ESSAYS ON ASYMMETRIC EMPLOYER LEARNING AND THE ECONOMICS OF EDUCATION

By

Michael David Bates

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Economics - Doctor of Philosophy

2015

ABSTRACT

ESSAYS ON ASYMMETRIC EMPLOYER LEARNING AND THE ECONOMICS OF EDUCATION

By

Michael David Bates

Chapter 1 adapts models of public and private employer learning to the market for teachers. It then use statewide, micro-level, administrative data from North Carolina to formulate value-added measures (VAMs) of teacher productivity. It exploits the adoption of VAMs of teacher performance by two of the largest school districts in the state, a shock to the available information for some, but not all, employers, to provide an initial direct test of asymmetric employer learning. Consistent with a shock to public information, for job moves within the district, this work finds that the adoption of value-added measures increases the probability that high-VAM teachers move to higher-performing schools. For moves out of the district, the impacts of policy are mitigated and even reversed by teachers with lower value-added measures becoming more likely to move to higher-performing schools. This adverse selection to plausibly less informed principals is consistent with asymmetric employer learning. Further, this chapter provides evidence that these moves lead to an increase the inequality in access to high quality teaching.

Chapter 2 examines worker mobility, and empirically tests whether all firms learn about workers' abilities at the same rate (symmetric learning) or whether current employers accumulate and use private information about their workers (asymmetric learning). The employer learning model allows for both public and private learning, and thus, nests symmetric learning as a special case. The model predicts that conditional on employees' easily observable reference groups, workers are adversely selected into job switches and layoffs on the basis

of difficult to observe characteristics, such as intellectual ability. Inversely, conditional on ability, the model predicts that as the mean ability of a worker's reference group increases, the likelihood of job separation increases. Under asymmetric private learning, these effects should become more pronounced over the length of continuous working spells. The same effects should diminish with experience, in the presence of public learning. This study uses data from the 1979 cohort of the National Longitudinal Survey of Youth to test the model. I find adverse selection on AFQT of workers who become unemployed, and conditional on AFQT score, workers with higher education from more selective institutions are are positively selected into job switches and moves from employment to unemployment during recessions. The evidence largely rejects symmetric learning in favor asymmetric learning.

Chapter 3 discusses estimation of multilevel/hierarchical linear models that include cluster-level random intercepts and random slopes. The random intercepts and slopes represent the effects of omitted cluster-level covariates that may be correlated with included covariates. The resulting correlations between random effects (intercepts and slopes) and included covariates lead to bias when using standard random-effects estimators. When applied to models with random slopes, the standard fixed-effects (FE) estimator does not rely on standard cluster-level exogeneity assumptions, but requires an "uncorrelated variance assumption" that the variances of unit-level covariates are uncorrelated with their random slopes. This work proposes a "per-cluster regression" (PC) estimator that is straightforward to implement in standard software, and shows analytically that it is unbiased for all regression coefficients under cluster-level endogeneity and violation of the uncorrelated variance assumption. The PC, RE, and an augmented FE estimator are applied to a real dataset and evaluated in a simulation study that demonstrates that the PC estimator performs well in practice.

ACKNOWLEDGMENTS

I would first like to express my sincere gratitude to Todd Elder for his guidance and support throughout the process of this project. I am also very grateful to David Arsen, Mike Conlin, Scott Imberman, and Jeffrey Wooldridge. Each has taught me an incredible amount and has given me wonderful and thoughtful advice. Thanks also goes to Soren Anderson and Michael Waldman, for their helpful comments and feedback. I also want to acknowledge Katherine Castellano, Sophia Rabe-Hesketh, and Anders Skrondal with whom the third charter is co-authored.

I received great assistance from many classmates including Andrew Bibler, Paul Burkander, Hassan Enayati, Chris Khawand, Daniel Litwok, Michelle Maxfield, Mike Naretta, Brian Stacy, Amanda Stype, Paul Thompson,Valentin Verdier, Kelly Vosters, and Greg Walsworth, and I thank them for their many contributions. I also thank Barbara Schneider, Kara Bonneau, the North Carolina Education Research Data Center, as well as representatives of Guilford County Schools, Winston Salem/Forsyth Community Schools, Charlotte-Mecklenburg Schools, and Cumberland County Schools for assisting in accessing valuable data. I also want to acknowledge that this research was supported financially by the Institute of Education Sciences Grant R305B090011 to Michigan State University.

Lastly, I wish to thank my family. Specifically, I want to thank my parents, Peter and Teri, for their continued belief in me and support throughout my education. Also, I wish to thank my siblings, Alex, Emily, and Liz, who each inspire me to get the most from life. Most importantly, I want to thank my wife, Amanda, for her unwavering love and support, and with my daughter, Iris, and son, Adam, for bringing unfathomable joy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Public and Private Learning in the Market for Teachers: Evidence from the Adoption of Value-Added Measures

1.1 Introduction

Gaps in information hinder the efficient allocation of workers across employers Spence (1973); Jovanovic (1979); Gibbons and Katz (1991a); Farber and Gibbons (1996); Altonji and Pierret (2001). While a large literature focuses on informational asymmetries between workers and employers, a more recent literature focuses on asymmetric information between current and prospective employers. Empirical work uses these models of asymmetric employer learning to explain empirical facts, such as wage dynamics with respect to job tenure versus experience, variability of wages after a job loss, and selection of mobile or promoted workers on easy or difficult to observe characteristics Schönberg (2007); Pinkston (2009); DeVaro and Waldman (2012); Kahn (2013). If the current employer enjoys an informational advantage over other prospective employers, it becomes a monoposonist of that information, permitting persistent gaps between workers' wages and their marginal products of labor. Furthermore, workers may not flow to the employers at which they would be most productive. Despite these important implications and the intuitive appeal of the theory, there is little direct evidence of asymmetric employer learning. This is in part due to the absence of direct measures of productivity, and more importantly, due to a lack of exogenous variation in the informational landscape in which employers operate.

In this paper, I adapt models of public and private employer learning to the market for middle and elementary school teachers. I then use statewide, micro-level, administrative data from North Carolina to formulate value-added measures (VAMs) of teacher productivity. VAMs calculate how much a teachers' students learn in comparison to how much those students are expected to learn. There are several methods for estimating VAMs. In econometric terms, I estimate teacher fixed effects in the regression of student test scores on student covariates including past test scores. Lastly, I exploit the adoption of VAMs of teacher performance by two of the largest school districts in the state, a shock to the available information for some, but not all, potential employers, to provide an initial direct test of asymmetric employer learning.

The adoption of VAMs in North Carolina provides a rich context for examining employer learning. Each of the two large districts that adopted VAMs did so in different ways and separately from the rest of the state. This provides three different informational landscapes: one in Guilford County Schools (to be referred to as Guilford), where the teacher, the current (or retaining) principal, and any hiring principal within the district were given direct access to the teacher's VAMs; one in Winston Salem/Forsyth Community Schools (to be referred to as Winston-Salem), in which only teachers and their current principals received value-added reports; and lastly, in the rest of the state, where the information structure remained relatively constant. These releases of statistical measures of teacher effectiveness by some, but not all employers, provide unique tests of public and private learning hypotheses.

This study examines how the relationship between of teacher quality and the probability of moving schools changes with the adoption of VAMs of teacher effectiveness. If VAMs are informative, they provide teachers with a public signal of their ability. Thus, the model predicts that VAMs increase the likelihood that effective teachers move from one school to another within the district. If the information spreads easily through the market there should be no difference between the impacts of VAMs for moves within-district and teacher transitions out of Guilford and Winston-Salem. However, if retaining principals keep teachers' VAMs private, ineffective teachers may become more likely to move out-of-district. Thus, the asymmetric employer learning model predicts adverse selection of teachers out-of-district. Lastly, I investigate whether private or public learning previously prevailed. Prior public learning implies smaller effects for more experienced teachers about whom employers already know relatively more. Prior private learning implies that the release of VAMs would even the balance of information more so for teachers with relatively more years in a given school, all else being equal.

Using differences-in-differences analysis, I find that by releasing VAMs to teachers and principals, both districts increase the probability that high-VAM teachers will move within district to a higher-performing school. I estimate that the release of VAMs increases the probability that a teacher with a one standard deviation higher VAM moves within-district to higher-performing schools by about 10%. I find that the selection of mobile teachers becomes significantly more negative for teachers moving to another school outside of Guilford and Winston-Salem after they adopt VAMs. The policy leads teachers who are a full standard deviation below average to become 15% more likely to move from Guilford to a higher-performing school in the rest of the state. In Winston-Salem, the effect of the policy on the probability that a high-VAM teacher moves to a higher-performing school is 60% smaller for teachers moving out-of-district than it is for teachers moving within-district. The fact that we see positive selection to principals with access to the information and much smaller effects and even negative selection for moves to those without access to the VAMs is consistent with asymmetric employer learning.

In the primary education context, questions of efficiency and equity are of particular importance. Previous research finds wide variation in the quality of teachers Rivkin et al. (2005); Chetty et al. (2011, 2014). Yet, at the point of hire, detecting good teachers is difficult, since easily observable teacher characteristics, such as educational attainment and college selectivity, are not highly correlated with teacher effectiveness Rivkin et al. (2005); Staiger and Rockoff (2010). Informational gaps may lead schools and districts to hire relatively ineffective teachers, while passing on more capable ones. Thus, asymmetric information can have significant ramifications for the students they serve Chetty et al. (2011, 2014).

After the date of hire, while principals typically do not observe a direct measure of a teachers' effectiveness, they can observe their teachers in action and inspect student outcomes. However, the quality of a teacher may remain difficult for the employing school to uncover, and harder still for other schools to learn. The amount of uncertainty in the market, and with whom the uncertainty lies, can differentially affect not only the initial sorting, but also the resorting of teachers across schools.

Persistent informational gaps between teachers' true effectiveness and employers' perceptions of it may lead schools to undervalue effective teachers and allow ineffective teachers to impede the progress of their pupils. In contrast, complete and public information allows better teachers more choice over where to teach. When teachers are given VAM reports, the VAMs provide them a new credible way to signal their ability.

In the teacher labor market, wages are typically set rigidly and are not tied to performance.[1] Thus, the implications of employer learning are felt primarily through teacher mobility from one school to another. There is a large body of work, which examines teacher preferences Boyd et al. (2008); Jackson (2009); Boyd et al. (2013). They find that teachers in general prefer to teach in schools that are closer in proximity to their homes, higher performing, and for white teachers, schools with a lower percentage of black students. Consequently, while providing good teachers more choice, better information may also exacerbate the divide in access to high quality education. The degree to which information stays exclusively with current principals theoretically may mitigate these effects. This work provides the first examination of whether the release of VAMs leads to further sorting of teachers to schools. Rising inequity may be an important consequence of the policy that has been previously overlooked.

The possibility of growing inequity in access to effective teaching is particularly important given the speed at which states and school districts are adopting VAMs. The entire state of North Carolina adopted teacher-level VAMs in the 2013 school year. As of May, 2014, 38 states have required teacher evaluations to incorporate teachers' impacts on student achievement on standardized exams. Even among the remaining states, many large school districts have already incorporated VAMs into evaluations of their teachers. While these policies have been controversial, the debate has previously ignored the signaling impact of VAMs on the distribution of effective teachers across schools. By examining changes in the sorting of teachers, I evaluate the impact of the information on the distribution of teacher quality across schools. The rising mobility of effective teachers to high-performing schools and the rise in the correlation between teacher VAMs and school-wide student performance

[1] There are exceptions to this. In Section ??, I discuss two policies (ABC growth and Strategic Staffing) that deviate from this standard wage rigidity. The ABC growth program provides incentives to every teacher in schools that make their growth targets. Strategic staffing policies offer incentives to teach at hard-to-staff schools.

in Winston-Salem in particular, evidences rising inequity in access to high quality education as a result of VAM adoption.

1.2 Setting

Shocks to the information available on workers' productivity are rare. Shocks to the information of some, but not all, employers in a market are rarer still. The release of teacher performance measures to principals working within the school district, but not to those in the rest of the state, offers an opportunity to examine whether plausibly valuable personnel information spreads throughout the market.

Guilford County Schools (Guilford) contracted with SAS (originally called "Statistical Analysis System") to receive teacher EVAAS (Education Value-Added Assessment System) measures of teacher effectiveness in 2000. These measures are based on the model developed by Sanders et al. (1997) under the name "Tennessee Value-Added Assessment System" (TVAAS). In fact, the adoption of VAMs by Guilford accompanied the transition of TVAAS to EVAAS, as the system came under the management of SAS, which originated at North Carolina State University. The district gave teachers, their current principals, and hiring principals within the district direct access to these teacher value-added measures (VAMs). Because all hiring principals in Guilford can directly access a teacher's VAM, the introduction of VAMs provides a shock to the available information to all principals within the district.

The rest of the state of North Carolina adopted EVAAS measures of school effectiveness in 2008, but there was no new teacher-specific information provided. Winston-Salem/Forsyth Community Schools (Winston-Salem) took an additional step, providing SAS with student-teacher matches necessary to receive the same teacher specific measure of effectiveness already present in Guilford. In Winston-Salem, only the teacher and his principal directly received the VAM reports. The VAM reports were not given directly to any other principals within the district or otherwise.

Though the Winston-Salem dispersed the VAMs in a more restricted way, the introduction of VAMs in Winston-Salem is theoretically also public. As in Grossman (1981) and Milgrom (1981), each teacher contemplating moving within the district has as incentive to voluntarily disclose his score. Because all principals in the district know that the VAM exists, if a teacher chooses not to reveal his score, hiring principals within-district may well assume that he is as good as the average teacher who chooses not to reveal his score. Thus, all teachers with scores above that average have an incentive reveal their scores. Consequently, the average score of those who do not disclose drops until only teachers with scores at the minimum are indifferent between revealing and keeping the information private. If teachers act as predicted, all teachers voluntarily disclose their EVAAS reports, and the VAMs alter the information available to both current and hiring principals within Winston-Salem, just as they do in Guilford. This shock to the public information allows teachers with higher VAMs than their resumés may otherwise suggest to signal their ability to prospective employers.

The setting and incentives teachers face differ when moving out of the Guilford and Winston-Salem districts. Perhaps most importantly, it is possible that hiring principals in the rest of the state are unaware of the existence of an applying teacher's EVAAS report. In which case, a teacher may withhold his signal and leave the principal's expectation of his ability unchanged.[2] This informational asymmetry may be avoided by principals thoroughly

[2] In which case, only those whose VAMs are higher than would otherwise be expected would choose to reveal, and only out-of-district principals hiring those teachers would be aware of their VAMs' presence. Furthermore, for teachers whose VAM is worse than would be expected by their resumés, moving out of district may be an attractive choice, leading to more negative selection of teachers moving from districts that adopt VAMs.

researching from where their applicants are coming. In which case, the same predictions as were formulated for within-district moves would apply. However, such acquisition of information is costly, and principals may forgo it. Thus, the test between symmetric and asymmetric learning hinges on whether the adoption of VAMs leads the selection of out-of-district mobile teachers to be significantly more negative than its effects on the selection of within-district movers.

Since principals in both Guilford and Winston-Salem received training about the measures, VAMs likely served as a more salient signal for principals within the adopting districts than for those in the rest of the state. Out-of-district principals may have put particularly low weight on the measures in 2000, when Guilford initially contracted with SAS. At that point, only two years after the creation of EVAAS, No Child Left Behind was still a year away from passage, and VAMs were largely absent from education policy discussions. The salience of the signal was likely less of an issue for teachers moving from Winston-Salem, considering school-level EVAAS measures were implemented across the entire state the same year. This may lead the learning results for out-of-district moves to be more pronounced for Guilford than they are for teachers leaving Winston-Salem.

To summarize the basic intuition of the model in Section 1.4, if VAMs provide meaningful information to all principals in the district, and teachers in general prefer to teach at better schools, after districts release VAMs, good teachers will be more likely to move to higher-performing schools. It is also possible that current principals become less able to keep quiet which teachers are really good, while passing off the worse teachers to unwitting employers. Table A.1 shows exactly this general pattern for moves within Guilford and Winston-Salem. In both districts, the average VAM of teachers who move within the district increases sharply after releasing VAMs. For moves out of these districts, the average VAM of moving teachers drops following the adoption of the policy. These means are not conditional on any easily observable characteristics, and so it is difficult to say whether the changes in information are driving these patterns. However, the increases of 0.259 and 0.119 standard deviations of average VAMs of movers within Guilford and Winston-Salem respectively suggests that releasing VAMs within the district allows high-VAM teachers to move more easily to other schools. The 0.290 and 0.143 drop in average VAMs of moving out of Guilford and Winston-Salem is indicative of low-VAM teachers moving to plausibly less informed principals outside of the district.

1.3 Employer Learning, VAMs, and Teacher Mobility

This is the first study directly testing a general model of public and private learning by exploiting information shocks to a large, relevant labor market. However, There is a robust extant literature building models of employer learning and fitting them to stylized empirical facts.

Farber and Gibbons (1996) provides the seminal model and test for employer learning. They assume that employers cannot directly observe the ability of potential workers and must rely on correlates to infer workers' expected value to the firm. They treat a subset of worker characteristics as easily observable to all, another as easily observable to the market (and not to researchers), and yet another subset of potential correlates with productivity as easily observable to the econometricians (but not the market). This literature typically uses the percentile from a cognitive ability assessment, the Armed Forces Qualification Test (AFQT) from the National Longitudinal Survey of Youth of 1979 (NLSY79), as this relatively strong correlate with productivity that is veiled to the the market at the time of hire, but is visible to researchers. By assuming a competitive marketplace and that employers all learn at the same rate, in the Farber and Gibbons (1996) model wages perfectly track the employers' learning process. Altonji and Pierret (2001) adopt a similar foundation in their examination of statistical discrimination as does Lange (2007) in his study of the speed at which employers learn. Each finds that the correlation between wages and AFQT score increases with experience, while the correlation between wages and easily observable characteristics falls over time.

Recent work in the economics of education presents evidence that principals also learn about teacher quality over time. While Staiger and Rockoff (2010) and Rivkin et al. (2005) point to the difficulty in identifying effective teachers at the point of hire, Jacob and Lefgren (2008) presents evidence that principals' evaluations are positively correlated with VAMs of teacher effectiveness, but not perfectly. They find that principals are better at identifying the most and least effective teacher. The fact that they observe slightly higher correlations for principals who have known their teachers for longer is further suggestive of a gradual learning process.[3] The strongest evidence of principals learning about teacher quality comes from Rockoff et al. (2012). They present experimental evidence that teacher VAMs provide significant information on which principals update their prior beliefs. It is important to note that in this experiment, only teachers' current principals receive VAM reports, not the teachers themselves or principals of other schools within the district. Surveys of participating principals show that those who randomly received more precise VAM reports were more responsive to the information, than were principals receiving noisier VAM reports.[4] These

[3]Chingos and West (2011) provide further evidence that principals hone in on the effectiveness of their teachers. They find that principals classify their teachers on the basis of effectiveness, and move them accordingly. Principals of schools under accountability pressure are more likely to move effective teachers into and less effective teachers out of high-stakes teaching assignments.

[4]Rockoff et al. (2012) also finds that providing VAMs to principals cause less effective teachers to leave at a higher rate. While the authors do not directly link these results to either learning hypothesis, these

results are consistent with the Bayesian updating model used in Farber and Gibbons (1996); Altonji and Pierret (2001), and Lange (2007).

Schönberg (2007); Pinkston (2009); Kahn (2013), and Bates (2015) each relax the symmetric learning assumption and allow for private employer learning. Also, each use the NLSY79 to test their models against empirical features of the data. Their cumulative evidence regarding asymmetric learning is mixed. Whereas, Schönberg (2007) finds that learning is largely symmetric, Pinkston (2009) finds that learning is largely asymmetric. Their disagreement hinges on whether information passes through job-to-job transitions, with Pinkston (2009) finding that the correlation between wages and ability moves more closely with respect to continuous working spells than with experience. Both Schönberg (2007) and Bates (2015) that workers are only adversely selected into mobility in job-to-unemployment transitions, whereas asymmetric learning also predicts such selection for job-to-job moves as well. However, Bates (2015) also demonstrates positive selection into mobility on the basis of education, noting that consistent with asymmetric learning, those who attend more competitive colleges are more likely to both switch employers and be laid off. Consistent with asymmetric employer learning, Kahn (2013) finds that movers' wages are more volatile in the immediate aftermath of a transition than are the wages of those who remain in place.[5]

Only DeVaro and Waldman (2012) depart from the use of the NLSY. They use administrative personnel files from a large firm to examine promotion decisions based on private and public information. In support of asymmetric employer learning, they find that conditional on private performance reviews, those with more education are more likely to be

results in the experimental context are consistent with asymmetric employer learning.

[5]Kahn (2013) also considers differences between workers who enter a position during recessions as opposed to economic expansions, with the idea that there is less variation in the ability of entrants during recessions. She also uses variation in the amount of exposure an occupation has outside the firm, assuming that learning is more symmetric in more exposed occupations. Also, the effects are larger for those who enter a job during an economic expansion and for those in more insular occupations.

promoted than are those with less education. They also present evidence that larger wage increases accompany promotions of less educated workers than accompany promotions of higher-educated workers. This, they argue, is due to the fact that promotions are a stronger public signal for those with lower, easily observable characteristics.

A common criticism of much of the earlier literature asks what AFQT scores are really telling us. There is little evidence that AFQT scores are related to productivity in many jobs held by the largely low-skilled respondents of the NLSY. Similarly, if employers care greatly about AFQT scores, they would simply administer the test themselves. By using a more direct measure of productivity than the assumed correlates, this study avoids such criticism. More importantly, the stylized empirical facts given as evidence of asymmetric learning are consistent with the theoretical model, but are susceptible to alternative explanations. For instance, post-move wage volatility may be explained by differences in job match quality, education may provide more higher level skills leading to faster promotion, and symmetric learning may explain why large wage increases accompany promotions of less-educated workers. The absence of direct asymmetric information shocks has prevented the previous literature from examining whether the informational advantages persist and influence worker mobility patterns in equilibrium. This work uses the release of worker-level performance data to some, but not all, employers as a unique natural experiment, to test the degree to which the information spreads among employers, whether mobility responds in accordance with theory, and the type of learning that had previously prevailed.

Furthermore, while there is a large literature examining the mobility patters of higher- or lower-VAM teachers, none have previously considered the signaling effects of VAMs on teacher mobility and the distribution of teacher quality within the market. Students in poor, low-achieving schools face teachers who are in general less experienced, less educated, and less effective than their counterparts in more affluent and higher achieving schools Lankford et al. (2002); Clotfelter et al. (2005); Sass et al. (2012).[6] Though their is significant churn within the teacher labor market, Hanushek et al. (2005); Krieg (2006); Goldhaber et al. (2007) and Boyd et al. (2008) each note that higher VAM teachers tend to stay in the profession longer than do their less effective counterparts, and high-VAM teachers are no more likely to transfer between schools than their counterparts.[7] There is more disagreement about distributional effects of this turnover. Boyd et al. (2008) finds that, conditional on moving, high-VAM teachers are more likely to move to high-performing schools than are low-VAM teachers, whereas Hanushek et al. (2005) and Goldhaber et al. (2007) find no evidence of this resorting of teachers. While, descriptions of where effective teachers have traditionally moved from and to have important implications for education inequity, they have little power to predict how the adoption of VAMs will alter the allocation of teachers across schools.

Work closely examining teachers' preferences over work environment offers insight into how teacher mobility patterns may change with the introduction of VAMs. Jackson (2009) and Boyd et al. (2013) analyzes teachers each find that on average white teachers prefer not to teach in schools with a large proportion of black students. Boyd et al. (2013) also find that teachers prefer schools that are closer, are suburban, and have a smaller proportion of students in poverty.

If VAMs provide new and credible information to principals, this new signal may expand the number of schools willing to hire high-VAM teachers. Taking the estimated preferences from Jackson (2009) and Boyd et al. (2013) as given, this expanded choice set may lead high-VAM teachers to move to schools that have lower proportions of minorities, are more

[6] Sass et al. (2012) also notes that there is huge variation in teacher quality within high poverty schools.

[7] Boyd et al. (2008) finds that ineffective teachers are more likely to leave the profession only in their first year of teaching.

affluent, and are higher achieving. While this earlier literature points at the possibility, it has not directly examined the question of rising inequality in the allocation of teacher quality as a result of VAM adoption. Guilford and Winston-Salem's early release of VAMs, allows this work to explore this previously ignored consequence of the actively debated policy.

1.4 Model

This section develops a model to provide predictions for which workers move, and where they go–and how each may change in response to an information shock. Please see Appendix E for proofs of these predictions. The model builds on the model of asymmetric employer learning presented in Pinkston (2009), which in turn builds upon the canonical models of symmetric learning presented in Farber and Gibbons (1996) and extended in Altonji and Pierret (2001).

1.4.1 Model Structure

Teachers receive two job offers in the first period and take the highest offer. Each subsequent period, teachers receive one outside offer from either a principal within or outside of the current district with a given probability. Principals face rigid budget constraints, which translate to a fixed number of positions. Principals with a vacancy who encounter a teacher present the teacher with an offer reflecting their expectations about the effectiveness of the teacher, which is based upon the information available. I itemize the information structure below:

1. True effectiveness is given by, $\mu = m+\epsilon$, where m is the population mean of productivity

among a worker's reference group and $\epsilon \sim N(0, \sigma_\epsilon)$.[8]

2. The public signal is given by $R_x = \mu + \xi_x$, where $\xi \sim N(0, \sigma_\xi(x))$, and $\frac{\partial \sigma_\xi(x)}{\partial x} < 0$.

3. Private signal:

   (a) For hiring principals (denoted by the superscript h), the private signal is given by $P^h = \mu + \tau^h$ where $\tau^h \sim N(0, \sigma_\mathcal{T}(0))$. $\sigma_\mathcal{T}(0)$ is fixed over time.

   (b) For a retaining principal (denoted by the superscript r), the private signal is given by $P^r_t = \mu + \tau^r_t$ where $\tau^r_t \sim N(0, \sigma_\mathcal{T}(t))$ and $\frac{\partial \sigma_\mathcal{T}(t)}{\partial t} < 0$.

4. The VAM serve as an additional piece of information that may alter both the mean and precision of the public or private signal depending on whether it is available to both bidding principals. It has the form $V = \mu + \nu$, where $\nu \sim N(0, \sigma_\nu)$.

   (a) When both principals are informed by VAMs, the public signal becomes $R_{x\nu} = \frac{\sigma_\nu R_x + \sigma_\xi(x) V}{\sigma_\nu + \sigma_\xi(x)}$. The variance of $R_{x\nu}$ is denoted as $\sigma_\xi(x\,V)$.

   (b) When only the retaining principal is informed by VAMs, her private signal becomes $P^r_{t\nu} = \frac{\sigma_\nu P^r_t + \sigma_\mathcal{T}(t) V}{\sigma_\nu + \sigma_\mathcal{T}(t)}$. The variance of $P^r_{t\nu}$ is denoted as $\sigma_\mathcal{T}(t\,V)$. The hiring principal's signal remains unchanged.

5. The noise of each signal is orthogonal to the noise of the other signals.[9]

It is important to understand the context of this labor market for teachers. In formulating the model, I will highlight areas in which this market is peculiar and the model structures that

[8] The normality assumptions are not necessary, but are useful in deriving the comparative statics.

[9] The orthogonality assumptions are also not necessary to derive the following predictions. However, relaxing these require a less restrictive, though more complicated set of assumptions, outlining the direction and magnitude of correlations between the errors of the signals.

accompany them. However, the information structure is standard, based upon a Bayesian updating model with the modification that employers receive two signals rather than one. I assume that teachers know their effectiveness (μ), but cannot credibly reveal it. There are two broad classifications of principals: those who are hiring (denoted by the superscript h); and those who are retaining teachers (denoted by the superscript r). I further distinguish between within-district hiring principals who can access the incoming VAMs, and out-of-district hiring principals, whose information does not change. As a teacher begins her career, all principals begin with the prior belief that she is as good as the average teacher with her same characteristics (m). The teacher encounters two principals, both of whom are hiring principals in this first period, to whom she may privately signal her ability akin to an interview, (denoted by P_0^h where 0 indicates no additional private information).

Over time, teachers may draw on their experience to bolster their public signal denoted by R_x (for examples consider resumés and networks of references). Any information (x) that is credibly revealed to both prospective employers produces more precise public signals. Experience serves as a proxy for additional information, as is typical in the literature. If there is public learning, generally the variance of the public signal ($\sigma_\xi(x)$) will shrink with teacher experience $\left(\frac{\partial \sigma_\xi(x)}{\partial x} < 0\right)$. However any new public information directly produces this effect.

Through interactions, observations, and/or attention to student outcomes, principals may obtain private information unavailable to rival employers (t). Retaining principals' signals (P_t^r) are composed of information that is unavailable to the other prospective employer. Years of tenure with the current employer serve as proxy for this accumulated, private information, as is typical in the literature. If such private learning occurs, while hiring principals' private signals from interviewing the teacher have a constantly high variance

($\sigma_{\mathcal{T}}(0)$), the precision of the current principal's signal ($\sigma_{\mathcal{T}}(t)$) increases the longer a teacher works in the school. With any accumulation of private information, $\sigma_{\mathcal{T}}(t) < \sigma_{\mathcal{T}}(0)$ for all $t > 0$. In order to nest symmetric learning within the more flexible model, I maintain that that even in this special case, employers receive a private signal each period, but the variance of the signal is constant over tenure ($\sigma_{\mathcal{T}}(t) = \sigma_{\mathcal{T}}(0)$ for all $t > 0$).

VAMs enter the learning model as an additional piece of information that may enter either the public or private signal. Whether VAMs influence public or private signals depends on whether VAMs are accessible to both principals (as certainly occurs for moves within the unrestricted Guilford County school district and theoretically occurs in the restricted Winston-Salem district) or are accessible to only current principals (as is more likely to occur when competing principals are from different districts). If VAMs enter retaining principals' private signal, $P^r_{t\nu} = \frac{\sigma_\nu P^r_t + \sigma_{\mathcal{T}}(t)V}{\sigma_\nu + \sigma_{\mathcal{T}}(t)}$ replaces P^r_t. If VAMs enter both principals' public signal, $R_{x\nu} = \frac{\sigma_\nu R_x + \sigma_\xi(x)V}{\sigma_\nu + \sigma_\xi(x)}$ replaces R_x. The introduction of VAMs alter these expectations by changing both the content of the signal and the signal's precision, and thus the weight that principals ascribe to it.

1.4.2 Bidding

In many public education systems, strict salary schedules determines teachers' pay. In North Carolina, the state sets a base salary schedule that depends exclusively upon easily observable characteristics, such as education and experience.[10] Districts typically supplement this base amount with a percentage of the base schedule. In general, this means that a given teacher will earn the same salary regardless of where and what he is teaching within the

[10] As of 2014, North Carolina will move to paying teachers in part based upon teachers' VAMs.

district.[11] Further, cumbersome dismissal processes result in teachers initiating much of the mobility. While principals cannot adjust salaries to influence whether a teacher stays, principals may influence school staffing through non-pecuniary position attributes, such as planning time, teaching assignments, or additional requirements. Boyd et al. (2008, 2013), and Jackson (2009) each provide evidence that teachers have strong preferences over non-wage job attributes.

Initially, teachers take the position that offers the highest total compensation (C_{isd}), which is comprised of salary (w_d) set by district d, characteristics of school s (S_{sd}), and characteristics of position i (J_{isd}). Thus, $C_{isd} = w_d + S_{sd} + J_{isd}$.

For simplicity, I assume that each principal presents a sealed bid for the teacher and pays the minimum of the two bids. In such sealed-bid, second-price auctions, principals' optimal strategy is to offer the their expectation of the teacher's effectiveness (assuming that principals seek to maximize teacher effectiveness within their schools).[12] [13] Principals formulate these expectations by averaging over their prior belief of quality (m), the public signal (R_x), and their private signal (P_0^h). They weight each signal by its precision relative to the other signals, similar to a standard Bayesian updating model. As public information becomes more complete, hiring principals give less weight to their prior beliefs and private noisy signals from interviews, and more weight to the public signal. Thus, letting $Z_{NV}^h =$

[11] In Section 1.7, I discuss both the ABC growth and strategic staffing policies, which deviate from this general case. The ABC growth program provides incentives to every teacher in schools that make their growth targets. Strategic staffing policies offer incentives to teach at hard-to-staff schools. The bonuses attached to such positions varied formulaically and outside principals' discretion.

[12] Previous versions modeled open continuous bidding, which permits the adoption of optimal bidding strategies from Milgrom and Weber (1982). This allows each school to update the optimal bid conditioning on the rival's bidding behavior. However, both bidding processes result in the same predictions.

[13] Prior work shows principals care about teacher effectiveness, particularly in schools under accountability pressure. Other work shows that high-VAM teachers also lead to a wide array of better future outcomes for their students, giving further reason to suggest principals may maximize these short-run measures of effectiveness.

$\sigma_T(0)\sigma_\xi(x) + \sigma_T(0)\sigma_\epsilon + \sigma_\epsilon\sigma_\xi(x)$, if uninformed of a teacher's VAM (subscript NV), a hiring principal's optimal maximum bid (b^{h*}_{isdNV}) is given by equation 1.1.

$$b^{h*}_{isdNV} = \frac{\sigma_T(0)\sigma_\xi(x)}{Z^h_{NV}} m + \frac{\sigma_T(0)\sigma_\epsilon}{Z^h_{NV}} R_x + \frac{\sigma_\epsilon\sigma_\xi(x)}{Z^h_{NV}} P^h_0. \tag{1.1}$$

If there is public learning, as experience increases, more public information leads to a more precise public signal. As $\sigma_\xi(x)$ declines, hiring principals place less weight on their prior beliefs and noisy private information, and more weight on the public signal.

A principal seeking to retain her teacher, who is uninformed of his VAM, has an optimal bid (b^{*r}_{isdNV}) with very a similar form to that shown is equation 1.1. Equation 1.2 shows her optimal bid, letting $Z^r_{NV} = \sigma_T(t)\sigma_\xi(x) + \sigma_T(t)\sigma_\epsilon + \sigma_\epsilon\sigma_\xi(x)$.

$$b^{r*}_{isdNV} = \frac{\sigma_T(t)\sigma_\xi(x)}{Z^r_{NV}} m + \frac{\sigma_T(t)\sigma_\epsilon}{Z^r_{NV}} R_x + \frac{\sigma_\epsilon\sigma_\xi(x)}{Z^r_{NV}} P^r_t. \tag{1.2}$$

Retaining principals provide more weight to their private information (P^r_t), if they obtain more useful information than is publicly available. This is reflected by $\sigma_T(t)$ which shrinks with additional private information as opposed to $\sigma_T(0)$ from equation 1.1, which remains constant for hiring principals.

The introduction of VAMs alters the information available to principals, but the optimal bids that incorporate VAMs have similar form to those shown in equations 1 and 2. Whether the VAMs are public or private are particularly important for determining retaining principals' expectations of a given teacher in the adopting districts.

If a principal's rival is from outside of the district and uninformed of the measure, the retaining principal incorporates the VAM into her private signal. The new private signal

($P^r_{t\nu}$) becomes the precision-weighted average of the prior private information and the new VAM. Thus, the optimal bid of a retaining principal, who has access to her teacher's VAM and whose rival does not have access to the VAM (denoted by the subscript RV) is shown in equation 3 where $Z^r_{RV} = \sigma_\mathcal{T}(t\,V)\sigma_\xi(x) + \sigma_\mathcal{T}(t\,V)\sigma_\epsilon + \sigma_\epsilon\sigma_\xi(x)$.

$$b^{r*}_{isdRV} = \frac{\sigma_\mathcal{T}(t\,V)\sigma_\xi(x)}{Z^r_{RV}}m + \frac{\sigma_\mathcal{T}(t\,V)\sigma_\epsilon}{Z^r_{RV}}R_x + \frac{\sigma_\epsilon\sigma_\xi(x)}{Z^r_{RV}}P^r_{t\nu}. \tag{1.3}$$

Equation 3 is similar to equation 2 except for the replacement of P^r_t by $P^r_{t\nu}$ and of $\sigma_\mathcal{T}(t)$ by $\sigma_\mathcal{T}(t\,V)$. In expectation, the magnitude of the private signal will not change with the introduction of VAMs. However, the precision of the cumulative private information must increase.

Lemma 1: The precision of the private signal increases with the incorporation of VAMs into the private signal ($\sigma_\mathcal{T}(t\,V) < \sigma_\mathcal{T}(t)$).

Proof: Under the orthogonality assumptions, $var(P_{t\nu}) \equiv \sigma_\mathcal{T}(t\,V) = \frac{\sigma_\nu^2\sigma_\mathcal{T}(t)+\sigma_\nu\sigma_\mathcal{T}(t)^2}{(\sigma_\nu+\sigma_\mathcal{T}(t))^2} = \frac{\sigma_\nu\sigma_\mathcal{T}(t)}{\sigma_\nu+\sigma_\mathcal{T}(t)}$. $\frac{\sigma_\mathcal{T}(t)(\sigma_\nu+\sigma_\mathcal{T}(t))}{\sigma_\nu+\sigma_\mathcal{T}(t)} - \frac{\sigma_\nu\sigma_\mathcal{T}(t)}{\sigma_\nu+\sigma_\mathcal{T}(t)} = \frac{\sigma^2_\mathcal{T}(t)}{\sigma_\nu+\sigma_\mathcal{T}(t)}$, and $\frac{\sigma^2_\mathcal{T}(t)}{\sigma_\nu+\sigma_\mathcal{T}(t)} > 0$, by property of variances.

This decrease in the variance of the private signal decreases the weight retaining principals place on their prior beliefs and the public information, while increasing the relative weight they place on their now fuller private information.

Turning back to hiring principals' expectations of teacher quality, if a hiring principal is uninformed of VAMs (or their existence), her expectation of the teacher's quality would remain unchanged from those presented in equation 1. Thus, the introduction of VAMs exacerbate informational asymmetries between prospective employers.

In contrast, if both bidding principals are informed of a teacher's VAM, as is likely the case when both principals are from one of the adopting districts after the policy takes effect, the VAM enters the principals' public signal of teacher quality. Letting $Z^r_{HV} = \sigma_\tau(t)\sigma_\xi(x\,V) + \sigma_\tau(t)\sigma_\epsilon + \sigma_\epsilon\sigma_\xi(x\,V)$, equation 4 provides the retaining principal's optimal bid when the hiring principal may also access a teacher's VAM (denoted with the subscript HV).

$$b^{r*}_{isdHV} = \frac{\sigma_\tau(t)\sigma_\xi(x\,V)}{Z^r_{HV}}m + \frac{\sigma_\tau(t)\sigma_\epsilon}{Z^r_{HV}}R_{x\nu} + \frac{\sigma_\epsilon\sigma_\xi(x\,V)}{Z^r_{HV}}P^r_t. \tag{1.4}$$

Equation 4 is similar to equation 2 with the exception that R_x is replaced by $R_{x\nu}$, as VAMs enter the public signal. While in expectation the magnitude of the public signal is the same with or without VAMs, the variance of the public signal must change as a result.

Lemma 2: The precision of the public signal increases with the incorporation of VAMs into the public signal ($\sigma_\xi(x\,V) < \sigma_\xi(x)$).

Proof: Under the orthogonality assumptions, $var(R_{x\nu}) \equiv \sigma_\xi(x\,V) = \frac{\sigma_\nu^2\sigma_\xi(x)+\sigma_\nu\sigma_\xi(x)^2}{(\sigma_\nu+\sigma_\xi(x))^2} = \frac{\sigma_\nu\sigma_\xi(x)}{\sigma_\nu+\sigma_\xi(x)}.\ \frac{\sigma_\xi(x)(\sigma_\nu+\sigma_\xi(x))}{\sigma_\nu+\sigma_\xi(x)} - \frac{\sigma_\nu\sigma_\xi(x)}{\sigma_\nu+\sigma_\xi(x)} = \frac{\sigma_\xi^2(x)}{\sigma_\nu+\sigma_\xi(x)}.\ \frac{\sigma_\xi^2(x)}{\sigma_\nu+\sigma_\xi(x)} > 0$, by property of variances.

For equation 4, this means that retaining principals will shift weight that they had previously placed on the private signal onto the new more complete 'publically' available information.

If access to the VAMs is shared between employers, the VAMs enter the public signal of hiring principals, just as they enter the public signal of retaining principals. Letting $Z^h_{HV} = \sigma_\tau(0)\sigma_\xi(x\,V) + \sigma_\tau(0)\sigma_\epsilon + \sigma_\epsilon\sigma_\xi(x\,V)$, equation 5 provides the hiring principal's optimal bid when she may also access a teacher's VAM (subscripted HV).

$$b^{h*}_{isdHV} = \frac{\sigma_T(0)\sigma_\xi(x\,V)}{Z^r_{HV}}m + \frac{\sigma_T(0)\sigma_\epsilon}{Z^r_{HV}}R_{x\nu} + \frac{\sigma_\epsilon\sigma_\xi(x\,V)}{Z^r_{HV}}P^h_0. \tag{1.5}$$

The difference between equations 1 and 5 are in the public signal and its variance. Using the finding from lemma 2, that the variance of the public signal drops with the introduction of VAMs, once hiring principals may access a teacher's VAM, they place less weight upon their prior beliefs and less weight upon their noisy private information they glean from the application process, and place more weight on the public information that now includes a teacher's VAM. For bids in which both principals become informed of a teacher's VAM, the information between prospective employers becomes more symmetric, and their expectations converge, as both hiring and retaining principals shift weight onto the information that they share.

1.4.3 Mobility under Asymmetric Information

The teacher labor market generally moves in the summer between school years. At that time, teachers may sample two offers, an update from their current school and one outside offer. Teachers move to the school that offers the highest bid.[14] Accordingly, the probability of a move is:

$$P(M) = P\left[b^{h*}_{isd} - b^{r*}_{isd} > 0\right]. \tag{1.6}$$

Such school-to-school transfers are motivated in general by a hiring principal valuing the teacher more so than does the retaining principal. Letting ψ stand for the composite error

[14] For simplicity, I model mobility decisions as a spot market. A fixed transition cost or idiosyncratic teacher preferences over schools may be added without additional assumptions.

term and substituting in the bids from presented in equations 1 and 2 allows equation 6 to be written in the form presented in equation 1.7.[15]

$$P(M) = P\left[\psi > \sigma_{\xi}(x)\left(\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)\right)(\mu - m)\right]. \tag{1.7}$$

While the VAMs and who has access to them alters the informations on which principals operate, the general form of equation 1.7 remains the same, making it useful for illustration. Such transitions may occur due to extreme private signals. However, this may happen even if both principals receive the same private signal due to differences in how each principal weighs the signals she receives.

For such mobility, it is apparent from equation 1.7 that all else equal, the probability of a move is inversely related to true effectiveness. Intuitively, due to their additional knowledge of teacher effectiveness, the current school should value the true effectiveness of the teacher more than the outside market. Because the outside market has less information about true effectiveness, the outside schools should place more weight on the easily observed correlates with teacher effectiveness than the current school, which inform the prior belief (m).

The primary investigation in this study explores how mobility changes with the adoption of VAMs. The availability of VAMs to some prospective employers, but not others, provides a rare test for the model laid out above. As described in Section 1.2, both districts' adoption of VAMs, theoretically provide a shock to the information of all principals within the district. There are two primary ways of thinking about the impact of VAMs in the model. The first is more in keeping with the prior employer learning literature. VAMs serve as difficult-to-observe measures of teacher quality. Researchers may use VAMs to proxy directly for μ about

[15] See Subsection E in the Appendix for algebraic transformations.

which employers are learning. In this framework, the model offers predictions of whether better or worse teachers move as response to adopting these VAMs. Equation 1.8 takes this broad view.[16]

$$
\begin{aligned}
&\frac{\partial E\left[b^{h*}_{HV} - b^{r*}_{HV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,\mu\right]}{\partial \mu} = \\
&\sigma_\epsilon(\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t))((\sigma_\xi(x) - \sigma_\xi(x\,V))(\sigma_\mathcal{T}(t)\sigma_\epsilon\sigma_\mathcal{T}(0)\sigma_\xi(x)\sigma_\xi(x\,V) \\
&+ \sigma_\xi(x\,V)\sigma_\epsilon^2\sigma_\xi(x)\sigma_\mathcal{T}(0) + \sigma_\xi(x\,V)\sigma_\mathcal{T}(t)\sigma_\epsilon^2\sigma_\xi(x) + (\sigma_\xi(x\,V) + \sigma_\xi(x))\sigma_\mathcal{T}(t)\sigma_\epsilon^2\sigma_\mathcal{T}(0)).
\end{aligned}
\tag{1.8}
$$

Under the assumption that $\sigma_\mathcal{T}(0) > \sigma_\mathcal{T}(t)$, which is fundamental to asymmetric employer learning and by $\sigma_\xi(x) > \sigma_\xi(xV)$, which was shown in lemma 2, $\frac{\partial E\left[b^{h*}_{HV} - b^{r*}_{HV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,\mu\right]}{\partial \mu}$ > 0. Therefore, the model predicts that providing VAMs to both principals, as occurred within both districts, should raise the probability that good teachers move, all else equal.

Under the second interpretation, EVAAS VAMs enter the two districts directly as new signals. Accordingly, the model offers predictions on the differential effects of the policy on the probability of moving for teachers receiving different signals, all else equal. After some algebra, equation 1.9 takes this more narrow view.[17]

$$
\frac{\partial E\left[b^{h*}_{HV} - b^{r*}_{HV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,V\right]}{\partial V} = \frac{1}{Z^h_{HV} Z^r_{HV}} \frac{\sigma_\xi(x)}{\sigma_\nu + \sigma_\xi(x)} > 0 \tag{1.9}
$$

Therefore, while the interpretations are subtly different, the comparative statics with respect to VAMs after the policy takes effect are the same. Within the districts, where both principals are aware of the signals once they are implemented, the model predicts teachers who receive a high-VAM signal become more likely to transfer schools.

[16] See Appendix E provides the relevant algebraic transformations.
[17] See Appendix E for the relevant algebraic transformations.

Recall from Section 1.2, that if principals in other districts know of the existence of VAMs for teachers from Winston-Salem and Guilford, the policy would theoretically alter their information as well. The previous prediction would apply to out-of-district moves as well. However, it is plausible that principals in other districts were uninformed about the policy. In which case, the adoption of VAMs in Guilford and Winston-Salem would make the balance of information more asymmetric, in the event that a teacher contemplates moving to another school outside Winston-Salem or Guilford. If the hiring principal is uninformed of the VAM, VAMs enter retaining principals' private signals.

The same two interpretations of VAMs' role apply here. Again beginning with the broader view of VAMs as a measure of μ, equation E demonstrates the model's predictions with respect to teachers' underlying abilities on the probability of moving to uninformed principals.[18]

$$\frac{\partial E\left[b^{h*}_{RV} - b^{r*}_{RV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,\mu\right]}{\partial \mu} =$$

$$\frac{-\sigma_\xi(x)^2\sigma_\epsilon}{Z^h_{RV}Z^r_{RV}Z^h_{NV}Z^r_{NV}}(\sigma_\tau(t) - \sigma_\tau(t\,V))(\sigma_\tau(0)^2\sigma_\epsilon^2 + \sigma_\tau(0)^2\sigma_\xi(x)^2 + \sigma_\tau(0)^2\sigma_\epsilon\sigma_\xi(x) + \sigma_\xi(x)^2\sigma_\epsilon^2. \tag{1.10}$$

Under lemma 1, $\sigma_\tau(t) > \sigma_\tau(t\,V)$, which implies that $\frac{\partial E\left[b^{h*}_{RV} - b^{r*}_{RV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,\mu\right]}{\partial \mu} < 0$. Therefore, the model predicts that after the release of VAMs to retaining principals, the likelihood of moving out-of-district will decrease with increases in teacher quality, and vice versa.

Under the more narrow view of VAMs as only pertaining to the signal itself, again the

[18] See Appendix E for the relevant algebraic transformations.

predictions remain consistent. Equation 1.11 presents the partial derivative of the expected difference in the differences between employers bids with respect to the VAM signal itself.[19]

$$\frac{\partial E\left[b^{h*}_{HV} - b^{r*}_{HV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,V\right]}{\partial V} = \frac{-\sigma_{\xi}(x)\sigma_{\epsilon}\sigma_{\mathcal{T}}(t)}{Z^{r}_{RV}(\sigma_{\nu} + \sigma_{\mathcal{T}}(t))} < 0 \tag{1.11}$$

Here the model predicts adverse selection of out-of-district moves on the basis of VAMs, all else equal. It is important to note that good (or high-VAM) teachers may choose to reveal their EVAAS report to principals in other districts in an effort to move out-of-district. Accordingly, the furthering of information asymmetries between employers may not universally apply to out-of-district moves. However, as long as some low-VAM teachers are able to move out-of-district without being penalized by their EVAAS report (or their unwillingness to reveal it), the model predicts more negative (smaller in magnitude or negative) effects of VAM on the probability of moving out-of-district after policy implementation than are produced for moves within-district. Thus, the test between symmetric and asymmetric learning is whether effects of the policy on the selection of out-of district movers are significantly more negative than the effects of adopting VAMs on the selection of within-district movers.

1.5 Data and Estimation

In this section, I describe both the data and methods used to generate VAMs of teacher effectiveness, and the effects of the district policies on the teacher mobility. Subsection 1.5.1 describes the generation of VAMs. Subsection 1.5.2 describes the estimation sample. Subsection 1.5.3 describes the difference-in-differences estimation approach used to identify the effects of the new information on the mobility decisions of teachers and principals.

[19] See Appendix E for the relevant algebraic transformations.

1.5.1 Value-Added Measures

While there are other valuable dimensions of teaching, many schools and districts care a great deal about teachers' abilities to raise their students' performance on standardized assessments. This study relies on administrative, longitudinal data, which links students to their teachers and was generously provided by the North Carolina Education Research Data Center (NCERDC) to estimate teachers' abilities to do just that. Though a robust source of data, unfortunately, the NCERDC does not contain the exact VAMs issued to each teacher within the treatment districts, and neither district agreed to release them. Consequently, this study will measure the student gains on the North Carolina End of Grade exams attributable to each teacher.

There are two primary ways to go about this. The first is to attempt to model the exact measures that teachers and principals receive. This is primarily useful in explaining the teachers' and principals' observed behavior. The second is to model teacher effectiveness as accurately as possible. This is primarily useful in evaluating the consequences of the policy. To illustrate this distinction, suppose that the EVAAS score were totally uninformative. Observing mobility based on them would clearly illustrate the impact of the additional signal, but would offer no insight into the effect on educational equity. In contrast, using a measure of true effectiveness provides direct policy implications and is also useful in testing the learning hypotheses. Accordingly, I prefer this second, broader approach, which is tied more closely to the employer learning framework, which relies on the error in variables that proxy for underlying ability. This study follows earlier studies of employer learning in supposing that the researcher may access information originally unavailable to market participants.[20]

[20] Whereas Farber and Gibbons (1996); Altonji and Pierret (2001); Lange (2007); Schönberg (2007), and Pinkston (2009) use AFQT score as a strong correlate with productivity about which employers must learn,

In my preferred specification, I model teacher effectiveness rather than attempting to replicate the EVAAS measure. An element of feasibility also enters this preference. The EVAAS system is proprietary, and the exact data and methods used are not disclosed. Furthermore, SAS uses two different proprietary models, and for large school districts it is unclear which is used. Of course, in actuality, the resulting measures from either approach are likely be highly correlated, and in Section 1.7, I check the robustness of my results against other specifications.[21] In this context, the VAMs need not totally encompass a teacher's effectiveness. Here, VAMs only need to be stronger correlates with teacher effectiveness than are other correlates with productivity, such as educational attainment and level of certification.[22] While VAMs likely do not measure all traits that principals may seek in their teachers, they do directly measure one component of teaching quality that is important to principals and policy makers.

My preferred measure of VAMs is what Guarino et al. (2012) call the Dynamic OLS (DOLS) estimator presented in equation 1.12. According to Guarino et al. (2012), this DOLS estimator is more robust to nonrandom student assignment, a frequent criticism of the often used Empirical Bayes estimator, which assumes random assignment of students to teachers.[23]

I use the VAM described above in this capacity.

[21] Rose et al. (2012) finds a .91 correlation between one EVAAS measure and Dynamic OLS.

[22] The extant literature supports this claim. As Rivkin et al. (2005) show, easily observed teacher characteristics are not highly correlated with teacher effectiveness. Experimental evidence from Hinrichs (2013) suggests that GPA matters little to schools in hiring decisions, and that the strongest determinant of receiving a positive response from a school is whether the teacher holds an in-state certificate. However, Jacob and Lefgren (2008) find large agreement between principal evaluations of teachers and VAMs, at least in the tails of the distributions of both measures. Furthermore, recent work shows significant correlation between teachers' VAMs and many important future outcomes for their students, including educational attainment, earnings, and probability of incarceration Chetty et al. (2011, 2014).

[23] Given teachers' preferences found in Jackson (2009) and Boyd et al. (2013), it seems unlikely that teacher effects would be uncorrelated with student-level covariates.

$$A_{ijt} = \tau_{\mathbf{t}} + \mathbf{A_{ijt-1}}\boldsymbol{\beta_0} + \mathbf{X_{it}}\boldsymbol{\beta_1} + \boldsymbol{VAM_j} + e_{it} \quad (1.12)$$

Here, A_{ijt} represents student i's mathematics achievement in teacher j's class in year t. Including $\mathbf{A_{it-1}}$ allows for the correlation of previous math and reading test performances with current performance. Additionally, $\mathbf{X_{it}}$ is a vector including demographic attributes of individual students, such as grade, race, gender, special needs, and gifted status. It is $\mathbf{VAM_j}$, a vector of teacher indicators, which is of primary interest for this study. Acknowledging that VAMs can be somewhat unstable in any single year, my preferred estimates use data from each year a teacher is teaching 4^{th} through 8^{th} grade during my sample period. This allows me to gain the most precise estimate of teachers' true underlying ability, μ.

1.5.2 Estimation Sample

This study restricts attention to the 5,986,132 elementary and middle school student, year observations from 1997 through 2011 to construct the VAMs for 134,219 teachers who teach 4^{th} through 8^{th} grade. I link these data to education, licensing, and work history data of 67,062 lead teachers without teaching assistants for whom the records are complete. These teachers are dispersed across the 2,966 schools in 117 school districts. I further restrict the sample to only those teachers teaching 4^{th} through 8^{th} grade at the time of observation, since they are the only elementary and middle-school teachers to receive VAMs. This restriction pares down my sample from 416,135 teacher-year observations to 236,018. At the teacher level, the data includes the teachers' race, gender, institution of higher education, degrees earned, experience, and tenure at a given school. Each of these are easily observable to all schools and many are likely used to filter job candidates. I use performance at the school in which the teacher currently works as an additional, easily observable, possible correlate with effectiveness. Table A.2 provides summary statistics for my estimation sample.

The districts that adopt VAMs do not differ substantially from state averages in achievement or percent of student receiving proficiency on the state standardized exams. Given that both districts include urban centers, they do have a higher proportion of Black students and teachers than does an average district in the state. While teachers come from colleges of comparable selectivity, across districts, in Winston-Salem, a larger share of the teaching-force holds an advanced degree. However, on the basis of VAMs, teaching quality in both districts is very close to the state average.

1.5.3 Estimation Strategy

The regression based differences-in-differences approach allows me to isolate mobility based on underlying effectiveness from mobility based on correlates with effectiveness. Furthermore, easily observable, lower correlates with effectiveness may become less tied to the probability of moving after the introduction of VAMs. I estimate the following specification:

$$y^*_{jdt} = \boldsymbol{T_t} + \mathbf{d_d} + VAM_j\mathbf{G_{1dt}} + \mathbf{X_{jdt}}\mathbf{G_{2dt}} + \xi_{jdt}, \tag{1.13}$$

$$\mathbf{G_{hdt}} = \gamma_{h1} + \mathbf{TreatDist_{jd}}\boldsymbol{\gamma_{h2}} + \mathbf{Post_t}\boldsymbol{\gamma_{h3}} + \mathbf{TreatDist_{jd}} \times \mathbf{Post_t}\boldsymbol{\gamma_{h4}}, h = 1, 2,$$

where y^*_{jdt} is the latent probability of a job change for teacher j in district d and in year t. I only observe the binary outcome of when a move occurs. $\mathbf{T_t}$ represents year effects, $\mathbf{d_d}$

represents district fixed effects, and $\mathbf{X_{jdt}}$ is a vector of teacher and school characteristics including teacher experience, tenure,[24] race, highest degree earned and selectivity of bachelor degree granting institution, as well as percent of students who are Black and percent of students testing above proficiency at the school level. $\mathbf{G_{1dt}}$ and $\mathbf{G_{2dt}}$ capture the differences in the effects of VAMs on mobility based on whether VAMs were available for teacher j in district d, at time t. Interactions with treatment district indicators separate permanent differences in the impacts of VAMs and other characteristics from confounding the effect of treatment, while interactions with indicators for post years do the same for statewide changes in the effects at the times the policies take effect. Thus, the identifying variation comes from the differences between adopting districts and the rest of the state in the differences in the predictive power of VAMs on the probability of moving schools between pre- and post-policy years.

Keeping in mind previously estimated teacher preferences and more importantly potential differences in information available, I examine the six types of job changes separately: within district moves, within district moves to higher-performing schools, within district moves to lower-performing schools, out-of-district moves, out-of-district moves to higher-performing schools, and out-of-district moves to lower-performing schools. Given that teachers initiate most moves, moves to worse schools are likely driven by largely by idiosyncratic teacher preferences. Due to the indirect mechanism by which hiring principals in Winston-Salem obtain teachers' VAMs and the potential additional salience of VAM signals to principals outside the district during Winston-Salem's later adoption, I separate treatment by district.

Given how the districts distributed VAMs, it seems clear that the new information would

[24] Because tenure is generated and censored for job matches beginning prior to 1995, an indicator of whether the current match existed in 1995 is included in all regressions.

be public between two principals in Guilford. Perhaps to a lesser extant the same holds for Winston-Salem. Accordingly, the model predicts $\boldsymbol{\gamma_{14WD}} > 0$ (where $\boldsymbol{\gamma_{14WD}}$ is the effect of the interaction of VAM with receiving treatment on the probability of moving within-district). Furthermore, because there would be more information available on more experienced teachers, if there previously been some degree of public learning, the model predicts the effects to diminish with teacher experience. Likewise, if there had previously been private learning, the learning model predicts the shock to public information to have larger ramifications for teachers with more tenure at a given school all else equal. In later specifications, I interact VAM with experience and the difference-in-differences, $\mathbf{G}$, interactions.

When comparing the expectations of a retaining principal within one of the treatment districts to a hiring principal in another district there is some ambiguity as to whether VAMs provide a more precise expectation for both principals or only the current one. Thus, the symmetric learning model for out-of-district moves predicts $\boldsymbol{\gamma_{14OD}} = \boldsymbol{\gamma_{14WD}}$ (where $\boldsymbol{\gamma_{14OD}}$ is the effect of the interaction of VAM with receiving treatment on the probability of moving out-of-district). If current principals can keep information from employers in other districts, the signal improves the precision of the current principal's signal about the true quality of the teacher, while the expectation of the out-of-district principal is unaffected. In which case, the asymmetric learning model would apply predicting $\boldsymbol{\gamma_{14WD}} > \boldsymbol{\gamma_{14OD}}$ and possibly $\boldsymbol{\gamma_{14OD}} < 0$ for out-of district moves.

This type of movement may have important implications for the distribution of teacher quality across schools. If better teachers are more able to signal their true quality, and do so in general to move to better schools, the divide in teacher quality between the worst and best schools may widen. Accordingly, I estimate equation 1.13 substituting percent of students proficient in the school taught at the subsequent year, for the binary variable of whether teachers move. Again, if VAMs are informative, and teachers do in general prefer to teach at better schools, $\gamma_{\mathbf{14SQ}} > 0$ in this regression as well. ($\gamma_{\mathbf{14SQ}}$ is the effect of the interaction of VAM with receiving treatment on the proficiency levels of the school where the teacher works the subsequent year.) Similar to the probability of moving to a better school, we may expect these effects to be somewhat muted for teachers moving later in their careers, in which case hiring principals may already have more complete information.

There are two distinct issues that complicate the estimation of standard errors in this study. First, the policy variation occurs at the district level. As a result, the errors may be correlated for teachers moving from or within the same district. The appropriate response to this single issue is to cluster the standard errors at the district level. The second, issue results from the fact that the teacher VAMs are estimated. By simply clustering the standard errors, the VAMs are treated as though they are known, and thus, they do not account for the inherent variability due to estimation error. Were this a singular issue, it would be appropriate to bootstrap the student data to account for this estimation error. It may seem natural to then cluster-bootstrap at the district level. However, this samples all students for a every teacher in a sampled district, and as a result, does not actually address the estimation error. In fact, the standard errors from the cluster bootstrap are smaller than the non-bootstrap clustered standard errors by about a factor of ten.

Accordingly, I adopt a sampling approach that accounts for both the estimation error of VAM and the clustered nature of the data. First, I sample districts randomly with replacement just as with the standard cluster-bootstrap. I then conduct stratified sampling at the teacher level, such that for every teacher who was originally sampled, I randomly sample student/year observations with replacement. In so doing, this provides generally more conservative standard errors across parameters. The standard errors on the effects of the policy on the relationship between VAMs and the probability of moving schools are comparable to the standard bootstrapped standard errors, and the standard errors on all other estimated coefficients are comparable to the non-bootstrapped district-clustered standard errors. Table F.7 in the Appendix ?? presents all standard errors for Table A.3 for comparison. Throughout the remainder of this paper, I present the more conservative district-clustered-teacher-stratified-bootstrap standard errors (CSB SEs).

1.6 Results

1.6.1 Mobility and Sorting

Table A.3 presents the estimated impact of revealing EVAAS reports of teacher effectiveness on the relationship between teachers' VAMs and the probability a teacher moves to another school. Given the evidence that teachers prefer to teach in schools with higher-performing students, Table A.3 decomposes effects by whether the receiving school has higher or lower-performing students.[25] The test between symmetric and asymmetric employer learning focuses on how the effects of VAMs on the probability of moving within-district differ from the effects of VAMs on the probability of moving out-of-district after the treatment districts adopt the measures of teacher quality. Panel A restricts attention to within-district moves, and Panel B presents evidence from out-of-district moves.

The first row presents the the relationship between VAMs and the probability of each type of move in the rest of the state, regardless of any districts adopting the policy. In

[25] Primary effects of VAMs on different types of moves as well as on future school performance further supports this distinction. I define a move to a higher performing school as a move in which the school taught at the following year has a higher percentage of students who achieve proficiency than the current school. Proficiency rates are demeaned by year statewide averages, while a move to a lower-performing school is defined in the reverse way.

general, there is little relationship between VAMs and the probability of moving within or out of the district. However, when discerning between moves to more and less proficient schools a familiar pattern emerges. From columns 2 and 3 of Panel A, a teacher with a standard deviation higher VAM is about 0.3 percentage points more likely to move to a higher-performing school and 0.2 percentage points less likely to move to a lower-performing school within the district. Panel B exhibits the same pattern regarding moves to schools outside of the current district. A one standard deviation increase in VAM before the policy takes effect raises the probability of moving to a higher-performing school by about a tenth of a percentage point and lowers the probability of moving to lower-performing school by about the same magnitude.

Within both Guilford and Winston-Salem, the release of VAMs intensifies this pattern. From the coefficient on the interactions between policy treatment and VAMs in both districts, a standard deviation increase in a teacher's VAM leads to about a half of a percentage point increase in the probability of moving within district after the district released the value-added information. While the magnitudes of the effects are very close between districts, they are only statistically significant beyond the 95% confidence level for Guilford. Column 2 illustrates that these results are driven by moves to higher-performing schools, as the model predicts. From column 2, the estimated coefficients imply that the adoption of VAMs raises the probability that a teacher with one standard deviation higher VAM will move to a higher-performing school by over 14% (p-value .011) in Guilford and nearly 18% (p-value .009) in Winston-Salem. Column 3 reveals little change in the effects of VAMs on the probability of moving to a lower-performing school within district. The similarity of the point estimates on the impact of VAMs post-treatment between Guilford and Winston-Salem provides no evidence that relying upon teachers to voluntarily disclose their VAMs to hiring principals mitigates the effects.

From Section 1.4, the effect of the policy should be no different whether teachers move to schools within or outside of the district, under the symmetric learning hypothesis. However, asymmetric employer learning predicts the policy to give principals in Guilford and Winston-Salem an informational advantage over principals in other districts. This translates into smaller selection effects for teachers moving to other districts than for within-district moves, and these effects may even be negative. The second column of Panel B presents changes in the effect of teacher quality on the probability of moving to a better, out-of-district school after the adoption of VAMs. Again, these changes in selection of mobile workers are consistent with the employer learning model.

The change in selection of teachers leaving Guilford provides the strongest evidence of growing informational asymmetries between employers. In Guilford, a teacher who has a standard deviation lower VAM, is a full percentage point more likely to move out-of-district. This same teacher is about a half a percentage point more likely to move to a better school out-of-district (p-value 0.001). There is also a statistically significant effect on the probability of moving to lower-performing schools out of Guilford. While the model does not predict this type of movement, it is not surprising. Low VAMs may lead current principals to devalue some of their teachers, who may respond by moving to lower-performing schools that are not privy to their value-added scores.

In Winston-Salem, the difference between within- and out-of-district moves is less pronounced, though still consistent with private employer learning. While in Winston-Salem, a teacher with one standard deviation higher VAM is more likely to move to a higher-performing school out-of-district after the policy takes effect, the point estimate is only 38% of that from moving within-district and is no longer statistically significant. Were outside principals informed of the signal, we would expect the same positive effects found in the second column of Panel A to be present in in the second column of Panel B.

The fact that effects are more negative in Guilford than Winston-Salem, may be explained by differences in the salience of the signals between teachers moving from Guilford as opposed to those moving from Winston-Salem. Guilford's adoption of the EVAAS measures of teacher effectiveness occurred in 2000. It is unlikely that at that time principals in other districts had much understanding of the measures, or their reliability. In contrast, the rest of the state adopted school-level EVAAS reports simultaneously with Winston-Salem's adoption of teacher level VAMs. Given this difference in contexts, high VAM teachers from Winston-Salem may have been better able to use their VAMs to obtain positions outside of Winston-Salem, than would a comparable teacher moving earlier from Guilford. In Winston-Salem, the increase in high-VAM teachers' ability to signal their effectiveness may mitigate any effects from relatively low VAM teachers exploiting the informational asymmetry. The mitigated effects of VAM for those moving out of Winston-Salem in addition to the negative selection of teachers moving away from Guilford evidences informational asymmetries between potential employers within as opposed to outside of the district.

Turning to the implications of such mobility for educational equity in general, Table A.4 presents the results of how the sorting of teachers to schools changes with the implementation of the policy. The coefficient on VAM describes the relationship between teachers' VAMs and the proficiency level of the school they teach at the subsequent year in the rest of the state. Across both columns, a one standard deviation increase in a teacher's VAM leads to about a quarter of a percentage point increase in the percent of students who are proficient in the school in which he teaches the subsequent year. The result that students in better schools also get better teachers is consistent with findings in Boyd et al. (2005) and Boyd et al. (2008).

Column 1 examines the effect of the policy on sorting for all teachers in the sample who remain teaching in North Carolina the following year. Column 2 restricts the sample to those who remain within their current district. The second column may be more informative for predicting the effects in the rest of the state after the adoption of EVAAS VAMs becomes statewide. Theoretically, the effects may be more pronounced for the state as a whole, because the costs of moving out of state are in general higher than those of moving out of a school district. The difference in results from Table A.3 between within- and out-of-district moves imply more positive correlations between teacher VAMs and school performance among those who remain in district than overall, as a result of the policy. Table A.4 reflects those patterns. Including teachers who move within and out of district, it seems from column 1 that releasing VAMs of teacher effectiveness does little to change the distribution of teacher quality across schools. However, turning to the sample of teachers who remain in the same district in column 2, while there is no evidence of sorting in general rising in Guilford as a result of the policy, in Winston-Salem, on average I find a teacher with one standard deviation higher VAM will be at a school that has 0.2 percentage points higher proficiency rates after the district releases VAMs. In Winston-Salem, this translates to about a 70% increase in the correlation between teacher quality and student performance as a result of the policy. This large effect for Winston-Salem taken together with the mobility patterns from Table A.3 evidence rising inequality in the distribution of effective teachers as an unintended consequence of VAM adoption.

1.6.2 Observables

In addition to predicting mobility dynamics with respect to teacher VAMs, the model presented in Section 1.4 also offers predictions regarding easily observable covariates with teacher effectiveness. In instances where the VAMs shock the available public information, the model predicts principals would place less emphasis on easily observable covariates with teacher effectiveness, such as degree attainment and college selectivity. In cases where VAMs exacerbate informational asymmetries between current and hiring principals, the same covariates expectedly receive additional emphasis on the probability of a move.

To provide ease of interpretation, I generate an index of easily observable teacher quality by taking the fitted values from the OLS regression of teacher VAMs on teacher covariates. I include as components of this index, an indicator for having an advanced degree, a vector of indicators for Barron's College Competitiveness index, years of experience, years of tenure, an indicator for whether tenure is censored, race, gender, and a vector of year indicators.[26]

In general, those with high observable characteristics are more likely to move within district. That result is driven by moves to higher-performing schools, while those with lower observable characteristics are more likely to move to lower-performing schools. For moves out-of district, the positive relationship between the index and the probability of moving to a better school offsets the negative relationship between the index and the probability of moving to a lower-performing school. These relationships are expected given the sorting of teachers based on observable characteristics as shown in Jackson (2009) among others.

The first two columns of Table A.5 do not bear out the predictions for within district moves. While noisy, the point estimates of the effects of the teacher index on the prob-

[26] The VAMs used in this analysis are the residuals from the projection of my standard VAMs on the components of the index.

ability of moving schools within-district after the adoption of VAMs are positive, though only statistically significantly so for moves to better schools within Guilford. While not expected, this result may be explained by the additional churn that accompanies the adoption of VAMs particularly for moves to better schools within Guilford. More positions may become available as a result of high-VAM teachers moving to better schools, and low-VAM teachers moving out of district. As a result, those with good observables find it easier to move in addition to those with high VAMs. Heterogeneous openness among principals to VAMs may also contribute.[27] In which case, as high-VAM teachers move to principals that value VAMs those with other favorable attributes move to the principals who value those characteristics.

The change in the relationship between the index and the probability of moving out-of-district with the adoptions of VAMs is more supportive of the model. Whereas movers out of Guilford are adversely selected on the basis of the hard-to-observe VAM, they are positively selected on the basis of this index of easily observable measures of teacher quality. This is true across moves to higher or lower performing schools, and provides further evidence that the moving teachers with a high index, but low VAM were able to keep their VAM private, while utilizing their otherwise strong resumés to move to uninformed principals. Given that it is plausible that more teachers moving from Winston-Salem could inform out-of-district principals of their VAMs, results in either direction may make sense. Accordingly, the results for moves out of Winston-Salem are not very informative. While the results for moves out of Guilford are reassuring, cumulatively, the evidence from changes in the relationship between the index of easily observable teacher characteristics, and the probability of moving schools

[27] Informal conversations with principals in Winston-Salem and Guilford indicate this may be the case, as two current lower elementary principals that I spoke with indicated that teachers' VAMs played a limited role in their hiring decisions.

is too mixed to draw definitive conclusions.[28]

1.6.3 Differential Effects With Respect to Experience and Tenure

The final piece of primary analysis examines the effects of the policy on the correlation between teacher VAMs and the probability of moving with respect to years of experience and tenure. If teachers are able to draw upon each year of experience to better demonstrate how good they are through resumés, references, or any other device, the release of VAMs would not serve as much of a shock for teachers about whom there already exists a great deal of information. The model predicts that if there is substantial public learning prior to VAM adoption, the effects of the policy should be less dramatic for more experienced teachers. While Table A.6 exhibits this relationship for teachers moving out of the district, the same is not true for teachers moving within district. Taking the point estimates literally, a teacher with 5 more years of experience and one standard deviation higher VAM is twice as likely to move within Guilford to a better school after the release of VAM, than is a less experienced, but otherwise similar teacher. In Winston-Salem, the estimates on this triple interaction are too noisy to draw reliable inference. While the observed pattern of stronger effects for more experienced teachers may seem strange, this pattern may occur if it takes time to realize that moving is worthwhile or if releasing VAMs allow a built up stock of more

[28] In unreported regressions, with the exception of out-of-Guilford moves the results shown in Table A.5 are very sensitive to the variable composition of the teacher quality index. Table F.8 in Appendix **??** demonstrates that these results are also sensitive to the covariates included in the index. The regressions in Table F.8 includes measures of quality in the index of teacher quality, since it is likely that other principals use sending-school quality as an important signal of the teacher's quality. In which case, percent of students at current school who are on grade level and who are Black are reasonable to include in the index. In Table F.8, the coefficient estimates on each of the interaction terms, which are of primary interest, carry the predicted sign. However, the coefficient estimates on the index for the rest of the state have the opposite sign as predicted. This inconsistency is likely due to current school quality affecting the probability both through teachers' willingness to move as well as principals' willingness to hire them. It remains noteworthy that teachers in good school with other high observables, are even less likely to move within district after the district adopts VAMs.

experienced teachers who could not previously signal their quality to move. From columns 3 and 4, in both districts, each additional year of experience mitigates the negative selection of inexperienced teachers moving out of the district. For Guilford and Winston-Salem, 5 years of additional experience cuts the effect of VAM on the probability of moving to a better school outside the district by 15% and 20%, respectively. The same general pattern holds with regard to interactions with tenure, though the standard errors on the coefficient estimates for interactions with tenure are larger. Were private learning already prevalent in the market, the model predicts the effects of the policy to be larger for those who have taught at the same school for longer, all else being equal. This is consistent with the results in columns 1 and 2. While these results largely suggest prior private learning, the mixed evidence on public learning makes me hesitant to draw definitive conclusions on the prior learning environment.

1.7 Robustness

In the following section, I examine the robustness of the effects of VAM adoption. Section 1.7.1 considers changes in effects when using only prior years of student data when constructing VAMs. Section 1.7.2 considers whether other district policies that paid teachers to work in hard-to-staff schools impact the estimated effects. Appendix E includes sevel additional robustness exercises including consideration of teacher mobility in accordance with the state ABC growth bonus-pay system; within-district, year-by-year analysis of the changing effects of VAMs on mobility and sorting; and consideration of alternate functional forms for the mobility analysis, such as normal Maximum Likelihood Estimationas well as competing risks regression to examine the possibility of correlated errors between types of moves. [29]

1.7.1 Sensitivity to VAM Construction

The possibility that teachers may have different VAMs after moving to other schools, may present issues for using VAMs constructed from student data from a teacher's entire career. This could result from moves leading to higher match quality between teachers and schools, as Jackson (2013) finds. It may also result from transitory adjustment costs, giving a theoretically ambiguous direction of potential bias.[30]

Consequently, in Table A.7, I allow teachers VAM scores to vary each year, using only data from the current and previous years to construct a teacher's VAM in any given year. The main effects hold, though they are in general somewhat exaggerated in Winston-Salem and smaller in Guilford. Still, the adoption of VAMs raises the probability that good teachers move to better schools. Whereas in Winston-Salem, the effect grows to a full percentage point, in Guilford, a teacher with an one standard deviation higher VAM becomes 0.3 percentage points more likely to move to better school post-policy. From the middle column of Panel B, the negative selection of teachers moving out of Guilford falls to just 30% of the estimate given in Table A.3. Panel C in Table A.7 corresponds with Table A.4. While the effect on teacher sorting doubles in Winston-Salem, the results become more negative and statistically insignificant in Guilford.

While it is possible subsequent match quality increases for teachers from Guilford and

[29] Because job mobility is often localized I also restricted analysis to districts which share a border with Guilford and Winston-Salem. The results from this restriction were noisy and uninformative, and are unreported here.

[30] More closely approximating the information that teachers and principals receive is another rationale for restricting the data used in generating teacher VAMs. In which case using Empirical Bayes estimation provides what is believed to be a closer approximation to the algorithm used in creating the EVAAS measures. Table F.10 in Appendix **??** provides results using Empirical Bayes estimation on the restricted sample of student test scores in calculating teacher VAMs. The results are very similar.

decreases for teachers in Winston-Salem, I believe measurement error may provide a more plausible explanation. In Guilford, the effect of VAM prior to the their release is identified off of just two years of data. As a result, the estimates of teachers VAMs are noisier for this period as well as in the immediate aftermath of the policy. Measurement error in the primary variable of interest may attenuate the estimates in Guilford where there is little data prior to the adoption of the policy, while the effects in Winston-Salem become relatively stronger.

One way of getting around this issue is to use a fixed number of years prior to the current period when constructing VAMs. Unfortunately, the adoption of VAMs by Guilford comes just three years into the student data sample. Since the construction of VAMs requires at least one prior year of student data, this gives just two years at which I could fix my VAM estimate. Not only would this force a noisier estimate of each teacher's VAM for the entire sample, it also provides merely one year of data prior to the adoption of the policy in Guilford. To demonstrate the changes of the estimates with varying the number of years of data used in constructing VAMs, I drop Guilford from the analysis and vary the number of prior years of data I use to construct the VAMs from 2 to 8. Table A.8 demonstrates that though the relationship between years used and the effect of the interaction of the policy in Winston-Salem and VAM is not monotonic as the sample used varies, the estimates using more years of data are clearly the largest. This further suggests correlated measurement error presents a problem for this approach.

1.7.2 Strategic Staffing

A possible complication arises due to alternate teacher compenstion plans. District strategic staffing policies, which aim to attract more capable teachers to teach in and stay at hard-to-staff schools may be problematic because they occured in treatment districts during the sample period and could potentially alter teacher preferences over schools.[31] Charlotte-Mecklenburg Schools (CMS) and Winston-Salem were by far the earliest adopters of these initiatives with CMS beginning its Equity Plus program in 1999 and Winston-Salem following suit in 2000. By 2012 each major district in North Carolina adopted some program to attract teachers to hard-to-staff schools. In CMS, teachers received a signing bonus to enter a targeted school and teachers with a masters degree could receive up to $2,500 per year to remain in the school. A smaller incentive was offered to teachers enrolled in masters programs, though the district also offered tuition reimbursement. Winston-Salem awarded 20% of the district salary supplement ($500-$1,500) to each teacher in targeted schools. Furthermore the entire state offered $1,800 bonuses to math, science, and special education teachers who taught in high poverty or low achieving schools during the three year period 2002-2004. In 2007, Guilford adopted its own strategic staffing program, in which bonuses ranged from $5,000-$25,500 depending on subject taught, grade level, and VAM. Cumberland County Schools gave stipends to 30 "master teachers" across their 10 most difficult school. In 2008, CMS began tailoring their plan more towards targeting better teachers and Winston-Salem, followed suit in 2012. These programs may reverse which schools are most desirable to teachers. With large enough incentives, high-VAM teachers may opt to work at low performing school, which is in fact the intent of the policy.

Table A.9 reports similar information as is provided in Table A.3, with the difference that the binary dependent variable in Table A.9 is equal to one if a move occurs and the receiving school is not classified as strategic staffing. As might be expected, the results are quite similar to those in Table A.3, as teachers working in strategic staffing schools comprise just

[31] "Strategic Staffing" is the official term for later policies with the same objectives. Earlier policies had a variety of different names; Equity Plus (1 and 2), Focus School, and Mission Possible.

4% of the sample. However, the policy has a much larger effect on the correlation between VAMs and the probability of moving within Winston-Salem. Column 2 shows that releasing VAMs raises the probability that a teacher with one standard deviation higher VAM will move within Winston-Salem by a full percentage point, which is nearly double the effect found when examining all schools together. Also, the effect of the policy on the correlation between VAMs and the probability of moving out of Winston-Salem drops by 40%, when restricting analysis to moves to non-strategic staffing schools. Both changes serve to widen the gap in the estimates between moves within and out of Winston-Salem, providing further evidence of private learning.

Table A.10 presents the impacts of the policy on teacher sorting within-district and within-district among non-strategic staffing schools. Column 1 in Table A.10 is identical to column 2 in Table A.4. I include it here for ease of comparison. The third columns restrict the sample further to non-strategic staffing schools. Moving from column 1 to 2, in both districts, the point estimated effect of the policy on the degree to which high-VAM teachers sort into high performing schools becomes more positive. For Guilford, the coefficient becomes positive, though neither practically nor statistically significantly so. In Winston-Salem, the point estimate of the sorting effects more than triple. Table A.10 provides no evidence that strategic staffing policies are driving the earlier results. If anything, it seems that these pay policies may have muted what would otherwise have been much larger impacts of releasing VAMs.

1.8 Conclusion

If employers are unable to learn accurate information about their teaching force over time, their subsequent personnel decisions regarding teachers would be no better at identifying effective teachers than at the point of hire. If learning is entirely asymmetric, that is other schools are no better able to tell the effectiveness of an experienced applicant than of a novice applicant, effective teachers become trapped in schools in which they do not wish to teach, while principals shuffle their less capable teachers to other schools in what the documentary Waiting for Superman terms "The Lemon Dance" Guggenheim (2011). The release of value-added measures of teacher effectiveness does seem to provide actionable information to those who are aware of them. The evidence above suggests that the new information provides effective teachers with more mobility, while "The Lemon Dance" becomes focused on the uninformed.

Additionally, the evidence from subsequent teacher sorting suggests that the increase in mobility leads to increased inequity in the distribution of teacher quality across schools. Despite the fact that 38 states have adopted VAMs of teacher effectiveness, and often contentiously, this signaling role of the measures has avoided discussion. The policy implication of this finding is not to universally avoid using VAMs. However, it would be useful to provide policy makers an estimate of the cost of retaining high-VAM teachers in hard-to-staff schools. The analysis excluding strategic staffing schools implies that the sorting may have been larger without the incentives to induce teachers to work in lower-performing schools. As mentioned in Section 1.7.2, several districts in North Carolina are implementing a range of staffing policies designed to induce teachers to work in low-performing schools. Some incorporate VAMs into the incentive schemes.

Clotfelter et al. (2011) and Glazerman et al. (2012) have examined the question of attracting teachers to understaffed schools. Further work is needed to estimate the costs and effectiveness of these policies in retaining effective teachers in low-performing schools, which may cost substantially less. As states and districts continue to adopt teacher VAMs, policy makers should be aware of the potential consequences of these policies on educational equity, as well as the costs of offsetting these effects.

Chapter 2

Job Separation Under Asymmetric Employer Learning

2.1 Introduction

Employers gamble with each new hire, perhaps particularly so as applicants begin their careers. There is significant uncertainty about how reliable, hard working, and capable new hires will be. However, after workers have been with a particular firm for some time, it makes sense that employers accumulate information about these traits and how to use them. At the same time, some workers may draw on their acquired experience to bolster their resumés, while continually looking for other opportunities. Despite the intuitive appeal of this setting, the question of whether all firms in the market learn about the workers at the same rate (symmetric learning) or whether current employers enjoy the benefits of private information about their workers (asymmetric learning) is empirically largely unsettled.

There are several implications stemming from the process by which current and prospective employers learn. First,information gaps between firms may permit persistent wage gaps based on easily observable characteristics, such as race and education, even when workers are equally productive Schönberg (2007); Pinkston (2009). Secondly, the preservation of informational advantage may distort promotion decisions DeVaro and Waldman (2012). Lastly, the disparate types of learning have different predictions for worker mobility. This study focuses on these implications.

Under symmetric learning, workers' reference groups should have no bearing on the probability they experience job separations, since all firms should place the same importance on the workers' characteristics that they can easily observe. Likewise, gaps in information cannot explain worker mobility on the basis of hard to observe characteristics, since each employer is equally well informed under this framework.

In contrast, under asymmetric learning, information about a worker's true productivity is more valuable the further that true ability is from the average worker with the same easily observable characteristics. This heterogeneity in the value of information about a given worker's productivity leads different workers to have different probabilities of leaving. Workers who have higher ability than the average of their respective reference groups will be less likely to be bid away by another firm or laid off by the retaining firm. Inversely, those with high ability reference groups are more attractive to outside firms. By examining how ability and reference groups influence the probability of job separations, this study explores evidence of the presence of asymmetric employer learning.

I extend existing models of asymmetric employer learning to develop predictions regarding the selection of workers into job switches and layoffs. The predicted selection differs depending on whether worker covariates are easy or difficult to observe. I then develop predictions for how this selection will change over experience as opposed to continuous working spells. I test these predictions using the Armed Forces Qualification Test (AFQT) from the National Longitudinal Survey of Youth of 1979 (NLSY79) as a hard to observe worker characteristic and race and education as important characteristics that are easy to observe. Finally, I test the robustness of results under alternate specifications and use a control function approach to handle possible endogeneity in experience and continuous working spells.

I find that in general, consistent with asymmetric employer learning, workers are negatively selected into mobility on the basis of their AFQT score, while they are positively selected on the basis of their reference groups. This selection on AFQT is driven completely by job-to-unemployment transitions, whereas asymmetric learning predicts such negative selection through both job-to-job and job-to-unemployment moves. Though the interactions of AFQT with experience and working spell duration carry the signs predicted by asymmetric learning, they are statistically insignificant. The results regarding education are more consistent with asymmetric employer learning. Workers with higher and more selective education are more likely to both transition between jobs, as well as for moves from employment to unemployment during economic recessions. Further, the selection on the basis of education becomes more positive with respect experience and more negative with respect to working spell duration, as asymmetric learning predicts.

The rest of the study proceeds as follows. Section 2 situates this work in the existing literature. Section 3 lays out my extension of Pinkston's [2009] model of asymmetric employer learning. Section 4 describes the estimation strategies used. Section 5 describes the data, provides definitions of variables, and descriptive statistics. Section 6 provides empirical results, and Section 7 summaries the evidence.

2.2 Employer Learning Models and Evidence

This work follows a long line of predecessors, such as Farber and Gibbons (1996); Altonji and Pierret (2001); Lange (2007), Schönberg [2007], and Pinkston (2009), which examines how employers learn about their employees over time. Starting with Farber and Gibbons

[1996], these studies presume that worker ability is heterogeneous. Employers cannot directly observe the ability of potential workers and must rely on correlates to infer workers' expected value to the firm. Further, they treat a subset of characteristics as easily observable to all, another as easily observable to the market (and not to researchers), and yet another subset of potential correlates with productivity as easily observable to the econometricians (but not the market). This literature typically uses the percentile from a cognitive ability assessment, the AFQT, as this relatively strong correlate with productivity that is veiled to the the market at the time of hire, but is available to researchers throughout the workers' careers. This study does so as well.

Much of this earlier work assumes that all firms within the market learn at the same rate. In which case, all prospective employers learn more about workers' abilities over time and wages become more strongly linked to previously unobserved, strong correlates with productivity and less tied to the easily observed characteristics. Farber and Gibbons [1996], Altonji and Pierret [2001], and Lange [2007] each find this pattern regarding wages, education, and AFQT scores.

However, it seems reasonable that employers may learn more about their employees than do outside firms. Indeed, both Gibbons and Katz (1991b) and DeVaro and Waldman (2012) provide evidence of such informational asymmetries between firms though neither is couched in a leaning framework. Schönberg (2007); Pinkston (2009), and Kahn (2013) develop models of employer learning that allow employers to learn at different rates. They test these models for the presence of asymmetric learning in the labor market using the NLSY79, and find conflicting evidence.

Schönberg develops an initial model for asymmetric employer learning, and develops tests for its hypotheses. Her model assumes that the retaining firm perfectly observes the productivity of its workers in the second period under asymmetric learning, but this information is lost whenever a job match terminates. Perfect learning in the second period implies that there are no informational benefits from additional tenure. This is a testable assertion, and Lange (2007) presents evidence of continuous learning, finding that while about half the learning occurs in the first three years, 20 years later the variance of the error continues to decline. Further, a winner's curse befalls outside firms as they draw the least productive workers from their retaining firms. Allowing differences in match quality to motivate more productive workers to leave attenuates the severity of the winner's curse, and theoretically leads to volatile wages early in a job match.

In accordance with her model, Schönberg finds little evidence of asymmetric learning. Her empirical work focuses only on white male workers, abstracting away from the implications of employer learning for statistical discrimination. She finds that only white college educated job leavers are negatively selected on the basis of AFQT score, and that adverse selection is driven by job-to-unemployment transitions, which she contends weakens the case for asymmetric employer learning. Furthermore, Schönberg finds that the impact of schooling on wages decreases with experience and remains relatively constant with tenure. Consequently, she concludes that learning in the market is largely symmetric.

Kahn (2013) extends Schönberg's framework to test whether job movers experience more volatile wage patterns after a transition than do those who remain in place. Using the NLSY, she considers differences between workers who enter a position during recessions as opposed to economic expansions, with the idea that there is less variation in the ability of entrants during recessions. She also uses variation in the amount of exposure an occupation has outside the firm, assuming that learning is more symmetric in more exposed occupations. Kahn finds that movers' wages are more volatile in the immediate aftermath of a transition than are the wages of those who remain in place. Also, the effects are larger for those who enter a job during an economic expansion and for those in more insular occupations. These features, she argues, are supportive of the asymmetric learning hypothesis. Perhaps more applicable to this work, Kahn finds evidence of adverse selection on the basis of AFQT and years of education of movers in general and finds these effects are stronger for occupations that communicate less with the outside world. Again, she argues this evidence supports the asymmetric learning model.

The alternate model Pinkston [2009] develops allows the precision of the signal that current employers receive to increase with time, rather than imposing a discreet jump. He also allows race to influence the impact of ability on workers' wage progressions. Furthermore, Pinkston allows information about workers' productivity to be transmitted from incumbent employers to new employers during job-to-job transitions. Consequently, he examines wage dynamics over spells of continuous working as opposed to tenure, which Schönberg uses. Additionally, Pinkston notes, as has earlier research, that length of working spell and experience may contain information about the quality of the job match and the productivity of the worker. To address this issue, he uses the residuals from the regression of working spell length on workers' career-average spell length, total duration of the current job, and indicators for missing values of duration, as instrumental variables for working spell, and uses the residuals from the regression of experience on potential experience to instrument for current experience. Pinkston finds that AFQT becomes a larger factor in wage determination as working spell increases while the impact of race and education diminish with working spell. Both features are suggestive of asymmetric information.

Since Pinkston [2009] considers the implication of asymmetric learning on wage development, this study provides a natural extension of his model to examine evidence regarding job separation. Whereas Pinkston imposes an exogenous rate of job destruction in his model for asymmetric learning, this work models and tests the implications of asymmetric learning on the margin of job switches and layoffs.

Though Schönberg, [2007] also addresses job separation in her study of asymmetric employer learning, this work pushes past her analysis in a few key areas. Firstly, the model developed here respects the gradual nature of the learning process, and allows information to accumulate throughout job-to-job transitions. This flexibility allows me to formulate theoretical predictions and empirical tests of the dynamics of worker selection on the basis of easy and difficult-to-observe characteristics with respect to experience and working spell length. Secondly, while we both examine the implications of hard to observe characteristics on mobility, I also develop and test predictions regarding the impact of the workers' reference groups, conditional on individual AFQT scores, on the probability of job separation. Furthermore, Farber and Gibbons (1996); Altonji and Pierret (2001); Lange (2007), Schönberg, [2007] and Pinkston (2009) each use years of education to proxy for easily observable information. However, there is wide variation quality of those years of education. Which college an individual attends may provide more information about that person's cognitive ability than the quantity of education obtained. Consequently, I incorporate college selectivity into my analysis to construct more robust reference groups. Lastly, I provide a theoretical rationale for asymmetric information to impact selection in layoffs, and examine whether the impact of AFQT score and reference group membership on job-to-unemployment transitions differs between different economic conditions.

Table B.1 illustrates this basic story. In general, those in panel B with lower AFQT scores than the average person with same educational attainment at the same quality of institution, experience a higher rate of job separation than those in panel A whose AFQT scores are higher than their reference groups' averages. This is true across reference groups and separation types, with the exception of the less educated in job-to-job transitions.

Restricting attention to educational attainment doesn't tell a consistent learning story. College graduates generally enjoy lower mobility rates than college attendees. Also, those with high school diplomas and relatively low AFQT scores are more likely to enter unemployment, than are college attendees with lower AFQT scores than the average college attendee. However, there are significant differences in average AFQT across levels of education that this table does not take into account.

Despite differences in average AFQT between those who attend competitive college as opposed to noncompetitive schools, differences in mobility between these two groups is completely consistent with asymmetric employer learning. Across all categories, those who attend competitive colleges are both more likely to switch firms or move from employment to unemployment than those who attend less competitive institutions. This surprising result fits perfectly in line with asymmetric employer learning. As Section 3 shows in more detail, under asymmetric employer learning, outside firms place more emphasis on this public signal of worker ability than do retaining firms. This raises the outside offers for workers who attend and graduate from competitive colleges, thus raising the probability that they will be bid away to an outside firm. Further, higher outside offers drive up the wages of workers with relatively high public signals making their wages closer to their expected productivities. Thus, there little room to buffer against economic downturns, and in the event of recessions, employers let go of these workers.

2.3 Model

This section develops a model to demonstrate the implications of employer learning on the selection of workers into mobility. It builds largely upon the model presented in Pinkston [2009], which in turn builds upon models presented in Farber and Gibbons (1996) and Altonji and Pierret (2001). Please see the Appendix for proofs of predictions.

2.3.1 Framework

Employers care about the productivity of their workers, which is composed of the worker's underlying ability (μ) and the quality of the match between a given firm and worker (φ_f). Each period, employers learn workers' fixed characteristics (m), a public signal (R_x where x indexes experience), a private signal (P_f where f may stand for either the retaining firm (r) or the hiring firm (h)), and the expected match quality for that period ($E(\varphi_f)$).

Before receiving a private signal, all firms share a prior belief that a worker's expected ability equals the average ability (m) of other workers with the same easily observable characteristics. The public signal is analogous to a resumé, while the private signal is informed initially by an interview and later by the daily activities of the employee during work hours. Thus, the private signal to the outside firm (P_h) is not subset of the information the retaining firm observes (P_r), which means that hiring firms can profitably compete for workers against better informed retaining firms. With each signal, employers or perspective employers update their prior beliefs according to the value and precision of the signals they receive.

To allow the nesting of symmetric learning, workers and firms must learn about each possible match quality component equally, whether or not the worker is currently working for a given employer. When discussing bidding, I will use $E(\varphi_f)$, because φ_f may change over time within a match according to the economic climate. Firms and workers have an expectation over φ_f, which is specific to the particular match, but after the bidding is settled and wages set, the surrounding economic conditions are realized, and φ_f adjusts. This timing is important when modeling layoffs under downward wage rigidity.

When workers first enter the market, they receive two offers and go to the highest bidder. Each subsequent period, they continue to receive two offers, one from their retaining firm and one from an outside firm. The firms then bid on the worker as in a standard English auction. The firm with the highest bid gets the worker and pays him the highest bid of the rival. The following assumptions provide more structure to the described model:

1. Unobserved ability of the worker, $\mu = m + \epsilon$, where $\epsilon \sim N(0, \sigma_\epsilon)$.

2. The private signal, $P_r = \mu + \tau$ where $\tau \sim N(0,\sigma_\mathcal{T}(t))$, and $\frac{\partial \sigma_\mathcal{T}(t)}{\partial t} < 0$. t indexes tenure throughout the period of continuous employment. For hiring employers $t = 0$, and for current employers $t > 0$. $\sigma_\mathcal{T}(t) < \sigma_\mathcal{T}(0)$ for all $t > 0$.

3. The public signal, $R_x = \mu + \xi$, where $\xi \sim N(0, \sigma_\xi(x))$, and $\frac{\partial \sigma_\xi(x)}{\partial x} < 0$. Here x indexes experience.

4. Unobserved productivity, $\rho = \mu + \varphi$, where $\varphi \sim N(0, \sigma_\varphi)$ is the match quality between the worker and the firm. Firms' and workers' expectation of match quality during the bidding process $E(\varphi) \sim N(\varphi, \sigma_{E\varphi})$.

5. Errors are orthogonal to one another.

The optimal bids are precision weighted averages of the signals employers receive and the expected ability of workers with the same education and of the same race. In this model, the difference between the current and outside employer is that the current employer receives a more precise signal of their workers' productivity than do outside firms ($\sigma_T(t) < \sigma_T(0)$ for all $t > 0$). Assuming continuousness in the bidding process, the openness of an English Auction allows each firm to learn that the other firm values the worker at least as much as it does during the auction. Thus, the private signal receives double weight. Letting $W = \sigma_\xi(x)\sigma_T(0) + \sigma_T(0)\sigma_\epsilon + 2\sigma_\xi(x)\sigma_\epsilon$ and $W' = \sigma_\xi(x)\sigma_T(t) + \sigma_T(t)\sigma_\epsilon + 2\sigma_\xi(x)\sigma_\epsilon$, from Milgrom and Weber (1982), the retaining firm's optimal bid, (b_r), and the outside firm's optimal bid, (b_h), are given below:

$$b_r = E[\mu|R_x,\ P_r,\ P_h = P_r] = \frac{\sigma_\xi(x)\sigma_T(t)}{W'}m + \frac{\sigma_T(t)\sigma_\epsilon}{W'}R_x + \frac{2\sigma_\xi(x)\sigma_\epsilon}{W'}P_r + E(\varphi_r) \tag{2.1}$$

$$b_h = E[\mu|R_x,\ P_r = P_h,\ P_h] = \frac{\sigma_\xi(x)\sigma_T(0)}{W}m + \frac{\sigma_T(0)\sigma_\epsilon}{W}R_x + \frac{2\sigma_\xi(x)\sigma_\epsilon}{W}P_h + E(\varphi_h) \tag{2.2}$$

Notice that if the retaining firm's signal is more precise than that of the hiring firm (if $\sigma_T(t){<}\sigma_T(0)$), the retaining firm places relatively less weight on the reference group and public information and relatively more weight on the private signal. Further, as the public signals become more precise, the hiring firm places less emphasis on their private signal in favor of the public information. Thus, if both firms receive the same private signal, the differences in weighting will likely lead to different optimal bids.

It is important to note that because the valuation of the losing firm is revealed during the open auction, in the event that the outside firm wins the auction, it also captures the retaining firm's private valuation of the worker. Consequently, information accumulates throughout job-to-job transitions rather than resetting with each new employment spell. In this way, the information acts less as specific human capital, as in Becker (1962), and is more analogous to general human capital. However, if the worker is forced to endure a period of unemployment between spells, the market loses the accumulated private information.

Under symmetric learning, the optimal bids take a very similar form, though in this special case all signals are public. The optimal bid of the retaining firm is the weighted average of the prior belief and the public signal of the worker's ability, plus the expected quality of the match. Again, employers weigh the signal and prior belief in accordance to the relative precision of each. Thus, the optimal bids of the retaining firm (b_r) and outside firm (b_h) are respectively shown in equations 2.3 and 2.4:

$$b_r = E[\mu|R_x] + \varphi_r = \frac{\sigma_\xi(x)}{\sigma_\xi(x)+\sigma_\epsilon}m + \frac{\sigma_\epsilon}{\sigma_\xi(x)+\sigma_\epsilon}R_x + E(\varphi_r) \tag{2.3}$$

$$b_h = E[\mu|R_x] + \varphi_h = \frac{\sigma_\xi(x)}{\sigma_\xi(x)+\sigma_\epsilon}m + \frac{\sigma_\epsilon}{\sigma_\xi(x)+\sigma_\epsilon}R_x + E(\varphi_h) \tag{2.4}$$

Notice that if employers learn about workers' true ability over time, the variance of the public signal ($\sigma_\xi(x)$) decreases. Thus, employers place less weight on their prior belief and more weight on the public signal.

2.3.2 Job Switches

Assuming each firm plays its optimal strategy, the probability that a worker switches firms ($P(J)$) is equal to the probability that the outside firm has a higher optimal bid than the retaining firm.[1] The differences between firms in the precision and weighting of private information provides clear predictions of selection on the basis of both hard and easy-to-observe worker characteristics. After some algebra, and allowing ψ_J to stand for the composite error term (including the difference in match quality), the difference between the hiring and retaining firms' optimal bids can be written as:[2] [3]

$$P(J) = P[b_h - b_r > 0] = P\left\{\psi_J < \sigma_\xi(x)[\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t)](m - \mu)\right\} \tag{2.5}$$

Because a current employer has a clearer view of a worker's underlying ability, the retaining firm places more emphasis on it than do other firms. Therefore, even if the current and prospective employers receive equivalent relatively high private signals, the weighting will lead the current employer to have a higher optimal bid. Incorporating the normality and orthogonality assumptions above and allowing σ_{ψ_j} to stand for the variance of ψ_J, the derivative of equation 2.5 with respect to ability (μ) provides the following:[4]

[1] J references job-to-job transitions.

[2] $\psi_J \equiv E(\varphi_r) - E(\varphi_h) - \frac{2}{WW'}((\sigma_\epsilon\sigma_\mathcal{T}(t) + \sigma_\xi(x)\sigma_\mathcal{T}(t) + 2\sigma_\epsilon\sigma_\xi(x))\sigma_\epsilon\sigma_\xi(x)\tau_h - (\sigma_\epsilon\sigma_\mathcal{T}(0) + \sigma_\xi(x)\sigma_\mathcal{T}(0) + 2\sigma_\epsilon\sigma_\xi(x))\sigma_\epsilon\sigma_\xi(x)\tau_r + \sigma_\epsilon^2\sigma_\xi(x)(\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t))\xi)$.

[3] Please see Appendix H for algebra.

[4] $\sigma_{\psi_J} = var(\psi_J) = 2\sigma_{E\varphi} + \frac{4}{W^2W'^2}(W'^2\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\mathcal{T}(0) + W^2\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\mathcal{T}(t) + \sigma_\epsilon^4\sigma_\xi(x)^2(\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t))^2\sigma_\xi(x))$, and $\sigma_{E\varphi} = var(E(\varphi_r)) = var(E(\varphi_h))$.

$$\frac{\partial P(J)}{\partial \mu} = -\phi \left\{ \frac{2\sigma_{\xi}(x)\sigma_{\epsilon}}{WW'\sqrt{\sigma_{\psi_j}}} \sigma_{\xi}(x)[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)](m - \mu) \right\} \frac{2\sigma_{\xi}(x)\sigma_{\epsilon}}{WW'\sqrt{\sigma_{\psi_j}}} \sigma_{\xi}(x)[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)] < 0 \qquad (2.6)$$

$\phi\{.\}$, being the normal probability density function, is positive, as is each variance. Thus, as long as the precision of the private signal shrinks the longer a worker is with the retaining firm ($\sigma_{\mathcal{T}}(0) > \sigma_{\mathcal{T}}(t)$), which is fundamental to asymmetric employer learning, equation 2.6 shows that as ability (μ) increases the probability of a move decreases, all else equal.

This flexible learning model allows for further predictions about the evolution of this selection over time, specifically with regard to increases in length of continuous working spells as opposed to experience in the market. Intuitively, it makes sense that selection on the basis of ability would most pronounced when there are the greatest asymmetries in information between employers. This occurs when a worker has been continuously working for a long period of time, and information has accumulated with one employer and/or transferred to another through the bidding process associated with a job-to-job move. More formally, the cross-partial of $\frac{2\sigma_{\xi}(x)\sigma_{\epsilon}}{WW'\sqrt{\sigma_{\psi_j}}} \sigma_{\xi}(x)[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)](m - \mu)$ with respect to working spell length (t) and ability (μ) is negative.[5] Inversely, these effects are smaller when there are small asymmetries, such as when a worker has sufficient experience in the market for his ability to be apparent to all prospective employers. Thus, the cross-partial of $\frac{2\sigma_{\xi}(x)\sigma_{\epsilon}}{WW'\sqrt{\sigma_{\psi_j}}} \sigma_{\xi}(x)[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)](m - \mu)$ with respect to experience (x) and ability (μ) is positive.[6]

While Schönberg (2007) notes the predicted adverse selection on the basis of underlying ability for job switches, this work introduces an examination of selection into mobility on

[5] Please see Appendix A for proof. Note that this is not the cross partial of the probability of a job-to-job move, but is rather the scaled regression coefficient on the interaction between ability and working spell length.

[6] Please see Appendix A for proof.

the basis of reference group or easy to observe worker characteristics. The derivative of equation 2.5 with respect to average ability of the reference group is unsurprisingly nearly identical to equation 2.6, it simply has the opposite sign.

$$\frac{\partial P(J)}{\partial m} = \phi\left\{\frac{2\sigma_{\xi}(x)\sigma_{\epsilon}}{WW'\sqrt{\sigma_{\psi_j}}}\sigma_{\xi}(x)[\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)](m-\mu)\right\}\frac{2\sigma_{\xi}(x)\sigma_{\epsilon}}{WW'\sqrt{\sigma_{\psi_j}}}\sigma_{\xi}(x)[\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)] > 0 \tag{2.7}$$

From equation 2.7, conditional on individual ability, as the average of ability among workers in the same reference group (m) increases, the probability of a job-to-job move should increase. This is due to prospective employers applying more weight to public information than does the current employer. I will not duplicate the above close examination of the dynamics of this selection on the basis of reference group with respect to working spell duration and experience, due to the closeness of equations 2.6 and 2.7. The inverse of the dynamics with regard to experience and working spell duration is also true. As working spell length increases workers with high reference groups become even more likely to switch employers.[7] Conversely, with increases in experience, those with high reference groups become less likely to transfer firms.[8]

Under symmetric learning, the probability of a job switch ($P(J)$) is again the probability the outside firm has a higher optimal bid than the retaining firm. The primary difference here is that there are no private signals. Thus the difference in bids shown in equation 2.8 simplifies to equation 2.9.

$$P\left\{\frac{\sigma_{\xi}(x)}{\sigma_{\xi}(x)+\sigma_{\epsilon}}m+\frac{\sigma_{\epsilon}}{\sigma_{\xi}(x)+\sigma_{\epsilon}}R_x+E(\varphi_h)-\left[\frac{\sigma_{\xi}(x)}{\sigma_{\xi}(x)+\sigma_{\epsilon}}m+\frac{\sigma_{\epsilon}}{\sigma_{\xi}(x)+\sigma_{\epsilon}}R_x+E(\varphi_r)\right]>0\right\} \tag{2.8}$$

[7] See Appendix A

[8] See Appendix A for proof.

$$P(J) = P\left[E(\varphi_h - \varphi_r) > 0\right] \tag{2.9}$$

Notice that both individual ability and the average ability of the reference group are eliminated from the equation, since the market uniformly weights the easy and difficult-to-observe information.

In order for selection of job switchers to persist under symmetric learning, change in match quality must differ with ability.[9] Given the finding in Kahn (2013) among others, that job switching leads to large wage gains for young workers, job switching may be a desirable outcome, at which high ability workers may be more adept. In which case, positive selection on the basis of ability may be expected in under symmetric employer learning. This same positive correlation between worker ability and difference in match quality would produce ambiguity in the predictions regarding ability and the probability of job-to-job transitions under asymmetric employer learning.

2.3.3 Layoffs

Asymmetric employer information may also provide meaningful predictions regarding the probability of layoffs, as in Gibbons and Katz (1991b). Whereas Gibbons and Katz (1991b) examine wage penalties of layoffs as opposed to plant closings, this study provides further rationale for their findings by examining the easy and difficult-to-observe worker characteristics of those who are laid off.

A voluntary move from employment to unemployment is very different from the same

[9] This would be a violation of assumption 5 above.

move were it unilaterally decided by the employer. There are no immediate predictions from either employer learning model considering voluntary moves into unemployment, but there are different implications of asymmetric and symmetric learning regarding layoffs. Unfortunately for researchers, it is often difficult to discern quits from layoffs. It is helpful to decompose the probability of a job-to-unemployment transition, $P(JU)$, into the probability of a layoff, $P(L)$, and the probability of a quit $P(Q)$. During recessions, Davis et al. (2006) and Elsby et al. (2009) each show an increased inflow of workers into unemployment during recessions, which is driven by an increase in layoffs large enough to dominates a decrease in the number of quits. Comparing the magnitude of selection effects on the probability of a job-to-unemployment separations between economic recessions and expansions, provides insight into how selection differs for layoffs and quits.

There is also no broadly accepted theoretical justification for layoffs. It seems that there exists a range of lower wages in which workers would prefer to work until they could move to another firm at a higher wage rather than enduring a period of unemployment. Firms, it seems, should prefer keeping a worker as long as the wage is less than the worker's productivity. Even with relatively large economic fluctuations it seems there is likely to be overlap.

However, we observe nominal downward wage rigidity during economic downturns. Campbell and Kamlani (1997) report that human resources personnel most commonly list fear of the most productive workers leaving as their primary motivation for using layoffs rather than wage reductions. This begs the question, why would firms care which workers left, if each worker is paid their marginal product of labor? Asymmetric information provides one such rationale.

In the model presented above, the expected marginal profit from a given worker is his conditional expected productivity, net of his wage. In the case of symmetric learning, this is the difference between the expected match quality at the retaining firm and the expected match quality at the firm with the next highest bid.[10]

Under asymmetric learning, firms will not necessarily keep their most productive workers, but rather their most profitable workers; those who outperform their observable characteristics. The expected productivity is $E[\mu|S_x, S_f, v_h]$ whereas the wage is, $w = min\{E[\mu|R_x, P_r, P_h = P_r], E[\mu|R_x, P_r = P_h, P_h]\}$. Consequently, expected profits on a given worker are given below.

$$\begin{aligned} E[\pi|R_x, P_r, P_h] =& \frac{\sigma_\tau(0)\sigma_\tau(t)\sigma_\xi(x)}{Q} m + \frac{\sigma_\tau(t)\sigma_\xi(x)\sigma_\epsilon}{Q} P_h + \frac{\sigma_\tau(0)\sigma_\xi(x)\sigma_\epsilon}{Q} P_r + \frac{\sigma_\tau(t)\sigma_\tau(0)\sigma_\epsilon}{Q} R_x \\ &+ E(\varphi_r) - \left(\frac{\sigma_\tau(0)\sigma_\xi(x)}{Q'} m + \frac{\sigma_\tau(0)\sigma_\epsilon}{Q'} R_x + \frac{2\sigma_\epsilon\sigma_\xi(x)}{Q'} P_h + E(\varphi_h) \right) \end{aligned} \tag{2.10}$$

where $Q = \sigma_\tau(0)\sigma_\tau(t)\sigma_\xi(x) + \sigma_\tau(t)\sigma_\xi(x)\sigma_\epsilon + \sigma_\tau(0)\sigma_\xi(x)\sigma_\epsilon + \sigma_\tau(t)\sigma_\xi(x)\sigma_\tau(0)$ and $Q' = \sigma_\tau(0)\sigma_\xi(x) + \sigma_\tau(0)\sigma_\epsilon + 2\sigma_\epsilon\sigma_\xi(x)$. In expectation, the errors are zero leaving the simpler equation 2.11, with a very similar form to equation 2.5.[11]

$$E\left\{E\left[\pi|R_x, P_r, P_h\right]\right\} = \frac{\sigma_\xi(x)\sigma_\epsilon}{Q'Q} \left[\sigma_\xi(x)\sigma_\tau(0)\left(\sigma_\tau(0) - \sigma_\tau(t)\right)(\mu - m)\right]. \tag{2.11}$$

If $\sigma_\tau(t) < \sigma_\tau(0)$ (the basic assumption of asymmetric learning), and $\mu > m$, the firm will enjoy positive expected profits on the worker. This surprising result comes from the fact that retaining firms act as a monopsonistic consumers of the information they acquire about

[10] Assuming continuousness in match quality, this difference goes to zero, the longer a worker is in the market.

[11] See Appendix A

the productivity of their workers.[12] It is important for deriving predictions regarding layoffs that their workers differ in their profitability, and their profitability depends not only on their productivity, but also upon the observable characteristics.

It is possible that firms overbid on a worker after receiving overly favorable signals. Subsequent signals hone in on the true productivity and the expected value is exceeded by the previous wage. Ordinarily, such overbidding seems rare. However, in the context of economic fluctuations, it is more understandable for wages to exceed expected productivity. Allowing the match component (φ_r) of a workers' productivity to depend on the economic climate provides a mechanism by which economic conditions impact profitability, if match quality is realized after firms determine wages. The realization of a lower than expected match may lead wages to exceed productivity during economic downturns. This would be particularly more likely for workers whose wages were already close to their productivity. In the presence of downward wage rigidity, the probability that a firm lays off a worker is the probability that the expected profits from the worker (given the signals and revealed current match quality) are negative. More formally, allowing ψ_L to be the composite error term, the probability of a layoff $P(L)$, is given by equation 2.12 below:

$$P(L) = P\left\{\varphi_r - E(\varphi_h) + \psi_L > \frac{\sigma_\xi(x)\sigma_\epsilon}{Q'Q}\left[\sigma_\xi(x)\sigma_\tau(0)\left(\sigma_\tau(0) - \sigma_\tau(t)\right)(m - \mu)\right]\right\}. \quad (2.12)$$

Similar to job switches, equation 2.12 depends on the difference between ability and reference group quality, the difference in precision of the employers' signals, and is scaled by the

[12] Theoretically, a fixed cost of dismissing workers may prevent infinite hiring and dismissal and allow economic profits to be zero, even if there is heterogeneity in the difference between workers' wage and their marginal product of labor. Adding such a cost complicates algebraic derivations, but does not change predictions.

precision of the public signal. Again imposing the normality and orthogonality assumptions and taking the derivative with respect to ability gives the following:[13]

$$\begin{aligned}\frac{\partial P(L)}{\partial \mu} =& -\phi\left\{\frac{\sigma_\xi(x)\sigma_\epsilon}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\left[\sigma_\xi(x)\sigma_\mathcal{T}(0)\left(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)\right)(m-\mu)\right]\right\} \\ &\frac{\sigma_\xi(x)\sigma_\epsilon}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\sigma_\xi(x)\sigma_\mathcal{T}(0)\left(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)\right)<0.\end{aligned} \tag{2.13}$$

Equation 2.13 illustrates that as ability (μ) increases the probability of layoff should fall. This is perfectly in line with Gibbons and Katz (1991b). Taking the derivative with respect to mean reference group ability (m) gives perhaps a more surprising result:[14]

$$\begin{aligned}\frac{\partial P(L)}{\partial m} =& \phi\left\{\frac{\sigma_\xi(x)\sigma_\epsilon}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\left[\sigma_\xi(x)\sigma_\mathcal{T}(0)\left(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)\right)(m-\mu)\right]\right\} \\ &\frac{\sigma_\xi(x)\sigma_\epsilon}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\sigma_\xi(x)\sigma_\mathcal{T}(0)\left(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)\right)>0.\end{aligned} \tag{2.14}$$

Conditional on individual ability, as workers' reference groups (m) are in general more productive, the more likely they are to be laid off. This is because, high reference group workers' wages are bid higher by outside firms, which place significant weight on the reference group.

Again, the learning framework allows for further predictions concerning the evolution of this selection over experience and working spell length. Just as with job-to-job transitions, the cross partial of $\frac{\sigma_\xi(x)\sigma_\epsilon}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}[\sigma_\xi(x)\sigma_\mathcal{T}(0)\left(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)\right)(m-\mu)]$ with respect to working spell length and ability is negative, and is positive with respect to working spell length and reference group. This implies that with increases in working spell duration, the adverse selection into unemployment on the basis of ability should become stronger (more negative),

[13] Please see Appendix A

[14] Please see Appendix A for proof.

and the positive selection on the basis of reference group should also become stronger (more positive). [15] The cross partial of $\frac{\sigma_{\xi}(x)\sigma_{\epsilon}}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}[\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\left(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)\right)(m-\mu)]$ with respect to experience and ability is positive, and is negative with respect to experience and reference group. Thus the selection on the basis of both ability and reference group weakens with increases in experience. [16] The inverse is also true. For workers whose reference groups are generally more capable than they are, increases in working spell length make it more likely they will be laid off. With respect to experience the selection of higher reference groups into unemployment should become weaker (more negative) with increases in experience.

Given that such broad economic downturns are exogenous to an individual's ability, the predictions regarding selection into job-to-unemployment separations during recessions may be more insulated from possible correlation between ability and match quality than is the case for job-to-job transitions.

2.4 Estimation

Equation 2.5 and equation 2.12 are conveniently structured for normal maximum likelihood (probit) estimation. While the model is structured to estimate the probability of separation, I only observe the binary indicator for whether a separation occurred ($s = t$), and whether it was a job-to-job ($s = j$) or job-to-unemployment separation. I further distinguish job-to-unemployment separations that occur during economic recessions ($s = r$) as opposed to economic expansions ($s = e$) to explore selection into layoffs as opposed to quits.

Following earlier research, I use age-adjusted, *AFQT* percentile scores as the hard-to-

[15] Please see Appendix A for proof.
[16] Please see Appendix A for proof.

observe, strong correlate with ability (μ). I model the reference group (m) is two ways. First, I construct average adjusted $AFQT$ scores by educational attainment,[17] college selectivity,[18] and race. However, this assumes that each easily observable characteristic influences mobility decisions through the same mechanism and to the same degree. I explore the validity of this restriction by also allowing each covariate to enter separately. For simplicity, I substitute $\overline{AFQT}$ for m for the remainder of this discussion. Allowing y_s^* to be the latent probability of separation (where s indexes the separation type), I estimate the following including experience (Exp) and working spell duration ($WrkSpl$), which play a central role in learning and human capital accumulation:[19]

$$y_s^* = \Phi \left\{ \beta_{s0} + \beta_{s1} AFQT + \beta_{s2} \overline{AFQT} + \beta_{s3} WrkSpl + \beta_{s4} Exp \right\}. \tag{2.15}$$

From above, for both job-to-job moves (y_j) and job-to-unemployment separations during recessions (y_r), the model predicts $\beta_1 < 0$ and $\beta_2 > 0$.

Since the model predicts each to change as spell length and experience increase, I also estimate each variable with interaction terms with both length of spell and experience in estimation. Inserting the appropriate $AFQT$ scores for μ and m, leads to the following equation to be estimated:

$$y_s* = \Phi \left\{ \left(\boldsymbol{\beta_s} + \boldsymbol{\delta_s} WrkSpl + \boldsymbol{\gamma_s} Exp \right) \boldsymbol{X} \right\}, \tag{2.16}$$

where $\boldsymbol{X}$ is a vector containing $AFQT$, $\overline{AFQT}$, $WrkSpl$, and Exp; $\boldsymbol{\beta_s}$ is a vector of main

[17] Educational attainment is grouped by high school graduates with no college, those who have some college, and those with at least a four year degree.

[18] I group by Barron's 7 bins of college competitiveness, and add a separate bin as an indicator for if the institution was not listed on Barron's.

[19] All regressions also include an indicator for urbanicity of the labor market and a vector of year indicators.

effect coefficients; $\boldsymbol{\delta_s}$ is a vector of coefficients on interactions of each variable with working spell duration; and $\boldsymbol{\gamma_s}$ is a vector of coefficients on interactions of each variable with experience.[20] Referring back to Section 3.2 and 3.3, the model predicts for both job switches and layoffs all else equal, adverse selection on the basis of $AFQT$, and that selection should become more negative with increases in working spell duration (δ_{1j}, $\delta_{1r} < 0$), and should become more positive with increases in experience (γ_{1j}, $\gamma_{1r} < 0$). Further, the model predicts positive selection on the basis of $\overline{AFQT}$, which should become stronger with increases in length of working spell (δ_{2j}, $\delta_{2r} > 0$), and weaker with increases in experience (γ_{2j}, $\gamma_{2r} > 0$).

2.5 Data

I use the National Longitudinal Survey of Youth of 1979 (NLSY79), since it contains the dates of hiring and termination of each job respondents held and more importantly contains workers' AFQT scores to proxy for the underlying ability of each worker. I end the sample in 2000 to bring my estimates in line with the existing literature and to reduce issues related to non-random attrition, which begins to become more problematic in subsequent years. Women are excluded to minimize instances of job separation due to child rearing. The remaining sample is composed of observations of 6,403 males over 22 years. Additionally, I exclude 452 men for whom the NLSY79 contains no AFQT score. I also drop all individuals for whom there is missing data for more than a quarter of the time periods. Following Pinkston [2009], I further restrict the analysis to men who obtained at least a high school degree, dropping 889 men who did not complete their secondary education.[21]

[20] Naturally, γ_{s3} and δ_{s4} are redundant and only one is included in estimation.

[21] This last restriction is significant. The interpretation of the role of the reference group's mean AFQT score differs depending on whether they are included in the sample. However, it seems that employer learning may have a relatively smaller role in the determination of job separations in this population. It is possible that

Since all individuals completed the AFQT in 1979, these scores are useful in demonstrating the composition of the population who choose different levels of education. However, I should note that the AFQT test was administered when participants were between the ages of 14 and 22. Consequently, though the AFQT is meant to measure the aptitude of the individual, scores may also reflect some differences in the amount and quality of participants' education received prior to the administration of the assessment. In order to correct at least for developmental influences and the quantity of education available, the AFQT scores used in the analysis are age adjusted following Altonji and Pierret (2001) and Pinkston (2009). I subtract the average percentile score of all those who were the same age when they took the test and divide by the standard deviation.

Table B.2 provides the average standardized AFQT percentile scores for workers of each race with each level of education, which will proxy for the average ability of each worker's reference group in one specification. First, notice that as education increases so does the average AFQT. This is as expected since we generally think that education is less cognitively taxing on those with higher intellectual ability. Were the marginal benefit to additional education equal across people (an unlikely assumption), those for whom the cost was lower would choose to obtain more. Second, notice that even within levels of educational attainment, the mean AFQT percentile score differs significantly across races. The mean AFQT score for a Black, high-school graduate is approximately .7 standard deviations lower than that of White high-school graduates. Among college graduates, the mean AFQT score for Black men is more than a full standard deviation lower than the mean for White participants.

I construct reference AFQT using the NLSY79. Since the model assumes that workers'

---

the population who chose not to persist through high school may also choose not to persist in employment as well.

abilities, for which the AFQT proxies, is distributed normally around the mean of their reference group, using the estimated mean provides ease of interpretation. Rather than guessing the predicted sign of race and education within the analysis the model gives direct predictions. However, this also imposes the restriction that each dimension of the reference group affects the probability of separation in the same way. Consequently, I also conduct the analysis using indicators for race and highest grade completed for education.

Race is fixed within the data, while educational attainment varies for many individuals during the survey period. Since I am attempting to create a measure for the expectation of employers, an argument can be made for using either a fixed measure of educational attainment or allowing education to vary over time. I prefer to treat education as fixed. First, for many individuals their education stays below 12 for several years and then jumps to 12 when participants are much older. Secondly, the model treats education as a dimension of a reference group and it seems likely that, employers may draw different information from a degree obtained later in life. Consequently, education is measured as the highest grade completed at the age of 25. College competitiveness is measured according to the Barron's index of degree granting institution or most recent school attended at the age of 25.

As mentioned, the NLSY79 records employment status covering this 22 year period of observation. Analysis is restricted to periods in which the participants were working at least 30 hours in a week and to jobs initiated after the survey began. After matching employers across years and NLSY job lines, using employer start and stop dates, this study constructs measures of experience, tenure, working spell, and job separations.

Experience is measured as the number of quarters an individual reports working up to the current period. Because employers may infer additional information about the worker from experience, following Pinkston (2009), I use potential experience instead of actual experience for all single-step estimation and as an instrument for actual experience in the control function estimation. Length of working spell is defined as the difference between the current quarter and earliest date of hire over a period in which the respondent worked without experiencing a job-to-unemployment separation. This avoids a reset each time a participant reports switching in and out of the same employment spell or with job-to-job moves. Each of these occur in the weekly arrays. Table B.3 provides summary statistics of the work histories of those included in the sample.

As mentioned in Bratsberg and Terrell (1998), the tenure variable recorded in the NLSY79 is inconsistent in accounting for the start and stop week of jobs. Consequently, I generate tenure using the difference between the start date and beginning date of each quarter individual reports working for a particular employer, subtracting periods the worker reports being temporarily out of work or on active call in the military. Terminal tenure uses the date the respondent reports leaving the employer. As noted in Light (2005), the young workers in the NLSY79 are highly mobile. Table B.4 provides a rough distribution of the terminal tenure length of job spells within the sample measured in quarters. Notice that roughly 55% of all employment relationships end within the first year and 76% end within the first two years. From the third year onward the drop-off is less dramatic.

Job separations serve as the primary outcome variable, and following Schönberg [2007], I decompose separations into job-to-job and job-to-unemployment transitions. Because it seems that asymmetric learning may have different predictions for job-to-unemployment transitions between different states of the economy, I further separate analysis on moves to unemployment between economic expansions and contractions. Separations are taken directly from the quarter in which respondents reported to leave their primary employer. Separations in which the respondent reported working for a new employer during the same or next quarter without reporting to have looked for a job or spent more than a full quarter out of the labor force, I define as job-to-job moves. Separations during which the respondent reports having looked for a job, I define as job-to-unemployment transitions. Table B.5 provides this breakdown in job separations. Overall, it seems the quarterly termination rate is nearly twelve percent with the majority of moves coming from job-to-unemployment transitions.

Because job separations are public events, employers and workers reveal significant information whenever one occurs. Further, there is evidence that the information communicated differs depending on the type of separation that occurs Gibbons and Katz (1991b). Accordingly, I provide a brief examination of wage changes after a worker leaves a firm. Table B.6 shows the differences in workers' average wages over the duration of the two jobs on either end of a job separation. Ideally, I would want the beginning and ending wage, but the wage information is recoded yearly for most of the sample and every two years from 1994 to 2000. Since the separation data is quarterly, there is significant measurement error and the table should be interpreted accordingly.

The most immediate pattern seems to be that the wage gains in moving jobs seems more beneficial if the move does not include a period of unemployment. This is consistent with the "Layoffs and Lemons" story presented in Gibbons and Katz (1991b). It is also consistent with asymmetric employer learning. As employers learn about employees it is akin to the accumulation of human capital. In a job to unemployment move, this information is lost and workers are accordingly penalized through their wages. Secondly, the gains from mobility are almost strictly increasing in AFQT conditional on education. The fact that those with higher than average AFQT scores have higher wage growth is suggestive of employer learning in general. However, looking at the difference in wage changes between job-to-job and job-to-unemployment transitions, there is no evidence that the relative benefits of job moves are largest for those who have higher than average AFQT scores, as the asymmetric learning model suggests to be the case.

2.6 Empirical Analysis

2.6.1 Primary Results

The most basic prediction of asymmetric employer learning is that conditional on reference group, workers who experience job separations are adversely selected. Symmetric learning offers no rationale for an individual's AFQT to impact the probability of job separations, because all employers (both current and hiring) equally weigh hard to observe characteristics. Columns 3 and 4 of Table B.7 provide results from the most simple test of this hypothesis. Both in explicitly conditioning on reference groups and implicitly doing so through mean AFQT, I find that having one standard deviation higher AFQT score decreases the probability of job separation in a given quarter by 0.61 to 0.65 percentage points (p-values less than .001). Given that the base probability of separating is about 12% within sample, this is about a 5% increase in the probability of terminating an employment match.

Secondly, asymmetric learning predicts that conditional on the individual's AFQT, as the mean AFQT of the reference group increases, the individual should be more likely to leave. Similarly, easily observed correlates with productivity should be positively related to the probability of separation. This is because the outside market places more weight on workers' easily observed characteristics relative to the current employer. Table B.7 reveals exactly this relationship. From column 3, the estimated effect of the average AFQT of the respondent's reference group is positive, and statistically significantly so (p-value$<$0.05). Perhaps more importantly, without conditioning on the individual's AFQT in the first column, the AFQT of the reference group is significantly negatively related to the likelihood of separation (p-value<0.001). This suggests that while in general those with more desirable observable characteristics enjoy job security, the retaining firm allows (or encourages) those with lower AFQT within each group to leave.

As the reference group depends on race, educational attainment, and college selectivity, it is unsurprising that the inclusion of an individual's AFQT makes the relationship between each easily observable covariate and the probability of moving more positive.[22] From Column 4, the indicator for attending a competitive college increases the probability of separation by 0.8 percentage points (p-value<0.001), whereas educational attainment is essential unrelated to the probability of job separation. Regarding race, while White respondents have higher AFQT scores on average than Black respondents at each level of education, Hispanic respondents have lower AFQT scores in general. However, indicators for White and Hispanic are both positively correlated with job separations conditional on AFQT. Given that these results come from estimating equations that do not account for differences in job separations, they are more useful to provide a setting rather than direct evidence.

Table B.8 is split into two panels. Panel A provides estimates of the average partial effects (APEs) of the individual's and the references group's AFQT on the probability of job-to-job moves and job-to-unemployment separations both during economic expansions and recessions. Panel B provides APEs for individual AFQT, educational attainment, college competitiveness, race, length of work spell, and time in the labor market separately for each type of move. The model provides the most clear predictions for job-to-job transitions

[22] Hispanic is the only covariate that has its point estimate fall. However, Hispanics have lower AFQT scores than do Black respondents, meaning that this change is also in accordance with the change in the point estimate of the summary measure.

and job-to-unemployment separations during recessions. Accordingly, the majority of the following discussion will center on these types of transitions.

The third and fourth columns of each panel reveal that the adverse selection on the basis of an individual's AFQT score is driven by job-to-unemployment transitions, with the strongest effects during recessions. I find that a one standard deviation increase in AFQT conditional on observables is associated with a full percentage point decrease in the probability of experiencing a move from employment to unemployment during a recession. During economic expansions, the same increase in an individual's AFQT leads to around a 0.6 percentage point drop in the probability of separation. Each of these results are statistically significant with p-values less than 0.001, though they are not statistically different from one another. These results are suggestive of asymmetric learning. As in Gibbons and Katz (1991b), the rationale here is that firms layoff their least profitable workers. During job-to-job moves, though the the asymmetric learning hypothesis predicts similar adverse selection, in keeping with Schönberg [2007], I find no evidence of negative selection for these types of moves.

Neither Gibbons and Katz (1991b) nor Schönberg (2007) analyze the selection with regard to reference group, conditional on individual ability, though asymmetric employer learning clearly predicts positive selection into job switches and layoffs on the basis of easily observable covariates. Consistent with the asymmetric learning model, the second and third column of Panel A indicate positive selection into job-to-job and to a lesser degree job-to-unemployment transitions during recessions. Being a member of a reference group with a one standard deviation higher AFQT raises the probability of a job-to-job transition by 0.6 percentage point or about 9 percent (p-value$<$0.001). The finding that those with higher reference groups are more likely to be bid away conditional on the individual's AFQT suggests that the outside market values the reference group more so than does the current employer.

The point estimate for reference group's AFQT is also positive for job-to-unemployment transitions during recessions though not statistically significantly so. Taken literally, this positive point estimate indicates that those with better observables are more likely to move from employment to unemployment during a recession, conditional on individual ability. This suggests that there is less rent between workers' wages and marginal products, presumably because the outside market bids up their wages on the basis of the reference group. The resulting small buffer cannot insulate these high-reference-group workers from general productivity shocks.

Turning to the components of reference groups, White respondents are conditionally more likely than Black respondents to experience a separation in each environment with the largest point estimates coming from job-to-job transitions and job-to-unemployment separations during recession. These are exactly where the asymmetric learning model predicts them to be largest. While the estimated effect of being White is statistically significant for job-to-job moves, it is too noisy during recessions to draw meaningful inference. Hispanic respondents on the other hand are conditionally less likely to be bid away by another firm than are Black respondents all else equal, though not significantly so. Given that the average AFQT score for Hispanic respondents is lower than for Black respondents, this is in accordance with theory. On the other hand, during recessions, I estimate being Hispanic to lead to a 3 percentage point higher probability of experiencing unemployment. The magnitude of this unpredicted result raises questions as to whether asymmetric learning is driving the relationship between race and job mobility. This more complicated relationship between race and mobility lead me to prefer the more transparent approach of including each covariate separately for the remainder of the analysis.

Selection on the basis of education provides more consistent evidence supporting asymmetric employer learning. Attending a competitive colleges raises the probability of separation, all else equal. These results are driven by job switches and job-to-unemployment separations during recessions. Attending a competitive, very competitive, highly competitive, or most competitive undergraduate institution increases the probability of switching jobs by about 0.4 percentage points and the probability of a job-to-unemployment transition during a recession by 1.4 percentage points. While statistically insignificant, the point estimates indicate education is positively selected in job switches and job-to-unemployment transitions during recessions, where the asymmetric learning model has the strongest predictions. During economic expansions where the model has weaker predictions, this selection is reversed. In unreported regressions excluding college selectivity, educational attainment is statistically significantly positive for both job-to-job and job-to-unemployment transitions during recessions. These findings that more educated workers are more likely to switch jobs and be laid off further support the asymmetric learning hypothesis and seem to be a primary driver of the results regarding reference group.

The time dynamics of the effects of AFQT and education provide additional suggestive evidence of asymmetric learning. As employers learn, they should more accurately identify and retain the most profitable workers and layoff the least profitable. Thus, these effects should grow stronger over time. From the Pinkston's [2009] model, due to the bidding structure, current employers reveal their accumulated information to outside firms in the event of a job-to-job transfer. Accordingly, this extension of the model predicts that as a continuous working spell lengthens, all else equal, information asymmetries grow between retaining and hiring firms, and the effect of an individual's AFQT score should become more powerful. In contrast, information is revealed to all the longer the participant is in the market (working or not), and the smaller these asymmetries should become. Thus, the model predicts the selection to become weaker with increases in potential experience, all else equal. In Table B.9, I include a full set of interactions with length of working spell and potential experience to uncover these dynamics.

I find that in keeping with the asymmetric learning model's predictions, the effects of AFQT seem to grow more negative with increases in the length of continuous working spells and more positive with potential experience. Across all columns, the scaled coefficients (SCs) on the interactions of AFQT with spell length are negative, strengthening the selection on AFQT with increases in length of working spell. However, the point estimates are noisy and far from statistical significance. As predicted, the coefficients on the interactions between potential experience and AFQT are positive across all columns, indicating that the selection weakens, the longer the individual is in the market. Again, these dynamics are not statistically significantly different from zero. Dynamics in the effect of AFQT on the probability of separation regarding both increases in continuous working spell and increases in potential experience are consistent with asymmetric employer learning, but the results are too noisy to be very informative.

The dynamics regarding race are mixed, further questioning whether employer learning is driving the relationship between race and mobility. Regarding education, again the results are more strongly supportive of asymmetric employer learning. Scaled coefficient estimates on the interaction between college competitiveness and length of working spell are statistically significantly positive in all columns, but for job-to-job moves. The interaction between years of education and working spell duration is statistically significantly positive across all columns. These results indicate that the the selection into worker mobility on the basis of educations is getting more positive the longer the individual is continuously working. These results are consistent with idea that outside firms place more value on these easily observable covariates than do current employers.

In contrast, the scaled coefficient estimates on the interaction between college competitiveness and quarters of potential experience are statistically significantly negative in all columns, except for job-to-job moves, and the interaction between years of education and potential experience is statistically significantly negative across all columns. These results indicate that the selection on the basis of education weakens as potential experience increases and the market learns more about true worker productivity.

2.6.2 Robustness Results

Unlike linear estimation, normal maximum likelihood estimation is inconsistent in the presence of heteroskedasticity **?**. Accordingly, Table B.10 and Table B.11 present the estimated APEs and scaled coefficients respectively to compare with those using probit estimation. In the model laid out above, the composite error term, ψ, depends on τ, ξ, and v. It follows that $var(\psi)$ is a function of the variance of the incumbent employer's private signal, $\sigma_{\mathcal{T}}(t)$, and the variance of the public signal, $\sigma_{\xi}(x)$. Since $\sigma_{\mathcal{T}}(t)$ and $\sigma_{\xi}(x)$ are in turn functions of working spell length and experience, the asymmetric employer learning model predicts the conditional variance of the probability of a job-to-job transition to change with spell length and experience as well.

Above, I model the error term of the binary model of job separations as a function of the noise of the signals that each employer receives. Because the noise of the signals decreases with experience and working spell length, I model the variance of that error term as:

$$ln\left(var(\psi)\right) = \theta_0 + \theta_1 WrkSpl + \theta_2 Exp. \tag{2.17}$$

The assumptions that $\frac{\partial \sigma_\xi(x)}{\partial x} < 0$, which is basic to employers learning gradually over time, and that $\frac{\partial \sigma_\mathcal{T}(t)}{\partial t} < 0$, which is basic to asymmetric employer learning, imply that $\frac{\partial \sigma_\psi}{\partial x} < 0$ and $\frac{\partial \sigma_\psi}{\partial t} < 0$.[23] In other words, asymmetric employer learning predicts that the conditional variance of job-to-job transitions should decrease with experience and length of working spell. I estimate the effect of working spell length and experience on the variance implicitly using heteroskedastic probit estimation. Panel 2 of Table B.10 provides the estimated effects of spell length and potential experience on the conditional variance. The estimated effects of working spell duration and potential experience on the conditional variance large do not bear out these predictions, though they are statistically significant. Thus, it remains important to examine the stability of the earlier findings under this alternate specification.

Turning to the selection into job separation from Table B.10, the estimated APEs remain virtually unchanged from those in Table B.8 using standard probit estimation. The main results and inference hold, with a small change in that years of educational attainment now reaches the 95% confidence threshold in statistical significance for job-to-job moves. From Table B.11, the effects again remain largely stable with two key exceptions. First, the interaction between AFQT and working spell duration becomes statistically significantly negative for job-to-unemployment separations during recessions, giving additional support to the asymmetric learning hypothesis. Second, though they do not switch sign, the scaled coefficient estimates for the interaction between competitive college attendance and potential experience become noisier and lose statistical significance for separations in general and job-to-unemployment during recession. In general, it is reassuring to see the results remain this constant.

Pinkston [2009] expresses concern about possible endogeneity of experience and work

[23] Please see Appendix A, A, A, and A for proofs.

spell duration. Theoretically, each contains information that employers may use to form their initial expectations, which may bias the estimated effects of experience, working spell, and their interactions. Consequently, he takes an instrumental variables approach akin to that taken in Altonji and Shakotko (1985). However, since his model incorporates a public signal it perhaps does not make sense to throw out all of this information. It seems that the bias in the interaction terms are the main concern. Failing to control for the public information contained in experience and working spell length may lead to bias in the estimated effects of AFQT scores and reference group membership. Consequently, I use a control function approach to try to control for the additional public information contained in experience and spell length, while avoiding bias in the interaction terms. To clarify, I perform first stage regressions of working spell of average working spell in the sample, number of time observations and current job duration and experience on potential experience. I then interact those residuals with each variable that is interacted with spell length and experience. To account for the fact that the residuals are estimated, I bootstrap all standard errors over 500 repetitions. One benefit of this control function approach is that the residuals account for endogeneity in the interaction terms and provide an immediate statistical test for the presence of endogeneity.

Table B.12, Table B.13, and Table B.14, provide the average partial effects (APEs), effects on the conditional variance, and scaled coefficients respectively from heteroskedastic normal MLE using the control function approach described above. First, the APEs of the residuals are generally significant under either specification, indicating the presence of endogeneity. In both panels of Table B.12, the APEs of AFQT remain significantly negative only for job-to-unemployment transitions during economic expansions, though in no case do the estimated effects change sign. The selection on education is more robust. Rather than college competitiveness carrying the predictive power, it shifts to years of education. For both job switches and moves to unemployment during recessions, years of education is positively and significantly selected, while college competitiveness only approaches statistical significance for job-to-job transitions.

As before, Table B.13 reveals that contrary to prediction, the conditional variance largely increases in both experience and working spell length. It is actually increases in the control function residuals that cause the conditional variance to shrink.

From Table B.14, the scaled coefficient estimates of the interaction between AFQT and working spell length is no longer strictly negative across types of job separations, nor are the scaled coefficients of the interaction between AFQT and experience strictly positive moving across columns. However, none are statistically significant either, leaving the dynamics of AFQT uninformative regarding working spell duration and experience. The dynamics are more consistent regarding education. From Table B.14, a similar pattern as was shown in Table B.9 and Table B.11 persists. Across all columns, as working spell lengthens, educational attainment becomes more positively selected. Also consistent with theory, educational attainment becomes less important as experience increases. Regarding college selectivity, while the coefficients on the interaction with experience loses statistical significance, the interactions with working spell length continue to be positive in general and for job-to-unemployment transitions. The fact that the more educated workers from more selective school are positively selected into mobility, and that this effect strengthens the longer the worker is continuously employed and weakens with additional experience, suggests that other firms value these signals more so than does the current employer. The sensitivity of the APE of AFQT, as well as the lack of dynamics regarding AFQT indicate that perhaps there are other aspects of productivity, about which employers are primarily learning.

2.7 Conclusion

To summarize, I find that consistent with asymmetric learning, those with higher levels of education, conditional on individual ability are more likely to be laid off or bid away by another firm. This selection strengthens as continuous working spells increases and weakens with experience. These results are robust to alternate specifications. To my knowledge these are novel empirical facts that are difficult to explain in the absence of private employer learning.

I only find evidence of adverse selection on the basis of AFQT in job-to-unemployment separations. In keeping with the asymmetric learning hypothesis, the magnitude of this effect is largest during recessions, though in accounting for endogeneity of experience and working spell duration the magnitude of this effect drops and inference becomes tenuous. The evidence of on these effects strengthening with working spell length is sensitive, and the dynamics with regard to experience are merely suggestive. Taken cumulatively, the evidence for asymmetric employer learning is strong and consistent regarding workers' education and merely suggestive from the analysis of AFQT. Perhaps this is indicative that there are more dimensions of productivity about which employers are learning, and further work is needed to explore what those might be.

Chapter 3

Handling Correlations between Covariates and Random Slopes in Multilevel Models

3.1 Introduction

We consider linear regression models for clustered data that include cluster-specific random intercepts and slopes. Such models are called multilevel models, mixed models, random-coefficient models, or hierarchical linear models. If the models are viewed as "structural models," the perspective taken in this paper, the regression coefficients represent structural or causal parameters, and the error terms represent the effects of omitted covariates. If there are omitted confounders that are correlated with included covariates, then the error terms are correlated with the included covariates. These correlations lead to omitted-variable bias. This paper focuses on estimation methods that avoid bias due to omitted cluster-level confounders, also referred to as "cluster-level endogeneity." An alternative view of models, not taken here, is that regression coefficients merely represent associations between included variables, or linear projections in the case of linear models, in which case the error terms are orthogonal to the covariates by construction (see Spanos (2006) for a discussion of "structural" versus “statistical” models).

Research on addressing cluster-level endogeneity in multilevel models has traditionally been confined to correlations between unit-level covariates (i.e., covariates that vary over units) and random *intercepts* that vary over clusters in which units are nested. This constitutes a type of “cluster-level” endogeneity as it involves correlation with a cluster-level random error term. For instance, in estimating the effect of Catholic schooling on student achievement controlling for student socioeconomic status (SES), one may worry that school-level omitted variables, such as school resources, may be correlated with SES. Left unaddressed, this endogenity may lead us to mis-attribute the impacts of these omitted variables to the effect of SES. This bias may in turn spill over to other coefficients. To address this type of endogeneity, **?** shows that consistent estimators of the coefficients of unit-level covariates can be obtained by a fixed-effects approach. However, with standard fixed-effects estimators, coefficients of cluster-level covariates (i.e., covariates that only vary over clusters) cannot be estimated. The Hausman and Taylor (1981) instrumental-variable estimator resolves this limitation and is consistent for the coefficients of both unit- and cluster-level covariates under appropriate assumptions (see Castellano et al., 2014).

Endogeneity in the form of correlations between unit-level covariates and random *slopes* varying over clusters may also arise. Referring back to the Catholic schooling example, when controlling for students' SES, the slope of SES (or SES achievement gradient) may vary between schools, due to interactions between SES and omitted school-level covariates, such as school resources. If the omitted variables are negatively correlated with the SES achievement gradient, then the random slopes will be negatively correlated with SES.

Remarkably, such endogeneity is rarely considered. One exception is Frees (2004) who extends the Mundlak approach to handle random slopes. Another is Wooldridge (2005) who shows under seemingly benign conditions that traditional fixed-effects estimation of random-intercept models is robust against correlations between unit-level covariates and random slopes. However, neither of these approaches permits estimation of the coefficients of cluster-level covariates even if the covariates are exogenous (i.e., not endogenous). This limitation is overcome by Kim and Frees (2007) who use generalized method of moments estimation to extend the Hausman-Taylor approach to multilevel models with random slopes. However, their method is difficult to implement, making the fixed-effects approach more feasible in practice.

Unfortunately, a key assumption required for the fixed-effects approach may be violated in important applications. Specifically, the within-cluster variance of a unit-level covariate must be uncorrelated with the random slope of that covariate, which we refer to as the "uncorrelated variance assumption" throughout this paper. Turning back to the Catholic schooling example, it is possible that more diverse schools (schools with high variance of SES) may be better equipped to mitigate the effects of SES (lower the SES achievement gradient) than schools that are more homogeneous (schools with low variance of SES). Such a situation would directly violate the uncorrelated variance assumption.

In this paper, we investigate estimation of the coefficients of unit-level and cluster-level covariates in multilevel models in the presence of two sources of endogeneity; (nonzero) correlations between unit-level covariates and *both* the random intercept and random slopes. Throughout, we assume that covariates are uncorrelated with the unit-level error term and that cluster-level covariates are uncorrelated with the random intercepts and random slopes. We propose a simple "per-cluster regression" (PC) approach that is unbiased and consistent for coefficients for both unit-level and cluster-level covariates under both forms of cluster-level endogeneity and violation of the uncorrelated variance assumption. We contrast its performance to the "standard" random-effects (RE) estimator and what we call the "augmented fixed-effects" (FE+) approach, which extends the fixed-effects approach to provide estimates of the coefficients of cluster-level covariates. In Section 3.2, we first introduce our motivating empirical example and specific model of interest and then present our general model. In Section 3.3, we discuss the traditional random- and fixed-effects estimators and their conditions for unbiasedness. In Section 3.4, we introduce new estimators, namely the augmented fixed-effects estimator and the per-cluster estimator, and show under what assumptions they are unbiased. We provide conditions for consistency for all four estimators in Appendix I. All estimators are applied to a dataset in Section 3.5, and Section 3.6 investigates performance of the estimators using a simulation study.

3.2 Motivation and Multilevel Model

3.2.1 Motivating Example and Specific Model

As a motivating example, we consider the effect of private schooling on student achievement. We use the Raudenbush and Bryk (2002) data from the 1982 High School and Beyond (HSB) Survey because it is familiar in education and it is in the public domain, allowing us to provide data and commands for all estimators in Appendix I.

This two-level dataset provides us with an estimation sample of 7,185 students (units) nested in 160 schools (clusters), 70 of which are Catholic (private), and the remaining of which are public. The number of observations per school ranges from 14 to 67 students (Mean=45, SD=12). We use a mathematics standardized test score for student i in school j as our response variable, y_{ij}, which has a mean of 12.75 and a standard deviation of 6.88. Our primary variables of interest are w_j, a binary indicator for whether school j is Catholic, and x_{ij}, a continuous index of students' socioeconomic status, composed of parental education, parental occupation, and parental income. This index has a mean of zero and a standard deviation of 0.78.

We write the model using a two-stage formulation, similar to that used in Raudenbush and Bryk (2002). The first stage is the Level-1 model:

$$y_{ij} = \eta_{0j} + \eta_{1j} x_{ij} + \epsilon_{ij}. \tag{3.1}$$

This is a simple regression of the student mathematics test scores y_{ij} on their socio-economic status x_{ij}, where the intercept η_{0j} and slope η_{1j} can vary between schools, as indicated by the j subscript. Each student's test score can deviate from the school-specific regression line by a random error term ϵ_{ij}.

The school-specific intercepts and slopes become (unobservable) outcomes in the Level-2 models:

$$\eta_{0j} = \gamma_0 + \gamma_1 w_j + u_{0j} \eta_{1j} = \beta_1 + \beta_2 w_j + u_{1j}.$$

The mean intercept and slope of SES, for the population of schools, depend on whether the schools are Catholic or public (w_j). The intercepts γ_0 and β_1 in these models therefore represent the population means of the intercepts and slopes of SES for public schools, whereas the slopes γ_1 and β_2 represent the differences in population means of the intercepts and slopes between Catholic and private schools, respectively. The Level-2 models have errors u_{0j} and u_{1j} to allow each school's intercept and slope to vary within the sub-populations of Catholic and private schools. Assumptions regarding the error terms are discussed in subsequent sections.

Substituting the Level-2 models into the Level-1 model, we obtain the reduced form of the model:

$$y_{ij} \;=\; \gamma_0 + \gamma_1 w_j + \beta_1 x_{ij} + \beta_2 w_j x_{ij} + u_{1j} x_{ij} + u_{0j} + \epsilon_{ij}. \tag{3.2}$$

We see that β_2 is the coefficient of a cross-level interaction between the student-level covariate x_{ij} and the school-level covariate w_j.

In this setting, it is likely that there are omitted school-level variables that affect student achievement, and hence enter the random intercept u_{0j}, and are correlated with student SES. If these omitted school-level variables also interact with student SES, then they enter the random slope u_{1j}, and the slope is then correlated with SES. Ignoring such endogeneity may lead to bias when estimating the coefficients of this model.

3.2.2 General Multilevel Model

The general model we consider in this paper is for two-level data, such as the HSB data described above. In the cross-sectional case, units (Level 1) are typically individuals nested within clusters (Level 2), such as schools, hospitals, or neighborhoods. In the longitudinal case, units refer to measurement occasions nested within individuals, who constitute the "clusters." Clusters are indexed j, with $j = 1, \ldots, J$, and units are indexed ij, with $i = 1, \ldots, n_j$. The general model includes unit-level covariates that vary between units within clusters (and between clusters), as well as cluster-level covariates that vary between clusters but are constant within the same cluster. Same-level and cross-level interactions may also be included, where cross-level interactions are unit-level covariates. Some unit-level covariates may have random slopes.

The general model can be written as

$$\mathbf{y}_j = \mathbf{W}_j\boldsymbol{\gamma} + \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j. \tag{3.3}$$

Here $\mathbf{y}_j = (y_{1j}, \ldots, y_{n_j j})'$ is the vector of responses for cluster j, the $n_j \times (P+1)$ matrix $\mathbf{W}_j$ includes all P cluster-level covariates and its first column is a vector of ones, $\mathbf{1}_{n_j}$, for the intercept; the $n_j \times R$ matrix $\mathbf{X}_j$ includes all R unit-level covariates; the $n_j \times (R_1+1)$ matrix $\mathbf{Z}_j$ includes all $R_1 \leq R$ unit-level covariates in $\mathbf{X}_j$ that have random slopes and its first column is $\mathbf{1}_{n_j}$ for the random intercept. Finally, the $\mathbf{u}_j$ are random effects or cluster-level error terms (one random intercept and R_1 random slopes) and the $\boldsymbol{\epsilon}_j$ are unit-level error terms. The $\mathbf{u}_j$ are assumed to be independent of the $\boldsymbol{\epsilon}_j$, and the clusters are independent in the sense that error terms as well as covariates are independent across clusters. Other assumptions made regarding $\mathbf{u}_j$ and $\boldsymbol{\epsilon}_j$ depend on what estimators are used and are discussed in Sections 3.3 and 3.4. Sometimes all unit-level covariates have random slopes, so that $R_1 = R$, but typically $R_1 < R$, so that the slopes of some unit-level covariates do not vary between clusters, giving rise to the term "mixed-effects" (mixed random and "fixed" effects) model.

For the specific model for the HSB data in Equation 3.2, $P = 1$ so that $\mathbf{W}_j$ has n_j rows and $P + 1 = 2$ columns, with each row equal to $(1, w_j)$. The corresponding coefficients are $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)'$. The matrix of $R = 2$ unit-level covariates is $\mathbf{X}_j = (\mathbf{x}_j \; w_j\mathbf{x}_j)$, where $\mathbf{x}_j = (x_{1j}, \ldots, x_{n_j j})'$, and $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. The first unit-level covariate has a random slope, so $R_1 = 1$, $\mathbf{Z}_j = (\mathbf{1}_{n_j} \; \mathbf{x}_j)$, and $\mathbf{u}_j = (u_{0j}, u_{1j})'$. Finally, the unit-level errors are $\boldsymbol{\epsilon}_j = (\epsilon_{1j}, \ldots, \epsilon_{n_j j})'$.

The model in Equation 3.3 can be expressed more compactly by combining all covariates (i.e., $\mathbf{X}_j$ and $\mathbf{W}_j$) into a single matrix $\mathbf{V}_j$ and likewise their corresponding coefficients ($\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$) into a single vector $\boldsymbol{\delta}$:

$$\mathbf{y}_j = \mathbf{V}_j\boldsymbol{\delta} + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j. \tag{3.4}$$

It will often be convenient to stack the covariates for all J clusters in the matrix $\mathbf{V}$.

Writing this general model using a two-stage formulation, analogous to Equations 3.1 and 3.2, requires additional notation, so we defer this until Section 3.4.2 where this formulation is useful for explaining the "per cluster regression" (PC) approach.

3.3 Standard Estimators and Conditions for Unbiasedness

3.3.1 Exogeneity and Endogeneity

Throughout this paper, we assume unit-level exogeneity or strict exogeneity given the random effects,

$$\mathrm{E}\left(\boldsymbol{\epsilon}_j|\mathbf{V}_j,\mathbf{u}_j\right) \;=\; \mathbf{0}. \tag{3.5}$$

It follows that $\mathrm{E}\left(\boldsymbol{\epsilon}_j|\mathbf{V}_j\right) = \mathbf{0}$ and $\mathrm{E}\left(\boldsymbol{\epsilon}_j'\mathbf{V}_j\right) = \mathbf{0}$ and that each element of $\boldsymbol{\epsilon}_j$ is uncorrelated with each element of the covariate vectors.

The assumption of cluster-level exogeneity can be expressed as

$$\mathrm{E}\left(\mathbf{u}_j|\mathbf{V}_j\right) \;=\; \mathbf{0}. \tag{3.6}$$

When this assumption is violated, there is cluster-level endogeneity.

We assume that cluster-level exogeneity holds for the cluster-level covariates

$$\mathrm{E}\left(\mathbf{u}_j|\mathbf{W}_j\right) \;=\; \mathbf{0} \tag{3.7}$$

and discuss the problem that unit-level covariates may be cluster-level endogenous,

$$\mathrm{E}\left(\mathbf{u}_j|\mathbf{X}_j\right) \;\neq\; \mathbf{0}.$$

Cluster-level endogeneity occcurs if, for example, $\mathrm{E}\left(\mathbf{X}_j'\mathbf{1}_{n_j}u_{0j}\right) \neq \mathbf{0}$ or in other words, if the cluster sums or means of the unit-level covariates are correlated with the random intercepts.

In this paper, we consider random-effects, "augmented fixed-effects," and (our proposed) "per-cluster regression" approaches for estimating the model in Equation 3.2 when $\mathbf{X}_j$ is correlated with both the random intercept u_{0j} and the random slopes u_{rj}, $r = 1, \ldots, R_1$. We describe each of these estimators and under which conditions they produce unbiased estimates of the regression coefficients $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$ in the following subsections.

3.3.2 Random-effects estimators

For random-effects (RE) estimators, the random effects $\mathbf{u}_j$ are assumed to have zero means and covariance matrix $\boldsymbol{\Psi}$ given the covariates. They are uncorrelated across clusters, and they are also uncorrelated with the unit-level error term $\boldsymbol{\epsilon}_j$. The elements ϵ_{ij} of $\boldsymbol{\epsilon}_j$ have zero means and variance θ given the covariates, and are mutually uncorrelated.

It follows from these assumptions that the mean and covariance structure of $\mathbf{y}_j$ given $\mathbf{V}_j$ becomes

$$\mathrm{E}(\mathbf{y}_j|\mathbf{V}_j) \;=\; \mathbf{V}_j\boldsymbol{\delta},$$

and

$$\boldsymbol{\Sigma}_j \equiv \mathrm{Var}(\mathbf{y}_j|\mathbf{V}_j) \;=\; \mathbf{Z}_j \boldsymbol{\Psi} \mathbf{Z}_j' + \theta \mathbf{I}_{n_j}. \tag{3.8}$$

Under the exogeneity assumptions in Equations 3.5 and 3.6, unbiased and consistent estimators for the parameters of the model in Equation 3.4 can be obtained using maximum likelihood (ML), restricted maximum likelihood (REML), or feasible generalized least squares (FGLS). The FGLS estimator can be expressed in closed form as

$$\widehat{\boldsymbol{\delta}}_{\mathrm{RE}} = \boldsymbol{\delta} + \left(J^{-1} \sum_{j=1}^{J} \mathbf{V}_j' \widehat{\boldsymbol{\Sigma}}_j^{-1} \mathbf{V}_j \right)^{-1} \left(J^{-1} \sum_{j=1}^{J} \mathbf{V}_j' \widehat{\boldsymbol{\Sigma}}_j^{-1} (\mathbf{Z}_j \mathbf{u}_j + \boldsymbol{\epsilon}_j) \right), \tag{3.9}$$

where $\widehat{\boldsymbol{\Sigma}}_j$ is an estimator of $\boldsymbol{\Sigma}_j$, obtained by substituting estimators of $\boldsymbol{\Psi}$ and θ into Equation 3.8, and $\sum_j \mathbf{V}_j' \widehat{\boldsymbol{\Sigma}}_j^{-1} \mathbf{V}_j$ is assumed to be nonsingular with probability 1.

The conditional expectation of the GLS estimator $\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}$ (assuming $\boldsymbol{\Sigma}$ is known), given $\mathbf{V}$, is

$$\mathrm{E}(\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}|\mathbf{V}) = \boldsymbol{\delta} + \left(J^{-1} \sum_{j=1}^{J} \mathbf{V}_j' \boldsymbol{\Sigma}_j^{-1} \mathbf{V}_j \right)^{-1} \left(J^{-1} \sum_{j=1}^{J} \mathbf{V}_j' \boldsymbol{\Sigma}_j^{-1} [\mathbf{Z}_j \mathrm{E}(\mathbf{u}_j|\mathbf{V}) + \mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V})] \right).$$

Note that $\mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V}) = \mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V}_j)$ and $\mathrm{E}(\mathbf{u}_j|\mathbf{V}) = \mathrm{E}(\mathbf{u}_j|\mathbf{V}_j)$ since clusters are assumed to be independent. Furthermore, unit-level exogeneity implies that $\mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V}_j) = \mathbf{0}$ and cluster-level exogeneity implies that $\mathrm{E}(\mathbf{u}_j|\mathbf{V}_j) = \mathbf{0}$. Conditional unbiasedness, $\mathrm{E}(\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}|\mathbf{V}) = \boldsymbol{\delta}$, hence follows, and using the law of iterated expectations, $\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}$ is (unconditionally) unbiased, $\mathrm{E}(\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}}) = \boldsymbol{\delta}$.

Unfortunately, due to the nonlinear nature of the FGLS estimator, this unbiasedness result does not automatically apply when estimates $\widehat{\boldsymbol{\Sigma}}_j$ are plugged in for $\boldsymbol{\Sigma}_j$. Under the above assumptions, a sufficient assumption for unbiasedness, $\mathrm{E}(\widehat{\boldsymbol{\delta}}_{\mathrm{FGLS}}) = \boldsymbol{\delta}$, is that the error terms have symmetric distributions Kakwani (1967). This result also applies to ML and REML, given that these estimators can be expressed as iterative versions of the FGLS estimator Don and Magnus (1980). Note that for the empirical example and simulation study, we use REML, following the tradition of its use in the education research literature.

3.3.3 Fixed-effects estimators

In econometrics, the term fixed-effects (FE) estimator refers to an estimator that does not rely on cluster-level exogeneity, and we adopt this terminology here. Some of the estimators can be derived by treating the random effects as fixed and others by eliminating the random effects, but in either case the effects are typically viewed as random. The traditional FE approaches have been developed for random-intercept models, with $\mathbf{Z}_j = \mathbf{1}_{n_j}$ and $\mathbf{u}_j = u_{0j}$, to handle violation of the exogeneity assumption $\mathrm{E}(u_{0j}|\mathbf{V}_j) = 0$ **?**.

We define the FE estimator in terms of the de-meaning (or group-mean-centering) transformation $\mathbf{Q}_j \equiv \mathbf{I}_{n_j} - \mathbf{1}_{n_j}(\mathbf{1}'_{n_j}\mathbf{1}_{n_j})^{-1}\mathbf{1}'_{n_j}$, where $\mathbf{Q}_j\mathbf{1}_{n_j} = \mathbf{0}$, and define $\ddot{\mathbf{y}}_j \equiv \mathbf{Q}_j\mathbf{y}_j$, $\ddot{\mathbf{X}}_j \equiv \mathbf{Q}_j\mathbf{X}_j$, $\ddot{\mathbf{Z}}_j \equiv \mathbf{Q}_j\mathbf{Z}_j$, and $\ddot{\boldsymbol{\epsilon}}_j \equiv \mathbf{Q}_j\boldsymbol{\epsilon}_j$. Pre-multiplying Equation 3.3 by $\mathbf{Q}_j$, the de-meaned model becomes

$$\ddot{\mathbf{y}}_j = \ddot{\mathbf{X}}_j\boldsymbol{\beta} + \ddot{\mathbf{Z}}_j\mathbf{u}_j + \ddot{\boldsymbol{\epsilon}}_j,$$

Note that $\ddot{\mathbf{W}}_j\boldsymbol{\gamma} = \mathbf{0}_{n_j}$ because the columns of $\mathbf{W}_j$ are constant. The first column of $\ddot{\mathbf{Z}}_j$ is $\mathbf{0}_{n_j}$ because the first column of $\mathbf{Z}_j$ is $\mathbf{1}_{n_j}$, so that the random part of the model does not depend on u_{0j}. The FE estimator can be obtained by applying a pooled OLS (POLS) estimator (pooling over the Level-2 units) to the de-meaned model, giving

$$\widehat{\boldsymbol{\beta}}_{\text{FE}} = \boldsymbol{\beta} + \left(J^{-1} \sum_{j=1}^{J} \ddot{\mathbf{X}}_j' \ddot{\mathbf{X}}_j \right)^{-1} \left(J^{-1} \sum_{j=1}^{J} \ddot{\mathbf{X}}_j' (\ddot{\mathbf{Z}}_j \mathbf{u}_j + \ddot{\boldsymbol{\epsilon}}_j) \right),$$

where $\sum_j \ddot{\mathbf{X}}_j' \ddot{\mathbf{X}}_j$ is assumed to be nonsingular with probability 1.

To derive the conditions for unbiasedness, stack the $\ddot{\mathbf{X}}_j$ for each cluster j in $\ddot{\mathbf{X}}$. Consider the conditional expectation of $\widehat{\boldsymbol{\beta}}_{\text{FE}}$, given $\ddot{\mathbf{X}}$,

$$\text{E}(\widehat{\boldsymbol{\beta}}_{\text{FE}} | \ddot{\mathbf{X}}) = \boldsymbol{\beta} + \left(J^{-1} \sum_{j=1}^{J} \ddot{\mathbf{X}}_j' \ddot{\mathbf{X}}_j \right)^{-1} \left(J^{-1} \sum_{j=1}^{J} \ddot{\mathbf{X}}_j' [\ddot{\mathbf{Z}}_j \text{E}(\mathbf{u}_j | \ddot{\mathbf{X}}) + \text{E}(\ddot{\boldsymbol{\epsilon}}_j | \ddot{\mathbf{X}})] \right). \quad (3.10)$$

Keep in mind that $\ddot{\mathbf{Z}}_j$ is a subset of $\ddot{\mathbf{X}}_j$, and note that $\text{E}(\mathbf{u}_j|\ddot{\mathbf{X}}) = \text{E}(\mathbf{u}_j|\ddot{\mathbf{X}}_j)$ and $\text{E}(\ddot{\boldsymbol{\epsilon}}_j|\ddot{\mathbf{X}}) = \text{E}(\ddot{\boldsymbol{\epsilon}}_j|\ddot{\mathbf{X}}_j)$ because clusters are independent. Furthermore, unit-level exogeneity implies that $\text{E}(\ddot{\boldsymbol{\epsilon}}_j|\ddot{\mathbf{X}}_j) = \mathbf{0}$, and cluster-level exogeneity implies that $\text{E}(\mathbf{u}_j|\ddot{\mathbf{X}}_j) = \mathbf{0}$, so $\widehat{\boldsymbol{\beta}}_{\text{FE}}$ is conditionally unbiased, $\text{E}(\widehat{\boldsymbol{\beta}}_{\text{FE}}|\ddot{\mathbf{X}}) = \boldsymbol{\beta}$. Finally, using the law of iterated expectations, it follows that $\widehat{\boldsymbol{\beta}}_{\text{FE}}$ is (unconditionally) unbiased, $\text{E}(\widehat{\boldsymbol{\beta}}_{\text{FE}}) = \boldsymbol{\beta}$.

(Wooldridge, 2005, 2010, Sec.11.7.3) considers FE estimation for a special case of Equation 3.3 without cluster-level covariates $\mathbf{W}_j$. He derives the conditions required for consistency of the traditional FE estimator, and we briefly derive the analogous results for the general model in Equation 3.3 in Appendix I. The condition for consistency, $p\text{lim}\, \widehat{\boldsymbol{\beta}}_{\text{FE}} = \boldsymbol{\beta}$, is

$$\text{E}(\ddot{\mathbf{X}}_j' \ddot{\mathbf{Z}}_j \mathbf{u}_j) = \mathbf{0}. \quad (3.11)$$

As pointed out by (Wooldridge, 2010, p. 382), this assumption allows $\mathbf{u}_j$ to be correlated with the "permanent" components of $\mathbf{X}_j$, but not the "idiosyncratic" components $\ddot{\mathbf{X}}_j$. Condition 3.11 is implied by the more easily interpretable condition for unbiasedness,

$$\mathrm{E}(\mathbf{u}_j|\ddot{\mathbf{X}}_j) \;=\; \mathbf{0},$$

because $\mathrm{E}(\ddot{\mathbf{X}}_j'\ddot{\mathbf{Z}}_j\mathbf{u}_j|\ddot{\mathbf{X}}_j) = \ddot{\mathbf{X}}_j'\ddot{\mathbf{Z}}_j\mathrm{E}(\mathbf{u}_j|\ddot{\mathbf{X}}_j)$.

We now look at the assumption in Equation 3.11, which is required for consistency and unbiasedness, in more detail for our specific model for the empirical example in Equation 3.2 to understand how it can be violated. In that model, $\ddot{\mathbf{X}}_j = (\ddot{\mathbf{x}}_j \;\, w_j\ddot{\mathbf{x}}_j)$ and $\ddot{\mathbf{Z}}_j = (\mathbf{0}_{n_j} \;\, \ddot{\mathbf{x}}_j)$. The condition can therefore be written as two non-trivial equations,

$$\mathrm{E}\left[\begin{pmatrix} \ddot{\mathbf{x}}_j' \\ w_j\ddot{\mathbf{x}}_j' \end{pmatrix} \ddot{\mathbf{x}}_j u_{1j}\right] = \mathbf{0}.$$

Concentrating on the first equation, we obtain

$$\mathrm{E}(\ddot{\mathbf{x}}_j'\ddot{\mathbf{x}}_j u_{1j}) = \mathrm{E}\left[\left(\sum_{i=1}^{n_j} \ddot{x}_{ij}^2\right) u_{1j}\right] = (n_j-1)\mathrm{E}(s_j^2 u_{1j}) = 0, \tag{3.12}$$

where s_j^2 is the sample variance of x_{ij} for cluster j. In other words, the condition is violated if the within-cluster variance of x_{ij} is correlated with the random slope u_{1j}. In our empirical example, it seems reasonable that more diverse schools (larger s_j^2) may be better suited to mitigate the effects of SES (smaller u_{1j}) than more homogeneous schools. We will therefore consider the problem of non-zero correlation between the within-cluster variance of x_{ij} and its random slope u_{1j} and will refer to Equation 3.11 as the "uncorrelated variance assumption." Note, however, that Equation 3.11 can also be violated in other ways. For instance, in longitudinal data, the covariate value at the initial time-point, x_{1j}, may be correlated with

u_{1j}.

3.4 New Estimators and Conditions for Unbiasedness

3.4.1 Augmented fixed-effects estimation

The FE approach does not permit estimation of the coefficients, $\boldsymbol{\gamma}$, of any cluster-level covariates, but it can be "augmented" so that it does. The augmented fixed-effects (FE+) estimator we use is similar to the estimator proposed by Hausman and Taylor (1981) to estimate effects of cluster-level covariates for two-level models with only random intercepts—and no random slopes. As pointed out by Castellano et al. (2014), the estimator discussed here has been invented and re-invented several times (e.g., Wiley, 1975; Raudenbush and Willms, 1995; Ballou et al., 2004). However, the conditions for unbiasedness for random-coefficient models have not previously been considered.

Step 1: Estimation of $\boldsymbol{\beta}$

In the first step, $\boldsymbol{\beta}$ (coefficients for all unit-level covariates) is estimated by FE. The estimator is unbiased under the uncorrelated variance assumption in Equation 3.11 and the unit-level exogeneity assumption in Equation 3.5.

Step 2: Estimation of $\boldsymbol{\gamma}$

In the second step, we obtain quasi-residuals $\mathbf{r}_j$,

$$\mathbf{r}_j \equiv \mathbf{y}_j - \mathbf{X}_j \widehat{\boldsymbol{\beta}}_{\text{FE}},$$

and estimate $\boldsymbol{\gamma}$ (coefficients for the cluster-level covariates) in the regression of $\mathbf{r}_j$ on $\mathbf{W}_j$ by POLS. Substituting Equation 3.3 for $\mathbf{y}_j$, we obtain

$$\mathbf{r}_j \;=\; \mathbf{W}_j\boldsymbol{\gamma} + \mathbf{X}_j(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{\text{FE}}) + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j.$$

The POLS estimator of $\boldsymbol{\gamma}$ can be expressed as

$$\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + \left(J^{-1}\sum_{j=1}^{J}\mathbf{W}_j'\mathbf{W}_j\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\mathbf{W}_j'\left[\mathbf{X}_j(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{\text{FE}}) + \mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j\right]\right), \tag{3.13}$$

where $\sum_j \mathbf{W}_j'\mathbf{W}_j$ is assumed to be nonsingular with probability 1.

The conditional expectation of the POLS estimator $\widehat{\boldsymbol{\gamma}}$ given $\mathbf{V}$ becomes

$$\begin{aligned}\mathrm{E}(\widehat{\boldsymbol{\gamma}}|\mathbf{V}) =& \boldsymbol{\gamma} + \left(J^{-1}\sum_{j=1}^{J}\mathbf{W}_j'\mathbf{W}_j\right)^{-1} \\ &\left(J^{-1}\sum_{j=1}^{J}\mathbf{W}_j'\left\{\mathbf{X}_j[\boldsymbol{\beta} - \mathrm{E}(\widehat{\boldsymbol{\beta}}_{\text{FE}}|\mathbf{V})] + \mathbf{Z}_j\mathrm{E}(\mathbf{u}_j|\mathbf{V}) + \mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V})\right\}\right).\end{aligned}$$

Conditional unbiasedness, $\mathrm{E}(\widehat{\boldsymbol{\gamma}}|\mathbf{V}) = \boldsymbol{\gamma}$, follows since (i) $\mathrm{E}(\widehat{\boldsymbol{\beta}}_{\text{FE}}|\mathbf{V})] = \boldsymbol{\beta}$ under the uncorrelated variance assumption, (ii) $\mathrm{E}(\mathbf{u}_j|\mathbf{V}) = \mathrm{E}(\mathbf{u}_j|\mathbf{V}_j)$ and $\mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V}) = \mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V}_j)$ due to independence of the clusters, and (iii) $\mathrm{E}(\mathbf{u}_j|\mathbf{V}_j) = \mathbf{0}$ and $\mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V}_j) = \mathbf{0}$ under the exogeneity assumptions. Using the law of iterated expectations, we finally obtain (unconditional) unbiasedness $\mathrm{E}(\widehat{\boldsymbol{\gamma}}) = \boldsymbol{\gamma}$.

To implement the FE+ approach for the empirical example, the two steps are: (1) estimate β_1 and β_2 by FE and (2) regress quasi-residuals, $r_{ij} \equiv y_{ij} - \widehat{\beta}_{1\text{FE}}x_{ij} - \widehat{\beta}_{2\text{FE}}w_jx_{ij}$, on w_j to obtain estimates of γ_0 and γ_1.

3.4.2 Per-cluster regression estimation

In this section, we define our proposed per-cluster regression estimator, or, in shorthand, the "PC" estimator. This estimator is best understood by using two-stage formulation (see Section 2.1) of the general model in Equation 3.3, which requires some new notation.

The columns of the matrix $\mathbf{X}_j$ of unit-level covariates can be ordered so that the matrix can be decomposed as $\mathbf{X}_j = (\mathbf{X}_{1j}\ \mathbf{X}_{2j}\ \mathbf{X}_{3j})$, where $\mathbf{X}_{1j}$ are the R_1 unit-level covariates that have random slopes, so that $\mathbf{Z}_j = (\mathbf{1}_{n_j}\ \mathbf{X}_{1j})$, $\mathbf{X}_{2j}$ are R_2 cross-level interactions between unit-level covariates in $\mathbf{X}_{1j}$ and cluster-level covariates in $\mathbf{W}_j$, and $\mathbf{X}_{3j}$ are the R_3 remaining unit-level covariates (which may include cross-level interactions between covariates in $\mathbf{W}_j$ and other covariates in $\mathbf{X}_{3j}$). Correspondingly, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\beta}_3')'$ and $R = R_1 + R_2 + R_3$. The two-stage model can then be written as

$$\mathbf{y}_j = \mathbf{X}_{3j}\boldsymbol{\beta}_3 + \mathbf{Z}_j\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j \tag{3.14}$$

$$\eta_{rj} = \mathbf{w}_{rj}'\boldsymbol{\alpha}_r + u_{rj}, \quad r = 0, \ldots, R_1, \tag{3.15}$$

where $\boldsymbol{\eta}_j$ are cluster-specific coefficients. Here, Equation 3.14 is referred to as the Level-1 or unit-level model, and Equations 3.15 for the $R_1 + 1$ cluster-specific coefficients of the Level-1 model are referred to as the Level-2 or cluster-level models. When $r = 0$, Equation 3.15 is the model for the cluster-specific intercept η_{0j}, $\mathbf{w}_{0j}'$ is a row of $\mathbf{W}_j$ (with $\mathbf{W}_j = \mathbf{1}_{nj} \otimes \mathbf{w}_{0j}'$), and $\boldsymbol{\alpha}_0 = \boldsymbol{\gamma}$. When $r > 0$, Equation 3.15 is the model for the cluster-specific slope of the rth column of $\mathbf{X}_{1j}$. The vector, $\mathbf{w}_{rj}$, includes only those cluster-level covariates that interact with the rth column of $\mathbf{X}_{1j}$. The first element of $\boldsymbol{\alpha}_r$ is β_{1r} and the other elements are subsets of $\boldsymbol{\beta}_2$. Note that we include Level-2 models only for the coefficients of those unit-level covariates that have random slopes.

Equations 3.14 and 3.15 are a generalization of the two-stage formulation of the model for the HSB example in Equations 3.1 and 3.2. In that model, there are $R_1 = 1$ covariates $\mathbf{X}_{1j} = \mathbf{x}_j$ with random slopes and $\boldsymbol{\beta}_1 = \beta_1$. There are also $R_2 = 1$ cross-level interactions between unit-level covariates in $\mathbf{X}_{1j}$ and a cluster-level covariate w_j, namely $\mathbf{X}_{2j} = w_j\mathbf{x}_j$ with $\boldsymbol{\beta}_2 = \beta_2$. There are no remaining columns of $\mathbf{X}_j$, so $R_3 = 0$, and there is no $\mathbf{X}_{3j}$ or $\boldsymbol{\beta}_3$. Further, $\boldsymbol{\eta}_j = (\eta_{0j}, \eta_{1j})'$ and $\mathbf{w}_{0j} = \mathbf{w}_{1j} = (1, w_j)'$, $\boldsymbol{\alpha}_0 = (\gamma_0, \gamma_1)'$, and $\boldsymbol{\alpha}_1 = (\beta_1, \beta_2)'$.

Step 1: Estimation of $\boldsymbol{\beta}_3$

(Wooldridge, 2005, 2010, Sec.11.7.2) considers the special case of this model without cluster-level covariates, i.e., with $\boldsymbol{\eta}_j = \boldsymbol{\beta}_1 + \mathbf{u}_j$ and describes estimation of $\boldsymbol{\beta}_3$ (coefficients for all unit-level covariates without random slopes) by an extension of the de-meaning transformation used in FE estimation. Instead of pre-multiplying by the de-meaning operator $\mathbf{Q}_j$, we pre-multiply the Level-1 model by the projection matrix

$$\mathbf{M}_j \equiv \mathbf{I}_{n_j} - \mathbf{Z}_j(\mathbf{Z}_j'\mathbf{Z}_j)^{-1}\mathbf{Z}_j'.$$

Defining $\dot{\mathbf{y}}_j = \mathbf{M}_j\mathbf{y}_j$, $\dot{\mathbf{X}}_{3j} = \mathbf{M}_j\mathbf{X}_{3j}$, and $\dot{\boldsymbol{\epsilon}}_j = \mathbf{M}_j\boldsymbol{\epsilon}_j$, and noting that $\dot{\mathbf{Z}}_j = \mathbf{M}_j\mathbf{Z}_j = \mathbf{0}$, gives

$$\dot{\mathbf{y}}_j \;=\; \dot{\mathbf{X}}_{3j}\boldsymbol{\beta}_3 + \dot{\boldsymbol{\epsilon}}_j.$$

The POLS estimator of $\boldsymbol{\beta}_3$, denoted $\widehat{\boldsymbol{\beta}}_{3\text{CML}}$, can be expressed as

$$\widehat{\boldsymbol{\beta}}_{3\text{CML}} = \boldsymbol{\beta}_3 + \left(J^{-1}\sum_{j=1}^{J}\dot{\mathbf{X}}_{3j}'\dot{\mathbf{X}}_{3j}\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\dot{\mathbf{X}}_{3j}'\dot{\boldsymbol{\epsilon}}_j\right),$$

where $\sum_j \dot{\mathbf{X}}'_{3j}\dot{\mathbf{X}}_{3j}$ is assumed to be nonsingular with probability 1. If the ϵ_{ij} are normally distributed, this estimator also corresponds to the conditional maximum likelihood estimator (CML), conditioning on the sufficient statistics $\mathbf{Z}'_j\mathbf{y}_j$ for the "nuisance parameters" $\boldsymbol{\eta}_j$ Verbeke et al. (2001).

The conditional expectation of $\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}$, given all covariates $\mathbf{V}$, becomes

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V}) = \boldsymbol{\beta}_3 + \left(J^{-1}\sum_{j=1}^{J}\dot{\mathbf{X}}'_{3j}\dot{\mathbf{X}}_{3j}\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\dot{\mathbf{X}}'_{3j}\mathrm{E}(\dot{\boldsymbol{\epsilon}}_j|\mathbf{V})\right).$$

Unit-level exogeneity, which implies that $\mathrm{E}(\dot{\boldsymbol{\epsilon}}_j|\mathbf{V}) = \mathbf{0}$, is a sufficient condition for conditional unbiasedness $\mathrm{E}(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V}) = \boldsymbol{\beta}_3$.

Step 2: Estimation of $\boldsymbol{\eta}_j$

Next, form quasi-residuals as

$$\mathbf{r}_j \equiv \mathbf{y}_j - \mathbf{X}_{3j}\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}$$

and then obtain OLS estimates $\breve{\boldsymbol{\eta}}_j$ for the regressions of $\mathbf{r}_j$ on $\mathbf{Z}_j$ for each cluster, $j = 1, \ldots, J$,

$$\breve{\boldsymbol{\eta}}_j = (\mathbf{Z}_j\mathbf{Z}'_j)^{-1}\mathbf{Z}'_j\mathbf{r}_j \;=\; \left(n_j^{-1}\sum_{i=1}^{n_j}\mathbf{z}_{ij}\mathbf{z}'_{ij}\right)^{-1}\left(n_j^{-1}\sum_{i=1}^{n_j}\mathbf{z}_{ij}r_{ij}\right), \tag{3.16}$$

where $\mathbf{z}'_{ij}$ is the ith row of $\mathbf{Z}_j$, and $\sum_{i=1}^{n_j}\mathbf{z}_{ij}\mathbf{z}'_{ij}$ is nonsingular with probability 1, which requires that $R_1+1 \leq n_j$. This step gives rise to the name "per-cluster regressions." Identical estimates of $\boldsymbol{\beta}_3$ and $\boldsymbol{\eta}_j$ are obtained by treating $\boldsymbol{\eta}_j$ as fixed parameters in Equation 3.14 via the inclusion of interactions between dummy variables for clusters and the columns of $\mathbf{Z}_j$.

The estimator in Equation 3.16 can alternatively be expressed as

$$\breve{\boldsymbol{\eta}}_j \;=\; \boldsymbol{\eta}_j + (\mathbf{Z}_j\mathbf{Z}_j')^{-1}\mathbf{Z}_j' \left[\mathbf{X}_{3j}(\boldsymbol{\beta}_3 - \widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}) + \boldsymbol{\epsilon}_j\right], \qquad (3.17)$$

and the conditional expectation of $\breve{\boldsymbol{\eta}}_j$, given $\mathbf{V}$, becomes

$$\mathrm{E}(\breve{\boldsymbol{\eta}}_j|\mathbf{V}) = \boldsymbol{\eta}_j + (\mathbf{Z}_j\mathbf{Z}_j')^{-1}\mathbf{Z}_j' \left\{\mathbf{X}_{3j}[\boldsymbol{\beta}_3 - \mathrm{E}(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V})] + \mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V})\right\},$$

where the $\mathbf{Z}_j\mathbf{Z}_j'$ are assumed to be nonsingular with probability 1. Because $\mathrm{E}(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V}) = \boldsymbol{\beta}_3$ from Step 1, it follows that $\mathbf{X}_{3j}[\boldsymbol{\beta}_3 - \mathrm{E}(\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}|\mathbf{V})] = \mathbf{0}$. It follows from unit-level exogeneity that $\mathrm{E}(\boldsymbol{\epsilon}_j|\mathbf{V}) = \mathbf{0}$, and therefore the $\breve{\boldsymbol{\eta}}_j$ are conditionally unbiased; $\mathrm{E}(\breve{\boldsymbol{\eta}}_j|\mathbf{V}) = \boldsymbol{\eta}_j$.

Step 3: Estimation of $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_1$, and $\boldsymbol{\beta}_2$

The remaining regression coefficients $\boldsymbol{\gamma}$ (for cluster-level covariates), $\boldsymbol{\beta}_1$ (for unit-level covariates with random slopes), and $\boldsymbol{\beta}_2$ (for cross-level interactions involving unit-level covariates with random slopes) are now estimated. These coefficients are contained in the vectors $\boldsymbol{\alpha}_r$, $r = 0, \ldots, R_1$, in Equation 3.15. We write each Level-2 equation for all clusters using the following vector notation. Let $\boldsymbol{\eta}_r^* = (\eta_{r1}, \ldots, \eta_{rJ})'$ and $\mathbf{u}_r^* = (u_{r1}, \ldots, u_{rJ})'$ and let $\mathbf{W}_r^*$ have J rows $\mathbf{w}_{rj}'$, $j = 1, \ldots, J$. The Level-2 equation for each $\boldsymbol{\eta}_r^*$ can then be written as

$$\boldsymbol{\eta}_r^* \;=\; \mathbf{W}_r^*\boldsymbol{\alpha}_r + \mathbf{u}_r^*, \quad r = 0, \ldots, R_1.$$

Denoting the vector of estimates $\breve{\boldsymbol{\eta}}_r^* \equiv (\breve{\eta}_{r1}, \ldots, \breve{\eta}_{rJ})'$, the model can be written as

$$\breve{\boldsymbol{\eta}}_r^* \;=\; \mathbf{W}_r^*\boldsymbol{\alpha}_r + \mathbf{u}_r^* + \breve{\boldsymbol{\eta}}_r^* - \boldsymbol{\eta}_r^*$$

We estimate $\boldsymbol{\alpha}_r$ by applying OLS to the regression of $\breve{\boldsymbol{\eta}}_r^*$ on $\mathbf{W}_r^*$, giving

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_r &= \boldsymbol{\alpha}_r + (\mathbf{W}_r^{*\prime}\mathbf{W}_r^*)^{-1}\mathbf{W}_r^{*\prime}(\mathbf{u}_r^* + \breve{\boldsymbol{\eta}}_r^* - \boldsymbol{\eta}_r^*) \\ &= \boldsymbol{\alpha}_r + \left(J^{-1}\sum_{j=1}^{J}\mathbf{w}_{rj}\mathbf{w}_{rj}'\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\mathbf{w}_{rj}(u_{rj} + \breve{\eta}_{rj} - \eta_{rj})\right),\end{aligned}$$

where, for each r, $\sum_j \mathbf{w}_{rj}\mathbf{w}_{rj}'$ is assumed to be nonsingular with probability 1.

The conditional expectation of $\widehat{\boldsymbol{\alpha}}_r$, given $\mathbf{V}$, is

$$\mathrm{E}(\widehat{\boldsymbol{\alpha}}_r|\mathbf{V}) \;=\; \boldsymbol{\alpha}_r + \left(J^{-1}\sum_{j=1}^{J}\mathbf{w}_{rj}\mathbf{w}_{rj}'\right)^{-1}\left(J^{-1}\sum_{j=1}^{J}\mathbf{w}_{rj}\left[\mathrm{E}(u_{rj}|\mathbf{V}) + \mathrm{E}(\breve{\eta}_{rj}|\mathbf{V}) - \eta_{rj}\right]\right).$$

It follows from cluster-level exogeneity that $\mathrm{E}(u_{rj}|\mathbf{V}) = 0$ and from the results for Step 2 that $\mathrm{E}(\breve{\eta}_{rj}|\mathbf{V}) = \eta_{rj}$. Hence, $\mathrm{E}(\widehat{\boldsymbol{\alpha}}_r|\mathbf{V}) \;=\; \boldsymbol{\alpha}_r$, and using the law of iterated expectations, we see that the estimator is unbiased; $\mathrm{E}(\widehat{\boldsymbol{\alpha}}_r) \;=\; \boldsymbol{\alpha}_r$.

For the special case of our model with $\boldsymbol{\eta}_j = \boldsymbol{\beta}_1 + \mathbf{u}_j$, the estimator for $\boldsymbol{\beta}$ becomes the sample mean of $\breve{\boldsymbol{\eta}}_r^*$ and that estimator has been proposed by (Wooldridge, 2010, equation (11.80)). In models in which $\mathbf{X}_j = (\mathbf{X}_{1j}\ \mathbf{X}_{2j})$, or $R_3 = 0$, the first step can be skipped and $\mathbf{r}_j = \mathbf{y}_j$. Our empirical illustration is an example of the latter special case. Accordingly, we first estimate η_{0j} and η_{1j} in model (3.2) for each cluster j by regressing $\mathbf{y}_j$ on $\mathbf{x}_j$ using OLS, giving unbiased estimates $\breve{\eta}_{0j}$ and $\breve{\eta}_{1j}$. Identical estimates are obtained by OLS with dummy variables for clusters and interactions between these dummy variables and x_{ij}. Next, $\breve{\eta}_{0j}$ and $\breve{\eta}_{1j}$ are both regressed on w_j using OLS. In the regression for $\breve{\eta}_{0j}$, the OLS estimator for the intercept is unbiased for γ_0 and the OLS estimator for the coefficient of w_j is unbiased for γ_1. In the regression for $\breve{\eta}_{1j}$, the OLS estimator for the intercept is unbiased for β_1 and the OLS estimator for the coefficient of w_j is unbiased for β_2. If we did not include the cross-level interaction term, $x_{ij}w_j$, in our model, there would be no β_2, $R_1 = R$ and we would regress $\breve{\eta}_{1j}$ on just the intercept, i.e., find its sample mean, to obtain the unbiased estimate of β_1.

3.5 Empirical Example

To ground comparisons of our estimators of interest, we apply each to the HSB data introduced in Section 3.2.1. Table D.1 provides estimates of the regression coefficients for Equation 3.2. All estimates were obtained using standard commands in Stata 13 StataCorp (2013), such as `mixed` and `xtreg` (see Appendix I). Note that the RE estimate of the correlation between the random intercept and slope was 1, a relatively frequent occurrence in random-coefficient models Chung et al. (2014).

Castellano et al. (2014) show that positive correlation between a random intercept and a student-level covariate leads to overestimation of the coefficient of the covariate. Indeed, from the HSB data results presented in Table D.1, we see that RE produces the largest estimate of the coefficient of SES, 2.958, approximately 6% higher than the closest estimate (FE+). The indicator variable w_j for Catholic schools is positively correlated with SES and therefore over-estimation of the coefficient of SES is accompanied by underestimation of the coefficient of w_j, with RE producing the smallest estimate of γ_1, at 2.130.

While the differences in the FE+ and RE estimates of γ_1 may be practically significant, they are close in magnitude to the estimated standard errors of the coefficient estimates. FE+ produces estimates of both β_1 and γ_1 that lie between the estimates produced by RE and PC, which is intuitive given that FE+ relies only on the uncorrelated variance assumption, whereas RE additionally requires exogeneity, and PC requires neither assumption. PC gives the smallest estimated effect of SES on math achievement scores ($\widehat{\beta}_1 = 2.772$) and the largest estimated effect of Catholic schooling ($\widehat{\gamma}_1 = 2.253$). These estimates differ by about 6% from the RE counterparts, enough to give practitioners pause.

The small difference between estimates of β_1 from FE+ and PC provides evidence that, in this case, we may be able to ignore the possibility that the within-school variance in SES is correlated with the random slope. In fact, the within-school standard deviation of SES has a correlation of only 0.04 with the estimated residuals from the regression of $\breve{\eta}_{1j}$ on w_j in the final step of the PC approach.

3.6 Simulation Study

We now conduct a simulation study to investigate the performance of the RE, FE+, and PC estimators. In particular, we are interested in the amount of bias for RE and FE+ when the respective assumptions of cluster-level exogeneity and uncorrelated variance are violated. We also evaluate all three estimators, RE, FE+, and PC, by their root mean square errors and consider performance of the estimated standard errors. We use Stata 13 StataCorp (2013) throughout.

3.6.1 Data Generation Process

We generate the data using our model of interest in Equation 3.2. We first draw the school-level variables for each of $J = 100$ clusters. The random intercepts u_{0j} and random slopes u_{1j} are drawn from a bivariate normal distribution with zero means and variance-covariance matrix defined by variances $\psi_0 = 0.4^2$ and $\psi_1 = 0.25^2$ and correlation $\rho = 0.5$, giving the covariance $\psi_{10} = 0.05$. We specify these variances to reflect those found in our empirical example. The exogenous school-level covariate w_j is drawn independently from a normal distribution with mean 1.7 and variance $\sigma_w^2 = 1$.

We then generate the student-level covariate x_{ij} as

$$x_{ij} \;=\; b_0 u_{0j} + b_1 u_{1j} + b_2 w_j + a e_{ij}, \quad e_{ij} \sim N(0, \sigma_j), \tag{3.18}$$

where

$$a \;=\; \sqrt{1 - \psi_0 b_0^2 - \psi_1^2 b_1^2 - \sigma_w^2 b_2^2 - 2 b_0 b_1 \psi_{10}}.$$

Here, $b_0 = 1.33$, $b_1 = 2.13$, and $b_2 = 0.20$ so that x_{ij} is positively correlated with the random intercept, random slope, and school-level covariate w_j. Finally, we generate y_{ij} according to Equation 3.2 with $\gamma_0 = 1$, $\gamma_1 = 3$, $\beta_1 = 1$, and $\beta_2 = 2$.

The key assumption under which we want to assess the performance of the competing estimators is that the sample within-cluster variance s_j^2 of x_{ij} is uncorrelated with the random slope u_{1j}. Thus, the population within-cluster standard deviation, σ_j, is of particular importance. Accordingly, the uncorrelated variance assumption factor in this simulation has 2 levels: when it holds, $\sigma_j = 1$, and when it is violated, $\sigma_j = \exp(u_{1j})$.

Although our empirical example involves schools, which tend to have large numbers of students, both RE and FE are commonly used with classrooms serving as clusters. Furthermore, there are numerous relevant applications with longitudinal data where we often find even smaller cluster sizes. Thus, we also vary cluster size, primarily considering clusters sizes of 4 and 20. For simplicity, we set cluster sizes equal across clusters, $n_j = n$. We fully cross the cluster size and uncorrelated-variance-assumption factors, yielding four primary simulation conditions defined by: (large/small n) $\times$ (uncorrelated variance assumption holds/violated). To further determine the effect of cluster size when the uncorrelated variance assumption is violated, we also consider a range of clusters sizes from 4 to 50: $n = 4, 8, 14, 20, 50$.

All conditions are replicated 500 times. Due to occasional lack of variation of x_{ij} within some small clusters, the PC approach fails for some replications. The lowest number of successful replications is 489, which occurs when the variance of x_{ij} is correlated with the random slopes, and we have only 4 observations in each cluster. For all simulation conditions with a cluster size of 20, all 500 replications are successful.

3.6.2 Results

We evaluate the performance of each of our three estimators (RE, FE+, and PC) of the fixed regression coefficients in our model of interest (Equation 3.2) across our four simulation conditions. The estimated bias and root mean square error (RMSE) are given in Table D.2. Appendix G provides supplemental tables for each coefficient that also include the mean standard errors, standard deviations of the estimates, and the ratios of these values.

3.6.2.1 Bias

For β_1, the coefficient of the endogenous student-level covariate x_{ij}, there are three main results. First, the PC estimator is unbiased across all conditions even when the uncorrelated variance assumption is violated. Figure C.1 clearly illustrates this finding as the empirical distributions of the errors (i.e., estimate − parameter) of the PC estimator (the solid curves) are centered on 0 in all four panels, where each panel represents one of the four simulation conditions.

Second, the RE estimator is biased regardless of whether the uncorrelated variance assumption holds, whereas the FE+ estimator is biased only when this assumption is violated.

This result for RE is expected given that the RE estimator relies on the assumption of both unit- and cluster-level exogeneity (see Section 3.2), and cluster-level exogeneity is violated in all four conditions with the nonzero correlation between x_{ij} and both the random intercept and its random slope. We do note, however, that violation of the uncorrelated variance assumption exacerbates the magnitude of the RE estimator's bias: for the small cluster size condition ($n = 4$), the estimated bias is 1.28 times as large, and for the larger cluster size ($n = 20$), the estimated bias more than doubles as shown in the first column of results in Table D.2. In contrast, the FE+ approach only requires unit-level exogeneity assumptions, and thus produces unbiased estimates under cluster-level endogeneity as long as there is no correlation between the random slopes and within-cluster variance of x_{ij}. This is evident in Figure C.1 by observing that the curves for FE+ (dashed) are more similar to those for PC (solid) in the left-hand plots (for uncorrelated variance simulation conditions) and more similar to the curves for RE (dot-dashed) in the right-hand plots (for correlated variance simulation conditions).

Thirdly, the estimated bias for β_1 is larger than that for the other two regression coefficients, which is not surprising given that x_{ij} is the source of the endogeneity. For instance, as shown in Table D.2, the estimated bias of $\widehat{\beta}_{1_{RE}}$ ranges from 6.2%-21.3% of the true value. The next largest estimated bias is -0.053 for $\widehat{\gamma}_{1_{RE}}$ under the small clusters and uncorrelated variance condition, which is only 1.8% of the coefficient's true value ($\gamma_1 = 3$).

The coefficient of the interaction term, β_2, is the least affected by the simulation conditions. We only find statistically significant bias (at the 5% level) for $\widehat{\beta}_{2_{RE}}$ for the small cluster size condition—both when the uncorrelated variance assumption holds and when it is violated. Even in these cases, as given in Table D.2, the estimated bias is rather small relative to the magnitude of the true value ($\beta_2 = 2$): it is 0.4% of the parameter value when the condition holds and 0.6% when it is violated. (Plots of the empirical distributions of the estimation errors for β_2 are given in Figure I.1 in Appendix I.)

The estimated biases of the estimators for the coefficient γ_1 of the exogenous school-level covariate w_j follow similar patterns as for the coefficient β_1 of the endogenous student-level covariate x_{ij}. Just as for β_1, the PC estimator is unbiased across all conditions, the FE+ estimator is biased only when the variance of x_{ij} is correlated with the random slope u_{1j} (i.e., uncorrelated variance assumption violated), and the RE estimator is biased regardless of whether the uncorrelated variance assumption is violated. These findings are clearly illustrated in Figure C.2 by comparing the centers of the empirical distributions of errors for all estimators across all conditions: the PC curve (solid) is always centered on 0, whereas the RE curve (dot-dashed) is always centered below 0, and the FE+ curve (dashed) is centered below 0 only for the correlated variance conditions in the right-hand panels. Just as with β_1, the FE+ estimator's bias for γ_1 does not vary with cluster size—it's estimate is about 0.8% of the true parameter value for both $n = 4$ and $n = 20$ as seen in Table D.2. Cluster size affects the RE estimator's bias for γ_1 as it did for β_1: as cluster size increases, the bias decreases. When the assumption holds, this bias decreases by about 63% going from $n = 4$ to $n = 20$, and by about 45% when the assumption is violated (see Table D.2).

Given that β_1 was most affected by the violation of the uncorrelated variance assumption, we further investigated the effect of cluster size on this regression coefficient. Figure C.1 gives the estimated bias for each estimator across cluster sizes of 4, 8, 14, 20, and 50. The PC (solid) curve hugs the $y = 0$ line. The FE+ and RE curves cross at $n = 20$: as cluster size increases, the RE estimator's bias decreases (dot-dashed curve), whereas the FE+ estimator's bias is not as affected by cluster size, shown by its dot-dashed curve staying relatively constant across the range of cluster sizes. Thus, cluster size has a differential effect on the bias of the estimators.

When using bias to evaluate the estimators, our simulation study provides strong evidence that our proposed PC estimator outperforms the other estimators.

3.6.2.2 Precision and RMSE

As is often the case, there is a trade-off between bias and precision, which depends in part on the size of the clusters. The rank ordering of the estimators by their standard deviation (SD) is approximately the same for the three regression coefficients with slight differences between the smaller and larger cluster size conditions. Accordingly, we discuss how the precision of the estimators depends on cluster size, without distinguishing among the coefficients.

For the smaller cluster sizes of $n = 4$, RE produces the estimates with the smallest variances, followed by FE+, and PC produces the most variable estimates. This is clearly illustrated by comparing the widths of the empirical distributions of errors in Figure C.1 or C.2 for each estimator: the RE curves are the narrowest and the PC curves are the widest. For instance, for $n = 4$ and when the uncorrelated variance assumption is violated, the SD of $\widehat{\beta}_{1_{PC}}$ over 500 replications is about 0.266, whereas for RE, the SD is less than half that at about 0.114 (see Tables I.1, I.2, and I.3 in Appendix I for all SD values).

For the larger cluster size of $n = 20$, RE always yields the smallest variances, but the variances are not much smaller than those for FE+ and PC, which tend to have about equal variances. For instance, for β_1, for the large clusters and uncorrelated variance condition shown in the lower, left-hand panel of Figure C.1, it is difficult to discern any differences in the widths of the distributions. Indeed, the SD for RE, in this case, is about 0.078 and the SDs for both FE+ and PC are 0.080.

With regard to precision, RE consistently outperforms FE+ and PC for all the regression coefficients and across all the simulation conditions. However, given the tradeoff between bias and precision, it is useful to evaluate the estimators with regard to their RMSEs, which takes both bias and imprecision into account. Given that the estimates of β_1 are the most affected by the simulation conditions and that precision depends on cluster size, we consider the RMSEs as a function of the extended range of cluster sizes for β_1 in Figure C.4. Just as with bias in Figure C.3 , Figur C.4 shows that the FE+ and RE curves cross with RE outperforming FE+ as cluster size increases. This figure also shows that, for the smallest cluster size of 4, the RMSE for PC is large and similar to that of RE. However, with clusters of at least 8, the PC estimator outperforms both RE and FE+ with regards to RMSE, providing strong evidence in favor of the PC estimator.

3.6.2.3 Standard Error Estimation

As a final point, we evaluate the estimators in terms of how well their estimated standard errors (SEs) approximate the sampling SDs. We again focus on the most affected regression coefficient β_1. Figure C.5 displays this ratio of mean SE to SD over the extended range of cluster sizes—similar to Figures C.3 and C.4. If the SE estimation works well, this ratio should equal one. We see that both the PC (solid curve) and RE (dot-dashed curve) approaches provide good SE estimates. In contrast, for the FE+ approach, the SEs are severely underestimated as the cluster size increases. Although both the FE+ and PC approaches treat estimated coefficients from previous steps as known in the subsequent step, it appears that underestimation of the SE is a larger problem for the FE+ approach. Accordingly, we recommend using either analytically derived or bootstrap SEs for the FE+ approach. These could also be used for the PC approach, and may be necessary if Step 1 is required.

3.7 Conclusion

Given the popularity of multilevel models, studies that investigate potential biases for key parameters and provide simple solutions are clearly important.

We have shown that commonly used random- and fixed-effects estimators are biased in the presence of correlation between random-effects and the within-cluster variance of unit-level covariates. Further, such bias can spill over to the estimation of coefficients of other covariates. We have proposed a new per-cluster regression estimator that avoids such bias, produces good estimates of SEs, and generally has low RMSE. Consequently, we recommend broad use of per-cluster regression when working with longitudinal or nested cross-sectional data when the clusters are sufficiently large. Stata code for applying this method to the HSB data is provided in Appendix I. In instances where the cluster sizes are small relative to the number of random effects, or where estimates for the random part of the model are of interest, we recommend using per-cluster regression as part of a sensitivity analysis for alternative estimators.

Per-cluster methods have been used in the past for linear multilevel models (Burstein et al., 1978, p. 369) and multilevel structural equation models Chou et al. (2000). Per-cluster methods can also be used for nonlinear multilevel models, such as probit models with random intercepts Borjas and Sueyoshi (1994) and logit models with random intercepts and slopes Korn and Whittemore (1979). However, the purpose of that work was to develop simple estimators and not to address endogeneity concerns. For our proposed PC estimator for linear models, it might appear to be inefficient to use OLS in the final step, not taking into account that the intercepts and slopes are estimated with different precision for different clusters. However, FGLS approaches, such as those discussed by Berkey et al. (1998), suffer from similar biases as RE estimators, as we confirmed in simulations (not shown).

An alternative approach for handling endogeneity, proposed for random-intercept models by Allison and Bollen (1997) and Teachman et al. (2001) is to model the unit-level covariates jointly with the responses using structural equation modeling and allow them to be correlated with the random intercept. This approach can be generalized to random-coefficient models but becomes infeasible for large cluster sizes.

In summary, we have demonstrated that our proposed, simple-to-implement per-cluster regression approach outperforms standard estimators when estimating regression coefficients in multilevel models under violations of both the cluster-level exogeneity and uncorrelated variance assumptions. We recommend that researchers consider the validity of the uncorrelated variance assumption and add the PC method to their toolbox when investigating effects of covariates in cross-sectional and longitudinal analyses.

APPENDICES

Appendix A

Tables for Chapter 1

Table A.1: Average VAM of Teachers moving within and out of Winston-Salem and Guilford

		Panel A: Within District Movers			Panel B: Out of District Movers		
		1998-1999	2000-2007	2008-2010	1998-1999	2000-2007	2008-2010
Guilford	Mean	-0.166	0.093	0.246	0.116	-0.174	-0.125
	N	101	463	104	48	206	34
Winston-Salem	Mean	0.009	-0.088	0.031	-0.528	-0.100	-0.243
	N	188	275	63	26	121	21
Rest of State	Mean	-0.069	0.020	0.052	-0.116	-0.118	-0.109
	N	1882	6793	1966	962	4230	833

Note: VAMs are measured in standard deviations. Guilford first adopted VAMs in 2000. Winston-Salem first adopted VAMs in 2008.

Table A.2: Sample Summary Statistics

	Guilford		Winston-Salem		Rest of North Carolina	
	Mean	SD	Mean	SD	Mean	SD
Scaled Score	250.38	71.71	249.23	68.86	252.36	70.49
Percent Proficient	0.75	0.14	0.74	0.15	0.76	0.13
Share of Black Students	0.42	0.24	0.36	0.24	0.29	0.24
Share of Black Teachers	0.25	0.43	0.21	0.41	0.15	0.36
Share of Hispanic Teachers	0.01	0.09	0.00	0.04	0.00	0.06
Share of Teachers with Advanced Degrees	0.30	0.46	0.36	0.48	0.29	0.45
College Selectivity (Barron's)	3.95	1.43	3.92	1.68	3.93	1.44
Experience	11.59	9.76	13.36	9.71	12.19	9.85
Tenure	3.23	3.05	3.59	3.26	3.68	3.35
Job Moves	0.09	0.28	0.08	0.28	0.08	0.27
Within-District Moves	0.06	0.24	0.06	0.24	0.05	0.22
Out-of-District Moves	0.03	0.16	0.02	0.14	0.03	0.16
Left NCPS	0.06	0.23	0.04	0.20	0.06	0.24
VAM	0.02	1.01	0.01	0.99	0.00	1.00
N	11,239		8,295		216,484	

Note: VAM is measured in standard deviations with the mean centered at 0.
Tenure is generated, and is censored for those already working at a given school in 1995.

Table A.3: Probability of Moving Schools Within and Out of District

	Panal A: Within-District Moves			Panal B: Out-Of-District Moves		
VARIABLES	Total	To a higher performing school	To a lower performing school	Total	To a higher performing school	To a lower performing school
VAM	**0.0016**	**0.0032*****	**-0.0016****	**0.0002**	**0.0014****	**-0.0012****
	[0.00129]	[0.00091]	[0.00074]	[0.00096]	[0.00072]	[0.00058]
VAM x Treatment GCS	**0.0058****	**0.0051****	**0.0007**	**-0.0103*****	**-0.0054*****	**-0.0049*****
	[0.00265]	[0.00199]	[0.00151]	[0.00261]	[0.00195]	[0.00156]
VAM x Treatment WSF	**0.0052***	**0.0060*****	**-0.0008**	**0.0009**	**0.0023**	**-0.0014**
	[0.00286]	[0.00229]	[0.00194]	[0.00241]	[0.00208]	[0.00129]
Treatment GCS	**-0.0040**	**-0.0050**	**0.0010**	**-0.0162*****	**-0.0232*****	**0.0070*****
	[0.00851]	[0.00571]	[0.00679]	[0.00374]	[0.00233]	[0.00268]
Treatment WSF	**0.0555*****	**0.0475*****	**0.0080*****	**-0.0020**	**0.0147*****	**-0.0167*****
	[0.00499]	[0.00372]	[0.00299]	[0.00274]	[0.00224]	[0.00178]
Observations	236,018	236,018	236,018	236,018	236,018	236,018

CSB standard errors from 500 repetitions appear in brackets. All regressions include teacher level covariates and interactions with treatment indicators, as well as year and district fixed effects. *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Effects on Sorting

VARIABLES	Total	Within District
VAM	0.0028***	0.0024***
	[0.00033]	[0.00033]
VAM x Treatment GCS	-0.0005	-0.0000
	[0.00074]	[0.0007]
VAM x Treatment WSF	0.0007	0.0017*
	[0.00114]	[0.00102]
Treatment GCS	-0.0195***	-0.0157***
	[0.00211]	[0.00216]
Treatment WSF	0.0290***	0.0231***
	[0.00172]	[0.00168]
Observations	209,424	202,943

CSB standard errors from 500 repetitions appear in brackets. All regressions use a linear functional form, include teacher level covariates, and their interactions with treatment indicators. *** p<0.01, ** p<0.05, * p<0.1

Table A.5: Effects of teacher quality index on the probability of moving

	Within-District Moves			Out-of-District Moves		
Variables	Total	To a higher performing schools	To a lower performing schools	Total	To a higher performing schools	To a lower performing schools
VAM	0.0018	0.0039***	-0.0021***	-0.0002	0.0014**	-0.0016***
	[0.00111]	[0.00078]	[0.00073]	[0.00091]	[0.00068]	[0.00053]
Teacher Quality Index (TQ Index)	0.005**	0.0071***	-0.0021**	-0.0005	0.0031***	-0.0035***
	[0.00233]	[0.00173]	[0.00105]	[0.00186]	[0.00115]	[0.00096]
VAM x Treatment GCS	0.0083***	0.0069***	0.0014	-0.0109***	-0.0053***	-0.0056***
	[0.00237]	[0.00177]	[0.0014]	[0.00249]	[0.00189]	[0.00145]
VAM x Treatment WSF	0.0063**	0.0062***	0.0000	0.0001	0.0018	-0.0017
	[0.00248]	[0.00199]	[0.00193]	[0.00212]	[0.00189]	[0.00115]
TQ Index x Treatment GCS	0.0040	0.0043**	-0.0003	0.0076***	0.0061***	0.0015*
	[0.00246]	[0.00153]	[0.00145]	[0.00116]	[0.00088]	[0.00088]
TQ Index x Treatment WSF	0.0029	0.0027	0.0002	-0.0011	-0.0026***	0.0015**
	[0.00254]	[0.00192]	[0.00131]	[0.00097]	[0.00078]	[0.00063]
Treatment GCS	0.0142**	0.0253***	-0.0111***	-0.0120***	-0.0132***	0.0011
	[0.00595]	[0.00449]	[0.00405]	[0.00258]	[0.00167]	[0.00189]
Treatment WSF	-0.0015	0.0091***	-0.0106***	0.0118***	0.0177***	-0.0059***
	[0.00383]	[0.00242]	[0.00253]	[0.00251]	(0.00136]	[0.00139]
Observations	236,018	236,018	236,018	236,018	236,018	236,018

CSB standard errors from 500 repetitions appear in brackets. All regressions use a linear functional form, and include teacher level covariates and interactions with treatment indicators. The VAMs used in this analysis are the residuals from the projection of my standard VAMs on the components of the index. *** p<0.01, ** p<0.05, * p<0.1

Table A.6: Differential Effects With Respect to Experience and Tenure

VARIABLES	Within District		Out of District	
	Total	Higher Performing	Total	Higher Performing
VAM	-0.0001	0.0028*	-0.0001	0.0023
	[0.0023]	[0.00161]	[0.00244]	[0.00173]
Experience x VAM	-0.0000	0.0000	-0.0000	-0.0000
	[0.00011]	[0.00008]	[0.00011]	[0.00008]
Tenure x VAM	0.0020**	0.0006	0.0006	0.0005
	[0.0008]	[0.00059]	[0.00073]	[0.00058]
VAM x Treatment GCS	0.0033	0.0050	-0.0181***	-0.0095*
	[0.00568]	[0.00465]	[0.00693]	[0.00514]
Experience x VAM x Treatment GCS	0.0016***	0.0010***	0.0002	0.0003
	[0.00026]	[0.0002]	[0.00032]	[0.00026]
Tenure x VAM x Treatment GCS	0.0056***	0.0004	0.0008	0.0014
	[0.00179]	[0.00146]	[0.00217]	[0.00178]
VAM x Treatment WSF	-0.0003	-0.0010	-0.0073	-0.0051
	[0.00551]	[0.00431]	[0.00503]	[0.00452]
Experience x VAM x Treatment WSF	0.0003	0.0005	0.0002	0.0002
	[0.00043]	[0.00036]	[0.00029]	[0.00025]
Tenure x VAM x Treatment WSF	0.0028***	0.0009*	0.0004	0.0004
	[0.00078]	[0.00055]	[0.00053]	[0.00046]
Observations	236,018	236,018	236,018	236,018

CSB standard errors from 500 repetitions appear in brackets. All regressions use a linear functional form, and include teacher level covariates and interactions with treatment indicators. *** p<0.01, ** p<0.05, * p<0.1

Table A.7: Probability of moving schools within-district using restricted data VAM

Panel	A: Within-District Moves			B: Out-Of-District Moves			C: School Quality Growth	
VARIABLES	Total	To a higher performing school	To a lower performing school	Total	To a higher performing school	To a lower performing school	Total	Within District
VAM	0.0003	0.0011	-0.0008	-0.0013	-0.0006	-0.0007	0.0005	0.0004
	[0.00109]	[0.00097]	[0.00063]	[0.00079]	[0.00056]	[0.00043]	[0.00032]	[0.00033]
VAM x Treatment GCS	0.0034	0.0030	0.0004	-0.0027	-0.0016	-0.0011	-0.0015	-0.0010
	[0.00249]	[0.002]	[0.00152]	[0.00201]	[0.00167]	[0.00102]	[0.00083]	[0.00076]
VAM x Treatment WSF	0.0061*	0.0099***	-0.0038*	0.0019	0.0025	-0.0005	0.0025*	0.0037***
	[0.00312]	[0.00241]	[0.00216]	[0.00247]	[0.00224]	[0.00122]	[0.00131]	[0.00109]
Treatment GCS	-0.0034	-0.0042	0.0008	-0.0137***	-0.0220***	0.0082***	-0.0196***	-0.0156***
	[0.00848]	[0.00545]	[0.00717]	[0.00365]	[0.00243]	[0.00275]	[0.0022]	[0.00225]
Treatment WSF	0.0555***	0.0486***	0.0068**	-0.0017	0.0151***	-0.0168***	0.0299***	0.0241***
	[0.00533]	[0.00386]	[0.0033]	[0.00283]	[0.00217]	[0.0019]	[0.00165]	[0.00165]
Observations	236,018	236,018	236,018	236,018	236,018	236,018	209,424	202,943

CSB standard errors from 500 repetitions appear in brackets.
All regressions include teacher level covariates and interactions with treatment indicators.
*** p<0.01, ** p<0.05, * p<0.1

Table A.8: Effect of VAMs constructed using various number of years on the probability of moving to a "better" school

VARIABLES	2yr VAM	3yr VAM	4yr VAM	5yr VAM	6yr VAM	7yr VAM	8yr VAM
VAM	0.0020***	0.0023***	0.0024***	0.0023***	0.0025***	0.0027***	0.0040***
	[0.00054]	[0.0005]	[0.00051]	[0.00073]	[0.00076]	[0.00072]	[0.00083]
VAM x Treatment Winston-Salem	0.0103***	0.0087***	0.0076***	0.0064**	0.0099***	0.0118***	0.0150***
	[0.00241]	[0.00233]	[0.00245]	[0.00287]	[0.00293]	[0.003]	[0.00323]
Treatment Winston-Salem	0.0555***	0.0540***	0.0550***	0.0480***	0.0427***	0.0457***	0.0407***
	[0.00382]	[0.00373]	[0.00362]	[0.00385]	[0.00396]	[0.00427]	[0.00434]
Observations	207,673	189,531	170,598	151,067	131,567	111,786	94,884

CSB standard errors from 500 repetitions appear in brackets. All regressions use a linear functional form, and include teacher level covariates and interactions with treatment indicators. Observations from GCS are omitted from the above analysis. *** p<0.01, ** p<0.05, * p<0.1

Table A.9: Probability of Moving to Non-Strategic-Staffing Schools

	Panal A: Within-District Moves			Panal B: Out-Of-District Moves		
VARIABLES	Total	To a higher performing school	To a lower performing school	Total	To a higher performing school	To a lower performing school
VAM	0.0014	0.0031***	-0.0018**	0.0002	0.0013*	-0.0011*
	[0.00127]	[0.00086]	[0.00076]	[0.00098]	[0.00072]	[0.00059]
VAM x Treatment GCS	0.0043*	0.0041**	0.0002	-0.0111***	-0.0054***	-0.0057***
	[0.00244]	[0.00197]	[0.00148]	[0.00248]	[0.00194]	[0.0014]
VAM x Treatment WSF	0.0100***	0.0103***	-0.0004	-0.0007	0.0014	-0.0021**
	[0.00233]	[0.00176]	[0.00148]	[0.00208]	[0.00196]	[0.00113]
Treatment GCS	-0.0118	-0.0084	-0.0034	-0.0158***	-0.0238***	0.0079***
	[0.00848]	[0.00552]	[0.00728]	[0.00362]	[0.00221]	[0.00272]
Treatment WSF	0.0241***	0.0390***	-0.0149***	-0.0027	0.0114***	-0.0141***
	[0.0049]	[0.00345]	[0.00287]	[0.00255]	[0.00233]	[0.00142]
Observations	236,018	236,018	236,018	236,018	236,018	236,018

CSB standard errors from 500 repetitions appear in brackets.
All regressions include teacher level covariates and interactions with treatment indicators.
*** p<0.01, ** p<0.05, * p<0.1

Table A.10: Effects on Sorting Within District Excluding Strategic-Staffing Schools

VARIABLES	Total	Within all district schools	Within non-strategic staffing schools
VAM	0.0028***	0.0024***	0.0026***
	[0.00033]	[0.00033]	[0.00034]
VAM x Treatment GCS	-0.0005	-0.0000	0.0009
	[0.00074]	[0.0007]	[0.00072]
VAM x Treatment WSF	0.0007	0.0017*	0.0020*
	[0.00114]	[0.00102]	[0.00114]
Treatment GCS	-0.0195***	-0.0157***	0.0029
	[0.00211]	[0.00216]	[0.00222]
Treatment WSF	0.0290***	0.0231***	0.0196***
	[0.00172]	[0.00168]	[0.0018]
Observations	209,424	202,943	197,364

CSB standard errors from 500 repetitions appear in brackets.
All regressions include teacher level covariates and interactions with treatment indicators. *** p<0.01, ** p<0.05, * p<0.1

Appendix B

Tables for Chapter 2

Table B.1: Job separations by type, reference group, and AFQT relative to reference group

Panel A: Above average reference group AFQT score

	Separations	Job-to-Job	Job-to-Unemployment	
	Full Sample	Full Sample	Recession	Expansion
Highschool Graduate	0.110	0.047	0.106	0.056
Noncompetitive College Attendee	0.112	0.052	0.096	0.053
Competitive College Attendee	0.122	0.056	0.120	0.059
Noncompetitive College Graduate	0.091	0.044	0.100	0.041
Competitive College Graduate	0.102	0.048	0.119	0.046
Total	0.109	0.048	0.107	0.053

Panel B: Below average reference group AFQT score

	Separations	Job-to-Job	Job-to-Unemployment	
	Full Sample	Full Sample	Recession	Expansion
Highschool Graduate	0.131	0.046	0.134	0.078
Noncompetitive College Attendee	0.119	0.046	0.116	0.067
Competitive College Attendee	0.127	0.053	0.114	0.068
Noncompetitive College Graduate	0.098	0.048	0.098	0.044
Competitive College Graduate	0.108	0.053	0.110	0.048
Total	0.124	0.048	0.126	0.070

Table B.2: AFQT Percentiles (Standardized by Age) by Race and Education

	Count	Mean	SD
Black High school Graduate	372	-0.349	0.718
Black Uncompetitive College Attendee	122	-0.005	0.773
Black Competitive College Attendee	65	0.454	0.757
Black Uncompetitive College Graduate	15	0.488	0.891
Black Competitive College Graduate	54	0.932	0.754
Hispanic High school Graduate	722	-0.746	0.582
Hispanic Uncompetitive College Attendee	200	-0.345	0.645
Hispanic Competitive College Attendee	89	-0.173	0.784
Hispanic Uncompetitive College Graduate	47	0.032	0.861
Hispanic Competitive College Graduate	82	0.530	0.842
White High school Graduate	1508	0.152	0.804
White Uncompetitive College Attendee	391	0.585	0.799
White Competitive College Attendee	253	0.963	0.696
White Uncompetitive College Graduate	134	1.130	0.708
White Competitive College Graduate	524	1.336	0.607
Total	4578	0.203	0.969

Table B.3: Work History

	Mean	SD	Min	Max
Experience	36.652	22.325	0.000	92.385
Potential Experience	45.038	24.262	-4.000	103.000
Tenure	14.027	15.493	0.003	85.134
Working Spell	21.188	20.007	0.003	92.070
Observations	232388			

All variables are measured in quarters.

Table B.4: Terminal Tenure

Year	N	Share	Mean	SD	Min	Max
1	15096	0.556	1.887	0.952	0.003	3.997
2	5538	0.204	5.661	1.143	4	7.999
3	2364	0.087	9.749	1.119	8	11.997
4	1317	0.049	13.852	1.168	12.005	15.996
5	827	0.030	17.832	1.171	16.003	19.997
6	522	0.019	21.899	1.154	20.002	23.986
7	366	0.013	25.791	1.189	24.007	27.989
8	293	0.011	29.885	1.128	28.008	31.993
9	200	0.007	33.793	1.166	32.013	35.997
10	158	0.006	37.941	1.206	36.013	39.986
>10	457	0.017	52.597	10.409	40.014	85.134
Total	27138	1	6.717	9.208	0.003	85.134

Table B.5: Job Separations

	Mean	SD
Recessions		
Job Separation	0.178	0.382
Job-to-Unemployment Move	0.116	0.321
Job-to-Job Transition	0.062	0.240
Observations	29557	
Expansions		
Job Separation	0.108	0.310
Job-to-Unemployment Move	0.062	0.241
Job-to-Job Transition	0.046	0.210
Observations	202831	
Full Sample		
Job Separation	0.117	0.321
Job-to-Unemployment Move	0.069	0.253
Job-to-Job Transition	0.048	0.214
Observations	232388	

Table B.6: Nominal Wage Change with Separation

	Job Separations			Job-to-Job			Job-to-Unemployment			% Difference in Wage Changes
	N	Mean	SD	N	Mean	SD	N	Mean	SD	Between JTJ and JTU Moves
High School Graduates										
$AFQT - \overline{AFQT}_{HSG} < -0.75$	1855	0.95	3.56	742	1.23	3.70	1113	0.77	3.46	59.6%
$-0.75 < AFQT - \overline{AFQT}_{HSG} < 0.75$	9832	0.88	3.64	4087	1.08	3.88	5745	0.73	3.45	47.5%
$0.75 < AFQT - \overline{AFQT}_{HSG}$	1975	1.11	4.08	951	1.29	4.38	1024	0.94	3.77	37.4%
Uncompetitive College Attendees										
$AFQT - \overline{AFQT}_{CA} < -0.75$	601	0.92	4.09	251	1.30	4.32	350	0.65	3.89	101.2%
$-0.75 < AFQT - \overline{AFQT}_{UCA} < 0.75$	2056	1.12	4.02	993	1.28	4.36	1063	0.96	3.67	33.8%
$0.75 < AFQT - \overline{AFQT}_{UCA}$	445	1.43	4.47	229	1.98	4.93	216	0.85	3.84	132.3%
Competitive College Attendees										
$AFQT - \overline{AFQT}_{CG} < -0.75$	276	0.99	4.38	123	1.50	4.61	153	0.59	4.15	154.2%
$-0.75 < AFQT - \overline{AFQT}_{CCG} < 0.75$	901	1.41	4.14	434	1.67	4.18	467	1.16	4.09	43.8%
$0.75 < AFQT - \overline{AFQT}_{CCG}$	209	1.15	4.40	105	1.24	4.76	104	1.06	4.04	16.7%
Uncompetitive College Graduates										
$AFQT - \overline{AFQT}_{UCG} < -0.75$	204	1.67	4.38	116	1.83	4.36	88	1.46	4.42	25.1%
$-0.75 < AFQT - \overline{AFQT}_{UCG} < 0.75$	1050	1.55	4.43	511	2.03	4.73	539	1.10	4.07	84.7%
$0.75 < AFQT - \overline{AFQT}_{UCG}$	195	1.44	5.14	95	1.89	6.33	100	1.02	3.66	85.5%
Competitive College Graduates										
$AFQT - \overline{AFQT}_{CCG} < -0.75$	501	1.70	4.58	268	1.78	4.80	233	1.61	4.31	10.9%
$-0.75 < AFQT - \overline{AFQT}_{CCG} < 0.75$	2924	1.79	4.84	1534	1.97	5.09	1390	1.59	4.55	23.3%
$0.75 < AFQT - \overline{AFQT}_{CCG}$	441	1.97	4.90	201	1.90	4.67	240	2.03	5.10	-6.6%
Total	23465	1.15	4.05	10640	1.40	4.34	12825	0.94	3.78	49.0%

Table B.7: Changes in the effects of easy to observe characteristics on the probability of moving

VARIABLES	(1)	(2)	(3)	(4)
AFQT			-0.649***	-0.611***
			(0.0921)	(0.0910)
Reference AFQT	-0.333***		0.320*	
	(0.0995)		(0.136)	
Years of Education		-0.217***		-0.0982
		(0.0524)		(0.0553)
Competitive College		0.621**		0.822***
		(0.218)		(0.220)
White		0.504**		0.820***
		(0.192)		(0.197)
Hispanic		1.056***		0.800***
		(0.211)		(0.214)
Work Spell	-0.227***	-0.225***	-0.225***	-0.223***
	(0.00478)	(0.00479)	(0.00479)	(0.00480)
Potential Experience	-0.140***	-0.141***	-0.145***	-0.145***
	(0.00762)	(0.00762)	(0.00764)	(0.00764)
Observations	232,388	232,388	232,388	232,388

Standard errors are in parentheses. Average Partial Effects are from Normal MLE (Probit). Regressions include full set of year indicators and urbanicity.
*** p<0.001, ** p<0.01, * p<0.05

Table B.8: Effects of easy and difficult to observe characteristics on the probability of moving

Panel A: Mean AFQT of reference group summarizing observable characteristics

	Job Separations	Job-to-Job	Job-to-Unemployment	
Variables	Full Sample	Full Sample	Recession	Expansion
AFQT	-0.644***	0.0245	-0.940***	-0.638***
	(0.0925)	(0.0625)	(0.259)	(0.0750)
Reference AFQT	0.424**	0.621***	0.321	-0.303**
	(0.137)	(0.0921)	(0.394)	(0.111)
Potential Experience	-0.147***	-0.0600***	-0.290***	-0.0621***
	(0.00765)	(0.00520)	(0.0231)	(0.00610)
Working Spell	-0.222***	-0.0759***	-0.115***	-0.155***
	(0.00484)	(0.00320)	(0.0230)	(0.00401)
Observations	232388	232388	29557	202831

Panel B: Including observable characteristics

	Job Separations	Job-to-Job	Job-to-Unemployment	
Variables	Full Sample	Full Sample	Recession	Expansion
AFQT	-0.610***	0.0423	-1.100***	-0.596***
	(0.0911)	(0.0617)	(0.255)	(0.0738)
Years of Education	-0.0595	0.0704	0.112	-0.219***
	(0.0562)	(0.0372)	(0.165)	(0.0464)
Competitive College	0.863***	0.359*	1.399*	0.337
	(0.222)	(0.147)	(0.642)	(0.183)
White	0.878***	0.436**	1.066	0.349*
	(0.198)	(0.134)	(0.564)	(0.159)
Hispanic	0.808***	-0.197	3.018***	0.653***
	(0.216)	(0.150)	(0.625)	(0.170)
Potential Experience	-0.169***	-0.0630***	-0.334***	-0.0732***
	(0.00787)	(0.00531)	(0.0243)	(0.00632)
Working Spell	-0.270***	-0.100***	0.181***	-0.189***
	(0.00749)	(0.00495)	(0.0362)	(0.00603)
Observations	232388	232388	29557	202831

Standard errors are in parentheses. Average Partial Effects are from Normal MLE (Probit). Regressions include full set of indicators for year, and urbanicity as well as interactions between each covariate and potential experience and length of work spell. *** $p<0.001$, ** $p<0.01$, * $p<0.05$

Table B.9: Dynamics with regard to working spell duration and experience

Variables	Job Separations Full Sample	Job-to-Job Full Sample	Job-to-Unemployment Recession	 Expansion
AFQT	-4.735***	0.592	-5.970**	-7.124***
	(0.942)	(1.221)	(2.193)	(1.317)
Years of Education (Education)	4.492***	4.200***	2.242	3.058***
	(0.600)	(0.761)	(1.434)	(0.843)
Competitive College (Competitive)	7.157**	0.941	5.180	9.812**
	(2.368)	(3.013)	(5.559)	(3.337)
White	7.630***	8.868***	2.602	4.543
	(2.045)	(2.668)	(4.766)	(2.855)
Hispanic	7.262**	-3.939	13.76**	9.257**
	(2.285)	(3.044)	(5.276)	(3.121)
Potential Experience (Expp)	0.897***	0.924***	1.369	0.964***
	(0.212)	(0.273)	(0.819)	(0.287)
Working Spell (Spell)	-4.110***	-2.646***	-3.426*	-4.494***
	(0.282)	(0.352)	(1.414)	(0.373)
AFQT x Spell	-1.018	-2.034	-20.47	-1.281
	(3.383)	(4.241)	(17.48)	(4.433)
AFQT x Expp	4.212	0.399	10.01	5.124
	(2.362)	(3.085)	(9.618)	(3.097)
Competitive x Spell	23.54**	-0.528	144.0***	33.46**
	(8.601)	(10.76)	(43.63)	(11.34)
Competitive x Expp	-15.78**	7.495	-55.71*	-29.20***
	(5.988)	(7.661)	(24.58)	(8.013)
Education x Spell	18.23***	9.553***	46.07***	19.53***
	(2.118)	(2.637)	(10.84)	(2.807)
Education x Expp	-20.12***	-13.06***	-32.05***	-19.65***
	(1.537)	(1.963)	(6.248)	(2.062)
White x Spell	-4.970	7.684	-91.90*	-10.09
	(7.388)	(9.352)	(41.27)	(9.514)
White x Expp	-5.773	-14.72*	63.39**	-0.106
	(4.963)	(6.477)	(21.22)	(6.518)
Hispanic x Spell	15.15	23.58*	114.7*	-0.247
	(8.522)	(10.92)	(45.50)	(10.88)
Hispanic x Expp	-13.62*	-4.387	-40.39	-8.717
	(5.350)	(7.119)	(23.00)	(6.902)
Observations	232388	232388	29557	202831

Standard errors are in parentheses. Scaled coefficients are from Normal MLE (Probit). Suppressed coefficient estimates include a full set of year indicators, urbanicity, and square and interaction terms for potential experience and working spell duration. * p<0.05, ** p<0.01, *** p<0.001.

Table B.10: Heteroskedastic probit APEs and effects on the conditional variance

Panel A: Heteroskedastic Probit APEs				
	Job Separations	Job-to-Job	Job-to-Unemployment	
Variables	Full Sample	Full Sample	Recession	Expansion
AFQT	-0.635***	0.0414	-1.100***	-0.604***
	(0.0906)	(0.0616)	(0.253)	(0.0733)
Years of Education	-0.0450	0.0750*	0.0690	-0.216***
	(0.0559)	(0.0371)	(0.164)	(0.0462)
Competitive College	0.917***	0.353*	1.465*	0.363*
	(0.221)	(0.147)	(0.637)	(0.182)
White	0.880***	0.434**	0.941	0.365*
	(0.197)	(0.134)	(0.560)	(0.158)
Hispanic	0.853***	-0.203	3.079***	0.684***
	(0.215)	(0.150)	(0.626)	(0.169)
Potential Experience	-0.170***	-0.0651***	-0.347***	-0.0768***
	(0.00801)	(0.00534)	(0.0240)	(0.00650)
Working Spell	-0.168***	-0.0957***	0.282***	-0.150***
	(0.0111)	(0.00554)	(0.0406)	(0.00822)
Panel B: Effects on the conditional variance				
lnsigma2				
Working Spell	1.224***	-0.628**	1.244***	0.819***
	(0.0786)	(0.228)	(0.302)	(0.0936)
Potential Experience	1.769***	0.478*	2.487***	1.890***
	(0.0627)	(0.226)	(0.242)	(0.0752)
Observations	232388	232388	29557	202831

Standard errors are in parentheses. APEs and effects on the conditional variance are from normal heteroskedastic MLE (Probit). Suppressed coefficient estimates include a full set of year indicators and urbanicity.
* p<0.05, ** p<0.01, *** p<0.001.

Table B.11: Heteroskedastic probit scaled coefficients

	Job Separations	Job-to-Job	Job-to-Unemployment	
Variables	Full Sample	Full Sample	Recession	Expansion
AFQT	-4.923**	0.529	-5.104	-7.705***
	(1.541)	(1.284)	(3.387)	(2.235)
Years of Education (Education)	7.545***	4.557***	-3.347	8.294***
	(1.006)	(0.865)	(2.343)	(1.462)
Competitive College (Competitive)	5.571	0.859	-4.319	13.06*
	(3.952)	(3.170)	(8.916)	(5.755)
White	4.800	8.750**	7.201	-1.756
	(3.345)	(2.767)	(7.334)	(4.877)
Hispanic	5.403	-4.512	12.96	3.601
	(3.724)	(3.212)	(8.068)	(5.328)
Potential Experience (Expp)	4.981***	0.548	2.255	6.054***
	(0.534)	(0.330)	(1.753)	(0.773)
Working Spell (Spell)	-6.746***	-2.049**	-12.07***	-8.317***
	(0.963)	(0.652)	(3.352)	(1.212)
AFQT x Spell	-22.68	-3.541	-81.05*	-16.48
	(11.66)	(4.539)	(40.11)	(14.66)
AFQT x Expp	0.618	1.367	9.521	-6.480
	(5.929)	(3.594)	(21.45)	(8.076)
Education x Spell	68.80***	11.56***	142.7***	70.67***
	(7.463)	(3.230)	(26.51)	(9.426)
Education x Expp	-55.53***	-15.33***	-47.07***	-66.83***
	(3.949)	(2.868)	(13.45)	(5.588)
Competitive x Spell	123.0***	-4.049	358.3***	123.8***
	(29.59)	(11.39)	(104.9)	(37.12)
Competitive x Expp	-29.00	10.23	-59.61	-65.11**
	(14.92)	(9.115)	(52.14)	(20.48)
White x Spell	-29.32	7.649	-216.4*	-28.75
	(25.62)	(10.01)	(92.19)	(32.06)
White x Expp	29.40*	-14.16	128.6**	39.35*
	(12.73)	(7.617)	(47.44)	(17.30)
Hispanic x Spell	84.69**	22.54	271.6*	55.19
	(29.79)	(12.10)	(107.7)	(36.76)
Hispanic x Expp	-16.63	-3.713	-40.49	11.58
	(13.53)	(8.439)	(51.88)	(18.16)
Observations	232388	232388	29557	202831

Standard errors are in parentheses. Scaled coefficients are from normal heteroskedastic MLE (Probit). Suppressed coefficient estimates include a full set of year indicators, urbanicity, and square and interaction terms for potential experience and working spell duration.
* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table B.12: APEs from heteroskedastic probit control function

VARIABLES	Separations Full Sample	Job-to-Job Full Sample	Job-to-Unemployment Recessions	 Expansions
AFQT	-0.130	0.0727	-0.162	-0.172*
	(0.091)	(.061)	(.24)	(.068)
Years of Education	0.131*	0.122***	0.384*	0.0200
	(0.057)	(.038)	(.162)	(.048)
Competitive College	0.0159	0.266	0.114	-0.255
	(0.229)	(.137)	(.629)	(.181)
White	0.202	0.388**	-0.137	-0.00288
	(0.206)	(.132)	(.527)	(.161)
Hispanic	-0.160	-0.215	1.093	-0.0538
	(0.207)	(.145)	(.588)	(.167)
Experience	-0.162**	-0.0552	-1.361***	-0.0661
	(0.059)	(.043)	(.225)	(.051)
Working Spell Length	-0.629***	-0.131***	0.151	-0.548***
	(0.018)	(.008)	(.246)	(.015)
Experience Residuals	0.114***	0.0813***	0.241***	0.0149
	(0.059)	(.042)	(.226)	(.05)
Working Spell Residuals	0.871***	0.0711***	0.397***	0.788***
	(0.017)	(.007)	(.219)	(.016)
Observations	232,388	232,388	29,557	202,831

Bootstrapped standard errors in parentheses from 500 repetitions. Suppressed coefficient estimates include urbanicity and a full set of year indicators. Results from heteroskedastic normal MLE control function. *** $p<0.001$, ** $p<0.01$, * $p<0.05$

Table B.13: Control function effects on the conditional variance

VARIABLES	Separations Full Sample	Job-to-Job Full Sample	Job-to-Unemployment Recessions	 Expansions
Working Spell	2.740***	-0.466	1.986	2.467***
	(0.095)	(0.377)	(2.426)	(0.126)
Experience	1.525***	1.998***	2.530**	1.721***
	(0.062)	(0.246)	(0.89)	(0.099)
Spell Residuals	-3.107***	-0.569	-1.825	-2.687***
	(0.098)	(0.52)	(3.108)	(0.121)
Exp. Residuals	-0.704***	-0.940***	0.844	-1.038***
	(0.134)	(0.186)	(0.974)	(0.197)
Observations	232,388	232,388	29,557	202,831

Bootstrap standard errors in parentheses from 500 repetitions. Results from heteroskedastic normal MLE control function.
*** p<0.001, ** p<0.01, * p<0.05.

Table B.14: Scaled coefficients from heteroskedastic probit control function

VARIABLES	Job Separations Full Sample	Job-to-Job Full Sample	Job-to-Unemployment Recessions	Job-to-Unemployment Expansions
AFQT	-3.95	2.015	-9.171	-6.685*
	(2.303)	(2.157)	(7.092)	(3.302)
Years of Education (Education)	9.969***	5.494***	2.061	7.844***
	(1.108)	(1.322)	(3.04)	(1.604)
Competitive College (Competitive)	-19.23***	-4.931	-20.61	-12.52
	(5.484)	(5.196)	(17.694)	(7.716)
White	3.737	4.770	1.277	-0.516
	(5.291)	(5.51)	(15.722)	(7.42)
Hispanic	-11.03	-12.19*	-5.962	-7.791
	(5.761)	(5.672)	(16.808)	(8.277)
Experience (Exp)	2.038*	2.437**	-5.471*	1.596
	(0.817)	(0.879)	(2.443)	(1.251)
Working Spell (Spell)	-0.971	-3.694***	6.538	-0.991
	(1.219)	(0.879)	(5.158)	(1.893)
Experience Residuals (Exp_R)	-5.975***	-4.478***	1.665	-1.988
	(1.055)	(1.059)	(2.389)	(1.260)
Work Spell Residuals (Spell_R)	6.853***	1.007	-0.0566	6.670***
	(1.283)	(0.978)	(0.0424)	(1.534)
AFQT x Spell	18.35	-9.439	95.38	31.04
	(18.374)	(11.542)	(99.753)	(26.449)
AFQT x Exp	2.666	2.644	-40.95	-0.458
	(9.096)	(9.076)	(54.993)	(12.691)
AFQT x Spell_R	-22.27	0.371	-54.78	-18.86
	(15.543)	(10.018)	(68.606)	(22.342)
AFQT x Exp_R	20.04	4.252	-23.82	17.30
	(12.187)	(10.658)	(38.821)	(18.316)
Education x Spell	51.66***	13.25*	136.5***	77.48***
	(9.409)	(6.757)	(39.819)	(14.437)
Education x Exp	-55.57***	-19.68**	-91.28***	-69.86***
	(6.645)	(6.128)	(20.108)	(10.244)
Education x Spell_R	19.41***	0.754	20.24	18.49***
	(3.679)	(3.837)	(19.295)	(5.552)
Education x Exp_R	7.007	16.72**	1.005	0.899
	(6.645)	(6.128)	(20.108)	(10.244)
Competitive x Spell	245.7***	36.59	254.3	220.3***
	(42.404)	(30.864)	(244.782)	(58.798)
Competitive x Exp	-38.64	6.609	-62.52	-63.39
	(22.983)	(22.268)	(116.892)	(34.428)
Competitive x Spell_R	-283.7***	-58.58*	-140.6	-227.7***
	(34.161)	(26.935)	(181.335)	(45.493)
Competitive x Exp_R	50.71	-32.99	33.03	88.46*
	(26.961)	(24.693)	(77.702)	(39.179)
White x Spell	-50.54	13.63	-659.7**	-85.52
	(39.859)	(25.061)	(260.562)	(53.549)
White x Exp	27.57	-4.672	492.2**	49.87
	(19.298)	(19.672)	(136.897)	(27.121)
White x Spell_R	68.28*	-7.429	-25.74	86.50
	(33.516)	(22.274)	(138.333)	(46.274)
White x Exp_R	-17.13	-5.039	191.4*	-5.884
	(28.204)	(29.762)	(89.196)	(42.548)
Hispanic x Spell	103.1*	56.75	86.96	85.99
	(44.173)	(30.232)	(200.865)	(61.664)
Hispanic x Exp	-22.10	-4.676	75.22	-13.73
	(20.745)	(23.182)	(119.415)	(28.386)
Hispanic x Spell_R	-39.85	-44.02	-111.0	-25.68
	(37.954)	(26.084)	(156.889)	(53.164)
Hispanic x Exp_R	-27.63	-15.86	169.8	-22.07
	(29.033)	(30.752)	(96.193)	(45.675)
Observations	232,388	232,388	29,557	202,831

Bootstrapped standard errors are in parentheses. Scaled coefficients are from normal heteroskedastic MLE (Probit). Suppressed coefficient estimates include a full set of year indicators, urbanicity, and square and interaction terms for potential experience, working spell duration, and control function residuals of each. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Appendix C

Figures for Chapter 3

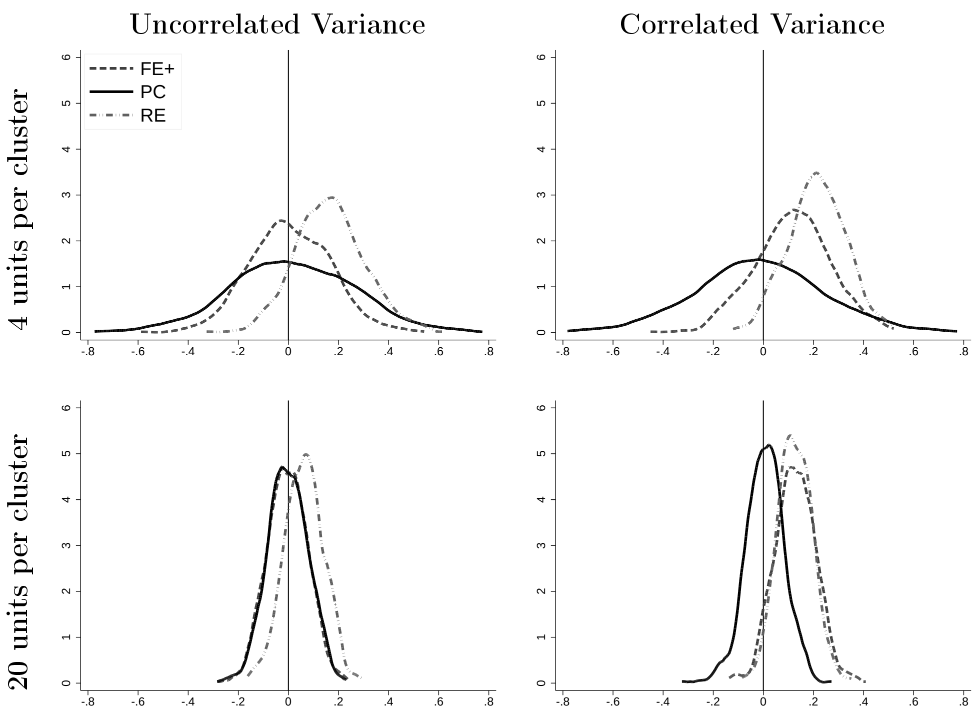

Figure C.1: Kernel density plots of estimation errors, $\widehat{\beta}_1 - \beta_1$, for coefficient of x_{ij} across replications for all methods when the uncorrelated variance assumption holds (left panels) and when it is violated (right panels). *Note.* FE+ = Augmented Fixed-Effects; PC = Per-Cluster; RE = Random-Effects.

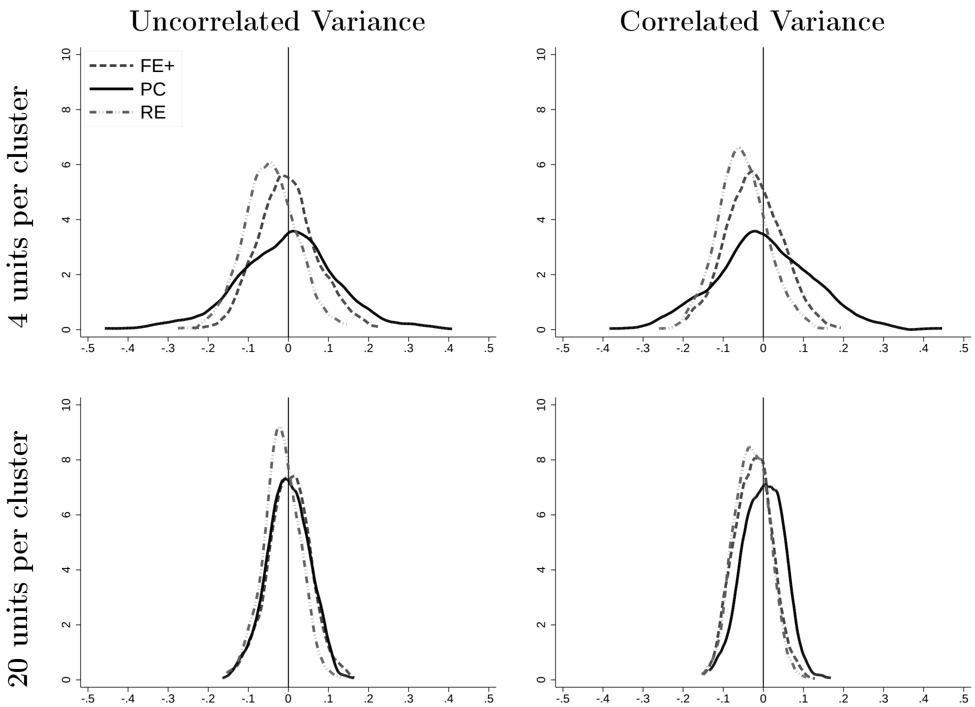

Figure C.2: Kernel density plots of estimation errors, $\widehat{\gamma}_1 - \gamma_1$, for coefficient of w_j across replications for all methods when the uncorrelated variance assumption holds (left panels) and when it is violated (right panels). *Note.* FE+ = Augmented Fixed-Effects; PC = Per-Cluster; RE = Random-Effects.

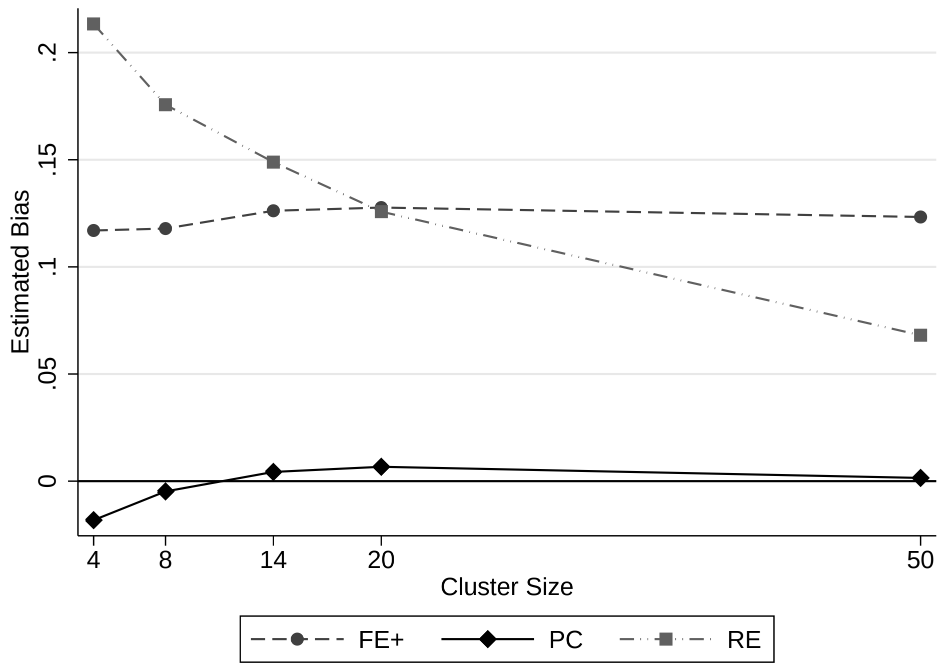

Figure C.3: Estimated bias for coefficient β_1 of x_{ij} versus cluster size. *Note.* FE+ = Augmented Fixed-Effects; PC = Per-Cluster; RE = Random-Effects.

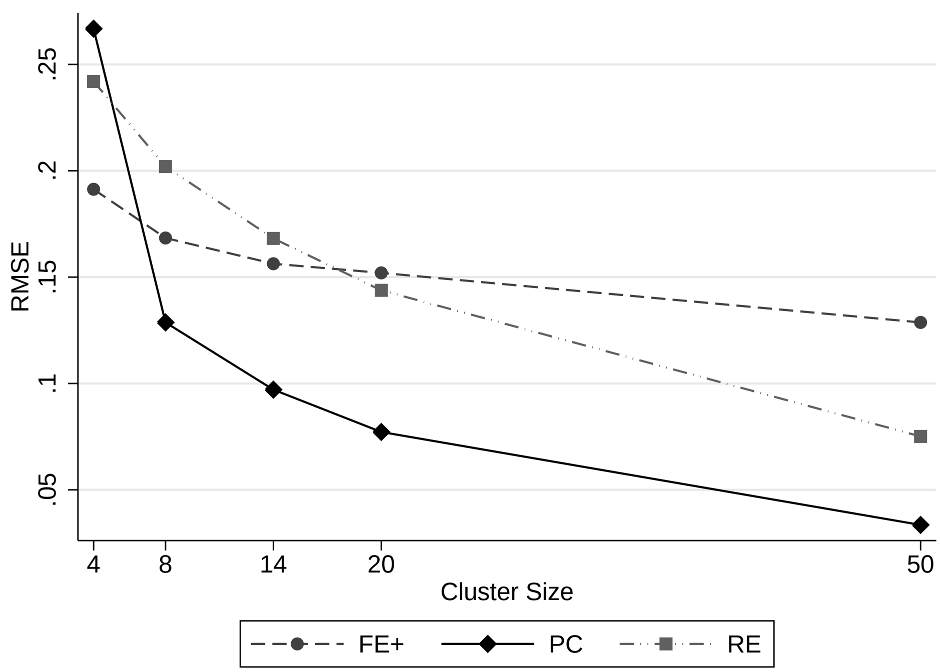

Figure C.4: Estimated root mean square error (RMSE) for coefficient β_1 of x_{ij} versus cluster size. *Note.* FE+ = Augmented Fixed-Effects; PC = Per-Cluster; RE = Random-Effects.

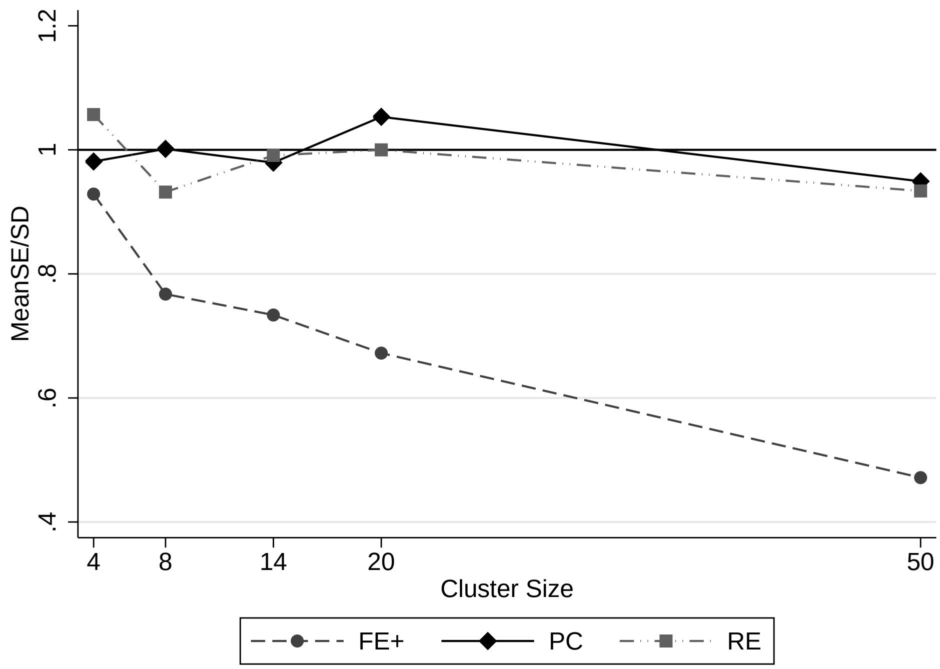

Figure C.5: Ratio of mean standard errors (Mean SE) divided by standard deviations (SD) of estimates versus cluster size. *Note.* FE+ = Augmented Fixed-Effects; PC = Per-Cluster; RE = Random-Effects.

Appendix D

Tables for Chapter 3

Table D.1: Estimates from different methods using High School and Beyond Data.

	Random Effects (RE)		Augmented Fixed Effects (FE+)		Per-Cluster Regression (PC)	
	Est	(SE)	Est	(SE)	Est	(SE)
γ_0 [Constant]	11.752	(0.232)	11.769	(0.205)	11.615	(0.271)
γ_1 [Catholic]	2.130	(0.346)	2.186	(0.337)	2.253	(0.406)
β_1 [SES]	2.958	(0.143)	2.782	(0.145)	2.772	(0.169)
β_2 [SES $\times$ Catholic]	-1.313	(0.216)	-1.349	(0.218)	-1.303	(0.234)

Note. All estimates are significantly different from 0 at the 0.05 level.

Table D.2: Comparing methods for estimating the coefficients.

		β_1 $[x_{ij}]$		β_2 $[x_{ij} \times w_j]$		γ_1 $[w_j]$	
Simulation Condition	Method	100× Bias	100× RMSE	100× Bias	100× RMSE	100× Bias	100× RMSE
Small Clusters	RE	16.6*	21.6	0.8	7.0	-4.5*	8.0
& Uncorrelated	FE+	0.6	16.2	0.0	8.2	-0.3	7.4
Variance	PC	1.9	25.5	-0.6	13.3	-0.5	12.7
Small Clusters	RE	21.3*	24.2	1.2	6.1	-5.3*	8.2
& Correlated	FE+	11.7*	19.1	0.2	7.9	-2.4*	8.2
Variance	PC	-1.8	26.7	0.7	13.4	0.2	12.2
Large Clusters	RE	6.2*	10.0	0.2	3.9	-1.7*	4.9
& Uncorrelated	FE+	-0.3	8.0	0.2	4.0	0.2	5.3
Variance	PC	-0.2	8.0	0.1	4.1	0.0	5.2
Large Clusters	RE	12.6*	14.4	-0.1	3.5	-2.9*	5.2
& Correlated	FE+	12.8*	15.2	-0.3	4.2	-2.5*	5.2
Variance	PC	0.7	7.7	-0.3	4.0	0.0	5.0

Note. Small Clusters: $n_j = n = 4$, Large Clusters: $n_j = n = 20$;
Uncorrelated Variance: $\sigma_j^2 = 1$, Correlated Variance: $\sigma_j = \exp(u_{1j})$.

RMSE=root-mean-square error;

Mean SE=mean of the standard error estimates over the replications;

SD=standard deviation of the coefficient estimates over the replications;

*Estimated bias differs significantly from 0 at the 0.05 level.

Appendix E

Additions for Chapter 1

Comparative Statics

The probability of transferring schools if given by the following equation (equation 6 in text):

$$P(M) = P[b^{h*} - b^{r*} > 0]$$

Base Probability of Moving

For simplicity, these first derivations adopt the notation of bidding in the absence of VAMs. Substituting the hiring and retaining principals bids provides the following:

$$P(M) =$$
$$P[\frac{\sigma_\tau(0)\sigma_\xi(x)}{Z^h_{NV}}m + \frac{\sigma_\tau(0)\sigma_\epsilon}{Z^h_{NV}}R_x + \frac{\sigma_\epsilon\sigma_\xi(x)}{Z^h_{NV}}P^h_0 - (\frac{\sigma_\tau(t)\sigma_\xi(x)}{Z^r_{NV}}m + \frac{\sigma_\tau(t)\sigma_\epsilon}{Z^r_{NV}}R_x + \frac{\sigma_\epsilon\sigma_\xi(x)}{Z^r_{NV}}P^r_t) > 0], \tag{E.1}$$

where $Z^h_{NV} = \sigma_\tau(0)\sigma_\xi(x) + \sigma_\tau(0)\sigma_\epsilon + \sigma_\epsilon\sigma_\xi(x)$ and $Z^r_{NV} = \sigma_\tau(t)\sigma_\xi(x) + \sigma_\tau(t)\sigma_\epsilon + \sigma_\epsilon\sigma_\xi(x)$

After some algebra, equation E.1 becomes the following:

$$=P\{\frac{\sigma_\xi(x)}{Z^h_{NV}Z^r_{NV}}[(m-\mu)\sigma_\xi(x)(\sigma_\tau(0) - \sigma_\tau(t)) + (\sigma_\epsilon\sigma_\tau(t) + \sigma_\xi(x)\sigma_\tau(t) + \sigma_\epsilon\sigma_\xi(x))\tau^h_0$$
$$- (\sigma_\epsilon\sigma_\tau(0) + \sigma_\xi(x)\sigma_\tau(0) + \sigma_\epsilon\sigma_\xi(x))\tau^r_t + \sigma_\epsilon(\sigma_\tau(0) - \sigma_\tau(t))\xi] > 0\}$$

Letting $\psi \equiv (\sigma_\epsilon\sigma_\tau(t) + \sigma_\xi(x)\sigma_\tau(t) + \sigma_\epsilon\sigma_\xi(x))\tau^h_0 - (\sigma_\epsilon\sigma_\tau(0) + \sigma_\xi(x)\sigma_\tau(0) + \sigma_\epsilon\sigma_\xi(x))\tau^r_t + \sigma_\epsilon(\sigma_v - \sigma_\tau(t))\xi$, be the composite error term, simplifies the above, to equation 1.7 from within text, presented below:

$$P(M) = P\left\{\psi > \sigma_\xi(x)[\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t)](\mu - m)\right\}.$$

Under the assumptions that τ^r, τ^h and ξ are each orthogonal to one another,

$$\begin{aligned}
\sigma_\psi \equiv var(\psi) =& var[(\sigma_\epsilon\sigma_\mathcal{T}(t) + \sigma_\xi(x)\sigma_\mathcal{T}(t) + \sigma_\epsilon\sigma_\xi(x))\tau_0^h \\
&- (\sigma_\epsilon\sigma_\mathcal{T}(0) + \sigma_\xi(x)\sigma_v + \sigma_\epsilon\sigma_\xi(x))\tau_t^r + \sigma_\epsilon(\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t))\xi] \\
&= \sigma_\mathcal{T}(t)(\sigma_\epsilon\sigma_\mathcal{T}(0) + \sigma_\xi(x)\sigma_\mathcal{T}(0) + \sigma_\epsilon\sigma_\xi(x))^2 \\
&+ \sigma_\mathcal{T}(0)(\sigma_\epsilon\sigma_\mathcal{T}(t) + \sigma_\xi(x)\sigma_\mathcal{T}(t) + \sigma_\epsilon\sigma_\xi(x))^2 + \sigma_\xi(x)\sigma_\epsilon^2(\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t))^2
\end{aligned} \tag{E.2}$$

Assuming normality of the error terms, the probability of a school-to-school transition may be written as:

$$\begin{aligned}
P(M) &= \Phi\left\{\frac{-1}{\sqrt{\sigma_\psi}}\left[\sigma_\xi(x)[\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t)](\mu - m)\right]\right\} \\
&= \Phi\left\{-\beta_{xt}(\mu - m)\right\}.
\end{aligned} \tag{E.3}$$

Comparative statics for within-district moves with respect to teacher effectiveness (μ)

Assuming the probability of moving schools is monotonically increasing in the difference between b^{h*} and b^{r*}, the sign of $\frac{\partial P[b_{HV}^{h*} - b_{HV}^{r*} > 0 | m\ \mu] - P[b_{NV}^{h*} - b_{NV}^{r*} > 0 | m\ \mu]}{\partial\mu}$ is implied by the sign of $\frac{\partial E[b_{HV}^{h*}-b_{HV}^{r*}-(b_{NV}^{h*}-b_{NV}^{r*})|m\,\mu]}{\partial \mu}$. Here, the subscript HV denotes that hiring principals may access a teacher's VAM, while the subscript NV denotes that there are no VAMs informing the bidding. The difference between hiring and retaining principals' bids without the presence of VAMs is given by equation E.1 and is given by equation E.4 when both principals may access the VAMs.

$$\begin{aligned} b_{HV}^{h*} - b_{HV}^{r*} =& \frac{\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x\,V)}{Z_{HV}^{r}} m + \frac{\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}}{Z_{HV}^{r}} R_{x\nu} + \frac{\sigma_{\epsilon}\sigma_{\xi}(x\,V)}{Z_{HV}^{r}} P_0^h \\ &- \left(\frac{\sigma_{\mathcal{T}}(t)\sigma_{\xi}(x\,V)}{Z_{HV}^{r}} m + \frac{\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}}{Z_{HV}^{r}} R_{x\nu} + \frac{\sigma_{\epsilon}\sigma_{\xi}(x\,V)}{Z_{HV}^{r}} P_t^r \right). \end{aligned} \tag{E.4}$$

The expectation of that difference given prior beliefs and the underlying ability in the presence of VAMs is given by equation E.5:

$$E[b_{HV}^{h*} - b_{HV}^{r*}|m\,\mu] = \frac{1}{Z_{HV}^{h} Z_{HV}^{r}} (m-\mu)\sigma_{\xi}(x\,V)^2 \sigma_{\epsilon}(\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)). \tag{E.5}$$

The expectation of difference between bids given prior beliefs and the underlying ability without VAMs is given by equation E.6:

$$E[b_{NV}^{h*} - b_{NV}^{r*}|m\,\mu] = \frac{1}{Z_{HV}^{h} Z_{HV}^{r}} (m-\mu)\sigma_{\xi}(x)^2 \sigma_{\epsilon}(\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)). \tag{E.6}$$

Let $A_1 = (m-\mu)\sigma_{\xi}(x\,V)^2\sigma_{\epsilon}(\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t))$

Let $A_0 = (m-\mu)\sigma_{\xi}(x)^2\sigma_{\epsilon}(\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t))$

$$E[b_{HV}^{h*} - b_{HV}^{r*} - (b_{NV}^{h*} - b_{NV}^{r*})|m\mu] = \frac{A_1}{Z_{HV}^{h} Z_{HV}^{r}} - \frac{A_0}{Z_{NV}^{h} Z_{NV}^{r}} = \frac{A_1 Z_{NV}^{h} Z_{NV}^{r} - A_0 Z_{HV}^{h} Z_{HV}^{r}}{Z_{HV}^{h} Z_{HV}^{r} Z_{NV}^{h} Z_{NV}^{r}} \tag{E.7}$$

Examining the numerator:

$$
\begin{aligned}
A_1 Z^h_{NV} Z^r_{NV} - A_0 Z^h_{HV} Z^r_{HV} =& (m-\mu)\sigma_\epsilon(\sigma_\tau(0)-\sigma_\tau(t))(\sigma_\xi(x\,V)^2 \\
& (\sigma_\tau(t)\sigma_\xi(x)^2\sigma_\tau(0)+\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x)+\sigma_\epsilon\sigma_\xi(x)^2\sigma_\tau(0) \\
& +\sigma_\tau(t)\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon+\sigma_\tau(t)\sigma_\epsilon^2\sigma_\tau(0)+\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon^2 \\
& +\sigma_\tau(t)\sigma_\xi(x)^2\sigma_\epsilon+\sigma_\tau(t)\sigma_\epsilon^2\sigma_\xi(x)+\sigma_\epsilon^2\sigma_\xi(x)^2 \\
& -\sigma_\xi(x)^2(\sigma_\tau(t)\sigma_\xi(x\,V)\sigma_\tau(0)\sigma_\xi(x\,V)+\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x\,V) \\
& +\sigma_\epsilon\sigma_\xi(x\,V)\sigma_\tau(0)\sigma_\xi(x\,V)+\sigma_\tau(t)\sigma_\xi(x\,V)\sigma_\tau(0)\sigma_\epsilon \\
& +\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\epsilon+\sigma_\epsilon\sigma_\xi(x\,V)\sigma_\tau(0)\sigma_\epsilon+\sigma_\epsilon\sigma_\xi(x\,V)\sigma_\epsilon\sigma_\xi(x\,V) \\
& +\sigma_\tau(t)\sigma_\xi(x\,V)\sigma_\epsilon\sigma_\xi(x\,V)+\sigma_\tau(t)\sigma_\epsilon\sigma_\epsilon\sigma_\xi(x\,V))
\end{aligned} \tag{E.8}
$$

$$
\begin{aligned}
=& (m-\mu)\sigma_\epsilon(\sigma_\tau(0)-\sigma_\tau(t))(\sigma_\xi(x\,V)^2(\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x) \\
& +\sigma_\tau(t)\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon+\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\epsilon+\sigma_\epsilon\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon+\sigma_\tau(t)\sigma_\epsilon\sigma_\epsilon\sigma_\xi(x)) \\
& -\sigma_\xi(x)^2(\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x\,V)+\sigma_\tau(t)\sigma_\xi(x\,V)\sigma_\tau(0)\sigma_\epsilon+\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\epsilon \\
& +\sigma_\epsilon\sigma_\xi(x\,V)\sigma_\tau(0)\sigma_\epsilon+\sigma_\tau(t)\sigma_\epsilon\sigma_\epsilon\sigma_\xi(x\,V))
\end{aligned}
$$

$$
\begin{aligned}
=& (m-\mu)\sigma_\epsilon(\sigma_\tau(0)-\sigma_\tau(t))((\sigma_\xi(x\,V)-\sigma_\xi(x))(\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x)\sigma_\xi(x\,V) \\
& +\sigma_\xi(x\,V)\sigma_\epsilon^2\sigma_\xi(x)\sigma_\tau(0)+\sigma_\xi(x\,V)\sigma_\tau(t)\sigma_\epsilon^2\sigma_\xi(x)+(\sigma_\xi(x\,V)+\sigma_\xi(x))\sigma_\tau(t)\sigma_\epsilon^2\sigma_\tau(0)).
\end{aligned}
$$

$$\frac{\partial A_1 Z^h_{NV} Z^r_{NV} - A_0 Z^h_{HV} Z^r_{HV}}{\partial \mu} = -\sigma_\epsilon(\sigma_\tau(0) - \sigma_\tau(t))$$

$$\begin{aligned}&((\sigma_\xi(x\,V) - \sigma_\xi(x))(\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x)\sigma_\xi(x\,V)\\ &+ \sigma_\xi(x\,V)\sigma_\epsilon^2\sigma_\xi(x)\sigma_\tau(0) + \sigma_\xi(x\,V)\sigma_\tau(t)\sigma_\epsilon^2\sigma_\xi(x)\\ &+ (\sigma_\xi(x\,V) + \sigma_\xi(x))\sigma_\tau(t)\sigma_\epsilon^2\sigma_\tau(0)).\end{aligned} \tag{E.9}$$

$\frac{\partial E[b^{h*}_{HV} - b^{r*}_{HV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,\mu]}{\partial \mu}$ is

$\frac{1}{Z^h_{HV} Z^r_{HV} Z^h_{NV} Z^r_{NV}} \frac{\partial A_1 Z^h_{NV} Z^r_{NV} - A_0 Z^h_{HV} Z^r_{HV}}{\partial \mu}$.

$\frac{1}{Z^h_{HV} Z^r_{HV} Z^h_{NV} Z^r_{NV}}$ is positive, as it is purely a function of variances. As a fundamental component of asymmetric employer learning, it is assumed that $\sigma_\tau(0) - \sigma_\tau(t) > 0$. Under lemma 2, $\sigma_\xi(x\,V) - \sigma_\xi(x) < 0$. All other terms are positive variances, which implies that $\frac{\partial E[b^{h*}_{HV} - b^{r*}_{HV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,\mu]}{\partial \mu} > 0$, which in turn implies that the probability of moving increases with increases in μ.

Comparative statics for within-district moves with respect to VAMs (V)

In determining the comparative statics with regard to the VAM signal, I seek to sign $\frac{\partial E[b^{h*}_{HV} - b^{r*}_{HV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,V]}{\partial V}$. From equation E.4:

$$\begin{aligned}b^{h*}_{HV} - b^{r*}_{HV} =& \frac{\sigma_\tau(0)\sigma_\xi(x\,V)}{Z^r_{HV}} m + \frac{\sigma_\tau(0)\sigma_\epsilon}{Z^r_{HV}} R_{x\nu} + \frac{\sigma_\epsilon\sigma_\xi(x\,V)}{Z^r_{HV}} P^h_0\\ &- \left(\frac{\sigma_\tau(t)\sigma_\xi(x\,V)}{Z^r_{HV}} m + \frac{\sigma_\tau(t)\sigma_\epsilon}{Z^r_{HV}} R_{x\nu} + \frac{\sigma_\epsilon\sigma_\xi(x\,V)}{Z^r_{HV}} P^r_t\right)\\ =& \frac{1}{Z^h_{HV} Z^r_{HV}}[\sigma_\xi(x\,V)\sigma_\epsilon(\sigma_\tau(0) - \sigma_\tau(t))(\sigma_\xi(x\,V)(m-\mu) + \sigma_\epsilon \frac{\sigma_\nu\xi + \sigma_\xi(x)\nu}{\sigma_\nu + \sigma_\xi(x)})\\ &+ \tau^h Z^r_{HV}\sigma_\xi(x\,V)\sigma_\epsilon - \tau^r_t Z^h_{HV}\sigma_\xi(x\,V)\sigma_\epsilon]\end{aligned}$$

Substituting in the VAM (V) and prior public signal (R_x) separately provides equation E.10

$$
\begin{aligned}
=&\frac{1}{Z^h_{HV}Z^r_{HV}}[\sigma_\xi(x\,V)\sigma_\epsilon(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t))(\sigma_\xi(x\,V)(m-(1+\sigma_\epsilon)\mu)+\sigma_\epsilon\frac{\sigma_\nu R_x+\sigma_\xi(x)V}{\sigma_\nu+\sigma_\xi(x)})\\
&+\tau^h Z^r_{HV}\sigma_\xi(x\,V)\sigma_\epsilon-\tau^r_t Z^h_{HV}\sigma_\xi(x\,V)\sigma_\epsilon]
\end{aligned}
\tag{E.10}
$$

Turning back to the probability of moving in absence of VAMs,

$$
\begin{aligned}
b^{h*}_{NV}-b^{r*}_{NV}=&\frac{\sigma_\mathcal{T}(0)\sigma_\xi(x)}{Z^h_{NV}}m+\frac{\sigma_\mathcal{T}(0)\sigma_\epsilon}{Z^h_{NV}}R_x+\frac{\sigma_\epsilon\sigma_\xi(x)}{Z^h_{NV}}P^H_0\\
&-\left(\frac{\sigma_\mathcal{T}(t)\sigma_\xi(x)}{Z^r_{NV}}m+\frac{\sigma_\mathcal{T}(t)\sigma_\epsilon}{Z^r_{NV}}R_x+\frac{\sigma_\epsilon\sigma_\xi(x)}{Z^r_{NV}}P^R_t\right)\\
&=\frac{1}{Z^h_{NV}Z^r_{NV}}[\sigma_\xi(x)\sigma_\epsilon(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t))(\sigma_\xi(x)(m-\mu)+\sigma_\epsilon\xi)\\
&+\tau^h Z^r_{NV}\sigma_\xi(x)\sigma_\epsilon-\tau^r_t Z^h_{NV}\sigma_\xi(x)\sigma_\epsilon]
\end{aligned}
\tag{E.11}
$$

Combining equation E.10 with equation E.11 and taking the expectation conditional on prior beliefs and VAMs provides equation E.12:

$$
\begin{aligned}
E[b^{h*}_{HV}-b^{r*}_{HV}-(b^{h*}_{NV}-b^{r*}_{NV})|m\,V]=&\frac{1}{Z^h_{HV}Z^r_{HV}}[\sigma_\xi(x\,V)\sigma_\epsilon(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t))(\sigma_\xi(x\,V)\\
&(m-(1+\sigma_\epsilon)\mu)+\sigma_\epsilon\frac{\sigma_\nu\mu+\sigma_\xi(x)V}{\sigma_\nu+\sigma_\xi(x)})]\\
&-\frac{1}{Z^h_{NV}Z^r_{NV}}\sigma_\xi(x)\sigma_\epsilon(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t))(\sigma_\xi(x)(m-\mu)
\end{aligned}
\tag{E.12}
$$

Taking the derivative with respect to VAMs (V) provides equation 1.9 from the text.

$$
\frac{\partial E\left[b^{h*}_{HV}-b^{r*}_{HV}-(b^{h*}_{NV}-b^{r*}_{NV})|m\,V\right]}{\partial V}=\frac{1}{Z^h_{HV}Z^r_{HV}}\frac{\sigma_\xi(x)}{\sigma_\nu+\sigma_\xi(x)}>0
$$

As $\frac{1}{Z^h_{HV} Z^r_{HV}} \frac{\sigma_\xi(x)}{\sigma_\nu + \sigma_\xi(x)}$ is function of variances, it must be positive. Meaning that releasing VAMs raises the probability that high-VAM teachers move schools.

Comparative statics for out-of-district moves with respect to teacher effectiveness (μ)

Assuming the probability of moving schools is monotonically increasing in the difference between b^{h*} and b^{r*}, the sign of $\frac{\partial P[b^{h*}_{RV} - b^{r*}_{RV} > 0 | m\ \mu] - P[b^{h*}_{NV} - b^{r*}_{NV} > 0 | m\ \mu]}{\partial \mu}$ is implied by the sign of $\frac{\partial E[b^{h*}_{RV} - b^{r*}_{RV} - (b^{h*}_{NV} - b^{r*}_{NV}) | m\ \mu]}{\partial \mu}$. Here, the subscript RV denotes that only retaining principals may access a teacher's VAM, while the subscript NV denotes that there are no VAMs informing the bidding. The difference between hiring and retaining principals' bids without the presence of VAMs is given by equation E.1, and is given by equation E.13 when both principals may access the VAMs.

$$
\begin{aligned}
b^{h*}_{RV} - b^{r*}_{RV} =& \frac{\sigma_\tau(0)\sigma_\xi(x)}{Z^r_{RV}} m + \frac{\sigma_\tau(0)\sigma_\epsilon}{Z^r_{RV}} R_x + \frac{\sigma_\epsilon \sigma_\xi(x)}{Z^r_{RV}} P^h_0 \\
& - \left(\frac{\sigma_\tau(t\ V)\sigma_\xi(x)}{Z^r_{RV}} m + \frac{\sigma_\tau(t\ V)\sigma_\epsilon}{Z^r_{RV}} R_x + \frac{\sigma_\epsilon \sigma_\xi(x\ V)}{Z^r_{RV}} P^r_{t\nu} \right).
\end{aligned} \tag{E.13}
$$

The expectation of that difference given prior beliefs and the underlying ability in the presence of VAMs is given by equation E.14:

$$
E[b^{h*}_{RV} - b^{r*}_{RV} | m\ \mu] = \frac{1}{Z^h_{RV} Z^r_{RV}} (m - \mu) \sigma_\xi(x)^2 \sigma_\epsilon (\sigma_\tau(0) - \sigma_\tau(t\ V)). \tag{E.14}
$$

The expectation of difference between bids given prior beliefs and the underlying ability without VAMs is again given by equation E.6:

$$E[b^{h*}_{NV} - b^{r*}_{NV}|m\,\mu] = \frac{1}{Z^h_{HV} Z^r_{HV}}(m-\mu)\sigma_\xi(x)^2\sigma_\epsilon(\sigma_\tau(0)-\sigma_\tau(t)).$$

Combining equation E.14 with equation E.6 gives the following:

$$\begin{aligned}
E[b^{h*}_{RV} - b^{r*}_{RV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,\mu] = {} & \frac{(m-\mu)\sigma_\xi(x)^2\sigma_\epsilon}{Z^h_{RV} Z^r_{RV} Z^h_{NV} Z^r_{NV}}[(\sigma_\tau(0)-\sigma_\tau(t\,V)) \\
& (\sigma_\tau(t)\sigma_\xi(x)^2\sigma_\tau(0) + \sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x) + \sigma_\epsilon\sigma_\xi(x)^2\sigma_\tau(0) \\
& + \sigma_\tau(t)\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon + \sigma_\tau(t)\sigma_\epsilon^2\sigma_\tau(0) + \sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon^2 \\
& + \sigma_\tau(t)\sigma_\xi(x)^2\sigma_\epsilon + \sigma_\tau(t)\sigma_\epsilon^2\sigma_\xi(x) + \sigma_\epsilon^2\sigma_\xi(x)^2) \\
& - (\sigma_\tau(0)-\sigma_\tau(t))(\sigma_\tau(t\,V)\sigma_\xi(x)\sigma_\tau(0)\sigma_\xi(x) \\
& + \sigma_\epsilon\sigma_\xi(x)\sigma_\tau(0)\sigma_\xi(x) + \sigma_\tau(t\,V)\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon \\
& + \sigma_\tau(t\,V)\sigma_\epsilon\sigma_\tau(0)\sigma_\epsilon + \sigma_\epsilon\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon \\
& + \sigma_\tau(t\,V)\sigma_\xi(x)\sigma_\epsilon\sigma_\xi(x) + \sigma_\tau(t\,V)\sigma_\epsilon\sigma_\epsilon\sigma_\xi(x) \\
& + \sigma_\tau(t\,V)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x) + \sigma_\epsilon\sigma_\xi(x)\sigma_\epsilon\sigma_\xi(x))] \\
& = \frac{(m-\mu)\sigma_\xi(x)^2\sigma_\epsilon}{Z^h_{RV} Z^r_{RV} Z^h_{NV} Z^r_{NV}}(\sigma_\tau(t)-\sigma_\tau(t\,V)) \\
& (\sigma_\tau(0)^2\sigma_\epsilon^2 + \sigma_\tau(0)^2\sigma_\xi(x)^2 + \sigma_\tau(0)^2\sigma_\epsilon\sigma_\xi(x) + \sigma_\xi(x)^2\sigma_\epsilon^2)
\end{aligned} \tag{E.15}$$

Taking the derivative of equation E.15 with respect to true effectiveness (μ), gives what is referred to in text as equation E.

$$\frac{\partial E\left[b^{h*}_{RV} - b^{r*}_{RV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,\mu\right]}{\partial \mu} = \frac{-\sigma_{\xi}(x)^2\sigma_{\epsilon}}{Z^h_{RV}Z^r_{RV}Z^h_{NV}Z^r_{NV}}(\sigma_{\mathcal{T}}(t) - \sigma_{\mathcal{T}}(t\,V))(\sigma_{\mathcal{T}}(0)^2\sigma^2_{\epsilon} + \sigma_{\mathcal{T}}(0)^2\sigma_{\xi}(x)^2 + \sigma_{\mathcal{T}}(0)^2\sigma_{\epsilon}\sigma_{\xi}(x) + \sigma_{\xi}(x)^2\sigma^2_{\epsilon}).$$

Lemma 1 demonstrates that $\sigma_{\mathcal{T}}(t) - \sigma_{\mathcal{T}}(t\,V) > 0$. All other terms are positive variances, implying that $\frac{\partial E[b^{h*}_{RV} - b^{r*}_{RV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,\mu]}{\partial \mu} < 0$, which in turn implies that the probability of out-of-district transitions increases with declines in teacher effectiveness (μ).

Comparative statics for out-of-district moves with respect to VAMs (V)

In determining the comparative statics with regard to the VAM signal, I seek to sign $\frac{\partial E[b^{h*}_{RV} - b^{r*}_{RV} - (b^{h*}_{NV} - b^{r*}_{NV})|m\,V]}{\partial V}$. Turning back to the probability of moving in absence of VAMs, equation E.11 provides:

$$b^{h*}_{NV} - b^{r*}_{NV} = \frac{1}{Z^h_{NV}Z^r_{NV}}[\sigma_{\xi}(x)\sigma_{\epsilon}(\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t))(\sigma_{\xi}(x)(m-\mu) + \sigma_{\epsilon}\xi) + \tau^h Z^r_{NV}\sigma_{\xi}(x)\sigma_{\epsilon} - \tau^r_t Z^h_{NV}\sigma_{\xi}(x)\sigma_{\epsilon}]$$

In the case where only retaining principals may access a teacher's VAM, as is plausible for out-of-district moves, the difference between hiring and retaining principals bids is given by equation E.16:

$$
\begin{aligned}
b_{RV}^{h*} - b_{RV}^{r*} =& \frac{\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)}{Z_{RV}^{r}} m + \frac{\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}}{Z_{RV}^{r}} R_x + \frac{\sigma_{\epsilon}\sigma_{\xi}(x)}{Z_{RV}^{r}} P_0^h \\
& - \left(\frac{\sigma_{\mathcal{T}}(t\,V)\sigma_{\xi}(x)}{Z_{RV}^{r}} m + \frac{\sigma_{\mathcal{T}}(t\,V)\sigma_{\epsilon}}{Z_{RV}^{r}} R_x + \frac{\sigma_{\epsilon}\sigma_{\xi}(x\,V)}{Z_{RV}^{r}} P_{t\nu}^{r} \right) \\
& = \frac{1}{Z_{RV}^{h} Z_{RV}^{r}} [\sigma_{\xi}(x)\sigma_{\epsilon}(\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}\,\nu}(t\,V))(\sigma_{\xi}(x)(m-\mu) + \sigma_{\epsilon}\xi) \\
& + \tau^h Z_{HV}^{r} \sigma_{\xi}(x)\sigma_{\epsilon} - \sigma_{\xi}(x)\sigma_{\epsilon} Z_{RV}^{h} \frac{\sigma_{\nu}\tau_t^r + \sigma_{\mathcal{T}}(t)\nu}{\sigma_{\nu} + \sigma_{\mathcal{T}}(t)}] \\
& = \frac{1}{Z_{RV}^{h} Z_{RV}^{r}} [\sigma_{\xi}(x)\sigma_{\epsilon}(\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}\,\nu}(t\,V))(\sigma_{\xi}(x)(m-\mu) + \sigma_{\epsilon}\xi) \\
& + \tau^h Z_{RV}^{r} \sigma_{\xi}(x)\sigma_{\epsilon} - \sigma_{\xi}(x)\sigma_{\epsilon} Z_{RV}^{h} \frac{\sigma_{\nu}\tau_t^r + \sigma_{\mathcal{T}}(t)(V-\mu)}{\sigma_{\nu} + \sigma_{\mathcal{T}}(t)}]
\end{aligned}
\tag{E.16}
$$

The derivative of equation E.16 with respect to the VAM signal (V) is referred to in text as equation 1.11, and is presented below:

$$
\frac{\partial E\left[b_{HV}^{h*} - b_{HV}^{r*} - (b_{NV}^{h*} - b_{NV}^{r*}) | m\,V \right]}{\partial V} = \frac{-\sigma_{\xi}(x)\sigma_{\epsilon}\sigma_{\mathcal{T}}(t)}{Z_{RV}^{r}(\sigma_{\nu} + \sigma_{\mathcal{T}}(t))} < 0
$$

As equation 1.11 is the negative of a function of variances, it is less than zero. Thus after VAMs are released, as a teacher's VAM decreases, the probability of moving out of district increases.

Robustness: Year interactions with VAM

The primary threat to validity for difference-in-difference analysis is differential trends. The tables below provide year interactions with the VAM within both treatment districts as well as the rest of the state. While the estimates are too noisy to say anything conclusive, the pre-policy trends do not seem diverge in a way that would bias up my results. It is also noteworthy that is both districts there is a spike in the correlation of VAM with the probability of moving within-district soon after the policy takes effect.

Robustness: Mobility based on ABC Growth Policies

In the 1996/1997 school year the state of North Carolina began rewarding teachers who worked in schools in which the students made substantial growth. The state awarded bonuses of either $750 or $1,500 based on whether the school achieved growth in student test scores beyond predetermined tiered thresholds. These bonuses were given to all teachers in qualifying schools. For additional detail about the policy please see Vigdor et al. (2008) and Ahn and Vigdor (2012).

As a result, teaching in high growth schools may be additionally attractive to teachers since the bonuses depended upon school performance. Table F.4 is comparable to Table A.3 except that the dependent variable here is whether the teacher moves to higher (lower) growth school as opposed to a higher (lower) performing school within and out of district. The total within and out-of districts mobility estimates in columns 1 and 4 of Table A.3 are unaffected, and so they are omitted.

When examining this alternate school attribute on which teachers may sort, the primary findings remain intact. The within district mobility is driven by moves to more favorable schools for both districts. Though the results are attenuated here as a teacher with a full standard deviation higher VAM is 0.3 percentage point more likely to move within district to a higher ABC growth school for teachers whose VAMs are released, the estimates remain statistically significantly positive for both districts. Though these estimates are not statistically different from the estimated effect on the probability of moving to higher performing schools, perhaps they suggest that school performance may be a stronger motivator for teacher mobility than student growth.

The estimated effects for moves outside the district are remarkably close between Table A.3 and Table F.4. The adverse selection of movers out of Guilford County Schools holds for moves to both better and worse schools, while moves from Winston-Salem to better schools remain unrelated to teachers' VAMs after the policy takes effect.

Normal Maximum Likelihood Estimation

The results in Table A.3 are from a linear probability model, which are more straight forward both computationally and in interpretation. Taking the normality and orthogonality assumptions from Section 1.4 seriously would suggest normal Maximum Likelihood Estimation (probit estimation). As noted in Ai and Norton (2003), the functional form of probit estimation incorporates an interaction term, even when one is not specifically modeled. As a result, if the researcher is interested in estimating the average partial effect (APE) of an interaction additionally programming is necessary. Table F.5 in Appendix **??** provides the APEs in accordance with Ai and Norton (2003). Comparison between Table A.3 and Table F.5 provides very similar results.

Competing Risks Analysis

By performing separate regressions for each type of school transfer, the above analysis treats each type of move as independent of the others. However, it is possible that the propensity of a teacher to move within-district to a higher-performing school is related to the propensity of moving to a higher-performing school in another district. The same could be said with any combination of outcomes. To test the sensitivity of my earlier results to these possibilities, I adopt a competing risks approach, as proposed by Fine and Gray (1999).

Competing risks survival analysis models the subdistribution hazard ($\lambda_E(t)$) of a particular type of event, such as a move within a school district ($E = WD$), as a function of an unspecified baseline hazard ($\lambda_{E0}(t)$), as well as a vector of time-varying covariates ($\mathbf{Z}(t)$). [1]

$$\lambda_{WD}(t|\mathbf{Z}) = \lambda_{WD0}(t)exp\{\mathbf{Z}(t)\boldsymbol{\beta}_0\}, \tag{E.17}$$

In the context of this study, time at risk (t) is defined as the difference between the current year and the year at which the teacher first appears matched with the current school.[2] $\mathbf{Z}(t)$ is a vector including all covariates used in Table A.3, with the exception of tenure, which is perfectly correlated with t. I additionally include district averages of all within-district-varying covariates to control for unobserved, district-wide effects, as in Mundlak (1978)[3].

Table F.6 reports the coefficient estimates for each type of transfer between schools.

[1] Gray (1988) defines the subdistribution hazard as, $\lambda_{WD}(t) = lim_{\Delta t \to 0} \frac{P(t<T\leq t+\Delta t, E=WD | t\leq T \bigcup t<T, E\neq WD)}{\Delta t}$, where T is the timing of the event occurrence of which there are different types.

[2] I use teacher to school matches as the basis of this survival analysis. Though this forces me to assume independence of matches, it allows me to retain the original sample making it easier to compare the results.

[3] Unreported regression results show little difference depending on whether or not district averages are included

Accordingly, $\beta \times 100$ may be interpreted as the percent change in the marginal probability of a particular type of mobility due to a one unit change in the covariate. Columns 1 and 4, examine transfers within and out of the district respectively, with the other broad type of transfer serving as a competing risk. Columns 2, 3, 5, and 6, examine transfers to higher and lower-performing schools, within and out of the district, with the other types of transfers serving as competing risks.

In this framework, results remain remarkably consistent. From columns 1 and 2, the probability of moving within-district for a teacher with a one standard deviation higher VAM score increases by 9% with the release of teacher VAMs, and for moves within-district to better school, the probability increases by 13%. Both effects are significantly different from zero and are within a percentage point estimates shown in Table A.3. From columns 4 and 5, a teacher with a one standard deviation lower VAM becomes 33.6% (29.5%) more likely to move out of Guilford (to a higher-performing school) after the policy takes effect. In Winston-Salem, the results from Table A.3 are muted for total within-district mobility. Column 1 shows a smaller point estimate than appears in Table A.3, and the the estimate loses statistical significance. The impact of the policy in Winston-Salem on moves to higher performing schools within district are more stable. The introduction of VAMs raises the probability that a teacher with a one standard deviation higher VAM moves to a higher-performing school by about 11%, though the significance level drops with this specification. For out-of-district moves to higher-performing schools, the point estimate corresponds with a 15% increase in the probability a high-VAM teacher moves out of Winston-Salem to a higher-performing school, though this estimate is very noisy and should be interpreted accordingly. In general, while the public and private learning results are further verified in Guilford with this competing risks analysis, the same cannot be said for Winston-Salem.

Appendix F

Supplemental Tables for Chapter 1

Table F.1: The effects of VAM on the probability of moving schools within-district by year.

	Total			To a more proficient school		
VARIABLES	Rest of NC	Guilford	Winston-Salem	Rest of NC	Guilford	Winston-Salem
year 1998 x VAM	0.0009	0.0012	0.0043	0.0021***	0.0006	-0.0003
	[0.00077]	[0.00269]	[0.00513]	[0.00061]	[0.00236]	[0.00267]
year 1999 x VAM	0.0022**	0.0023	-0.0001	0.0044***	0.0048**	0.0041
	[0.00083]	[0.00316]	[0.00587]	[0.00059]	[0.00242]	[0.00393]
year 2000 x VAM	0.0035***	0.0205***	-0.0007	0.0023***	0.0155***	-0.0042*
	[0.00079]	[0.00252]	[0.00311]	[0.00065]	[0.00156]	[0.00253]
year 2001 x VAM	0.0019**	0.0048	-0.0020	0.0035***	0.0030	0.0012
	[0.00079]	[0.00332]	[0.00298]	[0.00058]	[0.00262]	[0.00211]
year 2002 x VAM	0.0035**	-0.0044	0.0024	0.0055***	-0.0011	0.0107***
	[0.00096]	[0.00268]	[0.00535]	[0.00073]	[0.00205]	[0.00378]
year 2003 x VAM	0.0004	-0.0054	0.0041	0.0027***	-0.0013	0.0042
	[0.00089]	[0.00467]	[0.00486]	[0.00073]	[0.00329]	[0.00445]
year 2004 x VAM	0.0010	0.0020	-0.0088**	0.0016***	-0.0073**	-0.0043
	[0.00106]	[0.00446]	[0.00403]	[0.0008]	[0.00296]	[0.00358]
year 2005 x VAM	0.0015	0.0128***	-0.0160***	0.0040***	0.0190***	-0.0080**
	[0.00099]	[0.00300]	[0.00423]	[0.00075]	[0.00273]	[0.00297]
year 2006 x VAM	0.0047***	0.0169***	0.0100***	0.0055***	0.0158***	0.0037*
	[0.00087]	[0.00563]	[0.00308]	[0.00061]	[0.00521]	[0.00193]
year 2007 x VAM	0.0027***	0.0189***	-0.0133***	0.0039***	0.0147***	-0.0078**
	[0.00081]	[0.00355]	[0.00478]	[0.00056]	[0.00282]	[0.00366]
year 2008 x VAM	0.0029***	0.0057*	0.0005	0.0032***	0.0114***	0.0019
	[0.00092]	[0.00342]	[0.00469]	[0.00069]	[0.00247]	[0.00370]
year 2009 x VAM	0.0034***	0.0036	0.0110*	0.0032***	0.0046**	0.0173***
	[0.00118]	[0.00325]	[0.00579]	[0.00091]	[0.00233]	[0.00473]
year 2010 x VAM	-0.0001	0.0123***	0.0002	0.0009	0.0121***	0.0004
	[0.00095]	[0.00326]	[0.00489]	[0.00073]	[0.00274]	[0.00431]
Observations	216,484	11,239	8,295	216,484	11,239	8,295

Standard errors are bootstrapped at the student-year level and appear in brackets.
All regressions include teacher level covariates and interactions with year indicators.
*** p<0.01, ** p<0.05, * p<0.1

Table F.2: The effect of VAM on the probability of moving schools out-of-district by year.

	Total			To a more proficient school		
VARIABLES	Rest of NC	Guilford	Winston-Salem	Rest of NC	Guilford	Winston-Salem
year 1998 x VAM	0.0017***	0.0098***	-0.0079**	0.0023***	0.0076***	-0.0059***
	[0.0005]	[0.00212]	[0.0032]	[0.00039]	[0.00178]	[0.00187]
year 1999 x VAM	-0.0004	0.0065**	-0.0026*	0.0011**	0.0064***	-0.0033***
	[0.00057]	[0.00267]	[0.00136]	[0.00049]	[0.00243]	[0.00096]
year 2000 x VAM	0.0006	0.0013	0.0063***	0.0015***	0.0033***	0.0033*
	[0.00057]	[0.00157]	[0.00215]	[0.00045]	[0.00126]	[0.00195]
year 2001 x VAM	-0.0022***	0.0025	-0.0069***	-0.0005	0.0063***	-0.0070***
	[0.00057]	[0.00152]	[0.00202]	[0.00044]	[0.00112]	[0.00163]
year 2002 x VAM	-0.0033***	-0.0025	0.0106***	0.0000	0.0015	0.0146***
	[0.00063]	[0.00261]	[0.00203]	[0.00042]	[0.00167]	[0.00187]
year 2003 x VAM	-0.0011	-0.0016	-0.0141***	0.0017***	-0.0004	-0.0091***
	[0.00071]	[0.00282]	[0.00367]	[0.00052]	[0.0028]	[0.00346]
year 2004 x VAM	-0.0037***	0.0099***	0.0054	-0.0005	0.0080***	0.0092***
	[0.00073]	[0.00206]	[0.0034]	[0.00056]	[0.00172]	[0.00281]
year 2005 x VAM	-0.0001	-0.0038*	-0.0024	0.0011**	0.0033**	-0.0005
	[0.00064]	[0.00197]	[0.00212]	[0.00047]	[0.00164]	[0.00176]
year 2006 x VAM	-0.0011	-0.0095***	-0.0001	0.0017***	-0.0018	-0.0013
	[0.00071]	[0.00372]	[0.003]	[0.00048]	[0.00262]	[0.00276]
year 2007 x VAM	-0.0016**	-0.0223***	0.0011	0.0003	-0.0040***	0.0063*
	[0.00081]	[0.00367]	[0.00358]	[0.00061]	[0.00114]	[0.00352]
year 2008 x VAM	-0.0017**	-0.0079***	-0.0054	0.0006	0.0001	-0.0000
	[0.00064]	[0.00185]	[0.00414]	[0.00047]	[0.00099]	[0.0035]
year 2009 x VAM	0.0006	-0.0023	0.0047***	-0.0004	0.0000	0.0047***
	[0.00051]	[0.00089]	[0.00149]	[0.00035]	[0.00012]	[0.00148]
year 2010 x VAM	-0.0021***	-0.0058***	-0.0011	-0.0006	-0.0054***	-0.0011
	[0.00058]	[0.00156]	[0.00113]	[0.00051]	[0.00103]	[0.00112]
Observations	216,484	11,239	8,295	216,484	11,239	8,295

Standard errors are bootstrapped at the student-year level and appear in brackets.
All regressions include teacher level covariates and interactions with year indicators.
*** p<0.01, ** p<0.05, * p<0.1

Table F.3: The effect of VAM on teacher sorting within-district by year.

VARIABLES	Rest of NC	Guilford	Winston-Salem
year 1998 x VAM	0.0025***	0.0045**	-0.0014
	[0.00021]	[0.00071]	[0.00146]
year 1999 x VAM	0.0026***	0.0013	0.0021
	[0.00021]	[0.00109]	[0.00156]
year 2000 x VAM	0.0019***	0.0041***	0.0007
	[0.0002]	[0.00069]	[0.00084]
year 2001 x VAM	0.0051***	0.0038***	0.0077***
	[0.00026]	[0.00097]	[0.00146]
year 2002 x VAM	0.0046***	0.0031***	0.0072***
	[0.0002]	[0.00072]	[0.00164]
year 2003 x VAM	0.0031***	0.0043***	0.0052***
	[0.00019]	[0.00099]	[0.001]
year 2004 x VAM	0.0023***	-0.0006	0.0005
	[0.00021]	[0.00109]	[0.00212]
year 2005 x VAM	0.0102***	0.0109***	0.0096***
	[0.00032]	[0.00097]	[0.00126]
year 2006 x VAM	0.0047***	0.0009	-0.0014
	[0.00027]	[0.00161]	[0.00089]
year 2007 x VAM	0.0046***	0.0049***	0.0031**
	[0.00026]	[0.00105]	[0.00133]
year 2008 x VAM	0.0016***	0.0031***	0.0005
	[0.00025]	[0.00112]	[0.00127]
year 2009 x VAM	-0.0003	0.0055***	0.0053***
	[0.00042]	[0.00097]	[0.00146]
year 2010 x VAM	0.0033***	0.0050***	0.0045***
	[0.00027]	[0.00104]	[0.00145]
Observations	185,977	9,616	7,35

Standard errors are bootstrapped at the student-year level and appear in brackets. All regressions include teacher level covariates and interactions with treatment indicators.
*** p<0.01, ** p<0.05, * p<0.1

Table F.4: Probability of moving to higher or lower growth schools

	Panal A: Within-District Moves		Panal B: Out-Of-District Moves	
VARIABLES	**To a higher ABC growth school**	**To a lower ABC growth school**	**To a higher ABC growth school**	**To a lower ABC growth school**
VAM	0.0024***	-0.0006	0.0008	-0.0005
	[0.00073]	[0.00077]	[0.00056]	[0.0006]
VAM x Treatment GCS	0.0031**	0.0013	-0.0048***	-0.0052***
	[0.00152]	[0.00153]	[0.00139]	[0.002]
VAM x Treatment WSF	0.003**	0.0017	0	0.0014
	[0.0015]	[0.00155]	[0.00131]	[0.001]
Treatment GCS	0.0074*	-0.0023	0.0057***	-0.0129***
	[0.00385]	[0.00612]	[0.00187]	[0.00219]
Treatment WSF	0.0156***	0.0074**	-0.001	-0.0093***
	[0.00206]	[0.00297]	[0.00126]	[0.00209]
Observations	236,018	236,018	236,018	236,018

CSB standard errors from 500 repetitions appear in brackets.
All regressions include teacher level covariates and interactions with treatment indicators.
*** p<0.01, ** p<0.05, * p<0.1

Table F.5: Probability of moving schools using normal maximum likelihood estimation.

	Panal A: Within-District Moves			Panal B: Out-of-District Moves		
VARIABLES	**Total**	**To a higher performing school**	**To a lower performing school**	**Total**	**To a higher performing school**	**To a lower performing school**
VAM	0.0022**	0.0030***	-0.0011	-0.0011	0.0005	-0.0018***
	[0.00114]	[0.00079]	[0.00068]	[0.00083]	[0.0006]	[0.0005]
VAM x Treatment GCS	0.0046*	0.0040**	0.0021	-0.0117***	-0.0065***	-0.0053***
	[0.0025]	[0.00172]	[0.00185]	[0.00274]	[0.00203]	[0.0017]
VAM x Treatment WSF	0.0029	0.0038*	-0.0010	0.0002	0.0026	-0.0020
	[0.00268]	[0.00193]	[0.00221]	[0.00313]	[0.00238]	[0.00324]
Treatment GCS	0.0110***	0.0112***	0.0001	-0.0009	-0.0036**	0.0027***
	[0.00268]	[0.0019]	[0.00177]	[0.0019]	[0.00161]	[0.00101]
Treatment WSF	-0.0149***	-0.0103***	-0.0080***	0.0022	-0.0011	-0.0226***
	[0.00441]	[0.00369]	[0.0031]	[0.00493]	[0.00342]	[0.00679]
Observations	236,018	236,018	236,018	236,018	236,018	236,018

CSB standard errors from 500 repetitions appear in brackets.
All regressions include teacher level covariates and interactions with treatment indicators.
*** p<0.01, ** p<0.05, * p<0.1

Table F.6: Changes in the marginal probability of each type of transfer between schools

	Panal A: Within-District Moves			Panal B: Out-Of-District Moves		
VARIABLES	Total	To a higher performing school	To a lower performing school	Total	To a higher performing school	To a lower performing school
VAM	**0.03**	**0.09*** **	**-0.07** **	**0.01**	**0.08** **	**-0.10** **
	[0.021]	[0.024]	[0.030]	[0.028]	[0.035]	[0.042]
VAM x Treatment GCS	**0.09** **	**0.13** **	**0.10**	**-0.41*** **	**-0.35*** **	**-0.40** **
	[0.045]	[0.051]	[0.076]	[0.104]	[0.111]	[0.164]
VAM x Treatment WSF	**0.04**	**0.11* **	**-0.08**	**0.02**	**0.15**	**-0.21**
	[0.050]	[0.068]	[0.095]	[0.116]	[0.141]	[0.238]
Treatment GCS	**0.01**	**0.22** **	**-0.23** **	**0.24** **	**-0.12**	**0.49*** **
	[0.116]	[0.107]	[0.113]	[0.122]	[0.130]	[0.160]
Treatment WSF	**0.56*** **	**0.27* **	**0.87*** **	**-0.87*** **	**0.18**	**-7.22*** **
	[0.118]	[0.145]	[0.144]	[0.167]	[0.219]	[0.587]
Observations	236,018	236,018	236,018	236,018	236,018	236,018

CSB standard errors from 500 repetitions appear in brackets.

All regressions include teacher level covariates and interactions with treatment indicators.

*** p<0.01, ** p<0.05, * p<0.1

Table F.7: Probability of moving schools using alternate standard errors

	Within-District Moves			Out-of-District Moves		
	Total	**To higher performing schools**	**To lower performing schools**	**Total**	**To higher performing schools**	**To lower performing schools**
VAM	**0.0016**	**0.0032**	**-0.0016**	**0.0002**	**0.0014**	**-0.0012**
	(0.00139)	(0.00091)	(0.00083)	(0.00084)	(0.00057)	(0.00050)
	{0.00056}	{0.0004}	{0.00036}	{0.00039}	{0.00031}	{0.00022}
	[0.00129]	[0.00091]	[0.00074]	[0.00096]	[0.00072]	[0.00058]
VAM x Treatment GCS	**0.0058**	**0.0051**	**0.0007**	**-0.0103**	**-0.0054**	**-0.0049**
	(0.00168)	(0.00115)	(0.00091)	(0.00090)	(0.00061)	(0.00057)
	{0.00262}	{0.00204}	{0.00153}	{0.00192}	{0.00164}	{0.00106}
	[0.00265]	[0.00199]	[0.00151]	[0.00261]	[0.00195]	[0.00156]
VAM x Treatment WSF	**0.0052**	**0.006**	**-0.0008**	**0.0009**	**0.0023**	**-0.0014**
	(0.00147)	(0.00094)	(0.00125)	(0.00084)	(0.00068)	(0.00051)
	{0.00323}	{0.00255}	{0.00204}	{0.00186}	{0.00167}	{0.00096}
	[0.00286]	[0.00229]	[0.00194]	[0.00241]	[0.00208]	[0.00129]
Treatment GCS	**-0.004**	**-0.005**	**0.001**	**-0.0162**	**-0.0232**	**0.007**
	(0.00829)	(0.00608)	(0.00537)	(0.00402)	(0.00319)	(0.00214)
	{0.00583}	{0.00436}	{0.00444}	{0.00261}	{0.00114}	{0.0024}
	[0.00851]	[0.00571]	[0.00679]	[0.00374]	[0.00233]	[0.00268]
Treatment WSF	**0.0555**	**0.0475**	**0.008**	**-0.002**	**0.0147**	**-0.0167**
	(0.00579)	(0.00417)	(0.00311)	(0.00258)	(0.00199)	(0.00184)
	{0.00314}	{0.00253}	{0.00215}	{0.0029}	{0.0022}	{0.00171}
	[0.00499]	[0.00372]	[0.00299]	[0.00274]	[0.00224]	[0.00178]
Observations	236,018	236,018	236,018	236,018	236,018	236,018

Clustered standard errors in parentheses. Bootstrapped standard errors in braces. District-cluster-bootstrapped-teacher-stratified standard errors in brackets.

Table F.8: Probability of moving including alternate index of teacher quality

	Within-District Moves			Out-of-District Moves		
Variables	Total	To a higher performing schools	To a lower performing schools	Total	To a higher performing schools	To a lower performing schools
VAM	0.0017	0.0036***	-0.0020*	-0.0003	0.0012	-0.0015**
	[0.00172]	[0.00116]	[0.00102]	[0.00110]	[0.00076]	[0.00062]
Teacher Quality Index (TQ Index)	-0.0375**	-0.0917***	0.0542***	-0.0319***	-0.0395***	0.0076**
	[0.01836]	[0.01406]	[0.00718]	[0.00657]	[0.00622]	[0.00299]
VAM x Treatment GCS	0.0086***	0.0061***	0.0025**	-0.0113***	-0.0059***	-0.0054***
	[0.00205]	[0.00138]	[0.00113]	[0.00114]	[0.00080]	[0.00069]
VAM x Treatment WSF	0.0051***	0.0046***	0.0005	-0.0004	0.0008	-0.0012*
	[0.00175]	[0.00120]	[0.00155]	[0.00100]	[0.00078]	[0.00063]
TQ Index x Treatment GCS	-0.0103	-0.0102	-0.0001	0.0181***	0.0148***	0.0033
	[0.01934]	[0.01522]	[0.00762]	[0.00558]	[0.00477]	[0.00329]
TQ Index x Treatment WSF	-0.0680***	-0.0381***	-0.0300***	-0.0208***	-0.0269***	0.0061**
	[0.00943]	[0.00735]	[0.00466]	[0.00501]	[0.00402]	[0.00261]
Treatment GCS	0.0178***	0.0114***	0.0064***	-0.0029**	-0.0031***	0.0002
	[0.00513]	[0.00416]	[0.00161]	[0.00135]	[0.00115]	[0.00073]
Treatment WSF	-0.0096***	-0.0042*	-0.0054***	0.0065***	0.0075***	-0.0010*
	[0.00358]	[0.00226]	[0.00174]	[0.00124]	[0.00114]	[0.00061]
Observations	236,018	236,018	236,018	236,018	236,018	236,018

Bootstrapped standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table F.9: Probability of moving schools using Empirical Bayes VAM

	Panal A: Within-District Moves			**Panal B: Out-Of-District Moves**		
VARIABLES	**Total**	**To a higher performing school**	**To a lower performing school**	**Total**	**To a higher performing school**	**To a lower performing school**
VAM	**0.0006**	**0.0028*** **	**-0.0022*** **	**-0.0006** **	**0.0014*** **	**-0.0020*** **
	[0.00042]	[0.00032]	[0.00027]	[0.0003]	[0.00023]	[0.00019]
VAM x Treatment GCS	**0.0048*** **	**0.0059*** **	**-0.0011**	**-0.0130*** **	**-0.0078*** **	**-0.0051*** **
	[0.00206]	[0.00162]	[0.00121]	[0.00148]	[0.00111]	[0.00097]
VAM x Treatment WSF	**0.0066*** **	**0.0085*** **	**-0.0020**	**0.0009**	**0.0023**	**-0.0013**
	[0.00276]	[0.00216]	[0.00166]	[0.00166]	[0.00143]	[0.00084]
Treatment GCS	**-0.0048**	**-0.0055*** **	**0.0007**	**-0.0174*** **	**-0.0245*** **	**0.0072*** **
	[0.00408]	[0.00109]	[0.00409]	[0.00121]	[0.00098]	[0.00064]
Treatment WSF	**0.0553*** **	**0.0471*** **	**0.0082*** **	**-0.0022**	**0.0144*** **	**-0.0167*** **
	[0.00232]	[0.00173]	[0.00162]	[0.00194]	[0.00193]	[0.00028]
Observations	236,018	236,018	236,018	236,018	236,018	236,018

Standard errors are bootstrapped at the student-year level and appear in brackets.
All regressions include teacher level covariates and interactions with treatment indicators.
*** p<0.01, ** p<0.05, * p<0.1

Table F.10: Probability of moving schools using restricted-data, Empirical Bayes VAM

	Panal A: Within-District Moves			Panal B: Out-Of-District Moves		
VARIABLES	**Total**	**To a higher performing school**	**To a lower performing school**	**Total**	**To a higher performing school**	**To a lower performing school**
VAM	**0.0015****	**0.0000**	**-0.0015*****	**-0.0021*****	**-0.0011*****	**-0.0010*****
	[0.00063]	[0.00053]	[0.0004]	[0.00046]	[0.00038]	[0.0003]
VAM x Treatment GCS	**0.0035**	**0.0037**	**-0.0001**	**-0.0063*****	**-0.0041****	**-0.0023***
	[0.00327]	[0.00245]	[0.00206]	[0.00244]	[0.00202]	[0.00123]
VAM x Treatment WSF	**0.0090*****	**0.0129*****	**-0.0039****	**0.0020**	**0.0019**	**0.0001**
	[0.00282]	[0.00219]	[0.00179]	[0.00193]	[0.00171]	[0.00086]
Treatment GCS	**-0.0032**	**-0.0040**	**0.0008**	**-0.0162**	**-0.0239*****	**0.0077***
	[0.00902]	[0.00698]	[0.00719]	[0.00451]	[0.00168]	[0.00438]
Treatment WSF	**0.0555*****	**0.0476*****	**0.0078*****	**-0.0021**	**0.0147*****	**-0.0167*****
	[0.00265]	[0.00195]	[0.00181]	[0.00204]	[0.00201]	[0.00031]
Observations	236,018	236,018	236,018	236,018	236,018	236,018

Standard errors are bootstrapped at the student-year level and appear in brackets.
All regressions include teacher level covariates and interactions with treatment indicators.
*** p<0.01, ** p<0.05, * p<0.1

Appendix G

Supplemental Figures for Chapter 1

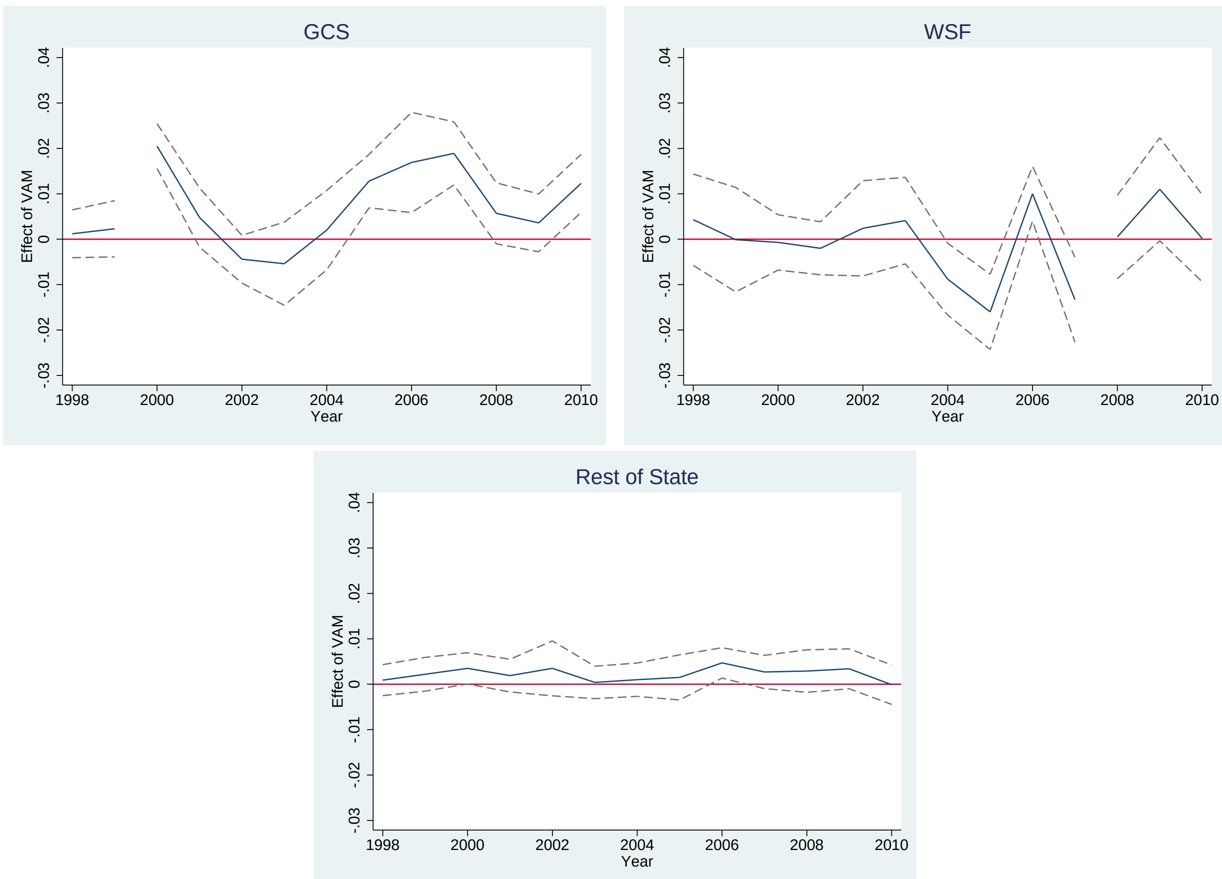

Figure G.1: The effects of VAM on the probability of moving schools within-district by year.

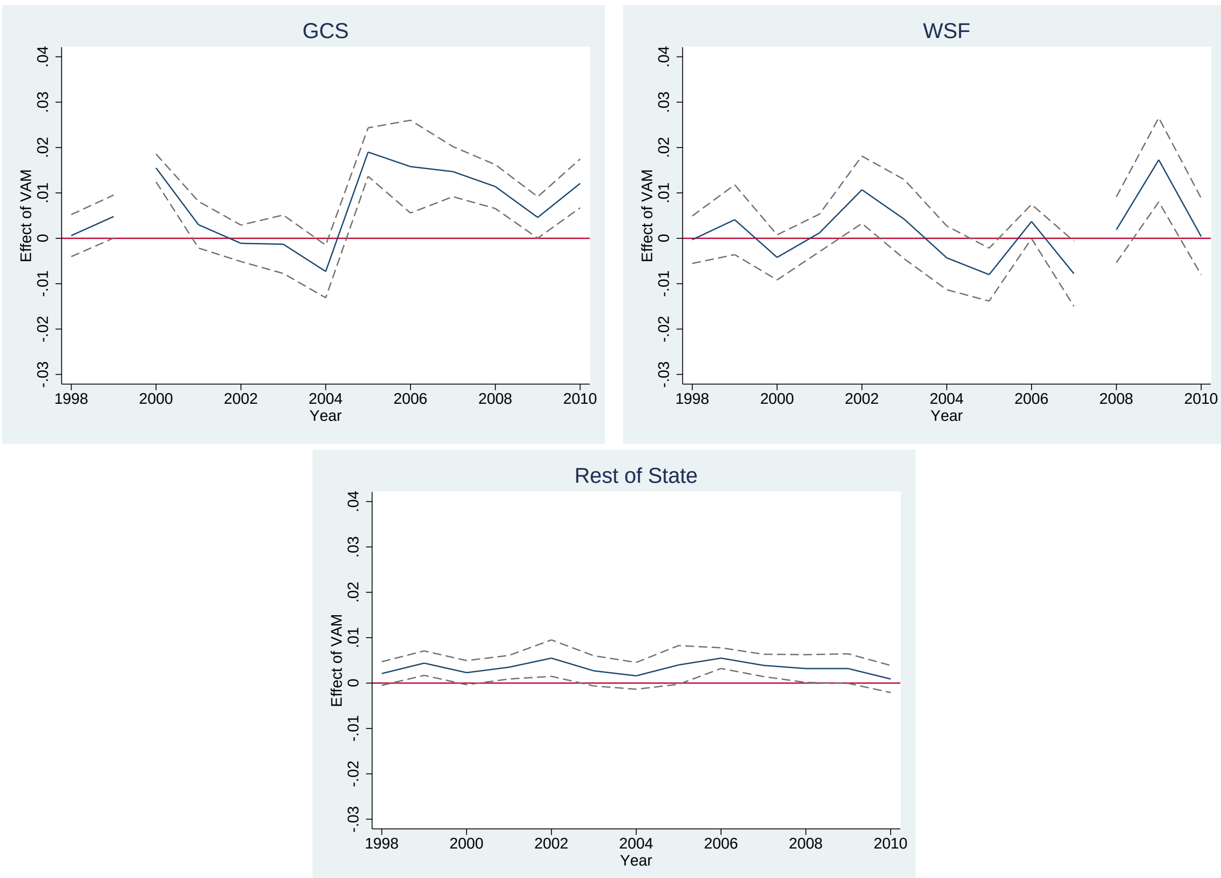

Figure G.2: The effects of VAM on the probability of moving to a a better school within-district by year.

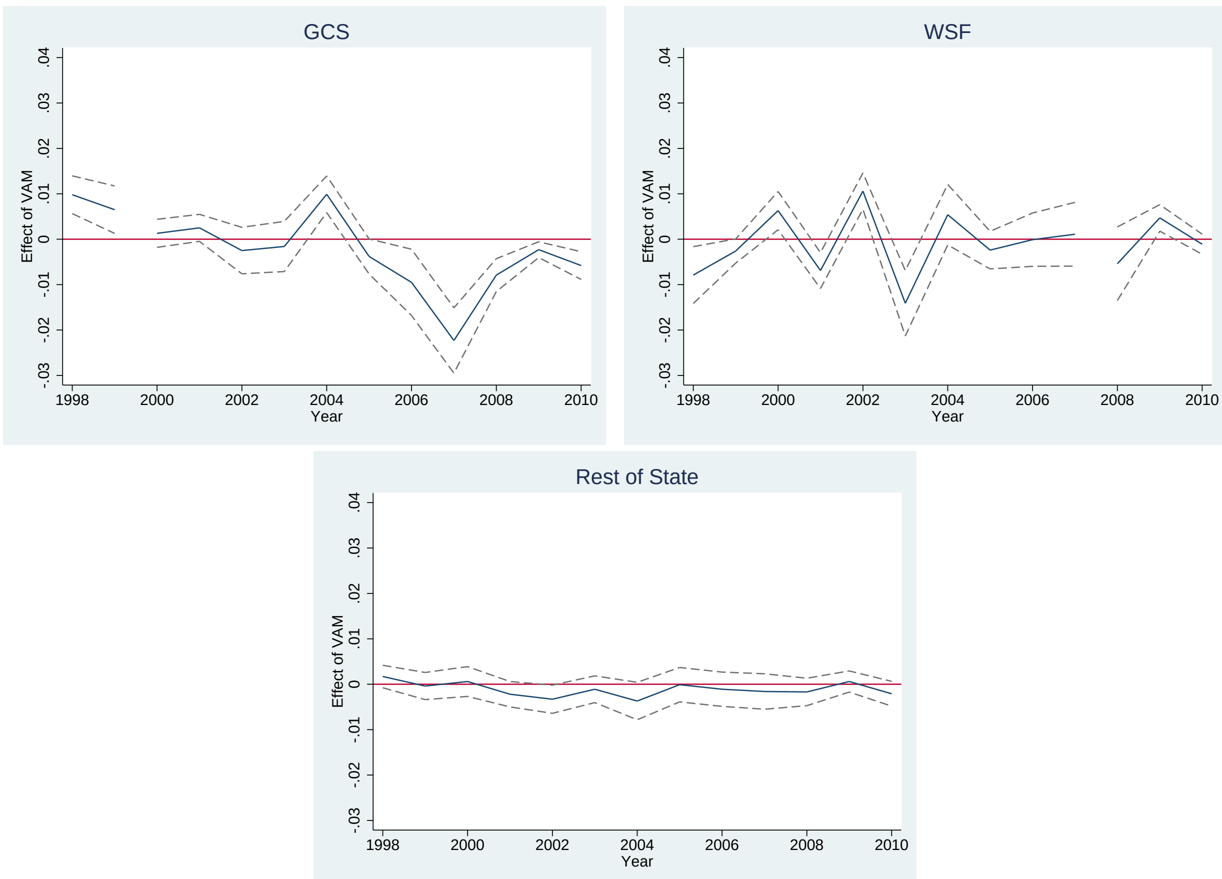

Figure G.3: The effect of VAM on the probability of moving schools out-of-district by year.

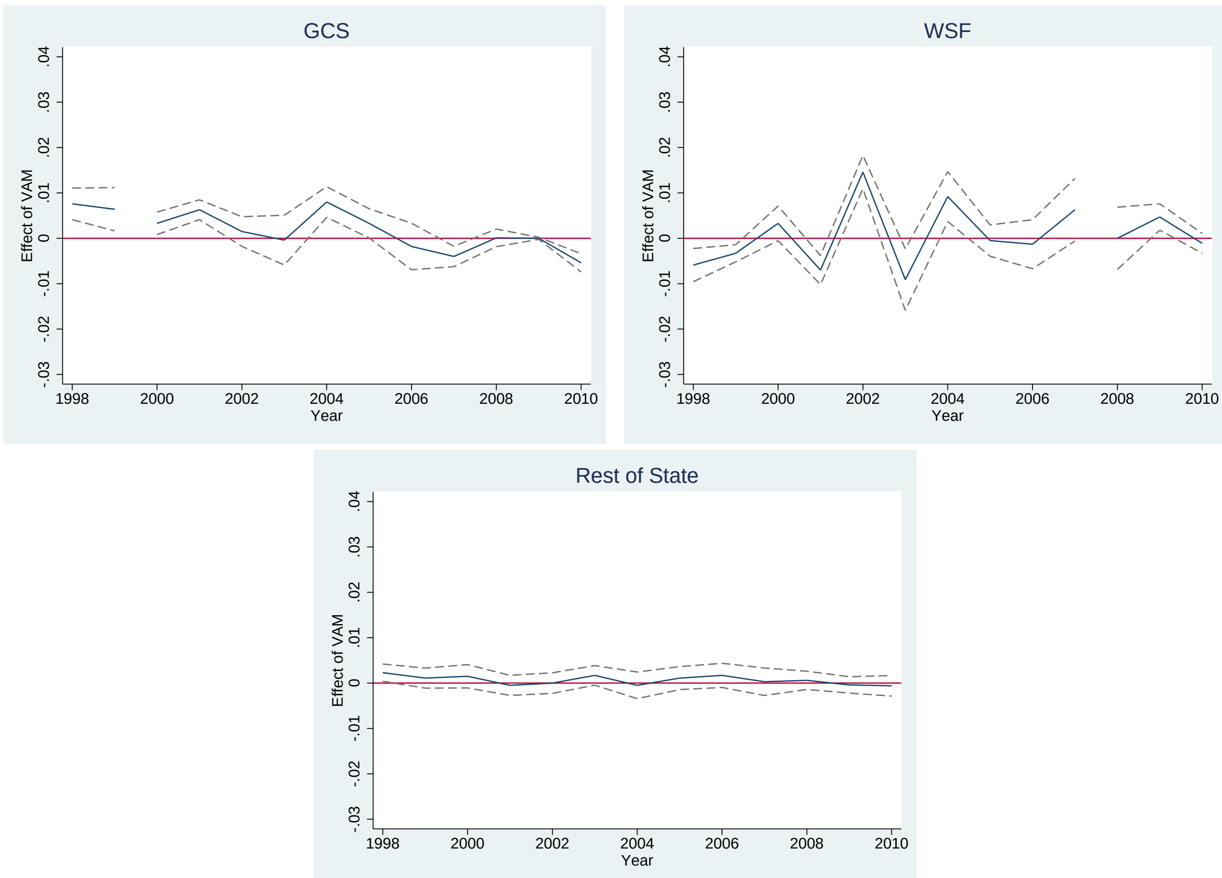

Figure G.4: The effects of VAM on the probability of moving to a better school out-of-district by year.

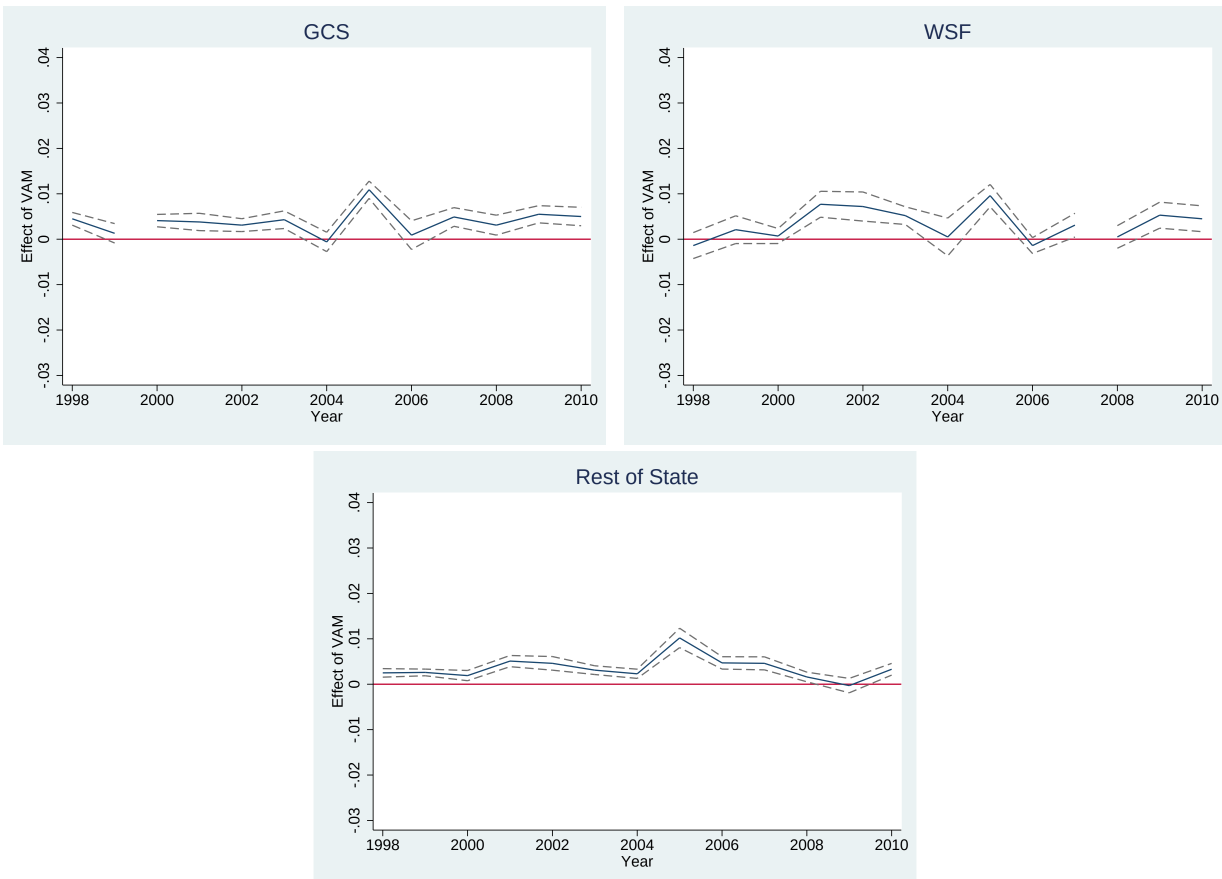

Figure G.5: The effect of VAM on teacher sorting within-district by year.

Appendix H

Additions for Chapter 2

Job-to-Job selection with respect to ability and reference group quality

As stated in Section 3.1 the optimal bids for the current (b_r) and hiring firm (b_h) are given by

$$b_r = E[\mu|R_x,\ P_r,\ P_h = P_r] = \frac{\sigma_\xi(x)\sigma_\tau(t)}{W'}m + \frac{\sigma_\tau(t)\sigma_\epsilon}{W'}R_x + \frac{2\sigma_\xi(x)\sigma_\epsilon}{W'}P_r + E(\varphi_r),$$

$$b_h = E[\mu|R_x,\ P_r = P_h,\ P_h] = \frac{\sigma_\xi(x)\sigma_\tau(0)}{W}m + \frac{\sigma_\tau(0)\sigma_\epsilon}{W}R_x + \frac{2\sigma_\xi(x)\sigma_\epsilon}{W}P_h + E(\varphi_h),$$

where $W = \sigma_\xi(x)\sigma_\tau(0) + \sigma_\tau(0)\sigma_\epsilon + 2\sigma_\xi(x)\sigma_\epsilon$ and $W' = \sigma_\xi(x)\sigma_\tau(t) + \sigma_\tau(t)\sigma_\epsilon + 2\sigma_\xi(x)\sigma_\epsilon$. Thus the probability of a job-to-job transition is the outside firm's optimal bid exceeds that of the retaining firm, as shown below:

$$P(J) = P[b_h - b_r > 0]$$

Substituting in the bids and the definition of each signal provides the following:

$$= P[\frac{\sigma_\xi(x)\sigma_\tau(0)}{W}m + \frac{\sigma_\tau(0)\sigma_\epsilon}{W}(\mu + \xi) + \frac{2\sigma_\xi(x)\sigma_\epsilon}{W}(\mu + \tau_h)$$
$$-(\frac{\sigma_\xi(x)\sigma_\tau(t)}{W'}m + \frac{\sigma_\tau(t)\sigma_\epsilon}{W'}(\mu + \xi) + \frac{2\sigma_\xi(x)\sigma_\epsilon}{W'}(\mu + \tau_r)) > E(\varphi_r) - E(\varphi_h)],$$

which after some algebra gives:

$$P\{\frac{2}{WW'}[(m-\mu)\sigma_\xi(x)^2\sigma_\epsilon(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t))+W\tau_h$$

$$-W'\tau_r+\sigma_\epsilon^2\sigma_\xi(x)(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t))\xi]>E(\varphi_r)-E(\varphi_h)\}.$$

Letting $\psi_J \equiv E(\varphi_h - E(\varphi_r)) + \frac{2}{WW'}(W\sigma_\epsilon\sigma_\xi(x)\tau_r - W'\sigma_\epsilon\sigma_\xi(x)\tau_h + \sigma_\epsilon^2\sigma_\xi(x)(\sigma_\mathcal{T}(t)-\sigma_\mathcal{T}(0))\xi)$ be the composite error term, provides the simplification:

$$P(J)=P\left\{\frac{2}{WW'}\sigma_\xi(x)^2\sigma_\epsilon[\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)](m-\mu)>\psi_J\right\}.$$

Imposing the normal and orthogonality assumptions provide:

$$P(J)=\Phi\left\{\frac{2}{\sqrt{\sigma_{\psi_J}}WW'}\left[\sigma_\xi(x)^2\sigma_\epsilon[\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)](m-\mu)\right]\right\},$$

where $\Phi\{.\}$ is the normal CDF, $\sigma_\psi = var(\psi) = 2\sigma_\varphi + \frac{4}{W^2W'^2}(W'^2\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\mathcal{T}(0)$

$+W^2\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\mathcal{T}(t)+\sigma_\epsilon^4\sigma_\xi(x)^2(\sigma_\mathcal{T}(t)-\sigma_\mathcal{T}(0))^2\sigma_\xi(x))$

and $\sigma_\varphi = var(E(\varphi_r)) = var(E(\varphi_h))$.

The derivative of the probability of job-to-job transitions with respect to ability (μ) is:

$$\frac{\partial P(J)}{\partial\mu}=-\phi\left\{\frac{2\sigma_\xi(x)\sigma_\epsilon}{WW'\sqrt{\sigma_{\psi_j}}}\sigma_\xi(x)[\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)](m-\mu)\right\}\frac{2\sigma_\xi(x)\sigma_\epsilon}{WW'\sqrt{\sigma_{\psi_j}}}\sigma_\xi(x)[\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)]$$

$\phi\{.\}$, being the normal pdf, is positive, as is each variance. Thus, as long as the precision of the private signal shrinks the longer a worker is with the retaining firm ($\sigma_\mathcal{T}(0) > \sigma_\mathcal{T}(t)$), $\frac{\partial P(J)}{\partial\mu} < 0$.

The derivative of the probability of job-to-job transitions with respect to reference group quality (m) is:

$$\frac{\partial P(J)}{\partial m} = \phi\left\{\frac{2\sigma_\xi(x)\sigma_\epsilon}{WW'\sqrt{\sigma_{\psi_j}}}\sigma_\xi(x)[\sigma_\tau(0)-\sigma_\tau(t)](m-\mu)\right\}\frac{2\sigma_\xi(x)\sigma_\epsilon}{WW'\sqrt{\sigma_{\psi_j}}}\sigma_\xi(x)[\sigma_\tau(0)-\sigma_\tau(t)]$$

Under the same conditions, $\frac{\partial P(J)}{\partial m} > 0$.

Job-to-Job dynamics with respect to working spell duration

One of the main indicators of asymmetric employer learning is the evolution of these selection effects as information accumulates the longer a worker is continuously employed. In order to examine the dynamics of this selection over working spell duration, I focus on the interactions of individual ability with working spell duration and reference group with working spell duration respectively. Below, I will first find the derivative of the probability of a job-to-job move, P(J), with respect to working spell duration, in order to find the predicted sign of the scaled coefficients on the interaction terms.

The scaled coefficient on the interaction of individual ability with working spell duration, δ_{1j}, is given below:

$$\delta_{1j} = \phi\{.\}\frac{\partial^2\{.\}}{\partial t\partial\mu} = \phi\{.\}\left[\frac{\partial[.]}{\partial\mu}\left((-\frac{1}{2})\sigma_\psi^{\frac{-3}{2}}\frac{\partial\sigma_\psi}{\partial t}\frac{2}{WW'} + \frac{\partial\frac{2}{WW'}}{\partial t}\sigma_\psi^{\frac{-1}{2}}\right) + \frac{\partial^2[.]}{\partial t\partial\mu}\sigma_\psi^{\frac{-1}{2}}\frac{2}{WW'}\right],$$

where $\phi\{.\}$ stands for $\phi\left\{\frac{2\sigma_\xi(x)\sigma_\epsilon}{WW'\sqrt{\sigma_{\psi jj}}}\sigma_\xi(x)[\sigma_\tau(0)-\sigma_\tau(t)](m-\mu)\right\}$, and [.] stands for $\left[\sigma_\xi(x)^2\sigma_\epsilon[\sigma_\tau(0)-\sigma_\tau(t)](m-\mu)\right]$.

The scaled coefficient on the interaction of average reference group ability with working spell duration, δ_{2j}, is given below:

$$\delta_{2j} = \phi\{.\}\frac{\partial^2\{.\}}{\partial t \partial m} = \phi\{.\}\left[\frac{\partial[.]}{\partial m}\left((-\frac{1}{2})\sigma_{\psi}^{\frac{-3}{2}}\frac{\partial\sigma_{\psi}}{\partial t}\frac{2}{WW'} + \frac{\partial\frac{2}{WW'}}{\partial t}\sigma_{\psi}^{\frac{-1}{2}}\right) + \frac{\partial^2[.]}{\partial t\partial m}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}\right].$$

Interaction between working spell duration and ability

I want to show:

$$\delta_{1j} = \phi\{.\}[-\sigma_{\xi}(x)^2\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)](\frac{\partial\frac{2}{WW'}}{\partial t}\sigma_{\psi}^{\frac{-1}{2}} - \sigma_{\psi}^{\frac{-3}{2}}\frac{\partial\sigma_{\psi}}{\partial t}\frac{1}{WW'}) + \frac{\partial^2[.]}{\partial t\partial\mu}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}] < 0.$$

Below I take each part of the scaled coefficient on the interaction between working spell duration and ability separately before signing the entire expression.

Starting with the first term:

$$-\sigma_{\xi}(x)^2\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)]\frac{\partial\frac{2}{WW'}}{\partial t}\sigma_{\psi}^{\frac{-1}{2}} =$$

$$2\sigma_{\xi}(x)^2\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)]\sigma_{\psi}^{\frac{-1}{2}}\frac{\partial\sigma_{\mathcal{T}}(t))}{\partial t}\left[(\sigma_{\xi}(x) + \sigma_{\epsilon})W^{-1}W'^{-2}\right].$$

Since W, W' and σ_{ψ} are each sums of variances, the assumption that $\frac{\partial\sigma_{\mathcal{T}}(t))}{\partial t} < 0$, which is key to asymmetric employer learning, implies that the first term is also negative.

Moving to the second term, recall that:

$\sigma_\psi =$

$2\sigma_\varphi + \frac{4}{W^2 W'^2}(W'^2\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\tau(0) + W^2\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\tau(t) + \sigma_\epsilon^4\sigma_\xi(x)^2(\sigma_\tau(t) - \sigma_\tau(0))^2\sigma_\xi(x)).$

Thus,

$-\sigma_\xi(x)^4\sigma_\epsilon^3[\sigma_\tau(0) - \sigma_\tau(t)]\sigma_\psi^{\frac{-3}{2}}\frac{\partial\sigma_\psi}{\partial t}\frac{1}{WW'} = \sigma_\xi(x)^2\sigma_\epsilon(\sigma_\tau(0) - \sigma_\tau(t))\frac{(1)}{WW'}\sigma_\psi^{\frac{-3}{2}}\frac{\partial\sigma_\tau(t))}{\partial t}\{$

$\frac{8}{W^2W'^2}\sigma_\epsilon^2\sigma_\xi(x)(\sigma_\tau(0) - \sigma_\tau(t))(\frac{1}{W'}(\sigma_\xi(x) + \sigma_\epsilon)(\sigma_\tau(0) - \sigma_\tau(t)) + 1)$

$+\frac{4}{W'^3}\left[2(\sigma_\xi(x) + \sigma_\epsilon)\sigma_\tau(t) - W'\right]\}.$

As $\frac{\sigma_\xi(x)^2\sigma_\epsilon(\sigma_\tau(0) - \sigma_\tau(t))}{WW'}\sigma_\psi^{\frac{-3}{2}}\frac{\partial\sigma_\tau(t))}{\partial t}\frac{4\sigma_\epsilon\sigma_\xi(x)}{W'^3}$, cannot be eliminated here, I will discuss a sufficient condition under which this source of ambiguity may be eliminated below.

Finally, examine the last term.

$\frac{\partial[.]}{\partial\mu} = \frac{2\sigma_\xi(x)\sigma_\epsilon}{WW'\sqrt{\sigma_{\psi_j}}}\sigma_\xi(x)[\sigma_\tau(0) - \sigma_\tau(t)]$, which means that,

$\frac{\partial^2[.]}{\partial t\partial\mu}\sigma_\psi^{\frac{-1}{2}}\frac{2\sigma_\xi(x)^2\sigma_\epsilon}{WW'} = \sigma_{\psi_J}^{\frac{-1}{2}}\frac{2}{WW'}\frac{\partial\sigma_\tau(t))}{\partial t}.$

Again, the assumption that $\frac{\partial\sigma_\tau(t))}{\partial t} < 0$, which is central to asymmetric learning, implies that the selection on ability should grow stronger (more negative) with increases in working spell duration.

Thus, the interaction term between working spell duration and ability is given by the following:

$\delta_{1j} = \sigma_\xi(x)^2\sigma_\epsilon\sigma_\psi^{\frac{-1}{2}}\frac{2}{WW'}\frac{\partial\sigma_\tau(t))}{\partial t}\{[\sigma_\tau(0) - \sigma_\tau(t)][(\sigma_\xi(x) + \sigma_\epsilon)W'^{-1}$

$+\sigma_\psi^{-1}\frac{4}{W^2W'^2}\sigma_\epsilon^4\sigma_\xi(x)^3(\sigma_\tau(0) - \sigma_\tau(t))(\frac{1}{W'}(\sigma_\xi(x) + \sigma_\epsilon)(\sigma_\tau(0) - \sigma_\tau(t)) + 1)$

$+\frac{2}{W'^2}\left[\frac{2}{W'}(\sigma_\xi(x) + \sigma_\epsilon)\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\tau(t) - \sigma_\epsilon^2\sigma_\xi(x)^2\right]] + 1\}$

With sufficient variability in match quality, this entire term is negative, meaning that as working spell duration increases, selection on the basis of ability should become more negative.

Interaction between working spell duration and reference group

I want to show:

$$\delta_{2j} = \phi\{.\}[\sigma_\xi(x)^2\sigma_\epsilon[\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t)](\frac{\partial \frac{2}{WW'}}{\partial t}\sigma_\psi^{\frac{-1}{2}} - \sigma_\psi^{\frac{-3}{2}}\frac{\partial \sigma_\psi}{\partial t}\frac{1}{WW'}) + \frac{\partial^2[.]}{\partial t \partial \mu}\sigma_\psi^{\frac{-1}{2}}\frac{2}{WW'}] > 0.$$

The algebra is equivalent to the previous subsection, just with the opposite sign. Thus, the scaled coefficient of the interaction between working spell duration and experience is given by following.

$$\delta_{2j} = -\sigma_\xi(x)^2\sigma_\epsilon\sigma_\psi^{\frac{-1}{2}}\frac{2}{WW'}\frac{\partial\sigma_\mathcal{T}(t))}{\partial t}\{[\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t)][(\sigma_\xi(x) + \sigma_\epsilon)W'^{-1}$$
$$+\sigma_\psi^{-1}\frac{4}{W^2W'^2}\sigma_\epsilon^4\sigma_\xi(x)^3(\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t))(\frac{1}{W'}(\sigma_\xi(x) + \sigma_\epsilon)(\sigma_\mathcal{T}(0) - \sigma_\mathcal{T}(t)) + 1)$$
$$+\frac{4}{W'^2}\left[\frac{1}{W'}(\sigma_\xi(x) + \sigma_\epsilon)\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\mathcal{T}(t) - \sigma_\epsilon^2\sigma_\xi(x)^2\right]] + 1\}$$

Under the same assumptions made previously, this term is positive meaning that as working spell duration increases, the selection of mobile workers on the basis of their reference group should become stronger (more positive).

Job-to-Job dynamics with respect to experience

The scaled coefficient on the interaction of individual ability with experience, γ_{1j}, is given below:

$$\gamma_{1j} = \phi\{.\}\frac{\partial^2\{.\}}{\partial x \partial \mu} = \phi\{.\}\left[\frac{\partial[.]}{\partial \mu}\left((-\frac{1}{2})\sigma_{\psi}^{\frac{-3}{2}}\frac{\partial \sigma_{\psi}}{\partial x}\frac{2}{WW'} + \frac{\partial \frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\right) + \frac{\partial^2[.]}{\partial x \partial \mu}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}\right],$$

where $\phi\{.\}$ stands for $\phi\left\{\frac{2\sigma_{\xi}(x)\sigma_{\epsilon}}{WW'\sqrt{\sigma_{\psi}jj}}\sigma_{\xi}(x)[\sigma_{\tau}(0) - \sigma_{\tau}(t)](m-\mu)\right\}$, and [.] stands for $\left[\sigma_{\xi}(x)^2\sigma_{\epsilon}[\sigma_{\tau}(0) - \sigma_{\tau}(t)](m-\mu)\right]$.

The scaled coefficient on the interaction of average reference group ability with experience, γ_{2j}, is given below:

$$\gamma_{2j} = \phi\{.\}\frac{\partial^2\{.\}}{\partial x \partial m} = \phi\{.\}\left[\frac{\partial[.]}{\partial m}\left((-\frac{1}{2})\sigma_{\psi}^{\frac{-3}{2}}\frac{\partial \sigma_{\psi}}{\partial x}\frac{2}{WW'} + \frac{\partial \frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\right) + \frac{\partial^2[.]}{\partial x \partial m}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}\right].$$

In the following subsections, I take the derivative of $P(J)$ with respect to experience, and then derive these scaled coefficients for the interactions with ability and reference group respectively.

The derivative of $P(J)$ with respect to experience

The derivative of $P(J)$ with respect to experience is:

$$\frac{\partial P(J)}{\partial x} = \phi\{.\}\left[\frac{\partial[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'} + \frac{\partial \frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}[.] - \frac{1}{2}\sigma_{\psi}^{\frac{-3}{2}}\frac{\partial \sigma_{\psi}}{\partial x}\frac{2}{WW'}[.]\right],$$

where $\phi\{.\}$ stands for $\phi\left\{\frac{2\sigma_{\xi}(x)\sigma_{\epsilon}}{WW'\sqrt{\sigma_{\psi_j}}}\sigma_{\xi}(x)[\sigma_{\tau}(0) - \sigma_{\tau}(t)](m-\mu)\right\}$, and [.] stands for $\left[\sigma_{\xi}(x)^2\sigma_{\epsilon}[\sigma_{\tau}(0) - \sigma_{\tau}(t)](m-\mu)\right]$. The following takes each component separately and resolves any conflicting signs.

$\frac{\partial[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}$,

$$\frac{\partial\,[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'} = \sigma_{\psi}^{\frac{-1}{2}}\frac{4}{WW'}\frac{\partial\sigma_{\xi}(x)}{\partial x}\left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)](m-\mu)\right].$$

$\frac{\partial\sigma_{\xi}(x)}{\partial x} < 0$ and $\sigma_{\mathcal{T}}(0) > \sigma_{\mathcal{T}}(t)$ by assumptions intrinsic to public and private employer learning respectively. Therefore, if $\mu > m$, $\frac{\partial[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'} > 0$, since all other terms are sums of variances, and positive by definition.

$\frac{\partial\frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}[.]$

$$\frac{\partial\frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}[.] = (-2)\sigma_{\psi}^{\frac{-1}{2}}[\sigma_{\xi}(x)^2\sigma_{\epsilon}(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))(m-\mu)]\frac{\partial\sigma_{\mathcal{T}}(t)}{\partial x}$$

$$\left[(\sigma_{\mathcal{T}}(0)+2\sigma_{\epsilon})W^{-2}W'^{-1} + (\sigma_{\mathcal{T}}(t)+2\sigma_{\epsilon})W^{-1}W'^{-2}\right].$$

Under the same learning assumptions, the sign of $\frac{\partial\frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}[.]$, also depends upon the whether ability (μ) is higher than reference group quality (m), although here it is in the opposite direction. $\frac{\partial\frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}$.

$\frac{\partial[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'} + \frac{\partial\frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}[.]$,

The conflict in the sign of the first two terms is resolved below.

$$\frac{\partial[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}+\frac{\partial\frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}[.] = \sigma_{\psi}^{\frac{-1}{2}}\frac{4}{WW'}\frac{\partial\sigma_{\xi}(x)}{\partial x}\left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)](m-\mu)\right]-2\sigma_{\psi}^{\frac{-1}{2}}$$
$$[\sigma_{\xi}(x)^2\sigma_{\epsilon}(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))(m-\mu)]\frac{\partial\sigma_{\xi}(x)}{\partial x}\left[(\sigma_{\mathcal{T}}(0)+2\sigma_{\epsilon})W^{-2}W'^{-1} + (\sigma_{\mathcal{T}}(t)+2\sigma_{\epsilon})W^{-1}W'^{-2}\right]$$

$$= \sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}\frac{\partial\sigma_{\xi}(x)}{\partial x}\left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)](m-\mu)\right](2-\sigma_{\xi}(x)((\sigma_{\mathcal{T}}(0)+2\sigma_{\epsilon})W^{-1}$$
$$+(\sigma_{\mathcal{T}}(t)+2\sigma_{\epsilon})W'^{-1})$$

$$= \sigma_{\psi}^{\frac{-1}{2}} \frac{2}{W^2 W'^2} \frac{\partial \sigma_{\xi}(x)}{\partial x} \left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)](m-\mu)\right]$$

$$(2WW' - \sigma_{\xi}(x)((\sigma_{\mathcal{T}}(0) + 2\sigma_{\epsilon})W' + (\sigma_{\mathcal{T}}(t) + 2\sigma_{\epsilon})W$$

$$= \sigma_{\psi}^{\frac{-1}{2}} \frac{2}{W^2 W'^2} \frac{\partial \sigma_{\xi}(x)}{\partial x} \left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)](m-\mu)\right] ($$

$$2\sigma_{\mathcal{T}}(t)\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(0) + 2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x) + 4\sigma_{\epsilon}\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(0)$$

$$+2\sigma_{\mathcal{T}}(t)\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon} + 2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2\sigma_{\mathcal{T}}(0) + 4\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}^2 + 4\sigma_{\mathcal{T}}(t)\sigma_{\xi}(x)^2\sigma_{\epsilon}$$

$$+4\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2\sigma_{\xi}(x) + 8\sigma_{\epsilon}^2\sigma_{\xi}(x)^2 - \sigma_{\xi}(x)(\sigma_{\mathcal{T}}(0) + 2\sigma_{\epsilon})(\sigma_{\epsilon}\sigma_{\mathcal{T}}(t) + \sigma_{\xi}(x)\sigma_{\mathcal{T}}(t)$$

$$+2\sigma_{\epsilon}\sigma_{\xi}(x)) - \sigma_{\xi}(x)(\sigma_{\mathcal{T}}(t) + 2\sigma_{\epsilon})(\sigma_{\epsilon}\sigma_{\mathcal{T}}(t) + \sigma_{\xi}(x)\sigma_{\mathcal{T}}(t) + 2\sigma_{\epsilon}\sigma_{\xi}(x))$$

$$= \sigma_{\psi}^{\frac{-1}{2}} \frac{2}{W^2 W'^2} \frac{\partial \sigma_{\xi}(x)}{\partial x} \left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)](m-\mu)\right] ($$

$$2\sigma_{\mathcal{T}}(t)\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(0) + 2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x) + 4\sigma_{\epsilon}\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(0)$$

$$+2\sigma_{\mathcal{T}}(t)\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon} + 2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2\sigma_{\mathcal{T}}(0) + 4\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}^2 + 4\sigma_{\mathcal{T}}(t)\sigma_{\xi}(x)^2\sigma_{\epsilon}$$

$$+4\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2\sigma_{\xi}(x) + 8\sigma_{\epsilon}^2\sigma_{\xi}(x)^2 - (\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}\sigma_{\mathcal{T}}(t) + \sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(t)$$

$$+2\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}) - (2\sigma_{\xi}(x)\sigma_{\epsilon}^2\sigma_{\mathcal{T}}(t) + 2\sigma_{\epsilon}\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(t) + 4\sigma_{\epsilon}^2\sigma_{\xi}(x)^2)$$

$$-(\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}\sigma_{\mathcal{T}}(t) + \sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(t) + 2\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon})$$

$$-(2\sigma_{\xi}(x)\sigma_{\epsilon}^2\sigma_{\mathcal{T}}(0) + 2\sigma_{\epsilon}\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(0) + 4\sigma_{\epsilon}^2\sigma_{\xi}(x)^2))$$

$$= \sigma_{\psi}^{\frac{-1}{2}} \frac{2}{W^2 W'^2} \frac{\partial \sigma_{\xi}(x)}{\partial x} \left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)](m-\mu)\right] ($$

$$2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x) + 2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2\sigma_{\mathcal{T}}(0) + 2\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}^2 + 2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2\sigma_{\xi}(x)) > 0,$$

if $\mu > m$. Since the negative term is larger in magnitude than the positive term, $\frac{\partial[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}} \frac{2}{WW'} + \frac{\partial \frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}} [.] > 0$, under the above assumptions.

Ambiguity of $(-\frac{1}{2})\sigma_{\psi}^{\frac{-3}{2}}\frac{\partial\sigma_{\psi}}{\partial x}\frac{2}{WW'}[.]$

$$(-\tfrac{1}{2})\sigma_{\psi}^{\frac{-3}{2}}\frac{\partial\sigma_{\psi}}{\partial x}\frac{2}{WW'}[.] = [\sigma_{\xi}(x)^2\sigma_{\epsilon}(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))(m-\mu)]\frac{(-1)}{WW'}\sigma_{\psi}^{\frac{-3}{2}}$$
$$\frac{\partial\sigma_{\xi}(x)}{\partial x}[\frac{(-8)}{W^2W'^2}\left[\frac{1}{W}(\sigma_{\mathcal{T}}(0)+2\sigma_{\epsilon})+\frac{1}{W'}(\sigma_{\mathcal{T}}(t)+2\sigma_{\epsilon})\right]$$
$$(W'^2\sigma_{\epsilon}^2\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(0)+W^2(\sigma_{\epsilon}^2\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(t)+\sigma_{\epsilon}^4\sigma_{\xi}(x)^3(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))^2)$$
$$+\frac{4}{W^2W'^2}[2W'(\sigma_{\mathcal{T}}(0)+2\sigma_{\epsilon})\sigma_{\epsilon}^2\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(0)+2W'^2\sigma_{\epsilon}^2\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)+2W(\sigma_{\mathcal{T}}(t)$$
$$+2\sigma_{\epsilon})\sigma_{\epsilon}^2\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(t)+2W^2\sigma_{\epsilon}^2\sigma_{\xi}(x)\sigma_{\mathcal{T}}(t)+3\sigma_{\epsilon}^4\sigma_{\xi}(x)^2(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))^2)]]$$

$$= [\sigma_{\xi}(x)^2\sigma_{\epsilon}(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))(m-\mu)]\frac{(-1)}{WW'}\sigma_{\psi}^{\frac{-3}{2}}$$
$$\frac{\partial\sigma_{\xi}(x)}{\partial x}[\frac{8}{W^3}\sigma_{\epsilon}^3\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)^2+\frac{8}{W'^3}\sigma_{\epsilon}^3\sigma_{\xi}(x)\sigma_{\mathcal{T}}(t)^2+\frac{4}{W^3W'^3}\sigma_{\epsilon}^4\sigma_{\xi}(x)^2(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))^2$$
$$[2\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}+3\sigma_{\mathcal{T}}(0)\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2+2\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}^2+2\sigma_{\xi}(x)\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2$$
$$-(\sigma_{\mathcal{T}}(0)+2\sigma_{\epsilon})(\sigma_{\mathcal{T}}(t)+2\sigma_{\epsilon})\sigma_{\xi}(x)^2]]$$

The ambiguity in the third term is produced by the following, which I will define as A :

$$A \equiv \sigma_{\psi}^{\frac{-3}{2}}\frac{-1}{W^4W'^4}\left[\sigma_{\xi}(x)^2\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)](m-\mu)\right]\frac{\partial\sigma_{\xi}(x)}{\partial x}$$

$$[8\sigma_{\epsilon}^3\sigma_{\xi}(x)(W'^3\sigma_{\mathcal{T}}(0)^2+W^3\sigma_{\mathcal{T}}(t)^2)+4\sigma_{\epsilon}^5\sigma_{\xi}(x)^2(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))^2($$

$$2\sigma_{\mathcal{T}}(t)\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)+3\sigma_{\mathcal{T}}(t)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}+2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}\sigma_{\xi}(x)+2\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}\sigma_{\xi}(x))].$$

Resolution of ambiguity in the third term

$\frac{\partial[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}+\frac{\partial\frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}[.]+A>0$

From Appendix A,

$$\frac{\partial\,[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}+\frac{\partial\frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}[.]=\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{W^2W'^2}\frac{\partial\sigma_\xi(x)}{\partial x}[.]\,($$

$$2\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x)+2\sigma_\tau(t)\sigma_\epsilon^2\sigma_\tau(0)+2\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon^2+2\sigma_\tau(t)\sigma_\epsilon^2\sigma_\xi(x)).$$

Adding the negative A to this term gives:

$\frac{\partial[.]}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}\frac{2}{WW'}+\frac{\partial\frac{2}{WW'}}{\partial x}\sigma_{\psi}^{\frac{-1}{2}}[.]+A=\sigma_{\psi}^{\frac{-3}{2}}\frac{4}{W^4W'^4}\frac{\partial\sigma_\xi(x)}{\partial x}[.]\,[$

$\sigma_\psi W^2W'^2(\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x)+\sigma_\tau(t)\sigma_\epsilon^2\sigma_\tau(0)+\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon^2+\sigma_\tau(t)\sigma_\epsilon^2\sigma_\xi(x))$

$-[2\sigma_\epsilon^3\sigma_\xi(x)^2(W'^3\sigma_\tau(0)^2+W^3\sigma_\tau(t)^2)+\sigma_\epsilon^5\sigma_\xi(x)^3(\sigma_\tau(0)-\sigma_\tau(t))^2($

$2\sigma_\tau(t)\sigma_\tau(0)\sigma_\xi(x)+3\sigma_\tau(t)\sigma_\tau(0)\sigma_\epsilon+2\sigma_\tau(t)\sigma_\epsilon\sigma_\xi(x)+2\sigma_\tau(0)\sigma_\epsilon\sigma_\xi(x))]$

$=\sigma_{\psi}^{\frac{-3}{2}}\frac{4}{W^4W'^4}\frac{\partial\sigma_\xi(x)}{\partial x}\left[\sigma_\xi(x)\sigma_\epsilon[\sigma_\tau(0)-\sigma_\tau(t)](m-\mu)\right](2\sigma_\varphi W^2W'^2+$

$4(W'^2\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\tau(0)+W^2\sigma_\epsilon^2\sigma_\xi(x)^2\sigma_\tau(t)+\sigma_\epsilon^4\sigma_\xi(x)^3(\sigma_\tau(0)-\sigma_\tau(t))^2)$

$(2\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x)+2\sigma_\tau(t)\sigma_\epsilon^2\sigma_\tau(0)+2\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon^2+2\sigma_\tau(t)\sigma_\epsilon^2\sigma_\xi(x))$

$-[2\sigma_\epsilon^3\sigma_\xi(x)^2(W'^3\sigma_\tau(0)^2+W^3\sigma_\tau(t)^2)+\sigma_\epsilon^5\sigma_\xi(x)^3(\sigma_\tau(0)-\sigma_\tau(t))^2($

$2\sigma_\tau(t)\sigma_\tau(0)\sigma_\xi(x)+3\sigma_\tau(t)\sigma_\tau(0)\sigma_\epsilon+2\sigma_\tau(t)\sigma_\epsilon\sigma_\xi(x)+2\sigma_\tau(0)\sigma_\epsilon\sigma_\xi(x))]$

$=\sigma_{\psi}^{\frac{-3}{2}}\frac{4}{W^4W'^4}\frac{\partial\sigma_\xi(x)}{\partial x}\left[\sigma_\xi(x)\sigma_\epsilon[\sigma_\tau(0)-\sigma_\tau(t)](m-\mu)\right](2\sigma_\varphi W^2W'^2$

$+8W'^3\sigma_\epsilon^3\sigma_\xi(x)^2\sigma_\tau(0)^2+8W'^2\sigma_\epsilon^4\sigma_\xi(x)^3\sigma_\tau(0)\sigma_\tau(t)$

$+8W^3\sigma_\epsilon^3\sigma_\xi(x)^2\sigma_\tau(t)^2+8W^2\sigma_\epsilon^4\sigma_\xi(x)^3\sigma_\tau(0)\sigma_\tau(t)+4\sigma_\epsilon^4\sigma_\xi(x)^3(\sigma_\tau(0)-\sigma_\tau(t))^2)$

$(2\sigma_\tau(t)\sigma_\epsilon\sigma_\tau(0)\sigma_\xi(x)+2\sigma_\tau(t)\sigma_\epsilon^2\sigma_\tau(0)+2\sigma_\xi(x)\sigma_\tau(0)\sigma_\epsilon^2+2\sigma_\tau(t)\sigma_\epsilon^2\sigma_\xi(x))$

$-[2\sigma_{\epsilon}^{3}\sigma_{\xi}(x)^{2}(W'^{3}\sigma_{\mathcal{T}}(0)^{2}+W^{3}\sigma_{\mathcal{T}}(t)^{2})+\sigma_{\epsilon}^{5}\sigma_{\xi}(x)^{3}(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))^{2}($

$2\sigma_{\mathcal{T}}(t)\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)+3\sigma_{\mathcal{T}}(t)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}+2\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}\sigma_{\xi}(x)+2\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}\sigma_{\xi}(x))]$

$=\sigma_{\psi}^{\frac{-3}{2}}\frac{4}{W^{4}W'^{4}}\frac{\partial\sigma_{\xi}(x)}{\partial x}\left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)](m-\mu)\right](2\sigma_{\varphi}W^{2}W'^{2}+6W'^{3}\sigma_{\epsilon}^{3}\sigma_{\xi}(x)^{2}\sigma_{\mathcal{T}}(0)^{2}$

$+8W'^{2}\sigma_{\epsilon}^{4}\sigma_{\xi}(x)^{3}\sigma_{\mathcal{T}}(0)\sigma_{\mathcal{T}}(t)+6W^{3}\sigma_{\epsilon}^{3}\sigma_{\xi}(x)^{2}\sigma_{\mathcal{T}}$

$(t)^{2}+8W^{2}\sigma_{\epsilon}^{4}\sigma_{\xi}(x)^{3}\sigma_{\mathcal{T}}(0)\sigma_{\mathcal{T}}(t)+\sigma_{\epsilon}^{4}\sigma_{\xi}(x)^{3}(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))^{2})(6\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)$

$+5\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^{2}\sigma_{\mathcal{T}}(0)+6\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}^{2}+6\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^{2}\sigma_{\xi}(x)).$

Interaction between experience and ability

I use the derivative of $P(J)$ with respect to experience to sign the scaled coefficient for the interaction of experience and ability. Taking the derivative of $\frac{\partial\{.\}}{\partial x}$ with respect to μ gives the following.

$\gamma_{1j}=-\phi\{.\}\sigma_{\psi}^{\frac{-3}{2}}\frac{4}{W^{4}W'^{4}}\frac{\partial\sigma_{\xi}(x)}{\partial x}\left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)]\right]$

$(2\sigma_{\varphi}W^{2}W'^{2}+6W'^{3}\sigma_{\epsilon}^{3}\sigma_{\xi}(x)^{2}\sigma_{\mathcal{T}}(0)^{2}$

$+8W'^{2}\sigma_{\epsilon}^{4}\sigma_{\xi}(x)^{3}\sigma_{\mathcal{T}}(0)\sigma_{\mathcal{T}}(t)+6W^{3}\sigma_{\epsilon}^{3}\sigma_{\xi}(x)^{2}\sigma_{\mathcal{T}}$

$(t)^{2}+8W^{2}\sigma_{\epsilon}^{4}\sigma_{\xi}(x)^{3}\sigma_{\mathcal{T}}(0)\sigma_{\mathcal{T}}(t)+\sigma_{\epsilon}^{4}\sigma_{\xi}(x)^{3}(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))^{2})(6\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)$

$+5\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^{2}\sigma_{\mathcal{T}}(0)+6\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}^{2}+6\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^{2}\sigma_{\xi}(x)).$

Under the assumptions that $\frac{\partial\sigma_{\xi}(x)}{\partial x}<0$, which is fundamental to pubic employer learning, and $\sigma_{t}(t)<\sigma_{\mathcal{T}}(0)$, which is fundamental to private employer learning, $\gamma_{1j}>0$. The selection on the basis of ability weakens (becomes more positive) with increases in experience.

Interaction between experience and reference group

I also use the derivative of $P(J)$ with respect to experience to sign the scaled coefficient for the interaction of experience and reference group ability. Taking the derivative of $\frac{\partial\{.\}}{\partial x}$ with respect to m gives the following.

$$\gamma_{2j} = \phi\left\{.\right\}\sigma_{\psi}^{\frac{-3}{2}} \frac{4}{W^4 W'^4} \frac{\partial \sigma_{\xi}(x)}{\partial x} \left[\sigma_{\xi}(x)\sigma_{\epsilon}[\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t)]\right]$$

$$(2\sigma_{\varphi}W^2W'^2 + 6W'^3\sigma_{\epsilon}^3\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}(0)^2$$

$$+8W'^2\sigma_{\epsilon}^4\sigma_{\xi}(x)^3\sigma_{\mathcal{T}}(0)\sigma_{\mathcal{T}}(t) + 6W^3\sigma_{\epsilon}^3\sigma_{\xi}(x)^2\sigma_{\mathcal{T}}$$

$$(t)^2 + 8W^2\sigma_{\epsilon}^4\sigma_{\xi}(x)^3\sigma_{\mathcal{T}}(0)\sigma_{\mathcal{T}}(t) + \sigma_{\epsilon}^4\sigma_{\xi}(x)^3(\sigma_{\mathcal{T}}(0) - \sigma_{\mathcal{T}}(t))^2)(6\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)$$

$$+5\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2\sigma_{\mathcal{T}}(0) + 6\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}^2 + 6\sigma_{\mathcal{T}}(t)\sigma_{\epsilon}^2\sigma_{\xi}(x)).$$

Under the same assumptions, $\gamma_{2j} < 0$, which means that the selection on the basis of reference group also weakens, as in this case, the selection effects become more negative with increases in experience.

Expected Profits

$E[\pi|R_x, P_r, P_h] = E[\mu|R_x, P_r, P_h] - w.$

For workers with the retaining firm,$w = E[\mu|R_x, P_r = P_h, P_h]$ and $E[\pi|R_x, P_r, P_h] = E[\mu|R_x, P_r, P_h] - E[\mu|R_x, P_r = P_h, P_h]$

$$E[\pi|R_x, P_r, P_h] = \frac{\sigma_{\mathcal{T}}(0)\sigma_{\mathcal{T}}(t)\sigma_{\xi}(x)}{Q}m + \frac{\sigma_{\mathcal{T}}(t)\sigma_{\xi}(x)\sigma_{\epsilon}}{Q}P_h + \frac{\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)\sigma_{\epsilon}}{Q}P_r + \frac{\sigma_{\mathcal{T}}(t)\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}}{Q}R_x$$

$$+E(\varphi_r) - \left(\frac{\sigma_\tau(0)\sigma_\xi(x)}{Q'}m + \frac{\sigma_\tau(0)\sigma_\epsilon}{Q'}R_x + \frac{2\sigma_\epsilon\sigma_\xi(x)}{Q'}P_h + E(\varphi_h)\right)$$

where $Q = \sigma_\tau(0)\sigma_\tau(t)\sigma_\xi(x) + \sigma_\tau(t)\sigma_\xi(x)\sigma_\epsilon + \sigma_\tau(0)\sigma_\xi(x)\sigma_\epsilon + \sigma_\tau(t)\sigma_\tau(0)\sigma_\epsilon$

and $Q' = \sigma_\tau(0)\sigma_\xi(x) + \sigma_\tau(0)\sigma_\epsilon + 2\sigma_\epsilon\sigma_\xi(x)$. Substituting in the definition of each signal gives the following:

$$= \frac{\sigma_\tau(0)\sigma_\tau(t)\sigma_\xi(x)}{Q}m + \frac{\sigma_\tau(t)\sigma_\xi(x)\sigma_\epsilon}{Q}(\mu + v) + \frac{\sigma_\tau(0)\sigma_\xi(x)\sigma_\epsilon}{Q}(\mu + \tau) + \frac{\sigma_\tau(t)\sigma_\tau(0)\sigma_\epsilon}{Q}(\mu + \xi)$$

$$+E(\varphi_r) - \left(\frac{\sigma_\tau(0)\sigma_\xi(x)}{Q'}m + \frac{\sigma_\tau(0)\sigma_\epsilon}{Q'}(\mu + \xi) + \frac{2\sigma_\epsilon\sigma_\xi(x)}{Q'}(\mu + v) + E(\varphi_h)\right).$$

This simplifies to an expression similar to that shown for the probability of job-to-job moves.

$$= \frac{\sigma_\xi(x)\sigma_\epsilon\sigma_\tau(0)}{Q'Q}[\sigma_\xi(x)\left(\sigma_\tau(0) - \sigma_\tau(t)\right)(\mu - m)] + \frac{1}{Q'Q}$$

$$[Q'\sigma_\epsilon(\sigma_\tau(0)\sigma_\xi(x)\tau_r + \sigma_\tau(t)\sigma_\xi(x)\tau_h + \sigma_\tau(t)\sigma_\tau(0)\xi) - Q\sigma_\epsilon(2\sigma_\xi(x)\tau_h + \sigma_\tau(0)\xi)] + E(\varphi_r - \varphi_h)$$

Note that in expectation $\tau_h = \tau_r = \xi = \varphi = 0$, the expression above simplifies to $= \frac{\sigma_\xi(x)\sigma_\epsilon\sigma_\tau(0)}{Q'Q}\left[\sigma_\xi(x)\left(\sigma_\tau(0) - \sigma_\tau(t)\right)(\mu - m)\right] > 0$

Layoff selection with respect to ability and reference group quality

Under nominal wage rigidity, the probability of a layoff is given by:

$$P(L) = P\left[E[\pi|R_x, P_r, P_h, \varphi] < 0\right]$$

$$P(L) = P\left\{\psi_{L\varphi} > \frac{\sigma_{\xi}(x)\sigma_{\epsilon}}{Q'Q}\left[\sigma_{\xi}(x)\sigma_{\tau}(0)\left(\sigma_{\tau}(0) - \sigma_{\tau}(t)\right)(m-\mu)\right]\right\}$$

where $\psi_{L\varphi} = \frac{1}{Q'Q}[(Q'\sigma_{\tau}(0)\tau_r - Q\tau_h + \sigma_{\epsilon}\sigma_{\tau_h}(\sigma_{\tau}(t) - \sigma_{\tau}(0))\xi] + \varphi_r - E(\varphi_h)$,

$Q = \sigma_{\tau}(0)\sigma_{\tau}(t)\sigma_{\xi}(x) + \sigma_{\tau}(t)\sigma_{\xi}(x)\sigma_{\epsilon} + \sigma_{\tau}(0)\sigma_{\xi}(x)\sigma_{\epsilon} + \sigma_{\tau}(t)\sigma_{\tau}(0)\sigma_{\epsilon}$,

and $Q' = \sigma_{\tau}(0)\sigma_{\xi}(x) + \sigma_{\tau}(0)\sigma_{\epsilon} + 2\sigma_{\epsilon}\sigma_{\xi}(x)$.

Imposing the normal and orthogonality assumptions provide:

$$P(L) = \Phi\left\{\frac{\sigma_{\xi}(x)\sigma_{\epsilon}}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\left[\sigma_{\xi}(x)\sigma_{\tau}(0)\left(\sigma_{\tau}(0) - \sigma_{\tau}(t)\right)(m-\mu)\right]\right\}$$

where $\sigma_{\psi_{L\varphi}} = Q^{-2}Q'^{-2}[Q'^2(\sigma_{\tau}(0)^2\sigma_{\tau}(t) + Q^2\sigma_{\tau}(0) + \sigma_{\epsilon}^2\sigma_{\tau_h}^2(\sigma_{\tau}(t) - \sigma_{\tau}(0))^2\sigma_{\xi}(x)] + \sigma_{\varphi} + \sigma_{E\varphi}$

Taking the derivative with respect to ability (μ) gives the following:

$$\frac{\partial P(L)}{\partial \mu} = -\phi\left\{\frac{\sigma_{\xi}(x)\sigma_{\epsilon}}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\left[\sigma_{\xi}(x)\sigma_{\tau}(0)\left(\sigma_{\tau}(0) - \sigma_{\tau}(t)\right)(m-\mu)\right]\right\}$$

$$\frac{\sigma_{\xi}(x)\sigma_{\epsilon}}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\left(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)\right)<0.$$

The equation above illustrates that as ability (μ) increases the probability of layoff should fall, as long as $\sigma_{\mathcal{T}}(0) > \sigma_{\mathcal{T}}(t)$. Taking the derivative with respect to reference group quality (m) provides:

$$\frac{\partial P(L)}{\partial m}=\phi\left\{\frac{\sigma_{\xi}(x)\sigma_{\epsilon}}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\left[\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\left(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)\right)(m-\mu)\right]\right\}$$

$$\frac{\sigma_{\xi}(x)\sigma_{\epsilon}}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\sigma_{\xi}(x)\sigma_{\mathcal{T}}(0)\left(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t)\right)>0.$$

Under the assumption that $\sigma_{\mathcal{T}}(0) > \sigma_{\mathcal{T}}(t)$, as mean reference group increases the probability of layoff increases as well $\left(\frac{\partial P(L)}{\partial m} > 0\right)$.

Layoff dynamics with respect to working spell duration

Again, I use interactions of working spell duration and experience with ability and reference group to explore the evolution of these selection effects over time. I first take the derivative of $P(L)$ and then use this to find and sign the scaled coefficient on these interactions.

The derivative of $P(L)$ with respect to working spell length is given by the following:

$$\frac{\partial P(L)}{\partial t}=\phi\{.\}\left[\frac{\partial\frac{1}{QQ'}}{\partial t}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}[.]+(-\frac{1}{2})(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}}\frac{\partial\sigma_{\psi_{L\varphi}}}{\partial t}\frac{1}{QQ'}[.]+\frac{\partial[.]}{\partial t}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}\right],$$

where $\phi\{.\}$ stands for $\phi\left\{\frac{\sigma_\xi(x)\sigma_\epsilon}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\left[\sigma_\xi(x)\sigma_\tau(0)\left(\sigma_\tau(0)-\sigma_\tau(t)\right)(m-\mu)\right]\right\}$, and $[.]$ stands for $\left[\sigma_\xi(x)\sigma_\tau(0)\left(\sigma_\tau(0)-\sigma_\tau(t)\right)(m-\mu)\right]$. Below, I examine each term separately to make the algebra more manageable.

First term:

$$\frac{\partial\frac{1}{QQ'}}{\partial t}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}[.] = (-1)(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\tau(0)(\sigma_\tau(0)-\sigma_\tau(t))(m-\mu)]$$

$$\frac{\partial\sigma_\tau(t))}{\partial t}\left[(\sigma_\tau(0)\sigma_\xi(x)+\sigma_\xi(x)\sigma_\epsilon+\sigma_\xi(x)\sigma_\epsilon)Q'^{-1}Q^{-2}\right]$$

Turning to the second term, recall that:

$$\sigma_{\psi_{L\varphi}} = Q^{-2}Q'^{-2}(Q'^2\sigma_\tau(0)^2\sigma_\tau(t)+Q^2\sigma_\tau(0)+\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_\tau(t)-\sigma_\tau(0))^2\sigma_\xi(x))+\sigma_\varphi+\sigma_{E\varphi}.$$

$$Q = \sigma_\tau(0)\sigma_\tau(t)\sigma_\xi(x)+\sigma_\tau(t)\sigma_\xi(x)\sigma_\epsilon+\sigma_\tau(0)\sigma_\xi(x)\sigma_\epsilon+\sigma_\tau(t)\sigma_\tau(0)\sigma_\epsilon,$$

$$Q' = \sigma_\tau(0)\sigma_\xi(x)+\sigma_\tau(0)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x).$$

Thus,

$$(-\frac{1}{2})(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}}\frac{\partial\sigma_{\psi_{L\varphi}}}{\partial t}\frac{1}{QQ'}[.] = (-\frac{1}{2})[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\tau(0)(\sigma_\tau(0)-\sigma_\tau(t))(m-\mu)]\frac{1}{QQ'}\sigma_\psi^{\frac{-3}{2}}\frac{\partial\sigma_\tau(t))}{\partial t}$$

$$[(-1)Q^{-3}[2(\sigma_\tau(0)\sigma_\xi(x)+\sigma_\xi(x)\sigma_\epsilon+\sigma_\xi(x)\sigma_\epsilon)\sigma_\tau(0)^2\sigma_\tau(t)-Q\sigma_\tau(0)^2]$$

$$+\frac{-2\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_\tau(t)-\sigma_\tau(0))\sigma_\xi(x)}{Q^2Q'^2}[Q^{-1}(\sigma_\tau(0)\sigma_\xi(x)+\sigma_\xi(x)\sigma_\epsilon+\sigma_\xi(x)\sigma_\epsilon)+(\sigma_\tau(0)-\sigma_\tau(t))]]$$

Third term:

$$\frac{\partial\,[.]}{\partial t}\sigma_{\psi_{L\varphi}}^{\frac{-1}{2}}\frac{1}{QQ'} = (-1)(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}\frac{\partial\sigma_{\mathcal{T}}(t))}{\partial t}\left[\sigma_{\xi}(x)^2\sigma_{\epsilon}\sigma_{\mathcal{T}}(0)(m-\mu)\right].$$

Recombining the terms:

$$\frac{\partial P(L)}{\partial t} = -\phi\{.\}\sigma_{\psi_{L\varphi}}^{\frac{-1}{2}}\frac{\sigma_{\xi}(x)^2\sigma_{\epsilon}\sigma_{\mathcal{T}}(0)(m-\mu)}{QQ'}\frac{\partial\sigma_{\mathcal{T}}(t))}{\partial t}[1+(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))$$

$$[(\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)+\sigma_{\xi}(x)\sigma_{\epsilon}+\sigma_{\mathcal{T}}(0)\sigma_{\epsilon})Q^{-1}]$$

$$+\sigma_{\psi}^{-1}[(-1)Q^{-3}[2(\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)+\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}+\sigma_{\xi}(x)\sigma_{\epsilon})\sigma_{\mathcal{T}}(0)^2\sigma_{\mathcal{T}}(t)-Q\sigma_{\mathcal{T}}(0)^2]$$

$$+\frac{-2\sigma_{\epsilon}^2\sigma_{\mathcal{T}}(0)^2(\sigma_{\mathcal{T}}(t)-\sigma_{\mathcal{T}}(0))\sigma_{\xi}(x)}{Q^2Q'^2}[Q^{-1}(\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)+\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}+\sigma_{\xi}(x)\sigma_{\epsilon})+(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))]]$$

Interaction between working spell duration and ability

The scaled coefficient of the interaction between working spell length and ability (δ_{1L}) on the probability of layoff is given below.

$$\delta_{1L} = \phi\{.\}\sigma_{\psi_{L\varphi}}^{\frac{-1}{2}}\frac{\sigma_{\xi}(x)^2\sigma_{\epsilon}\sigma_{\mathcal{T}}(0)}{QQ'}\frac{\partial\sigma_{\mathcal{T}}(t))}{\partial t}[1+(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))$$

$$[(\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)+\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}+\sigma_{\xi}(x)\sigma_{\epsilon})Q^{-1}]$$

$$+\sigma_{\psi}^{-1}[(-1)Q^{-3}[2(\sigma_{\mathcal{T}}(0)\sigma_{\xi}(x)+\sigma_{\mathcal{T}}(0)\sigma_{\epsilon}+\sigma_{\xi}(x)\sigma_{\epsilon})\sigma_{\mathcal{T}}(0)^2\sigma_{\mathcal{T}}(t)-Q\sigma_{\mathcal{T}}(0)^2]$$

$$+\frac{-2\sigma_{\epsilon}^2\sigma_{\mathcal{T}}(0)^2(\sigma_{\mathcal{T}}(t)-\sigma_{\mathcal{T}}(0))\sigma_{\xi}(x)}{Q^2Q'^2}[Q^{-1}(\sigma_{\mathcal{T}}(0)\sigma_{\mathcal{T}}(0)+\sigma_{\xi}(x)\sigma_{\epsilon}+\sigma_{\xi}(x)\sigma_{\epsilon})+(\sigma_{\mathcal{T}}(0)-\sigma_{\mathcal{T}}(t))]]$$

$\delta_{1L} < 0$ follows from the assumption that $\frac{\partial \sigma_T(t))}{\partial t} < 0$ with the additional sufficient conditional that $\sigma_T(0)\sigma_T(t)\sigma_\xi(x) + \sigma_T(0)\sigma_T(t)\sigma_\epsilon + \sigma_T(t)\sigma_\xi(x)\sigma_\epsilon > \sigma_T(0)\sigma_\xi(x)\sigma_\epsilon$

Interaction between working spell duration and reference group

$$\delta_{2L} = -\phi\{.\}\sigma_{\psi_{L\varphi}}^{\frac{-1}{2}} \frac{\sigma_\xi(x)^2\sigma_\epsilon\sigma_T(0)}{QQ'} \frac{\partial \sigma_T(t))}{\partial t}[1 + (\sigma_T(0) - \sigma_T(t))$$

$$[(\sigma_T(0)\sigma_\xi(x) + \sigma_T(0)\sigma_\epsilon + \sigma_\xi(x)\sigma_\epsilon)Q^{-1}]$$

$$+\sigma_\psi^{-1}[(-1)Q^{-3}[2(\sigma_T(0)\sigma_\xi(x) + \sigma_T(0)\sigma_\epsilon + \sigma_\xi(x)\sigma_\epsilon)\sigma_T(0)^2\sigma_T(t) - Q\sigma_T(0)^2]$$

$$+\frac{-2\sigma_\epsilon^2\sigma_T(0)^2(\sigma_T(t) - \sigma_T(0))\sigma_\xi(x)}{Q^2Q'^2}[Q^{-1}(\sigma_T(0)\sigma_T(0) + \sigma_\xi(x)\sigma_\epsilon + \sigma_\xi(x)\sigma_\epsilon) + (\sigma_T(0) - \sigma_T(t))]]$$

$\delta_{2L} > 0$ follows from the same assumptions.

Layoff dynamics with respect to experience

I take the derivative of $P(L)$ with respect to experience in order to find the scaled coefficient on the interaction between experience and and ability and reference group for the probability of layoff.

The derivative of $P(L)$ with respect to experience is given by the following:

$$\frac{\partial P(L)}{\partial x} = \phi\{.\}\left[\frac{\partial\,[.]}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'} + \frac{\partial \frac{1}{QQ'}}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}[.] + (-\frac{1}{2})(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}}\frac{\partial \sigma_{\psi_{L\varphi}}}{\partial x}\frac{1}{QQ'}[.]\right],$$

where $\phi\{.\}$ stands for $\phi\left\{\frac{\sigma_\xi(x)\sigma_\epsilon}{Q'Q\sqrt{\sigma_{\psi_{L\varphi}}}}\left[\sigma_\xi(x)\sigma_T(0)\left(\sigma_T(0) - \sigma_T(t)\right)(m-\mu)\right]\right\}$ (and is positive by definition of normal PDF), and [.] stands for $\left[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\mathcal{T}(0)\left(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)\right)(m-\mu)\right]$ (and under the asumption $\sigma_\mathcal{T}(0) > \sigma_\mathcal{T}(t)$, is negative if $\mu > m$). The following takes each component separately and resolves any conflicting signs.

$\frac{\partial[.]}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}$

$$\frac{\partial\,[.]}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'} = (\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}\frac{\partial\sigma_\xi(x)}{\partial x}\left[2\sigma_\xi(x)\sigma_\epsilon\sigma_\mathcal{T}(0)\left(\sigma_\mathcal{T}(0)-\sigma_\mathcal{T}(t)\right)(m-\mu)\right]$$

$\frac{\partial\frac{1}{QQ'}}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\,[.]$

Recall $Q = \sigma_\mathcal{T}(0)\sigma_\mathcal{T}(t)\sigma_\xi(x)+\sigma_\mathcal{T}(t)\sigma_\xi(x)\sigma_\epsilon+\sigma_\mathcal{T}(0)\sigma_\xi(x)\sigma_\epsilon+\sigma_\mathcal{T}(t)\sigma_\mathcal{T}(0)\sigma_\epsilon$and $Q' = \sigma_\mathcal{T}(0)\sigma_\xi(x)+\sigma_\mathcal{T}(0)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x)$

$$\frac{\partial\frac{1}{QQ'}}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\,[.] = (-1)\frac{1}{QQ'}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}$$

$$[.]\,\frac{\partial\sigma_\xi(x)}{\partial x}\left[Q^{-1}(\sigma_\mathcal{T}(0)\sigma_\mathcal{T}(t)+\sigma_\mathcal{T}(t)\sigma_\epsilon+\sigma_\mathcal{T}(0)\sigma_\epsilon)+Q'^{-1}(\sigma_\mathcal{T}(0)+2\sigma_\epsilon)\right]$$

Resolving conflict between 8.7.2 and 8.7.1

$$\frac{\partial\,[.]}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}+\frac{\partial\frac{1}{QQ'}}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\,[.] = (\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}\frac{\partial\sigma_\xi(x)}{\partial x}\,[.]$$

$$\left[2-\sigma_\xi(x)\left(Q^{-1}(\sigma_\mathcal{T}(0)\sigma_\mathcal{T}(t)+\sigma_\mathcal{T}(t)\sigma_\epsilon+\sigma_\mathcal{T}(0)\sigma_\epsilon)+Q'^{-1}(\sigma_\mathcal{T}(0)+2\sigma_\epsilon)\right)\right]$$

$$= (\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}\frac{\partial\sigma_\xi(x)}{\partial x}\,[.]\,\sigma_\xi(x)^{-1}$$
$$[Q^{-1}(Q-\sigma_\xi(x)(\sigma_\mathcal{T}(0)\sigma_\mathcal{T}(t)+\sigma_\mathcal{T}(t)\sigma_\epsilon+\sigma_\mathcal{T}(0)\sigma_\epsilon)+Q'^{-1}(Q'-\sigma_\xi(x)((\sigma_\mathcal{T}(0)+2\sigma_\epsilon))]$$

$$= (\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}} \frac{1}{QQ'} \frac{\partial \sigma_{\xi}(x)}{\partial x} [.] \, \sigma_{\xi}(x)^{-1}[Q^{-1}\sigma_{\tau}(t)\sigma_{\tau}(0)\sigma_{\epsilon} + Q'^{-1}\sigma_{\tau}(0)\sigma_{\epsilon}]$$

$(-\frac{1}{2})(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}} \frac{\partial \sigma_{\psi_{L\varphi}}}{\partial x} \frac{1}{QQ'} [.]$ is ambiguous

Recall

$$\sigma_{\psi_{L\varphi}} = Q^{-2}Q'^{-2}(Q'^{2}\sigma_{\tau}(0)^{2}\sigma_{t}(t) + (\sigma_{t}(t)\sigma_{\xi}(x) + \sigma_{t}(t)\sigma_{\epsilon} + 2\sigma_{\epsilon}\sigma_{\xi}(x))^{2}\sigma_{\tau}(0)^{3}$$
$$+\sigma_{\epsilon}^{2}\sigma_{\tau}(0)^{2}(\sigma_{t}(t) - \sigma_{\tau}(0))^{2}\sigma_{\xi}(x)) + \sigma_{\varphi} + \sigma_{E\varphi}, \text{ and}$$

$Q = \sigma_{\tau}(0)\sigma_{t}(t)\sigma_{\xi}(x) + \sigma_{t}(t)\sigma_{\xi}(x)\sigma_{\epsilon} + \sigma_{\tau}(0)\sigma_{\xi}(x)\sigma_{\epsilon} + \sigma_{t}(t)\sigma_{\tau}(0)\sigma_{\epsilon}$and $Q' = \sigma_{\tau}(0)\sigma_{\xi}(x) + \sigma_{\tau}(0)\sigma_{\epsilon} + 2\sigma_{\epsilon}\sigma_{\xi}(x)$

$$(-\tfrac{1}{2})(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}} \frac{\partial \sigma_{\psi_{L\varphi}}}{\partial x} \frac{1}{QQ'} [.] = (-\tfrac{1}{2})[\sigma_{\xi}(x)^{2}\sigma_{\epsilon}\sigma_{\tau}(0)(\sigma_{t}(0) - \sigma_{t}(t))(m-\mu)] \left(\frac{1}{QQ'}\right)^{-3} \sigma_{\psi}^{\frac{-3}{2}}$$
$$[2\frac{\partial \frac{1}{QQ'}}{\partial x}(Q'^{2}\sigma_{\tau}(0)^{2}\sigma_{t}(t) + (\sigma_{t}(t)\sigma_{\xi}(x) + \sigma_{t}(t)\sigma_{\epsilon} + 2\sigma_{\epsilon}\sigma_{\xi}(x))^{2}\sigma_{\tau}(0)^{3} + \sigma_{\epsilon}^{2}\sigma_{\tau}(0)^{2}(\sigma_{t}(t) - \sigma_{\tau}(0))^{2}\sigma_{\xi}(x))$$
$$+(2Q'(\sigma_{\tau}(0) + 2\sigma_{\epsilon})\sigma_{\tau}(0)^{2}\sigma_{t}(t) + 2(\sigma_{t}(t) + 2\sigma_{\epsilon})(\sigma_{t}(t)\sigma_{\xi}(x) + \sigma_{t}(t)\sigma_{\epsilon} + 2\sigma_{\epsilon}\sigma_{\xi}(x))\sigma_{\tau}(0)^{3}$$
$$+\sigma_{\epsilon}^{2}\sigma_{\tau}(0)^{2}(\sigma_{t}(t) - \sigma_{\tau}(0))^{2})]$$

$$= (-\tfrac{1}{2})[\sigma_{\xi}(x)^{2}\sigma_{\epsilon}\sigma_{\tau}(0)(\sigma_{t}(0) - \sigma_{t}(t))(m-\mu)] \left(\frac{1}{QQ'}\right)^{-4} \sigma_{\psi}^{\frac{-3}{2}} \frac{\partial \sigma_{\xi}(x)}{\partial x}$$
$$[-2\left[Q'(\sigma_{\tau}(0)\sigma_{t}(t) + \sigma_{t}(t)\sigma_{\epsilon} + \sigma_{\tau}(0)\sigma_{\epsilon}) + Q(\sigma_{\tau}(0) + 2\sigma_{\epsilon})\right]$$
$$[Q'^{2}\sigma_{\tau}(0)^{2}\sigma_{t}(t) + (\sigma_{t}(t)\sigma_{\xi}(x) + \sigma_{t}(t)\sigma_{\epsilon} + 2\sigma_{\epsilon}\sigma_{\xi}(x))^{2}\sigma_{\tau}(0)^{3} + \sigma_{\epsilon}^{2}\sigma_{\tau}(0)^{2}(\sigma_{t}(t) - \sigma_{\tau}(0))^{2}\sigma_{\xi}(x))]$$
$$+QQ'(2Q'(\sigma_{\tau}(0) + 2\sigma_{\epsilon})\sigma_{\tau}(0)^{2}\sigma_{t}(t) + 2(\sigma_{t}(t) + 2\sigma_{\epsilon})(\sigma_{t}(t)\sigma_{\xi}(x) + \sigma_{t}(t)\sigma_{\epsilon} + 2\sigma_{\epsilon}\sigma_{\xi}(x))\sigma_{\tau}(0)^{3}$$
$$+\sigma_{\epsilon}^{2}\sigma_{\tau}(0)^{2}(\sigma_{t}(t) - \sigma_{\tau}(0))^{2})]]$$

$$= (-\tfrac{1}{2})[\sigma_{\xi}(x)^{2}\sigma_{\epsilon}\sigma_{\tau}(0)(\sigma_{t}(0) - \sigma_{t}(t))(m-\mu)] \left(\frac{1}{QQ'}\right)^{-4} \sigma_{\psi}^{\frac{-3}{2}} \frac{\partial \sigma_{\xi}(x)}{\partial x}$$
$$[-2\left[Q'^{3}(\sigma_{\tau}(0)\sigma_{t}(t) + \sigma_{t}(t)\sigma_{\epsilon} + \sigma_{\tau}(0)\sigma_{\epsilon})\sigma_{\tau}(0)^{2}\sigma_{t}(t)\right]$$

$-2\left[Q'(\sigma_\tau(0)\sigma_t(t)+\sigma_t(t)\sigma_\epsilon+\sigma_\tau(0)\sigma_\epsilon)+Q(\sigma_\tau(0)+2\sigma_\epsilon)\right]$

$\left[(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))^2\sigma_\tau(0)^3+\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2\sigma_\xi(x))\right]$

$+QQ'(2(\sigma_t(t)+2\sigma_\epsilon)(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))\sigma_\tau(0)^3+\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2)]]$

$=(-\frac{1}{2})[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\tau(0)(\sigma_t(0)-\sigma_t(t))(m-\mu)]\left(\frac{1}{QQ'}\right)^{-4}\sigma_\psi^{\frac{-3}{2}}\frac{\partial\sigma_\xi(x)}{\partial x}$

$[-2\left[Q'^3(\sigma_\tau(0)\sigma_t(t)+\sigma_t(t)\sigma_\epsilon+\sigma_\tau(0)\sigma_\epsilon)\sigma_\tau(0)^2\sigma_t(t)\right]$

$+2\sigma_\tau(0)^3(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))$

$[QQ'(\sigma_t(t)+2\sigma_\epsilon)-\left[Q'(\sigma_\tau(0)\sigma_t(t)+\sigma_t(t)\sigma_\epsilon+\sigma_\tau(0)\sigma_\epsilon)+Q(\sigma_\tau(0)+2\sigma_\epsilon)\right]$

$(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))]$

$+\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2[QQ'$

$-2\sigma_\xi(x)[Q'(\sigma_\tau(0)\sigma_t(t)+\sigma_t(t)\sigma_\epsilon+\sigma_\tau(0)\sigma_\epsilon)+Q(\sigma_\tau(0)+2\sigma_\epsilon)]]]$

$=(-\frac{1}{2})[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\tau(0)(\sigma_t(0)-\sigma_t(t))(m-\mu)]\left(\frac{1}{QQ'}\right)^{-4}\sigma_\psi^{\frac{-3}{2}}\frac{\partial\sigma_\xi(x)}{\partial x}$

$[-2\left[Q'^3(\sigma_\tau(0)\sigma_t(t)+\sigma_t(t)\sigma_\epsilon+\sigma_\tau(0)\sigma_\epsilon)\sigma_\tau(0)^2\sigma_t(t)\right]$

$-2\sigma_\tau(0)^3(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))$

$[-\sigma_t(t)\sigma_\epsilon^2(\sigma_\tau(0)\sigma_\xi(x)\sigma_t(t)+\sigma_\epsilon\sigma_\tau(0)\sigma_t(t)+2\sigma_\epsilon\sigma_\xi(x)\sigma_t(t)+\sigma_\tau(0)^2\sigma_\xi(x)+\sigma_\epsilon\sigma_\xi(x)\sigma_\tau(0))$

$+Q(\sigma_\tau(0)+2\sigma_\epsilon)(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))]$

$+\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2[\sigma_\tau(0)^2\sigma_\epsilon^2\sigma_t(t)-(Q'\sigma_\tau(0)\sigma_\epsilon$

$+\sigma_\tau(0)\sigma_\xi(x)(\sigma_\tau(0)\sigma_t(t)+\sigma_t(t)\sigma_\epsilon+\sigma_\tau(0)\sigma_\epsilon)+2\sigma_\xi(x)Q(\sigma_\tau(0)+2\sigma_\epsilon))]]$

From Appendix A above, $(-\frac{1}{2})(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}}\frac{\partial\sigma_{\psi_{L\varphi}}}{\partial x}\frac{1}{QQ'}[.]$ ambiguous.

This term producing the ambiguity I define as B.

$B\equiv(-\frac{1}{2})[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\tau(0)(\sigma_t(0)-\sigma_t(t))(m-\mu)]\left(\frac{1}{QQ'}\right)^{-4}\sigma_\psi^{\frac{-3}{2}}\frac{\partial\sigma_\xi(x)}{\partial x}$

$\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2\sigma_\tau(0)^2\sigma_\epsilon^2\sigma_t(t)$

Resolving ambiguity in 8.7.4

From Appendix A,

$$\frac{\partial[.]}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}+\frac{\partial\frac{1}{QQ'}}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}[.]=$$
$$(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}\frac{\partial\sigma_{\xi}(x)}{\partial x}[.]\,\sigma_{\xi}(x)^{-1}[Q^{-1}\sigma_t(t)\sigma_\tau(0)\sigma_\epsilon+Q'^{-1}\sigma_\tau(0)\sigma_\epsilon].$$

Thus, $\frac{\partial[.]}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}+\frac{\partial\frac{1}{QQ'}}{\partial x}(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}[.]+B=$

$$(\sigma_{\psi_{L\varphi}})^{\frac{-1}{2}}\frac{1}{QQ'}\frac{\partial\sigma_{\xi}(x)}{\partial x}[.]\,\{\sigma_{\xi}(x)^{-1}[Q^{-1}\sigma_t(t)\sigma_\tau(0)\sigma_\epsilon+Q'^{-1}\sigma_\tau(0)\sigma_\epsilon]$$
$$+(-\tfrac{1}{2})[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\tau(0)(\sigma_t(0)-\sigma_t(t))(m-\mu)]\left(\frac{1}{QQ'}\right)^{-4}\sigma_{\psi}^{\frac{-3}{2}}$$
$$\frac{\partial\sigma_{\xi}(x)}{\partial x}\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2\sigma_\tau(0)^2\sigma_\epsilon^2\sigma_t(t)$$

$$=\sigma_{\psi_{L\varphi}}^{\frac{-3}{2}}\frac{1}{QQ'}\frac{\partial\sigma_{\xi}(x)}{\partial x}[.]\,\{\sigma_{\xi}(x)^{-1}[Q^{-1}\sigma_t(t)\sigma_\tau(0)\sigma_\epsilon+Q'^{-1}\sigma_\tau(0)\sigma_\epsilon]$$
$$+(-\tfrac{1}{2})[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\tau(0)(\sigma_t(0)-\sigma_t(t))(m-\mu)]\left(\frac{1}{QQ'}\right)^{-4}\sigma_{\psi}^{\frac{-3}{2}}\frac{\partial\sigma_{\xi}(x)}{\partial x}$$
$$\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2\sigma_\tau(0)^2\sigma_\epsilon^2\sigma_t(t)$$

$$=(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}}\frac{1}{Q^4Q'^4}\frac{\partial\sigma_{\xi}(x)}{\partial x}[.]\,\{[2(Q'^2\sigma_\tau(0)^2\sigma_t(t)+(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))^2\sigma_\tau(0)^3$$
$$+\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2\sigma_\xi(x))+\sigma_\varphi+\sigma_{E\varphi})$$
$$\sigma_\xi^{-1}(Q'\sigma_t(t)\sigma_\tau(0)\sigma_\epsilon+Q\sigma_\tau(0)\sigma_\epsilon]-\left[\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2\sigma_\tau(0)^2\sigma_\epsilon^2\sigma_t(t)\right]$$

$$=(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}}\frac{1}{Q^4Q'^4}\frac{\partial\sigma_{\xi}(x)}{\partial x}\left[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\tau(0)\left(\sigma_\tau(0)-\sigma_t(t)\right)(m-\mu)\right]\{$$
$$[2\left(Q'^2\sigma_\tau(0)^2\sigma_t(t)+(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))^2\sigma_\tau(0)^3\right)+\sigma_\varphi+\sigma_{E\varphi})$$
$$+\sigma_\epsilon^2\sigma_\tau(0)^2(\sigma_t(t)-\sigma_\tau(0))^2\left[Q'\sigma_t(t)\sigma_\tau(0)\sigma_\epsilon+Q\sigma_\tau(0)\sigma_\epsilon-\sigma_\tau(0)^2\sigma_\epsilon^2\sigma_t(t)\right]$$

$$=(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}}\frac{1}{Q^4Q'^4}\frac{\partial\sigma_{\xi}(x)}{\partial x}\left[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\tau(0)\left(\sigma_\tau(0)-\sigma_t(t)\right)(m-\mu)\right]\{$$
$$[2\left(Q'^2\sigma_\tau(0)^2\sigma_t(t)+(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))^2\sigma_\tau(0)^3\right)+\sigma_\varphi+\sigma_{E\varphi})$$

$$+\sigma_\epsilon^2\sigma_\mathcal{T}(0)^2(\sigma_t(t)-\sigma_\mathcal{T}(0))^2\left[\sigma_t(t)\sigma_\mathcal{T}(0)\sigma_\epsilon(\sigma_\mathcal{T}(0)\sigma_\xi(x)+2\sigma_\epsilon\sigma_\xi(x))+Q\sigma_\mathcal{T}(0)\sigma_\epsilon\right]\}$$

Interaction between experience and ability

I take the derivative of the $\frac{\partial\{.\}}{\partial x}$ with respect to ability (μ) to find the predicted evolution of selection into mobility with increases in experience. The analytical scaled coefficient γ_{1L} is given below.

$$\gamma_{1L} = -\phi\{.\}(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}}\frac{1}{Q^4Q'^4}\frac{\partial\sigma_\xi(x)}{\partial x}\left[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\mathcal{T}(0)\left(\sigma_\mathcal{T}(0)-\sigma_t(t)\right)\right]\{$$
$$[2\left(Q'^2\sigma_\mathcal{T}(0)^2\sigma_t(t)+(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))^2\sigma_\mathcal{T}(0)^3\right)+\sigma_\varphi+\sigma_{E\varphi})$$
$$+\sigma_\epsilon^2\sigma_\mathcal{T}(0)^2(\sigma_t(t)-\sigma_\mathcal{T}(0))^2\left[\sigma_t(t)\sigma_\mathcal{T}(0)\sigma_\epsilon(\sigma_\mathcal{T}(0)\sigma_\xi(x)+2\sigma_\epsilon\sigma_\xi(x))+Q\sigma_\mathcal{T}(0)\sigma_\epsilon\right]\}$$

Under the assumptions that $\frac{\partial\sigma_\xi(x)}{\partial x} < 0$ and $\sigma_\mathcal{T}(t) < \sigma_\mathcal{T}(0)$, $\gamma_{1L} > 0$. This means that the negative selection on the basis of ability of mobile workers decreases with increases in experience.

Interaction between experience and reference group

I take the derivative of the $\frac{\partial\{.\}}{\partial x}$ with respect to reference group (m) to find the predicted evolution of selection into mobility with increases in experience. The analytical scaled coefficient γ_{2L} is given below.

$$\gamma_L = \phi\{.\}(\sigma_{\psi_{L\varphi}})^{\frac{-3}{2}}\frac{1}{Q^4Q'^4}\frac{\partial\sigma_\xi(x)}{\partial x}\left[\sigma_\xi(x)^2\sigma_\epsilon\sigma_\mathcal{T}(0)\left(\sigma_\mathcal{T}(0)-\sigma_t(t)\right)\right]\{$$
$$[2\left(Q'^2\sigma_\mathcal{T}(0)^2\sigma_t(t)+(\sigma_t(t)\sigma_\xi(x)+\sigma_t(t)\sigma_\epsilon+2\sigma_\epsilon\sigma_\xi(x))^2\sigma_\mathcal{T}(0)^3\right)+\sigma_\varphi+\sigma_{E\varphi})$$
$$+\sigma_\epsilon^2\sigma_\mathcal{T}(0)^2(\sigma_t(t)-\sigma_\mathcal{T}(0))^2\left[\sigma_t(t)\sigma_\mathcal{T}(0)\sigma_\epsilon(\sigma_\mathcal{T}(0)\sigma_\xi(x)+2\sigma_\epsilon\sigma_\xi(x))+Q\sigma_\mathcal{T}(0)\sigma_\epsilon\right]\}$$

Under the same assumptions as above, $\gamma_{2L} < 0$. This means that the positive selection on the basis of ability of mobile workers becomes more negative with increases in experience.

Appendix I

Additions for Chapter 3

Consistency of Estimators

A.1 Random-effects estimator

Assume that $p\lim \widehat{\boldsymbol{\Sigma}}_j = \boldsymbol{\Omega}_j$, where $\boldsymbol{\Omega}_j$ is of full rank. A law of large numbers gives (see Equation 10), $p\lim \widehat{\boldsymbol{\delta}}_{\text{RE}} = \boldsymbol{\delta} + \mathbf{A}_j^{-1}\text{E}[\mathbf{V}_j'\boldsymbol{\Omega}_j^{-1}(\mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j)]$, where $\mathbf{A}_j \equiv \text{E}(\mathbf{V}_j'\boldsymbol{\Omega}_j^{-1}\mathbf{V}_j)$ is assumed to be of full rank. Consistency therefore requires that $\text{E}[\mathbf{V}_j'\boldsymbol{\Omega}_j^{-1}(\mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j)] = \mathbf{0}$. Under exogeneity, $\text{E}[\mathbf{V}_j'\boldsymbol{\Omega}_j^{-1}(\mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j)|\mathbf{V}_j] = \mathbf{V}_j'\boldsymbol{\Omega}_j^{-1}[\mathbf{Z}_j\text{E}(\mathbf{u}_j|\mathbf{V}_j) + \text{E}(\boldsymbol{\epsilon}_j|\mathbf{V}_j)] = \mathbf{0}$, because $\text{E}(\mathbf{u}_j|\mathbf{V}_j) = \mathbf{0}$ and $\text{E}(\boldsymbol{\epsilon}_j|\mathbf{V}_j) = \mathbf{0}$ from Equations 7 and 6, respectively. Since $\text{E}[\mathbf{V}_j'\boldsymbol{\Omega}_j^{-1}(\mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j)|\mathbf{V}_j] = \mathbf{0}$ for all $\mathbf{V}_j$, it follows that $\text{E}[\mathbf{V}_j'\boldsymbol{\Omega}_j^{-1}(\mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j)] = \mathbf{0}$, so $p\lim \widehat{\boldsymbol{\delta}}_{\text{RE}} = \boldsymbol{\delta}$.

The consistency result also holds for ML and REML because for both, the estimates of the regression coefficients can be obtained by substituting the corresponding covariance matrix estimate into the FGLS estimator. Consistency of $\widehat{\boldsymbol{\delta}}_{\text{RE}}$ does not require that $\boldsymbol{\Omega}_j = \boldsymbol{\Sigma}_j$, so the estimator is consistent even if the covariance structure is misspecified.

A.2 Fixed-effects estimator

From Equation 11 and a law of large numbers, $p\lim \widehat{\boldsymbol{\beta}}_{\text{FE}} = \boldsymbol{\beta} + \mathbf{A}_j^{-1}\text{E}[\ddot{\mathbf{X}}_j'(\ddot{\mathbf{Z}}_j\mathbf{u}_j + \ddot{\boldsymbol{\epsilon}}_j)]$, where $\mathbf{A}_j \equiv \text{E}(\ddot{\mathbf{X}}_j'\ddot{\mathbf{X}}_j)$ is assumed to be of full rank. The requirement for consistency is that $\text{E}[\ddot{\mathbf{X}}_j'(\ddot{\mathbf{Z}}_j\mathbf{u}_j + \ddot{\boldsymbol{\epsilon}}_j)] = \mathbf{0}$. From unit-level exogeneity (Equation 6), $\text{E}(\ddot{\mathbf{X}}_j'\ddot{\boldsymbol{\epsilon}}_j) = \mathbf{0}$, so the remaining requirement is $\text{E}(\ddot{\mathbf{X}}_j'\ddot{\mathbf{Z}}_j\mathbf{u}_j) = \mathbf{0}$.

Augmented fixed-effects estimator

From Equation 14 and a law of large numbers, $p\lim\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma} + \mathbf{A}_j^{-1}\mathrm{E}[\mathbf{W}_j'(\mathbf{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j)]$, because $p\lim\widehat{\boldsymbol{\beta}}_{\mathrm{FE}} = \boldsymbol{\beta}$ under the uncorrelated variance assumption in Equation 12. Here $\mathbf{A}_j \equiv \mathrm{E}\left(\mathbf{W}_j'\mathbf{W}_j\right)$ is assumed to be of full rank so that the required exogeneity assumption for $p\lim\widehat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$ becomes $\mathrm{E}[\mathbf{W}_j'(\mathbf{Z}_j\mathbf{u}_j+\boldsymbol{\epsilon}_j)] = \mathbf{0}$. It follows from unit-level exogeneity (Equation 6) that $\mathrm{E}(\mathbf{W}_j'\boldsymbol{\epsilon}_j) = \mathbf{0}$, so the remaining requirement is that $\mathrm{E}(\mathbf{W}_j'\mathbf{Z}_j\mathbf{u}_j) = \mathbf{0}$. A sufficient condition for this remaining requirement is that $\mathrm{E}(\mathbf{u}_j|\mathbf{W}_j) = \mathbf{0}$, the assumption that cluster-level covariates are cluster-level exogenous that is stated in Equation 8.

Per-cluster regression estimation

For Step 1, $p\lim\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}} = \boldsymbol{\beta}_3$ Verbeke et al. (2001). For Step 2, from Equation 18 and a law of large numbers, $p\lim\breve{\boldsymbol{\eta}}_j = \boldsymbol{\eta}_j + \mathbf{A}_j^{-1}\mathrm{E}\left\{\mathbf{z}_{ij}[\mathbf{x}_{3ij}'(\boldsymbol{\beta}_3 - \widehat{\boldsymbol{\beta}}_{3\mathrm{CML}}) + \mathbf{z}_{ij}'\boldsymbol{\eta}_j + \epsilon_{ij}]\right\}$. Assume that $\mathbf{A}_j \equiv \mathrm{E}(\mathbf{z}_{ij}\mathbf{z}_{ij}')$ is of full rank. $\mathrm{E}[\mathbf{z}_{ij}\mathbf{x}_{3ij}'(\boldsymbol{\beta}_3 - \widehat{\boldsymbol{\beta}}_{3\mathrm{CML}})] = \mathbf{0}$ since $p\lim\widehat{\boldsymbol{\beta}}_{3\mathrm{CML}} = \boldsymbol{\beta}_3$ and $\mathrm{E}(\mathbf{z}_{ij}\epsilon_{ij}) = \mathbf{0}$ under unit-level exogeneity. The remaining requirement for consistency is that $\mathrm{E}(\mathbf{z}_{ij}\mathbf{z}_{ij}'\boldsymbol{\eta}_j) = \mathbf{0}$. For Step 3, $p\lim\widehat{\boldsymbol{\alpha}}_r = \boldsymbol{\alpha}_r + \mathbf{A}_j^{-1}\mathrm{E}\left[\mathbf{w}_{rj}(u_{rj} + \breve{\eta}_{rj} - \eta_{rj})\right]$, from Equation 19 and a law of large numbers. Assume that $\mathbf{A}_j \equiv \mathrm{E}\left(\mathbf{w}_{rj}\mathbf{w}_{rj}'\right)$ is of full rank. It follows from cluster-level exogeneity of the cluster-level covariates (Equation 8) that $\mathrm{E}[\mathbf{w}_{rj}u_{rj}] = \mathbf{0}$. Moreover, $\mathrm{E}(\mathbf{w}_{rj}(\breve{\eta}_{rj} - \eta_{rj})) = \mathbf{0}$ if the estimation errors are uncorrelated with $\mathbf{w}_{rj}$ or if $n_j \to \infty$.

Stata Code for HSB Example

Below we provide the Stata 13 StataCorp (2013) code we used to produce the results reported in Table D.1 for the High School and Beyond (HSB) dataset, distributed with the HLM software Raudenbush et al. (2011). The dataset, `hsb.dta`, can also be downloaded from the website for Rabe-Hesketh and Skrondal (2012):

`http://www.stata-press.com/data/mlmus3.html`.

```
*******************************************************************************
*       Initial Data Setup                                                    *
*******************************************************************************
use http://www.stata-press.com/data/mlmus3/hsb, clear
keep schoolid mathach sector ses
describe

/* The variables of interest are:
sector (wj) = cluster-level indicator for whether school is Catholic
ses (xij) = unit-level covariate that indexes students' socioeconomic status
mathach (yij) = outcome, students' performance on a math test
schoolid = cluster (school) identifier */

*Define cluster identifier
xtset schoolid

*Generate cross-level interaction term between ses and sector (xij*wj)
generate sesXsector = ses*sector
```

```
********************************************************************************
*       Random Effects (REML)                                                  *
********************************************************************************
/* Estimation is done in one step using covariates that are unit-level,
cluster-level, and interactions between the two. Note: ses is the covariate
with a random slope */

mixed mathach ses sector sesXsector || schoolid: ses, ///
   covariance(unstructured) reml


********************************************************************************
*       Augmented Fixed Effects (FE+)                                          *
********************************************************************************
***STEP 1 - FE
**Estimate coefficients of unit-level covariates using standard fixed effects
xtreg mathach ses sesXsector, fe

****STEP 2 - Regress quasi-residuals on cluster-level covariate
**Generate "newy" as residuals from the first stage regression
generate ynew = mathach - _b[ses]*ses - _b[sesXsector]*sesXsector

**Regress the residuals on the cluster-level covariate, with cluster-robust SEs
```

```
regress ynew sector, vce(cluster schoolid)


********************************************************************************
*       Per-Cluster Regression (PC)                                           *
********************************************************************************
***Step 1
**Not needed because there are no unit-level covariates that do not have
**random slopes (R3=0)


***Step 2
**For each cluster, regress outcome on unit-level covariate using OLS, saving
**estimates of the intercepts (a1) and coefficients (a2) in statsby_HSB.dta
statsby a1=_b[_cons] a2=_b[ses] , by(schoolid) saving(statsby_HSB, replace): ///
   regress mathach ses


**Merge estimates into dataset (after sorting data according to schoolid)
sort schoolid
merge m:1 schoolid using statsby_HSB


***Step 3
/* Part a: Regress intercept estimates (a1) on cluster-level covariate,
using 1 observation per cluster - OLS with robust SEs */


**Create indicator for 1 observation per cluster (it doesn't matter which one)
```

```
egen pickone = tag(schoolid)

**OLS for 1 observation per cluster, with robust SEs
regress a1 sector if pickone==1, vce(robust)

*=>estimated intercept (_cons) = estimated intercept of model (gamma0)
*=>estimated coefficient of sector = estimate coefficient of sector (gamma1)

/* Part b: Regress coefficient estimates (a2) on the cluster-level covariates,
using 1 observation per cluster - OLS with robust SEs */

regress a2 sector if pickone==1, vce(robust)

*=>estimated intercept (_cons) = estimated coefficient of ses (beta1)
*=>estimated coefficient of sector = estimated interaction parameter (beta2)
```

Supplemental Tables and Figures

Table I.1: Comparing methods for estimating the coefficient β_1 of x_{ij} ($\beta_1 = 1$).

Simulation Condition	Method	100× Bias	100× RMSE	100× Mean SE	100× SD	Mean SE / SD
Small Clusters	RE	16.6*	21.6	12.9	13.8	0.93
& Uncorrelated	FE+	0.6	16.2	14.9	16.2	0.92
Variance	PC	1.9	25.5	24.7	25.5	0.97
Small Clusters &	RE	21.3*	24.2	12.1	11.4	1.06
& Correlated	FE+	11.7*	19.1	14.1	15.2	0.93
Variance	PC	-1.8	26.7	26.1	26.6	0.98
Large Clusters &	RE	6.2*	10.0	7.5	7.8	0.96
& Uncorrelated	FE+	-0.3	8.0	6.0	8.0	0.74
Variance	PC	-0.2	8.0	7.9	8.0	0.99
Large Clusters &	RE	12.6*	14.4	7.0	7.0	1.00
& Correlated	FE+	12.8*	15.2	5.5	8.2	0.67
Variance	PC	0.7	7.7	8.1	7.7	1.05

Note. Small Clusters: $n_j = n = 4$, Large Clusters: $n_j = n = 20$;
Uncorrelated Variance: $\sigma_j^2 = 1$, Correlated Variance: $\sigma_j = \exp(u_{1j})$.
RMSE=root-mean-square error;
Mean SE=mean of the standard error estimates over the replications;
Mean SE=mean of the standard error estimates over the replications;
SD=standard deviation of the coefficient estimates over the replications;
*Estimated bias differs significantly from 0 at the 0.05 level.

Table I.2: Comparing methods for estimating the coefficient β_2 of $x_{ij} \times w_j$ ($\beta_2 = 2$)

Simulation Condition	Method	100× Bias	100× RMSE	100× Mean SE	100× SD	$\frac{\text{Mean SE}}{\text{SD}}$
Small Clusters	RE	0.8*	7.0	6.5	7.0	0.94
& Uncorrelated	FE+	0.0	8.2	7.6	8.2	0.93
Variance	PC	-0.6	13.3	12.6	13.3	0.95
Small Clusters	RE	1.2*	6.1	6.1	5.9	1.03
& Correlated	FE+	0.2	7.9	7.2	7.9	0.91
Variance	PC	0.7	13.4	13.3	13.4	0.99
Large Clusters	RE	0.2	3.9	3.8	3.9	0.97
& Uncorrelated	FE+	0.2	4.0	3.0	4.0	0.75
Variance	PC	0.1	4.1	4.0	4.1	0.99
Large Clusters	RE	-0.1	3.5	3.6	3.5	1.01
& Correlated	FE+	-0.3	4.2	2.8	4.2	0.68
Variance	PC	-0.3	4.0	4.1	4.0	1.04

Note. Small Clusters: $n_j = n = 4$, Large Clusters: $n_j = n = 20$;
Uncorrelated Variance: $\sigma_j^2 = 1$, Correlated Variance: $\sigma_j = \exp(u_{1j})$.

RMSE=root-mean-square error;

Mean SE=mean of the standard error estimates over the replications;

SD=standard deviation of the coefficient estimates over the replications;

*Estimated bias differs significantly from 0 at the 0.05 level.

Table I.3: Comparing methods for estimating the coefficient γ_1 of w_j ($\gamma_1 = 3$)

Simulation Condition	Method	100× Bias	100× RMSE	100× Mean SE	100× SD	Mean SE / SD
Small Clusters	RE	-4.5*	8.0	6.6	6.6	1.01
& Uncorrelated	FE+	-0.3	7.4	7.1	7.4	0.97
Variance	PC	-0.5	12.7	11.7	12.7	0.92
Small Clusters	RE	-5.3*	8.2	6.5	6.2	1.04
& Correlated	FE+	-2.4*	7.3	6.8	6.9	0.98
Variance	PC	0.2	12.2	11.2	12.2	0.92
Large Clusters	RE	-1.7*	4.9	4.5	4.6	0.99
& Uncorrelated	FE+	0.2	5.3	5.4	5.3	1.03
Variance	PC	0.0	5.2	5.2	5.2	0.99
Large Clusters	RE	-2.9*	5.2	4.3	4.3	1.00
& Correlated	FE+	-2.5*	5.2	4.9	4.6	1.05
Variance	PC	0.0	5.0	5.0	5.0	1.00

Note. Small Clusters: $n_j = n = 4$, Large Clusters: $n_j = n = 20$;
Uncorrelated Variance: $\sigma_j^2 = 1$, Correlated Variance: $\sigma_j = \exp(u_{1j})$.

RMSE=root-mean-square error;

Mean SE=mean of the standard error estimates over the replications;

SD=standard deviation of the coefficient estimates over the replications;

*Estimated bias differs significantly from 0 at the 0.05 level.

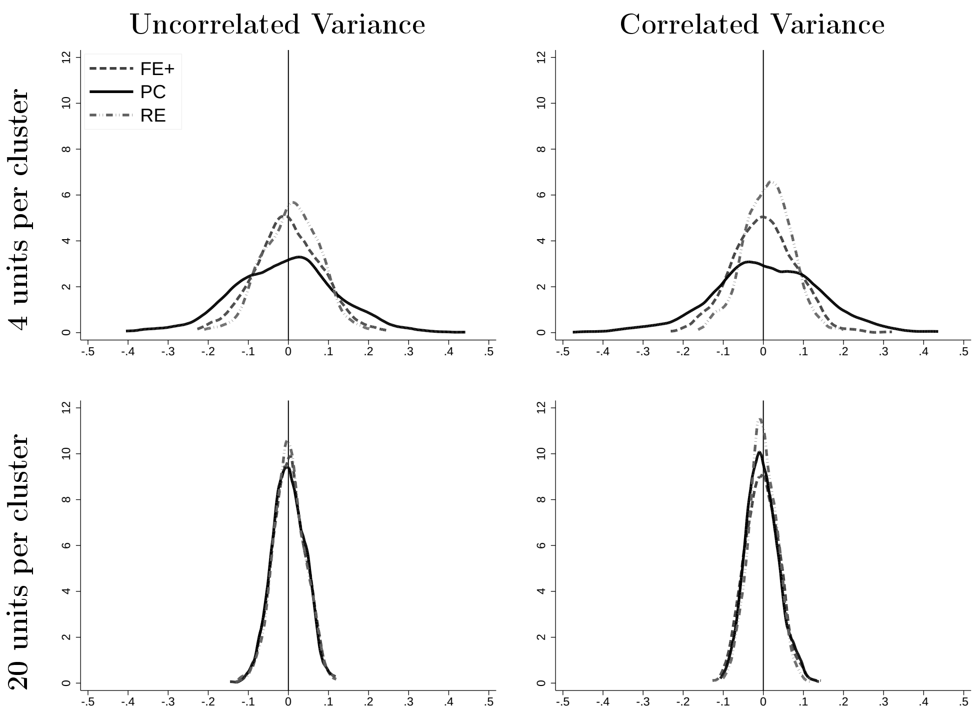

Figure I.1: Kernel density plots of estimation errors, $\widehat{\beta}_2 - \beta_2$, for coefficient of $x_{ij} \times w_j$ across replications for all methods when the uncorrelated variance assumption holds (left panels) and when it is violated (right panels). *Note.* FE+ = Augmented Fixed-Effects; PC = Per-Cluster; RE = Random-Effects.

BIBLIOGRAPHY

BIBLIOGRAPHY

Thomas Ahn and Jacob Vigdor. How salient are performance incentives in education? evidence from north carolina. Technical report, Working paper, Univ. Kentucky and Duke Univ, 2012.

Chunrong Ai and Edward C Norton. Interaction terms in logit and probit models. *Economics letters*, 80(1):123–129, 2003.

P. D. Allison and K. A. Bollen. Change score, fixed effects, and random component models: A structural equation approach. Paper presented at the Annual Meeting of the American Sociological Association., 1997.

Joseph G. Altonji and Charles R. Pierret. Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, 116(1):313–350, February 2001. ISSN 0033-5533, 1531-4650. doi: 10.1162/003355301556329. URL `http://qje.oxfordjournals.org/content/116/1/313`.

Joseph G. Altonji and Robert A. Shakotko. Do wages rise with job seniority? Working Paper 1616, National Bureau of Economic Research, May 1985. URL `http://www.nber.org/papers/w1616`.

D. Ballou, W. Sanders, and P. Wright. Controlling for student background variables in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29:37–65, 2004.

Michael David Bates. Job seperation under asymmetric employer learning. Working paper, University of California at Riverside, July 2015. URL `http://www.nber.org/papers/w5605`.

Gary S. Becker. Investment in human capital: A theoretical analysis. *Journal of Political Economy*, 70, 1962.

C. S. Berkey, D. C. Hoaglin, A. Antczak-Bouckoms, F. Mosteller, and G. A. Golditz. Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, 17:2537–2550, 1998.

G. J. Borjas and G. T. Sueyoshi. A two-stage estimator for probit models with structural group effects. *Journal of Econometrics*, 64:165–182, 1994.

Donald Boyd, Hamilton Lankford, Susanna Loeb, and James Wyckoff. Explaining the short careers of high-achieving teachers in schools with low-performing students. *The American Economic Review*, 95(2):166–171, May 2005. ISSN 0002-8282. doi: 10.2307/4132810. URL `http://www.jstor.org/stable/4132810`.

Donald Boyd, Pam Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. Who leaves? teacher attrition and student achievement. Working Paper 14022, National Bureau of Economic Research, May 2008. URL http://www.nber.org/papers/w14022.

Donald Boyd, Hamilton Lankford, Loeb, and James Wyckoff. Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, 31 (1):83–117, January 2013. ISSN 0734-306X. doi: 10.1086/666725. URL http://www.jstor.org/stable/10.1086/666725. ArticleType: research-article / Full publication date: January 2013 / Copyright Â© 2013 The University of Chicago.

Bernt Bratsberg and Dek Terrell. Experience, tenure, and wage growth of young black and white men. *The Journal of Human Resources*, 33(3):658–682, July 1998. ISSN 0022-166X. doi: 10.2307/146337. URL http://www.jstor.org/stable/146337. ArticleType: research-article / Full publication date: Summer, 1998 / Copyright Â© 1998 The Board of Regents of the University of Wisconsin System.

L. Burstein, R. L. Linn, and F. J. Capell. Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics*, 3: 347–383, 1978.

Carl M. Campbell and Kunal S. Kamlani. The reasons for wage rigidity: Evidence from a survey of firms. *The Quarterly Journal of Economics*, 112(3):759–789, August 1997. ISSN 0033-5533, 1531-4650. doi: 10.1162/003355397555343. URL http://qje.oxfordjournals.org/content/112/3/759.

K. E. Castellano, S. Rabe-Hesketh, and A. Skrondal. Composition, context, and endogeneity in school and teacher comparisons. *Journal of Educational and Behavioral Statistics*, in press, 2014.

Raj Chetty, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly Journal of Economics*, 126(4):1593–1660, November 2011. ISSN 0033-5533, 1531-4650. doi: 10.1093/qje/qjr041. URL http://qje.oxfordjournals.org/content/126/4/1593.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679, September 2014. ISSN 0002-8282.

Matthew M. Chingos and Martin R. West. Promotion and reassignment in public school districts: How do schools respond to differences in teacher effectiveness? *Economics of Education Review*, 30(3):419–433, June 2011. ISSN 0272-7757. doi: 10.1016/j.econedurev.2010.12.011. URL http://www.sciencedirect.com/science/article/pii/S0272775710001767.

C.-P. Chou, P. M. Bentler, and M. A. Pentz. Two-stage approach to multilevel structural equation models: Application to longitudinal data. In T. D. Little, K. U. Schnabel, and J. Baumert, editors, *Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples*, pages 33–49. Erlbaum, Mahwah, NJ, 2000.

Y. Chung, A. Gelman, S. Rabe-Hesketh, J. Liu, and V. Dorie. Weakly informative prior for point estimation of covariance matrices in hierarchical linear models. *Journal of Educational and Behavioral Statistics*, conditionally accepted, 2014.

Charles T. Clotfelter, Helen F. Ladd, and Jacob Vigdor. Who teaches whom? race and the distribution of novice teachers. *Economics of Education Review*, 24(4):377–392, August 2005. ISSN 0272-7757. doi: 10.1016/j.econedurev.2004.06.008. URL http://www.sciencedirect.com/science/article/pii/S0272775704001153.

Charles T Clotfelter, Helen F Ladd, and Jacob L Vigdor. Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education*, 6(3): 399–438, 2011.

Steven J. Davis, R. Jason Faberman, and John Haltiwanger. The flow approach to labor markets: New data sources and micro-macro links. Working Paper 12167, National Bureau of Economic Research, April 2006. URL http://www.nber.org/papers/w12167.

Jed DeVaro and Michael Waldman. The signaling role of promotions: Further theory and empirical evidence. *Journal of Labor Economics*, 30(1): 91–147, January 2012. ISSN 0734-306X. doi: 10.1086/662072. URL http://www.jstor.org/stable/10.1086/662072. ArticleType: research-article / Full publication date: January 2012 / Copyright Â© 2012 The University of Chicago.

F.J.H. Don and J.R. Magnus. On the unbiasedness of iterated GLS estimators. *Communications in Statistics – Theory and Methods*, 9:519–527, 1980.

Michael W. L. Elsby, Ryan Michaels, and Gary Solon. The ins and outs of cyclical unemployment. *American Economic Journal: Macroeconomics*, 1(1):84–110, January 2009. doi: 10.1257/mac.1.1.84.

Henry S. Farber and Robert Gibbons. Learning and wage dynamics. *The Quarterly Journal of Economics*, 111(4):1007–1047, November 1996. ISSN 0033-5533, 1531-4650. doi: 10.2307/2946706. URL http://qje.oxfordjournals.org/content/111/4/1007.

Jason P. Fine and Robert J. Gray. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446): 496–509, June 1999. ISSN 0162-1459. doi: 10.1080/01621459.1999.10474144. URL http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1999.10474144.

E. W. Frees. *Longitudinal and Panel Data: Analysis and Applications for the Social Sciences*. Cambridge University Press, Cambridge, 2004.

Robert Gibbons and Lawrence Katz. Layoffs and lemons. Working Paper 2968, National Bureau of Economic Research, December 1991a. URL http://www.nber.org/papers/w2968.

Robert Gibbons and Lawrence Katz. Layoffs and lemons. *Journal of Labor Economics*, 9(4):351–380, 1991b.

Steven Glazerman, Ali Protik, Bing-ru Teh, Julie Bruch, and Neil Seftor. Moving high-performing teachers: Implementation of transfer incentives in seven districts. Technical report, Mathematica Policy Research, 2012.

Dan Goldhaber, Betheny Gross, and Daniel Player. *Are Public Schools Really Losing Their Best? Assessing the Career Transitions of Teachers and Their Implications for the Quality of the Teacher Workforce. Working Paper 12*. National Center for Analysis of Longitudinal Data in Education Research, October 2007.

Robert J. Gray. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *The Annals of Statistics*, 16(3):1141–1154, September 1988. ISSN 0090-5364. URL http://www.jstor.org/stable/2241622.

Sanford J. Grossman. The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, 24(3):461–483, December 1981. URL http://www.jstor.org/stable/725273.

Cassandra Guarino, Mark D. Reckase, and Jeffrey M. Wooldridge. Can value-added measures of teacher performance be trusted? Technical Report 6602, Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit, 2012. URL http://www.econstor.eu/handle/10419/62407.

Davis Guggenheim. *Waiting for Superman*. Paramount Home Entertainment, Hollywood, CA, 2011.

Eric A. Hanushek, John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. The market for teacher quality. Working Paper 11154, National Bureau of Economic Research, February 2005. URL http://www.nber.org/papers/w11154.

J. A. Hausman and W. E. Taylor. Panel data and unobservable individual effects. *Econometrica*, 49:1377–1398, 1981.

Peter L. Hinrichs. What kind of teachers are schools looking for? evidence from a randomized field experiment. *Work in Progress*, May 2013.

C. Kirabo Jackson. Match quality, worker productivity, and worker mobility: Direct evidence from teachers. *Review of Economics and Statistics*, 2013.

Clement (Kirabo) Jackson. Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *The Journal of Labor Economics*, 27(2):213–256, 2009.

Brian A. Jacob and Lars Lefgren. Can principals identify effective teachers? evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1):101–136, January 2008. ISSN 0734-306X, 1537-5307.

Boyan Jovanovic. Job matching and the theory of turnover. *Journal of Political Economy*, 87(5):972–990, October 1979. ISSN 0022-3808. doi: 10.2307/1833078. URL http://www.jstor.org/stable/1833078. ArticleType: research-article / Issue Title: Part 1 / Full publication date: Oct., 1979 / Copyright Â© 1979 The University of Chicago Press.

Lisa B Kahn. Asymmetric information between employers. *American Economic Journal: Applied Economics*, 5(4):165–205, October 2013. ISSN 1945-7782, 1945-7790.

N. C. Kakwani. The unbiasedness of Zellner's seemingly unrelated regression equations estimators. *Journal of the American Statistical Association*, 62:141–142, 1967.

J. S. Kim and E. W. Frees. Multilevel modeling with correlated effects. *Psychometrika*, 72:505–533, 2007.

E. L. Korn and A. S. Whittemore. Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics*, 35:795–802, 1979.

John M. Krieg. Teacher quality and attrition. *Economics of Education Review*, 25(1):13–27, February 2006. ISSN 0272-7757. doi: 10.1016/j.econedurev.2004.09.004. URL http://www.sciencedirect.com/science/article/pii/S0272775704001281.

Fabian Lange. The speed of employer learning. *Journal of Labor Economics*, 25(1):1–35, January 2007. ISSN 0734-306X. doi: 10.1086/508730. URL http://www.jstor.org/stable/10.1086/508730. ArticleType: research-article / Full publication date: January 2007 / Copyright Â© 2007 The University of Chicago.

Hamilton Lankford, Susanna Loeb, and James Wyckoff. Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1):37–62, March 2002. ISSN 0162-3737, 1935-1062. doi: 10.3102/01623737024001037. URL http://epa.sagepub.com/content/24/1/37.

Audrey Light. Job mobility and wage growth: Evidence from the NLSY79. *Monthly Labor Review*, 128:33, 2005.

Paul R. Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391, October 1981. doi: 10.2307/3003562. URL http://www.jstor.org/stable/3003562.

Paul R. Milgrom and Robert J. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122, September 1982. ISSN 0012-9682. doi: 10.2307/1911865. URL http://www.jstor.org/stable/1911865. ArticleType: research-article / Full publication date: Sep., 1982 / Copyright Â© 1982 The Econometric Society.

Y. Mundlak. On the pooling of time series and cross-sectional data. *Econometrica*, 46:69–86, 1978.

Joshua C. Pinkston. A model of asymmetric employer learning with testable implications. *The Review of Economic Studies*, 76(1):367–394, 2009. doi: 10.1111/j.1467-937X.2008.00507.x.

S. Rabe-Hesketh and A. Skrondal. *Multilevel and Longitudinal Modeling using Stata (Third Edition), Volume I: Continuous Responses*. Stata Press, College Station, TX, 2012.

S. W. Raudenbush and A. S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage, Thousand Oaks, CA, 2nd edition, 2002.

S. W. Raudenbush and J. D. Willms. The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20:307–335, 1995.

S. W. Raudenbush, A. S. Bryk, Y. F. Cheong, R. Congdon, and M. du Toit. *HLM 7: Hierarchical Linear and Nonlinear Modeling*. Scientific Software International, Lincolnwood, IL, 2011.

Steven G. Rivkin, Eric A. Hanushek, and John F. Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, 2005. ISSN 1468-0262.

Jonah E. Rockoff, Douglas O. Staiger, Thomas J. Kane, and Eric S. Taylor. Information and employee evaluation: Evidence from a randomized intervention in public schools. *The American Economic Review*, 102(7):3184–3213, December 2012. doi: 10.1257/aer.102.7.3184.

Roderick A. Rose, Gary T. Henry, and Douglas L. Lauen. Comparing value-added models for estimating individual teacher effects on a statewide basis simulations and empirical analyses. Technical report, Consortium for Educational Research and Evaluation, North Carolina, August 2012.

William L Sanders, Arnold M Saxton, and Sandra P Horn. The tennessee value-added assessment system: A quantitative outcomes-based approach to educational assessment. *Grading teachers, grading schools: Is student achievement a valid evaluational measure*, pages 137–162, 1997.

Tim R. Sass, Jane Hannaway, Zeyu Xu, David N. Figlio, and Li Feng. Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72(2-3):104–122, September 2012. ISSN 0094-1190.

Uta Schönberg. Testing for asymmetric employer learning. *Journal of Labor Economics*, 25(4):651–691, October 2007. ISSN 0734-306X. URL http://www.jstor.org/stable/10.1086/522905.

A. Spanos. Where do statistical models come from? Revisiting the problem of specification. *IMS Lecture Notes – Monograph Series 2nd Lehmann Symposium – Optimality*, 49:98–119, 2006.

Michael Spence. Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374, August 1973. ISSN 0033-5533. doi: 10.2307/1882010. URL http://www.jstor.org/stable/1882010. ArticleType: research-article / Full publication date: Aug., 1973 / Copyright Â© 1973 Oxford University Press.

Douglas O. Staiger and Jonah E. Rockoff. Searching for effective teachers with imperfect information. *The Journal of Economic Perspectives*, 24(3):97–117, July 2010. doi: 10.1257/jep.24.3.97.

StataCorp. *Stata: Statistics/Data Analysis [computer software]*. College Station, TX, 2013.

J. Teachman, G. J. Duncan, W. J. Yeung, and D. Levy. Covariance structure models for fixed and random effects. *Sociological Methods & Research*, 30:271–288, 2001.

G. Verbeke, B. Spiessens, and E. Lesaffre. Conditional linear mixed models. *American Statistician*, 55:25–34, 2001.

Jacob L Vigdor et al. Teacher salary bonuses in north carolina. In *Conference paper, National Center on Performance Incentives.-0.026*, 2008.

D. E. Wiley. Another hour, another day: Quantity for schooling, a potent path for policy. In R. M. Hauser, W. H. Sewell, and D. Alwin, editors, *Schooling and Achievement in American Society*, pages 225–265. Academic Press, New York, 1975.

J. M. Wooldridge. Fixed-effects and related estimators for correlated random-coefficient and treatment-effect panel data models. *Review of Economics and Statistics*, 51:385–390, 2005.

J. M. Wooldridge. *Econometric Analysis of Cross-Section and Panel Data*. The MIT Press, Cambridge, MA, 2010.