Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Disentangling the Preferences of Teachers and Employers
Author(s): Donald Boyd, Hamilton Lankford, Susanna Loeb, and James Wyckoff
Source: *Journal of Labor Economics*, Vol. 31, No. 1 (January 2013), pp. 83-117
Published by: The University of Chicago Press on behalf of the Society of Labor Economists and the NORC at the University of Chicago
Stable URL: http://www.jstor.org/stable/10.1086/666725
Accessed: 18/03/2014 12:13

Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Disentangling the Preferences of Teachers and Employers

Donald Boyd, *University at Albany, SUNY*

Hamilton Lankford, *University at Albany, SUNY*

Susanna Loeb, *Stanford University*

James Wyckoff, *University of Virginia*

This article uses a game-theoretic, two-sided matching model and method of simulated moments estimation to study factors affecting the match of elementary teachers to their first jobs. We find that employers demonstrate preferences for teachers having stronger academic achievement (e.g., attended a more selective college) and for teachers living in closer proximity to the school. Teachers show preferences for schools that are closer geographically, are suburban, have a smaller proportion of students in poverty, and, for white teachers, have a smaller proportion of minority students. These results appear predictable but contradict findings from prior research estimating hedonic wage equations for teacher labor markets.

We are grateful to the New York State Education Department for the data employed in this article. We appreciate the many helpful comments provided by John Jones, Huaming Peng, Michael Sattinger, and Chris Taber and workshop partic-

[*Journal of Labor Economics,* 2013, vol. 31, no. 1]

0734-306X/2013/3101-0003$10.00

I. Introduction

The 3.1 million elementary and secondary public school teachers in the United States make up more than 8.5% of all college-educated workers 25–64 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 analyzing the sorting of teachers to schools has employed hedonic models, often yielding counterintuitive results (e.g., salaries are estimated to be lower in schools having relatively more students in poverty, with other observed attributes of schools and teachers held constant). The objective of this article is to develop and estimate an alternative model based on a game-theoretic two-sided matching model and a method of simulated moments estimator. In this way, 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 nonwhite students, particularly those in urban areas, often are taught by the least skilled teachers (see, e.g., Lankford, Loeb, and Wyckoff 2002), a factor that likely contributes to the substantial gaps in academic achievement based on student income and race/ethnicity. 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;[2] however, teacher preferences are likely to be particularly influential. 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 nonpecuniary job

ipants at Duke, Michigan, National Bureau of Economic Research, Stanford, Syracuse, Union College, SUNY–Albany, University of California, San Diego, and University of California, 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, the US Department of Education, and the National Center for the Analysis of Longitudinal Data in Education Research supported by US Department of Education grant R305A060018 to the Urban Institute provided financial support. The views expressed in the article are solely ours and may not reflect those of the funders. Any errors are attributable to us. Contact the corresponding author, Hamilton Lankford, at hamp@albany.edu.

[1] *Digest of Education Statistics* 2004 and US Census Bureau Educational Attainment in the United States 2000 Detailed Tables.

[2] Levin and Quinn (2003) discuss problems with hiring practices in many urban districts. Pflaum and Abramson (1990), Ballou (1996), and Ballou and Podgursky (1997) provide evidence that many districts are not hiring the most qualified candidates.

characteristics, such as school leadership, class size, preparation time, facilities, or characteristics of the student body, are important as well.[3]

A large literature suggests that teachers respond to wages.[4] Yet, as noted, research employing hedonic models to estimate the compensating wage differentials needed to attract teachers with particular attributes to schools with particular characteristics has produced counterintuitive results. As an alternative to the hedonic model, we develop and estimate an empirical framework based on the two-sided matching model extensively studied by game theorists (Roth and Sotomayer 1990). We argue that the two-sided matching model is an attractive alternative for analyzing the sorting of teachers across jobs and show how the underlying preferences of job candidates and employers can be estimated using the method of simulated moments and data characterizing observed job-worker matches.

We find that teachers demonstrate preferences for schools that are closer geographically, are suburban, have a smaller proportion of students in poverty, and, for white teachers, have a smaller proportion of minority students. Employers show preferences for teachers with stronger academic achievement, measured by having more than a bachelor's degree, the selectivity of their undergraduate college, and their score on the basic-knowledge teacher-certification exam and teachers living in closer proximity to the school. As predictable as these results are, they differ from estimates that suggest, for example, that employers do not value teacher skills and that teachers prefer schools having higher percentages of students in poverty.

Section II of the article briefly summarizes relevant features of teacher labor markets. A game-theoretic model of teacher-school match is presented in Section III. For comparison, hedonic models and match models with search are discussed in Section IV. Our strategy for estimating the game-theoretic two-sided matching model is summarized in Section V. Sections VI and VII discuss the data and model specifications employed and empirical results. Section VIII presents conclusions.

[3] 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.

[4] 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 (Manski 1987; Murnane, Singer, and Willett 1989; Dolton 1990; Rickman and Parker 1990; Theobald 1990; Dolton and Makepeace 1993; Hanushek and Pace 1995; Brewer 1996; Mont and Reece 1996; Theobald and Gritz 1996; Stinebrickner 1998, 1999, 2001; Dolton and van der Klaaw 1999). Baugh and Stone (1982), e.g., find that teachers are at least as responsive to wages in their decision to quit teaching as are workers in other occupations.

II. Features of Teacher Labor Markets

In previous research we have used New York teacher data to document various aspects of teacher labor markets and find a marked sorting of teachers across schools. For example, in schools in the highest quartile of student performance on the New York State fourth-grade English Language Arts Exam, only 3% of teachers are uncertified, 10% earned their undergraduate degree from least competitive colleges, and 9% of those who have taken a general-knowledge teacher-certification exam failed.[5] In contrast, in schools in the lowest quartile of performance, 22% of teachers are uncertified, 26% come from least competitive colleges, and 35% have failed a general-knowledge certification exam (Lankford et al. 2002). We find similar patterns when schools are grouped on the basis of student poverty, race/ethnicity, and limited English proficiency.

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 et al. 2002). We focus on these initial job matches and the sorting of teachers within local labor markets.

A surprisingly large number of individuals take their first teaching jobs very close to where they grew up. In New York State, over 60% of teachers first teach within 15 miles of the high school from which they graduated and 85% teach within 40 miles (Boyd et al. 2005a). This proximity has two important implications for modeling the sorting of teachers across jobs. First, because most teachers make job choices within a very limited geographic area, our empirical analysis focuses on the matching of teachers to jobs within relatively small geographic areas (metropolitan areas). Second, even within each of these local labor markets, work proximity is likely to affect teachers' rankings of alternative job opportunities, suggesting that models of teacher labor markets 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 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

[5] During the years studied teachers had the option of taking the National Teacher Examination (NTE) General-Knowledge Exam or the New York State Teacher Certification Examinations Liberal Arts and Science Exam (LAST). Scores on the NTE exam were rescaled to correspond to scores on the LAST.

policy matters (e.g., class size)—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 on 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 attributes of the jobs being filled also are largely fixed during this hiring season. District-level union contracts typically set some working conditions and teacher salaries for 3 years and require that all teachers having the same number of years of education and within-district experience earn the same salary. Typically, salary is unaffected by other teacher attributes or the characteristics of the schools in which they teach.[6] Other conditions of work also are either largely 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 process and collective bargaining. The inflexibility of salaries and many working conditions is especially restrictive in large urban districts and countywide districts in which there is considerable within-district variation in nonwage attributes across schools. On the supply side, the number and attributes of those entering the market each year need not reflect recent market shocks as a result of the time typically required for teacher preparation and certification. The current excess supply of individuals newly certified to teach, due to the recession, is a tangible example. These features of teacher labor markets lead us to 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.

III. A Model of Worker-Jobs Match

The allocation of workers to jobs is an example of a more general setting in which individuals in one group are matched with individuals, agents, or firms in a separate, second group.[7] Other examples include marriage and college attendance. In such cases, the matching is two-sided in that whether a particular match occurs depends on 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 are an individual's opportunities elsewhere, the more selective he

[6] Very recently some school districts have begun offering either one-time or continuing incentives for teachers to work in difficult-to-staff schools or teach shortage subjects. There is increasing interest and experimentation with these options, but there is little good evidence regarding the incentives necessary to attract high-quality teachers.

[7] These cases differ from the roommate problem in which those being matched come from the same group.

or she will be in evaluating a potential partner" (Burdett and Coles 1999, F307). Researchers have analyzed such settings employing game-theoretic two-sided match models, hedonic models, and match models with search. In this section we summarize the game-theoretic match model that underlies our empirical approach and discuss the other two approaches in Section IV.

Building on the work of Gale and Shapley (1962), the game-theoretic two-sided match literature focuses on one-to-one matching such as marriage and many-to-one matching such as college admission, the former being a special case of the latter. This framework has been used to analyze the matching of medical residents to hospitals, an application in many ways similar to the sorting of teachers across schools.[8] Because the two-sided matching model is the foundation for our empirical framework, much in that literature is relevant here.[9]

Consider an environment in which $C = \{c_1, \ldots, c_J\}$ is the set of J individuals seeking jobs and $S = \{s_1, \ldots, s_K\}$ is the set of K schools having jobs to be filled, $J \geq K$, assuming that each school has one vacancy. Each agent has a complete and transitive preference ordering over the agents on the other side of the market, with these orderings arising from candidates' preferences over job attributes and hiring authorities' preferences over the attributes of candidates. In our model, $u_{jk} = u(z_k^1 | q_j^2, \beta) + \delta_{jk}$ represents the utility from working in the kth school as viewed from the perspective of the jth candidate; z_k^1 is a vector of observed school attributes; vector q_j^2 represents observed attributes of the candidate that affect her assessment of the kth alternative; β is a vector of parameters; and δ_{jk} is a random variable reflecting unobserved heterogeneity in the attractiveness of school k for different individuals. Similarly, $v_{jk} = v(q_j^1 | z_k^2, \alpha) + \omega_{jk}$ represents the attractiveness of the jth candidate from the perspective of the hiring authority for school k, q_j^1 represents pertinent observed attributes of the candidate, z_k^2 represents the observed attributes of the school that affect the authority's assessment of candidate j, α is a vector of parameters, and the random error ω_{jk} reflects unobserved factors.

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 (e.g., the employer and worker negotiate the salary to be paid), most theoretical two-sided match models assume that utility is nontransferable; that is, how the surplus from any given match is split between a matching pair is predetermined. Given the features of teacher labor markets discussed

[8] The major difference is that the assignment of residents typically results from a centrally controlled allocation mechanism, whereas teacher labor markets involve decentralized job matching. However, this difference is not as great as one might first think since the Gale-Shapley algorithm employed in the centralized matching of residents to hospitals mimics one particular decentralized mechanism that yields a match equilibrium.

[9] See Roth and Sotomayer (1990) for a clear development of the model and a discussion of its properties.

(A) Candidates' benefits from alternative employment	(B) Schools' benefits from alternative candidates
$\begin{array}{c\|cccc} & s_1 & s_2 & \cdots & s_K \\ \hline c_1 & u_{11} & u_{12} & \cdots & u_{1K} \\ c_2 & u_{21} & u_{22} & \cdots & u_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_J & u_{J1} & u_{J2} & \cdots & u_{JK} \end{array}$	$\begin{array}{c\|cccc} & s_1 & s_2 & \cdots & s_K \\ \hline c_1 & v_{11} & v_{12} & \cdots & v_{1K} \\ c_2 & v_{21} & v_{22} & \cdots & v_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_J & v_{J1} & v_{J2} & \cdots & v_{JK} \end{array}$

FIG. 1.—Utility and rankings of candidates and schools

in Section II (e.g., wage posting), we maintain nontransferable utility in our empirical framework; the attributes $q_j \equiv (q_j^1,\ q_j^2)$ and $z_k \equiv (z_k^1,\ z_k^2)$, for all j, k, are fixed during the period in which workers are matched to jobs.

Suppose that C and S are known as are the fixed values of q_j and z_k. Given β and a set of random draws for the δ_{jk}, $u_{jk} = u(z_k^1|q_j^2,\ \beta) + \delta_{jk}$ implies the matrix of candidates' benefits represented in panel A of figure 1. For any row, the benefits to a candidate from being employed in each of the K schools imply her complete rankings of school alternatives.[10] Similarly α, a particular set of random draws for the ω_{jk}, and $v_{jk} = v(q_j^1|z_k^2,\ \alpha) + \omega_{jk}$ imply the matrix of school benefits represented in panel B.[11]

If each of the candidates unilaterally were able to choose where to teach, β in $u_{jk} = u(z_k^1|q_j^2,\ \beta) + \delta_{jk}$ could be estimated using a multinomial probit or

[10] Here it is assumed that the attractiveness of a particular job depends only on the current attributes of that job. A candidate's evaluation of a job could also depend on the chances of moving from that job to more attractive positions (e.g., intra- and interdistrict transfer possibilities that vary across initial positions) and a variety of other future considerations. The potential importance of accounting for transfer possibilities is underscored in a number of papers (e.g., Hanushek, Kain, and Rivkin 2004) showing that teachers make transfers that are both substantial in number and systematic in that teachers typically move to schools having higher test scores and relatively fewer poor and minority students. If candidates consider such possibilities when seeking their first teaching jobs, a school having undesirable attributes but offering new hires the opportunity to quickly transfer to schools having more desirable attributes will be more attractive to a job seeker than will a school having the same undesirable attributes and more limited transfer opportunities. Because our model does not account for such dynamics, bias is a potential problem. For example, suppose that x measures some school attribute in which the direct effect of an increase in x is to increase school attractiveness. If transfer opportunities are greater in schools having relatively lower values of x, our estimate of the coefficient for x will be biased downward compared to the actual effect of a change in x, ceteris paribus. Introducing dynamics into the analysis of job selection offers the possibility of disentangling such effects but goes beyond the scope of this article.

[11] To simplify the discussion, we assume that hiring authorities prefer hiring any of the candidates rather than leaving job openings unfilled and candidates prefer

logit random utility model; α could be estimated in a similar way if each hiring authority unilaterally chose among candidates. However, our empirical model is more complex for two reasons. First, it is the interaction of decisions of a candidate and a hiring authority that determines whether they match. Second, even though any such interaction would complicate the model, the decisions by the two parties considering whether to match crucially depend on the choices made by all other candidates and employers. In particular, a candidate's willingness to accept a particular match depends on her preferences as well as her choice set, that is, the set of schools willing to hire her. In turn, the willingness of employers to make the candidate an offer depends on whether they prefer to employ alternative candidates who are willing to fill their positions, and so on.

The set of job-worker matches will be stable if there is 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 and match with each other. Formally, suppose that candidate g is employed in job g', with candidate h and job h' similarly matched. The stability of these two pairings requires 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.[12] The conditions $1(u_{gg'} < u_{gh'})1(v_{hh'} < v_{gh'}) = 0$ and $1(u_{hh'} < u_{gh'})1(v_{gg'} < v_{hg'}) = 0$ are equivalent expressions, where $1(\,)$ is the indicator function, which equals one if the function argument is true and zero otherwise. Overall stability requires that the condition $1(u_{gg'} < u_{gh'})1(v_{hh'} < v_{gh'}) = 0$ hold for every candidate (g) and job (h') pairing not matched together.

As shown by Gale and Shapley (1962), a straightforward decentralized job match mechanism always will yield a stable matching.[13] Assuming that the

---

each of the possible jobs to the alternative of not taking a teaching job, assumptions that can be relaxed easily.

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

[13] 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 were not rejected 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.

rankings are strict and employers make offers to employees, the resulting stable matching will be both unique and employer optimal (i.e., all employers weakly prefer this allocation to all other stable matchings). An employee-optimal matching would result if candidates made offers to hiring authorities.[14]

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 differentiated only by age). Fox (2009a) develops a computationally appealing estimation strategy for estimating more general models falling within a broad class of matching games with transfers.[15] However, with transferable utility, the statistical approaches do not 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.

IV. Other Models of Worker-Jobs Match

In hedonic models it is assumed that there are continua of worker and job attributes and that these agents have complete and transitive preference orderings over the attributes of agents on the other side of the market.[16] Going beyond assuming that the observed allocation of workers to jobs is stable, hedonic models maintain that wages or some combination of other attributes, or both, are sufficiently flexible to clear the market, meaning that demand equals supply at each combination of worker and job attributes.[17] Together these assumptions imply that there is an equilibrium hedonic wage locus that supports the observed allocation of workers to jobs.[18] Further-

[14] The match mechanism need not rely on the Gale-Shapley algorithm. Roth and Vande Vate (1990) show that a very general decentralized mechanism will lead to a stable allocation.

[15] See Fox (2009b) for a discussion of these papers as well as the small number of other structural empirical work employing match models.

[16] 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.

[17] For example, Rosen (1974, 35) 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."

[18] While the hedonic wage locus is unique, this is not the case in the above two-sided match model. Given the attributes of workers and the nonwage attributes of

more, the choices made by agents are such that workers' indifference curves for job attributes are tangent to the hedonic wage function as are employers' indifference curves for worker attributes. In such settings, it is possible to estimate underlying preference parameters employing information characterizing realized worker-job matches and how wages vary over the observed combinations of worker and job attributes.

Most studies of teacher labor markets (e.g., 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 in an effort 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 teachers hired in jobs having particular attributes. However, estimation of such wage equations often yields anomalous results (e.g., Boyd et al. 2003, 31; Goldhaber, Destler, and Player 2010). For example, employing the same data used in this article, we find that salaries are higher in schools with higher proportions of minority students (which is consistent with a variety of studies examining teacher retention; e.g., Hanushek et al. 2004) but lower for schools with higher proportions of children in poverty and urban schools (which is inconsistent with the same retention analyses). Also, there appears to be no premium for stronger teacher qualifications.

Researchers have posited a number of reasons for similar counterintuitive hedonic results in other labor markets, including omitted variables (Lucas 1977; Brown 1980), simultaneity (McLean, Windling, and Neergaard 1978), measurement error, and labor market frictions (Hwang, Mortensen, and Reed 1998; Lang and Majumdar 2004). In the context of the frequently estimated linear-quadratic hedonic model due to Tinbergen (1956; also see Sattinger 1979), Epple (1987) considers identification and estimation challenges that arise from omitted variables, measurement error, and simultaneity. For example, he shows that ordinary least squares (OLS) estimates of this model will be inconsistent when there are unmeasured attributes or attributes measured with error. Ekeland, Heckman, and Nesheim (2004) point out that this linear-quadratic hedonic model is quite restrictive in most applications. These findings point out important issues regarding the hedonic specifications that have been employed to analyze teacher salaries and how those models have been estimated. First, the functional forms employed have been even more restrictive than the special case criticized by Ekeland et al. Second, empirical analyses employing OLS have not accounted for the fact that many potentially important teacher and school attributes either are unobserved or are measured with error. For example, our finding that salaries are lower in schools having relatively more

jobs, a particular matching of workers to jobs will be stable under a range of salary loci.

poor students merely could result from these schools employing teachers of lower quality not completely captured by the teacher attributes included in the model.

A large literature in labor economics employs two-sided matching models with search, which differ from game-theoretic match models in important ways.[19] In contrast to the assumption of full information and no market frictions in the game-theoretic framework, such frictions are central to labor search models of marriage and job match. The demand side of labor search models also is often 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 concerns the extent and nature of agent heterogeneity allowed. Game-theoretic match models typically require only that agents' ranking of match partners are complete and transitive, placing no restrictions on the extent of preference heterogeneity. In contrast, preference heterogeneity must be limited in search models in order to solve for search equilibria (Burdett and Coles 1999). Such limited heterogeneity would be quite restrictive in our analysis. For example, because of the importance of distance from home to jobs, teachers may rank the same job differently because of their location relative to the school.

In what follows, we develop and estimate a structural model drawing on the game-theoretic two-sided match literature.[20] The model can incorporate quite general preference heterogeneity, accounts for job candidates and employers each having relatively limited numbers of discrete choices, and allows for the possibility that teacher labor markets do not clear.[21] 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 underlying preference parameters reflecting teachers' evaluations of various job attributes as well as employers' preferences for attributes characterizing teachers. Given past anomalous results for estimated hedonic models of teacher salaries, it is of particular interest whether parameter estimates of our empirical two-sided matching model are more consistent with what would be expected.

[19] See Sattinger (1993), Rogerson, Shimer, and Wright (2005), and Eckstein and van den Berg (2007) for informative literature reviews.

[20] Other examples of the use of structural models in education research include analyses by van der Klaauw (2000) and Stinebrickner (2001).

[21] We analyze the matching of new elementary teachers to job openings in five labor markets (Albany-Schenectady-Troy, Buffalo, Rochester, Syracuse, and Utica-Rome metropolitan areas) in each of six years (1994–95 through 1999–2000). 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. Market thinness would be even more pronounced if one considers the markets for particular specializations (e.g., those certified to teach high school mathematics).

There is little doubt that several assumptions in our two-sided match model do not hold exactly. First, the assumption that each candidate (employer) is knowledgeable of, and considers, all employers (candidates) in the local labor market is likely to be violated. Search costs will limit the number of schools to which candidates apply as well as the number of applicants seriously considered by hiring authorities. For example, personal connections or familiarity may influence which job applicants are given serious consideration. Second, the deferred-acceptance procedure, integral to the Gale-Shapley algorithm (see n. 13), is violated to the extent that time-limited offers are made sequentially with candidates having to decide whether to accept such offers or decline in order to continue searching for better offers that might arise later.[22] Such considerations will result in the number of matched pairs actually considered by candidates and employers being smaller than the total number of possible pairings. Our two-sided matching model does not allow for such market frictions. Even so, we view the model and its assumptions as providing a plausible framework for the empirical analysis of teacher-school sorting, understanding that neither the conceptual model nor empirical results should be taken too literally.

V. Our Empirical Model

For the model specified above, the equilibrium matchings corresponding to the alternatives and rankings characterized in figure 1 are represented in the left side of figure 2. The right side characterizes these matches in terms of the resulting relationship between the attributes of candidates and the schools where they are employed.[23] The matches in figure 2 correspond to particular values of the random variables, the explanatory variables, and the model parameters.

We generalize the notation to allow for M local labor markets, $m = 1, 2, \ldots, M$, and T years, $t = 1, 2, \ldots, T$, by adding the subscripts m and t to the explanatory and random variables. To allow for multiple job openings in a school in any given year, we assume that vacancies within a school are identical, which does not seem overly restrictive given our focus on elementary schools, where there is a large degree of job homogeneity within schools. The pertinent theoretical underpinning for many-to-one matches parallels that for one-to-one matches discussed above.

Estimation is based on the method of simulated moments (MSM; see Pakes and Pollard 1989) with δ_{jk} and ω_{jk} assumed to be standard normal random variables that are uncorrelated with the observed attributes of teachers

[22] The deferred-acceptance algorithm used to demonstrate how a stable equilibrium in a game-theoretic two-sided match model could be achieved has a role similar to the *tâtonnement* process used to demonstrate a market-clearing process.

[23] 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.

$$
\begin{array}{cc}
\text{School-teacher matched pairs} & \text{Joint distribution of school and teacher attributes} \\
\left\{\begin{array}{c} (s_1, c_{1'}) \\ (s_2, c_{2'}) \\ \cdots \\ (s_K, c_{K'}) \end{array}\right\} &
\begin{bmatrix} z_1 & q_{1'} \\ z_2 & q_{2'} \\ \vdots & \vdots \\ z_K & q_{K'} \end{bmatrix} = [z\ q]
\end{array}
$$

FIG. 2.—Resulting matching of teachers and jobs

and schools.[24] Let $\tilde{z}_{mtj}$ represent the attributes of the job taken by teacher j hired in market m during period t.[25] The model structure, parameters α and β, as well as the distributions of the explanatory and random variables together imply the joint distribution of $\tilde{z}_{mtj}$ and q_{mtj} and the expected value of $\tilde{z}_{mtj}$ for candidate j, $E(\tilde{z}_{mtj}|q_{mtj};\ \theta)$. It follows that $E[\tilde{z}_{mtj} - E(\tilde{z}_{mtj}|q_{mtj};\ \theta)|q_{mtj}] = 0$; for a candidate having attributes q_{mtj}, the difference between the attributes of the school where the individual actually works, $\tilde{z}_{mtj}$, and the expected mean attributes, given q_{mtj}, is zero in expectation. In turn, this implies that $E(q_{mtj}[\tilde{z}_{mtj} - E(\tilde{z}_{mtj}|q_{mtj};\ \theta)]) = 0$; 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 analogue $\sum_t\sum_j q_{mtj}[\tilde{z}_{mtj} - E(\tilde{z}_{mtj}|q_{mtj};\ \theta)] = 0$ is employed in estimation.[26] Similarly, we use $\sum_t\sum_j q_{mtj}[\tilde{d}_{mtj} - E(\tilde{d}_{mtj}|q_{mtj};\ \theta)] = 0$, which relates the actual distances for employees to their expected values.

In contrast to the hedonic model, there being unobserved attributes of candidates and employers does not create problems for our empirical two-sided match model when those attributes are uncorrelated with the observed attributes. The preference equations specified allow for there being such unobserved quality attributes as does the moment conditions employed in estimation. As discussed below, estimated preference parameters

[24] It is discussed below that salary is likely to be correlated with unobserved school attributes reflected in the error terms, possibly resulting in the salary coefficient being biased and complicating the estimation of compensating differentials.

[25] Reflecting the two-sided match, $\tilde{z}_{mtj}$ from the perspective of this teacher is the same as z_{mtk} defined above, where the kth school employs the jth individual.

[26] Equivalently, we could have employed $E(z_{mtk}[\tilde{q}_{mtki} - E(\tilde{q}_{mtki}|z_{mtk};\ \theta)]) = 0$ and its sample analogue $\sum_t\sum_k\sum_i z_{mtk}[\tilde{q}_{mtki} - E(\tilde{q}_{mtk}|z_{mtk};\ \theta)] = 0$, which can be rewritten $\sum_t\sum_k n_{mtk}\, z_{mtk}[\bar{q}_{mtk} - E(\tilde{q}_{mtk}|z_{mtk};\ \theta)] = 0$. Here $\tilde{q}_{mtki}$ represents the attributes of the teacher newly employed during period t to fill the ith vacancy in school k, $i = 1, 2, \ldots, n_{mtk}$, and $\bar{q}_{mtk}$ is the mean attributes of the n_{mtk} new teachers employed by the kth school. The expression $\sum_t\sum_k n_{mtk} z_{mtk}[\bar{q}_{mtk} - E(\tilde{q}_{mtk}|z_{mtk};\ \theta)]$ will always equal $\sum_t \sum_j q_{mtj}[\tilde{z}_{mtj} - E(\tilde{z}_{mtj}|q_{mtj};\ \theta)]$. Thus, including both $\sum_t\sum_k n_{mtk} z_{mtk}[\bar{q}_{\ mtk}|z_{mtk};\ \theta)] = 0$ and $\sum_t\sum_j q_{mtj}[\tilde{z}_{mtj} - E(\tilde{z}_{mtj}|q_{mtj};\ \theta)]$ would be redundant.

must be interpreted with care when the unobserved attributes are correlated with observed attributes included in the empirical model.

Because of the difficulties in deriving and computing analytical expressions for $E(\tilde{z}_{mtj}|q_{mtj};\ \theta)$ and $E(\tilde{d}_{mtj}|q_{mtj};\ \theta)$, we compute their values using simulation. As described in appendix A, our MSM estimator, $\hat{\theta}$, is the value of θ that minimizes a quadratic form defined in terms of the moment conditions. The parameter estimates minimize the distance between empirical moments reflecting the actual distribution of school attributes across teachers and the corresponding theoretical moments implied by our model.

Two papers employing the MSM have substantial overlap with our application. Berry (1992) estimates an equilibrium game-theoretic model of market entry in the airline industry, with the moments reflecting the equilibrium number of firms operating at each airport. Sieg (2000) estimates a bargaining model of medical malpractice disputes, focusing on bilateral interactions between individual plaintiffs and defendants. The papers are pertinent in that simulated moments are obtained by repeatedly solving game-theoretic models for each of a large number of draws for the random variables in the model.

VI. Data and Model Specifications

Our analysis focuses on 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 employ data from a larger database of teachers, schools, and districts drawn from seven administrative data sets. The core data come from the Personnel Master File, part of the Basic Education Data System of the New York State Education Department. The annual records are linked to other databases that contain information about the qualifications of prospective and actual teachers as well as the environments in which these individuals make career decisions, including New York State data characterizing each school. Similar data have been used to study teacher labor markets in other states (e.g., Hanushek et al. 2004; Clotfelter et al. 2008; Harris and Sass 2009). Matched employer and employee data have proved useful in the analysis of labor markets more generally (Rosen 1986; Abowd and Kramarz 1999; Postel-Vinay and Robin 2002).

We estimate model I shown in (1). The jth teacher's utility associated with working in job k, u_{jk}, is a function of student poverty in the school measured by the proportion of kindergarten through sixth-grade students,

[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.

$$
\begin{aligned}
u_{jk} = {} & \beta_1 s\text{poverty}_k + [\beta_2 t\text{minority}_j + \beta_3(1 - t\text{minority}_j)]s\text{minority}_k \\
& + \beta_4\text{salary}_k + \beta_5\text{urban}_k + \beta_6\text{distance}_{jk} + \delta_{jk}, \\
v_{jk} = {} & \alpha_1\text{BA}_j + \alpha_2\text{score}_j + \alpha_5\text{highly selective}_j + \alpha_6\text{least selective}_j \\
& + \alpha_7 t\text{minority}_j + \alpha_8\text{distance}_{jk} + \omega_{jk},
\end{aligned} \tag{1}
$$

eligible for free lunch (*s*poverty), school racial composition measured by the proportion of students who are black or Latino (*s*minority), the starting salary for a teacher having a BA degree (salary),[28] a dummy variable indicating whether a school is in an urban district (urban), and the distance of the school from the teacher candidate (distance). The specification allows for the possibility that the effect of a school's racial composition varies depending on whether the teacher is black or Latino (*t*minority). The attractiveness of candidate *j* from the perspective of the hiring authority for school *k*, v_{jk}, is a function of a dummy variable indicating whether the individual has no more than a bachelor's degree (BA), her score on the first taking of the general-knowledge certification exam (score), college selectivity measured by dummy variables indicating whether the institution from which each individual earned her undergraduate degree was rated by Barron's as being highly selective or least selective,[29] a dummy variable indicating whether the candidate is black or Latino (*t*minority), and the candidate-school distance (distance).[30] Descriptive statistics are presented in table 1.

The model in (1) is consistent with model restrictions needed for identification, as discussed in appendix B and summarized here. First, distributional assumptions regarding the unobserved random errors δ_{jk} and ω_{jk} are needed for the revealed preferences implied by the observed candidate-job matches to provide any information regarding model parameters. Assumptions are also needed regarding the random-error covariance. We assume that δ_{jk} and ω_{jk} are independent, normal random errors standardized, with no loss of generality, to have zero means and standard deviations of one.[31] Second,

[28] Salaries are for 2000. If the 2000 salaries were not available, salary information for the most recent prior year was used after inflating the value using the average percentage change across districts having salaries in both years.

[29] The omitted category includes selective and somewhat selective colleges.

[30] For each of the labor markets, $\sum_t\sum_j q_{mtj}[\tilde{z}_{mtj} - E(\tilde{z}_{mtj}|q_{mtj};\ \theta)] = 0$ includes five school characteristics (*s*poverty, *s*minority, salary, urban, and distance) in $\tilde{z}_{mtj}$ and BA, score, highly selective, least selective, *t*minority, and 1 − *t*minority in q_{mtj}. (Both *t*minority and 1 − *t*minority are entered because q_{mtj} does not include a constant term.) Thus, estimation was based on 30 moments for each of the five markets.

[31] We explored the robustness of the empirical specifications included in the article and found few changes in coefficient estimates. As explained in the discussion of identification in app. B, even though the theoretical model places no restrictions on the structure of the error terms in our specification of an empirical two-sided

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

Variable	Mean	SD
Schools (N = 2,443):		
*s*minority	.210	.293
*s*poverty, K–6	.293	.265
Urban	.217	.293
Salary	32,458	2,607
Teachers (N = 5,028):		
*t*minority	.064	.246
*t*quality index	.00	1.00
BA or less	.505	.500
Score	260.217	18.441
Highly selective	.134	.340
Least selective	.041	.198
Distance to job (miles)	24.616	115.27
Distance if < 50 miles	8.638	8.349
Year:		
1995	.109	
1996	.123	
1997	.151	
1998	.139	
1999	.211	
2000	.267	
Metropolitan stastical areas/regions:		
Albany	.178	
Buffalo	.251	
Rochester	.350	
Syracuse	.167	
Utica-Rome	.055	

NOTE.—Only 6.6% of the sample traveled more than 50 miles to their jobs.

matching model, parameter identification crucially depends on such restrictions. The estimators of the models in table 2 below maintain that the stochastic errors δ_{jk} and ω_{jk} are independent, standard-normal random variables. As specification checks, alternative specifications were estimated (results are available from the authors). First, school random effects were introduced into the specification of candidates' preferences as a first step in accounting for school attributes observed by teacher candidates but omitted in our analysis. (This specification has limitations as it is maintained that the random errors are independent of the variables characterizing schools included in the preference equation. The problem, as discussed below, is that unobserved school characteristics are likely to be correlated with salary.) Second, in a model without school random effects, δ_{jk} and ω_{jk} were maintained to be based on independent draws from the student's t-distribution with four degrees of freedom, which has both thicker tails and a higher concentration of values close to the mode of zero compared to the standard-normal distribution. The random-effects model changes the results very little except that the percentage of minority students has an even stronger negative effect on the utility of white teachers, while the effect of being in an urban school is not as strong. The student's t-model is also quite similar.

teacher attributes such as *t*minority cannot enter $u(\ )$ additively but can enter when interacted with one or more school attributes (e.g., $\beta_2 t\text{minority}_j \cdot s\text{minority}_k$ in $u(\ )$). A similar restriction holds regarding how school attributes enter $v(\ )$. Third, as in bivariate discrete-choice models with partial observability, identification requires that one or more quantitatively important variables enter $u(\ )$ but not $v(\ )$, or the reverse. In certain cases, exclusion restrictions follow from reasonable a priori assumptions. For example, the salary paid may affect how candidates value a school but not the school's ranking of applicants receiving the same salary. Exclusion restrictions also follow from the above point regarding variable additivity (e.g., attributes of candidates, such as BA, can enter $v(\ )$ additively, but not $u(\ )$).

As noted above, most individuals take their first teaching jobs very close to where they grew up, possibly the result of multiple factors. In addition to the obvious preference for proximity, distance could proxy school familiarity, possibly reflecting individuals wanting to teach in an environment they know, the existence of informal networks (e.g., being attracted to a school because they know other individuals working there), as well as the availability of information regarding job openings, school environments, working conditions, and so on.

Disentangling the separate effects of such related factors would be quite informative but goes beyond the scope of the current analysis. Data limitations lead us to employ the distance from schools to the address given when individuals applied for certification, a point in time typically prior to when individuals apply for teaching jobs. We view this measure (d_{jk}) as being a useful proxy for some combination of the above factors. Because the proximity of candidates to the schools also could affect how employers evaluate candidates, for reasons similar to those suggested above regarding the preferences of candidates, the candidate-school distance measure, d_{jk}, is entered in $v(\ )$.[32] The distance variable employed is $\text{distance}_{jk} = \ln(d_{jk} + 1)$.

We also estimate a second specification, the difference being that the constant distance coefficient for candidates in model I is replaced using a random coefficient in model II (β_{6j}). Because the estimated distance effect for candidates in model I is large in magnitude, the random-effect specification is included to explore whether the importance of distance varies across teachers, possibly reflecting both observed and unobserved heterogeneity. In particular, we assume that the distribution of $-\beta_{6j}$ across teachers is lognormal, where the mean and standard deviation for the corresponding normal random variable, ε_j, are $\mu_j = \mu_o + \gamma\text{score}_j$ and σ^*, respectively; $-\beta_{6j} =$

[32] Zip codes were used to compute distances. The distance measure was censored at 50 if the distance to a school was greater than 50 miles. As a result, if the distances to all schools exceed 50 miles, distance is not a factor in the candidate's choice of jobs and drops out of the moment conditions.

$\exp(\mu_o + \gamma \text{score}_j + \varepsilon_j)$.[33] Here μ_o, γ, and σ^* are parameters and the certification exam score, score_j, is entered to explore whether there is observed teacher heterogeneity with respect to the effect of distance. When $\gamma = \sigma^* = 0$, model II reduces to model I.

VII. Empirical Results

The method of simulated moments parameter estimates are reported in table 2. Teacher candidates are estimated to prefer schools having smaller proportions of students who are poor and, for white teachers, those with a smaller percentage of African American or Latino students, suburban schools (even after accounting for school attributes correlated with student poverty and race), and schools closer in proximity. Employers value candidates having more than a bachelor's degree, those who score higher on a general-knowledge certification exam, those who graduated from more selective colleges, and candidates less distant from their schools. In the following discussion, we largely focus on model II.

Employers' Preferences for Job Candidates

Candidates' scores on a general-knowledge certification exam (score) are entered using a piecewise-linear specification with kink points at 230 and 260. (A score of 220 is needed to pass the exam, and the maximum score is 300.) The estimated coefficient for Score-1 is positive and statistically significant, while the coefficient estimates for both Score-2 and Score-3 are quite small in magnitude and statistically insignificant.[34]

Figure 3 shows how employers' evaluations of candidates are estimated to vary with the certification exam score, along with 95% confidence bounds. The graph provides information regarding an employer's estimated evaluation for any given score measured relative to the evaluation for a score of 220. For example, the value of an employer's preference equation is estimated to be larger by 0.241 for a teacher having a score of 240 compared to that for an otherwise identical teacher having a score of 220. The magnitude of this effect for what is roughly a one standard deviation change in the score is as large as 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 one-quarter as large as a one standard deviation increase in the error term. Since employers know only whether candidates pass the certification exam, possibly after multiple attempts, and are unaware of exam scores, the certifica-

[33] The minus sign in $-\beta_j$ is included since a lognormal random variable is positive and greater distance appears to reduce the attractiveness of a school.

[34] Score-1 equals score if score $\leq$ 230 and 230 otherwise. Score-2 equals zero if score $\leq$ 230, score $-$ 230 if 230 $<$ score $\leq$ 260, and 30 (= 260 $-$ 230) if score $>$ 260. Score-3 equals zero if score $\leq$ 260 and score $-$ 260 otherwise.

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

	Model I	Model II
Candidates' criterion function:		
Salary ($1,000s)	.0313	.0167
	(.0896)	(.0717)
Urban	−2.1827*	−.9253*
	(.6841)	(.2489)
*s*poverty, K–6	−1.4116*	−1.3143*
	(.3525)	(.3347)
*s*minority for nonwhite teachers	.9893	−.2169
	(.6062)	(.5002)
*s*minority for white teachers	−3.4303*	−4.8932*
	(.7336)	(.6677)
Distance $\ln(d_{jk}+1)$	−2.0679*	
	(.2001)	
Distance μ_o		2.5865*
		(.2891)
Distance γ		−.0074*
		(.0011)
Distance σ^*		.1797*
		(.0473)
Employers' criterion function:		
BA or less	−.0471*	−.0462*
	(.0145)	(.0120)
Score-1	.0292*	.0234*
	(.0059)	(.0055)
Score-2	.0035	.0007
	(.0023)	(.0018)
Score-3	−.0028	−.0004
	(.0018)	(.0018)
Highly selective college	.0442**	.0227
	(.0198)	(.0186)
Least selective college	−.1896*	−.2310*
	(.0275)	(.0337)
*t*minority	−.0018	.0140
	(.0588)	(.0400)
Distance $\ln(d_{jk}+1)$	−.3394*	−.3347*
	(.0148)	(.0197)
Objective	.5735	.4853

NOTE.—Standard errors are reported in parentheses. Score-1 equals score if score ≤ 230, 230 otherwise. Score-2 equals zero if score ≤ 230, score − 230 if 230 < score ≤ 260, and 30 (= 260 − 230) if score > 260. Score-3 equals zero if score ≤ 260 and score − 260 otherwise.
* .01 level of significance.
** .05 level of significance.

tion exam score must be a good proxy for one or more of the teacher attributes hiring authorities do observe and value.

In general, inferences regarding the relative size of various effects can be obtained by comparing the coefficient estimates across variables in a preference equation. For example, the estimated difference in an employer's

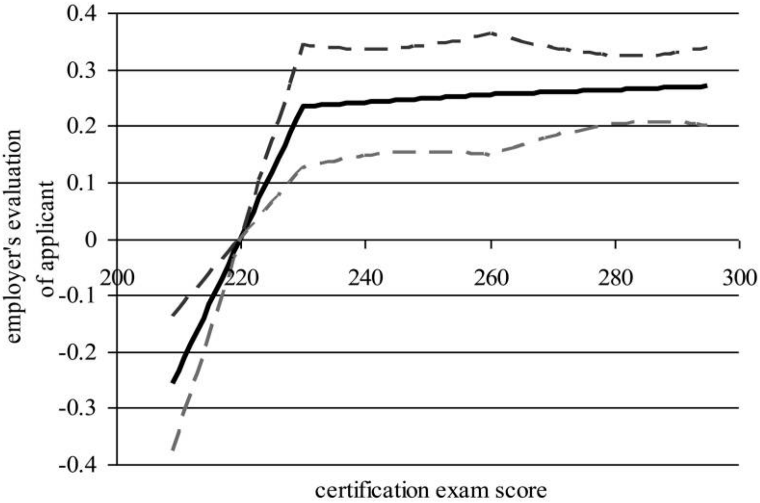

FIG. 3.—Estimated relative evaluations of applicants by hiring authorities, varying the general-knowledge teacher-certification exam score, point estimates (solid line), and 95% confidence bands.

evaluation of a candidate who graduated from a highly selective college, compared to that for an otherwise identical candidate who attended a least selective college, is over five times as large as the estimated difference for having at least a master's degree. Reflecting the relative importance of the certification exam score over its lower range, an employer is estimated to be indifferent between a 10-point increase in certification exam scores from 220 to 230 and the teacher having attended a highly selective college rather than one that is least selective.

These results indicate that hiring authorities favor teacher candidates having stronger qualifications. Ballou (1996) finds that a meaningful number of academically able college graduates who completed teacher preparation and applied for one or more teaching jobs do not end up teaching, while many of their lesser-qualified peers do. Analysis in this article cannot resolve this discrepancy as we do not observe candidates who do not take positions, data that could be quite informative.[35] If job candidates would prefer almost any teaching job to not teaching, the finding that some more able applicants do not find jobs would support Ballou's conclusion that hiring authorities do not value stronger academic qualifications. However, many

[35] The empirical framework can easily be generalized to include candidates who do not find, or are not willing to accept, jobs as well as job openings that are not filled because either no one is willing to accept the positions or employers are unwilling to hire the willing candidates.

of the candidates with stronger qualifications who do not end up teaching simply may be unwilling to teach in the schools where lesser-qualified candidates obtain jobs. In an effort to disentangle the choices of employers from the possible unwillingness of individuals to teach in the schools where they can get jobs, we have recently analyzed job-level applications data for New York City teachers seeking to transfer within the city system (Boyd et al. 2011). Among the applicants for individual jobs, we find that candidates having stronger qualifications are more likely to be hired.

Possibly reflecting personal connections affecting the interest of employers for candidates, models I and II indicate that the distance measure also is important. For example, a candidate being 11 rather than 5 miles distant from the school is estimated to lead to a reduction in attractiveness to an employer that is comparable to the reduction associated with a 10-point lower certification exam score, from 230 to 220. Because of the declining importance of distance, the 10-point score difference is estimated to have the same effect as the candidate being 21 versus 10 miles distant. The difference in candidate attractiveness associated with having graduated from a least selective, as opposed to a highly selective, college is comparable to a 13-mile increase in distance from 10 to 23 miles.

The estimated coefficient for the variable indicating the minority status of job candidates is both small in magnitude and not statistically significant, suggesting that employers are generally unconcerned with a teacher's race.

Candidates' Preferences for Job Attributes

White teachers prefer schools with a greater proportion of white students, and nonwhite teachers appear to be indifferent. On average, white teachers would be indifferent between an otherwise identical school that had a 10 percentage point higher minority student enrollment and one that had a 37 percentage point higher enrollment of students in poverty. This result, along with race not appearing to be important in the selection of teachers by hiring authorities, bears on long-standing questions going back to the work of Antos and Rosen (1974) concerning racial preferences in teacher labor markets.

After we account for the race and poverty of students, school proximity, and salary, urban schools are estimated to be relatively unattractive to those seeking teaching jobs. For example, the estimated coefficient for the urban dummy variable (−0.925) is slightly larger than the estimated effect for white teachers from a 19 percentage point difference in the proportion of minority students in a school—roughly one-third of the mean difference between urban and suburban schools in our sample.

The comparison of the estimated coefficient for salary with those for other school attributes has the potential of yielding estimates of compensating differentials for school working conditions. However, the estimated effect of salary is not statistically significant in either model I or II but was positive and statistically significant in some other model specifications. This lack of robustness and two additional considerations cause us to be cautious in making inferences using the estimated effect of changes in salary. Because the variation in salary across districts in our sample is quite small, questions arise regarding the extent to which any estimated effect of small salary differences can be extrapolated to larger salary changes. More importantly, observed salary differences are likely to reflect differences in unobserved working conditions. In such cases the estimated salary coefficient will reflect how a teacher's evaluation of a job changes as a result of a salary change that is accompanied by the correlated changes in those unobserved job attributes not accounted for by other school attributes included in the analysis. To the extent that salaries are higher in jobs having higher levels of unobserved attributes that are unattractive to those seeking teaching jobs, the estimated salary coefficient on average will understate the effect of an increase in salary, ceteris paribus.

The direct effect of salary on job attractiveness could be identified using an extension of the estimation strategy employed here where the moment conditions are modified to take advantage of credible instruments for salary. The challenge is identifying instrumental variables that meaningfully affect the salaries paid by districts but do not affect the attractiveness of jobs, other than through school variables included in the preference equation for candidates.

Distance.—Our most striking finding concerning teacher preferences is the importance of distance as a determinant of their evaluations of school alternatives.[36] For example, consider two schools, one having attributes equal to the averages for all those urban schools having job openings in year 2000 and 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 I indicate that a white teacher 1 mile from such an urban school would prefer teaching there, provided that the suburban alternative was at least 21 miles away.

The estimates of μ_o, γ, and σ^* in model II are all statistically significant. The sign of $\hat{\gamma}$ indicates that the magnitude of the effect of distance is smaller for teachers having greater qualifications, here proxied by the certification exam score.[37] Evaluated at the mean value of the certification exam

[36] In related work we have found that a school's geographical proximity is important in determining whether an individual teaching in a particular school decides to transfer to another school or to leave teaching altogether (Boyd et al. 2005b).

[37] The same pattern was found when qualifications were proxied using a composite teacher qualification index.

score (260.2), the estimates of μ_o, γ, and σ^* imply that the mean, median, mode, and standard deviation of β_{6j} are -1.98, -1.95, -1.89, and 0.36, respectively. Here the mean is close to the estimate of the distance coefficient for candidates in model I. However, the 0.36 standard deviation of the distance coefficient in model II indicates that there is significant dispersion with respect to the importance of school proximity for teachers.

As noted above, a potential advantage of the empirical model developed here is the ease with which both observed and unobserved preference heterogeneity can be taken into account, an important example being preferences associated with teacher-job proximity. The large magnitude of the estimated distance coefficient underscores that this is important. In fact, the apparent preference of candidates for teaching in schools nearby can help explain why there does not appear to be complete assortative matching of teachers and schools. Disregarding the importance of distance, the otherwise most attractive schools do not exclusively employ teachers having either the strongest qualifications or those most effective in improving student learning. Even though there is partial assortative matching, research shows that teachers having the strongest academic qualifications teach in a variety of settings. Similarly, studies estimating teacher value added find that within-school differences often exceed between-school differences. Such findings can be explained at least in part by the heterogeneity in the preferences of candidates for schools.

Even though care is needed when comparing the two estimated distance coefficients for candidates and employers, a comparison is possible. For the actual matches observed in our data, the mean and standard deviation for teacher-school distances are 8.6 and 8.3 miles, respectively.[38] If distance is increased by one standard deviation from the mean, the attractiveness of a school to a candidate is estimated to decline by 1.25 for a teacher whose distance coefficient is the mean value for model II (-1.98). With the random error in the preference equation for candidates having a standard deviation of one, a one standard deviation increase in distance has an estimated effect on a candidate's assessment of a school that is comparable to roughly a 1.25 standard deviation reduction in the random error. For employers, a one standard deviation increase in distance is estimated to have an effect on their assessment of a candidate comparable to roughly one-fifth of a standard deviation reduction in the random error of the preference equation. Thus, compared to the effects of all unmeasured factors in the two criterion functions, distance is more important in the assessments of schools by candidates than in the assessments of candidates by employers.

How well does our estimated model account for candidates' overall assessments of the available job opportunities and employers' assessments of

[38] This excludes the 6.6% of new teachers whose distance to their schools exceeded 50 miles.

job candidates, including both the nonstochastic and random components in the preference equations? As explained in appendix C, these questions can be addressed in a relatively straightforward manner. Let σ_v^2 represent the typical within-market-year variance in the assessment of candidates by employers, v_{jk}. In model I the explanatory variables entering employers' preferences for candidates explain 6.6% of this typical within-market-year variation in v_{jk}. In contrast, the school attributes entering the criterion function of candidates explain 48% of the typical within-market-year variance in the assessments of schools by candidates (e.g., the variance of u_{jk}).

VIII. 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 contrast, our simulated moments estimates of the two-sided matching model are consistent with the hypotheses that schools prefer teachers having stronger qualifications, and teachers prefer schools that are closer to home, have fewer poor students, and, for white teachers, have fewer minority students. While these results may appear predictable, they contradict many of the findings from prior research estimating hedonic wage equations for teacher labor markets. For example, prior studies, as well as our estimates of wage regressions employing the data used in this study, show little relationship between wages and teacher attributes. This result has been used to suggest that districts do not care about teacher quality. Similarly, it has been estimated that there is a negative relationship between student poverty in a school and teacher wages. The estimated effect of poverty using the two-sided matching approach is both theoretically plausible and consistent with empirical results in analyses of teacher retention.

The model can be extended to include the analysis of who becomes a teacher (e.g., which candidates find acceptable jobs) as well as teacher quits, transfers, and vacancy chains (i.e., employed teachers moving into vacancies, thereby creating vacancies that in turn must be filled). The model can be generalized to allow for candidates preferring not to teach rather than teaching in particular schools, and employers preferring to leave jobs open rather than hiring particular candidates, given information characterizing the candidates who sought, but did not take, teaching jobs and positions that were left vacant. In cases in which good instruments for salaries are available, these instruments can be employed in estimation to identify the marginal effect of changes in salary and, in turn, the compensating wage differentials needed so that workers with particular attributes would be indifferent between jobs having different characteristics as well as the pay increase needed to improve the quality of workers hired in jobs having particular attributes, which depends on preferences as well as the distributions of worker and job attributes. These extensions were not included in the current article because of data limitations and our desire to explore estimation of the basic model without adding further computational complexity. Given the plausibility of results for the basic empirical two-sided match model, there appears to be good reason to explore such extensions of the framework.

In summary, this article 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 also shows promise for estimating the preferences of both employers and employees in other labor markets not characterized by rapid adjustment of wages and flexible working conditions.

Appendix A

As noted in Section III, simulation is used to compute values of $E(\tilde{z}_{mtj}|q_{mtj};\ \theta)$ and $E(\tilde{d}_{mtj}|q_{mtj};\ \theta)$ in the moment conditions. Let $F(\tilde{z}_{mtj}|q_{mtj};\ \theta)$ and $F(\tilde{d}_{mtj}|q_{mtj}; \theta)$, respectively, represent these simulated values, 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^h_{mtjk} for the hth draw. Similarly, the hth draw includes randomly generated values for ω^h_{mtjk}. These randomly generated values are held constant throughout the estimation.

Step 2: For a given set of parameter values ($\theta = (\alpha, \beta)$) and the random numbers drawn in step 1, we compute the simulated moments as follows. The implied nonstochastic components of utility along with the values of δ^h_{jk} and ω^h_{jk} 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., $\tilde{z}^h_{mtj}$ and $\tilde{d}^h_{mtj}$ for each of the candidates hired in the hth 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) used in estimation:[39]

[39] In contrast to $\sum_t\sum_j[\tilde{d}_{mtj} - F(\tilde{d}_{mtj}|q_{mtj};\ \theta,\ H)] = 0$, which enters (A2), the moment condition $\sum_t\sum_j[\tilde{z}_{mtj} - F(\tilde{z}_{mtj}|q_{mtj};\ \theta,\ H)] = 0$ is not employed in estimation as the latter condition holds exactly for all values of θ.

$$F(\tilde{z}_{mtj}|q_{mtj};\ \theta,\ H) = \frac{1}{H}\sum_{h=1}^{H} \tilde{z}_{mtj}^{h} \approx E(\tilde{z}_{mtj}|q_{mtj};\ \theta),$$
$$F(\tilde{d}_{mtj}|q_{mtj};\ \theta,\ H) = \frac{1}{H}\sum_{h=1}^{H} \tilde{d}_{mtj}^{h} \approx E(\tilde{d}_{mtj}|q_{mtj};\ \theta), \tag{A1}$$

$$\psi_m = \sum_t \sum_j \begin{bmatrix} \psi_{mtj}^{a} \\ \psi_{mtj}^{b} \\ \psi_{mtj}^{c} \end{bmatrix} = \sum_t \sum_j \psi_{mtj} = 0, \tag{A2}$$

where

$$\psi_{mtj}^{a} \equiv q_{mtj}[\tilde{z}_{mtj} - F(\tilde{z}_{mtj}|q_{mtj};\ \theta,\ H)],$$
$$\psi_{mtj}^{b} \equiv q_{mtj}[\tilde{d}_{mtj} - F(\tilde{d}_{mtj}|q_{mtj};\ \theta,\ H)],$$
$$\psi_{mtj}^{c} \equiv [\tilde{d}_{mtj} - F(\tilde{d}_{mtj}|q_{mtj};\ \theta,\ H)].$$

Defining $\psi(\theta)$ to be a column vector containing the stacked values of ψ_1, $\psi_2, \ldots,\ \psi_5$ for the five markets, our method of simulated moment (MSM) estimator is $\arg\min_\theta \psi(\theta)'\psi(\theta)$.[40]

The asymptotic covariance matrix of this estimator is

$$V(\hat{\theta}\,) = \frac{1 + 1/H}{N}[D'\ D]^{-1}D'\Omega D[D'\ D]^{-1},$$

where N is the total number of job candidates included in the moment conditions and H is the number of simulations. (We employ $H = 1{,}000$ in all

[40] Note that

$$\psi' = \left[\ \sum_t \sum_{j=1}^{J_{1t}} \psi'_{1tj}\ \ \sum_t \sum_{j=1}^{J_{2t}} \psi'_{2tj}\ \cdots\ \sum_t \sum_{j=1}^{J_{Mt}} \psi'_{Mtj}\ \right]$$

defined above is equivalent to

$$\psi' = \sum_m \sum_t \sum_{j=1}^{J_{mt}} [\,d_m^1 \psi'_{mtj}\ \ d_m^2 \psi'_{mtj}\ \cdots\ d_m^M \psi'_{mtj}\,],$$

where d_m^z equals one if $m = z$ and zero otherwise. An alternative specification would be to employ

$$\psi' = \sum_m \sum_t \sum_j [\,\psi_{mtj}^{a'}\ \ \psi_{mtj}^{b'}\ \ \psi_{mtj}^{c'}\,] = \sum_m \sum_t \sum_j \psi'_{mtj}.$$

The primary difference is that there is a moment condition $\psi_{mtj}^{c} = 0$ in (A2) for the average distance traveled by candidates in each market m whereas $\Sigma_m\Sigma_t\Sigma_j\psi_{mtj}^{c} = 0$ in the alternative specification is a single-moment condition corresponding to the average distance of candidates across all markets.

simulations.) With $D \equiv E[\partial\psi(\theta_0)/\partial\theta']$, we employ simulation and numerical derivates to compute $\sum_t\sum_j[\partial\psi_{mtj}(\hat{\theta})/\partial\theta'] \approx D_m$ in D. The term Ω in $V(\hat{\theta})$ is the asymptotic variance of $\psi(\theta_0)$ shown in (A3):

$$\begin{aligned}\Omega = E[\psi\psi'] &= \begin{bmatrix} E\psi_1\psi_1' & 0 & \cdots & 0 \\ 0 & E\psi_2\psi_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E\psi_5\psi_5' \end{bmatrix} \\ &= \begin{bmatrix} \Omega_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega_5 \end{bmatrix}.\end{aligned} \tag{A3}$$

If the elemental moments, ψ_{mtj}, were independent across candidates (j) for each market-year, the mth diagonal block in Ω could be approximated 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. 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 $E(\tilde{z}_{mtj}|q_{mtj};\ \theta)$ and $E(\tilde{d}_{mtj}|q_{mtj};\ \theta)$. The terms $\tilde{z}_{mtj}^h$ and $\tilde{d}_{mtj}^h$ represent the school attributes and distance for the jth teacher's match in simulation h. Substituting these expressions for $\tilde{z}_{mtj}$ and $\tilde{d}_{mtj}$ in (A1) that characterize the jth teacher's actual match yields the expressions in (A4). These are based on the difference between the model-predicted match for simulation h and the simulated expected values $F(\tilde{z}_{mtj}|q_{mtj};\ \hat{\theta},\ H)$:

$$\psi_m^h = \sum_t\sum_j \begin{bmatrix} \psi_{mtj}^{ah} \\ \psi_{mtj}^{bh} \\ \psi_{mtj}^{ch} \end{bmatrix}, \tag{A4}$$

where

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

and $F(\tilde{d}_{mtj}|q_{mtj};\hat{\theta}, H)$. Averaging across the H draws, the simulated second moment of ψ_m is

$$\hat{\Omega}_m = \frac{1}{H}\sum_{h=1}^{H}\psi_m^h\psi_m^{h'} \approx E\psi_m\psi_m'.$$

Turning to the asymptotics of the estimator and the standard errors, reconsider the asymptotic covariance matrix

$$V(\hat{\theta}) = \frac{1 + 1/H}{N}[D' \, D]^{-1} D' \, \Omega \, D \, [D' \, D]^{-1},$$

where the total number of candidates, $N = \Sigma_m \Sigma_t J_{mt} = MT\bar{J}$, reflects the number of market-years (MT) included in the analysis as well as the average number of jobs/job seekers per market-year, $\bar{J}$. Because there is a parallel between this structure and that in panel data models, we are able to draw on asymptotic theory developed for panel data models—in particular the work of Phillips and Moon (1999).

Asymptotic properties of the estimator and standard errors depend on the extent and nature of any correlation between the ψ_{mtj}. As noted above, the nature of the candidate-job sorting in a particular market and year implies that the ψ_{mtj} are correlated within each market-year. Even though the basic model employed in this article assumes that the ψ_{mtj} are independent across years in a particular market, this need not be the case.[41] However, it is quite reasonable to assume that the ψ_{mtj} are independent across markets.

One can consider asymptotic properties associated with $\bar{J} \to \infty$, $T \to \infty$, or $M \to \infty$ as well as various combinations (e.g., a sequential limit where $T \to \infty$ and then $M \to \infty$ or a simultaneous limit with T and M increasing along some specified path such as $T(M) = cM$, where c is some constant). In our application, the independence of moments across markets implies that asymptotic results are straightforward when $M \to \infty$ for given $\bar{J}$ and T as well as $M \to \infty$ and then T, $\bar{J} \to \infty$. More generally, the asymptotic results associated with this particular sequential convergence will be the same as that resulting from any simultaneous path in which the J_{mt}, T, and M pass to infinity under some uniform integrability and uniform boundedness conditions (Phillips and Moon 1999).

Appendix B

Model Identification

Even though a complete analysis of identification for the two-sided matching model goes beyond the scope of this article, a number of useful insights follow from properties of related empirical models. Reconsider the case in which the gth (hth) candidate is employed in job g' (h'). Stability and the structure of revealed preferences imply that $1(u_{gg'} < u_{gh'})1(v_{hh'} < v_{gh'}) = 0$. Contrast this to the case in which matchings are one-sided. 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, $1(u_{gg'} < u_{gh'}) = 0$ for all $h', h' \neq g'$. Similarly, if the hiring authority filling job h' were able

[41] For example, correlation across years would follow from a model with school-level random effects that are constant across years.

to employ any candidate, the employer would hire candidate h only if $1(v_{hh'} < v_{gh'}) = 0$ for all $g, g \neq h$. Such decision rules underlie discrete-choice random-utility models 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 the candidate's preferences over jobs implies the expression

$$u(z^1_{g'}|q^2_g,\ \beta) + \delta_{gg'} > u(z^1_{h'}|q^2_g,\ \beta) + \delta_{gh'}$$

for all $h', h' \neq g'$. As discussed by Manski (1995, 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. 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 random errors in a multinomial probit model (e.g., the variance of at least one of the random errors must be fixed; see Dansie 1985; Bunch and Kitamura 1989; Bunch 1991; Keane 1992). 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 (see, e.g., Ben-Akiva and Lerman 1985). Consider the linear-in-parameters specification

$$u(z^1_{h'}|q^2_g,\ \beta) = \beta z^1_{h'} + \gamma q^2_g + (q^2_g)'\Lambda z^1_{h'}$$

for the case in which candidate g can freely choose among a given set of job openings; β and γ are vectors of parameters, and Λ is a conforming matrix of parameters. In this specification, γq^2_g 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^2_g is interacted with the attributes of alternatives or has coefficients that vary across alternatives. However, dropping γq^2_g from the equation is of no consequence. 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 matching.

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

$$y = y_1 y_2 = 1(\theta_1 x_1 + \eta_1 < 0)1(\theta_2 x_2 + \eta_2 < 0).$$

In this case, the identification of θ_1 and θ_2 crucially depends on 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 (see Poirier 1980).

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 $1(u_{gg'} < u_{gh'})1(v_{hh'} < v_{gh'}) = 0$, where there is one such condition for each candidate (g) and job (h') pair not actually matched. Comparing the utility expressions $u_{jk} = u(z_k^1|q_j^2, \beta) + \delta_{jk}$ and $v_{jk} = v(q_j^1|z_k^2, \alpha) + \omega_{jk}$ 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 certain cases, exclusion restrictions will follow from reasonable a priori assumptions. For example, one might reasonably assume that the starting salary at a school will affect how candidates value that alternative while that salary does not affect the school's relative evaluation of the candidates who apply. Exclusion restrictions also naturally arise in the two-sided matching model, even when there are no differences in the variables entering $u(\ )$ and $v(\ )$. Consider the linear-in-parameters second-order Taylor approximations $u(z_k|q_j\beta) = \beta z_k + q_j' \Lambda z_q$ and $v(q_j|z_k, \alpha) = \alpha q_j + z_k' \Psi q_j$. 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 $u(\ )$ linearly, just as z_k does not enter $v(\ )$ linearly, thus implying very general a priori exclusion restrictions in two-sided match models.[42]

This discussion of identification focuses 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 account only for the attributes of those entering matches, not the identities of those entering each candidate-job pairing, as accounted for in the condition $1(u_{gg'} < u_{gh'})1(v_{hh'} < v_{gh'}) = 0$. The identifying information contained in the structure of revealed preferences represents an upper bound with

[42] 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.

respect to identification within our generalized method of moments framework.

Appendix C

Let $\sigma^2_{v_{\cdot k}}$ represent the variance of v_{jk} across candidates as viewed from the perspective of the kth hiring authority. (Year and market indicators, i.e., t and m, are implicit.) In turn, σ^2_v is the mean value of $\sigma^2_{v_{\cdot k}}$ across schools, years, and markets, reflecting how employers' evaluations of candidates typically vary across candidates. What portion of this variance is explained by the typical within-market-year variation in the attributes of candidates included in the estimated model? Let $v_{jk} = \alpha X_{jk} + \omega_{jk}$, where X_{jk} is the vector of attributes characterizing the jth candidate, possibly interacted with variables characterizing the kth school to allow for heterogeneity in employers' evaluations of candidates. For school k in market m during period t, let Ψ_{kmt} represent the covariance of X_{jk} across the relevant universe of candidates.[43] The expected value of Ψ_{kmt} across schools, years, and markets, Ψ, represents the typical variation in the attributes of candidates weighted by school attributes in a typical market period. If X_{jk} is uncorrelated with the random error ω_{jk}, $\sigma^2_v = \alpha\Psi\alpha' + \sigma^2_\omega$, where $\sigma^2_\omega = 1$. Thus, estimates of α and Ψ yield an estimate of σ^2_v; $\hat{\sigma}^2_v = \hat{\alpha}\hat{\Psi}\hat{\alpha}' + \sigma^2_\omega$.

In a similar way, σ^2_u can be estimated along with the proportion of this variation that is explained by the variation in variables included in the preference equation for candidates.

References

Abowd, John M., and Francis Kramarz. 1999. The analysis of labor markets using matched employer-employee data. In *Handbook of labor economics*, vol. 3, pt. B, ed. Orley C. Ashenfelter and David E. Card. Amsterdam: North-Holland.

Antos, Joseph R., and Sherwin Rosen. 1975. Discrimination in the market for public school teachers. *Journal of Econometrics* 3, no. 2:123–50.

Ballou, Dale. 1996. Do public schools hire the best applicants? *Quarterly Journal of Economics* 111, no. 1:97–133.

Ballou, Dale, and Michael Podgursky. 1997. *Teacher pay and teacher quality*. Kalamazoo, MI: Upjohn Institute for Employment Research.

Baugh, William H., and Joe A. Stone. 1982. Mobility and wage equilibration in the educator labor market. *Economics of Education Review* 2, no. 3: 253–74.

Ben-Akiva, Moshe, and Steven R. Lerman. 1985. *Discrete choice analysis: Theory and application to travel demand*. Cambridge, MA: MIT Press.

[43] The value of Ψ_{kmt} will be the same for all k in a market period if none of the candidate attributes in X_{jk} are interacted with school attributes (i.e., $X_{jk} = X_j$).

Berry, Steven T. 1992. Estimation of a model of entry in the airline industry. *Econometrica* 60, no. 4:889–917.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, Matt Ronfeldt, and James Wyckoff. 2011. The role of teacher quality in retention and hiring: Using applications-to-transfer to uncover preferences of teachers and schools. *Journal of Policy Analysis and Management* 30, no. 1:88–110.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2002. Initial matches, transfers, and quits: The role of teachers' career decisions in the disparities in average teacher qualifications across schools. Working paper, Teacher Policy Research, Albany, NY.

———. 2003. Analyzing the matching of public school teachers to jobs: Estimating compensating differentials in imperfect labor markets. Working Paper no. 9878, National Bureau of Economic Research, Cambridge, MA.

———. 2005a. The draw of home: How teachers' preferences for proximity disadvantage urban schools. *Journal of Policy Analysis and Management* 24, no. 1:113–32.

———. 2005b. Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Review Papers and Proceedings* 95, no. 2:166–71.

Brewer, Dominic J. 1996. Career paths and quit decisions: Evidence from teaching. *Journal of Labor Economics* 14, no. 2:313–39.

Brown, Charles. 1980. Equalizing differences in the labor market. *Quarterly Journal of Economics* 94, no. 1:113–34.

Bunch, David S. 1991. Estimability in the multinomial probit model. *Transportation Research B* 25, no. 1:1–12.

Bunch, David S., and Ryuichi 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 no. 70, Transportation Center, University of California, Davis.

Burdett, Kenneth, and Melvyn Coles. 1999. Long-term partnership formation: Marriage and employment. *Economic Journal* 109, no. 456:F306–F334.

Choo, Eugene, and Aloysius Siow. 2006. Who marries whom and why. *Journal of Political Economy* 114, no. 1:175–201.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor. 2008. Would higher salaries keep teachers in high poverty schools? Evidence for a policy intervention in North Carolina. *Journal of Public Economics* 92, nos. 5–6:1352–70.

Dansie, B. R. 1985. Parameter estimability in the multinomial probit model. *Transportation Research B* 19, no. 6:526–28.

Dolton, Peter J. 1990. The economics of UK teacher supply: The graduate's decision. *Economic Journal* 100, no. 400:91–104.

Dolton, Peter J., and Gerald H. Makepeace. 1993. Female labour force participation and the choice of occupation. *European Economic Review* 37, no. 7:1393–1411.

Dolton, Peter J., and Wilbert van der Klaaw. 1999. The turnover of teachers: A competing risks explanation. *Review of Economics and Statistics* 81, no. 3:543–52.

Eckstein, Zvi, and Gerald J. van den Berg. 2007. Empirical labor search: A survey. *Journal of Econometrics* 136, no. 2:531–64.

Ekeland, Ivar, James Heckman, and Lars Nesheim. 2004. Identification and estimation of hedonic models. *Journal of Political Economy* 112, no. 1, pt. 2:S60–S109.

Epple, Dennis. 1987. Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. *Journal of Political Economy* 95, no. 1:59–79.

Fox, Jeremy T. 2009a. Estimating matching games with transfers. Working Paper no. 14382, National Bureau of Economic Research, Cambridge, MA.

———. 2009b. Structural empirical work using matching models. Unpublished manuscript, Department of Economics, University of Michigan.

Gale, David, and Lloyd S. Shapley. 1962. College admissions and the stability of marriage. *American Mathematical Monthly* 6, no. 1:9–15.

Goldhaber, Dan, Katharine Destler, and Daniel Player. 2010. Teacher labor markets and the perils of using hedonics to estimate compensating differentials in the public sector. *Economics of Education Review* 29, no. 1:1–17.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 1999. Do higher salaries buy better teachers? Working Paper no. 7082, National Bureau of Economic Research, Cambridge, MA.

———. 2004. Why public schools lose teachers. *Journal of Human Resources* 39, no. 2:326–54.

Hanushek, Eric A., and Richard R. Pace. 1995. Who chooses to teach (and why)? *Economics of Education Review* 14, no. 2:101–17.

Harris, Douglas N., and Tim R. Sass. 2009. The effects of NBPTS-certified teachers on student achievement. *Journal of Policy Analysis and Management* 28, no. 4:55–80.

Hwang, Hae-shin, Dale T. Mortensen, and W. Robert Reed. 1998. Hedonic wages and labor market search. *Journal of Labor Economics* 16, no. 4: 815–47.

Keane, Michael P. 1992. A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics* 10, no. 2:193–200.

Lang, Kevin, and Sumon Majumdar. 2004. The pricing of job characteristics when markets do not clear: Theory and policy implications. *International Economic Review* 45, no. 4:1111–28.

Lankford, Hamilton. 1999. *A descriptive analysis of the New York State and New York City teaching force*. Plaintiff's Exhibit P1482 for the New York Supreme Court case *Campaign for Fiscal Equity vs. the State of New York et al.*

Lankford, Hamilton, Susanna Loeb, and James Wyckoff. 2002. Teacher sorting and the plight of urban schools. *Education Evaluation and Policy Analysis* 24, no. 1:37–62.

Levin, Jessica, and Meredith Quinn. 2003. Missed opportunities: How we keep high-quality teachers out of urban classrooms. New Teacher Project, New York.

Lucas, Robert E. B. 1977. Hedonic wage equations and psychic wages in the returns to schooling. *American Economic Review* 67, no. 4: 549–58.

Manski, Charles 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 *Public sector payrolls*, ed. D. A. Wise. Chicago: University of Chicago Press.

———. 1995. *Identification problems in the social sciences.* Cambridge, MA: Harvard University Press.

McLean, Robert A., Wayne R. Windling, and Paul R. Neergaard. 1978. Compensating wage differentials for hazardous work: An empirical analysis. *Quarterly Review of Economics and Business* 18, no. 3:97–107.

Mont, Daniel, and Daniel I. Rees. 1996. The influence of classroom characteristics on high school teacher turnover. *Economic Inquiry* 34, no. 1: 152–67.

Murnane, Richard J., Judith D. Singer, and John B. Willett. 1989. The influences of salaries and "opportunity costs" on teachers' career choices: Evidence from North Carolina. *Harvard Educational Review* 59, no. 3: 325–46.

Pakes, Ariel, and David Pollard. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57, no. 5:1027–57.

Pflaum, Susanna W., and Theodore Abramson. 1990. Teacher assignment, hiring, and preparation: Minority teachers in New York City. *Urban Review* 22, no. 1:17–31.

Phillips, Peter C., and Hyungsik R. Moon. 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67, no. 5:1057–1111.

Porier, Dale J. 1980. Partial observability in bivariate probit models. *Journal of Econometrics* 12, no. 2:209–17.

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica* 70, no. 6:2295–2350.

Rickman, Bill D., and Carl D. Parker. 1990. Alternative wages and teacher mobility: A human capital approach. *Economics of Education Review* 9, no. 1:73–79.

Rogerson, Richard, Robert Shimer, and Randall Wright. 2005. Search-theoretic models of the labor market: A survey. *Journal of Economic Literature* 43, no. 4:959–88.

Rosen, Sherwin. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82, no. 1: 34–55.

———. 1986. The theory of equalizing differences. In *Handbook of labor economics*, vol. 1, ed. Orley C. Ashenfelter and Richard Layard. Oxford: Elsevier.

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

Roth, Alvin E., and John H. Vande Vate. 1990. Random paths to stability in two-sided matching. *Econometrica* 58, no. 6:1475–80.

Sattinger, Michael. 1979. Differential rents and the distribution of earnings. *Oxford Economic Papers* 31, no. 1:60–71.

———. 1993. Assignment models of the distribution of earnings. *Journal of Economic Literature* 31, no. 2:831–80.

Sieg, Holger. 2000. Estimating a bargaining model with asymmetric information: Evidence from medical malpractice disputes. *Journal of Political Economy* 108, no. 5:1006–21.

Stinebrickner, Todd R. 1998. An empirical investigation of teacher attrition. *Economics of Education Review* 17, no. 2:127–36.

———. 1999. Estimation of a duration model in the presence of missing data. *Review of Economics and Statistics* 81, no. 3:529–42.

———. 2001. A dynamic model of teacher labor supply. *Journal of Labor Economics* 19, no. 1:196–230.

Theobald, Neil D. 1990. An examination of the influences of personal, professional, and school district characteristics on public school teacher retention. *Economics of Education Review* 9, no. 3:241–50.

Theobald, Neil D., and R. Mark Gritz. 1996. The effects of school district spending priorities on the exit paths of beginning teachers leaving the district. *Economics of Education Review* 15, no. 1:11–22.

Tinbergen, Jan. 1956. On the theory of income distribution. *Weltwirtschaftliches Archiv* 77:155–73.

van der Klaauw, Wilbert. 2000. On the use of expectations data in estimating structural dynamic models: An analysis of career choices. Working paper, Department of Economics, University of North Carolina, Chapel Hill.