Pension Enhancements and Teacher Retirement

Wei Kong

Shawn Ni

Michael Podgursky

Weiwei Wu

Pension Enhancements and Teacher Retirement

Wei Kong
University of Missouri – Columbia

Shawn Ni
University of Missouri – Columbia

Michael Podgursky
CALDER/University of Missouri – Columbia

Weiwei Wu
University of Missouri – Columbia

Contents

Acknowledgments

Abstract

We examine how pension rule changes affected teacher retirement by estimating an option-value retirement model on a large cohort of late career Missouri public school teachers from 1994 to 2008. In so doing we offer potential solutions to several statistical challenges that arise in estimating structural models of retirement on large panel data sets. The first concerns modeling the formation of teacher expectations of future pension rules. The second is bias induced by baseline sample selection: in baseline cohorts we only observe teachers who are still working. This bias also evolves with pension rule changes. A third challenge arises from maximum-likelihood 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 in large panels. We address these challenges by incorporating policy expectations and sample selection directly into estimation of the likelihood. 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, and b) dramatic reductions in computation cost. Counter-factual simulations of the estimated structural model suggest that Missouri's pension enhancements led to earlier retirement by about 0.4 years on average for the 1994 cohort and by more than one year in a steady state. Enhancements increased steady state pension liabilities by 16 percent for senior teachers.

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

JEL codes: I21, J26, J38

1   Introduction

During the 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 almost every 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). Since pension rules affect retirement decisions and the length of teaching careers they also affect school staffing and can exacerbate (or ameliorate) "teacher shortages." (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 an increasing 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 1990s 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 predicting the effects of 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 Missouri administrative personnel data on public K-12 teachers from 1994 to 2008 to examine the implications of the pension-enhancement legislation. Missouri public school teachers, like nearly all public school employees, are covered by a defined benefit (DB) pension system in a state-wide educator plan – the Public School Retirement System (PSRS). Administrative data for teachers allow for a more precise 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.

Estimating a dynamic structural retirement model in the presence of pension rule changes presents three general challenges that have not been addressed in the 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. 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 and 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 this 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, computing likelihood values using teacher-level panel data is costly. 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 only 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 resulted in *earlier* retirement. For the baseline 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 a steady state, the enhancement reduced a typical career by more than one year. We also find the enhancements benefited senior teachers unevenly. Overall, the enhancements raised pension wealth (and hence plan liabilities) by 16 percent for senior teachers.

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 retirement eligible. The normal retirement age is one way that eligibility is determined. Minimum retirement ages vary across plans, typically between the ages of 55 - 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) 60 years of age with at least 5 years of teaching experience, b) 30 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 × FAS × rf, where S is service years (essentially years of experience in the system), FAS is final average salary (calculated as the average of the highest three years of salary,) rf is the replacement factor. The cost of living allowance (COLA) is capped at a percent of the initial retirement annuity pursuant to the rules of the pension system. The replacement factor is .025 (2.5%) up to 30 years. Thus, a teacher with 30 years experience and a final average salary of $60,000$ would receive a 30 × $60,000$ × 0.025= $45,000$ annuity. There are several other minor adjustments to the formula. A "25 and out" option 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 have changed over time. In the 1990s, state and local pension funds expe-rienced increases in their funding ratios (i.e., the ratio of assets to liabilities). The actuarial

surpluses (or small deficits) were used to justify legislation 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 (Koedel, Ni, Podgursky, 2014). In most states teachers' benefits were automatically and retroactively adjusted for active teachers 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.

Table 1 describes the series of enhancements that occurred in the Missouri PSRS. In 1992 the replacement factor was 0.021, final average salary was calculated based on the highest five years of earnings, and early retirement was possible through the "55 and 25" rule. The "55 and 25" rule allowed 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) she had to have accrued at least 25 years of system service. In 1994 the replacement rate was raised from 0.021 to 0.023. By 2002 the replacement factor had been raised from 0.023 to 0.025, the final-average-salary calculation changed from the highest five to highest three years of earnings, and the "25 and out" and "rule of 80" provisions had been incorporated into the system (the "rule of 80" is a more-flexible version of the "55 and 25" rule whereby retirement with full benefits can occur if age + experience sums to 80). In addition, the cap on the 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 de-cision depends on expectations of future pension rules since these rules determine retirement benefits. Our sample period spans the years of enhancements in Table 1. After a couple of

pension enhancements, teachers may have forecasted additional gains. Because teachers may be forward looking, the effects of future rule changes on current retirement decisions 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 who retired on 7/1/1998 may not have anticipated 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 (which covers 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 current pension rules and the expectations of future rules.

The Stock-Wise option-value model assumes that a teacher chooses the year of retirement to maximize the expected present value of the utility of the salary and retirement benefit flows given current information, but does not take into account the value of options in the future, as she is assumed to do in a dynamic programming setting. Several studies examine retirement decisions by simulating or estimating structural models through dynamic programming (e.g., Rust and Phelan 1997, French and Jones 2011, Knapp et. al., 2016).[1]

---

[1]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 omission of Social Security (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 the strong serial correlation of preference errors (based on empirical evidence presented later in this paper). In addition, solving the selection bias problem in the baseline sample 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. Dealing with these two statistical problems is much more difficult in a dynamic programming framework. For example, numerical solutions to a dynamic programming problem with the complications in this paper will involve significant approximation error. Finally, we see no simple way to use grouped (age, experience) cell data rather than teacher panel data in a dynamic programming framework. Fortunately, Lumsdaine et al. (1992) find that the option-value model yields similar predictions to a dynamic programming model. After weighing the trade-offs, we choose to use the simpler option-value model in this study. As will be noted below, we believe the good in-sample fit for our model provides justification for our empirical methodology.

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 beginning of period t, and $r(a, e, t)$ be the retirement probability in period t, $(t = 1, 2, ..., \Lambda)$. Struc-tural models such as the Stock-Wise option-value model of retirement dictates how $r(a, e, t)$ changes with a change in rules. In the sample of senior teachers 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. For the 1994 sample $a^l = 47, e^l = 5$. The teacher distribution in period $t > 1$ is given by

$$N(a, e, t) = N(a - 1, e - 1, t - 1)[1 - r(a - 1, e - 1, t - 1)], \tag{1}$$

7

for $a \geq a^l + t - 1$ and $e^l + t - 1 \leq e \leq a + t - a^l - 1$.

Starting from an initial distribution $N(a, e, 1)$ in the beginning of period 1, in the beginning of period $t > 1$ the remaining teachers are

$$N(a + t - 1, e + t - 1, t) = N(a, e, 1)s_t(a, e),$$

where $s_t(a, e) = \prod_{j=1}^{t-1}[1 - r(a+j-1, e+j-1, j)]$ is the survival rate of a teacher with initial age and experience (a, e) until the beginning of period t.

The probability of a teacher with initial age and experience (a, e) in period 1 retiring in period $t > 1$ is $G_{1,t}(a, e) = r(a + t - 1, e + t - 1, t)s_t(a, e)$. The probability she remains teaching at the end of the sample period is $s_\Lambda(a, e) = 1 - \sum_{t=1}^{\Lambda} G_{1,t}(a, e)$. 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 (estimated below).

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 (follow (M) for years 1995 and 1996, then 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 have anticipated more generous pension rules in the future. So, after 1997, we assume teachers give the current rule a weight p and rule of the next year a weight $1 - p$. This case is labeled as adaptive (A1). One may interpret the assumption (A1) as follows: with probability p a teacher does not pay attention to the news on pension rules or she is committed to making a retirement decision based on the current rules, and with probability $1 - p$ she is informed regarding the rule changes in the next AY and has the flexibility to make her 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 Myopic Expectations (M)

We introduce the model of retirement 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 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 lifespan). 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 of teacher'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. With this setting one dollar of salary is

worth k_s dollars of pension benefit in same period 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:

$$\nu_s = \rho \nu_{s-1} + \epsilon_s. \tag{3}$$

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:

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

where

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

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. We find the logarithm of salary is accurately predicted by a cubic function of experience. 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 Perfect Foresight (P)

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

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

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 denoted by (P1).

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

$$g_t^{P1}(m, R_t) = \sum_{s=t}^{m-1} \pi(s|t)\beta^{s-t}(k_s(1 - c_s(R_{t+1}, m))Y_s)^\gamma + \sum_{s=m}^{T} \pi(s|t)\beta^{s-t}(B_s(R_{t+1}, m))^\gamma$$
$$- \sum_{s=t}^{T} \pi(s|t)\beta^{s-t}(B_s(R_t, t))^\gamma.$$

The probability that the teacher retires in period t is given by (6) where $g_t^M(.,.)$ is replaced by $g_t^{P1}(.,.)$.

3.2.3 Adaptive Expectations (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 empirically 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

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

The retirement probability in Model (A1) can be computed similarly as that in Models (M) and (P).

The structural parameters are $\boldsymbol{\theta} = (\gamma, \kappa, \kappa_1, \beta, \sigma, \rho)$ for Models (M) and (P1); and $\boldsymbol{\theta} = (\gamma, \kappa, \kappa_1, \beta, \sigma, \rho, p)$ for Model (A1).

3.3   Likelihood by Age-Experience Cells

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

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

Here we omit the superscript on $g_t(m^\dagger, R_t)$ with respect to expectations of 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 $(n_i < \Lambda)$ 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 $\{\boldsymbol{\nu}_{1,n_i-1} > -\mathbf{f}_{1,n_i-1}^+ \bigcap (f_{n_i}^+ \leq -\nu_{n_i}) | in\ the\ initial\ sample\}$. If the teacher did not retire at the

12

end of the sample then $n_i = \Lambda$ and $\boldsymbol{\nu}_{1,\Lambda} > -\mathbf{f}^+_{1,\Lambda}$. Denote the probability of that the teacher observed in period 1 retires in n_i as

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

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

Denote the data on teacher i as $\mathbf{y}_i$, and data on all teachers $(i = 1, 2, .., I)$ as $\mathbf{Y}$. The likelihood of the sample data $\mathbf{Y}$ is

$$L(\mathbf{Y}; \boldsymbol{\theta}) = \prod_{i=1}^{I} \int_{A_i} \phi(\boldsymbol{\nu}_{1,n_i}) d\boldsymbol{\nu}_{1,n_i}, \tag{9}$$

where $\phi(.)$ denotes multivariate normal density distribution of $N(0, \boldsymbol{\Sigma}_i)$. The covariance matrix of $\boldsymbol{\nu}_{1,n_i}$, $\boldsymbol{\Sigma}_i$, depends on σ and ρ. 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 for a 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. In each period (from 1 to Λ), a senior teacher in our sample chooses between irreversible retirement or continuing teaching. If she chooses the latter both her age and experience gain by one. The retirement decision of teacher i is $y_{it} \in \{0, 1\}$, $t = 1, .., \Lambda$; where $y_{it} = 1$ if the teacher retires in period t, and $y_{it} = 0$ otherwise. Hence $\sum_{t=1}^{\Lambda} y_{it} = 1$ if the teacher retires and $\sum_{t=1}^{\Lambda} y_{it} = 0$ 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 by $G_{1t}(a, e)$. The likelihood function of parameter $\boldsymbol{\theta}$ of teacher i is

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

13

The likelihood of the whole sample is

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

We denote $N(a,e,1)$ as the number of teachers with (age, experience) (a,e) in the beginning of the initial period, we now denote the counts of retirement of teachers with initial (a,e) in period t as $R(a,e,t) = \sum_{i|a_i=a,e_i=e}y_{it}$, then

$$
L(\mathbf{Y};\boldsymbol{\theta}) = \prod_{t=1}^{\Lambda}\prod_{a,e}[G_{1t}(a,e)]^{R(a,e,t)}[1-\sum_{t=1}^{\Lambda}G_{1t}(a,e)]^{N(a,e,1)-\sum_{t=1}^{\Lambda}R(a,e,t)}. \tag{10}
$$

The alternative expression of the likelihood (10) means 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 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.)

Age and experience for teachers included in the sample are right-censored. We observe retirement age and experience for teachers who retired during our panel and do not observe the retirement year of the teachers who continued to teach at the end of the sample period (the latter is roughly 6% of the sample.) However, this censoring does not result in biased parameter estimates.

More consequential issue are left-truncated data. 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. There is no simple solution to the sample truncation 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.

For the "stayers" in 1994 (where $t = 1$), $\nu_0 \geq -\frac{g_0(m^\dagger, R_0)}{K_0(m^\dagger)}$. This implies that ν_0 differs from the unconditional stationary distribution $N(0, \frac{\sigma^2}{1-\rho^2})$. Without taking into account this sample selection bias, one would draw the initial value ν_1 from the unconditional stationary distribution. This would result in over-prediction of retirement in the initial years. The longer a teacher has become retirement eligible in the initial sample (say J years ago) the more likely her preference shock in the initial period has a large value.

In the present setting, selection (of "early leavers") occurs when ν_0 falls below a threshold (that is a function of J). Hence the sample selection in this setting is an example of the familiar "initial condition problem" in dynamic panel data models (Heckman, 1981; Wooldridge, 2005). In a typical initial condition context, ν_0 is correlated with latent teacher-specific effects and observations prior to the initial year may be missing for some teachers, but the whole sample of teachers is observed. In our context we do not have any data on the "early leavers" and we are thus unable to apply off-the-shelf solutions to the initial condition problems for nonlinear panel data models. In our problem, the sample selection process is the same as the decision model of retirement, but widely-used two-step procedures for sample selection (e.g., Heckman 1979) are not applicable here because we do not have additional data to separately estimate the selection process.

We solve the problem of missing "early leavers" by estimating the model conditioning on the probability that retirement-eligible teachers are in the initial sample. The option-value model depicts how the sample selection depends on the preference errors prior to the initial sample period, $(\nu_{-J}, ..., \nu_0)$. The likelihood of the sample can then be computed by integrating out these preference errors as well as the preference errors in the sample period as in (9). The computational cost of the exercise can be prohibitively high for a general problem, but not so for the present problem for reasons explained below.

The retirement probability conditional on being observed in the sample is

$$prob(retiring\ in\ period\ n|in\ initial\ sample) = \frac{prob(retiring\ in\ period\ n, in\ initial\ sample)}{prob(in\ initial\ sample)}. (11)$$

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 $\nu_{-J} \sim N(0, \frac{\sigma^2}{1-\rho^2})$. 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

$$\boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0}. \tag{12}$$

The precise formula for the statement in (11) is

$$\begin{aligned} G_{1,n} &= prob[((\boldsymbol{\nu}_{1,n-1} > -\mathbf{f}^+_{1,n-1}) \bigcap (f^+_n \leq -\nu_n)) | \boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0}] \\ &= \frac{prob[(\boldsymbol{\nu}_{-J,n-1} > -\mathbf{f}^+_{-J,n-1}) \bigcap (f^+_n \leq -\nu_n)]}{prob(\boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0})}. \end{aligned} \tag{13}$$

$Algorithm1(t_1, t_2)$ 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 (13), $prob[(\boldsymbol{\nu}_{-J,n-1} > -\mathbf{f}^+_{-J,n-1}) \bigcap (f^+_n \leq -\nu_n)]$, 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. The algorithm is also given in Appendix 1. The probability in the denominator in (13), $prob(\boldsymbol{\nu}_{-J,0} > -\mathbf{f}^+_{-J,0})$ can be computed using Algorithm$2(-J, 0)$.

The algorithms are based on the Geweke–Hajivassiliou–Keane (GHK) simulator (see e.g., Börsch-Supan and Hajivassiliou 1993), which is more efficient in computing high-dimensional integrations than brute-force Monte Carlo simulations. A new feature in the application of the GHK algorithm here is that 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.

In these algorithms one may note that the left-side truncation of the preference shocks

16

before the initial sample period, $(\epsilon_{-J}, .., \epsilon_0)$, 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 instead of using classical methods.

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.

(Insert Table 2 here.)

Obtaining the MLE estimates requires repeated evaluations of the likelihood of the panels with a large number of teachers. The likelihood evaluation involves numerical integrations of dimensions of the length of the sample period (in this study 14 plus the years of retirement eligibility prior to 1994), and if likelihood is evaluated for each teacher, the high-dimensional integrations are needed 12,871 times for one likelihood evaluation of the whole sample. We achieve relatively low computational cost in two ways. First, we use Algorithm1 and Algorithm2 for efficient computation of high-dimensional integrations. Second, we track the counts of (age,experience) cells instead of teacher-level panel data. The number of cells are fixed by the range of age and experience. Teachers in the 1994 cohort with age 47-64 and 5-46 years of experience are grouped into 513 cells ($18 \times 42 = 756$ cells subtracting cells where the experience is too high for the age, assuming the minimum age for a novice teacher is 22) 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 it would take about 25 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.

The retirement decisions depend on the expectations of future pension rules. Table 3 shows MLE estimates of structural parameters, $(\beta, \kappa, \kappa_1, \gamma, \sigma, \rho, p)$, 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 0.955, or 4.5% annual 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 for the retirement decision: e.g., the teacher's health or family-related factors. 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 large and positive. Recall the parameter k_s measures the disutility of working. The disutility of working depends on age. The estimates of κ implies that at age 60, one dollar of salary is equivalent to about seventy cents of pension benefit.

For female and male teachers the likelihood of adaptive expectation case (A1) is larger than other cases, as we would 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 the survival

rate and the age/experience distribution of retiring teachers. 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.

4.2   In-Sample Goodness of Fit

We begin with overall employment versus retirement. 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. Unless gender is specified, the tables and figures in the paper pertain to data or simulation results combined by gender. Figure 3 plots the observed fraction of the 1994 cohort who remain teaching each year from 1995 to 2008, and the simulated survival rates under different models of pension rule expectations. The simulated survival rates from all three models of expectations track the observed rate quite closely.

(Insert Figure 3 here.)

Table 4 reports the year-by-year prediction of the share of retiring teachers based on MLE estimates under expectation models (M), (A1), and (P1). Among the three models (A1) does best in terms of the in-sample MSE. For most years, the prediction errors for all three models are within a reasonable range, except for 1998 and 2001. Closer examination of the misfits of the model in these two years reveals the difference in the nature of the expectation models and their limitations.

(Insert Table 4 here.)

As noted in Table 1, in 1997 the rule "25 and out" was introduced to allow early retirement. There is a large increase in observed retirement in 1997 over that of 1996, followed by a slight decrease in 1998. This is expected if some teachers chose early retirement following the new rule. The myopic Model (M) under-predicts retirements in 1997 and substantially over-predicts retirements in 1998. The misfits are concentrated in cells that are directly affected by the introduction of "25 and out" (the cells with age below 55 and experience between 25 and 30.) In contrast, the perfect foresight model slightly over-predicts retirements

19

in 1997. The perfect foresight model fits better than the myopic model. As we noted in Section 2.2, the "25 and out" rule was announced early in the AY 1997. The prediction of the adaptive expectation model (A1) is in between of the two extreme models. Its prediction errors in 1997 and 1998 have the same signs of prediction errors as Model (M), but are much smaller in magnitude.

Unlike the misfit of 1997, which is easily explained by expectations, in 2001 all three models substantially under-predict retirement. One may wonder whether this is because the model poorly predicts responses to the introduction of the "rule of 80" in year 2000. However, that is not the case since in our cohort of teachers, by year 2000 most teachers who satisfy the "rule of 80" met the regular retirement requirements in the absence of the rule anyway. Hence the new rule is not expected to have a significant effect. Closer examination at the fit of age-experience cells in 2001 shows that the under-prediction is largely because of cells with experience over 30 years and age 53-58 that are legible for retirement without the "rule of 80" and not because of the cells made retirement legible by the rule.

We speculate that the three models under-predict retirement in 2001 because there are important elements missing in all three models. In early 2000, the U.S. stock market peaked and started to decline after the several years of unusually high returns. During the dot.com bubble the annual returns to NASDAQ composite index were 39.9% in 1995; 22.7% in 1996; 21.6% in 1997; 39.6% in 1998; 85.6% in 1999; and -39.3% in 2000. As we noted earlier, the pension enhancements nationwide were correlated with the stock market run-up in the 1990s. It is likely that teachers associated the end of the bull market with the end of major future enhancements. Such expectation would be reasonable and turn out to be correct (even though a minor enhancement was implemented in 2002.) The expectation of the end of enhancements pushes more teachers to retirement than our models predict. If we introduce a time fixed effect into the model it would eliminate the misfit of 2001 for all models, but it is hard to justify and unhelpful for out-of-sample prediction.

Overall, the models fit well for most years during the sample period. To examine how well the estimated models fit data in more detail, we also compare the observed and simulated age distribution of retiring teachers every year from 1995 to 2008. The plots (available upon request) show that age-experience distributions simulated under different expectations

20

of pension rules are very similar. In addition, the simulated distributions almost perfectly match the observed distribution for teachers who choose to teach each year.

Figure 4 shows that all models capture the pattern of the joint age-experience distributions for retired teachers quite well. This suggests that the option value model captures not only the dynamics of the number of retirements, but also the age and experience of the teachers at the time of retirement.

(Insert Figure 4 here.)

Recall that our data in the initial year (AY94) only includes teachers who are still teaching in the initial year and excludes those who already retired before then. Without adjusting for sample selection bias, one would draw the initial value ν_0 from the unconditional stationary distribution $N(0, \frac{\sigma^2}{1-\rho^2})$ and compute the unconditional probability of retirement. This will result in over-prediction of retirement in the initial years for reasons explained in Section 3.4. Figure 5 reports the observed and predicted survival rates with and without adjustment for sample selection bias under (A1) expectation of pension rules. In the first three years, without adjusting for sample selection the model predicts lower survival rates (and predicts more retirements). This pattern holds for any model of pension rule expectation and for any set of parameters. In Figure 5 the survival rate simulated from Model (A1) adjusted for sample selection bias fits the data much better than the model without adjustment.

(Insert Figure 5 here.)

4.3 Out-of-Sample Goodness of Fit

In this section, we examine the out-of-sample 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 aged 47-64 with at least 5 years of 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. However, the simulation results will differ because of differences in estimated parameters estimated based on different expectation

models.

Table 5 compares the observed and simulated survival rates. The last three columns of Table 5 reports the residuals of simulated survival rates. The out-of-sample predictions under myopic expectations tend to error on the side of under-prediction of survival rates. Overall, the models fit the data well. In particular, the adaptive expectation model provides a good fit to the actual values. It is not surprising that the adaptive model would provide better in-sample fit, since it involves an extra parameter. However, the resulting core behavioral parameters that emerge from the A1 model seem to provide better out-of-sample fit as well (even in the absence of plan changes).

(Insert Table 5 here.)

5 The Effect of Pension Enhancements

Pension enhancements change the gain in the 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 "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, but this turns out have a minor effect.) 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 also 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

$$g_t(R, m) = \sum_{s=t}^{m-1} \pi(s|t)\beta^{s-t}(k_s(1-c)Y_s)^\gamma + \sum_{s=m}^{T} \pi(s|t)\beta^{s-t}(B_s(R,m))^\gamma$$
$$- \sum_{s=t}^{T} \pi(s|t)\beta^{s-t}(B_s(R,t))^\gamma.$$

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

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

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 estimates 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.

Lowering the eligibility requirement on age and/or experience leads to earlier retirement. But the short-term effects 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 the "rule of 80". In this section, we conduct counterfactual analysis on how teacher retirement would differ from the historical data if no "25 and out", "rule of 80", or any enhancements took place.

To quantify the effects of pension enhancements on retirement of teachers in 1994 cohort, we simulate the retirement probability for the next 30 years. The baseline case is the 1994 cohort that experienced the enhancements which occurred in the sample period. Table 6 reports the simulated average retirement age and experience in different counterfactual cases under different expectation assumptions. The average retirement age and experience in the baseline case are lower than in all counterfactual cases meaning the pension enhancements induced teachers 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

24

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 and/or experience thresholds that are affected by the rule changes. Because the rule changes only affected part of the sample, the simulated short-term effects understate the full effect of the rule changes.

Appendix 2 shows that as the retired teachers are replaced by new senior entrants (those with age of 47 or experience of 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 7 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 6, the averages over teachers in the steady state are 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 larger. The enhancements cause teachers to retire 1.06-1.20 years earlier, while the effect of enhancements is only 0.36-0.50 years for the 1994 cohort. The larger effects are due to the simulated long-term effects when all new entrants are subject to the pension enhancements.

(Insert Tables 6-7 here.)

25

5.2.1 Differential Financial Effects

Pension enhancements, especially two policy changes, "25 and out" and "rule of 80", have different effects on the expected pension benefit for senior teachers with similar age and experience. Figure 6 reports the distribution of the percentage change of expected pension wealth under four scenarios compared to the case of no enhancement: (i) with "25 and out" (without other enhancements); (ii) with "the rule of 80" (without other enhancements); (iii) increase in the replacement rate (without other enhancements); and (iv) all of the enhancements. 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 effects of rule "25 and out" on pension wealth of teachers with experience less than 25 years differ by age and experience. A teacher with (age, experience) of $(52, 24)$ would have to wait 3 years (until she hits $(55,27)$) to qualify for regular retirement without the "25 and out" provision; and only one year with "25 and out" for early retirement (when she hits $(51,25)$). Obviously introduction of "25 and out" expands her choice set. In contrast, a teacher in the cell $(54, 24)$ would only wait one year (until she hits $(55, 25)$) to qualify for regular retirement; and introduction of the "25 and out" rule does not affect her at all.

Moreover, the "25 and out" rule expands the choice sets for teachers in the age-experience cells that lead to $(52, 24)$ (i.e., $(51, 23)$, $(50, 22)$, etc.), but does not affect the choice set of cells that lead to $(54,24)$ (i.e, $(53, 23)$, $(52, 22)$, etc.). The differential effects are continuous across the (age, experience) plane. So the cells with increased pension wealth form a small mountain (with the ridge along the path of $(52, 24)$, $(51, 23)$, $(50, 22)$, etc.).

It's worth noting that whether a new retirement option raises expected pension wealth depends on the balance of the benefit/collection-years 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. For teachers in the cells with positive pension wealth gains the effect of the additional collection years dominates the effect of early retirement penalty.

Introduction of the "rule of 80" broadens eligibility of regular retirement and creates pen-

26

sion wealth gains along two ridges on the (age, experience) plane in Figure 6. The additional flexibility has differential effects on the expected pension wealth of teachers in different (a, e) cells. Per rules in Table 1, prior to enhancements, a regular retirement requires $a \geq 60$, or $e \geq 30$, or $a \geq 55$ *and* $e \geq 25$. Consider the benchmark age-experience combinations on the (a, e) plane: $(60, 20), (55, 25), (50, 30)$. All three cells satisfy the conditions of regular retirement in the absence of the "rule of 80". However, cells on a straight line connecting $(60, 20)$ and $(55, 25)$ (hence satisfy the "rule of 80") do not qualify for regular retirement without the "rule of 80". For example, teachers with $(57, 23)$ need to wait 2 years (to hit $(59, 25)$ before they qualify for regular retirement. Introduction of the "rule of 80" makes the $(57, 23)$-teacher immediately eligible for regular retirement. The "rule of 80" also shortens the years-to-eligibility of the teachers by 2 years for cells leading to $(57, 23)$ (i.e., $(56, 22)$, $(55, 21), (54, 20)$, etc.). But the rule does not change the years-to-eligibility of the teachers for cells leading to $(60, 20)$. This differential gains in expected pension wealth form one ridge (the one on the left) in Figure 6.

The same scenario occurs for cells on a line connecting $(55, 25)$ and $(50, 30)$. Without the "rule of 80", teachers with $(53, 27)$ or $(52, 28)$ need to wait 2 years (to hit $(55, 29)$ or $(54, 30)$) before they qualify for regular retirement. Introduction of the "rule of 80" shortens the years-to-eligibility by 2 years for these cells and for cells leading to them (i.e., $(52, 26)$, $(51, 25)$, etc.; or $(51, 27), (50, 26)$, etc.); and forms the other ridge (on the right) in Figure 6.

(Insert Figure 6 here.)

5.2.2 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 retirement rate $r(a, e, t)$. By definition $\sum_{a,e} N(a, e, t) = 1$. Let the present value of pension wealth for each teacher retiring at (a, e, t) be $P(a, e, t)$. Then total pension cost of the teachers is $\bar{P}_t = \sum_{a,e} N(a, e, t) r(a, e, t) P(a, e, t)$.

Pension enhancements may result in uneven gains in pension wealth for retiring teachers

with different age-experience combinations. This may alter retirement decisions for some teachers, resulting in changes in $r(a,e,t)$. The change in retirement probabilities leads a change in the age-experience distribution in the long run. Hence pension enhancements affect pension cost by altering $N(a,e,t)$, $r(a,e,t)$, and $P(a,e,t)$. In various parts of this paper we discuss how pension rules affect these components. The main focus of the paper is on how pension rules affect retirement probability $r(a,e,t)$. Given retirement probability $r(a,e,t)$ and current pension wealth $P(a,e,t)$, one can compute expected pension wealth. Section 5.2.1 discussed the differential effects on expected pension wealth. The model in Appendix 2 quantifies the dynamics of the distributions $N(a,e,t)$, given $r(a,e,t)$.

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 superscript $*$). Then the change in pension cost in the short run is $\bar{P}^*_t - \bar{P}_t$; and that in the steady state should be a constant (either positive or negative) that depends on the parameters and the rules.

Table 8 reports pension wealth and the retirement rate for senior teachers in a steady state, under different counterfactual scenarios. The total number of teachers is normalized to one. The stationary distribution is the snapshot of the senior teachers aged 47-80 and with at least 5 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 $(\bar{P}^*/\bar{P} - 1)$. The table shows that in the long term, without "25 and out" the pension cost would be lowered by rough 2% as compared to current rules; and without the "rule of 80" the pension cost would be lowered by a slightly lower magnitude. Another important pension enhancement is the increase of replacement factor that raises pension wealth for all teachers. If none of the enhancements were made from the 1994 rules, the pension cost would be about 14% lower, i.e., the enhancements raised the steady-state liabilities of the pension plan by roughly 16% per senior teacher. These estimates are robust with respect to the parameters estimated under different assumptions on teacher expectations on pension rules.

(Insert Table 8 here.)

6 Conclusion

In this paper we examined the effect of pension rule enhancements during the 1990s on retire-ment behavior of Missouri public school teachers. We estimated an option-value retirement model after overcoming several statistical challenges. The resulting models exhibit very good in-and out-of sample fit. We then used the estimated model to evaluate the effects of 1990's pension enhancements on retirement behavior. Our simulations show that the enhancements lead to earlier retirement of senior teachers – roughly 0.4 years for the 1994 cohort and by more than one year in a steady state. Steady state liabilities grew by roughly 16 percent per senior teacher. 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. Sweeping rule changes such as switching from a DB plan to a defined contribution (DC) plan may also be considered. Structural methods such as those developed in this paper would be useful in estimating the costs and benefits of such high-stakes changes.

The methodology developed is readily extended to analysis of teacher (and non-teacher) 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 option-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: Major Changes in PSRS Pension Rules

Academic Year	Replacement Factor	COLA	Retirement Age and Experience
1992	0.021	0.56	Age $\geq$ 55 *and* Exp$\geq$25, or
			Age$\geq$60 and Exp$\geq$5, or Exp$\geq$30
1994	0.023	0.56	same
1995	0.023	0.65	same
1996	0.023	0.65	Salary includes employer paid health insurance
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$80
2001	0.025	0.80	same
2002	0.0255 if Exp$\geq$31	0.80	same

Note: If a teacher satisfies one of conditions for regular retirement rules, she may choose regular retirement. If a teacher does not satisfy regular retirement, she may choose the rule of "25 and out" with a reduced annuity after AY1997. COLAs in the table are the maximum annual cost of living allowances relative to the initial annual pension benefit. Before 1995, the teacher contribution rate was 10%. During 1996-2004, contribution was 10.5%. It was increased by 0.5% annually 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		
	(M)	(A1)	(P1)	(M)	(A1)	(P1)
β	0.956	0.962	0.958	0.957	0.954	0.962
	(0.004)	(0.004)	(0.004)	(0.006)	(0.011)	(0.006)
γ	0.656	0.708	0.634	0.615	0.666	0.701
	(0.016)	(0.017)	(0.018)	(0.029)	(0.031)	(0.032)
σ	2658.229	4874.581	2231.140	2543.740	4612.812	3966.185
	(460.277)	(863.353)	(421.438)	(812.413)	(1509.311)	(1358.974)
ρ	0.532	0.484	0.478	0.416	0.422	0.479
	(0.007)	(0.007)	(0.007)	(0.015)	(0.018)	(0.010)
κ	0.690	0.612	0.669	0.774	0.747	0.584
	(0.019)	(0.020)	(0.020)	(0.027)	(0.037)	(0.025)
$\kappa 1$	0.453	1.057	0.401	0.392	1.405	0.870
	(0.142)	(0.145)	(0.140)	(0.180)	(0.262)	(0.229)
p		0.284			0.477	
		(0.034)			(0.065)	
log-likelihood	-22130.020	-21995.200	-22175.660	-7637.431	-7526.444	-7641.678

Note: The standard errors are in parentheses. The sample includes Missouri PSRS teachers with age 47-64 and 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 (A1).

Table 4: Observed and Simulated Retirement Rate by Year

Retirement Year	Observed	(M)	(A1)	(P1)	(M)-Obs.	(A1)-Obs.	(P1)-Obs.
1995	0.0524	0.0608	0.0612	0.0589	0.0084	0.0089	0.0065
1996	0.0671	0.0630	0.0639	0.0666	-0.0041	-0.0032	-0.0005
1997	0.0914	0.0750	0.0883	0.1014	-0.0164	-0.0030	0.0100
1998	0.0851	0.1062	0.0972	0.0894	0.0211	0.0121	0.0044
1999	0.0927	0.0914	0.0960	0.0951	-0.0013	0.0033	0.0024
2000	0.0942	0.0945	0.0965	0.0944	0.0003	0.0024	0.0002
2001	0.1067	0.0896	0.0876	0.0825	-0.0171	-0.0191	-0.0242
2002	0.0778	0.0768	0.0750	0.0707	-0.0010	-0.0028	-0.0071
2003	0.0641	0.0658	0.0658	0.0636	0.0017	0.0017	-0.0005
2004	0.0576	0.0567	0.0570	0.0555	-0.0009	-0.0007	-0.0021
2005	0.0534	0.0477	0.0481	0.0478	-0.0056	-0.0053	-0.0056
2006	0.0420	0.0398	0.0400	0.0403	-0.0023	-0.0020	-0.0017
2007	0.0320	0.0323	0.0322	0.0330	0.0003	0.0002	0.0010
2008	0.0235	0.0255	0.0250	0.0262	0.0020	0.0016	0.0027
Not retired	0.0601	0.0751	0.0661	0.0746	0.0150	0.0060	0.0144
MSE(10^{-5})	-	-	-	-	9.1078	4.7215	7.0551

Note: We use administrative data during 1994-2008 for Missouri PSRS teachers with age 47-64 and at least 5 years of experience in 1994. The observed retirement rates are the number of retired teachers each year (column 2 in Table 2) divided by the total number of teachers (12871). Simulated retirement rates are the number of retired teachers each year divided by the total number of teachers. Simulations are based on the different estimates of expectation assumptions: myopic (M), one step perfect foresight (P1), adaptive expectation (A1). The last three columns report the residuals of predicted retirement rates (the predicted rates minus the observed ones.)

Table 5: Out-of-Sample Observed and Simulated Survival Rates

Year	Observed	(M)	(A1)	(P1)	(M)-Obs.	(A1)-Obs.	(P1)-Obs.
2012	0.8938	0.8824	0.8865	0.8902	-0.0114	-0.0073	-0.0036
2013	0.7969	0.7918	0.7943	0.8010	-0.0051	-0.0026	0.0041
2014	0.6991	0.7063	0.7074	0.7175	0.0072	0.0083	0.0184

Note: We use administrative data during 2011-2014 for Missouri PSRS teachers with age 47-64 and 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 (A1). The last three columns report the residuals of predicted survival rates minus the observed ones.

Table 6: Simulated Average Retirement Age and Experience: 1994 Cohort

		Avg Retirement Age	Avg Retirement Experience
M	Rules for the 1994 cohort	58.51	27.84
	No 25 and out	58.66 (0.15)	27.99 (0.15)
	No rule of 80	58.57 (0.06)	27.90 (0.06)
	No Enhancements	58.87 (0.36)	28.20 (0.36)
A1	Rules for the 1994 cohort	58.35	27.68
	No 25 and out	58.51 (0.16)	27.84 (0.16)
	No rule of 80	58.44 (0.09)	27.77 (0.09)
	No Enhancements	58.73 (0.38)	28.06 (0.38)
P1	Rules for the 1994 cohort	58.42	27.75
	No 25 and out	58.64 (0.22)	27.97 (0.22)
	No rule of 80	58.52 (0.10)	27.85 (0.10)
	No Enhancements	58.92 (0.50)	28.25 (0.50)

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 (A1). Teachers in 1994 cohort are Missouri PSRS teachers who are age 47-64 with at least 5 years of experience in 1994. Reported in the parenthesis are the average simulated age and experience under counter-factual scenarios subtracting that under the enhancements experienced by the 1994 cohort.

Table 7: Simulated Average Retirement Age and Experience: Steady State

		Avg Retirement Age	Avg Retirement Exp
M	With All Enhancements	57.00	25.04
	No 25 and out	57.55 (0.55)	25.59 (0.55)
	No rule of 80	57.14 (0.14)	25.18 (0.14)
	No Enhancements	58.19 (1.19)	26.23 (1.19)
A1	With All Enhancements	57.11	25.16
	No 25 and out	57.56 (0.45)	25.61 (0.45)
	No rule of 80	57.30 (0.19)	25.35 (0.19)
	No Enhancements	58.17 (1.06)	26.22 (1.06)
P1	With All Enhancements	57.28	25.32
	No 25 and out	57.81 (0.53)	25.85 (0.53)
	No rule of 80	57.46 (0.18)	25.50 (0.18)
	No Enhancements	58.48 (1.20)	26.52 (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 (A1). Reported in the parenthesis are the average simulated age and experience under counter-factual scenarios subtracting that under the case of all enhancements.

Table 8: Pension Wealth and Retirement Rate Under Counterfactual Scenarios: Steady State

		Effect on Pension Wealth	Retirement rate
M	With All Enhancements	-	0.0895
	No 25 and out	-2.45%	0.0852
	No rule of 80	-1.76%	0.0883
	No Enhancements	-14.56%	0.0808
A1	With All Enhancements	-	0.0885
	No 25 and out	-1.84%	0.0851
	No rule of 80	-1.91%	0.0871
	No Enhancements	-13.84%	0.0809
P1	With All Enhancements	-	0.0872
	No 25 and out	-1.97%	0.0834
	No rule of 80	-1.87%	0.0859
	No Enhancements	-14.06%	0.0789

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 of least 5 years at the steady state. Effect on Pension Cost is the percentage of pension cost reduction under counterfactual cases from 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 (A1).
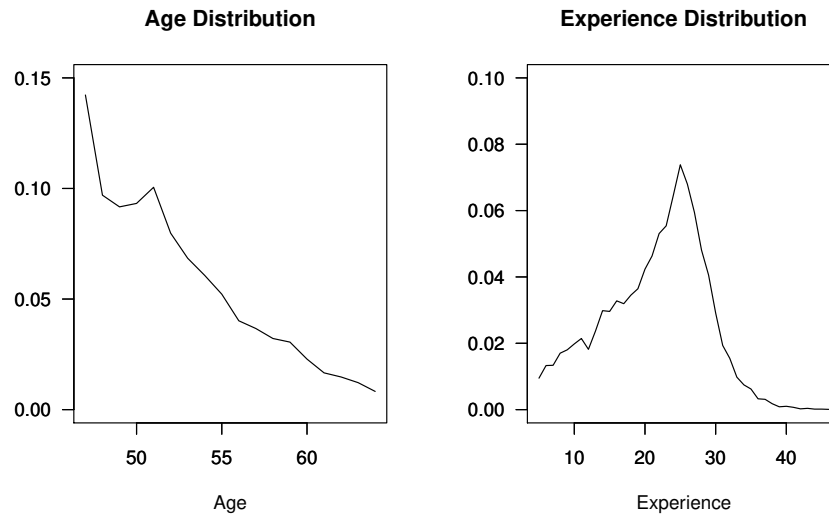
Figure 1: Age/Experience Distribution of the Initial Sample

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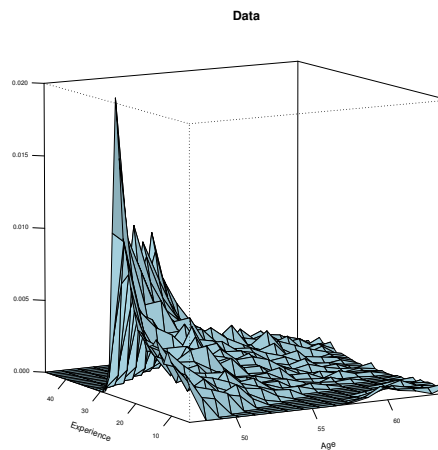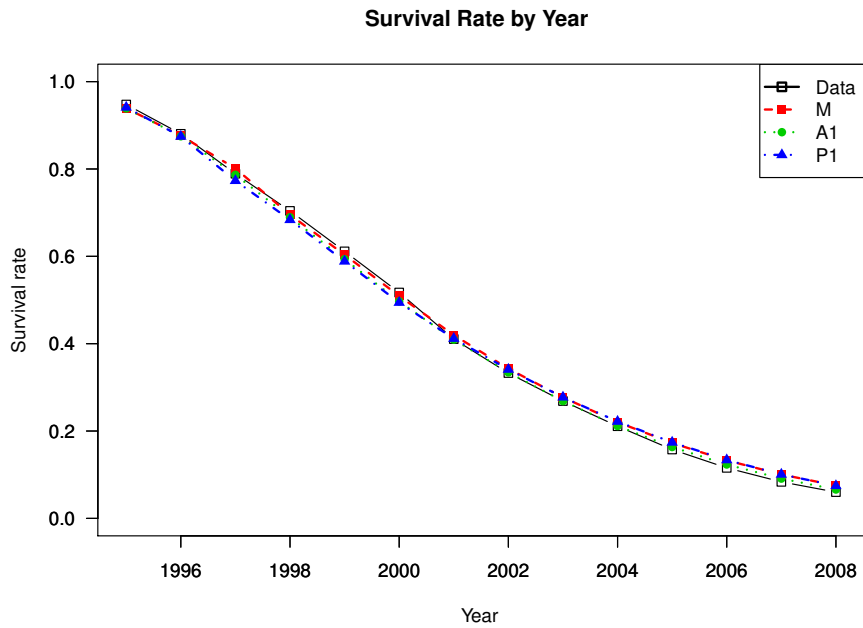

Figure 2: Age-Experience Distribution of the Initial Sample

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

36

Survival Rate by Year

Figure 3: Observed and Simulated Survival Rates

Note: Observed survival rate is based on the administrative data during 1994-2008 for Missouri PSRS teachers with age 47-64 and 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 (A1).
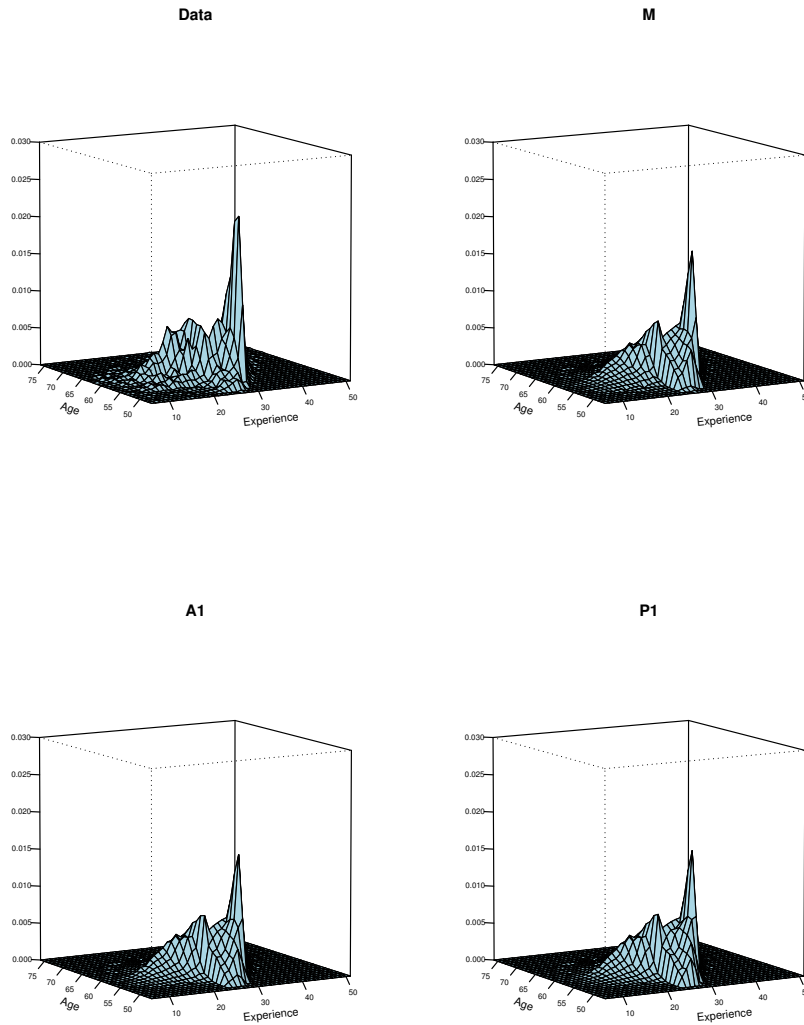
Figure 4: Observed and Simulated Joint Age-Experience Distributions for Retired Teachers

Note: Observed age-experience distribution is based on all teachers of the 1994 cohort at the time of retirement. Simulated age-experience distributions are based on the different estimates of expectation assumptions, including myopic (M), one step perfect foresight (P1), adaptive expectation (A1).
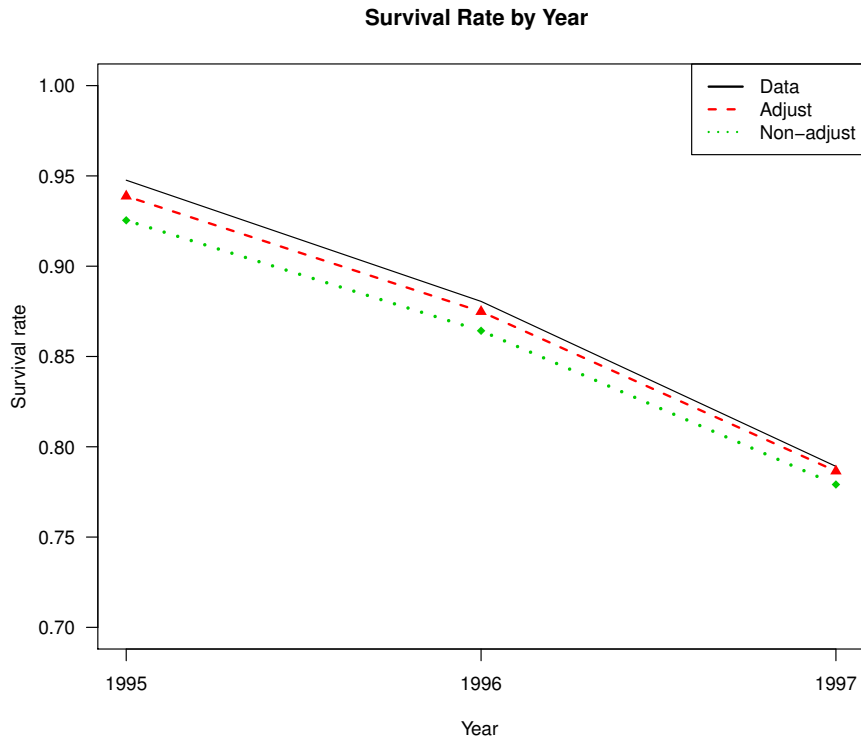
Figure 5: Observed and Predicted Survival Rates With and Without Selection Bias Correction

Note: Observed survival rate is based on the administrative data during 1994-2008 for Missouri PSRS teachers with age 47-64 and at least 5 years of experience in 1994. Simulated survival rate labeled "adjusting" is based on the estimated parameters from adaptive expectation (A1) and adjusted for the sample selection bias. Simulated survival rate "Non-adjusting" is based on the same estimated parameters but without adjusting for the sample selection bias.
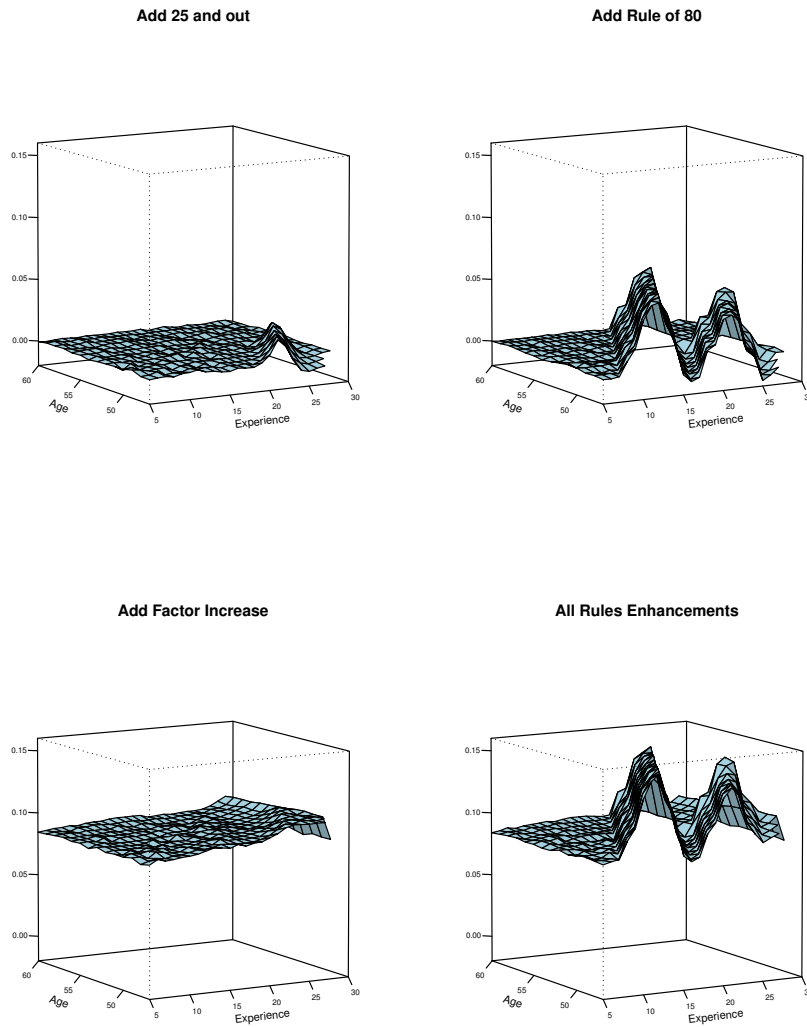
Figure 6: Distribution of Percentage Change of Pension Wealth

Note: The percentage change of expected pension wealth in each case of pension enhancement is the fraction of pension wealth gain over that with no enhancement of cases 1 to 3. 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 enhancements of Cases 1-3.

Appendix

Appendix 1. Algorithms for Computing Likelihood

Denote the normal $N(m, \sigma^2)$ truncated at (a, b) as $TN_{(a,b)}(m, \sigma^2)$. For all t, f_t^+ is defined in (8).

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 $TN_{(-f_{t_1}^+, \infty)}(0, \frac{\sigma^2}{1-\rho^2})$. This is equivalent to drawing $\epsilon_{t_1}^{(k)}$ 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 $\nu_{t-1}^{(k)}$ (or $\epsilon_{t_1}^{(k)}, .., \epsilon_{t-1}^{(k)}$) draw $\epsilon_t^{(k)}$ from $TN_{(-f_t^+ - \rho\nu_{t-1}^{(k)}, \infty)}(0, \sigma^2)$, which implies $\nu_t^{(k)} = \rho\nu_{t-1}^{(k)} + \epsilon_t^{(k)} > -f_t^+$,

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

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 $(1 \leq j \leq t)$ drawn from the truncated conditional distributions.

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

where $\Phi(.)$ is the cdf of the standard normal. The probability can be computed using the K simulated sequences of ν's as

$$
G_{-J,n} \approx \frac{1}{K}\sum_{k=1}^{K}[\prod_{t=-J+1}^{n-1}\Phi(\frac{f_t^+ + \rho\nu_{t-1}^{(k)}}{\sigma})]\Phi(-\frac{f_n^+ + \rho\nu_{n-1}^{(k)}}{\sigma}). \tag{14}
$$

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 use

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 $prob((\rho\nu_{t-1} + \epsilon_t > -f_t^+)|\nu_{t-1}) = 1 - \Phi(\frac{-f_t^+ - \rho\nu_{t-1}}{\sigma}) = \Phi(\frac{f_t^+ + \rho\nu_{t-1}}{\sigma})$. The retirement probability conditional on being observed in the sample is

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

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

$$
prob(\boldsymbol{\nu}_{-J,0} > -\mathbf{f}_{-J,0}^+) \approx \frac{1}{K} \sum_{k=1}^{K} [\prod_{t=-J+1}^{0} \Phi(\frac{f_t^+ + \rho\nu_{t-1}^{(k)}}{\sigma})]. \tag{15}
$$

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 senior teachers is fixed. Without losing generality, we let the fixed number be 1.

2. All senior 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 is set at 80 in the policy

42

simulations.)

5. The minimum experience in the sample is e^l ($e^l = 5$ for the sample used in the study), $e^l \leq e \leq a - a_0$. The last inequality says that a teacher's experience e can not exceed $a - a_0$ because her starting age of teaching can not be lower than 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 senior 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 $e^l + 1 \leq e \leq a^l - a_0$. Denote $\sum_{a=a^l}^{a^h-1} f(a) = a^*$, $\sum_{e=e^l+1}^{a^l-a_0} h(e) = e^*$. We count the (a^l, e^l) cell 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 senior 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

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

By assumption 1, the vacancies due to retirements are filled in the next period by new senior 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 as the Stock-Wise

option-value model or a dynamic programming model dictate 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 the size of the (a, e) cell in period t can be traced back using (1) to a cell with minimum age or minimum experience.

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 of 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 $t - 1$ is $\mathbf{r}(t - 1)'\mathbf{x}_{t-1}$.

44

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 senior entrant teachers as

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

The number of non-entrant teachers from the remaining teachers in the cell $(a-1, e-1)$ is $N(a, e, t) = (1 - r(a-1, e-1, t-1))N(a-1, e-1, t-1)$. We write $\begin{pmatrix} \mathbf{w}(e^l+1, t) \\ \vdots \\ \mathbf{w}(a^h - a_0, t) \end{pmatrix} =$ $\mathbf{B}(t-1)\mathbf{x}_{t-1}$, 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 $\mathbf{vr}(t-1)'$ 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

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

with 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 $r(a, e, t-1)$. 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 following, we will show that the initial distribution no longer matters in the long-run, 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.

Facts:

(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 (i) attrition is less than unity prior to the maximum age, (ii) any age-experience cell eventually leads to retirement, and (iii) 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, 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). The Career Decisions of Young Men. Journal of Political Economy, vol. 105, no. 3, 473-522.

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 Benefits from 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, 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.