NATIONAL
CENTER *for* ANALYSIS *of* LONGITUDINAL DATA *in* EDUCATION RESEARCH

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

A program of research by the American Institutes for Research with Duke University, Northwestern University, Stanford University, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

How Teachers Respond to Pension System Incentives: New Estimates and Policy Applications

Shawn Ni
Michael Podgursky

WORKING PAPER 147 • NOVEMBER 2015

Acknowledgements

This research was made possible in part by generous support from the Laura and John Arnold Foundation, the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), and an anonymous foundation. Research supported by CALDER is funded through Grant R305C120008 to the American Institutes for Research from the Institute of Education Sciences, U.S. Department of Education.

The authors wish to acknowledge research assistance from Mark Ehlert, Eric Parsons, and Chao Wu, and research support from the Center for Analysis of Longitudinal Data in Education Research (CALDER) at the American Institutes for Research, and the John and Laura Arnold Foundation. They also thank the Missouri Department of Elementary and Secondary Education for allowing use of their data for research.

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. The views expressed here are those of the authors and should not be attributed to their institutions, data providers, or the funders. Any and all errors are attributable to the authors.

CALDER • American Institutes for Research
1000 Thomas Jefferson Street N.W., Washington, D.C. 20007
202-403-5796 • www.caldercenter.org

How Teachers Respond to Pension System Incentives: New Estimates and Policy Applications

Shawn Ni & Michael Podgursky

CALDER Working Paper No. 147
November 2015

Abstract

Rising costs of public employee pension plans are a source of fiscal stress in many cities and states and have led to calls for reform. To assess the economic consequences of plan changes it is important to have reliable statistical models of employee retirement behavior. The authors estimate a structural model of teacher retirement using administrative panel data. A Stock-Wise option value model provides a good fit to the data and predicts well out-of-sample on the effects of pension enhancements during the 1990s. The structural model is used to simulate the effect of alternatives to the current defined benefit plan.

Keywords: teacher pensions, school staffing, school finance.

JEL codes: I21, J26, J38.

I. Introduction

Pension costs are a major source of fiscal stress for many states and local governments, including school districts, which account for nearly half of state and local workers. Most plans enrolling teachers report large unfunded liabilities.[2] In March 2004, employer costs for teacher pensions averaged 11.9% of salaries. By January 2014 these costs had risen to 18.7%. By contrast, private sector retirement costs for professionals and administrators over the same period have remained relatively stable at about 11% of salaries.[3] Reform of teacher pensions has been widely discussed in legislatures and in the education policy community. Changes have been made (usually for new hires) in several states (National Center on Teacher Quality, 2012). However, reliable estimates of the fiscal and staffing effects of such changes require, in turn, reliable behavioral models of retirement, which is the focus of this study.

A large literature in labor economics has analyzed the effect of incentives in pension systems on the timing of retirement decisions, labor turnover, and staffing (e.g., Friedberg and Webb, 2005; Asch et al. 2005; Ippolito, 1997; Stock and Wise, 1990; Gustman and Steinmeier, 1986, 2005). However, none of this literature pertains to teachers. While there have been many studies of the effect of current compensation on teacher turnover and mobility (e.g., Murnane and Olsen, 1990; Stinebrickner, 2001; Hanushek et al. 2004; Podgursky et al. 2004), the literature on teacher pensions and their labor market effects is slender (Furgeson, et al., 2006; Brown, 2009; Costrell and McGee, 2010; Friedberg and Turner, 2010). The issue of teachers and pensions takes on particular importance since teacher quality has been shown to have a major effect on student achievement (Rivkin et al. 2005 and Chetty et al. 2011).

[2]National Center on Teacher Quality (2012). Novy-Marx and Rauh (2011) argue that the true liabilities of these plans are much larger than the reported actuarial values.

[3]Costrell and Podgursky (2009), updated at
http://www.uark.edu/ua/der/People/Costrell/Employer − Contributions − Update.pdf

To date, none of the papers examining teachers estimate structural models that are standard in the empirical retirement literature (e.g., Stock and Wise, 1990; Berkovec and Stern, 1991). Given concerns about the fiscal state of the pension funds and staffing schools with qualified teachers, a study of the effect of teacher pension plan incentives on teacher retirement behavior has obvious policy relevance. This is a large market, with roughly 3.2 million public school teachers. In addition, other professional staff (e.g., counselors and administrators) are in the same systems, yielding a total closer to 3.7 million. While the rules of defined benefit (DB) pension systems vary from state to state, the general structure of these systems are similar, as are the teachers themselves. Thus we believe that the results of a single state study like this one would generalize to a much larger universe.[4]

However, an analysis of teacher retirement has more general research interest. The administrative data about the teachers and their pension plans in state data systems are of high quality and an excellent resource for research on the behavioral effects of plans incentives. The rules of the teacher pension plans are also readily available to outside researchers. These pension rules subject teachers to large, sharp, and exogenous incentives that allow researchers to study behavioral responses. Moreover, these rules have changed over time in ways that are readily documented.[5]

State administrative data files provide reliable data on teacher employment histories, salaries, and the exact timing of retirement. These administrative panel data are of high quality compared to the household survey data that have been used in some other studies.[6] In

[4]Teachers in 23 states participate in consolidated state retirement plans with other state and local workers. The remaining states are like Missouri, where educators have their own plan. See National Council on Teacher Quality, 2012, Figure 4.

[5]For example, "rule of 80" permits regular retirement when age+experience $\geq$ 80. While one might expect experience and age to have independent effects on retirement, there is no reason to expect an effect of the sum of the two passing a threshold of 80 to affect retirement, independent of pension rules. There are other such rules which produce sharp discontinuities in pension wealth accrual. See Costrell and Podgursky (2009b) for further discussion.

[6]There are tradeoffs. These administrative data are rich in information about the teachers, their employers, and their work histories. Unfortunately, our data file has no information about the teacher's household.

modeling retirement in other markets, a worker's information on future wages or salaries may substantially differ from that known to the researcher. In contrast, the salaries of teachers are determined by schedules that are highly predictable. Thus, teachers and teacher data potentially offer a good laboratory for testing decision models commonly used in retirement research.

In this paper, we show that structural models of teacher retirement fit the data well and are a useful tool for analyzing policy alternatives. The empirical regularities on which reduced-form models rely are the outcomes of pension plan incentives. If those incentives change in fundamental ways – which they invariably do when major plan redesigns occur – the empirical regularities change, possibly in complicated ways. Identification of "deep parameters" provide a basis for researchers and policy-makers to simulate the behavioral effects of changes in these plans. Transitions from final average salary DB plans to a defined contribution (DC) or hybrid plans is a good example. The former plans introduce powerful pull and push incentives to concentrate retirement at certain experience or age combinations associated with "peak value" pension wealth (Lazear, 1983, Costrell and Podgursky 2009a). These incentives shape observed retirement patterns. Reduced-form models fit to these retirement patterns are uninformative about what retirement patterns would look like in a system with smoother pension wealth accrual and no peak value.

In this paper, we estimate a dynamic option value model developed by Stock and Wise (1990). We report parameter estimates and show that the model fits our data very well. We then use the estimated parameters to predict out-of-sample to earlier periods when pension rules were changed (enhanced) and find that the predictions of changes in retirement patterns are quite accurate. Finally, we use the estimated structural parameters to simulate the effect of several DC alternatives.

In particular, we have no information about spousal income, or even whether the teacher is married.

Institutional Background

Missouri public school teachers, like nearly all public school employees, are covered by a DB pension system. In fact, Missouri public school teachers are in three different DB systems. Teachers in the St. Louis and Kansas City districts, less than ten percent of teachers statewide, are covered by Social Security and are in their own pension systems. The rest of the public school teachers in the state are not covered by the Social Security system (as teachers) and are in a state-wide educator plan–the Public School Retirement System (PSRS).[7] Our focus in this paper is on teachers in the PSRS plan.

Under the current rules, Missouri teachers become eligible for a full pension if they meet one of three conditions: a) they are sixty years of age with at least five years of teaching experience, b) thirty years of experience (and any age), or c) the sum of age and years of service equals or exceeds 80 ("rule of 80"). Benefits at retirement are determined by the following formula (some variant of which is nearly universal in teacher DB systems):

$$Annual\ Benefit = S \times FAS \times R \tag{1}$$

where S is service years (essentially years of experience in the system), FAS is final average salary calculated as the average of the highest three years of salary, and R is the replacement factor. Teachers earn 2.5% for each year of teaching service up to 30 years. Thus, a teacher with 30 years experience and a final average salary of $\$60,000$ would receive $30 \times \$60,000 \times 0.025 = \$45,000$. There are several other minor adjustments to the formula in (1). In order to provide teachers with assistance in purchasing health insurance, the district contribution

[7] Missouri teachers are not unique in this regard. Public school teachers in a number of large states are entirely or mostly outside of Social Security (e.g., California, Texas, Illinois, Ohio). The BLS reports that 72 percent of public school are covered by Social Security. State and local employees were not covered by the 1935 Social Security Act. Amendments in the early 1950's permitted these employees to enter the system. Some groups of teachers (as a group) chose to enter, whereas others did not. The result is a complicated mosaic. Usually, all teachers in a state are in or out (e.g., California out, Florida in, see Costrell and Podgursky (2009b)). The Stock-Wise model used in this paper can be adapted to incorporate Social Security.

to individual teacher health insurance is included in FAS. Thus, if the average of the highest three salary years was $\$60,000$ and the average contribution to health insurance was $\$3,000$ annually, then FAS would equal $\$63,000$. Second, there is a "25 and out" option that permits retirement at a reduced rate if teachers have 25 or more years of experience. Finally, the value of R used in formula (1) 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%.

The rules of the pension system changed numerous times between 1992 and 2001. These rule changes made the system more generous for teachers and are widely acknowledged to have passed in response to the booming stock market returns earned by the fund during the 1990's. The more uneven stock market performance since 2001 has tempered enthusiasm by the legislature for further generosity and there have been no further enhancements or significant changes since then.

We will be estimating our structural model under the post-2002 rules. However, since we will be evaluating the predictive power of our model under prior rules, we briefly review rule changes prior to 2002. Table 1 chronicles a number of significant rule changes over this period. At the beginning of the period, 1991-92, regular retirement occurred at 30 years, the replacement rate (R) in equation (1) was 2.1%, final average salary was computed as the average of the five highest years of earnings, and cost of living allowance (COLA) increases were capped at 65% of the initial retirement annuity. Over the next decade all of these rules were liberalized. The most important change for regular retirement was the introduction of the "rule of 80" in 2000. The replacement rate rose to 2.5% by 1998 and 2.55% for years above 30 in 2001. District contributions toward teacher health insurance were added to the calculation of FAS in 1996. Another remunerative enhancement occurred in 1999, when calculation of final average salary was changed from the highest five years to the highest

three years. Finally, the COLA cap increased from 65% to 80% in steps over the period.

II. Modeling the Retirement Decision

Our focus is on the timing of retirement. We assume that an experienced educator who is teaching in the current year has two choices: teach next year or retire.[8] Applying the Stock-Wise (SW) model to teacher retirement, we first write the teacher's expected utility in period t as a function of expected retirement in year m (with $m = t, \cdots, T$ and $T = 101$ is an upper bound on age). In period t, the expected utility of retiring in period m is the discounted sum of pre- and post-retirement expected utility

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

where $0 < k_s < 1$ captures the disutility of working, Y is real salary, c is the teacher's contribution rate to the pension, and B is the real pension benefit. 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 $N(0, \sigma^2)$. This specification assumes that the disutility of work, k_s, does not depend on age. This is a problematic assumption that is at variance with our data. Following Stock and Wise, we relax this assumption by allowing k_s to change monotonically with age: $k_s = \kappa(\frac{60}{age})^{\kappa_1}$. The retirement decision in year t can thus be formulated as choosing $m = t, \cdots, T$ that maximizes $\mathbb{E}_t V_t(m)$.

The retirement decision is irreversible. Once a teacher retires, she cannot return to the same job.[9] Because the future is uncertain and the teacher is risk averse, there is a value

[8] In this context "retire" can also mean stop teaching and collect a pension at a future date rather than immediately.

[9] Thus, we are ruling out the option of a teacher retiring and returning to a PSRS-covered job ("double-dipping"). PSRS rules make it very difficult to return to full time covered employment and collect a pension, although part-time teaching employment (less than half time) is an option.

associated with keeping the retirement option open, hence this is termed an "option value" model.[10]

With a fixed salary schedule there are two sources of uncertainty: the uncertainty of survival and uncertainty in preference shocks. The latter would include, for example, changes in own or spouse's health. To make survival uncertainty explicit, 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:

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

where

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

is the difference in expected utility between retiring in year $m > t$ and retiring now (in year t). Because the teacher's future salary and pension benefits are predictable, in the empirical analysis we replace the expected salary and benefit in $g_t(m)$ with a forecast based on historical data. In the last term in (4), $K_t(m) = \sum_{s=t}^{m-1} \pi(s|t)(\beta\rho)^{s-t}$ depends on unknown parameters and the AR(1) error term ν_t given in (3). Let $m_t^\dagger = argmax\ g_t(m)/K_t(m)$, then the probability that the teacher retires in period t ($G_t(m) \leq 0$ for all $m > t$) is $Prob(\frac{g_t(m^\dagger)}{K_t(m^\dagger)} \leq -\nu_t)$. The details of the model and the MLE estimation methods are reported in the appendix.

[10]The option value model above assumes that a teacher chooses the year of retirement that maximizes the expected present value of the utility of the salary and benefit flows given current information. In a dynamic programming setting, a teacher evaluates the expectation of the value of salary and benefit flow under present and future optimal choices. Hence the option value model does not take into account the value of options in the future. The gain from this is a simpler derivation of the empirical model. Stern (1997) shows that the option value model may yield different results from those obtained by dynamic programming. Lumsdaine et al. (1992) argue that it is not obvious that the more sophisticated dynamic programming model is more realistic for modeling actual retirement decisions. They find that the predictive performance of the option value model is comparable to that of a dynamic programming approach. As we will see below, the SW option value model fits our data very well.

A "peak value" approach has been used in some applied retirement studies (e.g., Coile and Gruber, 2007; Friedberg and Webb, 2005). It can be treated as a special case of the SW model in which the teacher chooses the timing of retirement to maximize the present value of her expected pension wealth. "Peak value" behavior implies the following restrictions: $\kappa = 0, \gamma = 1, \sigma = \rho = 0$. Setting the discount rate β to be the inverse of one plus the real interest rate, the peak value model corresponds to the SW model where the objective becomes finding the peak year m that maximizes pension wealth $\mathbb{E}_t \sum_{s=m}^{T} \beta^{s-m} B_s$, where the expectation is with respect to survival probability.

Data

The data used for estimation of the option value model consists of a cohort of 16,792 Missouri teachers aged 47-58 at the beginning of the 2002-03 school year. We tracked this cohort of teachers forward to the 2008-09 school year. Descriptive statistics on this sample are found in Table 2. In the base year 2002 eighty percent of teachers in the sample are female and had an average of 19.8 years of teaching experience. Over the six year panel, roughly half of the teachers in the cohort retired.[11]

Estimates

Table 3 reports maximum-likelihood estimates of the structural parameters in the retirement model: $\kappa, \kappa_1, \beta, \gamma, \sigma, \rho$. We begin with the pooled estimates in the first column. All of the parameter estimates are statistically significant and of reasonable magnitude. The parameter β reflects the rate of time preference for the teacher, the β estimate of 0.965 implies a 3.5% annual discount rate. The parameter k measures the value of work versus leisure (retirement) time. Recall that the disutility of working is modeled as $k_s = \kappa(\frac{60}{age})^{\kappa_1}$. If $k_s = 1$ then there is no disutility associated with teaching. Our estimates are $\kappa = 0.640$

[11]In an earlier version of the paper, we used a sub-sample of teachers of age 50-55 years old and found similar estimates.

and $\kappa_1 = 0.976$, which imply that the disutility of teaching rises with age. At age 55, one dollar of salary yields the same utility as 70 cents in the retirement benefit. By age 65, this drops to 59 cents. We find that allowing for age-dependency in the disutility of teaching substantially improves the fit of the model.

The point estimate of γ is significantly less than unity, indicating risk aversion. The large value of σ indicates a good deal of heterogeneity in preferences. This is not surprising since there are no covariates in the model. One might expect various household and personal factors such a spouse's pension, health, and preferences for teaching to affect the timing of retirement. These and other factors are picked up in σ. In addition, these omitted factors are not transient but tend to persist over time, as indicated by large and significant values for ρ.[12]

Table 3 also reports estimates for males and females separately. The point estimates are fairly similar, with the exception of κ_1. In both cases the data support the model with age-dependent disutility of working. The preference parameter κ_1 of male teachers is 1.513 while that for female teachers is 1.109. This suggests that as male teachers age, their disutility for teaching relative to retirement rises more quickly than for female teachers. This may reflect different non-teaching opportunities. It may also reflect different mortality rates. The mortality rate of males in the general population is higher (0.748% at age 55) than that of females (0.434% at the same age). Since the DB rules are unisex, this predicts earlier

[12]The parameter estimates are comparable to those reported by Stock and Wise (1990) on a sample of older salesmen of an unidentified firm. Their estimates vary with model specifications, with γ being in the range of 0.7 to 0.8, and β in the range of 0.7 to 0.9 (which implies salesmen are much less patient than teachers), and one dollar of working generates the same utility as 60 cents of pension benefits. They found the unobserved heterogeneity is persistent, with ρ being about 0.7. Along with the stratification by gender in Table 3 we have estimated the model on other subsamples, expecting a further drop in σ. While the basic model fits subsamples very well, some parameters move about. Interestingly, the σ on subgroups generally does not decline by an appreciable amount. Our interpretation of this finding is that σ mostly represents variation in individual unobservables such as health, spouse's circumstances, and preferences for teaching rather than observables like race or school factors on which we stratify. Hence there is no reason for σ to fall when the sample is stratified on observables.

retirement for males.

As noted above, a number of articles in the literature have estimated peak value models rather than a full structural model. We also consider the peak value constrained version of the model: $k = 0, \gamma = 1, \sigma = \rho = 0, \beta = 0.96$. The constraints are easily rejected under a likelihood-ratio test.[13] In practice, while retirements do concentrate in the neighborhood of the peak value of pension wealth, the majority do not. Many concentrate at other pension rule kinks (e.g., 25 and out). Others continue to teach beyond peak value.

Goodness of Fit: In- and Out-of-Sample

The in-sample goodness of fit is quite good. Figures 1 and 2 plot the actual and forecast distribution by experience and age for the teachers who retired or continued employment to the end of the period. Visual examination shows that the model provides an excellent fit to the profiles of retiring and non-retiring teachers for each year and for those who remained employed at the end of the panel. The χ^2 tests on the equality of the observed and predicted distributions by age or experiences in Figures 1 and 2 easily accept the null. Figure 3 shows that the model nicely mimics the joint distribution of age and experience for retirees and non-retirees as well, in particular the "rule of 80" ridge (i.e., age + experience = 80).

These plots use the pooled-sample estimates in Table 3. Using these estimates, we also examined the profile fit for various subsamples (e.g., men versus women in high and low poverty schools) and the fit remained quite good, which suggests that the parameters estimated for the entire sample work well for subgroups. Since the pooled estimates perform well within sample and for subgroups, we use these for the out-of-sample analysis and the policy simulations below.

As noted in the introduction, a structural model is useful in analyzing the effect of

[13] A χ^2 test on the likelihood ratio of the constrained peak value SW model versus the unconstrained SW model overwhelmingly rejects the constraints.

major changes in retirement plans. The current patterns of retirement reflect strong, but rather arbitrary, incentives built into plan rules. For example, the "rule of 80" provision creates a ridge of increased retirement probability along the age+experience = 80 line if one plots retirement rates against age and experience. A similar spike in retirements occurs at 25 years experience. There is no obvious efficiency rationale for these kinks in the intertemporal budget constraint and it is likely that a more rational retirement plan would eliminate them in favor smoother life-cycle benefit accrual. Thus it is important to have a model that can yield accurate behavioral predictions in the absence of such kinks and discontinuities, or when these kinks are moved around in the age -experience space. Unfortunately, we cannot test the former but we can test the latter. That is, we can test the forecasting ability of the model against a very different set of plan design incentives during the 1990's.

Table 1 reports the enhancements to the pension plan during the 1990's. Koedel, et al. (2014) document the pension wealth gains generated by these enhancements. We use the estimated parameters from the pooled sample in Table 3 to forecast the annual retirements of teachers aged 50-62 after each of the enhancements between 1995 to 1999. This provides a robust test of the predictive validity of the model because it is "out of sample" in two respects. First, this is a different sample of teachers. Second, it is a very different set of plan parameters.[14] Figures 4a and 4b plot the actual and predicted distribution of retiring teachers by age and experience under the different, and less generous, DB plan rules during the 1990's. The structural estimates on the 2002-08 sample provide an excellent fit to the age and experience distribution of the retiring teachers. Figures 4c and 4d plot the observed frequencies and predicted retirement probabilities of teachers given the age or experience in 1995. Figures 5a-5h plot the age and experience profiles for 1996-1999 retirees, which reflect

[14]Because the measurement of unobserved heterogeneity σ depends on salary and the salary during 1995-1999 is lower than that during 2002-2008, one would expect a lower value of σ as well. Instead of using the estimated value of 3660 based on the 2002-2008 sample, we adjusted σ as $3660(\frac{salary\ year\ t}{salary\ 2002})^{\gamma}$ for simulation of year t between 1995-1999 as per the first term in equation (2) in the text.

piecemeal introduction of the enhancements. The model fits the experience and age profiles very well, similar to those in Figures 4a and 4b. The χ^2 tests of equality of observed and predicted distributions easily accept the null for Figures 5 and Figures 4a and 4b. Figures 4c and 4d show that predicted probabilities of retirement track observed rates, but the model tends to over-predict retirement.

The in- and out-of-sample predictions of retirement probability are compared with the observed rates in Table 4. For the in-sample prediction of 2002-2008 data, the first row of Table 4 shows that the model predicted a 45.0 percent retirement rate over the sample period. The actual rate was 45.3 percent. The good overall match masks a slight over-prediction of retirement in the earlier part of the sample (year 2003 and 2004).[15] The remaining rows of Table 4 show that the out-of-sample predictions of retirement in the 1990's are higher than the observed ones. Besides the general problem of model mis-specification (e.g., in the parametric form of the utility function), there are three potential specification issues that may explain this over-prediction.

First, the patterns of retirement suggest that there may be multiple types of teachers who differ in preference for teaching. About ten percent of teachers continue teaching even when pension wealth declines in experience. In Figure 4d, the 1995 data show that teachers with

[15]The over-prediction of retirements in 2003 and 2004 is likely an artifact of our sampling scheme (i.e., teachers aged 47-58 and employed 2002). Our base-year cohort of employed teachers *includes* teachers who were eligible for retirement but who chose to wait, but obviously *excludes* those who chose to retire. Thus, it is not surprising that our model slightly over-predicts retirement in 2003 and 2004, but the fit improves as most of these oversampled base-year stayers leave the sample over time. Evidence for this interpretation is found in the fact that the base year (2002) value of A+E is 80.2 years for 2004 retirees (who chose to retire in 2003). Thus a large number of these teachers could have retired in 2002 but did not (the 2003 retirees have an average A+E of 81.6, according to Table 2). However, average A+E in 2002 falls to 75.0 and 73.9 years for 2007 and 2008 retirees, and drops to 66.4 years for the teachers who were still working and not retired by 2008. There no simple solution to this sample censoring problem in our panel since starting with a younger base-year sample (e.g., 40-45 in 2002) means that the vast majority of the teachers would still have been employed 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.

31 or 32 years of experience are less likely to retire than those with 30 years of experience, but the model predicts that the probability of retiring should increase with experience after 30 years. It appears that the preferences of a small fraction of teachers who stay after passing the "peak" of pension wealth differ from the rest of the population. Without taking into account of the presence of these "stayers", the model over-predicts retirement of experienced teachers. This bias is likely to be empirically small though, because teachers who continue teaching past the "peak" are relatively few in number. This is fortunate since estimation of a SW model with heterogenous preferences is computationally very challenging.[16]

Another possible source of bias is sample selection induced by the pension rules. The sample excludes some teachers who prefer earlier retirement. For example, a teacher with 25 years of experience in 1994 and pre-disposed toward early retirement may have already separated from teaching in 1994 and is not included in the 1995 sample. Hence the sample of teachers with more than 25 years of experience contains few who are likely to retire early. Consequently the model over-predicts retirement of teachers with more than 25 years of experience. After 1995, the "25 and out" rule reduces the cost of retiring before the peak of pension wealth, and the theory predicts higher retirement of teachers with 25 to 30 years of experience. But the observed retirement in this range in 1996-1999 is considerably less than the prediction, perhaps because the "early leavers" are already gone. From 1996 to 1999, gradual enhancements made pension benefits higher for teachers with 25-30 years experience (the pension wealth accrual peak becomes a plateau). After each enhancement the new rules predict more retirement in the 25-30 experience range for the same sample of teachers. But

[16]Suppose there are n teachers whose time preference may be one of the two parameters κ_1 or κ_2. To estimate κ_1 and κ_2 using the data on teacher i, y_i, we need to entertain the probability that the likelihood of y_i is either $l(y_i|\kappa_1)$ or $l(y_i|\kappa_2)$. Because there are 2^n possible combinations of these assignments, the likelihood of the whole teacher sample is quite difficult to evaluate. We could introduce a model with two types of teachers (e.g., relative to Type 1, Type 2 teachers are less adverse to teaching, hence with a smaller κ_1.) We can estimate the probability of a teacher being Type 1 or 2, along with the parameters associated with each type. The MLE of such a model maybe obtained via an EM algorithm. However, the computational cost is prohibitively high given the model and the sample size. So we will have to leave this to future research.

the teachers who favor earlier retirement left before the enhancements. Those who remained in the teaching force were those who chose not to retire in the previous years. Hence the sample selection creates more bias in 1996-99 than in 1995. The presence of sample selection bias implies that the fit in the model of a panel over multiple years is on average better than the prediction over the first year of the data. This is exactly what Table 4 shows.

A third potential source of bias is teachers expectation of future pension enhancements. The out-of-sample model predictions are made under the assumption that the teachers expect the current rules are unchanged in the future. However, teachers expecting enhancements in the near future may postpone retirement. It is difficult to model how teachers form expectations on future rule changes, but it is possible that the frequent enhancements experienced in the 1990's may have created the expectation of more enhancements in the future. If that is the case, then the model would over-predict near-term retirement. In addition, the expectation on pension rules may play an important role because the retirement decision is likely planned ahead of time. The out-of-sample simulations are made under the assumption that a teacher makes the retirement decision instantaneously after an pension enhancement. Without allowing for the time of retirement planning, the model may over-predict retirement of the following year.

The biases induced by sample selection and expectation on future rules are absent in the policy simulations in the next section, where we assume a fixed policy is in place for a long period of time. For prediction of a long horizon (say 20 years), the bias in initial sample selection should have a much smaller influence on the model prediction, and the expectation of future rule changes is absent by assumption.

III. Simulating Pension Plan Alternatives

In this section we use the structural estimates to explore the behavioral effects of pension

plan changes. Given the lively policy debate in this area, there are many options one might explore. Some states are considering a switch to DC plans in total, or partially in "hybrid" plans, to reduce fiscal exposure as well as eliminate incentives for early retirement. A structural model like SW is well-suited for exploring alternatives to DB plans. Indeed, Stock and Wise use their model estimates to simulate the effect of a conversion to a DC plan. We will consider several variants of a DC conversion and compare them to the current DB plan. Before laying out those alternatives we first show how a DC-type plan can be introduced into the option-value model.[17]

Conversion to DC

We consider the following hypothetical DC plan: teachers contribute a mandatory fixed percent (c) of salary. This is matched by an equivalent annual employer contribution into each teacher's account. A teacher's account accumulates with annual contributions and nominal investment returns of r on the fund balance. We treat this as a guaranteed return (e.g., as with TIAA or a "cash balance" pension plan).[18] The account is portable and teachers can withdraw from the account at any age without penalty. When a teacher retires, the contribution to the account stops and an insurance company provides an actuarially fair annuity B (in real dollars) equal to the cash value in the teacher's account. Assume that a teacher aged a holds a DC account worth W_t in year t, which generates an expected nominal flow of an annuity B_{t+n} in the nth year of retirement up to a maximum life T, $(t+n \leq T)$. The annual inflation rate is i. The retiree survives to $t+n$ with conditional probability $\pi(t+n|t)$. The expected account value and the expected payment evolve as:

$$W_{t+n} = W_{t+n-1}(1+r) - B_{t+n}, \quad B_{t+n} = \pi(t+n|t)(1+i)^n B.$$

[17]Researchers have used peak value models estimated on DB plan participants to simulate DC conversions (e.g., Friedberg and Webb, 2005; Costrell and McGee, 2010). A problem with this approach is that DC plans never reach a peak value so the simulation of DC alternatives is necessarily ad hoc.

[18]This is a somewhat stylized DC plan, since we abstract from any risk associated with the investments made by the teacher and assume a guaranteed rate of return.

We set $W_T = 0$, (as is in the DB plan case, $T = 101$.) It follows that:

$$B = \frac{W_t}{\sum_{n=1}^{T-a} \pi(t+n|t)(\frac{1+i}{1+r})^n.} \tag{5}$$

In the policy scenarios below we will be considering the effect of a conversion from the current DB to a DC plan. Thus, we need to determine the DC account value for a teacher who is in the current DB plan. We consider the following scenario. All teachers in the DB plan in 2002 have a cash balance W (or a fixed fraction thereof) based on the current rules of the DB plan. Further accrual of pension wealth under the old plan is frozen. Going forward the value in this account grows by the nominal interest rate (on the fund balance) and further annual contributions from teachers and districts.

With this initial value in the DC plan, the teacher considers whether to retire or continue to work as in the SW model: a teacher's expected utility in period t is a function of expected retirement in year m (with $m = t, \cdots, T$.) In period t, the expected utility of retiring in period m is the discounted sum of pre- and post retirement expected utility of (2).

For a teacher retiring at year m, the benefit B_s is set at B given in (5) with W_t replaced by the real value $\frac{W_m}{(1+i)^{m-t}}$. Note that the nominal account value in year $m > t$ is the value of accumulated contributions plus the compound return of the wealth in period t: $W_m = W_t(1+r)^{m-t} + \sum_{k=t+1}^{m} 2cY_k(1+r)^{m-k}$.

Because the DC rules are simpler than the DB rules, we are able to formalize the marginal condition for retirement under the DC rules and thereby gain some intuition about the tradeoff between teaching and retirement. Suppose in the absence of unobserved preference shifters the teacher with salary Y_t and pension wealth W_t is indifferent between retiring in year $t+1$ (with a constant real pension flow of B starting in year $t+1$) or t (with pension flow $\tilde{B}$ starting in year t.) Then

$$(k_t(1-c)Y_t)^\gamma + \sum_{s=t+1}^{T} \beta^{s-t}\pi(s|t)B^\gamma = \tilde{B}^\gamma + \sum_{s=t+1}^{T} \beta^{s-t}\pi(s|t)\tilde{B}^\gamma, \tag{6}$$

where $B = \frac{W_t(1+r)+2cY_t}{1+\sum_{s=t+1}^{T}\pi(s|t)(\frac{1+i}{1+r})^{s-t}}$ and $\tilde{B} = \frac{W_t}{\sum_{s=t+1}^{T}\pi(s|t)(\frac{1+i}{1+r})^{s-t}}$. Denote the constants $b_1 = \sum_{s=t+1}^{T}\pi(s|t)(\frac{1+i}{1+r})^{s-t}$, and $b_2 = \sum_{s=t}^{T}\beta^{s-t}\pi(s|t)$, then condition (6) can be written as

$$[b_1 k_t(1-c)\frac{Y_t}{W_t}]^{\gamma} + b_2(1+r+2c\frac{Y_t}{W_t})^{\gamma} = (1+b_2)(\frac{b_1}{1+b_1})^{\gamma}. \tag{7}$$

(7) implies that for a given age, under the DC plan a teacher chooses to retire when the ratio of salary to pension wealth is lower than a constant. The dynamics of the pension wealth/salary ratio is given by $\frac{W_t}{Y_t} = (1+r)(\frac{W_{t-1}}{Y_t-1})(\frac{Y_{t-1}}{Y_t}) + 2c\frac{Y_{t-1}}{Y_t}$. The pension wealth/salary ratio is increasing in the return to savings and increases over time as real salary growth slows down at the later stage of a teacher's career. At some point the ratio $\frac{W_t}{Y_t}$ is large enough to render the LHS lower than the RHS of (7).

Because the pension annuity B is increasing in initial pension wealth, the level at which pension wealth is set in the year of initial conversion from DB to DC plans affects the retirement decision. For teachers at or near the "peak value" of pension wealth, this can be a very attractive option–the DC plan eliminates the penalty on working after reaching the peak value under the current rules (i.e., the "pushing out" effect of the current rules). However, the DC plan does not necessarily postpone retirement. For some teachers, it is optimal to retire earlier under the DC than under the current rules. Whether this is the case depends on the teacher's age, experience, and the initial 2002 pension wealth lump sum payment. As condition (7) shows, under the DC plan a teacher retires when the salary/pension wealth ratio is below a threshold. The higher the initial pension wealth, the earlier the retirement under the DC plan.

The contrast between retirement incentives under DC and DB plans can be illustrated in the context of the option value model. Under a DB plan, because the pension accrual can change sharply by age and experience, the expected gain from retirement at an optimal retirement year $m^{\dagger}$ over retirement in the current period $\frac{g_t(m^{\dagger})}{K_t(m^{\dagger})}$ in (4) can vary greatly by

the current age and experience. Under a DC plan, the wealth W accumulates smoothly over time, and the timing of retirement only matters marginally. Hence $\frac{g_t(m^\dagger)}{K_t(m^\dagger)}$ does not vary sharply by a small change in age and experience. Given the same distribution in preference shocks, the retirement probability and profile of retiring teachers are both more "smoothed out" under a DC plan than under the DB plan.

Policy Simulations of Behavioral Effects

The teacher's contribution rate c was 10.5% in 1990's and has since increased to 14.5%. We will experiment with different contribution rates in the simulations below. The inflation rate is assumed to be $i = 3\%$. Given the fiscal challenges with public sector pension plans, we consider two policy-relevant funding scenarios. In a "full conversion" scenario, at the time of conversion, the senior teachers in our sample (recall, aged 48-57) teachers get the full actuarial value of their DB pension wealth. We also consider a "haircut" scenario in which these senior teachers lose 10% of their DB pension wealth at the time of conversion.[19] Such a policy may be necessitated financially and may be acceptable to senior teachers who benefit from the DB to DC conversion.

We analyze four specific policies:

Policy A: the current DB rules;

Policy B: $r = 6.5\%$, conversion to a DC plan with the full 2002 pension wealth and contribution rate $c = 14\%$;

Policy C: $r = 4\%$, conversion to a DC plan with the full 2002 pension wealth and contribution rate $c = 10\%$;

Policy D: $r = 4\%$, conversion to a DC plan with a 10% "haircut" in the 2002 pension wealth and contribution rate $c = 14\%$.

[19]This reduction in pension wealth may come about not because of a cut in the initial retirement annuity (B), but rather a cut in future COLA adjustments. COLA adjustments are sometimes seen by courts as having weaker legal protections than the initial annuity set by formula (1). See Munnell (2014).

The estimate of real discount factor $\beta = 0.965$ in Table 3 implies an annual discount rate of 3.5%. With the 3% inflation and the 6.5% nominal return, Policy B takes the same nominal rate of return and contribution rate as the 2002 DB plan. So it is most relevant for comparison with the DB plan. Our calculations show that the DC Policy B renders a substantial welfare gain for the late career teachers over the DB Policy A, by eliminating the penalty from working past the peak year of DB pension wealth. Hence the DC policy gives the teacher the flexility regrading the retirement date and eliminates the "push out" incentive. Her expected utility under the DC policy is higher than that under the DB policy, and remains so even after a moderate haircut at the time of conversion from DB to DC plans.

In examining the goodness of in-sample fit of the model in the previous section, we were constrained to the six year window of our panel data. In simulating the effect of these policies, there is no reason to restrict our time horizon so narrowly, thus we extend the forecast horizon to 20 years, by which time nearly all of these teachers will have retired.

Figure 6 plots the predicted survival rate (the percentage of the 2002 teachers who remain teaching) over the next 20 years under the alternative pension scenarios. Under all the DC changes the teachers are more likely to continue teaching than under the current DB plan. The model predicts that by the year 2020 about 6% the teachers in our 2002 sample would still remain in teaching force, compared to 14% under the DC Policy B, and 18% under DC Policies C and D. The 10% "haircut" in initial pension wealth makes teachers more likely to continue teaching, as noted in the discussion above. Fixing the initial pension wealth while raising the contribution rate from 10% to 14% initially increases the survival rate and eventually decreases it; but these effects are quantitatively small.

Figure 7 plots the predicted experience and age distributions of retiring teachers over the 20 year horizon. The left panel shows that the predicted retirement ages under various

DC plans are much less concentrated than those under the current DB rules. Under the DC plans the percentage of retiring teachers at younger ages is similar to the current DB plan. However far more retiring teachers are over age 60 under the DC plans. The right panel depicts a similar picture on the predicted experience of retiring teachers. Under the DC rules the retirement experience is much more dispersed than under the current DB rules. The predicted percentage of teachers retiring at low experience is similar under the DC or DB rules. But under the DC rules, far fewer teachers would retire with 25-31 years of experience than under the current DB plan.

The left panel of Figure 8 plots the joint distribution of retirement age and experience over the 20-year horizon under the current DB rules and the right panel the joint distribution of age-experience under the DC Policy D (with a 10% "haircut" in the 2002 pension wealth and with contribution rate 14%.) Consistent with the plot of the marginal distributions of age and experience, under the DB plan the joint age-experience distribution is more concentrated than that under the DC rules. In particular, the joint distribution under the current DB plan has a ridge that follows the "rule of 80" line. Along the ridge, the retirement age and experience are negatively related. Under the DC plans, the retirement age and experience are positively related: the teachers retiring at age 60 have more teaching experience than those retiring at age 55.

IV. Conclusion

Policy discussions about teacher quality and teacher "shortages" often focus on recruitment and retention of young teachers. However, attention has begun to focus on the incentive effects of teacher retirement benefit systems, particularly given their rising costs and their large unfunded liabilities. In this paper we estimate a structural model of retirement for teachers and use it to estimate the effect of pension rules on the timing of retirement. The

model fits the data very well, and nicely mimics the sharp spikes associated with certain age and experience combinations. It also does a good job predicting the effect of enhancements enacted during the 1990's. We use the model to simulate the effect of enacting various types of DC alternative plans. A DC (or cash balance) alternative plan would greatly ameliorate the spikes and smooth out retirements.

As states consider reform of teacher pension plans, structural econometric models of retirement behavior can be of great value in estimating the labor market and fiscal consequences of plan changes. The virtue of the approach used in this paper is its simplicity. Longitudinal data files on teachers containing age, experience and salary are routinely constructed by state education agencies and used by researchers studying teacher retention and mobility. The rules of pension systems (and modifications thereof) are readily available. Structural models like the one estimated in this paper can be used to explore revenue-neutral and utility-enhancing plan designs. In the case of retrenchments, it can be used to assess the consequences for school staffing and overall welfare effects. Behavioral econometric models can also enhance the reliability of actuarial studies of the fiscal solvency of these plans – a topic of interest in several large states.

APPENDIX: MLE Estimation of the Option Value Model

The expected gain from retirement at year m over retirement in the current period is

$$\begin{aligned} G_t(m) &= \mathbb{E}_t V_t(m) - \mathbb{E}_t V_t(t) \\ &= \mathbb{E}_t \sum_{s=t}^{m-1} \beta^{s-t}(k_s Y_s)^\gamma + \mathbb{E}_t \sum_{s=m}^{T} \beta^{s-t}(B_s)^\gamma - \mathbb{E}_t \sum_{s=t}^{T} \beta^{s-t}(B_s)^\gamma + \mathbb{E}_t \sum_{s=t}^{m-1} \beta^{s-t}(w_s - \xi_s). \end{aligned}$$

For a teacher alive in year t we denote the probability of survival to period $s > t$ as $\pi(s|t)$. Now

$$\begin{aligned} G_t(m) &= \sum_{s=t}^{m-1} \pi(s|t)\beta^{s-t}\mathbb{E}_t(k_s Y_s)^\gamma + \sum_{s=m}^{T} \pi(s|t)\beta^{s-t}\mathbb{E}_t(B_s)^\gamma - \sum_{s=t}^{T} \pi(s|t)\beta^{s-t}\mathbb{E}_t(B_s)^\gamma \\ &+ \sum_{s=t}^{m-1} \pi(s|t)(\beta\rho)^{s-t}(w_t - \xi_t). \end{aligned}$$

The sum of the first three terms is a function of current salary and experience, and is denoted by $g_t(m)$. The last term $\sum_{s=t}^{m-1} \pi(s|t)(\beta\rho)^{s-t}(w_t - \xi_t)$ is unobservable and is denoted $K_t(m) = \sum_{s=t}^{m-1} \pi(s|t)(\beta\rho)^{s-t}$ (which depends on unknown parameters) times an error term $\nu_t = w_t - \xi_t$, which follows $\nu_t = \rho\nu_{t-1} + \epsilon_t$ where ϵ_t is assumed to be $N(0, \sigma^2)$. Let $m_t^\dagger = argmax\ g_t(m)/K_t(m)$,, the probability that teacher retires in period t ($G_t(m) \leq 0$ for all $m > t$) is $Prob(\frac{g_t(m^\dagger)}{K_t(m^\dagger)} \leq -\nu_t)$.

The likelihood can be specified under the normality assumption on ν_t and given rules for predicting future earnings. We assume salary is predictable under an estimated nonlinear (a third order polynomial) function of experience.[20] For estimation of the model, if a teacher $i \in \{1, .., I\}$ retires in period t, $d_{it} = 1$, otherwise $d_{it} = 0$. After retirement the teacher is dropped out of the sample. For cross-section data with a teacher i observed only in period

[20] Missouri teachers, like nearly all public school teachers, are paid according to salary schedules that set pay based on years of teaching experience and education credentials (frequently terminating in an MA). Thus it is not unrealistic to treat teacher pay as a function of teaching experience, assuming all teachers move from the BA column on the schedule over to the MA column with the passage of time. Because we focus on late-career teachers, the degree-related salary adjustment is largely absent in the sample. The fairly deterministic advancement over well-defined district salary schedules underlies the salary growth assumption in the text.

t, the likelihood is

$$L(\gamma, \kappa, \kappa_1, \beta, \sigma, \rho \mid \mathbf{Y}, \mathbf{B}, \mathbf{D}) \propto \prod_{i=1}^{I} \Phi(\frac{g_t(m_t^\dagger)}{K_t(m_t^\dagger)}/\sigma_\nu)^{d_{it}}(1 - \Phi(\frac{g_t(m_t^\dagger)}{K_t(m_t^\dagger)}/\sigma_\nu))^{1-d_{it}},$$

where $\Phi(.)$ is the cumulative density function of standard normal and σ_ν is the standard deviation of ν_t. For panel data the likelihood is made more complicated by the serial correlation of ν_t. Suppose a teacher is observed for period $t, t+1, .., t+n$ and she retired in $t+n$, then the likelihood is the probability of the joint event $\pi(\frac{g_t(m_t^\dagger)}{K_t(m_t^\dagger)} > -\nu_t, .., \frac{g_{t+n-1}(m_{t+n-1}^\dagger)}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}, \frac{g_{t+n}(m_{t+n}^\dagger)}{K_{t+n}(m_{t+n}^\dagger)} < -\nu_{t+n})$. By the definition of conditional probability, one can view this joint probability as products of a sequence of conditionals:

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

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

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

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

The log likelihood is

$$logL(\gamma, \kappa, \kappa_1, \beta, \sigma, \rho \mid \mathbf{Y}, \mathbf{B}, \mathbf{D}) = \sum_{i=1}^{I} log\pi_i(\boldsymbol{\nu}_{t,t+n} \in A_i) = \sum_{i=1}^{I} log \int_{A_i} \phi(\boldsymbol{\nu}_{t,t+n}) d\boldsymbol{\nu}_{t,t+n}, \quad (8)$$

where for teacher i retiring in period $t+n$, $\boldsymbol{\nu}_{t,t+n} \in A_i$ if $\frac{g_t(m_t^\dagger)}{K_t(m_t^\dagger)} > -\nu_t, .., \frac{g_{t+n-1}(m_{t+n-1}^\dagger)}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}, \frac{g_{t+n}(m_{t+n}^\dagger)}{K_{t+n}(m_{t+n}^\dagger)} < -\nu_{t+n}$, and $\phi(.)$ denotes multivariate normal density distribution of $N(0, \mathbf{\Sigma})$. An obstacle to evaluating the likelihood is the large computational time of n -dimensional integration. Even for a moderate size n (say 5), deterministic methods for numerical integration can be prohibitively costly. In this study, we solve the problem through Monte Carlo simulation. The covariance matrix $\mathbf{\Sigma}$ permits a Cholesky decomposition $\mathbf{\Sigma} = \mathbf{VV}'$ ($\mathbf{V}$ is lower triangular.)

The algorithm for computing $\int_{A_i} \phi(\boldsymbol{\nu}_{t,t+n}) d\boldsymbol{\nu}_{t,t+n}$ via frequency simulation is as follows: (1) Draw $\mathbf{e}^{\{j\}}$ from $N(0, \mathbf{I}_{n+1})$ ($j = 1, \cdots, J$) and let $\boldsymbol{\nu}_{t,t+n}^{\{j\}} = \mathbf{Ve}^{\{j\}}$. (2) Use the frequency $\frac{1}{J}\sum_{j=1}^{J} I(\boldsymbol{\nu}_{t,t+n}^{\{j\}} \in A_i)$ to approximate $\int_{A_i} \phi(\boldsymbol{\nu}_{t,t+n}) d\boldsymbol{\nu}_{t,t+n}$. $I(\boldsymbol{\nu}_{t,t+n}^{\{j\}} \in A_i) = 1$ if $\boldsymbol{\nu}_{t,t+n}^{\{j\}} \in A_i$ and $I(\boldsymbol{\nu}_{t,t+n}^{\{j\}} \in A_i) = 0$ otherwise. This method yields accurate approximation of the likelihood if the number of draws J is large enough. But for a sample of a large number of teachers, the computational cost is high if we use a large number of draws for each teacher.

An alternative approach to the above Monte Carlo frequency simulation for computing likelihood is the Geweke–Hajivassiliou–Keane (GHK) simulator. For a longer data panel the GHK simulator is more efficient than the MC approach for frequency of retire-

ment. For the present problem, we obtain the MLE of the model parameters by using the version of the GHK simulator proposed by Börsch-Supan and Hajivassiliou (1993). The Cholesky decomposition of the covariance matrix $\mathbf{\Sigma}$, relates the conditions on the $n+1-$dimensional vector of correlated errors $\boldsymbol{\nu}_{t,t+n}$ to a condition on $n+1$ iid standard normal errors $\mathbf{e} = (e_t, e_{t+1}, e_{t+2}, .., e_{t+n})' \sim N(\mathbf{0}, \mathbf{I}_{n+1})$. In the context of the present model, the GHK algorithm express the probability of the joint event such as $\frac{g_t(m_t^\dagger)}{K_t(m_t^\dagger)} > -\nu_t, .., \frac{g_{t+n-1}(m_{t+n-1}^\dagger)}{K_{t+n-1}(m_{t+n-1}^\dagger)} > -\nu_{t+n-1}, \frac{g_{t+n}(m_{t+n}^\dagger)}{K_{t+n}(m_{t+n}^\dagger)} < -\nu_{t+n}$, associated with the correlated errors $\boldsymbol{\nu}_{t,t+n}$, to a sequence of conditional events associated with iid standard normal errors $\mathbf{e}$. In doing so, it transforms the problem of simulating the probability of the joint event involving $\boldsymbol{\nu}_{t,t+n}$ to a problem of sequentially simulating the probability of $n+1$ events involving $n+1$ independent random variables e_{t+i} (for $i = 0, 1, .., n$.) In other words, the GHK algorithm transforms the problem of numerically computing a $n+1$-dimensional integration to $n+1$ one-dimensional integrations. The computational cost of n one-dimensional integrations is much less than one n-dimensional integration, especially when n is relatively large. We experimented with both MC simulation of frequency of joint distribution $\boldsymbol{\nu}_{t,t+n}$ and the GHK method. The two methods yield very similar estimates but the GHK method takes about 4 hours to reach convergence to the MLE on a 3.2 GHz PC for the data sample of all teachers, which is about one-fifth of the computation time using the method of frequency simulation.

The MLE is obtained using an IMSL subroutine based on grid search, with upper- and lower bounds on each parameter. For instance, the parameter σ is bounded in $(1000, 10000)$. A reasonably constrained search helps to reduce the computation time. Our experiments show that varying the bounds on the parameters may give rise to different MLE estimates, but does not materially affect the overall fit and predictions of the model.

The MLE estimates are used for the goodness of fit and policy simulations. For the in-sample goodness of fit for all teachers aged 47-58 in 2002 (our baseline sample), we use the

estimated parameters of the structural model and the information on these teachers in 2002 to generate the probability that each teacher took one of the following 7 actions: retired in year 2003, retired in 2004, ..., retired in 2008, and remained in teaching workforce in 2008. The probabilities are obtained through Monte Carlo simulation. Specifically, for each teacher in the 2002 sample, regardless of the actual retirement decision the teacher took, we draw 6 serially correlated error terms ϵ_t ($t = 2002, ..., 2007$). If according to the SW model, with the realized error terms of ϵ_{2002} and given the age, salary, and experience, the teacher chooses to retire in 2002, then for that draw the teacher is recorded as retired in 2003. If the model predicts that teacher chooses not to retire in 2002, then we project the 2003 salary and add one year to the age and experience. If the model predicts retirement given the ϵ_{2003} draw and the new state variables, then the teacher is recorded as retired in 2004. We repeat the process to 2007. If model predicts the teacher chooses not to retire up to 2007, then the teacher is recorded as a non-retiree at the end of the sample. For each teacher we replicate the above experiment a large number of times ($100,000$, changing it to $1,000,000$ produces the same results). The frequency of the simulated retirement decisions give rise to the predicted probabilities. We aggregate the probabilities over the teachers in the 2002 sample to obtain the aggregate predicted retirement. We present aggregated predicted and actual retirement by age, experience, and age by experience. Comparisons of the observed and predicted distributions of the retirees (at the year they decide to retire) and non-retirees (in 2008) are used to gauge the fit of the model. The simulations under the DC policies for the 2002 cohort are similar to the in-sample simulations except we use the DC rules to simulate retirement decisions and extend the forecasting horizon to 20 years. The out-of-sample forecasts of 1995-1999 are based on a similar procedure and with a forecasting horizon of one year.

References

Asch, Beth, Steven J. Haider, and Julie Zissimopoulos. 2005. Financial Incentives and Retirement: Evidence from Federal Civil Service Workers. *Journal of Public Economics* 89 no. 2-3:427-440.

Berkovec, James and Stern, Steven. 1991. Job Exit Behavior of Older Men, *Econometrica* 59 no. 1:189-210.

Brown, Kristine. 2009. The Link Between Pensions and Retirement Timing: Lessons from California Teachers. National Center on Performance Incentives. Vanderbilt University.

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 no. 3:347-368.

Chetty, Raj, John N. Friedman and Jonah E. Rockoff. 2011. The Long-term Impacts of Teachers: Teacher Value-Added and Student Outcome in Adulthood. Working Paper 17699, National Bureau of Economic Research.

Coile, Courtney and Jonathan Gruber. 2007. Future Social Security Entitlements and the Retirement Decision. *Review of Economics and Statistics* 89 no. 2:234-246.

Costrell, Robert, and Joshua McGee. 2010. Teacher Pension Incentives, Retirement Behavior, and Potential for Reform in Arkansas. *Education Finance and Policy*, 5 no. 4:492-518.

Costrell, Robert M. and Michael J. Podgursky. (2009a). "Peaks, Cliffs and Valleys: The Peculiar Incentives in Teacher Retirement Systems and Their Consequences for School Staffing." Education Finance and Policy, 4(2), 175-211.

Costrell, Robert, and Michael Podgursky. (2009b). "Teacher Retirement Benefits." Education Next, 9(2) (Spring), 58-63.

Friedberg, Leora and Sarah Turner. (2010). "Labor Market Effects of Pensions and Implications for Teachers." Education Finance and Policy, 5(4), 463-491.

Friedberg, Leora, and Anthony Webb. (2005). "Retirement and the Evolution of Pension Structure." Journal of Human Resources. 40(2), 281-308.

Furgeson, Joshua, Robert Strauss, and William Vogt. (2006). "The Effects of Defined Benefit Pension Incentives and Working Conditions on Teacher Retirement Decisions." Education Finance and Policy 1(3), 316-48.

Gustman, Alan L. and Thomas L. Steinmeier (1986). "A Structural Retirement Model. Econometrica, 54 (3), 555-584

Gustman, Alan L. and Thomas L. Steinmeier (2005). "The Social Security Early Entitlement Age in a Structural Model of Retirement and Wealth. Econometrica, 54 (3), 555-584

Hanushek, Erik A. Kain, John F. and Rivkin, Steven G. (2004). "Why Public Schools Lose Teachers." Journal of Human Resources, 39(2), 326-354

Ippolito, Richard A. (1997). Pension Plans and Employee Performance: Evidence, Analysis, and Policy. Chicago: University of Chicago Press.

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

Lazear, Edward (1983) "Incentive Effects of Pensions." NBER Working Paper 1126. Cambridge, MA.

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

Munnell, Alicia H. (2014). "States Cut COLAs for Public Pensions." Market Watch, May 22.

Murnane, Richard J., and Olsen, Randall J. (1990). "The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from North Carolina." Journal of Human Resources, 25(Winter), 106-124.

Novy-Marx, Robert, and Joshua Rauh (2011). "Public Pension Promises: How Big Are They and What Are They Worth?" The Journal of Finance, 66, 12111249.

Podgursky, Michael, Ryan Monroe, Donald Watson. (2004). "Teacher Pay, Mobility, and Academic Quality." Economics of Education Review, 23, 507-518.

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

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

Stinebrickner, Todd. (2001). "A dynamic model of teacher labor supply." Journal of Labor Economics, 19(1) (January), 196-230.

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

Table 1: PSRS Pensions Rule Changes

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

Note: The table lists important changes in pension benefit rule of the state-wide educator plan–the Public School Retirement System (PSRS) in Missouri from 1995 to 2002. The "25 and out" rule in 1996 permits retirement at a reduced benefit factor (replacement rate) R in formula (1) if teachers have 25 or more years of experience, with the following benefit factors: 2% for teachers with 25 years of experience, 2.05% for 26 years, 2.1% for 27 years, 2.15% for 28 years and 2.2% for teachers with 29 years of experience. The "25 and out" rule in 1998 raises the benefit factors to 2.2% for 25 years, 2.25% for 26 years, 2.3% for 27 years, 2.35% for 28 years and 2.4% for teachers with 29 years of experience. The "rule of 80" permits regular retirement when age+experience $\geq$ 80.

Table 2: Sample Averages by the Year of Retirement

Sample Year	Number of teachers	Age	Experience	Male
Base year				
All 2002	16792	51.62	19.79	0.20
Retirement year				
2003	979	53.78	27.86	0.28
2004	1271	54.24	27.92	0.24
2005	1473	54.92	27.32	0.23
2006	1353	55.64	27.26	0.21
2007	1317	56.05	26.95	0.20
2008	1213	56.80	26.89	0.19
Not Retired by 2008				
Not Retired	9186	55.73	20.66	0.17

Note: Missouri teachers aged 47-58 in 2002. "All 2002" is the total cohort of 16792 teachers in the base year; and age and experience are the averages in the base year. The rows with retirement year labels 2003-2008 are contemporaneous averages for teachers who retired in that year. The row for 'Not retired' are the contemporaneous averages for teachers who remained employed at the end of the sample period. Male=1 for male teachers.

Table 3: MLE Estimates of Structural Parameters

	Pooled Sample	Female	Male
β	0.965	0.957	0.969
	(0.026)	(0.037)	(0.069)
κ	0.640	0.671	0.674
	(0.013)	(0.028)	(0.025)
κ_1	0.976	1.109	1.513
	(0.060)	(0.036)	(0.228)
γ	0.716	0.663	0.676
	(0.032)	(0.019)	(0.079)
σ	3660.166	2886.944	2603.229
	(69.778)	(109.127)	(157.750)
ρ	0.643	0.520	0.629
	(0.052)	(0.033)	(0.133)
log-likelihood	-21213.733	-16688.576	-4531.550
Number of teachers	16792	13482	3310

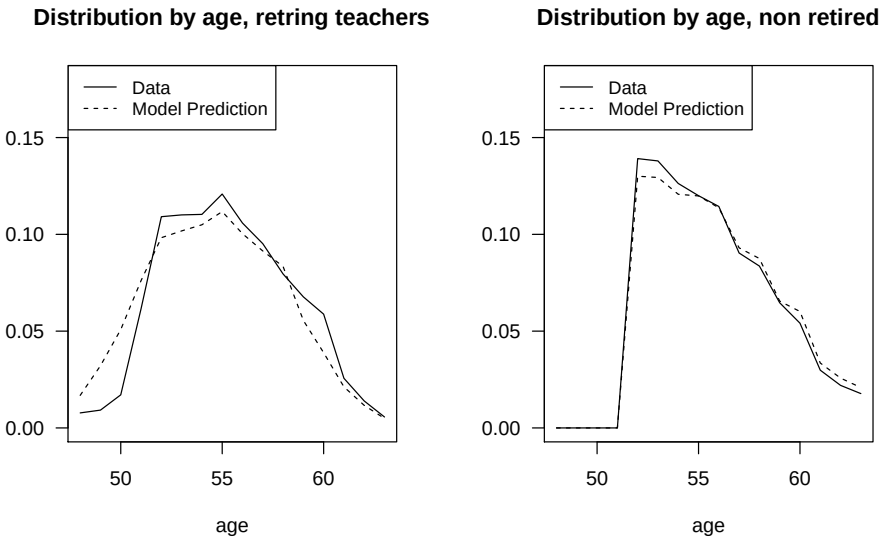
Note: The standard errors are in parentheses. Missouri PSRS teachers aged 47-58 in 2002. The sample period is 2002-2008. The likelihood is evaluated using the "GHK" algorithm described in the appendix.

Table 4: Observed and Predicted Fraction of Retiring Teachers

sample	number of teachers	observed	predicted
In Sample 2002-2008	16,792	0.453	0.450
Out of Sample 1995	9,584	0.078	0.096
Out of Sample 1996	10,125	0.098	0.126
Out of Sample 1997	11,219	0.085	0.123
Out of Sample 1998	12,127	0.090	0.131
Out of Sample 1999	13,059	0.092	0.131

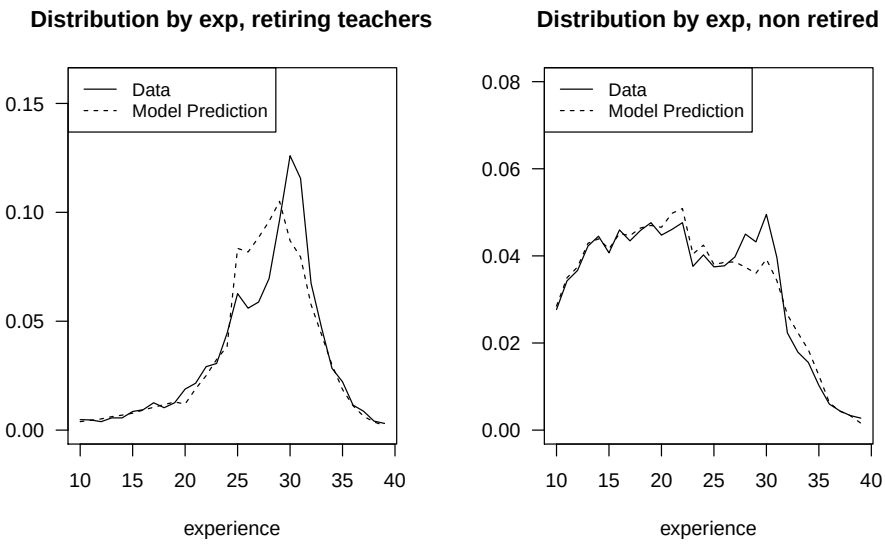
Note: The first column reports the total number of teachers in the beginning of the sample period. The second column of the table reports the percentage of the teachers in the first column retired by the end of the sample period, the third column reports the average of the simulated probability of these teachers' retirement. The simulation is based on the Monte Carlo study described in the last paragraph of the appendix. The out-of-sample teachers are 50-62 years old in the beginning in each of sample year from 1995 to 1999.

Figure 1: Observed and Predicted Age Distributions of Retired and Non-retired Teachers



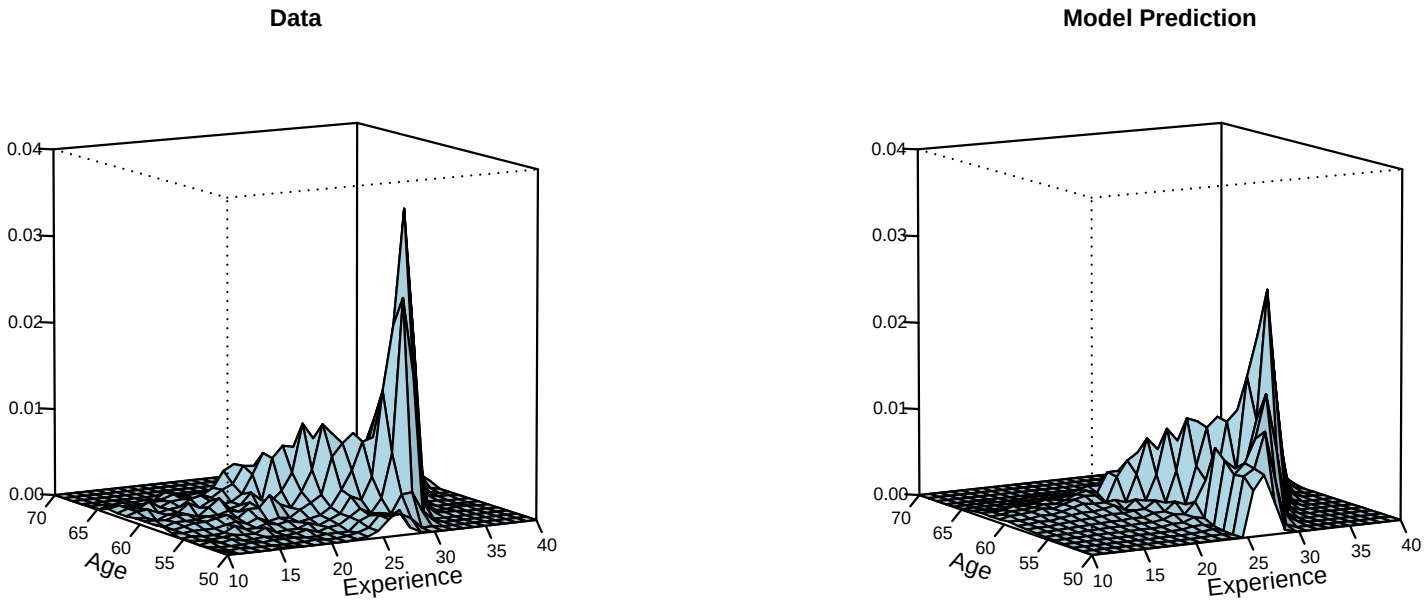
Note: The observed age pertains to all teachers at the year of retirement (for the left panel) or the non-retired at the end of the sample period (for the right panel). The model prediction is the in-sample prediction based on the estimates in first column of Table 3.

Figure 2: Observed and Predicted Experience Distributions of Retired and Non-retired Teachers



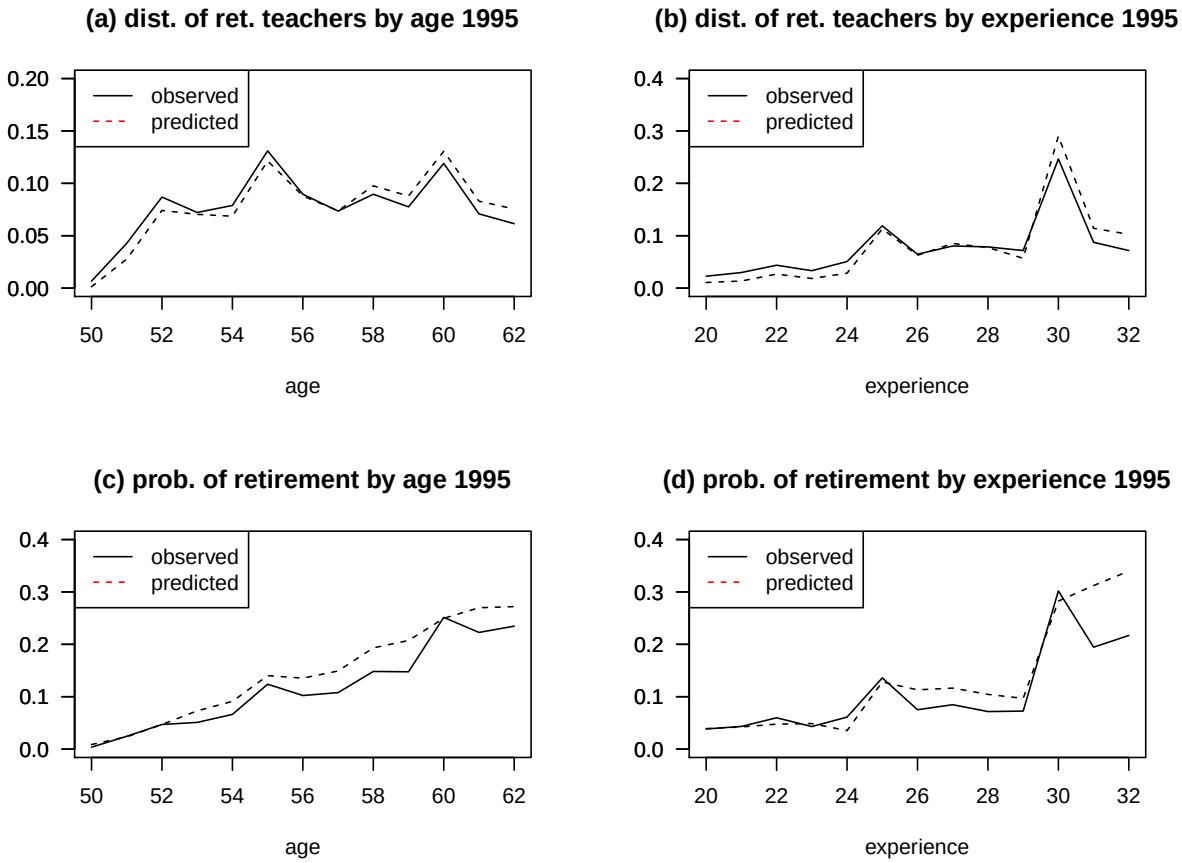
Note: The observed experience pertains to all teachers at the year of retirement (for the left panel) or the non-retired at the end of the sample period (for the right panel). The model prediction is the in-sample prediction based on the estimates in first column of Table 3.

Figure 3: Observed and Predicted Joint Retirement Age-Experience Distribution for Teachers at the Time of Retirement



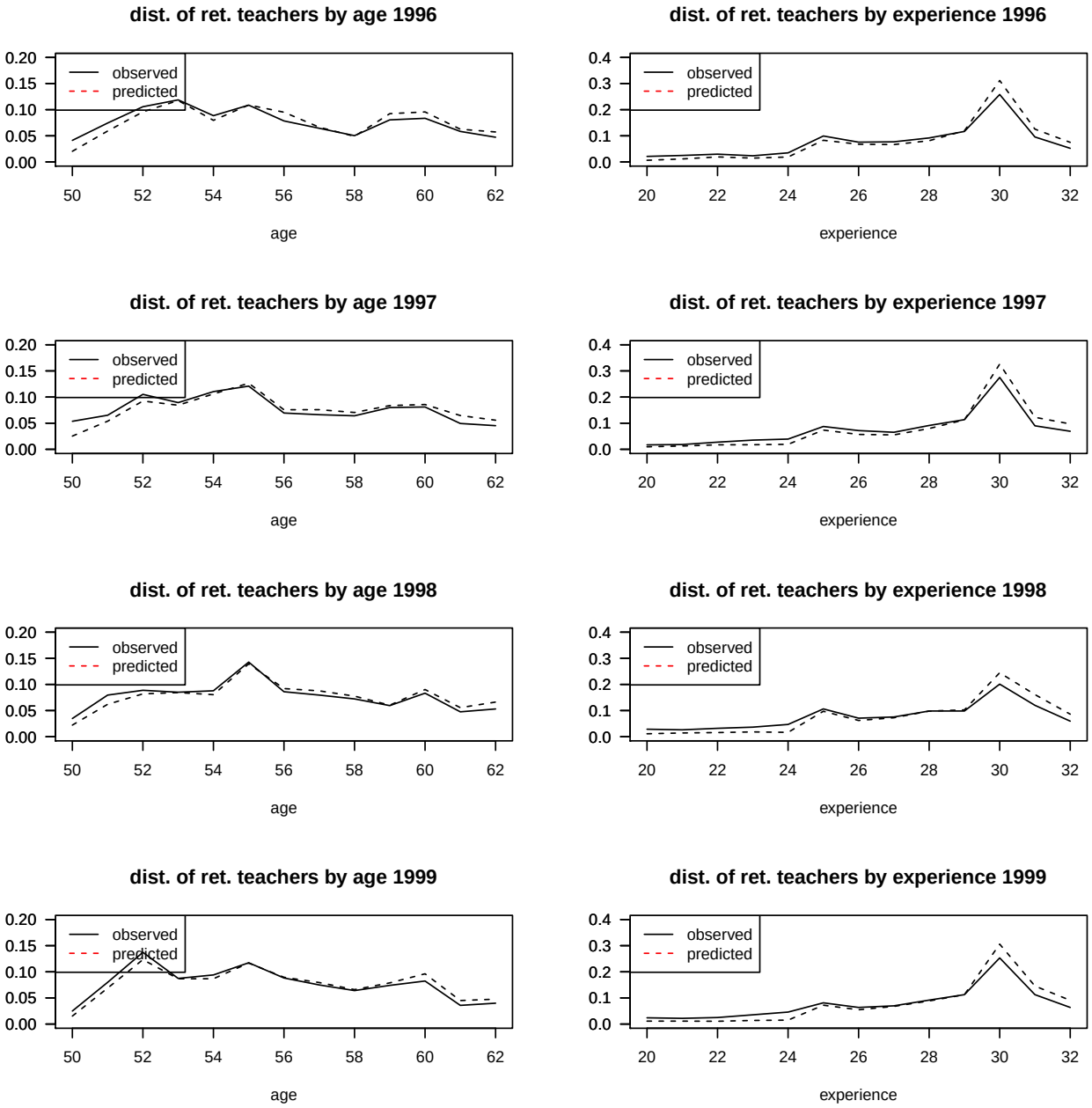
Note: The plot on the left is the observed age–experience distribution of all teachers the 2002 cohort at the year of retirement. The plot on the right is the in-sample model prediction of the age–experience distribution of all teachers the 2002 cohort at the year of retirement. The simulation on the right is based on the estimates in the first column of Table 3.

Figure 4: Observed and Predicted Distributions of Retiring Teachers in 1995



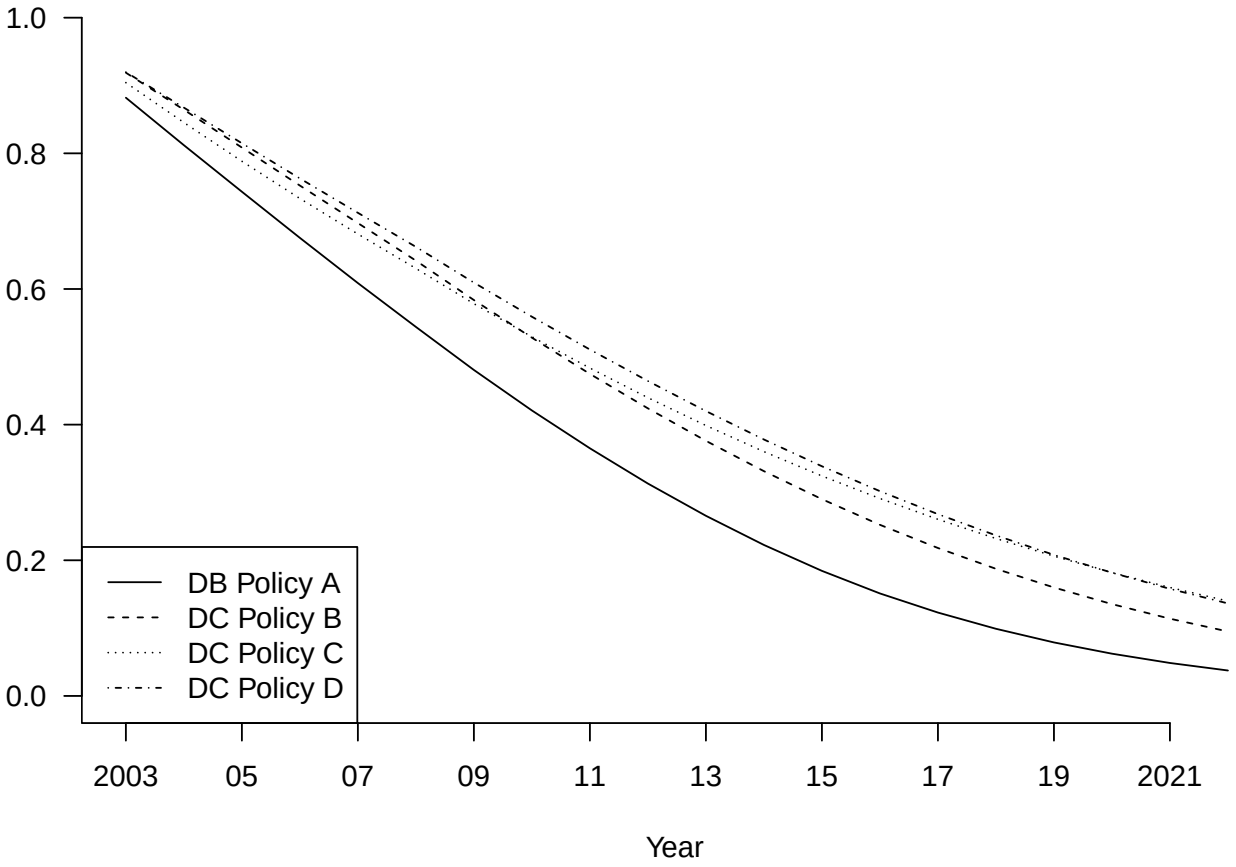
Note: Figure 4a and 4b: the observed 1995 distribution and the out-of-sample predicted distribution of retiring teachers by age and experience under the 1995 DB rules. Figures 4c and 4d plot the observed and predicted retirement probabilities of teachers given the age or experience in 1995. The out-of-sample simulation is based on the estimates in first column of Table 3.

Figure 5: Observed and Predicted Distributions of Retiring Teachers 1996-1999



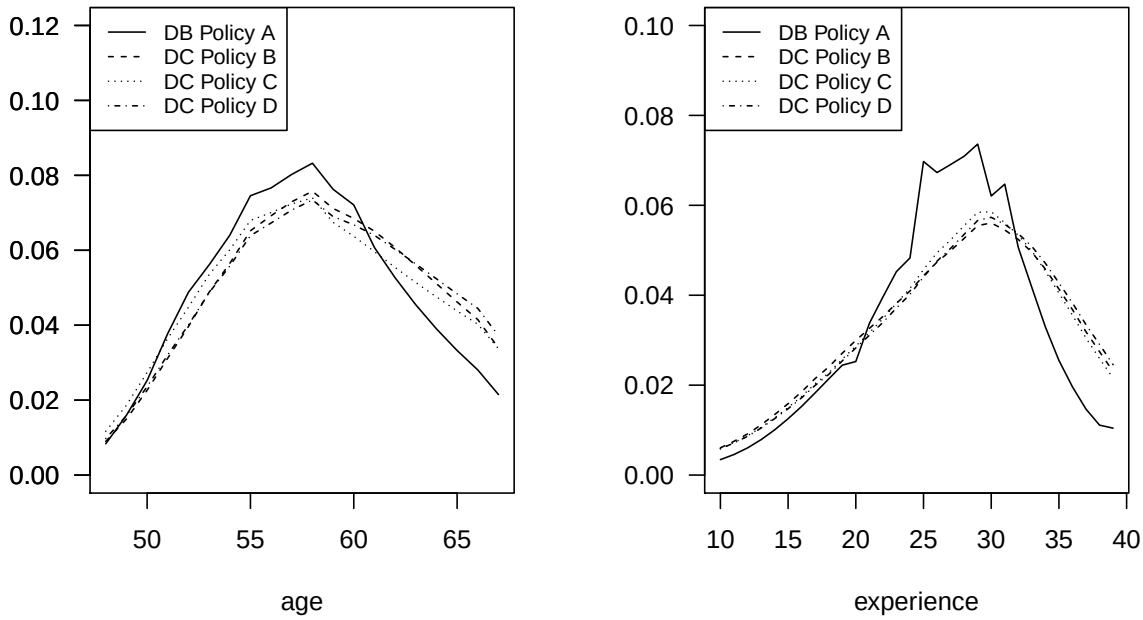
Note: The observed 1996-1999 distribution and the out-of-sample predicted distribution of retiring teachers by age (on the left) and experience (on the right) under the DB rules of the respective years. The out-of-sample simulation is based on the estimates in first column of Table 3.

Figure 6: Predicted Survival Rate Under Alternative Policies 2003-2022.



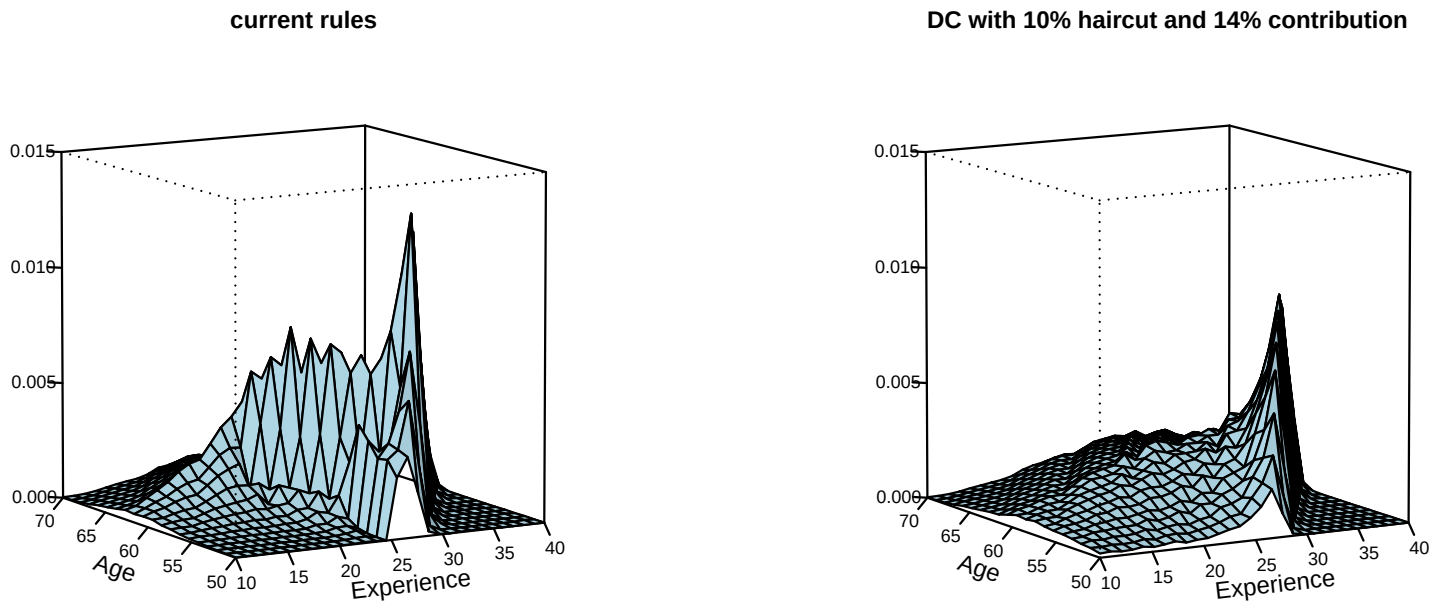
Note: The simulated survival rates are based on the estimates in the first column of Table 3, under alternative pension rules, 20 year prediction.
Policy A: the current DB plan;
Policy B: $r = 6.5\%$, conversion to a DC plan with the full 2002 pension wealth and contribution rate $c = 14\%$;
Policy C: $r = 4\%$, conversion to a DC plan with the full 2002 pension wealth and contribution rate $c = 10\%$;
Policy D: $r = 4\%$, conversion to a DC plan with a 10% "haircut" in the 2002 pension wealth and contribution rate $c = 14\%$.

Figure 7: Predicted Retirement Age and Experience Distributions Under Alternative Policies, 2003-2022.



Note: The simulation is based on the estimates in the first column of Table 3, 20 year prediction. Current DB rules assume a nominal return of 6.5%.
Policy A: the current DB rules;
Policy B: $r = 6.5\%$, conversion to a DC plan with the full 2002 pension wealth and contribution rate $c = 14\%$;
Policy C: $r = 4\%$, conversion to a DC plan with the full 2002 pension wealth and contribution rate $c = 10\%$;
Policy D: $r = 4\%$, conversion to a DC plan with a 10% "haircut" in the 2002 pension wealth and contribution rate $c = 14\%$.

Figure 8: Predicted Joint Distribution of Retirement Age-Experience Under Alternative Policies 2003-2022.



Note: The simulation is based on the estimates in the first column of Table 3, 20 year prediction. The left plot is the joint age-experience distribution of retiring teachers under Policy A (the current DB rules.) The right plot is the joint age-experience distribution of retiring teachers under Policy D ($r = 4\%$, conversion to a DC plan with a 10% "haircut" in the 2002 pension wealth and contribution rate $c = 14\%$.)