COMPARISONS AND EXTENSIONS OF REDUCED FORM AND STRUCTURAL APPROACHES TO RETIREMENT DECISION OF MISSOURI PUBLIC SCHOOL TEACHERS

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

CHAO WU

Dr. Ni, Xiaoguang & Dr. Podgursky, Michale J., Dissertation Supervisors

MAY 2019

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

COMPARISONS AND EXTENSIONS OF REDUCED FORM AND STRUCTURAL APPROACHES TO RETIREMENT DECISION OF MISSOURI PUBLIC SCHOOL TEACHERS

presented by Chao Wu,

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

Dr. Ni, Xiaoguang & Dr. Podgursky, Michael J.

Dr. Koedel, Cory R.

Dr. Podgursky, Michael J.

Dr. Ni, Xiaoguang

Dr. Tsoi, Allanus

ACKNOWLEDGMENTS

This research was conducted at the Department of Economics. It was financed by the Economics and Policy Analysis Research Center.

I am grateful for the opportunity to work at the Economics and Policy Analysis Research Center. Rich data and programming sources have given me the chance and access to explore data and try different models. Many colleagues at the center have provided me with enormous assistance. Dr. Mark Ehlert not only provided me with work instructions but also trained me to build good research habits. Margie has always given me a hand whenever I have been confronted with programming difficulties.

Many professors and students played a substantial role in shaping this work. Among them all, I am deeply grateful to my advisor, Xiaoguang Ni. He has guided me through these years and has had a major impact on this thesis. Thank you for guiding me through the subtleties of writing as well, often with a big dose of patience.

I am also very grateful to my other committee members - Dr. Cory R. Koedel, Dr. Michael J. Podgursky, and Dr. Allanus Tsoi - for reading the thesis and providing comments.

I want to express my heartfelt gratitude to all the teaching and research staff who have taken some time to discuss and enrich my work. The Department of Economics has provided me with a very stimulating environment with the extraordinary quality of its academic staff. That experience will leave a mark beyond this thesis.

A very special word of thanks go to my parents for their unconditional love. The last words go to my husband, Xiaoxing Lu, and my two loving daughters, Lu, Ann Q. and Lu, Belle W. for their support and companionship. This thesis is dedicated to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

There have been growing calls to reform teachers' pension systems. Designing of pension reforms requires modeling retirement decisions first. My thesis compares different models of teachers' retirement decisions.

Chapter one reviews studies published after 1982 on retirement incentives, methodologies and data sets. Retirement wealth and Social Security in retirement are the main points of focus; however, their effects are debatable. Probit models, dynamic models, and option value models are among the most widely applied methodologies. Retirement research is restricted by the lack of data. Large surveys on retirement, such as the Health and Retirement Study (HRS), are quite rare.

Panel data on Missouri public school teachers provides a good opportunity to study how a pension incentive affects retirement. In chapter two, I study how teachers' retirement is influenced by pension rules of the Missouri Public School Retirement System (PSRS). I estimate a simple probit model on Missouri public school teachers' administrative data. The probit model fits the data well, however, its coecients depend on the existing pension rules.

To further predict behavioral responses to changes in pension rules, I estimate a structural model of retirement decisions (the option value model by Stock and Wise (1990)) by maximizing likelihood (ML). In chapter three, I focus on a number of technical issues in ML estimation. I show that the commonly used frequency simulator for likelihood evaluation is computationally costly when the data set has a large number of teachers and long panels. In addition, the maximum obtained by a hill-climbing algorithm may not be global when the objective function is ill-shaped. Therefore, I propose to evaluate the option value model's likelihood by using the GHK simulator, in place of the frequency simulator, and obtain robust estimates via simulated annealing instead of the hill-climbing algorithm. The results suggest that the GHK simulator is much more efficient than the frequency simulator and that

the aggregate behavior that is predicted on the basis of the estimates of the GHK simulator is reasonably robust.

The last chapter compares the fit of the probit model with the option value model and discusses the pros and cons of the two competing models for teachers' retirement decisions.

Chapter 1

Literature Review on Retirement Decision

1.1 Introduction

Understanding employees' retirement behavior is essential for government or institutions to establish policies. There are two renowned puzzles in retirement studies for U.S. workers: one involves the two retirement spikes at ages 62 and 65, which coincide with the early entitlement age and full retirement age for the Social Security (SS) benefit, respectively; however, there is no big incentive to retire at the early entitlement age as future benefits increase to an actuarially fair amount beyond that age (Gustman and Steinmeier, 2005). The other one is that consumption spending drops at retirement, which contradicts the assumption of continuous consumption over retirement in the simple, one-good life-cycle model. Many studies attempted to provide reasonable explanations to these puzzles.

I use three threads to link these studies. The first one summarizes potential economic incentives for retirement decisions covered in these studies, the second categorizes model methodologies, and the last one reviews commonly applied data sets.

The U.S. retirement system is briefly introduced in section 2. Conclusions are drawn at the end.

1.2 The U.S. Retirement System

Retirement plans can be roughly classified as defined benefit (DB) plans and defined contribution (DC) plans, depending on how benefits are determined. In a DB plan, benefits are calculated using a fixed formula that factors in participants' final pay and years of service with their employer. In contrast, DC benefits come from the amount of money contributed by a participant themselves and investment performance in their retirement account.

In the past two decades, the trend has shifted from DB plans to DC plans in private retirement arrangements. Among private sector workers, DB plan coverage fell from 83% in 1980 to 31% in 2008 and further to 18% in 2011. The shift implies that individuals take more responsibility in adequately saving for retirement and investing their savings. On the other hand, most employees in the public sectors are covered by DB plans and the coverage slowly increases. The U.S. Census Bureau (2016) reports that the total membership of state and locally administered pension systems exceeds 20 million, which is a 3.7% increase since 2014. Total beneficiaries from those pensions increased by 4.3%, from about 9,559,956 members in 2014 to 9,971,726 members in 2015. Friedberg and Owyang (2002) describe the development of the two plans in detail.

In addition to the typical DB pension plans, SS is an important source of most Americans' retirement income. SS was launched in 1935. Originally, the act required that all workers in commerce and industry (except railroads) under the age of 65 to be covered. Over the years, SS has expanded to almost all workers, except for some state and local government employees. In 1956, benefits (reduced for early retirement) were made available to women between the ages of 62 and 65. Men began to receive the same treatment in 1961. In 2016, approximately 89% of workers aged 21-64 were covered by SS; overall, around 97% of American workers were covered by SS. SS accounts for over half of the income of 62% of SS beneficiaries aged 55 and older.

There were three amendments imposing great effects on SS structure. Before 1972, monthly SS benefits were based on average nominal monthly earnings (AME), generally excluding some years of low earnings. In calculating the AME, earnings were truncated at SS taxable maximum. A fixed, progressive benefit formula was applied to the AME to derive the Primary Insurance Amount. Until 1972, Congress changed the benefit formula to raise benefits, making adjustments for inflation on an ad hoc basis. In 1977, amendments were enacted to eliminate over-indexed benefits. Meanwhile, average indexed monthly earnings replaced AME and came to be applied in a new benefit formula. The 1977 amendments covered only those who were aged 60 or younger at that time. To phase in the lower benefits, a special five-year transition period was adopted for individuals born between 1917 and 1921, known as the notch babies. The last major SS reform occurred in 1983 on account of an imminent shortage of funds. The most notable change is that, of the normal retirement age, which increases from the current age of 65 to 67. This takes effect for people aged 62 in 2022 and thereafter. The retirement benefit claimed earlier still remains at age 62; however, the amount of benefits received at age 62 or later would be smaller if the normal retirement age is higher.

Table 1.1: Effects on age of retirement

Variable	Effect on age of retirement
For an Income-Maximizer	
Increase in pension intercept	Earlier
Increase in slope of pension function	Ambiguous
Increase in earnings	Later
For a Utility-Maximizer	
Variable	Effect on age of retirement
Increase in pension intercept	Earlier
Increase in slope of pension function	Ambiguous
Increase in earnings	Ambiguous

1.3 Economic Incentives

1.3.1 Pre-1982 Research

Mitchell and Fields (1982) review studies on the effects of pensions and earnings on retirement. The framework of their review is the life cycle theory of work and leisure allocation. According to Mitchell and Fields, an agent aims to maximize either his lifetime utility, which is a function of consumption and leisure, or his lifetime income. Comparative static analysis is mainly applied to analyze how retirement age would change by the control of some instruments. Mitchell and Fields summarize three economic factors that have received serious consideration: wage stream on the main job, private pension benefits, and SS benefits. The theoretical effects of these three factors on retirement age are summarized in Table 1.1.

Empirically, corresponding to the above three variables, Mitchell and Fields conclude the following:

- Current wage: no statistically significant effect of changes. No impact according to Gustman and Steinmeier (1981), and no statistically insignificant effect from Hurd and Bosking (1981) nor from Quinn (1977).

- Social Security: ambiguous effect. Boskin (1977) and Boskin and Hurd (1978) find very large social security effects. Reimers (1977) and Kotlikoff (1979) find that higher benefit levels have no impact.

- Private pensions: ambiguous effect. Quinn (1977) concludes that people who are eligible for private pensions are significantly more likely to retire. This conclusion is reiterated by Gordon and Blinder (1980), as well as by Gustman and Steinmeier (1981), although no study reports significant levels. On the other hand, studies by Reimers (1977), Burkhauser and Quinn (1980), and Kotlikoff (1979) reveal that people with pensions are significantly less likely to leave the labor force.

1.3.2 Post-1982 Research

One can analyze retirement decisions beyond the framework of life cycle theory. More methodologies are introduced in the following section. Moreover, economic incentives to retire are extended beyond the three variables mentioned above.

In psychological theory, Beehr (1986) first propose a comprehensive model on retirement behavior that incorporates personal factors (e.g., health and economic well-being) and environmental forces (e.g., job characteristics and leisure interests). Those influential factors and forces can be classified into two categories, the "push" and the "pull." Push factors are defined as negative considerations inducing workers to retire, such as poor health or dislike of one's job. Pull factors are typically positive considerations that attract workers to the workforce, such as the desire to pursue job satisfaction. The same event can be rated as either a push factor or a pull factor by different workers because of context. One example of this is retirement incentive programs. More generous retirement benefits can cause some employees to postpone retirement, while these benefits encourage others to retire earlier. With the availability of more informative data sets and the development of research methodology, more economic factors have been included in discussion beyond the above three factors since the 1980s. In the following sections, the “push” or “pull” effects of these factors will be discussed individually.

Current Wage

Particularly in DB plans, an employee’s current wage determines future retirement benefits. In studies after 1982, both the substitution effect and the income effect caused by wage still bring ambiguous changes to retirement. French (2005) points out that wage elasticity on average working hours varies with age. The elasticity is 0.19-0.37 at age 40. The number increases with age and reaches 1.04-1.44 by age 60. Kopecky (2011) attributes the rise in the retirement of males aged 65 and over to two main factors: wage and price of leisure. According to his research, the above two factors can explain 87% of the increase in the retirement rate of men over the age of 65.

It is worthy noting that if economic incentives are extended beyond wealth, wage may cause a selection problem as it interacts with other variables such as health. Grossman (1992) predicts that individuals who have higher expected lifetime wages have good health but not vice versa. In this case, wage does not have a legitimate influence on retirement decisions.

Social Security

As mentioned in the history of SS, SS plays an important role in American workers’ retirement. However, SS is expected to face a long-term financial imbalance that would force sharp benefit cuts in 2034 unless the government makes changes. Both falling fertility rates and labor force growth mean lower collection of payroll taxes to fund the system. On the other hand, the retirement of the large baby boomer population increases benefit costs. There are substantial documents that aim to find the economic relationship between SS and retirement behavior, and to simulate behavior change by simulating policy change. One reform in action is the postponment of the full entitlement age from 65 to 67.

In 1984, Hurd and Boskin claim that increased SS benefits during the early 1970s accounts a good explanation for the decline in elderly male labor force participation during that period. Gustman and Steinmeier (1986) and Rust and Phelan (1997) find that SS benefits have a strong negative effect on male labor supply. Using data from the Current Population Survey, Krueger and Pischke (1992) conclude that SS benefits growth can explain less than one sixth of the decline in male labor force participation rates during the 1970s. Anderson, Gustman and Steinmeier (1999) conclude that increases in pensions and SS can account for about a quarter of the total trend toward earlier retirement observed from 1960 to 1980 but have no effect on retirement for those above age 65. Lumsdaine, Stock, and Wise (1994) find that while changes in pension plans have a significant effect on retirement, the effect of changes due to SS is modest. French (2005) estimates that reducing SS benefits by 20% would cause workers to delay their exit from the labor force by only three months. Other studies, including Burtless (1986), Burtless and Moffitt (1984), Fields and Mitchell (1984), Hausman and Wise (1985), Sueyoshi (1989), Stock and Wise (1990) and Krueger and Pischke (1992), conclude that SS plays a modest but important role.

In addition to the effects from SS income, French and Jones (2011) claim that eliminating two years' worth of SS benefits increases workers' work years by 0.076 years. Coile and Gruber (2007) find that a one-standard-deviation change to SS structure can have important effects on retirement decisions.

Private Pensions

If each dollar of retirement wealth is weighed equally by a retiree, then private pension and SS can be treated the same way as retirement income effects. However, if not, either because the retiree understands private pension incentives better than SS incentives or because SS and private pension incentives are valued differently, then it is important to estimate the impact of SS and private pension separately.

Fields and Mitchell (1984) indicate a larger role for private pensions when compared with SS. In contrast, Diamond and Hausman (1984) suggest that retirement is actually more responsive to SS than it is to private pensions.

In the 1970s, the primarily explored data set is the Retirement History Survey (RHS). However, private pension information in that survey is limited. The more informative HRS survey allows a wider investigation of the effect of pensions on retirement. Gruber (2005) and Coile and Gruber (2007) find that retirement decisions respond, in a significantly positive way, to the level of retirement wealth. The response is roughly equal to a comparable change in SS and pension incentives. Samwick (1994) emphasizes that it is private pension, instead of SS, that primarily determines the change in retirement wealth.

Lumsdaine et al. (1994) find that while changes in pension plans have a significant effect on retirement, SS has only a modest effect. However, their conclusion is based on sampling from one company, which is not representative of the whole retirement population.

Chan and Stevens (2008) indicate that there are differences between administrative and self-reported pension data. Well-informed individuals are far more responsive to pension incentives than are average individuals. Ill-informed individuals seem to respond systematically to their own misperception of pension incentives.

Health

In addition to financial incentives, health is the top factor that has been discussed. There are two strands of studies on the relationship between health and retirement. One is constructing the health factor in a structural model, usually by treating it in the same manner as other monetary variables in the objective utility function. The other is based on reduced form regressions. Those studies draw mixed conclusions on the effect of health on retirement.

Hurd and Boskin (1984) find that bad health has a strong influence on early retirement. Bazzoli (1985) claims that the impact of health has been overestimated in the extant literature. Sickles and Taubman (1986) indicate that retirement decisions are strongly affected by health status. SS and pension payments have positive effects on healthiness. healthiness. McGarry (2004) finds that poor health has a large effect on labor force attachment. The gap between the expected probability of continued work in fair or poor health and that in excellent health is 8.2%. Dwyer and Mitchell (1999) use more objective measures to disclose that the influence of health problems on retirement plans is strong. Moreover, they find that men in poor health are expected to retire one to two years ahead of time than those in good health. Rust and Phelan (1997) find that unhealthy individuals are roughly twice as likely as healthy individuals to apply for SS benefits at the early retirement age. In contrast, Fields and Mitchell (1984), Moffitt (1987), Burtless (1986), and Krueger and Pischke (1992) conclude that large increases in real SS benefits over the past four decades have little effect on behavior. French (2005) finds that health fails to explain the decline of labor force participation around retirement age even though the labor force participation rate of healthy individuals is above that of unhealthy individuals aged 40 and over.

Health Insurance

The earliest health insurance studies mainly focus on estimating the effect of a retiree's health insurance on retirement. Subsequent studies have branched out to consider other types of health insurance, such as Medicare. Similar to the treatment of health, health insurance is also commonly included in both structural models and reduced form models.

Gruber and Madrian (2002) provide a comprehensive literature review of over 50 papers written since 1990 on the relationship between health insurance and labor supply, as well as job mobility. They conclude that health insurance is a central determinant of retirement decisions, although the effect is ambiguous. Besides, they specifically review the relationship between health insurance and labor supply decisions made by low-income mothers. According to them, the relationship is not statistically significant.

Among the studies in the past 10 years, Ferreira and Santos (2013) suggest that most of the changes in the retirement profile by age, observed in the second part of the 20th century, can be explained by the changes in SS during that period. The introduction of Medicare accounts also plays a role. By focusing on the retirement behavior of spouses instead of individuals, Blau and Gilleskie (2006, 2008) indicate that health insurance has a modest impact on the labor force behavior of older married couples. What is more, the impact differs between married men and married women.

With reggard to the entitlement change of Medicare, French (2011) concludes that raising Medicare eligibility age from 65 to 67 leads to individuals aged 60-69 working for an additional 0.074 years, whereas Blau and Gilleskie (2006) imply that the change in age eligibility has little influence on employment.

Spouse

During the first half of the last century, relatively few studies examine retirement decisions in a household context. The subject of retirement research is a single individual, paticularly a male over age 50. With dramatic changes in the labor force behavior of both older men and older women during the post-war period, research is increasingly considering a couple as a unit.

In early studies, such as Sickles and Taubman (1986), a spouse's retirement status is set as an exogenous variable. Hurd (1990) proposes that the retirement of husbands and wives is a joint process and that both spouses usually retire within a short period of one another; however, the causes are not clearly identified. Gustman and Steinmeier (1994) construct a structural model in which couples retire together. Their estimates of a spouse's retirement are jointly significant. Gustman and Steinmeier (2004) introduce a new variable - the measurement of how much each spouse values being able to spend time in retirement with the other - into their structural model. They claim that the new variable accounts for much of the apparent interdependence of retirement decisions within a household. Coile (2004) finds that men and women respond similarly to their own retirement incentives. In contrast, their response to each other's incentives is different. In sum, men are very responsive to their wives' incentives but not vice versa. In that paper, Coile highlights "spillover effects," which come from spouses' financial incentives. Without accounting for the spillover effect, policy reforms would be inaccurate.

Tax

Diamond and Gruber (1999) explore implicit tax rates on continued work. They find that, "for married men with non-earning spouses, there is little net tax on continued work around the age of early eligibility for SS, but the tax becomes quite large at the normal retirement age." Laitner and Silverman (2012) discuss tax policy in their life cycle model and claim that tax reform can extend working lives and lower the federal deficit.

Other Demographic Factors

Other demographic factors include, but are not limited to, age and gender. Age is often treated in a reduced form model as an explanatory variable in three forms: continuous form, dummy variable, and categorized variable. For example, Gustman and Steinmeier (2004) set age as a continuous variable in the coefficient of leisure preference in a structural model, whereas Coile and Gruber (2001, 2007) as well as, Panis et al. (2002), include age dummies in reduced form retirement equations. The dummy coefficient for age 62 is strong and significant. Blau (1994) includes both a polynomial function of age and dummy variables for ages 62 and 65 in his probit model. Gruber and Madrian (1993) put age dummies from 55 to 64 in a probit model. Warren and Oguzoglu (2010) also incorporate age dummies from 56 to 70 to study Australian retirement decisions.

With regard to gender, Lumsdaine et al. (1994) find that women are perhaps slightly more likely to take early retirement (between the ages of 55 and 60) than men. However, at most other ages, there is no significant difference in the retirement behavior of men and women. Gustman and Steinmeier (2004) note that husbands are more influenced by whether their spouse is retired or not than wives are.

1.4 Methodology

1.4.1 Reduced Form Model

Retirement decision models can be separated into structural and reduced form models. The primary advantage of the reduced form approach lies in its functional specification flexibility. Various potential explanatory variables can be incorporated. Most articles published in the 1970s and 1980s consider reduced form models in which retirement decisions are a function of economic variables such as SS wealth. To a large extent, the literature has been reviewed in Mitchell and Fields (1982).

Hazards Model

Hausman and Wise (1985) extend the traditional hazards model by specifying disturbances following a continuous time Brownian motion (or Wiener) process. This innovation allows the hazards model to have much in common with qualitative choice models. Even though implementing this kind of hazards model is not as "forward-looking" as the nonlinear budget constraint model, updated information is mainly achieved through age variables, rather than other variables, in future years. They also introduce a new variable, accrual pension wealth, into the model to measure pension incentives in a dynamic approach.

In Sueyoshi's (1989) hazards model, he includes partial retirement, defined as "employment at a job less than 35 hours of work per week or less than 46 weeks," into implicitly assumed retirement, which he defines as "not employed in a full-time job." The same category can also be found in Gutsman (1986).

Blau (1994) further extends the retirement definition from permanent retirement to retirement of three different types: unemployed, part-time employed, and full-time employed. Labor force status transition is examined in detail by using the hazards model. Of all elements in the budget constraint, only SS benefits have large effects on labor force transitions at older ages. Despite the importance of SS benefits, changes in benefits over time account for only a small part of the decline in male labor force participation in the 1970s.

Recently, Manoli et al. (2015) apply a proportional hazards model, including the variable of SS wealth accrual, to estimate both SS wealth and one-year accrual in SS wealth elasticities.

Probit Model/Logit Model

The probit/logit retirement probability regression models directly estimate economic incentives' effects on retirement probability; however, once the model specification deviates from the "true" economic incentives, its predictive ability weakens. Similar to a hazards model, forward-looking variables capture some important features of the option value model (which will be elaborated later), and the dynamic programming rules are included in recent probit models.

Lumsdaine et al. (1992) include age, current SS value, present SS value, accumulated SS, pension wealth, SS accrual, pension wealth accrual if a person works for another year, and expected wage income in their probit model. Using a probit model, Gruber and Madrian (1993) find a significant effect of post-retirement health insurance on retirement by considering a variable of "continuation of coverage" mandates. Samwick (1998) introduces the option value in his probit model. Retirement wealth, retirement wealth accrual, dummy variable of pension coverage, wage, tenure, financial assets, housing net equity and other demographic variables are also contained in their probit model to explore a unique data set. There are many more innovative variables examined in the probit model to investigate the effects on retirement probability in recent years. For instance, individuals' knowledge of financial incentives is introduced by Chan and Stevens in 2008.

In their 2007 publication, Coile and Gruber introduce a new variable, "peak value," into an explicit linear regression model. The variable is defined as "the value of continuing to work until the future year when SS wealth is maximized, or the difference between the expected present discounted value (PDV) of SS wealth at its highest possible value in the future and the expected PDV of SS wealth if one retires this year." "Peak value" is an extension of "accrual value," introduced by Hausman and Wise in 1985, that involves looking ahead by more than just one year. Besides "old" variables, such as expected PDV; the common measure for retirement benefits (net present discounted value excluding tax/subsidy); individuals' characteristics (age, education, race, and so on); and expected earnings next year - Coile and Gruber compare the contribution of "peak value" with two other measures of financial incentives: accrual value and option value. The linear structure makes SS decomposition and private pension incentive effects straightforward so that comparing effects from different pension sources is possible. The three "accrual" measures' performance can be compared. They find that both the peak value and the option value models perform much better than an accrual-only model. These results also suggest that changes to SS structure can have an important influence on the retirement decision-making. However, as they admit, their modeling is still far from "explaining" retirement decisions, such as the reasons for the enormous "spikes" to exist. It is also found that retirement is roughly equally responsive to a comparable change in SS and pension incentives.

Friedberg, Leora, and Webb (2003) adopt the same model that is introduced by Coile and Gruber (2001) with the classification of DB and DC plans. They claim that the spread of 401(k) and other DC plans in place of DB plans brings substantial changes in retirement patterns via major differences incurred in pension wealth accrual. DC plans accrue pension wealth smoothly, while traditional DB plans boost pension wealth sharply at older ages and turn negative afterward. Their estimates reveal that workers with DC plans are retiring significantly later. According to their findings, "Retirement patterns begin to diverge at around age 55 and accelerate around age 60, when most workers with DB plans begin to experience negative accruals." Instead of explaining the reason for the drastic decrease in labor force participation of the elderly since 1950, they provide reasons for the rebounding of employment rates since 1997. Studies taking the same approach include Blundell et al. (2002) on retirement in Britain and Warren and Oguzoglu (2010) on retirement in Australia.

1.4.2 Structural Model

As Lucas pointed out in 1976, unlike reduced form methods, the structural approach allows models to simulate policy change. Both high non-linearity of pension accrual with age and uncertainty, and the sequential nature of the retirement decision-making process, are effectively captured by the structural approach. With the increasing feasibility of estimating structural models by directly using maximum likelihood techniques, retirement behavior analysis has progressed rapidly from reduced form estimates to a structural model. The structural approach's usefulness depends on the accuracy of the underlying model specification.

Structural models studying retirement can be roughly categorized as static models, option value models, and dynamic programming models. One subgroup of the static models category is the lifetime budget constraint model. The retirement decision is subjected to a life-long view, and retirement is treated in much the same way as leisure. The lifetime budget constraint approach, adopted by Burtless in 1986, inherits the previous view with a standard lifetime labor-leisure budget constraint; however, the budget constraint in his model is discontinuous or kinked with annual earnings replaced by cumulative lifetime compensation. Burtless' publication also incorporates the unforeseen changes in SS benefits in 1969 and 1972, which were overlooked in previous studies. Fields and Mitchell (1984) and Gustman and Steinmeier (1986) extend the general framework by employing the form of indirect utility function and modeling full and partial retirement separately. In their frameworks, households are assumed to have perfectly smooth consumption by borrowing and lending without limit. Based on my findings, Gustman and Steinmeier's model is the first to be able to simulate the "spikes" at ages 62 and 65. The other subgroup is the hazards model, measuring the transition probability of retirement over time (e.g., Hausman and Wise, 1985; Blau, 1994). Hazards models capture the effects caused by recent changes in current variables, such as varying health and market wage rates in a natural way. Unfortunately, hazards models are not grounded in a utility-maximizing theory of behavior and cannot be easily extended to measure the impact of expected future changes on relevant economic variables.

A dynamic programming model is based on recursive representation of a value function. In a dynamic programming frame, retirement is an optimal stopping time on a finite horizon. With its parameters capturing teachers' preference, a dynamic programming model depicts the retirement decision. A well-known disadvantage of a dynamic programming model is the computational cost involved in solving the Bellman Equation. An alternative is to approximate the dynamic programming rule by simplifying the disturbance term's covariance structure to reduce the number of dimensions. Stern (1990) proposes a simulation technique. Berkovec and Stern (1991) assume that the disturbance term follows an extreme value distribution to study transition among full-time work, part-time work, and retirement and estimate the model using the Method of Simulated Moments. They also include an additive individual and a job-specific random effect in their model. Daula and Moffitt (1991) study retention in the military. In their dynamic model, the disturbances are assumed to be normally distributed. Stinebrickner (1998, 2001) designs an approximation method. Rust et al. (2001) and Karlstrom (2004) establish a simplified, two-stage framework for estimation and find that it performs well.

Rust and Phelan (1997) develop a general dynamic programming model. In their model, future health status, health expenditure, working status, and marital status are uncertain. The disturbance terms follow i.i.d. extreme value distributions. Blau and Gilleskie (2006) follow Rust and Phelan (1997)'s approach to study couples' joint employment decisions and add pension plans to their study. French (2005) estimates a partial equilibrium life cycle model of retirement behavior in which health and wages are stochastic. French and Jones (2011) revise aforementioned model by allowing the saving option to smooth consumption. Consequently, individuals continue to work until they are entitled to pension benefits. Van der Klaauw and Wolpin (2008) introduce subjective expectation to the model because "subjective data provide useful information about the decision process in the same way as do objective data." Laitner and Silverman (2012) apply the framework to Consumer Expenditure Survey (CEX). Kopecky (2011) uses an overlapping generational model with an endogenous retirement decision in order to study pension system reform.

Option Value/Peak Value Model

Along with his colleagues, Wise (Stock and Wise (1990), Lumsdaine and Wise (1990), and Lumsdaine, Stock, and Wise (1992, 1994, 1996)) is known for proposing the "option value" model: retirement decisions are posited as a function of the difference between the expected lifetime utility of retirement at the current date and at a future date when one's expected lifetime utility is maximized. Pension incentives are measured by the return to work in the current year, which is relative to retiring at some future optimal date rather than "year-to-year accrual." In other words, the nonlinearity future accrual can represent the fact that individuals consider not only the accrual to the next year but the entire future path of incentives. The option value model incorporates the advantages of the traditional hazards model and considers potential compensation many years in the future as does the nonlinear budget constraint approach. The option value model reveals that changes in key firm pension plan provisions have very substantial effects on retirement, while scheduled changes in SS provisions have modest effects, contradicting the conclusions drawn by Coile and Gruber (2007).

There are limitations to option value models. If wages are endogenous in retirement decisions, that is, correlated in some way with an underlying taste for retirement, then retirement income effects cannot be identified by wage variation. Even though the leisure parameter can be a source of heterogeneity, the limitation could not be mitigated except when the utility function is correctly specified. The method of capturing the heterogeneity that may bias these estimates is proposed by Coile and Gruber (2001) by including a set of flexible controls for earnings directly in a linear regression model. They also augment "option value" to "peak value" so as to capture the option value of continued work even before pension entitlement ages are reached.

In a series of papers by Coile and Gruber (2000, 2001, 2007) introduce a new variable “peak value.” It refers to “the value of continuing to work until the future year when SS wealth is maximized, or the difference between the expected PDV of SS wealth at its highest possible value in the future and the expected PDV of SS wealth if one retires this year.” “Peak value” is an extension of “accrual value” introduced by Hausman and Wise in 1985 that involves looking ahead by more than just one year. The reason for this is discussed in the next section.

When compared with the dynamic model, it is easier to incorporate and compute more flexible correlation assumptions using this category of models. In the meantime, as Lumsdaine et al. (1992) point out, this kind of value is based on the maximum of the expected present values of future stream (whether it is wealth or utility) retirement occurs now versus each potential future age. The stochastic dynamic programming rule considers, instead, the expected value of the maximum of current versus future options. As they claim, “The expected value of the maximum of a series of random variables will be greater than the maximum of the expected values.” Thus, if this difference is large, this kind of value model will underestimate the value of postponing retirement.

1.5 Data Set

Data are critical for empirical studies. For studies of retirement decisions of the U.S. population, the following data sets are commonly employed.

1.5.1 Retirement History Survey (RHS)

RHS was conducted from 1969 until 1979. RHS is a biennial, longitudinal, panel study covering approximately 11,000 individuals from over 8,000 households, the heads of which were between the ages of 58 and 63 in 1969. A series of six follow-up surveys were initiated to obtain information on these persons at two-year intervals through 1979. Detailed information on annual wages rates, potential SS and pension benefits, assets, and retirement age were included. Information such as private pension wealth, health and physical or mental function was not considered. Information on pensions was limited to whether an individual was eligible for a pension upon retirement. Thus, there was not enough information to construct a reliable measure of expected pension benefits.

Most articles in Mitchell and Fields' review adopt RHS (called "LRHS" in their paper). Researchers who study RHS after 1982 include Hausman and Wise (1985), Burtless (1986), Gutsman (1986), Sueyoshi (1989), Blau (1994), and Rust and Phelan (1997). These works by the aforementioned researchers continue to attempt to model SS benefits' potential role in determining retirement. The general strategy followed by those studies is to measure individuals' retirement incentives by extracting information on potential SS benefit determinants (earnings histories) or ex-post benefit levels across individuals from RHS. Then retirement models are estimated as a function of these incentive measures. Even though the modeling techniques differ substantially across studies, the conclusions drawn are fairly similar: SS affects retirement, but the effect is not enough to explain the rapid decrease of old men since 1950 and "spikes" at ages 62 and 65.

1.5.2 Health and Retirement Survey (HRS)

Private pension is an important component of retirement income. However, constrained by data availability, earlier studies reviewed by Mitchell and Field treat private pension as a "dummy variable of industry" or "eligibility dummy." Different from the SS benefit, which is standard statewide, private pension varies across industries and cannot be fully captured by a representative plan or a dummy variable. This deficiency in data is remedied by the availability of HRS since 1992.

HRS updates RHS with a new cohort of older people who will be followed for at least 10 years. Originally, the HRS target population was limited to those born between 1931 and 1941. Since 1998, HRS' objective has been broadened to include information regarding the whole U.S. population over age 50. For practical reasons, new cohorts are added every six years rather than at each wave of data collection. The survey began with an initial cohort of 12,652 individuals from 7,607 households, with at least one household member born between 1931 and 1941. Respondents' spouses were also interviewed. The current representative sample size increases to more than 26,000. Data on income, work, assets, pension plans, health insurance, disability, physical health and functioning, cognitive functioning, and healthcare expenditure are collected every two years. If a respondent is covered by a pension, HRS also requests for a detailed pension plan description from his employer. Beginning in 2012, genetic information from consenting participants is added to its database. HRS has the richest data available from a retirement survey until now, including detailed information on earnings histories and firm pension plans. Not surprisingly, HRS is the most frequently referenced retirement study source in the past decade. Exploration of HRS is still underway. Studies exploring HRS include Gustman and Steinmeier (1999), Friedberg, Leora and Webb (2003), Coile and Gruber (2001, 2007), and Kopecky (2011).

1.5.3 Studies With Survey on Firms

Before the creation of HRS, researchers compiled information on private pensions by focusing on some firms as an alternative to relying on large survey data. Using the information on the firms' pension rules, one may study how private pension plans affect retirement decisions. Back in 1976 and 1979, Burkhauser surveys auto workers in his study and concludes that, "workers retired significantly earlier when they faced a larger decline in their private pension wealth as retirement was postponed." Wise and his colleagues (Stock and Wise (1990), Lumsdaine and Wise (1990), and Lumsdaine, Stock, and Wise (1992, 1994, 1996)) are representations for conducting surveys on firms.

1.5.4 Other Data Sets

Some other data sets contain information on private pensions. Samwick (1998) combines Surveys of Consumer Finances (SCF) (1983, 1986) and the Pension Provider Survey (PPS) to derive pension plan information. He also identifies the limitation of survey data on firms: "First, the sample is not necessarily representative of the working population nearing retirement age, especially those without pensions. Second, personnel records do not contain information on other potential determinants of retirement such as health status, household composition, and wealth. Third, firm reported data could only determine whether an employee left the firm and not whether he left the labor force. The distinction is critical when evaluating the extent to which the growth of pensions has attributed to the decline in labor force participation at older ages."

SCF is a triennial survey of the balance sheets, pension, income, and other demographic characteristics of U.S. families. The survey also gathers information on the use of financial institutions. In total, 4,500 families are interviewed in the main study. However, selection of these families might be biased as study participation was strictly voluntary. The PPS interviewed plan providers for every worker who reported being covered by a pension in the SCF (1983). Only 525 individuals had matching PPS-reported pension information. Using SCF data, Gustman and Steinmeier (1988) thoroughly investigate empirical issues related to pensions. In their research, SCF is compared with incomplete data from other data sets on pensions.

The PPS surveyed households' pension sponsors to obtain a full summary of each plan description. Once either a respondent or the spouse/partner of the respondent reveals that he had some pension coverage from his current job, the SCF interviewer contacts the pension provider(s) for copies of the official summary of plan descriptions, and this information was coded for use with the main survey data.

In a special "notch generation" study, Krueger and Pischke (1992) focus on retirees in the late 1970s and early 1980s. For this cohort, SS benefits are greatly reduced relative to what they would have expected on the basis of the early-to-mid-1970s SS reform. As RHS was discontinued in 1979, and did not include individuals born in the notch cohort, Krueger and Pischke create an aggregate panel data set from the Current Population Survey (CPS) every March from 1976 to 1988. The survey provides a comprehensive body of data on the labor force, employment, unemployment, persons not in the labor force, hours of work, earnings, and other demographic and labor force characteristics. With the natural experiment, the authors find that the growth in SS benefits in the 1970s explains less than one sixth of the decline in the male labor force participation rate observed in that time period. Gruber and Mardrian (1993) connect CPS with Survey of Income and Program Participation to investigate the effect of health insurance on retirement. They claim that "continuation of coverage" mandates significantly increase retirement. Diamond and Gruber (1999) use CPS to study tax effect on SS benefits.

To analyze retirement decisions of a family, Hurd (1990) uses the New Beneficiary Survey (NBS), while Gustman and Steinmeier (2000) rely on information from National Longitudinal Survey of Mature Women (NLS). NBS is sponsored by the Social Security Administration. This national, cross-sectional survey was fielded from October through December 1982, using a sample drawn from the Social Security Administration's Master Beneficiary Record. The sample included retired workers, disabled workers, and aged wives and widows who received a first benefit payment from mid-1980 through mid-1981. NLS is a survey for a joint retirement study. The survey is sponsored by the U.S. Bureau of Labor Statistics and follows the same sample of individuals from specific birth cohorts over time. The survey collected data on labor market activity, schooling, fertility, program participation, health, and other factors. It had a large number of observations on working husbands and wives of retirement age. Its main disadvantage was that it was a choice-based cross section, which limited the scope of the analysis that can be undertaken.

Panel Study of Income Dynamics (PSID) is used by French (2005). PSID is a longitudinal panel survey of American families, conducted by the Survey Research Center at the University of Michigan. It is the world's longest running household panel survey. With over 40 years of data on the same families and their descendants, the PSID measures economic, social, and health factors of the life cycles of families over multiple generations. Data have been collected from the same families and their descendants since 1968.

1.6 Conclusion

This chapter reviews literature explaining two puzzles noted by Mitchell and Fields: the drop in labor force participation since 1950 and the "spikes" in hazards ratio of retirement at ages of 62 and 65. Although the role of more variables, such as private pension and health care, explored as more data, became available in recent years, the two puzzles are still not fully explained.

One reason for the limited success of retirement research is that the variation in benefits is often difficult to measure. Retirement decisions may depend on the prospect of future earnings and other pension benefits that are not easy to measure. The availability of data sets, such as HRS, provides researchers with more useful data for future research. However, the study of the retirement problem of a large population usually requires working with noisy measurements of pension benefits and incomplete information on the prospect of future salary and employment.

For a number of reasons, public school teachers' retirement problem offers a special opportunity to study the effect of pension incentives on retirement decisions. The teacher pension rules cover a large number of teachers in the system. The pension rules are clearly specified and usually depend on measurable variables, such as age and experience. Hence, there are few measurement errors in the pension benefit. In addition, teacher employment and salary are predictable and retirement is voluntary; hence, teacher retirement is not caused by forced separation. Because the measurable pension incentives differ across teachers, researchers can estimate the effects of the pension incentives by studying teachers' retirement patterns with different ages and experience.

Chapter 2

Probit Model on Retirement Decisions of Missouri Public School Teachers

2.1 Introduction

The ongoing trend of transferring from DB plans to DC plans in the retirement in private sectors has fomented the growing voice for reform of the public school teachers' pension system. The first reason to choose public teachers as a target sample in this study is that the group itself is a large population. Teachers' pensions cover 3.2 million full-time-equivalent (FTE) teachers as of fall 2017. Studying retirement decisions made by teachers not only provides a platform to study how monetary and non-monetary factors affect a teacher's decision to leave, it also creates an opportunity for schools to improve teaching efficiency. Second, teachers' pension rules are clearly defined by the pension systems covering a large number of teachers. Many teachers do not participate in the SS program. There are 13 states - including the state of Missouri, which is my primary focus - and the District of Columbia, where teachers are non-participants. Retirement benefits for participants in DC plans and some DB plans - such as SS - are based on their historical earnings. For teachers whose benefit does not include SS, typically the last three years of salary payment determines their retirement benefits. The latter should receive more attention for financial consideration. Last but not least, the administrative panel data of teachers are generally of very high quality when compared with the household survey data used in some other studies.

Many studies find that teachers' pension system impacts retention and pension is a significant component of education budget (e.g., Barro and Buck, 2010). In recent years, there has been mounting pressure to reform DB plans (e.g., Costrell and McGee, 2010). Reliable estimates of the effects of alternative rules are needed to select an appropriate pension policy. To derive accurate estimates, finding a reliable model is a prerequisite and is also the purpose of this study.

As summarized in my first chapter, current popular methodologies currently applied to retirement studies fall under three categories: probit models, option value models, and dynamic programming models. Among these three approaches, the probit model is flexible in adding a new potential explanatory variable and straightforward in demonstrating the variable's contribution to retirement probability. In addition, the probit model is computationally much simpler than the other two models. The dynamic programming model can well capture both the non-linearity of pension accrual with age and experience, as well as the sequential nature of the retirement decision-making process. On the other hand, stochastic dynamic programming problems generally have no analytical form of the likelihood function and therefore, it is hard to find analytical solutions. The option value model, discussed later, is more tractable than the dynamic programming model but more complex than the probit model. The probit model, as a fundamental and popular model, is discussed in this chapter. I will examine its performance in fitting the retirement data of Missouri public school teachers.

For Missouri public school teachers involved in my study, salary is exogenous and can be accurately predicted by salary schedules exclusively related to and highest educational attainment. Pension wealth is also predictable at each age when a teacher decides to leave the workforce following PSRS pension rules.

2.2 Previous Literature

Since Samwick (1998) introduces the idea of an "option value" in his probit model to capture the forward-looking features of retirement, option value related variables are commonly incorporated into the probit model. Coile and Gruber (2001) propose a probit model that includes one of the following financial incentive measures - one-year accrual, option value, and peak value - to test each one's performance. They define the peak value incentive measure as an appropriately discounted difference in expected pension wealth if one retires at a future optimal age versus retiring today (at time t). Their model demonstrates that both the peak value model and the option value model perform much better than accrual-only models. Several subsequent studies use the same approach. For example, Friedberg, Leora, and Webb (2003) adopt the same model with the classification of DB and DC plans, and Friedberg and Webb (2005) explore the role of private pensions further. Blundell et al. (2002) employ the same method to examine retirement in Britain, and Warren and Oguzoglu (2010) expand the analysis of retirement issues to Australia.

In addition to those option value related variables, other explanatory variables that are commonly incorporated into probit models to study retirement decisions include the following: dummy variables of pension coverage, wage, tenure, financial assets, housing net equity, and other demographic variables, such as age, education, and race.

Even though teacher retirement is an important topic, related publications are limited, partially due to restricted available data resources on teacher populations (Costrell and Podgurksy, 2009; Friedberg and Turner, 2010; Ni and Podgurksy, 2016). Taking the probit approach, Clark et al. (2006) focus on the manner in which new hires make the choice between a state DB pension and a DC plan. Costrell and McGee (2010) borrow Coile and Gruber's research method to examine Arkansas teachers' retirement decisions. One-year accrual is incorporated in addition to peak value in their probit model.

2.3 Missouri's Public Teacher Retirement Rules and Data

2.3.1 Missouri's State Teachers' Retirement Rules

Teachers from Missouri public schools, similar to public schools in many other states, are covered by a DB plan. Except for St. Louis and Kansas City, all other teachers in the state of Missouri are under one single retirement system, called the Public School Retirement System (PSRS), which is the focus of my paper. PSRS was established in Missouri in July 1946, while the Public School Retirement Act of Missouri became effective in August 1945. PSRS receives state and local administrative unit contributions annually, as well as contributions from employees, and makes investments. The pension trust amounts to about $43.5 billion in assets as of 2017.

Entitlement to full retirement requires at least one of the following conditions to be met:

1. The individual must be above the age of 60 and have at least 5 years of working experience; or

2. The individual must have over 30 years of working experience; or

3. The sum of the age of the individual and her working years is greater than or equal to 80 ("rule of 80")

Full pension benefits are determined by the following equation:

$$AnnualBenefit = Exper \times Y \times R$$

Where Exper is years of experience up until the beginning of the retirement year; Y is calculated by the average of the three highest consecutive annual salaries; and R is a benefit factor. The factor equals 0.025 when $Exper \leq 30$ and 0.0255 once $Exper \geq 31$. Meanwhile, there is a "25 and out" option, which allows a teacher to retire when she has over 25 years of experience. In particular, R is 0.022 for $Exper = 25$, and the factor increases by 0.0005 for each additional year of experience up to 29 years. Last, if a teacher intends to retire at an earlier age, i.e., $age \leq 60$, an additional discount factor is applied, 0.9101, 0.8293, 0.7566, 0.6909, 0.6315, and 0.5777 for each single year less than age 60.

The first cost-of-living adjustment (COLA) was applied in 1975 when retirees began drawing on their annuities. The January 2018 COLA was 1.63%. Current Missouri law states that COLAs cannot exceed 5% per year, and the dollar amount of the COLAs in a teacher's lifetime cannot exceed 80% of her original monthly retirement benefit. Figure 2.1 and Figure 2.2 depict the historical contribution rates (the contribution rate is the same for both employees and employers) and benefit factors in the PSRS pension calculation formula since the very beginning of PSRS.

2.3.2 Data Description

All the data used in this study are from administrative records of all Missouri public school teachers. Those records are collected annually. Teachers' birth dates, years of experience, gender, salary, teaching specialty, building assignment, SS number, FTE, date of leaving, as well as the percentage of students eligible for free lunch in a school

Figure 2.1: PSRS Contribution Rate

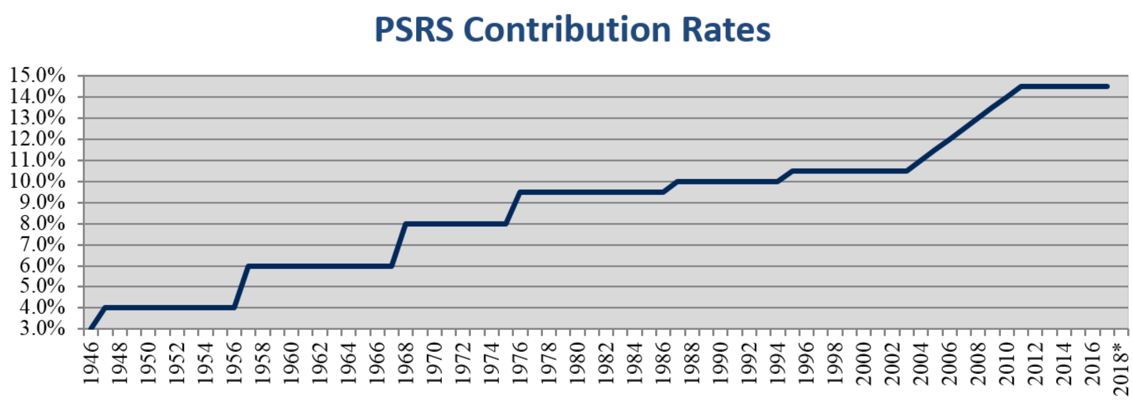
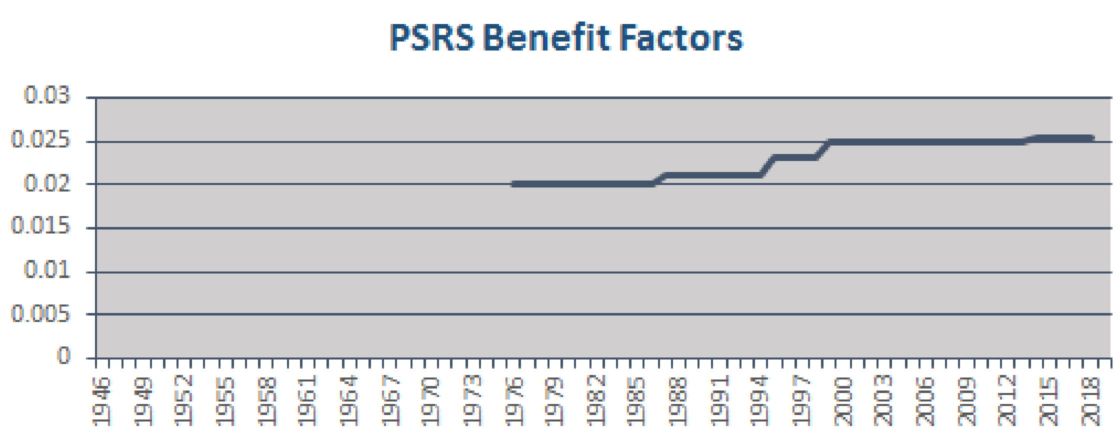


Figure 2.2: PSRS Factor in Retirement Formula



Note: Benefit factor is introduced since 1979.

are observed. Only full-time teachers and teachers who exit the labor force for the first time are studied. If a person who is labeled as "retired" re-enters the labor force, her later observation is not used.

In 2002, a cohort of 9,605 teachers aged 50-55 is tracked from the 2002-03 school year up until the 2007-08 school year. All teachers have continuous records until they leave the work force, resulting in 45,644 teacher-year observations. Hundreds of teachers retire each year. Approximately 4,793 teachers retire over six years, and 4,812 teachers remain in their positions until the last school year. Table 2.1 summarizes those retired teachers' main characteristics by year and those of non-retired teachers in the last year separately. Descriptive statistics of key variables are shown in Table 2.2. The variable "Retired" is a dummy variable for retirement status. When a teacher retires, the variable equals 1; otherwise it equals 0. The average age of retired teachers is approximately 56. Gender equals 1 for male teachers and 0 for female teachers. "Pctlunch" denotes the percentage of students eligible for free lunch in a school. Among those 9,605 teachers, approximately 20% are males, the average work experience is around 24 years and the average "smoothed" salary over the three top paying years before retirement is $49,395.

Table 2.1: Description of Cohort Data

Year	Retired	Age	Experience	Male
2003	640	53.09	28.53	0.29
2004	862	53.58	28.60	0.26
2005	972	54.42	28.10	0.24
2006	822	55.39	27.73	0.21
2007	778	56.39	26.86	0.20
2008	719	57.44	26.37	0.19
Not Retired	4812	57.04	20.61	0.17

Over the six years, roughly half of the teachers in the cohort retire. The retirement distribution by age and experience is shown in Figure 2.3 from two different angles. Most retirement occurs when those teachers reach 30 years of experience and/or age

Table 2.2: Descriptive Statistics for Cohort

Variable	Mean	Standard Deviation	Minimum	Maximum
Retired	0.499	0.50	0	1
Age	56.038	2.204	50	60
Experience	24.149	7.502	5	39
Salary	49395	11906	22494	98944
Gender	0.199	0.399	0	1
Pctlunch (%)	34.637	20.285	0	100
Number of Teachers	9,605			

55. Figure 2.4 depicts the retirement hazard rates (the proportion retiring at each age is conditional on not retiring at the previous age) for the whole sample, male and female teachers alike. Unlike the retirement spikes at ages 62 and 65 that are commonly observed in nationwide survey data sets such as RHS and HRS, Missouri public school teachers demonstrate unique retirement patterns. Generally, the hazard rate of retirement goes up with age and increases relatively rapidly at age 55 and age 60. The only exception is in the case of men; when their age goes from 57 to 59, the hazard rate goes down from its highest point (17.7%) to 13.7% and then decreases slightly to 13.5%. Afterwards the rate climbs up again to 17.8% at age 60. The retirement hazards rate for women nearly follows the same pattern as that it does for the whole population.

2.4 Probit Model

A probit framework created by Coile and Gruber (2001) has been widely employed in empirical research on retirement in recent years. In their model, a new, forward-looking incentive, peak value, is introduced. Together with other control covariates, the peak value variable is intended to capture the heterogeneity in workers' taste. In this chapter, I use their regression framework to examine Missouri public school teachers' retirement decisions. The main difference between their model and mine is

Figure 2.3: Retirement Distribution by Age and Experience

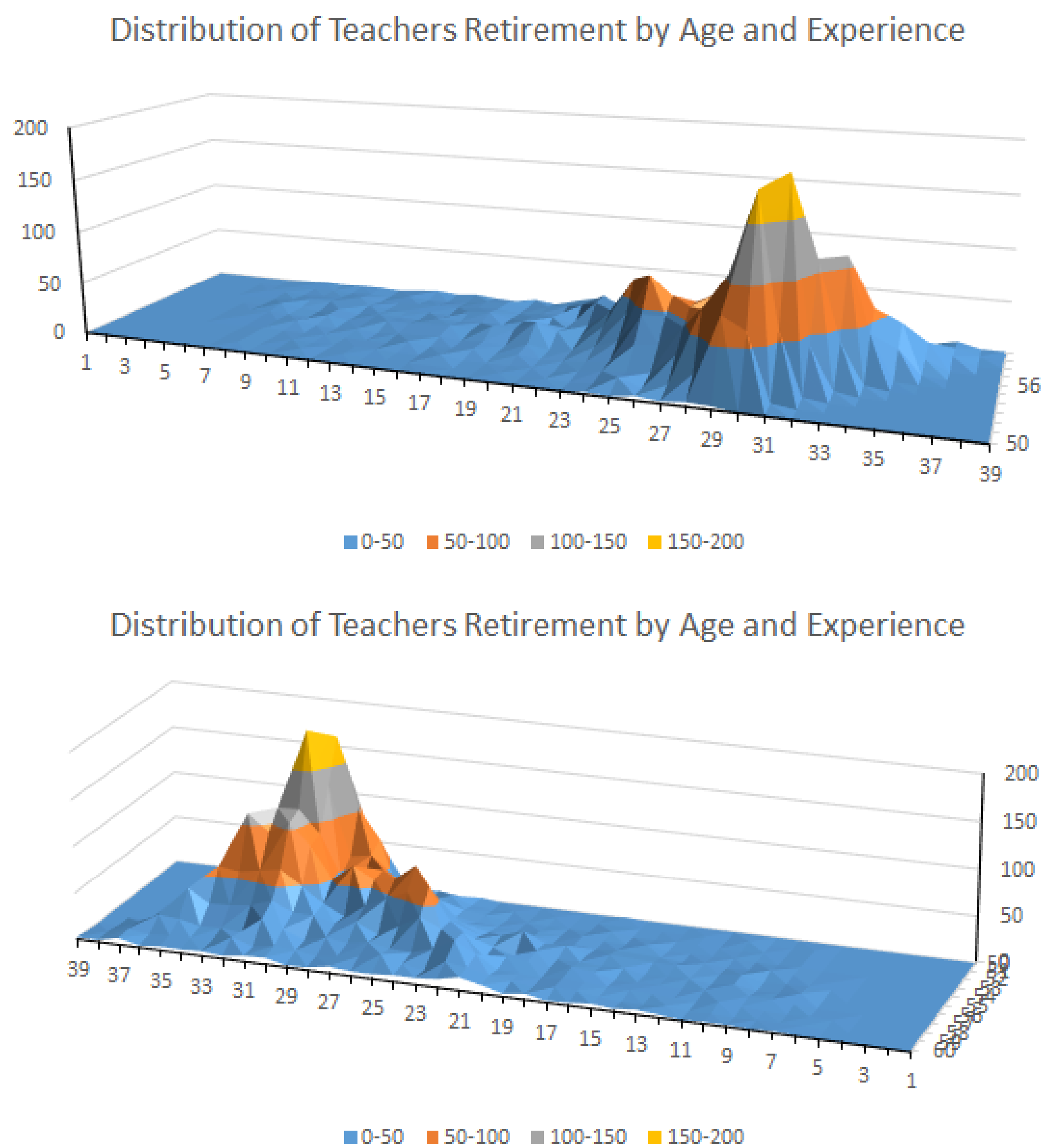
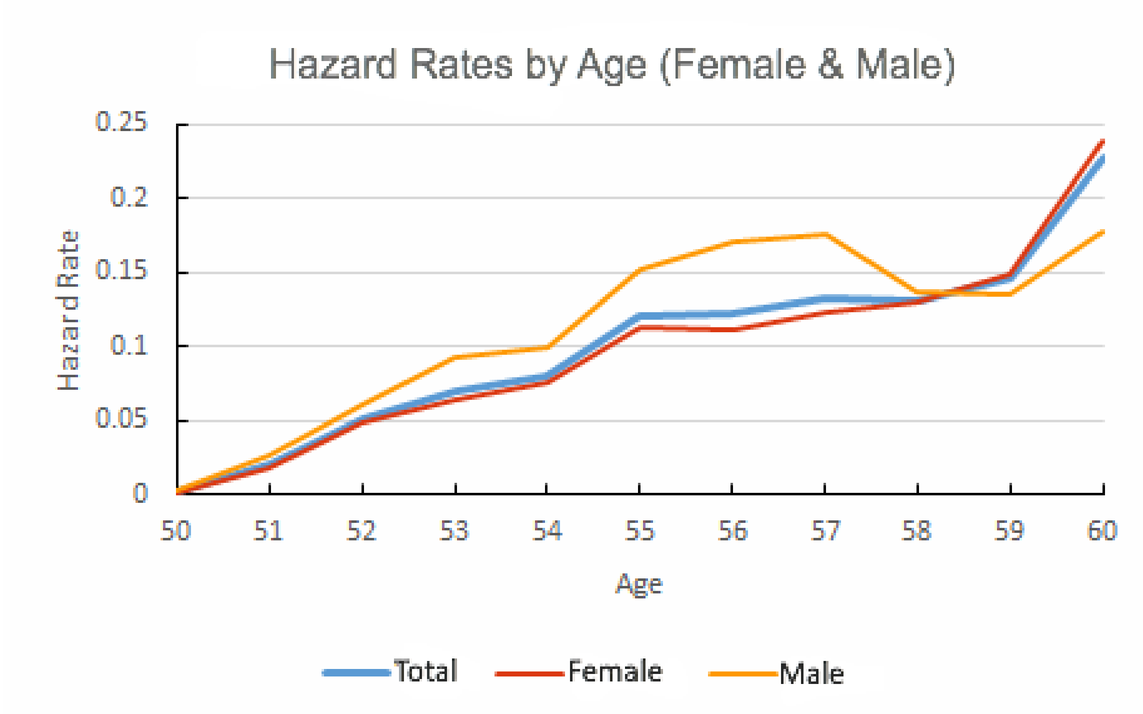


Figure 2.4: Retirement Hazard Rate, Female & Male



that PSRS teachers are not in SS; hence, their pension is the only source of retirement income. Following their definition, "peak value" is the difference between pension at its maximum expected value and current value. If the optimum value of pension is reached, then the peak value would be the difference between the pension received this year and that in the next year. Table 2.3 shows the age pattern for peak value. For the studied age period 50-60, the peak value declines with age.

Table 2.3: Peak Value by Age

Age	Counts	Median	10th%	90th%	Std Dev
50	1847	219444.4	95204.26	344415.34	96080.03
51	3671	202199.95	63304.39	327595.33	99979.46
52	5141	187187.29	49413.3	305979.13	99773.58
53	6245	176039.71	38653.24	288277.11	96691.33
54	6951	161475.3	30880.24	267087.75	91511.05
55	7638	143866.19	21945.79	245037.48	85609.7
56	5605	126891.74	15734.34	218570.03	77264.23
57	3878	106534.1	11097.55	190289.1	67397.52
58	2553	84237.79	7054.28	160970.15	57352.34
59	1462	63371.8	3978.02	127887.7	46228.96
60	653	47596.11	1164.61	99434.45	35675.11

Apart from peak value, the probit model also includes age, experience, wage, rules, and some other characteristics such as explanatory variables. The model specification is expressed in the following form:

$$\begin{aligned} \Pr(retired = 1)_{it} =& \Phi(\beta_0 + \beta_1 PV_{it} + \beta_2 AGE_{it} + \beta_3 EXP_{it} + \beta_4 AGE_{it} \\ & + \beta_5 AGEEXP_{it} + \beta_6 RULE_{it} + \beta_7 SCH_{it}) \end{aligned}, \tag{2.1}$$

where $\Phi(.)$ is the cumulative density of standard normal distribution; PV represents calculated peak value; AGE denotes a teacher's age, which will be tested in different forms; EXP is a quadratic function of a teacher's experience; AGEEXP represents the interaction between age and experience; WAGE is a control for earnings; and RULE is a set of dummies for each eligible condition for retirement; and "SCH" controls school-level variables for that teacher, i.e., free lunch percentage. Different specifications of these explanatory variables will be explored in the next section.

2.5 Regression Results

2.5.1 The Whole Sample

Initially, I exclude school-level information in my model. Table 2.4 provides estimates when age takes different specifications: quadratic form and dummies form for each age from 50 to 60 years old (which is the form taken by Coile and Gruber). The first column lists estimates for the probit model with quadratic form, and the second one is for the one with age dummies. For both forms, all variables are significant, even at the 5% level, and all coefficients are comparable. Judged by log-likelihood value, the model with age dummies performs better than the quadratic form. A negative coefficient for peak value implies a higher peak value and lowers the odds of retirement, which is consistent with Coile and Gruber (2011). Controlling all other variables, retirement probability increases with salary. Both models show a rising pattern of retirement propensity with age. The coefficients for age dummies increase in age. This means an increasing preference for leisure with age. Increased experience also contributes to a higher retirement probability. The coefficient for the male dummy variable demonstrates the relative higher level for males than females when other conditions are the same.

Dummies for each retirement rule are further incorporated to demonstrate the rules' effects. Each dummy variable consequently corresponds to each qualification of full entitlement (age over 60 and experience above 5 years; over 30 years of experience; rule of 80). The estimates are provided in Table 2.5. The baseline is Model 1, which is the same as that in Table 2.4 with quadratic specification of age (dummy variable for Rule 1 is a linear combination of intercept and age dummies. To avoid dropping intercepts and to make covariates comparable, quadratic specification is used for comparison). Model 2 contains three dummy variables for the retirement rule. All estimates are statistically significant, even at the 1% tolerance level. All estimates for

Table 2.4: Retirement Probit with Different Age Specifications (Marginal Effects)

Variable	Quadratic Age	Age Dummies
Constant	-44.2558	-1.2091
	(4.868)	(0.1282)
Peak Value	-0.00000474	-0.00000471
	(0.0000003341)	(0.0000003303)
Wage	0.000005768	0.000005705
	(0.0000009341)	(0.0000009326)
Male	0.0471	0.0467
	(0.022)	(0.022)
Age	1.2892	
	(0.1661)	
Experience	0.5235	0.4942
	(0.0499)	(0.0487)
Age squared	-0.00958	
	(0.00145)	
Experience squared	-0.00075	-0.00077
	(0.000189)	(0.00019)
Age*Experience	-0.00816	-0.00762
	0.000861	(0.000844)
Age 50		-3.0254
		(0.2631)
Age 51		-2.192
		(0.2236)
Age 52		-1.7516
		(0.1974)
Age 53		-1.5498
		(0.1732)
Age 54		-1.407
		(0.1495)
Age 55		-1.0996
		(0.1265)
Age 56		-0.9469
		(0.1062)
Age 57		-0.7382
		(0.0891)
Age 58		-0.6188
		(0.0766)
Age 59		-0.4261
		(0.0717)
N	45,644	45,644
Log-likelihood	-12497.1865	-12447.6355

Notes: The estimates are based on a regression model with two different age specifications; sample data are teachers aged 50-55 in school year 2002-2003; standard errors are in parentheses.

explanatory variables that are common in the two models have the same sign. When compared with the Model 1 estimates, all old explanatory variables' contributions are reduced in Model 2 as those rule dummies partially explain retirement decisions.

Missouri public school teachers data include school level information, percentage of students eligible for free or reduced-price lunch, which might effect retirement decision. However, this variable is insignificant.

2.5.2 Simulation Comparison

Figures 2.5, 2.6, and 2.7 plot actual and simulated retirement distributions by the probit model, including three dummies for pension rules, age, experience, and the combination of age and experience. In all three graphs, the red lines represent actual retirement distribution, and the purple lines refer to simulated distribution by the probit model. For age distribution, the simulated line almost coincides with the line-representing actual distribution, particularly for ages beyond 55. For ages below 55, the probit model still fits well except for two kinks. Similar fit pattern occurs for experience distribution. The probit model fits comparatively better for experience over 31 than it does for experience below 31.

Simulation tends to smooth retirement probability out; therefore, kinks are also ignored in simulated experience simulation. Once age and experience are summed up, there is almost no difference between simulation and actual lines when the sum of age and experience is more than 88 and less than 68. Between the two ages, the simulated line is also smoothed. As a whole, all simulations perform quite well.

2.6 Conclusion

As an important method for studying retirement issues, the probit model excels in its ability and flexibility to incorporate potential explanatory variables. By introduc-

Table 2.5: Retirement Probit with Rule Dummies (Marginal Effects)

Variable	Model 1	Model 2
Constant	-44.2558	-40.2107
	(4.868)	(5.6067)
Peak Value	-0.00000474	-0.00000296
	(0.000000334)	(0.0000003972)
Wage	0.000005768	0.000003084
	(0.0000009341)	(0.000001001)
Male	0.0471	0.0513
	(0.022)	0.0221
Age	1.2892	1.1545
	0.1661	0.1963
Experience	0.5235	0.5149
	0.0499	0.0482
Age squared	-0.00958	-0.00861
	0.00145	0.00175
Experience squared	-0.00075	-0.00208
	0.000189	0.000257
Age*Experience	-0.00816	-0.00731
	0.000861	0.000862
Rule1		0.3404
		(0.0712)
Rule2		0.2933
		(0.0344)
Rule3		0.4087
		(0.0392)
N	45,644	45,644
Log-likelihood	-12497.1865	-12389.3705

Notes: 1. The estimates are based on a regression model with and without dummy variables for retirement rules; sample data are teachers aged 50-55 in school year 2002-2003; standard errors are in parentheses. 2. Rule 1: age is over 60 and working experience is over 5 years; Rule 2: over 30 years of working experience; Rule 3: sum of age and working years equals or exceeds 80 (“rule of 80”)

Figure 2.5: Observed and Simulated Age Distribution of Retirement

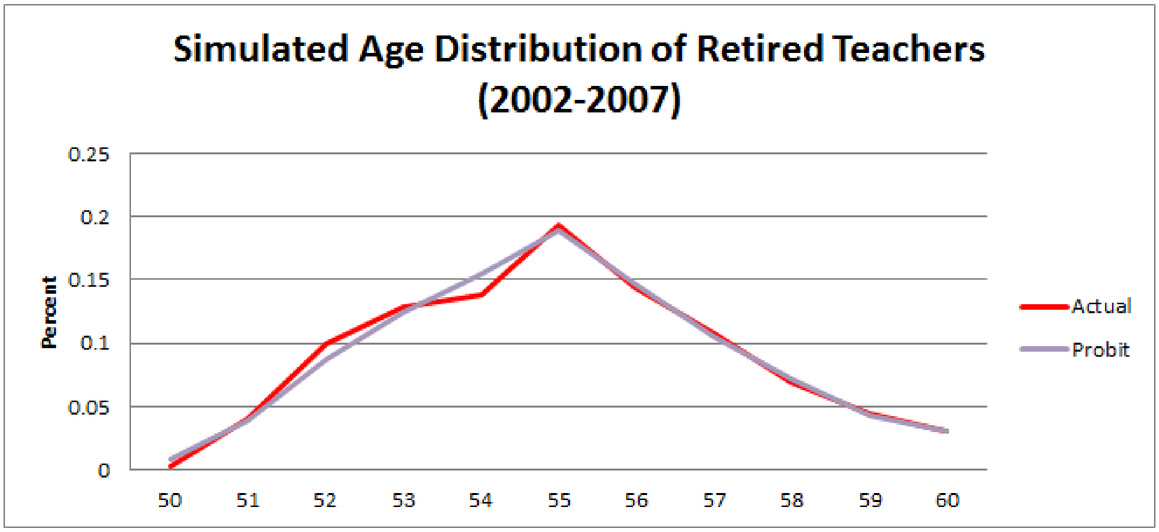


Figure 2.6: Observed and Simulated Experience Distribution of Retirement

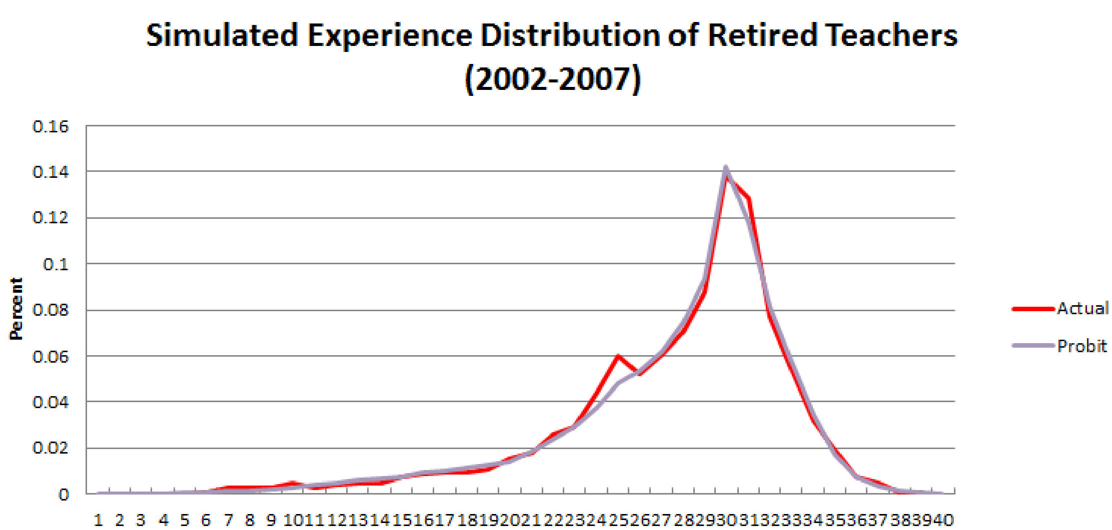
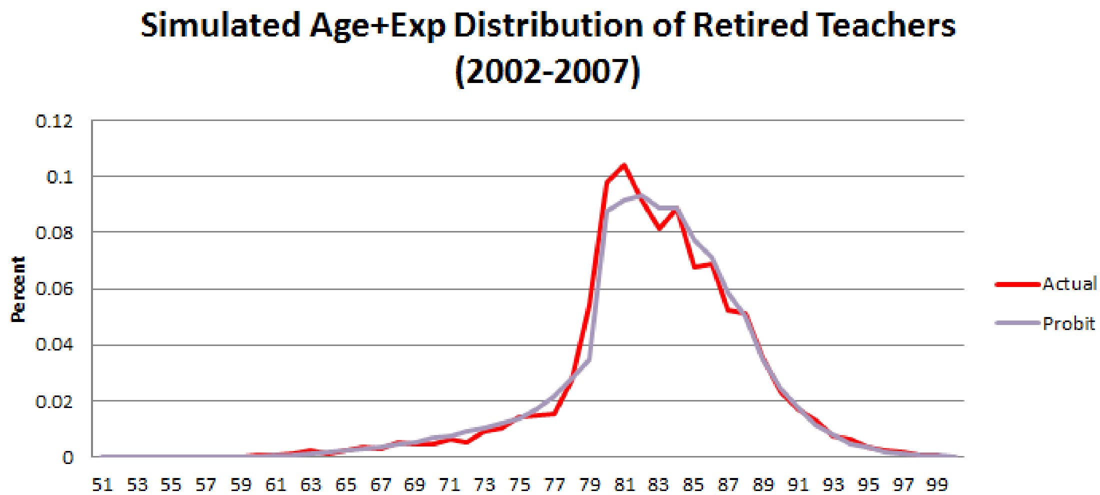


Figure 2.7: Observed and Simulated Age&Experience Distribution of Retirement



ing a new, forward-looking financial incentive, peak value, and incorporating wealth directly into their regression framework, Coile and Gruber (2000, 2001) attempt to separate impact from SS incentive on retirement decision from other pension incentives and wages. The Missouri public school teachers studied in this paper are exclusively entitled to a pension, and wage follows the wage schedule determined by age and experience. Therefore, Coile and Gruber's regression framework is expected to lend itself to selecting out the influence coming from pension on retirement for the sample of Missouri public school teachers.

Judging by the simulated distribution, no matter whether over age, experience, or a combination of age and experience, the probit model fits well with actual retirement distribution except for a few turning points. When compared with enriched large survey data such as HRS, Missouri teachers' pension data are still lacking in some important information that is widely accepted to affect retirement decisions, such as health and spousal information. With more information included in the future, retirement behavior might be better explained.

Chapter 3

Structural Model on Retirement Decision of Missouri Public School Teachers

3.1 Introduction

The previous chapter illustrates that the probit model excels in its simplicity of calculation and straightforwardness in demonstrating the relationship between explanatory variables and probability. However, as the "Lucas critique" implies, estimating "deep parameters" better predicts the effect of a policy experiment. When compared with reduced form models that study responses to some certain pension incentives, structural models estimate hidden parameters governing individual behavior; thereby, the effect of policy change can be more accurately captured. The benchmark model studied in this chapter is called the option value model, introduced by Stock and Wise (1990) (hereafter referred to as SW). The idea is to compare the difference between utility from retirement at the optimal date and utility from retirement today in order to determine when retirement should happen. By comparing the utility values derived from retiring at the two different times, the option value model captures a forward-looking behavior for each individual. The SW model is a structural model because the underlining utility function reflects preference and the parameters in it are independent of pension rules.

The SW option value model differs from the probit model proposed by Coile and Gruber (2001). Coile and Gruber (2001) point out two limitations of the SW option value model. The first one is that most of the variation in the option value comes from variation in wages, instead of retirement income itself as the option value is a function of both pension incentives and wages. Wage variation may be due to occupational heterogeneity. The second one is that the nonlinear structure of the option value fails to isolate SS from other retirement income sources. The responsiveness of retirement to SS and private pension might differ, and only SS is relevant for SS policymaking. Therefore, they put forth two approaches to surmount the aforementioned two limitations. One includes a set of flexible controls for earnings directly in the model to capture sample heterogeneity; the other introduces an alternative forward-looking measure that is not primarily driven by wage differences, "peak value," to reflect variation in retirement income instead of wages.

However, Missouri public school teachers are only entitled to one source of pension from PSRS and without SS. Therefore, it is unnecessary to identify SS from other sources of retirement income. As for the second limitation, teachers' salaries follow a fixed schedule, which is determined by experience. Variation in pension wealth, which is determined by teachers' age and experience. Hence, Coile and Gruber's concerns regarding the option value are not relevant to Missouri public school teachers' retirement. Moreover, PSRS pension rules create a plateau in pension wealth, which makes the peak value hard to define. The option value is suitable for analyzing teachers' retirement decisions.

When compared with the SW paper that studies pension rules for employees in firms, there are several big differences in examining the teachers' pension system.

Firstly, as previously mentioned, teachers' future salaries can be accurately predicted salary schedules exclusively by experience. On the contrary, employees' future salaries in firms are difficult to predict. The econometrician's prediction of salary data may differ from the workers' expectations. This would result in model misspecification. Secondly, there may exist selection bias in the sample used for salary prediction. For example, employees with longer experience might move to other firms, while the remaining employees with less experience could underestimate salary for experienced workers. In contrast, the clearly defined pension rules and salary table for Missouri public school teachers make analyzing their pension incentives relatively simple.

In the option value model, the Maximum Likelihood Estimation (MLE) is applied to estimate parameters in the model. If the objective function has multiple dependent variables, it is costly, or even impossible, to directly solve a high-dimensional integration. The frequency simulator is one alternative to approximate the likelihood. To ensure consistency of the estimated likelihood, the number of Monte Carlo simulations is required to run to infinity in theory. To ensure consistency of estimation, the sample size needs to be large; however, running one round of evaluation of likelihood is time-consuming. Hajivassiliou et al. develop a more efficient simulation algorithm than the frequency simulator in a series of papers in 1990s (e.g., Borsch-Supan and Hajivassiliou, 1994). With the new simulator, known as the GHK simulator, the required number of Monte Carlo simulations can be reduced significantly.

Once equipped with an algorithm to evaluate the likelihood function for any given parameter, the next step is to find an algorithm to search for parameter values that maximize the likelihood function. Finding a global maximum is for odd-shaped likelihood functions. If the function is well-conditioned (e.g., unimodal), it is not difficult to find the optimum; however, for multimodal likelihood functions with ridges and plateaus, traditional algorithms such as those for gradient evaluation do not lend themselves to detecting the optimum. The algorithm might not converge under the limited convergence condition. Even if it does, the resultant optimum found is not guaranteed to be global. Therefore, verification of optimum is necessary. A method called simulated annealing (SA) was introduced by Goffe, Ferrier and Rogers (1994) to find an optimum and to do a robustness check for the global optimum. This study seeks to apply an efficient simulator and a robust estimation method to obtain MLE on a large sample. Comparing alternative models and simulating pension policies can also be efficiently and reliably performed and evaluated with the GHK simulator and SA.

3.2 The Option Value Model

The option value model captures the ever-changing and forward-looking behavior of an employee. The option value model can be interpreted as a suboptimal solution of dynamic programming models. The biggest difference between the two lies in the manner in which they treat uncertainty. The option value rule compares the utility of retiring now with the maximum value of expected utilities at a future retirement time, while the dynamic programming compares the expected value of the maximum utilities of current versus future options. The dynamic programming approach considers the value of future options, while the option value model does not. However, the option value model allows for easier derivation of the likelihood than dynamic programming.

"Option value" refers to the value of the optimal time to retire. A teacher has two options each year: continuing teaching or retiring. According to the option value model, the choice is made on the basis of a comparison of the expected discounted utility of the rest of her life for the two options. Suppose year r is the first year after retirement, the expected discounted utility from present time (t) until the last receiving pension date S (upper bound of life) is derived from two parts. The first part is gotten from wage earnings while teaching, which is, $\sum_{s=t}^{r-1} \beta^{s-t} U_w(Y_s)$. The other part is derived from discounted retirement benefits after retirement, expressed as $\sum_{s=r}^{S} \beta^{s-t} U_r(B_s(r))$. The total weighted discounted utility received over the remaining life ($V_t(r)$) is as follows:

$$V_t(r) = \sum_{s=t}^{r-1} \beta^{s-t} U_w(Y_s) + \sum_{s=r}^{S} \beta^{s-t} U_r(B_s(r)). \tag{3.1}$$

Furthermore, suppose utility has a constant relative risk aversion form, with additive individual disturbance terms distributed independently of income and age. Then, the total expected discounted utility equals the following:

$$\mathrm{E}_t V_t(r) = \mathrm{E}_t \{ \sum_{s=t}^{r-1} \beta^{s-t} [(k_s Y_s)^\gamma + \omega_s] + \sum_{s=r}^{S} \beta^{s-t} [(B_s)^\gamma + \xi_s] \}, \tag{3.2}$$

where γ is a parameter for risk aversion; k_s captures disutility of labor; and ω and ξ are independent random variables following AR(1). If retirement happens next year, then the expected utility fully comes from pension benefits received at each year after retirement:

$$\mathrm{E}_t V_t(t) = \mathrm{E}_t \sum_{s=t}^{S} \beta^{s-t} [(B_s(r))^\gamma + \xi_s]. \tag{3.3}$$

If there is any $r > t$ that makes $\mathrm{E}_t V_t(r) < \mathrm{E}_t V_t(t)$, the teacher will retire at current time t. Let $r^* > t$ be the future year when $\mathrm{E}_t V_t(r)$ reaches the highest expected value (the value is called option value), that is to say:

$$r^* = \underset{r \in \{t+1.t+2,\ldots,S\}}{\arg\max} \; \mathrm{E}_t V_t(r). \tag{3.4}$$

The expected gain at year t, from postponing retirement to year r^* is given by the following:

$$G_t(r^*) = \mathrm{E}_t V_t(r^*) - \mathrm{E}_t V_t(t). \tag{3.5}$$

If $G_t(r^*) \leq 0$, it implies that there is no expected gain from continued teaching and that the teacher will retire now. Otherwise, she postpones her retirement. In particular, $G_t(r^*)$ can be expressed as follows,

$$\begin{aligned}
G_t(r^*) &= \mathrm{E}_t \sum_{s=t}^{r^*-1} \beta^{s-t}[(k_s Y_s)^\gamma + \omega_s] + \mathrm{E}_t \sum_{s=r^*}^{S} \beta^{s-t}[(B_s(r^*))^\gamma + \xi_s] \\
&- \mathrm{E}_t \sum_{s=t}^{S} \beta^{s-t}[(B_s(r^*))^\gamma + \xi_s] \\
&= \sum_{s=t}^{r^*-1} \beta^{s-t}\pi(s|t)[\mathrm{E}_t(k_s Y_s)^\gamma] + \sum_{s=r^*}^{S} \beta^{s-t}\pi(s|t)\mathrm{E}_t[B_s(r)]^\gamma \\
&- \sum_{s=t}^{S} \beta^{s-t}\pi(s|t)[E_t(B_s(r^*))^\gamma] + \sum_{s=t}^{r^*-1} \beta^{s-t}\pi(s|t)\mathrm{E}_t(\omega_s - \xi_s) \\
&\equiv g_t(r^*) + K_t(r^*)v_t,
\end{aligned} \tag{3.6}$$

where $\pi(s|t)$ is the expected mortality rate for time s at time t, $g_t(r^*) = \sum_{s=t}^{r^*-1} \beta^{s-t}\pi(s|t)[E_t(kB_s(r^*))^\gamma] + \sum_{s=r^*}^{S} \beta^{s-t}\pi(s|t)[E_t kB_s(r^*)]^\gamma - \sum_{s=t}^{S} \beta^{s-t}\pi(s|t)[E_t(kB_s(r))^\gamma]$; $K_t(r) = \sum_{s=t}^{r^*-1} \beta^{s-t}\pi(s|t)$ aggregates all pre-determined components and $v_t = E_t(\omega_s - \xi_s)$ is a stochastic term. $K_t(r^*)$ cumulates all deflators that yield present value at time t of the future expected values of the random components of utility. The further r^* is in the future, the larger $K_t(r^*)$ is. Intuitively, people feel more uncertain about the distant future. As both ω and ξ are assumed following the AR(1) process, v_t can be treated as follows:

$$v_s = \rho v_{s-1} + \varepsilon_s, \varepsilon_s \ i.i.d. N(0, \sigma_\varepsilon^2). \tag{3.7}$$

$G_t(r^*) \leq 0$ is thereby equivalent to $g_t(r^*)/K_t(r^*) \leq -v_t$.

Salary is assumed to be predictable under an estimated nonlinear (a third-order polynomial) function of working experience.[1] The option value model can be regarded

[1] This is a rational assumption for teachers whose salaries are administrative. Experience is the only factor determining their salaries.

as a special "probit model" with a nonlinear expression: when $g_t(r^*)/K_t(r^*) \leq -v_t$ for a teacher, she chooses to retire now (at time t), an index variable d_{it} equals 1; 0 otherwise. Once the teacher retires, she drops out of the sample pool. If we only use the data of time t, the likelihood distribution for the whole sample size I is given by the following:

$$L(\gamma,\kappa,\beta,\sigma,\rho|\mathrm{X}) = \prod_{i=1}^{I} \Phi(\frac{g_t(r_t^*)}{K_t(r_t^*)}/\sigma_v)^{d_{it}}(1-\Phi(\frac{g_t(r_t^*)}{K_t(r_t^*)}/\sigma_v))^{1-d_{it}}, \tag{3.8}$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal. The actual estimate is based on the sample of multiple years. Assuming that the observation period is n if a teacher does not retire over the whole period, then the probability of not retiring from t till $t+n$ is the probability of a joint event:

$$\begin{aligned}
&\Pr[R=t+n] = \Pr[g_t(r_t^*)/K_t(r_t^*) > -v_t, ...,\\
&g_{t+n-1}(r_{t+n-1}^*)/K_{t+n-1}(r_{t+n-1}^*) > -v_{t+n-1},\\
&g_{t+n}(r_{t+n}^*)/K_{t+n}(r_{t+n}^*) > -v_{t+n}]\\
&= \Pr[g_{t+n}(r_{t+n}^*)/K_{t+n}(r_{t+n}^*) > -v_{t+n}|g_t(r_t^*)/K_t(r_t^*) > -v_t, ...,\\
&g_{t+n-1}(r_{t+n-1}^*)/K_{t+n-1}(r_{t+n-1}^*) > -v_{t+n-1}]\\
&\cdot \Pr[g_{t+n-1}(r_{t+n-1}^*)/K_{t+n-1}(r_{t+n-1}^*) > -v_{t+n-1}|g_t(r_t^*)/K_t(r_t^*) > -v_t, ...,\\
&g_{t+n-2}(r_{t+n-2}^*)/K_{t+n-2}(r_{t+n-2}^*) > -v_{t+n-2}]\\
&\cdot ... \cdot\\
&\cdot \Pr[g_{t+1}(r_{t+1}^*)/K_{t+1}(r_{t+1}^*) > -v_{t+1}|g_t(r_t^*)/K_t(r_t^*) > -v_t]\\
&\cdot \Pr[g_t(r_t^*)/K_t(r_t^*) > -v_t]
\end{aligned} \quad . \tag{3.9}$$

Similarly, the probability of retiring at time $t+n$ is

$$
\begin{aligned}
&\Pr[R=t+n] = \Pr[g_t(r_t^*)/K_t(r_t^*) > -v_t, ..., \\
&g_{t+n-1}(r_{t+n-1}^*)/K_{t+n-1}(r_{t+n-1}^*) > -v_{t+n-1}, \\
&g_{t+n}(r_{t+n}^*)/K_{t+n}(r_{t+n}^*) < -v_{t+n}] \\
&= \Pr[g_{t+n}(r_{t+n}^*)/K_{t+n}(r_{t+n}^*) > -v_{t+n} | g_t(r_t^*)/K_t(r_t^*) > -v_t, ..., \\
&g_{t+n-1}(r_{t+n-1}^*)/K_{t+n-1}(r_{t+n-1}^*) > -v_{t+n-1}] \\
&\cdot \Pr[g_{t+n-1}(r_{t+n-1}^*)/K_{t+n-1}(r_{t+n-1}^*) > -v_{t+n-1} | g_t(r_t^*)/K_t(r_t^*) > -v_t, ..., \\
&g_{t+n-2}(r_{t+n-2}^*)/K_{t+n-2}(r_{t+n-2}^*) > -v_{t+n-2}] \\
&\cdot ... \cdot \\
&\cdot \Pr[g_{t+1}(r_{t+1}^*)/K_{t+1}(r_{t+1}^*) < -v_{t+1} | g_t(r_t^*)/K_t(r_t^*) > -v_t] \\
&\cdot \Pr[g_t(r_t^*)/K_t(r_t^*) > -v_t]
\end{aligned}
. \qquad (3.10)
$$

Denote $\boldsymbol{v_{t,t+n}} = (v_t, ..., v_{t+n})'$, the covariance of $\boldsymbol{v_{t,t+n}}$ is given by:

$$
\Sigma_{\boldsymbol{v}} = \frac{\sigma^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-3} & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-4} & \rho^{n-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \cdots & 1 & \rho \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & \rho & 1 \end{pmatrix} \qquad (3.11)
$$

Thus, the log-likelihood function for not retiring till time $t+n$ be expressed as follows:

$$
l(\gamma, \kappa, \beta, \sigma, \rho | \mathrm{X}) = \sum_{i=1}^{I} \log \Pr_i(\boldsymbol{v_{t,t+n}} \in A_i) = \sum_{i=1}^{I} \log \int_{A_i} \phi(\boldsymbol{v_{t,t+n}}) d\boldsymbol{v_{t,t+n}}, \qquad (3.12)
$$

where the supporting set A_i is $\{g_t(r_t^*)/K_t(r_t^*) > -v_t, ..., g_{t+n-1}(r_{t+n-1}^*)/K_{t+n-1}(r_{t+n-1}^*) > -v_{t+n-1}, g_{t+n}(r_{t+n}^*)/K_{t+n}(r_{t+n}^*) > -v_{t+n}\}$, $\phi(\cdot)$ denotes the multivariate normal density function for the column vector $\boldsymbol{v_{t,t+n}}$, that is to say, $N(0,\Sigma)$. Likewise, the supporting set A_i for the log-likelihood function of retiring at time $t+n$ equals $\{g_t(r_t^*)/K_t(r_t^*) > -v_t, ..., g_{t+n-1}(r_{t+n-1}^*)/K_{t+n-1}(r_{t+n-1}^*) > -v_{t+n-1}, g_{t+n}(r_{t+n}^*)/K_{t+n}(r_{t+n}^*) < -v_{t+n}\}$.

3.3 Data

Data for analysis in this chapter are the same as those applied in the probit model in chapter 2. At the beginning of 2002, the age and experience combinations of target teachers are illustrated in figure 3.1 from two different angles. As the lowest working age is set at 20, the gap between age and experience must be over 20. The distribution spans the age range of 50-55 and experience of 1-35 years. The distribution is relatively even, except for the area of over 25 years of experience. The highest frequency level of 2.65% lies in the cell of age 50 and experience of 28 years. Figure 3.2 depicts the retirement ratio in every cell of age and experience by year. Generally, the retirement ratio does not exceed 0.6. Before the number of years of experience reaches 25, the retirement ratio is quite low. After that point, the retirement ratio surges. Figure 3.3 demonstrates the non-retirement rate in the last year (2007-2008) from two perspectives.

Figure 3.1: Distribution of All Teachers in 2002 from Two Angles

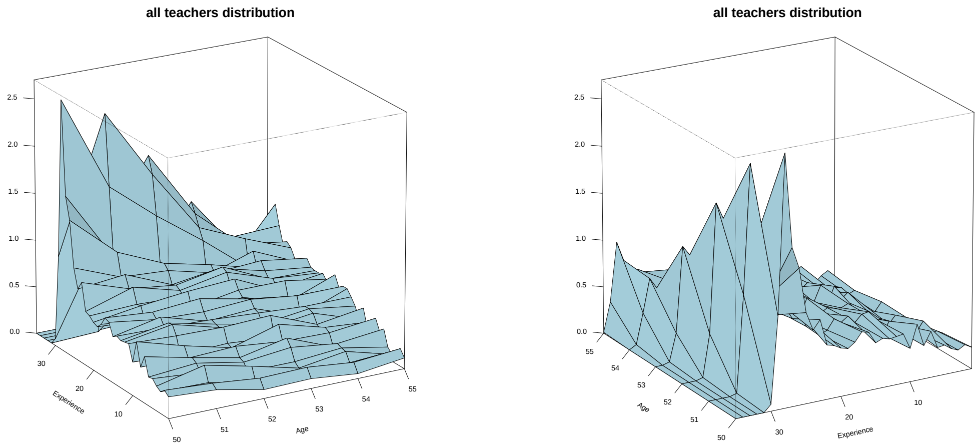


Figure 3.2: Distribution of Teachers Retirement Probability in Each Year

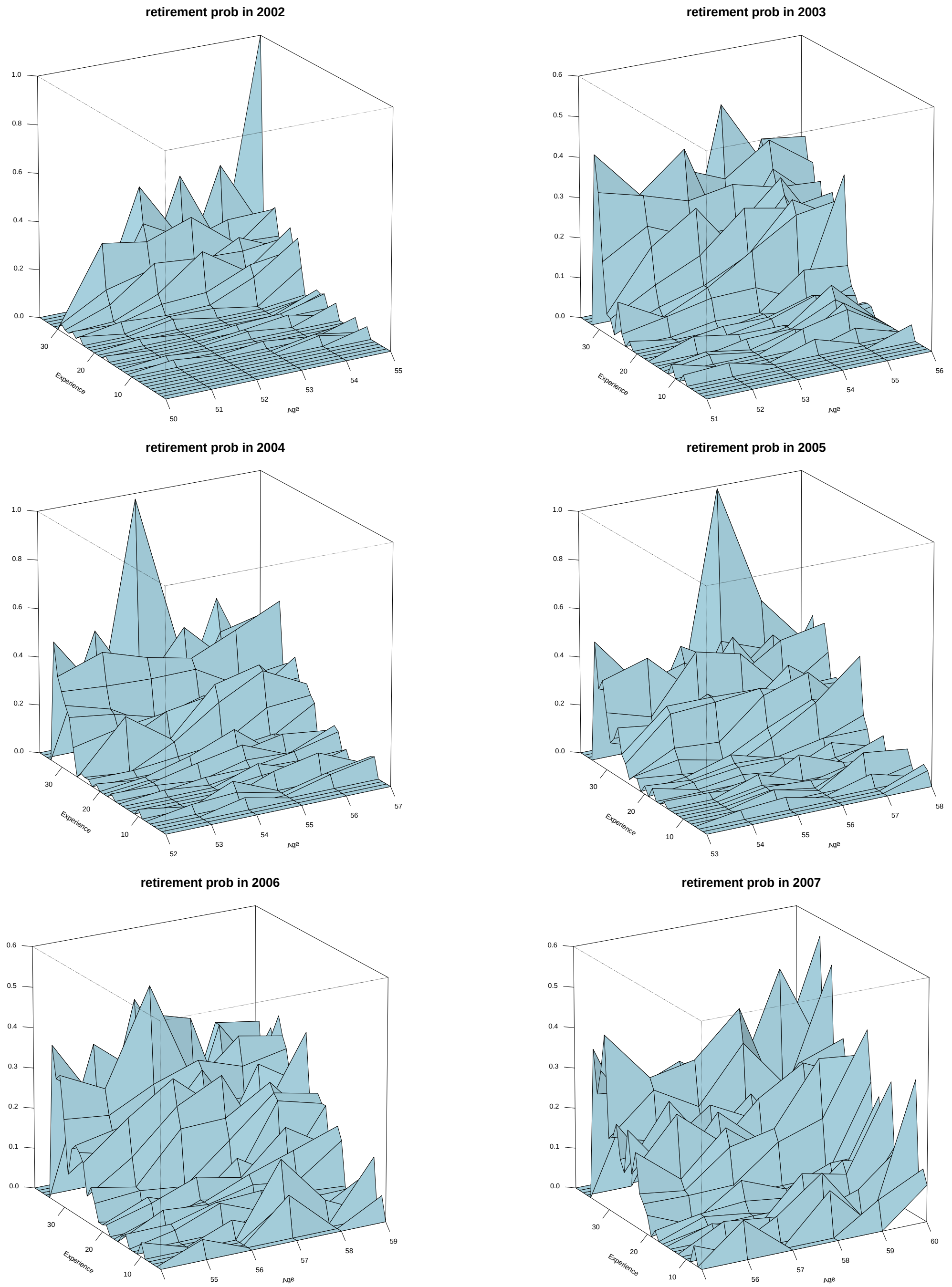
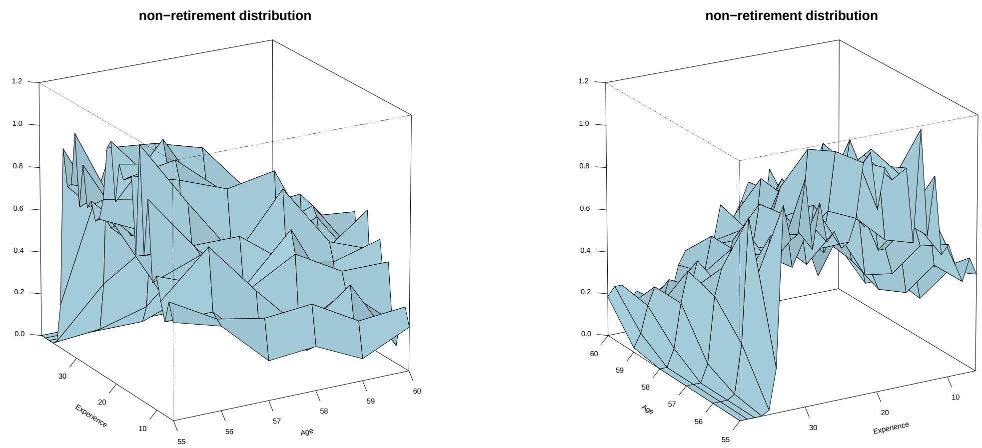


Figure 3.3: Distribution of All Teachers not Retired in Year 2007 from Two Angles



3.4 Estimation Method

3.4.1 Computation of Likelihood by the Frequency Simulator

Directly maximizing the joint likelihood function for a multinomial model involves multidimensional integration, $\int\limits_{A_i} \phi(\boldsymbol{v_{t,t+n}}) d\boldsymbol{v_{t,t+n}}$, which is prohibitively costly to compute. A commonly applied simulation approach uses the frequency simulator through a Monte Carlo simulation, which is also applied in the SW model. The detailed steps are as follows:

1. Decompose variance matrix Σ by Cholesky method into LL'.

2. Draw a series of vector $\boldsymbol{e}^{(j)}, (j = 1, ...n)$ from a standard multivariate normal distribution $N(0, I)$.

3. Then multiply the above two terms: $\boldsymbol{v_{t,t+n}}^{(j)} = L\boldsymbol{e}^{(j)}$. The average value $\frac{1}{n}\sum\limits_{j=1}^{n} I(\boldsymbol{v_{t,t+n}}^{(j)} \in A_i)$ can be used to approximate the probability $\log \int\limits_{A_i} \phi(\boldsymbol{v_{t,t+n}}) d\boldsymbol{v_{t,t+n}}$, where I is an index function.

This simulation method involves taking the average of counts in the third step; therefore, it is named as the "frequency simulator" by Börsch-Supan and Hajivas-siliou

(1994). In the SW model, five parameters, $\kappa, \beta, \gamma, \sigma$ and ρ are estimated after the aforementioned three steps. κ represents the value of work in contrast to leisure time, β denotes the rate of time preference for teachers, γ shows a teacher's risk aversion tendency, σ is the variance of error term, and ρ refers to the covariance between two consequent error terms.

There are many different model specifications worth studying. First, as Stock and Wise (2000) point out, to capture the increasing value of retirement, κ_s can be specified as $\kappa(\frac{60}{age})^{\kappa_1}$, where $\kappa 1$ is another coefficient for disutility decreasing with age. When $\kappa_1 = 0$, κ_s collapses to a constant. Second, as Furgeson, Strauss and Vogt (2006) indicate, work disutility might be a U-shaped function with age. The reason is that, "younger teachers have a higher disutility of work because of a desire to have children," and elder teachers tend to retire due to health issues. In this case, the coefficient is assumed to be $\kappa_s = \kappa(400/(age-40)^2)^{\kappa_1}$. Finally, when γ equals 1, the objective function turns out to be based on dollar value instead of utility value.

Table 3.1 reports estimated results for each parameter in different models. All models are estimated separately for 10,000 Monte Carlo draws (N). Model 1 is the basic SW model with five parameters (no κ_1, κ_s is constant); Model 2 refers to the U-shaped utility case; and Model 3 is the general case in the SW paper, where κ_s decreases with age. In the second model, the new discount factor is actually $(20/(age-40))^{2\kappa_1}$. Thus, κ_1 should be multiplied by 2 to compare with that in Model 3. None of the parameters in the first three models differ much, except for variance σ, suggesting that the variance σ is sensitive to model setting. The value of "Log-likelihood" is the sum of the log value of estimated probability across all sampled teachers throughout all six years. The lower the absolute value is, the better the model fit. When γ is set as the fixed value 1, as shown in Model 4, the absolute value of log-likelihood increases by approximately 50%, and the estimate of σ almost goes up by 60% when compared with that from Model 3. To further probe key

Table 3.1: Parameter Estimates on the Basis of Retirement Decisions in Three Models

Model	κ	κ_1	β	γ	σ	ρ	Log-likelihood
1	0.698		0.964	0.742	7198.523	0.553	-12909.625
2	0.682	0.948	0.947	0.59	1501.472	0.443	-12761.893
3	0.623	2.271	0.960	0.689	3018.812	0.540	-12825.600
4	0.470	4.038	0.968	1.000	45245.322	0.886	-19671.820
5	1.000	2.568	0.950	1.000	96988.844	0.597	-13224.510
6	0.470	4.038	0.968	1.000	96624.603	0.886	-17135.352

parameters, more parameters are fixed, while the number of parameters stays the same as that in Model 3. In Model 5, γ, β, and κ are assigned as 1.0, 0.95 and 1.0 respectively. In Model 5, the estimated σ almost doubles even though the absolute value of log-likelihood decreases by around 50% when compared with that in Model 4. In the last model, all parameters remain the same as the results derived in Model 4 except for σ, then σ also almost doubles and the absolute value of log-likelihood decreases slightly.

When σ is large, the model is less informative. Roughly speaking, the probability that a teacher retires in period t ($G_t(r) \leq 0$ for all $r > t$) is $Prob(\frac{g_t(r^\dagger)}{K_t(r^\dagger)} \leq -\nu) = \Phi(\frac{g_t(r^\dagger)}{\sigma_\nu K_t(r^\dagger)})$, where $\nu \sim N(0, \sigma_\nu^2)$. If you replace the values of other parameters, then $\sigma_\nu = \sigma/\sqrt{(1-\rho^2)} = 45245/\sqrt{(1-0.886^2)} \approx 97,576$. Suppose $r_t^\dagger = 1$, then $K(1)$ is approximately 1. The denominator of the function in Φ is about 10,000. Therefore, the probability of $Prob(\frac{g_t(r^\dagger)}{K_t(r^\dagger)} \leq -\nu)$ should be 0.5 in almost all cases. If the model is noninformative, the probability of retiring should be around 0.5 for $n = 1$. At the other extreme, where σ is close to 0, the likelihood of not retiring reaches 0, which means that every teacher will retire at the same time. The simulated probability of retiring for the first 20 observations in the last round of draws by the computer is shown in Table 3.2. As observed, it is true that the probability is around 0.5 in Model 4, when the variance of error term is high. In Model 3, the probability varies for each individual. Does that mean that the SW model is a good fit for some teachers but not for everyone? Naturally, the next step is to break down the non-conditional

Table 3.2: Part of Simulated Probability of Retiring under Different Parameters

No fixed parameters (Model 3)	$\gamma = 1$(Model 4)
0.0007	0.5054
0.2898	0.4989
0.0038	0.4990
0.3444	0.4962
0.2689	0.5020
0.3538	0.4965
0.0131	0.5041
0.3956	0.4922
0.4051	0.5041
0.3584	0.5016
0.2810	0.5080
0.3720	0.4953
0.4108	0.5090
0.2719	0.4976
0.4249	0.5041
0.2786	0.4995
0.3033	0.4990
0.3565	0.4944
0.2854	0.5002
0.2921	0.4935

probability of retirement by year to conditional probability for each sample.

The average probability of retiring/non-retiring for all sampled teachers across each year is summarized in Table 3.3 (still take estimates in Model 3). The last number in each row represents the average probability of retirement in previous years. For example, for teachers who retired in 2003, the average conditional probability of not retiring in school year 2002-2003 is 0.748, and the average probability of retiring in 2003-2004 is 0.313. In the last row, the numbers refer to the average conditional probability of not retiring by year for teachers who continue to work until year 2007.

There are several interesting findings worth pointing out. Firstly, theoretically, the actual conditional probability of not retiring is 1 for each teacher in every previous year before retiring, and the actual conditional probability of retiring is also 1 in the last retiring year. The estimated average probability gained from the model is no less

Table 3.3: Simulated Average Probability of Retiring/Non-retiring by Year

Retiring Year	year					
	2002	2003	2004	2005	2006	2007
2002	0.308					
2003	0.748	0.313				
2004	0.799	0.749	0.314			
2005	0.842	0.794	0.741	0.321		
2006	0.878	0.838	0.790	0.743	0.316	
2007	0.902	0.868	0.823	0.779	0.730	0.326
Non-ret	0.955	0.938	0.914	0.886	0.853	0.813

than 0.7 for each year. The fit of non-retiring is accurate overall. Conversely, the estimated conditional probability of retiring is only approximately 0.3, which is much lower than the true value of 1. At any age and experience, the observed fraction of teachers retiring in any year is no larger than 0.3. The low predicted probability is consistent with the observed fraction of those retiring in any year. In other words, the model does well in fitting the fraction of retirement among teachers with a given age and experience level. However, it does not predict the retirement of a given teacher. The large value of variance σ also reveals the presence of too much noise in the model. The situation can be improved in two ways: richer data, including more teacher characteristics such as marital status, health information, and spouse information and a richer model with more variables. Secondly, the probability of non-retiring teachers increases vertically in the table, which is reasonable, considering that teachers who are more inclined to retire tend to do so in earlier years. The remaining teachers are those who are relatively less likely to retire. Thirdly, when converting retirement probability (the last number in each row value in Table 3.3) to non-retirement probability, the numbers in each row decrease from left to right. This implies that the tendency to retire increases with age for a teacher.

A comprehensive and straightforward way of demonstrating estimated probability is to draw those simulations across each age-experience cell. The following figures present the distribution by year in detail. In the years prior to retirement, most estimated non-retirement probabilities are close to 1, although some are approximately 0.6. Distribution graphs of the estimated non-retirement probability in previous working years of the group of teachers who retire in the same year have a similar pattern; however, the top surface tends to lower each year as teachers increasingly become likely to retire as they age. The estimated retirement rate is less than 0.6 except in 2007, where the ratio is slightly above 0.6. The top surface rises by year in the series of distribution graphs of estimated retirement probability. Figure 3.3 depicts the estimated probability of not retiring from year 2002 until year 2007, where the top surface also goes down by year.

Figure 3.4: Estimated Probability Distribution in 2002

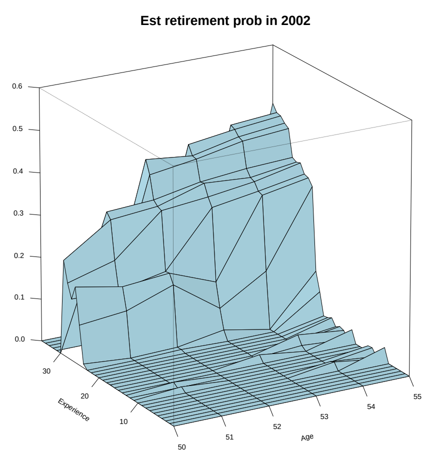


Figure 3.5: Estimated Probability Distribution in 2003

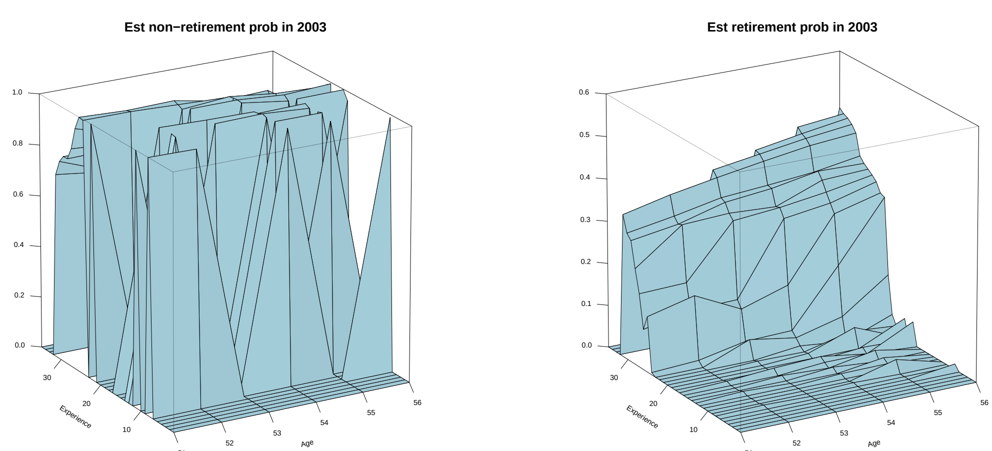


Figure 3.6: Estimated Probability Distribution in 2004

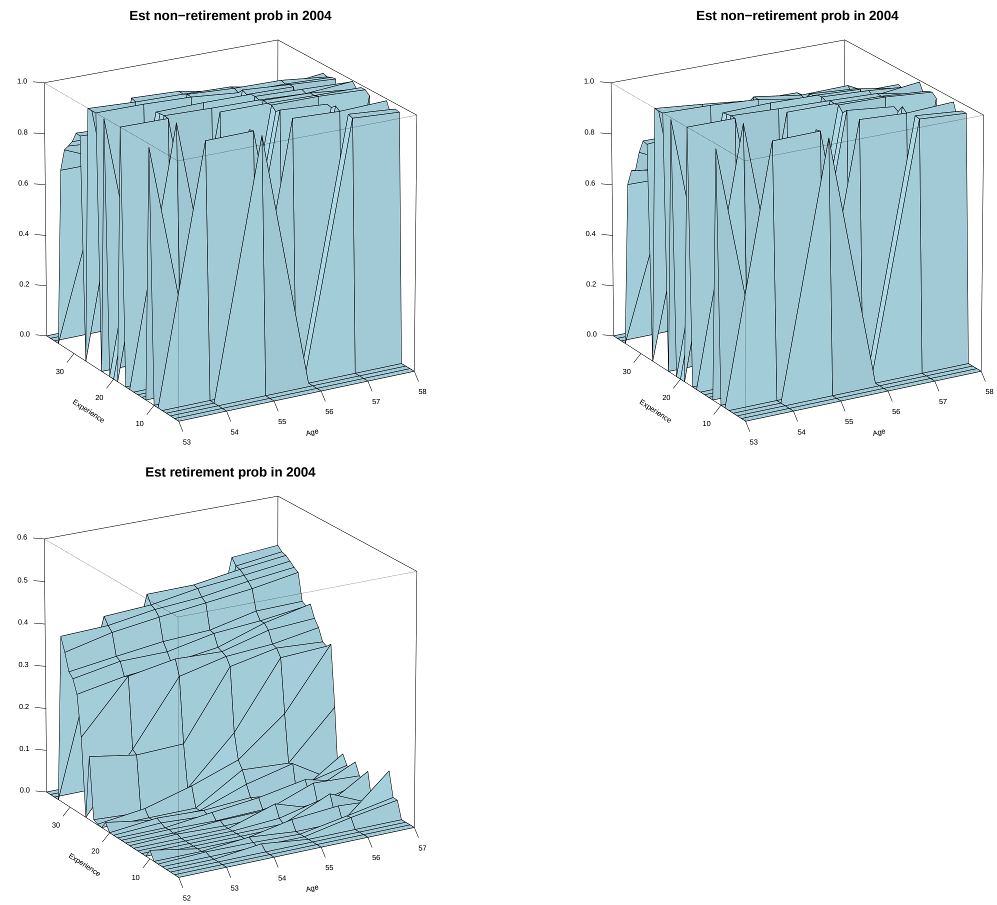


Figure 3.7: Estimated Probability Distribution in 2005

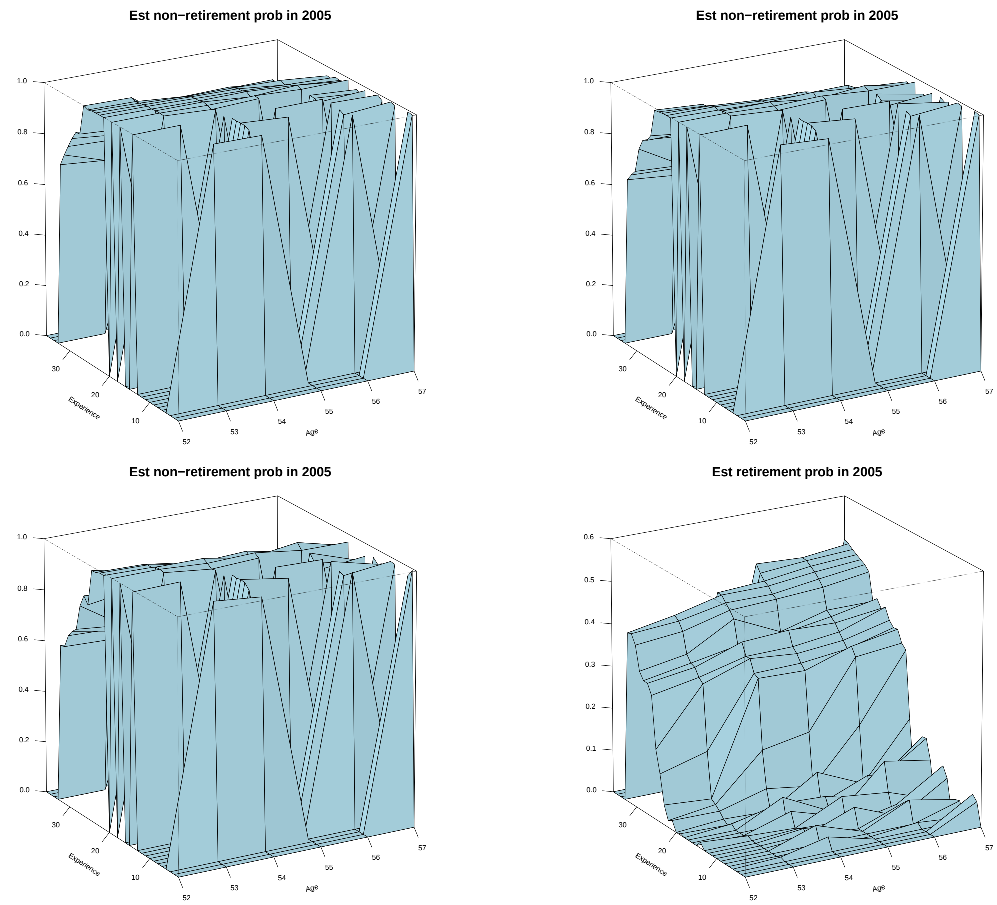


Figure 3.8: Estimated Probability Distribution in 2006

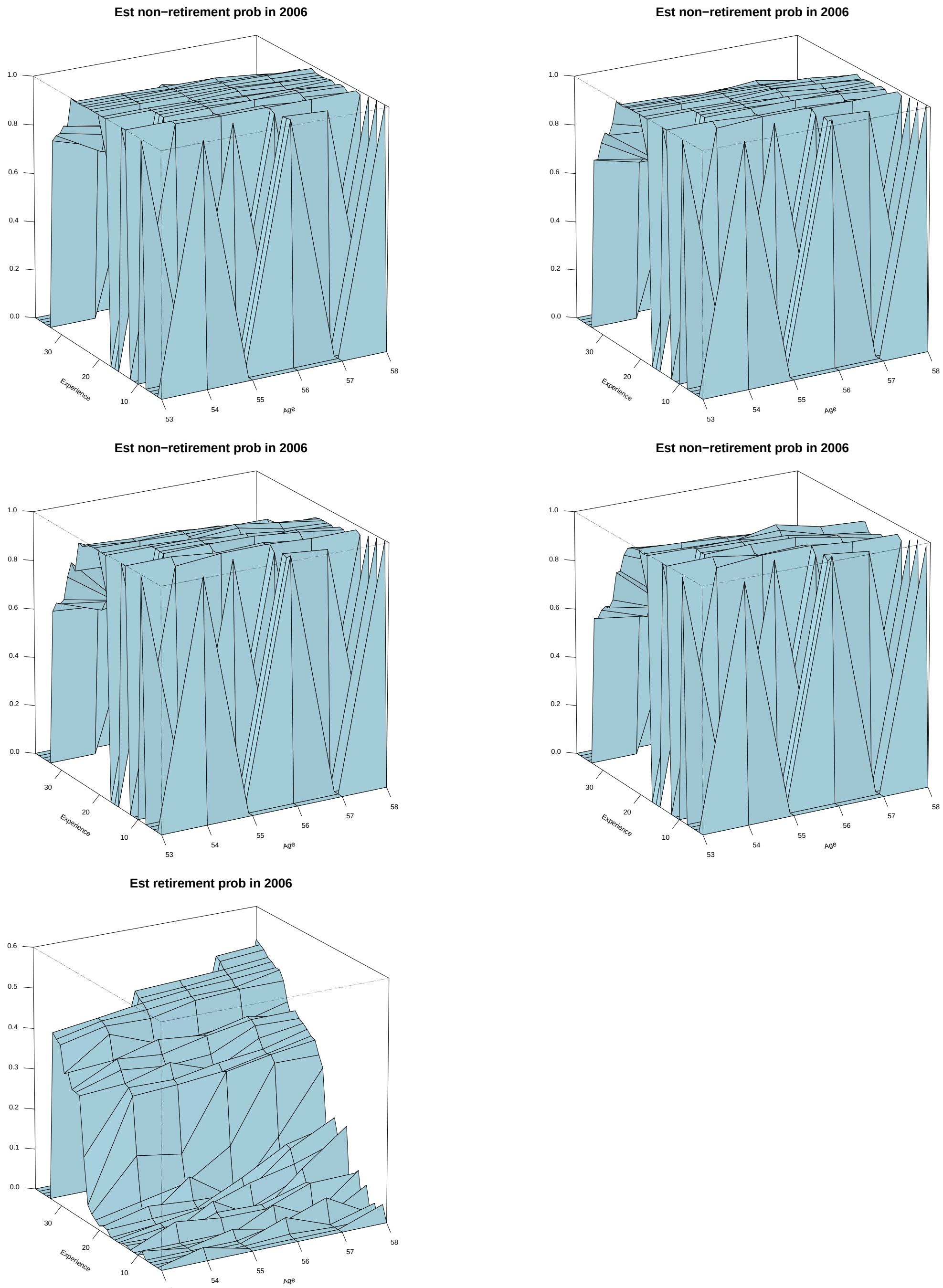


Figure 3.9: Estimated Probability Distribution in 2007

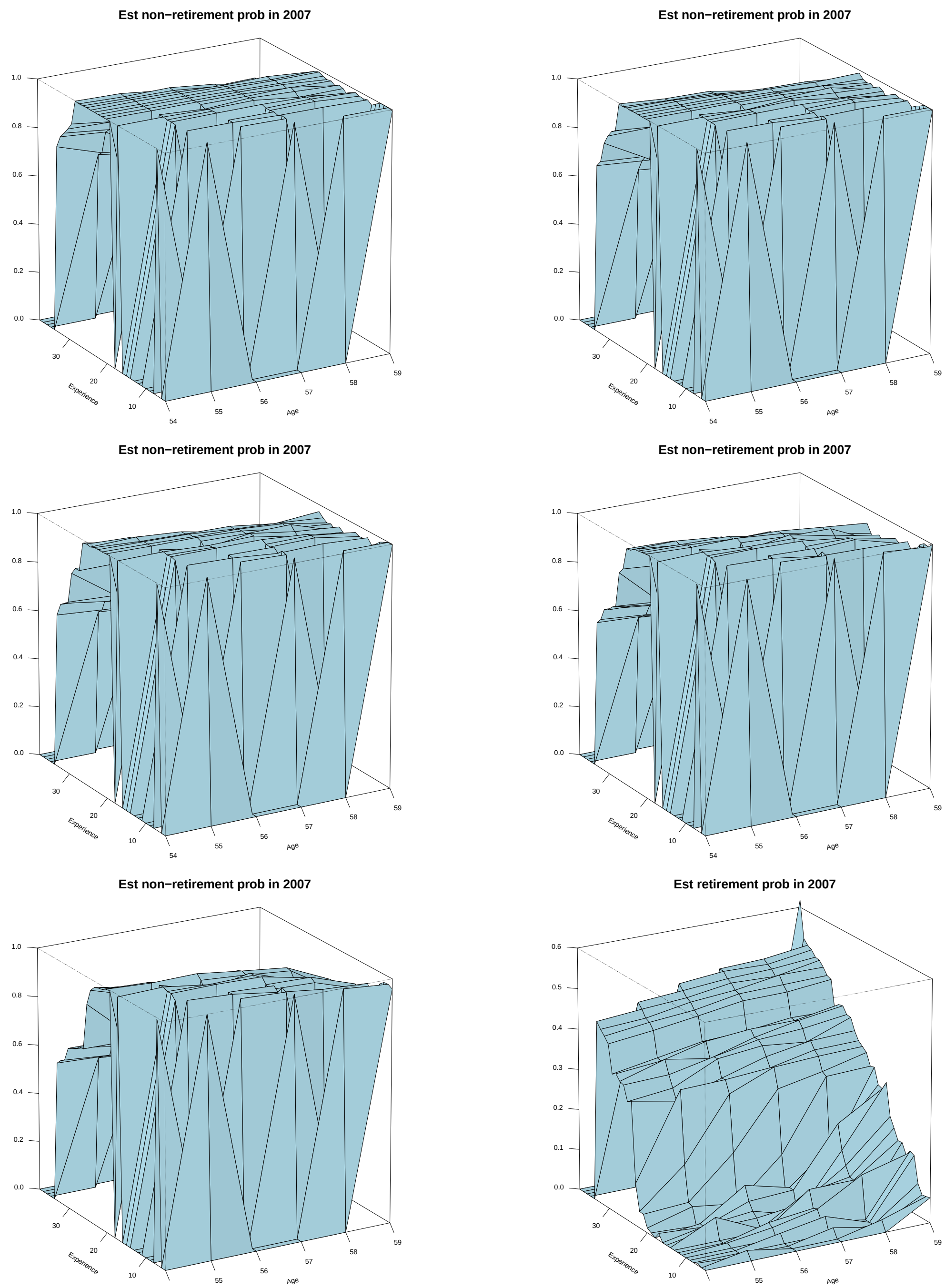
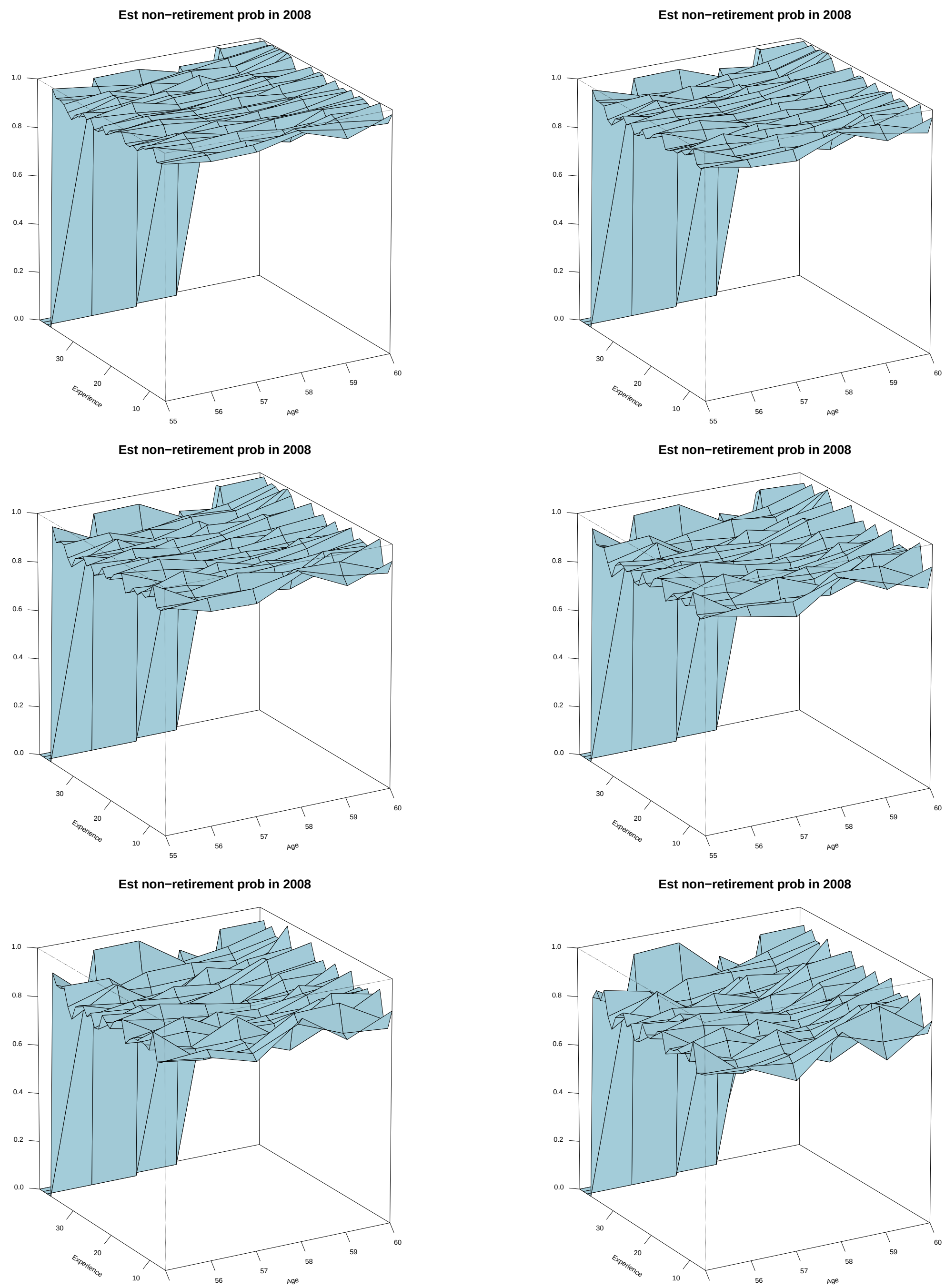


Figure 3.10: Estimated Non-retirement Probability Distribution in 2008



Non-retirement is well-estimated by the SW model, but retirement is not. The reason for this shortcoming in the SW model is perhaps that there are only a few cells of age and experience with a retirement frequency over 0.35. Because age and experience are essentially the only available pieces of information, the data are not informative enough for a more accurate outcome. The error term serves smoothly in the structure of the SW model; there is no room for a sharp "switching" from 0 to 1 over time in the model. Therefore, the non-retirement rate decreases gently and continuously. Teachers' sudden departure from the workforce cannot be fully captured by the SW model. The improvement may come in two ways: richer data that include more teacher characteristics and a richer model with more variables.

The frequency simulator adopted above has major drawbacks. First, the average count is a discontinuous function, which impedes numerical optimization and makes proving consistency and asymptotic normality of simulation estimators difficult. Second, yielding consistent likelihood estimates requires both sample size and the number of Monte Carlo draws per observation goes to infinity. Particularly for a small probability, a large number of draws are required. Börsch-Supan and Hajivassiliou (1994) propose a more efficient estimate method, the GHK simulator, to compute maximum likelihood to compute maximum likelihood. When compared with the frequency simulator, the GHK simulator ensures unbiasedness and smoothness, as well as a substantially smaller variance of the simulated probabilities even for highly interdependent error terms. Moreover, this method can be generalized to some non-normal distributions.

Data in this paper include 4,812 retired teachers and 4,793 non-retired teachers with 45,644 records over six years. The large sample requires long computation time. Moreover, error terms over those years are highly correlated for every individual teacher. Therefore, to more accurately and efficiently estimate the SW model, using the GHK simulator here is more suitable than using the frequency simulator. Taking the specification in Model 3 in Table 3.1, estimation results and running time by the FORTRAN program under different Monte Carlo draws (N) are reported in Table 3.4. The absolute value of the log-likelihood value does not improve inordinately until N increases to 10,000; however, the running time explodes rapidly. For example, when

$N = 100,000$, the running time increases to nearly three days. The running time here is only the CPU working time. When N = 100, the evaluation time exceeds the maximum; that is, the MLE does not converge, and the optimum cannot be obtained numerically.

Table 3.4: Estimation Result by the Frequency Simulator

Parameters	$N = 100$	$N = 1,000$	$N = 10,000$	$N = 100,000$
κ	0.742	0.656	0.623	0.629
β	0.968	0.961	0.960	0.961
γ	0.684	0.675	0.689	0.677
σ	3222.761	2912.950	3018.812	2666.807
ρ	0.610	0.543	0.540	0.543
$\kappa 1$	3.101	1.758	2.271	1.973
Log likelihood	-13319.338	-12934.400	-12825.600	-12821.478
Time(s)	?	11360.227	32059.375	256158.281

3.4.2 Likelihood Computation by the GHK Simulator

Introduction

Given a simple linear multivariate latent variable model as follows,

$$u_i = X_i\beta + \varepsilon_i, i = 1, ..., I, \tag{3.13}$$

where $\vec{u} \in \mathscr{R}^I$ is a latent variable and $\vec{\varepsilon} \in \mathscr{R}^I$ with a covariance matrix Ω. X is a $I \cdot K$ matrix of observed explanatory variables, and β is a K-dimensional parameter vector. Then the probability of choosing option i over other option j is given by the following:

$$P(y = 1|\, x) = P(u_i > u_j) = P(\varepsilon_j - \varepsilon_i \leq X_i\beta - X_j\beta) = P(A \cdot \varepsilon \leq b) \text{for} j \neq i. \tag{3.14}$$

where y is a dichotomous variable of value of 0 or 1. If $y = 1$, then school i is selected; otherwise, it is not. The $I \cdot I$ matrix A may depend on y and $X\beta$. A more general expression of the above probability function is the following:

$$P(y, X; \beta) = P(a \leq A \cdot \varepsilon \leq b), \tag{3.15}$$

where a and $b \in \mathscr{R}^I_\infty$ are the limiting vectors.

Let L be the Choleski decomposition of variance matrix of $A\varepsilon$, that is,

$$LL' = A\Omega A', \tag{3.16}$$

which also means that $L \cdot \vec{e} = A \cdot \vec{\varepsilon}$ ($\vec{e}$ is the unit vector). Then,

$$\vec{\varepsilon} \sim N(0, \Omega) \text{ s.t. } a \leq A \cdot u \leq b, \tag{3.17}$$

can be transformed to drawing a random vector:

$$\vec{e} \sim N(0, I) \text{ s.t. } a \leq L \cdot e \leq b. \tag{3.18}$$

Instead of drawing the whole $A \cdot u$ subject to a and b, as in the frequency simulator, the GHK simulator decomposes the drawing to the following recursive restrictions:

$$\begin{aligned} e_1 \sim N(0,1) \text{ s.t. } & a_1 \leq l_{11} \cdot e_1 \leq b_1 \\ & \Leftrightarrow a_1/l_{11} \leq e_1 \leq b_1/l_{11}, \end{aligned} \tag{3.19}$$

$$\begin{aligned} e_2 \sim N(0,1) \text{ s.t. } & a_2 \leq l_{21} \cdot e_1 + l_{22} \cdot e_2 \leq b_2 \\ & \Leftrightarrow (a_2 - l_{21} \cdot e_1)/l_{22} \leq e_2 \\ & \leq (b_2 - l_{21} \cdot e_1/l_{22}), \end{aligned} \tag{3.20}$$

$$\text{etc.} \tag{3.21}$$

Consequently, the likelihood function can be expressed as follows:

$$\begin{aligned}
l(y, X; \beta, \Omega) = &\ P(a_1/l_{11} \leq e_1 \leq b_1/l_{11}) \\
&\cdot P((a_2 - l_{21} \cdot e_1)/l_{22} \leq e_2 \leq (b_2 - l_{21} \cdot e_1)/l_{22}|e_1) \ldots \\
&\cdot P((a_I - l_{11} \cdot e_1 - ... - l_{Ii-1} \cdot e_{i-1})/l_{II} \leq e_I \\
&\leq (b_I - l_{I1} \cdot e_1 - ... - l_{Ii-1} \cdot e_{i-1})/l_{II}|e_1 \cdot, ..., e_{i-1}) \\
&= Q_1 \cdot Q_2(e_1) \cdot Q_3(e_1, e_2) ... Q_I(e_1, ..., e_{I-1}).
\end{aligned} \tag{3.22}$$

Then, the average value $\left(\frac{1}{R}\right) \sum_{r=1}^{R} \prod_{i=1}^{I} Q_i(e_{1r}, ..., e_{i-1,r})$, with e_{ir} drawn from a truncated standard normal distribution, approximates the above likelihood function, and R denotes the number of replications. The way to derive a univariate truncated normal variate $Z \in (a, b)$ is to draw a random variable X from a univariate uniform distribution on [0,1]; then:

$$Z \equiv G^{-1}(X) = \Phi^{-1}[((\Phi(b) - (\Phi(a)) \cdot X + \Phi(a)], \tag{3.23}$$

where Φ denotes the univariate normal cumulative distribution function.

The GHK Simulator in Retirement Model

The variance matrix of error terms, $\vec{v} = (-v_t, ..., -v_{t+n})$, in the SW model is given by the following:

$$\Sigma = \frac{\sigma^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \ldots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \ldots & \rho^{n-3} & \rho^{n-2} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \ldots & \rho & 1 \end{pmatrix}, \tag{3.24}$$

Taking $n = 6$ as an example (which implies that a teacher retires in school year 2007-2008), [2] the Cholesky decomposition matrix is given by the following:

$$L = \begin{pmatrix} \frac{\sigma}{\sqrt{1-\rho^2}} & & & \\ \frac{\sigma\rho}{\sqrt{1-\rho^2}} & \sigma & & \\ \vdots & \dots & \vdots & \\ \frac{\sigma\rho^5}{\sqrt{1-\rho^2}} & \sigma\rho^4 & \dots & \sigma \end{pmatrix} = \begin{pmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \dots & \vdots & \\ l_{61} & l_{62} & \dots & l_{66} \end{pmatrix}; \tag{3.25}$$

and the derived truncated normal distribution e_i is as follows:

$$\begin{aligned} e_1 \sim N(0,1) \text{ s.t. } & v_1 = l_{11} \cdot e_1 \le g_1(r_1^*)/K_1(r_1^*) \\ & \Leftrightarrow e_1 \le g_1(r_1^*)/(K_1(r_1^*)l_{11}), \\ e_2 \sim N(0,1) \text{ s.t. } & v_2 = l_{21} \cdot e_1 + l_{22} \cdot e_2 \le g_2(r_2^*)/K_2(r_2^*) \\ & \Leftrightarrow e_2 \le (g_2(r_2^*)/K_2(r_2^*) - l_{21} \cdot e_1)/l_{22} \\ \dots & \\ e_6 \sim N(0,1) \text{ s.t. } & g_6(r_6^*)/K_6(r_6^*) \le l_{61} \cdot e_1 + \ldots + l_{66} \cdot e_6 = v_6 \\ & \Leftrightarrow (g_6(r_6^*)/K_6(r_6^*) - l_{61} \cdot e_1 - \ldots - l_{65} \cdot e_5)/l_{66} \le e_6 \end{aligned} \tag{3.26}$$

After drawing a uniform variant X from $(0,1)$, the series of e_i from truncated normal distribution is simulated by the following:

$$\begin{aligned} & e_1 = \Phi^{-1}[\Phi(g_1(r_1^*)/K_1(r_1^*)l_{11})X] \\ & e_2 = \Phi^{-1}[\Phi(g_2(r_2^*)/K_2(r_2^*) - l_{21} \cdot e_1/l_{22})X] \\ & \dots \\ & e_6 = \Phi^{-1}[(1 - \Phi(g_6(r_6^*)/K_6(r_6^*) - l_{61} \cdot e_1 - \ldots - l_{65} \cdot e_5/l_{66}))X \\ & \quad + \Phi(g_6(r_6^*)/K_6(r_6^*) - l_{61} \cdot e_1 - \ldots - l_{65} \cdot e_5/l_{66})] \end{aligned} \tag{3.27}$$

[2] If a teacher do not retire till school year 2007-2008 then $n = 7$.

Therefore, the likelihood function of retiring at time $t+n$ can be written as follows:

$$
\begin{aligned}
L(y, X; \beta, \Sigma) = \Pr(e_1 \leq g_1(r_1^*)/(K_1(r_1^*)l_{11})) \\
\cdot \Pr(e_2 \leq (g_2(r_2^*)/K_2(r_2^*) - l_{21} \cdot e_1)/l_{22}|e_1) \ldots \\
\cdot \Pr((g_6(r_6^*)/K_6(r_6^*) - l_{61} \cdot e_1 - \ldots - l_{65} \cdot e_5)/l_{66} \leq e_6|e_5) \\
= \Phi[g_1(r_1^*)/K_1(r_1^*)l_{11}] \cdot \Phi[g_2(r_2^*)/K_2(r_2^*) - l_{21} \cdot e_1/l_{22})] \\
\ldots \\
(1 - \Phi[(g_6(r_6^*)/K_6(r_6^*) - l_{61} \cdot e_1 - \ldots - l_{65} \cdot e_5/l_{66}) \leq e_6],
\end{aligned}
\tag{3.28}
$$

and the simulated likelihood is estimated by $\hat{l} = \sum l/R$.

Still taking the Model 3 specification of the SW model as an example, when comparing of estimates and costing time by the GHK simulator and the frequency simulators with different times of Monte Carlo draws (N) are listed in the following table:

Table 3.5: Estimates of the GHK vs the Frequency Simulator Applied in the Model 3

The GHK Estimates	$N = 10$	$N = 100$	$N = 1,000$
κ	0.623	0.631	0.639
β	0.959	0.960	0.962
γ	0.663	0.682	0.695
σ	2983.134	2635.503	2889.515
ρ	0.510	0.556	0.582
$\kappa 1$	3.013	2.448	1.812
Log likelihood	-12912.655	-12807.564	-12813.651
Computation Time(S)	8705.683	10399.261	29064.063
The Frequency Estimates	$N = 1,000$	$N = 10,000$	$N = 100,000$
κ	0.656	0.623	0.629
β	0.961	0.960	0.961
γ	0.675	0.689	0.677
σ	2912.950	3018.812	2666.807
ρ	0.543	0.540	0.543
$\kappa 1$	1.758	2.271	1.973
Log likelihood	-12934.400	-12825.600	-12821.478
Computation Time(S)	11360.227	32059.375	256158.281

The upper part and the lower parts of Table 3.5 report estimates derived by using the GHK simulator and the frequency simulator separately. Estimated parameters by the two simulators do not differ inordinately except for the log-likelihood. As log-likelihood is a log function, a difference of 1 in value means the difference in multiplication of e. After only 100 draws, the GHK simulator can make the value of log-likelihood almost the same as what the frequency simulator does after 100,000 draws but with a 25-fold shorter running time. The advantage of the GHK simulator would be even greater if the sample of teachers is extended to more states or over a longer time span.

3.5 Simulated Annealing

3.5.1 Introduction

It is usually straightforward to obtain MLE when there are few unknown parameters and the objective function is well-behaved. However, if the number of parameters is large and the objective function is ill-behaved, particularly for a likelihood function without specific expression, conventional algorithms might fail to find the optimum. Often even after running a large number of steps and starting from different values, the algorithms may not converge. Even if they do, those algorithms do not guarantee that the optimum detected is a global one instead of a local one.

For Missouri public school teachers' data, because of complicated and nonlinear retirement rules, the objective option value function does not have a single peak value. The flat ridge is found to be a feature of the objective function. As demonstrated in the just preceding section, the SW model includes six parameters, which are insensitive to model setting except σ. SA is effective and robust in finding a global optimum of those unpleasant functions than other traditional algorithms are.

Generally, a conventional searching method, such as a simplex algorithm, always starts at a point and continues searching in the best direction, with the best step length calculated, until predefined stopping criteria are achieved after iterative steps. History does not provide updated information for next steps. In contrast, simulated annealing (SA), proposed in Kirkpatrick, Gelett, and Vecchi (1983), is a generic, simulation-based, metaheuristic approach to the global optimization problem. SA explores the whole domain of objective function by moving both uphill and downhill, with information getting updated after each step. It takes root in metallurgy annealing, a technique involving heating and cooling a metal to arrive at a low energy state. Heating and cooling fluctuations in energy allow the annealing system to escape from local energy minima to achieve a global one, which might not be achieved if the metal is cooled too rapidly. Put simply, slowly sprinkling on a heated metal is more effective in cooling it down than pouring water. In global optimization, the concept of SA is extended as a slow decrease in the probability of accepting worse solutions when the algorithm explores the entire solution space. One drawback of SA is its lengthy running time. Combing the previously discussed efficient GHK simulator can alleviate this issue.

Let T^0 be the initial temperature and $f(X)$ be the objective function to maximize, where X denotes the set of parameters. Starting from a random point X^0, a succession of parameter candidates is generated around the starting point by varying each coordinate of X^0 in turn. Particularly, new points are derived by the following:

$$x_i{}' = x_i{}^0 + rv_i, \tag{3.29}$$

where r is randomly chosen from a unit uniform distribution and v_i is an entrant of V, the step length for X. Compare the objective function value at both the starting point f_0 to that at the new point f'. If f' is greater, then accept the new parameter

$X^{'}$, and replace X^0 with $X^{'}$, otherwise $X^{'}$ is accepted according to the Metropolis criterion. The probability of acceptance is as follows:

$$p = e^{(f^{'} - f^0)/T^0}, \tag{3.30}$$

where T^0 is the parameter for initial temperature mentioned before. The SA algorithm always begins with some high initial temperature so that the probability of acceptance p is not small. If p is larger than $p^{'}$, a random number drawn from a unit uniform distribution, then the new parameter $X^{'}$ is accepted, otherwise it is declined.

After N_s rounds of changing all elements of X, [3] if the acceptance ratio is more than half, step length V is enlarged to decrease the percentage and vice versa. The temperature is reduced by $T^{'} = r_T \cdot T$ after N_T times through the aforementioned cycles, where $0 < r_T < 1$. [4] In each round of function evaluations, parameters and function values are updated if the new function value is better. If the difference between the last new function value in the last N_ϵ and the previous best value is smaller than a preset tolerance value ϵ, the search stops.

In summary, X^0, N_s and N_T are fixed external parameters that are chosen on the basis of researchers' experience. Researchers may opt for high T^0 and large V to begin with. In the process of SA climbing up and down, V is adjusted after every N_s rounds. Afterward, the temperature is decreased after every N_T rounds. Eventually, as Goffe et al. point out, SA searches both uphill and downhill; hence, there is a higher probability of finding a global minimum. Furthermore, unlike some traditional algorithms using differentiation techniques, SA does not require the function to be differentiable around the optimum. The changing vector length during climbing process reveals important information regarding the function. Because of climbing in two directions and containing more varying parameters, SA requires much longer

[3] If the length of vector X is N, then the times of function evaluations is $N_s \cdot N$.

[4] The times of function evaluations is $N_s \cdot N_T \cdot N$.

computation time. The aforementioned features have been demonstrated in detail in their study.

In Goffe et al. (1999), three common conventional algorithms, from the IMSL Math/Library for multivariate optimization, are chosen to be compared with SA because of quality and availability: UMPOL is a simplex algorithm that minimizes a function using a direct search polytope algorithm by replacing the worst point among $n + 1$ points; UMCGF is a conjugate gradient algorithm that minimizes a function by using a conjugate gradient algorithm and a finite-difference gradient; UMINF is a quasi-Newton algorithm using a quasi-Newton method and a finite-difference gradient numerical derivatives. In this section, one more traditional algorithm, BCPOL, is added to this study to be compared with the above mentioned three traditional algorithms and further illustrate SA's superiority over traditional algorithms. BCPOL is the algorithm applied with both the frequency and the GHK simulator in the last section to estimate the option value model. The BCPOL algorithm minimizes a function, subject to bounds on the variables, using a direct search complex algorithm. BCPOL is similar to UMPOL but replaces the worst point from $2n$ points. The complex algorithm stands for constrained simplex, and so the procedure is analogous to the simplex procedure for unconstrained problems, except that the search is in the permissible area. The complex algorithm does not offer any improvements over the simplex algorithm except for the consideration of constraints. Before applying SA to the SW model, some popular optimization problems are computed by using the four algorithms and SA to illustrate how SA works and to compare its performance with those traditional algorithms.

3.5.2 Examples

The Judge Function

Judge et al.(1985, pp. 956-957) consider minimization problem for a nonlinear function:

$$\min_{\theta_1,\theta_2} \sum_{i=1}^{N} (\theta_1 + \theta_2 x_{2i} + {\theta_2}^2 x_{3i} - y_i)^2, \tag{3.31}$$

where (θ_1, θ_2) is the parameter vector. The model is actually a transformation of a simple linear model with two parameters, as follows:

$$y_i = \theta_1 + \theta_2 x_{2i} + {\theta_2}^2 x_{3i} + e_i, \tag{3.32}$$

Table 3.6: Data for the Judge Function

i	y_i	x_{1i}	x_{2i}	x_{3i}
1	4.284	1.000	0.286	0.645
2	4.149	1.000	0.973	0.585
3	3.877	1.000	0.384	0.310
4	0.533	1.000	0.276	0.058
5	2.211	1.000	0.973	0.455
6	2.389	1.000	0.543	0.779
7	2.145	1.000	0.957	0.259
8	3.231	1.000	0.948	0.202
9	1.998	1.000	0.543	0.028
10	1.379	1.000	0.797	0.099
11	2.106	1.000	0.936	0.142
12	1.428	1.000	0.889	0.296
13	1.011	1.000	0.006	0.175
14	2.179	1.000	0.828	0.180
15	2.858	1.000	0.399	0.842
16	1.388	1.000	0.617	0.039
17	1.651	1.000	0.939	0.103
18	1.593	1.000	0.784	0.620
19	1.046	1.000	0.072	0.158
20	2.152	1.000	0.889	0.704

Source: The Theory and Practice of Economics, 2nd ed.
1985, (975-976)

Given the value of the 20 records in Table 3.6, the above four traditional algorithms and SA return estimates for θ as follows:

Table 3.7: The Frequency of Correctly Finding Maximum of The Judge Function out of 100 Tries

Algorithm		UMPOL (Simplex)	UMCGF (conjugate gradient)	UMINF (quasi-Newton)	BCPOL (Complex)	SA (simulated annealing)
Solutions	@20.482	42	47	45	39	0
	@16.082	58	53	55	61	100

For comparison, the results given by Goffe et al. are as follows:

Table 3.8: The Frequency of Correctly Finding Maximum of The Judge Function out of 100 Tries in Goffe et al. (1994)

Algorithm		UMPOL (Simplex)	UMCGF (conjugate gradient)	UMINF (quasi-Newton)	SA (simulated annealing)
Solutions	@20.482	40	48	48	0
	@16.082	60	52	52	100

The Judge function has two local minima of 20.482 and 16.082, and 16.082 is the global minimum. All algorithms are run 100 times, with starting points randomly chosen from a uniform distribution on the interval (-100, 100). Most settings for the four traditional algorithms take the default values given in the IMSL Fortran Numerical Library, and the settings of key inputs in SA are reported in Table 3.9 :

Table 3.9: Input in SA

Input	Notation	Value
T^0	initial temperature	5000
ϵ	convergence criteria	1.0E-8
N_s	# times through function evaluated before V adjusts	20
C_i	parameter to control adjustment step of V	2.0(i=1,2)
r_T	temperature reduction factor	0.85
N_ϵ	# times ϵ tolerance is achieved before termination	4
N_T	# times through N_s loop before T reduces	100
V_i	step length	100.0(i=1,2)
$MAXEVL$	maximum times of function evaluations	2000

The frequency results tables clearly illustrate each algorithm's performance. The chance for the four traditional algorithms to find the global minimum is almost the same, 50%, while SA does not miss any opportunity to find the right global optimum. As Goffe et al. state, the simple example cannot fully demonstrate that SA is superior as narrowing the interval of starting points drawn can help traditional algorithms find the global minimum eventually. Therefore, another more complicated example is presented below to address the advantages of SA.

The Rosenbrock Function

The Rosenbrock function, introduced by Howard H. Rosenbrock (Rosenbrock, H. H., 1960), is broadly used as a performance test for optimization algorithms. The function is non-convex, and its global minimum is inside a long, narrow, parabolic-shaped flat valley. It is easy to find the valley but not the global minimum. The function is defined as follows:

$$f(x) = \sum_{i=1}^{N-1} \left[(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2\right] \tag{3.33}$$

When $N = 2$, the function is given by the following,

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2. \tag{3.34}$$

The global minimum is at $(x, y) = (1, 1)$, where $f(x, y) = 0$ and the plot of the Rosenbrock function looks as follows:

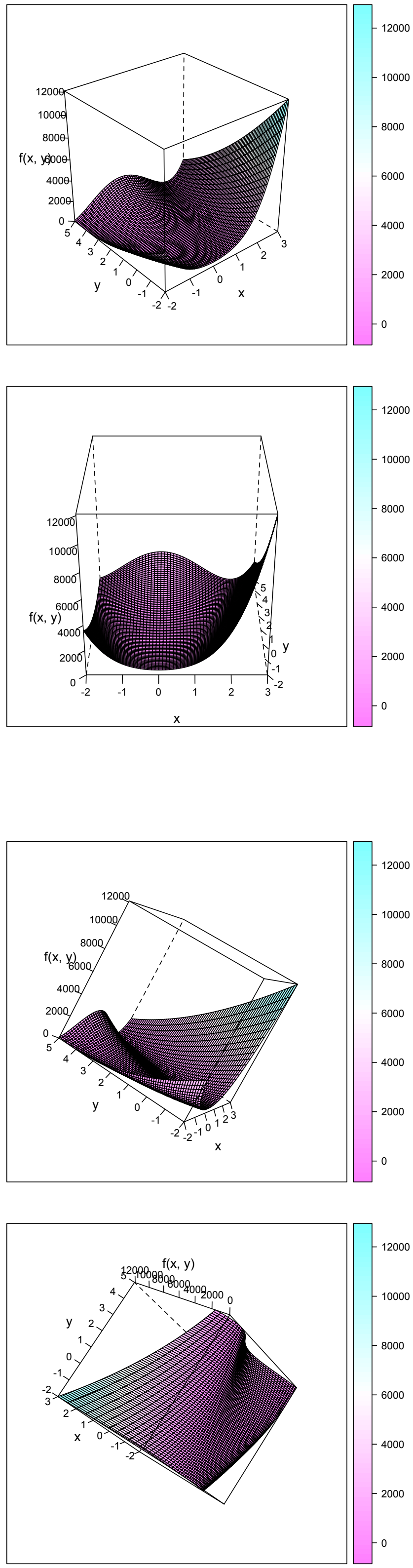


Figure 3.11: Different Views of the Graph of the 2-D Rosenbrock Function

When dimension N increases to 4, the function changes to the following:

$$\begin{aligned} f(x,y) = {} & [(1-x_3)^2 + 100(x_4 - {x_3}^2)^2] \\ & + [(1-x_2)^2 + 100(x_3 - {x_2}^2)^2] \cdot \\ & + [(1-x_1)^2 + 100(x_2 - {x_1}^2)^2] \end{aligned} \tag{3.35}$$

There are two minima; the global minimum is at the point $(1,1,1,1)$ and a local minimum is near $(-1,1,1,1)$. [5] Applying the four traditional algorithms and SA to the two Rosenbrock functions, the chances of finding the global minimum and consuming time are reported in the following two tables:

[5]This function has exactly one minimum of all ones for $N \leq 3$ and exactly two minima for $4 \leq N \leq 7$, the global minimum of all ones and a local minimum near $(x_1, x_2, \ldots, x_N) = (-1, 1, \ldots, 1)$

Table 3.10: Number of Failures out of 1000 tries and Time of Running the Rosenbrock Function by Traditional Algorithms

Algorithm	Number of Parameters	Initial XGuess Range	# of Failure	Running Time
UMPOL	$N = 2$	$[-1, 1]$	0	0.1092s
		$[-10, 10]$	0	0.1404s
		$[-100, 100]$	31	0.2028s
	$N = 4$	$[-1, 1]$	331	0.3276s
		$[-10, 10]$	657	0.3432s
		$[-100, 100]$	870	0.3744s
UMCGF	$N = 2$	$[-1, 1]$	0	0.1716s
		$[-10, 10]$	0	0.2808s
		$[-100, 100]$	0	0.4524s
	$N = 4$	$[-1, 1]$	60	0.4680s
		$[-10, 10]$	101	0.5928s
		$[-100, 100]$	117	0.8268s
UMINF	$N = 2$	$[-1, 1]$	0	0.156s
		$[-10, 10]$	0	0.2496s
		$[-100, 100]$	0	0.8112s
	$N = 4$	$[-1, 1]$	129	0.2808s
		$[-10, 10]$	215	0.4680s
		$[-100, 100]$	237	1.1856s
BCPOL	$N = 2$	$[-1, 1]$	0	0.4056s
		$[-10, 10]$	0	0.5148s
		$[-100, 100]$	24	1.2480s
	$N = 4$	$[-1, 1]$	1000	1.1232s
		$[-10, 10]$	1000	1.2324s
		$[-100, 100]$	1000	1.3260s

Note: All settings in those functions adopt default value given by IMSL.

Table 3.11: Updated Running Performance by BCPOL Algorithm with Narrower Bounds of Parameters

Algorithm	Dimension	Initial XGuess Range	# of Failure	Running Time
BCPOL	$N = 2$	$[-1, 1]$	0	0.1404s
		$[-10, 10]$	0	0.1716s
		$[-100, 100]$	0	0.2184s
	$N = 4$	$[-1, 1]$	80	0.4056s
		$[-10, 10]$	130	0.4368s
		$[-100, 100]$	139	0.4680s

Note: Bounds of all XGuess have changed from $\pm 10^{25}$ to ± 100.

The four traditional algorithms' performance is listed in sequence for both cases of $N = 2$ and $N = 4$. The interval where the starting point for every parameter x_i is drawn expands from $[-1, 1]$ to $[-10, 10]$ and then to $[-100, 100]$. As this example is more complicated than the last one, all algorithms are run 1,000 times instead of 100 to make the data meaningful. The last two columns refer to the times of failure to find the global minimum and the running time for computer to return results.

When the Rosenbrock function is two-dimensional, its expression is relatively simple. It is not difficult for any of the four algorithms to find the global minimum, particularly when the parameters' starting point is not far away from the point where the function reaches the smallest. Unsurprisingly, the number of times any of the four fail to find the global minimum equals 0 for "Initial XGuess Range" is either $[-1, 1]$ or $[-10, 10]$. Once the "Initial XGuess Range" expands to $[-100, 100]$, those algorithms' performance starts to differ. All inputs in those IMSL routines are still set as the default value. The UMPOL algorithm fails to find the global minimum 31 times. UMCGF does not fail at all even when the initial value of each parameter is now drawn from $[-100, 100]$, UMCGF still performs perfectly. When "Initial XGuess Range" enlarges from $[-1, 1]$ to $[-10, 10]$, the default values of inputs causes the computer to state, "ERROR 1 from UMCGF. The line search of an integration was abandoned. An error in the gradient may be the cause." Changing the gradient tolerance value from $1.0E-6$ to $1.0E-10$ eliminates the problem. The new gradient tolerance also works for the guess interval of $[-100, 100]$. The big decrease in gradient tolerance criteria in some way demonstrates the feature of the Rosenbrock functions introduced at the beginning: a relatively flat valley around its global minimum. The performance of UMINF is quite similar to that of the UMCGF algorithm. It is not difficult to comprehend the similarity, considering that the calculations in the two routines does not differ too much for the low dimension. For the two narrower intervals of "XGuess," the default input value works; however, when the "Initial XGuess Range" is $[-100, 100]$, all input values, such as the maximum number of function evaluations, gradient evaluations, and iterations need to be adjusted for them to be much larger than the corresponding default values, otherwise the algorithm fails to converge. The BCPOL algorithm performs approximately the same as UMPOL does. It misses the global minimum 24 times out of 1,000 times when "Initial XGuess Range" falls in $[-100, 100]$ but does not miss any for the other two intervals.

When the dimension of the Rosenbrock function increases to four, the situation becomes more complicated, and each algorithm's performance. Generally speaking, the rate of failure to find the global optimum goes up as the "Initial XGuess Range" expands. The simplex method in the UMPOL routine still does not require modifying any input value to make the program run. On the other hand, this simple method cannot be said to be successful in finding the global minimum. With the expansion of the "Initial XGuess Range" from $[-1, 1]$ to $[-100, 100]$, the number of times of the global minimum is not found increases from 331 to 870, with the failure rate being high as 87%. The UMCGF algorithm generates much better results. Still, the rate of failure increases with the expansion of the "Initial XGuess Range"; however, the highest failure rate is only 11.7% under a new gradient tolerance of $1.0E - 10$. Though UMINF does not work as well as UMCGF, the maximum failure rate is an acceptable 23.7%, which is acceptable. When $N = 4$, the maximum number of function evaluations, the gradient evaluations and iterations need to be adjusted for all three cases of different intervals so that the IMSL routine in UMINF gets running. The BCPOL algorithm also returns a good result: its failure rates are 8%, 13%, and 13.9% for the three corresponding intervals.

In both UMPOL and BCPOL routines, there are two key required arguments in function inputs, XLB (lower bounds on the variables) and XUB (upper bounds on the variables). [6] When $N = 2$, both XLB and XUB can be set as low as $-10E25$ and

[6] The two can actually be claimed as the only two controllable inputs since the other input IBTYPE, indicating the types of bounds on variables, usually does not have too much room to reset.

as high as $10E25$. The two routines can find the global minimum easily. Once N is set at 4, the values of XLB and XUB matter for BCPOL but not for UMPOL. When XLB and XUB are still set at $-10E25$ and $10E25$, respectively, the failure rates in the BCPOL routine are 100% for any initial XGuess regardless of the interval from which it is drawn. Once the bounds are set as -100 and 100, the failure rate can decrease remarkably from 100% to less than 15% as shown in Table 3.10. In the meantime, the BCPOL algorithm can find the global minimum in all trials when $N = 2$. One thing to be noted is that further narrowing the interval does not imply imply an improvement in the program's performance. When the bounds of XGuess are changed to 1 and 1, even the failure rate increases. The different effect brought about by constraints on the BCPOL routine and the UMPOL routine is implied. Another notable feature is that, when $N = 4$, the error points in all algorithms except UMPOL are focused around the F-value of 3.70, another local minimum of the Rosenbrock function. In UMPOL, however, erroneous points roam very wild.

As for running time, the pattern is that it goes up with the increasing of the parameter N and the expanding of the XGuess interval. The running time for the BCPOL algorithm, when XLB and XUB are $-10E25$ and $10E25$ respectively, is relatively longer than that of others. Once the bounds are narrowed, the BCPOL running time is shortened rapidly. The running time in the UMINF routine is relatively higher than the other three.

After viewing the performance of the four traditional algorithms, let us examine how SA works. Among those setting inputs for the SA algorithm, the convergence criteria ϵ is subject to the objective function's estimated value. If the function value itself is quite large, then SA is hard to converge with the small ϵ. C_i and N_ϵ are suggested to be set at 2 and 4 separately according to experience (Corata, et al. 1987). The value of the initial V_i does not matter as V_i depends on T^0 and will be adjusted after every step. As Goffe et al. (1994) point out an overly low initial temperature makes the step length too small, and the area containing the optimum may be missed. If it is too high, then the step length is too large and an excessively large area is searched. Therefore, three key inputs, T_0, r_T, and N_T, are controlled here to check their influences.

Table 3.12: Number of Failures out of 1000 tries and Time of Running the 2-D Rosenbrock Function by SA under $T^0 = 200,000$

		N_T	# of Failure	Running Time
$N = 2$, $T^0 = 200,000$	$r_T = 0.5$	5	394	3.2916s
		10	116	3.6972s
		15	56	4.3212s
		20	28	4.9764s
	$r_T = 0.85$	5	4	6.8640s
		10	0	9.3913s
		15	0	15.3037s
		20	0	19.5781s

Similar to Table 3.10 for traditional algorithms, Table 3.12 reports the failure rate and running time by SA under different input values of r_T and N_T with $T^0 = 200,000$ and $N = 2$ for the simplest version of the Rosenbrock function, $N = 2$. SA runs 1,000 times, with starting points randomly chosen from a uniform distribution on the interval $(-100, 100)$. When temperature reduction factor r_T equals 0.5, the rate of failure goes down from 394 to 28 as N_T, and the number of times going through one N_s loop before T reduces, goes up from 5 to 20 by 5. Once r_T increases to 0.85, the rate of failure reduces substantially. Even when $N_T = 5$, the rate of failure is only four, and once N_T increases to 10, SA no longer misses any global optimum.

Table 3.13: Number of Failures out of 1000 tries and Time of Running the 4-D Rosenbrock Function by SA under $T^0 = 200,000$

		N_T	# of Failure	Running Time
$N = 4$, $T^0 = 200,000$	$r_T = 0.5$	5	571	9.4069s
		10	293	9.4849s
		15	198	11.5753s
		20	166	13.6345s
	$r_T = 0.85$	5	203	13.2289s
		10	91	25.5842s
		15	36	33.8366s
		20	23	44.8347s
		25	8	55.0216s
		30	8	64.2724s
	$r_T = 0.95$	30	0	165.1271s

Table 3.13 presents SA performance when evaluating the Rosenbrock function with four variants. The law states that the rate of missing the global optimum goes down with increasing r_T and N_T still holds. When $r_T = 0.5$ and $N_T = 5$, the failure rate of 57.1% is larger than that of the four traditional algorithms except for UMPOL. Once N_T increases to 10, the failure rate reduces almost by half. The rate decreases by around 32% as N_T increases another five. When N_T increases to 20, the failure rate continues to decrease to nearly 16%; however, the decreasing speed slows down. Until now, the failure rate is comparable to that of by the four traditional algorithms under the same conditions when initial XGuess is drawn from $[-100, 100]$.

When r_T rises to 0.85, the failure rate drops from 20.3% at $N_T = 5$ down to 9.1% at $N_T = 10$, then to 3.6% at $N_T = 15$, further to 2.3% at $N_T = 20$, and last to 0.8% at $N_T = 25$. In contrast, none of the four traditional algorithms has a failure rate below 10%. As N_T finally increases from 25 to 30, the failure rate does not change any more but remains at 0.8%. While r_T moves to 0.95, SA no longer misses any global optimum.

To investigate the effects of initial temperature T^0, SA is run again under a new initial temperature parameter, $T^0 = 1,000,000$:

Table 3.14: Number of Failures out of 1000 tries and Time of Running the 4-D Rosenbrock Function by SA under $T^0 = 1000,000$

		N_T	# of Failure	Running Time
$N = 4$, $T^0 = 1,000,000$	$r_T = 0.5$	5	621	10.0309s
		10	373	10.4677s
		15	307	12.8701s
		20	246	14.4925s
	$r_T = 0.85$	5	292	18.7825s
		10	173	29.4374s
		15	112	41.8239s
		20	75	52.7595s
		25	8	64.8964s
		30	8	75.1301s
	$r_T = 0.95$	30	6	162.5530s

Under the new initial temperature, the time required to complete SA searching still goes up as N_T increases. When comparing the two tables under two different initial temperatures T^0, it is interesting to discover that the higher temperature does not necessarily incur a lower rate of failure. On the other hand, it also illustrates that there is no certain rule on how T^0 should be set. T^0 needs to be tuned by trial so that 50% of all moves are eventually accepted.

A vital drawback of SA is its long running time. Most of the four algorithms previously discussed have running times of less than one second, particularly for $N = 2$ and the initial XGuess drawn from $[-1, 1]$. However, the running time skyrockets in SA. The shortest time for SA (when $T^0 = 200,000$, $r_T = 0.5$ and $N_T = 5$) solving the two-dimensional Rosenbrock function is around 3.29s, almost 32 times that of the fastest running time by the four traditional algorithm (UMPOL with initial XGuess falling into $[-1, 1]$). SA running time goes up as N_T increases. When r_T increases from 0.5 to 0.85, the running time more than doubles for $N_T = 5$; more than triples for N_T is either 10 or 15; and nearly quadruples for $N_T = 20$. When N changes to 4, all running time more than doubles when compared with $N = 2$ in Table 3.12 under the same r_T and N_T. When r_T is 0.95, the running time is as high as 165.1271 seconds, which is more than 1,000 times than that of any traditional algorithm for $N = 4$. Even though the failure rate does not go down with the rise in initial temperature T^0, searching time increases a little bit for most combinations of r_T and N_T.

It is not difficult to imagine that, for more complicated and ill-shaped functions (e.g., the Rosenbrock function of higher dimension), the failure rate of traditional algorithms will keep increasing if those algorithms successfully converge, and SA is much more robust in ensuring that a global optimum is found after setting appropriate inputs. The significantly longer running time is the main concern for SA.

3.5.3 SA for the Stock-Wise Model

The difference of more than 1,000 times in running time between traditional algorithms and the SA algorithm may not be noticeable for a simple question. For example, the longest running time to solve the Rosenbrock function here is only around 165 seconds. However, for an oddly-behaved model applied to a large sample, such as the SW model applied to teachers' data discussed in the last chapter, an effective calculation method must be adopted if the SA is to be introduced to ensure optimum robustness. As stated earlier, the frequency simulator requires a large sample size and a large number of Monte Carlo draws to make estimates consistent. In the SW model with the frequency simulator, after the number of Monte Carlo draws (N) reaches 10,000, the BCPOL algorithm converges and the running time is 32,059.375 seconds, which is more than eight hours. When N increases to 100,000, the running time is 256,158.281 seconds, which is nearly 71 hours. Upon applying SA instead of BCPOL to solve the optimization in the SW model simulated by the frequency simulator, the running time should be at least 1,000 times of 71 hours (almost 3,000 days), [7] which is not operational. The frequency simulator has to be replaced with a more effective

[7] Conclude from the difference in computing speed by BCPOL and SA in the example of Rosenbrock function

simulator, particularly when more modifications of SW also need to be tested and compared.

I also attempt to run the other three traditional algorithms (reported in Table 3.5) to estimate the SW model. The UMINF algorithm claims that no better result can be found except by decreasing the step tolerance, and the other two traditional algorithms fail to converge under various rational settings.[8] UMPOL shows that the maximum function evaluation is exceeded. Even when the value is set at 5,000, the algorithm fails to converge. UMCGF displays this message, " The line search of an integration was abandoned. An error in the gradient may be the cause." Adjusting the gradient tolerance from 10 to 0.1 does not alter the result.

As mentioned earlier, σ is the parameter that is most sensitive to model settings. Figure 3.12 demonstrates the relationship between the negative log-likelihood value and the values of σ when other parameters are fixed, which are the same as those in Model 3. The distribution is convex. The absolute value of log-likelihood, as shown in the y-axis, goes down sharply as σ goes up from 0. When σ goes up from around 1,000, the decreasing speed of the minus log-likelihood becomes very slow. Its minimum is reached when σ is around 3,000. After the minimum, the minus log-likelihood gradually climbs. As it is unrealistic to plot more than 3D graphs, the graph of two selected parameters and the minus log-likelihood function can reveal some characteristics of the objective function in the SW model. Figure 3.13 represents the relationship between the minus log-likelihood value and the value of both κ_1 and σ from four different angles. It clearly represents that similar to the Rosenbrock function with $N = 4$, the option value model applied to the Missouri public teachers' pensions also poses the feature of a long and flat valley; in fact, it is even longer and flatter. Almost after σ reaches 2,000 until 4,000, the graph is almost like a platform, and parameter κ_1 shows almost no impact on the altering function value.

[8]Those calculations are based on parameters from Model 3.

Figure 3.12: The Value of Minus Log-likelihood of SW Model using PSRS data at Different Values of σ with the Other Parameters are Fixed as in Model 3

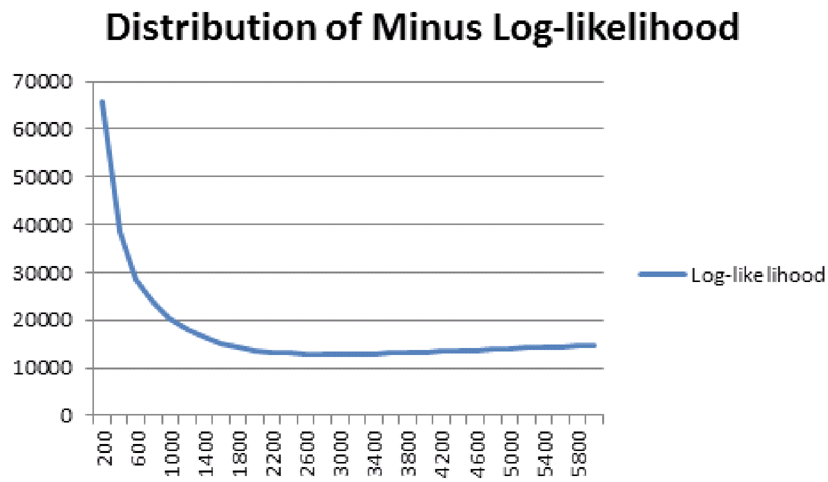
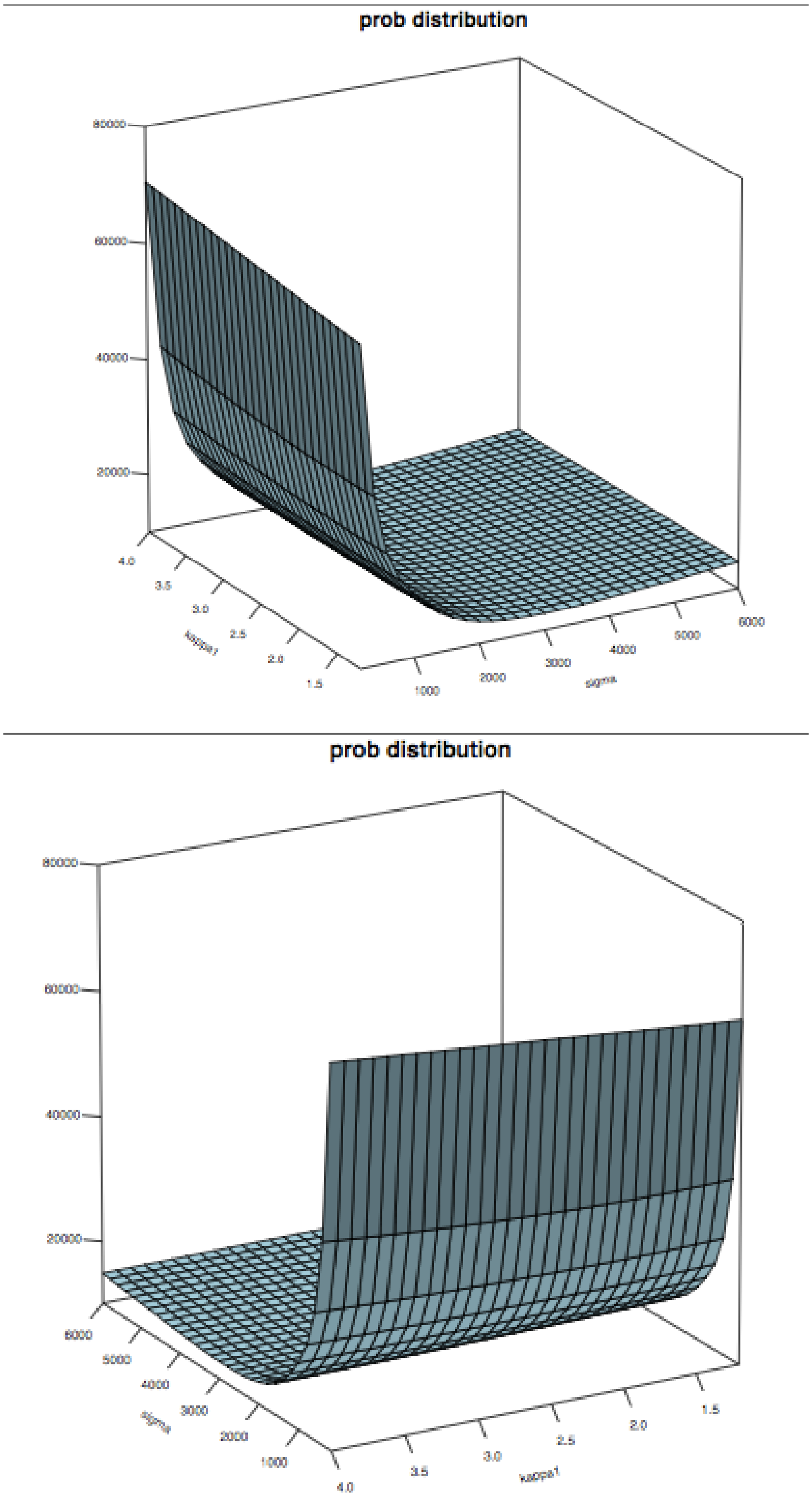
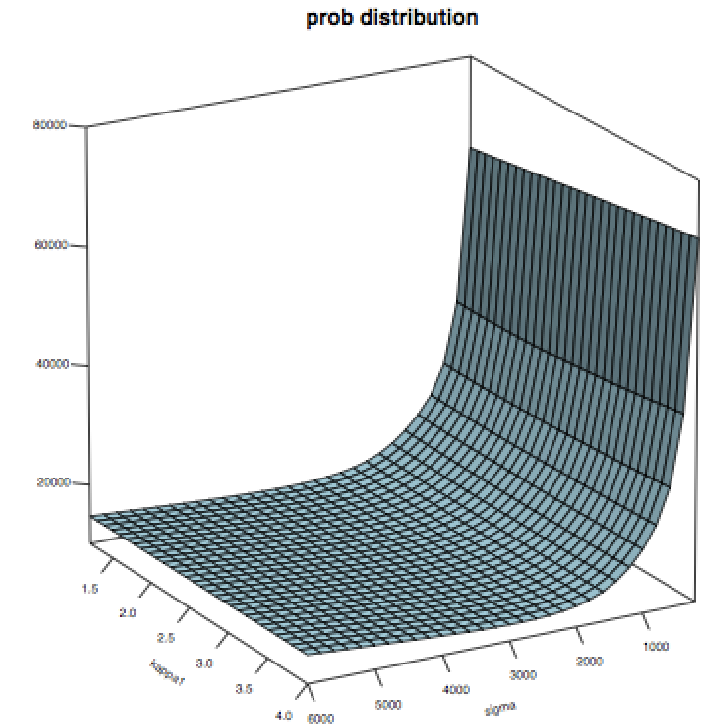
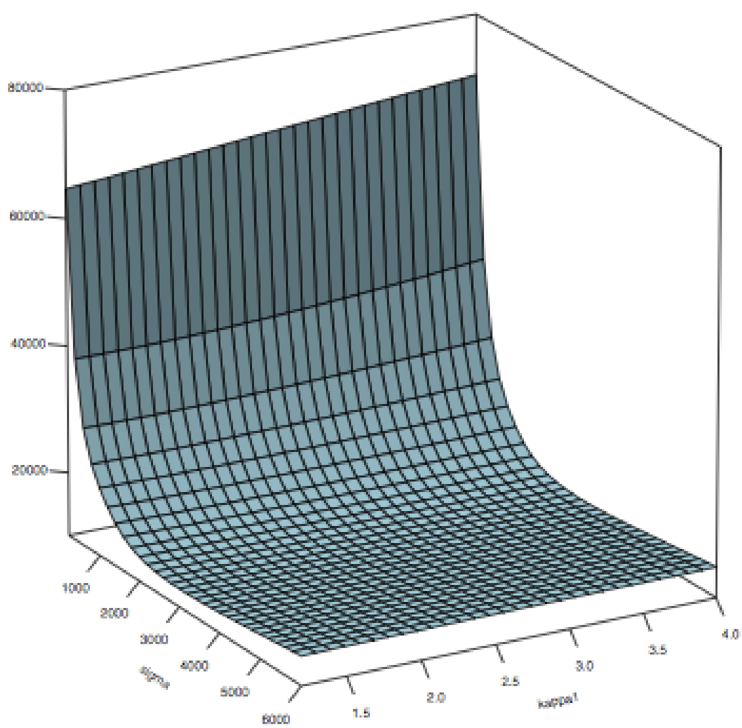


Figure 3.13: The Value of Minus Log-likelihood of SW Model using PSRS data at Different Values of κ_1 and σ with the Other Parameters are Fixed as in Model 3







The SA algorithm also fails to converge within a reasonably large number of function evaluations. The final log-likelihood values bounce around $-12,640$. When compared to the value of $-12,809.212$, which is based on estimates derived by the

GHK simulator, SA makes significant improvements as judged by the log-likelihood value. The vital question is whether those different sets of estimates by SA make a difference toward describing teachers' retirement behavior. Table 3.15 lists four sets of estimated parameters randomly retrieved from results given by the SA algorithm and returning relatively larger log-likelihoods than those given by the GHK simulator. The following figures compare actual distributions of age, experience, and the sum of age and experience with distributions simulated by the four different sets of SA estimates. I conclude from those figures that, even though the SA algorithm that is applied in the option value model bounces among different values instead of pointing at one global optimum, simulation by the four sets of SA estimates almost coincide with each other in those graphs regardless of age, experience, and the sum of age and experience. It is worth noting that, unlike the deterministic objective functions in the two earlier examples illustrating SA, the likelihood in the option value model is computed through stochastic simulation. Hence, because of sampling errors, the same set of parameters may produce different numerical values of the likelihood. This may result in difficulty in convergence for SA. Nevertheless, SA does find a better fit than the BCPOL algorithm.

Table 3.15: Comparison of Estimation Results between GHK Simulator and SA algorithm

Estimation Method	κ	β	γ	σ	ρ	$\kappa1$	Log-likelihood
GHK	0.631	0.960	0.682	2635.503	0.556	2.448	-12807.564
SA1	0.742	0.979	0.651	4241.710	0.441	3.720	-12637.630
SA2	0.795	0.970	0.637	3600.079	0.429	3.810	-12651.885
SA3	0.771	0.972	0.614	2278.390	0.473	3.257	-12664.982
SA4	0.749	0.973	0.656	4886.210	0.417	2.917	-12671.604

Note: The initial log-likelihood is calculated under the GHK estimates when $N = 100$ in table.

Figure 3.14: Comparison between Retirement Frequency by the Sum of Age and Experience Based on Actual Values and GHK Estimates and Different SA Estimates

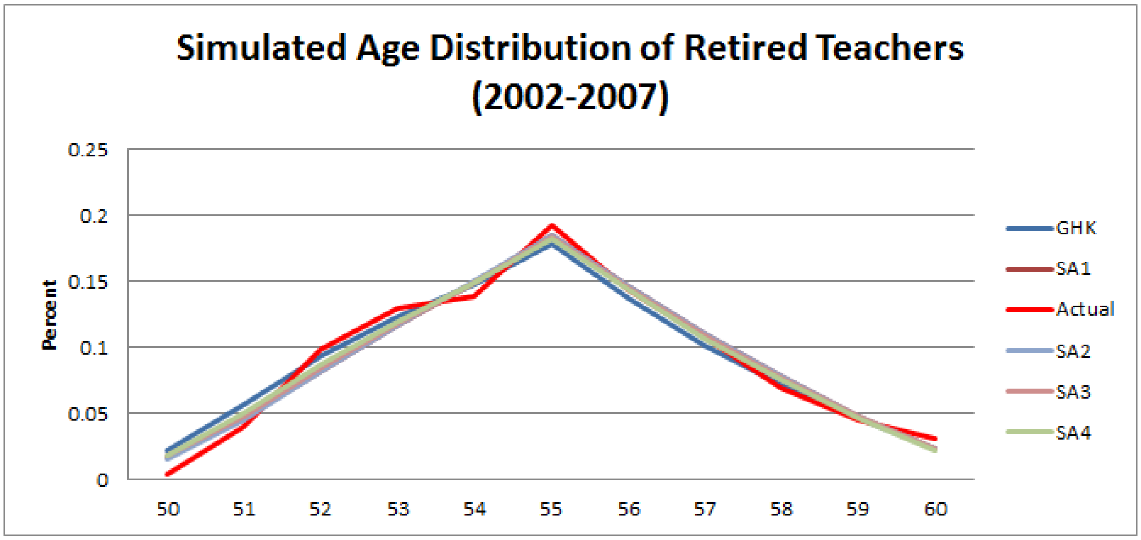


Figure 3.15: Comparison between Retirement Frequency by Age Based on Actual Values and GHK Estimates and Different SA Estimates

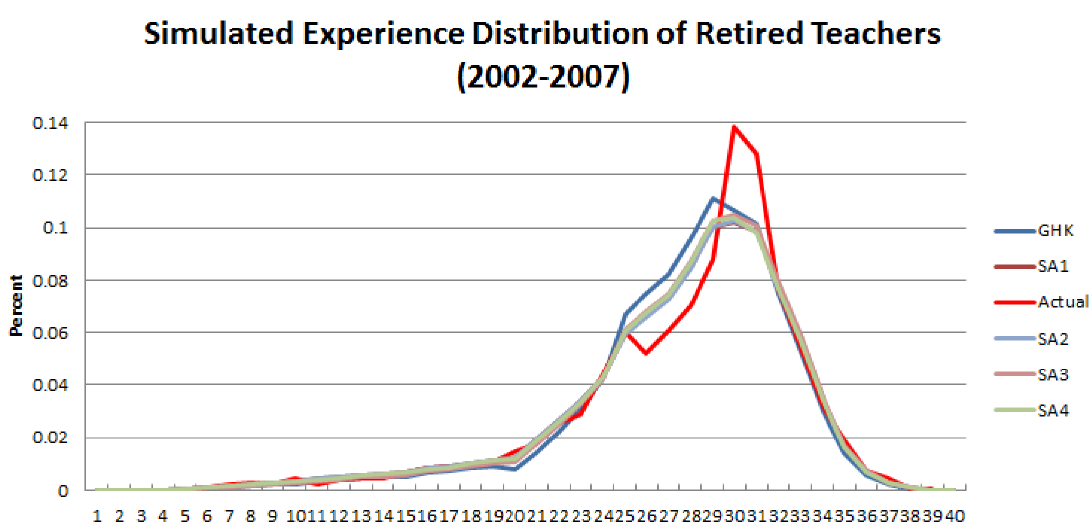
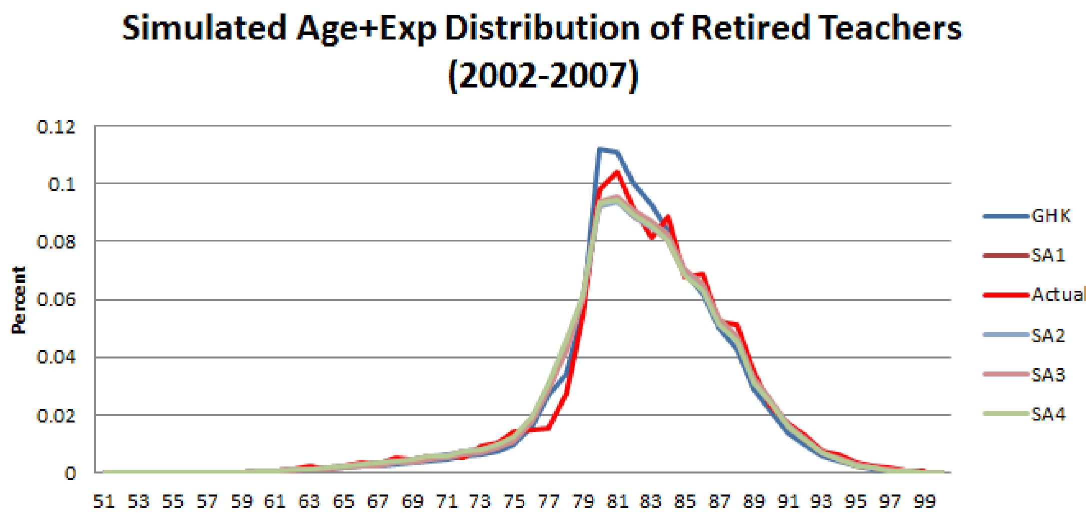


Figure 3.16: Comparison between Retirement Frequency by Experience Based on Actual Values and GHK Estimates and Different SA Estimates



3.6 Conclusion

Policy issues on whether and how to reform teachers' pension systems to facilitate attracting and retaining good teachers is under debate. When compared with the reduced form model, a structural model is more appropriate for finding answers to policy simulation for its parameters are independent of pension rules. The option value model proposed by Stock and Wise (1990) is a great benchmark model for combining the advantages of both hazards models and dynamic programming models.

When the studied sample size explodes, and the objective function is ill-behaved, the frequency simulator adopted in MLE to solve the option value model is still time-consuming. To improve efficiency in calculating the likelihood function, the GHK simulator is introduced to replace the frequency simulator. Meanwhile, to maximize the likelihood function and to ensure that the optimum detected is a global optimum, a new searching method, SA algorithm, is introduced. After comparing SA with four other traditional algorithms in maximizing different problems, I conclude that SA is more likely to detect a global optimum in an optimization problem but with a longer processing time. Equipped with the GHK simulator and SA, more behavioral models can be efficiently and reliably tested, and in this paper, teachers' retirement behavior can be more accurately simulated. In estimating the SW model on Missouri public school teachers' data, SA fails to converge but returns a different estimated result from the BCPOL algorithm, and it facilitates understanding regarding a "sensitive" parameter in this model; however, this study also reveals that the age and experience distribution of retired teachers does not change much when certain key parameters change.

Chapter 4

Comparison and Conclusion

In this paper, I apply both the probit and the option value models to study retirement issues of Missouri public school teachers. The option value model is a kind of structural model wherein parameters are independent of pension rules, making it suitable for policy simulation. Because of nonlinear objective functions and complex features of PSRS retirement rules, as well as large samples, the SW model is time-consuming. I apply the GHK simulator to efficiently evaluate the likelihood for each selected parameter. In addition, a robust optimization method, SA, is applied to ensure that the selected parameter reaches a global maximum.

Hanushek and Maritato (1996) once propose an interesting phenomenon: "It is puzzling that some structural models, such as those of Lumsdaine, Stock, and Wise (1994) and Gustman and Steinmeier (1993a), fail to capture a significant Medicare/health insurance effect as anecdotal evidence and media attention suggest that concern over health insurance is on the minds of many individuals nearing retirement. Reduced form models have been more successful at capturing a larger effect, at the expense of policy inference. Using a probit model, Gruber and Madrian (1993) find a significant effect of post-retirement health insurance on retirement, exploiting state cross-sectional variation in continuation of coverage laws." Similar to their findings, as demonstrated by the following figures, upon plotting models fit by both the probit approach and the option value approach, the former is found to fit the actual data better than the latter. The three figures list actual and simulated distribution with the option value approach, along with all GHK and SA methodologies, and also by the probit approach over age, experience, and the combination of both age and experience. The observed data (red line) matches the actual distribution more closely as opposed to the probit approach (purple line). Particularly for experience distribution, the superiority is obvious. However, the probit model tends to smooth out retirement odds over all ages, and the option value model is able to catch some turning points. The uniqueness of teachers only covered by PSRS: a single retirement system; a fixed wage schedule; and clearly defined retirement rules might be attributed to the probit model's good performance as addressed in this study. Nevertheless, the structural model simulates the retirement path by drawing a sequence of preference shocks and using the SW model to generate a whole path of retirement decisions. The only piece of information used is the initial age and experience. The probit model predicts retirement for the next year, conditional on whether the teacher is teaching this year. When a teacher retires she is out of the sample. Therefore, the prediction by the probit model is a one-step-ahead forecast, which is much easier than predicting the whole sample period. The policy question of interest is predicting the long-term effect under different rules, for which the probit model has little use.

There are still some parts of retirement behavior that are not fully captured by these two approaches. More information about Missouri public school teachers, such as health status, spouse information and race, and some other information about schools, such as class size, school grading may further contribute to the explanatory ability of models.

Figure 4.1: Observed and Simulated Age Distribution of Retirement

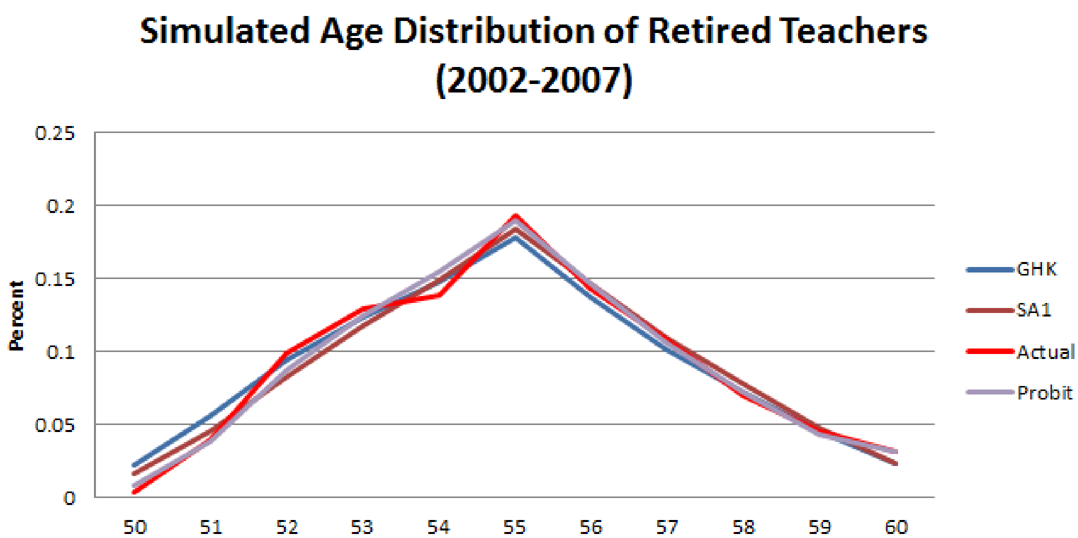


Figure 4.2: Observed and Simulated Experience Distribution of Retirement

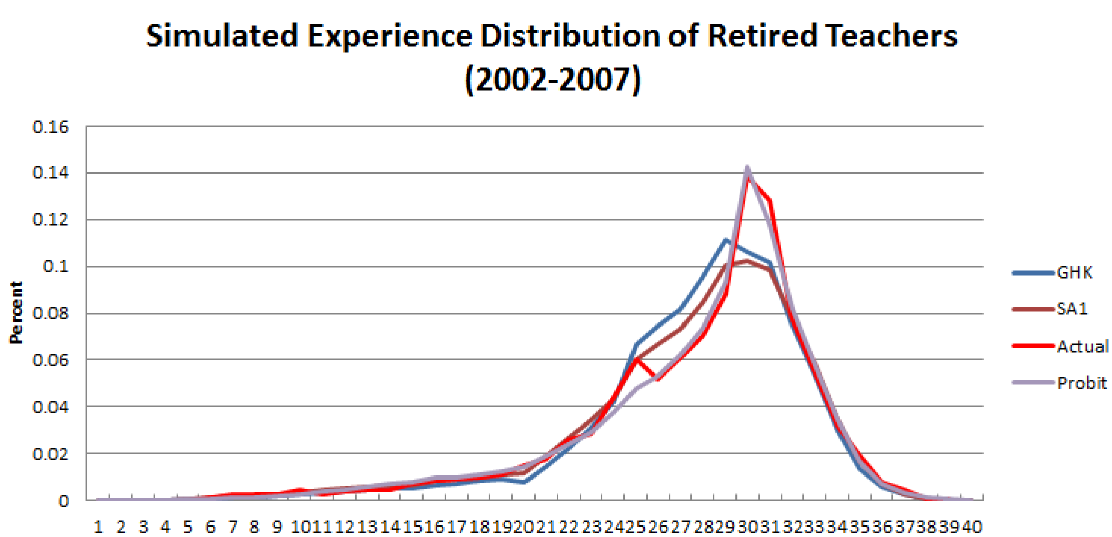
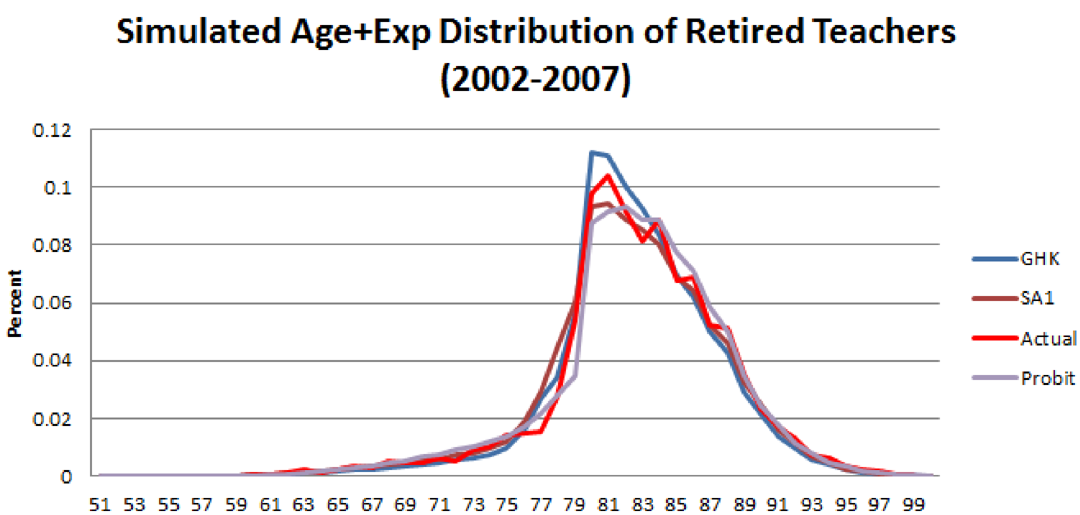


Figure 4.3: Observed and Simulated Age&Experience Distribution of Retirement



Bibliography

[1] Aaronson, Daniel and French, Eric, "The Effect of Part-Time Work on Wages: Evidence from the Social Security Rules," *Journal of Labor Economics 22* (2004), 329-352.

[2] Anderson, Patricia M., Gustman, Alan L. and Steinmeier, Thomas L., "Trends in Male Labor Force Participation and Retirement: Some Evidence on the Role of Pensions and Social Security in the 1970s and 1980s," *Journal of Labor Economics 17* (1999), 757783.

[3] Barr, Nicholas, "Reforming Pensions: Myths, Truths, and Policy Choices," *International Social Security* (2002), 3-36.

[4] Barro, Josh, Buck, Stuart, "Underfunded Teacher Pension Plans: It's Worse than You Think," *Civic Report no. 61* (April 2010).

[5] Bazzoli, Gloria J., "The Early Retirement Decision: New Empirical Evidence on the Influence of Health," *The Journal of Human Resources 20* (1985), 214-234.

[6] Beehr, Terry A., "The Process of Retirement: A Review and Recommendations for Future Investigation," *Personnel Psychology* (March 1986), 31-55.

[7] Blau, David M., "Labor Force Dynamics of Older Men," *Econometrica 62* (January 1994), 117156.

[8] Blau, David M. and D. Gilleskie., "Health Insurance and Retirement of Married Couples," *The Journal of Human Resources 14* (1979), 63-75.

[9] Belloni, Michele, "The Option Value Model in the Retirement Literature: The Trade-off between Computational Complexity and Predictive Validity," *ENEPRI Research Report No. 50* (April 2008).

[10] Borsch-Supan Alex, and Hajivassiliou, Vassilis A., "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Econometrica 58* (1994), 347-368.

[11] Burkhauser, Richard V., "The Pension Acceptance Decision of Older Workers," *Journal of Applied Econometrics 21* (2006), 935-953.

[12] Burtless, Gary, "Social Security, Unanticipated Benefit Increases, and the Timing of Retirement," *Review of Economic Studies 53* (October 1986), 781805.

[13] Burtless, Gary and Moffitt, Robert, "The Effect of Social Security Benefits on the Labor Supply of the Aged" In Aaron, H. and G. Burtless, eds., *Retirement and Economic Behavior* (1984), Washington: Brookings Institution.

[14] Chan, Sewin, and Stevens, Ann, "What You Don't Know Can't Help You: Pension Knowledge and Retirement Decision-Making," *The Review of Economics and Statistics 90* (2008), 253-266.

[15] Coile, Courtney, and Jonathan Gruber, "Future Social Security Entilements and The Retirement Decision," *Themes in the Economics of Aging*, (D. Wise, ed. Chicago: University of Chicago Press) (2001).

[16] Coile, Courtney, and Jonathan Gruber, "Future Social Security Entirlements and The Retirement Decision," *The Review of Economics and Statistics 89* (May 2007), 234-246.

[17] Costrell, M. Robert, McGee, B. Josh, "Teacher Pension Incentives, Retirement Behavior, and Potential for Reform in Arkansas", *Education Finance and Policy* (Oct 2010), 492-518.

[18] Corana, A., M. Marchesi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the ?simulated annealing algorithm", *ACM Transactions on Math-ematical Software* (1987, 13), 262-280.

[19] Daula, Thomas V. and Robert A. Moffitt, "Estimating a Dynamic Programming Model of Army Reenlistment Behavior," *In Military Compensation and Personnel Retention: Models and Evidence*, ed. C. L. Gilroy, D. K. Home, and D. Alton Smith. Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences (1991).

[20] Diamond, Peter and Gruber, Jonathan, "Social Security and Retirement in the United States," *NBER working paper no. 7258* (1999).

[21] Diamond, Peter and Hausman Jerry, "Retirement and Unemployment Behavior of Older Men," in H. Aaron and G. Burtless (Eds.), Retirement and Economic Behavior (Washington: Brookings Institution (1984).

[22] Dwyer, Debra, Mitchell, Olivia, "Health problems as determinants of retirement: are self-rated measures endogenous?," *Journal of Health Economics* (1999) ,173193.

[23] Ferguson, Joshua, Strauss, Robert P., Vogt, William B., "The effect of defined benefit pension incentives and conditions on teacher retirement decisions," *Education Finance and Policy 1 (3)* (2006),316-348.

[24] Ferreira, Pedro C. and Santos, Marcelo R., "The effect of social security, health, demography and technology on retirement," *Review of Economic Dynamics s 16* (2013),350370

[25] Fields, Gary S. and Mitchell, Olivia S., "Economic Determinants of the Optimal Retirement Age: An Empirical Investigation," *Journal of Human Resources 19* (1984), 245-262.

[26] French, Eric, "The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behaviour," *Review of Economic Studies 72* (2005), 395-427.

[27] French, Eric, and Jones, John B., "The Effects of Health Insurance and Self-insurance on Retirement Behavior" *Econometrica 79* (2011), 693-732.

[28] Friedberg, Leora, and Michael Owyang., "Not Your Fathers Pension Plan: The Rise of 401(k) and Other Defined Contribution Plans," *Federal Reserve Bank of St. Louis Review 84(1)* (2002), 23-34.

[29] Friedberg, Leora and Webb, Anthony, "Retirement and the Evolution of Pension Structure," *NBER working paper no. 9999* (2003).

[30] Goffe, William L., Gary D. Ferrier, and John Rogers, "Global Optimization of Statistical Functions with Simulated Annealing," *Journal of Econometrics 60* (1994) 65-99.

[31] Grossman,M., "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy 80* (1972), 223-225.

[32] Gruber, Jonathan, and Brigitte C. Madrian. "Health Insurance Availability and The Retirement Decision," *NBER working paper no. 4469* (1996).

[33] Gustman, Alan L. and Thomas L. Steinmeier, "A Disaggregated Structural Analysis of Retirement by Race, Difficulty of Work and Health," *The Review of Economics and Statistics 67* (August 1986), 509-513.

[34] Gustman, Alan L. and Thomas L. Steinmeier, "An Analysis Of Pension Benefit Formulas, Pension Wealth And Incentives From Pensions," *NBER working paper no. 2535* (1988).

[35] Gustman, Alan L. and Thomas L. Steinmeier, "Retirement in A Family Context: A Structural Model for Husbands and Wifes," *NBER Working Paper No. 4629* (1994).

[36] Gustman, Alan L. and Thomas L. Steinmeier, "What People Dont Know About Their Pensions and Social Security: An Analysis Using Linked Data From the Health and Retirement Study," *NBER Working Paper No. 7368* (1999).

[37] Gustman, Alan L. and Thomas L. Steinmeier, "Social Security, Pensions And Retirement Behavior Within The Family," *Journal of Applied Econometrics 19* (2004), 723-737.

[38] Gustman, Alan L. and Thomas L. Steinmeier, "The Social Security Early Entitlement Age in A Strutural Model of Retirement and Wealth" *Journal of Public Economics 89* (2005) 441-463.

[39] Hausman, Jerry A, and David A. Wise, "Social Security, Health Status, and Retirement," *in David A. Wise (Ed.), Pensions, Labor, and Individual Choice* (Chicago: University of Chicago Press, 1985).

[40] Hurd, Michael D. and Boskin Michael J., "The Effect of Social Security on Retirement in the Early 1970s," *The Quarterly Journal of Economics 99* (Nov 1984), 767-790.

[41] Judge, George G., W.E. Griffiths, R. Carter Hill, Helmut Liitkepohl, and Tsoung-Chao Lee, *The theory and practice of econometrics, 2nd ed.*(Wiley, New York, NY, 1995).

[42] Karlstrom, Anders, Palme, Marten and Svensson, Ingemar. "A Dynamic Programming Approach to Model the Retirement Behavior of Blue-collar Workers in Sweden," *Journal of Applied Econometrics 19* (2004) 795-807.

[43] van der Klaauwa, Kenneth Wolpinb, "Social Security and the Retirement and Savings Behavior of Low-income Households," *Journal of Econometrics 45* (2005), 21-42.

[44] Kopecky, Karen A., "The Trend in Retirement," *International Economic Review 52* (2011), 287-316.

[45] Krueger, Alan, and Jom-Steffen Pischke. "The Effect of Social Security on Labor Supply: A Cohort Analysis of the Notch Generation," *Journal of Labor Economics 10* (Oct 1992) 412-37.

[46] Laitner, John and Dan Silverman, "Consumption, retirement and social security: Evaluating the efficiency of reform that encourages longer careers," *Journal of Public Economics 96* (2012), 615634.

[47] Lumsdaine, Robin L., and David A. Wise. "Aging and Labor Force Participation: A Review of Trends and Explanation," *NBER working paper no. 3420* (1990).

[48] Lumsdaine, Robin L., Stock, James H., and David A. Wise. "Pension Plan Provisions and Retirement: Men and Women, Medicare and Models," *NBER working paper no. 7344* (1994).

[49] Lumsdaine, Robin L., Stock, James H.,and David A. Wise. "Three Models of Retirement: Computational Complexity versus Predictive Validity," *NBER working paper no. 7097* (1992).

[50] Lumsdaine, Robin L., Stock, James H.,and David A. Wise. "Retirement Incentives: The Interaction between Employer-Provided Pensions, Social Security, and Retiree Health Benefits," *NBER working paper no. 8649* (1996).

[51] Manoli, Day, Mullen, Kathleen J. and Mathis Wagner, "Policy Variation, Labor Supply Elasticities, and A Structural Model of Retirement," *Economic Inquiry 53* (October 2015), 17021717.

[52] McGarry, K., "Health and retirement: do changes in health affect retirement expectations?" *Journal of Human Resources 39* (2004), 624648.

[53] Mitchell, Olivia S., and Gary S. Fields, "The Effects of Pensions and Earnings on Retirement," *Research in Labor Economics 5* (Greenwich, CT: JAI Press, 1982).

[54] Ni, Shawn, and Michael Podgursky, "Teacher Pension Incentives and the Timing of Retirement," *Working Paper* (Nov 2011).

[55] Pang, Gaobo, Warshawsky, Mark and Weitzer, Ben, "The Retirement Decision: Current In?uences on the Timing of Retirement among Older Workers," *Working Paper no. 2008-04* (Univ. Pennsylvania, Pension Res. Council, 2008).

[56] Piekkola, Hannu, and Deschryvere, Matthias"Option Values for Retirement Effects of Public Incentives to Postpone Retirement in Finland, Belgium and Germany," *ENEPRI Research Report* (Aug 2005).

[57] Quinn, Joseph, "Retirement Patterns and Bridge Jobs in the 1990s," *Policy Brief for the Employee Benefit Research Institue* (Washington, DC: EBRI,1999).

[58] Rosenbrock, H. H., "An automatic method for finding the greatest or least value of a function," *The Computer Journal* (1960, 3), 175?184 *Policy Brief for the Employee Benefit Research Institute* (Washington, DC: EBRI,1999).

[59] Rust, John and Phelan, Christopher, "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica 65* (1997), 781831.

[60] Rust, John, Buchinsky, Moshe and Benitez-Silva, Hugo, "An Empirical Model of Social Insurance at the End of the Life Cycle," *manuscript University of Maryland* (2001).

[61] Samwick, Andrew A., "New Evidence on Pensions, Social Security, and the Timing of Retirement," *Journal of Public Economics 70* (November 1998), 207-236.

[62] Sickles, Robin C. and Taubman, Paul, "An Analysis of the Health and Retirement Status of the Elderly," *Econometrica 54* (Nov 1986), 1339-1356.

[63] Stinebrickner, Todd R., "Serially Correlated Wages in a Dynamic, Discrete Choice Model of Teacher Attrition',' *UWO Department of Economics Working Papers* (1998, 9821), University of Western Ontario, Department of Economics.

[64] Stinebrickner, Todd R,"A Dynamic Model of Teacher Labor Supply," *Journal of Labor Economics* (January 2001), University of Chicago Press, vol. 19(1), 196-230.

[65] Stock, James H., and David A. Wise, "Pensions, and the Option Value of Work and Retirement," *Econometrica 58* (Sepetember 1990a), 1151-1180.

[66] Stock, James H., and David A. Wise, "The Pension Inducement to Retire: An Option Value Analysis," in David A. Wise,(Ed.), *Issues in the Economics of Aging* (Chicago: University of Chicago Press, 1990b).

[67] Stern, Steven., "Approximate Solutions to Stochastic Dynamic Programs." *Econometric Theory* (1990, 13), 392-405.

[68] Sueyoshi, Glenn T., "Social Security and the Determinants of Full and Partial Retirement: A Competing Risks Analysis," *NBER working paper no. 3113* (1989).

[69] Warren, Diana and Umut, Oguzoglu, "Retirement in Australia: A Closer Look at the Financial Incentives," *Australian Economic Review* (2010), 357-375 .

VITA

I was born in China and received both my Bachelor and Master degree in Economics from Renmin University of China. During the period pursuing PHD degree in Economics at the University of Missouri-Columbia, I was awarded Harry Gunnison Brown Graduate Student Fellowship. My fields of study are Econometrics and Quantitative Microeconomic Policy Analysis.