Working Paper

Teacher Pension Workshop: Connecting Evidence-Based Research to Pension Reform

Pension Enhancements and Teacher Retirement Behavior

Wei Kong, Shawn Ni, Michael Podgursky, Weiwei Wu

RAND Education

WR-1252
May 2018

For more information on this publication, visit www.rand.org/pubs/working_papers/WR1252.html

Published by the RAND Corporation, Santa Monica, Calif.

Pension Enhancements and Teacher Retirement Behavior

Wei Kong, Shawn Ni, Michael Podgursky, Weiwei Wu

Department of Economics
University of Missouri – Columbia

May 17, 2018

Abstract

We examine how pension rule changes affected teacher retirement by estimating an option value retirement model using data from a large cohort of late career Missouri public school teachers. In so doing we offer potential solutions to several statistical challenges that often arise in estimating structural models of retirement using large panel data sets. The first challenge concerns modelling the formation of expectations of future pension rules on the part of late career teachers. The second challenge is bias induced by baseline sample selection: in baseline cohorts we only observe teachers who are still working. This is a bias that also evolves with pension rule changes. A third challenge arises from estimation using large panels of micro-data on individual teachers. The teacher-level data can be difficult to obtain and the likelihood of teacher-data is costly to compute. We address these challenges by incorporating policy expectations and sample selection directly into estimation of the likelihood function. We also show that the likelihood can be efficiently estimated by using teacher-data grouped by age and experience cells, which permits: *a*) estimating structural models of teacher retirement with data that are more widely available; *b*) use of longer panels, and *c*) dramatic reductions in computation cost. Counterfactual simulations of the estimated structural model suggest that Missouri's pension enhancements reduced the average retirement age by about 0.3 years for the 1994 cohort and by more than one year in a steady state.

Keywords: teachers' pensions, sample selection bias, expectation of policy rules

*JEL codes:*I21, J26, J38.

The authors thank the Laura and John Arnold Foundation for research support, and helpful comments from participants in the applied microeconomics seminar at Cambridge University and the Teacher Pension Workshop organized by the Rand Corporation: "Connecting Evidence-Based Research to Pension Reform." The usual disclaimers apply.

1 Introduction

During the late 1990s, pension benefits were enhanced for public K-12 teachers in many states. These enhancements caused a significant increase in pension liabilities (Koedel, et al. 2014), yet analysis of the policy's effects on the labor supply of late career teachers is limited. In this study, we examine the effect of pension rule enhancements using a large administrative panel data set for public school teachers in Missouri, who experienced pension enhancements each year from 1995 to 2002.

This study is motivated by several policy concerns. First, teachers retire earlier than other comparable professionals (Harris and Adams 2007; Kim, et al., 2017), and retaining experienced teachers in high-need districts can have positive effects on student achievement (Rivkin et al. 2005). Pension rules can affect retirement decisions and the length of teaching careers (Costrell and McGee, 2010; Brown, 2013; Knapp, et al., 2016; Ni and Podgursky, 2016).

Second, unfunded pension liabilities have caused considerable fiscal stress for many state and local governments and have generated a growing literature highlighting the need for reform (Novy-Marx and Rauh, 2011; Malanga and McGee, 2018). As calls for pension reform intensify, there is a growing need to develop behavioral models that can reliably predict retirement behavior in the presence of changes in pension rules. Because the pension enhancements occurred consecutively in a short time span of the 1990's in Missouri (and other states) and are expected to have long-lasting effects, commonly used regression-based tools for policy analysis (e.g., regression discontinuity or difference-in-differences) are not well suited for estimating their effects and, importantly, are not useful for estimating the retirement effects of pension reforms not yet implemented.

A structural model such as the option value model (Stock and Wise, 1990) of forward-looking teachers is ideal for evaluation of the pension rule changes in the 1990s and prediction of the effects of future pension reforms. The Stock-Wise option value model contains fixed parameters on teacher preferences independent of the pension rules. Such a structural model is well-suited to analyze teacher behavior under the time-varying pension rules and to explore the effects of counter-factual retirement policies (Ni and Podgursky, 2016).

We use an option value model for retirement decisions based on administrative personnel data on public K-12 teachers from academic year 1994 to 2008 to examine the implications of the pension- enhancement legislation in Missouri. Missouri public school teachers, like nearly all public school employees, are covered by a defined benefits (DB) pension system in a state-wide educator plan-the Public School Retirement System (PSRS). The administrative data for teachers allow for a clean identification of the labor supply effects of pension rules, because the late career teachers are tenured (hence exits are voluntary) and future teacher salaries are highly predictable.

Missouri's pension rule changes present three general challenges in estimating an option value model that have not been addressed in the pension research literature. Since these problems arise in nearly all applied studies of retirement, we believe that the solutions employed in this paper represent useful contributions to the empirical retirement literature. The first problem concerns how one models teacher expectations of future pension enhancements. Pension rules change over time. Under a myopic expectation assumption teachers expect the current rules to be unchanged in the future, and all future enhancements are surprises. However, in reality if teachers expect enhancements in the near future they may postpone retirement, which leads a myopic model to over-predict near-term retirement. An alternative is perfect foresight, in which pension enhancements in the near future are perfectly forecasted. Other options are a hybrid of the two expectation models, or adaptive expectations that weigh by the probabilities of the current rules and the rules of the near future. We compare the fit of these competing models of expectations in our panel data set.

A second estimation challenge is caused by the presence of sample selection bias in the baseline sample. A sample selection bias arises due to the fact that in the initial period some teachers were eligible for retirement but we only observe those who chose to continue working. The initial sample selection bias affects the remaining teachers in each subsequent year, where the sample selection evolves with pension rule changes. We model unobserved factors affecting retirement as serially correlated preference errors. A positive value of the error means the teacher has unobserved reasons that favor staying (given age and experience). The sample selection bias is addressed by deriving the distribution of the preference errors in the initial period as a function of age and experience conditional on being observed in the initial sample.

The third challenge is the high cost in using large panels of teachers tracked for many years. There are two aspects of the high cost. First, it may be difficult or impossible to acquire teacher-level data in states that are reluctant to share such data with researchers. However, it is routine for state and local pension plans to prepare aggregated data files for actuarial cost calculations. Of course with aggregated data, individual confidentiality is more readily maintained. Second, the likelihood using teacher-level panel data is costly to compute. In this paper, we develop an algorithm that allows for efficient computation of the likelihood by using grouped data by age and experience cells. Instead of tracking individual teachers (as in Stock and Wise, 1990 and Ni and Podgursky, 2016), we track the counts of (age, experience) cells. The algorithm utilizes the fact that teacher retirement DB rules are dependent on age and experience. There is no loss of information if no additional covariates are used (e.g., characteristics of the teacher, school or district). This procedure makes it computationally feasible to exploit large panels of teachers (or other employees) from administrative data sets tracked over many years. Since these types of teacher data are available for many states, our cell-based approach facilitates more widespread estimation and simulation using structural retirement models.

We find that the estimated model fits Missouri data very well in- as well as out-of- sample. From a policy perspective, we find that the pension enhancements enacted during the 1990s had the effect of shortening teacher careers. For the cohort of teachers in the 1994 sample, we estimate that enhancements reduced the average career by 0.3 years. This estimate understates the long-term effects because the enhancements did not affect all teachers in the 1994 cohort. We also considered the long run in a steady state, where the retired teachers are replaced by senior teachers. In the steady state, the enhancement reduced a typical career by more than one year. Finally, we find the enhancements benefited late career teachers unevenly, even when they have similar age and experience.

2 Teacher Pension Rules and Enhancements

In a DB plan, it typically takes 3-5 years for teachers to become vested in the system. Once vested, a teacher can collect her pension upon becoming collection eligible. The normal retirement age is one way that collection eligibility is determined. Minimum retirement ages vary across plans, typically between the ages of 55 and 65. Retirement eligibility can also be based on service years (e.g., 30 years of service), or combinations of age and experience. There are also early-retirement provisions in most systems that allow individuals to retire and begin collecting a reduced benefit prior to normal retirement.

In this paper we focus on Missouri teachers in the state pension plan. Under the current rules, Missouri teachers become eligible for a full pension if they meet one of three conditions: a) they are sixty years of age with at least five years of teaching experience, b) thirty years of experience (and any age), or c) the sum of age and years of service equals or exceeds 80 ("rule-of-80"). Benefits at retirement are determined by the following formula (some variant of which is nearly universal in teacher DB systems): Annual Benefit=S $\times$ FAS $\times$ rf, where S is service years (essentially years of experience in the system), FAS is final average salary (calculated by the average of the highest three years of salary,) rf is the replacement factor. There is a cost of living allowance (COLA). Its increases were capped at a percent of the initial retirement annuity pursuant to the rules of the pension system. Teachers earn 2.5% for each year of teaching service up to 30 years. Thus, a teacher with 30 years experience and a final average salary of $\$60,000$ would receive a $30 \times \$60,000 \times 0.025 = \$45,000$ annuity. There are several other minor adjustments to the formula. There is a "25-and-out" option that permits retirement at a reduced rate if teachers have 25 or more years of experience. Finally, the replacement factor rf is 2.5% for experience up to 30 years and 2.55% for experience of 31 or more years. The 2.55% at 31 years is paid on the 30 inframarginal years as well. Thus the increase in the annuity for the 31st year is $2.55 + .05\,(30) = 4.05\%$.

2.1 History of Pension Rule Changes

The pension rules are not fixed. Between 1995 and 1999, state and local pension funds experienced increases in their funding ratios. The actuarial surpluses (or small deficits) were used to justify legislation in most states that enhanced pension-benefit formulas for public workers. Educator pensions were among the most-actively enhanced. For example, the National Conference of State Legislators (NCSL) reports that educator pensions were enhanced in more than half the states. In most states teachers' benefits were automatically and retroactively adjusted to reflect the enhancements at the time of their enactment without additional required contributions. Therefore, teachers whose retirement plans happened to coincide with the timing of the benefit enhancements were able to collect the more-generous pensions even though their lifetime contributions were structured to fund a much less remunerative flow of benefits. The benefits by the enhancement accrued to teachers immediately upon enactment of the enhancement legislation because the enhancements were implemented retroactively. Table 1 describes the series of enhancements that occurred in the Missouri PSRS. In 1995 the formula factor was 0.023, final average salary was calculated based on the highest five years of earnings, and early retirement was possible through the 55-25 rule. The 55-25 rule allowed for a teacher to retire and collect benefits without penalty if two conditions were met: (1) the teacher had to be at least 55 years old, *and* (2) he/she had to have accrued at least 25 years of system service. By 2002 the formula factor had been raised from 0.023 to 0.025, the final-average-salary calculation changed from the highest five to highest three years of earnings, and the 25-and-out and Rule-of-80 provisions had been incorporated into the system (Rule-of-80 is a more-flexible version of the 55-25 rule whereby retirement with full benefits can occur if age + experience sums to 80). In addition, the cap on post-retirement cost-of-living adjustments (COLAs) was raised from 65 to 80 percent of the baseline annual pension payment, and a retroactive bonus was added for teachers who reached their 31st year of system service.

(Insert Table 1 here.)

2.2 Expectations Regarding Pension Rule Changes

Note that pension benefits are determined by the rules in place at the time of retirement. After retirement the benefit will not be adjusted if pension rules change. The retirement decision depends on expectations of future pension rules since these rules are important to determine retirement benefits. Our sample period spans the years in Table 1. After a couple of pension enhancements, teachers may forecast additional enhancements. Because teachers may be forward looking, the effects of the rule changes depend on whether they are anticipated. The typical official retirement date is July 1. The year we use is the academic year (AY). Hence AY t starts on July 1 of the calendar year $t-1$. The rule R_t, effective in year t, applies to retirement filed before the end of AY t. In most cases the pension rule changes were effective on July 1. Some of the important changes in rules were introduced at the beginning of the AY. For example, the benefit rate was raised to 2.5% from 2.3% on 7/1/1998 (the start of AY 1999); teachers retired on 7/1/1998 may not know the rule change when they made the decision prior to the retirement date. Similarly, the rule-of-80 was introduced on 7/1/1999 (the start of AY 2000). In some cases major changes were known to teachers in the middle of the AY. For instance, the 25-and-out rule effective 7/1/1996 (and covers the AY 1997) was introduced on 8/28/1995. Teachers who considered retirement in 1996 had knowledge of the rule change almost one year ahead.

3 Simulating Retirement Decisions under Changing Pension Rules

Our focus is on the timing of retirement. We assume that an experienced educator teaching in the current year has two choices: teach next year or retire (stop teaching and collect a pension immediately or at a future date.) The salary schedule (as a function of experience) is known and fixed. The retirement decisions depend on the expectations of future pension rules.

The option value model of Stock-Wise assumes that a teacher chooses the year of retirement that maximizes the expected present value of the utility of the salary and benefit flows given current information but does not take into account the value of options in the future, as does the corresponding problem in a dynamic programming setting. A wealth of studies examine retirement decisions by simulating or estimating structural models through dynamic programming (e.g., Rust and Phelan 1997, French and Jones 2011). An earlier study by Gustman and Steinmeier (1986) estimates a life-cycle model with time-invariant preference errors. The model in this study borrows heavily from Stock and Wise, and is simplified by the lack of need to model Social Security (because PSRS teachers are not in Social Security). Compared to dynamic programming, the simplicity of the option-value model affords several key benefits. First, it allows for lower computation cost of the likelihood function in the presence of complications, such as the consecutive changes in policy in the 1990s and serial correlation of preference errors (based on empirical evidence later presented in this paper). In addition, solving the initial condition requires accurately computing the probability that a teacher appears in the initial sample. Following Stock and Wise we also assume normality in preference errors, which allows us to compute the likelihood of the initial condition jointly with the likelihood of sample of panel data. Numerical solutions to a dynamic programming problem with the complications in this paper will involve approximation error.[1] Lumsdaine et al. (1992) argue that the option-value model yields similar prediction as dynamic programming. We choose to focus the option-value model after weighing the tradeoffs.

3.1 Tracking the Panel Data of (Age, Experience) Cells

Let $N(a, e, t)$ be the number of teachers with (age, experience) (a, e) at the end of period t, and $r(a, e, t)$ be the retirement probability in period t. Structural models such as the Stock-Wise model of retirement choice dictates how $r(a, e, t)$ changes with a change in rules. We assume $a \geq a^l$ and $e^l \leq e \leq a - a^l$, where a^l is a minimum age in the sample, e^l is the minimum experience in the sample. For the 1994 sample $a^l = 47, e^l = 5$. The teacher

[1] Emprical evidence strongly support serial correlation in the preference errors. Since the retirement decision depends on the preference error, the serial correlation implies that the distribution of preference errors is age- and experience-dependent. Because dynamic programming problems are solved by backwards recursion, the computation cost is high when distribution of preference errors depends on age and experience in a complicated way. Innovations in numerical approximation methods may reduce the computational cost of dynamic programming problems (e.g., see Keane and Wolpin 1997 and Stinebrickner 2000). The Stock-Wise model is solved forwards, which is much less costly.

distribution in period t is given by

$$N(a, e, t) = N(a-1, e-1, t-1)[1 - r(a, e, t)],$$

for $a \geq a^l + t$ and $e^l + t \leq e \leq a + t - a^l$.

Starting from an initial distribution in the beginning of $N(a, e, 0)$ in period 1, in the beginning of period $t \geq 1$

$$N(a + t, e + t, t) = N(a, e, 0)s_t(a, e),$$

where

$$s_t(a, e) = \prod_{j=1}^{t}[1 - r(a + j - 1, e + j - 1, j)]$$

is the survival rate after period t of a teacher with initial age and experience (a, e). Again, $a \geq a^l + t$ and $e^l + t \leq e \leq a + t - a^l$.

The probability of a teacher with initial age and experience (a, e) in period t retiring in period $t > 1$ is

$$G_{1,t}(a, e) = r(a + t - 1, e + t - 1, t)s_{t-1}(a, e).$$

The probability she remains teaching at the end of the sample period is $s_T(a, e) = 1 - \sum_{t=1}^{T} G_{1,t}(a, e)$. The retirement hazard $r(a + t - 1, e + t - 1, t)$ conditional on the initial (a, e), is determined by a structural model for a given set of parameters. Later in the paper we estimate the parameters of a Stock-Wise option value model using the data for Missouri teachers.

One way to summarize the data is by tracking the joint probability of retirement of each cell with the initial (a, e), for period $t = 1, 2, .., T$, and write the likelihood function as

$$\prod_{a \geq a^l, e^l \leq e \leq a - a^l} \prod_{t=1}^{T} [G_{1,t}(a, e)]^{N(a,e,0)G_{1,t}(a,e)}[s_T(a, e)]^{N(a,e,0)s_T(a,e)}. \quad (1)$$

The empirical counterpart of the quantity (1) is the frequency of teachers with initial age and experience in period $t = 0$ who retire in year $t = 1, 2, ..., T$; (and the implied probability of teaching after T.) The observed number of teachers in year t is $\hat{N}(a, e, t)$.

The total retirement in period $t \geq 1$ from the sample is given by

$$R(t) = \sum_{a \geq a^l + t - 1} \sum_{e \geq e^l + t - 1}^{a - a^l + t - 1} N(a - 1, e - 1, t - 1) r(a, e, t).$$

3.2 The Option Value Model under Different Expectations of Future Pension Rules

We will use the following notation for expectations of future pension rules. Recall by our earlier notation, rule R_t, effective in year t, applies to retirement filed before the end of AY t. So when a teacher makes a retirement decision in period t (AY t) on whether to retire at the beginning of AY $t + 1$, we assume she uses one of the following to calculate the pension benefit.

(M) Myopic (rule R_t).

(P1) One step perfect foresight (rule R_{t+1}).

(A1) Adaptive expectation 1 (No change for years 1995 and 1996, then they follow (M) with probability p and follow (P1) with probability (1-p)).

The adaptive learning in expectations on future rule changes is based on the following assumption: Because 1995-1996 are the first two years of major pension rule enhancements, teachers may not expect continuous enhancements at the time. We assume retirement decisions in 1995 and 1996 are based on the current pension rules in 1995-1996. But by 1997, after consecutive enhancements, teachers may anticipate more generous pension rules in the future. So, after 1997, we assume teachers give the current rule weight p and rule of the next year with weight $1 - p$. This case is labeled as (A1). One may interpret the assumption that with probability p a teacher does not pay attention to the news on pension rules or is committed to retirement based on the current rules, and with probability $1 - p$ she is informed with the rule changes in the next AY and has the flexibility to make a retirement decision based on the new information. Hence under (A1): in t=1995 and 1996, teachers calculate pension benefits using rule R_t; in $1997 \leq t \leq 2002$ teachers calculate expected pension benefits using rules R_t with probability p and rules R_{t+1} with probability $1-p$; and after $t \geq 2003$ teachers calculate pension benefits using rule R_t (which is the same as R_{t+1}).

3.2.1 Retirement decision under myopic expectations of pension rules (M)

We introduce the model under the case of myopic expectations and assume that the teachers calculate the expected future pension benefit based only on the current rules. Applying the Stock-Wise (SW) model to teacher retirement, we first write the teacher's expected utility in period t as a function of expected retirement in year m (with $m = t, \cdots, T$ and $T = 101$ is an upper bound on age). Denote year t the prevailing pension rules as R_t, the teachers' contribution rate as c_t. In year t, the expected utility of retiring in period m is the discounted sum of pre- and post-retirement expected utility

$$\mathbb{E}_t V_t^M(m, R_t) = \mathbb{E}_t\{\sum_{s=t}^{m-1} \beta^{s-t}[(k_s(1-c_t)Y_s)^\gamma + w_s] + \sum_{s=m}^{T} \beta^{s-t}[(B_s(R_t, m))^\gamma + \xi_s]\}, \qquad (2)$$

where Y_s is real salary in period s. $B_s(R_t, m)$ is the real pension benefit collected in year s under the rules of year t, R_t, if the teacher retires in year $m \geq t$. The salary is a function the teachers's experience, and the pension benefit depends on the teacher's age, experience, and the pension rules. The superscript "M" indicates the myopic expectation of future pension rules. For notational simplicity we do not specify the age and experience of the teacher. The parameter k_s captures the dis-utility of working. We assume k_s to be decreasing with age: $k_s = \kappa(\frac{60}{age})^{\kappa_1}$, where age is the age in period s. One dollar of salary is worth k_s dollar of pension benefit and we expect $0 < k_s < 1$.

The unobserved innovations in preferences are AR(1): $w_s = \rho w_{s-1} + \epsilon_{ws}$, $\xi_s = \rho \xi_{s-1} + \epsilon_{\xi s}$. Denote the error terms $\nu_s = w_s - \xi_s$, $\epsilon_s = \epsilon_{ws} - \epsilon_{\xi s}$. Then it follows that:

$$\nu_s = \rho \nu_{s-1} + \epsilon_s. \qquad (3)$$

We assume ϵ_s is iid $N(0, \sigma^2)$. The retirement decision in year t is choosing $m = t, \cdots, T$ that maximizes $\mathbb{E}_t V_t^M(m, R_t)$.

This is termed an "option value" model since the retirement decision is irreversible. Because the future is uncertain and the teacher is risk averse, there is a value associated with continuing teaching and keeping the retirement option open.

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

$$G_t^M(m, R_t) = \mathbb{E}_t V_t(m, R_t) - \mathbb{E}_t V_t(t, R_t) = g_t(m, R_t) + K_t(m)\nu_t, \tag{4}$$

where

$$\begin{aligned} g_t^M(m, R_t) = & \sum_{s=t}^{m-1} \pi(s|t)\beta^{s-t}\mathbb{E}_t(k_s(1-c_t)Y_s)^\gamma + \sum_{s=m}^{T} \pi(s|t)\beta^{s-t}\mathbb{E}_t(B_s(R_t, m))^\gamma \\ & - \sum_{s=t}^{T} \pi(s|t)\beta^{s-t}\mathbb{E}_t(B_s(R_t, t))^\gamma \end{aligned} \tag{5}$$

is the difference in expected utility between retiring in year $m > t$ and retiring now (in year t). A closer look at the last two terms in (5) sheds light on the trade-off on delaying retirement. By retiring in year $m > t$, the teacher receives a higher annuity: ($B_s(R_t, m) > B_s(R_t, t)$ for all $s \geq m$), but she receives the benefit for $m - t$ fewer years, and draws a salary with an increasingly high discount rate for the disutility of working.

Because the teacher's future salary and pension benefits are very predictable, in the empirical analysis we replace the expected salary and benefit in $g_t(m, R_t)$ with a forecast based on historical data. In the last term in (4), $K_t(m) = \sum_{s=t}^{m-1} \pi(s|t)(\beta\rho)^{s-t}$ depends on unknown parameters and the AR(1) error term ν_t given in (3). Let $m_t^\dagger(R_t) = argmax\ g_t(m, R_t)/K_t(m)$, then the probability that the teacher retires in period t ($G_t(m, R_t) \leq 0$ for all $m > t$) is

$$Prob(\frac{g_t^M(m^\dagger, R_t)}{K_t(m^\dagger)} \leq -\nu_t). \tag{6}$$

3.2.2 Under perfect foresight of pension rules (P)

Under perfect foresight of future pension rules the expected utility in period t of retiring in period m becomes

$$\mathbb{E}_t V_t^P(m, R_t) = \mathbb{E}_t \{ \sum_{s=t}^{m-1} \beta^{s-t}[(k_s(1-c_s(R_m, m))Y_s)^\gamma + w_s] + \sum_{s=m}^{T} \beta^{s-t}[(B_s(R_m, m))^\gamma + \xi_s]\}, \quad (7)$$

The superscript "P" indicates the perfect foresight expectation of future pension rules. A more empirically plausible expectations of pensions rules are foresight of the next year. A 1-year-ahead perfect foresight (P1) means a teacher retiring on 7/1/1998 perfectly anticipated the increase of benefit rate to 2.5% from 2.3% when she made the decision before 7/1/1998.

Under (P1), the gain in expected utility by retiring in year $m \geq t+1$ over retiring now (in year t) $\mathbb{E}_t V_t(m, R_t) - \mathbb{E}_t V_t(t, R_t)$ can be written as

$$\begin{aligned} g_t^{P1}(m, R_t) = \quad & \sum_{s=t}^{m-1} \pi(s|t)\beta^{s-t}\mathbb{E}_t(k_s(1-c_s(R_{t+1}, m))Y_s)^\gamma + \sum_{s=m}^{T} \pi(s|t)\beta^{s-t}\mathbb{E}_t(B_s(R_{t+1}, m))^\gamma \\ & - \sum_{s=t}^{T} \pi(s|t)\beta^{s-t}\mathbb{E}_t(B_s(R_t, t))^\gamma. \end{aligned}$$

3.2.3 Under adaptive expectation of pension rules (A1)

Suppose in period t the teacher assumes that there is a probability p that the current rule R_t prevails in a future period $m > t$, and a $1-p$ probability that in period m the rule of the next year R_{t+1} replaces the current rule. So the myopic case corresponds to $p = 1$ and the perfect foresight of the next year's rule corresponds to $p = 0$. We label this adaptive expectation of future pension rules by a superscript "A1".

The case (A1) is a combination of the case (M) and the case (P1). In the setting of the previous example, $\mathbb{E}V_t(m, R_t)$ in year $t \in [1995, 1996]$ for any $(m > t)$ $p_t = 1$. After 1996, $\mathbb{E}V_t(m, R_m)$ in year t $(t > 1996)$ for any $(m > t)$ $p_t = p$; and $t(t > 2002)$ for any $(m > t)$ $p_t = 1$ (which is equivalent to $p_t = 0$ in the absence of rule changes.)

Under assumption (A1) the difference in expected utility between retiring in year $m > t$

and retiring now (in year t) is

$$
\begin{aligned}
&\mathbb{E}_t V_t^{A1}(m,R)\\
&= p\mathbb{E}_t\{\sum_{s=t}^{m-1}\beta^{s-t}[(k_s(1-c_s(R_t,m))Y_s)^\gamma + w_s] + \sum_{s=m}^{T}\beta^{s-t}(B_s(R_t,m))^\gamma\}\\
&+(1-p)\mathbb{E}_t\{\sum_{s=t}^{m-1}\beta^{s-t}[(k_s(1-c_s(R_{t+1},m))Y_s)^\gamma + w_s] + \sum_{s=m}^{T}\beta^{s-t}B_s(R_{t+1},m))^\gamma + \xi_s]\},
\end{aligned}
$$

3.3 Likelihood by Age-Experience Cells

The condition (6) is affected by the time-varying pension rules. Denote

$$
f_t^+ = \frac{g_t(m^\dagger,R_t)}{K_t(m^\dagger)} = max_{m\geq t}\{\frac{g_t(m,R_t)}{K_t(m)}\}. \tag{8}
$$

We omit the superscript on $g_t(m^\dagger,R_t)$ with respect to pension rules. Suppose a teacher i is observed for period $1,2,..,n_i$.

Denote her preference error as $\boldsymbol{\nu}_{1,n_i} = (\nu_1,\cdots,\nu_{n_i})'$ and $\mathbf{f}^+_{1,n_i} = (f_1^+,...,f_{n_i}^+)'$. If she retired in n_i ($n_i < T$) then the observations on teacher i imply $\boldsymbol{\nu}_{1,n_i} > -\mathbf{f}^+_{1,n_i}$. $\boldsymbol{\nu}_{1,n_i} \in A_i \subset R^{n_i}$, where A_i is defined by the joint event $\{\boldsymbol{\nu}_{1,n_i-1} > -\mathbf{f}^+_{1,n_i-1}\bigcap(f^+_{n_i} \leq -\nu_{n_i})\}$. If the teacher did not retire at the end of the sample then $n_i = T$ and $\boldsymbol{\nu}_{1,T} > -\mathbf{f}^+_{1,T}$. Denote the probability of that the teacher observed in period 1 retires in n_i as

$$
G_{1,n_i} = prob(\boldsymbol{\nu}_{1,n_i-1} > -\mathbf{f}^+_{1,n_i-1}\bigcap(f^+_{n_i} \leq -\nu_{n_i})|\textit{in the initial sample}).
$$

The probability of not retiring is $prob(\boldsymbol{\nu}_{1,T} > -\mathbf{f}^+_{1,T}|\textit{in the initial sample}) = 1 - \sum_{j=1}^{T} G_{1,j}$.

The likelihood of teacher $i = 1,2,..,I$ data is

$$
L(\gamma,\kappa,\kappa_1,\beta,\sigma,\rho \mid \mathbf{Y},\mathbf{B},\mathbf{D}) = \prod_{i=1}^{I}\int_{A_i}\phi(\boldsymbol{\nu}_{1,n_i})d\boldsymbol{\nu}_{1,n_i}, \tag{9}
$$

where $\phi(.)$ denotes multivariate normal density distribution of $N(0,\boldsymbol{\Sigma}_i)$, and $\boldsymbol{\Sigma}_i$ is the covariance matrix of $\boldsymbol{\nu}_{1,n_i}$. Evaluating the likelihood involves computing high dimensional integration for each teacher. In practice, the number of teachers I is quite large for large states.

Computing the likelihood of the sample can be quite costly.

For a teacher's retirement decision, the observable variables are few (age, experience, gender). If age and experience are sufficient statistics of teacher's retirement incentive under a given set of pension rules then instead of tracking the decisions of each teacher, we track the (age, experience) cells, based on the following analysis. The sample period is from period 1 to T. The retirement decision of teacher i is $y_{it} \in \{0,1\}$, $t = 1,..,T$; where $y_{it} = 1$ if the teacher retires in period t, and $y_{it} = 0$ otherwise. Hence $\sum_{t=1}^{T} y_{it} = 1$ if the teacher retires and $\sum_{t=1}^{T} y_{it} = 0$ if she remains teaching at the end of the sample period.

Denote the probability of retiring in period t, conditional on age and experience in the initial sample period, and other information x by $G_{1t}(a,e,x)$. The likelihood function of parameter $\boldsymbol{\theta}$ of teacher i is

$$L_i(\boldsymbol{\theta}, \mathbf{y}_i) = \prod_{t=1}^{T} G_{1t}(a_i, e_i, x_i)^{y_{it}} (1 - \sum_{t=1}^{T} G_{1t}(a_i, e_i, x_i))^{1-\sum_{t=1}^{T} y_{it}}.$$

The likelihood of the whole sample is

$$\begin{aligned} L(\boldsymbol{\theta}, \mathbf{y}) \quad &= \prod_{i=1}^{I} L_i(\boldsymbol{\theta}, \mathbf{y}_i) = \prod_{i=1}^{I}\prod_{t=1}^{T} G_{1t}(a_i, e_i, x_i)^{y_{it}} (1 - \sum_{t=1}^{T} G_{1t}(a_i, e_i, x_i))^{1-\sum_{t=1}^{T} y_{it}} \\ &= \prod_{t=1}^{T} \prod_{i|a_i=a, e_i=e, x_i=x} G_{1t}(a,e,x)^{\sum_{i|a_i=a,e_i=e,x_i=x} y_{it}} (1 - \sum_{t=1}^{T} G_{1t}(a,e,x))^{\sum_{i|a_i=a,e_i=e,x_i=x}(1-\sum_{t=1}^{T} y_{it})}. \end{aligned}$$

We denote $N(a,e,x,0)$ as the number of teachers with (age, experience, other information) (a,e,x) in the initial period, we now denote the counts of retirement of teachers with initial (a,e,x) in period t as $R(a,e,x,t) = \sum_{i|a_i=a,e_i=e,x_i=x} y_{it}$, then

$$L(\boldsymbol{\theta}, \mathbf{y}) = \prod_{t=1}^{T} \prod_{a,e,x} G_{1t}(a,e,x)^{R(a,e,x,t)} (1 - \sum_{t=1}^{T} G_{1t}(a,e,x))^{N(a,e,x,0)-\sum_{t=1}^{T} R(a,e,x,t)}. \qquad (10)$$

The alternative expression of the likelihood (10) suggests that we only need to track the retirement counts of (a,e) cells, instead of tracking the panel data of individual teachers.

3.4 Adjusting for Sample Selection Bias by Using the Distribution of Initial Preference Errors

Figures 1 and 2 plot marginal distributions of the initial sample of the 1994 cohort by age and by experience and the joint distribution of age and experience. The plots show that about 15% of the teachers in 1994 are eligible for retirement.

(Insert Figures 1-2 here.)

In the option value model, the retirement condition for every teacher in each year is condition (6). The preference shock ν_t follows an AR(1) process (3). The key question is how to assign the initial value ν_1. The sample data in the initial year 1994 include all teachers of age 47-64 in that year. The data set *includes* teachers who were eligible for retirement but who chose to wait, but *excludes* those who chose to retire prior to 1994 (where $t = 1$). For the 'stayers' in 1994

$$\nu_0 \geq -\frac{g_0(m^\dagger, R_0)}{K_0(m^\dagger)}. \tag{11}$$

This implies that the initial value ν_1 differs from the unconditional stationary distribution $N(0, \frac{\sigma^2}{1-\rho^2})$. Without taking into this sample selection bias, one would draw the initial value ν_1 from the unconditional stationary distribution. This will result in over-prediction of retirement in the initial years.

There no simple solution to the sample censoring problem in our panel since starting with a younger base-year sample (e.g., 40-45 in 1994) means that the majority of the teachers would still be ineligible to retire at the end of the panel, and "early leavers" would have been over-represented among the retirees. Moreover, with a younger cohort some teachers are more likely to have left the sample for reasons other than retirement.

To compute the probability of retiring in period $n > 0$, we need to take into account the fact that a teacher in period 1 was eligible for retirement J periods ago has a preference error in period 0 different from the unconditional distribution $N(0, \frac{\sigma^2}{1-\rho^2})$. The longer a teacher becomes retirement eligible in the initial sample the more likely her preference shock in the initial period has a large value.

The retirement probability that adjusts for the sample selection is the conditional probability

$$prob(retiring\ in\ period\ n|in\ initial\ sample) = \frac{prob(retiring\ in\ period\ n, in\ initial\ sample)}{prob(in\ initial\ sample)}. \quad (12)$$

The probability $prob(retiring\ in\ period\ n, in\ initial\ sample)$ is computed under the assumption that in the first year of eligibility, the preference error $\nu_{-J} \sim N(0, \frac{\sigma^2}{\sqrt{1-\rho^2}})$, instead of the assumption $\nu_0 \sim N(0, \frac{\sigma^2}{\sqrt{1-\rho^2}})$. Note J depends on the age and experience of the initial year.

Denote $\boldsymbol{\nu}_{-J,0} = (\nu_{-J}, ..., \nu_0)', \boldsymbol{\nu}_{-J,n-1} = (\nu_{-J}, ..., \nu_{n-1})', \mathbf{f}^+_{-J,0} = (f^+_{-J}, ..., f^+_0)', \mathbf{f}^+_{-J,n-1} = (f^+_{-J}, ..., f^+_{n-1})'$. The condition for the teacher in the initial sample is

$$\boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0}. \quad (13)$$

The precise formula for the statement in (12) is

$$\begin{aligned} G_{1,n} &= prob[((\boldsymbol{\nu}_{1,n-1} > -\mathbf{f}^+_{1,n-1}) \bigcap (f^+_n \leq -\nu_n)) | \boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0}] \\ &= \frac{prob[(\boldsymbol{\nu}_{-J,n-1} > -\mathbf{f}^+_{-J,n-1}) \bigcap (f^+_n \leq -\nu_n)]}{prob(\boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0})}. \end{aligned} \quad (14)$$

A procedure for computing the denominator in (14) is Algorithm1(t_1, t_2) that computes the probability of retirement in t_2 for a teacher with the first eligible year t_1. The algorithm is in Appendix 1.

The numerator in (14), $prob[(\boldsymbol{\nu}_{-J,n-1} > -\mathbf{f}^+_{-J,n-1}) \bigcap (f^+_n \leq -\nu_n)]$, can be computed using Algorithm 1$(-J, n)$.

Algorithm2(t_1, t_2) computes the probability of staying in t_2 for a teacher with the first eligible year t_1 (for a given pension rule and pension expectation, and (a, e, x) in the in initial year.) The algorithm is also given in Appendix 1. The probability $prob(\boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0})$ can be computed using Algorithm2$(-J, 0)$.

In Algorithm1 one may note that the left-side truncation of $\epsilon_{-J}, .., \epsilon_0$ shifts ν_0 to the right. Hence the initial preference error for a retirement eligible teacher in the initial year has a positive (instead of zero) mean. By estimating the probability that a retirement-eligible teacher is in the initial sample, we solve the initial condition problem by using the institutional knowledge in this nonlinear setting rather than classical methods.

The likelihood of a teacher in the initial sample retiring in period n can be computed as the ratio of (15) and (16). For teacher in each (a, e) cell in the initial period 1, the probability of retiring in period $1 \leq n < T$ can be computed via Algorithm1$(1, n)$; and the probability of staying teaching at the end of sample period T can be computed via Algorithm2$(1, T)$.

4 MLE Estimation and Diagnostics

4.1 MLE Estimation Results

The option value model described in the previous section is estimated on a cohort of 12,871 Missouri PSRS teachers aged 47-64 and with five or more years of experience in the 1993-1994 academic year. We track the cohort forward to the 2008 academic year. Table 2 reports descriptive statistics on this sample. In the base year 1994 about 74% of teachers in the sample are female, with average age 52.15 and an average of 21.48 years of teaching experience. Over the 14-year panel, roughly 94% of the teachers in the cohort retired. Based on the analysis in the previous section, we only need to track the retirement counts of (age, experience) cells, instead of the panel data of individual teachers.

(Insert Table 2 here.)

The retirement decisions depend on the expectations of future pension rules. Table 3 shows MLE estimates of structural parameters, β, κ, κ_1, γ, σ, ρ, p, in the retirement models for females and males separately under the following expectation assumptions: myopic (M), one step perfect foresight (P1), and adaptive expectation (A1). All cases are adjusted for the sample selection bias.

For all cases, the estimates of all parameters are economically plausible. The left half of Table 3 is the MLE estimates of structural parameters for females and the right half is for males. The estimates of β, the discount factor, which measures the time value, are very similar for all the cases (around 0.950-0.957, or 4.3%-5.0% discount rate.) The parameter

γ captures the risk preference or elasticity of intertemporal substitution for teachers. The estimates of γ are less than one, meaning the teachers are risk averse or prefer smoothing the income flow over time. The parameter σ measures the heterogeneity of unobserved preference errors for teachers. The preference errors intend to capture unmeasured factors that are relevant in deciding whether to retire: for example, the teacher's health, family-related factors, or peer effects. The persistence of the preference errors is captured by the parameter ρ. The estimates of ρ differ by expectations of pension rules, but are all positive. Recall the parameter k_s measures the disutility of working. The specification of the disutility of working can be a constant or a decreasing function of age. The estimates of κ implies that at age 60, one dollar of salary is equivalent to about seventy cents of pension benefit.

The rank order of likelihoods under different expectation models is the same for female and male teachers. The likelihood of adaptive expectation case (A1) is significantly larger than other cases, as we expect. The adaptive case (A1) is more flexible than the myopic expectation and perfect foresight.

(Insert Table 3 here.)

We also use the estimates in Table 3 to simulate the retirement probability under different models and compute aggregate statistics of interest to policy makers such as survival rate, age distribution and experience distribution. We compare the fit of these statistics of interest under each model of expectation, as a supplement to the overall fit measured by the likelihoods.

Computation Time

In this study, we track the counts of (age,experience) cells. The number of cells are fixed by the range of age and experience. For example, the teachers in the 1994 cohort who are age 47-64 with 5-46 years of experience are grouped into 756 (18×42) cells and tracked forward for 14 years. The computation time is independent of the number of teachers since the number of cells is fixed. If we use teacher-level data we need to simulate 12,871 teachers and track each teacher for 14 years. This would take about 17 times the number of computations compared to a cell-based approach. The computation time increases in the number of teachers and the length of the sample period. For a sample of teachers in a much larger state and/or a longer sample period, the computation time using teacher level data can be much higher.

4.2 In-Sample Goodness of Fit

For teachers in each age-experience cell in 1994, we simulate the probability of retiring in each year from 1995 to 2008, and not retiring in 2008. We aggregate the simulated probabilities of each cell and compare them to statistics based on observed retirement data.

The in-sample fit is good, particularly under the adaptive expectation assumption. Figure 3 plots the observed fraction of the 1994 cohort who remain teaching (the survival rates) each year from 1995 to 2008, and the simulated survival rates under different expectations. The simulated survival rates are very close to the observed.

(Insert Figure 3 here.)

In Figure 4, we plot the difference between observed and simulated survival rates. The survival rate simulated under the myopic case persistently under-predicts the observed survival rate. One explanation for this under-prediction of the survival rate (or over-prediction of retirement in the earlier years of the sample) under the myopic model is that after a series of pension enhancements in the early part of the sample period, teachers may anticipate additional enhancements, or be more attentive to discussions of future pension rule changes. Thus, the actual perceived benefit of delaying retirement is higher than that of the myopic (M) model.

(Insert Figure 4 here.)

In the first several years adaptive and perfect foresight models fit better than the myopic model. By 1997, the differences between simulated and observed survival rates are small for Models (P1). However, after 2000, Model (P1) persistently over-predicts the survival rate. Compared to these extreme cases, the in-sample prediction errors of Model (A1) is close to zero before 2000. After 2000, it also slightly over-predicts the survival rate. In addition, we also report the mean squared errors (MSE) of the prediction errors in survival rates in Table 4. The table shows that the adaptive model (A1) match the data better than other cases.

(Insert Tables 4-5 here.)

Figure 5 shows that the model predicts the joint age-experience distributions for retired teachers well. Table 5 also reports the MSE for the predicted age-experience distributions under different expectations for retired teachers. Based on the MSEs under different expectations, both model (M) and model (A1) perform well.

(Insert Figure 5 here.)

Generally speaking, the models fit well for most years during the sample period, with some mismatches in some years. To examine how well the estimated models fit data in more detail, we also compare the observed and simulated age distribution of retiring teachers at each year from 1995 to 2008. The plots (available upon request) show that age distributions simulated under different expectations of pension rules are very similar. In addition, the simulated distributions almost perfectly match the observed distribution for teachers who choose to stay each year.

4.3 Sample Selection Bias

In our data set, teachers are initially observed in the 1993-1994 school year. The baseline sample only includes teachers who are still teaching in the initial year and excludes those who already retired before the first year we observe. Without adjusting for sample selection bias, one would draw the initial value ν_1 from the unconditional stationary distribution and the probability of retiring in n_i is:

$$G_{1,n_i} = prob(\boldsymbol{\nu}_{1,(n_i-1)} > -\mathbf{f}^+_{1,(n_i-1)} \bigcap (f^+_{n_i} \leq -\nu_{n_i})).$$

This will result in over-prediction of retirement in the initial years for reasons explained in Section 3.4. Figure 6 reports the observed and predicted survival rates with and without adjusting for sample selection bias under (A1) expectation of pension rules. Overall, the survival rate simulated from the model adjusting for sample selection bias fits the data better than the model without adjustment. In particular, in the first three years the survival rate simulated from model without adjusting for sample selection bias is lower than the actual values. In other words, without adjusting for sample selection the model over-predicts retirement in the first three years.

(Insert Figure 6 here.)

4.4 Out-of-Sample Goodness of Fit

In this section, we examine the predictive performance of the estimated models. We simulate retirement decisions using parameters estimated based on 1994-2008 sample in Table 3 to predict the retirement behavior for a 2010-11 cohort with age 47-64 and with five or more years experience. We track this cohort forward to 2013-14. There were no pension rule enhancements during this period (or to the present time), and it is reasonable to assume that teachers did not expect pension rule changes during the period. Hence there is no difference in rules between myopic and perfect foresight expectations. The simulation results may differ because of the difference in parameters estimated based on different expectation models.

Table 6 compares the observed and simulated survival rates. Table 7 reports the difference between simulated and observed survival rates. The simulated survival rates under myopic expectations tend to under-predict the actual values. Overall, the models fit the data well. In particular, the simulated survival rate under adaptive expectation provides a very good fit to the actual values. One reason that the adaptive parameters fit the out-of-sample data better is that the parameter p in the (A1) model fits policies specific to the sample period, and affected the other parameters that are jointly estimated.

(Insert Tables 6-7 here.)

5 The Effect of Pension Enhancements

Pension enhancements change the gain in option value of working versus retiring. The question we seek to answer is how would the teachers alter their retirement in the absence of the enhancements observed during the sample period. The observed enhancements are generally of two types. One allows for earlier retirement, e.g., 25-and-out and the rule-of-80. This type of enhancement leads to earlier retirement. Another is an increase in the retirement benefit. Some new benefits seem designed to induce later retirement (for example, raising the replacement factor from 2.5% to 2.55% if the teacher retires with at least 31 years of experience. A more expensive cross-board increase in the replacement factor (from 2.3% to 2.5%) for all teachers taking regular retirement has a less obvious effect on retirement. The increase in replacement factor raises pension wealth for any given level of experience and should "pull" teachers towards retiring to the peak value year. Intuitively the "pull" effect should help to delay retirement. In the following we show that raising the replacement increases the "push" effect after passing the pension peak.

To analyze the "push" effect of an across-the-board benefit raise, we consider, for simplicity, the case of perfect foresight, with a constant contribution rate, and fixed pension rule R. Then the deterministic gain of staying until m over retiring in current period t is

$$\begin{aligned} g_t(R,m) = \quad & \sum_{s=t}^{m-1} \pi(s|t)\beta^{s-t}\mathbb{E}_t(k_s(1-c)Y_s)^\gamma + \sum_{s=m}^{T} \pi(s|t)\beta^{s-t}\mathbb{E}_t(B_s(R,m))^\gamma \\ & - \sum_{s=t}^{T} \pi(s|t)\beta^{s-t}\mathbb{E}_t(B_s(R,t))^\gamma . \end{aligned}$$

Assume the real benefit is roughly constant over time so $B_s(R,.)$ can be denoted as $B(R,.)$. For senior teachers who qualify for regular retirement and pass the peak of pension wealth, under each rule $B(R,m)/B(R,t)$ is roughly $(1+rf)^{m-t}$ (where rf is the replacement factor.) Suppose there are two pension benefit rules, R^l and R^h, with high and low benefits: $B(R^h,m) = (1+\tau)B(R^l,m)$ for any $m \geq t$. Then the difference between the net benefits of retiring in period m instead of in period t is

$$\begin{aligned} & g_t(R^h,m) - g_t(R^l,m) \\ & = \underbrace{(\sum_{s=m}^{T} \pi(s|t)\beta^{s-t}\mathbb{E}_t[(B(R^h,m))^\gamma - (B(R^l,m))^\gamma])}_{pension\ of\ h - pension\ of\ l\ if\ retire\ in\ m} - \underbrace{(\sum_{s=t}^{T} \pi(s|t)\beta^{s-t}([(B(R^h,t))^\gamma - (B(R^l,t))^\gamma])}_{pension\ of\ h - pension\ of\ l\ if\ retire\ in\ t} \\ & \approx (B(R^l,t))^\gamma[(1+\tau)^\gamma - 1]\{(1+rf)^{(m-t)\gamma}(\sum_{s=m}^{T} \pi(s|t)\beta^{s-t}) - (\sum_{s=t}^{T} \pi(s|t)\beta^{s-t})\} \\ & < (B(R^l,t))^\gamma[(1+\tau)^\gamma - 1][((1+rf)^\gamma\beta)^{(m-t)} - 1](\sum_{s=t}^{T} \pi(s|t)\beta^{s-t}). \end{aligned}$$

The last inequality follows from the fact that $\sum_{s=m}^{T} \pi(s|t)\beta^{s-t} < \beta^{m-t}(\sum_{s=t}^{T} \pi(s|t)\beta^{s-t})$. The inequality implies that $g_t(R^h,m) - g_t(R^l,m) < 0$ if $(1+rf)^\gamma\beta < 1$. From the Missouri pension rules and the parameter estimate reported in Table 3, $rf = 0.025, \gamma \approx 0.7, \beta \approx 0.96$, which implies $g_t(R^h, m) - g_t(R^l, m) < 0$. Dividing the difference in $g(.,.)$ by $K_t(m) = \sum_{s=t}^{m-1} \pi(s|t)(\beta\rho)^{s-t}$ does not change the sign. Hence raising the benefit rate reduces the welfare gain from staying and raises the retirement probability at t, through a stronger "push" effect. The numerical simulation shows that on net, the "push" effect dominates the "pull" effect for Missouri teachers, and the rise in the replacement factor leads to earlier retirement on average.

Since the calculation above compares fixed pension rules, it approximates the long-term effects of switching from R^l to R^h. The short-term effect may differ. Consider the case of an anticipated raise in benefit in the next several years, the gain from staying increases and current retirement should decrease.

Relaxing the eligibility requirement on age and/or experience leads to earlier retirement. But the short-term for a given cohort differs quantitatively from long-term effects for a steady state population of senior teachers. In the short-term, as the requirements were relaxed, a fraction of the cohort who would have retired using the new rules already passed the age/experience threshold stipulated by the new rules. This is particularly relevant for the rules introduced later in the sample period (such as the rule-of-80.) But in the long-term all senior teachers have an opportunity to retire under the new rules. Hence, in the long term the average retirement age should be reduced more than in the short term. In the following section we analyze the short- and long-term effects of the pension enhancements.

5.1 The Effects of 25-and-Out and Rule-of-80 on the 1994 Cohort

We now quantify the effects of specific pension enhancements in PSRS during 1994-2002, including 25-and-out and rule-of-80. In this section, we conduct counterfactual analysis how teacher retirement would differ from the historical data if no 25-and-out, rule-of-80, or any enhancements ever took place.

To see the effects of pension enhancements on retirement of teachers in 1994 cohort, we simulate the retirement probability for the next 30 years. Table 8 reports the observed and simulated average retirement age and experience in different counterfactual cases under different expectation assumptions. In contrast, the average retirement age and experience in the case with the observed enhancements is smaller than that in counterfactual cases. In other words, pension enhancements prompt teachers to retire earlier. Under the more generous pension system, teachers could receive more retirement benefit earlier so that teachers may choose to retire earlier. Among these different counterfactual cases, the scenario of no enhancements has the most significant impact on average retirement age and experience since it removed two rules, 25-and-out and rule-of-80, and maintained a lower replacement factor 2.3%. The effect of rule-of-80 is smaller compared to other cases because the rule was introduced in 2000. At that time, many teachers have retired and a sizable proportion of teachers could qualify for the regular retirement in the absence of the rule. Hence only a small proportion of teachers would be affected by this rule. However, 25-and-out was introduced in 1997, much earlier than rule-of-80. So it could affect more teachers and the effect should be larger.

5.2 Steady State Estimates of Pension Enhancements

The above empirical analysis is conducted by tracking the 1994 cohort for 30 years. During the sample period, the pension rule changes were introduced for a cohort many of whom had already passed the age/experience thresholds that are affected by the rule changes. Because the rule changes only affected part of the sample, the simulated short-term effects understates the full effect of the rule changes.

Appendix 2 shows that as the retired teachers are replaced by new entrants (those with age=47 or experience=5), under each pension rule in the long run the distribution of senior teachers converges to a stationary distribution in age and experience. By comparing the stationary distributions under different pension rules we obtain the long-term effect of a change of pension rules. Table 9 reports the average retirement age and experience of senior teachers in a steady state. We use different estimates of expectation assumptions to simulate the retirement for these senior teachers. As expected, compared to Table 8, the averages over teachers in the steady state is smaller than for the 1994 cohort. And the differences between the counterfactual scenarios and the case with all enhancements for teachers in the steady state are bigger than the ones for 1994 cohort. The enhancements cause teachers to retire 1.13-1.22 years earlier, while the effect of enhancements is only 0.35-0.49 years for the 1994 cohort. The bigger effects are due to the simulated long-term effects when all new entrants are subject to the pension enhancements.

(Insert Tables 8-9 here.)

5.2.1 Estimated fiscal effects of pension enhancements

The pension enhancements have short-run and long-run effects on teacher retirement and fiscal costs. The fiscal effects are evaluated in the following framework. Let the total number of teachers be fixed and let $N(a,e,t)$ be the fraction of teachers with (age, experience) (a,e) in period t, and with a given attrition rate $r(a,e,t)$. By definition $\sum_{a,e} N(a,e,t) = 1$. Let the present value of pension wealth for attrition occurred at each (a,e,t) be $P(a,e,t)$. Then overall pension cost of the teachers is $\bar{P}_t = \sum_{a,e} N(a,e,t)r(a,e,t)P(a,e,t)$.

The simulations of a sweeping change in policy rules give rise to a change in attrition rates $r(a,e,t)$. The model in Appendix 2 quantifies the dynamics of the distributions $N(a,e,t)$, given the $r(a,e,t)$. Structural models such as the Stock-Wise model of career choice dictates how $r(a,e,t)$ changes with a change in rule R. Suppose under the new policy $*$ the attrition of cell (a,e,t) changes from the current $r(a,e,t)$ to $r^*(a,e,t)$, and label average quantities corresponding to policy $*$ by a subscript $*$. Then the fiscal cost in the short run is $\bar{P}_t^* - \bar{P}_t$; and that in the long run should approach a function of t (either positive or negative).

Table 10 reports the pension wealth and retirement rate for senior teachers in a steady state, under different counterfactual scenarios. The total number of teachers is normalized to be one. The stationary distribution is the snapshot of the senior teachers aged 47-80 with 5-45 years of experience at the steady state. The pension cost changes are measured by the percentage of pension wealth under different counterfactual cases compared to the case with all of the enhancements.

(Insert Table 10 here.)

Pension enhancements, especially two policy changes, 25-and-out and rule-of-80, have different effects on the amount of expected pension benefit for senior teachers. Figure 7 compares the average pension wealth with no enhancements with that under four scenarios: (i) with 25-and-out (but without other enhancements); (ii) with rule-of-80 (but without other enhancements); (iii) the increase in the replacement rate (without other enhancements); and (iv) with all of the enhancements. Figure 8 reports the distribution of percentage change of expected pension wealth under the four scenarios comparing to the case of no enhancement. The percentage change of expected pension wealth is the ratio of pension wealth gains under each scenario of pension enhancements over the pension wealth with no enhancements.

The rule 25-and-out affects teachers who are young with 21-24 experience. Whether to take advantage of the rule depends on the balance of the following tradeoff. By retiring earlier a teacher can collect pension benefits for more years; however the amount of annual benefit is lower than regular retirement because of the early retirement penalty. The overall effect depends on which side dominates. Teachers in the cells with positive pension wealth gains receive more expected pension benefits. For them, the effect of the additional collection years dominates the effect of early retirement penalty. While teachers in the cell (47,24) receive slightly lower expected pension benefits since for them the effect of early retirement penalty dominates.

The introduction of the rule-of-80 broadens eligibility of retirement without penalty. The additional flexibility affects teachers' expected pension wealth in more than one way. Some teachers receive more expected pension wealth with the rule-of-80 because under the rule allow teachers retire earlier and receive full pension benefits for more years. For example, teachers in the cell (56,22) need to continue to work for three more years to receive full benefits without enhancement. However, they only need to continue to work for one more year with the rule-of-80.

Some teachers may receive slightly less expected pension wealth by choosing to retire under the "rule-of-80." The additional flexibility in retirement may induce earlier retirement, causing a slightly lower expected pension wealth for some cells. It is worth noting that while the rule-of-80 may slightly lower the expected pension wealth it still raises the expected utility. It is also worth noting that if the rule induces earlier retirement for teachers in certain cells it raises the cost of replacement. So the overall pension cost may still increase, in addition to the increase in the cost of salary.

Another important pension enhancement is the increase of replacement factor from 2.3% to 2.5%. By increasing the replacement factor, pension wealth increases for all teachers.

Table 10 shows that for a senior teacher population of unity the enhancements in the 1990s raised the pension cost by about $\$7,000$ per year at the steady state. For the whole population of senior teachers (roughly $13,000$) the total increase in pension cost is $91 million per annum (in 1995 dollars).

(Insert Figures 7-8 here.)

6 Conclusion

In this paper we examined the effect of pension rule enhancements during the 1990s on retirement behavior of Missouri public school teachers. We estimated an option-value retirement model after overcoming several challenges. The resulting option value models provided very good in-and out-of sample fit. We then used the estimated model to evaluate the effects the pension enhancements on retirement behavior. Our simulations show that the enhancements lead to earlier retirement of senior teachers, by about 0.3 years for the 1994 cohort and by more than one year in the steady state. Since teachers already retire at ages considerably younger than comparable professionals, many retirement plans are significantly underfunded, and complaints of teacher shortages have become commonplace, reversing some or all of these enhancements, or using other plan incentives to encourage longer teaching careers may be appropriate. Structural methods such as those developed in this paper would prove useful in estimating the costs and benefits of such changes.

The methodology developed here can be extended to analysis of pensions in other states. Given the diversity of teacher pension rules (e.g., some of the states are in Social Security, some have hybrid DB/DC plans) it is of interest to examine whether the optional value model accurately captures retirement behavior in these systems as well. Many states have now recognized the need to retain experienced teachers and fiscal pressures have forced some states to implement less generous plans for new teachers. The methods developed in this paper can be used to analyze the workforce and fiscal effects of such changes.

Table 1: PSRS Pension Rule Changes

Academic Year	Replacement Factor	COLA	Retirement Age and Experience
1994	0.023	0.56	Age $\geq$ 55 and Exp$\geq$25, or Age$\geq$60 and Exp$\geq$5, or Exp$\geq$30
1995	0.023	0.65	same
1997	0.023	0.75	Add 25-and-out early retirement (with Exp $\geq$25)
1999	0.025	0.75	25-and-out formula factors increased
2000	0.025	0.75	Add rule-of-80, Age+Exp$\geq$80
2001	0.025	0.8	same
2002	0.0255 if Exp$\geq$31	0.8	same

Note: There are three main rules for retirement: regular retirement, 25-and-out, and age-reduced. At the time of retirement, if teachers satisfy one of conditions for regular retirement rules, then they follow regular retirement. Before the 1999-2000 school year, regular retirement rules include Age $\geq$ 55 and Exp$\geq$25, or Age$\geq$60 and Exp$\geq$5, or Exp$\geq$30. After 2000, rule-of-80 was added to regular retirement. If teachers did not satisfy regular retirement, the rule of 25-and-out with a reduced annuity can be applied since 1997 academic year. Before 1995, the teacher contribution rate was 10%. During 1996-2004, contribution was 10.5%. It increased 0.5% each year from 2005-2012.

Table 2: Sample Summary Statistics

1994 cohort	Number of Teachers	Age	Experience	Male
All 1994	12871	52.15	21.48	0.26
Retirement Year				
1995	674	57.04	27.15	0.34
1996	864	57.50	27.30	0.32
1997	1176	55.98	27.74	0.33
1998	1095	56.71	27.73	0.32
1999	1193	56.85	27.64	0.29
2000	1212	56.99	27.66	0.29
2001	1373	57.49	27.68	0.26
2002	1001	58.11	27.74	0.23
2003	825	58.72	27.24	0.22
2004	742	59.34	27.86	0.20
2005	687	60.10	27.17	0.19
2006	541	60.67	27.12	0.20
2007	412	61.23	27.60	0.18
2008	302	62.17	27.36	0.14
Not Retired by 2008	774	62.28	26.88	0.17

Note: The sample is Missouri PSRS public school teachers aged 47-64 with at least 5 years of experience in the 1993-1994 school year. "All 1994" denotes the cohort of 12871 teachers in the base year. The rows with specific retirement year present the average age and average experience of teachers who retired in that year. The row with "Not retired by 2008" are teachers who remained employed at the end of the sample period. Male=1 for male teachers.

Table 3: MLE Estimates of Structural Parameters

Parameters	Female			Male		
Expectations	(M)	(A1)	(P1)	(M)	(A1)	(P1)
β	0.954	0.957	0.955	0.950	0.955	0.956
	(0.004)	(0.005)	(0.005)	(0.008)	(0.009)	(0.008)
γ	0.653	0.660	0.643	0.600	0.673	0.656
	(0.015)	(0.017)	(0.017)	(0.028)	(0.030)	(0.032)
σ	2792.201	3939.323	2811.374	3176.834	4525.896	2974.686
	(490.033)	(702.488)	(517.593)	(1038.241)	(1518.990)	(1037.865)
ρ	0.546	0.476	0.503	0.391	0.471	0.532
	(0.006)	(0.008)	(0.007)	(0.017)	(0.012)	(0.011)
κ	0.660	0.674	0.681	0.777	0.709	0.690
	(0.017)	(0.021)	(0.019)	(0.034)	(0.032)	(0.029)
$\kappa 1$	0.572	0.998	0.478	0.729	0.984	1.012
	(0.136)	(0.141)	(0.136)	(0.210)	(0.231)	(0.251)
p		0.575			0.581	
		(0.037)			(0.049)	
log-likelihood	-22201.81	-21992.97	-22186.69	-7604.50	-7560.12	-7598.75

Note: The standard errors are in parentheses. Missouri PSRS teachers are age 47-64 with at least 5 years of experience in 1994. The sample period is 1994-2008. The likelihood is evaluated using the "GHK" algorithm described in the appendix. Expectation assumptions are: myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1).

Table 4: MSE for Different Predicted Survival Rates

Expectation	(M)	(A1)	(P1)
MSE	2.4031	0.8972	2.7663

Note: The unit of numbers is 10^{-4}.
We use administrative data during 1994-2008 for Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 1994. MSE (Mean squared errors) are for different predicted survival rates under myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1).

Table 5: MSE for Different Predicted Age-Experience Distributions

Expectation	(M)	(A1)	(P1)
Retired	2.3050	2.5930	3.4530
Non-retired	11.4000	9.9050	11.5600

Note: The unit of numbers is 10^{-6}.
We use administrative data during 1994-2008 for Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 1994. MSE (Mean squared errors) are for different predicted age-experience distributions under myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1).

Table 6: Observed and Simulated Survival Rates

Year	Data	(M)	(A1)	(P1)
2012	0.8939	0.8778	0.8863	0.8885
2013	0.7970	0.7847	0.7950	0.7991
2014	0.6991	0.6976	0.7090	0.7155

Note: We use administrative data during 2011-2014 for Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 2011. Simulated survival rates are based on the different estimates of expectation assumptions: myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1).

Table 7: Difference between Observed and Simulated Survival Rate

Year	(M)	(A1)	(P1)
2012	-0.0161	-0.0076	-0.0053
2013	-0.0123	-0.0019	0.0022
2014	-0.0015	0.0099	0.0164

Note: We use administrative data during 2011-2014 for Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 2011. Simulated survival rates are based on the different estimates of expectation assumptions: myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1).

Table 8: Simulated Average Retirement Age and Experience: 1994 Cohort

		Average Retirement Age	Average Retirement Experience
M	Observed rules for the 1994 cohort	58.34	27.67
	No 25 and out	58.50 (0.16)	27.84 (0.16)
	No rule of 80	58.41 (0.06)	27.74 (0.06)
	No Enhancements	58.69 (0.35)	28.02 (0.35)
A1	Observed rules for the 1994 cohort	58.40	27.74
	No 25 and out	58.56 (0.16)	27.89 (0.16)
	No rule of 80	58.48 (0.08)	27.81 (0.08)
	No Enhancements	58.78 (0.37)	28.11 (0.37)
P1	Observed rules for the 1994 cohort	58.45	27.78
	No 25 and out	58.67 (0.22)	28.00 (0.22)
	No rule of 80	58.54 (0.09)	27.87 (0.09)
	No Enhancements	58.94 (0.49)	28.27 (0.49)

Note: The simulated average retirement age and experience in different counterfactual cases are based on the different estimates of expectation assumptions: myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1). Teachers in 1994 cohort are Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 1994. The average simulated age and experience under counter-factual scenarios subtracting that under the enhancements experienced by the 1994 cohort are reported in the parenthesis.

Table 9: Simulated Average Retirement Age and Experience: Steady State

		Avg Retirement Age	Avg Retirement Exp
M	With All Enhancements	56.89	24.93
	No 25 and out	57.48 (0.59)	25.52 (0.59)
	No rule of 80	57.03 (0.14)	25.07 (0.14)
	No Enhancements	58.11 (1.22)	26.15 (1.22)
A1	With All Enhancements	57.15	25.19
	No 25 and out	57.63 (0.47)	25.67 (0.47)
	No rule of 80	57.31 (0.16)	25.35 (0.16)
	No Enhancements	58.24 (1.13)	26.28 (1.13)
P1	With All Enhancements	57.25	25.29
	No 25 and out	57.80 (0.55)	25.84 (0.55)
	No rule of 80	57.40 (0.15)	25.45 (0.15)
	No Enhancements	58.45 (1.20)	26.50 (1.20)

Note: The simulated average retirement age and experience in different counterfactual cases are based on the different estimates of expectation assumptions, including myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1). The average simulated age and experience under counter-factual scenarios subtracting that under the case of all enhancements are reported in the parenthesis.

Table 10: Pension Wealth and Retirement Rate Under Counterfactual Scenarios: Steady State

		Pension Cost	Minus the Cost of Enhancements	Retirement rate
M	With All Enhancements	52628.35		0.0903
	No 25 and out	51365.86	-0.0240	0.0858
	No rule of 80	51691.87	-0.0178	0.0892
	No Enhancements	45030.71	-0.1444	0.0814
A1	With All Enhancements	52382.46		0.0882
	No 25 and out	51379.44	-0.0191	0.0847
	No rule of 80	51436.66	-0.0181	0.0870
	No Enhancements	45081.98	-0.1394	0.0803
P1	With All Enhancements	51911.38		0.0875
	No 25 and out	50850.37	-0.0204	0.0835
	No rule of 80	51013.62	-0.0173	0.0863
	No Enhancements	44614.05	-0.1406	0.0791

Note: The total number of teachers is normalized to one. The stationary distribution is the distribution of teachers over the cells of age 47-80 and experience 5-45 years at the steady state. Minus the Cost of Enhancements is the percentage of pension cost under counterfactual cases minus that of the pension cost with all enhancements. The calculation in different counterfactual cases are based on the different estimates of expectation assumptions, including myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1).

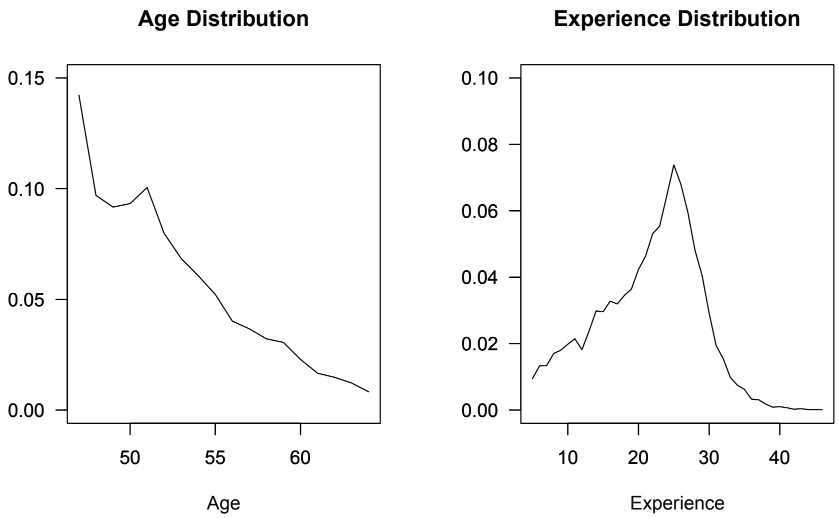

Figure 1: Age/Experience Distribution of Initial Sample

Note: The initial sample of the 1994 cohort are 12871 Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 1994.

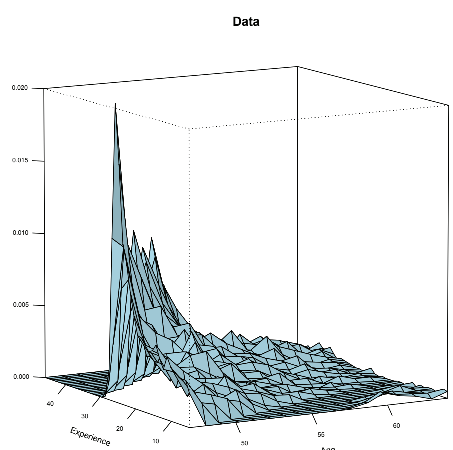

Figure 2: Age and Experience Distribution of Initial Sample

Note: The initial sample of the 1994 cohort are 12871 Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 1994.

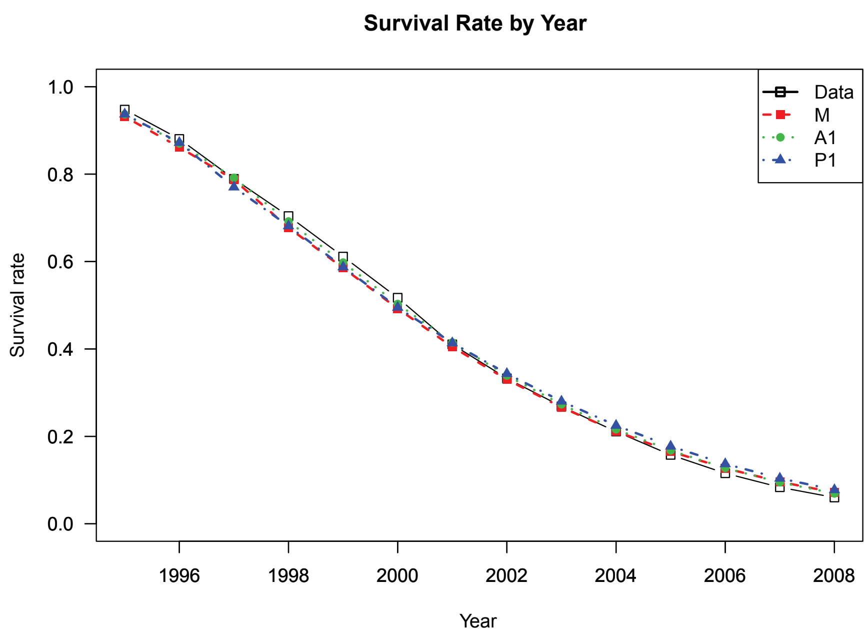

Figure 3: Observed and Simulated Survival Rates

Note: Observed survival rate is based on the administrative data during 1994-2008 for Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 1994. Simulated survival rates are based on based on the different estimates of expectation assumptions, including myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1).

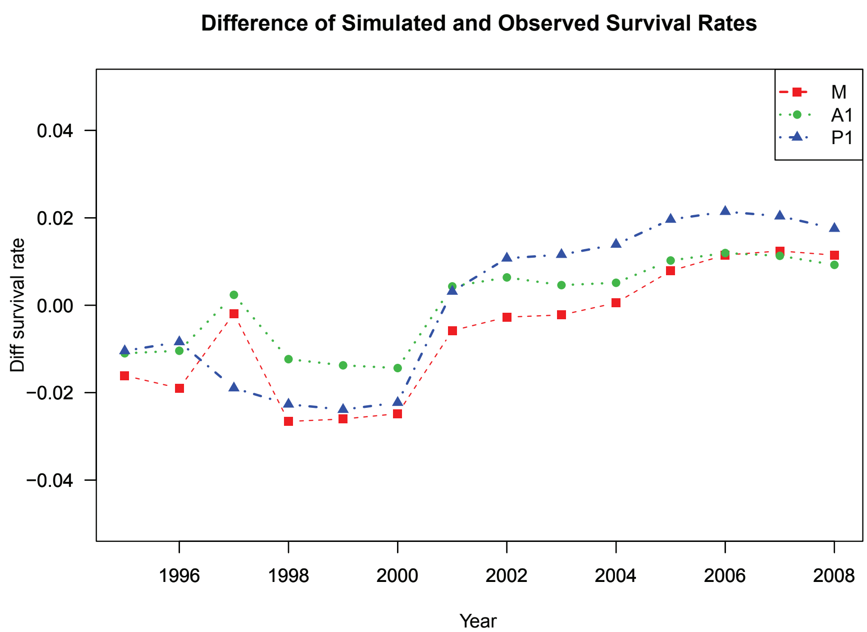

Figure 4: Difference between Simulated and Observed Survival Rates

Note: Observed survival rate is based on the administrative data during 1994-2008 for Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 1994. Simulated survival rates are based on the different estimates of expectation assumptions, including myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1).

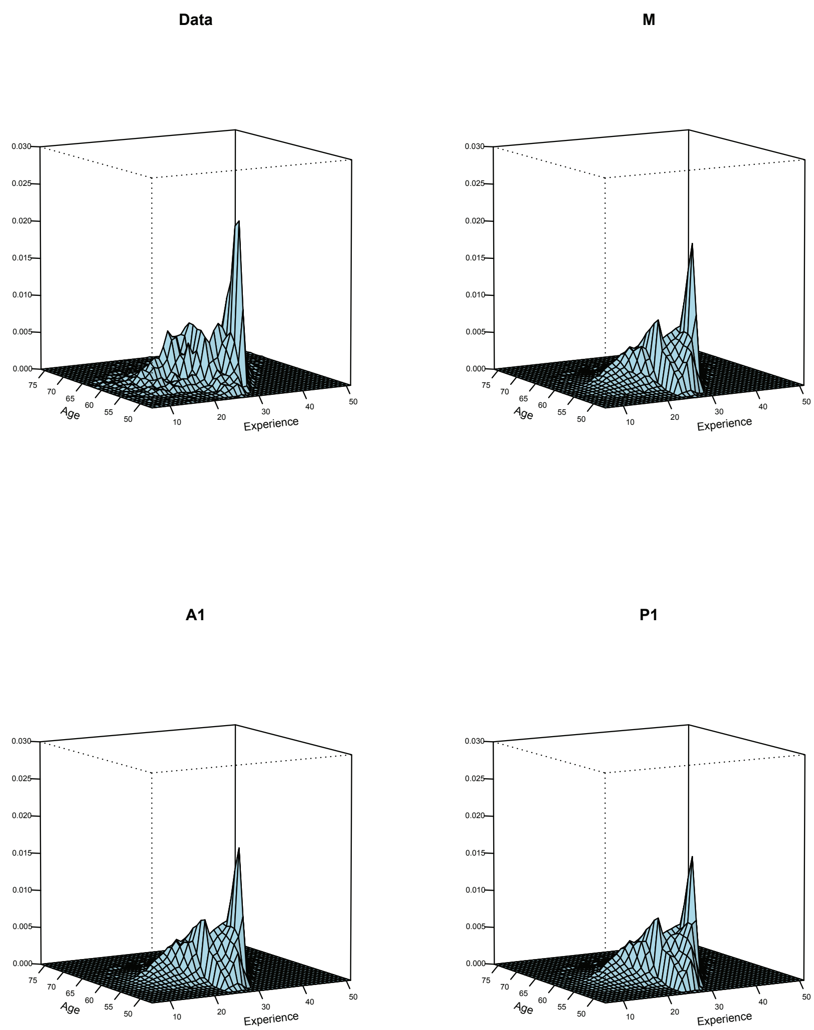

Figure 5: Observed and Simulated Joint Retirement Age-Experience Distributions for Retired Teachers

Note: Observed age-experience distribution is based on all teachers of the 1994 cohort at the time of retirement (for the left panel) or the non-retired teachers at the end of the sample period (for the right panel). Simulated age-experience distributions are based on the different estimates of expectation assumptions, including myopic(M), one step perfect foresight(P1), adaptive expectation 1(A1).

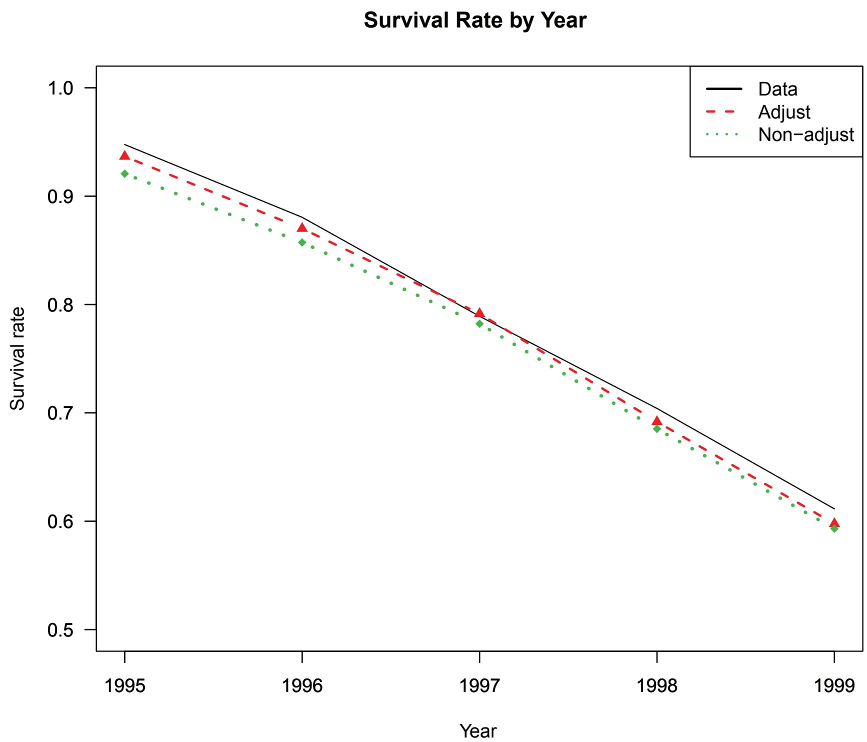

Figure 6: Observed and Predicted Survival Rates With and Without Selection Bias Correction

Note: Observed survival rate is based on the administrative data during 1994-2008 for Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 1994. Simulated survival rate with adjusting is based on the estimates of adaptive expectation 1(A1) and adjusted for the sample selection bias. Simulated survival rate without adjusting is based on the estimates of adaptive expectation 1(A1) without correction.

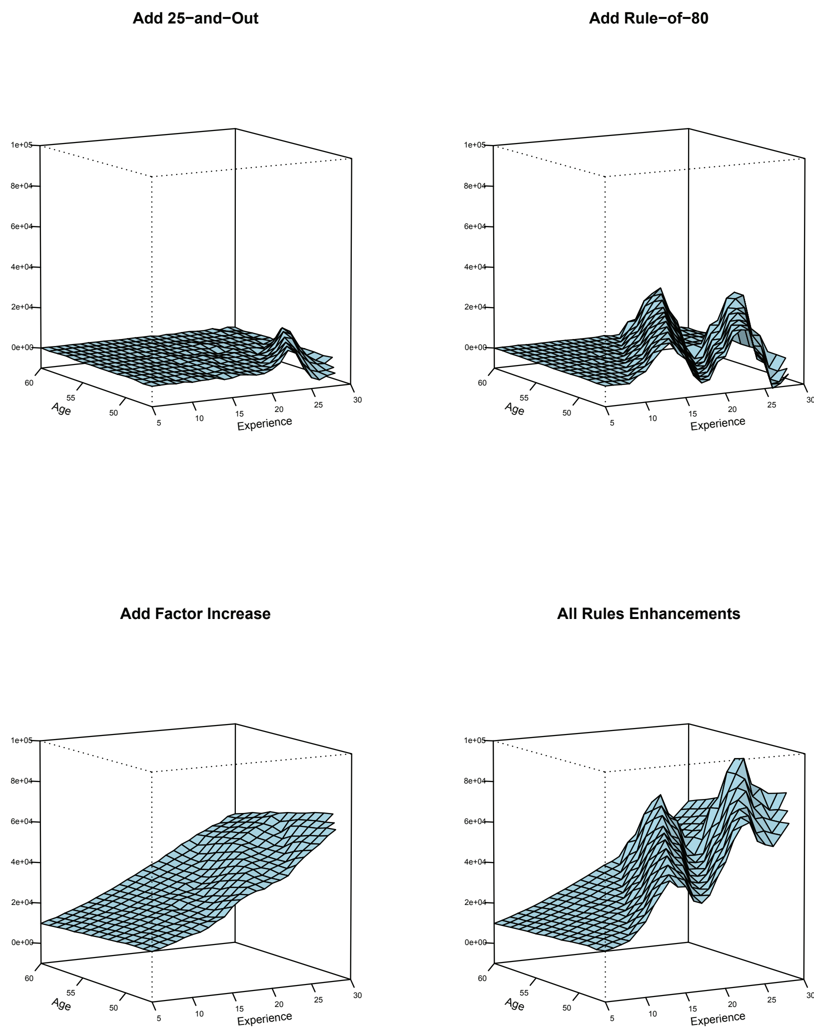

Figure 7: Distribution of Pension Wealth Gains

Note: The pension wealth gains are increases in pension wealth by different enhancements over that with no enhancement. Case 1: adding the rule 25-and-out; Case 2: adding the Rule-of-80; Case 3: adding the increase of replacement factor; Case 4: adding all pension enhancements.

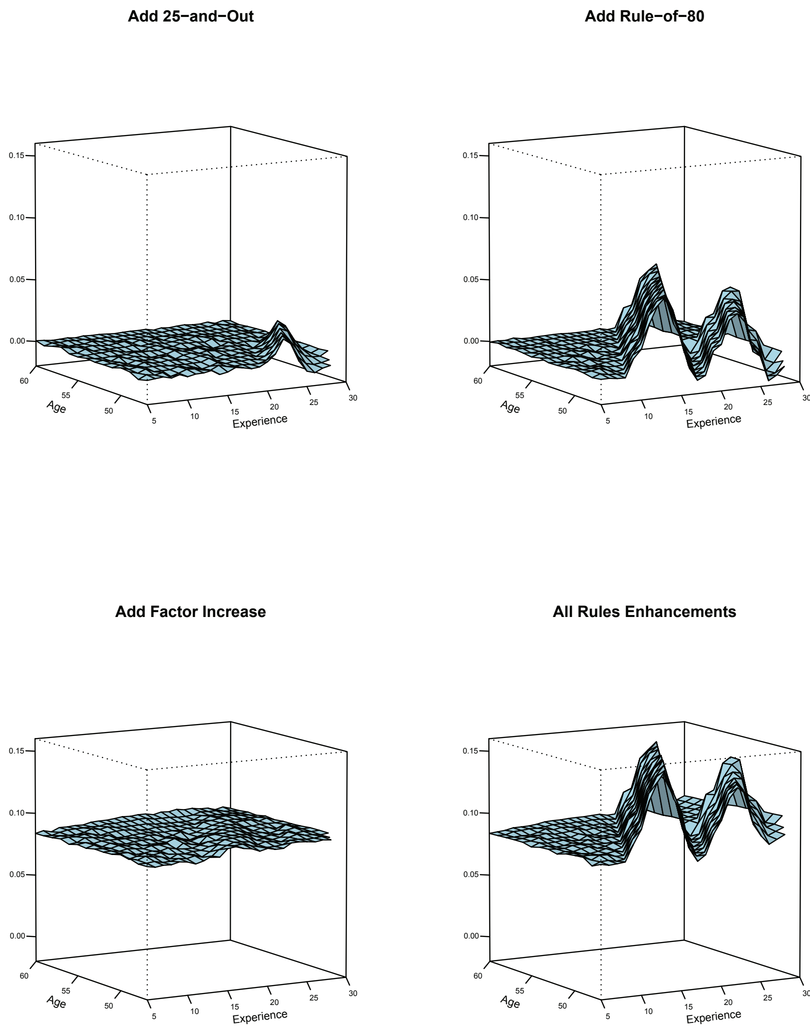

Figure 8: Distribution of Percentage Change of Pension Wealth

Note: The percentage change of expected pension wealth is the ratio of pension wealth gains under the case of pension enhancements and the pension wealth with no enhancement.The pension wealth gains are increases in pension wealth by different enhancements over that with no enhancement. Case 1: adding the rule 25-and-out; Case 2: adding the Rule-of-80; Case 3: adding the increase of replacement factor; Case 4: adding all pension enhancements.

7 Appendix

Appendix

Appendix 1. Algorithms for Computing Likelihood

Denote the normal $N(m, \sigma^2)$ truncated at (a, b) as $TN_{(a,b)}(m, \sigma^2)$.

Algorithm1(t_1, t_2) computes the probability of retirement in t_2 for a teacher with the first eligible year t_1:

1. Starting in period t_1, obtain K ν_{t_1}s that satisfy $\nu_{t_1} > -f^+_{t_1}$, by drawing from the left truncated $TN_{(-f^+_{t_1},\infty)}(0, \frac{\sigma^2}{\sqrt{1-\rho^2}})$. This is equivalent to drawing $\epsilon^{(k)}_{t_1}$ from $TN_{(-f^+_{t_1}\sqrt{1-\rho^2},\infty)}(0, \sigma^2)$, $k = 1, .., K$.

2. For $t_1 < t < t_2$, conditional on $\nu^{(k)}_{t-1}$ (or $\epsilon^{(k)}_{t_1}, .., \epsilon^{(k)}_{t-1}$) draw $\epsilon^{(k)}_t$ from $TN_{(-f^+_t-\rho\nu^{(k)}_{t-1},\infty)}(0, \sigma^2)$, which implies $\nu^{(k)}_t = \rho\nu^{(k)}_{t-1} + \epsilon^{(k)}_t > -f^+_t$,

3. In period t_2 conditional on $\nu^{(k)}_{t_2-1}$ draw $\epsilon^{(k)}_{t_2}$ from $TN_{(-\infty,-f^+_{t_2}-\rho\nu^{(k)}_{t_2-1})}(0, \sigma^2)$, which implies $f^+_{t_2} < -\nu^{(k)}_{t_2} = -\rho\nu^{(k)}_{t_2-1} - \epsilon^{(k)}_{t_2}$.

The preference error in period t

$$\nu_t = \rho^t\nu_0 + \rho^{t-1}\epsilon_1 + ... + \rho\epsilon_{t-1} + \epsilon_t$$

is conditional on the initial error ν_{-J} and the shocks ϵ_j $(1 \le j \le t)$ drawn from the truncated conditional distributions.

$$\begin{aligned} prob(G_{-J,n}) &= \prod_{t=-J+1}^{n-1} prob((-\rho\nu_{t-1} + \epsilon_t > -f^+_t)|\nu_{t-1}) \times prob((\rho\nu_{n-1} + \epsilon_n \le -f^+_n)|\nu_{n-1}) \\ &= \prod_{t=-J+1}^{n-1} \Phi(\frac{f^+_t + \rho\nu_{t-1}}{\sigma})\Phi(-\frac{f^+_n + \rho\nu_{n-1}}{\sigma}). \end{aligned}$$

The probability can be computed using the K simulated sequences of ν's as

$$G_{-J,n} \approx \frac{1}{K}\sum_{k=1}^{K}[\prod_{t=-J+1}^{n-1} \Phi(\frac{f^+_t + \rho\nu^{(k)}_{t-1}}{\sigma})]\Phi(-\frac{f^+_n + \rho\nu^{(k)}_{n-1}}{\sigma}). \tag{15}$$

The algorithm is based on the same idea of the Geweke–Hajivassiliou–Keane (GHK) simulator (see e.g., Börsch-Supan and Hajivassiliou 1993). The new twist is the initial condition is adjusted based on institutional knowledge of the pension rules. With time varying rules, $f(t)^+$ depends on the expectation of pension rules. Note that Algorithm1 does not produce unbiased draws of ν_0.

Algorithm2(t_1, t_2)

Take the first 2 steps of Algorithm1(t_1, t_2). Step 1 is the same. In Step 2 replace $t_1 < t < t_2$ by $t_1 < t \leq t_2$, and

Step 2'. For $t_1 < t \leq t_2$, conditional on $\nu_{t-1}^{(k)}$ (or $\epsilon_{t_1}^{(k)}, .., \epsilon_{t-1}^{(k)}$) draw $\epsilon_t^{(k)}$ from $TN_{(-f_t^+ - \rho\nu_{t-1}^{(k)}, \infty)}(0, \sigma^2)$.

The probability of staying teaching in period t is $prob((\rho\nu_{t-1} + \epsilon_t > -f_t^+)|\nu_{t-1}) = 1 - \Phi(\frac{-f_t^+ - \rho\nu_{t-1}}{\sigma}) = \Phi(\frac{f_t^+ + \rho\nu_{t-1}}{\sigma})$.

$$\begin{aligned} prob(\boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0}) &= \prod_{t=-J+1}^{0} prob((-\rho\nu_{t-1} + \epsilon_t > -f_t^+)|\nu_{t-1}) \\ &= [\prod_{t=-J}^{0} \Phi(\frac{f_t^+ + \rho\nu_{t-1}}{\sigma})]. \end{aligned}$$

The probability can be computed using the K simulated sequences of ν's as

$$prob(\boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0}) \approx \frac{1}{K}\sum_{k=1}^{K}[\prod_{t=-J+1}^{0} \Phi(\frac{f_t^+ + \rho\nu_{t-1}^{(k)}}{\sigma})]. \tag{16}$$

Appendix 2. Pension Rules and the Dynamics of the Distributions of Senior Teachers

The appendix explores the dynamic effects on the distribution of senior teachers by pension rule changes that lead to changes in attrition rates for certain age-experience cells.

The assumptions are:

1. The total number of teachers is fixed. Without losing generality, we let the fixed number be 1.

2. All teachers teach without interruption prior to leaving.

3. Let $N(a,e,t)$ be the fraction (or the number) of teachers with (age, experience) (a,e) in period t, and with a given attrition rate $r(a,e,t)$. We assume $a \geq a^l$ (a^l is a minimum age, 47 for the sample in the study). Let a_0 be the youngest age to be a novice teacher (say 22).

4. The attrition is 1 if age hits an upper limit $a = a^h$ (a^h may be 75.)

5. The minimum experience in the sample is e^l ($e^l = 5$ for the sample used in the study), $e^l \leq e \leq a - a_0$.

6. When attrition occurs in period t, they are replaced by either teachers with of age a minimum experience e^l, of size $N(a, e^l, t+1)$, or teachers with experience e and minimum age a^l in the sample, of size $N(a^l, e, t+1)$. The new entrants work for at least one year. The age distribution of the minimum experience or minimum age teachers is given by $\frac{N(a,e^l,t)}{N(a^l,e^l,t)} = f(a)$, for $a = a^l, .., a^h - 1$; and $\frac{N(a^l,e,t)}{N(a^l,e^l,t)} = h(e)$ for $e^l + 1 \leq e \leq a - a_0 - 1$. Denote $\sum_{a=a^l}^{a^h-1} f(a) = a^*, \sum_{e=e^l+1}^{a^l-a_0} h(e) = e^*$. The (a^l, e^l) cell is counted in $f(a)$ but not in $h(e)$ without losing generality.

With this notion, the attrition of teachers of age $a = a^h$ and experience e in period t is $\sum_{e=e^l}^{a^h-a_0} N(a^h, e, t)$. The attrition of the new entrant teachers is $\sum_{a=a^l}^{a^h-1} N(a, e^l, t) r(a, e^l, t) + \sum_{e=e^l+1}^{a^l-a_0} N(a^l, e, t) r(a^l, e, t)$. The total attrition in period t is given by

$$
\begin{aligned}
s(t) = \quad & \sum_{a=a^l+1}^{a^h} \sum_{e=e^l+1}^{a-a_0} N(a-1, e-1, t) r(a-1, e-1, t) + \sum_{e=e^l}^{a^h-a_0} N(a^h, e, t) \\
& + \sum_{a=a^l}^{a^h-1} N(a, e^l, t) r(a, e^l, t) + \sum_{e=e^l+1}^{a^l-a_0} N(a^l, e, t) r(a^l, e, t).
\end{aligned}
$$

By assumption 1, the vacancies due to retirements are filled in the next period by new entrant teachers of size $s(t)$. The size of replacement teachers in period $t+1$ with minimum experience e^l is $\sum_{a=a^l}^{a^h-1} N(a, e^l, t+1) = \sum_{a=a^l}^{a^h-1} f(a) N(a^l, e^l, t+1)$. The replacement teachers in period $t+1$ with minimum age a^l is $\sum_{e=e^l+1}^{a^h-a_0-1} N(a, e^l, t+1) = \sum_{e=e^l+1}^{a^l-a_0} h(e) N(a^l, e^l, t+1)$. Hence $N(a^l, e^l, t+1) = \frac{s(t)}{a^*+e^*}$. The age-specific size of minimum-experience new entrant teachers of the next period is $N(a, e^l, t+1) = f(a) N(a^l, e^l, t+1) = \frac{s(t)f(a)}{a^*+e^*}$, and experience-specific size of minimum-experience new entrant teachers of the next period is $N(a^l, e, t+1) = h(e) N(a^l, e^l, t+1) = \frac{s(t)h(e)}{a^*+e^*}$ (for $e^l + 1 \leq e \leq a^l - a_0$.)

Pension rules affect attrition rates $r(a,e,t)$. Structural models such Stock-Wise or a dynamic programming model of career choice dictates how $r(a,e,t)$ changes with change in rules. The short-run effect of a pension rule change on the distribution of the teaching force workforce can be computed using the formula above. In a steady state all functions N, r are time-independent.The long-run effect concerns the stationary distribution of teachers (if it exists).

Let the attrition rate by time-independent: $r(a,e,t) = r(a,e)$, $(a = a^l, .., a^h, e^l \leq e \leq a - a_0)$. Then starting from an initial distribution (a,e) in period 0, then

$$\begin{aligned} N(a,e,t) &= N(a-1,e-1,t-1)[1-r(a-1,e-1)] \\ &= N(a-2,e-2,t-2)[1-r(a-1,e-1)][1-r(a-2,e-2)] \\ &= N(a-e^l,e^l,0)[1-r(a-1,e-1)]..[1-r(a-e^l,e^l)]. \end{aligned} \tag{17}$$

Here $N(a-e^l,e^l,t)$ is the fraction of new entrant teachers with age $a-e^l$ in period 0.

Denote $\mathbf{w}_0(a^l,t) = \begin{pmatrix} N(a^l,e^l+1,t) \\ \vdots \\ N(a^l,a^l-a_0,t) \end{pmatrix}$.

Then denote $\mathbf{w}(e,t)$ as the vector teacher shares with experience e of stacked up by age. The dimension varies by the level of experience.

$\mathbf{w}(e^l,t) = \begin{pmatrix} N(a^l,e^l,t) \\ N(a^l+1,e^l,t) \\ \vdots \\ N(a^h-1,e^l,t) \end{pmatrix}$, $\mathbf{w}(e^l+1,t) = \begin{pmatrix} N(a^l,e^l+1,t) \\ N(a^l+1,e^l+1,t) \\ \vdots \\ N(a^h,e^l+1,t) \end{pmatrix}$,....,

$\mathbf{w}(a^h-a_0-1,t) = \begin{pmatrix} N(a^h-1,a^h-a_0-1,t) \\ N(a^h,a^h-a_0-1,t) \end{pmatrix}$, $\mathbf{w}(a^h-a_0,t) = N(a^h,a^h-a_0,t)$.

Let $\mathbf{x}_t$ as the vector with stacked up shares of teachers by age and experience. The vector $\mathbf{x}_t$ represents the distribution of teachers in period t.

$$\mathbf{x}_t = \begin{pmatrix} \mathbf{w}_0(a^l,t) \\ \mathbf{w}(e^l,t) \\ \mathbf{w}(e^l+1,t) \\ \vdots \\ \mathbf{w}(a^h-a^l,t) \end{pmatrix}.$$

By definition the sum of all elements of $\mathbf{x}_t$ is unity.

Denote the vector of attrition rate of $\mathbf{x}_{t-1}$ as $\mathbf{r}_{t-1}$. By definition, the element of $\mathbf{r}_{t-1}$ corresponding to $N(a,e,t-1)$ is $r(a,e,t-1)$. The attrition in year t can be written as

$$s(t) = \mathbf{r}(t-1)'\mathbf{x}_{t-1}.$$

Denote the vector that weighs the cells of the new-entrant teachers as $\mathbf{v} = \begin{pmatrix} \frac{h(e^l+1)}{a^*+e^*} \\ \vdots \\ \frac{h(a^l-a_0)}{a^*+e^*} \\ \frac{f(a^l)}{a^*+e^*} \\ \vdots \\ \frac{f(a^h-1)}{a^*+e^*} \end{pmatrix}$.

One can write the vector of the new entrant teachers as

$$\begin{pmatrix} \mathbf{w}_0(a^l,t) \\ \mathbf{w}(e^l,t) \end{pmatrix} = \mathbf{v}\mathbf{r}(t-1)'\mathbf{x}_{t-1}.$$

The non-new entrant teachers $N(a,e,t) = (1-r(a-1,e-1,t-1))N(a-1,e-1,t-1)$ from the remaining teachers in the cell $(a-1,e-1)$. Hence $\begin{pmatrix} \mathbf{w}(e^l+1,t) \\ \vdots \\ \mathbf{w}(a^h-a_0,t) \end{pmatrix} = \mathbf{B}(t-1)\mathbf{x}_{t-1}$, where the elements of the row of $\mathbf{B}(t-1)$ corresponding to $N(a,e,t)$ are all 0's, except for the single element that corresponds to $N(a-1,e-1,t-1)$, which equals $1-r(a-1,e-1,t-1)$.

The sum of the vector $\mathbf{1}'\mathbf{v}$=$\frac{\sum_{a=a^l}^{a^h-1} f(a)}{a^*+e^*} + \frac{\sum_{e=e^l+1}^{a^h-a_0-1} h(e)}{a*+e^*} = 1$. It follows that the sum of the column of matrix $\mathbf{v}\mathbf{r}(t-1)'$ corresponding to element $N(a-1,e-1,t-1)$ (whose age is below a^h) in $\mathbf{x}_{t-1}$ is $r(a-1,e-1,t-1)$, and 1 otherwise.

Putting these components together, the dynamics of the attrition and replacement can be summarized by the following relationship

$$\mathbf{x}_t = \mathbf{A}_{t-1}\mathbf{x}_{t-1}.$$

The transition matrix

$$\mathbf{A}_{t-1} = \begin{pmatrix} \mathbf{vr}(t-1)' \\ \mathbf{B}(t-1) \end{pmatrix}.$$

In period t, all elements of $\mathbf{A}_t$ is nonnegative and each column of $\mathbf{A}_t$ sums to unity.

The short-run and long-run effect of pension rules change

A change in pension rules in period $t-1$ changes attrition rates $r(a, e, t-1)$ in certain cells (a, e). The short-run effect depends on the changes in the attrition rates and the initial distribution $\mathbf{x}_0$. The time t distribution is given by

$$\mathbf{x}_t = (\prod_{i=0}^{t-1} \mathbf{A}_i)\mathbf{x}_0.$$

A once-and-for-all policy change in period 0 with initial distribution $\mathbf{x}_0$ is

$$\mathbf{x}_t = \mathbf{A}^t \mathbf{x}_0,$$

where $\mathbf{A}$ is the transition matrix corresponding to the new policy. In the long-run, the initial distribution no longer matters, and the effect of change in the pension rules is captured by the shift in the stationary distribution of teachers.

The stationary distribution of teachers is a vector of fixed share for each (age,experience) cell $N(a, e)$, with $\sum_{a=a^l}^{a^h} \sum_{e=e^l}^{a-a_0} N(a, e) = 1$ and a constant total attrition in each period.

Fact

(a). The stationary distribution is uniquely determined by the attrition $r(a, e)$ and the relative share of new entrant teachers $f(a)$ and $h(e)$.

(b). Starting from an arbitrary distribution $N(a, e, 0)$ $(a = a^l, .., a^h, e^l \leq e \leq a - a^l)$, $N(a, e, T) \to N(a, e)$ as $T \to \infty$.

For part (a), note the system with constant attrition rates can be written as $\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1}$. One can treat $\mathbf{A}$ as the transition matrix for a Markov Chain, and vector $\mathbf{x}_t$ as the probability distribution over states of teachers' age and experience. Because (1) attrition is less than unity prior to the maximum age, (2) any age-experience cell eventually leads to retirement, and (3) the replacement age distribution assigns a positive share to each age at the minimum experience, all states are positive recurrent and the chain is irreducible. Hence there is a unique stationary distribution.

For part (b), it is known that because the sum of each column of elements of the transition $\mathbf{A}$ equals 1, unity is an eigenvalue of $\mathbf{A}$ and other eigenvalues of $\mathbf{A}$ are less than unity. Consider the spectral decomposition $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$, where diagonal matrix $\mathbf{D}$ are the eigenvalues of $\mathbf{A}$. Let $\mathbf{x}_0$ be the vectorized distribution $N(a, e, 0)$, since all elements of the diagonal matrix $\mathbf{D}$ is less or equal to 1, $\mathbf{x}_{t+1} - \mathbf{x}_t = \mathbf{V}(\mathbf{D}^{t+1} - \mathbf{D}^t)\mathbf{V}^{-1}\mathbf{x}_0 \to \mathbf{0}$ for any $\mathbf{x}_0$.

References

Brown, Kristine (2013). “The Link Between Pensions and Retirement Timing: Lessons from California Teachers.” Journal of Public Economics 98, 1-14.

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

Costrell, Robert, and Joshua McGee (2010). “Teacher Pension Incentives, Retirement Behavior, and Potential for Reform in Arkansas.” Education Finance and Policy, 5(4), 492-518.

French, Eric, and John Bailey Jones (2011). ”The Effects of Health Insurance and Self-Insurance on Retirement Behavior.” Econometrica vol. 79 no. 3, 693-732.

Gustman, Alan. L. and Steinmeier Thomas L., ”A Structural Retirement Model”, Econometrica, May, 1986, Vol. 54, No. 3, pp. 555-584.

Harris, Douglas and Scott Adams (2007). “Understanding the Level and Causes of Teacher turnover: A comparison with other professions.” Economics of Education Review, 26, 325-337.

Keane, Michael P., and Kenneth I. Wolpin (1997). he Career Decisions of Young Men.?Journal of Political Economy, vol. 105, no. 3, 473?22.

Kim, Dongwoo, Cory Koedel, Shawn Ni, Michael Podgursky, and Weiwei Wu (2017). “The Late-Career Teacher Retention.” Working paper, University of Missouri 17-08.

Knapp, David, Kristine Brown, James Hosek, Michael Mattock, Beth Asch (2016). “Teacher Retirement Benefits and Teacher Retention: A Structural Modeling Approach.” Santa Monica, CA: RAND.

Koedel, Cory, Shawn Ni, and Michael Podgursky (2014). “Who Benefit form Pension Enhancement?” Education Finance and Policy, 9, 165-192.

Lumsdaine, Robin. L, James H. Stock, and David A. Wise (1992). “Pension Plan Provisions and Retirement: Men and Women, Medicare, and Models”, NBER Working Paper 4201.

Malanga, Steven and Josh McGee (2018). “Garden State Crowd Out: How New Jersey’s Pension Crisis Threatens the State Budget.” New York: Manhattan Institute (January)

Ni, Shawn and Michael Podgursky (2016). “How Teachers Respond to Pension System Incentives: New Estimates and Policy Applications.” Journal of Labor Economics, 34, 1075-1104.

Novy-Marx, Robert, and Joshua Rauh (2011). “Public Pension Promises: How Big Are They and What Are They Worth?” Journal of Finance, 66, 1211-1249.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain (2005). “Teachers, Schools, and Academic Achievement.” Econometrica, 73(2), 417-458.

Rust, J. and Phelan, C., ”How Social Security and Medicare affect Retirement Behavior in a World of Incomplete Markets”, Econometrica, Vol. 65, No. 4, 1997, pp. 781-831.

Stern, Steven (1997). “Approximate Solutions to Stochastic Dynamic Programs.” Econometric Theory, 13, 392-405.

Stinebrickner, Todd R. (2000). “Serially Correlated Variables in Dynamic, Discrete Choice Models.” Journal of Applied Econometrics, 15 (6), 595-624.

Stock, James and David Wise (1990). “Pensions, the Option Value of Work, and Retirement.” Econometrica, 58(5) (September), 1151-1180.