

Teacher Turnover in Wisconsin

Michael Chirico

May 13, 2017

Abstract

Given the consistently-affirmed importance of teacher quality to student success, understanding teacher churn is crucial to formulating and evaluating teacher labor market policy. This paper replicates and extends the analysis of Hanushek, Kain, and Rivkin (2004) over a longer and more recent time period in Wisconsin and confirms all of its major findings, namely that while inter-district pay differentials are a significant determinant of turnover and that school quality measures are much better predictors of all three types of churn – within and between school districts and out of local public schools. We further implement and validate a new technique for inferring contracted wage schedules from observed cross-sectional wage data.

Introduction

Good teachers have large impacts on student achievement¹. It is therefore imperative for public schools to be able to attract and retain high-quality teachers. Of preeminent concern for policymakers, then, is the strength of the various manipulable levers at their disposal for influencing teacher labor markets. More specifically, state and local education administrators would like to identify the policy implications of various tools on three types of teacher mobility: intra-district switching, where due to the collectively bargained nature of most teachers' salaries, only nonpecuniary considerations matter, inter-district switching, where teachers move to another school district in the same state, and exo-district switching, where teachers

¹See, e.g., Rockoff (2004).

leave the public teaching workforce entirely².

In this paper, we consider pecuniary and non-pecuniary predictors of various types of teacher churn in replicating the analyses of Hanushek, Kain, and Rivkin (2004) (HKR) in a new context (Wisconsin) and time horizon (2000 - 2010). The headline results of Hanushek, Kain, and Rivkin (2004) were that “teacher mobility is much more strongly related to characteristics of the students, particularly race and achievement, than to salary, although salary exerts a modest impact once compensating differentials are taken into account.” We confirm the pith of this conclusion, namely that student characteristics are a much better predictor of turnover than are wage differentials, though we come to different conclusions regarding more specific points. To wit, while we do find strong evidence that the socioeconomic makeup of a teacher’s district predicts turnover (and that there is heterogeneity in this effect by teacher race), the evidence we find for the importance of wages and student achievement is far from compelling.

We explore to the extent possible potential contributors to this discrepancy in results; most salient are the measurable differences between Texas, where Hanushek, Kain, and Rivkin (2004) conduct their study, and Wisconsin. Wisconsin is a largely rural state – its largest city/metropolitan area, Milwaukee, currently has roughly 600,000 residents (1,500,000 including the metropolitan area), making it around the 30th-largest city in the United States³. By contrast, Texas has six cities larger than this, with El Paso (#6 in Texas) being the nearest in size to Milwaukee. Though the non-urban parts of Texas are themselves sparsely populated and distinctly rural, the more uniform lack of major population centers in Wisconsin is likely to be reflected when considerably different preferences among local vis-à-vis urban residents for various aspects of potential teaching positions are aggregated.

To the end of exploring the pecuniary aspects of teacher turnover, we start by expanding upon the efforts of Hanushek, Kain, and Rivkin (2004) to infer teachers’ tenure-wage paths from teacher-level data on pay. Most unionized teachers are paid according to a salary

²Policies that affect the supply and quality of new teachers to the profession may also be of considerable importance to replenishing and improving the stock of teachers over time, but we do not consider these channels in this work. See Harris and Sass (2011), Wayne et al. (2008), and Boyd et al. (2009).

³In fact, Milwaukee is the only city in Wisconsin considered to be “large” for NCES reporting purposes.

schedule (specifying wages as an increasing function of tenure and certification) explicated in contracts collectively bargained at the district level. With this easily-obtained information in hand, teachers are able to infer their future potential wage trajectories at their own and other potential district employers. Lacking the physical contract faced by the teachers, an econometrician armed only with administrative data reporting actual wages in a given year must use some imputation techniques to deduce the underlying wage structure. We explore the utility of natural Constrained B-Splines (COBS) to this end. COBS are an enhanced version of the traditional semiparametric splines technique enhanced by the ability to impose a monotonicity constraint on the resultant curve which allows the fit to incorporate more local information from nearby experience cells.

The fidelity and utility of the resulting fitted contract curves are supreme. In both large and small districts, COBS produces a plausible tenure-wage arc which enables us to examine counterfactual wage levels for mobile teachers. By comparing the fit to a small number of wage tables obtained from actual contracts, we also learn that using COBS may be preferable to an attempt to use actual wage tables, as the data-derived curves can reveal latent progress of teachers towards further certification, an aspect which is commonly observed in salary tables but rarely included in teacher-level data. Thus, empirically-derived curves can hew closer to the true wage path that rationally forward-thinking teachers are likely to consider when mulling over career options.

Literature Review

Because the potential policy implications of turnover in the teaching profession (from human capital and equity/distributional perspectives both) are far-reaching and polypartisan, the literature on turnover-related topics in education is extensive. As relates to this paper, there are five broad (and often overlapping) categories of inquiry: the relationship between turnover and wages, which has tended to focus on “opportunity wages” outside of the field of education; the relationship between turnover, school demographics, and other nonpecuniary benefits, which has tended to focus on distributional inequalities—whether teachers with

certain characteristics are more or less likely to be teaching certain disadvantaged groups; the relationship between turnover and teacher quality as measured by student performance, usually value added (VA); collective bargaining agreements in education, focusing by and large on the implications (or lack thereof) of seniority-preferential clauses; and the recent phenomenon of specific retention incentives, the provisioning of wage bonuses to teachers willing to teach in high-needs schools.

One of the earliest papers attempting to rigorously investigate turnover was a panel study of teachers in Michigan by Murnane and Olsen (1990), who used college degree field wages outside of education as opportunity wages, finding the expected lower exit rate for teachers with higher wages in teaching relative to the authors' defined alternative. Dolton and Van der Klaauw (1999) use panel data on university graduates in the United Kingdom to estimate a competing risks model of the decision to leave teaching entirely, finding results in line with Murnane and Olsen (1990). Returning to panel studies in the US, Loeb and Page (2000) use PUMS data to get an idea of teacher relative wages in many states and find that dropout rates fall when teacher relative wages are high. Stinebrickner (2002) also uses panel data (this time NLS-72) to track both teachers and non-teachers, focusing in particular on young teachers who leave the profession for long stints, and finds that the best predictor of female exit is recent childbearing, which is an important consideration for all work related to teacher turnover because such a high percentage (76 nationwide) of teachers are female. Lastly, Hanushek, Kain, and Rivkin (2004) focuses on teachers in Texas and emphasizes that the characteristics of students are much stronger factors in predicting teacher exit than are wages (while also affirming the statistical significance of pay).

While wages have been found consistently to have some measurable effect on teacher turnover, it is impossible to explain within-district migration (which constitutes a large portion of switching—as much as 50%) through wage-only channels because contracts are fixed at the district level. As such, another strand of literature has chosen to focus on the nonpecuniary aspects of the decision to take a teaching job—school environment/rapport, student enthusiasm, neighborhood characteristics, etc.—usually by directing attention to a single district so that any wage-based considerations are stifled, as is the case for Boyd et al. (2005) and Engel,

Jacob, and Curran (2014). Boyd et al. (2005) track early-career teachers in New York City as they quit or transfer out of the city, and most importantly finds that commuting time is an important, often overlooked aspect of location preference. Engel, Jacob, and Curran (2014) leverages a unique data set from Chicago Public School job fairs which affords them a rather strong measure of teachers' demand for vacancies, neutralizing the influence of school administration's behavior on turnover (through poor match selection or other means). The authors contribute evidence that the school's neighborhood (perhaps due to ambient crime or other reputational effects good and bad) is a better predictor of teachers' preference than distance from home, going somewhat against the grain of Boyd et al. (2005). Scafidi, Sjoquist, and Stinebrickner (2007) examine statewide data from Georgia, but ignore wage effects, choosing instead to focus on disentangling the contributions of low student achievement and minority status to turnover; they find that minority status is the more salient associate of teacher exit.

The key element missing from all of the above studies is perhaps the most important consideration in the issue of teacher turnover—teacher quality. None of the studies above have student-teacher matched data, and so are unable to directly associate student outcomes with any given teacher. If, with respect to any measure of quality you would like, we find that transitioning teachers are identical to their replacements, the issue of teacher turnover is not, in fact, much of an issue. Thus, the recent trend in the literature to incorporate measures of teacher quality (in large part made possible by a trend towards administrative records allowing students to be linked to teachers and tracked over time) in considerations of teacher turnover has made big strides in addressing the most policy-relevant questions to be asked. The most common and widely accepted measure of teacher quality is VA⁴ (in its various guises), and the literature has begun to incorporate such measures into studies of teacher turnover. Hanushek and Rivkin (2010) consider VA as a measure of teacher productivity, and ask if common results of labor search theory (namely that turnover falls with tenure and that turnover is negatively associated with match-specific productivity) continue to hold in

⁴The most commonly cited expositions on value-added, its validity, and so on are probably Rivkin, Hanushek, and Kain (2005), an extensive exploration of the predictive powers of empirical Bayes VA measures; and Chetty, Friedman, and Rockoff (2014a) and Chetty, Friedman, and Rockoff (2014b), the largest-scale study of long-term inferences based on VA.

the education labor market. In fact, the authors find that the teachers most likely to switch schools are those with low measured match quality, and especially that those who leave teaching entirely are those with the lowest match quality. The results are more pronounced for schools with high proportions of low-SES students, which has strong policy implications, as it appears the best teachers in high needs schools are the least likely to change jobs. Goldhaber, Gross, and Player (2007) performs a similar analysis with the longitudinal data of North Carolina and comes to similar conclusions, strengthening the robustness of the results. Lastly, Goldhaber, Lavery, and Theobald (2015) examine the inequity in the distribution of teacher quality by high-needs groups in Washington state, and find that for all three measures of quality (teacher experience, licensure exam score, and VA), the distribution of teachers favors the less needy (as measured by free/reduced-price lunch status, minority status, and low prior academic achievement).

The aforementioned papers have tended to keep the collective bargaining aspect of salary determination for teachers out of the spotlight, if largely for reasons of data restrictions. Nevertheless, it stands to reason to believe that the rigid structure of union-negotiated contracts could serve to contribute in a large way to teacher turnover. Ballou and Podgursky (2002) give much descriptive evidence of the shape of the wage-tenure profile, rooted in a data set collected by the Department of Defense and published by the AFT. They find that seniority premia in education largely mirror those in more traditional white collar professions, that steeper profiles are associated with less turnover, and that district financial and demographic conditions alone are insufficient to explain variation in contracts. Another common (and recently quite controversial, as evidenced by the contention in the ongoing contract negotiations in Philadelphia) feature of union-negotiated teacher contracts are seniority privileges—preferential treatments granted to teachers in voluntary and involuntary transfers. Moe (2006) codes contracts from 158 districts in California according to the strength of seniority rights therein guaranteed to teachers and finds that such rights are associated with the distribution of teachers across schools (measuring quality as experience and certification) in a way that serves to harm minorities. Revisiting California with a slightly different sample and definition of the “determinacy” of the contracts with respect to seniority,

Koski and Horng (2007) come to the opposite conclusion—that there is no such relationship. As a rebuttal, Anzia and Moe (2014) pin the difference in results on the exclusion in Moe (2006) of small school districts, where it appears that the entrenchment of bureaucracy falters and the rigidity of contract language wane, a claim which they support by repeating their analysis with the inclusion of an interaction for district size—indeed, for small districts the result of Koski and Horng (2007) holds, while the insight of Moe (2006) holds in larger districts. Cohen-Vogel, Feng, and Osborne-Lampkin (2013) use data from Florida and their results align with those of Koski and Horng (2007) (though they neglect to nuance their results by district size).

Finally, an emerging strand of literature is looking at the potential for transfer bonuses and retention incentives to positively affect student outcomes. Fulbeck (2014) analyzes a scheme in place in Denver whereby teachers who choose to transfer to high-needs schools (low-performing) are given recurring bonus pay, and those initially stationed there are given retention incentives. She concludes that recipients of incentives are significantly less likely to switch jobs, as driven by a reduction in district exit rates and especially by teachers whose incentive payments exceed \$5,000. Glazerman et al. (2013) evaluate the Talent Transfer Initiative, a randomized controlled trial conducted in 10 districts whereby high-performance teachers were given \$20,000 over the course of two years as reward for transferring the identified high-needs schools, and conclude that there were significant effects on teacher retention as well as on student outcomes.

Two highly germane papers investigate the impact in Wisconsin on teachers of Governor Scott Walker’s flagship policy, Act 10, which severely limited the scope for collective bargaining in the state. Litten (2016) uses differences in contract renewal dates surrounding the policy’s enactment to evince the effect of unionization on teachers’ wages, and finds the lack of union bargaining power reduced teacher compensation by 8%. Biasi (2017) constructs value-added measures from grade-level test results and concludes that the move to individually-negotiated salaries in some districts had a significant impact on teacher quality and student outcomes in such districts, while also cautioning that most of these gains are competition-based, so that scaling up the system state-wide would have an impact limited to a boost from the exit of

low-quality teachers.

Data

The State of Wisconsin’s Department of Public Instruction (DPI) releases annual Salary, Position & Demographic reports through the WISEstaff data collection system. These reports represent “a point-in-time collection of all staff members in public schools as of the 3rd Friday of September. . .” (Public Instruction 2017b), and will serve as the primary source of data on teachers in this paper. Data are available at the position-teacher level cross-sectionally, with each entry in a given year corresponding to one of possibly several positions/assignments held by each school district employee⁵. Identifiers in each file permit unique identification of an employee within a given year, but this identifier does not follow teachers between years⁶. To overcome this substantial hurdle to identifying teacher mobility, data are first fed through the matching algorithm described in further detail in the Appendix. Essentially, we are aided by the availability of various imperfect identifiers which should be more stable over time, most crucially teachers’ first and last names and year of birth. By building on these covariates and incorporating some limited fuzzy matching techniques, we construct a panel of teachers spanning the 1994-95 academic year (AY) through AY2015-16⁷ consisting of 3,588,614 teacher-position-year observations. The matching algorithm necessitates elimination of 26,304 (0.7%) observations over all 21 years on account of belonging to teachers who could not be uniquely identified in a given year of data due to exact overlap of their first name, last name, and birth year fields with another teacher in the data⁸.

Specific to the exercise at hand, with data reliability and precision in mind, we make the following series of further restrictions on the data. The introduction of Wisconsin Act 10

⁵Many teachers (and other district employees) serve in multiple roles within a school/district, for example as a coach, part-time program aide, or department head. Each of these is filed as a separate observation in the DPI system, though salary information is given at the teacher as opposed to the assignment level.

⁶From AY2011-12, a field called the File Number appears to allow longitudinal tracking of teachers. We use this in part to validate the matching algorithm; see the Appendix.

⁷For brevity, we herein refer to academic years by the spring year, e.g., AY2003-04 will be simply 2004.

⁸Technically, we use a slightly modified version of the name strings in making these eliminations which, for example, eliminates initials – see Appendix.

introduced a substantial structural break in the labor market for Wisconsin teachers, so we include only data from 2000-2010 to avoid conflating the effects of this policy on teacher turnover with the earlier functioning of the labor market (i.e., we do not want to mix the results from distinct equilibria of the teacher labor market, but would instead prefer to analyze the pre- and post-Act-10 markets separately). We drop all employees who are not full-time, full-year regular teachers of a major core subject (all-purpose elementary teachers or English/Math) at a single regular public school with a Bachelor’s or Master’s degree and fewer than 35 years’ recorded experience; taken together, these restrictions eliminate 79% of employees, the lion’s share of which come from eliminating substitutes/support staff and teachers of on-core subjects⁹. We then eliminate teachers with missing information on their subsequent school or district and teachers with instability in their recorded ethnicity, as well as teachers not categorized as white, black, or Hispanic, eliminating a further 0.2% of all employees¹⁰. Finally, we drop teachers’ multiple positions by keeping only the highest-intensity position for each teacher, as measured by full-time equivalency, resulting in a final count of 282,797 teacher-year observations – 49,325s in 449 districts and 2,296.

The data used for the incorporation of counterfactual salary calculations is largely the same, but with a few noteworthy differences. First, as noted in Footnote 9, the main turnover data eliminated some teachers who transitioned in and out of being categorized as a full-time teacher due to the muddling effects thereof on defining turnover. This concern not being relevant to constructing the salary schedules, it is not imposed for this data. Next, because all regular teachers are covered by the same collective bargaining agreement, the salary imputation data is less restrictive with respect to the subject codes excluded from the data, and generally includes any teacher not in special education. This data also ignores instability

⁹We also eliminate any teacher who appears in any role besides “Teacher” in any year. In particular, this eliminates a nontrivial number of educators who begin their career with an “ease-in” period, take a mid-career “leave” speckled with a transition to substitute teaching – perhaps during their child’s infancy – or end it with a “soft retirement” period, during which they act as a substitute teacher at some point in the midst of a career otherwise focused on teaching. Such teachers often have part-time roles at several local schools, which introduces sufficient ambiguity in the definition of mobility so as to obscure interpretation of results, so we opt for a stricter definition of full-time teaching than is completely necessary.

¹⁰Wisconsin teachers are predominantly white (96%). As noted in the Appendix, we also use the panel data to correct noise found in recorded ethnicity and gender over time for some teachers. In the final sample, 427 and 345 teachers had their ethnicity and gender (respectively) adjusted in some year(s).

in recorded gender and ethnicity within a teacher.

Finally, the salary data loses observations that are present in the turnover data based on a series of cuts which are either required for COBS to function, or else substantially increase the reliability of its output. The most noteworthy/far-reaching of these numerical restrictions is to eliminate any teachers working in districts where there are not at least 20 total teachers in each degree track for that year. While ultimately arbitrary, this number is reasonable to limit the potential effect of an individual teacher on an exercise determining 35 levels of pay with minimal functional form restrictions. The other numerical flags require both the BA & MA track to be represented at a district, for at least 7 distinct levels of experience to be represented within a degree track, and for at least 5 unique values of the two measures of pay (salary and fringe benefits) to be available in each degree track; all teachers at districts failing at least one of these tests is dropped.

The sum total of all of these restrictions leaves us with an analysis sample of 356,265 teacher-year observations to be used to estimate pay scales, made up of 65,069 individual teachers in 209 districts over 11 years. In total, there is sufficient data to fit 3,708 $y_t(\tau, c, d)$ curves, an average of roughly 100 observations per curve. Ultimately around 22% of teachers have missing salary information¹¹, mostly in rural districts or other districts with only one or two schools and a small number of students.

We supplement the WISEstaff data set in several ways to incorporate information about other characteristics of schools and districts in Wisconsin. To get school- and district-level measures of socioeconomic makeup (percentage of students who are black or Hispanic or eligible for free/reduced lunches) and community type/urbanicity, we tap the Universe Surveys from the National Center for Education Statistics' Common Core of Data, which provide this

¹¹HKR include like-minded restrictions, but combine teachers of different certification within an experience level, despite the headline importance of this factor to teacher pay – median pay at a given level of experience is on average 17 higher for those with a Master's degree.

information on a yearly basis for all years in the study^{12,13}. At the district level, we also use this data to compute class size and the size of the student body.

Lastly, we turn to DPI’s public data again to get school- and district-level performance metrics. While Hanushek, Kain, and Rivkin (2004) were able to obtain school- and district-level average scale scores on a standardized test in Texas, such a metric is not publicly available in Wisconsin for all years. Instead, we calculate student proficiency rates for each school and district as the percentage of test-takers deemed to be at grade level in mathematics or reading in a given year on the Wisconsin Knowledge and Concepts Examination (WKCE), which is administered to 4th, 8th, and 10th-grade students.

Salary Scale Imputation with Constrained B-Splines

For many years, the ubiquitous characteristic of collectively bargained teachers’ contracts has been the salary table, which gives a mapping from the calendar year, a teacher’s experience (their length of tenure at the current district), and their certification (typically Master’s vs. Bachelor’s degree) to their wage. This table gives current teachers a clear understanding of how their pay will advance as a function of their labor inputs, and thereby gives forward-looking potential teachers and potential migrant teachers a clear understanding of their would-be pay arcs under a district-switching decision-making framework, especially given that this information is typically openly available.

It would behoove an econometrician seeking to understand education labor market dynamics, then, to incorporate this information on future pay into their statistical modeling framework. Unfortunately, this data is typically not available in a format lending itself to easy analysis at

¹²The method of recording urbanicity by the Common Core switched from being “metropolitan-centric” to being “urban-centric” for Wisconsin from 2006 (Sable 2009). We map the corresponding codes to match those used by HKR as well as possible, and use the data file from 2006, which has both types of code for all US districts, to confirm that the pre- and post-2006 correspondence is by-and-large working as intended. For a small number of districts/schools with missing urbanicity codes in certain years, we use information about that entity from other years to inform urbanicity.

¹³Further, the WKCE data does not include a standard deviation field even in those years when the school average scale score is available, precluding any attempt to standardize test scores and put the data here on equal footing with that of HKR.

scale – whether locked inside idiosyncratically formatted and sporadically-available contract PDFs or hidden behind large-scale freedom of information act inquiries, the temporal and financial costs of scraping such data into a usable form can be substantial.

Much more common in empirical settings is access to teacher-year-level salary data of the form $y_{i,t} = y(\tau_{i,t}, c_{i,t}, d_{i,t}) + \varepsilon_{i,t}$, where $\tau_{i,t}$, $c_{i,t}$ and $d_{i,t}$ are the tenure, certification, and district of teacher i in year t , and $\varepsilon_{i,t}$ represents unaccounted factors affecting the wage (e.g., not all teachers work full time, some teachers split their time among duties yielding different pay levels, and many teachers supplement their income with additional duties like coaching). Here we consider one approach and some empirical lessons for trying to estimate the underlying mapping $y(\tau, c, d)$ from such data.

There are a multitude of inference/imputation techniques suitable to the inference of a latent function of unknown parametric form available in the statistician/econometrician’s palette. The powerful flexibility of nonparametric approaches (local regression, splines, Random Fourier Feature expansions) is a double-edged sword; as it happens, in this particular setting, even if we know linearity is not a reasonable functional form restriction, we do know some very basic properties of the underlying tenure-wage curves that will be violated in general by uninformed estimation techniques. In particular, we know that such tenure-wage curves are non-decreasing and that they are non-negative, i.e., $y(\tau', c, d) \geq y(\tau, c, d)$ whenever $\tau' \geq \tau$, and $y(0, c, d) \geq 0$.

He and Ng (1999) introduce a linear programming approach to incorporating monotonicity, curvature, and pointwise restraints to quantile regression spline estimation techniques, and Ng and Maechler (2007) present an overview of the R package `cobs` which gives an efficient implementation of this approach (COBS standing for Constrained B-Splines; B-splines are computationally-efficient basis functions for degree k splines). The basic idea of quantile regression spline estimation is to swap out the standard squared loss function for a quantile-dependent weighted absolute loss function to target conditional quantiles instead of conditional means. Monotonicity, point, and curvature restrictions enter as penalized terms to the objective function; `cobs` expresses this in a fashion which facilitates the application of

standard linear programming techniques for efficiency, and handles internally the issues of knot selection and penalty parameter assignment through cross-validation.

We implement and fine-tune this general approach with an eye to being as minimally-invasive as possible. The first innovation is required by the poor performance of standard COBS fit in extrapolation. Data sparsity in smaller districts means that it is often the case that only a small range of τ values are observed in a given year-certification-district. Monotonicity constraints are only built into the B-spline routine internally; the underlying basis functions may produce decreasing fits outside the observed range of data. To overcome this, we take a cue from the literature tackling Runge’s Phenomenon (Runge 1901), wherein polynomial approximations tend to exhibit extreme oscillations in extrapolation. This issue is one of the motivations behind natural cubic/smoothing splines (see, e.g., Friedman, Hastie, and Tibshirani 2001; Wahba 1990; Green and Silverman 1993; or de Boor 1978), which handle this issue by using a simple linear basis function outside the outermost interpolating knots. We incorporate this technique of linear extension only when necessary by testing the COBS fit for monotonicity; τ values failing this constraint are replaced by extending the final non-decreasing fit values through the end of the range of extrapolation.

Next, a major shortcoming of COBS for this context is its limit to one-dimensional spline fits; while techniques for nonparametric B-spline fits are available in arbitrary dimensions (see de Boor 1978), at present COBS is only capable of imposing monotonicity on one dimension of a curve. In our context, however, $y(\tau, c)$ is increasing not only with respect to τ , but also with respect to c (as, without fail until only very recently in Wisconsin, certification was rewarded with a Master’s premium, typically a percentage increase in wage). One solution would be to generalize the implementation of COBS to handle a second dimension by simply adding penalty terms along this dimension¹⁴. We abandon this approach because of the categorical nature of the certification dimension – there are not numerical units to the difference between having a Master’s vs. Bachelor’s degree. The assignment of such a number required by this

¹⁴Not to mention the empirical reality that the two-lane dichotomy is in fact false – it is very common, nearly ubiquitous, for contracts to offer separate lanes for teachers with the same completed certification, but different levels of progress towards completing further certification. As this dimension is impossible to glean from our data, we exclude it from the imputation exercise.

approach would itself become an implementation hyperparameter, meaning that the ultimate fit would itself be sensitive to the particular choice of continuous representation.

Instead, we use a two-step procedure to fit the Bachelor’s and Master’s pay tracks in serial. In the first step, we fit the Bachelor’s career track as a typical one-dimensional COBS fit. In the second step, we first construct Master’s premia for each observation by subtracting out the predicted Bachelor’s pay corresponding to each observed level of tenure for a teacher with a Master’s degree. We then use COBS to fit a non-decreasing Master’s premium curve over all tenure levels on these residuals, before finally adding the Master’s premium and Bachelor’s fit curves to get the overall Master’s fit curve. Monotonicity of the result is guaranteed by forcing upwards monotonicity on the Master’s premium, a restriction in line with the empirical observation that Master’s degree pay is often simply a fixed-percentage rise over the corresponding Bachelor’s pay.

Our implementation was also aided by the imposition of weak concavity on the $y(\tau, BA)$. While not a theoretically-assured functional form restriction¹⁵, concavity improves the goodness of fit notably. Small-sample district-level observations and simple reduced-form regressions of wages versus quadratic forms in experience support this shape’s validity. A variety of contracts obtained from a database for teachers in nearby Michigan also meet this condition, and the decrease in marginal returns to experience is also commonly found in the wider study of labor markets¹⁶. We do not impose this restriction on the fit for the Master’s premium (the only restrictions there being non-negativity at 0 and, as mentioned, an increasing relationship with tenure).

As an illustrative example of the patterns in the data we wish to quantify and formalize, we turn briefly now to Milwaukee Public Schools, the largest district in Wisconsin with roughly 32,387.73 teachers per year. Figure 1 depicts key moments of the empirical distribution of salary in 4 years at Milwaukee Public Schools, broken down by tenure and certification.

¹⁵In fact, in reality tenure-wage curves are often piecewise convex – year-over-year rises are specified as a percentage bump which eventually levels off to either linear increase or maxes out and flattens. Nevertheless, the degree of convexity in that section of the curve tends to be low, which leads COBS to fit a good linear approximation there. The lack of a concavity restriction on the Master’s pay track allows fit curves which fit this pattern for this lane.

¹⁶See, e.g., Heckman, Lochner, and Todd (2003).

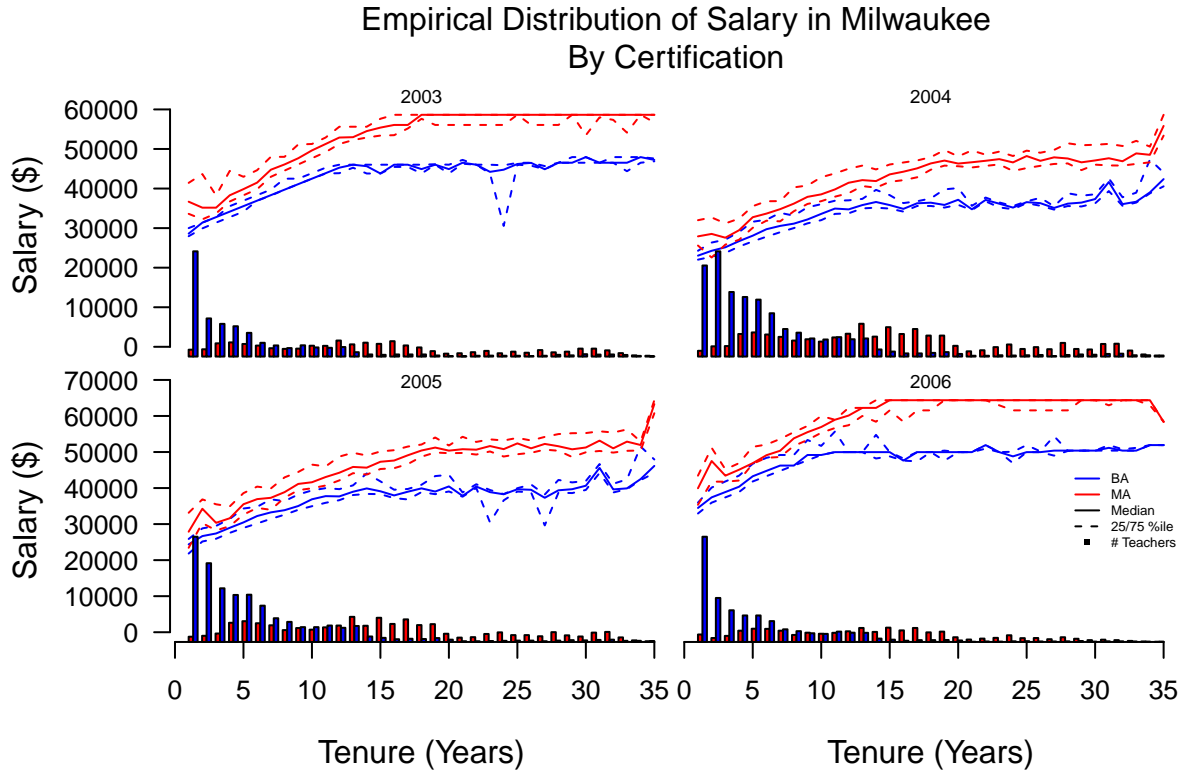


Figure 1: Pay in Milwaukee, 2003-2006

The central lines on each plot (Bachelor's pay track in blue, Master's pay track in red) are the empirical median levels of pay, and thus give a rough approximation to $y(\tau, c)$. The dashed-line intervals on either side represent the 25th and 75th percentiles.

Notably, these intervals and the medians themselves tend to get quite noisy at later stages in the career, especially for the Bachelor's track. This fact that reflects the almost universal certification of teachers by about 15 years into their career. This is reflected in the bar graph below each set of curves, which shows the distribution of teachers in each certification track by tenure. Almost all new teachers start with only a Bachelor's degree; the relative presence of Master's degrees grows over time as more teachers certify mid-career.

We can also note two more key empirical facts from this plot. First, the vanishing presence of teachers in both certification tracks leads the empirical median to be a poor approximation of $y(\tau, c)$ since it frequently fails to respect the fundamental monotonicity constraint discussed above. With respect to tenure, this tends to affect the Bachelor's track later in the career

as more teachers certify, and the Master’s track very early in the career before teachers certify. The monotonicity with respect to c of the median wage is mostly maintained here for Milwaukee, but this is not always the case; our estimation procedure is thus careful to impose these restrictions internally.

Second, structural breaks are an important empirical phenomenon in this context. Each time a contract is renegotiated at a district, the tenure-wage curves can potentially change shape dramatically. It is with this in mind that we refrain, given our ignorance with respect to when such structural breaks occur, from combining information from adjacent years in fitting a given year’s curve, an approach which would substantially enhance the statistical power available to fit contracts for sparsely-populated districts. Such a structural break is apparent in Milwaukee, for example, between 2003 and 2004 and between 2005 and 2006, where the shape of the Master’s pay scale has shifted notably. While we eschew, for example, full Bayesian estimation of structural breaks in a given district, such techniques are applicable and worthy of future exploration.

Goodness of Fit

Returning to the motivating example illustrated in Figure 1, we turn first to the performance in Milwaukee, where, given the relatively large sample size, performance is expected to be very good. Indeed this is the case, as seen in Figure 2. The COBS fit has retained all the salient features of the empirical median return to experience and certification, while simultaneously improving over this nonparametric conditional median by ironing out nonmonotonicities found empirically as a result of small-sample bias.

Perhaps more telling is the goodness of fit in minimally small districts. Four such examples are featured in Figure 3. These four districts just barely satisfy the sample restriction that at least 20 teachers be present in both the BA and MA pay track (each has fewer than 42 teachers, and only in a single year); for this reason, rather than plot the empirical median, we simply present the full distribution of wage, experience, and certification in these districts. Here again the COBS fit captures the essence of the wage-tenure curve even in

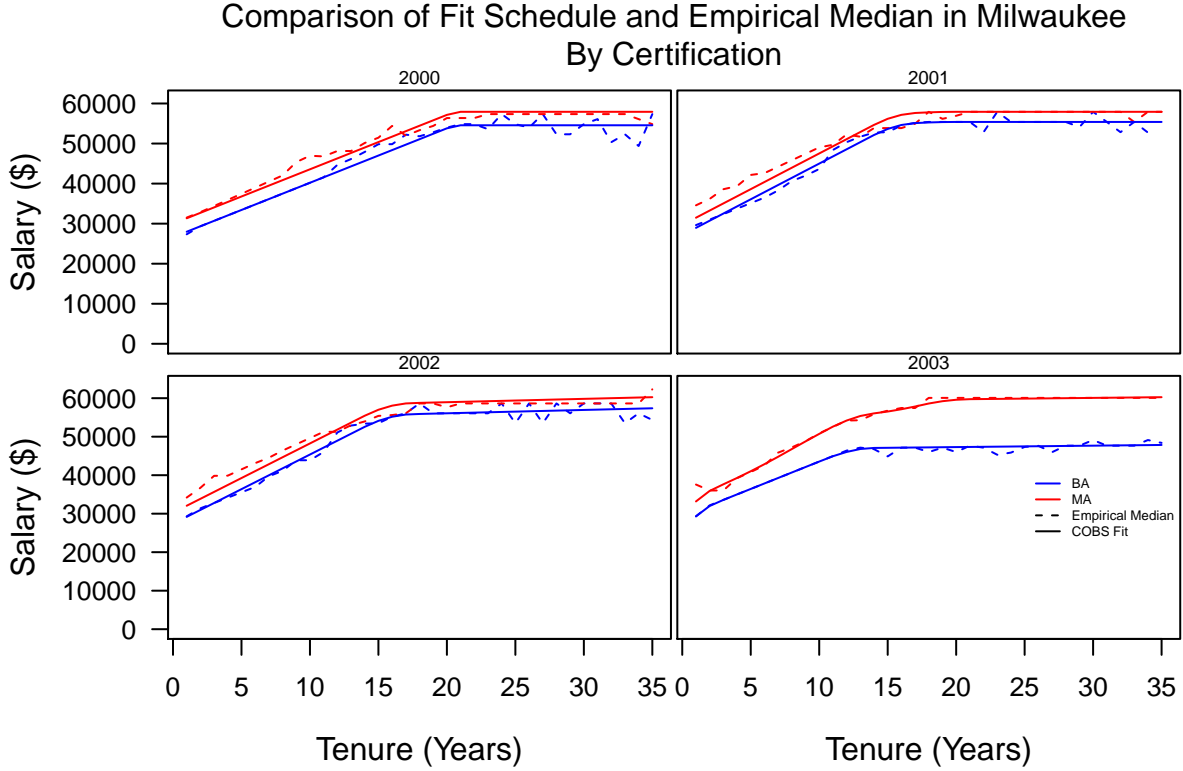


Figure 2: Estimation Results for Milwaukee, 2003-2006

these sparsely-staffed districts. Both Montello and Manawa evince the importance of the non-negativity constraint on extrapolated values of the Master’s premium – given the absence of teachers so certified prior to the fifth year of experience, some supplementary discipline is necessary to prevent the tail of this curve from dipping below that for the Bachelor’s lane.

A final check on the validity of the imputation procedure would be to compare fit schedules side-by-side with the true schedules, e.g. through root mean-squared error. As mentioned, this is typically a difficult undertaking on a mass scale since the true schedules may be hard to come by in a parseable electronic format. Usually, a smaller-scale version of this exercise would be possible through sampling, say, 5-10% of districts at random and spending the time to extract actual schedules by hand for this purpose. Unfortunately, with the passage of Act 10 and the abandonment of collective bargaining in many districts, electronic copies of legacy contracts became hard to come by – none of the large districts we contacted (nor their former union representatives) had access to old copies of contracts they were willing to share, nor could we find any but a very small number of these contracts online. We present here the

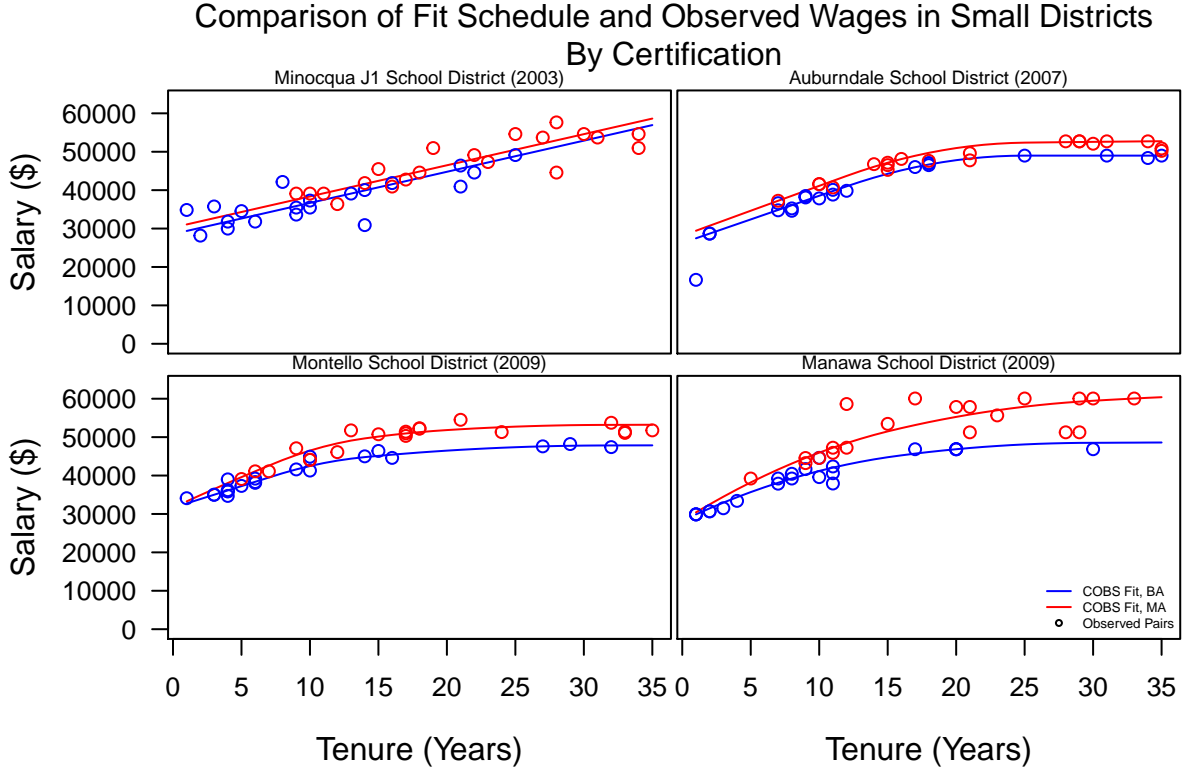


Figure 3: Estimation Results for Selected Sparse Districts

comparison of COBS-produced fit to true schedule in three district-year combinations for which we could actually obtain the true schedule¹⁷.

As seen in Figure 4, the resulting fit is generally superb¹⁸. Only for the Bachelor's track in Monona Grove in 2009 does the COBS-fit curve depart substantially from the true contracted schedule. Moreover, this departure is likely attributable to the oversimplification taken in this paper of restricting pay to follow only two "lanes" (Bachelor's and Master's degrees), when in reality districts often differentiate among holders of these degrees by rewarding those with more credit-hours of supplementary coursework under their belts in pursuit of continued learning or a higher degree – in fact, such coursework is often required of Bachelor's-certified

¹⁷These contracts and a few others from outside of the study time frame are available upon request.

¹⁸In terms of objective measures of the fit, the mean absolute error is \$1,762, while the overall median error is \$812. This is evidence against the assumption built into the COBS routine of 0-median errors $\varepsilon_{i,t}$, and is understandable – it is not uncommon for teachers to earn supplementary pay from coaching or extra teaching duties that would push them above their salary-schedule-dictated pay grade. With a more complete set of training data, one could potentially account for this by treating the quantile of the data targeted by COBS (.5 by default, i.e., COBS is median-targeted) as a hyperparameter to be fit by cross-validation to prevent overfitting (see Stone 1974 or Friedman, Hastie, and Tibshirani (2001)).

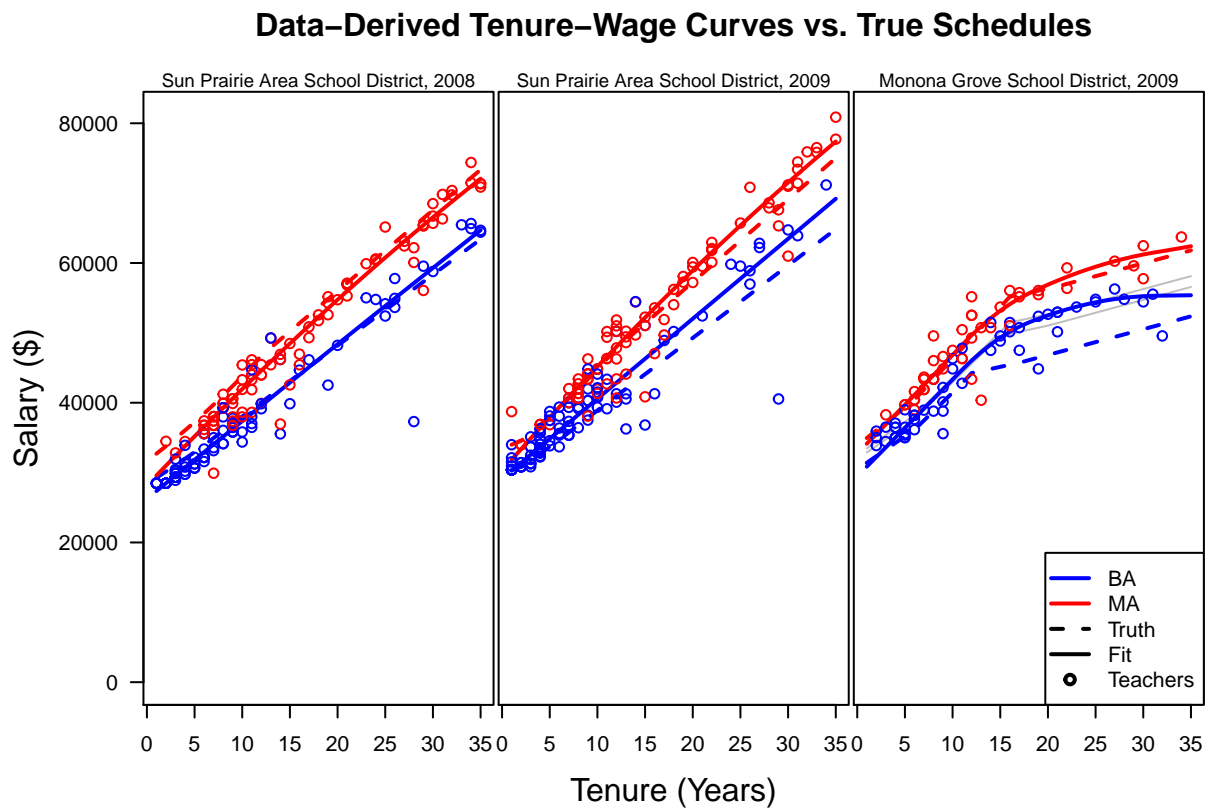


Figure 4: Comparison of True Contracted Schedule with Output of Imputation

teachers, which means it is likely that the later-career Bachelor-certified teachers observed in Monona Grove are actually being paid according to a higher lane. This is exactly what is depicted by the gray lines on the Monona Grove plot, which show the BA+12 and BA+24 lanes are more representative of instructors at the later stages of their career in Monona Grove (the data lack any way of detecting a given teacher’s extracurricular credit accumulation).

We take the above as strongly affirming the utility of COBS as a tool for constructing wage-tenure curves from teacher-level salary data. It is able to gloss over noise-induced non-monotonicities in the empirical median, not just with the rich data found in urban districts, but also in sparsely-populated districts. Moreover, as explored in the case of Monona Grove School District, COBS can be seen as doing a good job of capturing an aspect of the data which is still latent (namely, the degree of progress towards further certification), and of being closer to the “true” schedules that teachers use to make mobility decisions (since it is likely that teachers are able to anticipate extra income from holding multiple roles and factor this into their assessment of a wage offer). Lastly, the COBS routine is computationally attractive – embarrassingly parallelizable and implemented very efficiently, the whole routine runs in a few minutes.

Turnover in Wisconsin

Teacher Experience	Percent of Teachers Who				Number of Teachers
	Remain in Same School	Change Schools Within District	Switch Districts	Exit Wisconsin Public Schools	
1-3 years	79.7	7.1	6.0	7.2	41,042
4-6 years	86.6	5.6	3.3	4.5	37,770
7-11 years	90.6	5.0	1.8	2.6	54,623
12-30 years	92.4	4.1	0.6	2.9	129,002
>30 years	80.3	6.4	1.1	12.2	20,360
All	88.6	5.1	2.0	4.3	282,797

Table 1: Year-to-year Transitions of Teachers by Experience, 2000-10

We move now to the core focus of our analysis, examining the distinguishing features of turnover in the teacher labor market in Wisconsin. Table 1 replicates Table 1 of Hanushek,

Kain, and Rivkin (2004), and as HKR found in Texas, most turnover in Wisconsin is happening within districts and out of the profession^{19,20}. In Wisconsin, the fraction of teachers transitioning among districts is vanishingly small after a “burn-in” period of roughly 6 years – only 1% of such teachers do so (compared with 3.1% for the comparable group in HKR), but is still relatively highest among the youngest teachers – roughly twice as high for the “probationary” teachers (1-3 years’ experience) as for teachers with 7-11 years’ experience in both states.

By contrast, movement patterns within districts in the two states are very similar, lending weight to teachers “earning their stripes” within a district to be able to choose the best schools as a privilege of seniority. As expected, we also observe a U-shaped pattern in teachers exiting Wisconsin public schools, which jibes with there being two types of quits. Early-career quitters change to private schools, change state of residence, or change professions; late-career quitters retire – especially evident among teachers with more than 30 years’ experience, a group which sees a mass exodus of fully 10 percent of its teachers annually. Results not included here break down the exit rates by experience level, where this dichotomy is even more dramatic – first-year exit rates are about 8 percent and quickly level off at around 2 percent before spiking again past around 25 years.

As examined further below, the low rate of switches between districts appears to be owing to the generally more rural nature of Wisconsin vis-à-vis. Texas. To wit, Milwaukee is the only major urban area in the state, and its population (2010 Census) of 594,833 would rank 7th in Texas. This means that two major types of movers in the HKR data – Large Urban - Large Urban and Suburban - Large Urban – are limited within the state to ending up in a relatively minor metropolitan area. HKR don’t provide any results disaggregated by city,

¹⁹This and subsequent analyses were greatly facilitated by several facilities of the R programming language, for which due credit must be given to R Core Team (2016), RStudio Team (2017), Dowle and Srinivasan (2017), Xie (2016), Leifeld (2013), Dahl (2009), Henningsen and Toomet (2011), Zeileis and Hothorn (2002), Zeileis (2004), Zeileis (2006) and Croissant (2012).

²⁰We also note that some “turnover” identified by teachers not appearing at the same school in the following year is in fact spurious – Public Instruction (2011) identifies a number of instances of school districts merging during the timeframe of our analysis and hence disappearing from the data altogether. We take care to reset the district and school switch identifiers off for these 82 teachers if they appear in the newly-formed district in the subsequent period.

precluding any attempts to compare these numbers more comparably to those that would obtain from eliminating the largest cities in Texas.

Moving from the aggregate numbers to begin to examine heterogeneity in turnover, Table 2 replicates HKR Table 2, and reverberates its most important conclusions. HKR argue that there is little support for the idea that scores of young teachers are using large urban schools as a training ground before “settling down” with easier assignments in the suburb, based on the general low level of turnover from Large Urban districts. We affirm the scarcity of transitions from districts in Milwaukee, while also noting that such a path is certainly present, as evidenced by the plurality of those who do leave Large Urban districts ending up in a Suburban district in both settings. HKR also observe that the likelihoods of remaining in the same school and of quitting are roughly the same for urban and suburban teachers, an observation which we can confirm in Wisconsin. We further note that while Table 2 only presents a cross-sectional picture, the career-long trend reaffirms this – only 3.2% of teachers starting their careers at a large urban district ever work at a suburban district. Lastly, we echo the suggestion of HKR that this phenomenon cannot be driven purely by demand-side constraints – in our time period of observation, we observe only 1,459 urban teachers change districts, whereas 3,211 teachers were hired in suburban districts, though of course this does not rule out arguments based for example on stricter screening of applicants transferring from urban districts.

We note, however, that though tales of flight from troubled urban districts are apparently anecdotal, they are far from apocryphal. To wit, while 50 percent of districts have a net inflow (arrivals less departures) of four or fewer teachers (in absolute value), Milwaukee’s net outflow was 533 teachers, and the five highest-inflow districts, all suburbs of Milwaukee or districts adjacent the main university town of Madison, saw in total an inflow of 229 teachers in this time. This being a two-sided market, this state of affairs is perhaps largely attributable to the dynamic nature of student populations at these districts – but these, as well, are reflective of the appeal of the districts to parents (and teachers as parents).

As mentioned in the discussion of Table 1, the major difference with respect to quantities

Origin Community	Percent of Teachers Who Move to				Number Teachers Changing Districts	Percent of Origin Teachers	Change in Share of Teachers 2000-06
	Large Urban	Small Urban	Suburban	Rural			
I. All teachers							
Large Urban	17.4	15.8	48.7	18.1	819	2.7	0.4%
Small Urban	3.7	13.4	44.8	38.1	640	1.2	0.1%
Suburban	3.4	16.0	44.2	36.4	1,408	1.9	3.6%
Rural	0.6	11.2	24.2	63.9	2,794	2.2	-4.1%
II. Probationary teachers (1-3 years experience)							
Large Urban	15.8	17.9	47.6	18.8	437	5.0	
Small Urban	5.1	14.0	46.2	34.8	271	3.9	
Suburban	4.3	16.2	41.0	38.4	561	5.5	
Rural	0.3	10.9	24.4	64.5	1,204	8.0	

Table 2: Destination Community Type for Teachers Changing Districts, by Origin Community Type and Teacher Experience Level

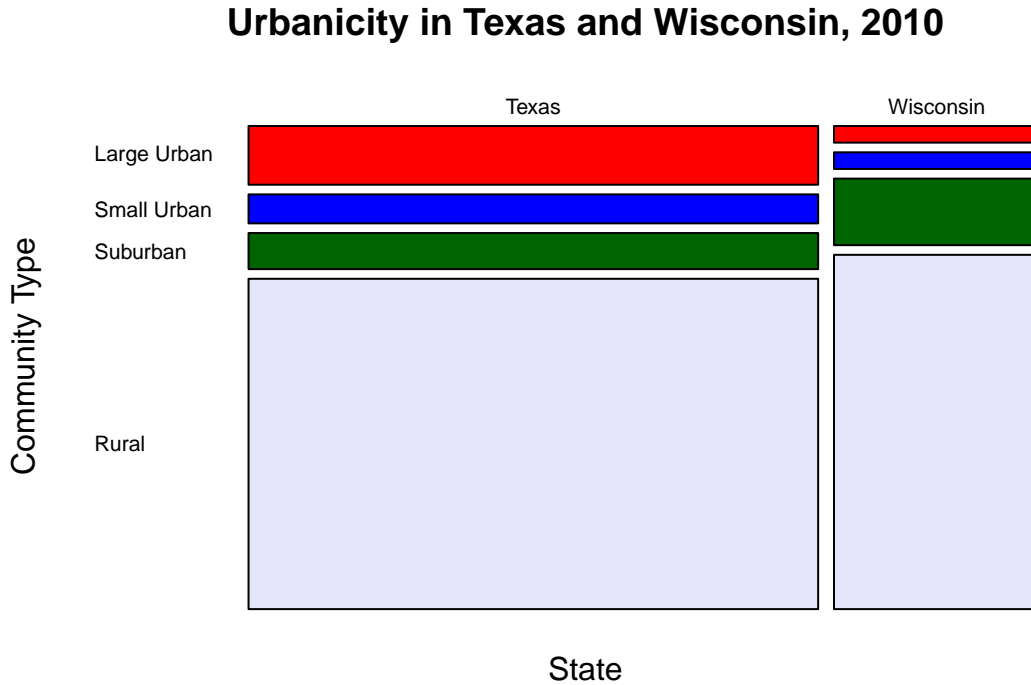


Figure 5: Comparison of the Prevalence of Different Community Types

observed in Texas appears to be driven by differences in the urban landscape between Texas and Wisconsin²¹. This is supported by the overall similarity of magnitudes of transition rates to community types besides Large Urban in the two papers. Figure 5 depicts this difference in landscape by comparing the distribution of community types in Texas and Wisconsin in 2010 (bar widths reflect the relative quantity of districts in Texas and Wisconsin). While both states are majority-rural, the non-rural part of Texas is comparatively urbanized, whereas more than 90% of Wisconsin districts are non-urban.

Returning to Table 2, we see that, as in HKR, the “stickiest” community type is Rural – over 60% of Rural teachers remain Rural in both papers, and even fewer Rural Wisconsin teachers end up in a big city than is the case for Texas. This may reflect the similarity in prevalence of rural districts in the two states and a natural similarity in preferences of rural teachers and districts. Lastly, we also find that the community type transition patterns of younger teachers as compared to all teachers are broadly similar.

²¹We also note a difference (as found in Table 2) in the relative shift in population among community types between the two states – Texas observed dramatic changes in its community type distribution over the period of study of only 4 years, while Wisconsin only saw some movement from Rural to Suburban communities over a longer period of 11 years.

	Men by Experience Class			Women by Experience Class			All Teachers	
	1-3 years	4-6 years	7-11 years	1-3 years	4-6 years	7-11 years	0-9 Years	
Base year salary (log)	-0.001 (0.009)	0.015 (0.011)	0.036 (0.015)	0.001 (0.005)	0.022 (0.007)	0.010 (0.010)	0.009 (0.003)	
Adjusted salary ^a (log)	0.011 (0.007)	0.003 (0.009)	0.024 (0.012)	0.002 (0.004)	0.014 (0.006)	0.015 (0.008)	0.008 (0.003)	
Percent proficient	4.2% (0.7%)	3.0% (0.8%)	2.5% (1.0%)	6.3% (0.4%)	5.7% (0.5%)	5.3% (0.6%)	5.4% (0.2%)	
Percent Hispanic	-1.4% (0.3%)	-0.3% (0.4%)	-0.2% (0.5%)	-1.7% (0.2%)	-1.6% (0.2%)	-1.1% (0.3%)	-1.4% (0.1%)	
Percent black	-5.4% (1.0%)	-2.1% (1.1%)	-3.8% (1.2%)	-8.6% (0.6%)	-6.8% (0.7%)	-6.9% (0.8%)	-7.0% (0.3%)	
Percent subsidized lunch	-7.4% (1.1%)	-3.7% (1.4%)	-4.4% (1.7%)	-9.5% (0.6%)	-7.0% (0.9%)	-7.6% (1.0%)	-7.9% (0.4%)	

Note: a. Adjusted salary is the residual of log salary by district and experience level on 12 regional indicators, three urbanicity indicators, and the district percentages proficient on the WKCE exam, black, Hispanic, and low income.

Table 3: Average Change in Salary and District Student Characteristics (and Standard Deviations) for Teachers Changing Districts, by Gender and Experience

Table 3 replicates Table 3 of HKR, and again confirms its most important insights. Raw salary differentials predict teacher mobility, but the average pay differential is not on average very large – only about \$325, or 1.7% higher than the counterfactually expected wage that would have obtained had the district-switching teacher remained in their current district²². This premium increases with age for both male and female teachers.

Potential Gains from Mobility

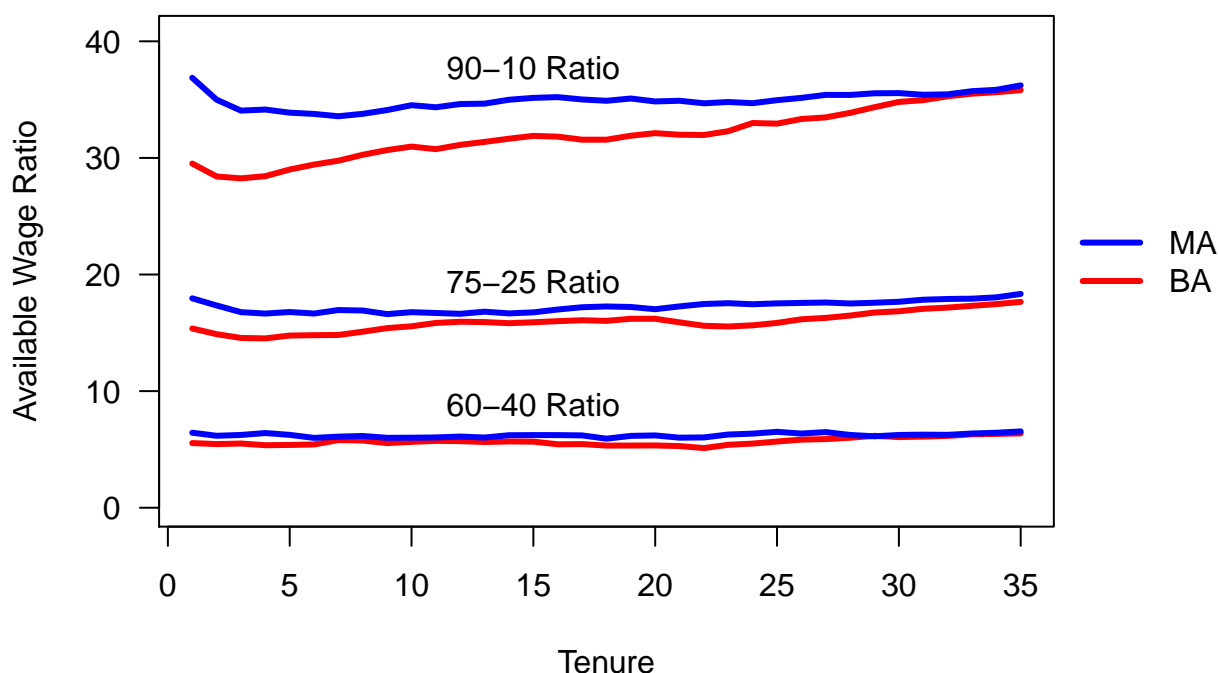


Figure 6: How Much Do Teachers Stand to Gain from Changing Districts throughout Their Careers?

One potential explanation of the weakness of the wage results is that there simply is not sufficient heterogeneity among available contracts to generate mobility incentives. Figure 6 demonstrates that this is not likely the case. No matter their current experience or certification level, a teacher in a district paying the 25th percentile of wages for that experience-certification cell would gain on average 17% by changing to a district at the 75th percentile. Especially for younger teachers, this potential gain would accumulate annually to become a hefty sum over the course of the career – discounting the average annual gain for Master’s-certified

²²There are 777 teachers in the data who skipped one or more years before reappearing at different school or district (perhaps representing leaves of absence for retraining or re-adjustment). For such teachers, the counterfactual subsequent experience and reference curves are taken from their next year in the data, rather than from simply incrementing their experience by one.

teachers at 6% and adding over 20 years, this means roughly \$100,000 is on the table; results are more dramatic for teachers at districts further in the tails of the wage distribution.

Attempting to isolate the influence of district characteristics on wage effects, HKR suggest comparing the differential leverage of residual wages (the residuals being the unexplained part of a Mincer-type regression) to get a more focused estimate of the association between wages and mobility²³. We run a similar regression, but evaluate separate regressions not just for each level of experience, but also for each certification track. This leads to a boost in the overall fraction of explained variance from 60% cited by HKR to 87% here; as in HKR, other included covariates are consistently significant, suggesting their strong independent correlation with salary levels.

Unlike HKR, we find the demographic-independent wage differentials to be no more important than the uncontrolled raw wages, with the predicted wage improvement amounting to 0.8%. In further contrast to HKR, we find a positive relationship between experience and residual wage differentials, with mid-career district switchers experiencing roughly 1.4% higher wages upon arrival to their new employer, by contrast to the null relationship for probationary teachers. This pattern is consistent across the dimension of certification which was ignored by HKR, suggesting the opposite result cannot be attributed to bias introduced by movement patterns of Bachelor's- vs. Master's-certified instructors.

Student demographic differentials are very important for predicting teacher turnover, a finding which held in Texas as it does in Wisconsin. Most distinguished in all experience classes and for both genders are changes in measures of student performance, student poverty and the percentage of black students – district switchers end up at schools with 5% more students at grade level overall, an effect which is stronger for female teachers and for young teachers. They also end up on average with about 8% fewer students (school-wide) eligible for subsidized lunch and 7% fewer black students. While this finding would need to be bolstered with experimental or quasi-experimental evidence, it hints at the potentially limited scope of

²³HKR mention they failed to adjust the standard errors associated with the adjusted wage differentials to account for the fact that they involve residuals from a regression. We explored accounting for this by bootstrapping the regression through resampling teachers and recalculating residuals, but little changes as a result, so we present the naive standard errors for simplicity.

teacher labor market policies intended to ameliorate teacher supply problems in hard-to-serve districts – schools can much more easily exert influence over their compensation policies than they can dictate their student bodies, but the latter appears more efficacious (see Fulbeck 2014 and Glazerman et al. (2013)).

Long-Distance Moves

One major aspect of teacher mobility glossed over by HKR is geographic separation. A wide variety of frictions may be geospatially-related or -generated – social and professional networks tend to be concentrated locally; there are typically substantial fixed costs involved in moving (real estate closing fees, moving expenses, etc.); preferences may depend on climate/geography; and so on. As a first pass at exploring how long-distance moves may differ in nature from those over short distances, we reproduce in Table 4 the analysis of Table 3 for only those moves where the distance between the origin and destination school exceeded 50 miles (a distance deemed sufficient to likely entail a physical move rather than simply an adjusted commute).

The preeminent distinction of long-distance moves is moderation – all average demographic differentials moderate towards zero, suggesting a diminution of the importance of these aspects in this population. The noteworthy exception to this trend is among probationary teachers – young males experience wage increases in an uprooting move, while young females experience declines for long moves. More detailed data would be needed to explore the mechanism at work behind this observation (in particular, none of the differences – male vs. female or short- vs. long-distance moves – have p values below .05), but one explanation is a higher willingness among bachelors to change scenery completely, while younger women may tend to be married and moving with their partners. In any case, the overall importance of wages in long-distance moves is close to zero, suggesting wage differentials are either of secondary or tertiary concern in the associated decision processes, or that there is insufficient heterogeneity in wages at such distances to generate enough moves so motivated, though the case of young male teachers does weaken the latter explanation.

	Men by Experience Class			Women by Experience Class			All Teachers	
	1-3 years	4-6 years	7-11 years	1-3 years	4-6 years	7-11 years	0-9 Years	
Base year salary (log)	0.020 (0.015)	0.012 (0.023)	-0.008 (0.033)	-0.019 (0.012)	0.009 (0.014)	0.009 (0.022)	-0.002 (0.007)	
Adjusted salary (log)	0.016 (0.011)	-0.032 (0.017)	-0.002 (0.031)	-0.005 (0.010)	0.002 (0.010)	0.014 (0.015)	-0.000 (0.005)	
Percent proficient	2.1% (1.0%)	3.1% (1.2%)	1.2% (1.5%)	4.3% (0.6%)	2.9% (1.0%)	4.7% (1.3%)	3.5% (0.4%)	
Percent Hispanic	-0.7% (0.5%)	0.3% (0.6%)	-0.3% (0.8%)	-1.3% (0.3%)	-1.2% (0.4%)	-1.1% (0.6%)	-1.0% (0.2%)	
Percent black	-1.7% (1.3%)	-0.4% (1.2%)	-2.1% (1.6%)	-4.3% (0.9%)	-2.8% (1.3%)	-5.0% (1.8%)	-3.3% (0.5%)	
Percent subsidized lunch	-5.9% (1.6%)	-5.6% (2.0%)	-2.6% (3.0%)	-7.2% (1.0%)	-3.8% (1.5%)	-6.1% (1.9%)	-5.9% (0.6%)	

Table 4: Average Change in Salary and District Student Characteristics (and Standard Deviations) for Teachers Changing to a district more than 50 Miles Away, by Gender and Experience

Supply and Demand for Subject Specialists

Another source of heterogeneity about which HKR have little to say is subject specialty. While it is true that all teachers on a given contract are typically paid independently of the subject they teach, teaching a hard-to-staff subject should lead to more bargaining power in the labor market (as such teachers are less easily replaced), so we would expect such teachers to transition to more attractive positions upon moving. Fully accounting for the demand side of labor markets would bestow higher confidence in results which ultimately depend on the strategic interaction of the two sides.

We are aided in trying to explore this aspect of the teacher labor market by the public availability of annual technical reports from DPI about various aggregate indicators for the health of supply and demand for educators in Wisconsin (the last published edition is Fischer, Swanger, and Skoning 2009). In addition to providing counts for the number of educators graduating from the in-state education programs broken down by subject area, the report uses a survey distributed to district administrators to give a score (based, for example, on the market tightness – applications per vacancy) in each Cooperative Educational Service Agency (CESA, the administrative unit for districts between the school district and DPI) rating the need for educators in various subject areas, including those in our study sample, Math, Reading, and Elementary.

Both Reading and Elementary are chronically over-supplied throughout the state, whereas the demand for math teachers varies considerably. In a given year, the market tightness for the former two subjects is roughly twice that in Math (e.g., it was 67.43 for Elementary, 28.65 for English/Speech/Theater/Journalism, and 24.22 for Mathematics). As a result, we expect to see some heterogeneity in labor market success of specialists in Math as compared to the other teachers in our sample. Table 5 explores some of the basic insights on subject matter heterogeneity. To mitigate the potential for degree holdings to skew results, we focus on Master’s holders and obfuscate gender differences for brevity. Actually, there is little in this table to support the hypothesis that math teachers are given a substantial advantage in the labor market – math teachers earn more (both in nominal and beyond-demographic pay),

	1-3 years		4-6 years		7-11 years		All	
	Non-Math	Math	Non-Math	Math	Non-Math	Math	Non-Math	Math
Base year salary (log)	-0.069 (0.030)	0.031 (0.042)	0.007 (0.019)	0.015 (0.053)	0.025 (0.015)	-0.005 (0.031)	0.001 (0.012)	0.011 (0.022)
Adjusted salary (log)	-0.008 (0.024)	0.068 (0.024)	0.031 (0.014)	0.047 (0.047)	0.034 (0.013)	0.019 (0.017)	0.025 (0.009)	0.041 (0.015)
Percent proficient	10.8% (2.0%)	8.5% (3.4%)	6.2% (1.3%)	1.4% (2.5%)	4.4% (1.0%)	2.2% (1.9%)	6.2% (0.8%)	3.7% (1.5%)
Percent Hispanic	-4.2% (1.0%)	1.4% (1.7%)	-2.1% (0.6%)	-1.1% (1.5%)	-0.8% (0.5%)	-1.4% (1.3%)	-1.9% (0.4%)	-0.6% (0.9%)
Percent black	-15.0% (3.2%)	-11.8% (5.3%)	-6.6% (1.8%)	0.4% (3.3%)	-5.1% (1.2%)	-4.2% (2.5%)	-7.5% (1.1%)	-4.9% (2.1%)
Percent subsidized lunch	-17.6% (3.3%)	-7.3% (5.7%)	-7.9% (2.3%)	-2.9% (5.2%)	-4.9% (1.6%)	-4.8% (3.4%)	-8.4% (1.3%)	-4.9% (2.6%)

Table 5: Average Change in Salary and District Student Characteristics (and Standard Deviations) for Teachers with Master's Degrees Changing Districts, by Subject Area and Experience

but this result is not significant. Further, English teachers are advantaged in ending up at less economically disadvantaged and higher-performing districts.

	District Average Characteristics		Campus Average Characteristics	
	Large Urban to Suburban	Suburban to Suburban	Large Urban to Suburban	Suburban to Suburban
Base year salary (log)	-0.056 (0.012)	0.018 (0.007)	—	—
Adjusted salary (log)	-0.004 (0.005)	0.011 (0.006)	—	—
Average Student Characteristics				
Percent proficient	37.9% (0.6%)	0.9% (0.4%)	35.1% (1.2%)	0.1% (0.6%)
Percent Hispanic	-11.3% (0.4%)	-0.6% (0.2%)	-7.3% (1.3%)	-0.4% (0.2%)
Percent black	-56.9% (0.8%)	-0.6% (0.3%)	-59.7% (1.8%)	-0.5% (0.4%)
Percent subsidized lunch	-55.7% (1.2%)	-1.7% (0.5%)	-61.1% (1.3%)	-1.6% (0.7%)

Table 6: Average Change in Salary and in District and Campus Student Characteristics (and Standard Deviations) for Teachers with 1-10 Years of Experience Who Change Districts, by Community Type of Origin and Destination District

Table 6, which parallels Table 4 of HKR, again uncovers a labor market functioning similar to that in Texas. In particular, while HKR find Large Urban - Suburban district switchers penalize themselves in pay but are rewarded in demographic-adjusted pay, Wisconsin teachers lose out on both measures when leaving Large Urban districts, albeit the residual pay penalty is much lower than that of nominal pay. This difference does not appear to be attributable to HKR's exclusion of certification as a conditioning variable, as the pattern here differs insignificantly by degree.

The other results of HKR are confirmed in even more dramatic fashion. There is strong evidence of selection on the student performance metric, which does vary quite widely in suburban districts. Teachers leaving Milwaukee tend to end up at districts with 38% more students deemed to be at grade level on the state standardized test. On the other hand, teachers leaving Large Urban districts (i.e, Milwaukee) for the suburbs experience a precipitous drop of 57% black students and 56% subsidized lunch eligibility. This is practically

a tautological result, as the student demographics outside of urban areas in Wisconsin are pretty uniformly non-minority – about 90% of suburban districts have fewer than 10% black students, and about 60% have fewer than 2% black students, whereas Milwaukee is about 60% black. Similarly, teachers leaving Milwaukee for the suburbs have little choice but to end up in a district with far fewer economically disadvantaged students – whereas 73% of Milwaukee students are eligible, the median percentage in suburban schools is 12%.

The direction of these effects are preserved among suburban-to-suburban moves, suggesting the importance of these factors even in areas where there is a wider array demographically of destination districts. We also find evidence of selection into economically better-off districts among suburban switchers, but the magnitude of this difference is attenuated with respect to that reported by HKR. We do not find patterns of selection on student performance as strongly as was found in HKR. This may be a reflection of the crudeness of the proficiency measure as compared to the more variable raw scale score measures used by HKR. Lastly, we confirm the finding of HKR that there does not appear to be evidence that teachers are able to select into the more desirable schools within their target districts – the differences in campus-level characteristics are almost identical to the differences in district-level characteristics. This is likely a reflection of supply-side constraints, as the choicest appointments in a district may be awarded to long-serving serving teachers (promotion from within), as well as suburban districts perhaps having only a small number of schools at which to teach a given grade level/subject.

HKR examine the state of Texas, which features substantially more ethnic heterogeneity than does Wisconsin. As a result, they are better-equipped to identify heterogeneity in preferences by teacher ethnicity. In Wisconsin, however, only 2,372 of the 49,325 teachers are non-white, so our results are underpowered relative to HKR. For completeness, Table 7 presents these results, which parallel HKR Table 5. Given how few observations we have of black or Hispanic teachers switching districts, we eschew any temptation to interpret these results. Only black switchers within districts provide enough records to interpret meaningfully. In Wisconsin, we find that, in contrast to white within-district switchers, black teachers tend to migrate to economically better-off and higher-performing schools (white teachers also

	Between District Moves		Within District Moves	
	Black Teachers	Hispanic Teachers	Black Teachers	Hispanic Teachers
Percent proficient	10.7% (3.4%)	8.0% (5.6%)	2.7% (0.9%)	2.2% (1.3%)
Percent Hispanic	3.2% (1.4%)	-14.8% (7.3%)	1.0% (0.9%)	-7.7% (2.3%)
Percent black	-21.1% (5.0%)	-0.6% (5.0%)	-2.1% (1.4%)	-0.3% (2.0%)
Percent subsidized lunch	-19.1% (7.7%)	-15.5% (6.7%)	-3.5% (0.8%)	-4.7% (1.3%)
Number of teachers	81	37	638	228

Table 7: Average Change in District and Campus Student Characteristics (and Standard Deviations) for Black and Hispanic Teachers with 1-10 Years of Experience who Change Campuses

select on percentage of black students). This could simply be a reflection of differences in initial district choice by black vis-à-vis white teachers – the median proficiency at a black teacher’s first district is 36%, compared to 64% for white teachers (71% and 22% for reduced lunch eligibility, respectively).

To the end of examining heterogeneity in the impact of school and district characteristic differentials on teacher mobility, HKR present their Table 6, which breaks down the three exit rates for each (weighted) quartile of the covariate distribution. We replicate that analysis here in Table 8. Saliently, our results for the correlation of school characteristics for within-district movers are qualitatively identical to those found in Texas and similar in magnitude, which gives a stronger indication that we have identified some fundamental nonpecuniary mechanisms driving sorting among schools in a district.

Differences with respect to the results in Texas begin to emerge for the other destinations of school leavers (other districts and other professions). As noted in Table 1, overall rates of switching districts are quite low compared to Texas and national averages; conditional on this, the patterns of movement by quartile of residual salary exhibit a similar pattern to that in Texas, with teachers in the lowest quartile about 28% more likely to change districts than teachers in the highest residual pay quartile. By contrast to HKR, however, who found the opposite association, we find the same trend (at attenuated magnitudes) with respect

Quartile of Distribution	Probability Teachers Move to New School within District	Probability Teachers Move to New District	Probability Teachers Exit Public Schools
<hr/>			
Residual salary			
Highest	—	1.5%	4.1%
3rd	—	1.8%	5.0%
2nd	—	1.8%	4.9%
Lowest	—	1.9%	4.1%
Percent proficient			
Highest	4.5%	1.9%	4.2%
3rd	4.6%	2.3%	4.2%
2nd	5.2%	1.7%	4.4%
Lowest	6.1%	2.1%	4.6%
Percent eligible for reduced-price lunch			
Highest	7.1%	2.1%	5.3%
3rd	5.6%	1.7%	3.8%
2nd	4.1%	2.0%	3.9%
Lowest	3.6%	2.2%	4.4%
Percent Black			
Highest	7.3%	2.1%	5.9%
3rd	4.9%	1.6%	4.2%
2nd	4.7%	1.9%	3.8%
Lowest	3.4%	2.4%	3.4%
Percent Hispanic			
Highest	7.6%	1.8%	5.5%
3rd	4.4%	2.0%	4.1%
2nd	4.3%	2.0%	4.0%
Lowest	4.0%	2.3%	3.7%

Table 8: School Average Transition Rates by Distribution of Residual Teacher Salary and Student Demographic Characteristics (data weighted by number of teachers in school)

to leaving Wisconsin public schools, suggesting salary considerations are also important for teachers considering options outside of public school teaching (or in other states).

We also find fairly strong patterns in quitting associated with subsidized lunch eligibility and with the ethnic makeup of schools, with teachers at the most economically advantaged schools 8% less likely to exit teaching; similar numbers obtain for both the quantity of black and of Hispanic students. For teachers moving within districts, we observe similar patterns.

Regression Results

Having identified some key patterns in moments of the data, we now move on to try and separate the confounding effects of each of these and other factors in affecting teacher turnover with the aim of identifying more fundamentally the association between salient district and school characteristics on teacher turnover. Table 9 provides the main coefficients of interest from a simple linear probability regression model predicting leaving a district (i.e., either switching districts or exiting teaching); this corresponds to HKR Table 7.

By contrast to the strength implied in earlier results, the importance of student achievement has dwindled in the regression specification, and only comes out as independently significant for probationary teachers. The same goes for base salary differentials – in contrast to HKR, the evidence we find in favor of an independent influence of salary on turnover rates is sparse and concentrated among young teachers²⁴. This does not appear to be due to imprecision – the magnitude of HKR’s standard errors follows closely those found for the Wisconsin data, despite our smaller sample sizes.

HKR also found little independent evidence in favor of student economic status factoring in to teachers’ mobility decisions, but in fact this is the source of our strongest effects. As mentioned above, it is possible that the crude nature of the proficiency measure is only weakly identified, and that some of the unaccounted for part of student performance is being

²⁴HKR also mention results not printed in their paper suggesting a paucity of evidence suggesting class size is an important factor in teacher turnover decisions; we give tepid support to this statement, as class size does indeed appear to be related to turnover, but somewhat weakly and only for younger teachers.

	Teacher Experience				
	1-3 years	4-6 years	7-11 years	12-30 years	>30 years
First year base salary (log)	0.03 (0.03)	-0.09** (0.03)	-0.04* (0.02)	0.01 (0.01)	-0.12 (0.06)
First year base salary (log) * female	-0.07* (0.03)	0.09** (0.03)	0.02 (0.02)	-0.02 (0.01)	0.12* (0.06)
Campus average student characteristics					
Percent proficient	-0.10* (0.05)	0.03 (0.04)	-0.02 (0.02)	0.00 (0.01)	-0.07 (0.06)
Percent eligible for subsidized lunch	-0.09** (0.03)	-0.07** (0.03)	-0.07*** (0.02)	-0.01 (0.01)	0.07 (0.04)
Percent Black	0.04 (0.04)	0.22*** (0.04)	0.15*** (0.03)	0.09*** (0.02)	0.14* (0.07)
Percent Hispanic	0.13* (0.06)	0.16*** (0.05)	0.04 (0.03)	-0.04* (0.02)	-0.19* (0.08)
Interactions					
Black * percent Black	-0.21** (0.08)	-0.13 (0.07)	-0.03 (0.05)	-0.03 (0.04)	-0.21 (0.12)
Hispanic * percent Black	-0.19*** (0.06)	-0.19*** (0.05)	-0.12* (0.05)	-0.12** (0.04)	0.02 (0.34)
Black * percent Hispanic	0.14 (0.25)	-0.23 (0.23)	-0.14 (0.14)	0.03 (0.11)	-0.53 (0.45)
Hispanic * percent Hispanic	0.14 (0.28)	-0.07 (0.23)	-0.21 (0.20)	0.26 (0.19)	0.61 (1.07)
Observations	33,108	30,244	43,509	98,753	15,217

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 9: Estimated Effects of Starting Teacher Salary and Student Demographic Characteristics on the Probability that Teachers Leave School Districts, by Experience (linear probability models; Huber-White standard errors in parentheses)

captured in other coefficients, especially subsidized lunch eligibility and student race. Even more compelling would be to associate student performance (and other school/district-level characteristics) more finely with the set of students actually faced by a given teacher.

The results in HKR about the differential effects of student body makeup are largely similar to those we find in Wisconsin. White and nonwhite teachers have opposite and significant correlations between the quantity of minority students in their origin district and their likelihood of leaving it. These results tend to modulate towards zero with experience, regardless of teacher or student race category, and suggest a degree of assortative matching on ethnicity among districts in Wisconsin (though the patterns for whites differ sharply from those of nonwhites, the patterns for black and Hispanic teachers are hard to distinguish).

To account in a rudimentary way for district-specific hiring policies, HKR move on to their Table 8 which repeats Table 7 (our Table 9) with district fixed effects. HKR note that the patterns in responsiveness to wages are the same, though attenuated; that coefficients involving student ethnicity are qualitatively unaffected; and that schools with high achievement continue to exhibit lower propensities for turnover. Our results, presented in Table 10, are similar in that they closely resemble the results without fixed effects, but with noted attenuation and weaker precision.

The most notable difference relative to Table 9 is the weakening of results regarding the importance of student characteristics for white teachers. While partially attributable to a decline in precision, this adjustment suggests much of the discovered correlation between student characteristics and exit probability for white teachers can be chalked up to district-to-district heterogeneity in preferences or hiring policies.

Finally, the conflation of switching districts and exiting teaching may mask important heterogeneity between these two choices. To separate these competing exit risks, HKR construct Table 9, which gives coefficients from a multinomial logit model with three choices – remain in district, switch districts, and exit teaching. We repeat that analysis here in Table 11, with the caveat that, given the sparsity in racial variation present among Wisconsin teachers, we are unable to identify the full model specified by HKR and mirrored above in

	Teacher Experience				
	1-3 years	4-6 years	7-11 years	12-30 years	>30 years
First year base salary (log)	0.01 (0.03)	-0.13** (0.04)	-0.05** (0.02)	0.00 (0.01)	-0.17** (0.07)
First year base salary (log) * female	-0.07* (0.03)	0.09** (0.03)	0.03 (0.02)	-0.02 (0.01)	0.13* (0.06)
Campus average student characteristics					
Percent proficient	-0.12 (0.08)	-0.04 (0.06)	-0.10* (0.04)	0.01 (0.02)	-0.09 (0.10)
Percent eligible for subsidized lunch	-0.07 (0.08)	-0.05 (0.06)	-0.03 (0.04)	-0.02 (0.02)	-0.21* (0.11)
Percent Black	0.29 (0.28)	0.46 (0.26)	0.54** (0.18)	0.03 (0.10)	0.38 (0.48)
Percent Hispanic	0.04 (0.18)	0.04 (0.15)	-0.23* (0.09)	-0.07 (0.05)	0.13 (0.25)
Interactions					
Black * percent Black	-0.19* (0.08)	-0.15* (0.07)	-0.06 (0.05)	-0.04 (0.04)	-0.17 (0.12)
Hispanic * percent Black	-0.17** (0.06)	-0.18*** (0.05)	-0.12* (0.05)	-0.11** (0.04)	-0.09 (0.33)
Black * percent Hispanic	0.06 (0.26)	-0.15 (0.25)	0.02 (0.15)	0.09 (0.11)	-0.73 (0.48)
Hispanic * percent Hispanic	0.14 (0.28)	-0.04 (0.23)	-0.12 (0.21)	0.26 (0.20)	0.26 (1.10)
Observations	33,108	30,244	43,509	98,753	15,217

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 10: Estimated Effects of Starting Teacher Salary and Student Demographic Characteristics on the Probability that Teachers Leave School Districts with District Fixed Effects, by Experience (linear probability models; Huber-White standard errors in parentheses)

	Teacher Experience			
	1-3 years	4-6 years	7-11 years	12-30 years
I. Switch Districts				
First year base salary (log)	0.66 (0.53)	-0.27 (0.74)	-0.81 (0.71)	1.36 (0.91)
First year base salary (log) * female	-1.12* (0.52)	-0.10 (0.73)	-0.36 (0.73)	-1.83* (0.91)
Percent proficient	-1.38* (0.60)	0.16 (0.80)	-1.05 (0.89)	1.20 (1.03)
Percent eligible for subsidized lunch	-0.57 (0.40)	-0.91 (0.57)	-1.98** (0.64)	-0.62 (0.74)
Percent Nonwhite	0.15 (0.40)	2.11*** (0.54)	2.87*** (0.58)	3.19*** (0.66)
Nonwhite * percent Nonwhite	-2.45*** (0.56)	-3.68*** (0.87)	-3.08** (1.04)	0.11 (1.29)
II. Exit Teaching				
First year base salary (log)	0.27 (0.52)	-2.29*** (0.57)	-0.76 (0.64)	0.05 (0.44)
First year base salary (log) * female	-0.63 (0.51)	2.43*** (0.61)	0.73 (0.65)	-0.32 (0.44)
Percent proficient	-0.46 (0.57)	0.42 (0.68)	0.02 (0.71)	0.12 (0.45)
Percent eligible for subsidized lunch	-0.88* (0.39)	-1.50** (0.49)	-2.31*** (0.53)	-0.57 (0.31)
Percent Nonwhite	1.24*** (0.35)	3.17*** (0.42)	2.36*** (0.46)	0.90** (0.29)
Nonwhite * percent Nonwhite	-0.97* (0.42)	-1.75*** (0.45)	-1.37** (0.47)	-0.93* (0.38)

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 11: Multinomial Logit Estimated Effects of Teacher Salary and Student Demographic Characteristics on the Probabilities That Teachers Switch School Districts or Exit Teaching Relative to Remaining in Same District

Tables 9 and 10. In light of this, and in light of the apparent similarity in Wisconsin in the behavior of black and Hispanic teachers described above, we specify the multinomial logit model in terms of a more parsimonious coefficient set. Namely, we distinguish between white and nonwhite teachers and white and nonwhite students (instead of among white, black, and Hispanic students and teachers).

We continue to see little evidence favoring the salience of wage considerations for Wisconsin teachers; the strongest suggestions found here point to the importance of wages for older male teachers in exiting teaching, a result which is generally opposed to that found by HKR in Texas, where salaries were generally important, but only for the propensity to change districts. Also as in the regression specifications, the prominence of student proficiency found by HKR fails to make a notable appearance in Wisconsin.

With respect to the importance of student demographics, our results again point to the same effects found in Texas. White teachers seem to be spurred to change districts or exit teaching by highly black student populations; the reverse is true of nonwhite teachers, who can be drawn to remain in high-minority districts. Subsidized lunch eligibility’s strong effect observed in the combined specification is found here to be concentrated more among those leaving teaching than those changing districts.

Conclusion

That salary incentives appear to play such a limited role in driving teacher churn is bound at first glance to be a disappointment for policymakers. The most powerful predictors of turnover in educators in Wisconsin are all basically beyond the control of administrators, who have no readily-manipulated direct lever for assigning students to schools²⁵. HKR found school quality (as measured by average standardized test performance) to be of key importance for attracting/retaining teachers, but we found no evidence that student proficiency (as measured

²⁵There is evidence (e.g., Richards 2014) that catchment area manipulation (educational gerrymandering) is being used by some schools to select their student populations, but the equilibrium outcome of the strategic interactions of districts competing for the most “desirable” students is far from clear.

by attainment levels on standardized tests) is a factor in the turnover decision for Wisconsin teachers. Regardless, manipulating school performance is famously difficult²⁶, and is in fact the original goal administrators often have in mind when they turn to labor market policies in the first place, so that telling administrators they can improve teacher retention by improving student performance amounts essentially to circular reasoning.

The upside is that this paper is far from settling the debate about welfare-maximizing teacher turnover policies. Limitations in our data prevent us from associating to teachers anything but crude measures of their productivity; measures such as experience, certification, and race are famously poor predictors of teacher quality measures such as value-added. We are thus unable to provide any input to the question of whether *high-quality* teachers have patterns of mobility which resemble that of the teaching population as a whole, or whether heterogeneity in their preferences can be used to devise appropriate policies.

Appendix: Longitudinal Teacher Panel from Unlinked Cross-Sectional Cuts

DPI WISEstaff data does not include a longitudinal identifier for teachers, so we resorted to an alternative approach to matching teachers from year to year. The essence of this algorithm relies on the inclusion of four fields in the DPI data – `first_name`, `last_name`, `nee` (former last name) and `birth_year`. By matching teacher using these identifiers, it is possible to construct with high accuracy a panel of teachers from simple teacher cross-sections²⁷.

²⁶See, for example, the widespread cheating scandals on standardized tests by teachers in Chicago (Jacob and Levitt 2003), Atlanta, and Philadelphia as an example of the lengths professionals feel they need to go to effect change in testing outcomes.

²⁷The code for this process was done using R and especially helped by the `data.table` package (Dowle and Srinivasan 2017). The code to reproduce this entire project can be found at https://github.com/MichaelChirico/wisconsin_teachers. The script for the algorithm is `teacher_match_and_clean.R`. The `README` file gives steps for full reproduction, including retrieving the raw data.

Step 0: Pre-Processing

Prior to beginning the matching process, a number of steps are taken to improve the quality of the raw data. The first is to incorporate as many of the errata mentioned in Public Instruction (2017a) as possible. All name variables are then converted to lower case, after which we extract maiden names (identified for those missing a DPI-supplied entry in `nee` as the part of the `last_name` field that appears in between parentheses or surrounding a hyphen or forward slash). Generally, it appears the maiden name comes in the data *before* the hyphen, but we create the `nee2` field to identify potential matches to the post-hyphen name as well. A search was done of the data for irregular characters (punctuation or numbers) which allowed several obvious typos to be resolved (e.g., `10is a dewey` was easy to resolve as being `lois a dewey`), and this is implemented next.

We then create a “clean” version of the name fields which strips away all whitespace, initials (lone letters), and punctuation. At this stage, all observations which identically match another in the same year from the viewpoint of the algorithm – namely, those that match exactly another observation on the cleaned first and last names and year of birth – are removed from the data since it would be impossible to tell such teachers apart. A more ambitious treatment would attempt to use other cues found in the duplicated records (ethnicity, subject/position cues, etc.) to separate such teachers, but the marginal cost of doing so was found to exceed the potential benefit considerably for the exercise at hand (recall that only 0.7% of total observations are lost in this fashion). Finally, teachers in the first year of data (1994-95) are assigned an ID starting from 1 using the within-year identifier provided by DPI.

Steps 1-21: Matching

The algorithm proceeds by iterating over years of the data. In each year Y , matches are found serially by progressively adjusting the criteria for considering two observations to be from the same teacher as follows:

1. Match anyone who stayed in the same school – i.e., match any teacher found in a

- year $Y' < Y$ with the same `first_name`, `last_name`, and `birth_year` at the same `district` and `school`.
2. Find within-district switchers – those who match on all but the `school` field from Step 1.
 3. Find out-of-district switchers – those who match on all but the `district` field from Step 2.
 4. Find teachers that appear to have been married, but have not moved – their `nee` field in Y is matched to the `last_name` fields in $Y' < Y$, but otherwise the fields from Step 1 are all matched. We create a flag for teachers matched in this fashion called `married`.
 5. Repeat Step 2 for those who appear to have married.
 6. Repeat Step 3 for those who appear to have married. 7-9. Repeat Steps 4-6 using the `nee2` field instead of `nee`. 10-18. Repeat Steps 1-9 using the “cleaned” version of the first name field that had non-alphabetic characters removed, `first_name_clean`. We create a flag for teachers matched in this fashion called `mismatch_inits`.
 7. Match individuals in the same school assigned to the same position (identified in `position_code`) but with different years of birth to overcome potential noise in year of birth (most commonly, the year of birth is missing in some years). We create a flag for teachers matched in this fashion called `mismatch_yob`.
 8. Match individuals in the same *district* assigned to the same position but with different years of birth. We do not extend this logic to find district switchers since the potential for erroneous match assignment in such a case is too great, and we neglect to extend the algorithm to use other cues from the data to facilitate matching in such cases.
 9. Assign new IDs to all teachers in Y not matched in the first 20 steps, incrementing from the highest ID recorded thus far.

To help ensure we are matching to the most important observation of each teacher, matching is always done to a teacher’s highest-FTE observation within a year (particularly important for Steps 19-20). Further, it is sometimes the case that a given tuple of search keys matches more than one teacher in the prior data; if so, these rows are simply ignored for that step and such a teacher will go unmatched unless they are uniquely pinned down in a subsequent step.

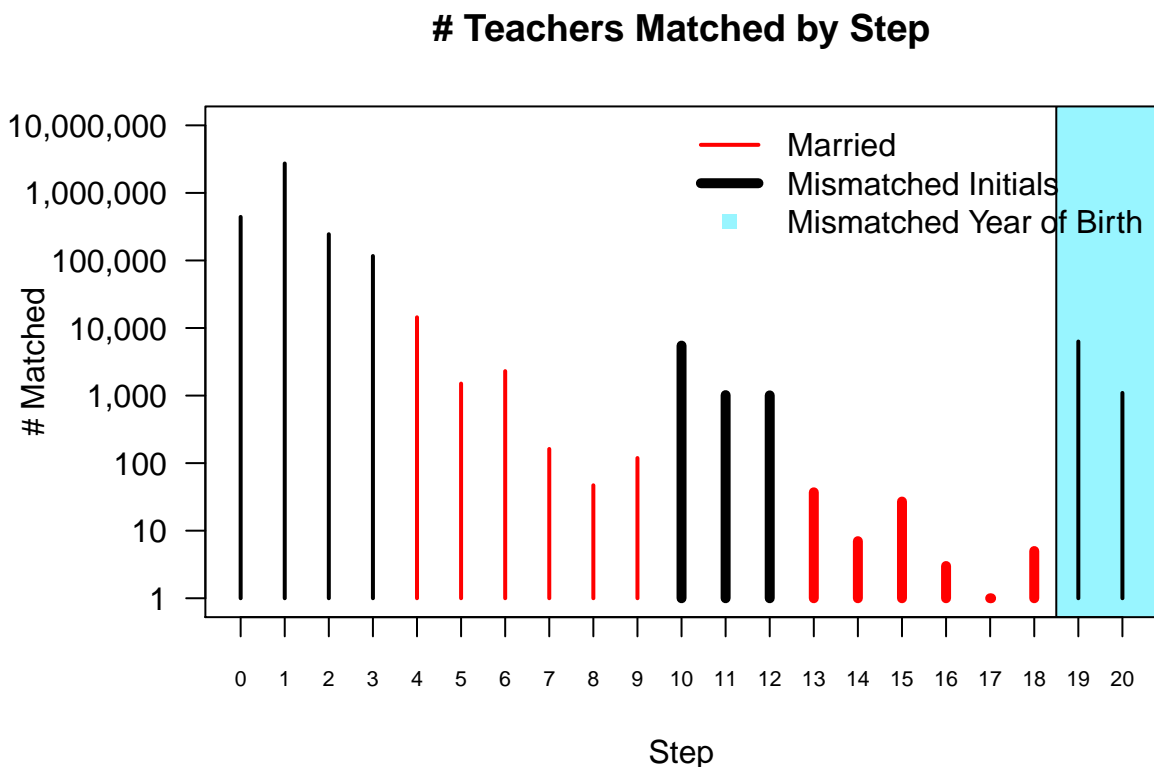


Figure 7: Frequency of Matching by Step

Figure 7 Shows that teachers are most commonly matched in the first three steps, meaning data fidelity issues are not *per se* devastating. The real benefits of the algorithm are in the subsequent steps, as a result of which an additional 33,538 teachers are matched than would have been if only the original first and last name fields as found in the raw data were used.

Step 22: Post-Processing

The panels implied by the matched IDs created by the algorithm still have a substantial amount of data quality issues which we can only address once teachers' multiple observations are associated, mainly having to do with instability in certain observable characteristics which should be constant over time. First, we cascade forward maiden names (if a teacher has non-missing `nee` in a period Y and it becomes missing in $Y' > Y$, we replace it in Y' with its value in Y); the same is done for the certification field, `highest_degree` (just as a teacher cannot erase marriage from their past, so can they not make a degree disappear).

Next, we correct instability in the `ethnicity` (and `gender`) fields when possible according to three steps: 1) it is sometimes missing, in which case we simply overwrite it with the other values found for that teacher; 2) at least 70% of a teacher’s observations use the same ethnicity (or gender); or 3) there are at least five people that share a last name with an ethnicity-ambiguous teacher, at least 70% of whom have one ethnicity (or gender), the idea being that names like Xu or Gutierrez are strongly associated with a particular ethnicity. This type of correction is uncommon enough not to warrant an appeal to a more sophisticated approach commonly found in natural language processing applications, e.g. training a classifier such as a random forest (Breiman 2001) to predict ethnicity as accurately as possible.

Lastly, we synergize the year of birth field for those matched in Steps 19 or 20 by assigning the one that appears most frequently for each teacher; in the case of ties, we use a regression-to-the-mean-type logic and assign the year which brings the teacher closer to the median age observed in the data. More data-driven approaches (conditioning the target median on the teacher’s employer, position, year in the data, or using social security data to determine the maximum-likelihood year of birth for a given first name, etc.) were again eschewed for expediency.

Validity Check: `file_number`

Starting in the 2011-12 release, the DPI data begins to consistently record a field called `file_number` for teachers which generally acts as a time-consistent ID (from verbiage gleaned from Public Instruction 2017a, it appears this corresponds to a teaching license number). We looked for instances of multiple file numbers and are content that the algorithm is performing well – only 78 teacher IDs were found to be associated with more than one `file_number`, with almost all of them having been matched on Steps 1 - 3 (what should be the highest-accuracy steps). Given a number of apparent transcription mistakes (i.e., `file_number` differing by one digit in some years) and that the `file_number` does appear to change on occasion, even these 78 could be an overstatement of the number of incorrectly matched individuals.

References

- Anzia, Sarah F, and Terry M Moe. 2014. "Collective Bargaining, Transfer Rights, and Disadvantaged Schools." *Educational Evaluation and Policy Analysis* 36 (1). SAGE: 83–111.
- Ballou, Dale, and Michael Podgursky. 2002. "Returns to Seniority Among Public School Teachers." *Journal of Human Resources*. University of Wisconsin Press, 892–912.
- Biasi, Barbara. 2017. "Unions, Salaries, and the Market for Teachers: Evidence from Wisconsin." SSRN. doi:<http://dx.doi.org/10.2139/ssrn.2942134>.
- Boyd, Donald J, Pamela L Grossman, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2009. "Teacher Preparation and Student Achievement." *Educational Evaluation and Policy Analysis* 31 (4). SAGE Publications Sage CA: Los Angeles, CA: 416–40.
- Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2005. "Explaining the Short Careers of High-Achieving Teachers in Schools with Low-Performing Students." *The American Economic Review* 95 (2). American Economic Association: 166–71.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1). Springer: 5–32.
- Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *The American Economic Review* 104 (9). American Economic Association: 2593–2632.
- . 2014b. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *The American Economic Review* 104 (9). American Economic Association: 2633–79.
- Cohen-Vogel, Lora, Li Feng, and La'Tara Osborne-Lampkin. 2013. "Seniority Provisions in Collective Bargaining Agreements and the 'Teacher Quality Gap'." *Educational Evaluation and Policy Analysis* 35 (3). SAGE: 324–43.
- Croissant, Yves. 2012. "Estimation of Multinomial Logit Models in R: The Mlogit Package." *R*

- Package Version 0.2-2*. <http://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf>.
- Dahl, David B. 2009. "Xtable: Export Tables to LaTeX or HTML." *R Package Version 1.8.2*, 1–5.
- de Boor, Carl. 1978. *A Practical Guide to Splines*. Vol. 27. Springer-Verlag New York.
- Dolton, Peter, and Wilbert Van der Klaauw. 1999. "The Turnover of Teachers: A Competing Risks Explanation." *Review of Economics and Statistics* 81 (3). MIT Press: 543–50.
- Dowle, Matt, and Arun Srinivasan. 2017. *Data.table: Extension of 'Data.frame'*. <http://r-datatable.com>.
- Engel, Mimi, Brian A Jacob, and F Chris Curran. 2014. "New Evidence on Teacher Labor Supply." *American Educational Research Journal* 51 (1). SAGE: 36–72.
- Fischer, Thomas A, Wayne H Swanger, and Stacey Skoning. 2009. "Supply & Demand 2008: Data Trends of Education Personnel in Wisconsin Public Schools." Wisconsin Educator Supply; Demand Project.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1. Springer.
- Fulbeck, Eleanor S. 2014. "Teacher Mobility and Financial Incentives: A Descriptive Analysis of Denver's Procomp." *Educational Evaluation and Policy Analysis* 36 (1). SAGE: 67–82.
- Glazerman, Steven, Ali Protik, Bing-Ru Teh, Julie Bruch, and Jeffrey Max. 2013. "Transfer Incentives for High-Performing Teachers: Final Results from a Multisite Randomized Experiment." *National Center for Education Evaluation and Regional Assistance*. ERIC.
- Goldhaber, Dan, Betheny Gross, and Daniel Player. 2007. "Are Public Schools Really Losing Their Best? Assessing the Career Transitions of Teachers and Their Implications for the Quality of the Teacher Workforce." *National Center for Analysis of Longitudinal Data in*

Education Research. CALDER.

Goldhaber, Dan, Lesley Lavery, and Roddy Theobald. 2015. “Uneven Playing Field? Assessing the Teacher Quality Gap Between Advantaged and Disadvantaged Students.” *Educational Researcher* 44 (5). SAGE: 293–307.

Green, Peter J, and Bernard W Silverman. 1993. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. CRC Press.

Hanushek, Eric A, and Steven G Rivkin. 2010. “Constrained Job Matching: Does Teacher Job Search Harm Disadvantaged Urban Schools?” National Bureau of Economic Research.

Hanushek, Eric A, John F Kain, and Steven G Rivkin. 2004. “Why Public Schools Lose Teachers.” *Journal of Human Resources* 39 (2). University of Wisconsin Press: 326–54.

Harris, Douglas N, and Tim R Sass. 2011. “Teacher Training, Teacher Quality and Student Achievement.” *Journal of Public Economics* 95 (7). Elsevier: 798–812.

He, Xuming, and Pin Ng. 1999. “COBS: Qualitatively Constrained Smoothing via Linear Programming.” *Computational Statistics* 14 (3). Springer: 315–38.

Heckman, James J, Lance J Lochner, and Petra E Todd. 2003. “Fifty Years of Mincer Earnings Regressions.” National Bureau of Economic Research.

Henningsen, Arne, and Ott Toomet. 2011. “MaxLik: A Package for Maximum Likelihood Estimation in R.” *Computational Statistics* 26 (3): 443–58. doi:10.1007/s00180-010-0217-1.

Jacob, Brian A, and Steven D Levitt. 2003. “Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating.” *The Quarterly Journal of Economics* 118 (3). Oxford University Press: 843–77.

Koski, William S, and Eileen L Horng. 2007. “Facilitating the Teacher Quality Gap? Collective Bargaining Agreements, Teacher Hiring and Transfer Rules, and Teacher Assignment

Among Schools in California.” *Education* 2 (3). MIT Press: 262–300.

Leifeld, Philip. 2013. “texreg: Conversion of Statistical Model Output in R to LaTeX and HTML Tables.” *Journal of Statistical Software* 55 (8): 1–24. <http://www.jstatsoft.org/v55/i08/>.

Litten, Andrew. 2016. “The Effects of Public Unions on Compensation: Evidence from Wisconsin.” https://drive.google.com/file/d/0BwL-PvOgW_Ordzk5QXloZGM1d1k/view.

Loeb, Susanna, and Marianne E Page. 2000. “Examining the Link Between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-Pecuniary Variation.” *Review of Economics and Statistics* 82 (3). MIT Press: 393–408.

Moe, Terry M. 2006. “Bottom-up Structure: Collective Bargaining, Transfer Rights, and the Plight of Disadvantaged Schools.” *Education Working Paper Archive*. ERIC.

Murnane, Richard J, and Randall J Olsen. 1990. “The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching: Evidence from North Carolina.” *Journal of Human Resources*. University of Wisconsin Press, 106–24.

Ng, Pin, and Martin Maechler. 2007. “A Fast and Efficient Implementation of Qualitatively Constrained Quantile Smoothing Splines.” *Statistical Modelling* 7 (4). SAGE: 315–28.

Public Instruction, Wisconsin Department of. 2011. “School District Name Changes - Since 1994.” dpi.wi.gov/sites/default/files/imce/sms/doc/rg_sdnamechanges.doc.

———. 2017a. “Data Errata - Data Changes After Dpi Publication.” <https://dpi.wi.gov/cst/data-collections/data-errata>.

———. 2017b. “School Staff: Salary, Position & Demographic Reports.” <https://dpi.wi.gov/cst/data-collections/staff/published-data>.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna,

- Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richards, Meredith P. 2014. "The Gerrymandering of School Attendance Zones and the Segregation of Public Schools: A Geospatial Analysis." *American Educational Research Journal* 51 (6). SAGE: 1119–57.
- Rivkin, Steven G, Eric A Hanushek, and John F Kain. 2005. "Teachers, Schools, and Academic Achievement." *Econometrica* 73 (2). Journal of the Econometric Society: 417–58.
- Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *The American Economic Review* 94 (2). JSTOR: 247–52.
- RStudio Team. 2017. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>.
- Runge, Carl. 1901. "Über Empirische Funktionen Und Die Interpolation Zwischen Äquidistanten Ordinaten." *Zeitschrift Für Mathematik Und Physik* 46 (224-243): 20.
- Sable, Jennifer. 2009. "Documentation to the Nces Common Core of Data Local Education Agency Universe Survey: School Year 2006-07 (Nces 2009-301)." *U.S. Department of Education*. Washington, DC: National Center for Education Statistics.
- Scafidi, Benjamin, David L Sjoquist, and Todd R Stinebrickner. 2007. "Race, Poverty, and Teacher Mobility." *Economics of Education Review* 26 (2). Elsevier: 145–59.
- Stinebrickner, Todd R. 2002. "An Analysis of Occupational Change and Departure from the Labor Force: Evidence of the Reasons That Teachers Leave." *Journal of Human Resources*. University of Wisconsin Press, 192–216.
- Stone, Mervyn. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions." *Journal of the Royal Statistical Society. Series B (Methodological)*. JSTOR, 111–47.
- Wahba, Grace. 1990. *Spline Models for Observational Data*. SIAM.
- Wayne, Andrew J, Kwang Suk Yoon, Pei Zhu, Stephanie Cronen, and Michael S Garet. 2008.

“Experimenting with Teacher Professional Development: Motives and Methods.” *Educational Researcher* 37 (8). SAGE: 469–79.

Xie, Yihui. 2016. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <http://yihui.name/knitr/>.

Zeileis, Achim. 2004. “Econometric Computing with Hc and Hac Covariance Matrix Estimators.” *Journal of Statistical Software* 11 (10): 1–17. <http://www.jstatsoft.org/v11/i10/>.

———. 2006. “Object-Oriented Computation of Sandwich Estimators.” *Journal of Statistical Software* 16 (9): 1–16. <http://www.jstatsoft.org/v16/i09/>.

Zeileis, Achim, and Torsten Hothorn. 2002. “Diagnostic Checking in Regression Relationships.” *R News* 2 (3): 7–10. <https://CRAN.R-project.org/doc/Rnews/>.