A Dynamic Model of Teacher Labor Supply

Todd R. Stinebrickner, *University of Western Ontario*

The labor supply decisions of certified elementary and high school teachers are examined using data from a general longitudinal survey. A significant decrease in the teaching participation rate takes place over time after teacher certification. Previously unavailable marital and fertility variables provide new insight into reasons for this decrease. Descriptive statistics indicate that high-ability teachers choose to teach a smaller proportion of time than other teachers. A dynamic, discrete-choice model, which accommodates Serial Correlation in the wage process for teachers, is used to analyze responsiveness of the overall sample and different types of teachers to two potential types of wage increases.

I. Introduction

The attrition rates of certified elementary and high school teachers have received significant attention in recent education policy debates. Interest in this topic is in part motivated by the possibility of increases in teacher demand. For example, a recent study by the United States Department of Education projects that, due to demographic changes in the population, the number of public and private classroom teachers will increase by 350,000 between 1995 and 2007.[1] Further, proposed governmental policies, which would mandate smaller class sizes at certain grade levels,

The author would like to thank Steven Stern, Michael Brien, William Johnson, John Bound, Chris Swann, Jonathan Skinner, David Breneman, Ralph Stinebrickner, and numerous seminar participants. The author also appreciates the hospitality that he received at the Osceola Research Institute.

[1] This represents a 12% increase. See the U.S. Department of Education report *Projections of Education Statistics to 2007* (Hussar and Gerald 1997). Also see the

[*Journal of Labor Economics*, 2001, vol. 19, no. 1]

0734-306X/2001/1901-0007$02.50

would lead to an additional increase in teacher demand. However, ensuring that schools have an ample supply of teachers is not the only reason that this issue is important. If it is the most talented or most effective teachers who are leaving the classroom, teacher attrition can be harmful. Studies, such as Murnane (1975), Murnane, and Phillips (1981), and Hanushek, Rivkin, and Kain (1998), suggest that a given teacher may become more effective during her early years in the field. If this is the case, high attrition rates in early years can significantly diminish the number of teachers who stay in teaching long enough to reach their most productive years. In addition, the salary structure in most public schools depends only on the amount of teaching experience and the amount of post-bachelor education that a person has accumulated. This wage structure rigidity may be harmful from the standpoint of retaining academically gifted teachers who, on average, would be expected to receive better nonteaching opportunities than other teachers.[2]

The goal of this article is to examine the effect that personal factors and wages have on a teacher's decision to enter and leave the teaching profession after the certification decision has been made. Although previous studies have had similar objectives, the approach taken in this article differs significantly from most previous work in terms of both the data and modeling approach that are used.[3] Most previous studies have utilized "teacher specific" data that are collected from a particular school district or the educational system in a particular state.[4] A benefit of these data is that they typically contain information on a large number of teachers. However, these data suffer from some significant limitations. First, because the data are collected directly from an educational system, no information is available about a person's labor activities before or after her teaching spell in the particular geographical region. For example, after a person leaves a teaching job in the particular state or school district, the data do not provide information about whether the person is teaching in

associated special report *Here Come the Teenagers: A Back to School Special Report on the Baby Boom Echo* (U.S. Department of Education, 1997).

[2] Brewer (1996) suggests that the rigidity in the teaching wage structure is less significant once potential teacher mobility to administrative positions within the educational system is taken into account.

[3] Previous work in the area of teacher decisions includes Eberts (1987); Murnane and Olsen (1989, 1990); Murnane, Singer, and Willett (1989); Theobald (1990); Dolton and van der Klaauw (1995, 1999); Brewer (1996); Gritz and Theobald (1996); Mont and Rees (1996); Theobald and Gritz (1996); van der Klaauw (1996*a*); and Stinebrickner (1998*a*, 1999*a*, 1999*b*).

[4] For example, data of North Carolina and Michigan teachers are used by Murnane and Olsen (1989, 1990) and Murnane et al. (1989). Mont and Rees (1996) use data from the State of New York. Gritz and Theobald (1996) and Theobald and Gritz (1996) use data from the state of Washington.

a different geographical area, is working in a nonteaching occupation, or has left the workforce altogether. A second limitation is that these data typically contain very little personal information about teachers. For example, the data do not contain information about a teacher's marital status or number of children, and sometimes they do not contain measures of academic ability.

In this article, these data limitations are avoided by using a general longitudinal study to construct individual-level data about certified teachers. The new data allow insight into the extent of teacher attrition. For example, past research, which used the previously discussed types of data, consistently found that many teachers have relatively short first spells in teaching. Studies that use teacher-specific data and focus on first spells in teaching may potentially overstate the problem of teacher attrition because some people who are assumed to be exiting the teaching profession may actually continue to teach in a different geographical region and some people who do leave teaching may return to teaching for subsequent teaching spells after relatively short absences from the teaching occupation. However, simple descriptive statistics in this article show that the teaching participation rate (i.e., the proportion of eligible teachers who are teaching in a particular year) does decline significantly in the years after certification, even when these two additional factors are taken into account.

Perhaps more important, the data used in this article also allow new insight into the reasons for teacher attrition. There seems to be a perception in the general public and in previous literature that the most important cause of teacher attrition is that teachers are drawn away from teaching by the attractiveness of nonteaching opportunities. For example, the theoretical models used to motivate the empirical work in Murnane and Olsen (1989, 1990) assumed that teachers who leave teaching do so to begin work at their best nonteaching job alternative. Similarly, recent work by Gritz and Theobald (1996) suggested that a teacher's "decision to remain in his or her current teaching assignment depends upon the perceived benefits of this choice relative to alternative career opportunities" (p. 478). However, given that starting teachers tend to be young and a large percentage of starting teachers are female, it seems very possible that many individuals who leave teaching jobs may leave the workforce altogether.[5] Indeed, simple descriptive statistics in this article suggest that this is the case. Although decreases in the teaching participation rate over time after certification are accompanied by both increases in the partici-

[5] Meitzen (1986) finds differences in male and female job quitting behavior. For an additional analysis of quit behavior, see Weiss (1984).

pation rate at nonteaching occupations and increases in the proportion of individuals who are not working, the latter are more important.

This article uses a dynamic, discrete choice framework to model the relationship between personal characteristics, wages, and the decision process of certified teachers.[6] In each school year, the model allows the certified teacher to choose a labor activity from among a set of teaching, nonteaching, and leisure options. Data on wages and individual choices allow the model to separately identify how wages and nonpecuniary utility in the various options affect this decision. Given the finding that the proportion of individuals who are out of the workforce increases in the years after certification, it is not surprising that previously unavailable marital and children variables are important determinants of nonpecuniary utility. It is important to note that this article includes these variables exogenously. As will be discussed in more detail, this assumption of exogeneity is incorrect to the extent that individuals make childbearing and marital decisions jointly with labor decisions.

An important policy issue involves the effect that wage increases have on the decisions of the overall group of teachers and also on the decisions of teachers with different levels of ability. Ideally, a measure of each person's teaching effectiveness would be observed directly in the data. Although this is not the case, the data do include a measure of the teacher's academic ability, the person's score on the Scholastic Aptitude Test (SAT). While SAT scores do not directly measure teaching effectiveness, it seems reasonable to believe that scores on this test are correlated with teaching effectiveness. By relating teacher decisions directly to underlying preferences, the structural model is conducive to examining the effects of potential wage policy changes. In particular, to illustrate the sensitivity of teacher labor supply to wage changes, this article will examine two different types of wage increases. The first, which corresponds to the often heard normative statement "teachers should earn more money," involves a uniform wage increase for all teachers. The second deviates from the traditional, rigid wage structure by allowing more flexibility in the wage offers that can be made to academically gifted teachers, who, in general, may have better nonteaching opportunities than other teachers.[7] Given the large amount of taxpayer resources that are spent on teacher salaries, this preimplementation analysis of the effects of policy changes is valuable.

The remainder of the article is organized as follows. Section II discusses

[6] The only other work in the area of teacher labor supply that uses this modeling approach is van der Klaauw (1996*a*). See Eckstein and Wolpin (1989) for a survey of applications well suited to this type of modeling approach.

[7] This is a deviation of an idea originally offered by Kershaw and McKean (1962); they proposed wage premiums for individuals in certain subject areas.

the longitudinal data that are used for the estimation of the model and presents descriptive statistics. The structural model of teacher decisions is described in Section III. Given the objective of simulating the effects of potential changes in the teaching wage structure, it is desirable to model the teaching wage process with as much realism as possible. Section III also describes the autoregressive process that is used, and a related appendix describes the econometric methods that are developed to address the well-known difficulties that arise during estimation as a result of relaxing the traditional dynamic, discrete choice modeling assumption that unobservables are serially uncorrelated. The structural results and policy simulations are discussed in Section IV, and Section V concludes the article.

II. Data

A. Description of the Sample

The National Longitudinal Study of the High School Class of 1972 (NLS-72) is used to estimate the model. The first wave of this survey, which was completed in 1972, includes interviews with 22,652 students who were expected to graduate from high school in that year. Included in the first wave is information on aptitude tests, such as the Scholastic Aptitude Test (SAT). Follow up surveys were taken in 1973, 1974, 1976, 1979, and 1986. Thus, for each person, the survey contains detailed information about work experience, education, marriage, and fertility for approximately 14 years after the person graduated from high school.[8]

Further, 832 individuals who were certified to teach in elementary or secondary schools were sent supplemental questionnaires that asked questions about their teaching experiences. The final sample used in this article consists of 450 of these individuals who became certified to teach at some point between 1975 and 1985.[9] The people who were sent the teaching supplement but who do not appear in the final sample were excluded for a variety of reasons. Some had missing observable characteristics. Others had crucial missing information that made the construction of job or personal histories impossible.

Since this article involves the career choices of individuals only after

[8] Since survey waves did not occur in every year, some of the survey waves ask the individual retrospective questions that cover several years of the individual's life. One consequence of this is that the estimation algorithm must take into account that the individual is not asked about wages for every year in which she works.

[9] It should be noted that because teacher certification is used as a selection criterion for the sample, the estimates in the model do not represent the preferences of the entire population of 1972 high school graduates or even the subset of 1972 high school graduates who also graduated from college.

Table 1
Descriptive Statistics

Variable	Mean	Standard Deviation
Number of years individual is observed (after certification)	9.0	4.1
Mathematics Standardized Aptitude Test (SAT) score	476.2	93.2
Percent female	72.5	...
Number of children (in first year of certification)	.2	.5
Number of children (in 1986)	1.1	1.1
Percent with at least one child (in first year of certification)	12.3	...
Percent with at least one child (in 1986)	63.2	...
Percent married (in first year of certification)	36.4	...
Percent married (in 1986)	77.7	...
Percent married in at least one period	81.4	...
Number of years of post-bachelor education (as of 1986)	1.4	1.2
Years of teaching experience (as of 1986)	4.3	3.6
Years of nonteaching experience (as of 1986)	2.9	3.3

they become certified, which usually requires a minimum of 4 years of training, the data contain between 1 and 11 years of work histories and personal information for each person.[10] Table 1 summarizes this information.

In each of the data years, it is observed whether the person chooses a teaching job, a nonteaching job, or not to work. Of the aggregated 4,041 years of data, .48 of the years are spent teaching, .32 working in nonteaching jobs, and the other .20 not working. However, there are substantial differences in these proportions by sex. For females, the proportions are .46, .29, and .24 respectively. For males, the proportions are .51, .40, and .09.

If teacher attrition is indeed a problem, these proportions may not be constant over time after certification and examining how these proportions change over time may provide some descriptive insight into the extent and reason for teacher attrition. For the females in the sample, figure 1 shows that, in each year after certification, the teaching participation rate is greater than the participation rate in other nonteaching occupations. The proportion of individuals who are not working is lower than both of these participation rates in almost all years. However, the relationship among the proportion of women who are teaching, the proportion of women who are working in nonteaching jobs, and the proportion of women who are not working is not constant over time after certification. Between the second year after certification and the ninth year after certification, the participation rate in teaching falls by nearly

[10] A small number of people finished college in 3 years. For these people, 11 years of data are observed. For most people, 10 or fewer years of data are observed.

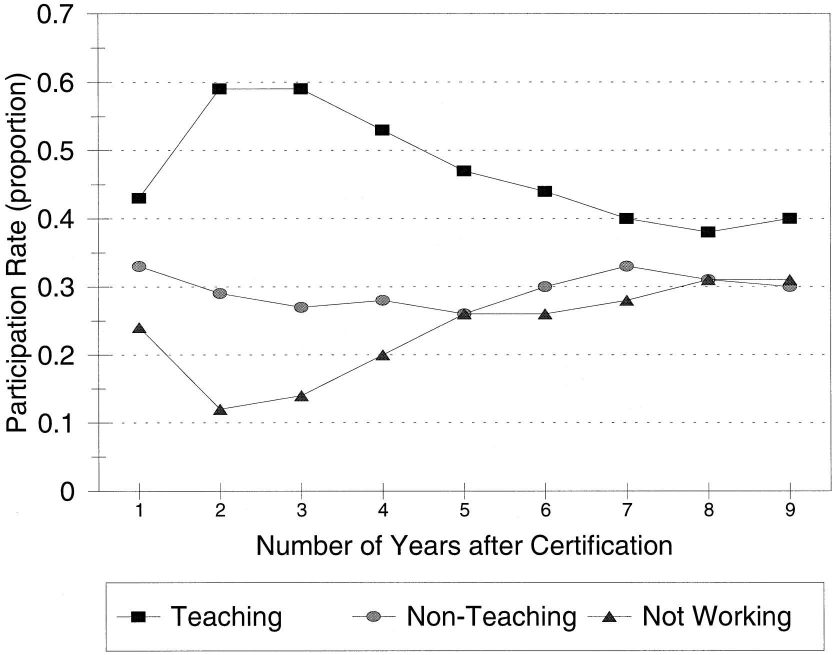

FIG. 1.—Participation rates of females over time after certification

.20.[11] However, the participation rate in nonteaching occupations remains largely unchanged during this same period. Instead, figure 1 suggests that a large increase in the proportion of women who are out of the workforce plays a more important role in the declining teacher participation rate that is found for women between the second and ninth years. Figure 2 shows that the teaching participation rate for males also decreases between the second and ninth years after certification. However, unlike the findings for females, figure 2 shows that the participation rate in nonteaching occupations increases in the years after certification, whereas the proportion of males who are out of the workforce remains relatively constant.

The aggregated proportions in figure 1 and figure 2 do not reveal much information about the trends of particular individuals. For example, a

[11] This significant decrease occurs despite the fact that fig. 1 allows for the possibility that some teachers may change geographical regions but remain in teaching and also takes into account that some people have multiple spells in teaching. The teaching participation rate in the first year after certification is substantially lower than in the second year. Especially for people who become certified as part of a bachelor degree program, this first year is likely to be a transitional period.

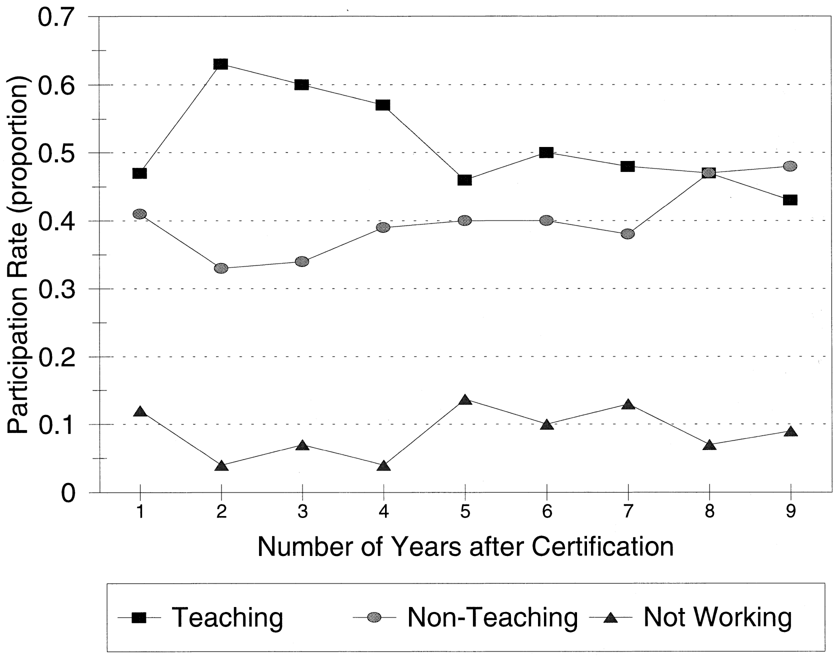

FIG. 2.—Participation rates of males over time after certification

yearly teaching participation rate of .50 is consistent with every teacher choosing teaching in half of the years. However, it is also consistent with 50% of all teachers never choosing to teach and the other 50% choosing to teach in every year. In these data, .76 of females and .75 of males choose to teach in at least one year.[12] The teaching duration for those who decide to teach can also be examined. In particular, figure 3 shows Kaplan-Meier survivor functions associated with the first spell in teaching for the 248 females and the 92 males who enter teaching at some point. In this work, the end of the first spell in teaching occurs when a person chooses to take a nonteaching job or not to work at all; changing teaching jobs does not represent the end of a spell. The survivor function evaluated at a duration of t years is the probability that an individual will teach more than t full years before quitting.[13] Thus, in the first teaching spell, the probability that a female teacher will teach more than 4 full years before leaving teaching is .48, and the probability that a male teacher will stay in teaching

[12] Care must be taken in interpreting this number because some individuals became certified near the end of the sample period.

[13] For example, the survivor function evaluated at one year measures the probability that the person will return to teaching for her second year.

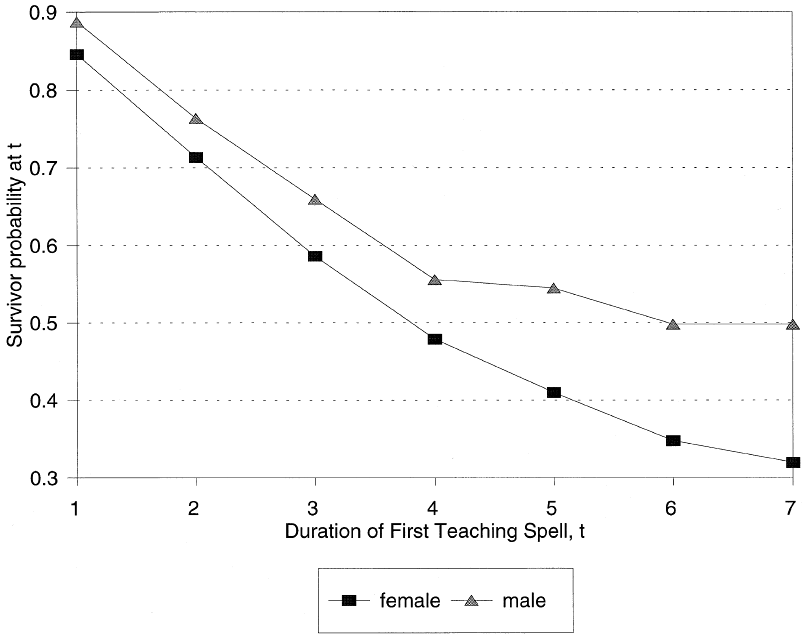

FIG. 3.—Kaplan-Meier survivor functions: Length of first teaching spell by sex

more than 4 years is .56. The increasing gap between the survivor functions in figure 3 indicates that, relative to men, women become more likely to exit teaching as the number of years after certification increases.[14]

A challenge for the structural model in this article is to explain why the decisions of males and females are quite similar during the first years after certification but become increasingly different as the number of years after certification increases. The reality that the relative behavior of men and women changes over time suggests that it is unlikely that the differences in the behavior of men and women can be entirely attributed to differences in permanent factors, such as a person's "love" for teaching. One possible explanation of the trends in figures 1–3 is that men and women react differently to family changes that occur during the years

[14] Figure 3 does not indicate the reason for teaching exits. However, Stinebrickner (1999*c*) found that .40 of the teachers in this sample with uncensored teaching spells worked in a new occupation the year after exiting the teaching occupation. The other .60 of exiting teachers left the workforce entirely. Some of the first teaching spells are right censored because individuals are still teaching in their first spell at the end of the sample. For censored spells, the exit reason is not observed.

after certification. Using data from a general longitudinal survey allows this possibility to be explored in detail.

B. Trends in the Labor Supply of Different Types of Teachers

The quality of teachers has been shown to be an important input to student learning (Bishop 1996; Hanushek et al. 1998). Unfortunately, the data do not provide a direct measure of teaching effectiveness. However, test scores assessing the teacher's academic ability are observed and have been shown to be a teacher characteristic that increases student learning (Hanushek 1971; Strauss and Sawyer 1986; Ferguson 1991; Monk 1992; Ehrenberg and Brewer 1993).[15] For the remainder of the article the term "ability" will be used to refer to the academic ability of the teacher.

It is of interest to compare the trends in the labor supply of teachers with differing levels of ability. As noted, the teaching supplement includes information about teaching certification. This is valuable because it allows the sample to include not only people who actually decide to teach but also people who are qualified to teach but choose alternative career paths. Thus, two distinct aspects of the relationship between ability and occupational choice can be examined: the effect of ability on the likelihood that a certified person actually enters teaching and the effect of ability on the duration of teaching spells for those who enter teaching. However, this article does not account for the reality that the decision to become certified is endogenous. Therefore, the effect that changing teaching wage structures has on the certification decision will not be examined. It is important that the policy conclusions in the article take this into account.

The measure of ability that is used in this analysis is the individual's score on the mathematics SAT score.[16] Using SAT scores as the grouping

[15] For example, Strauss and Sawyer (1986) find that having teachers with higher scores on the National Teaching Examination (NTE), which Ayers and Qualls (1979) find to be positively correlated with SAT scores, has a substantial effect on whether or not students fail to demonstrate independently measured reading and mathematics skills. Murnane, Singer, Willett, Kemple, and Olsen (1991) do not conclude that the research link is quite as convincing as Bishop (1996) finds it. Nonetheless, the trend to increase the stringency of minimum standardized test scores for new teachers seems to indicate a general belief among policy makers that recruiting teachers with higher academic ability is important.

[16] It is likely that retaining and recruiting teachers with high mathematics SAT scores will be an important priority given student deficiencies on standardized tests of basic math and science. In general, the retention of teachers with high mathematics SAT scores may be more difficult than the retention of teachers with high verbal SAT scores because the mathematics SAT score is a strong predictor of opportunity costs. For example, Murnane, Willett, and Levy (1995) find that males graduating from high school in 1972 with strong basic math skills have significantly higher hourly earnings at age 24 than do males graduating in that year with average mathematics skills. Using data of this article, simple wage

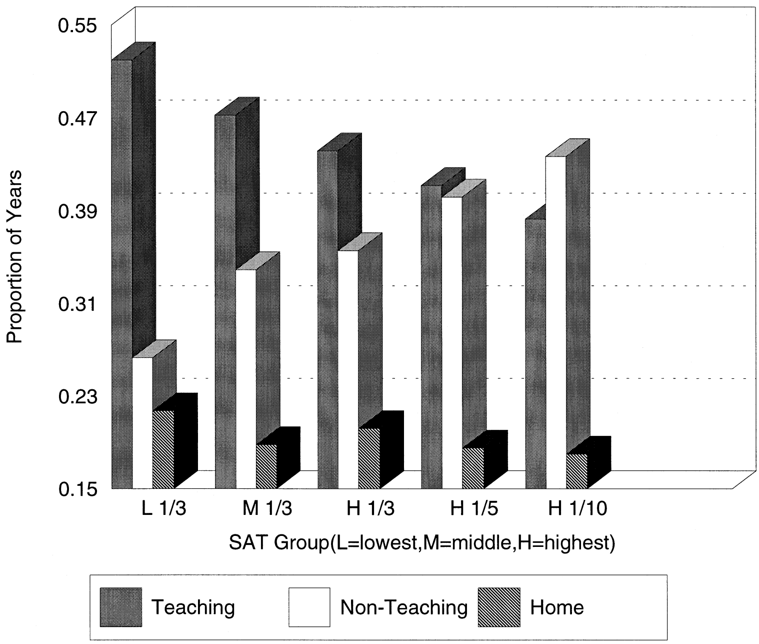

FIG. 4.—Proportion of aggregate years in each option by SAT group

criterion, the occupational trends of different ability groups are examined. The first three entries in figure 4 show the proportion of aggregate person years spent in teaching jobs, nonteaching jobs, and the nonwork (home) option for three mutually exclusive SAT groups: teachers in the lowest third of the sample in terms of SAT scores, teachers in the middle third, and teachers in the highest third. Notice that the proportion of periods spent in teaching decreases monotonically across the three groups as the ability level increases, therefore showing that high-ability teachers do choose teaching less frequently than do other teachers under the current rigid wage structure. Two additional groups, teachers in the highest fifth and teachers in the highest tenth, are added to give more detail about teachers at the top of the ability distribution. Members of these groups choose teaching even less frequently than do those in the other groups.

regressions indicate that teachers with high mathematics SAT scores do receive higher nonteaching wages than do other teachers. However, this nonteaching wage premium is not present for individuals with high verbal SAT scores. Using combined mathematics and verbal scores would tend to obscure some of the problems and solutions associated with retaining teachers with high mathematics SAT scores.

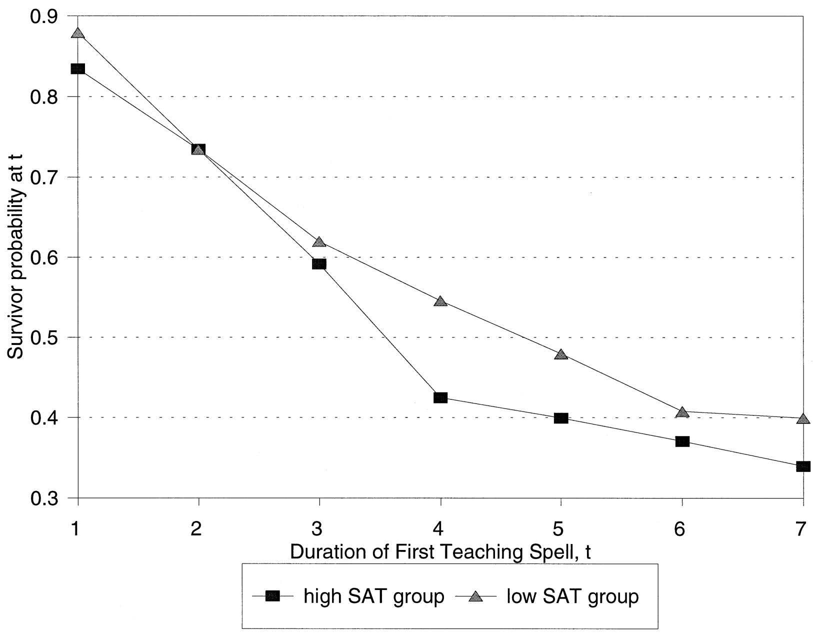

FIG. 5.—Kaplan-Meier survivor functions: Length of first teaching spell by SAT

Both the effect of ability on the decision to enter teaching (conditional on certification) and the effect of ability on teaching duration contribute to the decreasing trend across groups. To see this, consider differences in occupational choices between two arbitrarily chosen ability groups: teachers in the top third in terms of mathematics SAT scores and teachers in the bottom two-thirds in terms of mathematics SAT scores. The "high"-ability group chooses teaching .44 of the time, while the "low"-ability group chooses teaching .50 of the time. Low-ability individuals are more likely to enter teaching in at least one year during the sample period than are the high-ability individuals; .78 of the low-ability group participate in teaching for at least one year, while only .72 of the high ability group participate in teaching for at least one year. The Kaplan-Meier survivor rates, given in figure 5 for the two groups, show that the survivor function for the high-ability group is always below the survivor function for the low-ability group. The definition of the "high" and "low" groups will be used again in Section IV to examine the effect of alternative wage policies.

III. The Dynamic Model

At the beginning of each school year, a particular teacher examines the set of teaching and nonteaching opportunities that are available to her for the current year and makes an employment decision by choosing the most desirable option. In each period, the person always receives one new teaching job offer and one nonteaching job offer and also has the option of choosing the leisure (home) option and not working.[17] In addition, if the person was employed as a teacher in the previous period, she has the option of returning to the previously held job. Thus, the set of available employment options in a given year depends in part on the person's employment status in the previous year.

In order to make her employment decision, the person assesses the total wage and nonwage benefits, or utility, that she would receive from each of her options and chooses the option with the highest utility. There are several theoretical reasons why individuals do not make identical labor decisions. First, different people do not receive identical job offers. This is the case because both the person's observable characteristics (e.g., SAT scores, gender, and type of bachelor degree) and randomness in the labor market can affect the wages and nonpecuniary job characteristics that compose the job offers that a person receives. Second, different people may derive different levels of utility from identical sets of employment (and leisure) opportunities. One reason for this disparity is that observable characteristics may play a role in determining the amount of enjoyment that a person receives from being in each of the professions (e.g., an academically gifted person may find the teaching profession more or less stimulating than a teacher with less academic ability, a person with children may get more enjoyment out of the academic work schedule associated with teaching than a person without children).[18] However, it is also likely that each individual has a distinct level of appreciation for the teaching, nonteaching, and home options (e.g., a "love" for teaching), which cannot be captured by observable characteristics. This "unobserved heterogeneity" is assumed to be known by the individual and constant over the person's work horizon but, because it is unobserved in the data, is unknown to the econometrician. In addition, it is reasonable to assume that the person experiences some yearly fluctuation (randomness) in how much she enjoys being in each of the professions.

Let W_{it}^{j} and Q_{it}^{j} represent the wage and nonwage utility (in wage equivalents), respectively, for person i at time t for some employment

[17] The data do not provide any information that would directly identify the probability that a person receives a teaching or nonteaching offer in a particular year. Rather than attempting to identify these probabilities from the functional form of the model, it is assumed that the person receives one new teaching and one new nonteaching offer in each period. The model does imply that the teacher sometimes receives bad offers, which, from the viewpoint of the teacher, could be very similar to receiving no offer.

[18] The term "profession" is used loosely. The three "professions" in this article are the teaching profession, the nonteaching profession, and the home (leisure) profession.

option j. Let $j = E$, $j = N$, and $j = H$ denote a teaching job, a nonteaching job, and the home option, respectively. The total current period utility, U^j_{it}, that i receives in t by choosing j is assumed to be additive in W^j_{it} and Q^j_{it}:

$$U^j_{it} = W^j_{it} + Q^j_{it} = (\alpha^j_W X_{it} + \nu^j_{it}) + (\alpha^j_Q X_{it} + \mu^j_i + \varepsilon^j_{it}). \tag{1}$$

Equation (1) shows that W^j_{it} and Q^j_{it} are specified to be linear functions of the set of observable characteristics, X_{it}, of person i at time t.[19] Thus, α^j_W and α^j_Q represent the effect that these observable characteristics have on the average wage and average nonwage utility that a person receives from option j.[20] The unobservables ν^j_{it} and ε^j_{it} are error terms that represent the randomness in the wage and nonwage utility, respectively, of the person in option j, and μ^j_i represents the person's unobserved taste (heterogeneity) for option j.[21]

The teacher knows α^j_W and α^j_Q and, therefore, the average wage, $\alpha^j_W X_{it}$, and average nonwage utility, $\alpha^j_Q X_{it} + \mu^j_i$, that she would receive by choosing option j in any period t if she had observable characteristics X_{it}.[22] In the current period, the person is assumed to know also the realizations of the error terms, ν^j_{it} and ε^j_{it}, and, therefore, the exact values of the wage, W^j_{it}, the nonpecuniary utility, Q^j_{it}, and the total current period utility, U^j_{it}, from each option.[23] Thus, if the person cared only about her current period utility or if current period decisions did not affect the person's future utility, the person's

[19] The relevant elements of X_{it} may be different for the wage and nonpecuniary equations. In Sec. IV, when the elements of X_{it} are specified, the elements of α^j_W and α^j_Q, which are constrained to be zero, will be discussed.

[20] It would be desirable to include nonpecuniary school characteristics in the model. However, this adds complexity to the model estimation because (1) the characteristics are endogenously determined and must be considered state variables, (2) the characteristics are correlated with each other and with the wage, and (3) the characteristics are sometimes missing. Solutions to these problems are discussed in Stinebrickner (1999*b*). Since school characteristics are not included, α^E_Q represents both the effect of observable personal characteristics on how much a person likes teaching and the effect of personal characteristics on the average type of job (in terms of the nonpecuniary working conditions) that a person obtains.

[21] The unobserved heterogeneity term μ^j_i is assumed to be distributed $N(0, \sigma^2_1)$, $N(0, \sigma^2_2)$, $N(0, \sigma^2_3)$ for $j = E$, N, and H, respectively.

[22] Some of the person's observable characteristics in X_{it} (e.g., years of teaching experience) depend on choices that the individual makes. That is, they are endogenous state variables.

[23] This does not seem to be a particularly unreasonable assumption because, for a particular job in the teaching profession, it is likely that an individual would have a wage contract for the current year, would know the nonpecuniary school characteristics associated with the job for the current year, and would know how she currently felt about teaching.

employment decision would be made by choosing the option j with the highest current period utility, U^j_{it}. However, in this model neither of these is the case. The former is not true because, instead of being myopic, the individual is assumed to care about the discounted utility that she will receive over a finite work horizon. The latter is not true for two reasons. First, as mentioned earlier, current period decisions influence future utility through their determination of which options will be available to the person in the future. Second, current period decisions also influence a set of state variables that affects the utility that a person receives from these available future options. For example, the person knows that one result of deciding to teach in the current period is that she will accumulate an additional year of teaching experience. Given the rigidity in the teaching wage structure, the extra year of experience is likely to have strong, positive effects on future teaching wages.[24]

However, although the teacher is assumed to know with certainty the utility associated with each option in the current period, this is not a good assumption about future periods; in reality, the teacher cannot know exactly what types of jobs (wages and nonpecuniary characteristics) will be offered to her in the future or exactly how she will feel about the teaching, non-teaching, and home options. The model captures this uncertainty by assuming that ν^j_{it} and ε^j_{it} are random variables whose future realizations are not known (although the individual does know the distributions). This uncertainty implies that, for an option j, the individual cannot compute the exact discounted utility that she will receive over her finite work horizon but, instead, calculates and makes decisions based on the discounted expected utility, or value, of the option over her finite work horizon.

The value function of an option, which describes how the value of the option depends on the person's state variables, can be written as

$$V^j_{it} = U^j_{it} + \beta E \max[\{V_{it+1} | d_{it} = j\}], \tag{2}$$

where β is the discount rate, E is the expectation operator, max is the maximum function, and the set $\{V_{it+1}|d_{it} = j\}$ includes the value functions associated with the set of available options for the person in the next period, $t + 1$, conditional on the choice in time t, d_{it}, being equal to j.[25]

Although the individual is assumed to make decisions based on the value functions given by equation (2), the econometrician cannot calculate (2). Given distributional assumptions about the error terms in the model,

[24] To the extent that the amount of enjoyment in teaching depends on how long the person has been in the field, it may also affect the nonpecuniary rewards of teaching in the future.

[25] Although not explicitly written, V^j_{it} is a function of state variables.

the econometrician can calculate the second term on the right side of equation (2). However, the econometrician cannot calculate the value of U^j_{it} known by the individual because the exact nonpecuniary utility, Q^j_{it}, that person i receives from option j is not observed by the econometrician.[26] The econometrician can calculate the expected value functions: $\bar{V}^j_{it} = V^j_{it} - \varepsilon^j_{it}$.[27]

The goal of the maximum likelihood estimation procedure is to estimate the values of α^j_W, α^j_Q, and the parameters defining the distributions of the random variables ν^j_{it}, ε^j_{it}, and μ_i associated with each option j. Much of the identification in the model comes from observing the sequence of choices that each person makes. At each point in time, conditional on all current period wages and the values of unobserved heterogeneity, the econometrician can use the set of available expected value functions $\{\bar{V}^j_{it}|d_{it-1}\}$ to compute the probability, P_{it}, of the choice that is actually chosen by person i.[28]

In order to separately identify wage effects from nonpecuniary utility, information from observed wages must also be used. Although our sample consists of individuals from a single birth year, we are able to separate the influence that teaching experience has on wages from the time trend in wages that is present over the sample period. This is the case because the majority of individuals do not choose to teach in all possible years after certification, which ensures that variation exists in accumulated teaching experience among the individuals who choose to teach in each particular calendar year.[29] Because wages are likely to play an important role in the labor decisions of teachers, the model in this article makes an effort to realistically describe the wage process of teachers. Given this focus, assuming that the error terms in the teaching wage

[26] A person's utility is not observable. In terms of the model, this comes from the assumption that the random shock to nonpecuniary utility, ε^j_{it}, is not observed.

[27] To do so, the recursive nature of eq. (5) implies that a set of expected value functions $\{\bar{V}_{it+1}|d_{it} = j\}$ must be solved at time $t + 1$ before each of the expected value functions $\bar{V}^j_{it}$ can be solved at time t. Thus, the standard solution technique involves a backward recursion algorithm.

[28] Under the assumption that ε^j_{it} is independently and identically distributed (i.i.d.) extreme value for all j (see Rust 1987; Berkovec and Stern 1991),

$$P_{it} = \frac{\exp\{\bar{V}^{d_{it}}_{it}\}}{\sum_j \exp\{\bar{V}^j_{it}\}},$$

where the summation is over all options that a person considers at time t given choice at time $t - 1$.

[29] Variation also exists because not all individuals become certified in the same calendar year.

equation are serially uncorrelated (i.e., that ν_{it}^{E} is independent of ν_{it+1}^{E}) does not seem desirable. First, some schools are generally higher paying than others, and a person is likely to take this into account when thinking about the future wages that she would receive if she returned to the same job in the future. Second, the time period covered by the data was one in which real teaching wages fluctuated substantially, in part because of changes in inflation. It is likely that a person would observe how the nominal wages paid by her particular school responded to price changes in a particular year and would use this information to update her expectations about the future wages that she would receive if she returned to the same job in the future. To allow the model to capture this type of forward-looking behavior, the unobservable in the teaching wage equation is assumed to follow an autoregressive (1) (AR[1]) process across years in a particular teaching job:

$$\nu_{it+1}^{E} = \rho \nu_{it}^{E} + e_{it+1}, \tag{3}$$

where $e_{it+1} \sim N(0, \sigma_e^2)$.

In dynamic, discrete choice models, error terms are traditionally assumed to be uncorrelated across time. Relaxing this assumption to allow the AR (1) specification in equation (3) creates significant complications from the standpoint of computing the value functions in equation (2). The approximation method designed to deal with this problem is described in the appendix and in greater detail in Stinebrickner (1998*b*).[30]

At a particular point in time, the likelihood contribution for person *i*, conditional on any unobserved wages and unobserved heterogeneity, is the joint probability of the choice that is made and any wage that is observed. The conditional likelihood contribution for the person over the sample period is the product of the period-specific conditional likelihood contributions. Finding the unconditional likelihood contribution requires integrating the conditional likelihood contribution over the distributions of the unobserved wages and unobserved heterogeneity. This is discussed in detail in Stinebrickner (1998*b*).

[30] The wage error in the first year of a teaching job is assumed to have a mean of zero and to be independent across periods: $\nu_t^E \sim N(0, \sigma_E^2)$ if t is the first year in a new teaching job. Clearly, it would also be desirable to allow the unobservable in the nonteaching wage equation to be serially correlated. However, in order to keep the model tractable, the traditional assumption of independence is assumed: $\nu_t^N \sim N(0, \sigma_N^2)$ for all t.

IV. Results

A. Model Estimates

Table 2 shows coefficient estimates for two specifications assuming a discount factor, β, of .95.[31] The first column indicates the personal characteristics that are in the vector X_{it} and, therefore, are assumed to enter the wage and nonwage utility equations. Although it would be desirable to estimate the model separately for males and females, this is not done because of the relatively small sample. However, because the descriptive statistics in Section II raised the possibility that males and females might react differently to family changes that take place in the years after certification, the specification does include interaction terms that allow marital and children variables to have different effects for males and females. It is assumed that the number of children and marital status of a teacher do not affect the person's wage. Thus, the coefficients of α^j_W corresponding to these variables are restricted to be zero. It is also assumed that the average nonpecuniary benefits from teaching do not depend on the year, so that the coefficients of α^j_Q corresponding to the time trend variables are restricted to be zero.

The estimates in the second column (specification 1) are obtained under the traditional assumption that the unobservable in the teaching wage equation is independent across periods (i.e., $\rho = 0$ in eq. [3]). The estimates in the third column (specification 2) show the results when this assumption of independence is relaxed. The log likelihood function is significantly better in the second column, −4,233.30, than in the first column, −4,686.11. There are several reasons why including serial correlation in teaching wages may improve the log likelihood function value. First, allowing serial correlation in teaching wages may increase the teaching wage probabilities in the model.[32] Second, allowing serial correlation in the model allows individuals to take into account more realistic information about future wages when making decisions. On average, this is likely to increase the choice probabilities in the model.[33] The final reason has to do with missing data. Recall that wages are often missing for some subset of a person's teaching career. When wages are allowed to be serially correlated, missing wages are more accurately "replaced" during the model estimation. If wages play an impotant role in the decisions of

[31] The model did not converge when I attempted to jointly estimate the discount factor. This is often the case with these type of models and occurs in Rust (1987) and Berkovec and Stern (1991).

[32] The likelihood contributions from wages, or wage "probabilities," involve the evaluation of a density function.

[33] Technically, it is really the joint probability of the wages and choices that is of relevance here.

Table 2
Structural Model Estimates

Variable X_{it}	Specification 1	Specification 2
Teaching wage (α_W^E):		
Constant	5.405*	5.718*
	(.069)	(.046)
Year (1975 = 1, 1976 = 2, etc.)	−.151*	−.270*
	(.017)	(.013)
Year × year	.008*	.015*
	(.001)	(.001)
Dummy for male	−.015	.007
	(.019)	(.012)
Mathematics SAT/100	−.020*	−.009
	(.010)	(.007)
Years of post-bachelor education	.046*	.053*
	(.009)	(.009)
Years of teaching experience	.050*	.019*
	(.006)	(.005)
Years of nonteaching experience	−.001	.006
	(.010)	(.009)
Autoregressive coefficient (ρ)	. . .	.943*
		(.008)
Nonteaching wage (α_W^N):		
Constant	4.509*	4.517*
	(.092)	(.094)
Year (1975 = 1, 1976 = 2, etc.)	−.011	−.015
	(.026)	(.027)
Year × year	.003	.002
	(.002)	(.002)
Dummy for male	.077*	.078*
	(.029)	(.029)
Mathematics SAT/100	.025*	.024*
	(.012)	(.012)
Years of post-bachelor education	.039*	.039*
	(.010)	(.008)
Years of teaching experience	−.039*	−.036*
	(.009)	(.010)
Years of nonteaching experience	.068*	.061*
	(.009)	(.009)
Teaching nonpecuniary utility (α_Q^E):		
Constant	−3.445*	−3.384*
	(.157)	(.175)
Dummy for male	.039*	.078*
	(.102)	(.115)
Mathematics SAT	−.018	−.032
	(.030)	(.031)
Years of teaching experience	−.085*	−.081*
	(.007)	(.007)
Years of nonteaching experience	−.057*	−.065*
	(.011)	(.009)
Number of children	−.438*	−.468*
	(.025)	(.031)
Number of children × male	.432*	.452*
	(.049)	(.062)
Dummy for marriage	−.428*	−.426*
	(.063)	(.070)
Marriage × male	.371*	.350*
	(.095)	(.115)

Table 2 (*Continued*)

Variable X_{it}	Specification 1	Specification 2
Nonteaching nonpecuniary utility (α_Q^N)		
Constant	−3.757*	−3.860*
	(.157)	(.173)
Dummy for male	−.114	−.072
	(.105)	(.116)
Mathematics SAT	−.027	−.025
	(.030)	(.032)
Years of teaching experience	−.049*	−.060*
	(.009)	(.010)
Years of nonteaching experience	−.044*	−.035*
	(.009)	(.009)
Number of children	−.408*	−.418*
	(.026)	(.032)
Number of children × male	.396*	.408*
	(.049)	(.062)
Dummy for marriage	−.463*	−.448*
	(.063)	(.070)
Marriage × male	.471*	.445*
	(.097)	(.112)
Variance terms:		
σ_1 (heterogeneity teaching)	.072*	.086*
	(.016)	(.017)
σ_2 (heterogeneity nonteaching)	.083*	.132*
	(.017)	(.019)
σ_3 (heterogeneity home)	.920*	.909*
	(.041)	(.053)
σ_E (teaching wage—new job)	.371*	.352*
	(.005)	(.004)
σ_e (teaching wage—old job)	.384*	.310*
	(.006)	(.004)
σ_N (nonteaching wage)	.460*	.453*
	(.007)	(.006)
τ	.417	.441*
	(.018)	(.017)
Log likelihood function	−4,686.11	−4,233.30

NOTE.—The numbers are estimates from a specification in which the discount factor, β, is set to .95. The numbers in parentheses are asymptotic standard errors. Specification 1 is the model that does not accommodate serially correlated wage error terms ($\rho = 0$). Specification 2 is the model that allows serial correlation ($\rho \neq 0$), σ_N is the standard deviation of V_t^N, and σ_E is the standard deviation of V_t^E if t is the first year in a new teaching job (see n. 30), σ_e is the standard deviation of e_{it} (see eq. 3) and $(\tau^2\pi^2)/6$ is the variance of the ϵs in eq. (1).

* Denotes an asymptotic t-ratio greater than 2.

teachers, utilizing more accurate wage information leads to better prediction of individual decisions.[34]

The coefficients in the wage equations measure the effect that increasing a particular variable has on real weekly log wages. The numbers in parentheses

[34] For example, individuals are more likely to remain in teaching jobs that have higher wages. When serial correlation is allowed, the imputed wages reflect whether previous wages at the job suggest that the job is a high-paying job or a low-paying job. This is not the case when wages are serially uncorrelated.

represent asymptotic standard errors. Not surprising, given the rigidity of the traditional teaching wage structure, the post-bachelor education level and the years of teaching experience are significant determinants of average teaching wages.[35] The negative coefficient on the year and the positive coefficient on the year-squared variable represent a significant U-shaped time trend for teaching wages. This is consistent with previous literature, which has shown that real teaching wages decreased substantially after 1972 until beginning a recovery in the early 1980s.[36]

Table 2 indicates that, on average, individuals with high SAT mathematics scores do not receive higher wages in teaching jobs but do receive significant wage premiums in nonteaching jobs. Thus, these estimates support the notion that lower teaching participation rates for academically gifted teachers stem in part from the ability of these teachers to obtain higher than average wages in the nonteaching sector but not in the teaching occupation.

The estimates in table 2 also reveal that, although males and females receive similar wages in teaching jobs, males have significantly higher-paying nonteaching opportunities. At first glance, this seems inconsistent with the finding in Section II that men have higher overall participation rates in teaching than women. One potential explanation of this discrepancy is that, because teaching has typically been dominated by females (especially in elementary schools), only men who are especially dedicated to teaching will choose to become certified. However, the fact that the male coefficient is insignificant in the nonpecuniary utility equations in table 2 suggests that behavioral differences between men and women are not largely driven by differences in this type of permanent factor. Instead, table 2 suggests that previously unavailable marital and fertility information is very important in explaining male-female differences. The "number of children" variable, which represents the effect of children for females, indicates that women with children are significantly less likely to be working in either teaching jobs or nonteaching jobs than women without children (i.e., they get negative nonpecuniary utility from either type of job relative to not working). The "children × male" variable is an interaction term that represents the additional effect of children for males. Not surprisingly, the effect of this interaction variable is large and posi-

[35] The models were also estimated allowing the experience variables to enter in a quadratic fashion. These specifications did not provide additional insight into the job-switching behavior of individuals. For example, teaching wages appear to increase linearly with teaching experience (i.e., the effect of squared teaching experience on teaching wages was small and insignificant).

[36] See U.S. Department of Education, National Center for Education Statistics, *The Condition of Education, 1994.* Since everyone in the sample graduated from high school in the same year, the effect of age cannot be differentiated from the time trend.

tive in both the teaching and nonteaching (nonpecuniary) utility equations. Therefore, for both specifications, the total effect of children for males (children + [children × male]) is small and insignificant in both the teaching and nonteaching, nonpecuniary equations. Similarly, being married makes a woman much less likely to work but has little effect on males. Thus, although the pecuniary opportunity costs of teaching are higher for males, women who are married or have children appear to have much higher nonpecuniary opportunity costs of working in general. As noted earlier, the marital and fertility variables are assumed to be exogenous. To the extent that individuals make labor supply decisions, marital decisions, and childbearing decisions jointly, this is incorrect. For example, if job unhappiness makes individuals more likely to leave work to have children, the effect of children on job exits would be overstated.[37]

Recall that the descriptive statistics in Section II indicated that the decisions of males and females are quite similar in the first years after certification but become quite different as the number of years after certification increases. The estimated effects of the marital and fertility variables suggest that the model can explain this finding. As shown in table 1, the majority of individuals in the sample are not married and do not have children when they initially become certified to teach. Thus, the model predicts very little difference in the behavior of men and women at the early stages of their careers. However, as families are created or enlarged in the years after certification, the model predicts that women become relatively less likely to work in either the teaching occupation or a nonteaching occupation and become more likely to be out of the workforce altogether. To more formally examine the extent to which the model is able to capture changes in the behavior of males and females that take place after certification, the estimated structural model was used to simulate the choices of the individuals in the sample, and these simulated choices were used to compute the simulated participation rates which correspond to the descriptive figures 1 and 2. Figure 6 shows that the model very accurately predicts the decreasing teaching participation rate that is observed over time for women and that it accurately predicts that this decrease takes place almost exclusively because of an increase in the proportion of women who are out of the workforce. Figure 7 shows that, for men, the model is able to successfully predict a decreasing participation rate in teaching, an increasing participation rate in nonteaching jobs, and a constant proportion of individuals who are out of the workforce.[38] However, the magnitudes of the predicted decrease in the teaching par-

[37] In this scenario, the people with children would be more likely to leave teaching even if they did not have children. Note that van der Klaauw (1996*b*) estimates a female labor supply model with an endogenously determined labor supply.

[38] For men, the changes in the participation rate in teaching and the participa-

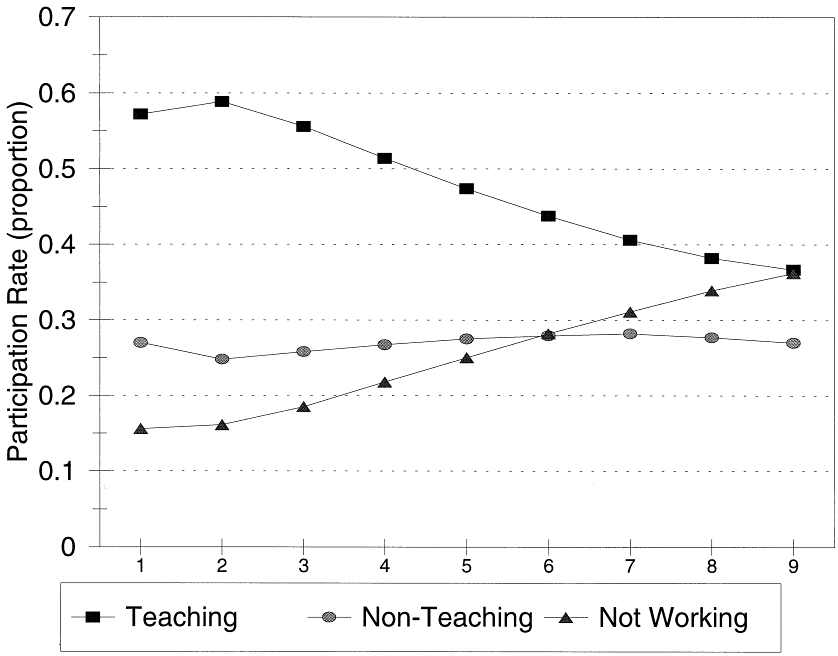

FIG. 6.—Simulated participation rates of females over time after certification

ticipation rate and the predicted increase in the nonteaching participation rate are somewhat smaller than what is observed in figure 2.

B. Specification Tests

The extent to which the two specifications fit the data can be examined using the residuals from the model. Let p_{ijt} be the probability that person i makes choice j in time t.[39] Let e_{ijt} represent the residual from the choice for the ith person at time t. Then, $e_{ijt} = 1 - p_{ijt}$ if person i chooses choice j at time t, and $e_{ijt} = 0 - p_{ijt}$ if person i does not choose choice j at time t. Let N represent the total number of residuals (choices) in the model, and let e represent the $N \times 1$ column vector of residuals. To get an idea of how well the models fit the data, a weighted sum of squared residuals (WSSR), $e^t\Omega^{-1}e$, can be computed, where the weighting matrix Ω is the

tion rate in nonteaching occupations is driven by the finding that teaching wages fell relative to nonteaching wages over the period of the sample.

[39] Recall that the upper case P_{it} was used earlier to represent the probability of the person's observed choice.

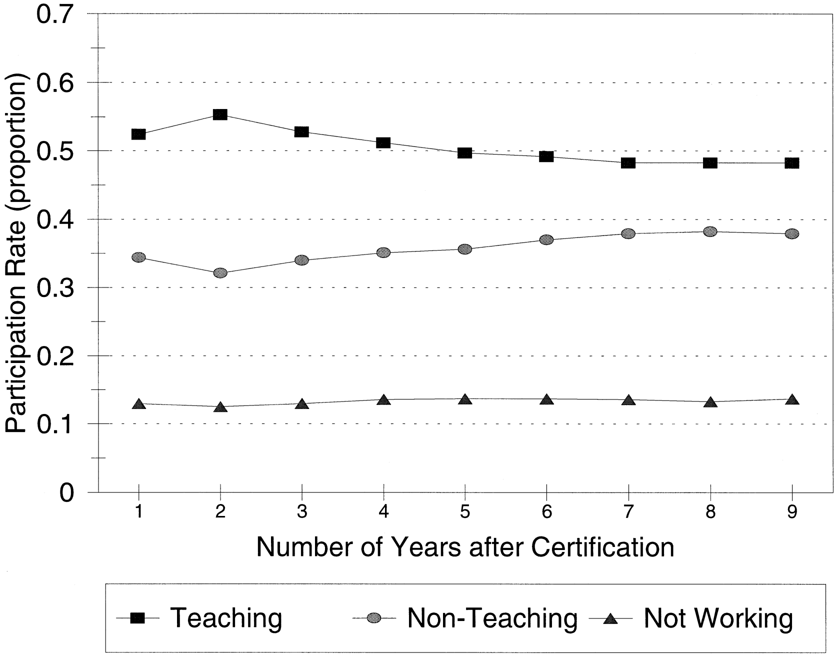

FIG. 7.—Simulated participation rates of males over time after certification

covariance matrix of the residuals.[40] It is reasonable to think of this statistic as having a chi-square distribution with $N = 9{,}835$ degrees of freedom. Large values of the test statistic (relative to the χ^2 distribution) would suggest that the model does not fit the data well.

The model specification that accommodates serial correlation can be compared with the model specification that does not accommodate serial correlation if the same weighting matrix Ω is used in both cases when computing equation (4). When the weighting matrix comes from the model that does not accommodate serial correlation, the WSSR is 7,928 for the model that allows serial correlation, and the WSSR is 8,959 for the model that does not allow serial correlation. When the weighting matrix comes from the model that accommodates serial correlation, the WSSR is 217,877 for the model that allows serial correlation, and the WSSR is 113,650,553 for the model that does not allow serial correlation. However, a very small number of outliers are responsible for almost the entire size of these statistics. When these outliers are removed, the WSSR's associated with the latter weighting matrix

[40] The row vector e is the transpose of e.

are reduced to 8,724 and 11,176, respectively.[41] Thus, the specification tests show that the specification that accommodates serial correlation performs substantially better than the specification that does not accommodate serial correlation and, with the exception of a very small number of outliers, it fits the data well.[42]

C. Policy Simulations

The sensitivity of individual decisions to wage changes is of policy importance. To illustrate the model's predictions with respect to this issue, the effects of two changes in the teacher wage structure are now examined. The first, denoted "policy 1," corresponds to the often heard normative statement that "teachers should earn more money." It involves a uniform wage increase of 25% for all teachers under the traditional, rigid wage structure. The second policy, "policy 2," explores the possibility of implementing a wage policy that deviates from the traditional wage structure. It raises the average teacher wage by 25% but does so by basing the amount of the raise on the ability level of the teacher. In particular, the wage increase that an individual receives depends linearly on the person's SAT score with the lowest SAT person receiving no additional wage. While it is not the intention of this article to argue that this specific policy would be optimal for schools, the evaluation of this policy does provide an illustration of the potential benefits of reducing the rigidity of the teaching wage structure in some manner.[43]

For each of the individuals in the data, it is necessary to simulate their choices for a particular wage policy. In each period (starting in the first period and proceeding forward) error terms are drawn from their estimated distributions, value functions are compared to find which choice the person would choose, and the state variables in the model (e.g., the

[41] Three outliers were removed for the specification with serial correlation and 10 outliers were removed for the specification without serial correlation. The unweighted sum of squared residuals, $e^t e$, is 1,458 for the model with serial correlation and 1,800 for the model without serial correlation.

[42] For example, to reject the null hypothesis that the model fits the data at the .05 level of significance, the WSSR must be greater than 10,066. It is often the case that goodness-of-fit tests of this type are rejected. For example, see Berkovec and Stern (1991).

[43] There is evidence that experiments with wage structures based on "merit" pay have not been traditionally successful. This is so largely because teachers perceive that the subjective evaluation process that determines pay increases is somewhat arbitrary and unfair (see e.g., Murnane and Cohen 1986; Murnane et al. 1991; Cohn 1996). However, if it is reasonable to believe that there is a "shortage" of academically gifted teachers in schools, relating pay to some sort of test scores would seem to be more similar to the more commonly accepted practice (suggested by Kershaw and McKean 1962) of paying higher wages to teachers in certain subject areas where shortages may be present.

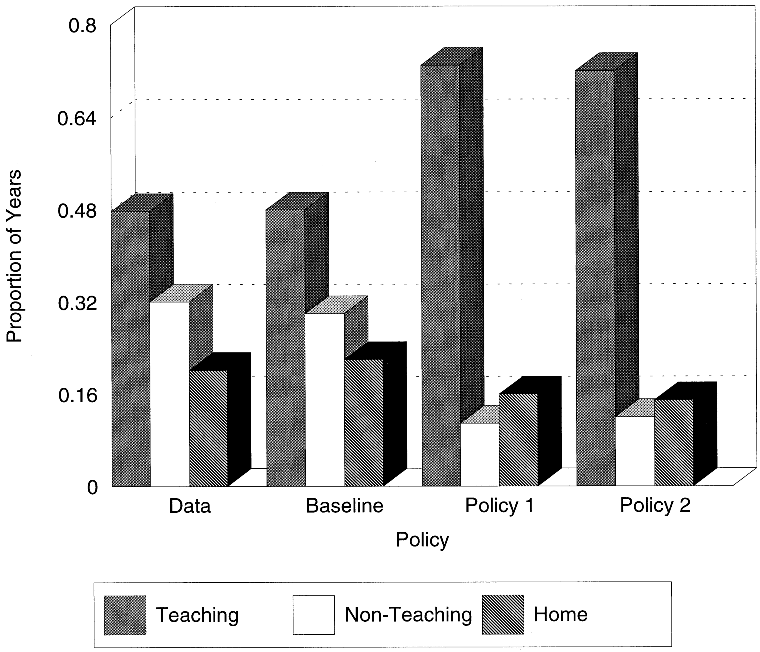

FIG. 8.—Proportion of aggregate years in each option for full sample for different policies

experience variables) are appropriately updated given the simulated choice. This simulation procedure is first performed on the unchanged set of estimates from the model to get a baseline set of results (these baseline simulations were used earlier to produce figures 6 and 7) and then performed on two sets of parameters that have been modified to be consistent with the desired policy changes.

The effects of policy 1 and policy 2 on the overall sample can be compared by examining several figures. Figure 8 shows that policy 1 and policy 2 have very similar effects on the proportion of time that individuals in the sample as a whole choose teaching; the aggregated proportion in teaching increases from .48 in the baseline simulation to .72 under both policy 1 and policy 2.[44] The increase in the proportion is due to two factors. First, the proportion of individuals who choose to teach in at least one period increases from .75 under the baseline simulation to .94 under both policy 1 and policy 2. Second, figure 9, which shows Kaplan-Meier survivor functions for the

[44] It should be noted that the exogeneity assumption for marital and fertility variables could lead to an understatement of responsiveness. For example, this would occur if women with higher wages would decide to have fewer children or would be less likely to be married.

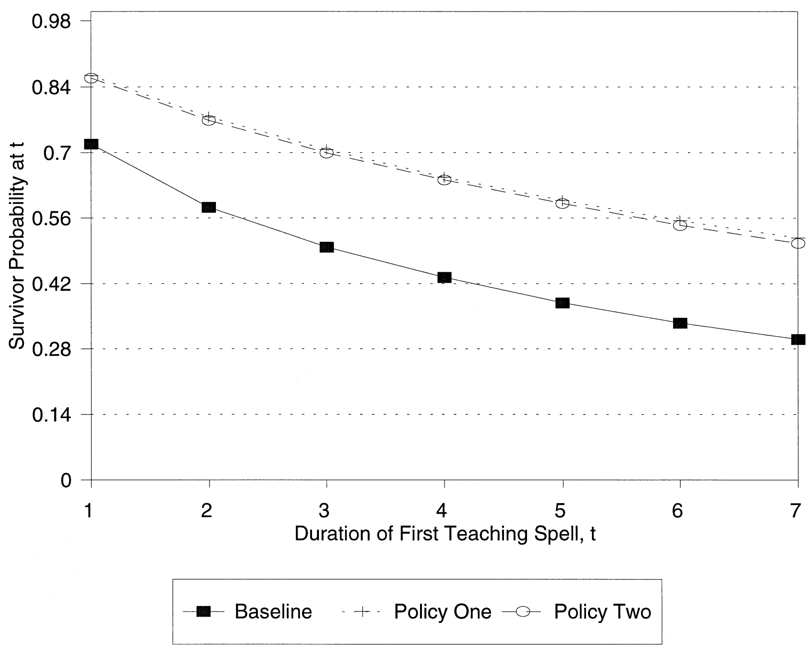

FIG. 9.—Survivor functions for full sample for different policies

overall sample, indicates that teachers who enter teaching have longer average spells under policy 1 and policy 2 than under the baseline simulation. Figure 8 shows that both policies lead to large reductions in the proportion of years spent in nonteaching jobs but have relatively little effect on the proportion of time spent out of the workforce. Thus, it is not surprising that males are more responsive to the policies than females. The proportion of years spent teaching for males increases from .51 in the baseline simulation to .80 under both policy 1 and policy 2. The proportion of years spent teaching for females increases from .47 in the baseline simulation to .70 under policy 1 and .69 under policy 2.

Despite the preceding similarities between the effects of policy 1 and policy 2, one must be careful in concluding that these two policies have identical overall benefits. Figure 10 shows a measure of the labor supply of teachers with SAT scores in the top one-third of the sample relative to the labor supply of teachers in the lower two-thirds of the sample for the data, the baseline simulation, policy 1, and policy 2. In particular, it shows the percentage of periods that each option is chosen by the low-ability group divided by the percentage of periods that the option is chosen by the high-ability group. For example, the ratio for the teaching option under the unchanged (baseline) wage structure, .88, is found by dividing the teaching proportion of the high-ability group, .44, by the teaching

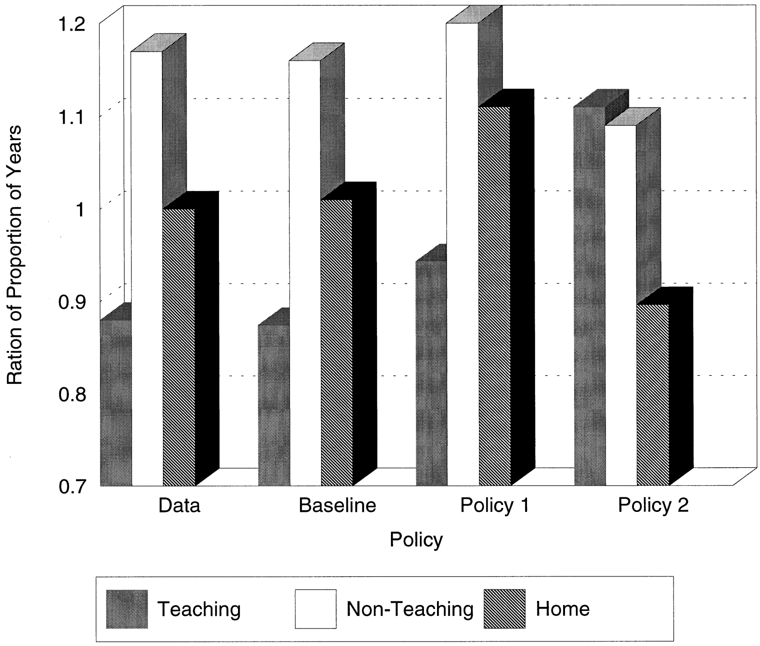

FIG. 10.—Relative labor supply of high-SAT group. Figure shows the proportion of aggregate years that individuals in the low-SAT group choose a particular option divided by the proportion of aggregate years that individuals in the high-SAT group choose the option.

proportion of the low-ability group, .50. The ratio for policy 1, .94, suggests that the uniform wage increase leads to a somewhat small increase in the relative labor supply of high-ability teachers. However, policy 2 is much more successful from this standpoint. The ratio, 1.11, suggests that high-ability teachers would choose to teach more often than low-ability teachers under this policy.[45]

V. Conclusion and Policy Implications

In this article, data from a general longitudinal survey and the use of a dynamic, discrete choice model allow analyses of teacher attrition that were not previously possible. The article examines both the labor supply of the overall sample of teachers and the labor supply of teachers of varying academic ability.

In terms of overall teacher labor supply, descriptive statistics show that

[45] Relative to policy 1, policy 2 leads to a larger proportion of high-ability people who choose to teach in at least one period, and it leads to longer durations for the high-ability individuals that enter teaching.

the proportion of individuals who choose to teach decreases significantly over time after certification even after controlling for geographical mobility and multiple spells in teaching. Previously unavailable information about the labor activities of individuals when they are not teaching suggests that it may be incorrect to believe that the primary cause of teacher attrition is the relative attractiveness of nonteaching occupations. Instead, the decrease in the teaching participation rate over time after certification is accompanied by large increases in the proportion of individuals who are not working. The structural model suggests that marital status and number of children are very important predictors of exits out of the workforce for women. Thus, because women who have been certified for a longer period of time are more likely to be married and tend to have more children, these variables provide an explanation for much of the decrease in the teaching participation rate that occurs over time after certification. These findings suggest the importance of future research that examines whether the important role that is found for family variables has changed since the end of the NLS-72 sample period. From the standpoint of predicting future teacher shortages, these findings also suggest the importance of future studies that examine the extent to which female teachers who leave teaching for family reasons return to teaching at a later time (e.g., when children enter school).[46] The utility-maximizing model is conducive to analyzing the sensitivity of teacher labor supply to wage increases. Simulations suggest that the overall labor supply of currently certified teachers is responsive to wage increases.

Descriptive statistics in the article show that individuals with higher academic ability, as measured by scores on the SAT examination, teach in a smaller proportion of years than other teachers. Estimates of the wage equations in the model support the notion that lower teaching participation rates for academically gifted teachers stem in part from the ability of these teachers to obtain wage premiums in the nonteaching sector but not in the teaching occupation. From the standpoint of improving the ability composition of the teaching workforce, simulations suggest that policies that reduce the rigidity of the teaching wage structure by allowing wage increases to be correlated with the opportunity costs of teachers are likely to be more promising than traditional, uniform wage increases.

Although this article does not examine an individual's certification decision, influencing the set of individuals who choose to become certified is likely to be a primary motivation for raising the wages of teachers. It would seem to be quite likely that wage increases would be successful from this

[46] Stinebrickner (1999*c*) finds that relatively few teachers in the NLS-72 who leave teaching following the birth of a child return in the next several years. However, the data do not allow an examination of whether these teachers return when their children enter school.

standpoint if the goal is simply to increase the pool of applicants. From the standpoint of increasing the number of academically gifted individuals who are teaching, a particular wage increase will tend to be successful if it either has beneficial consequences on the ability composition of those seeking certification or if it increases the number of certified teachers who are looking for jobs and schools hire the best applicants.[47] However, if encouraging academically gifted teachers to seek certification and enter teaching is the primary purpose of a wage increase, it is important to note that the decisions of individuals who are currently teaching will be important in determining the amount of time that will be needed before turnover of current teachers occurs and the new teachers can be hired into the schools. The simulations in this article suggest that a substantial increase in wages would significantly slow the rate at which current teachers leave.

Appendix

Approximation Methods to Accommodate Serially Correlated Wage Errors

Both years of teaching experience and years of nonteaching experience are state variables. In addition, because the teaching wage error term, ν_t^E, is assumed to be serially correlated, its current value provides information about the distribution of ν_{t+1}^E, and it must also be considered a state variable. Given the recursive nature of the value functions in equation (2), value functions are traditionally solved by backward recursion techniques for all combinations of the state variables that arise in a particular period.

An approximation method is necessary in this case because ν_t^E is a continuous state variable. For illustration, consider computing the value at time t of teaching in a job that was started at time b:

$$\bar{V}^E(t,\ Y_t,\ \nu_t^{Eb}) = U_t^E - \varepsilon_t^{Eb} + \beta E \max[\{V_{t+1}|d_t = Eb\}], \qquad \text{(A1)}$$

where the value function has been written as an explicit function of a vector of the two experience variables, Y_t, and the wage error, ν_t^{Eb}, associated with the teaching job. The starting date of the teaching job, b, is used here to differentiate between a teaching job that was accepted in the past and a new teaching opportunity that arrives in the current period. The difficulty is in computing the last term in equation (A1), which represents the discounted expected future utility. The expectation in this term is with respect to all stochastic components of the value functions in the set $\{V_{t+1}|d_t = Eb\}$, whose time $t + 1$ realizations are not known at time t. If the person is currently teaching in period t (i.e., $d_t = Eb$, where $b \leq t$), she will have two teaching options, one nonteaching job option,

[47] Ballou (1996) finds that schools may not necessarily hire the most qualified applicants. If teachers are selected randomly from the pool of applicants, simply having a larger pool of applicants will not increase quality.

and the option of not working at time $t + 1$. Thus, computing $E\max[\{V_{t+1}|d_t = Eb\}]$ requires integrating $\max[\{V_{t+1}|d_t = Eb\}]$ over the distributions of the nonpecuniary unobservables, ε_{t+1}^{Eb}, ε_{t+1}^{Et+1}, ε_{t+1}^{N}, ε_{t+1}^{H}, and the pecuniary unobservables, ν_{t+1}^{Eb}, ν_{t+1}^{Et+1}, and ν_{t+1}^{N}. Given the assumption that the ε's are i.i.d. extreme value, it is well known that, conditional on the pecuniary errors, the expectation will have a closed-form solution. Denote this conditional expectation as $EZ(t + 1, Y_{t+1}|\nu_{t+1}^{EB}, \nu_{t+1}^{Et+1}, \nu_{t+1}^{N})$. Then the unconditional expectation is given by

$$E\max[V_{t+1}|d_t| = Eb]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} EZ(t,\bullet|\nu_{t+1}^{Eb}, \nu_{t+1}^{Et+1}, \nu_{t+1}^{N})\,g(\nu_{t+1}^{Eb}, \nu_{t+1}^{Et+1}, \nu_{t+1}^{N}) \times d\nu_{t+1}^{Eb}\, d\nu_{t+1}^{Et+1}\, d\nu_{t+1}^{N}, \qquad \text{(A2)}$$

where g is the joint distribution of the time $t + 1$ wage errors (in this article, the errors are assumed to be independent of each other). Using Hermite Gaussian quadrature integral approximation methods, this integral can be approximated as

$$\approx \frac{1}{\pi^{1.5}}\sum_{k=1}^{p} a_k \sum_{y=1}^{p} a_y \sum_{z=1}^{p} a_z EZ^E(t, \bullet|\sqrt{2}\sigma_e m_z + \rho\nu_t^{Eb}, \sqrt{2}\sigma_E m_y, \sqrt{2}\sigma_N m_k), \qquad \text{(A3)}$$

where $a_1, \ldots, a_p$ are the quadrature weights and $m_1, \ldots, m_p$ are the quadrature points associated with the Hermite approximation (see Stroud and Secrest 1966). Thus, approximating $\bar{V}^E(t, Y_t, \nu_t^{Eb})$ requires the knowledge of $\bar{V}^E(t + 1, Y_{t+1}, \nu_{t+1}^{Eb})$ for the p values of ν_{t+1}^{Eb} in the set $\{\sqrt{2}\sigma_e m_z + \rho\nu_t^{Eb}: z = 1, 2, \ldots, p\}$. A repetition of this argument reveals that the calculation of $\bar{V}^E(t + 1, Y_{t+1}, \nu_{t+1}^{Eb})$ for each of these p values of ν_{t+1}^{Eb} requires the knowledge of $\bar{V}^E(t + 2, Y_{t+2}, \nu_{t+2}^{Eb})$ for p values of ν_{t+2}^{Eb}. In general, the number of values of ν_{t+i}^{Eb} that must be calculated in order to calculate the value function $\bar{V}^E(t, Y_t, \nu_t^{Eb})$ grows exponentially with i and is given by $(p)^i$.[48]

Because the number of values of the teaching wage error that must be calculated is extremely large in some periods, the amount of computer time necessary to solve value functions for all possible combinations of the state variables by the standard backward method becomes infeasible.

[48] Thus, the set of possible teaching wage errors at time t depends in part on the set of possible teaching wage errors in time $t - 1$. It also depends on certain estimation issues, such as the wage errors that are observed in accepted jobs at time t and the appropriate distributions of wage errors needed to adjust for all rejected wage offers and any unobserved accepted wage offers.

The interpolation algorithm that is implemented to deal with this problem involves, for each time t, specifying the range of error terms $[\nu_t^L, \nu_t^U]$ that could arise in each period t. This interval is then discretized into a finite set, R_t^ν, of "grid points" that are equally spaced a distance of Δ apart:

$$R_t^\nu = \{[\nu_t^U - \nu_t^L]/2 \pm j^*\Delta: j = 1, 2, \ldots, B\}, \tag{A4}$$

where B is the smallest integer value of B^* for which $[\nu_t^U - \nu_t^L]/2 + B^*\Delta > \nu_t^U$.

Once the grid points have been determined for all time periods, value functions can be solved for all combinations of the grid points and possible values of Y_t (for all time periods) by backward recursion. However, to do this requires some interpolation. Suppose that the value of teaching, $\bar{V}^E(t, Y_t, r_t^\nu)$, is being solved for a particular value of Y_t and some grid point $r_t^Y \in R_t^\nu$. Then as discussed earlier, the calculation of $\bar{V}^E(t, Y_t, r_t^\nu)$ requires the knowledge of $\bar{V}^E(t + 1, Y_t + 1, \nu_{t+1}^{Et})$ for the p values of ν_{t+1}^{Et}, given by $\{\sqrt{2}\sigma_e m_z + \rho r_t^\nu: z = 1 \ldots p\}$.[49] However, because during the $t + 1$ stage of the backward recursion process value functions will not have been calculated for all values of the wage error that could arise at $t + 1$ (rather, only a much smaller set of grid points R_{t+1}^ν), $\bar{V}^E(t + 1, Y_{t+1}, \nu_{t+1}^{Et})$ will not have been calculated for any of the p values in this set. Let $\hat{\nu}_{t+1}^{Et}$ represent any one of the p values of ν_{t+1}^{Et}. Then $\bar{V}^E(t + 1, Y_t + 1, \hat{\nu}_{t+1}^{Et})$ is interpolated as a weighted average of $\bar{V}^E(t + 1, Y_t + 1, r_{t+1}^{\nu,1})$ and $\bar{V}^E(t + 1, Y_t + 1, r_{t+1}^{\nu,2})$, where $r_{t+1}^{\nu,1}$ and $r_{t+1}^{\nu,2}$ are the closest two grid points to $\hat{\nu}_{t+1}^{Et}$ such that $r_{t+1}^{\nu,1} < \hat{\nu}_{t+1}^{Et} < r_{t+1}^{\nu,2}$.[50]

Notice that through a reduction of the spacing, Δ, between grid points, $r_{t+1}^{\nu,1}$ and $r_{t+1}^{\nu,2}$, can be made arbitrarily close to $\hat{\nu}_{t+1}^{Et}$. Thus, this source of interpolation error can be made arbitrarily small. Stinebrickner (1996) shows that by reducing Δ and increasing the number of quadrature points, p, the approximated value functions converge to the true value functions. Testing in Stinebrickner (1996) shows that the approximation method works well for relatively large values of Δ and small values of p. For example, on average, each parameter obtained using $p = 3$ (as in this article) differs from its counterpart obtained using $p = 6$ by only .03 of its standard error. On average, each parameter estimate obtained using $\Delta = .4$ (as in this case) differs from its counterpart obtained using $\Delta = .05$ by .09 of its standard error.

Once value functions are solved in all time periods for all combinations

[49] It also requires $\bar{V}^E(t + 1, r_t^Y + 1, \nu_{t+1}^{Et+1})$ for the p values of ν_{t+1}^{Et+1} given by $\{\sqrt{2}\sigma_E m_y: y = 1 \ldots p\}$.

[50] The approximation method used here is similar in spirit to that of Keane and Wolpin (1994), who compute the expected future utility component of value functions for a subset of the possible time $t + 1$ state points and use these computed values to fit a single interpolating regression that can provide an approximate value for the expected future utility associated with any other time $t + 1$ state point.

of Y and the elements of R_t^v, the likelihood function can be evaluated. As mentioned earlier, conditional on any unobserved wages and unobserved heterogeneity, the likelihood contribution at time t for person i is the joint probability of the choice that is made and any wage that is observed. This can be written as the product of the probability of the observed wage and the choice probability conditional on the observed wage. Computing the choice probability associated with a particular observed wage requires the knowledge of the value function evaluated at the wage error that is determined by the observed wage. This value function can be approximated in a manner similar to that described above in the solution process.

References

Ayers, Jerry, and Qualls, Glenda. "Concurrent and Predictive Validity of the National Teacher Examinations." *Journal of Educational Research* 73 (November/December 1979): 86–92.

Ballou, Dale. "Do Public Schools Hire the Best Applicants?" *Quarterly Journal of Economics* 111, no. 1 (February 1996): 97–133.

Berkovec, James, and Stern, Steven. "Job Exit Behavior of Older Men." *Econometrica* 59, no. 1 (January 1991): 189–210.

Bishop, John. "Incentives to Study and the Organization of Secondary Instruction." In *Assessing Educational Practices,* edited by William Becker and William Baumol. Cambridge, MA: MIT press, 1996.

Brewer, Dominic. "Career Paths and Quit Decisions: Evidence from Teaching." *Journal of Labor Economics* 14, no. 2 (April 1996): 313–39.

Cohn, E. "Methods of Teacher Remuneration: Merit Pay and Career Ladders." In *Assessing Educational Practices,* edited by W. Becker and W. Baumol. Cambridge, MA: MIT Press, 1996.

Dolton, Peter, and van der Klaauw, Wilbert. "Leaving Teaching in the UK: A Duration Analysis." *Economic Journal* 105 (1995): 431–44.

———. "The Turnover of UK Teachers: A Competing Risk Explanation." *Review of Economics and Statistics* 81, no. 3 (August 1999): 543–49.

Eberts, Randall. "Union-Negotiated Employment Rules and Teachers Quits." *Economics of Education Review* 6, no. 1 (1987): 15–25.

Eckstein, Zvi, and Wolpin, Kenneth I. "Dynamic Labor Force Participation of Married Women and Endogenous Work Experience." *Review of Economic Studies* 65, no. 3 (1989): 375–90.

Ehrenberg, Ronald, and Brewer, Dominic. "Did Teachers' Verbal Ability and Race Matter in the 1960's?" *Economics of Education Review* 14, no. 1 (1995): 1–21.

Ferguson, Ronald F. "Racial Patterns in How School and Teacher Quality Affect Achievement and Earnings." *Challenge: A Journal of Research on Black Men* 2, no. 1 (May 1991): 1–20.

Gritz, Mark R., and Theobald, Neil D. "The Effects of School District Spending Priorities on Length of Stay in Teaching." *Journal of Human Resources* 31, no. 3 (Summer 1996): 477–512.

Hanushek, Eric A. "Teacher Characteristics and Gains in Student

Achievement: Estimation Using Micro-Data." *American Economic Review* 61, no. 2 (1971): 280–88.

Hanushek, Eric A.; Rivkin, Steven; and Kain, John F. "Teachers, Schools and Academic Achievement." Working Paper no. 6691. Cambridge, MA: National Bureau of Economic Research, 1998.

Hussar, William J., and Gerald, Debra E., *Projections of Education Statistics to 2007.* Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 1997.

Keane, Michael P., and Wolpin, Kenneth I. "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence." *Review of Economics and Statistics* 76, no. 4 (November 1994): 648–72.

Kershaw, Joseph A., and McKean, Roland N. *Teacher Shortages and Salary Schedules.* New York: McGraw-Hill, 1962.

Meitzen, Mark. "Differences in Male and Female Job-Quitting Behavior." *Journal of Labor Economics* 4, no. 2 (April 1986): 151–67.

Monk, David. "Subject Area Preparation of Secondary Mathematics and Science Teachers and Students Achievement." *Economics of Education Review* 13, no. 2 (1994): 125–45.

Mont, Daniel, and Rees, Daniel. "The Influence of Classroom Characteristics on High School Teacher Turnover." *Economic Inquiry* 34, no. 1 (January 1996): 152–67.

Murnane, Richard J. *The Impact of School Resources on the Learning of Inner City Children.* Cambridge, MA: Ballinger, 1975.

Murnane, Richard, and Cohen, David. "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." *Harvard Education Review* 56 (1986): 1–17.

Murnane, Richard J., and Olsen, Randall J. "The Effects of Salaries and Opportunity Costs on Duration in Teaching: Evidence from Michigan." *Review of Economics and Statistics* 71, no. 2 (May 1989): 347–52.

———. "The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from North Carolina." *Journal of Human Resources* 25 (Winter 1990): 106–24.

Murnane, Richard J., and Phillips, Barbara R. "What do Effective Teachers of Inner City Children Have in Common?" *Social Science Research* 10 (1981): 83–100.

Murnane, Richard J.; Singer, Judith D.; and Willett, John B. "The Influences of Salaries and Opportunity Costs on Teachers' Career Choices: Evidence from North Carolina." *Harvard Educational Review* 59 (1989): 345–46.

Murnane, Richard J.; Singer, Judith D.; Willett, John B.; Kemple, James J.; and Olsen, Randall J. *Who Will Teach? Policies That Matter.* London: Harvard University Press, 1991.

Murnane, Richard J.; Willett, John B.; and Levy, Frank. "The Growing Importance of Cognitive Skills in Wage Determination." *Review of Economics and Statistics* 77, no. 2 (May 1995): 251–66.

Rust, John. “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher.” *Econometrica* 55, no. 5 (September 1987): 999–1033.

Stinebrickner, Todd R. “An Empirical Investigation of Teacher Attrition.” *Economics of Education Review* 17, no. 2 (April 1998): 127–36. (*a*)

———. “Serially Correlated Wages in Dynamic, Discrete Choice Models.” Unpublished manuscript. London: University of Western Ontario, 1998. (*b*)

———. “Estimation of a Duration Model in the Presence of Missing Data.” *Review of Economics and Statistics* 81, no. 3 (August 1999): 529–42. (*a*)

———. “Using Latent Variables in Dynamic, Discrete Choice Models: The Effect of School Characteristics on Teacher Decisions.” In *Research in Labor Economics,* vol. 18, edited by Solomon Polachek. Greenwich, CT: JAI Press, 1999. (*b*)

———. “An Analysis of Occupational Change and Departure from the Labor Force: Evidence of the Reasons That Teachers Quit.” Working paper. London: University of Western Ontario, 1999. (*c*)

Strauss, Robert P., and Sawyer, Elizabeth A. “Some New Evidence on Teacher and Student Competencies.” *Economics of Education Review* 5, no. 1 (1986): 41–48.

Stroud, A. H., and Secrest, Don. *Gaussian Quadrature Formulas.* Englewood Cliffs, NJ: Prentice Hall, 1966.

Theobald, Neil D. “An Examination of the Influence of Personal, Professional, and School District Characteristics on Public School Teacher Retention.” *Economics of Education Review* 9, no. 3 (1990): 241–50.

Theobald, Neil D., and Gritz, Mark R. “The Effects of School District Spending Priorities on the Exit Paths of Beginning Teachers Leaving the District.” *Economics of Education Review* 15, no. 1 (February 1996): 11–22.

U.S. Department of Education, National Center for Education Statistics. *The Condition of Education.* Washington DC: U.S. Department of Education, 1994.

U.S. Department of Education, *Here Come the Teenagers: A Back to School Report on the Baby Boomers.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, Educational Resources Information Center, 1997.

van der Klaauw, Wilbert. “Expectations and Career Decisions: An Analysis of Teaching Careers Using Expectations Data.” Unpublished manuscript. New York: New York University, 1996. (*a*)

———. “Female Labor Supply and Marital Status Decisions: A Life Cycle Model.” *Review of Economic Studies* 63 (April 1996): 199–236. (*b*)

Weiss, Andrew. “Determinants of Quit Behavior.” *Journal of Labor Economics* 2 (July 1984): 371–87.