomes, e.g. De Beyer & Knight (1989) and Case *et al.* (2009).

Channel 6

On the other hand, earnings or accumulated earnings (wealth) may influence on occupational choice as well. Banerjee & Newman (1993) argue that poor youth work perform wage labor, whereas wealthier young people are more likely to become entrepreneurs or managers. King (1973) notes that individuals with large initial wealth are going to choose occupations with higher risk and which are higher paying. Ravesteijn *et al.* (2013) conclude that initial wealth, education, and health are important factors in the selection of occupation.

Even though the empirical analyses based on reduced form models have explored the relationships among health, occupation, and labor income from all these possible views, they shed little light on the causal effects. Figure 2.1 shows there is a loop that links three items. Only focusing on part of the picture may risk missing variables in the model. In constructing regressions, researchers must decide which variables serve as regressors. As noted earlier, some studies regress health on income and some regress income on health, depending on the assumed exogeneity of the regressors. In some cases researchers make use of instruments when there is no apparent justification for exogeneity. But valid instruments are difficult to come by. In addition, many reduced form models use a linear functional form to describe the relationship between them, which may not be realistic.

Therefore, this chapter builds a structural model in which health, occupation, and earnings are endogenously decided in the model. Using the model, we can explore the dynamic relationships among them.

My theoretical model is built upon the existing theory on the health-SES relationship. A pioneer in health economics, Grossman (1972) introduces health into a utility-maximizing model as a durable good that depreciates over time but can be recovered using medical care. The consumer maximizes the life time utility through the allocation of time and consumption. The consumer's health is endogenously decided. Muurinen (1982) generalizes Grossman's model. She points out that income has three main determinants: wealth, health, and human capital. She argues that rich people use their assets to generate income, and highly educated people use their knowledge to earn income. Meanwhile, poor people who are likely less educated can only earn their income by working in manual labor jobs that are taxing on their health.

This study builds a life-cycle model to explore the dynamic correlations among health, occupation, and earnings observed from PSID data. The simulations from the calibrated model fit the data very well. Based on the good fit of the model, I conduct several simulation-based counter-factual experiments and conclude that the occupations are mainly decided by people's education level. "Manual" occupations leads to higher health deterioration rate than "non-manual" occupations, while "non-manual" occupations generates higher labor income than "manual" occupations. Given an occupation, people's health status affects the labor supplied and labor income. On the other hand, Labor income affect health level via occupational choices, which is the main source for the health-income correlation.

People from wealthy families can get sufficient education. High education make them more likely to find a "non-manual" job, which generates a relative good health level in the rest of their life. Both good health and good occupation give them higher labor income during their lifetime, which reinforces income inequality. On the other hand, those less educated people from poor families are more likely enter the "manual" occupations, and then their health deteriorates faster. With poor health, there is a large cost for them migrating to "non-manual" occupations. Therefore, their health deteriorates faster. Over time poor health and low paying occupation trap them in the lower labor income group. The next section reports the correlations among health, occupation, working hours and labor income in the PSID data in various ways. Section three introduces the life-cycle model and the methodology for solving and calibrating the model. Section four shows how well the model fits data. Section five reports the results of several simulated-based counter-factual experiments. The last section concludes.

2.2 Data

This section describes the individual level data in two ways: by year and by age. Plotting data by year shows the effects of business-cycle or other macro-level shocks. Plotting variables against age illustrate the change over the life cycle.

All data used in this study are from the PSID, the longest running longitudinal household survey in the world. Beginning in 1964, households in the U.S. were surveyed every one or two years. Some households are missed at each survey wave, and some households are newly separated from the original households. On the average, about 5000 households are interviewed in each survey year. In each survey year, the "head" of the household is interviewed for employment, income, wealth, expenditures, health, marriage, childbearing, child development, philanthropy, education, and numerous other topics. The "head" is defined as the main income earner of the household. Some survey questions are about the whole family, some of the questions are about the "head" or other members in the household. In this study, I only focus on the head of the households.

I collated the survey data from 1984 through 2013, since the health level variable responses were first available in 1984. PSID was changed to a bi-year survey after 1997, so I have data in every year before 1997 and in every two years after that.

- **Health:** a self-assessed health status of the "head" of each household. (1=poor, 2=fair, 3=good, 4=very good, 5=excellent.)
- **Initial Health:** a recalled health status before the age 17 of the "head". (1=poor, 2=fair, 3=good, 4=very good, 5=excellent.)
- **Labor Income:** the logarithm of labor income of the "head" of the household in 2010 dollars, also called "earning" in this chapter.
- **Wage Rate:** wage earned per hour by the "head" of the household in 2010 dollars.
- **Hours:** the total number of hours the "head" of the household works in a year.
- **Age:** the age of the household "head" in the survey year.

- **Occupation:** the occupation for the main job of the "head". Before 2001, a 3-digit occupation code was used from the 1970 Census of Population. After 2003 it is coded according to the 2000 Census of Population. The coding rules differ in the two censuses and there are too many categories in the occupation codes from either census. In order to combine the two coding rules and reduce the number of categories, the occupations are grouped into three categories (more grouping details are in the appendix):

    1. Non-manual workers.
    2. Manual workers.
    3. Not working.

- **Education:** the completed education level of the "head". It is divided into 4 groups:

    1. No high school degree (14.69%).
    2. High school degree (51.01%).
    3. College degree (19.01%).
    4. Post college degree (15.29%).

Education levels do not change over time for household "heads". The percentages of people with corresponding education levels are in the parentheses.

2.2.1 Variables' Change over Time

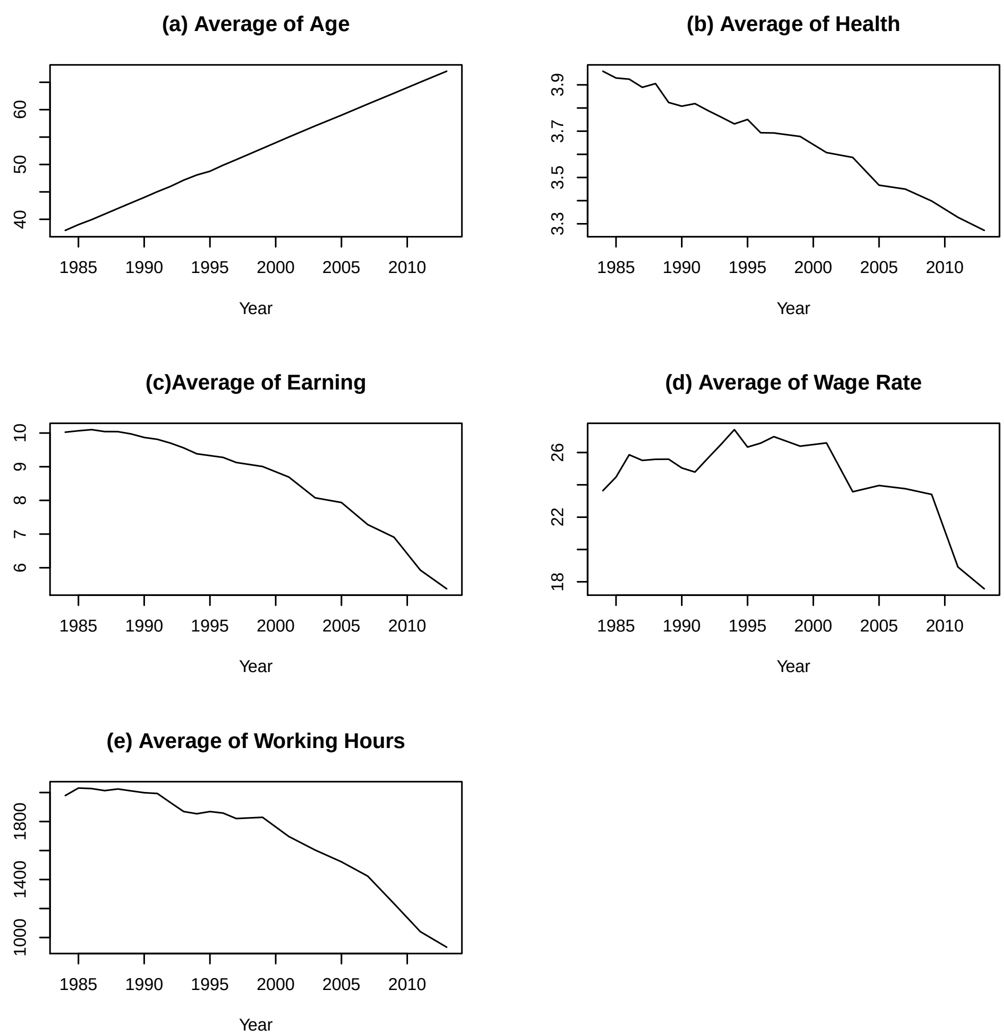
I build a panel data from PSID that contains health, occupation, labor supply, and earnings. This panel data contains the heads of all traceable households from 1984 to 2013 and the heads of the households remain unchanged during these years. So this panel data can describe the dynamic of health and labor income for a fixed group of people over time. There are 994 heads of households in this panel.

The averages of age, health, earning, wage rate, and working hours are reported in Figure 2.2. The health status of this panel deteriorates over time. Labor income and working hours slightly increase for the first several years then decrease. Wage rate drops in 1991 and 2002, possibly due to the business cycle shocks at that time. The final drop of the average wage rate around 2010 is because most of the workers in the panel had exit labor force and reported zero wage rates when they were older than 65.

The bars in Figure 2.3 show the structure of people with different health status and occupations. The percentage of people in the upper panel who reports "Excellent Health" declines over time, while the number of people who report "Poor Health" increases. Overall, the health status of this group is declining, which is the same as what we see in Figure 2.2(b).

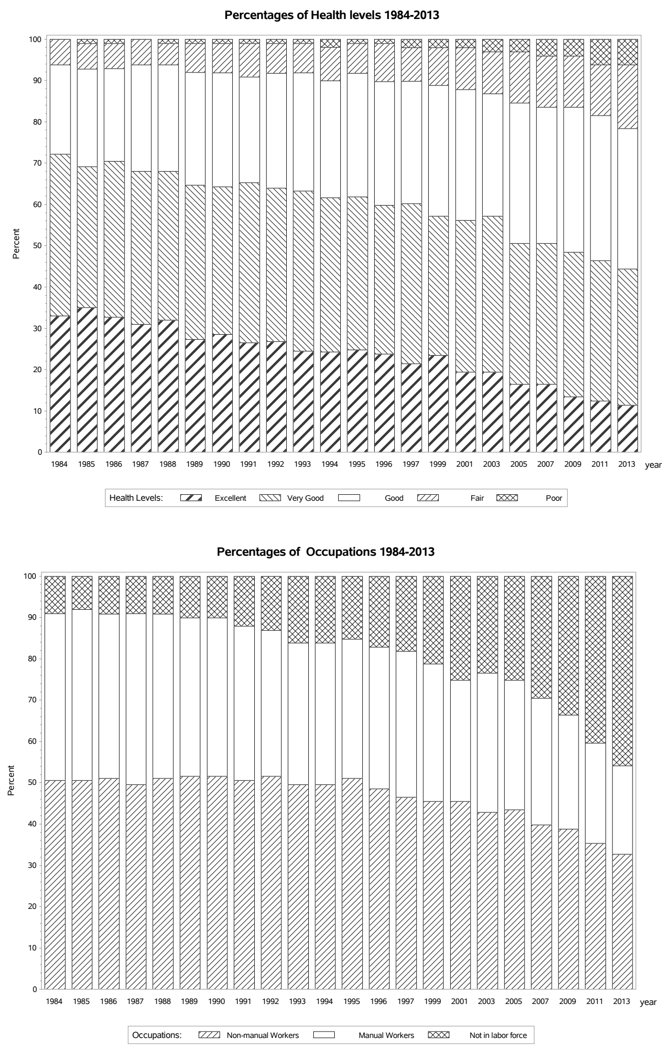
The bars in the lower panel of Figure 2.3 show the percentage of people with different occupations. As people get old, a larger fraction reports "Not working", while the fractions of the other two occupations decline.

Figure 2.2: **Averages of age, health, earning, wage rate, and working hours of the data sample 1984-2013**



Notes: This figure plots the cross sectional averages of age, health status, earning, wage rate, and working hours of the household "head" in each year from 1984-2013. Sample size in each year: 994.

Figure 2.3: **Percentages of Health Levels and Occupations 1984-2013**



Notes: The bars are the percentages of each health status or occupation category of the observed sample in 1984-2013.

2.2.2 Variables' Change over Age

In order to show the dynamic relationship between health and other SES variables during the life cycle, I plot health status, occupations, labor income (earning), wage rate, and working hours against age. In 1984, the first observed year of this panel data, people were in different ages, therefore I can plot variables against age. For example, for those people who were 26 years old in 1984, their labor incomes from 1984 to 2011 are plotted as the labor incomes from age 26 to age 53.

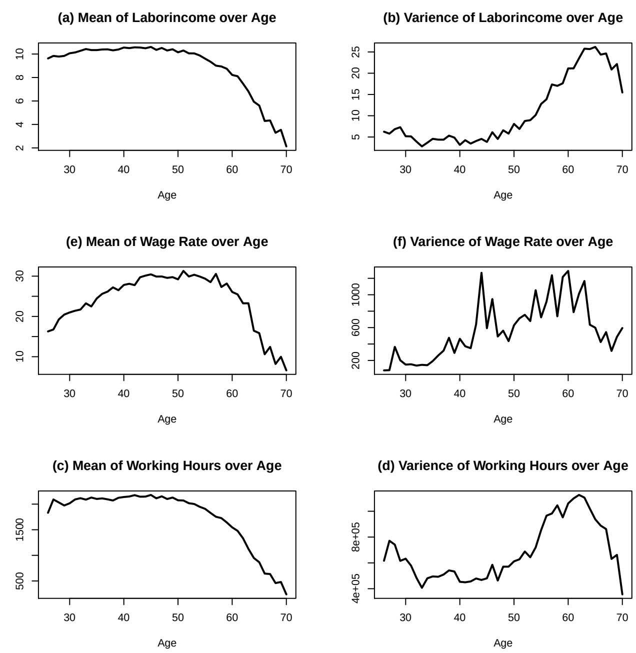
Figure 2.4: **Health and Occupation Percentages over Age**

Notes: The percentages of different health levels and occupations over age.

Health status and occupations are categorical variables. So I report the percentage of each category. Figure 2.4 (a) gives the percentages of each health level over age. The percentage of "Excellent" and "Very Good" health decline over age, but the percentage of "Very Good" health decreases more slowly than that of "Excellent" health since some of the "Excellent" health people go into "Very Good" health at old ages. The percentages of other health statuses have an increasing trend over age. Overall, the health level declines over age.

Figure 2.4 (b) plots the percentages of each occupation over age. The percentage of "Non-manual" workers is bell shaped and peaks between ages 45-50. The percentage of "Not working" increases dramatically at age 60-65 since most of people retire at that age.

Figure 2.5: **Earning, Working Hours, and Wage Rate over Life Cycle**



Notes: The sample average and variance of Log(labor income), wage rate, and working hours over age.

Figure 2.5 plots the means and variances of labor income[1], wage rate and working hours over the life cycle. Labor income gradually increases before age 52 then quickly declines between age 60 to age 70, the common retirement ages. The variance of labor income peaks at age 62-68, since some of the people are still working while

[1]Labor income always means the logarithm of labor income in the study.

some of them have already retired with zero labor income. The mean of wage rate is hump-shaped, similar to the finding of previous studies. The variance of wage rate is hump-shaped as well, but not as smooth. The shapes of the mean and the variances of working hours are almost the same as those of labor income. The peak of the variance of working hours is also between age 58 and age 68, which is the retirement period. The pattern of labor income is mainly decided by the shape of working hours.

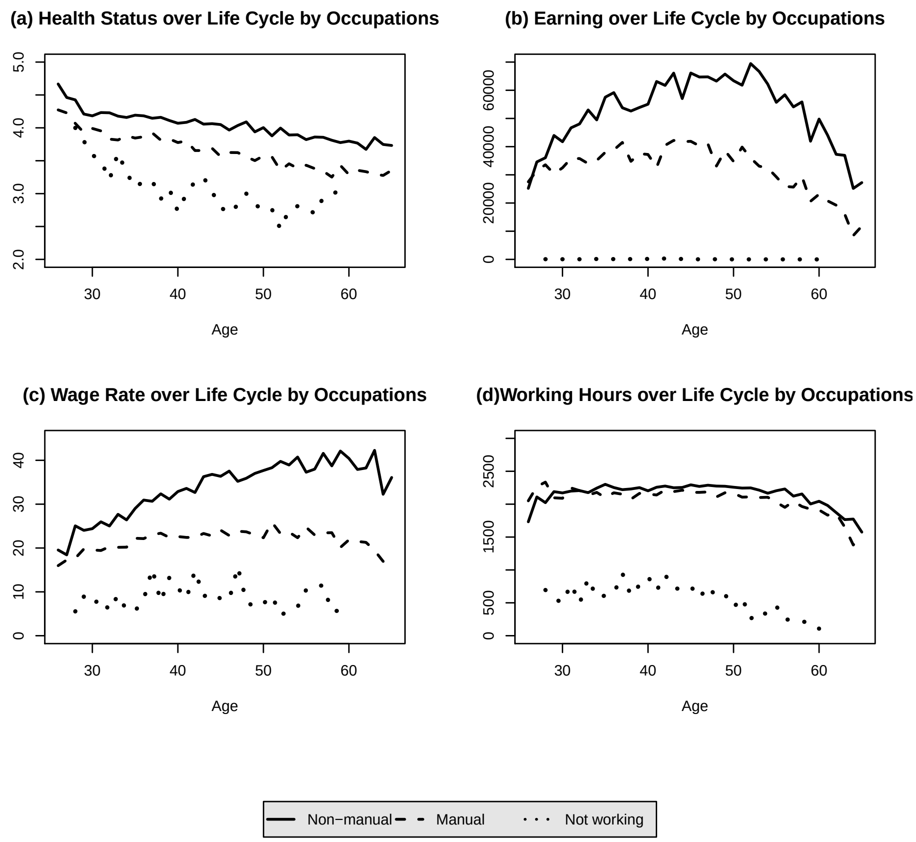
2.2.3 Data Plots by Different Occupations

This section examines the health level, earning, wage rate, and working hours for different occupations.

Figure 2.6 plots the sample average of health status, earning, wage rate, and working hours over the life cycle for different occupations. Non-manual workers have better health level, higher wage rate, more working hours, and higher earning during the whole life cycle compared with other workers in other occupations. At the same time, people reported as "Not working" have worse health level, lower labor income, lower wage rate, and fewer work hours during the whole life cycle. The "not working" group does have positive labor income, possibly because of part-time jobs or inaccurate answers on the question of occupation during the survey.

In Figure 2.6(a), the health level gap between non-manual workers and manual workers widens as age increases, consistent with Case & Deaton (2005). In Figure 2.6(c), the wage rate of the non-manual worker is higher and growing faster than others, while the working hours of non-manual workers are almost the same as that of manual workers in (d). Consequently, the difference of labor income increases over age. Figure 2.6(b) plots the labor income without taking logarithm. In short, occupation is correlated with health and labor income.

Figure 2.6: **Averages of Variables over Life Cycle for Different Occupations**



Notes: The sample average of health status, Labor income, wage rate, and working hours for different occupation groups. Unlike other figures, the earning in (b) is the labor income without taking logarithm.

2.2.4 Data Plots for Different Health Levels

The observations can be either grouped by initial health or by current health. Initial health is the recalled health status of the household “head” before the age 17, which may be viewed as exogenous. Current health is the health status at each period. Throughout the discussion health status refer to the current health status, unless specified otherwise.

Initial Health

Initial health is the health level before the samples enter the labor market, so it is not affected by occupation and future health, but it may affect the future health level, labor income, and occupational choices.

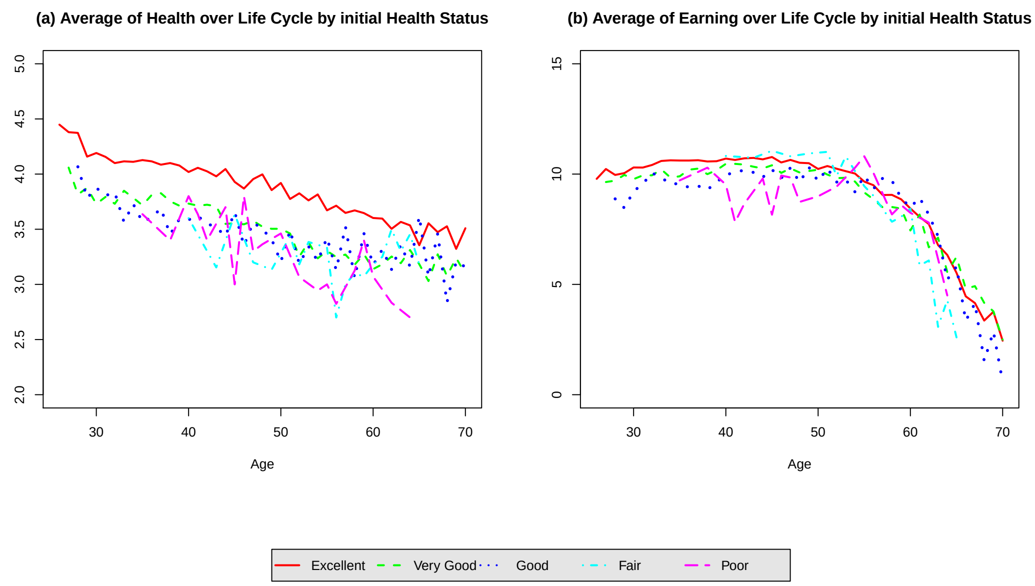
Figure 2.7 plots the average of future health status and earning for different initial health level. For people with “Excellent” initial health, the average health level is better than others during the whole life cycle. However, the average of life time log earning is not clearly distinguishable for those people from different initial health status groups.

The occupational choice varies for different initial health levels. Figure 2.8 shows the percentages of occupations given different initial health status. About 50% of people with “Excellent” initial health work in non-manual occupations. However, less than 30% of people with “Poor” initial health work in the non-manual occupations. The percentage of “Not working” decreases while the initial health status gets better.

Current Health

Now I break the labor income profile down by different contemporary health status, from poor health ($H = 1$) to excellent health ($H = 5$). Figure 2.9 plots the mean

Figure 2.7: **Average of Health and Earning over Life Cycle by Initial Health Status**

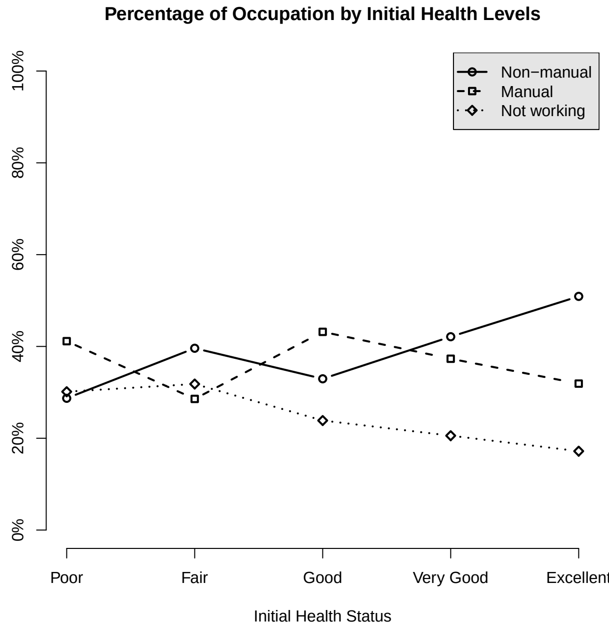


Notes: The sample average of their current health status and Log(labor income) for different initial health status group.

and variance of labor income for different health statuses. Obviously, people with "Excellent" health or "Very Good" health have higher life time labor income than those people with "Poor" health. At the first half of the life cycle, some of the poor health people are working and some of them are not, so the variances at that period are quite high. But after age 60, almost all "Poor" health people have low labor income, which explains the small variance. The opposite pattern of variance is observed for those people with "Excellent" health or "Very Good" health. This is because, all of them are working at their young ages and some of them are still working while others retire at their old ages.

I also break down the wage rate profile and the working hours profile by contem-

Figure 2.8: **Percentage of Occupation by Initial Health Status**

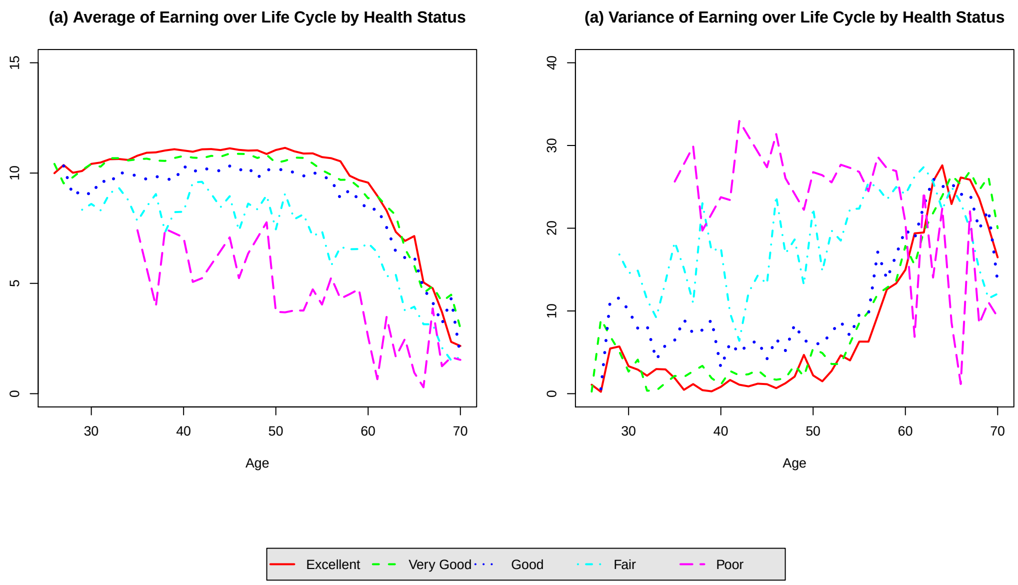


Notes: The observed data is grouped by initial health status. For each different initial health status group, I report the percentage of people in different occupations. For each initial health group, the sum of the percentages of different occupations is 1.

porary health status in Figure 2.10, which shows that better health status correlates with higher wage rates and longer working hours during the whole life time. A comparison of Figure 2.10 and Figure 2.9(a) suggests that the pattern of labor income mainly decided by the pattern of working hours.

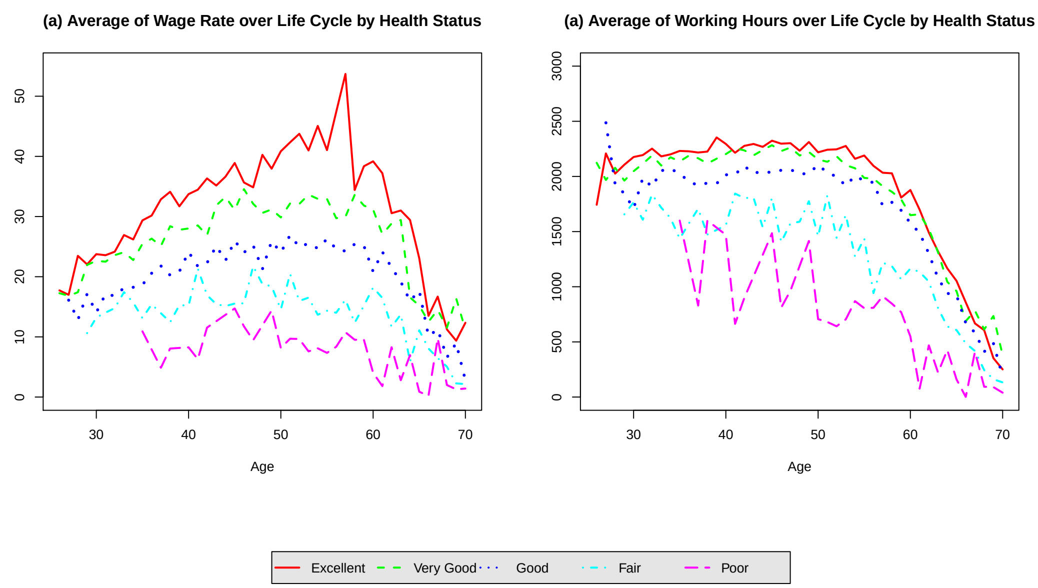
In short, from the PSID data, there are strong correlations among occupation, health, working hours, and labor income. But it is hard to determine the causality. This study proposes a model to endogenize all these variables and model the correlations.

Figure 2.9: **Labor Income for Different Health Status**



Notes: The sample average and variance of Log(labor income) for people with different health status.

Figure 2.10: **Wage and Working Hours for Different Health Status**



Notes: The sample average of wage rate and working hours for people with different health status.

2.3 A structural model on health, occupational choice, labor supply, and earnings

The number of households in the economy is normalized to a unit mass. The household maximizes its life time utility. In each period, there is an instantaneous utility function $u(c_t, h_t, l_t)$ where t indexes age: $t = age - 25 \in (1, 60)$ under the assumption that people start to work at the age of 26 and no one works after 85. c_t is consumption and h_t is the health status of the household. l_t is the labor supply in terms of hours. $L - l_t$ is the leisure time. I assume $L = 6000$ hours, which is the total usable time of an average person in a year.

The utility function is set as

$$
\begin{aligned}
u(c, h, l) &= u(c^*) = \alpha(h_t)\tfrac{c*^{1-\rho}-1}{1-\rho} \\
c^* &= c^{\kappa_1}(L-l)^{\kappa_2}, \\
\alpha(h_t) = \alpha_i & \quad \text{if } h_t = i \ and\ \alpha_1 = 1.0
\end{aligned}
$$

where $\rho = 0.9$, $\kappa_1 = \frac{1}{3}$, and $\kappa_2 = \frac{2}{3}$ given by Prescott (1986). [2]

Another important factor in this model is the occupation. Occupation (o_t) has $N_o = 3$ states, the same as the occupation groups defined for PSID.

1. $o_t = 1$: Non-manual workers.
2. $o_t = 2$: Manual workers.
3. $o_t = 3$: Not working.

In the first period ($t = 1$), the household has an endowed occupation, o_1. In the following periods, the household makes the occupational decision according to the

[2] Prescott (1986) suggests ρ to be a number close to 1. I set $\rho = 0.9$ and also tried several other values, but there is no significant differences on the final results.

preference for the current job and the cost of changing to a new job. The cost of changing occupation is given by a cost matrix:

$$COST(e) = \begin{pmatrix} .00 & cost_{12}(e) & .00 \\ cost_{21}(e) & .00 & .00 \\ cost_{31}(e) & cost_{32}(e) & .00 \end{pmatrix} \tag{2.1}$$

In the cost matrix, $cost_{o_t o_{t+1}}$ is the cost incurred in utility when occupation change from o_t to o_{t+1}. Usually, to continue working at a job or change to "Not working" only have a negligible cost, so the diagonal and the third column of the cost matrix are zeros. But if the household want to jump between a non-manual job and a manual job or jump from "Not working" to any other jobs, there will be a certain amount of cost in learning new skills. The cost matrix of changing occupation varies by the household's education levels. Their education does not change over time and is fixed at four levels $N_e = 4$:

1. $e = 1$: no high school degrees.
2. $e = 2$: have high school degree but no college degrees.
3. $e = 3$: college degree.
4. $e = 4$: post college degree.

Health status h_t takes $N_h = 5$ possible states (1=poor health, 2=fair health, 3=good health, 4=every good health, 5=excellent health). The evolution of health is decided by a transition matrix, where

$$prob(h_t = k_1' | h_{t-1} = k_1) = p(k_1', k_1, t, o_{t-1}) \tag{2.2}$$

The transition probability depends on age. At different phases of life, people have different health transition matrix. The transition probability also depends on the occupation in the previous period (o_{t-1}). Many studies emphasize the effect of occupation on health, e.g. Case & Deaton (2005). This study estimates the occupation-time depending on health transition matrices non-parametrically from PSID data, by grouping observations into different occupation-time groups. The whole life time (T periods) are divided into three phrases: young ($t \in (1, 15)$), mid-age ($t \in (16, 25)$), and old ($t \in (26, T)$. In each occupation-time group, the 5×5 transition matrix is calculated.

The wage rate (hourly pay) of household (w_t) depends on the health status h_t, age t (a proxy for experience), occupation o_t, education level e and a normal shock. The functional form of the wage rate is estimated by data from PSID and it is given as:

$$
\begin{aligned}
log(w_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_{3i} \sum_{i=1}^{4} 1\{H_t = i\} + \beta_{4i} \sum_{i=1}^{2} 1\{o_t = i\} \\
+\beta_{5i} \sum_{i=1}^{2} t \cdot 1\{o_t = i\} + \beta_{6i} \sum_{i=1}^{2} t^2 \cdot 1\{o_t = i\} + \beta_{7i} \sum_{i=1}^{3} 1\{e = i\} + \epsilon \\
\epsilon \sim N(0, 0.6^2)
\end{aligned}
\tag{2.3}
$$

The "β"s in the wage function are estimated with PSID data. [3]

Given the wage rate, once I know the labor supply l_t, then the total labor income of the household is $w_t l_t$.

At old age, with worse health, the labor income of the household head is low relative to the retirement benefits, such as Social Security income and pension benefits. Therefore, at this point the household head would retire ($l_t = 0$ and $o_t = 3$). The retirement benefit (SS) of the household depends on the age of household head t, and the employment status l_t and o_t. Since people start to retire after age 50 ($t = 50 - 25 = 25$) and they can receive the full social security payment after the age

[3]Notes: more details and β estimates are in Appendix 2.7.2.

of 67 ($t = 67 - 25 = 42$), I set $SS(t, l_t) = (\frac{t}{42})^2 \times 40000$ if $l_t = 0$ and $o_t = 3$, when $t > 25$. If they are unemployed ($l_t = 0$ and $o_t = 3$) in early ages $t \leq 25$, I assume there is an unemployment insurance payment to them, $UI(t, l_t, o_t) = \frac{1}{2}exp(11.0)$, which is 50% of the average labor income.

The budget constraint for the household is:

$$c_t = w(t, h_t, O_t)l_t + SS(t, l_t, o_t) + UI(t, l_t, o_t) \tag{2.4}$$

If the "head" works, $c_t = w(t, h_t, O_t)l_t$. By maximizing the utilities, the endogenous labor supply is $l_t = \kappa_1 \times L = 2000$ hours. If the "head" doesn't work, the $c_t = SS(t, l_t, o_t)$ if $t > 25$, while $c_t = UI(t, l_t, o_t)$ if $t \leq 25$.

Death rate $d(t, h)$ is a function of age and health. [4] Therefore, survival probability is also a function of health at different ages:

$$\beta_t(h_t) = 1 - d(t, h).$$

The household chooses $c_t > 0, o_{t+1} \in (1, .., N_o), l_t \in (0, \kappa_1 L)$ that solves:

$$\begin{aligned} V_t(h_t = k_1, o_t = k_2, e) = \quad & max \left\{ u(c_t, h_t, l_t) - cost_{o_t o_{t+1}}(e) \right. \\ & \left. + \beta_{t+1} E_t V_{t+1}(h_{t+1}, o_{t+1}, e) \right\} \end{aligned} \tag{2.5}$$

[4]See Appendix 2.7.3 for more details.

where the conditional expectation

$$\begin{aligned} & E_t V_{t+1}(h_{t+1}, o_{t+1}, e) \\ & = \sum_{k_1'=1}^{N_h} p_{h_{t+1}|o_t} V_{t+1}(h_{t+1} = k_1', o_{t+1}, e) \end{aligned} \tag{2.6}$$

with constraints (2.1), (2.2), (2.3), and (2.4).

The terminal period T health is given by (2.2) with $t = T$.

The household chooses $c_T \geq 0, l_T \in (0, \kappa_1 L), o_{T+1} = o_T$, and value function is given by solving

$$V_T(h_T, o_T, e) = max\, u(c_T, H_T, l_T)$$

with constraints (2.2), (2.3), and (2.4).

Overall, in the model, households wake up with given H_t, o_t, and e in each period t. They make decisions on the labor supply l_t and occupational choice for next period o_{t+1} to maximize the expected life time utility. The labor supply generates the consumption in period t, while o_{t+1} affects the health level in period $t+2$.

2.3.1 Computation

The solution of this life-cycle model is a series of policy functions, from $t = 1$ to T, respect to state variables: (H_t, O_t, e). In each period, the solution contains a consumption function $(c_t(H_t, O_t, e))$, a labor supply function $(l_t(H_t, O_t, e))$ and a function of the occupational choice for the next period $(O_{t+1}(H_t, O_t, e))$. There are 5 statuses for health H_t, 3 categories for occupation O_t, and 4 levels of education e respectively. The choice set is $\{0, \kappa_1 L\}$ for labor supply.

The life-cycle model with terminal period T is a finite dynamic optimization problem. So, the model is solved backward from the last period $t = T$, then solve the problem at $t = T - 1$ to $t = 1$ period by period.

2.3.2 Calibration

I calibrate the health parameters in the utility function and the parameters in the cost matrix of changing occupation. In the utility function, the parameters for the health status are $\alpha(h_t) = \alpha_i$ if $h_t = i$. Intuitively, the healthier the people, the higher utility they have, which means $\alpha_1 \leq \alpha_2 \leq \alpha_3 \leq \alpha_4 \leq \alpha_5$. According to Figure 2.10 and Figure 2.9, the income gap between "Poor" Health people and "Fair" health people is bigger than the gaps between the other two levels. So I set: $\alpha_1 = 1$,$\alpha_2 = 5$,$\alpha_3 = 7$,$\alpha_4 = 9$,$\alpha_5 = 11$.

The cost matrix of changing occupation varies by education levels. Intuitively, with higher education, the costs of changing occupations are lower than others. See Table 2.1 for numbers in detail.

Table 2.1: **Cost Matrix of Changing Occupations**

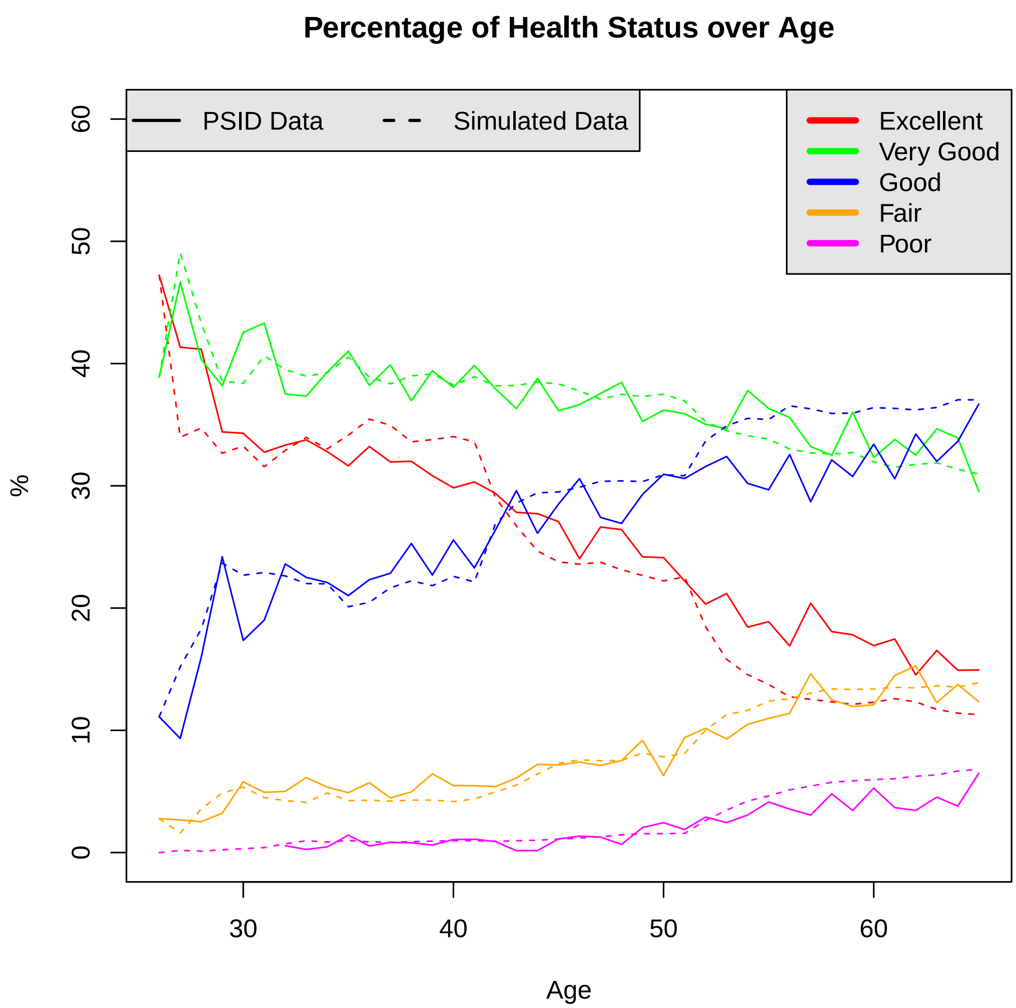
Edu		o_{t+1}		
	o_t	1	2	3
e=1	1	0	50	0
	2	800	0	0
	3	1200	300	0
e=2	1	0	40	0
	2	600	0	0
	3	1000	200	0
e=3	1	0	20	0
	2	400	0	0
	3	600	100	0
e=4	1	0	10	0
	2	200	0	0
	3	300	50	0

Notes: The cost of changing occupations varies by different education levels.

2.4 Goodness of Fit

The model fit can be examined by comparing the observed data and the simulated data. In this section, I simulate data from the life cycle model and compare the simulated data with the observed data from PSID.

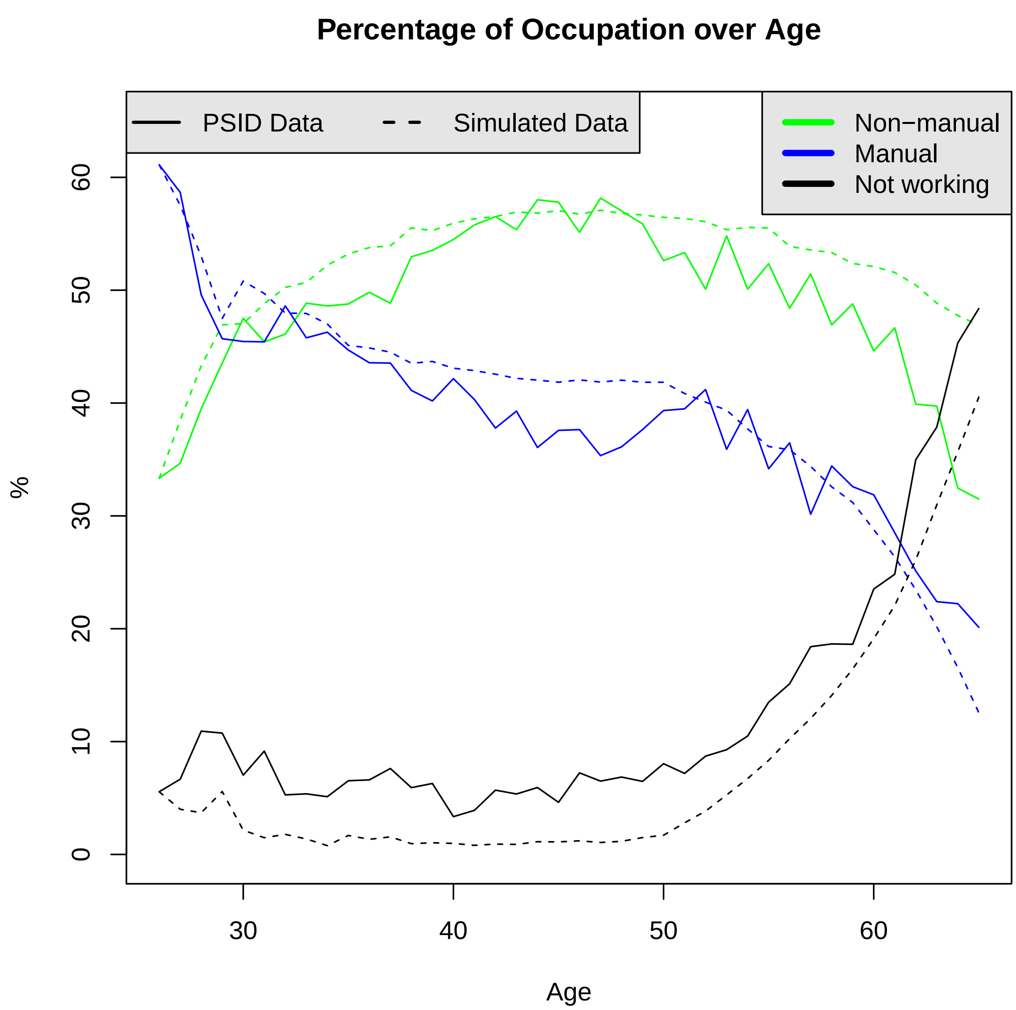
Figure 2.11: **Percentages of Different Health Levels over Age**



Notes: The percentage of each health level during the life time from both the PSID data and the simulated data.

Figure 2.11 compares the observed and the simulated percentages of different health levels over the life cycle. Different colors represent different health levels while the dashed lines represent the simulated data and the solid lines represent the data in PSID. Basically, the percentages of all health levels are matched very well over the life cycle.

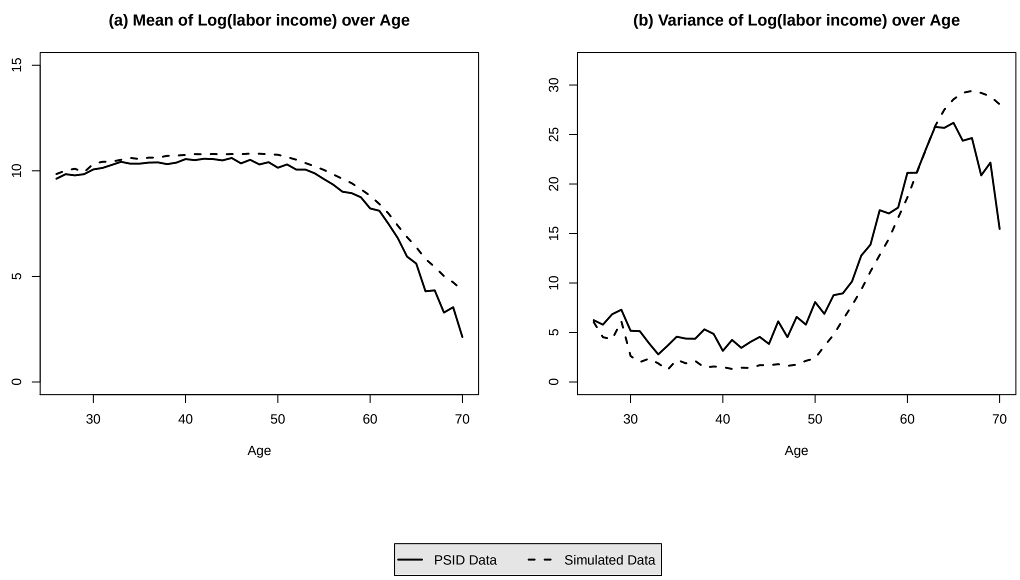
Figure 2.12: **Percentages of Different Occupations over Age**



Notes: The percentage of each occupation during the life time from both the PSID data and the simulated data.

Figure 2.12 compares the observed and the simulated percentages of different occupations over the life cycle. Different colors represent different occupations while the dashed lines are from the simulated data and the solid line are from PSID. Also, the percentage patterns of all occupations are matched very well.

Figure 2.13: **Earnings over the Life Cycle**



Notes: The mean and variance of labor income during the life time from both the PSID data and the simulated data.

Figure 2.13 (a) and (b) plot the mean and variance of labor income over the life cycle respectively. Both the observed data and the simulated data show that the variances are increasing between age 50-67, which is the period for making decisions on retirement.

Overall, the model fits the observed data fairly well, which is the base for the counter-factual experiments in the next section. Still, there are two potential ways to modify the model for improving the its fit.

- Add saving into the model. If saving is allowed in each period, then people can make optimal decisions considering inter-period borrowing. To do this, another state variable, asset (accumulated saving) should be added into the model.

Asset is a continuous variable, which requires more computational resources when I discretize it into fine grids.

- Estimate the life-cycle model. So far, many parameters in the model are calibrated, but they can be estimated by using the method of simulated moments (MSM). Reasonable estimates may generate more reliable simulations.

2.5 Experiments

Based on the excellent fit shown in the last section, I conduct several counter-factual experiments. Many questions can be answered by counter-factual experiments. One of the goals of this study is to explore the effects of initial health and initial occupation conditions. If people start with different occupations and/or different health statuses, their health, occupation, and labor income may be different in their life time. Secondly, I examine what happens if there is a health shock for a typical person in their mid-age. Thirdly, I control the health level of the population to see the influence of health on people's labor income. Last but not the least, I close with some of the connecting channels between health and labor income and re-examine the correlation between them.

2.5.1 Effects of Initial Health and Initial Occupation

I assume people enter the labor market at age 26 ($t = 1$) with different initial health statuses and different initial occupations. The life cycle model predicts the profile of health, occupation, and labor income in the rest of the periods. Since education is an exogenous factor decided outside the model and affects the occupation decisions. I analyze the counter-factual experiments under two extreme cases of education: without high school degree ($e = 1$) and with post-college education ($e = 4$). For both education levels, I simulate the lifetime profiles of health, occupation, and labor income with different combinations of initial health and initial occupation. Since health is an ordered discrete data, I only consider two initial health levels: "Poor" ($h_1 = 1$) and "Excellent" ($h_1 = 5$). Table 2.2 lists the details of all combination of the initial conditions.

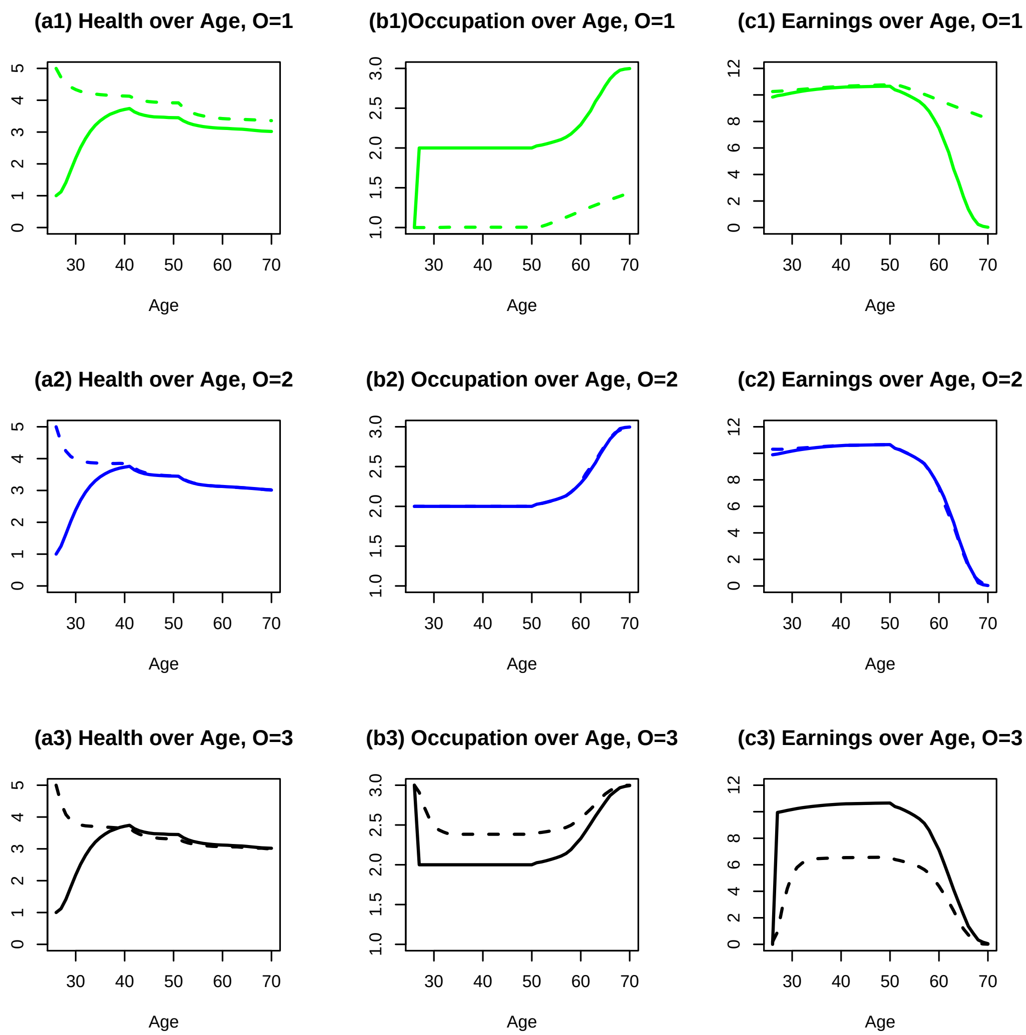
Table 2.2: **Initial Conditions**

Edu=1 No high school degree						Edu=4 Post college					
Occ=1 Non-manual		Occ=2 Manual		Occ=3 Not working		Occ=1 Non-manual		Occ=2 Manual		Occ=3 Not working	
H=1 Poor	H=5 Excellent	H=1 Poor	H=5 Excellent	H=1 Poor	H=5 Excellent	H=1 Poor	H=5 Excellent	H=1 Poor	H=5 Excellent	H=1 Poor	H=5 Excellent

Figure 2.14 shows the cases with Edu=1 (no high school degree). Three different row represent three different starting occupation. In each plot, the solid line is the case started with "Poor" health ($H_1 = 1$) while the dashed line represent the case with "Excellent" initial health ($H_1 = 5$). The first column is the profile of health, the second column is the profile of occupation, and the third column is the profile of log(labor income) over age.

The first column of Figure 2.14 shows that regardless the initial occupation, if people have "Excellent" health ($H_1 = 5$), the health level declines over age. If people start with a "Poor" health, their health levels can be improved before age 40 then decline. Finally, whatever the initial health is, the health levels converge to $H = 3$. The occupation paths tell us that low educated people (Edu=1) mainly work in manual jobs. Except for those people who have excellent initial health ($H_1 = 5$) and non-manual initial occupation ($O_1 = 1$), they work at non-manual occupations for most of their life and have a higher income than others at the old ages. The second column and the third column together show that the labor income is mainly decided by the labor supply type. For example, the lower labor income in (c3) is caused

Figure 2.14: **Simulated Health, Occupation, and Earnings over Age, Edu=1**

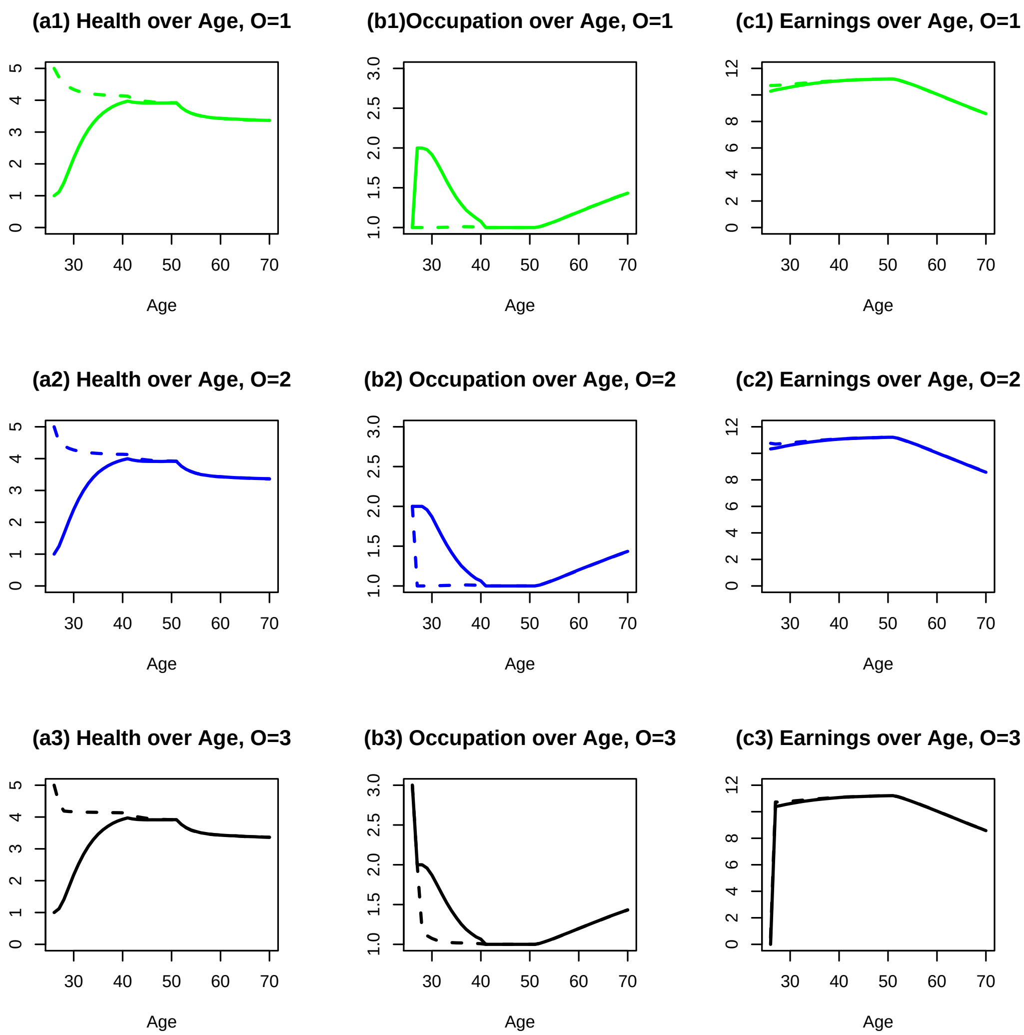


Notes: Solid line are the cases that have initial "Poor" health, while dashed lines are the cases have initial "Excellent" health. Three different rows represent three different starting occupations. The first column is the profile of health, the second column is the profile of occupation, and the third column is the profile of log(labor income) over age. I simulate 100,000 times and plot the means.

by the fact that in some of the simulations, the "head" was "not working" ($o = 3$). Health also generates a difference in labor income directly, but it is very small. The income differences at young ages in (c2) are caused by the different contemporary health levels at those periods in (a2).

Figure 2.15 shows the cases with Edu=4 (post college). Three different row repre-

Figure 2.15: **Simulated Health, Occupation, and Earnings over Age, Edu=4**



Notes: Solid line are the cases that have initial "Poor" health, while dashed lines are the cases have initial "Excellent" health. Three different rows represent three different starting occupations. The first column is the profile of health, the second column is the profile of occupation, and the third column is the profile of log(labor income) over age. I simulate 100,000 times and plot the means.

sent three different starting occupation. In each plot, the solid line is the case started with "Poor" health ($H_1 = 1$) while the dashed line represent the case with "Excellent" initial health ($H_1 = 5$). The first column is the profile of health, the second column is the profile of occupation, and the third column is the profile of log(labor income) over age.

The second column of Figure 2.15 and Figure 2.14 show that highly educated people are more likely to work in the "non-manual" job and while those people with lower education are more likely to be in the "manual" occupations. The education level decides the occupational regardless of the initial occupation. For example, in (b2) of Figure 2.15, even if people start with a "manual" job, they move to a "non-manual" job very soon.

From either Figure 2.14 or Figure 2.15, the labor income patterns are mainly decided by the occupation and labor supply patterns. Basically, a "non-manual" job gives highest labor income, a "manual" job gives lower labor income, while "not working" generates the lowest labor income.

2.5.2 Health Shocks for Typical Persons

Now, let's consider the question: if there is a health shock to a typical person in his mid-age, how would the occupation and labor income be affected by the health shock? It is hard to imagine positive health shocks in people's mid-age, so I only consider the negative health shocks in this section. Many severe heavy diseases or accident injuries can be regarded as negative health shocks.

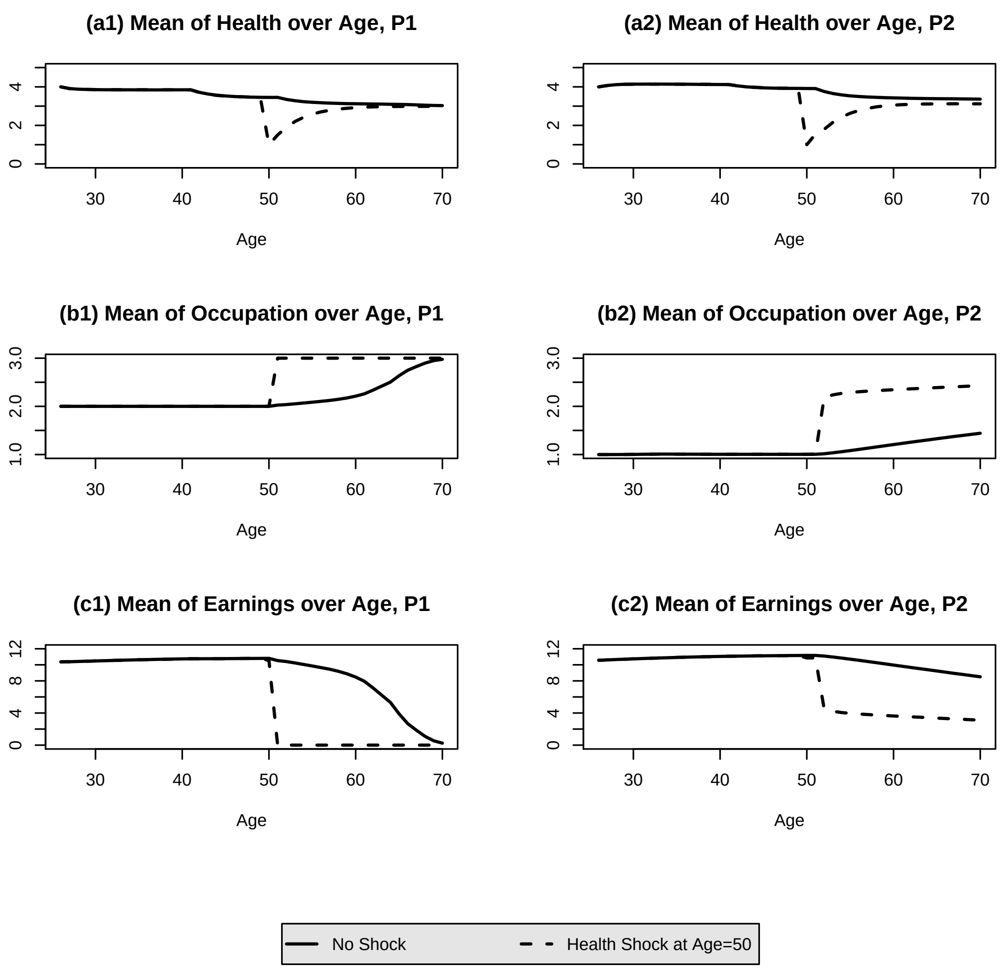
In order to examine the effects of the health shocks, firstly, I simulate the health profile, occupation profile, and labor income profile for a typical person, which would be used as the baseline case without health shocks. Secondly, I put a very negative health shock ($H = 1$) at age 50 and simulate the whole life profile for health, occupation, and labor income again. By comparing the life profiles in the baseline setting and the setting with health shocks, the effects of health shock can be explored.

It is hard to define a typical person, so I consider two typical persons:

- **Person 1 (P1)**: A person with a high school degree ($e = 2$) who has a "manual" job as his initial occupation ($o_1 = 2$) and whose initial health is “Very Good” ($H_1 = 4$).

- **Person 2 (P2)**: A person with a college degree ($e = 3$) who has a “non-manual” job as his initial occupation ($o_1 = 1$) and whose initial health is “Very Good” ($H_1 = 4$).

Figure 2.16: **Effects of Negative Health Shock at Age 50**



Notes: The cases with shocks are shown by dash-lines. The plots in the first column are the simulated health, occupation, and labor income for person 1, and the plots in the second column are the simulated health, occupation, and labor income for person 2. I simulate for 100,000 times and plot the means.

Figure 2.16 shows the effects of health shocks clearly. For person 1, when a very negative health shock comes at age 50, he moves to “Not working” ($o = 3$) for the rest of his life, even if his health level can be recovered back to “Good” ($H = 3$). The labor income drops to zero because of not working. For person 2, when there is a negative health shock, he has more chances to move to “not working” ($o = 3$) and with zero labor income. So the mean of the simulated occupation is much higher than the one without health shocks. At the same time, the labor income drops correspondingly. In short, health status has significant effects on labor supply and labor income.

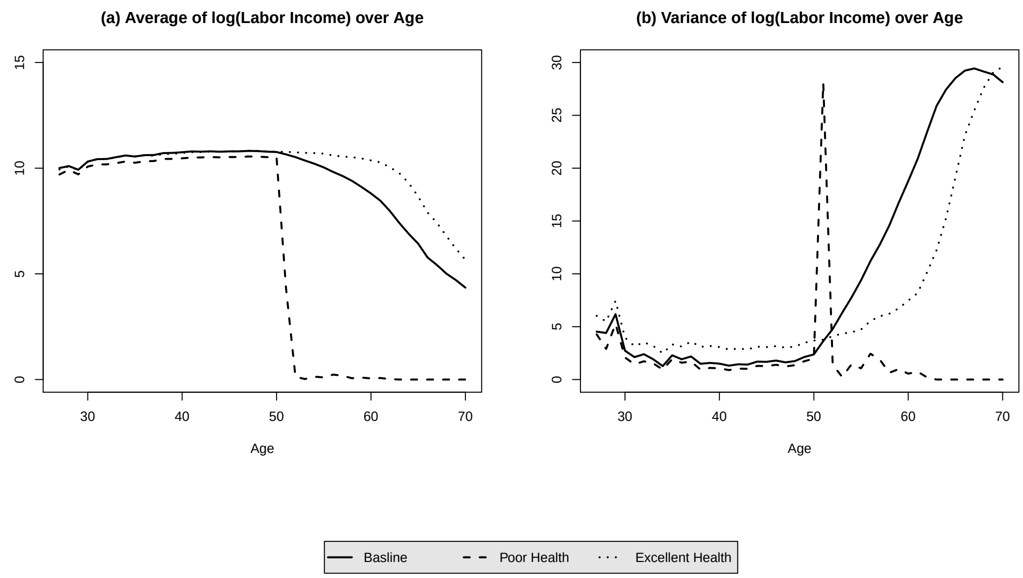
By comparing the solid lines in (b1) and (b2) in Figure 2.16, we know person 2 works in “non-manual” occupations for the most of the time while person 1 works in “manual” jobs for most of his life without the health shock. Also, in (a1) and (a2), the health of person 1 deteriorates faster than the health of person two. So we can also conclude that “manual” occupations lead to a higher health deterioration rate than “non-manual” occupations.

The labor income in (c2) of Figure 2.16 is higher than the labor income in (c1); thus, we can also conclude that “manual” occupations generate higher labor income than “non-manual” occupations.

2.5.3 Health Effects

In this section, I control people’s health level at excellent (H=5) and poor (H=1) for their whole life time, and compare these two scenarios with the baseline case. In the baseline case, people’s health levels are decided by the initial health and the occupation-specified deterioration rates. Figure 2.17 shows the average and variance of the simulated labor income in each of these scenarios. With poor health, people strongly prefer to exit the labor force soon after they are eligible to retire. In this case, the variance of labor income has a peak around age 50. On the other side, if people have excellent health all the time, some of them may keep working at their old ages, so the mean of labor income is higher, and the variance of labor income keeps rising in this case.

Figure 2.17: **Effects of Health**



Notes: Baseline: people's health levels are decided by the initial health and the occupation specified deterioration rates.
Poor health: assume the whole population have poor health during their life time.
Excellent health: assume the whole population have excellent health during their life time.

2.5.4 Health-Income Correlation

In this section, I explore the driving factors within the health-income correlation. In the baseline model, health status and labor income are mutually affected. First, health status is involved the utility function and the wage rate function, therefore people make their decisions accordingly. In other words, people's health status affects people's labor income. Secondly, people's labor income can provide financial support for changing their occupations. Different occupations lead to different health deterioration rates. So, health status is also affected by income levels.

In this section, I close one of the channels at a time to explore the driving factors in the health-income correlation. I try the following to counter-factual scenarios:

- [A] **Fixed Health:** Assume people keep their initial health status during their life time.

- [B] **Fixed Occupation:** Assume people always keep their initial occupation as their occupation during their life time. In this case, there are effects from labor income to health.

Table 2.3 shows the health-income correlation in the observed data and in different model settings. With the mutual effects in the baseline model, the correlation between health and labor income observed in the PSID data can be modeled: the correlation coefficient is 0.32 in the observed data, while it is 0.34 in the baseline model.

First, when the health statuses of all people are fixed, the Correlation Coefficient decreases from 0.34 to 0.28, which means about 17.65% of the health-income correlation is decided by the changes of health status. If there are no changes in health status, those people with relatively bad health are not going to earn less due to the faster deterioration of their health, and also, people cannot improve their health to earn more.

Secondly, when the occupations are fixed, the correlation coefficient decreases from 0.34 to 0.17, which means about 48.24% of the health-income correlation is decided by the changes of occupations. So based on this model, a large portion of correlation comes from labor income with the following mechanism: a higher labor income supports people for moving to a better occupation, and the better occupation can provide a lower health deterioration rate and more labor income in the future.

Conclusively, the channel from labor income to health is the main reason for generating a high correlation between health and labor income.

Table 2.3: **Health-Income Correlation**

	Var(health)	Var(log(labor income))	Correlation Coefficient
DATA	1.02	17.06	0.32
Baseline	1.04	14.52	0.34
Fixed Health [A]	0.86	10.90	0.28
Fixed Occupation [B]	1.13	9.97	0.17

Notes: Baseline: health status and labor income are mutually affected.
Fixed Health Status [A]: Assume people keep their initial health status during their life time .
Fixed Occupation [B]: Assume people always keep their initial occupation as their occupation during their life time.

2.6 Conclusions

This study builds a life-cycle model to explore the dynamic correlations among health, occupation, and earnings observed from PSID data. The simulations from the calibrated model fit the data very well. Based on the good fit of the model, I conduct several simulation-based counter-factual experiments and get the following conclusions:

- Given an occupation and an education level, people's health status converges to the same level regardless the initial health level.

- The occupations are mainly decided by the education level.

- Different occupations have different effects on health status. “Manual” occupations leads to higher health deterioration rate than “non-manual” occupations.
- Occupations have direct effects on labor income. “Non-manual” occupations generate higher labor income than “manual” occupations.
- Given an education level and an occupation, people’s health level affects the labor supply and labor supply decides the labor income directly.
- The main channel in deciding the health-income correlation is the channel that allows labor income to affect health via occupation.

The combinations of these conclusions explain the labor income difference and the reason why income inequality widens. People from wealthy families can get sufficient education. Highly educated people are more likely to be in “Non-manual” occupations, which generate relative good health level. Both good health and good occupation give them higher labor income during the life time, which reinforces income inequality. On the other end, for those less educated people from poor families, once they enter the “manual” occupations, their health deteriorates faster and with poor health, there is a large cost for them to jump out to “Non-manual” occupations. Then, their health levels deteriorates much faster. The poor health level and the low paying occupation give them lower life-time labor income. Policies for providing education for poor people can be effective in reducing income inequality.

2.7 Appendix

2.7.1 Details on Grouping Occupations

From 1974-2001, PSID had a question that asked, “What is your main occupation?”. And the answer is coded as the occupation code from the 1970 Census of Population. The grouping details are:

Group Name	Occupation Code	Rang Text
Non-manual:	1 - 195	Professional, Technical, and Kindred Workers
	201 - 245	Managers and Administrators, except Farm
	260 - 285	Sales Workers
	301 - 395	Clerical and Kindred Workers
Manual:	401 - 600	Craftsman and Kindred Workers
	601 - 695	Operatives, except Transport
	701 - 715	Transport Equipment Operatives
	740 - 785	Laborers, except Farm
	801 - 802	Farmers and Farm Managers
	821 - 824	Farm Laborers and Farm Foremen
	901 - 965	Service Workers, except Private Household
	980 - 984	Private Household Workers
	810	Wild code
Not working:	999	Refused
	0	Not working for money now

From 2003-2013, PSID asked “What was Head’s Occupation last year?”. The occupation codes are from the 2000 Census of Population. The grouping details are:

Group Name	Occupation Code	Rang Text
Non-manual:	1 - 43	Management Occupations
	50 - 73	Business Operations Specialists
	80 - 95	Financial Specialists
	100 - 124	Computer and Mathematical Occupations
	130 - 156	Architecture and Engineering Occupations
	160 - 196	Life, Physical, and Social Science Occupations
	200 - 206	Community and Social Services Occupations
	210 - 215	Legal Occupations
	220 - 255	Education, Training, and Library Occupations
	260 - 296	Arts, Design, Entertainment, Sports, and Media Occupations
	300 - 354	Healthcare Practitioners and Technical Occupations
	360 - 365	Healthcare Support Occupations
	470 - 496	Sales Occupations
	500 - 593	Office and Administrative Support Occupations
Manual:	370 - 395	Protective Service Occupations
	400 - 416	Food Preparation and Serving Occupations
	420 - 425	Building and Grounds Cleaning and Maintenance Occupations
	430 - 465	Personal Care and Service Occupations
	600 - 613	Farming, Fishing, and Forestry Occupations
	620 - 676	Construction Trades
	680 - 694	Extraction Workers
	700 - 762	Installation, Maintenance, and Repair Workers
	770 - 896	Production Occupations
	900 - 975	Transportation and Material Moving Occupations
	980 - 983	Military Specific Occupations
Not working:	999	refused
	0	did not work for money

2.7.2 Estimation of Wage Rate Function

The wage rate (hourly pay) of household (w_t) depends on the health status h_t, age t (a proxy for experience), occupation o_t, education level e, and a normal shock. The functional form of the wage rate is given as:

$$\begin{aligned} log(w_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_{3i}\sum_{i=1}^{4} 1\{H_t = i\} + \beta_{4i}\sum_{i=1}^{2} 1\{o_t = i\} \\ +\beta_{5i}\sum_{i=1}^{2} t\cdot 1\{o_t = i\} + \beta_{6i}\sum_{i=1}^{2} t^2\cdot 1\{o_t = i\} + \beta_{7i}\sum_{i=1}^{3} 1\{e = i\} + \epsilon \\ \epsilon \sim N(0, 0.6^2) \end{aligned} \tag{2.7}$$

The βs in the wage function are estimated with the pooled panel data from the PSID. The estimates are in Table 2.4

Table 2.4: **Estimates of the wage rate function**

	Estimate	Standard Error
β_0	3.8	0.16
β_1	-0.064	0.0055
β_2	0.00027	0.000045
β_{31}	-0.424	0.034
β_{32}	-0.361	0.02
β_{33}	-0.221	0.014
β_{34}	-0.097	0.013
β_{41}	-2.45	0.21
β_{42}	-2.14	0.22
β_{51}	0.15	0.0078
β_{52}	0.14	0.0082
β_{61}	- 0.0011	0.00007
β_{62}	-0.001	0.00008
β_{71}	- 0.444	0.020
β_{72}	-0.297	0.0153
β_{73}	- 0.038	0.0167

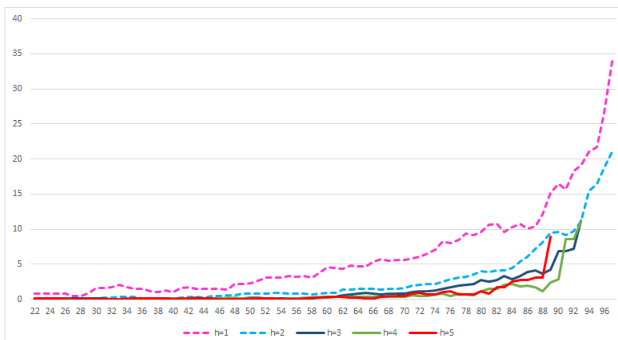
Notes: All βs are significant at 0.05 significant level.

2.7.3 Death Rate

The death probability over age increases exponentially for different health status in Figure 2.18. After taking log for the death rate, it can be modeled by linear functions. The estimated linear model based on the data from PSID is :

$$
\begin{aligned}
log(d(t,h)) = & -0.61 + 0.02t - 1.1I(h=2) - 1.5I(h=3) - 1.7I(h=4) - 1.65I(h=5) \\
& + 0.0095I(h=2)t + 0.01I(h=3)t + 0.01I(h=4)t + 0.008I(h=5)t
\end{aligned}
$$

Figure 2.18: **Death Rate over Ages for Different Health Status(%)**



Notes: The dead people in PSID are grouped by their death-age and death-health. Meanwhile, the samples in PSID are pooled to count a total number of people in each age-health cell. The death rate is the ratio between the total death and the total number of people in each age-health cell.

Bibliography

Adams, Peter, Hurd, Michael D, McFadden, Daniel, Merrill, Angela, & Ribeiro, Tiago. 2003. Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. *Journal of Econometrics*, **112**(1), 3–56.

Akay, Alpaslan. 2009. The Wooldridge method for the initial values problem is simple: what about performance?

Banerjee, Abhijit V, & Newman, Andrew F. 1993. Occupational choice and the process of development. *Journal of Political Economy*, 274–298.

Börsch-Supan, Axel, & Hajivassiliou, Vassilis A. 1993. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of econometrics*, **58**(3), 347–368.

Bound, John, Schoenbaum, Michael, Stinebrickner, Todd R, & Waidmann, Timothy. 1999. The dynamic effects of health on the labor force transitions of older workers. *Labour Economics*, **6**(2), 179–202.

Carson, Scott Alan. 2014. Nineteenth Century White Physical Activity and Calories: Socioeconomic Status and Diets.

Case, Anne, & Deaton, Angus S. 2005. Broken down by work and sex: How our health declines. *Pages 185–212 of: Analyses in the Economics of Aging*. University of Chicago Press.

Case, Anne, & Paxson, Christina. 2006 (August). *Stature and Status: Height, Ability, and Labor Market Outcomes*. Working Paper 12466. National Bureau of Economic Research.

Case, Anne, Paxson, Christina, & Islam, Mahnaz. 2009. Making sense of the labor market height premium: Evidence from the British Household Panel Survey. *Economics letters*, **102**(3), 174–176.

Chetty, Raj, Friedman, John N, & Rockoff, Jonah E. 2011. *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood*. Tech. rept. National Bureau of Economic Research.

Costrell, Robert M, & Podgursky, Michael. 2009. Teacher retirement benefits. *Education Next*, **9**(2).

De Beyer, Joy, & Knight, John B. 1989. The role of occupation in the determination of wages. *Oxford Economic Papers*, 595–618.

Ettner, Susan L. 1996. New evidence on the relationship between income and health. *Journal of health economics*, **15**(1), 67–85.

Frijters, Paul, Haisken-DeNew, John P, & Shields, Michael A. 2005. The causal effect of income on health: Evidence from German reunification. *Journal of health economics*, **24**(5), 997–1017.

Grossman, Michael. 1972. The demand for health: a theoretical and empirical investigation. *NBER Books*.

Haller, Archibald O, & Spenner, Kenneth I. 1977. Occupational income differentiation in status attainment. *Rural Sociology*, **42**(4), 517.

Harris, Douglas N, & Adams, Scott J. 2007. Understanding the level and causes of teacher turnover: A comparison with other professions. *Economics of Education Review*, **26**(3), 325–337.

Hausman, Jerry A, & Wise, David A. 1979. Attrition bias in experimental and panel data: the Gary income maintenance experiment. *Econometrica: Journal of the Econometric Society*, 455–473.

Heckman, James J. 1981. VThe Incidental Parameters Problem and the Problem of Ini (tial Conditions in Estimating a Discrete Time (Discrete Data Stochastic Process and Some Monte Carlo Evidence, V in Structural Analysis of Discrete Data with Econometric Appli" cations, in C. Manski and D. McFadden. *MIT Press, Cambridge, MA*, **179**, 195.

King, Allan G. 1973. Occupational choice, risk aversion, and wealth. *Indus. & Lab. Rel. Rev.*, **27**, 586.

Michaud, Pierre-Carl, & Van Soest, Arthur. 2008. Health and wealth of elderly couples: Causality tests using dynamic panel data models. *Journal of health economics*, **27**(5), 1312–1325.

Mullahy, John, & Sindelar, Jody L. 1994. *Alcoholism, work, and income over the life cycle*. Tech. rept. National Bureau of Economic Research.

Muurinen, Jaana-Marja. 1982. Demand for health: a generalised Grossman model. *Journal of Health Economics*, **1**(1), 5–28.

Ni, Shawn, & Podgursky, Michael. 2016. How teachers respond to pension system incentives: New estimates and policy applications. *Journal of Labor Economics*, **34**(4), 1075–1104.

Oster, Emily, Shoulson, Ira, & Dorsey, E. 2013. Limited life expectancy, human capital and health investments. *The American Economic Review*, **103**(5), 1977–2002.

Pagan, Jose A, & Davila, Alberto. 1997. Obesity, occupational attainment, and earnings. *Social Science Quarterly*, 756–770.

Prescott, Edward C. 1986. Theory Ahead of Business Cycle Measurement. *Quarterly Review*.

Ravesteijn, Bastian, van Kippersluis, Hans, & van Doorslaer, Eddy. 2013. The contribution of occupation to health inequality. *Research on economic inequality*, **21**, 311–332.

Rivkin, Steven G, Hanushek, Eric A, & Kain, John F. 2005. Teachers, schools, and academic achievement. *Econometrica*, **73**(2), 417–458.

Rosen, Harvey S, & Wu, Stephen. 2004. Portfolio choice and health status. *Journal of Financial Economics*, **72**(3), 457–484.

Schultz, T Paul. 2002. Wage gains associated with height as a form of health human capital. *American Economic Review*, 349–353.

Stevens, Gillian, & Featherman, David L. 1981. A revised socioeconomic index of occupational status. *Social Science Research*, **10**(4), 364–395.

Stock, J, & Wise, D. 1990. Pensions, the Option Value of Work, and Retirement. *Econometrica*, **58**(5), 1151–1180.

Wooldridge, Jeffrey M. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of applied econometrics*, **20**(1), 39–54.

VITA

Weiwei Wu was born in the People's Republic of China, on April 23, 1988. Between 2007 and 2011, he studied economics at Nankai University in Tianjin, China. He received an M.A. in economics from University of Missouri in Dec 2015, and a Ph.D. in economics with a minor in statistics from University of Missouri in May 2017.