The Pennsylvania State University
The Graduate School

ESSAYS ON THE ECONOMICS OF EDUCATION

A Dissertation in
Economics
by
Frank Erickson

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2019

The dissertation of Frank Erickson was reviewed and approved* by the following:

Mark J. Roberts
Professor of Economics
Dissertation Advisor, Chair of Committee

Paul L.E. Grieco
Associate Professor of Economics

Kalyan Chatterjee
Professor of Economics

John J. Cheslock
Associate Professor of Education

Barry W. Ickes
Professor of Economics
Head of the Department of Economics

*Signatures are on file in the Graduate School.

Abstract

This dissertation consists of three chapters on the economics of education. The first chapter empirically investigates assortative matching between elementary teachers and school districts through descriptive analyses of job matching and mobility. The second chapter continues the same research program by structurally estimating a model of teacher job choice. The final chapter goes in a different direction, using simulated data to study measurement of the 'value added' by teachers or other educational inputs to student test scores.

Contents

List of Figures

List of Tables

Preface

Chapter 3 of this dissertation is a coauthored work with Kaveh Akram and Robert H. Meyer that originated while we were working at the Value-Added Research Center in the University of Wisconsin–Madison, where Dr. Meyer served as the Director, Dr. Akram served as a Researcher, and I was an Assistant Researcher. The idea for the work was proposed by Dr. Meyer in a grant. The three of us together determined how to approach the questions involved, based on Dr. Meyer's expertise in ensemble estimators and value-added estimation, Dr. Akram's expertise in semiparametric methods, and a review of existing literature in statistics and econometrics. I took ownership of the writing, the particulars of the process for simulating data, and the implementation in code, working closely with Dr. Akram and iterating based on our regular discussions with Dr. Meyer.

Acknowledgments

I am deeply indebted to Edward Green and Mark Roberts for guidance and encouragement in this research project; and the Value-Added Research Center and my colleagues there for financial support and excellent guidance in the field of education research, its data, methods and policy concerns. In addition, Paul Grieco and Kalyan Chatterjee have provided invaluable comments at various stages of this project; and the North Carolina Education Research Data Center made data available to me and other researchers.

Chapter 3 of this dissertation is based upon work supported by the Institute of Education Sciences under IES grant # R305D100018, and partially fulfills Task 3 of that grant. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the IES.

Introduction

This dissertation consists of three chapters on the economics of education. The first chapter empirically investigates assortative matching between elementary teachers and school districts through descriptive analyses of job matching and mobility. The second chapter continues the same research program by structurally estimating a model of teacher job choice. The final chapter goes in a different direction, using simulated data to study measurement of the 'value added' by teachers or other educational inputs to student test scores.

CHAPTER 1: In the first chapter, I examine assortative matching in the public elementary school teacher labor market. The sorting is between teachers with better credentials and school districts that offer higher pay. Such sorting is of interest in education policy in the United States, where there are concerns about students' not having equitable access to good teachers and districts not being able to attract and retain good teachers (US Department of Education, 2009, 2016).

I study sorting as the outcome of a dynamic process as teachers move through their careers, since teacher quality evolves over time, through new degrees and more experience, and job matches have high persistence. I break down the teacher career paths into two parts, conducting descriptive analyses of patterns in teacher job mobility and credential growth.

I use statewide data from North Carolina between 1996 and 2011, where there is good reason to expect sorting. North Carolina districts are free to adjust the quality of their teaching workforce, through the provision of training or higher pay, but cannot adjust the quantity of teachers, since the state pays salaries only for a certain number of positions. With this data, I build career paths for new elementary teachers who stay in the profession for their first ten consecutive years. These career paths consist of yearly job histories, recording which district they worked in; and credential histories, recording their experience, degrees and certifications attained. I consider sorting on all three of these teacher attributes.

For districts, I primarily consider sorting according to pay differences, but also look at other attributes that may be relevant to teacher job mobility or credential growth. Pay differences in North Carolina during this period only arise from policies that district may set for paying supplements in excess of the base teacher salary that is funded by the state.

In my analysis, I focus on the labor market of the Research Triangle metropolitan area, which contains nine school districts which are near each other but very heterogeneous in pay and other attributes. I find that the higher-paying districts in the Triangle generally have more highly credentialed teachers. The highest-paying district has twice as many teachers holding post-bachelor degrees as the lowest-paying, and more than ten times as many teachers with advanced certificates. Fast-growing districts tend to go in the other direction, having teachers with worse credentials. One plausible explanation for the negative association between growth and credentials is that districts face frictions in filling new job openings, so a growing district has to recruit less well credentialed teachers to fill its growing workforce.

Looking at teachers' job mobility across Research Triangle districts, I find that the highest-paying districts hire teachers after they have acquired experience in other districts. Drawing an analogy to the labor market for athletes, we can think of these career flows as consistent with a story of low-paying districts implicitly taking a 'feeder' role or as suffering from 'poaching.'

Across the state, teachers' credentials grow faster in districts with more local school funding and better-educated parents, but these statistical relationships disappear when comparing districts within the Research Triangle. This might be because the districts are so close to each other that all teachers can reach all of the institutions of higher education in the metropolitan area.

CHAPTER 2: In the second chapter, I extend the empirical investigation of sorting in the teacher labor market using a structural model of teacher job choice with a simple model of district hiring rules. Estimates from this model allow me to evaluate how differences among teachers affect their labor-market matches and payoffs. Teacher preferences and district hiring both respond to district attributes, and I identify these two components of the model separately through an exclusion restriction, based on the belief that district growth affects hiring but not teachers preferences.

Instead of structurally modeling both sides of the labor market, I only consider teacher job choice, based on three assumptions. Teachers are hard enough to fire that I assume all job separations are by their choice. I further assume that districts' pay supplements and hiring practices do not evolve over the 15 years of interest. Finally, I assume that districts delegate their hiring processes so that teachers' receive job opportunities probabilistically based only on their own attributes. I model district hiring practices as stochastic screening thresholds applied to a 'teacher-quality residual,' a single index of teacher attributes capturing how teachers are rewarded through more job offers. It is 'residual' to the direct rewards provided by higher pay for certain teacher attributes.

With these assumptions, I model job mobility as a Markov Decision Process for teachers who at some point work in the Research Triangle (Rust, 1994). The teachers face transient payoff shocks for each district and movement costs associated with switching districts from one year to the next (Kennan and Walker, 2011; Fox, 2010). The teachers' job opportunities come in the form of a random choice set drawn each period, following Manski (1977). Credential acquisition from one year to the next is an exogenous process since my focus is on how teachers match with districts and, according to estimates from Chapter 1, credential acquisition is not very different across districts within the Research Triangle.

In addition to the credentials considered in the first chapter, I allow for one extra dimension of teacher heterogeneity in the teacher-quality residual, one that is observed by all participants in the labor market but not by the econometrician. I assume that this component of quality does not change across a teacher's career and integrate over it during estimation, following the approach of Kennan (2006) and discussion in Heckman (1981). The assumption that this component of quality is observable to market participants is consistent with the findings of Boyd et al. (2011), who find that employers can recognize good teachers. The need for such an unobserved component is suggested by Ballou (1996) and Hanushek and Rivkin (2010) who find that the hiring and performance of teachers are not adequately explained by their observable characteristics.

After estimating the parameters of the teachers' decision processes, I can evaluate how different teacher attributes are rewarded in the labor market. My main finding is that sorting on pay and job appeal in terms of other district attributes is relatively weak, since good teachers get much of their incremental payoff through a larger set of job offers rather than by sorting into the ex ante most appealing job.

The net discounted labor-market rewards to each dimension of teacher quality are relatively small across a teacher's career. A better teacher receives more and better job offers, but incurs large movement costs to take advantage of them.

The results suggest that the movement of teachers across districts leads to weak sorting, with most of the payoff to better teachers coming from district-specific shocks rather than pay and appeal differences. The finding that, while pay and appeal matter, their impact on behavior is relatively weak, is consistent with the findings of Kennan and Walker (2011) for interstate worker mobility in the United States. Along similar lines, the education statistics literature on teacher productivity has found that access to good teachers does not differ much across schools (Mansfield, 2012).

CHAPTER 3 (with Kaveh Akram and Robert H. Meyer): In this chapter, we consider statistical approaches to measuring the relative effectiveness of one teacher versus another in raising student test scores. Our focus is on approaches to estimating the vector of teacher effects that only use the ordinal information in test scores and so are 'robust' to monotone transformations of the scores. This project is motivated by caution regarding the assumptions underlying standard estimation of teacher value added. In particular, test scales are designed so that student proficiency can be measured; only rarely so that differences in scores between years measure growth; and never so that average differences should capture teacher proficiency. We compare robust and non-robust methods in terms of their mean-squared error under several test-rescaling scenarios using simulated data.

The robust methods we evaluate include ordinary least squares with dummies for each pretest value; semiparametric methods from Honoré and Powell (2001) and Robinson (1988); and a conditional-quantile estimator from Betebenner (2009). We also review robust methods by Härdle et al. (2004) and Ichimura (1993), but leave them out of our simulation exercise because they are too computationally costly to be run a large number of times.

Our results highlight the tradeoff between eliminating bias by weakening assumptions on functional form and reducing estimate precision. On balance, the robust methods lose little in terms of precision, while the methods that are vulnerable to test rescaling can suffer large misspecification bias.

Chapter 1

The Distribution of Teacher Quality across Heterogeneous School Districts: Descriptive Evidence from North Carolina

1.1 Introduction

Better teachers can make a big difference in student outcomes (Chetty et al., 2011). To design policy that ensures "equitable access" to good teachers (US Department of Education, 2009), we first need to understand the teacher labor market. If some school systems provide systematically better access to good teachers than others, we want to know whether this asymmetry is due to better attraction, retention or growth in human capital of teachers.

In this paper, I identify good teachers as those with better credentials and examine the extent to which they 'sort' into higher-paying school districts. I also look at job mobility and credential growth as processes that could lead to sorting. Throughout the analysis starting in section 1.7, I look at differences among public school districts in North Carolina's Research Triangle region with respect to elementary teachers.

Section 1.2 describes why sorting might be expected in this labor market and why I look at teacher credential and job dynamics. Section 1.3 introduces that data and how I use it. Sections 1.4, 1.5 and 1.6 explain my data modeling decisions on each side of the market and in building career histories. Section 1.7 discusses my analysis and results; and section 1.8 concludes.

1.2 The labor market for teachers

In this paper, I study the career paths of teachers working in North Carolina between 1995 and 2011. There are two constraints on employers in this market that could lead to systematic sorting of good teachers across school systems: employers' inability to customize pay and to adjust the size of their workforce. All else equal, these features should lead good teachers to sort into districts with higher pay and more attractive amenities, resulting in systematically unequal access for students.

First, pay cannot be customized for individual teachers. Employers can recognize good teachers (Boyd et al., 2011), but typically cannot reward them with higher compensation than their peers with similar experience, degrees and certificates. Because pay negotiation is not possible, matching with the right employer plays an important role in how good teachers are rewarded in the labor market.

Second, the quantity of teachers employed in a school system is inflexible, so only quality adjustment is available. Each school district has a fixed number of teaching positions, or 'allotments', that the state offers to fund according to its salary schedule. Hiring more or fewer than the allotted number of teachers would mean, respectively, much greater costs to the district or forgoing a large benefit.

Due to these constraints on employers, I expect more scope for sorting than might be found in a typical labor market, where each worker can be paid their marginal value no matter where they work.

Sorting might be realized in a couple different ways here. There may be better human resource growth in some school systems than others, which I study by estimating a model of credential acquisition. And there might be better attraction or retention of good teachers, which I study by looking at teacher credentials and flows between school systems. These analyses require data on 'career histories,' by which I mean individual teachers' credentials and employment over time, described in the next section.

1.3 Data on public school teacher salaries and work histories

I use data from North Carolina's public schools from 1995 to 2011, collected by the state's Department of Public Instruction (DPI); validated, documented and distributed by the North Carolina Education Research Data Center at Duke University; and used widely in education research.

The career history data cover all employees, their work activities and the state-financed portion of their payroll. I use salary codes and activity descriptors to select elementary teachers from the pool of all employees (in section 1.5.1); use other salary codes to construct their credential histories (in 1.6.1); and use employer indicators in both the activity and payroll data to build job histories (in 1.6.2).

I use data on teachers' education and licensing histories to validate the credential histories built with payroll salary codes in section 1.6.1.

I use publicly available data from the DPI for the state salary schedule (in section 1.5.2); and for the average salary supplement from each district (in 1.4.2). Ideally, I would use districts' full supplement schedules instead of their average supplements, but the former are not available.

I use counts of students by grade to compute teacher allotments set by the state (in section 1.4.1). During the analysis, I select other employer attributes from tables covering student performance and behavior, district financing, educational resources, and local demographics (in 1.7.1).

The next three sections explain how the data is used to construct the most important labor market variables, related to employers, in section 1.4; teachers in 1.5; and their career paths in 1.6.

1.4 School districts

On the employer side, the key labor market variables are the supplementary pay offered by the school system beyond teachers' state-financed salaries, and the number and growth in teaching positions allotted by the state. Before explaining how I construct these variables, I will briefly explain why districts are the appropriate unit of analysis on the employer side, instead of schools.

Looking at teacher careers at the school level, there are too few observations and usually too much noise. Individual staff or other local idiosyncrasies may have an outsize influence on whether a teacher wants or is offered a job at a particular school, while at the district level, such dynamics are more likely to average out. There is also noise in the size of the workforce, which is dictated by allotments at the district, but not the school level; as school catchment areas change; and as schools open and close. These may be interesting phenomena to investigate, but for this project I am starting with the simpler story of teacher sorting at the district level.

There are also practical issues if attempting to build job histories at the school level, as is done in section 1.6.2 for districts. Finally, supplementary pay is decided at the district level, and many attributes of school systems that we might want to look at, discussed in section 1.7.1, are not measured more finely than at the district level.

The next two sections explain my approach to capturing heterogeneity across districts in the elementary teacher labor market in quantity and price.

1.4.1 School district workforce size and growth

I am looking at the labor market for teachers between 1995 and 2011 and so want to track changes in employers' workforce over this period. I summarize the available information about the number of employees in each district as a level and a slope ('size' and 'growth').

The size of the teaching workforce in a district is determined by allotments set by the state, as mentioned in section 1.2. District allotments are not directly reported in the data, but can be approximated or substituted with a direct calculation of the number of employees. I measure workforce size as a count of employed elementary teachers in each district (as defined in 1.5.1), averaged over the years 1995-2011. This may imperfectly capture the number of alloted positions a district has, due to frictions in the adjustment of employment in response to changes in allotments.

For the workforce growth variable, I use the average growth in allotments according to the state's formula, which is based on how many students the district has, with some weighting applied for different grades.

1.4.2 School district pay supplements

Compensation differences across school systems arise from pay supplements that districts add to teachers' state-paid salaries. Both state salaries and district supplements are a function of a teacher's experience, highest-attained degree and national board certification.

While the state's salary schedule is known (and described in section 1.5.2), the districts' are not. Similarly, while salary payments from the state appear in the payroll data, those from districts do not.

Fortunately, mean teacher supplements (averaged over teachers in each district) are available in public reports from 2004 to 2011. The ranking of districts' supplements is pretty consistent for those years where it is reported. In my analysis, I use the values from the 2007-2008 academic year, which are representative. The endogenous allocation of teachers across districts makes average supplements a contaminated measure of differences in offered pay, but these averages are the best option available.[1]

With supplements and allotments, I have the key labor market variables on the district side. I now move on to processing data for teachers.

1.5 Elementary teachers

I restrict attention to elementary teachers because their labor market is simpler to characterize. Their responsibilities are fairly homogeneous across individuals and school districts, in contrast with middle and high school teachers, who may each have a different course load and compete in a smaller labor market (for example, for teachers who can teach both history and liberal arts) that is not constrained by state allotments.

In the next two subsections, I explain how elementary teachers are selected from the pool of all state education system employees in the data; and how the state determines their base salaries (which are paid alongside the district supplements discussed in section 1.4.2).

1.5.1 Selecting elementary teachers

I define an employee as an elementary teacher in a given year if they serve as the primary teacher for a classroom of students in an elementary school according to activity or payroll data (introduced in section 1.3).

Activities, including some not associated with teaching, are recorded per semester. Payroll data are reported per academic year. Employees often have multiple records in both files, which I denote as 'tasks'.

I designate an employee as an elementary teacher in a given year if any of their tasks, in payroll or activity data, satisfy certain criteria associated with elementary classroom teaching. For tasks in the payroll data, I use several budget codes which together indicate elementary classroom teachers funded by the state; and additionally require that the task be associated with at least two months of payroll. For tasks in the activity data, I use

[1] Some districts describe their current supplement schedules on their websites, but these descriptions are not always present and often provide incomplete information. In the appendix of Clotfelter et al. (2011), the authors describe a more complicated approach to characterizing supplement schedules.

classroom grades to identify elementary school responsibilities; and assignment types to recognize primary classroom teachers.

1.5.2 Elementary teacher state salaries

This section describes the state-financed salary schedule for elementary teachers. The schedule is not used directly in this paper, but provides context for teachers' credential growth, analyzed in section 1.7.3.

In North Carolina, the minimal requirements to work as a public school teacher are state certification and a bachelor's degree. Several further credentials yield higher state pay, as documented in public salary schedules: experience up to a maximum of around 30 years; three types of degrees beyond a bachelor's; and certification from the National Board of Professional Teaching Standards (hereafter, "National Board certification"). Teachers can be certified by the National Board only after their third year of experience. Its certifications are available in a number of categories that are treated equally by the state for compensation purposes.

The 2001 schedule for teachers with a bachelor's degree can be seen in Figure 1.1. There is an early period of slow growth; faster growth starting at three years of experience; and then moderate growth before pay goes entirely flat beyond around 30 years of experience. Besides 2001, the other years' salary schedules follow the same pattern, consisting of three stages with different rates of growth.

Schedules for teachers with more advanced degrees and National Board certification are simple shifts of the uncertified bachelor's schedule. A master's degree earns an extra 6.25% until 2000, and 10% thereafter. After accounting for a Masters', National Board certification adds 12%. Finally, those with more advanced degrees and doctorates earn \$126 and \$253 extra per month above a master's.

1.6 Career histories

An asymmetric distribution of good teachers could be realized through differing rates of credential growth and patterns of job mobility across school districts. To investigate these possibilities, I need to construct records of credentials and jobs, or 'career histories', for each elementary teacher.

I construct job and credential histories at the year level. It is rare for more than one job transition to occur within a year; and, while a teacher may acquire two different degrees within one year, I think it is reasonable to assume that only the higher degree affected their pay and job market situation.

For job histories, I assume teachers are assigned to one district at a time. While elementary teachers may in principle hold positions in multiple districts at once, those who do so are probably competing in an essentially different labor market from the typical classroom teacher (for example, as substitute teachers or traveling specialists).

To create career histories amenable to analysis, I need to assign each teacher a sequence of districts with pay grades that progress sensibly over time. For the majority of teachers, this is straightforward, but for those who change jobs or acquire new credentials, parsing out what happened from the data can be more complicated. Since these teachers are precisely the ones I am interested in, I try to resolve their data ambiguities instead of dropping observations.

Fortunately, the data is rich enough that almost all ambiguities in teachers' career paths can be resolved. Following a suggestion in the data provider's documentation, I primarily achieve this data cleaning step by validating payroll and activity data against each other. The next two sections describe the details of what that entails.

1.6.1 Credential histories

Credentials are captured by the teacher's pay grade, a string value in each record of payroll data that encodes the experience, degrees and national board certification applicable to a given task in a given year.[2]

I find 2.2 million payroll records for teachers and drop 790,000 that are associated with less than one month of work. This left about 6,900 teachers having multiple records with conflicting pay grades in at least one year. To resolve these conflicts, I kept records with the highest degree and then the record with the most years of experience.

This leaves at most one pay grade per teacher and year. The following sections explain how I validated these credential histories to ensure that they are internally consistent and complete.

Internally consistent credential histories

For teachers with records in more than one year, their credential histories are internally consistent if experience increases by zero or one each year and their degrees and National Board certification never decrease. Where inconsistencies cannot be resolved unambiguously using the credential histories alone, I also use separate data on when teachers acquired their degrees and teaching licenses.

I tackle experience before other credentials, finding that most inconsistencies are due to systematic and transparent data entry errors. I divide each teacher's history into spells in which experience is increasing. Most teachers have a single spell, but 3,000 have multiple, which means that their experience drops at some point, which should not happen.

Let T be the number of years of payroll records for a teacher, and $expr_t$ be a teacher's experience in a given year. 98% of these problematic records are for teachers at experience levels so high – above 30 years – that additional experience no longer increases their pay according to the state salary schedules discussed in section 1.5.2. For some of these, that are miscoded only in their final year in the data, I recode to $expr_T = expr_{T-1} + (year_T - year_{T-1})$. For others, that are miscoded as having 31 years of experience, I recode to $expr_t = \min(expr_{t-1} + (year_t - year_{t-1}), expr_{t+1})$ (where the first or second argument to the min is missing if $t = 1$ or $t = T$, respectively). Another hundred teachers have a jump downwards immediately after their first year, so I recode them at $expr_1 = \max(0, expr_2 - (year_2 - year_1))$. A few dozen other jumps downwards do not have such simple fixes and are carried over to the next step in cleaning.

[2] While my analysis is focused on elementary teachers, as noted in section 1.5.1, the steps described in this section are taken for all teachers in the data set to ensure that my cleaning process is general enough.

In the second step of cleaning teachers' experience histories, I collected spells where a teacher's experience grows at a rate less than or equal to the passage of time, as it should:

$$0 \leq \Delta expr_t \leq \Delta year_t, \tag{1.1}$$

Teachers with multiple spells include both those not fixed in the previous step as well as those whose experience increases more rapidly than time passes. There are 9,000 such teachers, some with more than one 'jump.' To resolve these, I select a single spell and work outward from it, lifting or dropping adjacent spells to make (1.1) hold. In particular, I change $\Delta expr_t$ to $\min(\Delta year_t, \max(\Delta expr_t, 0))$.

To select a spell to hold (while shifting the others), I first compare lengths. If the longest spell is more than twice the length of any other, it is selected. Next, I look at auxiliary data to define $minyr$, taken as the teacher's earliest license, degree or observation in the data. (Years of licensing and degree acquisition are taken from the teacher education and licensing data sets.) If the teacher had more experience in a spell than time had passed since they started in the profession, $expr_t > year_t - minyr + 1$, then other spells were preferred. For remaining ties, I take the most recent spell.

After these adjustments, I have sensible data on teachers' growth in experience over time. Next, I need to interpret teachers' growth in other credentials over time. For academic degrees, 1,000 records for 400 teachers have a decreasing sequence. In all but a few cases, I fix this by raising later degrees. The exceptions, for which I instead reduce earlier degrees, are recorded in payroll with a degree that the teacher either (i) acquires more than five years later according to their education history or (ii) only is recorded as having in the first year and never seemed to attain.

The final credential that should be nondecreasing is National Board certification, which fell for 400 teachers. I give all of them the benefit of the doubt, recognizing their certification in any year after its first appearance. Half of the losses of certification occurr in the teacher's final year; almost all the rest are listed as certified in the certification data set at some point; and the handful that do not meet these two criteria also looked legitimate.

After constructing an increasing sequence of credentials for the observations found in the payroll data, there still may be gaps in a teacher's credential history, as discussed in the next section.

Complete credential histories

For some teachers in some years, pay-grade data is missing and needs to be imputed. This may happen when a teacher takes a job out of state, for example. I fill in experience growth at the normal rate (of one year of experience per one year of time) when possible, and generally fill in degrees with their nearest observed value.

I split each teacher's history into spells in which pay-grade data was missing or not. For teachers with only one spell, I am done. With two or more spells, I define t^* as the first entry of the second spell. With missing data in the first spell, I set $expr_t = \max(expr_{t^*} - (y_{t^*} - y_t), 0)$, $deg_t = deg_{t^*}$ and $n_t = nb_{t^*}$, where deg indicates the teacher's highest degree; and nb, their National Board certification status. With missing data in the second spell, I set an upper limit, $uplim$, on experience at the first entry of the third spell, if any, or 99;

and then set $expr_t = \min(expr_{t^*-1} - (y_t - y_{t^*-1}), uplim)$, $deg_t = deg_{t^*}$ and $nb_t = nb_{t^*}$. The result is passed back recursively, imputing data for 16,000 teachers.

1.6.2 Job histories

To build job histories, I need to assign a single district as the teacher's employer for each year that they appear in the data as an elementary teacher. In this section, I use 'jobs' as a shorthand for teacher-year-district observations, so my task is to assign a unique job to each teacher-year observation.

For the 140,000 elementary teachers, I have records of 900,000 elementary jobs and 200,000 jobs with no elementary teaching tasks as defined in section 1.5.1. I assign the latter records to a dummy district representing the teacher being 'outside the market' of elementary teachers in the state's public schools. For the in-market records, around one thousand teacher-year combinations are associated with two districts, and only a handful have three. I drop teachers appearing in three districts, but retain those appearing in two, since the duplicate records likely represent teacher mobility across districts, which is of central interest to my research project.

With a teacher working in two districts – call them A and B – I look at where the teacher worked in adjacent years to decide:

- If the teacher was recorded as working at A both before and after, but B in neither of them, then A was kept.
- If A appeared before and B after, then A was kept.
- If A appeared before and neither of them after, then B was kept.
- If both appeared in one of the adjacent years, I zoomed out to look at the next year in that direction, and selected an entry using similar rules.
- It never happened that both A and B appeared both before and after (leading to three conflicts in a row), so I did not have to deal with this case.
- In a few dozen cases, neither appeared in adjacent years, and I made manual selections based on auxiliary data. For example, when only one district was a charter, the other was kept; and when only one entry showed up in the activity data, it was kept. In several cases it was still ambiguous, so I dropped the associated teachers.

These rules primarily rely on inferences based on internal consistency checks, similar to those discussed in 1.6.1; though I also choose some conventions to be apply to cases that cannot be unambiguously resolved from the available data.

When done with this preparatory work, I have 840,000 teaching jobs for 130,000 individuals who at some point serve as elementary teachers, with full credential and job histories that can be used for analysis.

1.7 Descriptive analysis on teacher sorting

In this section, I discuss results related to sorting of teachers into high-paying school districts. In most of the analysis, I select one cohort of teachers and focus on their career dynamics across districts in the Research Triangle region. I will briefly explain and justify these choices before getting to the substance of my analysis.

I analyze career dynamics for a cohort of around 3,000 elementary teachers entering the profession between 1996 and 2001 and staying in the profession for ten years (by accumulating ten years of experience in the ten years after entry). In looking at a single cohort, I can focus on the question of how these teachers sort across districts without extra complications arising from changes over time in compensation regimes, regulations or macroeconomic conditions.[3]

I am not looking at teachers' career paths beyond ten years in the labor market due to lack of data and the stabilization of teacher quality and mobility after the first few years. The education-statistics literature emphasizes that teacher quality grows rapidly in the first few years; and, in assessing equitable access to good teachers, the US Education Department attaches special significance to teacher quality differences in the first three years in the profession (US Department of Education, 2016). Table 1.1 shows that there is a spike in job mobility in the first four years, when teachers leave their jobs for new ones at an average rate of 10.0%, and that this flattens out in later years to an average rate of 5.2%.

I focus on differences in the labor market positions occupied by the nine districts in the Research Triangle region, depicted in Figure 1.2. These districts form a coherent local labor market, with close proximity and common membership in a metropolitan area as defined by the federal government. The region is fully interior to the state, and most teacher job mobility to or from these districts is with other parts of the state, so the statewide data I use has good coverage of the relevant labor market variables.

As a metropolitan labor market, the Research Triangle is not representative of the state as a whole. As shown in Table 1.2, the districts are larger, faster growing and better educated, and spend more on schools and teacher salary supplements. However, there is substantial differentiation among the nine districts that could lead to sorting, as discussed in 1.7.1 below.

Of the roughly 3,000 elementary teachers in the cohort entering between 1996 and 2001, 485 at some point work in the Research Triangle. They are somewhat better educated than their peers, as shown in Table 1.3. Looking at district averages in Table 1.4, the difference looks greater only because the smallest districts in the Research Triangle have the most highly credentialed teachers, as will be discussed in the next section.

1.7.1 Sorting on observables

If there is assortative matching in this market, its pattern should depend on district preferences over teacher attributes and teacher preferences over district attributes. In this section, I select attributes to look at and then compare their values across Research Triangle districts.

[3] Regarding teachers who do not stay in the profession, either by leaving the data set entirely or gaining experience at a slower rate, I exclude them because there is too little information available about their labor market activities. A teacher who leaves the data set could have either quit the profession or moved out of state, for example, and I would not be able to distinguish these two cases.

For district attributes, I selected relatively uncorrelated variables from the few dozen available by applying a clustering algorithm and selecting one variable from each cluster of correlated variables: the rate of short-term suspensions among students; the rate at which high-school seniors take the SAT (or the 'SAT uptake'); and the ratio of internet-connected computers to students.

For teacher attributes, I consider experience, degrees on entry and degrees and National Board certification in the current year. National Board certification is not available when teachers initially enter the profession; and so few teachers have degrees beyond a Masters' that I consider all advanced degrees together.

I collapse the selected district attributes over years, as shown in Table 1.5 alongside average salary supplements and workforce size and growth (introduced in 1.4); local educational funds and the share of parents with less than a high school education (which will be used in section 1.7.3's credential growth model). Looking at this table, we can see the large differentiation among these districts, with Chapel-Hill Carrboro and Wake paying more than twice as much as Harnett and Franklin in salary supplements, and having less than half as many parents with less than a high school education.

I collapse the selected teacher attributes over teacher-year observations in Table 1.6 alongside the average time (out of their first ten years) a teacher spent in the district. Looking at Tables 1.5 and 1.6 together, we see that higher pay is generally associated with better credentials. However, it is also clear that there is more to the story.

First, the association between pay and credentials is not perfect. Wake pays almost twice as much as Person and yet employs teachers with similar credentials. It may be that the differing growth rates of 0.65% in Person and 4.76% in Wake explain this. In a labor market model with frictions, fast-growing Wake needs to raise pay and lower its teacher quality standards to ensure that it fills its growing workforce (relative to where its pay and quality standards would be at a lower rate of growth).

Second, we see that credential growth varies a great deal for teachers in different districts. Orange hires two-thirds of its advanced-degree holders from teachers who enter the profession with their degrees; while Chapel Hill-Carrboro only hires 12% of its advanced-degree holders from that pool. From these tables alone, it is unclear if Chapel Hill-Carrboro is helping teachers acquire advanced degrees after they enter the profession or 'poaching' teachers after they improve their credentials. I investigate the possibility of credential growth differences across districts in section 1.7.3.

Third, we see that job mobility patterns are behind some teacher credential differences. Looking at Table 1.6, we see that Chapel Hill-Carrboro has the lowest average time spent and the highest average experience, indicating that it hires teachers later in their careers than the other districts do. The next section will investigate this feature of the Research Triangle labor market in greater detail.

1.7.2 Job mobility

Teachers in the selected cohort move between districts at a rate of 6.8% during their first ten years. Table 1.7 shows their movement statistics for the Research Triangle districts. The Department of Education regards year-to-year teacher retention as an important metric of equitable access to teachers (US Department of Education, 2016), and the rate at which teachers are retained in a given year can be seen in the 'Leave rate' column of the table.

Wake has a low share of leavers at 16.5%. This may partly be because its large size and fast growth (seen in Table 1.5) give teachers many options for internal moves. In contrast, the smaller and slower growing district of Orange has a much higher leaver share of 37.5%.

Chapel Hill-Carrboro stands out, with 86% of its jobs for such teachers filled by movers into the district; and only 7% by teachers who leave within their first ten years in the profession. In contrast, in Harnett 80% of teachers have started their careers in the district (instead of moving into it from elsewhere), and 37% eventually leave. This is consistent with sorting where districts like Harnett take teachers early in their careers and districts like Chapel Hill-Carrboro hire teachers later after their experience and other credentials have improved.

1.7.3 Credential growth

Acquisition of new credentials, either advanced degrees or National Board certification, could be important to sorting here in a number of ways. A district that sees high credential growth among teachers working there could be offering better professional development resources, which serve as a perquisite that draws teachers to the district. On the other hand, teachers might improve their credentials so that they are able to secure work in a more attractive different district in the future.

About 10,000 elementary-school teachers (out of 130,000 in the data) acquire new degrees on the job. Looking at Table 1.6, we see that the frequency with which advanced degrees are held by teachers differs a great deal across districts. In both Harnett and Chapel Hill-Carrboro, more than 80% of advanced degree holders acquire their degrees on the job; while in Franklin and Durham, fewer than 20% do.

Table 1.8 shows logit results where the probability of acquiring a new credential between academic years – an advanced degree or National Board certification – depends on the teacher's attributes (a polynomial in experience) and the district's (financing from local funds or the rate at which parents hold high-school diplomas).

I run separate regressions (1) and (2) for National Board certification conditional on whether the teacher has an advanced degree already, and vice versa in (3) and (4). For teachers able to acquire both, covered in (2) and (3), I do not model the probability of acquiring both at once, since the transition matrix for acquiring each passes a chi-squared test for independence (with a p value of 0.55 for observing these frequencies under the null). The final regression (5) applies to the special case of teachers who do not have National Board certification and cannot acquire it in the current year (since certification is only available after four years of experience).

For National Board certification, I included local school funds as a regressor because the certification process costs money and may be subsidized by the district. Based on the results, this seems to have a very weak effect.

For advanced degree acquisition, I included local parent education level as a proxy for the availability of educational resources, expecting there to be a negative relationship between the rate of parents with less than a high-school education and degree acquisition. The results instead point in the opposite direction, perhaps because teachers have greater incentive to acquire degrees in less educated districts to improve their options for job mobility, as discussed in the last section.

Table 1.9 shows results for the same regressions on the subset of teachers who worked in the Research Triangle at some point. The chi-squared test for independence in regressions (2) and (3) is passed here as well (with a p value of 0.91). The relationship between school funds and National Board certification is stronger here than in the statewide data.

The relationship between advanced degrees and local parents' education levels is weaker, perhaps because the educational opportunities are similar across districts in a single metropolitan area due to the commonly reachable institutions of higher education. The explanatory power of experience also disappears, perhaps for the same reason – acquiring a degree late in one's career to facilitate a move to a new district is less disruptive in a metropolitan area, where the new employer may still be within commuting distance from one's home.

These results highlight important caveats necessary when studying a metropolitan labor market and trying to extrapolate to a broader context. Geography means that the interaction across teachers' career paths between credential growth and job mobility is different in urban and rural markets.

1.8 Conclusion

In general, I expected to find sorting of better-credentialed teachers into higher-paying districts. Such labor market sorting along this single dimension is present, but imperfect, leaving open questions about how it comes about. To dig deeper in further work, one approach would be to add dimensions, considering a wider range of teacher credential differences that districts might have preferences over (like licensure test scores and degree-granting institutions); and of district amenities that teachers might value. There is good reason to believe that this approach would still miss important parts of the story of sorting here, however.

We have seen that asymmetric job mobility across districts underlies some sorting, with some districts tending to hire teachers only after they have acquired some experience elsewhere. Besides experience, acquisition of advanced degrees often occurs on the job and National Board certification always does, and we see these credentials being acquired at very different rates among teachers employed at different districts. Finally, district workforce growth seems to interact with job mobility in ways that fit a labor market model with frictions. For these reasons, I think the natural next step is to study such a model, accounting for credential growth, job mobility and districts' changing workforces jointly.

Figures

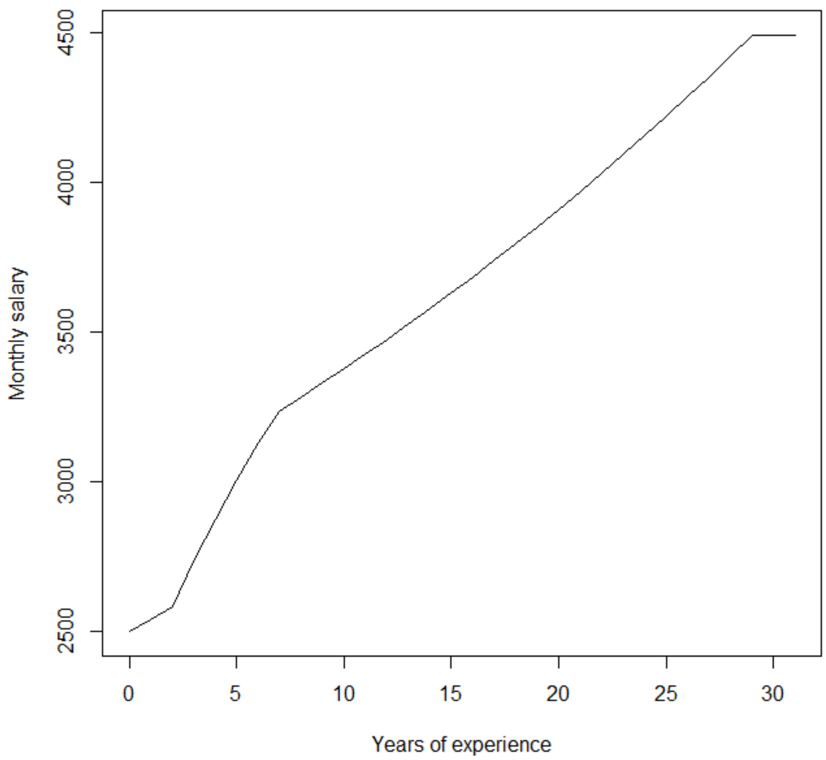


Figure 1.1: **Salary growth.** The state supported monthly salary schedule in 2001 for teachers with a Bachelor's and no National Board certification. Teachers are typically paid for ten months per year.

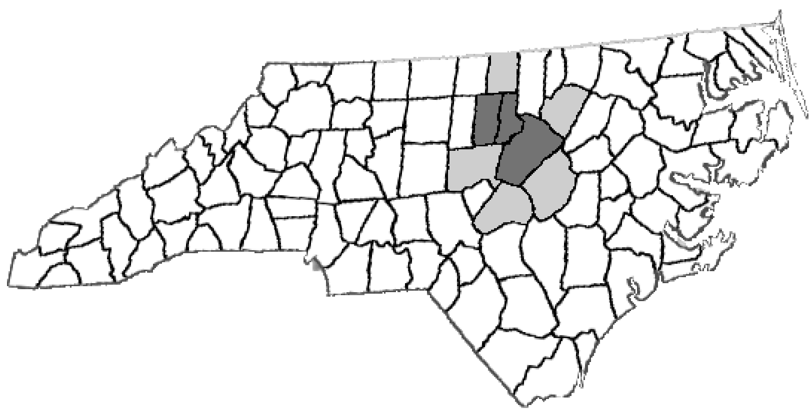


Figure 1.2: **The Research Triangle region.** The eight shaded counties form the Research Triangle, containing the nine public school districts that I study as a labor market for elementary teachers. The core counties are Orange, Durham and Wake (shaded more darkly from left to right); with Person and Chatham north and south of Durham; and Franklin, Johnston and Harnett running clockwise around Wake from the north. Each of the eight counties has its own school district, and the ninth district is Chapel Hill-Carrboro City Schools in Orange county. This definition of the Triangle comes from the US Office of Management and Budget which refers to it as the Raleigh-Durham-Cary Combined Statistical Area.

Tables

Years in market	N	Leave rate
1	485	10.7%
2	485	12.2%
3	485	9.1%
4	485	7.8%
5	485	5.8%
6	485	4.3%
7	485	5.8%
8	485	5.6%
9	485	4.1%
10	485	3.1%
11	385	8.1%
12	281	5.3%
13	203	4.9%
14	119	4.2%
15	49	6.1%

Table 1.1: **Teacher moves by their years in the market.** This table shows the rate at which the selected cohort of teachers who worked in the Research Triangle left one district for another. 'N' is the number of teachers used in the calculation of each rate.

Variable	Mean	SD	Middle 50%
Salary Supplement (2007-2008)	2319.72 4241.44	1503.69 1544.48	1346-2820 3147-5366
Workforce size	383.5 834.97	564.12 1122.67	105.54-441.06 266.53-836.71
Workforce growth (%)	0.47 1.8	1.58 1.49	-0.4-1.33 0.92-2.21
SAT Uptake (%)	53.99 63.97	9.54 15.41	47.96-58.28 49.77-73.6
Short-term Suspensions (per 1000)	2.08 1.71	1.08 0.66	1.26-2.76 1.36-2.17
Internet computers (per 100)	32.01 32.32	7.5 8.1	27.66-35.84 25.69-38.27
Local school funds ($1,000s)	2078.15 3148.68	864.56 1217.75	1544.59-2442.38 2084.44-3691.58
Local parents w/o HS (%)	18.4 14.54	5.43 5.25	14.25-21.26 13.69-18.39

Table 1.2: **Summary statistics for district attributes in North Carolina and the Research Triangle.** District means are taken over the years 1995-2011 (except for the salary supplement, which is taken from the 2001-2002 academic year), taking account of a few district mergers over the period; and then means, standard deviations and the middle 50% are computed across districts. Within each cell, the values on top are for all districts in the state; while those below are for the Research Triangle only.

Variable	State	RT
Advanced Degree On Entry	5.8%	6.8%
Advanced Degree In Year	14.5%	16.0%
National Board Certification	9.0%	9.2%
Count	2,946	485

Table 1.3: **Teacher characteristics.** This table shows statistics for elementary teachers in the selected cohort, with each teacher-year observation given equal weight. The 'State' column includes the full cohort, while 'RT' only includes those teachers who at some point work in the Research Triangle. 'Advanced degree' refers to any degree beyond a Bachelor's.

Variable	Mean	SD	Middle 50%
Adv. Deg. on Entry	6.98	12.88	0-7.34
(%)	10.04	12.59	3.09-11.89
Adv. Deg.	15.15	15.25	6.42-19.34
(%)	21.35	15.51	13.44-21.79
NB Cert.	8.84	10.12	2.2-11.85
(%)	11.65	10.51	3.78-15.06
Years Expr.	4.86	1.28	4.75-5.14
	5.14	0.51	4.84-5.24
Time Spent	7.11	2.08	6.06-8.52
	7.71	0.55	7.4-7.94

Table 1.4: **Summary statistics for teacher attributes in North Carolina and the Research Triangle.** District means are taken over teacher-year observations from 1995 to 2011 for the 2,946 teachers in the selected cohort, taking account of a few district mergers over the period; and then means, standard deviations and the middle 50% are computed across districts. Within each cell, the values on top are for all districts in the state; while those below are for the Research Triangle only. The new variables not shown in 1.3 are: years of experience; and total time spent in the district. The overall average experience is 5 by construction, since I am counting each of the teachers once at every experience level, 0-10.

District	Supp.	Size	Growth	SAT Uptake	STS	Comps.	Loc. Funds	Par. w/o HS
Harnett	$2,315.0	520	1.71%	49.8%	1.61	28.78	$1,748.7	15.3%
Franklin	$2,700.0	267	0.06%	49.1%	2.52	25.15	$2,026.8	20.0%
Person	$3,147.0	197	0.65%	53.7%	1.98	29.63	$2,084.4	18.5%
Chatham	$3,324.0	296	1.12%	62.4%	1.29	48.10	$3,167.1	18.4%
Johnston	$3,587.0	837	4.76%	47.1%	2.17	24.10	$2,350.2	17.7%
Orange	$5,163.0	235	0.92%	72.1%	1.36	39.82	$5,330.3	15.4%
Durham	$5,366.0	1,100	1.21%	73.6%	2.39	25.69	$3,483.4	13.7%
Wake	$6,115.0	3,715	3.53%	76.2%	1.67	31.35	$3,691.6	7.1%
CH-Carr.	$6,456.0	348	2.21%	91.9%	0.39	38.27	$4,455.6	4.8%

Table 1.5: **District attributes in the Research Triangle.** These are the values of the selected variables shown in Table 1.2 for each district in the Research Triangle, again computed by averaging over 1995-2011. The table is sorted by average salary supplement, the second column.

District	AdvDeg0	AdvDeg	NB	Expr	Time Spent
Harnett	2.1%	13.4%	2.2%	4.75	8.68
Franklin	11.9%	14.6%	3.8%	4.89	7.40
Person	4.2%	18.1%	10.9%	4.84	7.21
Chatham	1.7%	6.9%	3.4%	5.00	8.33
Johnston	4.6%	17.8%	4.7%	4.92	7.62
Orange	40.2%	59.8%	27.6%	5.24	7.94
Durham	18.7%	21.8%	7.3%	4.74	7.50
Wake	4.0%	13.0%	15.1%	5.58	7.79
CH-Carr.	3.1%	26.8%	29.9%	6.30	6.93

Table 1.6: **Teacher characteristics in the Research Triangle.** The columns show mean teacher attributes at each district for the same variables found in Table 1.4 for the subset of 485 active in the Research Triangle. The table is sorted by average salary supplement from Table 1.5.

District	Mover share	Move rate	Leave rate	Leaver share
Harnett	20.0%	2.4%	4.7%	36.9%
Franklin	39.1%	5.4%	6.6%	44.0%
Person	30.0%	4.2%	6.0%	39.4%
Chatham	21.1%	2.5%	3.1%	23.8%
Johnston	31.8%	4.3%	5.0%	34.0%
Orange	42.9%	6.0%	5.1%	37.5%
Durham	28.6%	4.0%	6.1%	41.5%
Wake	55.2%	7.5%	2.2%	16.5%
CH-Carr.	85.7%	12.6%	1.1%	7.1%

Table 1.7: **Teacher flows in the Research Triangle.** This table reports several frequencies computed over teacher-years, based on the career paths of the 485 teachers belonging to the cohort and at some point working in the Research Triangle. The mover share covers teachers who at some point moved to the district from elsewhere; the move rate covers teachers newly moved to the district; the leave rate covers teachers who immediately leave the district; and the leaver share covers teachers who eventually leave the district. The table is sorted by average salary supplement from Table 1.5.

Table 1.8: **Credential growth regression results.** These results are for the 2,946 teachers in the selected cohort. The first two models predict whether a teacher will acquire National Board certification in a given year, for those already having an advanced degree (1) and those who do not (2). The last three models predict acquisition of an advanced degree, for teachers eligible for but lacking NB certification (3); those who have NB certification (4); and those ineligible for NB certification because their experience is not great enough (5).

	Dependent variable:				
	nb		advdeg		
	(1)	(2)	(3)	(4)	(5)
experience	5.281^{***}	7.065^{***}	1.379^{**}	0.073	-0.096
	(1.729)	(0.924)	(0.615)	(0.077)	(0.318)
experience^2	-0.680^{***}	-0.976^{***}	-0.207^{**}		
	(0.258)	(0.141)	(0.101)		
experience^3	0.027^{**}	0.043^{***}	0.009^{*}		
	(0.012)	(0.007)	(0.005)		
Local school funds ($1,000s)	0.0001	0.0002^{***}			
	(0.0001)	(0.0001)			
Local parents w/o HS (%)			0.026^{***}	0.039^{**}	-0.013
			(0.008)	(0.019)	(0.029)
Constant	-15.309^{***}	-20.081^{***}	-6.679^{***}	-3.957^{***}	-4.581^{***}
	(3.737)	(1.942)	(1.167)	(0.730)	(0.693)
Observations	2,885	18,448	18,448	1,494	5,605
Log Likelihood	−662.619	−2,224.608	−2,410.699	−343.023	−237.418
Note:			*p<0.1; **p<0.05; ***p<0.01		

Table 1.9: **Credential growth regression results, for teachers who worked in the Research Triangle.** These results only cover the 485 teachers from the selected cohort who work in the Research Triangle at some point. The regression formulas used here are the same as in Table 1.8.

	Dependent variable:				
	nb		advdeg		
	(1)	(2)	(3)	(4)	(5)
experience	10.783^{**}	6.817^{***}	0.445	0.176	0.755
	(4.920)	(2.310)	(1.505)	(0.208)	(0.713)
experience^2	-1.535^{**}	-0.921^{***}	-0.036		
	(0.733)	(0.350)	(0.248)		
experience^3	0.068^{*}	0.040^{**}	0.0001		
	(0.035)	(0.017)	(0.013)		
Local school funds ($1,000s)	0.0003^{**}	0.0005^{***}			
	(0.0002)	(0.0001)			
Local parents w/o HS (%)			0.028	0.067	0.064
			(0.021)	(0.050)	(0.066)
Constant	-27.330^{**}	-20.795^{***}	-5.090^{*}	-5.102^{***}	-6.793^{***}
	(10.673)	(4.909)	(2.851)	(1.915)	(1.691)
Observations	537	2,966	2,966	257	913
Log Likelihood	-103.608	-368.806	-404.763	-53.115	-49.531
Note:					*p<0.1; **p<0.05; ***p<0.01

Chapter 2

The Distribution of Teacher Quality across Heterogeneous School Districts: Structural Evidence from North Carolina

2.1 Introduction

Providing students with equitable access to good teachers is a common policy goal. To what extent do labor market patterns lead to sorting of good teachers into some districts, limiting access for others?

In chapter 1, I summarized differences in teacher attributes and mobility across the nine public school districts in North Carolina's Research Triangle. The descriptive analysis there can only go so far. For example, the districts of Chatham and Harnett Counties had the highest retention rates, but we do not know if this is because teachers found these districts appealing; or because the cost of moving to more appealing districts was too great; or because their teachers did not meet other districts' highly-selective hiring criteria. The policy implications may be very different in each of these cases.

In this paper, 'teacher quality' refers to whatever teacher attributes are rewarded in the labor market. If teachers with some attribute receive stochastically higher lifetime payoffs than those without it, we can say that the former are higher quality in this sense. Such interpersonal payoff comparisons rely on a model where teachers have common preferences up to some idiosyncratic shocks. I also impose common preferences among districts, so that higher teacher quality is always valued.

These modeling choices force a degree of sorting, but leave open the possibility that it is washed out by idiosyncratic preference shocks and matching frictions. This turns out to be the case for the market studied here, a cohort of elementary teachers, described in chapter 1, examining how they sort across districts in the Research Triangle metropolitan area in North Carolina.

In this paper, I separate the preferences of teachers and the hiring behavior of school districts to estimate a Markov decision process (Rust, 1994) for teachers' employer-choice decisions. Districts make once-and-for-all decisions about their pay and selectivity; while teachers make constrained choices about where to work each year. I allow for an unobserved constant component of teacher quality, following Kennan and Walker (2011); and for teachers' unobserved choice sets (Manski, 1977).

The next section introduces the empirical model; section 2.3 describes the estimation procedure; section 2.4 explains the results; and section 2.5 concludes.

2.2 Empirical model of teacher job dynamics

I model how teachers match to school districts over the years of their careers. The timing is as follows. Districts set a menu of pay supplements and a hiring policy. Then, elementary-school teachers choose where to work in each of their first 10 years, with district hiring policies determining the choices available to them and an exogenous process controlling the evolution of the teacher's attributes. The core of the model is the teacher's job-choice problem, while the other features of the labor market are simplified.

District hiring rules interact with each teacher independently. We can think of the district telling its several hiring managers to apply a general level of selectivity when deciding to make a job available to a teacher. The hiring managers draw a match-quality shock for each teacher and offer a job if the total value is high enough, ignoring dynamic factors like the number of vacancies available. A larger district has more such hiring managers and therefore offers teachers more chances to move to it.

Districts see all teachers attributes, including some that may not be observed by the econometrician, captured in a scalar, τ, that is constant for each teacher over time and assumed to be standard normally distributed across teachers. Their hiring rules only depend on teacher attributes and not on their job histories.

Job histories similarly do not matter for teachers, only their current location, which factors into the available job choices and their movement costs. Teachers anticipate district pay supplements and stochastic hiring behavior when solving their optimization problems. The calendar year is not a part of the model, so state salaries, district pay supplements and district hiring rules are the same for every teacher in ever year.

Teachers make job choices each period, while credential growth evolves between periods. Although I am not estimating the job choice and credential growth processes jointly as the solution to a single optimization problem, we can think about teachers making credential growth decisions between periods based on district resources and where they stand in their careers.

All data mentioned below was first introduced in chapter 1, which provides details on how the selected district and teacher attributes vary across North Carolina and the Research Triangle.

The next three sections describe the the teacher's state vector, dynamic choice problem and state transition rules; section 2.2.4 describes district behavior that enters teachers' payoffs and choice sets; section 2.2.5 solves the teacher's problem; and 2.2.6 provides my identification arguments.

2.2.1 The teacher's state vector and other notation

The teacher's state vector is $\boldsymbol{x} = (\tau, AdvDeg_0, AdvDeg, NB, y, \ell)$. Their current location, $\ell \in \{-1, 0, 1..., J\}$, represents the job choice they made last year, and is either -1 if they are about to enter the market; 0 for the outside option; or one of the $J = 9$ districts.

All other components of the state, $\tilde{\boldsymbol{x}}$ in $\boldsymbol{x} = (\tilde{\boldsymbol{x}}, \ell)$, are teacher attributes visible to all market participants. τ is a teacher quality component unobserved by the econometrician. $AdvDeg_0$, $AdvDeg$, NB are dummies introduced in chapter 1 that are observable to the econometrician, respectively indicating whether the teacher entered the profession with an advanced degree (beyond the minimum-required Bachelor's); has an advanced degree now; and has certification from the National Board of Professional Teaching Standards now. $y \in \{0, ..., T\}$ is how long the teacher has been in the profession, up to $T = 10$ years.

During estimation, it is useful to break down teacher attributes between those that are observable and unobservable to the econometrician, $\tilde{\boldsymbol{x}} = (\tau, \tilde{\boldsymbol{x}}_o)$.

Attributes for district j are denoted by $\boldsymbol{z}_j$, all of which were introduced in chapter 1. Subsets of these attributes are denoted separately based on their roles in the model: $\boldsymbol{z}_{uj}$ in the teacher's flow utility; and $\boldsymbol{z}_{\zeta j}$ in the district's hiring rule.

2.2.2 The teacher's dynamic choice problem and payoffs

Each period, the teacher enters with a state $\boldsymbol{x} = (\tilde{\boldsymbol{x}}, \ell)$ consisting of their attributes $\tilde{\boldsymbol{x}}$ and location, ℓ, as described in section 2.2.1. Then they draw a vector of transient payoff shocks for each district, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_J)$; and a choice set, $C \subseteq \{0, ..., J\}$, consisting of districts in which they can take a job in the current year.

They choose the best option, $j^*(\boldsymbol{x}, \boldsymbol{\varepsilon}, C)$, solving

$$V(\boldsymbol{x}, \boldsymbol{\varepsilon}, C) = u(\boldsymbol{x}) + \max_{j \in C} \left\{ -\Delta(\boldsymbol{x}, j) + \beta \sum_{\boldsymbol{x}'} p(\boldsymbol{x}'|\boldsymbol{x}, j) EV(\boldsymbol{x}') + \varepsilon_j \right\} \tag{2.1}$$

where u is a flow payoff; Δ is a movement cost; β is the discount factor; $p(\boldsymbol{x}'|\boldsymbol{x}, j)$ is an exogenous transition rule detailed in section 2.2.3; and

$$EV(\boldsymbol{x}) := \mathbf{E}_{\boldsymbol{\varepsilon}, C} V(\boldsymbol{x}, \boldsymbol{\varepsilon}, C) \tag{2.2}$$

is the continuation value of state $\boldsymbol{x}$, or the 'integrated value function.'

Flow utility is a linear index of compensation and district attributes:

$$u(\boldsymbol{x}) = \alpha \left[w(AdvDeg, NB, y) + supp_\ell(AdvDeg, NB, y) \right] + \gamma_u' \boldsymbol{z}_{u\ell} \tag{2.3}$$

Compensation comes from both a state salary w; and a pay supplement rule set by the district, $supp_\ell$, discussed further in section 2.2.4. For the state funded salary, I use the schedule in the academic year 2001-2002. Chapter 1 explains how experience and other credentials are rewarded on this schedule. Due to data limitations explained in chapter 1, I approximate district supplement schedules with the mean supplement paid to teachers in the district in the 2007-2008 academic year, $supp_\ell(AdvDeg, NB, y) \approx \overline{supp}_\ell^{2007}$.

The district attributes entering payoffs, $\boldsymbol{z}_{u\ell}$, are introduced in chapter 1: per-student rates of short-term suspensions, of seniors taking the SAT, and of internet-connected computers.

All are fixed values per district, computed by taking the average over the available years of data, 1995-2011.

For the outside option, payoffs are normalized to zero, $u(\tilde{\boldsymbol{x}}, 0) = 0$; and without loss of generality, the payoff collected before entering the market is also zero $u(\tilde{\boldsymbol{x}}, -1) = 0$. A teacher in a terminal state, $y = T$, collects discounted flow payoffs associated with the next three years until $\overline{T} = 13$.

I parameterize movement costs as a function of the number of county borders crossed between the current location ℓ and the chosen district j, $d_{\ell,j}$.

$$\Delta(x, j) = \gamma_\Delta \log(1 + d_{\ell,j})/\log(2) \tag{2.4}$$

I divide by log(2) so that the coefficient corresponds to the cost of moving to an adjacent county. The distance between a Research Triangle district and the outside option is computed using the nearest outside district. Based on the map in chapter 1, Research Triangle districts are up to three counties away from each other and all only one county away from the outside option.

The discount factor is not identified and is fixed at $\beta = 0.9$. The choice-specific payoff shocks for each district are independently Gumbel-distributed, $\varepsilon_j \sim \mathit{Gumbel}(\gamma, 1)$. The Gumbel cumulative distribution function is $F_{Gumbel}(x; a, b) = 1 - \exp(-\exp((x-a)/b))$, and γ is the Euler-Mascheroni Constant. This distribution is often used in discrete-choice models because it leads to simple formulas for the optimal choice probabilities.

The choice set, C, is drawn conditional on the teacher's state with probability $\pi(C|\boldsymbol{x})$. It always includes the outside option $j = 0$, as well as the current district, if any, $j = \ell$. Other options are drawn independently across districts based on their stochastic hiring rules, $\{H_j\}_j$, characterized in section 2.2.4:

$$\begin{aligned} \pi(C|\boldsymbol{x}) =& \mathbb{1}(0 \in C, \ell \in C \cup \{-1\}) \times \\ & \prod_{j \in \{1,..,J\} \setminus \{\ell\}} \Pr(H_j = 1|\tilde{\boldsymbol{x}})^{\mathbb{1}(j \in C)} [1 - Pr(H_j = 1|\tilde{\boldsymbol{x}})]^{\mathbb{1}(j \notin C)} \end{aligned} \tag{2.5}$$

2.2.3 Teacher initial state and transitions

A new teacher's initial state, $\boldsymbol{x} = (\tau, AdvDeg_0, AdvDeg, NB, y, \ell)$, is as follows. They enter at dummy location $\ell = -1$, with starting experience of $y = 0$; an advanced degree or not, $AdvDeg_0 \in \{0, 1\}$, observable in the data; current credentials $NB = 0$ (since National Board certification is not possible with zero years of experience) and $AdvDeg = AdvDeg_0$; and unobserved quality component, τ, drawn from the standard normal distribution.

In each year, after a teacher has chosen a new location $j \in C$, their new state is drawn according to $p(\boldsymbol{x}'|\boldsymbol{x}, j)$, which has the following features. Their new location is equal to their choice, $\ell' = j$; experience grows by one, $y' = y + 1$; unobserved quality is never changing, $\tau' = \tau$; and current credentials grow exogenously based on the teacher's current location, ℓ.

For credential growth, I use regression estimates computed in chapter 1.

2.2.4 District behavior

Districts choose a level of selectivity in hiring and a salary supplement schedule. Selectivity affects the generation of a teacher's choice set in equation (2.1); and supplements enter a teacher's payoffs in equation (2.3).

The supplement schedule determines how much extra the teacher is paid above their state-financed salary. As noted in section 2.2.2, due to insufficient data, I am using the average salary supplement paid by the district rather than the full schedule. By this measure, supplements make up on 12.4% of total pay in the selected sample of teachers.

Districts want to be selective, rather than using the supplement schedule alone for a few reasons. First, the state-funded base salary puts a high price floor on total compensation, so the district is not free to reduce the quantity of applicants to zero by setting its supplements to zero. Second, some teacher attributes that the district may care about cannot be included on the supplement schedule, captured by τ and $AdvDeg_0$ from section 2.2.1, and the district must screen to select on those dimensions. Finally, the district is constrained to hire a certain number of teachers based on supported allotments from the state, as explained in chapter 1, so a district that wants to improve its workforce while constrained to keep its size constant will need to both pay more and be more selective.

The district's chosen level of selectivity, $\zeta_j \in \mathbb{R}$, stochastically determines whether a given teacher with attributes $\tilde{\boldsymbol{x}}$ can move to the district, $H_j = 1$. The full hiring rule is

$$H_j(\tilde{\boldsymbol{x}}_{it}, \boldsymbol{\nu}_{ijt}) = 1 \text{ iff } \exists k \in \{1, ..., \overline{n}_j\} \ s.t. \ h(\tilde{\boldsymbol{x}}_i) + \nu_{ijtk} > \zeta_j$$

where $\boldsymbol{\nu}$ is a vector of transient match-specific shocks; $h(\tilde{\boldsymbol{x}})$ is an index of teacher attributes; and selectivity, ζ_j, acts as a stochastic threshold that the teacher has $\overline{n}_j$ chances to reach, where $\overline{n}_j$ is an increasing function of the district's size.

Each component of the vector of shocks, $\boldsymbol{\nu}_{ijt} \in \mathbb{R}^{\overline{n}_j}$, is associated with one of these chances at a job offer to teacher i from district j in year t. The shocks are independent and identically distributed according to the standard normal cumulative distribution function, Φ, so the probability of an offer takes the following closed form:

$$\begin{aligned} \Pr(H_j = 1 | \tilde{\boldsymbol{x}}) &= \Pr(H_j(\tilde{\boldsymbol{x}}, \boldsymbol{\nu}) = 1) \\ &= 1 - (1 - \Phi(\zeta_j - h(\tilde{\boldsymbol{x}})))^{\overline{n}_j} \end{aligned} \tag{2.6}$$

I parameterize the other parts, capturing how all the districts' hiring rules commonly treat teacher attributes and respond to the districts' own attributes, as follows:

$$\begin{aligned} h(\tilde{\boldsymbol{x}}) &= \gamma_\tau \tau + \gamma_o'[AdvDeg_0; \ AdvDeg; \ NB; \ \min(y, 4); \ \max(0, y-4)] \\ \zeta_j &= \gamma_\zeta' \boldsymbol{z}_{\zeta j} \end{aligned}$$

If district hiring rules select for a particular teacher attribute in h, it means that the district was not able to adequately select for the attribute by putting a higher price on it in their supplement schedule $supp_j(AdvDeg, NB, y)$. This may either be because the attribute cannot be compensated directly in the supplement schedule, like if the teacher has a double major from a prestigious university; or because teachers with the desired attribute would not respond strongly enough to such a pay increase to justify it. In either case, we can view valuation of teacher attributes in h as the residual component of teacher quality that is not compensated directly on the state salary or district supplement schedules. I include all teacher attributes from $\tilde{\boldsymbol{x}}$ in h, with experience entering piecewise linearly to allow districts' marginal response to it to change, since the first few years of experience may be important, as discussed in chapter 1.

The district attributes, $\boldsymbol{z}_{\zeta j}$, are annual growth in student population and per-student rates of short-term suspensions, seniors taking the SAT, and internet-connected computers. All are fixed values per district, computed by taking the average over the available years of data, 1995-2011.

The number of chances each teacher gets from the district is set to

$$\overline{n}_j = round(n_{teachers}/250) \tag{2.7}$$

Here, the selection of one chance at a job offer for every 250 positions in the district is essentially a normalization, as this number cannot be identified separately. An increase in the number of chances at a job offer, $\overline{n}_j$, can be balanced out by raising the selectivity bar applied in each chance to rationalize the same data, so adjusting this number should simply induce a monotonic transformation of the districts' ζs.

2.2.5 Solution

Optimal teacher behavior is described by conditional choice probabilities, $\rho(\boldsymbol{x}, j), j = 0, .., J$. Fixing a choice set, C, we know that the probability of selecting $j \in C$ is $ev(x, j)/\sum_{k\in C} ev(x, k)$ where

$$ev(x, j) := \exp\left(-\Delta(x, j) + \beta \sum_{\boldsymbol{x}'} p(\boldsymbol{x}'|\boldsymbol{x}, j) EV(\boldsymbol{x}')\right)$$

is the expected continuation value of alternative j exponentiated (McFadden, 1974). Integrating over choice sets, with probabilities given in equations (2.5) and (2.6), we have

$$\rho(\boldsymbol{x}, j) = \sum_{C: j\in C} \pi(C|\boldsymbol{x}) \frac{ev(x, j)}{\sum_{k\in C} ev(x, k)}. \tag{2.8}$$

The expected value of the optimal choice is

$$m(x) = \sum_{C} \pi(C|\boldsymbol{x}) \log\left(\sum_{k\in C} ev(x, k)\right).$$

The full solution can be found by iteratively computing m, ev and $EV = u + m$ backwards from the terminal payoffs (Rust, 1994), where EV was defined in Equation (2.2).

2.2.6 Identification

The district attributes are a subset of those that enter districts' hiring rules in section 2.2.4: per-student rates of short-term suspensions, seniors taking the SAT, and internet-connected computers.

For the outside option, payoffs are normalized to zero, $u(\tilde{\boldsymbol{x}}, 0) = 0$; and without loss of generality, the payoff collected before entering the market is also zero $u(\tilde{\boldsymbol{x}}, -1) = 0$.

Several identification pitfalls are addressed through normalizations mentioned above. Common normalizations are applied to the discount factor; payoffs collected before entry; and payoffs for one choice in every choice set. In addition, in the model of district job offers, the relationship between district size and the rate of offers was fixed. Finally, the distribution of the latent scalar capturing unobserved teacher attributes, τ, was fixed. The payoff normalizations are without loss of generality, and I suspect that results would be sensitive to the relationship imposed in Equation (2.7) mostly due to my functional-form assumptions. The normal distribution for τ has some support in the education statistics literature on teacher quality. That leaves only the discount factor that may merit further exploration.

One other identification question is how hiring rules and teacher preferences over districts can be identified separately. I use an exclusion restriction between the two subsets of district attributes, $\boldsymbol{z}_{uj}$ and $\boldsymbol{z}_{\zeta j}$, associated with teacher payoffs from working in a district and district behavior, respectively. The restriction is based on the belief that some district attributes are relevant to the selectivity of their hiring rule but not to teacher preferences. District growth, measured by the change in the number of elementary-school teacher positions allotted by the state, satisfies this criterion.

2.3 Estimation

The goal in estimating the model is to explain the allocation of teachers across Research Triangle school districts in terms of how teacher attributes are rewarded in the labor market. The parameter vector is $\theta = (\gamma_\zeta, \gamma_o, \gamma_\tau, \gamma_u, \gamma_\Delta, \alpha)$. The first three elements are related to district hiring rules, while the latter three are related to teacher preferences.

For each teacher i and year of his or her career $t = 0, ..., T$, we see the observable part of the state x_{ito} and the decision taken j_{it}. Unobserved teacher attributes are captured by the remaining part of the state vector, τ. Following Kennan and Walker (2011), I discretize the distribution of τ into n_τ discrete points and integrate over it in estimation. As shown in Kennan (2006), the best discrete approximation to a strictly increasing distribution assigns equal weight to equally-spaced quantiles, with $\Phi(\tau_k) = (2k-1)/2K,\ k = 1, ..., K$ for the standard normal case considered here. I use $n_\tau = 5$ points of support, and so am using the median and quantiles at .1, .3, .7 and .9. In my exploration of the teacher's value function, I only look at τ values at these five points.

The log likelihood is

$$L(\theta) = \sum_i \log\left(\frac{1}{n_\tau}\sum_{k=1}^{n_\tau} L_{ik}(\theta)\right)$$

where we see that equal weight is assigned to each value of τ, per Kennan (2006).

The likelihood contribution of individual i, conditional on having unobserved attributes $\tau_i = \tau_k$ is

$$L_{ik}(\theta) = \prod_{t=0}^{T} \rho(x_{itk}, j_{it}) p(x_{i,t+1,k}|x_{itk}, j_{it}),$$

where $x_{itk} = (\tau_k, x_{ito})$ is i's state vector, with $t = y$ redundantly encoding the time since the teacher entered the market. Teacher job histories, solving a choice problem (2.1) conditional on district selectivity, are captured by the choice probabilities, ρ, given in equation (2.8). Teacher credential histories are captured by a credential growth model in chapter 1, and are captured jointly with other state-transition processes in p.

I minimize the likelihood by Nelder-Mead simplex search starting from a vector close to zero and addressing a few common numerical issues.[1]

To prevent the algorithm from stepping into territory where compensation is a bad, I estimate the log of the corresponding coefficient, α. If the true coefficient on compensation were somehow negative, my estimation procedure should bring that coefficient towards negative infinity.[2]

Parameters in the credential growth process are estimated separately in chapter 1, and serve as an input into the estimation procedure described above. For both the job choice and credential growth models, I use the same subset of 485 teachers who enter the labor market between 1996 and 2001 at some point worked in the Research Triangle.

2.4 Results

Subsection 2.4.1 discusses the model estimates, which suggest fairly weak sorting, since a teacher attribute that could be rewarded through pay is instead rewarded through districts' more generous hiring rules. This also highlights the usefulness of this approach compared to calculating wage differentials associated with worker attributes.

Section 2.4.2 checks the model in a couple ways. In subsection 2.4.2, we see that while the estimates' standard errors are large, the model does fit the data well; and that there is empirical justification for preferring its constrained-choice features over an analogous model without them.

2.4.1 Estimates

The coefficient estimates are reported in Table 2.1 along with the log likelihood and standard errors from 100 bootstrapped samples of the same size.

As argued in section 2.2.4, districts can bid for teacher attributes directly through pay, so what we see in the hiring rule is the residual part of teacher quality that districts only entice through job offers. National Board certification seems to be the largest factor here, equal in impact to a four-standard-deviations boost in the unobserved part of teacher quality. This tells us that incremental pay for being certified does not capture its full valuation in the labor market; and it suggests that certified teachers' job choices are too insensitive to direct pay enticements to be swayed by what districts are willing to pay.

[1] Nelder-Mead (NM) does not always terminate at the minimum, so I restart the algorithm until no further improvements are made. NM expects each dimension of its search space to have roughly the same effect on the objective function, so I 'scale' district attributes (demean them and divide by their standard deviations); pass compensation in units of $10 million dollars; and tell the algorithm that the scale of the movement cost should be some multiple greater than the other coefficients.

[2] The coefficient on unobserved teacher attributes γ_τ, is allowed to be positive or negative. The likelihood function is the same for any pair of values $\gamma_\tau = c$ and $\gamma_\tau = -c$, and one or the other simply changes the meaning of τ (between having a good or bad effect on hiring).

Unmeasured attributes, τ, explain a lot of the variation in new teachers' quality residual, as illustrated in Figure 2.3. Among more experienced teachers, τ's share falls as it is swamped by National Board certification (available only after the fourth year) and, to a lesser extent, possession of advanced degrees.

Among district attributes in the hiring rules, district growth has the expected effect of leading to a lower threshold. Two other district attributes – suspensions and SAT uptake – have matching signs in the hiring rules and teacher preferences, which makes sense as districts facing high demand can be more selective. The third district attribute in common, computers, changes sign between the two indices, with computer-rich districts hiring more selectively even though teachers like them less.

Before delving into what else the estimates say, the next subsection investigates their soundness.

2.4.2 Checking model suitability

Subsection 2.4.2 shows that the model fits the data well, and where it misses moments, there is reason to believe it is harmless for the questions of interest. In subsection 2.4.2, I report results for a simpler model that fits the data poorly.

Fit

To check the validity of my estimation procedure, I followed the approach in Kennan and Walker (2011): for each teacher, I took his or her initial credentials, $AdvDeg_0$; drew a random effect, τ, for 100 replicas of the teacher; and simulated the behavior of each replica forward for ten years using my estimates.

As in Kennan and Walker (2011), I ran the estimation procedure on the simulated data to recover an estimate of its generating parameter vector, θ_0^{sim}. I tested whether the simulated data was consistent with my original estimate, $\hat{\theta}_0$. It passed the test, suggesting that the estimation and simulation procedures do work.[3]

These simulated data also highlight the nature of the model's fit. Table 2.2 shows some moments of the true and simulated data. Simulated 'market shares' of the nine districts are usually off by less than two percent (either as a raw difference or as a fraction of the true value) in the simulations, with the error being larger for initial and terminal shares, particularly for the outside option, $j = 0$, as highlighted in the table.

The pattern of errors is driven by my sample-selection rule that teachers must work in the Research Triangle at some point. The model underpredicts teachers from outside the Triangle moving in, and balances this by overpredicting the movement rate for teachers already in the Triangle, as illustrated in Figure 2.1. The goal in this paper is to understand differences in hiring practices and appeal among districts in the sample, so this mismatch with respect to the outside option is probably innocuous.

---

[3] The test is $H_0 : \theta_0^{sim} = \hat{\theta}$ against $H_1 : \theta_0^{sim} \in \mathbb{R}^{d_\theta}$. The dimension of θ is $d_\theta = 15$ and χ^2_{15} has a critical value of 22.30713 when testing at a significance level of 0.1. The test statistic, two times the ratio of the log likelihoods, is $2 \times (248910.89 - 248901.78) = 18.22$, so the test is passed.

District selectivity

This section argues that the choice constraints added to the teachers' Markov decision process (2.1) to account for district hiring behavior are necessary to fit the data well. We might believe that it is unnecessary overhead if (i) the teachers' estimated behavior is little different when their constraints are removed; or (ii) when we remove the constraints and estimate the simpler 'baseline' model, the estimates are essentially the same as in the full model;

To investigate the first possibility, I considered the case where movement is unrestricted in every year. I recomputed the value function and conditional choice probabilities removing the choice-set restriction and holding preferences fixed (at their estimated values in Table 2.1). As in the verification exercise discussed in subsection 2.4.2, I simulated the behavior of 100 replicas of each teacher.

These simulated teachers making unconstrained choices spent 94% of their careers in Orange County or Chapel Hill. In the real data, about half of one percent of teacher-year observations are in these districts, a moment that simulations from the full model (described in subsection 2.4.2) matches well.[4]

To investigate the second possibility, I estimated the model with all choices driven by teacher preferences with no choice constraints. The resulting estimates are compared against those of the full model in Table 2.3. They are substantially different, with students' suspensions appearing to be a 'good,' and their taking the SAT as a 'bad.'

Together, these two exercises give strong grounds for rejecting a baseline model consisting of a Markov decision process in which choices are unconstrained.

2.4.3 Teacher payoffs

Value of a larger choice set

Taking the value function in later periods as given, we can ask how much one-time access to the full choice set would be worth. For a typical teacher in most years, it's worth around \$1000 and the odds of moving are increased by about 1%. For a new teacher, a full choice set is worth around \$72,000. The new teacher's gain is greater because he or she must enter the market; by convention, they face no movement costs for any option, including the outside option.[5]

We can also look at how payoffs would differ in a baseline model, where movement is unrestricted in every year, as in the exercise described in section 2.4.2. The average payoff in this simulation is \$1.7m higher for new teachers, coming from higher flow payoffs and less need for repeated moves.

[4] The baseline model drives teachers to these two districts in particular because the movement cost between them is zero according to the parameterization in section 2.2.2, since the two districts are within the same county borders. This allows the simulated teachers to move costlessly back and forth between these two districts in response to payoff shocks.

[5] For these comparisons, I looked at states starting in Franklin county for a teacher with a median value for their unobserved attributes, no National Board certification, over all years and with or without an advanced degree.

Teacher payoff breakdown

Let $\overline{V}(\boldsymbol{x}, j) = \beta \sum_{\boldsymbol{x}'} p(\boldsymbol{x}'|\boldsymbol{x}, j) EV(\boldsymbol{x}')$ be the continuation value of choice j in the teacher's choice problem seen in equation (2.1), where $EV(x) = u(x) + m(x)$ is the integrated value function. We can break down the value of the choice problem as follows:[6]

$$\begin{aligned} m(x) &= \mathbf{E}_{\varepsilon,C} \max_{j\in C}\{-\Delta(x,j) + \overline{V}(\boldsymbol{x},j) + \varepsilon_j\} \\ &= \sum_j \rho(x,j)\left(-\Delta(x,j) + \overline{V}(\boldsymbol{x},j)\right) + \sum_j \rho(x,j)\mathbf{E}_{\varepsilon,C}[\varepsilon_j | j \text{ is chosen}] \\ &\equiv -E\Delta(x) + \overline{EV}(x) + E\varepsilon(x) \end{aligned}$$

where ρ are the conditional choice probabilities solving m; Δ are movement costs; ε are transient shocks to the payoff of each choice; and the three terms on the final line are just some notational shorthand for corresponding items on the line above.

The final term, $E\varepsilon(x)$, is the expected benefit from payoff draws; it can be calculated from the other terms, which are all known. The value of this term is greater when the choice set is larger and when the values of the choices are closer together and less correlated. (They are entirely independent in the current specification.) When one available option has a much higher expected value than the others, the benefit from drawing the payoff shocks is much smaller.

We can also break down the value, m, and continuation value, $\overline{EV}$, into component parts – future amenities, pay and movement costs – keeping track of them during value-function iteration. Consider an example from section 2.4.3: a teacher is given one-time access to all districts, for a net (expected discounted) benefit of $m' - m = \$1000$. Current movement costs increase by $E\Delta' - E\Delta = \$4000$; future movement costs fall by $m'_\Delta - m_\Delta = \$400$; amenities and salary improve by $m'_{w+supp+z_u} - m_{w+supp+z_u} = \$600$; and the value of taking more payoff-shock draws is $E\varepsilon' - E\varepsilon = \$4000$.

The large role of movement costs and payoff shocks relative to observable attributes in explaining behavior is also found in Kennan and Walker (2011). They regard ε as movement-cost shocks, so that $-\Delta(x, j)$ is the expected cost of each *potential* move, while $E\Delta(x) - E\varepsilon(x)$ is the expected cost of the *best* move. In this sense, potential costs can be very large (over \$300,000 in their paper and this one), but realized costs will be much smaller, and possibly negative.

Movement costs and payoff shocks, which add friction and noise to the model, would play a smaller role if sorting on observables were stronger, with teachers consistently moving away from some districts and toward others. In estimation, because many moves go in each direction, the payoff shocks must be large; and, because the payoff shocks are large, the movement cost must also be, to maintain a low rate of movement.

Rewards for teacher qualifications

Teacher quality is rewarded through better pay and more job opportunities. Rewards from pay are transparent because the salary schedule for all counterfactual matches is public; and

[6] j is chosen if $j \in C$ & $\forall k \in C[-\Delta(x,j) + \overline{V}(\boldsymbol{x},j) + \varepsilon_j \geq -\Delta(x,k) + \overline{V}(\boldsymbol{x},k) + \varepsilon_k]$ (which generically holds for only a single $j \in C$, so tie-breaking does not matter for this expectation).

the value of having more choices was discussed in section 2.4.3. How do the rewards break down for each teacher attribute?

A teacher moved from the 10th to the 90th percentile of unobserved quality gains only about $3000 in lifetime discounted utility, less than a month's pay. Again, there is a large expected benefit from the extra draws, mostly drowned out by a large increase in the expected movement cost. They both rise by around $20,000, as the probability of moving increases by 5-10% (as a percentage of a 10th-percentile movement rate).

A new teacher with an advanced degree has about $33,000 more lifetime utility than one without. This is a little more then their first-year state-paid salary. $16,000 of the difference is from higher state pay; $2,000 from district amenities and supplements; $27,000 from payoff shocks; and -$12,500 is lost due to higher movement costs.

For unmeasured teacher quality, which is not rewarded on the salary schedule, the contribution of local factors is similarly small (5% of the $11,000 reward for a new teacher moving from its 30th to its 70th percentile). The remainder of the reward in each case (43% and 95%) is granted through more opportunities to move (among districts in a larger choice set). These results suggest that, while heterogeneity across districts does influence where teachers work, the sorting of teacher quality is relatively weak.

2.4.4 Sorting

How differently situated are the nine districts in terms of how they participate in the labor market? Figure 2.4 shows districts' appeal and job hiring thresholds based on the estimates in Table 2.1.

Consider the elasticity of demand for positions in the most attractive district, Chapel Hill. Its appeal is 2.5 standard deviations above the mean level and 2.2 standard deviations above its nearest competitor. If Chapel Hill dropped its hiring threshold entirely for new teachers, only 44% of them would choose to start there. This finding lines up with that of Mansfield (2012) that differences in teacher productivity across schools in North Carolina are not very great.

If we want to convince a teacher starting with an advanced degree to behave like one starting without, the most we would have to pay is a lump-sum transfer of $17,000 (= $33,000-$16,000 = total surplus - surplus earned independent of location). This is not infeasibly large.

These preliminary analyses suggest that (i) teacher preferences over this sample of heterogeneous districts are not strong enough to create much inequity; and (ii) if, nonetheless, policymakers wanted to reallocate teachers, the cost is not infeasible. North Carolina ran an incentive-pay program to retain specialist teachers in high-need schools for three school years from 2001-2004. Clotfelter et al. (2008) find that the low-cost program was, while it lasted, effective in changing mobility patterns.[7]

[7] This program was for qualified teachers of math, science or special education at the middle- or high-school level, so the elementary classroom teachers in my sample were not eligible. It paid only $1800 per year spent in a high-needs school, and reduced mobility among the targeted teachers by 17% (Clotfelter et al., 2008).

2.5 Conclusion

This paper provides a different perspective on the allocation of teacher quality across heterogeneous communities. Using job histories and assuming common preferences for teachers, I find that differences in their flow utility in different districts in the Research Triangle are large enough that high-quality teachers are rewarded by more appealing jobs, but not so large that they are self-segregated into the most attractive districts. Kennan and Walker (2011) similarly found that, while workers' migration among US states responds to differences in mean utility, the relationship is relatively weak. My findings are also consistent with Mansfield's (2012) that students' access to productive teachers (i.e., teachers who have a large positive effect on their students' test scores) does not differ much across schools.

I allowed for teachers facing choice-set restrictions that I do not observe, derived from districts' hiring rules. This proved useful in this market, where dropping the choice-set restriction leads to very different results. Allowing firms hiring rules to respond to persistent unmeasured teacher attributes also proves beneficial, improving the explanatory power of the teacher quality residual.

The teacher quality residual and district thresholds allow for rich small-scale counterfactuals that highlight the rewards to teacher attributes. These, in turn, reveal the extent to which sorting is important in the market.

While I focus on how teachers choose where to work in this paper, more might be learned about a teacher's quality by looking at his or her history of applications for National Board certification, which is tracked by the Educational Testing Service. My estimates suggest that, while the cost of applying is high, at over $2000, the labor-market rewards to successful certification are around ten times as large. The application fee is often partly subsidized by the district, so the decision to apply for certification is economically important to both sides of the market, and may reveal more about the allocation of teacher quality across districts. My model assumes perfect information and exogenously evolving human capital. Contrary to these assumptions, teachers may be acquiring certification to (coarsely) disclose their quality (Dranove and Jin, 2010); or to augment it. It is likely a combination of the two, as roughly half of first-time applicants fail (Goldhaber and Anthony, 2007); and the application process takes over 200 hours.[8]

[8] Using my estimated value function, I find that, for a teacher of median unobserved ability without an advanced degree in Wake county their fifth year, $15,000 is the expected discounted reward in pay; and $8000 is the nonpecuniary reward. The pay reward is 'expected' because the teacher may acquire an advanced degree in the future, and is more likely to do so when they have been National-Board certified.

On-the-job acquisition of advanced degrees would also be worth studying. Unlike National Board certification, few teachers who attempt a master's degree fail, I suspect. This lack of screening may be behind the much larger reward to certification (about 10 times as large, as seen in Table 2.1).

Figures

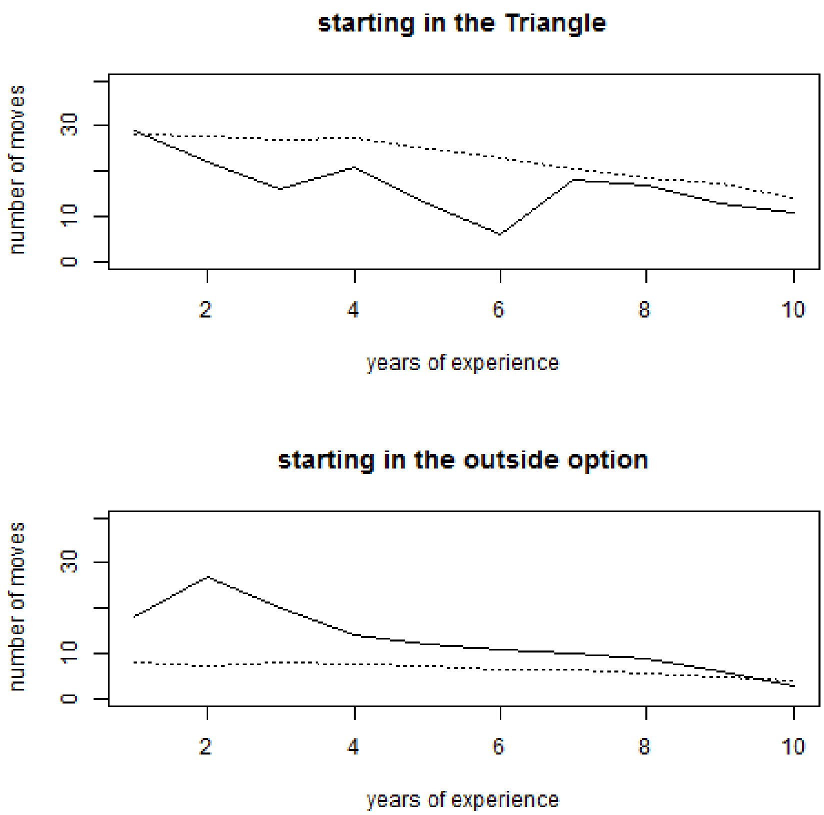


Figure 2.1: **The timing of moves, based on origin, in real and simulated data.** The solid lines represent the true rates of departure; while the dotted lines are the model's predictions (approximated by simulation). The panels are split according to origin; so, for example, moves tallied in the top panel could be destined for elsewhere within the Triangle.

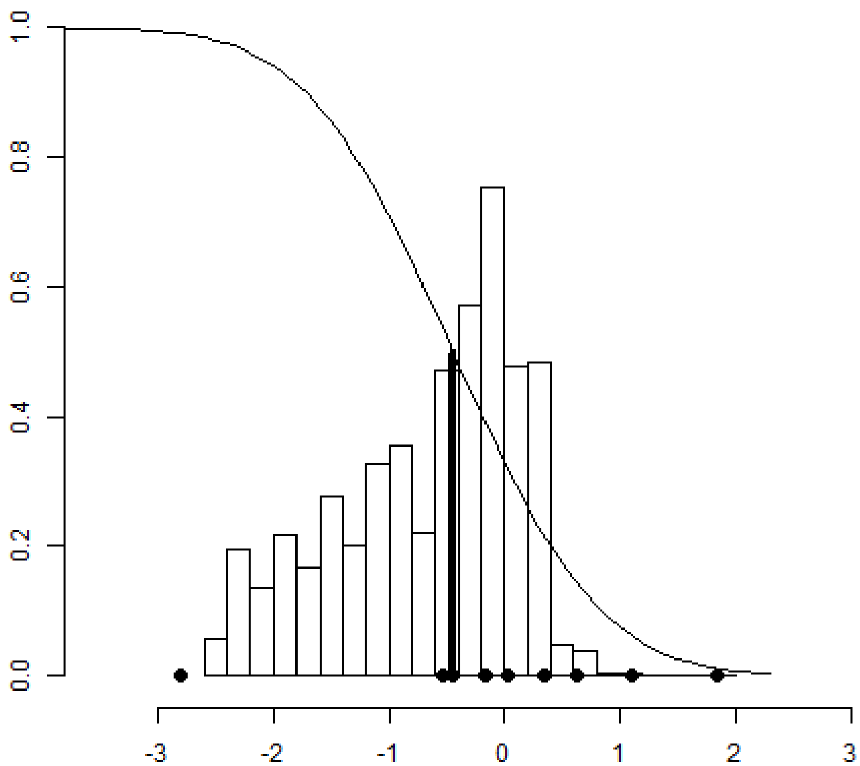


Figure 2.2: **The distribution of estimated teacher residuals, along with district thresholds.** The vertical bar is at the median teacher residual, and the curve represents the probability that the median-quality teacher would meet a threshold located at any point on the x axis.

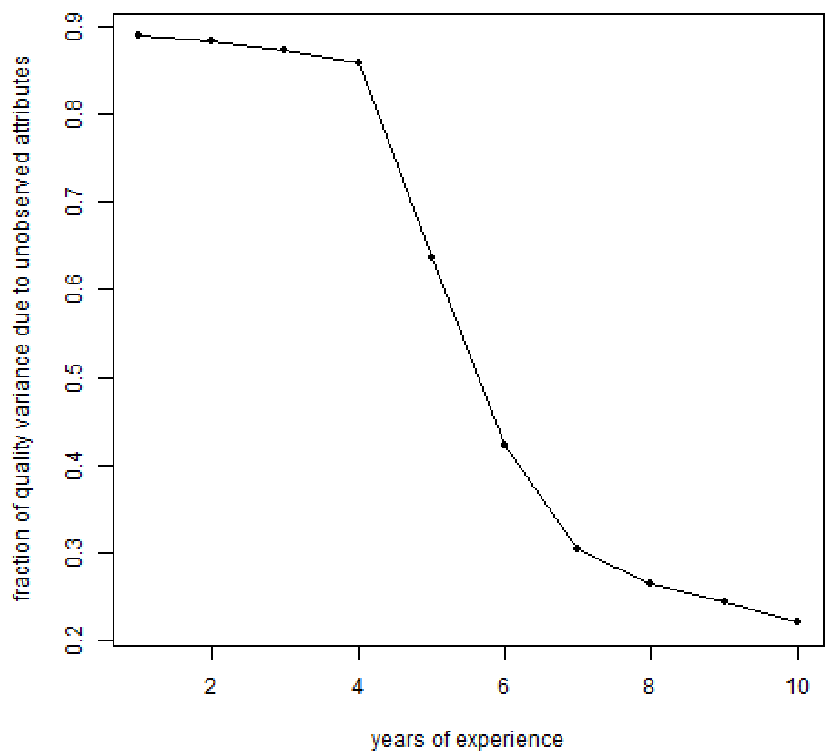


Figure 2.3: **Share of teacher residual attributed to unobserved attributes by years of experience.** The two orthogonal variance components are computed using the data and the coefficient estimates in Table 2.1.

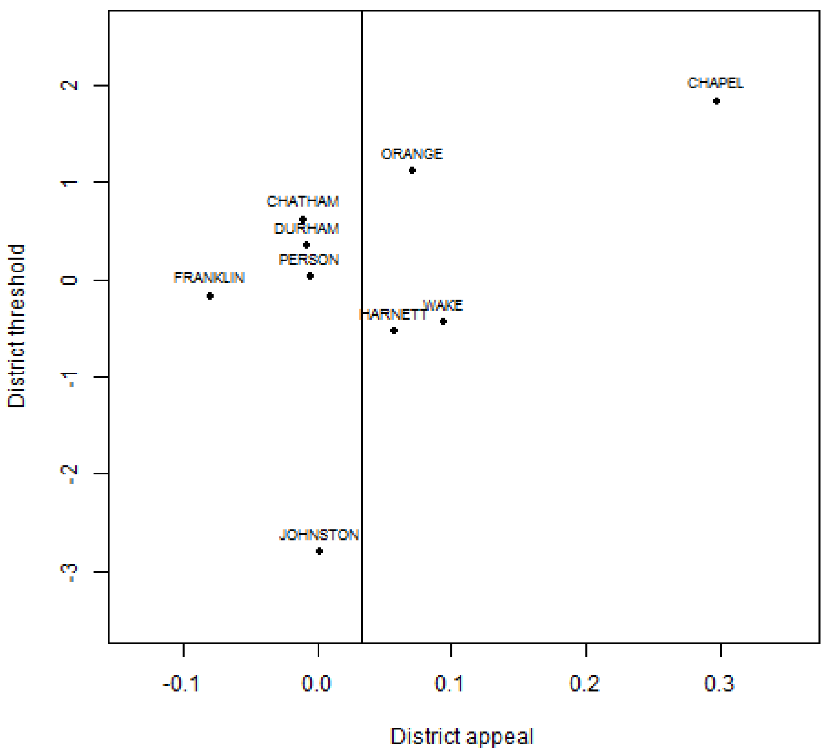


Figure 2.4: **Estimated district characteristics.** 'Appeal' is a linear combination of the district's average supplement and other attributes, listed in Table 2.1. The district's 'threshold' is a linear combination of the same district attributes and one or more that are excluded from 'appeal,' also listed in Table 2.1.

Tables

Variable	Estimate	Std. Err.
Hiring rules		
Teacher attributes		
Unobserved quality	0.2752	0.1981
Advanced degree on entry	0.3071	0.4478
Advanced degree	0.1188	0.5674
National Board certification	1.1502	0.6024
Experience for first 4 years	-0.0301	0.1194
Experience for later years	-0.3485	0.1020
District attributes		
Growth in allotments	-0.8743	0.4422
Short-term suspensions	-0.4002	0.2533
H.S. seniors taking SAT	0.6969	0.2503
Internet-connected computers	0.0213	0.2026
Teacher preferences		
District appeal		
Short-term suspensions	-0.1230	0.0496
H.S. seniors taking SAT	0.0151	0.0213
Internet-connected computers	-0.0637	0.0520
Money		
Salary/$100k	1.0926	0.2604
Movement costs		
log(1 + distance)/log(2)	3.9200	0.1464
n = 485	Log Likelihood	= -2516.881

Table 2.1: **Estimated parameters.** District attributes have been scaled so that their coefficients represent a one-standard deviation change.

		Real	Simulated
% of observations	with moves	6.5	6.3
% of teachers	never moving	49.9	53.3
	moving 2+ times	6.1	6.1
	taking j=0 in y=0	24.7	17.2
	taking j=0 in y=10	13.8	16.7

Table 2.2: **Moments of the real and simulated data.** The simulated moments are from 100 replicas of each teacher, simulated forward according to the model and estimated parameters reported in Table 2.1.

Variable	Baseline	Full
District appeal		
Short-term suspensions	0.0860	-0.1126
H.S. seniors taking SAT	-0.0442	0.0138
Internet-connected computers	-0.1214	-0.0583
Movement costs		
log(1 + distance)/log(2)	3.1378	3.5878

Table 2.3: **Estimated parameters, with and without constrained choices.** In the 'baseline' model, choice sets are not restricted. The coefficient on money is omitted here; and the other coefficients represent valuations in hundreds of thousands of dollars. The district quality components are measured in standard deviations of their per-student values.

Chapter 3

Growth and value-added models that are robust to alternative test scales

WITH KAVEH AKRAM AND ROBERT H. MEYER

3.1 Introduction

Student test scores can be useful when evaluating how well different parts of the education system are working. Unfortunately, some straightforward analytical approaches yield results that change with the test scale. In this paper, we review methods that are robust to any test-score rescaling. We investigate the tradeoffs of adopting each method through simulations.

Most methods discussed below can be tied to an explicit student growth model of the following form:

$$T(Y_1) = S(Y_0) + \beta_2 X + \alpha D + U, \tag{3.1}$$

where Y_0 is a vector of pretest scores, one for each student; X, a matrix of student characteristics; and D, a student-treatment assignment matrix. Each column of D sums to one, so each student receives exactly one treatment (eg, a teacher, classroom or school). The error term, U, captures idiosyncratic differences in student growth and has a mean of zero conditioned on Y_0, X, and D.

The model tells us, given certain inputs, Y_0, X, D and U, how much learning a student will attain. The focus of estimation is the 'ensemble' vector of treatment effects, α, and almost always considers non-experimental settings where D is correlated with Y_0 and X. While the rest of the paper refers to the student-growth interpretation of the model, it is appropriate whenever we have an ensemble of estimators and an ordinal variable.

When D indicates which teacher a student has been assigned to, the additively-separable effect, α_j, is referred to as teacher j's value added. We can use the same sort of model to study the effectiveness of other educational units and agents, such as schools and principals. For an overview of value-added modeling, see McCaffrey et al. (2004) and other articles in the same volume.

The linear-linear special case of model (3.1) is typically used in value-added analysis:

$$Y_1 = \beta_0 + \beta_1 Y_0 + \beta_2 X + \alpha D + U,$$

This linear-linear model has the weakness of assuming that student test scores are scaled in a way that is appropriate for measuring teacher value added, α. Under the weaker assumptions of our model (3.1), each teacher's performance is still reduced to a single number, but this number need not be measured as direct differences along the test score scale that was designed to measure student performance.

Besides our work, Ballou (2009) also investigates test-scaling in the context of value-added models, noting that effect estimates change when moving between standard and robust methods.

The robust models we have investigated fall into three classes:

- those where $T(\cdot)$ is unknown and $S(\cdot)$ is linear;
- those where $S(\cdot)$ is unknown and $T(\cdot)$ is linear;
- those where $T(\cdot)$ and $S(\cdot)$ are both unknown.

These methods are designed to evaluate the performance of one or more mutually-exclusive treatments (e.g., teachers) captured in the D matrix. We therefore discuss robustness with respect to the associated coefficient estimates, $\hat{\alpha}$, below. For methods that cannot be tied to explicit growth models (discussed below under "Conditional-percentile methods"), we use different notions of robustness.

Suppose we know the transformations up to finite parameter vectors, $S(\cdot) = S(\cdot; \theta_S)$ and $T(\cdot) = T(\cdot; \theta_T)$ with $\theta_S \in \mathbb{R}^{d_S}$ and $\theta_T \in \mathbb{R}^{d_T}$. For $d_T = 0$ (so that $T(\cdot)$ is known), we can estimate θ_S jointly with our other parameters by maximum likelihood. This is also feasible when $d_T > 0$ but we know the distribution of the error term (up to some finite parameter vector $\theta_U \in \mathbb{R}^{d_U}$). Along the same lines, when $T(\cdot)$ is known, we can assume that some flexible family of functions (e.g., Box-Cox, polynomial or splines) offers a reasonable approximation for $S(\cdot)$. The weakness of these approaches is that while they are straightforward, they require strong assumptions. Nonetheless, we do study a polynomial approximation for $S(\cdot)$ in our simulation below. A polynomial approximation of $S(\cdot)$ has been used in Chetty et al. (2011).

First, in sections 3.2 through 3.4, we cover the three special cases of (3.1) identified above; next, in section 3.5, we explain how we set up simulations to compare estimators for these cases; in section 3.6, we discuss our results; and we conclude in section 3.7 by bringing together our discussion of the tradeoffs involved with these methods.

3.2 Transformations of the pretest score

Suppose that posttest scores follow the formula

$$Y_1 = S(Y_0) + \beta_2 X + \alpha D + U, \tag{3.2}$$

where (as before) Y_0 is a vector of pretest scores; X, a matrix of student characteristics; and D, a student-teacher assignment matrix. The error term, U, captures idiosyncratic differences in student growth. This is the special case of equation (3.1) known as a partially-linear model.

This is a special case of equation (3.1) where units of measurement for the pretest are meaningful but the posttest only conveys ordinal information.

3.2.1 Ordinary least squares with pretest dummies

If we could condition on $S_i = S(Y_{0i})$, we could have a consistent estimator for α by running ordinary least squares for posttest on S_i and our other regressors. Although $S(\cdot)$ is not known, this conditioning can be achieved in some special cases. If the pretest is discrete, we can define a dummy variable for each value. If $S(\cdot)$ is monotone, using these dummies, we are actually conditioning on $S(Y_0)$ without knowing the correct scale. When $S(\cdot)$ is not monotone, conditioning on these dummies is different from conditioning on $S(\cdot)$ but the estimates for α remain consistent. Since these dummies only depend on Y_0 and are invariant with respect to $S(\cdot)$, this estimator is robust with respect to a transformation of the pretest.

3.2.2 Honoré and Powell's pairwise-differencing estimator

With discrete test scores, many students may have identical pretest scores. A pair of students with same pretest score, $Y_{0i} = Y_{0j}$, will also have the same transformed scores, $S(Y_{0i}) = S(Y_{0j})$. By differencing the data for such pairs of students, we eliminate $S(\cdot)$ from our regression equation and can ignore its possible nonlinearity while still estimating our other parameters consistently. That is, we can run the regression:

$$Y_i - Y_j = \beta_2(X_i - X_j) + \alpha(D_i - D_j) + U_i - U_j \quad \text{for all } i \neq j \text{ s.t. } Y_{0i} = Y_{0j} \tag{3.3}$$

A related trick can be used if the pretest is continuous and $S(\cdot)$ is continuous. Fix a student i with test score Y_i and consider the two students scoring just below and above, with scores $\underline{Y}_i$ and $\overline{Y}_i$. Asymptotically (as the number of students grows) $\underline{Y}_i$ and $\overline{Y}_i$ both approach Y_i, and so $S(\underline{Y}_i)$ and $S(\overline{Y}_i)$ both approach $S(Y_i)$ if $S(\cdot)$ is continuous. To take advantage of this insight, we can either (i) sort students by their pretest scores and take differences for adjacent students; or (ii) take differences for all pairs of students, and minimize the weighted sum of squared errors in (3.3) where the weight of each term in the objective function is related to how different Y_{0i} is from Y_{0j} (Honoré and Powell, 2001).

3.2.3 Robinson's estimator

From equation (3.2), we also have

$$E[Y_1|Y_0] = S(Y_0) + \beta_2 E[X|Y_0] + \alpha E[D|Y_0] \tag{3.4}$$

Taking the difference between (3.2) and (3.4), we have

$$Y_1 - E[Y_1|Y_0] = \beta_2\left(X - E[X|Y_0]\right) + \alpha\left(D - E[D|Y_0]\right) + U. \tag{3.5}$$

Robinson (1988) has shown that we can consistently estimate β_2 and α by substituting nonparametric estimates, $\widehat{E[Y_1|Y_0]}$, $\widehat{E[X|Y_0]}$, and $\widehat{E[D|Y_0]}$, into our regression equation, (3.5). To do so, first the conditional expectation functions are estimated nonparametrically using kernel methods, then β_2 and α are estimated by running a simple OLS on the transformed data, $(Y_1 - \widehat{E[Y_1|Y_0]})$, $(X - \widehat{E[X|Y_0]})$, and $(D - \widehat{E[D|Y_0]})$.

3.3 Transformations of the posttest score

Suppose that posttest scores follow the formula

$$T(Y_1) = \beta_0 + \beta_1 Y_0 + \beta_2 X + \alpha D + U, \tag{3.6}$$

where Y_0 is a vector of pretest scores; X, a matrix of student characteristics; and D, a student-teacher assignment matrix. The error term, U, captures idiosyncratic differences in student growth. This is a special case of equation (3.1).

3.3.1 Ordered models

By sacrificing much of the information contained in the posttest score, we can achieve consistent, albeit imprecise, estimates. In particular, suppose we divide our posttest scores into ordered categories (with pre-determined shares of the data). Then any subsequent analysis using our new ordered variable will be completely robust to transformations because we do not use the cardinal information contained in the raw posttest scores. From this point, parametric (e.g., ordered probit or logit) and semiparametric methods can be used to estimate the parameters.

First, we must choose how many ordered categories, K, we want for the posttest and the shares of observations that should fall into each category, $s_1, ..., s_K$, with $s_1 + ... + s_K = 1$. Our categorical variable can then be defined as

$$B_i = \min k \ s.t. \ F_{Y_1}(Y_{1i}) \leq s_1 + ... + s_k,$$

where $F_{Y_1}(Y_{1i})$ is the fraction of observations smaller than Y_{1i}. This discretization of the posttest is completely robust to strictly increasing transformations.

For convenience, define empirical quantiles $q_k = \max Y_{1i} \ s.t. \ B_i = k$, $k = 1, ..., K$. The model for our new categorical variable is

$$B_i = k \text{ if}$$
$$\begin{cases} \beta_0 + \beta_1 Y_{0i} + \beta_2 X_i + \alpha D_i + U \leq T(q_1) & \text{for} \quad k = 1, \\ T(q_{k-1}) < \beta_0 + \beta_1 Y_{0i} + \beta_2 X_i + \alpha D_i + U \leq T(q_k) & \text{for} \quad 1 < k < K, \\ T(q_{K-1}) < \beta_0 + \beta_1 Y_{0i} + \beta_2 X_i + \alpha D_i + U & \text{for} \quad k = K, \end{cases}$$

If we choose a distribution for the error term, U, and impose an additional normalization, we can rewrite these probabilities in terms of the distribution we have chosen and proceed to estimate our parameters via maximum likelihood or other methods (Gurland et al., 1960).

This model is sensitive to the distribution we choose for the error term. This problem can be side-stepped using semiparametric ordered estimators which are robust with respect to the distribution of the error term.

Ordered models are robust in the sense that a given model – pinned down by its shares, $s_1, ..., s_k$, a distribution for the error term, and a normalization – will yield exactly the same estimates, $(\hat{\alpha}, \hat{\beta})$, under any strictly increasing transformation of Y_1. They achieve this robustness by dispensing with much of the information in the raw posttest score, Y_1.

3.3.2 Ichimura's estimator

Since $T(\cdot)$ is strictly increasing, it has an inverse. Applying this inverse to both sides of our structural equation (3.1), we have

$$\begin{aligned} Y_1 &= T^{-1}(\beta_0 + \beta_1 Y_0 + \beta_2 X + \alpha D + U) \\ &= G(\beta_0 + \beta_1 Y_0 + \beta_2 X + \alpha D) + \varepsilon, \end{aligned}$$

for some function $G(\cdot)$ and an error term, ε, that has a mean of zero conditional on our regressors. This is called a single-index model because $G(\cdot)$ is an unknown function of a scalar term built up from our regressors.

Using Ichimura's (1993) Semiparametric Least Squares (SLS) estimator, our parameters (α, β) can be consistently estimated up to location and scale. In general, our nonlinear transformation $T(\cdot)$ makes our new error term, ε, heteroskedastic. This will cause our estimators to be inefficient. Ichimura's weighted SLS estimator can be used to deal with this problem.[1]

Unlike the ordered estimator which remains unchanged with any increasing transformation of the posttest, the SLS estimator changes with different transformations since it does not throw away the cardinal content of the data. However, it can be proven that the SLS estimator remains consistent with different transformations of the posttest.

The SLS estimator is computationally demanding, so we do not use it in our simulations below.

3.4 Transformations of both test scores

Suppose that posttest scores follow the formula

$$T(Y_1) = S(Y_0) + \beta_2 X + \alpha D + U, \tag{3.7}$$

where (as before) Y_0 is a vector of pretest scores; X, a matrix of student characteristics; and D, a student-teacher assignment matrix, $T(\cdot)$ and $S(\cdot)$ are unknown functions. The error term, U, captures idiosyncratic differences in student growth.

As we saw above, it is possible to have consistent estimates for α by employing estimators that are developed to deal with partially linear models. We also know that by using ordered estimators, we can consistently estimate α and this consistency does not depend on $T(\cdot)$ as long as it is strictly increasing. Thus, by using an ordered estimator that has a partially linear index, we can deal with unknown transformation of both pretest and posttest and we still can estimate α consistently (up to location and scale).

[1] If we assume that U is homoskedastic, we can use the following procedure. In the first stage, the parameters are estimated consistently using SLS. Using these consistent estimates, we have a consistent estimate of the index. The heteroskedastic variance then can be estimated as the conditional expectation of the residuals squared derived from the first stage of the estimation. Using these estimated variances, a weighted SLS estimator will provide us estimates that are consistent and more efficient.

3.4.1 Ordered models with pretest dummies

As we saw above, when the pretest is discrete and coarse, we can use dummy variables to have an estimator for α which is robust with respect to function $S(\cdot)$. We also learned above that we can have a robust estimator with respect to a strictly increasing function, $T(\cdot)$, if we use an ordered model. When these two methods are combined, we have an estimator for α that is robust to both pretest and posttest scales. Thus, an ordered model with dummy variables for each pretest value is employed.

3.4.2 Ordered estimators for partially-linear models

We can apply partially-linear differencing techniques to deal with the pretest transformation. At the same time, we can address an unknown transformation of the posttest by dispensing with the cardinal information in the raw scores. Honoré and Powell (2001) studied partially linear logit models and showed that the parameters can be estimated consistently (up to location and scale). We use their approach after (i) transforming the posttest into an (ordered) binary variable indicating whether the student's posttest is above the median and (ii) assuming the additively separable error term follows a logistic distribution.

Härdle et al. (2004) developed a semiparametric estimator that is applicable to ordered models with an index that includes a nonparametric function (and, more broadly, to "generalized additive models"). Härdle et al. (2004)'s estimator, by allowing for more than two ordered categories, is less vulnerable to identification problems than Honoré and Powell (2001)'s, where we cannot identify a teacher's effectiveness if all of her students are above (or all below) the median. On the other hand, since this estimator is computationally demanding, we are not currently using it in our simulations but we will use it in future studies.

3.4.3 Conditional-percentile methods

Several methods for evaluating performance in the educational system eschew explicit models of student growth like formula (3.1):

$$T(Y_1) = S(Y_0) + \beta_2 X + \alpha D + U \qquad (3.1)$$

As a result, they have no α and the notions of robustness used above do not apply. These models start by estimating the conditional percentile of each student's posttest (conditional on his or her pretest and perhaps other characteristics). From here, the conditional-percentile estimates are summarized at the level of the teacher or other educational treatment. We treat these summaries as analogous to the α parameters used up until now.

The conditional percentiles are themselves robust, in the sense that they will not change when either or both test score is passed through a strictly increasing transformation. As we will discuss below, some ways of estimating conditional percentiles are not robust to transformations. In this section, we discuss the robustness of the estimates themselves. The summaries will generally be robust if and only if the underlying estimates are robust. In our simulations, we evaluate the summaries, since they are the closest analogue to α.

Quantile regression

Quantile regression is often used to find conditional percentile estimates (e.g., Barlevy and Neal, 2011; Betebenner, 2009). Castellano and Ho (2012) contrasted quantile regression with a mean-regression approach to estimating conditional percentiles. The first step here is to estimate conditional quantile functions,

$$Q_\tau \left[Y_1 | Y_0 = y_0, X = x\right] := \inf \left\{ y : F_{Y_1|Y_0=y_0, X=x}(y_1) \geq \tau \right\},$$

for many values of τ between zero and one. When the posttest is continuously distributed, the quantile function is the inverse of the cumulative distribution function, $Q_\tau \left[Y_1 | Y_0, X\right] = F^{-1}_{Y_1|y_0,x}(\tau)$. Quantile regression proceeds by parametrizing the quantile regression function, e.g., by assuming $Q_\tau[Y_1|y_0, x] = \gamma_{0\tau} + \gamma_{1\tau} y_0 + \gamma_{2\tau} x$.

The left panel of Figure 3.1 illustrates how quantile curves are used to calculate a student's conditional percentile. This is for the special case where pretest is our only regressor. Given a finite vector $\tau_0 = 0 < ... < \tau_K = 1$, we can estimate student i's conditional percentile as any number in $[\tau_k, \tau_{k+1}]$ where the k^{th} and $k+1^{st}$ estimated quantile curves bound (y_{1i}, y_{0i}).[2]

Figure 3.1 also illustrates why conditional percentiles calculated in this way are robust to transformations of both the pretest and the posttest. The argument is as follows. If student i's posttest satisfies

$$Q_{\tau_1} \left[Y_1 | Y_0, X\right] < Y_{1i} < Q_{\tau_2} \left[Y_1 | Y_0, X\right]$$

for some $\tau_1 < \tau_2$, then

$$Q_{\tau_1} \left[T(Y_1) | S(Y_0), X\right] < T(Y_{1i}) < Q_{\tau_2} \left[T(Y_1) | S(Y_0), X\right].$$

However, parametric estimation of conditional quantiles is not robust to pretest transformations because the regression model is generally misspecified. In other words, the same parametric specification cannot fit $Q_\tau \left[Y_1 | Y_0, X\right]$ and $Q_\tau \left[T(Y_1) | S(Y_0), X\right]$. One way to diminish the effects of misspecification is allowing the conditional quantile function to be 'flexible.' For example if the conditional quantile functions are estimated with splines, the problem of misspecification is reduced. In short, conditional quantile functions can be used to calculate the conditional distribution of the posttest on the pretest. The estimated conditional distribution is robust to strictly increasing transformations of the posttest and pretest only if the conditional quantile functions are well estimated. On the other hand, we know that the same parametric specification cannot be used when both the dependent variable and the independent variable are transformed. By using 'flexible' functions (e.g. splines) and estimating many more parameters, one can reduce the problem of misspecification.

Empirical conditional percentiles

If the pretest takes on discrete values, then the conditional percentile for each student can be calculated directly, by comparing him or her to all other students with same pretest. This can be achieved by calculating the rank of this student amongst his or her peers. Since we are conditioning on each value of the pretest, this method is robust to pretest transformation.

[2] Quantile curves may cross; and this possibility may need to be addressed in a practical implementation.

Also since the rank of the student is not changed with an increasing transformation of the posttest, this method is robust to a rescaling of the posttest.

With thin pretest data or many conditioning variables (which, in turn, lead to thin data), quantile regression is preferred. The advantage of quantile regression is that it allows us to estimate the conditional percentile for observations where the conditioning data is thin since information from students with different test scores is used.

Kernel density estimation

When we cannot rely on the discreteness of the test score, the conditional percentiles can still be calculated nonparametrically using kernel density estimation methods. Since this method and empirical conditional percentiles are nonparametric, they are not subject to misspecification, unlike quantile regression.

One disadvantage of using kernel density estimation is that it is computationally slow, which might make it inappropriate for large datasets. The other disadvantage is that, as with empirical conditional percentiles, it suffers from the 'curse of dimensionality.' If we want to condition on a history of pretests or if we want to condition on demographic variables, the density in this high-dimensional space becomes thin very fast (as additional variables are added), introducing imprecision into the estimated density.

3.5 Simulation procedure

In this section, we describe our approach to evaluating the estimators introduced in the preceding sections, our Monte-Carlo simulations and other implementation details.

To get a good idea of how each estimator performs, we need it to run on many simulated data sets. A couple of our estimators – Ichimura's and Härdle's – are too slow for this and so are not considered in this section. Note that these two estimators' computational cost would not necessarily be an obstacle in a practical application where they only need to be run once.

The estimators that we do use illustrate the breadth of approaches mentioned above: pretest dummies, pairwise differencing and Robinson's estimator for pretest transformations; ordered models to address posttest transformations; ordered models with pretest dummies and pairwise differencing for transformations on both sides; and median conditional percentiles (also known as 'Student Growth Percentiles,' eg, in Betebenner (2009)).

We illustrate our estimators' performance relative to an ordinary least-squares (OLS) regression of posttest on pretest, other covariates and teacher-assignment dummies (without an intercept). Our primary performance criterion is minimization of mean squared error (MSE) across all teacher effects.

3.5.1 Performance criteria and identification issues

In the model for our OLS regression,

$$Y_1 = \beta_1 Y_0 + \beta_2 X + \alpha D + U,$$

the teacher-assignment coefficients, $\alpha_1, ..., \alpha_N$, are measured in units of the posttest, but cannot be compared to a baseline assignment (like not having a teacher). Because we know the units but have no 'zero' for the teacher effects, we say that their scale is identified but their location is not (or that the estimates are identified 'up to location').

Similarly, with a model that protects against pretest transformations as covered in section 3.2,

$$Y_1 = S(Y_0) + \beta_2 X + \alpha D + U, \tag{3.2}$$

where $S(\cdot)$ is an unknown function, we have identification up to location because our teacher effects still enter linearly into production of the posttest. To evaluate estimators based on this partially-linear model, we should demean the true (simulated) teacher effects as well as our estimates of them.

Consider again models of the following form, which are robust against posttest transformations,

$$T(Y_1) = S(Y_0) + \beta_2 X + \alpha D + U,$$

where $T(\cdot)$ is unknown and $S(\cdot)$ is either known or unknown, as discussed in sections 3.3 and 3.4, respectively. Here is the intuition for why we can no longer identify the scale of the teacher effects. A one-unit increase the effect of student i's teacher would increase her learning by $T(Y_{1i} + 1) - T(Y_{1i})$. Although we know the expression for this increase, by assumption we do not know $T(\cdot)$. Moreover, the increase depends on student i's original score unless $T(\cdot)$ is linear. To evaluate estimators for such models, we demean and rescale them by dividing their standard deviation.

For ease of comparison, we rescale all of our estimates, even those associated with the partially-linear model from section 3.2, where the scale of the estimates is meaningful. The mean square errors reported below are computed by comparing these demeaned and rescaled estimates against the demeaned and rescaled true effects.

3.5.2 Data generation

We generate a fixed number of teachers, drawing their true effects, $\alpha_1, ..., \alpha_N$, independently from a normal distribution with mean zero and standard deviation, $\sigma_\alpha = 0.5$. Teacher j is assigned a mean student pretest based on a random standard-normal of Z for $\mu_j = \rho\alpha_j/\sigma_\alpha + \sqrt{1-\rho^2}Z$. This leads to a correlation of ρ between μ and α. We have fixed this correlation at $\rho = 0.05$.

Students are simulated as follows. We take a draw for each student, $\widetilde{Y}_{0i}$, independently from a normal distribution with mean $\mu_{j(i)}$ so that 75% of the variance in $\widetilde{Y}_0$ is within-classroom. To get our pretest scores, we convert these draws into normal-curve equivalents and round to the nearest integer between −3 and 103. This leaves up to 107 unique pretest scores. The idiosyncratic component of student growth, U_i, is drawn independently from a logistic distribution scaled to have a standard deviation of $\sigma_U = 1$. We do not include any extra regressors, X.[3]

[3] The normal-curve equivalent (NCE) conversion of a random vector proceeds as follows. Demean and rescale the vector, then apply $f(x) = 50 + 21.06x$ to each element. Please note that $\widetilde{Y}_0$ is not assigned any structural interpretation and the NCE conversion of $\widetilde{Y}_0$ is simply a convenient way of generating discrete, approximately normally-distributed pretest scores.

Having drawn the random variables, Y_0, D and U, on the right-hand side of our structural equation,

$$T(Y_1) = S(Y_0) + \alpha D + U,$$

we then calculate the posttest scores:

$$Y_1 = T^{-1}\left(S(Y_0) + \alpha D + U\right).$$

Because U is drawn from a continuous distribution, the posttest takes on a continuum of values, unlike the pretest.

We examine nine combinations for $T(\cdot)$ and $S(\cdot)$, with three functions for each transformation. To ensure that our pretest transformation does not alter the relative importance of our right-hand side variables, we make sure that $S(Y_0)$ has a standard deviation of one. That is, we pick some tranformation, $\overline{S}(\cdot)$ and define $S(y) = \overline{S}(y)/\sigma_{\overline{S}(Y_0)}$, where $\sigma_{\overline{S}(Y_0)}$ is the standard deviation of the random variable $\overline{S}(Y_0)$. The chosen transformations are listed in table 3.1.

We generate 25 students for each teacher's classroom. Some of our estimators perform better in large data sets, so we report results with both $N = 30$ and $N = 100$ teachers.

3.5.3 Implementation details

We run our simulation for the three cases simultaneously: where $S(\cdot)$ is unknown, where $T(\cdot)$ and where $S(\cdot)$ and $T(\cdot)$ are both unknown. For each simulated student, let $t_i = T(y_{1i})$ and $s_i = S(Y_{0i})$. And let $n = 25N$ be the total number of students. Estimators meant to address the three cases are treated differently in terms of the 'data' they are given.

- Those estimators that are meant to be robust only to a transformation of the pretest are given $(t_1, y_{01}, d_1), ..., (t_n, y_{0n}, d_n)$ as data.
- Those estimators that are meant to be robust only to a transformation of the posttest are given $(y_{11}, s_1, d_1), ..., (y_{1n}, s_n, d_n)$ as data.
- Those estimators that are meant to be robust to both transformations are given the raw test score data, $(y_{11}, y_{01}, d_1), ..., (y_{1n}, y_{0n}, d_n)$.

In this way, we study each estimator in an environment where it is meant to be robust. Each of the three classes of estimator are compared against OLS estimates using the same data.

We have some discretion over the details of our estimators' implementation. The rest of this section describes the choices we have made.

In estimating an ordered model with a continuous outcome variable – in this case, the posttest score – we must decide (i) how to split up the outcome variable into ordered categories and (ii) what to assume about the error term, U. The ordered estimators below split the posttest into seven categories of equal frequency and assume (correctly) that U is distributed logistically. We are exploring alternative specifications in other work.

Robinson's estimator and two of our pairwise differencing estimators use a kernel on the pretest. For Robinson's estimator, which uses estimates of the conditional expectation functions of Y_1 and D conditioned on Y_0, we use the Gaussian kernel with an optimally-smoothing bandwidth. For the weighted pairwise-differencing estimators, we follow Honoré and Powell (2001) in using the biweight kernel with a bandwidth of $h_n = 0.9\sigma_{Y_0} n^{-1/5}$ (where $n = 25N$ is the number of students per simulation).

Finally, we run regressions with third-order polynomials on the pretest. While these estimators are not robust to arbitrary pretest transformations, they may attain a reasonably good fit and therefore reach a middle ground between robust estimators that may be inefficient and OLS estimates which may suffer from misspecification bias.

3.6 Simulation results

We ran the simulation 100 times each for $N = 30$ and $N = 100$ teachers. In the subsections below, we compare our estimators in terms of mean squared error (MSE). The 'mean squared error' reported takes the average over all value added vectors, α, which change from one run of the simulation to the next.

3.6.1 Pretest-only transformations

Tables 3.2 and 3.3, corresponding to simulation runs $N = 30$ and $N = 100$ teachers, list the MSEs for those estimators that address transformations of the pretest alongside estimates from OLS, which instead assomes that pretest enters linearly. All these estimators use Y_1, Y_0, and D as data.

When OLS is correctly specified, it is the efficient estimator and it achieves the lowest mean squared error. The mean squared error of the OLS estimator increases when the model is misspecified. As seen in these tables, the OLS with third-degree polynomial performs better in these cases although it also suffers from misspecification. Some other estimators in particular OLS with pretest dummies and Robinson's estimator can perform better than OLS under misspecification. OLS on pairwise-differenced data does not perform well in these cases since it suitable for continuously distributed pretest and the pretest in our simulations is discretized. The maximum-likelihood estimator on weighted pairwise-differences data performs relatively well and is better than OLS in some cases.

3.6.2 Posttest-only transformations

The mean squared error of the OLS estimator and the ordered logit estimator can be seen in tables 3.4 through 3.9. Since we want to investigate the robustness with respect to $T(\cdot)$, these estimators know $S(\cdot)$ and use Y_1, $S(Y_0)$, and D as data.

In Tables 3.4 and 3.5 where $T(\cdot)$ is linear, OLS does not suffer from misspecification and is the efficient estimator. However, it is important to note that the ordered logit estimator does not have a severe efficiency loss as the MSEs are very close.

In Tables 3.6 and 3.7, $T(y) = \ln(y)$ and it can be seen that the OLS estimator is very sensitive with respect to the transformation of the left-hand side function. The ordered logit estimator, on the other hand, is completely invariant with respect to $T(\cdot)$ since it uses exactly the same data in both cases. This property of the ordered estimator can be verified by comparing the MSEs in tables 3.4 and 3.6 or in tables 3.5 and 3.7. The MSEs are identical.

In Tables 3.8 and 3.9, $T(y) = \text{asinh}(y)$ and again we observe that the OLS estimator creates very huge MSEs both due to the face that it is biased and also that the nonlinear $T(\cdot)$ implies a heteroskedastic error term.

3.6.3 Transformations of both tests

Tables 3.10 through 3.15 list the MSE's for those estimators that address the transformations of both the pretest and the posttest. These estimators are meant to be robust to transformations on both sides and are therefore given the raw data (and compared to OLS results based on the same).

Tables 3.10 and 3.11 refer to designs where the left-hand side function in correctly specified. As seen before, the OLS estimator produces larger MSEs when $S(\cdot)$ is not linear. OLS with third-degree pretest polynomial although biased can produce smaller MSEs. Ordered logit with pretest dummies and median classroom Student Growth Percentile can produce MSEs that are smaller than OLS in some designs.

Tables 3.12 through 3.15 show how MSEs can change when the left-hand side function is also misspecified.

In tables 3.12 and 3.13, the left-hand side transformation function is logarithmic and in tables 3.14 and 3.15, $T(y) = \text{asinh}(y)$.

As expected, OLS creates much larger MSEs when the left-hand side function is misspecified. OLS with third-degree polynomial does not do much better than OLS. However, other estimators generally produce smaller MSEs. In particular, the mean squared error in ordered logit with pretest dummies and median classroom Student Growth Percentile are much smaller.

The logit estimators on pairwise-differenced data do not perform very well although they are generally better than OLS when $T(\cdot)$ is misspecified. We think this inefficiency is due to the fact that by creating a binary dependent variable, we are sacrificing too much information. Instead of using binary response model with a partially linear index, one can use an ordered model with an index that contains a nonparametric function. The approach of Härdle et al. (2004) may also be a good way to address this, though, as mentioned above, its computational cost made it unsuitable for comparison in the simulations we run in this paper.

3.7 Conclusion

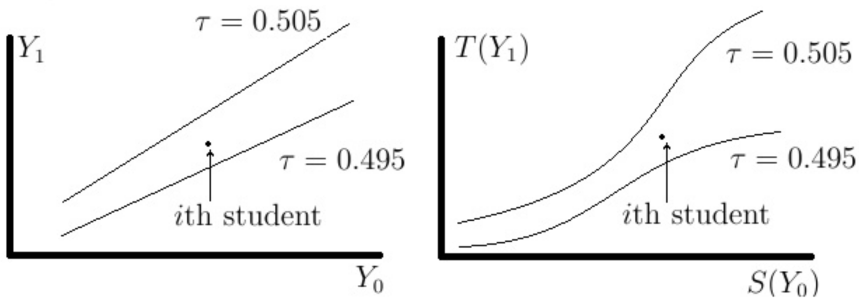
We studied the consequences of using incorrect scales on the estimates of teachers' effectiveness. In order to identify teachers' effectiveness when the correct scale is not known, we studied different estimators that are robust with respect to the scale of the pretest, the scale of the posttest, or both. Since the OLS estimator is efficient when the model is correctly specified, we also studied the cost of using each one of these estimators in terms of how much efficiency is sacrificed in order to produce consistent estimators. Our mean square error calculations suggest that the efficiency sacrificed is small for some robust estimators.

In this study we assumed that the true test score is observed. In the future we plan to study the behavior of the estimators presented in this paper in the presence of measurement error and also employ measurement error correction methods for these estimators. We also want to investigate the possibility of using empirical Bayes techniques to combine the OLS estimator and the robust estimators to minimize mean squared error. Also since the presence of the transformation function on the posttest implies that teachers have different effects on different students, we want to study how the estimators that are robust to posttest transformation tie in with the current literature on the differential effects.[4]

[4] The models, (3.6) and (3.7), with a transformation of the posttest score include 'differential effects' in a natural and parsimonious way. Consider a student with posttest score Y_{1i} assigned to a teacher with α. If this student is re-assigned to a teacher with α^*, the OLS model predicts an increase of $Y_{1i}^* - Y_{1i} = \alpha^* - \alpha$ in the posttest no matter where the student's original score, Y_{1i}, fell. In the more general models, the effect of such a change on the posttest score is allowed to be different for different students: we have $T(Y_{1i}^*) - T(Y_{1i}) = \alpha^* - \alpha$, so generally, $Y_{1i}^* - Y_{1i} \neq \alpha^* - \alpha$ for all students.

Figures

Figure 3.1: Conditional quantile functions



Tables

name	function
linear	$\overline{S}(y) = y$
inv. hyperbolic sine	$\overline{S}(y) = \text{asinh}(3y)$
shifted Φ	$\overline{S}(y) = \Phi(3y + 2)$
linear	$T(y) = y$
natural log	$T(y) = \ln(y)$
hyperbolic sine	$T(y) = \text{asinh}(y)$

Table 3.1: **Notation.** $\Phi(\cdot)$ is the standard-normal cumulative distribution function; $\text{asinh}(\cdot)$ is the inverse of the hyperbolic sine; and $\ln(\cdot)$ is the natural log.

MSEs of estimators for pretest transformations with $N = 30$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.1596	0.1690	0.2236
OLS with third-degree pretest polynomial	0.1596	0.1612	0.1776
OLS with pretest dummies	0.1807	0.1807	0.1807
Robinson's estimator	0.1625	0.1620	0.1635
OLS on pairwise-differenced data for identical pretests	0.2010	0.2010	0.2010
OLS on pairwise-differenced data for data sorted by pretest	0.5881	0.5724	0.5294
MLE on weighted pairwise-differenced data	0.1825	0.1825	0.1828

Table 3.2: These estimators are given the data $(y_{11}, y_{01}, d_1), ..., (y_{1n}, y_{0n}, d_n)$. OLS refers to ordinary least squares; and MLE refers to maximum-likelihood estimation.

MSEs of estimators for pretest transformations with $N = 100$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.1441	0.1528	0.2058
OLS with third-degree pretest polynomial	0.1445	0.1458	0.1596
OLS with pretest dummies	0.1510	0.1510	0.1510
Robinson's estimator	0.1446	0.1445	0.1449
OLS on pairwise-differenced data for identical pretests	0.1711	0.1711	0.1711
OLS on pairwise-differenced data for data sorted by pretest	0.3793	0.3715	0.3440
MLE on weighted pairwise-differenced	0.1654	0.1654	0.1654

Table 3.3: These estimators are given the data $(y_{11}, y_{01}, d_1), ..., (y_{1n}, y_{0n}, d_n)$. OLS refers to ordinary least squares; and MLE refers to maximum-likelihood estimation.

MSEs of estimators for posttest transformations, with $T(\cdot)$ linear and $N = 30$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.1596	0.1593	0.1584
Ordered logit	0.1631	0.1611	0.1606

Table 3.4: These estimators are given the data $(y_{11}, s_1, d_1), ..., (y_{1n}, s_n, d_n)$. OLS refers to ordinary least squares.

MSEs of estimators for posttest transformations, with $T(\cdot)$ linear and $N = 100$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.1441	0.1440	0.1437
Ordered logit	0.1511	0.1489	0.1487

Table 3.5: These estimators are given the data $(y_{11}, s_1, d_1), ..., (y_{1n}, s_n, d_n)$. OLS refers to ordinary least squares.

MSEs of estimators for posttest transformations, with $T(\cdot)$ natural log and $N = 30$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.7611	0.6781	0.5958
Ordered logit	0.1631	0.1611	0.1606

Table 3.6: These estimators are given the data $(y_{11}, s_1, d_1), ..., (y_{1n}, s_n, d_n)$. OLS refers to ordinary least squares.

MSEs of estimators for posttest transformations, with $T(\cdot)$ natural log and $N = 100$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.8359	0.7531	0.6489
Ordered logit	0.1511	0.1489	0.1487

Table 3.7: These estimators are given the data $(y_{11}, s_1, d_1), ..., (y_{1n}, s_n, d_n)$. OLS refers to ordinary least squares.

MSEs of estimators for posttest transformations, with $T(\cdot)$ inv. hyperb. sine and $N = 30$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.5065	0.4626	0.5759
Ordered logit	0.1631	0.1611	0.1606

Table 3.8: These estimators are given the data $(y_{11}, s_1, d_1), ..., (y_{1n}, s_n, d_n)$. OLS refers to ordinary least squares.

MSEs of estimators for posttest transformations, with $T(\cdot)$ inv. hyperb. sine and $N = 100$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.5755	0.5060	0.6291
Ordered logit	0.1511	0.1489	0.1487

Table 3.9: These estimators are given the data $(y_{11}, s_1, d_1), ..., (y_{1n}, s_n, d_n)$. OLS refers to ordinary least squares.

MSEs for robust estimators with $T(\cdot)$ linear and $N = 30$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.1596	0.1690	0.2236
OLS with third-degree pretest polynomial	0.1596	0.1612	0.1776
Ordered logit with pretest dummies	0.1855	0.1857	0.1852
Median classroom Student Growth Percentile	0.1919	0.1892	0.1907
Logit on pairwise-differenced data for identical pretests	0.3802	0.4044	0.3934
Logit on weighted pairwise-differenced data	0.3686	0.3777	0.3795

Table 3.10: These estimators are given the data $(y_{11}, y_{01}, d_1), ..., (y_{1n}, y_{0n}, d_n)$. OLS refers to ordinary least squares.

MSEs for robust estimators with $T(\cdot)$ linear and $N = 100$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.1441	0.1528	0.2058
OLS with third-degree pretest polynomial	0.1445	0.1458	0.1596
Ordered logit with pretest dummies	0.1587	0.1570	0.1567
Median classroom Student Growth Percentile	0.1804	0.1801	0.1800
Logit on pairwise-differenced data for identical pretests	0.4048	0.4039	0.3903
Logit on weighted pairwise-differenced data	0.4739	0.4894	0.4446

Table 3.11: These estimators are given the data $(y_{11}, y_{01}, d_1), ..., (y_{1n}, y_{0n}, d_n)$. OLS refers to ordinary least squares.

MSEs for robust estimators with $T(\cdot)$ being natural log and $N = 30$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.7611	0.6806	0.6042
OLS with third-degree pretest polynomial	0.7147	0.6679	0.6037
Ordered logit with pretest dummies	0.1855	0.1857	0.1852
Median classroom Student Growth Percentile	0.1915	0.1915	0.1878
Logit on pairwise-differenced data for identical pretests	0.3802	0.4044	0.3934
Logit on weighted pairwise-differenced data	0.3686	0.3777	0.3795

Table 3.12: These estimators are given the data $(y_{11}, y_{01}, d_1), ..., (y_{1n}, y_{0n}, d_n)$. OLS refers to ordinary least squares.

MSEs for robust estimators with $T(\cdot)$ being natural log and $N = 100$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.8359	0.7556	0.6551
OLS with third-degree pretest polynomial	0.8047	0.7470	0.6511
Ordered logit with pretest dummies	0.1587	0.1570	0.1567
Median classroom Student Growth Percentile	0.1802	0.1803	0.1844
Logit on pairwise-differenced data for identical pretests	0.4048	0.4039	0.3903
Logit on weighted pairwise-differenced data	0.4739	0.4894	0.4446

Table 3.13: These estimators are given the data $(y_{11}, y_{01}, d_1), ..., (y_{1n}, y_{0n}, d_n)$. OLS refers to ordinary least squares.

MSEs for robust estimators with $T(\cdot)$ being inverse hyperbolic sine and $N = 30$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.5065	0.4665	0.5855
OLS with third-degree pretest polynomial	0.5058	0.4664	0.5841
Ordered logit with pretest dummies	0.1855	0.1857	0.1852
Median classroom Student Growth Percentile	0.1887	0.1899	0.1872
Logit on pairwise-differenced data for identical pretests	0.3802	0.4044	0.3934
Logit on weighted pairwise-differenced data	0.3686	0.3777	0.3795

Table 3.14: These estimators are given the data $(y_{11}, y_{01}, d_1), ..., (y_{1n}, y_{0n}, d_n)$. OLS refers to ordinary least squares.

MSEs for robust estimators with $T(\cdot)$ being inverse hyperbolic sine and $N = 100$

Estimator \ $S(\cdot)$ is...	linear	inv. hyp. sine	shifted Φ
OLS	0.5755	0.5075	0.6361
OLS with third-degree pretest polynomial	0.5700	0.5069	0.6315
Ordered logit with pretest dummies	0.1587	0.1570	0.1567
Median classroom Student Growth Percentile	0.1807	0.1804	0.1818
Logit on pairwise-differenced data for identical pretests	0.4048	0.4039	0.3903
Logit on weighted pairwise-differenced data	0.4739	0.4894	0.4446

Table 3.15: These estimators are given the data $(y_{11}, y_{01}, d_1), ..., (y_{1n}, y_{0n}, d_n)$. OLS refers to ordinary least squares.

Bibliography

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

Ballou, D. (2009, October). Test scaling and value-added measurement. *Education Finance and Policy 4*(4), 351–383.

Barlevy, G. and D. Neal (2011, July). Pay for percentile. Working Paper 17194, National Bureau of Economic Research.

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice 28*(4), 42–51.

Boyd, D., H. Lankford, S. Loeb, M. Ronfeldt, and J. 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*(1), 88–110.

Castellano, K. E. and A. D. Ho (2012). Contrasting ols and quantile regression approaches to student "growth" percentiles. *Journal of Educational and Behavioral Statistics*.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. *National Bureau of Economic Research Working Paper Series No. 17699*, –.

Clotfelter, C., E. Glennie, H. Ladd, and J. Vigdor (2008). Would higher salaries keep teachers in high-poverty schools? evidence from a policy intervention in north carolina. *Journal of Public Economics 92*(5-6), 1352 – 1370.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2011, May). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy 6*(3), 399–438.

Dranove, D. and G. Z. Jin (2010, January). Quality disclosure and certification: Theory and practice. Working Paper 15644, National Bureau of Economic Research.

Fox, J. T. (2010). Estimating the employer switching costs and wage responses of forward-looking engineers. *Journal of Labor Economics 28*(2), pp. 357–412.

Goldhaber, D. and E. Anthony (2007, February). Can teacher quality be effectively assessed? national board certification as a signal of effective teaching. *Review of Economics and Statistics 89*(1), 134–150.

Gurland, J., I. Lee, and P. A. Dahm (1960). Polychotomous quantal response in biological assay. *Biometrics 16*(3), pp. 382–398.

Hanushek, E. A. and S. G. Rivkin (2010). The quality and distribution of teachers under the no child left behind act. *Journal of Economic Perspectives 24*(3), 133–50.

Härdle, W., S. Huet, E. Mammen, and S. Sperlich (2004). Bootstrap inference in semiparametric generalized additive models. *Econometric Theory 20*(02), 265–300.

Heckman, J. J. (1981). Heterogeneity and state dependence. In *Studies in labor markets*, pp. 91–140. University of Chicago Press.

Honoré, B. E. and J. L. Powell (2001, July). Pairwise difference estimators for nonlinear models. Working paper.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics 58*, 71 – 120.

Kennan, J. (2006). A note on discrete approximations of continuous distributions. Technical report, University of Wisconsin-Madison and NBER.

Kennan, J. and J. R. Walker (2011). The effect of expected income on individual migration decisions. *Econometrica 79*(1), 211–251.

Mansfield, R. K. (2012). Teacher quality and student inequality.

Manski, C. (1977). The structure of random utility models. *Theory and Decision 8*(3), 229–254.

McCaffrey, D. F., J. R. Lockwood, D. Koretz, T. A. Louis, and L. Hamilton (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics 29*(1), 67–101.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (Ed.), *Frontiers in econometrics*, pp. 105–142. New York: Academic Press.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica 56*(4), pp. 931–954.

Rust, J. (1994). Chapter 51 structural estimation of markov decision processes. Volume 4 of *Handbook of Econometrics*, pp. 3081 – 3143. Elsevier.

US Department of Education (2009, November). Race to the top executive summary.

US Department of Education (2016, October). 2013-2014 civil rights data collection: A first look.

Vita

Frank Erickson

Education

- Ph.D. Economics, Pennsylvania State University, February 2019
- B.A. Economics, University of Chicago, June 2007

Ph.D. Thesis

- Essays on the economics of education
  Committee Chair: Mark J. Roberts