Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets

Donald Boyd
University at Albany, SUNY

Hamilton Lankford
University at Albany, SUNY

Susanna Loeb
Stanford University

James Wyckoff
University at Albany, SUNY

September 2006

(revision of NBER working paper 9878)

We appreciate the many helpful comments provided by John Jones, Michael Sattinger and workshop participants at Duke, Michigan, NBER, Stanford, Syracuse, Union College, SUNY-Albany, UC-San Diego, and UC-Santa Cruz. Special thanks go to Charles Peck for his able consulting regarding many programming issues, especially the parallelization of computer code. The Smith Richardson Foundation, the Office of Educational Research and Improvement, U.S. Department of Education, and the New York State Department of Education provided financial support, but do not necessarily support the views expressed in this paper. Any errors are attributable to the authors.

Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets

Abstract

Although there is growing recognition of the contribution of teachers to students' educational outcomes, there are large gaps in our understanding of how teacher labor markets function. Low-income, low-achieving and non-white students, particularly those in urban areas, often are taught by the least skilled teachers. This sorting of teachers to jobs likely contributes to the substantial gaps in academic achievement among income and racial/ethnic groups of students. The objective of this paper is to develop and estimate a model that identifies the key factors explaining the allocation of teachers to jobs. The approach is based on a game-theoretic two-sided matching model and the estimation strategy employs the method of simulated moments. With this combination, we estimate how a range of factors affects the choices of individual teachers and hiring authorities, as well as how these choices interact to determine the equilibrium allocation of teachers across jobs. We find that employers demonstrate preferences for teachers with stronger academic achievement and for non-white teachers, while teachers show preferences for schools that are closer geographically, offer higher salaries, are suburban, and have a smaller proportion of students in poverty and of a different race. While these results may appear predictable, they contradict many prior wage equation estimates for teacher labor markets, possibly because important features of these markets are inconsistent with key assumptions of the wage equation framework. Although the paper focuses on worker-job match within teacher labor markets, many of the issues raised and the empirical framework employed are relevant in other settings as well.

I. Introduction

The 3.1 million elementary and secondary public school teachers in the United States make up more than 8.5 percent of all college-educated workers 25 to 64 years old.[1] Even though there is growing recognition of the contribution of these teachers to students' educational outcomes and later economic success, large gaps exist in our understanding of how teacher labor markets function. Most research on teacher labor markets has used models developed for the private sector. However, markets for public school teachers, as well as markets for many other public employees, differ in fundamental ways from those in the private sector. The objective of this paper is to develop and estimate a model that more accurately characterizes the institutional features of teacher labor markets. The approach is based on a game-theoretic two-sided matching model and the estimation strategy employs the method of simulated moments. With this combination, we estimate how factors affect the choices of individual teachers and hiring authorities, as well as how these choices interact to determine the equilibrium allocation of teachers across jobs.

Low-income, low-achieving and non-white students, particularly those in urban areas, often are taught by the least skilled teachers,[2] a factor that likely contributes to the substantial gaps in academic achievement among income and racial/ethnic groups of students. Such sorting of teachers across schools and districts is the result of a range of decisions made by individual teachers and school officials. Inefficient hiring and district assignment may contribute to the disparities in teacher qualifications across schools; however, teacher preferences are likely to be

[1] Digest of Education Statistics 2004 and U.S Census Bureau Educational Attainment in the United States 2000 Detailed Tables.
[2] For example, see Lankford, Loeb and Wyckoff (2002).

particularly influential.[3] Teachers differ fundamentally from other school resources. Unlike textbooks, computers, and facilities, teachers have preferences about whether to teach, what to teach, and where to teach. Salaries are one job attribute that likely affects sorting, but non-pecuniary job characteristics, such as class size, preparation time, facilities, or characteristics of the student body, are important as well.[4] A large literature suggests that teachers respond to wages,[5] yet research on the compensating wage differentials needed to attract teachers with particular attributes to schools with particular characteristics has not produced consistent results.

Applications of hedonic models to estimate compensating differentials in a wide variety of other labor markets also have yielded counter-intuitive results. Researchers have offered a number of explanations, including omitted variables, measurement error, simultaneity and market frictions. Each of these could explain the counter-intuitive estimates in the case of teacher labor markets. However, the inconsistencies, at least in part, could be the result of either the thinness of local teacher labor markets or features of teacher labor markets inconsistent with the hedonic framework. For example, hedonic models maintain that wages or other attributes of jobs and workers together are sufficiently flexible to assure that demand equals supply at each combination of worker and job attributes. However, this may not always hold for the sorting of new teachers across job openings leading up to the start of a school year. Salaries typically are

[3] Few studies have explored district-hiring practices, though Pflaum & Abramson (1990), Ballou (1996) and Ballou and Podgursky (1997) do provide evidence that many districts are not hiring the most qualified candidates. Schools also vary in the political power they exert, which may lead to differences in teacher qualifications among schools within the same district. Bridges (1996) found that when parents and students complained about poor teachers, the teachers were likely to be transferred to schools with high student-transfer rates, large numbers of students receiving free or reduced-price lunches, and large numbers of minority students.

[4] In Texas, Hanushek, Kain and Rivkin (1999) found teachers moving to schools with high-achieving students and, in New York City, Lankford (1999) found experienced teachers moving to high-socioeconomic status schools when positions became available.

[5] As a group, these studies show that individuals are more likely to choose to teach when starting teacher wages are high relative to wages in other occupations (Baugh and Stone, 1982; Brewer, 1996; Dolton, 1990; Dolton and van der Klaaw, 1999; Dolton and Makepeace, 1993; Hanushek and Pace, 1995; Manski, 1987; Mont and Reece, 1996; Murnane, Singer & Willett, 1989; Rickman and Parker, 1990; Stinebrickner, 1998, 1999, 2000; Theobald, 1990; Theobald and Gritz, 1996). Baugh and Stone (1982), for example, find that teachers are at least as responsive to wages in their decision to quit teaching, as are workers in other occupations.

fixed during this time, reflecting the collective bargaining agreements in place, and other attributes of jobs and workers may not be sufficiently flexible in the short-run to clear the market. The certification process slows the entry of new teachers into the market when there are market shocks. In the absence of market clearing, there will be workers or employers, or both, for whom the marginal evaluations of at least some of the attributes characterizing their job-match partners will differ from the marginal prices they face implicit in the wage function, violating a basic feature of the hedonic framework.

Market thinness as well as the absence of market clearing can lead to complex interdependencies in the choices made by job candidates and employers. We can analyze such an environment in a relatively straightforward manner using the standard two-sided matching model extensively studied by game theorists (Roth and Sotomayer, 1990). The contributions of this paper are to note the conceptual applicability of the game-theoretic, two-side matching model as an attractive alternative to the standard model, and to show how the underlying preferences of job candidates and employers in such a two-sided matching model can be estimated using the method of simulated moments and data characterizing observed job-worker matches.

We find that employers demonstrate preferences for teachers with stronger academic achievement and for non-white teachers, while teachers show preferences for schools that are closer geographically, offer higher salaries, are suburban, and have a smaller proportion of students in poverty and of a different race. As predictable as these results are, they differ from wage-equation estimates that are often used to suggest, for example, that employers do not value teacher skills and that teachers do not value salary.

The following two sections of the paper briefly summarize the data we employ and some key features of teacher labor markets. Section IV contrasts the hedonic approach with two alternative models of job match. We outline our conceptual framework and empirical approach in section V and discuss identification in section VI. Section VII presents estimates of several models, as well as simulations of policy changes based on the results of these models. Section VIII reports estimates of hedonic wage equations and discusses simulation results that help to clarify the difference between the wage-equation approach and our model. The final section concludes.

II. Data

The data we use for this analysis comes from a larger database of teachers and schools that links seven administrative datasets and various other information characterizing schools, districts, communities, and local labor markets in New York State. It includes information for every teacher and administrator employed in a New York public school at any time from 1969-70 through 1999-2000. The core data comes from the Personnel Master File (PMF), part of the Basic Education Data System of the New York State Education Department. In a typical year there are approximately 200,000 teachers identified in the PMF. We have linked these annual records through time, yielding detailed data characterizing the career history of each individual.

Several other databases that contain a range of information about the qualifications of prospective and actual teachers, as well as the environments in which these individuals make career decisions, substantially enrich this core data. For teachers this information includes age, gender, race/ethnicity, salary, experience (in the district, in NYS public schools, and total), years of education and degree attainment, and teacher certification exam scores of individual teachers and whether they passed on their first attempts. In addition, we identify the institutions from which individual teachers earned their undergraduate degrees and combine it with the Barron's ranking of college selectivity to construct variables measuring the selectivity of the college from which each teacher graduated and the location of the institution. Measures for schools and districts include enrollment, student poverty, racial composition, and district salary schedules, as well as many other measures. Using information on the zip code of residence when each teacher applied for certification and the zip code of each school, we create a “distance from home” measure for each school-teacher combination in our sample. For a sub-sample of teachers we know where they lived while in high school.

Our data is richer in its descriptions of teachers than other administrative datasets used to date, particularly in that it includes teachers' certification test scores and undergraduate institutions. It also allows us to match teachers to characteristics of the schools in which they teach in a way that most national longitudinal surveys, such as High School and Beyond or the National Longitudinal Survey of Youth, do not. This matching of employer and employee data has proved useful in the analysis of labor markets more generally (Abowd and Kramarz, 1999; Postel-Vinay and Robin, 2002; Rosen, 1986).

III. Features of Teacher Labor Markets

We have used the New York teacher data to document various aspects of teacher labor markets, a number of which are pertinent here. First, as noted above, there is a marked sorting of teachers across schools. For example, in schools in the highest quartile of student performance on the New York State 4^{th} Grade English Language Arts Exam, only three percent of teachers are uncertified, only ten percent earned their undergraduate degree from least competitive colleges, and only nine percent of those who have taken a general knowledge teacher certification exam failed.[6] In contrast, in schools in the lowest quartile of student performance, 22 percent of teachers are uncertified, 26 percent come from least competitive colleges, and 35 percent have failed a general-knowledge certification exam (Lankford, Loeb and Wyckoff, 2002). We find similar patterns when schools are grouped based on student poverty, race/ethnicity and limited English proficiency. These differences reflect urban-suburban differences in the qualifications of teachers as well as meaningful differences across schools within urban areas.

Differences in the qualifications of teachers are the result of the decisions of individuals and school officials that determine initial job matches and subsequent decisions that affect job quits, transfers and terminations. Of these, initial job matches appear particularly important in that they account for almost all of the urban-suburban differences in teacher qualifications as well as approximately half of the differences between schools within urban districts (Boyd, Lankford, Loeb, and Wyckoff, 2002). We focus on these initial job matches and the sorting of teachers within local labor markets.

We have also found that a surprisingly large number of individuals take their first teaching job very close to where they grew up. Over 60 percent of teachers first teach within 15 miles of the high school from which they graduated and 85 percent teach within 40 miles. Even of those who travel over 100 miles to college, most return home to teach (Boyd, Lankford, Loeb, and Wyckoff, 2005a). This proximity has two important implications for modeling the sorting of teachers across jobs. First, most teachers make job choices within a very limited geographic area. Because of this, our empirical analysis of job match presented below focuses on the matching of teachers to jobs within relatively small geographic areas (metropolitan areas) instead of across the entire state. Second, even within each of these local labor markets, work proximity

[6] Teachers in New York have had the option of taking the NTE General Knowledge Exam or the NYSTCE Liberal Arts and Science Exam. Throughout the paper "failure" refers to failing one of these exams on the first attempt.

is likely to affect teachers' rankings of alternative job opportunities. Teachers will rank otherwise identical jobs differently because of differences in the relative proximity of jobs to the teachers' own locations. These ranking differences suggest that an accurate model of teacher labor markets will need to incorporate this potentially important source of preference heterogeneity.

Other institutional features of teacher labor markets are pertinent as well. For example, the annual hiring cycle for teachers is such that most job openings are filled over several months leading up to the start of the school year. During this period, the total number of teaching positions in a local labor market is largely predetermined, reflecting enrollment levels and choices made by school officials regarding budgetary, programmatic and other policy matters (e.g., class size policies) – decisions typically made prior to the start of the hiring season. In turn, the net number of openings to be filled by individuals not currently teaching in the local labor market will depend upon the change in the total number of positions from the previous year and the number of teachers leaving the labor market (e.g., retirements). The total number of positions filled then will equal the number of hired individuals new to the local market plus the number of teachers making intra-market transfers.[7] During this hiring season, the attributes of the jobs being filled also are largely fixed. The labor contracts districts have in place determine salaries, and other conditions of work either will be exogenously determined (e.g., student-body composition and school location) or set by prior decisions made by school officials. In general, many policies, while not completely inflexible, are slow to change as a result of both the political

[7] A large majority of teaching positions are held by teachers unaffected by such annual hiring. With these individuals having tenure, school officials have almost no discretion regarding their continued employment. At the same time, teachers having more than a few years of experience only rarely make inter-district transfers. Because the salary schedules typically employed by districts reward prior teaching experience within the district, not total experience, a more experienced teacher transferring to another district would loose her within-district experience premium, which typically is quite large. For many of these individuals, the relevant employment decision is merely whether to continue teaching in the same district or leave teaching altogether.

process and collective bargaining. Given these institutional features, we view the matching of teacher candidates to job openings at the start of a school year as reflecting a short-run equilibrium in a setting with posted wages and nontransferable utility.

IV. Common Approaches for Modeling Sorting

This paper uses the observed matches of teachers to jobs to make inferences about the factors affecting the employment choices and underlying preferences of individual teachers and hiring authorities. Before describing our empirical sorting model in detail, we review several literatures pertinent to the study of the sorting of teachers across jobs. These include the hedonics literature and two literatures concerned with two-sided matching.

Hedonic Models: Most studies of teacher labor markets, such as Antos and Rosen (1975), employ hedonic models. Using data characterizing teachers and the jobs they hold (e.g., salaries) researchers estimate reduced-form wage equations. These in turn can be used to estimate the pay differential needed to compensate individuals for working in jobs with particular characteristics, as well as the pay increase needed to improve the quality of workers hired in jobs having particular attributes. However, estimates of such wage equations in teacher labor markets, as well as a range of other settings, have proven inconsistent.

Researchers have posited a number of reasons for counterintuitive wage equation results including omitted variables (Brown, 1980; Lucas, 1977), simultaneity (McLean, 1978), measurement error, and labor market frictions (Hwang, Mortensen and Reed, 1998; Lang and Majumdar, 2004). Recently, Ekeland, Heckman and Nesheim (2004) point out that the hedonic models typically estimated employ linear approximations having unappealing properties. Thus, these restrictive functional forms could be a contributing factor. In the case of teacher labor markets, omitted variables characterizing schools, students, and teachers, as well as the endogenous determination of pertinent school policies have been offered as possible explanations for the inconsistencies found. Furthermore, the functional forms for the hedonic wage equations for teacher salaries typically are even more restrictive than the special case criticized by Ekeland *et. al*.. In addition to these possible explanations, the counterintuitive results, at least in part, could be the result of teacher labor markets having characteristics inconsistent with key assumptions in the hedonic framework.

An estimated wage equation characterizes how wages vary across the worker-job matches observed. Such a linking of wages to job and worker attributes can be used to make inferences regarding the underlying preferences of workers in circumstances in which workers' indifference curves for job attributes are tangent to the hedonic wage function. Similar tangencies for employers are needed in order to make inferences regarding their willingness to tradeoff various attributes of workers. These tangencies in part follow from two assumptions standard in the formulation of hedonic models. First, markets clear such that demand equals supply at each combination of worker and job attributes.[8] Second, there are continuums of worker and job attributes.[9] In the case of teacher labor markets, these assumptions may not hold, even as first approximations.

Union contracts at the district level typically set teacher salaries for three or more years. Furthermore, they typically dictate that all teachers having the same number of years of

---

[8] For example, Ekeland, Heckman and Nesheim (2004) note that "[e]quilibrium in hedonic models requires that demand and supply be equated at each point of the support of *z*" where z is the set of job attributes. Similarly, Rosen (1974) assumes that the hedonic price relationship is "determined by some market clearing conditions: Amounts of commodities offered by sellers at every point on the [attribute] plane must equal amounts demanded by consumers choosing to locate there."

[9] More formally, the density functions characterizing the distributions of buyer and seller characteristics are assumed to be strictly positive in the interiors of their respective supports.

education and within-district experience earn the same salary, regardless of either their other attributes or the characteristics of the schools in which they teach. This limitation is especially restrictive in large urban districts and countywide districts in which there is considerable within-district variation in the non-wage attributes of schools. Lack of wage flexibility from year to year and across schools within districts brings into question whether wages could be the mechanism for short-run market clearing in the hedonic framework. In order for markets to clear, another school factor (e.g., fringe benefits or length of the work day) or worker attribute would need to be flexible enough to provide this mechanism. Worker attributes could be flexible either by the choices of a given set of worker (i.e. effort) or by flexibility in the pool of teachers. However, within the context of the sorting of new teachers across job openings at the start of a school year, it is questionable whether there is sufficient flexibility in some combination of the attributes of the jobs to be filled and teacher candidates to reasonably assure market clearing. [10]

Without market clearing, there will be workers or employers, or both, who are unable to find match partners having attributes they view as optimal. Here optimality is determined by their preferences and the hedonic wage function faced. Depending upon the circumstances, some type of equilibrium typically will result, which could involve worker queues, unfilled positions, etc.. The sorting of teachers across jobs might be similar to what would have occurred had wages or a combination of other attributes been sufficiently flexible to clear the market. However, those teachers and employers unable to optimize subject only to the hedonic wage function will have marginal evaluations of at least some of the attributes characterizing their

[10] Rather than the attributes of individual workers and employers being variable, some models assume that individual workers and employers have fixed attributes but the number of each type can vary (e.g., free entry of firms). Such a possibility is more relevant for teachers than schools. However, the entry/exit of teacher candidates having particular attributes, while possible in the long-run, seems less relevant as a mechanism for bringing about a short-run equilibrium. Furthermore, the free entry of teachers by itself need not assure that the full set of tangencies are satisfied.

match partners that differ from the marginal prices they face implicit in the wage function, violating a basic feature of the hedonic framework.[11]

In cases where there is market clearing, the assumption that there are continuums of worker and job attributes available over their relevant ranges implies that decision makers have complete flexibility in choosing optimal bundles of attributes. This is necessary to assure the tangencies that justify using the hedonic wage equation to make inferences regarding the underlying preferences of workers and employers. However, local teacher labor markets are thin in terms of the number of job openings and candidates available. As noted above, labor markets for teachers tend to be quite small geographically, so there may not be sufficient numbers of jobs and candidates in each local market to assure that the distributions of attributes characterizing employer and employee interacting in the market are approximately continuous. This is especially true when one considers the markets for particular specializations (e.g., those certified to teach high school mathematics).[12] With decision-makers facing a limited number of discrete alternatives, wages still could adjust so that there was market clearing in the sense that demand equaled supply for each worker-job combination. However, the implicit prices for each attribute implied by the market wage function need not equal the marginal evaluations of decision-makers. In this setting, discrete choice models, such as random utility models, are likely to be more applicable for the analysis of job choice (Freeman, 1979; Palmquist 1991).

Our intent here is not to make the case that teacher labor markets never clear in the short-run or that the marginal prices implicit in the wage equation never reflect either teachers'

[11] Again, market equilibrating prices would rule out this possibility and help assure optimality by both workers and employers. "[E]quilibrium prices are determined so that buyers and sellers are perfectly matched. No individual can improve his position and all optimum choices are feasible." (Rosen, 1974, Page 35)
[12] We consider the matching of new elementary teachers to job openings in five labor markets (metropolitan areas) in each of six years. For the median of these cases, 141 newly hired teachers took jobs in a total of 82 elementary schools. Market thinness is even more apparent when it is noted that an empirical analysis typically will include multiple attributes characterizing schools and job candidates.

marginal evaluations of job characteristics or employers' willingness to pay for teacher attributes. Rather, the point is that teacher labor markets are such that it is unlikely that these conditions *always* hold, which could help explain why the application of hedonic models in the analysis of teacher labor markets has yielded counter-intuitive results.

In what follows, we develop and estimate structural models drawing upon the game-theoretic two-sided matching literature. These models account for pertinent features of teacher labor markets, including the fact that job candidates and employers both face relatively limited numbers of discrete choices. The framework allows for the possibility that neither wages nor other attributes of teachers or jobs are sufficiently flexible to assure market clearing. We use the empirical framework to isolate the factors affecting the separate, but interdependent, choices made by job candidates and school officials. More specifically, we estimate the underlying preference parameters reflecting teachers' evaluations of various job attributes as well as employers' preferences for attributes characterizing teachers.

Two-sided matching: The two-sided matching literature is applicable to a broad range of settings in which individuals in one group are matched with individuals, agents or firms in a separate, second group. Examples include models of marriage, employment and college attendance.[13] In all of these cases, the matching is two-sided in that whether a particular match occurs depends upon separate choices made by the two parties. Furthermore, these choices are not made in isolation. "A worker's willingness to accept employment at a firm depends not only on the characteristics of the firm but also the other possible options open to the worker. The better an individual's opportunities elsewhere, the more selective he or she will be in evaluating a potential partner," (Burdett and Coles, 1999).

[13] These cases differ from the roommate problem where those being matched come from the same group. In two-sided match models all agents fall into one of two distinct groups and seek a match with one or more agents in the other group.

Within the two-sided matching literature, there are now a large number of papers that build upon the work of Gale and Shapley (1962) to consider one-to-one matching such as marriage or many-to-one matching such as employment and college-admission, the former being a special case of the latter. This game-theoretic two-sided matching framework has been widely discussed in the context of matching medical residents to hospitals. The major difference between the assignment of residents to hospitals and the sorting of teachers across schools is that the assignment of residents typically results from a centrally controlled allocation mechanism,[14] whereas teacher labor markets involve decentralized job matching. However, this difference is not as great as one might first think, since the Gale-Shapely algorithm employed in the centralized matching of residents to hospitals mimics one particular decentralized mechanism that yields a match equilibrium.[15] We further discuss this mechanism below in the context of our empirical model.

While a growing number of papers allow utility to be transferable so that the division of match surplus is determined endogenously at the time matches occur, most game-theoretic models assume that utility is *nontransferable*; that is, how the surplus from any given match is split between the matching pair is predetermined. Given the features of teacher labor markets discussed above (e.g., wage posting), we maintain nontransferable utility in our empirical framework.

In general, very little empirical work has been done estimating game-theoretic matching models. Choo and Siow (2006) estimate a static transferable utility model of the marriage market in which the number of person types is very limited (i.e., individuals are only

[14] Roth and Sotomayer (1990) provide a clear synthesis of both the theoretical literature and how the theoretical findings provide important insights regarding implications of the institutional features characterizing centralized matching algorithms, as well as factors that have contributed to the evolution of those features.

[15] A decentralized allocation need not rely on the Gale-Shapely algorithm. Roth and Vande Vate (1990) show a very general decentralized mechanism will lead to a stable allocation.

differentiated by age). Fox (2006) develops a computationally appealing estimation strategy for estimating more general models falling within a broad class of matching games with transfers. However, with transferable utility, the statistical approaches do no allow one to separately identify preference parameters for workers and employers, as the match production (utility) function estimated is the sum of the match production (utility) levels of the matched pair. In contrast, the empirical framework we develop allows us to separately estimate workers' preferences for particular job attributes as well as employers' preferences for individual worker attributes. In fact, sorting out how various factors separately affect the employment choices of teachers and school hiring authorities is the primary motivation for our analysis.

In addition to the game-theoretic studies, there is a large literature in labor economics employing two-sided matching models with search.[16] This research distinguishes itself in a number of respects. First, whereas almost all the game-theoretic models assume full information and no market frictions, such frictions are central to the labor-search models of marriage and job match. A second difference is that the demand side of the labor-search models often is characterized by free entry of profit maximizing firms, so that the number of jobs to be filled is not fixed as in the game-theoretic match literature. A third difference that is especially pertinent for our empirical analysis concerns the extent and nature of agent heterogeneity allowed in the models. Game-theoretic two-sided match models typically only require that each agent's ranking of match partners is complete and transitive, with no restrictions regarding the extent of preference heterogeneity. In contrast, the search models either maintain homogeneity of preferences or allow for only limited heterogeneity. Some models maintain *match heterogeneity*, where agents in each group are ex-ante identical but some matches are relatively

[16] See Rogerson, Shimer and Wright (2005) and Eckstein and van den Berg (forthcoming) for informative overviews of the theoretical and empirical literatures, respectively.

more productive, with the productivity of each possible match determined by a random draw from some known distribution. Other models maintain *ex-ante heterogeneity* where there are systematic differences across agents independent of the partners to whom they are matched, with all agents in one group having the same ranking of the potential partners in the other. For example, some workers may be more productive than others and some jobs may be more or less attractive. Such *ex-ante* heterogeneity is maintained by Wong (2003) in a structural estimation of marriage models.

Limitations on the degree of heterogeneity are needed in order to solve for the search equilibriums (Burdett and Coles, 1999). However, such limited heterogeneity would be quite restrictive if maintained in our analysis. For example, as evidenced by the discussion of the importance of distance from home to jobs, teachers may rank the same job differently because of their location relative to the school, violating *ex-ante heterogeneity*. We avoid this limitation by employing the game-theoretic approach, with the hope of incorporating market frictions into later work.

V. The Model

Consider an environment in which $C=\{c_1,\cdots,c_J\}$ represents the set of J individuals seeking teaching jobs and $S=\{s_1,\cdots,s_K\}$ represents the set of K schools having jobs to be filled, $J \geq K$. For now assume that each school has one job opening, though this assumption is relaxed in the empirical analysis. We assume that each agent has a complete and transitive preference ordering over the agents on the other side of the market and that these orderings arise from job candidates' preferences over job attributes and hiring authorities' preferences over the attributes of candidates.

Let u_{jk} represent the utility of working in the kth school as viewed from the perspective of the jth candidate where $u_{jk} = u(z_k^1, d_{jk} \mid q_j^2, \beta) + \delta_{jk}$. z_k^1 is a vector of observed attributes of the kth school pertinent to the jth individual and d_{jk} is the distance to the kth job for the candidate. Vector q_j^2 represents observed attributes of the jth candidate that affect the individual's assessment of the kth alternative and β is a vector of parameters. δ_{jk} is a random variable reflecting unobserved heterogeneity in the attractiveness of a particular school for different individuals. If no job match is entered, the individual's utility is u_{j0} which depends upon observed and unobserved attributes of the individual. Thus, the individual will always turn down a job offer if $u_{jk} < u_{j0}$. Here we assume that $u_{jk} > u_{j0}$ for all k and j but plan to allow for the more general case when, in further research, we extend the empirical model to consider all candidates, not just those actually obtaining jobs.

The hiring authority for the kth school is assumed to have preferences over the attributes of job candidates. Let $v_{jk} = v(q_j^1 \mid z_k^2, \alpha) + \omega_{jk}$ represent the attractiveness of the jth candidate from the perspective of the hiring authority for school k. The vector q_j^1 represents pertinent observed attributes of the jth candidate. The vector z_k^2 represents the observed attributes of the kth school that might affect the authority's assessment of the jth candidate. α is a vector of parameters. The random error ω_{jk} reflects unobserved factors. To simplify the analysis, we assume hiring authorities prefer all of the candidates to the alternative of leaving job vacancies unfilled. This assumption, combined with the assumption that there are sufficient numbers of willing candidates, implies that all job openings will be filled.

Consider a case where the sets C and S are known, as are the values of $q_j \equiv (q_j^1, q_j^2)$ for each candidate and $z_k \equiv (z_k^1, z_k^2)$ for each job. Given the vector of parameters β and a particular set of random variable draws for the δ_{jk}, the formula $u_{jk} = u(z_k^1, d_{jk} \middle| q_j^2, \beta) + \delta_{jk}$ implies the matrix of candidates' benefits represented in panel (A) of Figure 1. Each row shows the benefits that a particular candidate attributes to being employed in each of the K school alternatives. These rows of benefit values, in turn, imply candidates' complete rankings of school alternatives shown in panel (C). r_{jk}^c is the jth candidate's ranking of the kth school alternative. In a similar way, the vector of parameters α and a particular set of random variable draws for the ω_{jk}, together with the formula $v_{jk} = v(q_j^1 \middle| z_k^2, \alpha) + \omega_{jk}$, imply the matrix of school benefits represented in panel (B) of Figure 1 and the complete rankings of candidates by hiring authorities shown in panel (D). Each column of panel B shows the benefits to a particular school of having an opening filled by each of the alternative candidates. r_{jk}^s is the ranking of the jth candidate from the perspective of the kth employer.

If each of the candidates unilaterally were able to choose the school in which to teach, the framework summarized in panel A would imply that β in $u_{jk} = u(z_k^1, d_{jk} \middle| q_j^2, \beta) + \delta_{jk}$ could be estimated using data characterizing those choices and a standard multinomial probit or logit random utility model. Similarly, α could be estimated easily using the same type model, if each hiring authority unilaterally chose among candidates. However, the empirical model we employ is more complex for two reasons. First, it is the interaction of decisions made by a candidate and a hiring authority for a school that determines whether the two are matched. Second, even though any such interaction would complicate the model, the decisions made by the two parties considering whether to match crucially depend upon the choices made by all other candidates and employers. In particular, a candidate's willingness to accept a particular match depends upon her own preferences as well as her choice set, i.e., the set of schools willing to hire her given their own alternatives. In turn, whether employers make the candidate an offer will depend upon whether they prefer to employ alternative candidates who are willing to fill their positions, and so on.

Because our framework is an empirical application of the standard two-sided matching model extensively studied by game theorists, much in that literature directly applies to our analysis (Roth and Sotomayer, 1990). We assume that a decentralized job-match mechanism leads to a stable matching of teacher candidates to jobs. For a set of matches to be stable, there must be no candidate-employer pair currently not matched together who both would prefer such a new match rather than remain in their current matches. Otherwise, if allowed, the pair would break their current matches in order to match with each other. More formally, without loss of generality, suppose that candidate g is employed in job g', with candidate h and job h' similarly matched. For these two pairings to be stable, it must be the case that (1) $u_{gg'} > u_{gh'}$ or $v_{hh'} > v_{gh'}$, or both (i.e., either candidate g or employer h' prefers the status quo to the alternative of candidate g being employed in job h') and, similarly, (2) $u_{hh'} > u_{hg'}$ or $v_{gg'} > v_{hg'}$, or both.[17] Equivalent expressions for these two conditions are $1\left(u_{gg'} < u_{gh'}\right) 1\left(v_{hh'} < v_{gh'}\right) = 0$ and $1\left(u_{hh'} < u_{gh'}\right) 1\left(v_{gg'} < v_{hg'}\right) = 0$ where $1(\ )$ is the indicator function which equals one if the function argument is true and zero otherwise. Overall stability required that the condition

[17] Strict rankings of alternatives (i.e., no agent is indifferent between any two alternatives) are assumed here to simplify the discussion.

$1\left(u_{gg'} < u_{gh'}\right) 1\left(v_{hh'} < v_{gh'}\right) = 0$ hold for every candidate (g) and job (h') pairing not currently matched.

Our empirical framework maintains that the observed matching of teachers to schools is an employer-optimal stable matching (i.e., all employers weakly prefer this allocation to all other stable matchings). The following decentralized job-match mechanism is one process that leads to this employer-optimal matching. Each employer initially makes an offer to its highest ranked prospect. Job candidates receiving offers reject those that are dominated either by remaining unemployed or by better job offers, and "hold" their best offers if they dominate being unemployed. Employers whose offers are rejected make second round offers to their second highest ranked choices. Employers whose offers remain open stay in communication with these candidates but otherwise take no action. Job candidates receiving better offers inform employers that they are rejecting the less attractive positions previously held. In subsequent steps each employer having an opening with no outstanding offer makes an offer to its top candidate among the set of job seekers who have not already rejected an offer from the employer. Employees in turn respond. This *deferred acceptance procedure* continues until firms have filled all their positions with their top choices among those not having a better offer or have made unsuccessful offers to all their acceptable candidates. As shown by Gale and Shapley (1962), such an allocation mechanism always will yield a stable matching. Furthermore, if the rankings are strict, the resulting stable matching will be both unique and employer-optimal. Alternatively, a deferred acceptance procedure in which candidates made offers to hiring authorities would result in an employee-optimal matching.

The equilibrium employer-optimal stable matching corresponding to the alternatives and rankings characterized in Figure 1 is represented in the left side of Figure 2. The right side of

Figure 2 characterizes this matching in terms of the resulting relationship between the attributes of candidates and the schools where they are employed.[18] The matching of candidates to schools represented in Figure 2 corresponds to particular values of the model's random variables (δ_{jk} and ω_{jk}; $\forall\, j,k$), the explanatory variables (e.g., q_j and z_k) and the parameters of the model $\left(\theta = (\alpha, \beta)\right)$.

Given the implied rankings for candidates and employers, deriving such a stable matching is relatively easy using the Gale-Shapley matching algorithm. However, deriving closed-form expressions for the likelihood of observing any particular candidate-job matching or the probability distribution of any particular distribution of worker and job attributes is impossible. To compute the likelihood of a particular stable matching one would need to identify the set of all possible combinations of the random errors that would lead to that same stable matching. This would entail determining all possible combinations of the rankings of candidates and employers that would yield a particular matching and, in turn, all the combinations of random variable values that would lead to each of those sets of rankings. This is an impossible task, especially since it would have to be done repeatedly for various parameter values. Even if the ranges of the various random errors could be identified, computation of the corresponding likelihood would be impossible given that the implied integrals would have high dimensions and very complex regions of integration.[19] These complexities motivate our use of a method of simulated moments (MSM) estimation strategy.

Before discussing the MSM approach, it is first necessary to generalize the notation and framework. Whereas the above discussion was for a single market at one point in time, our

[18] Note that multiple worker-job matchings will yield the same distribution of matched attributes if either multiple candidates or multiple jobs have the same observed attributes.
[19] Berry (1992) makes a similar point in a game-theoretic model of entry in the airline industry.

empirical analysis considers M local labor markets, m = 1,2,…,M, and T years, t = 1,2,…,T. To account for this generalization, we need only add the subscripts "m" and "t" to the explanatory and random variables defined above. For example, q_{mtj} represents the attributes of candidate j first employed in market m during time period t. An assumption is needed to allow for multiple job openings in a single school in any given year. With our empirical analysis focusing on elementary schools where there is a large degree of homogeneity across teaching jobs, we assume that all job openings within a school are identical. As shown in the two-sided match literature, the pertinent theoretical underpinning for many-to-one matches parallels that for one-to-one matches discussed above.

Let $\tilde{z}_{mtj}$ represent the attributes of the job taken by teacher j newly hired in market m during period t. (Reflecting the two-sided match, $\tilde{z}_{mtj}$ from the perspective of this teacher is the same as z_{mtk} defined above where the k^{th} school employs the jth individual.) The structure of the two-sided matching model, values of parameters α and β and the distributions of the random variables δ_{jk} and ω_{kj} together imply the joint distribution of $\tilde{z}_{mtj}$ and q_{mtj}. This in turn implies the expected value of $\tilde{z}_{mtj}$ for the j^{th} job candidate, $E\left(\tilde{z}_{mtj} \mid q_{mtj}\,;\theta\right)$. It follows that $E\left[\tilde{z}_{mtj} - E\left(\tilde{z}_{mtj} \mid q_{mtj}\,;\theta\right) \middle| q_{mtj}\right] = 0$; for a candidate having attributes q_{mtj}, the difference between the attributes of the school where the individual works, $\tilde{z}_{mtj}$, and the expected mean attributes, given q_{mtj}, is zero in expectation. In turn, this implies that $E\left(q_{mtj}\left[\tilde{z}_{mtj} - E\left(\tilde{z}_{mtj} \mid q_{mtj}\,;\theta\right)\right]\right) = 0$; across teacher candidates, the difference between the actual and expected attributes of the school where individuals work is orthogonal to their own attributes. The sample analog of the last expression is $\sum_t \sum_j q_{mtj} \left[\tilde{z}_{mtj} - E\left(\tilde{z}_{mtj} \mid q_{mtj}; \theta \right) \right] = 0$,

which we employ in estimation.[20] Similarly, we use $\sum_t \sum_j q_{mtj} \left[\tilde{d}_{mtj} - E\left(\tilde{d}_{mtj} \mid q_{mtj}; \theta \right) \right] = 0$ which relates the actual distance for each employee to the corresponding expected value.

Implementing our estimation strategy is complicated by the fact that $E\left(\tilde{z}_{mtj} \middle| q_{mtj};\, \theta\right)$ and $E\left(\tilde{d}_{mtj} \middle| q_{mtj};\, \theta\right)$ are not easily computed; we cannot write out, much less compute, analytical expressions for these expected values. We instead compute values for these expressions using simulation. Our method of simulated moment estimation strategy is described in Appendix A. In short, the MSM estimator, $\hat{\theta}$, is the value of θ which minimizes a quadratic form defined in terms of the moment conditions. The parameter estimates minimize the distance between moments reflecting the empirical distribution of school attributes across teachers and the corresponding theoretical moments implied by our model.[21]

Even though we here focus on the sorting of first-time teachers, the model can be extended to include the analysis of who becomes a teacher and who quits or transfers. It is

---

[20] Equivalently, we could have employed $E\left(z_{mtk} \left[\tilde{q}_{mtki} - E\left(\tilde{q}_{mtki} \middle| z_{mtk}; \theta \right) \right] \right) = 0$ and its sample analog $\sum_t \sum_k \sum_i z_{mtk} \left[\tilde{q}_{mtki} - E\left(\tilde{q}_{mtk} \mid z_{mtk}; \theta \right) \right] = 0$ which can be rewritten $\sum_t \sum_k n_{mtk}\, z_{mtk} \left[\overline{q}_{mtk} - E\left(\tilde{q}_{mtk} \mid z_{mtk}; \theta \right) \right] = 0$. Here $\tilde{q}_{mtki}$ represent the attributes of the teacher newly employed during period t to fill the i[th] vacancy in school k, i = 1,2,…, n_{mtk}, and $\overline{q}_{mtk}$ is the mean attributes of the n_{mtk} new teachers employed by the kth school. $\sum_t \sum_k n_{mtk}\, z_{mtk} \left[\overline{q}_{mtk} - E\left(\tilde{q}_{mtk} \mid z_{mtk}; \theta \right) \right]$ will always equal $\sum_t \sum_j q_{mtj} \left[\tilde{z}_{mtj} - E\left(\tilde{z}_{mtj} \mid q_{mtj}; \theta \right) \right]$.

[21] An alternative strategy would be simulated maximum likelihood (SML) estimation, based on the underlying model described above and the "crude Monte Carlo simulator". For a given set of parameter values and R simulations of the matching model corresponding to each market-year, the simulator for the probability of observing a particular market outcome, Pr(.) would be the proportion of those simulations for which *every* worker-job match exactly matched those observed. However, the number of draws (simulations) needed to attain a given variance for the simulator "rises inversely with Pr(.), which makes it intractable when this probability is small." (Hajivassiliou and Ruud, 1994) In cases like ours where there often are one to three hundred, and in some cases more, teachers being matched to jobs in roughly half that many schools, the probability of any particular full set of matches is extremely small. Thus, SML estimation is not feasible in our application.

possible to include potential teachers in the matching process, not just those who took teaching jobs. Similarly, instead of modeling job matching only for new teachers, one could allow for vacancy chains. That is, when an opening becomes available because a teacher leaves the system or because the number of teachers in a school increases, we can allow current teachers to move into those spots, creating vacancies in their old schools.

Within the burgeoning set of papers employing the method of simulated moments, three papers have substantial overlap with our application. Epple and Sieg (2001) employ the method of simulated moments approach to estimate Tiebout equilibrium models of residential choice. Their moment conditions relate to the equilibrium, one-sided sorting of households to local communities. Berry (1992) has employed a simulation estimator to estimate an equilibrium game-theoretic model of market entry in the airline industry, with the simulated moments based on the equilibrium number of firms operating at each airport each year. Sieg (2000) has estimated a bargaining model of medical malpractice disputes. Even though this analysis focuses on bilateral interactions between individual plaintiffs and defendants, rather than a market-level analysis, the paper is pertinent in that the simulated moments are obtained by repeatedly solving a game-theoretic model for each of a large number of draws of the model's random variables, as is the case in Berry's analysis.

VI. Model Identification

Even though a complete analysis of identification for the two-sided matching model goes beyond the scope of this paper, a number of useful insights follow from properties of related empirical models. Reconsider the case where the g^{th} (h^{th}) candidate is employed in job g' (h'). As noted above, the assumption of stability and the structure of revealed preferences imply that

$1\left(u_{gg'} < u_{gh'}\right) 1\left(v_{hh'} < v_{gh'}\right) = 0$. Contrast this to the case where matchings are one-sided. For example, if candidate g were able to freely choose among the full set of job openings, individual g would choose job g' only if $u_{gg'} > u_{gh'}$ and, equivalently, $1\left(u_{gg'} < u_{gh'}\right) = 0$, $\forall\ h';\ h' \neq g'$. Similarly, if the hiring authority filling job h' were able to employ any candidate, the employer would hire candidate h only if $1\left(v_{hh'} < v_{gh'}\right) = 0$, $\forall\ g;\ g \neq h$. Such decision rules underlie the common discrete-choice random utility framework in which each decision-maker is free to choose any alternative from a predetermined, finite set of options. Findings regarding identification in standard random utility models of choice carry over to the case of two-sided choice.

Consider a one-sided job match in which teacher candidate g chooses job g'. This choice together with our characterization of candidate's preferences over jobs implies the expression $u(z^1_{g'}, d_{gg'} \mid q^2_g, \beta) + \delta_{gg'} > u(z^1_{h'}, d_{gh'} \mid q^2_g, \beta) + \delta_{gh'};\ \forall\ h',\ h' \neq g'$. As discussed by Manski (1995, p. 93), such inequalities provide no identifying power with respect to the nonstochastic component of utility, u( ), in general, and the parameter vector β, in particular, unless assumptions are made regarding the unobserved random variables.[22] Parametric models typically assume that the random errors are drawn from either a normal or logistic distribution and are statistically independent of the variables included in u( ). Identification also requires additional assumptions with respect to the covariance structure of the error terms. For example, it is not possible to estimate all the parameters in an unrestricted covariance matrix for the normal random errors in a multinomial probit model (e.g., the variance of at least one of the random

[22] At least minimal assumption regarding the random errors are needed even in the case of nonparametric estimation. For example see Klein and Spady (1993).

errors must be fixed).[23] In our analysis of two-sided matching, we also employ a parametric estimation strategy in that computation of the simulated moments is based on explicit assumptions regarding the distributions of the error terms.

Issues of identification also arise with respect to the nonstochastic component of utility in one-sided random utility models.[24] Consider the linear-in-parameters specification $u(z_{h'}^1, d_{gh'} \,|\, q_g^2, \beta) = \beta^1 z_{h'}^1 + \beta^2 d_{gh'} + \gamma\, q_g^2 + \left(q_g^2\right)' \Lambda\, z_{h'}^1$ for the case in which candidate g can freely choose among a full set of job openings and where β^1, β^2, and γ are vectors of parameters and Λ is a conforming matrix of parameters. In this specification, $\gamma\, q_g^2$ does not affect the individuals' relative rankings of alternatives, implying that γ cannot be identified. Thus, attributes of the candidate will affect the alternative chosen only to the extent that q_g^2 is interacted with the attributes of alternatives or has coefficients that vary across alternatives. However, dropping $\gamma\, q_g^2$ from the equation is of no consequence, a result which will prove useful below in the case of two-sided matching. In general, all the issues regarding identification in the case of one-sided choice carry over to the specification of the random utility equations in models of two-sided match.

In addition to the limitations common to identification of standard random utility models, the two-sided model has additional limitations similar to those in bivariate discrete choice models with partial observability. Consider a bivariate discrete choice model where $y_m^* = \theta_m x_m + \eta_m$ is a latent dependent variable and $y_m = 1\left(y_m^* < 0\right)$, m = 1,2. Compared to the case where y_1 and y_2 are both observed, identification is more difficult when the researcher only

[23] See Bunch and Kitamura (1989), Bunch (1991), Dansie (1985), and Keane (1992).
[24] For example, see Ben-Akiva and Lerman (1985).

observes the value of $y = y_1 y_2 = 1(\theta_1 x_1 + \eta_1 < 0)\, 1(\theta_2 x_2 + \eta_2 < 0)$. With only partial observability, the identification of θ_1 and θ_2 crucially depends upon whether exclusion restrictions are justified *a priori*; there must be one or more quantitatively important variables that enter x_1 or x_2, but not both.[25]

Similar exclusion restrictions are needed for identification in the two-sided matching model. Stable two-sided worker-job matches imply that the structure of revealed preferences is fully characterized by $1\left(u_{gg'} < u_{gh'}\right) 1\left(v_{hh'} < v_{gh'}\right) = 0$ where there is one such condition for each candidate (g) and job (h') pair not actually matched. Comparing the utility expressions $u_{jk} = u(z_k^1, d_{jk} \mid q_j^2, \beta) + \delta_{jk}$ and $v_{jk} = v(q_j^1 \mid z_k^2, \alpha) + \omega_{jk}$ that enter the above expression, one sees that either differences between the variables entering z_k^1 and z_k^2 or differences between the variables entering q_j^1 and q_g^2 would yield such exclusion restrictions. In our application, the assumption that distance enters u( ) but not v( ) is one such example. More generally, rather than having to make such *a priori* assumptions, exclusion restrictions naturally arise in the two-sided match model, even when there are no differences in the variables entering u( ) and v( ). Consider the linear-in-parameters second-order Taylor approximations $u(z_k \mid q_j, \beta) = \beta z_k + q_j' \Lambda z_k$ and $v(q_j \mid z_k, \alpha) = \alpha q_j + z_k' \Psi q_j$. When q_j is normalized to have a zero mean, β in u( ) captures the average effect of z_k on u( ). Given a similar normalization of z_k, α captures the average effect of q_j on v( ). As noted above for the case of one-sided matching, q_j does not enter u( )

[25] See Poirier (1980). Even when the model is identified, partial observability typically leads to a reduction in the precision of parameter estimates (Meng and Schmidt, 1985).

linearly, just as z_k does not enter v( ) linearly, thus implying very general *a priori* exclusion restrictions in two-sided match models.[26]

This discussion of identification has focused on the revealed preferences implied by the structural model, rather than the particular estimation strategy we employ. However, the moment conditions we employ in estimation only account for the attributes of those entering matches, not the identities of those entering each candidate-job pairing, as accounted for in the condition $1\left(u_{gg'} < u_{gh'}\right) 1\left(v_{hh'} < v_{gh'}\right) = 0$. The identifying information contained in the structure of revealed preferences represents an upper bound with respect to identification within our GMM framework.

VII. Estimates of Several Models

We estimate the two-sided matching model utilizing data characterizing the initial sorting of first through sixth grade teachers across schools in the Albany-Schenectady-Troy, Buffalo, Rochester, Syracuse, and Utica-Rome metropolitan areas for school years 1994-95 through 1999-2000.[27] We estimate the following utility functions, with certain parameters set equal to zero in some models.

$$u_{jk} = \beta_1\, salary_k + \beta_2\, Spoverty_k + \left[\beta_3\, Tminority_j + \beta_4 \left(1 - Tminority_j\right)\right] Sminority_k + \beta_5\, urban_k + \beta_6\, distance_{jk} + \delta_{jk}$$

$$v_{jk} = \alpha_1\, Tquality_j + \alpha_2\, Tminority_j + \omega_{jk} \qquad \textbf{(3)}$$

---

[26] Here the key assumption is that either z_k enters u( ) or q_j enters v( ), at least in part, additively. For example, representing u( ) generally as an nth-order Taylor approximation, z_k will enter u( ) linearly whenever the first derivative of the underlying function with respect to z_k is not zero at the point of expansion.

[27] With computational limitations necessitating that we exclude the New York City metropolitan area, our analysis includes the other large metropolitan areas in the state.

Thus, the j[th] teachers' utility associated with working in the k[th] job, u_{jk}, is assumed to be a function of the starting salary (*salary)*, the proportion of students in the school who are poor (*Spoverty*) as measured by eligibility for free lunch, the proportion of students who are black or Hispanic (*Sminority*), whether the school is in an urban area (*urban*), and *distance*. The specification allows for the possibility that the effect of a school's racial composition will vary depending upon whether the teacher is black or Hispanic (*Tminority*). *Distance* is measured from the school to the address given when the individual applied for certification, a point in time typically prior to when individuals apply for teaching jobs. While an alternative distance measure based on their location when in high school would be preferable because it is not endogenous to where teachers hope to teach, we do not have this for all teachers. If the distance to each district in the labor market where the individual took a job was greater than 50 miles, we equalized the distance measures for all job alternatives, so that distance is not a factor in the candidate's choice of jobs and will drop out of the moment conditions.

The attractiveness of the j[th] candidate from the perspective of the hiring authority for school k, v_{jk}, is a function of teacher qualifications (*Tquality*) and whether the individual is black or Hispanic. *Tquality* is measured as a scalar composite of (1) whether the teacher ever failed a liberal arts certification test; (2) the test score on the first taking of the certification exam; (3) the Barron's rating of his/her undergraduate institution; and (3) whether or not he/she has attained more than a Bachelor's degree.[28] Both equations have normal random errors that are standardized, with no loss of generality, to have standard deviations of one.

[28] We used principal component analysis to determine the weights used in constructing the composite. The eigenvalue is 1.65 and the weightings are 0.6773 for test score, 0.6087 for failing, 0.3089 for college selectivity and 0.1603 for higher degree.

Table 1 presents the sample statistics. Starting salaries average $32,458 with a small standard deviation of $2,607. On average 21 percent of students in a school were black or Hispanic and 29 percent were poor. Many more new teachers were hired in recent years. Few (6.4 percent) were black or Hispanic, and for those traveling less than 50 miles to their job, the average distance was only 8.6 miles.

Table 2 gives the method of simulated moments results for several different models. The models differ in terms of assumptions regarding the distributions of random errors, the explanatory variables included in the nonstochastic components of utility, and the moment conditions employed in estimation. In Models I-VI the error terms are independent, standard normal random variables. The MSM estimation of the parameters in models I, II, IV, and V rely on 75 moment conditions. For each of the labor markets, $\sum_t \sum_j q_{mtj} \left[\tilde{z}_{mtj} - E\left(\tilde{z}_{mtj} \mid q_{mtj};\theta \right) \right] = 0$ includes the four school characteristics (*salary*, *Spoverty*, *Sminority*, and *urban*) in $\tilde{z}_{mtj}$ and *Tquality*, *Tminority* and (1-*Tminority*) in q_{mtj}. Both *Tminority* and (1-*Tminority*) are entered because q_{mtj} does not include a constant term. These moments along with the three moment conditions in $\sum_t \sum_j q_{mtj} \left[\tilde{d}_{mtj} - E\left(\tilde{d}_{mtj} \mid q_{mtj};\theta \right) \right] = 0$ imply a total of 15 for each market. We employ the same set of moments in estimating the four models in order to isolate the effect of changing the variables entering the preference equations from the effect of changing the moment conditions entering the objective function used in defining the estimator. Models III and VI employ only moment conditions for the variables included as explanatory variables in the preference equations.

Note that all the estimated coefficients are of the expected signs and standard errors are quite small. For Model I, teacher qualifications have a positive effect on employer utility.

Salary has a positive effect on teacher utility; while percent minority, percent poor, urban, and distance all have negative effects. To interpret the size of these effects we can compare the coefficient estimates across variables or compare the size of the effect to the variance of the error (signal to noise). In Model I the estimate of the utility loss associated with teaching in a school having 30 percentage points more minority students (approximately one standard deviation) is 1.18, an effect that could be offset by roughly a $10,747 increase in salary. As noted above, the variation in salary is quite small; therefore, comparisons between non-salary variables may be more fruitful. The model suggests that the proportion of poor students has about half of the effect on utility as the proportion of minority students. On average, teachers would be indifferent between a school that had five percentage points higher minority student enrollment and one that had ten percentage points higher enrollment of students in poverty.

The employers criterion function for Model I is univariate so we can not compare coefficients. An alternative is to compare the coefficient to the error. Teacher qualifications as measured by test scores and college attended contributes to schools' assessments of potential teachers. A one standard deviation increase in qualifications increases utility by 0.205 points. With the error in this equation and the teacher qualifications factor both having standard deviations equal to one, the overall variance in utility is 1.042 (alpha squared times the variation in qualifications plus the variation of the random error), assuming that qualifications are orthogonal to the error. Thus, our qualifications measure appears to account for approximately four percent of the total variance in utility.

As noted above, a potential advantage of the empirical model developed here is the ease with which preference heterogeneity can be taken into account, in particular the heterogeneity resulting for teacher-job proximity. The large magnitude of the distance coefficient estimate underscores that this is important. To investigate the importance of accounting for distance further, the second and third column of Table 2 reports results with distance omitted from the model. The estimates in Model II come from a specification in which distance is included in the moment conditions, while the estimates in the third column come from a model without distance in the moment conditions. We find that with distance removed from both preferences and the moment conditions, the coefficients on both urban and percent minority change sign. The coefficient on the percent of minority students is no longer significant.

This change in the estimated coefficients is likely to result from the initial location of teachers. While urban areas are often net importers of teachers, their central location within labor markets actually makes the distances for teachers from their prior home to their new jobs somewhat lower on average for urban teachers than for suburban teachers. For this sample, excluding those who travel more than 50 miles, the average distance from home to job for urban teachers is 6.14 miles, compared with 9.05 miles for suburban teachers (the medians are 4.47 and 7.28, respectively). Thus, when not controlling for distance, the negative characteristics of local schools (as proxied by the percent of minority students and urban) are masked by the positive effect of proximity. Overall, distance is an important explanatory variable and provides important identification in the standard model.

The fourth model in Table 2 introduces the race/ethnicity of candidates into the utility function of the employers. This does not substantially change the coefficients for the other variables but does show that employers value minority candidates. They appear to be willing to tradeoff slightly more than one standard deviation in the quality index for a non-white teacher. Model IV also adds an interaction between the measure of school racial composition and whether or not a teacher is non-white. The estimates for the teachers' utility show that for non-white teachers the effect of the proportion of non-white students on utility is positive, while for white teachers it is negative. Distance continues to play an important role in this specification. Without adjustments for distance, teachers appear to prefer urban schools and their preferences for schools with a lower proportion of minority students is, generally, not as strong (See Models V and VI).

The final two models in Table 2 serve as specification checks to our identifying assumption in Models I-VI that all the error terms are independent, standard normal random variables. As discussed above, even though the theoretical model places no restrictions on the structure of the error terms in our specification of an empirical two-sided matching model, parameter identification crucially depends upon such restrictions. Model VII introduces school random effects into the specification of candidates' preferences as a first step in accounting for school attributes observed by teacher candidates but omitted in our analysis. Model VIII maintains that δ_{jk} and ω_{jk} are based on independent draws from the student's t distribution with four degrees of freedom. Because the variance of this student's t distribution equals 2, we divided the random draws by $\sqrt{2}$, yielding random variables with unit variances, consistent with the normalization maintained throughout our analysis. This transformation yields random variables with thicker-tails at the same time there is a higher concentration of values close to the mode of zero, compared to the standard-normal distribution employed in Models I-VI.

The random effects model changes the results very little except that the percent of minority students has an even stronger negative effect on teachers' utility, while the effect of being in an urban school is not as strong. Note that the school random effects are estimated to be slightly less than a third of the total variance. The Student's t model is also quite similar, except that non-white teachers no longer appear to affect employers' utility.

Going beyond illustrating the feasibility of estimating Gale-Shapley-type two-sided matching models using the method of simulated moments, the results in Table 2, especially Models IV and VII, provide insights into the choices of teachers and hiring authorities and how these choices interact to determine the allocation of teachers across jobs. Estimates of the coefficients for the racial composition of schools and teachers own race, as well as their interaction, for example, bear on long standing questions going back to the work of Antos and Rosen (1975) concerning racial preferences in teacher labor markets. We find that hiring authorities do value non-white candidates in their selection of teachers, which is consistent with the under-representation of non-white teacher candidates in the pool. In addition, white teachers prefer schools with a greater proportion of white students, while non-white teachers do not share this preference.

Distance: Our most striking finding concerning teacher preferences is the importance of distance as a determinant of teachers' evaluations of school alternatives.[29] For example, consider two schools, one having attributes equal to the averages for all those urban schools in our sample for year 2000, with the second school having attributes equaling the averages for the suburban schools hiring that same year. Compared to the representative suburban school, the urban school has far more minority students (a 60 percentage point difference), more students living in poverty (a 52.2 percentage point difference in free-lunch eligibility) and slightly lower starting salaries (-$221). In spite of these differences, the parameter estimates in Model IV indicate that a white teacher four miles from such an urban school would prefer teaching there, provided that the suburban alternative was at least 14 miles away.

[29] In related work we have found that a school's geographical proximity is important in determining whether an individual teaching there decides to transfer to another school or to leave teaching altogether (Boyd, Lankford, Loeb, and Wyckoff, 2005b).

The important of such distance measures in the choices made by teachers likely reflect more than merely a preference for proximity. In fact, our earlier findings indicate that similarity of place is part of the explanation. Other factors, such as the extent of informal networks and information availability, proxied by distance, might also be important.

The importance of distance might vary across teachers, possibly reflecting both observed and unobserved heterogeneity. We explore this possibility by employing a random coefficient specification for the coefficient β_6 in (3). In particular, consider the case where $-\beta_{6j}$ for the jth teacher is distributed as a log-normal random variable where the mean and standard deviation for the corresponding normal density are $\mu_j = \mu_o + \gamma q_j$ and $\sigma *$, respectively.[30] Here μ_o, γ and $\sigma *$ are parameters and q_j is an observed teacher attribute. This specification allows for the possibility that the distribution of unobserved heterogeneity varies with one or more observed teacher attributes. The specification in Model IX shown in Table 3 is the same as Model IV, also included, except that the fixed coefficient for distance in the latter model is replace by this random coefficient specification where q_j is the teacher-quality index discussed above.

The estimates of μ_o and $\sigma *$ are both statistically significant. Furthermore, allowing for the heterogeneity in the effect of distance (i.e., $\sigma * > 0$) results in a meaningful reduction in the value of the objective function. The estimated sign of γ is consistent with the magnitude of the effect of distance being smaller for teachers having greater qualifications, although the coefficient is not statistically significant. In the case where the teacher-quality index is evaluated at its means (i.e., $q_j = 0$), the estimates of μ_o and $\sigma *$ imply that the mean, median, mode and standard deviation of β_{6j} are -5.21, -4.63, -3.64 and 2.71, respectively. Here the median is

[30] The minus sign in $-\beta_{6j}$ is included since a lognormal random variable is positive and greater distance appears to reduce the attractiveness of a school.

almost identical to the estimate of the distance coefficient in Model IV, as well as the other three specifications in Table 2. However, there is strong evidence that there is significant dispersion with respect to the importance of school proximity for teachers.

Teacher Qualifications: The models in Table 2 also shed light on the preferences of hiring authorities. We find evidence that hiring authorities favor teachers with stronger qualifications. This result is pertinent to the issue raised by Ballou (1996) concerning whether schools and districts seek out stronger candidates. Ballou finds that some potential teachers with stronger qualifications do not get teaching jobs while many of their lesser qualified peers do. He uses this result as evidence that hiring authorities do not value these qualifications. However, his data does not allow him to distinguish the choices of employers from the unwillingness of teachers to teach in the schools where they can get jobs. Candidates with stronger qualification simply may not be willing to teach in the schools where lesser qualified candidates obtain jobs. Table 2 provides some evidence that hiring authorities do choose candidates with stronger qualifications when given the opportunity.

To explore employers' preferences for teach qualifications further, we employ a richer model specification shown in Model X where the scalar index of teacher quality is replaced by the following set of variables: a dummy variable indicating whether a teacher has no more than a BA degree, the individual's score on the liberal arts and sciences certification exam, a dummy variables indicating whether a teacher's undergraduate degree was from a highly selective college and a dummy variable indicating the individual graduated from an institution rated by Barrons as being least selective. In effect, the relative importance of these teacher attributes are estimated in Model X, rather than predetermined by us in the weights used to construct the teacher quality index.[31]

The estimates for the parameters in the cubic specification for the certification exam score make clear that a nonlinear specification is warranted. Figure 3 shows the score's estimated effect on the value of the employer's criterion function, $v()$, along with 95 percent confidence boards. The horizontal axis intercept results from normalizing the score in the cubic specification using the score cut-point (220) that determines whether test-takers pass the exam. Thus, the vertical axis measures the change in the value of the criterion function for employers from a teacher having the score shown on the horizontal axis in comparison to a score of 220. For example, the value of an employer's criterion function is estimated to be larger by 0.48 for a teacher having a score of 240 than for an otherwise identical teacher with a score of 220. The magnitude of this effect for what is roughly a one standard deviation change in the score is large relative to the effects of the other teacher attributes and is meaningful relative to the composite effect of all unmeasured factors captured by the error term; here a one standard deviation increase in the score has an effect roughly half as large as a one standard deviation increase in the error term.

Our results concerning the certification exam score is especially interesting given that employers do not know the scores of job applicants – only whether the individuals are certified, which requires passing the test. Thus, the liberal arts and sciences certification exam score must be a good proxy for one or more of the teacher attributes hiring authorities do observe and care about.

---

[31] Rather than q_j in $\mu_j = \mu_o + \gamma q_j$ representing the teach-quality index as in Model IX, q_j in Model X is specified to be the general knowledge certification exam score.

Changes in the exam score have little effect when scores are above 250. A possible explanation is that hiring authorities only care about minimum qualifications. However, an alternative explanation is that the certification exam is designed to measure minimum acceptable competency, as defined by the state, and does not measure higher achievement accurately. In fact, State officials have told us that differences in scores among higher scoring teachers are not informative. Whether or not this is the case, the distribution of scores below 250 is almost identical to that for the left side of a normal distribution; while, there is a marked compression of scores above 250. The maximum score is 300. Thus, the flat relationship for higher scores could be due to the test being a poor measure of teacher achievement at higher levels.

The estimated effects for college selectivity are somewhat mixed and the estimated effect of a teacher having no more than a BA is small and statistically insignificant.. Holding the certification exam score constant, attending a least-competitive college appears to reduce employers' valuation. However, the estimated effect of having attended a most competitive college is also negative and statistically significant, although the size of the effect is small relative to that for either the other Barrons dummy or the certification exam score. Here it is pertinent to note that when the certification exam score is dropped from the model (not shown), the estimated effect of having attended a selective college is positive (0.04) and statistically significant. In general, we find that the estimates of the two dummy variables characterizing college selectivity are sensitive to the functional form used to capture the certification exam score (e.g., quadratic vs. cubic).

Salary: Empirical two-sided match models like those we have estimated can be used to simulate the pay differentials needed to attract particular individuals to work in jobs having particular characteristics, as well as the pay increase needed to improve the quality of workers taking jobs with particular attributes. Current court cases addressing educational inequities and inadequacies, such as the Campaign for Fiscal Equity, Inc. case in New York State, demonstrate the widespread interest in these predictions.

Even though compensating differentials will shed light on how salary would have to be adjusted to attract an individual teacher to a particular job, such measures in themselves do not answer questions regarding the pay increase that would be needed for a particular school or district to improve the quality of teachers hired by a given amount or how more general reforms would change the distribution of teachers across schools. These questions must be answered in the context of the equilibrium matching of workers to jobs, which will depend upon the structure of preferences, the numbers of job openings and candidates, and the distributions of their attributes in the labor market considered.

Whether a particular salary increase will yield a new stable equilibrium satisfying a particular policy target will depend upon the realized values of all the random variables in the model. Not observing these values, one can define policy simulations in terms of the salary changes needed for the expected values of attributes characterizing newly hired teachers in the school(s) to satisfy the policy target. The methods we employed in estimation are directly applicable here. Let $\tilde{q}^*_{mtk}$ represent the teacher attributes included in the policy targets and $\tilde{q}^T_{mtk}$ denote the variable values characterizing those policy targets. For a particular estimated model, the change in the salary paid by the k^th^ school that would be needed so that the expected attributes of the teacher(s) hired would match the target $\tilde{q}^T_{mtk}$ equals D^*_{mtk}, which is implicitly defined in the expression $E\left(\tilde{q}^*_{mtk} \mid S_{mtk} + D^*_{mtk}, Z_{mtk}; \hat{\theta}\right) = \tilde{q}^T_{mtk}$ where S_{mtk} is the initial salary paid and Z_{mtk} represents the other attributes of job k. Carrying out such a policy simulation involves finding the value of D^*_{mtk} which results in the simulated value of this moment condition being satisfied.[32] Alternatively, simulations can be used to analyze the various effects of a given change in salary or other school attributes.

To illustrate the usefulness of our two-sided matching model in such policy simulations, we use the results from Table 2 to consider three alternative policies intended to redistribute teachers across schools so that the quality of teachers in urban schools more closely mirrors that for the metropolitan area as a whole: (1) the salary paid by the urban district to teachers in all district schools increases equally until the mean expected quality of teachers hired by the district equals the metropolitan average; (2) the district increases salaries by differing amounts across schools where the increases are the minima needed for the expected quality of teachers hired in each school to be at least as high as the metropolitan average; and (3) the salaries paid by the urban district are allowed to vary across schools so that the expected quality of teachers hired by each urban school equals the metropolitan average. In this third scenario it is possible for at least some urban schools to have salary reductions.

Our two-sided matching model provides a straight-forward method for estimating the salary changes need to achieve these policy goals. For each of the labor markets, we can simulate each of the salary change policies. Using student-t distribution estimates (Model VIII), we find that for the Syracuse metropolitan area a salary increase of $16,000 for all urban teachers would equalize the expected quality of urban and suburban teachers. Using the normal random errors estimates (Model IV), the predicted salary increase is slightly smaller, while using the random-effects estimates (Model VIII), the predicted salary increase is somewhat larger. If instead of increasing salaries in all urban schools by the same amount, salaries were increased by

[32] In such simulations, problems can arise in that there might not be a stable matching consistent with the policy target(s) or such an equilibrium might not be unique.

varying amounts only in schools not otherwise meeting the expected quality target, the average salary increase in the urban district drops to $13,900. The standard deviation in urban salary increases from zero to $11,180. If we, further, allowed salaries to drop in those schools that exceeded the expected quality goal, the needed average salary increase drops further to $10,410 with a standard deviation of $14,320.

Note that the three salary changes result in quite different distributions of teachers across schools. Prior to the policy change, urban schools had an average expected teacher quality of –0.19 compared with 0.09 in the suburbs. The standard deviation in the expected quality of urban teachers was 0.36. The first approach equalizes average expected quality between urban and suburban schools, while the standard deviation among urban schools drops to 0.21. The second and third approach dramatically reduce the variation across urban schools, from a standard deviation of 0.06 when no salaries are decreased to 0.003 (essentially zero) when salaries in some urban schools decrease.

The constant dollar increase in the salary paid all urban teachers resulted in a marked reduction in the standard deviation of expected quality within the urban district (0.36 to 0.206), even though the increase in all urban salaries did not affect the relative attractiveness of various urban schools. The explanation is that prior to the salary increase teachers found urban schools the relatively least attractive places to work in the metropolitan area, resulting in the teachers hired in those schools being drawn from the lower tail of the quality distribution, where the absolute differences in teacher quality are quite large. With the district-wide salary increase making all urban schools more attractive relative to suburban schools, more of the teachers hired in urban schools are drawn from the quality distribution closer to the mode where there is significantly less dispersion in teacher quality. The result is a reduction in the dispersion of teacher quality across urban schools and an increase within the suburbs.

Even though the general salary increase would reduce the dispersion in teacher quality between urban schools, substantial variation would remain. In fact, 40 percent of the metro-wide variation in expected teacher quality across hires corresponds to variation within the urban district. Targeted salary increases would eliminate both the systematic differences between urban and suburban schools and the systematic dispersion within the urban district – at a total cost less than that associated with the across the board salary increase.

An important caveat for these simulations is that there is relatively little salary variation in our data so that the simulations consider salary changes that far exceed the range observed in the data. Preferences for salary may be non-linear and, thus our estimates may not capture the true effects of substantial salary change. In addition, the simulations consider changing salary so as to redistribute a given set of teachers. In reality, increasing urban salaries likely would also attract new teacher candidates, many of whom may well live close to the urban district. In this case, achieving a given policy target might be less costly than indicated in our simulations.

Summary: To summarize this section, across a number of specifications, our estimation procedure produces estimates of positive effects of both teacher academic ability and non-white status on employers' utility. It also produces estimates of a positive effect of salary and of negative effects of urban status, percent of students in poverty, percent of students of a different race/ethnicity, and distance of school from home on teachers' utility. We have also demonstrated the usefulness of this approach for policy simulations. As we will show below, these consistent estimates of expected sign differ dramatically from those obtained through regression-based approaches.

VIII. Regression Model Estimates and Simulations

This section compares the above estimates with those from the traditional regression-based approach. Table 4 reports parameter estimates for the following two equations:

$$\begin{aligned} \text{salary} &= \beta_0 + \beta_1(\text{Tquality}) + \beta_2(\text{Sminority}) + \beta_3(\text{Spoverty}) + \beta_4(\text{urban}) + \varepsilon \\ \text{T}quailty &= \alpha_0 + \alpha_1(\text{salary}) + \alpha_2(\text{Sminority}) + \alpha_3(\text{Spoverty}) + \alpha_4(\text{urban}) + \eta \end{aligned} \qquad \textbf{(4)}$$

The first equation is an example of a first-stage (reduced-form) hedonic wage locus where a second-step could be used to estimate the underlying preference/technology parameters. We also include a specification having quality as the dependent variable because some studies have used this approach as an alternative to the traditional wage equation (Loeb and Page, 2001). Fixed effects for years and for metropolitan areas are included in columns II and IV of each panel. Estimates in column III include a dummy variable for whether or not the teacher is non-white and an interaction of non-white with the percent of minority students. Column IV estimates include measures of distance to job: both a continuous measure of distance for those who travel 100 miles or less to their job and a dummy variable for traveling farther.

The regression models produce typically inconsistent results. In the wage equation, salaries are higher in schools with higher proportions of minority students. Yet, there appears to be no premium for better teacher qualifications, and teachers are willing to take lower salaries to teach in schools with high proportions of children in poverty and in urban schools. In the quality equations, there is again no relationship between quality and salary; but at the same wage, schools with higher proportions of poor students appear to attract less-qualified teachers. This specification shows no relationship between qualifications and either urban or the percent of minority students. The one exception to this is for non-white teachers whose qualifications are lower in high proportion minority schools.[33] Clearly, it would be difficult to draw policy implications from these inconsistent results.

Given the wide use of the hedonic model, it is pertinent to investigate further why a wage model and our empirical two-sided matching model yield such different results. We do this by carrying out Monte Carlo simulations. Preferences are assumed to be as follows for 150 employers each having on average 3 openings and 450 teachers seeking those positions.

$$u_{jk} = \beta_1(\mathrm{Z}_1) + \beta_2(Z_2) + \beta_3(salary) + \beta_4(\text{distance}) + \sigma_\delta \delta_{jk}$$

$$v_{jk} = \alpha(Tquality) + \sigma_\omega \omega_{jk}$$

The locations of teachers and schools are represented by scalar variables L_T and L_S, respectively, so that a teacher's distance to a particular school equals $|L_S - L_T|$. The values of L_S, Z_1, Z_2 and *salary* for each school, the values of L_T and *Tquality* for each teacher, as well as the values of the errors terms δ_{jk} and ω_{jk}, were obtained by making 100 sets of independent random draws from the standard normal distribution. For given values of the preference parameters and the standard deviations σ_δ and σ_ω, the teacher-employer stable matching implied by the matching algorithm underlying our model was determined for each of the 100 draws. In turn, the following salary and quality equations were estimated for each draw and mean values of the parameter estimates were computed for the given values of the parameters and correlations.

$$salary = \gamma_0 + \gamma_1(Z_1) + \gamma_2(Z_2) + \gamma_3(Tquality) + \varepsilon$$

$$Tquality = \tau_0 + \tau_1(Z_1) + \tau_2(Z_2) + \tau_3(salary) + \xi$$

In this way, we investigate how differing (i.) the degree of correlation among the variables and (ii.) the preferences of teachers and schools affect parameter estimates in the salary and quality

[33] Estimation of hedonics equations for individual markets yields similarly unintuitive results.

equations. A number of general trends emerge, which are illustrated in Table 5 where the first number in each cell is the average of the parameter estimates and the second is the proportion of the 100 estimates that are statistically significant ($p < .05$).

First, when there is no correlation among variables in the model and teachers do not have preferences over distance, the wage equation gives coefficients that qualitatively reflect preferences. In Comparison 1, β_3, α, σ_δ, and σ_ω all equal 1 and β_4 equals zero. If β_1 and β_2 equal zero the mean estimates of γ_1, γ_2 and γ_3 in the wage equation equal -0.0003, -0.009 and 0.699, respectively. The estimates for γ_3 are statistically significant in all of the simulations, while those for γ_1 and γ_2 are significant 24 percent and 19 percent of the time. If β_1 and β_2 increase to 0.3 and 0.6, respectively, the estimates change to -0.13, -0.27, and 0.73. If β_1 and β_2 increase again to 0.6 and 1.2, the estimates change to –0.22, -0.45 and 0.78, respectively, with the estimates of γ_2 and γ_3 statistically distinguishable from zero in all simulations and the estimate of γ_1 statistically significant in all but one draw.

Second, when salary is correlated with another school characteristics, Z_1, the coefficient on Z_1 reflects that correlation, even if candidates do not value Z_1. Consider the same example as above, except with β_1 and β_2 equal to zero (Comparison 2). When the correlation between Z_1 and salary equals zero, the mean estimates of γ_1, γ_2 and γ_3 equal -0.0003, -0.009, and 0.70. The estimate of γ_1 is significant in 24 percent of the simulations. When the correlation is 0.3, the estimates are 0.16, -0.01 and 0.67; and γ_1 is significant in 97 percent of the simulations. When the correlation is 0.6, the coefficient estimates are 0.38, -0.01 and 0.54, not reflecting the underlying preferences for Z_1 at all. Furthermore, the estimate of γ_1 is significant in all of the simulations.

Third, when distance enters candidates' preferences, or similarly when a relative increase in noise raises the variance of the errors, the estimated coefficients in the wage equation drop in magnitude. This happens even when distance is not correlated with any other measure. Comparison 3 uses the reference parameter values (β_1, β_2, and β_3 equal 0.5, 0.5 and 1.0, respectively). If β_4 equals zero, the mean estimates of γ_1, γ_2 and γ_3 are –0.22, -0.23 and 0.74. When β_4 equals -0.5, the mean estimates change to –0.21, -0.22 and 0.72; and when β_4 equals –1.0, the mean estimates change to -0.20, -0.21 and 0.69. When β_4 equals -1.5, the mean estimates fall further to –0.19, -0.20 and 0.67. Increases in σ_δ and σ_ω also decrease the estimates in the wage equation (results not shown in Table 4). If β_1 and β_2 equal zero and β_3 equals 1.0, the mean estimate of γ_3 is 0.70 when the standard deviations of the errors equal 1.0. When the standard deviations drop to 0.5, the mean estimate of γ_3 increases to 0.90. When they increase to 1.5, $\tilde{\gamma}_3$ drops to 0. 50.

Fourth, if candidates prefer closer schools and schools that are closer to more qualified candidates systematically differ in their characteristics, then the estimated wage equation will misrepresent the value teachers place on these characteristics. Using the example in which β_1, β_2, β_3 and β_4 equal 0.5, 0.5, 1.0, and –1.0 respectively (Comparison 4), if the correlations both between qualifications and teacher location and between Z_1 and school location are zero then the mean estimated coefficient for Z_1 is -0.20 (statistically significant in 96 percent of simulations). When these correlations equal 0.3,[34] the mean estimated coefficient is -0.21. When the correlations equal 0.6, the mean estimate is -0.261, one-third larger in magnitude compared to the case where there is no such spatial proximity.

[34] As these correlations increase, teachers having greater qualifications, on average, live increasingly close to schools having higher values of Z_1.

Finally, while the quality equation more accurately reflects the underlying preferences than the wage equation in some instances, it is also subject to potential biases. For example, the increasing correlation between Z1 and salary has little effect on the predicted relationship between salary and quality in the quality equation, while it reduces the estimated relationship substantially in the wage equation (Comparison 2). Furthermore, in Comparisons 1 and 2 the tests of statistical significance for the quality equation yield results that more accurately reflect the underlying parameter values, more so than the tests of statistical significance for the wage equation. On the other hand, increased error such as that resulting from the importance of distance has an approximately equal effect on the estimated relationship in the salary and quality equation estimates (Comparisons 3). The same is true for increasing the geographical proximity of more qualified teachers and schools having higher values of Z_1 (Comparison 4).

In summary, the simulations show substantial differences between the standard wage model approach for estimating compensating differentials and our approach. It is not altogether clear which approach is closer to the truth.[35] The algorithm that we use for matching teachers to jobs relies on an unrealistic deferred acceptance procedure in which no teacher commits herself to a job until all other workers are in a match, teachers only apply to jobs in the small labor market in which they end up working in, and all teachers initially apply to all schools. However, even with these shortcomings, there are a number of reasons to believe that the new approach is likely to reveal more accurate estimates of worker and employer preferences. In particular, the new approach easily incorporates heterogenous preferences, allows for thin labor markets, and directly models preferences. These simulations, the theoretical considerations discussed in section IV, and the stark differences in the empirical results described above together suggest to

[35] Here one might distinguish between the standard two-step hedonic approach and the one-step structural estimation approach of Eckland, Heckman and Nesheim (2004).

us that the matching model employed here provides a preferable framework for analyzing teacher labor markets.

IX. Conclusion

Descriptive analyses of teacher labor markets point to a high degree of systematic sorting of teachers across schools. Yet, regression-based empirical models have not produced consistent estimates for understanding this sorting. In this paper we have used method of simulated moments estimates of two-sided matching models. This empirical matching model produces estimates in keeping with the hypotheses that schools prefer high ability teachers and teachers prefer higher wages and schools that are closer to home with fewer poor students. While these results may appear predictable, they contradict many of the findings from prior research on compensating differentials in the teacher labor market. For example, these prior studies, as well as our estimates of wage regressions with the data used in this study, often show little relationship between wages and teacher attributes. This result has been used to suggest that districts do not care about teacher quality and teachers do not care about wages. Similarly, a negative estimate of the relationship between student poverty rate and wages could be used to suggest that teachers prefer teaching in high-poverty schools. Yet, these estimates are driven simply by low wages in high-poverty districts and are unlikely to reflect an underlying preference for poverty. The two-sided matching approach finds a far more likely negative valuation of poverty on teachers' preferences.

Even though this paper focuses on worker-job match within the context of teacher labor markets, the issues raised and the empirical framework employed are relevant in the analysis of markets that fail to clear or are thin as well as situations in which idiosyncratic heterogeneity in preferences, such as that related to geography, is important. The theoretical points made in Section IV, the simulations in Section VIII, and the differences between the estimates for the regression models and two-sided matching models bring into question the common practice of employing regression-based wage models to estimate compensating differentials in such settings. There is reason to think that the wage locus in such cases may not reflect marginal evaluations on either side of the market. The empirical framework and estimation strategy developed in this paper may prove useful in such cases.

In summary, this paper is a step toward understanding the functioning of teacher labor markets and the factors that influence teachers' decisions about whether and where to teach and schools' decisions about which teachers to hire. The matching model shows promise for estimating compensating differentials and the preferences of both employers and employees in labor markets not characterized by perfect competition and the rapid adjustment of wages.

References

Abowd, John M. and Francis Kramarz (1999). ‘‘The Analysis of Labor Markets Using Matched Employer-Employee Data,’’ in O. Ashenfelter and D. Card, eds. *Handbook of Labor Economics*, Amsterdam, North Holland.

Antos, J., and S. Rosen (1975). “Discrimination in the Market for Public School Teachers,” *Journal of Econometrics, 3*(2) 123-50.

Ballou, D. (1996). “Do public schools hire the best applicants?” *Quarterly Journal of Economics, 111(*1), 97-133.

Ballou, D., and M. Podgursky (1997). *Teacher pay and teacher quality*. Kalamazoo, MI: E. Upjohn Institute for Employment Research.

Baugh, W.H., and J.A. Stone (1982). “Mobility and wage equilibration in the educator labor market,” *Economics of Education Review, 2*(3), 253-274.

Ben-Akiva, M. and S.R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA.

Berry, S. T. (1992) “Estimation of a Model of Entry in the Airline Industry,” *Econometrica* 60: 889-917.

Bertschek, I. and M. Lechner (1998). “Convenient Estimators for the Panel Probit Model,” Journal of Econometrics, 87(2), 329-371.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2002). “Initial Matches, Transfers, and Quits: The Role of Teachers’ Career Decisions in the Disparities in Average Teacher Qualifications Across Schools,” Working Paper.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2005a). “The Draw of Home: How Teachers’ Preferences for Proximity Disadvantage Urban Schools,” *Journal of Policy Analysis and Management*, 24(1), 113-132.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2005b). “Explaining the Short Careers of High-Achieving Teachers in Schools with Low-Performing Students,” *American Economic Review Papers and Proceedings*, 95(2), 166-171.

Brewer, D.J. (1996). “Career paths and quit decisions: Evidence from teaching.” *Journal of Labor Economics, 14*(2) (April), 313-339.

Bridges, E.M. (1996). “Evaluation for tenure and dismissal.” In Millman, J. & Darling-Hammond, L. (Eds.), *The New Handbook of Teacher Evaluation* (pp.147-157). Newbury Park: Sage Publications.

Brown, C (1980). “Equalizing Differences in the Labor Market,” *Quarterly Journal of Economics, 94*(1): 113-34.

Bunch, D.S. (1991), “Estimability in the Multinomial Probit Model,” *Transportation Research B, 25B*(1), 1-12.

Bunch, D.S. and R. Kitamura (1989), “Multinomial Probit Model Estimation Revisited: Testing Estimable Model Specifications, Maximum Likelihood Algorithms, and Probit Integral Approximations for Trinomial Models of Household Car Ownership,” Working Paper, University of California at Davis.

Burdett, K. and M. Coles (1999). “Long-term partnership formation: marriage and employment,” *The Economic Journal 109*, F306-F334.

Choo, E., and A. Siow (2006) “Who Marries Whom and Why,” *Journal of Political Economy*, 114(1), 175-201.

Dansie, B.R. (1985), “Parameter Estimability in the Multinomial Probit Model,” *Transportation Research B, 19B*(6), 526-528.

Dolton, P.J. (1990). “The economics of UK teacher supply: The graduate's decision,” *The Economic Journal, 100*(400): 91-104.

Dolton, P.J. and G.H. Makepeace (1993). “Female labour force participation and the choice of occupation,” *European Economic Review, 37*: 1393-1411.

Dolton, P.J. and W. van der Klaaw (1999). “The turnover of teachers: A competing risks explanation,” *Review of Economics and Statistics, 81*(3): 543-552.

Eckstein, Z. and G. J. van den Berg (forthcoming) “Empirical Labor Search: A Survey,” *Journal of Econometircs.*

Ekeland, I., J.Heckman and L. Nesheim (2004). “Identification and Estimation of Hedonic Models,” *Journal of Political Economy*, Part 2 Supplement 112(1): S60-109.

Epple, D. and H. Sieg (2001). “Interjurisdictional Sorting and Majority Rule: An Empirical Analysis,” *Econometrica, 69*(6): 1437-65

Freeman, A. M. (1979). “Hedonic Prices, Property Values and Measuring Environmental Benefits: A Survey of the Issues,” *Scandinavian Journal of Economics, 81*(2):154-173.

Gale, D. and L. Shapley (1962). “College Admissions and the Stability of Marriage,” *American Mathematical Monthly 69*: 9-15.

Hansen, L.P. (1983). “Large Sample Properties of Generalized Method of Moments Estimators,” Econometrica, 50(4), 1029-1054.

Hanushek, E.A., J.F. Kain, and S.G. Rivkin (1999). "Do higher salaries buy better teachers?" Working Paper.

Hanushek, E.A., and R.R. Pace (1995). "Who chooses to teach (and why)?" *Economics of Education Review, 14*(2), 101-117.

Hajivassiliou, V. and P. Ruud (1994). "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of Econometrics*, Vol. 4, R. Engle and D. McFadden, eds., Amsterdam: Elsevier, North-Holland, 2383-2441.

Hwang, H., D.T. Mortensen and W.R. Reed (1998). "Hedonic Wages and Labor Market Search," *Journal of Labor Economics, 16*(4): 815-47.

Keane, M.P. (1992), "A Note on Identification in the Multinomial Probit Model," *Journal of Business & Economic Statistics, 10*(2), 193-200.

Klein, R.W. and Spady, R.H. (1993). "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica,* 61(2), 387-421.

Lang, K. and S. Majumdar (2004). "The Pricing of Job Characteristics When Markets Do Not Clear: Theory and Policy Implications," *International Economic Review*, 45(4): 1111-28.

Lankford, H. (1999). *A descriptive analysis of the New York State and New York City teaching force. Report prepared for the New York Supreme Court case Campaign for Fiscal Equity v. New York State.*

Lankford, H., S. Loeb and J. Wyckoff (2002). "Teacher Sorting and the Plight of Urban Schools," *Education Evaluation and Policy Analysis, 24*(1): 37-62.

Loeb, S. and M. Page (2000). "Examining the Link between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-pecuniary Variation," *Review of Economics and Statistics*, 82(3): 393-408.

Lucas, R. (1977). "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling," *America Economic Review, 67*(4): 549-58.

Manski, C.F. (1987). "Academic ability, earnings, and the decision to become a teacher: Evidence from the National Longitudinal Study of the High School Class of 1972." In D.A. Wise (Ed.), *Public Sector Payrolls*. Chicago: University of Chicago Press: 291-312.

Manski, C.F. (1995). *Identification Problems in the Social Sciences*, Cambridge, MA:Harvard University Press.

McLean, H. M. (1978). "Compensating Wage Differentials for Hazardous Work: An Empirical Analysis," *Quarterly Review of Economics and Business, 18*(3): 97-107.

Meng, C.L. and Schmidt, P. (1985), "On the cost of partial observability in the bivariate probit model," *International Economic Review 26*, 71-86.

Mont, D., and D.I. Rees (1996). "The influence of classroom characteristics on high school teacher turnover," *Economic Inquiry, 34*(1): 152-167.

Murnane, R.J., J.D. Singer, and J.B. Willett (1989). "The influences of salaries and "opportunity costs" on teachers' career choices: Evidence from North Carolina," *Harvard Educational Review, 59*(3), 325-346.

Newey, W. and K. West (1987). "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28(3): 777-787.

Pakes, A. and D. Pollard (1989). "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5), 1027-57.

Palmquist, R. B. (1991). "Chapter IV: Hedonic Methods," in John B. Braden and Charles D. Kolstad (Eds). *Measuring the Demand for Environmental Quality,* Elsevier Science Publishers B.V.: 77-120.

Pflaum, S.W. and T. Abramson (1990). "Teacher Assignment, Hiring, and Preparation: Minority Teachers in New York City," *Urban Review 22*(1): 17-31.

Porier, D.J. (1980). "Partial Observability in Bivariate Probit Models," *Journal of Econometrics, 12*, 209-217.

Postel-Vinay, Fabien and Jean-Marc Robin (2002). "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity ", *Econometrica*,70 (6), 2295-2350.

Rickman, B.D., and C.D. Parker (1990). "Alternative wages and teacher mobility: A human capital approach," *Economics of Education Review, 9*(1): 73-79.

Rogerson, R., R. Shimer and R. Wright (2005) "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, 959-988.

Rosen, S. (1974). "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy 82*(1): 34-55.

Rosen, S. (1986). "The Theory of Equalizing Differences," in Ashenfelter, E. and R. Layard, *Handbook of Labor Economics Volume I*. Elsevier Science Publishers: 642-692.

Roth, A.E. and M.A. Sotomayer (1990) *Two-sided matching: A study in game-theoretic modeling and analysis*. Cambridge: Cambridge University Press.

Roth, A.E. and J.H. Vande Vate (1990) “Random Paths to Stability in Two-Sided Matching,” *Econometrica*, 58(6): 1475-1480.

Sattinger, M. (1979) “Differential Rents and the Distribution of Earnings,” *Oxford Economic Papers*, 31(1): 60-71.

Sieg, H. (2000) “Estimating a Bargaining Model with Asymmetric Information: Evidence from Medical Malpractice Disputes,” *Journal of Political Economy, 108*(5): 1006-1021.

Stinebrickner, T.R. (2000). An analysis of occupational change and departure from the labor force: Evidence of the reasons that teachers quit. Working paper.

Stinebrickner, T.R. (1999). “Estimation of a duration model in the presence of missing data.” *Review of Economics and Statistics, 81*(3), 529-542.

Stinebrickner, T.R. (1998). “An empirical investigation of teacher attrition,” *Economics of Education Review, 17*(2): 127-136.

Theobald, N.D. (1990). “An examination of the influences of personal, professional, and school district characteristics on public school teacher retention,” *Economics of education review, 9*(3), 241-250.

Theobald, N.D., and R.M.Gritz, (1996). “The effects of school district spending priorities on the exit paths of beginning teachers leaving the district,” *Economics of Education Review, 15*(1), 11-22.

Wong, L.Y., (2003) “Structural Estimation of Marriage Models,” *Journal of Labor Economics*, 21(3), 699-727.

Figure 1: Utility and Rankings of Candidates and Schools

(A)
Candidates' benefits from alternative employment

$$\begin{array}{ccccc} & s_1 & s_2 & \cdots & s_K \\ c_1 & u_{11} & u_{12} & \cdots & u_{1K} \\ c_2 & u_{21} & u_{22} & \cdots & u_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_J & u_{J1} & u_{J2} & \cdots & u_{JK} \end{array}$$

(B)
Schools' benefits from alternative candidates

$$\begin{array}{ccccc} & s_1 & s_2 & \cdots & s_K \\ c_1 & v_{11} & v_{12} & \cdots & v_{1K} \\ c_2 & v_{21} & v_{22} & \cdots & v_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_J & v_{J1} & v_{J2} & \cdots & v_{JK} \end{array}$$

(C)
Candidates' rankings of employers

$$\begin{array}{ccccc} & s_1 & s_2 & \cdots & s_K \\ c_1 & r_{11}^c & r_{12}^c & \cdots & r_{1K}^c \\ c_2 & r_{21}^c & r_{22}^c & \cdots & r_{2K}^c \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_J & r_{J1}^c & r_{J2}^c & \cdots & r_{JK}^c \end{array}$$

(D)
Schools' rankings of candidates

$$\begin{array}{ccccc} & s_1 & s_2 & \cdots & s_K \\ c_1 & r_{11}^s & r_{12}^s & \cdots & r_{1K}^s \\ c_2 & r_{21}^s & r_{22}^s & \cdots & r_{2K}^s \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_J & r_{J1}^s & r_{J2}^s & \cdots & r_{JK}^s \end{array}$$

Figure 2: Resulting Matching of Teachers and Jobs

School-teacher matched pairs

$$\begin{Bmatrix} (s_1, c_{1'}) \\ (s_2, c_{2'}) \\ \cdots \\ (s_K, c_{K'}) \end{Bmatrix}$$

Joint distribution of school and teacher attributes

$$\begin{bmatrix} z_1 & q_{1'} \\ z_2 & q_{2'} \\ \vdots & \vdots \\ z_K & q_{K'} \end{bmatrix} = \begin{bmatrix} z & q \end{bmatrix}$$

Figure 3
Estimated relative evaluations of applicants by hiring authorities, varying the general knowledge teacher certification exam score, point estimates and 95 percent confidence bands

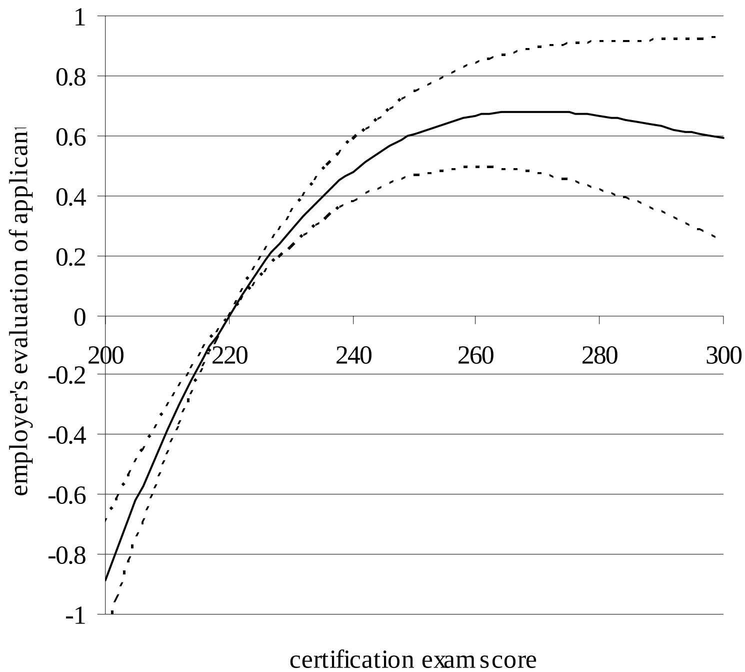

Table 1: Descriptive Statistics: Elementary Schools and K-6 Teachers Hired

Schools			Teachers		
Variable	Mean	Std Dev	Variable	Mean	Std Dev
Sminority	0.210	0.293	Tminority	0.064	0.246
Spoverty, K-6	0.293	0.265	Tquality Index	0.00	1.00
Urban	0.217	0.293	BA or less	0.505	0.500
Salary	32,458	2,607	Score	260.217	18.441
			Highly Selective	0.134	0.340
			Least Selective	0.041	0.198
			Distance to Job (miles)	24.616	115.27
			Distance if < 50 miles	8.638	8.349
N = 2443			N = 5028		
Year			MSAs/Regions		
1995	0.109		Albany	0.178	
1996	0.123		Buffalo	0.251	
1997	0.151		Rochester	0.350	
1998	0.139		Syracuse	0.167	
1999	0.211		Utica-Rome	0.055	
2000	0.267				

Note: Salaries are for 2000. If the 2000 salaries were not available due to districts operating out of contract, we used salary information for the most recent prior year and inflated the value using the average percent change across districts with salaries in both years. Only 4 percent of the sample traveled more than 100 miles to their job.

Table 2: Estimated Parameters in Employers' and Employees' Criterion Functions

	Model I	Model II	Model III	Model IV	Model V	Model VI	Model VII	Model VIII
Employers' Criterion Function								
Tquality Index	0.2053	0.2722	0.1773	0.1616	0.1226	0.2703	0.1923	0.2465
	(0.0080)	(0.0367)	(0.0211)	(0.0079)	(0.0052)	(0.0328)	(0.0055)	(0.0090)
Tminority				0.2011	0.2654	-0.9208	0.2736	0.0024
				(0.0457)	(0.0814)	(0.1304)	(0.0494)	(0.0613)
Candidates' Criterion Function								
Salary ($1000s)	0.1099	0.1468	0.2085	0.1837	0.1508	0.2240	0.1261	0.1661
	(0.0475)	(0.0382)	(0.1043)	(0.0654)	(0.0576)	(0.0333)	(0.0490)	(0.0342)
Urban	-1.5535	0.6597	0.4801	-2.3090	-1.0163	0.6803	-1.0091	-2.0708
	(0.4699)	(0.1304)	(0.2045)	(0.3747)	(0.3540)	(0.1047)	(0.3262)	(0.3808)
Spoverty, K-6	-2.0367	-1.9877	-2.2647	-0.8066	-2.2224	-0.8074	-0.6436	-1.2969
	(0.4072)	(0.3909)	(0.1714)	(0.1926)	(0.4674)	(0.1721)	(0.2455)	(0.2198)
Sminority	-3.9371	-1.0227	0.0068					
	(0.4738)	(0.1696)	(0.1565)					
Sminority for Non-White Teachers				2.3676	2.2070	0.3658	0.7207	1.3856
				(0.5454)	(0.6882)	(0.3597)	(0.6864)	(0.3591)
Sminority for White Teachers				-3.5916	-1.8890	-1.0708	-5.4174	-4.0778
				(0.4731)	(0.4951)	(0.1637)	(0.4169)	(0.4703)
Distance ln(D+1)	-4.5342			-4.6241			-4.5181	-4.3741
	(0.2367)			(0.3370)			(0.4472)	(0.3465)
School random effect variance							0.3178	
							(0.0327)	
Objective	0.7883	19.4054	0.0220	0.6488	19.2455	0.0432	0.6000	0.5248

Note: Standard errors reported in parentheses. Models II and V include distance in the moment conditions while Models III and VI do not. Models I-VI use normal random errors, Model VII uses school random effects, and Model VIII assumes a Student-t distribution for the errors.

Table 3: Estimated Parameters in Employers' and Employees' Criterion Functions

	Model IV	Model IX	Model X
Employers' Criterion Function			
BA or less			0.0059
			(0.0177)
(Score – 220)			0.0333
			(0.0034)
$(\text{Score} - 220)^2$			-5.10E-4
			(8.08E-5)
$(\text{Score} - 220)^3$			2.32E-6
			(7.41E-7)
Barrons-1			-0.0517
			(0.0227)
Barrons-4			-0.1595
			(0.0468)
Tquality Index	0.1616	0.2740	
	(0.0079)	(0.0091)	
Tminority	0.2011	0.1824	0.2721
	(0.0457)	(0.0547)	(0.0533)
Candidates' Criterion Function			
Salary ($1000s)	0.1837	0.1594	0.0404
	(0.0654)	(0.0687)	(0.0217)
Urban	-2.3090	-2.0627	-2.0885
	(0.3747)	(0.3055)	(0.3139)
Spoverty, K-6	-0.8066	-0.8339	-1.3273
	(0.1926)	(0.3504)	(0.1782)
Sminority for Non-White Teachers	2.3676	1.8506	2.3197
	(0.5454)	(0.6682)	(0.4703)
Sminority for White Teachers	-3.5916	-3.8511	-3.4403
	(0.4731)	(0.4780)	(0.3712)
Distance ln(D+1)	-4.6241		
	(0.3370)		
Distance – μ_o		1.5306	1.4291
		(0.0615)	(0.0832)
Distance – γ		-0.1161	-2.71E-4
		(0.0759)	(1.38E-3)
Distance – σ*		0.4891	0.6846
		(0.0900)	(0.0725)
objective	0.6488	0.6166	1.0174

Note: Standard errors reported in parentheses. The moment conditions for Models IV and IX are the same. In Models X the teacher quality index is replaced by the individual teacher attributes shown.

Table 4: Hedonic and "Quality Hedonic" Results

Variable	Salary				Quality			
	I	II	III	IV	I	II	III	IV
Salary					-0.013 (0.014)	-.018 (0.017)	-0.016 (0.017)	-0.018 (0.017)
Tquality Index	-0.012 (0.014)	-0.012 (0.011)	-0.011 (0.011)	-0.012 (0.011)				
Sminority	***1.30*** ***(0.11)***	***1.37*** ***(0.10)***	***1.35*** ***(0.10)***	***1.37*** ***(0.10)***	0.090 (0.113)	-0.16 (0.12)	-0.030 (0.124)	-0.16 (0.12)
Spoverty, K-6	***-1.16*** ***(0.11)***	***-1.12*** ***(0.09)***	***-1.12*** ***(0.09)***	***-1.12*** ***(0.09)***	***-0.62*** ***(0.11)***	***-0.52*** ***(0.12)***	***-0.51*** ***(0.12)***	***-0.52*** ***(0.12)***
Urban	***-0.21*** ***(0.08)***	***-0.18*** ***(0.07)***	***-0.18*** ***(0.07)***	***-0.18*** ***(0.06)***	0.025 (0.079)	0.12 (0.08)	0.11 (0.08)	0.12 (0.08)
Tminority			-0.058 (0.073)				0.13 (0.09)	
Sminority * Tminority			0.14 (0.12)				***-0.79*** ***(0.15)***	
Distance				-3.4E-4 (8.8E-5)				*2.0E-3* *(1.1E-3)*
Dist > 100 miles				-0.081 (0.057)				*0.12* *(0.70)*
Year fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
MSA fixed effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes
R^2	0.0328	0.3593	0.3595	0.3596	0.0189	0.0409	0.0495	0.0410

Note: Standard errors reported in parentheses

Table 5: Mean Estimates of Parameters in Wage and Quality Equations

Monte Carlo Simulations Varying Either the Preference Parameters or Correlations Between Explanatory Variables

	values of preference parameters					mean estimates of parameters in wage equation			mean estimates of parameters in quality equation		
	β_1	β_2	β_4	ρ_{Z_1S}	ρ_{qL}, ρ_{Z_1L}	$\tilde{\gamma}_1$	$\tilde{\gamma}_2$	$\tilde{\gamma}_3$	$\tilde{\tau}_1$	$\tilde{\tau}_2$	$\tilde{\tau}_3$
Comparison 1: Variation in the preference parameters for Z_1 and Z_2	**0.0**	**0.0**	0.0	0.0	0.0	-0.0003 0.24	-0.009 0.19	0.699 1.00	0.004 0.08	-0.002 0.08	0.692 1.00
	0.3	**0.6**	0.0	0.0	0.0	-0.131 0.84	-0.274 0.99	0.732 1.0	0.183 1.0	0.363 1.0	0.609 1.0
	0.6	**1.2**	0.0	0.0	0.0	-0.215 0.99	-0.447 1.0	0.776 1.0	0.282 1.0	0.564 1.0	0.470 1.0
Comparison 2: Variation in the correlation between Z_1 and salary	0.0	0.0	0.0	**0.0**	0.0	-0.0003 0.24	-0.009 0.19	0.699 1.0	0.004 0.08	-0.002 0.08	0.692 1.0
	0.0	0.0	0.0	**0.3**	0.0	0.165 0.97	-0.011 0.22	0.667 1.0	0.001 0.08	0.001 0.02	0.693 1.0
	0.0	0.0	0.0	**0.6**	0.0	0.380 1.0	-0.011 0.27	0.543 1.0	-0.006 0.04	0.002 0.05	0.696 1.0
Comparison 3: Variation in teachers' distance preference parameter	0.5	0.5	**0.0**	0.0	0.0	-0.219 0.98	-0.231 0.99	0.736 1.0	0.302 1.0	0.302 1.0	0.599 1.0
	0.5	0.5	**-0.5**	0.0	0.0	-0.208 0.99	-0.224 0.99	0.719 1.0	0.296 1.0	0.299 1.0	0.590 1.0
	0.5	0.5	**-1.0**	0.0	0.0	-0.195 0.96	-0.211 0.96	0.690 1.0	0.289 1.0	0.292 1.0	0.572 1.0
	0.5	0.5	**-1.5**	0.0	0.0	-0.186 0.95	-0.200 0.97	0.668 1.0	0.284 1.0	0.284 1.0	0.558 1.0
Comparison 4: Variation in the correlations between teachers' location and q and between school location and Z_1	0.5	0.5	-1.0	0.0	**0.0**	-0.195 0.96	-0.211 0.96	0.690 1.0	0.289 1.0	0.292 1.0	0.572 1.0
	0.5	0.5	-1.0	0.0	**0.3**	-0.211 0.97	-0.207 0.97	0.688 1.0	0.313 1.0	0.286 1.0	0.561 1.0
	0.5	0.5	-1.0	0.0	**0.6**	-0.261 1.0	-0.197 0.96	0.687 1.0	0.387 1.0	0.271 1.0	0.529 1.0

Appendix A

As noted in Section V, simulation is used to compute values of $E\left(\tilde{z}_{mtj} \middle| q_{mtj};\theta\right)$ in the moment condition $\sum_t \sum_j q_{mtj}\left[\tilde{z}_{mtj} - E\left(\tilde{z}_{mtj} \mid q_{mtj};\theta\right)\right] = 0$ as well as $E\left(\tilde{d}_{mtj} \middle| q_{mtj};\theta\right)$ in the moment conditions $\sum_t \sum_j q_{mtj}\left[\tilde{d}_{mtj} - E\left(\tilde{d}_{mtj} \mid q_{mtj};\theta\right)\right] = 0$ and $\sum_t \sum_j \left[\tilde{d}_{mtj} - E\left(\tilde{d}_{mtj} \mid q_{mtj};\theta\right)\right] = 0$. Let $F\left(\tilde{z}_{mtj} \mid q_{mtj};\theta\right)$ and $F\left(\tilde{d}_{mtj} \mid q_{mtj};\theta\right)$ represent the simulated values of $E\left(\tilde{z}_{mtj} \middle| q_{mtj};\theta\right)$ and $E\left(\tilde{d}_{mtj} \middle| q_{mtj};\theta\right)$ obtained using the following two-step approach.

Step 1: A random number generator generates H sets of independent draws for the random variables in the model. Each draw generates random numbers corresponding to the random variable(s) in each candidate's benefit equation for every school alternative, denoted by δ_{jk}^h for the hth draw, j = 1,2,…,J and k = 1,2,…,K. Similarly, the hth draw includes randomly generated values for the random error terms (ω_{jk}^h) in the equations characterizing the benefits to each employer associated with hiring each candidate. We hold these randomly generated values, as well as the observed attributes of candidates and schools, constant throughout the estimation. In the first set of models, we assume that δ_{jk}^h and ω_{jk}^h are independent standard-normal random variables and then consider a model with δ_{jk}^h and ω_{jk}^h drawn from the student's t distribution with four degrees of freedom. We also estimate a model with school random effects in the specification of candidates' preferences; $\delta_{jk}^h = \eta_{jk}^h + \eta_{\bullet k}^h$ where $\eta_{\bullet k}^h$ is a random effect for school k, which we assume to be independent of the white-noise random variable η_{jk}^h and the measured school attributes included in the model. Normalizing the variance of δ_{jk}^h to be one, the variances of $\eta_{\bullet k}^h$ and η_{jk}^h are σ^2 and $1-\sigma^2$ respectively.

Step 2: For a given set of parameter values $(\theta = (\alpha, \beta))$ and the random numbers drawn in Step 1 for a particular error structure, we compute the simulated moments as follows. The implied nonstochastic components of utility along with the values of δ_{jk}^{h} and ω_{jk}^{h} for a particular draw imply the individual rankings for candidates and employers. These combined with the Gale-Shapley matching algorithm imply the school-optimal stable matching and the resulting distribution of teacher and job attributes (e.g., $\tilde{z}_{mtj}^{h}$ and $\tilde{d}_{mtj}^{h}$ for each of the candidates hired in the h[th] simulated outcome for market m during period t). Repeating this step for each draw yields the approximations of the pertinent expected values in (A1) and the simulated moment conditions in (A2) and (A3) used in estimation.[36]

$$\begin{aligned} F\left(\tilde{z}_{mtj} \middle| q_{mtj};\theta,H\right) &= \tfrac{1}{H}\sum_{h=1}^{H}\tilde{z}_{mtj}^{h} \approx E\left(\tilde{z}_{mtj} \middle| q_{mtj};\theta\right) \\ F\left(\tilde{d}_{mtj} \middle| q_{mtj};\theta,H\right) &= \tfrac{1}{H}\sum_{h=1}^{H}\tilde{d}_{mtj}^{h} \approx E\left(\tilde{d}_{mtj} \middle| q_{mtj};\theta\right) \end{aligned} \qquad \text{(A1)}$$

$$\begin{aligned} \psi_{mtj}^{a} &\equiv q_{mtj}\left[\tilde{z}_{mtj} - F\left(\tilde{z}_{mtj} \mid q_{mtj};\theta,H\right)\right] \\ \psi_{mtj}^{b} &\equiv q_{mtj}\left[\tilde{d}_{mtj} - F\left(\tilde{d}_{mtj} \mid q_{mtj};\theta,H\right)\right] \\ \psi_{mtj}^{c} &\equiv \left[\tilde{d}_{mtj} - F\left(\tilde{d}_{mtj} \mid q_{mtj};\theta,H\right)\right] \end{aligned} \qquad \text{(A2)}$$

$$\psi_m = \sum_t \sum_j \begin{bmatrix} \psi_{mtj}^{a} \\ \psi_{mtj}^{b} \\ \psi_{mtj}^{c} \end{bmatrix} = \sum_t \sum_j \psi_{mtj} = 0 \qquad \text{(A3)}$$

Defining $\psi(\theta)$ to be a column vector containing the stacked values of $\psi_1, \psi_2, \ldots, \psi_5$ for the five markets, our method of simulated moment (MSM) estimator is $\arg\min_{\theta} \psi(\theta)' W \psi(\theta)$ where W is a symmetric, positive semidefinite weighting matrix. We employ the identity weighting matrix and the consistent estimator $\hat{\theta} = \arg\min_{\theta} \sum_m \psi_m(\theta)' \psi_m(\theta)$. The asymptotic covariance matrix of this estimator

---

[36] In contrast to $\sum_t \sum_j \left[\tilde{d}_{mtj} - F\left(\tilde{d}_{mtj} \mid q_{mtj};\theta,H\right)\right] = 0$ which enters (A3), the moment condition $\sum_t \sum_j \left[\tilde{z}_{mtj} - F\left(\tilde{z}_{mtj} \mid q_{mtj};\theta,H\right)\right] = 0$ is not employed in estimation as the latter condition holds exactly for all values of θ.

is $V\left(\hat{\theta}\right) = \frac{1+\frac{1}{H}}{n} \left[D' \quad D\right]^{-1} D'\Omega D \left[D' \quad D\right]^{-1}$ where $D \equiv E\left[\frac{\partial\psi\left(\theta_0\right)}{\partial\theta'}\right]$ and Ω is the asymptotic variance of $\psi(\theta_o)$ shown in (A4).

$$\Omega = E[\psi\psi'] = \begin{bmatrix} E\psi_1\psi_1' & 0 & \cdots & 0 \\ 0 & E\psi_2\psi_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E\psi_5\psi_5' \end{bmatrix} = \begin{bmatrix} \Omega_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega_5 \end{bmatrix} \qquad \text{(A4)}$$

We employ simulation and numerical derivates to compute $\sum_t\sum_j \frac{\partial \quad \psi_{mtj}(\hat{\theta})}{\partial \quad \theta'} \approx D_m$ in D.

If the elemental moments, ψ_{mtj}, were independent across t and j for each m, the m[th] diagonal block in Ω could be approximated simply, using the formula $\hat{\Omega}_m = \sum_t\sum_j \psi_{mtj}(\hat{\theta})\,\psi'_{mtj}(\hat{\theta})$. However, the ψ_{mtj} are correlated because the sorting mechanism jointly determines the matching of all workers to jobs in each market-year and, if present, the random school effects. Such correlation can be accounted for in a relatively straightforward manner by using simulation to approximate $\Omega_m = E[\psi_m\psi_m']$, as is done for $E\left(\tilde{z}_{mtj}\middle| q_{mtj};\theta\right)$ and $E\left(\tilde{d}_{mtj}\middle| q_{mtj};\theta\right)$. As defined above, $\tilde{z}_{mtj}^h$ and $\tilde{d}_{mtj}^h$ characterize the school attributes and distance for the j[th] teacher's match in simulation h. Substituting these expressions for $\tilde{z}_{mtj}$ and $\tilde{d}_{mtj}$ in (1) that characterize the j[th] teacher's actually match yields the expressions in (A5) and (A6). These are based on the difference between the model-predicted match for simulation h and the simulated expected values $F\left(\tilde{z}_{mtj} \mid q_{mtj};\hat{\theta},H\right)$ and $F\left(\tilde{d}_{mtj} \mid q_{mtj};\hat{\theta},H\right)$.

$$\begin{aligned}
\psi_{mtj}^{ah}(\hat{\theta}) &\equiv q_{mtj}\left[\tilde{z}_{mtj}^h - F\left(\tilde{z}_{mtj} \mid q_{mtj};\hat{\theta},H\right)\right] \\
\psi_{mtj}^{bh}(\hat{\theta}) &\equiv q_{mtj}\left[\tilde{d}_{mtj}^h - F\left(\tilde{d}_{mtj} \mid q_{mtj};\hat{\theta},H\right)\right] \qquad \text{(A5)} \\
\psi_{mtj}^{ch}(\hat{\theta}) &\equiv \quad\left[\tilde{d}_{mtj}^h - F\left(\tilde{d}_{mtj} \mid q_{mtj};\hat{\theta},H\right)\right]
\end{aligned}$$

$$\psi_m^h = \sum_t \sum_j \begin{bmatrix} \psi_{mtj}^{ah} \\ \psi_{mtj}^{bh} \\ \psi_{mtj}^{ch} \end{bmatrix} \quad \text{(A6)}$$

Averaging across the H draws, $\hat{\Omega}_m = \frac{1}{H}\sum_{h=1}^{H} \psi_m^h \psi_m^{h\prime} \approx E\psi_m \psi_m^{\prime}$ is the simulated second moment of ψ_m.

In the computation of $\hat{D} \approx E\left[\frac{\partial \psi(\theta_0)}{\partial \theta'}\right]$ as well as $\hat{\Omega}_m$, we employed H = 1000 draws for the full set of random errors in the model. The large number of draws used in simulating Ω_m is intended to largely eliminate simulation error in the calculation of $\hat{\Omega}_m$. H is large in the computation of the numerical derivatives in $\tilde{D}_m = \sum_t \sum_j \frac{\partial \, \psi_{mtj}(\hat{\theta})}{\partial \, \theta'}$ to reduce the coarseness in how $\psi_{mtj}(\theta)$ varies with small changes in θ. In contrast, the simulation of $\sum_m \psi_m(\theta)' \psi_m(\theta)$ is based on H = 100 draws since the law of large numbers reduces the overall simulation error and there is no problem with respect to differentiability of the objective function, as we used a combination of extensive grid searches and the simplex method in estimation. As shown by Pakes and Pollard (1989), even though the limiting function $E\psi(\theta)$ must be smooth with respect to θ, the simulated values of $\psi(\theta)$ employed in computing the criterion function employed in estimation need not be a smooth function of θ.