Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage*

Barbara Biasi, Chao Fu and John Stromme†

February, 2021

Abstract

We study the equity-efficiency implication of giving school districts control over teacher pay using an equilibrium model of the market for public-school teachers. Teachers differ in their comparative advantages in teaching low- or high-achieving students. School districts, which serve different student bodies, use *both* wage *and* hiring strategies to compete for their preferred teachers. We estimate the model using data from Wisconsin, where districts gained control over teacher pay in 2011. We find that, all else equal, giving districts control over teacher pay would lead to more efficient teacher-district sorting but larger educational inequality. Teacher bonus programs that incentivize comparative advantage-based sorting, combined with bonus rates favoring districts with more low-achieving students, could improve both efficiency and equity.

JEL Classification: I20, J31, J45, J51, J61, J63

Keywords: Teachers Salaries, Teacher-District Sorting, Education Efficiency and Equity, Market Equilibrium, Structural Estimation

*We thank seminar and conference participants at Bocconi University, The University of Chicago, Johns Hopkins University, Toulouse School of Economics, Tsinghua University, Yale University, AEA, and SEA for helpful comments. All errors are ours.

†Biasi: Yale School of Management and NBER; Fu: University of Wisconsin and NBER, cfu@ssc.wisc.edu; Stromme: University of Wisconsin.

1 Introduction

Education, as a production process, requires a substantial amount of interaction between a teacher and their students. Students with different learning abilities and needs may disagree on who the best teacher is, because some teachers may be better at stimulating high-achieving students, while others at helping low-achieving students. Therefore, it would be most efficient if teachers sort into teaching students according to teachers' comparative advantages.

Unfortunately, the typical salary structure for U.S. public school teachers fails to incentivize such sorting. Salaries follow rigid experience-education schedules set via collective bargaining; despite differences in the composition of students they serve, districts cannot use salary schemes to attract teachers better suited for their students. We label this regime the "rigid-pay regime." Associated with pay rigidity, teacher-district sorting often exhibits a vertical pattern, where teachers deemed better by various measures tend to teach in districts with more advantaged students.[1] Such vertical sorting can lead to both efficiency losses and large inequalities across children from different backgrounds.[2]

An alternative arrangement would be one where districts have the flexibility to design their own teacher pay schedules, which we label as the "flexible-pay regime."[3] This paper investigates the implication of the flexible-pay arrangement in a market equilibrium setting, where districts compete for their preferred teachers, and explores counterfactual policies to improve educational efficiency and equity.

To achieve this goal, we need a solid understanding of several key factors. The first is teachers' preferences over non-pecuniary aspects of their jobs (e.g., student composition) relative to monetary compensation, which govern how effectively teacher pay schemes can incentivize sorting. The second is school districts' preferences over various attributes of a teacher, which govern their hiring decisions, and if given the flexibility, their choices of teacher pay schedules. A third factor is the competition among districts for teachers, which needs to be accounted for when evaluating major policy reforms. Holding everything else fixed, a district will always be weakly better off with more flexibility. However, when equilibrium responses by all districts are taken into account, some districts may be worse off in the flexible-pay regime than they are in the rigid-pay regime.

An obstacle to understanding these factors is the lack of both flexibility and variation in teacher pay schedules. Due to pattern bargaining by a state's teachers' union, very similar and

[1]See, for example, Lankford et al. (2002); Ingersoll (2004); Clotfelter et al. (2005); Mansfield (2015); Jacob (2007).

[2]Some studies have found significant short-run and long-run effects of teachers on students, e.g., Chetty et al. (2014); Rivkin et al. (2005); Jackson (2018).

[3]Throughout the paper, flexible pay refers to a regime in which districts can choose their own teacher pay schemes; it does not necessarily mean that all districts will choose to reward teacher effectiveness in the equilibrium.

rigid pay schedules are often imposed on all districts in the state, which has made it difficult to infer districts' preferences, let alone how they would choose teacher pay if allowed to do so. A real-life exception provides us with an opportunity to gain more insight: In 2011, Wisconsin passed a law known as Act 10, which discontinued collective bargaining over teacher salaries and gave districts full autonomy over teacher pay.

Using post-Act 10 Wisconsin as a platform, we build and estimate an equilibrium model of the labor market for public school teachers. Teachers differ in their two-dimensional effectiveness in teaching low- and high-achieving students. A teacher cares about their wage and the characteristics of the district they work in, including its student composition (the fraction of students with low/high prior achievement).[4] A district cares about a teacher's contribution to its students' achievement, and may also care directly about a teacher's experience and education. Given its budget, the goal of a district is to fill its capacity with teachers it prefers the most, by setting a wage schedule and extending job offers. In particular, a wage schedule specifies how teachers are rewarded for their contribution to the district's student achievement, and for their experience and education. Districts simultaneously make wage and hiring decisions, given their beliefs about the probabilities of acceptance by different teachers and how these probabilities vary with their own wage offers. Among offers received, a teacher chooses their most preferred district, net of moving costs. An equilibrium requires districts' beliefs be consistent with decisions by all districts and teachers.

This model highlights a major trade-off embedded in a flexible-pay regime. On the one hand, given that student composition differs across districts, and teachers differ in their comparative advantages in teaching certain types of students, teacher-district matching is not necessarily a zero-sum game. Giving districts the flexibility to directly reward teacher contribution may encourage comparative advantage-based sorting and hence improve efficiency. On the other hand, districts make choices to maximize their own objectives without concerns about overall efficiency. With teacher pay at their disposal, advantaged districts may find it easier to attract teachers with absolute advantages in teaching both types of students, which would weaken comparative advantage-based sorting and exacerbate cross-district inequality. When this second force is strong, policy interventions favoring disadvantaged districts can be justified on grounds of both equity *and* efficiency.

To quantify the trade-off mentioned above and to design policy interventions, we first need to estimate our model and tackle a major identification challenge: The researcher observes only the accepted offers, making it hard to separate teacher preferences from district preferences. We estimate our model via indirect inference, where we design auxiliary models based on the following identification argument. First, with the mild assumption that district preferences for

[4]Throughout the paper, we use the words pay, wage and salary interchangeably.

teachers are weakly increasing in teacher attributes (experience, education, and effectiveness), we can infer from an observed match of district-teacher pair (d, i) that teachers who are weakly better in all attributes and weakly cheaper than i must have been eligible for a position in d. This observation allows us to infer a subset of feasible options each teacher must have faced. Teachers' observed choices among these options inform us of teacher preferences. In contrast, if one were to assume that all teachers had offers from all districts, the inferred "preferences" would be different. The discrepancy between the two sets of inferred preferences arises because certain districts did not make offers to certain teachers; this informs us of district preferences. Furthermore, given that districts have control over teacher pay, we can learn about district preferences from the degree to which a district's observed pay schedule favors or disfavors certain teacher groups and how it varies with district characteristics.[5]

We apply our model to administrative data from the Wisconsin Department of Public Instruction, consisting of three linked panel data sets of at the student, teacher, and district level. Extending the traditional value-added model, we define and estimate a teacher's two-dimensional effectiveness as their value added to test scores of their students with low and high prior scores, respectively. The data also allow us to track a teacher's employment history within the state's public school system, including their salaries and job characteristics. Our data cover eras both before and after Act 10. We use post-Act 10 data to estimate our model. With the estimated parameters, we validate the model by simulating the pre-Act 10 equilibrium under rigid pay and contrasting it with pre-Act 10 data. The model fits the data well in both eras.

Using the estimated model, we first examine the implication of giving districts control over teacher pay. Compared to the rigid-pay equilibrium, under th*e same initial market conditions*,[6] the flexible-pay equilibrium features more efficient teacher-district matching with a 0.08% improvement in overall student achievement. However, it enlarges the achievement gap between low- and high-achieving students and reduces student achievement in districts with higher fractions of low-achieving students.

These findings suggest that, under flexible pay, there is room for policy interventions favoring districts with more low-achieving students. We design a series of counterfactual state-funded teacher bonus programs under two bonus formulae (B1 and B2) to account for *both* efficiency *and* equity. Specifically, under both formulae, a teacher's bonus from the state is proportional to their contribution to student achievement, which may incentivize more efficient sorting because a teacher's contribution (and thus bonus) is higher when their comparative advantage better matches with a district's student composition. To account for equity, we adjust bonus rates

[5]For example, an experienced but ineffective teacher would prefer wages to be based mostly on experience rather than effectiveness, while a young but effective teacher would prefer the opposite.

[6]A comparison between pre-Act 10 data and post-Act 10 data is contaminated with differences in initial market conditions.

based on a district's student composition. Relative to B1, B2 has an additional feature that ties the state bonus to a district's wage schedule such that districts are incentivized to increase their own reward for teacher contribution. Our simulations show that at the same total cost, B2 programs with progressive bonus rates favoring districts with more low-achieving students would benefit both low- and high-achieving students, and narrow the achievement gap between the two groups. Moreover, student achievement improves more in districts with higher fractions of low-achieving students than it does in an average district. In summary, under flexible pay, carefully-designed intervention can induce more efficient and more equal teacher-district sorting.

Our paper contributes to the extensive body of work on the labor market for teachers. Most related to ours are papers studying this market via the lens of a structural model. A large subset of these studies focuses on the supply side. For example, Stinebrickner (2001a,b); Wiswall (2007); Lang and Palacios (2018) model individuals' dynamic choices between teaching and non-teaching options. Behrman et al. (2016) further break down the teaching option into teaching in one of three types of schools. Boyd et al. (2005) and Scafidi et al. (2007) use competing risks models to study teachers' preferences for schools and find that teachers prefer schools with fewer low-achieving and fewer minority students.[7]

A smaller subset of work considers both sides of the market. Boyd et al. (2013) estimate a two-sided matching model to disentangle teacher and school preferences, assuming that the observed teacher-school matches are stable. While Boyd et al. (2013) study a context of the rigid-pay regime, districts in our setting have control over teacher pay. We therefore explicitly model the competition among districts, which choose both wage and hiring strategies. Tincani (2021) estimates an equilibrium model where a representative private school sets teacher wages and tuition; workers choose among teaching in the public school (which is passive in her model), teaching in the private school, and non-teaching; and households choose between public and private schools.[8] Our paper and Tincani (2021) well complement each other. Tincani (2021) focuses on how a given wage function for public school teachers would induce reactions from the private school and affect teachers' and households' choices between public and private sectors. We are interested in efficiency and equity within the public sector, and we study how public school districts use wage and hiring strategies to compete with one another for better teachers.

Our paper also contributes to the literature on the effect of teacher pay on teachers' behavior and student outcomes (see Neal et al., 2011; Jackson et al., 2014, for reviews), and more specifically on teachers' mobility and educational inequality. Hanushek et al. (2004) find that teacher mobility is more related to student composition than salary, but salary has a modest impact. Some studies suggest that financial incentives can attract and retain teachers in dis-

[7]Dolton and Klaauw (1999) use a competing risk model to study teachers' decision to leave the profession.

[8]Mehta (2017) estimates an equilibrium model of charter school entry, school input choices, and student school choices.

advantaged schools (e.g., Clotfelter et al., 2008; Steele et al., 2010; Feng and Sass, 2018), while some other studies find little or no effect (e.g., Clotfelter et al., 2008, 2011; Russell, 2020). Using a randomized controlled trial in Pakistan, Brown and Andrabi (2020) find that performance pay induces positive teacher sorting. Similarly, Biasi (2020) shows that under Wisconsin Act 10, higher-quality teachers tend to move to districts that adopted flexible pay. Motivated by findings from Biasi (2020), we develop and estimate an equilibrium model to understand forces underlying the observed outcomes, which in turn allows us to study how counterfactual policies affect districts' wage and hiring decisions and equilibrium teacher-district matches.

Unlike the studies mentioned above, we allow for multi-dimensional teacher effectiveness in teaching different types of students, which leaves open the possibility that changing teacher-district sorting can improve both equity and efficiency. This consideration is supported by previous findings that teacher effectiveness might be specific to student composition. For example, Jackson (2013) demonstrates the importance of match quality between teachers and schools. Aucejo et al. (2019) and Graham et al. (2020) find significant complementarities between teachers and classroom composition, and show that reassigning teachers across classrooms would have sizable effects on teachers' contribution to learning.[9] Recognizing the importance of student-teacher match quality, we explore how education policies can affect districts' and teachers' decisions and thereby *induce* more efficient sorting in the equilibrium.

The rest of the paper is organized as follows. Section 2 describes the background; Section 3 describes the model; Section 4 explains our estimation strategy; Section 5 describes the data; Section 6 reports the estimation results; Section 7 conducts counterfactual experiments; and Section 8 concludes. Additional details are in appendices.

2 Background

Most US public school districts pay teachers according to "steps-and-lanes" schedules, which express a teacher's salary as a function of their experience and education (Podgursky, 2006). Movements along the "steps" (experience levels) and "lanes" (education degrees) of a schedule involve an increase in pay. In states with collective bargaining (CB), these schedules are negotiated between school districts and teachers' unions.[10] CB agreements usually prevent districts from adjusting pay at the individual level, which implies that pay is rigid and does not reward teachers for their effectiveness (Podgursky, 2006). Wisconsin introduced CB for public-sector employees in 1959 (Moe, 2013). Since then, teachers' unions have gained considerable power

[9]Other recent studies have considered heterogeneity in teacher effectiveness by student background characteristics (Lavy, 2016; Fox, 2016).

[10]Most states use CB; in states without CB these schedules are typically determined at the state level (e.g. Georgia). See, e.g., https://www.nctq.org/contract-database/collectiveBargaining

and have been involved in negotiations with school districts over key aspects of a teaching job. Until 2011, unions negotiated all teacher salary schedules, which were included in each district's CB agreement.

Facing a projected budget deficit of $3.6 billion, on June 29, 2011, the Wisconsin state legislature passed the Budget Repair Bill, also known as Act 10, which led to major reforms to public-sector employment in the state.[11] For public-school teachers, the most dramatic change was the exclusion of salary schedules from union negotiations. Under Act 10, unions are only allowed to negotiate base salaries (i.e., the starting pay for new employees), the annual growth rate of which is capped at the rate of inflation. Above and beyond base salaries, school districts are allowed ample flexibility to design teacher pay.[12]

Act 10 also introduced a series of other provisions, applied uniformly to all school districts in the state. First, Act 10 reduced employees' benefits reduced via an increase in employee contributions to pension and healthcare. Second, Act 10 made it harder for teachers' unions to operate: They are prohibited from automatically collecting dues from employees' paychecks and are required to re-certify annually with the majority of votes of all members. As a result, union membership dropped by nearly 50% in the 5 years after the passage of Act 10.

2.1 A Glance at the Market Before and After Act 10

We provide a first glance at the labor market for public school teachers in Wisconsin before and after Act 10, using data from the Wisconsin Department of Public Instruction. The data, which we describe in detail in Section 5, consist of three linked data sets at the teacher, student, and district level, respectively.

Variation in Teacher Salaries: Figure 1 shows that, prior to Act 10, teacher wage variation, as measured by the coefficient of variation (CV), was almost nonexistent within each district among teachers with similar experience and education.[13] After Act 10, wage variation increased as districts gained control over pay and could reward teachers directly for their effectiveness.

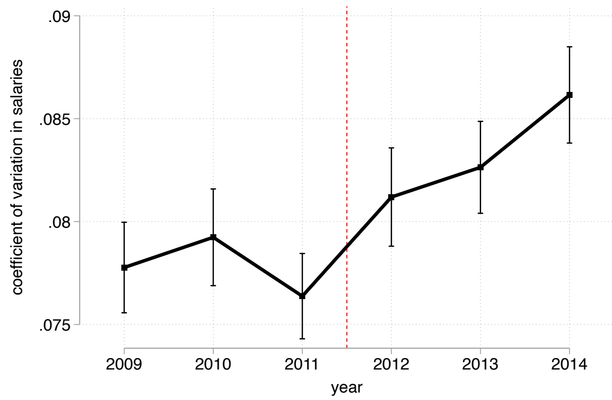
Teacher Mobility: Figure 2a shows that movements of teachers across districts are rare, but their frequency, i.e., the fraction of teachers employed in a district other than the one they worked for in the previous year, more than doubled after Act 10. Figure 2b compares the wage growth of movers relative to stayers both pre- and post-Act 10, controlling for teacher and

[11]On July 1, 2011 the state legislature also passed Act 32, which reduced state aid to school districts and decreased districts' revenue limits (the maximum revenue a district can raise through general state aid and local property taxes).

[12]For example, the 2015 employee handbook of the Mequon-Thiensville District states that "The District, in its sole discretion, may place employees at a salary it deems appropriate."

[13]We run a regression of wages on district-by-year and seniority-by-education fixed effects, from which we obtain the standard deviation of wage residuals and the mean wage, and hence their ratio, i.e., CV.

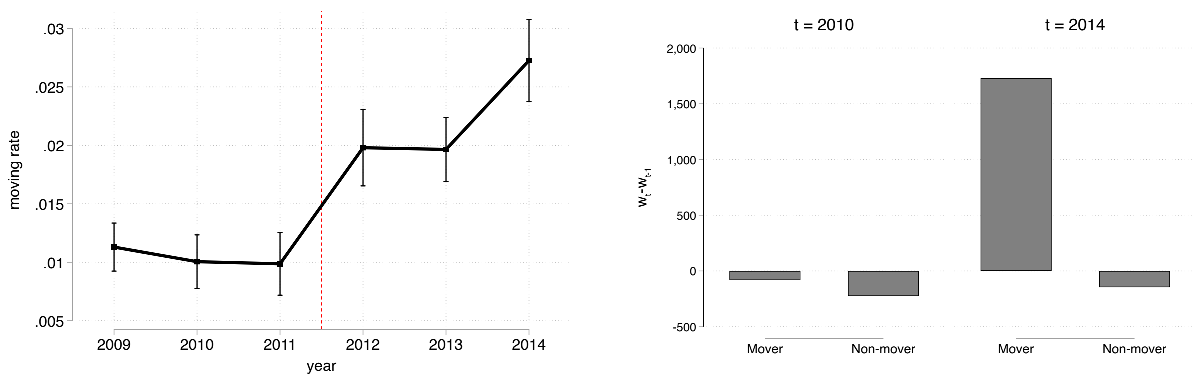
Figure 1: Variation in Teacher Salaries



CV of residuals from a regression of salaries on district-year and experience-education fixed effects.

year fixed effects.[14] Before Act 10, wage growth was small and negative in real terms for both movers and stayers. After Act 10, wage growth remained small and negative for stayers, but movers experienced significant wage growth ($1,750 at the median). This pattern is consistent with districts using wage strategies to compete for teachers after Act 10.

Figure 2: Teacher Mobility



(a) Share Changing District

(b) Relative Wage Growth

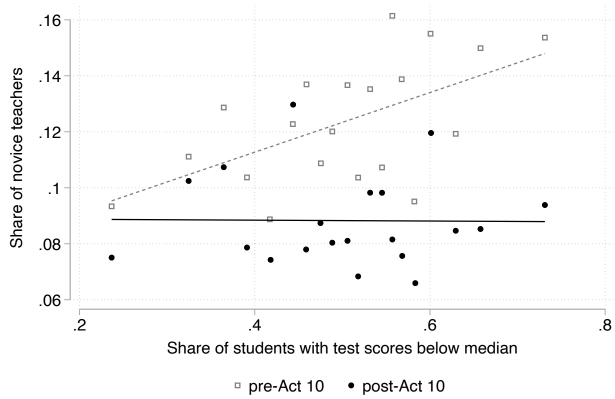
Panel a: share of teachers working in a different district relative to year t-1. Panel b: Median difference in salary residuals from a regression of salaries on teacher and year fixed effects.

Vertical Sorting of Teachers across Districts: Prior to Act 10, teachers with fewer than 3 years of experience, who tend to be less effective (Rockoff, 2004), were significantly more likely to work in districts serving larger shares of students who scored lower than the state median in math (Figure 3, lighter series). This relationship became much weaker after Act

[14]Specifically, let w_{it} be teacher i's real wage in year t, we regress w_{it} on teacher fixed effects and year fixed effects and obtain wage residuals ϖ_{it} from this regression. Let $\Delta_{it} = \varpi_{it} - \varpi_{it-1}$, Figure 2b shows the median Δ_{it} among those who moved across districts in t and the median Δ_{it} among those who stayed in the same district between $t-1$ and t, $t = 2010$ and 2014 are shown as examples.

10 (darker series). On the one hand, Figures 1 to 3 provide some suggestive evidence that,

Figure 3: Vertical Sorting



Avg. share of teachers with experience<= 3 yrs by vingile of within-district share of low-achieving students.

under flexible pay, districts used wage strategies to compete for teachers and teacher-district sorting became less vertical. On the other hand, these pre- versus post-Act 10 data patterns cannot be interpreted as the effect of giving districts control over teacher pay because market conditions differ in other aspects between the two eras (e.g., districts' budgets). To isolate the equilibrium impact of replacing rigid pay with flexible pay and, more importantly, to conduct counterfactual policy analysis, we resort to the following equilibrium model.

3 Model

We model a static equilibrium in the market for public school teachers, with a distribution of teachers and D school districts. Districts compete for their preferred teachers using wage and hiring strategies; each teacher chooses their most preferred district from those that offer them a job. Model primitives are as follows.

Teachers: A teacher is characterized by (x, c, d_0), where the vector $x = [x_1, x_2]$ includes experience and education; $c = [c_1, c_2]$ is the vector of one's effectiveness in teaching low- and high-achieving students, respectively;[15] and d_0 is the district one works in at the beginning of the model, where $d_0 \in \{1, ..., D\}$ for incumbent teachers working in district d_0, and $d_0 = 0$ for those who are yet to find a job on this market.

Districts: District d is characterized by $(q_d, \lambda_d, \kappa_d, M_d)$: q_d is a vector of district characteristics; λ_d is the fraction of students in d who are low-achieving (with prior test scores below the state median); κ_d is district d's capacity (number of teaching slots), and M_d is its budget. The sum of all slots across districts $\sum_d \kappa_d$ is equal to the total measure of teachers on the market.

[15]A teacher's effectiveness c is partly attributable to teacher characteristics x, as we specify in Section 5.1.2.

A teacher's total contribution to student achievement in district d is given by

$$TC(c, \lambda_d) \equiv \lambda_d c_1 + (1 - \lambda_d) c_2, \tag{1}$$

which, for the same teacher, varies across districts with districts' student composition λ_d.
Timing: The timing of the model is as follows:
1. Districts simultaneously choose their wage schedules $\{w_d(x, c)\}$ and job offers $\{o_d(x, c, d_0)\}$, where $o_d(x, c, d_0) = 1$ if d makes an offer to teacher (x, c, d_0), and 0 otherwise.
2. Each teacher observes their taste shocks and chooses their most preferred offer.

Notice that wages are assumed be blind to a teacher's origin d_0, which is consistent with real-life practice.[16] In contrast, job offers depend on d_0 because the current tenure system prevents a district from dismissing its tenured incumbent teachers.

3.1 Teacher's Problem

3.1.1 Teacher Preferences

For a teacher with (x, c, d_0), the net value of working in d is given by

$$V_d(x, c, d_0) + \epsilon_d \equiv w_d(x, c) + q_d\theta_0 + \theta_1 e^{\lambda_d} + \theta_2\lambda_d c_1 - \Gamma(d, d_0, x_1) + \epsilon_d, \tag{2}$$

where ϵ_d is an i.i.d. Type 1 extreme-value distributed taste shock with a scale parameter σ_ϵ. Wage enters with a normalized coefficient of 1, so that teacher preferences are measured in \$1,000. Teachers' preferences for district characteristics q_d are governed by the vector θ_0. The next two terms capture teachers' preferences for student composition (λ_d); these preferences may vary across teachers with different effectiveness in teaching low-achieving students (c_1).[17] $\Gamma(\cdot)$ is the cost of moving from d_0 to d, given by

$$\Gamma(d, d_0, x) = \begin{cases} 0 \text{ if } d_0 = 0 \\ I(d \neq d_0)(\delta_0 + x_1\delta_1) + I(z_d \neq z_{d_0})\delta_2 \text{ otherwise.} \end{cases}. \tag{3}$$

The cost is set to zero for teachers who are not already employed in any district $(d_0 = 0)$. For others, the cost of leaving their original district $(d \neq d_0)$ may vary with experience; in addition,

[16]Without this restriction, a district may want to pay incumbent teachers less than non-incumbent teachers with the same (x, c), since the former are easier to attract, due to teachers' moving costs. This restriction rules out such predictions, which are at odds with the data.

[17]It is disproportionally rare to see teachers move into districts with a high fraction of low-achieving students, suggesting that teachers' preference over λ_d might be convex. We therefore use e^{λ_d} in the utility function, which does not involve additional parameters. Indeed, we have estimated a model with a linear preference over λ_d, which does not fit the data well.

we allow for an additional cost if the two districts are not in the same commuting zone, where z_d denotes the commuting zone where d is located.

3.1.2 Teacher's Optimal Decision

Among all received offers $(o_d(x, c, d_0) = 1)$, a teacher chooses the one with the highest value:

$$\max_{d:o_d(x,c,d_0)=1} \{V_d(x, c, d_0) + \epsilon_d\}. \tag{4}$$

Let $d^*(x, c, d_0, \epsilon)$ be the teacher's optimal choice.

3.2 District's Problem

3.2.1 District Preferences

A teacher's value to district d is given by

$$xb_0 + b_1\lambda_d c_1 + b_2(1 - \lambda_d)c_2, \tag{5}$$

where b_0 captures districts' direct preferences for teacher experience and education; b_1 and b_2 capture how much a district cares about a teacher's contribution to its low- and high-achieving students, respectively.[18] We assume that $b \geq 0$, i.e., district preferences are weakly increasing in all teacher attributes, and we normalize b_1 to 1. A special case is $b_0 = 0$ and $b_1 = b_2$, in which Equation (5) is equivalent to $TC(c, \lambda_d)$, i.e., a district values a teacher only for their total contribution to its students. More generally, if b_1 and b_2 are large relative to b_0, districts would rank teachers differently depending on their student compositions λ_d; if b_0 is dominant, districts would largely agree on their rankings of teachers.

3.2.2 Choice Space for Wage Schedules

Because wage schedules are functions, the unrestricted choice space is of infinite dimensions. To keep the model tractable, we assume that a district's wage schedule is a linear combination of its pre-Act 10 experience-education wage schedule $W_d^0(x)$ and a teacher's contribution $TC(\cdot)$, given by

$$\omega_1 W_d^0(x) + \omega_2 TC(c, \lambda_d).$$

[18]Given that we only observe accepted offers, it is hard to separate out teachers' home bias from districts' direct preference over teachers' origins d_0. As such, we have ruled the latter out.

To avoid unrealistically high or low wages, we bound wages by $[\underline{w}, \overline{w}]$, such that[19]

$$w_d\left(x, c|\omega\right) = \max\left\{\min\left\{\omega_1 W_d^0\left(x\right) + \omega_2 TC\left(c, \lambda_d\right), \overline{w}\right\}, \underline{w}\right\}. \tag{6}$$

Under (6), a district's wage strategy boils down to a choice of $\omega = (\omega_1, \omega_2) \in \Omega$, where $\Omega \subset R^2_{\geq 0}$ is assumed to be discrete and finite.

Admittedly, the choice space implied by wage rule (6) is rather limited. However, as we show in Section 5.1.3, wages calculated under (6) match the observed wages very well. Moreover, this wage rule captures the essence of the wage-setting problem. In particular, if $\omega = (1, 0) \in \Omega$, teachers are paid on the rigid experience-education schedule, as is imposed on most U.S. school districts; if $\omega_2 > 0$, teachers are rewarded for their contribution, echoing the idea of performance pay. A central question is, if allowed to do so, how much would districts deviate from rigid pay and reward teachers for their contribution *in the equilibrium*? To answer this question, we first characterize districts' optimal decisions.

3.2.3 District's Optimal Decisions

Taking all the other districts' policies and teachers' decision rules as given, a district aims to fill its capacity with its most preferred teachers by making wage and job offer decisions, subject to its budget constraint. A district's problem can be solved in two steps: First, a district chooses a wage schedule $\omega = (\omega_1, \omega_2)$; second, it makes job offers conditional on ω. We solve the problem via backward induction.

Job Offers For a given wage schedule ω, district d's job offers $\{o_d\left(x, c, d_0|\omega\right)\}_{(x,c,d_0)}$ solve the following problem:

$$\begin{aligned}
\pi_d\left(\omega\right) = \max_{\{o_d(\cdot)\}} & \left\{\int o_d\left(x, c, d_0|\omega\right) h_d\left(x, c, d_0|\omega\right)\left[x b_0 + b_1 \lambda_d c_1 + b_2\left(1 - \lambda_d\right) c_2\right] dF\left(x, c, d_0\right)\right\} \qquad (7) \\
s.t. & \int o_d\left(x, c, d_0|\omega\right) h_d\left(x, c, d_0|\omega\right) dF\left(x, c, d_0\right) \leq \kappa_d, \\
& \int o_d\left(x, c, d_0|\omega\right) h_d\left(x, c, d_0|\omega\right) w_d\left(x, c|\omega\right) dF\left(x, c, d_0\right) \leq M_d \\
& o_d\left(x, c, d_0|\omega\right) = 1 \text{ if } x_1 \geq 3 \text{ and } d_0 = d,
\end{aligned}$$

where $h_d\left(x, c, d_0|\omega\right)$ is the probability that the teacher would accept the job *if* district d makes them an offer $\left(o_d\left(x, c, d_0|\omega\right) = 1\right)$, i.e., the probability that the teacher prefers d over all the

[19]Empirically, $\underline{w}$ ($\overline{w}$) is 0.3 standard deviations below (0.2 standard deviations above) the observed 1st (99th) wage percentile in the sample.

other districts that offer them a job. Teachers' decision rule in Equation (4) implies

$$h_d(x,c,d_0|\omega) = \frac{\exp\left(\frac{V_d(x,c,d_0)}{\sigma_\epsilon}\right)}{\exp\left(\frac{V_d(x,c,d_0)}{\sigma_\epsilon}\right) + \sum_{d' \in D \backslash d} o_{d'}(x,c,d_0) \exp\left(\frac{V_{d'}(x,c,d_0)}{\sigma_\epsilon}\right)}. \tag{8}$$

The first two constraints in (7) are for capacity and budget. The third constraint prohibits the district from dismissing its own tenured incumbent teachers, i.e., those with $d_0 = d$ and at least 3 years of experience, as is the case in Wisconsin. Let $\{o_d^*(x,c,d_0|\omega)\}$ be the optimal job offer decisions under wage schedule ω. Appendix A1 characterizes the solution to (7). In particular, district d would rank all teachers, except for tenured incumbents in d (because they are already guaranteed job offers from d). This ranking depends only on a teacher's value $xb_0 + b_1\lambda_d c_1 + b_2(1-\lambda_d)c_2$ and wage cost $w_d(x,c|\omega)$. Accounting for the acceptance probabilities by all teachers, including its tenured incumbents, district d would make offers to its n top-ranked teachers, where n is the maximum number of offers allowed by its capacity and budget.

Wage Schedule District d chooses ω to solve the following problem

$$\max_{\omega \in \Omega} \left\{ \frac{\pi_d(\omega)}{\kappa_d} - R(\omega) + \eta_\omega \right\}, \tag{9}$$

where $\pi_d(\omega)$ (given by (7)) is normalized by district capacity to make the scale comparable across districts with different capacities. $R(\cdot)$ captures some resistance or friction against deviating from $\omega_d = (1,0)$, i.e., a district's pre-reform wage schedule. We model $R(\cdot)$ as

$$R(\omega) = I(\omega \neq [1,0])(r_0 + r_1|\omega_1 - 1| + r_2\omega_2), \tag{10}$$

where r_0 captures the fixed cost of deviating from the rigid-pay schedule; r_1 and r_2 capture the incremental costs for larger deviations. Finally, η_ω is an i.i.d. extreme-value distributed shock associated with choosing ω, with a scale parameter σ_η.

3.3 Equilibrium

Definition 1 *An equilibrium is a tuple of decisions* $\{\{d^*(x,c,d_0,\epsilon)\}, \{\omega_d^*, \{o_d^*(x,c,d_0|\omega)\}\}_d\}$ *and belief* $\{\{h_d^*(x,c,d_0|\omega)\}_d\}$ *such that*
1) Given $\{\omega_d^*, \{o_d^*(\cdot|\omega_d^*)\}\}_d$, $d^*(x,c,d_0,\epsilon)$ *solves the teacher's problem, for all* (x,c,d_0,ϵ).
2) For all d, given $\{h_d^*(\cdot)\}$, ω_d^* *is an optimal wage decision and* $\{o_d^*(\cdot|\omega)\}$ *are optimal job offer decisions under* ω.
3) $\{h_d^*(\cdot)\}_d$ *is consistent with* $\{\{d^*(\cdot)\}, \{\omega_d^*, \{o_d^*(\cdot|\omega_d^*)\}\}_d\}$.

To solve its problem, it is sufficient for a district to know teachers' acceptance probabilities $\{h_d(x,c,d_0|\omega)\}$ (defined by (8)). Given $\{h_d(\cdot)\}$, knowledge about other districts' strategies is redundant. An equilibrium requires a consistent belief about $\{h_d(x,c,d_0|\omega)\}$. However, forming the exact belief about the high dimensional object $\{h_d(\cdot)\}$ is a daunting task for any decision maker. As a feasible alternative, we assume that districts make their decisions based on a simplified parametric belief about teachers' acceptance probabilities,[20] given by

$$\widetilde{h}_d(x,c,d_0|\omega) = \frac{1}{1+\exp\left(f(x,c,d_0,w_d,q_d,\lambda_d)\right)}, \tag{11}$$

where

$$\begin{aligned} f(\cdot) = {} & x\zeta_1 + \zeta_2\frac{c_1+c_2}{2} + \zeta_3\left(\frac{w_d(x,c|\omega)-\overline{w}(x,c)}{\sigma_{w(x,c)}}\right) + \zeta_4 q_d + \zeta_5 e^{\lambda_d} + \zeta_6\lambda_d c_1 \\ & + (1-I(d_0=0))\left[I(d\neq d_0)(\zeta_7+\zeta_8 x_1) + \zeta_9 I(z_d\neq z_{d_0})\right]. \end{aligned} \tag{12}$$

This simplified belief function captures the factors governing its counterpart $\{h_d(\cdot)\}$. The first two terms in (12) relate to the overall desirability of the teacher: A district should expect more competitors for a better teacher. The next term captures the idea that a district offering a more competitive wage should expect a higher acceptance rate. In particular, $\overline{w}(x,c)$ and $\sigma_w(x,c)$ are the cross-district average and standard deviation of wages for a teacher with attributes (x,c), according to the wage rules chosen by all districts in the equilibrium. We measure the competitiveness of a wage offer $w_d(x,c|\omega)$ by its standardized difference from the average $\overline{w}(x,c)$. The other terms in (12) mirror teachers' preferences over q_d and λ_d as in (2) and teachers' moving costs as in (3).

In the rest of the paper, we will study the market equilibrium with this simplified belief and replace $\{h_d(x,c,d_0|\omega)\}$ with $\left\{\widetilde{h}_d(x,c,d_0|\omega)\right\}$ in Definition 1. Solving for the equilibrium with the simplified belief boils down to finding $\{\zeta,\overline{w}(\cdot),\sigma_w(\cdot)\}$ that guarantee consistency between districts' belief $\widetilde{h}_d(\cdot)$ and teachers' acceptance rule $h(\cdot)$ given by Equation (8). Notice that $\{\zeta,\overline{w}(\cdot),\sigma_w(\cdot)\}$ are all equilibrium-specific and policy variant. For each counterfactual policy, we will search for the associated $\{\zeta,\overline{w}(\cdot),\sigma_w(\cdot)\}$ that guarantee belief consistency, using the equilibrium algorithm described in Online Appendix B1.

3.4 Discussion

For both tractability and data availability reasons, we abstract from several important aspects. First, because we only have data within Wisconsin's public school system, we focus on the

[20]Similar approaches have been used in the literature to approximate equilibrium objects that are too complex to compute exactly, e.g., Lee and Wolpin (2006) and Meghir et al. (2015).

competition among districts, while abstracting from their competition against teachers' outside options (e.g., private schools, pubic schools in other states and other occupations).[21] For the same reason, we do not model teachers' decisions to enter or exit the market, and we take the initial distribution of teachers on the market as pre-determined. Incorporating outside options in our framework would require additional data and modeling decision-making by outside employers, which we leave for future work. Although we cannot be certain about how incorporating teacher entry/exit may affect our findings, some studies suggest that tying public school teachers' wages to their effectiveness may improve the quality of the overall supply of teachers in both public and private schools (e.g., Tincani, 2021).[22] As such, the efficiency gain we find in our counterfactual policy experiments may be understated.

Second, given that wage schedules are set at the district level, we focus on the competition across districts while abstracting from the allocation of teachers across schools within a district.[23] Online Appendix B3 shows that the cross-district variation clearly dominates the within-district, cross-school variation in terms of both teacher wages and the share of low-achieving students. Introducing within-district competition into our framework would allow for a more complete view but would involve substantial complications.

Third, we take a district's student composition λ_d as given. In particular, we assume away potential households re-sorting across districts in response to our policy interventions.[24] In our data, the fraction of students moving across districts was very small and similarly so before and after Act 10.[25] Given that our counterfactual policies are much milder than Act 10, they are unlikely to significantly affect households' location choices. However, readers should still be aware of this limitation in interpreting our results.

Finally, we abstract from the effect of financial incentives on individual teachers' effort and effectiveness, which has been the focus of a large literature.[26] We complement this literature by focusing on a different channel via which financial incentives may improve education, i.e.,

[21] With a different focus, Dinerstein and Smith (2016) study private schools' responses to public school funding policies.

[22] Focusing on the supply side, Rothstein (2015) simulates individuals' dynamic self-selection into and out of teaching under different pre-set parameters, and finds that the effect of performance pay on selection is very small.

[23] Of the 411 districts in Wisconsin, 173 only have one public elementary school.

[24] For studies on household sorting and school district financing see, for example, Epple and Sieg (1999); Epple and Romano (2003); Ferreyra (2007); Epple and Ferreyra (2008).

[25] Between 2007 and 2016, 4.4% of Grade 4-6 students changed districts between two adjacent years on average. This fraction was *stable* before and after Act 10 (2011) at 4.2% in 2010, 4.3% in 2011, 4.2% in 2012 and 4.3% in 2013.

[26] Studies using data from outside of the US have found evidence that financial incentives for teachers affect student achievement (Muralidharan and Sundararaman, 2011; Duflo et al., 2012; Lavy, 2002; Atkinson et al., 2009; Glewwe et al., 2010). However, incentive programs implemented in the US have yielded mixed results, e.g., Epple and Ferreyra (2008); Fryer (2013); Imberman and Lovenheim (2015); Dee and Wyckoff (2015); Brehm et al. (2017).

financial incentives may incentivize more efficient teacher-district matching. To the extent that teachers may improve their effectiveness in response to financial incentives, our counterfactual policy results may understate the total policy effects.

4 Estimation

We estimate the model via indirect inference using post-Act 10 data, while holding out pre-Act 10 data for model validation. Indirect inference involves two steps: 1) compute from the data a set of "auxiliary models" that summarize the patterns in the data; and 2) repeatedly simulate data with the structural model, compute corresponding auxiliary models using the simulated data, and search for model parameters such that the auxiliary models from the simulated data match those from 1). In particular, let $\overline{\beta}$ denote our chosen set of auxiliary model parameters computed from data; let $\widehat{\beta}(\Theta)$ denote the corresponding auxiliary model parameters obtained from simulating a large dataset from the model (parameterized by Θ) and computing the same estimators. In our case, Θ consists of 20 parameters governing teacher preferences and district preferences. The estimated vector of structural parameters is the solution

$$\widehat{\Theta} = \text{argmin}_{\Theta} \left\{ [\widehat{\beta}(\Theta) - \overline{\beta}]' W [\widehat{\beta}(\Theta) - \overline{\beta}] \right\},$$

where W is a weighting matrix.

The estimation algorithm involves an outer loop searching for the parameter vector Θ and an inner loop solving the model for each given Θ (detailed in Online Appendix B1). While we need to find the fixed point for $\{\zeta, \overline{w}(\cdot), \sigma_w(\cdot)\}$ in our counterfactual policy simulations, we only need to find the fixed point for ζ during the estimation: Assuming that data were generated from an equilibrium, $\{\overline{w}(\cdot)\}$ and $\{\sigma_w(\cdot)\}$ can be derived directly from the *observed* district wage schedules $\{\omega_d^o\}_d$, where the superscript o denotes "observed."

4.1 Identification

A major identification challenge arises from the fact that among all offers made, the researcher observes only the accepted ones, i.e., the realized teacher-district matches. This makes it hard to separate teachers' preferences from districts' preferences. To overcome this obstacle, we fully exploit the following features, which guide our choice of auxiliary models.

4.1.1 Optimal Job Offers and Observed Matches

First, consider district d's job offer decisions. The marginal benefit of hiring a teacher consists of their contribution to district d's low-achieving students $\lambda_d c_1$ and high-achieving students

$(1-\lambda_d)\, c_2$, and the direct value of their education and experience x. The marginal cost consists of teacher wage $w_d\,(x, c|\omega_d^o)$ (calculated using wage rule (6) at the observed wage schedule ω_d^o) plus the shadow price of a slot.[27] If district d hired teacher i, who was not a tenured incumbent in d (hence the offer was for sure made based on d's preference instead of the non-dismissal constraint), then for any district preference parameter vector $b \geq 0$, a teacher j satisfying the following (sufficient) conditions was at least as preferable as i and hence must also have had an offer from d : 1) j had weakly higher c_1, c_2 and x than i,[28] and 2) $w_d\,(x_j, c_j|\omega_d^o) \leq w_d\,(x_i, c_i|\omega_d^o)$. With this argument, we can use observed matches ((i,d) in this example) to infer offers for other teachers (j in the example). We can then construct, for each teacher i, a subset of all the offers they received $O_i^s \subseteq \{d : o_d\,(x_i, c_i, d_{0i}) = 1\}$, which consists of inferred offers, the accepted offer, and, if i is tenured, the guaranteed offer from i's original employer d_{0i}.

If O_i^s is not a singleton, a teacher's choice within O_i^s informs us of teacher preferences, since all options in O_i^s were feasible.[29] In contrast, if one were to infer teacher preferences assuming that teachers had offers from *all* districts, the inferred "preferences" would be different. The discrepancy between the two sets of inferred preferences arises because certain districts did not make offers to certain teachers. One can learn about districts' preferences from this observation. In particular, districts' preference parameters have to generate not only the observed offers, but also the lack of offers from certain districts to certain teachers that would reconcile the aforementioned discrepancy.[30]

Remark 1 *The argument above relies on three maintained assumptions.*
A1: (x, c) are observable to all districts. With our data, it is difficult to separate preferences from information friction. We therefore rule the latter out.
A2: Districts cannot discriminate among teachers by factors other than (x, c).[31] *If some job offers were made for reasons other than (x, c), then the inferred O_i^s might include infeasible options for some teachers and thus introduce biases in the inferred teacher preferences based*

[27]The shadow price of a slot, being common between two potential hires, does not affect how a district ranks teachers. See Appendix A1 for more details.

[28]We assume that teacher experience (x_1) enters district preference as ordered categorical variables (0-2, 3-4, 5-9, 10-14, 15 years or more). Therefore, the comparison of teacher experience (x_1) here is based on these categories.

[29]For 5,170 out 6,600 teachers in our sample, O_i^s consists of at least two districts. Given that O_i^s is only a subset of feasible choices faced by i, our identification is facilitated by the I.I.A. property of the Type-1 extreme value distributed preference shocks. However, as Fox (2007) shows, under standard conditions, multinomial discrete-choice models are identified semiparametrically using a subset of choices.

[30]Districts' offer decisions, of course, depend on how they rank teachers. Noticeably, as seen in Equation (5) , the more districts value c relative to x, the more each district's ranking of teachers, and hence its offer decisions, would depend on the district's student composition (λ_d) .

[31]The identification argument holds as long as districts' decisions are based on (x, c) even if c measures teacher effectiveness with errors. The fundamental assumption is that districts value teachers for their (x, c), which we also observe.

on O_i^s. However, as long as most job offers are based on (x, c), such biases would be small and teacher preferences inferred from O_i^s would still be much closer to their true preferences than those inferred assuming that teachers had offers from all districts. In such cases, the essence of our identification strategy still holds, although the argument would not be as sharp.

A3: We have assumed away job posting costs, which is plausible because in reality districts post openings publicly on online platforms.[32] We also assume that teachers get offers without having to apply. This assumption does not affect our inference of teacher preferences. In particular, the following two cases would both imply that district d was not attractive enough to teacher j: 1) d made an offer to j and j did not accept; 2) j was eligible for a job position in d but did not apply.[33]

4.1.2 Wage Schedule and District Pre-Determined Conditions

Under Act 10, districts can choose how to reward teachers. Therefore, the extent to which wage schedules favor teachers with different (x, c) and how this relates to districts' pre-determined conditions are informative about district preferences.[34]

To see the intuition, notice that wage schedules can serve both to pull and to push teachers. To pull teachers with its preferred attributes (x, c), a district should choose a wage schedule that favors (x, c). The need to do so is higher when these teachers are not district incumbents because moving is costly for teachers, especially across commuting zones. Meanwhile, although a district cannot dismiss its tenured incumbents with undesirable (x', c'), it can push them out by choosing a wage schedule that disfavors (x', c'). Notice that for teachers with (x', c') who are not tenured incumbents in d, district d can avoid them simply by not offering them jobs. Therefore, the incentive to use a wage schedule disfavoring (x', c') is stronger if the district has more tenured incumbents with (x', c').

Finally, district preferences over teachers may not be sufficient to explain the data (24% of districts in our data kept their pre-reform wage schedules after Act 10). Districts' choices that are not explained by their preferences for teachers are attributed to the resistance cost $R(\cdot)$.

4.2 Auxiliary Models

Following the identification argument, we target the following auxiliary models *jointly*.

[32] See, for example, the Wisconsin Education Career Access Network (WECAN), https://wecan.education.wisc.edu/.

[33] If it is costly for teachers to apply for jobs (more so for jobs in districts other than one's initial district), then these costs would be absorbed in teachers' moving costs in our model.

[34] A district's pre-determined conditions include its student composition λ_d, characteristics q_d, capacity κ_d, budgets M_d, the composition of district incumbent teachers (those with $d_{i0} = d$) and that of teachers working in other districts but within the same commuting zone.

Aux 1 Coefficients from two regressions of the following form

$$y_{id} = \beta_1^m w\left(x_i, c_i|\omega_d\right) + I\left(d_{0i} > 0\right) \left[\begin{array}{c} I\left(d \neq d_{0i}\right) x_i \beta_2^m \\ +\beta_3^m I\left(z_d \neq z_{d_{0i}}\right) \end{array} \right] + q_d \beta_4^m + \beta_5^m e^{\lambda_d} + \beta_6^m c_{1i} \lambda_d + \psi_i + \varepsilon_{id}^m,$$

where $y_{id} = 1$ if teacher i is matched with district d, and 0 otherwise. The right-hand-side variables are the same as those entering teachers' preferences, including $w\left(x_i, c_i|\omega_d\right)$, the wage i would be paid by district d under wage rule (6). ψ_i is a teacher dummy that relates all $\left\{(i,d)\right\}_d$ observations associated with teacher i.[35] The two regressions differ in the number of observations, reflecting the identification argument in Section 4.1.1.

Aux 1a The first regression includes all teachers whose inferred subsets of offers O_i^s contain more than one district; an observation (i,d) is a teacher-district pair in these inferred subsets.[36]

Aux 1b The second regression includes every possible teacher-district pair.

Aux 2 Moments of district-level teacher characteristics (x, c_1, c_2) by district groups (quintiles of λ_d, quintiles of budget per slot, and urban/suburban status).[37]

Aux 3 Coefficients from regressions of wage schedule ω_{dn}, $n = 1, 2$, on district's pre-determined conditions, reflecting the identification argument in Section 4.1.2:

$$\omega_{dn} = \beta_{0n}^w + q_d \beta_{1n}^w + \beta_{2n}^w \lambda_d + \beta_{3n}^w \kappa_d + \beta_{4n}^w M_d + X_d \beta_{5n}^w + \beta_{6n}^w \overline{TC}_d + \beta_{7n}^w \sigma_{TC_d} + \beta_{8n}^w \overline{TC}_d^{tenure} + \beta_{9n}^w \overline{TC}_{Z_d} + \varepsilon_{dn}^w,$$

where coefficients β_{1n}^w to β_{4n}^w are associated with district characteristics and constraints; β_{5n}^w to β_{8n}^w are associated with the composition of district incumbents. In particular, X_d is the average x, $\overline{TC}_d$ (σ_{TC_d}) is the average (standard deviation) of TC among teachers with $d_{0i} = d$, and $\overline{TC}_d^{tenure}$ is the average TC among the district's tenured incumbents ($d_{0i} = d$ and $x_{1i} \geq 3$). Finally, $\overline{TC}_{Z_d}$ is the average TC of teachers originally working in other districts within d's commuting zone (i.e., $d_{0i} \neq d$, but $z_{d_{0i}} = z_d$).[38]

[35] Although conditional logit regressions would be a more intuitive way to summarize discrete choices, they are computationally too costly to run during the estimation. We use a linear regression with teacher dummies; these dummies capture the idea that the same teacher is choosing one district out of a given set of districts.

[36] To run this regression in the data, we take observed (i,d) matches and construct O_i^s following the procedure described in Section 4.1.1. To run this regression in the model, we take model simulated $((i,d)|\Theta)$ matches and use the same procedure to construct $O_i^s(\Theta)$.

[37] To calculate these district-level moments, we first calculate the averages of one-way teacher characteristics within each district, and then calculate the averages within each district group, treating each district as an observation.

[38] All else equal, teachers in nearby districts face lower costs for moving to d and therefore may be easier to attract than teachers in far-away districts. We do not include the average x of these teachers or the characteristics of near-by districts in our final specification of Aux 3, because they are insignificant when included.

Aux 4 Cross-district wage schedule moments: $E(\omega_1)$, $E(\omega_2)$, $E(\omega_1^2)$, $E(\omega_2^2)$, $E(\omega_1\omega_2)$, and $E[I((\omega_1,\omega_2)=(1,0))]$ (the fraction of districts using pre-Act 10 schedules).

5 Data

Our data, from the Wisconsin Department of Public Instruction (WDPI), consist of three linked data sets that provide information about teachers, students, and districts respectively.

Teacher-Level Data (PI-1202 Fall Staff Report) cover all individuals employed by WDPI between 2006 and 2016.[39] This panel provides information about teachers' education, years of teaching experience, total wages, full-time equivalency units, school and district identifiers, and grades and subjects taught.

Student-Level Data include scores for all public school students in Grades 3 to 8 in state standardized tests between 2007 and 2016, and their demographics.

District-level Information: Using student test score data, we calculate λ_d, the fraction of students in district d whose prior math scores were below the grade-specific state median. District characteristics q_d include indicators for urbanicity (urban, suburban or rural), and an indicator for being in a large metropolitan area, all based on the 2010 Census classification. Each district is assigned to one of 19 commuting zones z_d.

5.1 Empirical Definitions

To map our equilibrium model to the data, we use the following empirical definitions (more details are in Online Appendix B2).

5.1.1 The Market

Our model is in a static equilibrium setting. For estimation and counterfactual policy analyses, we use data in 2014, i.e., 3 years after Act 10; by then, all the CB agreements pre-dating Act 10 had expired (Feng and Sass, 2018), and districts had obtained full autonomy over teacher pay. To validate the estimated model, we simulate the market equilibrium under rigid-pay and initial conditions in 2010 data, i.e., the year preceding Act 10.

In both years, we focus on the market for non-substitute full-time public-school math teachers in Grades 4-6, for the following reasons. We exclude the few substitute and part-time

[39] All of our data are reported by academic years, and referenced by the calendar year of the spring semester (e.g. 2014 for 2013-14 academic year).

teachers because they face different types of contracts than regular, full-time teachers.[40] We exclude secondary-school teachers because they often teach multiple grades, making it hard to identify individual teacher contribution (Kane and Staiger, 2008; Chetty et al., 2014). Among elementary-school teachers, effectiveness measures are obtainable for teachers in Grades 4-6; and we restrict attention to those teaching the same subject (math), so that the effectiveness measures are comparable across teachers.[41] The estimation sample contains 411 districts and 6,600 teachers; and the validation sample contains 411 districts and 6,741 teachers.

By focusing on a subgroup of teachers, we have implicitly assumed that a district's capacity and budget constraints for these teachers do not interact with those for other teachers. This assumption will hold if, for example, a district commits certain resources for the math education of its Grade 4-6 students.

5.1.2 Teacher Characteristics

Teacher Effectiveness c_{i1} and c_{i2} are teacher i's contributions to the achievement of low- and high-achieving students, respectively. To obtain (c_{i1}, c_{i2}) for each i, we modify the student achievement model in Kane and Staiger (2008) to allow for two-dimensional effectiveness as follows:[42]

$$A_{kt} = \gamma Z_{kt}^{s} + \sum_{i:SG_{kt}=SG_{it}^{T}} \left[\begin{array}{c} I\left(\tau_k = 1\right)\left(\rho_1 x_{it} + v_{i1}\right) \\ +I\left(\tau_k = 2\right)\left(\rho_2 x_{it} + v_{i2}\right) \end{array} \right] + \varepsilon_{kt}, \tag{13}$$

where A_{kt} is student k's achievement (standardized math score) in year t; Z_{kt}^{s} includes a vector of student observables (including A_{kt-1}), a school-grade fixed effect, and a year fixed effect. In the summation, SG_{kt} (SG_{it}^{T}) denotes the school-grade student k (teacher i) belongs to in year t; τ_k denotes a student's type: $\tau_k = 1$ if k is low-achieving or k's prior score is below the grade-specific state median; conversely, $\tau_k = 2$ if k is high-achieving. For a student of achievement type $n \in \{1, 2\}$, teacher i's contribution is given by $\rho_n x_{it} + v_{in}$, where x_{it} denotes i's education and experience in year t, and v_{in} is the part unexplained by x_{it}. Assuming ε_{kt} is an i.i.d. idiosyncratic component, we estimate γ, ρ_1 and ρ_2 via OLS using data from 2007 to 2016; then, we use the Bayes estimator of Kane and Staiger (2008) to estimate v_{i1} and v_{i2}.

[40] Among all public school teachers teaching Grades 4-6 math in 2014 (2010), 2.0% (1.8%) were substitute teachers and 2.8% (3.9%) were part-time teachers.

[41] We need student test scores from the previous year to calculate teachers' effectiveness; and our test score data start from Grade 3. We choose math over English because previous studies have found that teacher effects on students are larger in math than in reading or language (e.g. Rivkin et al., 2005; Kane and Staiger, 2008; Chetty et al., 2014).

[42] Online Appendix B2.3.4 shows that our two-dimensional teacher effectiveness model explains approximately 20% more variation in test scores compared to the one-dimensional effectiveness model.

Finally, we construct teacher effectiveness (c_{i1}, c_{i2}) in our model as

$$c_{in} \equiv \widehat{\rho}_n x_{it^*} + \widehat{v}_{in},\ n \in \{1, 2\}, \tag{14}$$

where t^* is 2014 for the estimation sample and 2010 for the validation sample.[43]

Remark 2 *Besides c being two-dimensional, Equations* (13) *and* (14) *have two additional features.*

Feature 1: We allow c to vary directly with x because experience has been shown to affect teacher effectiveness (e.g., Wiswall, 2013; Rockoff, 2004).

Feature 2: In (13) *, a teacher contributes to all students in their school-grade.*

To estimate effectiveness with Feature 1 using our data, we have to assume Feature 2 because of data limitation: We can link students and teachers only up to the school-grade level. In an alternative achievement model where a teacher contributes only to students in their class, we can use our data to identify teacher effectiveness assuming that it is invariant to one's experience. Identification of both models exploits teacher turnover across school-grades and the assumption that ε_{kt} *and* v_{in} *are uncorrelated. Notice that this assumption allows for endogeneous district-teacher sorting (as is the case in our model), because we control for* Z^s_{kt}*, which includes school-grade fixed effects and year fixed effects.*[44] *We have estimated both achievement models. Online Appendix B2.3 shows details of the estimation and identification. It also shows that the estimated teacher effectiveness measures from the two achievement models are highly correlated. More importantly, auxiliary models Aux 1a and 1b, which provide key information for identifying our equilibrium model, are very similar using either type of effectiveness measures.*

Teacher's Origin District: For the estimation sample, we use teachers' employment histories between 2011 (when Act 10 was passed) and 2014, and define d_{0i} as i's last employer before 2014.[45] We follow the same procedure for the validation sample, using a teacher's employment history between 2007 and 2010.

[43] Following the literature, we measure c_{i1} and c_{i2} as residual contributions to standardized test scores; given that the mean of test scores is 0, c_{i1} and c_{i2} can be negative. In order to make sure that all teachers contribute (weakly) positively to a district's objective value (7) and that a district would not want to leave classrooms unstaffed, we replace c_1 and c_2 in (7) with $(c_1 - \underline{c}_1)$ and $(c_2 - \underline{c}_2)$, where $\underline{c}_1$ $(\underline{c}_2)$ is the minimum of c_1 (c_2) across all teachers in the sample. Notice that this re-scaling is innocuous because it does not affect how a district ranks teachers.

[44] Consistent with our model, this assumption relies on random allocation of teachers within a school-grade, while allowing for endogeneous matching of teachers across school-grades and hence across districts.

[45] For example, if i moved at most once between 2011 and 2014, d_{0i} is i's employer in 2011; if i worked in A in 2011, moved to B in 2013, and then to C in 2014, we set $d_{0i} = B$.

5.1.3 Wage Schedules and District Constraints

Pre-Act 10 Wage Schedules $\{W_d^0(x_i)\}_d$ are obtained using data from 2007 to 2011. Specifically, $W_d^0(x_i)$ is the predicted value from a regression of observed pre-Act 10 teacher real wages (in 2014 dollars) on indicators for experience groups and education groups, where the regression coefficients are allowed to differ across districts.[46]

Choice Set for Wage Schedules **(Ω):** We first construct a grid Ω^o such that wages $w_d(x_i, c_i|\omega)$ under (6) and $\omega \in \Omega^o$ provide a good coverage of the observed wage distribution. We then expand the grid range such that the model choice set $\Omega \supset \Omega^o$ to allow for the possibility that district choices may go out of the empirical range in counterfactual scenarios. We use the same Ω, which contains 6×8 different (ω_1, ω_2) combinations, throughout our estimation and simulations.[47]

District Wage Schedules: For each district, we find the grid point on Ω that best summarizes the observed wages (w_i^o) of teachers working in d ($d(i) = d$):

$$(\omega_{d1}^o, \omega_{d2}^o) = \arg \min_{(\omega_1,\omega_2)\in\Omega} \sum_{i:d(i)=d} \left(w_i^o - w_d(x_i, c_i|\omega)\right)^2,$$

where $w_d(x_i, c_i|\omega)$ is given by wage rule (6). The pair $(\omega_{d1}^o, \omega_{d2}^o)$ is treated as district d's wage schedule in the realized equilibrium. The resulting $\{(\omega_{d1}^o, \omega_{d2}^o)\}_d$ matches teachers' actual wages very well.[48]

District Capacity and Budget Constraints are binding in the equilibrium. Assuming data are generated from an equilibrium, κ_d is then the number of teachers in our sample working in d in year t, and M_d is the sum of wages ($w_d(x_i, c_i|\omega_d^o)$) among these teachers, where $t = 2014$ (2010) for the estimation (validation) sample.

5.2 Summary Statistics

Panel A of Table 1 shows summary statistics for all 6,600 teachers in the estimation sample, for non-tenured teachers ($x_1 < 3$), and for those with over 10 years of experience ($x_1 \geq 10$). Fifty-three percent of all teachers have a graduate degree; this share is 6% among non-tenured teachers and 68% among teachers with over 10 years of experience. On average, non-tenured teachers are less effective than more experienced teachers in terms of both c_1 and c_2. However,

[46]Among the specifications we have tried, we found this specification of $W_d^0(x_i)$, as detailed in Online Appendix B2.4.1, fits the wage data the best. The experience groups are 0, 1-2, 3-4, 5-9, 10-14 and 15 or more.

[47]Specifically, $\Omega = \{0.9, 0.95, 1, 1.05, 1.1, 1.15\} \times \{0, 10, 30, 50, 75, 100, 200, 225\}$, and $\Omega^o = \{0.9, 0.95, 1, 1.05, 1.1\} \times \{0, 10, 30, 50, 75, 100, 200\}$. Details are in Online Appendix B2.4.2.

[48]The estimated slope coefficient of a model of w_i^o as a function of $w_d(x_i, c_i|\omega_d^o)$ equals 0.98 (with a standard error of 0.001) and an R^2 of 0.99.

the overall correlation between experience (x_1) and either c_1 or c_2, not shown in the table, is only around 0.04.[49] The last row of Panel A shows that the correlation between c_1 and c_2 is 0.67, which implies the existence of both absolute and comparative advantages across teachers in teaching different types of students.

Panel B of Table 1 summarizes districts' characteristics and the composition of a district's incumbent teachers ($d_{0i} = d$). We present statistics for all the 411 school districts in the estimation sample and separately for districts belonging to the 1st and 4th quartiles of the distribution of λ_d (the fraction of low-achieving students). Districts with fewer low-achieving students are more likely to be located in suburban areas and to have larger capacity and per-teacher budgets (throughout the paper, all dollar values are in 2014 dollars). Incumbent teachers in these districts are more likely to be highly-educated.

Table 1: Teacher and District Characteristics (2014)

A. Teacher Characteristics	All	$x_1 < 3$	$x_1 \geq 10$
x_1: Experience	14.6 (9.2)	1.4 (0.5)	19.7 (6.9)
x_2: MA or above	0.53 (0.50)	0.06 (0.24)	0.68 (0.47)
$10c_1$	0.12 (0.29)	0.04 (0.37)	0.12 (0.26)
$10c_2$	0.11 (0.33)	0.02 (0.42)	0.12 (0.31)
Corr (c_1, c_2)	0.67	-	-
# Teachers	6,600	627	4,384
B. District Characteristics	**All**	**λ_d 1st Quartile**	**λ_d 4th Quartile**
Urban	0.04	0.02	0.03
Suburban	0.15	0.34	0.09
λ_d	0.50 (0.12)	0.34 (0.07)	0.65 (0.06)
Capacity	16.9 (30.5)	18.4 (15.9)	14.3 (43.9)
Budget/Capacity ($1,000)	50.9 (6.6)	53.0 (6.8)	48.9 (6.3)
Characteristics of District Incumbent Teachers ($d_0 = d$)			
Average experience	17.7 (4.8)	17.4 (4.5)	17.7 (5.7)
Share w/MA or above	0.56 (0.28)	0.64 (0.26)	0.47 (0.29)
Average $10c_1$	0.14 (0.11)	0.14 (0.11)	0.14 (0.12)
Average $10c_2$	0.14 (0.13)	0.13 (0.10)	0.12 (0.14)
# Districts	411	103	103

Means and std. deviations (in parentheses) of teacher (Panel A) and district (Panel B) characteristics.

Column 1 of Table 2 shows the OLS estimates from Aux 1a (Section 4.2), which summarize how teachers made their choices given their inferred subsets of offers O_i^s. Column 2 shows

[49] This is consistent with, for example, Rockoff (2004).

OLS estimates from Aux 1b, which would reflect teachers' preferences *only if* all teachers received offers from all districts. Some clear differences exist between the two columns. For example, Column 1 shows that teachers care about wages (Row 1) and that teachers who are more effective with low-achieving students are more willing to teach in districts with higher fractions of these students (Row 3); however, neither of these relationships exist in Column 2. These differences arise because certain teachers did not get offers from certain districts and are informative about district preferences.

Table 2: OLS of Teacher-District Match (2014)

Teacher's Choice Set	Inferred Offer Set^a	All Districts^b
wage	0.002 (0.0002)	-5$\times 10^{-6}$(2$\times 10^{-6}$)
e^{λ_d}	-0.004 (0.009)	-0.0001 (0.0001)
$c_1 \times \lambda_d$	0.52 (0.29)	-0.02 (0.006)
$I\,(d \neq d_0)$	-0.72 (0.02)	-0.80 (0.01)
$I\,(d \neq d_0) \times$ experience	-0.008 (0.001)	-0.008 (0.0005)
$I\,(z_d \neq z_{d_0})$	-0.06 (0.006)	-0.0006 (0.0001)
q_d : urban	0.01 (0.002)	0.003 (0.0002)
q_d : suburban	0.01 (0.002)	0.001 (0.0001)
q_d : large metro	0.11 (0.03)	0.01 (0.002)
# Obs	57,068	2,712,600

$a(b)$: OLS specified in Aux 1a (1b), teacher fixed effects included.

Robust standard errors are in parentheses.

Panel A of Table 3 summarizes districts' wage schedule choices. Districts' choices of ω_2 (rewards for teacher contribution) are more dispersed than their choices of ω_1. Although given the flexibility, 24% of districts continued to use their pre-reform wage schedules ($\omega = (1, 0)$) and only 50% of districts chose to reward teacher contribution ($\omega_2 > 0$). Panel B summarizes wages in the realized district-teacher matches. On average, more experienced teachers are paid more. Panel C compares districts' characteristics and the composition of each district's incumbent teachers among districts that did not reward teacher contribution and those that did. The difference is small, but districts with $\omega_2 > 0$ appear more disadvantaged: They are more likely to be in rural areas and have higher fractions of low-achieving students, smaller per-teacher budgets, and slightly weaker composition of incumbent teachers.[50]

[50]One possible explanation is the following: It would be difficult and costly for disadvantaged districts to compete for experienced and effective teachers. By setting a higher ω_2 (which implies a lower ω_1 to balance the budget), these districts can improve their attractiveness to young but effective teachers.

Table 3: District Wage Schedules (2014)

A. Summary stats of (ω_1,ω_2)		B. $w_d\left(x, c\vert\omega_d^o\right)$ in Realized Matches (\$1,000)		
ω_1 mean (std)	0.99 (0.04)	All Teachers: mean (std)		55.1 (11.6)
ω_2 mean (std)	31.3 (50.8)	Experience	< 3	37.2 (4.4)
$Corr\left(\omega_1,\omega_2\right)$	-0.19		$\in [3,4]$	41.0 (5.6)
Fr$((\omega_1,\omega_2)=(1,0))$	0.24		$\in [5,9]$	48.0 (6.4)
Fr$(\omega_2>0)$	0.50		≥ 10	56.5 (7.2)

C. District Characteristics by ω_2	$\omega_2 = 0$	$\omega_2 > 0$
Rural	0.80	0.83
λ_d>median	0.48	0.52
Budget/Capacity (\$1,000)	51.2	50.7
Incumbent Teachers in $d(d_{0i}=d)$		
Average experience	17.8	17.6
Share w/MA or above	0.57	0.55
Average $10c_1$	0.14	0.14
Average $10c_2$	0.14	0.13
# Districts	205	206

Standard deviations are in parentheses.

6 Estimation Results

6.1 Parameter Estimates

Table 4 shows estimated model parameters, with standard errors in parentheses.[51] Panel A shows estimated parameters governing teachers' preferences. For an average teacher, districts with higher fractions of low-achieving students (λ_d) are less desirable. However, teachers who are more effective in teaching low-achieving students are more willing to teach in these districts.[52] We also find that rural districts are less attractive than their urban counterparts. The rest of Panel A show that, on average, teachers face high moving costs, especially when moving across commuting zones. However, individuals compare the total value of each option when making their choices, including their preference shocks (governed by the scale parameter σ_ϵ). High average moving costs help explain the lack of teacher mobility in general, while preference shocks absorb idiosyncratic reasons for mobility. Our findings, i.e., both the average moving cost and the dispersion of preference shocks are large, are consistent with those in previous

[51]Standard errors are derived numerically via the Delta Method.

[52]For example, a teacher whose c_1 is at the 10th percentile ($c_1=-0.02$) would put a premium of \$4,227 on a district with $\lambda_d=0.3$ over an otherwise identical district with $\lambda_{d'}=0.7$; for a teacher whose c_1 is at the 90th percentile $(c_1=0.05)$, this permium is only \$1,817.

studies on worker mobility (e.g., Kennan and Walker, 2011). One possible explanation, which we abstract from, is the family joint location problem: The tied stayer would appear to have very high moving costs, while the tied mover would appear to have very low and even negative moving costs (e.g., Gemici, 2011).

Panels B and C show district-side parameter estimates, which tend to have larger standard errors than those in Panel A, since we have many more teacher observations (6,600) than district observations (411). Panel B suggests that districts significantly value a teacher's contribution to its students' achievement but do not value teacher experience and education per se. We also find that districts value a teacher's contribution to its low-achieving students slightly more than their contribution to its high-achieving students, although the difference between these two parameters is not significant. Panel C shows the cost a district faces for deviating from the pre-Act 10 wage schedule.[53]

Table 4: Parameter Estimates

A. Teacher Preference			B. District Preference		
wage (\$1,000)	1	normalized	c_1	1	normalized
e^{λ_d}	-5.33	(2.72)	c_2	0.90	(0.23)
$c_1 \times \lambda_d$	86.09	(21.53)	Yrs of experience:		
q_d : urban	14.10	(1.86)	1-2	0.006	(0.01)
q_d : suburban	14.03	(2.53)	3-4	0.011	(0.02)
q_d : large metro	18.34	(3.45)	5-9	0.014	(0.01)
$I(d \neq d_0)$	-90.97	(1.37)	10-14	0.022	(0.04)
$I(z_d \neq z_{d_0})$	-83.38	(91.45)	≥ 15	0.058	(0.07)
$I(d \neq d_0)\, x_1$	-2.33	(0.08)	MA or above	1.3×10^{-5}	(5.8×10^{-6})
σ_ϵ	19.98	(1.45)			
C. Wage Setting Cost $R(\cdot)$					
r_0: $I(\omega \neq [1,0])$	0.93	(0.72)	$r_1 : \lvert\omega_1 - 1\rvert$	16.59	(24.15)
r_2: $\omega_2/100$	0.76	(0.39)	σ_η	0.75	(0.33)

*Std errors (in parentheses) are derived numerically via the Delta Method.

6.2 Within-Sample Fit

Table 5 shows within-sample model fits of the coefficients from the two regressions specified in Section 4.2, Aux 1a on the left and Aux 1b on the right. Table 6 shows model fits for the

[53] Our model is silent on what causes these costs, which may arise, for example, from the resistance of teachers or school boards. It is possible that such resistance may fade off over time, in which case, our policy implications should be interpreted as short-run effects.

moments of district-level teacher characteristics (Aux 2). The model well replicates teacher-district sorting patterns in both tables.

Table 5: Model Fit: OLS of Teacher-District Match ($d^*(\cdot) = d$)

Teacher's Choice Set	Inferred Offer Set[a]		All Districts[b]	
	Data	Model	Data	Model
wage	0.002	0.003	-5×10^{-6}	-3×10^{-6}
e^{λ_d}	-0.004	-0.005	-0.0001	-0.00005
$c_1 \times \lambda_d$	0.52	0.25	-0.02	0.002
$I(d \neq d_0)$	-0.72	-0.87	-0.80	-0.96
$I(d \neq d_0) \times$ experience	-0.008	-0.002	-0.008	-0.001
$I(z_d \neq z_{d_0})$	-0.06	-0.06	-0.0006	0.0001
q_d : urban	0.01	0.01	0.003	0.002
q_d : suburban	0.01	0.01	0.001	0.002
q_d : large metro	0.11	0.54	0.01	0.03

$a(b)$: OLS specified in Aux 1a (1b), teacher fixed effects included: data vs model, post Act 10.

Table 6: Model Fit: Average District Employee Characteristics ($d^*(\cdot) = d$)

		Experience		Share MA or above		$10c_1$		$10c_2$	
District Group		Data	Model	Data	Model	Data	Model	Data	Model
λ_d :	Quintile 1	14.7	13.7	0.53	0.50	0.13	0.13	0.11	0.13
	Quintile 2	15.5	14.5	0.51	0.49	0.12	0.13	0.13	0.15
	Quintile 3	15.6	14.4	0.48	0.46	0.14	0.14	0.12	0.14
	Quintile 4	16.3	15.2	0.48	0.48	0.14	0.14	0.16	0.15
Budget/Capacity:	Quintile 1	11.5	11.5	0.29	0.33	0.14	0.15	0.12	0.14
	Quintile 2	14.8	13.8	0.38	0.38	0.11	0.14	0.12	0.14
	Quintile 3	15.9	14.8	0.48	0.46	0.13	0.13	0.12	0.13
	Quintile 4	17.7	16.0	0.59	0.55	0.13	0.14	0.13	0.14
Urban		14.2	15.2	0.57	0.59	0.10	0.11	0.09	0.09
Suburban		14.7	13.5	0.60	0.56	0.14	0.13	0.13	0.12

*Moments as specified in Aux 2: data vs model, post Act 10.

The upper panel of Table 7 shows model fits for the distribution of ω. Overall, the model fits the data well, although it underpredicts the dispersion of ω_2 and the fraction of districts choosing $\omega_2 = 0$. The lower panel shows model fits for district characteristics by whether or not they reward teacher contribution; these statistics are not directly targeted. Consistent with the data, the model predicts that districts with $\omega_2 > 0$ are slightly more disadvantaged.[54]

Table 7: Model Fit: District Wage Schedules

A. Summary Stats (ω_1,ω_2)	Data	Model		Data	Model
$E(\omega_1)$	0.99	0.99	$E(\omega_2)$	31.3	30.8
$E(\omega_1^2)$	0.98	0.98	$E(\omega_2^2)$	3562.2	3076.6
$E(\omega_1\omega_2)$	30.47	30.46			
$\text{Fr}((\omega_1, \omega_2) = (1, 0))$	0.24	0.30	$\text{Fr}(\omega_2 = 0)$	0.50	0.41
B. District Characteristics by ω_2	$\omega_2 = 0$		$\omega_2 > 0$		
	Data	Model	Data	Model	
Rural	0.80	0.81	0.83	0.81	
λ_d>median	0.48	0.49	0.52	0.50	
Budget/Capacity ($1,000)	51.2	51.1	50.7	50.9	
Incumbent Teachers in d					
Average experience	17.8	17.7	17.6	17.8	
Share w/MA or above	0.57	0.56	0.55	0.55	
Average $10c_1$	0.14	0.14	0.14	0.13	
Average $10c_2$	0.14	0.14	0.13	0.13	

Summary stats of (ω_1, ω_2) and district characteristics by ω_2: data vs model, post Act 10.

6.3 Model Validation

Using the parameter estimates in Table 4, we apply our model to pre-Act 10 data, when districts were restricted to use the rigid wage schedule. We simulate the model under rigid pay and initial conditions from 2010 data. Tables 8 and 9 are counterparts of Tables 5 and 6, and they contrast model-predicted 2010 equilibrium outcomes with 2010 data outcomes. Despite the nontrivial change in the policy environment, our model, estimated using post-Act 10 data, is able to match pre-Act 10 data well. This validation exercise increases our confidence in the model's ability to study counterfactual polices.

[54]Appendix Table A1 shows that the model captures the correlation between ω and district pre-determined conditions as summarized by Aux 3.

Table 8: Model Validation: OLS of Teacher-District Match (pre-Act 10)

Teacher's Choice Set	Inferred Offer Set[a]		All Districts[b]	
	Data	Model	Data	Model
wage	0.001	0.0003	-3$\times10^{-6}$	-3$\times10^{-6}$
e^{λ_d}	-0.006	-0.009	-0.0001	-0.0001
$c_1 \times \lambda_d$	0.33	0.38	0.002	0.003
$I(d \neq d_0)$	-0.96	-0.95	-0.98	-0.95
$I(d \neq d_0) \times$experience	-0.001	-0.001	-0.001	-0.002
$I(z_d \neq z_{d_0})$	-0.002	-0.019	-0.00003	-0.0003
q_d : urban	0.002	0.002	0.001	0.001
q_d : suburban	0.001	0.004	0.0002	0.0006
q_d : large metro	0.016	0.013	0.006	0.011

$a(b)$: OLS specified in Aux 1a (1b), teacher fixed effects included: data vs model, pre Act 10.

Table 9: Model Validation: Average District Employee Characteristics (Pre-Act 10)

		Experience		Share MA or above		$10c_1$		$10c_2$	
District Group		Data	Model	Data	Model	Data	Model	Data	Model
λ_d :	Quintile 1	16.1	15.3	0.56	0.54	0.10	0.11	0.11	0.13
	Quintile 2	16.4	16.1	0.51	0.50	0.10	0.10	0.11	0.12
	Quintile 3	17.6	17.1	0.46	0.47	0.10	0.11	0.13	0.14
	Quintile 4	17.5	17.1	0.52	0.54	0.09	0.10	0.10	0.11
Budget/Capacity:	Quintile 1	13.5	13.8	0.27	0.30	0.10	0.12	0.12	0.14
	Quintile 2	17.7	17.2	0.42	0.43	0.11	0.13	0.12	0.13
	Quintile 3	17.2	16.8	0.52	0.51	0.09	0.09	0.10	0.10
	Quintile 4	18.7	17.9	0.60	0.57	0.08	0.09	0.10	0.11
Urban		15.2	15.2	0.56	0.55	0.14	0.14	0.13	0.13
Suburban		15.6	15.0	0.62	0.60	0.08	0.09	0.10	0.12

*Moments as specified in Aux 2: data vs model, pre Act 10.

7 Counterfactual Experiments

We use our estimated model to first examine the educational equity-efficiency implication of flexible pay, and then to evaluate a set of counterfactual state bonus programs. Let $\Pr(i \text{ in } d|\Upsilon)$ be the equilibrium probability that teacher i works in district d in a given policy environment Υ, we pay special attention to the following metrics:[55]

[55]Teacher-district matching is probabilistic because of shocks $\{\epsilon_d\}$ to teachers' choices, and shocks $\{\eta_\omega\}$ to districts' wage choices. For a given policy, we calculate the expected equilibrium outcomes by numerically

1. Average total contribution among teachers working in a given group of districts $d \in D'$:

$$\frac{\sum_{d\in D'}\sum_i \Pr\left(i \text{ in } d|\Upsilon\right) TC\left(c_i, \lambda_d\right)}{\sum_{d\in D'}\sum_i \Pr\left(i \text{ in } d|\Upsilon\right)}, \tag{M1}$$

where $TC\left(c_i, \lambda_d\right) = c_{i1}\lambda_d + c_{i2}\left(1-\lambda_d\right)$ is teacher i's total contribution to students in d (if i works in d) and the numerator is the expected total contribution among teachers working in D'. The denominator is the expected total number of teachers working in D'.[56] Given that teacher contribution enters student achievement additively, an increase in M1 maps one-to-one into an increase in the average achievement for students in D'. Therefore, when $D' = D$, M1 measures the overall match *efficiency* in the market. Moreover, a policy will improve cross-district educational equity if it increases M1 more for high-λ_d districts, i.e., districts with higher fractions of low-achieving students, than it does for low-λ_d districts.

2. Average teacher contribution to low-achieving students in the state

$$\frac{\sum_d\sum_i \Pr\left(i \text{ in } d|\Upsilon\right) c_{i1}\lambda_d}{\sum_d\sum_i \Pr\left(i \text{ in } d|\Upsilon\right) \lambda_d}, \tag{M2.1}$$

and to high-achieving students in the state

$$\frac{\sum_d\sum_i \Pr\left(i \text{ in } d|\Upsilon\right) c_{i2}\left(1-\lambda_d\right)}{\sum_d\sum_i \Pr\left(i \text{ in } d|\Upsilon\right) \left(1-\lambda_d\right)}, \tag{M2.2}$$

where $c_{i1}\lambda_d$ and $c_{i2}\left(1-\lambda_d\right)$ are teacher i's contribution to low- and high-achieving students in district d (if i works in d), respectively. An increase in M2.1 (M2.2) maps one-to-one to an increase in the average achievement for low-achieving (high-achieving) students in the state. A policy will narrow the achievement gap between the two groups of students if it improves M2.1 more than it does M2.2.

7.1 Flexible Pay versus Rigid Pay

To examine the equity-efficiency implication of a regime switch from rigid pay to flexible pay, we contrast the baseline flexible-pay equilibrium (as described in Section 3) with the counterfactual equilibrium where all initial conditions are kept the same but the rigid wage schedule $\omega = (1, 0)$ is imposed on all districts.[57]

Column 1 of Table 10 presents outcomes in the flexible-pay equilibrium. The first three

integrating over $\{\eta_\omega\}$ and deriving teachers' choice probabilities analytically, detailed in Online Appendix B1.2.1. We use the same set of random shocks throughout our analysis.

[56] In equilibrium, $\sum_i \Pr\left(i \text{ in } d|\Upsilon\right)$ equals to d's capacity κ_d.

[57] One cannot directly measure the impact of flexible pay using the observed differences before and after Act 10, because those differences are contaminated by other factors (e.g., changes in district budgets).

rows report outcomes for all districts: average teacher total contribution to all students in the state (M1), to low-achieving students (M2.1) and to high-achieving students (M2.2). The next two rows report teacher total contribution to students in districts with higher fractions of low-achieving students (M1 for subsets of D). Column 2 reports percentage changes in these metrics associated with a shift from the rigid-pay regime to the flexible-pay regime. With such a shift, we find that 1) average teacher total contribution in the entire state increases by 0.08% (efficiency improves); 2) average teacher contribution to low-achieving students decreases, while contribution to high-achieving students increases, implying an enlarged achievement gap between the two groups of students; and 3) average teacher total contribution decreases in districts with higher fractions of low-achieving students.

Changes shown in Table 10, although small in magnitude, reflect a major trade-off between efficiency and equity.[58] In particular, flexible pay allows districts to directly reward teacher contribution, which encourages comparative advantage-based sorting and hence improves efficiency. However, all else equal, (most) teachers prefer working in districts with more high-achieving students. Under flexible pay, it is even easier for these districts, which also tend to have more resources (Table 1), to attract teachers at the cost of districts with more low-achieving students.[59] As a result, achievement gaps are enlarged across districts and between low- and high-achieving students.

Table 10: Flexible Pay vs Rigid Pay

	Flexible*	$\frac{\text{Flexible-Rigid}}{\text{Rigid}}$ (%)
M1: TC for all students in the state (efficiency)	0.113	0.08
M2.1 c_1 for all low-achieving students	0.115	-0.06
M2.2 c_2 for all high-achieving students	0.112	0.23
M1: TC in top quintile λ_d districts	0.108	-1.04
M1: TC in above median λ_d districts	0.107	-0.57

*Numbers in Column 1 are multiplied by 10 for easier reading.

7.2 State-Funded Bonuses

Results shown in Table 10 suggest that, under flexible pay, there is room for policy interventions favoring districts with more low-achieving students. Given that student composition

[58]The small magnitudes and the equity-efficiency tradeoff in our findings are in line with previous studies on *imposed* performance pay policies. For example, using data from North Carolina, Guarino et al. (2011) find that imposing across-the-board pay for performance based on school results have very small effects on teacher mobility and may exacerbate inequities in distribution of teacher qualification.

[59]Notice that a district has strictly more choices under flexible pay than it does under rigid pay, yet some districts lose in the flexible-pay regime. This highlights the importance of evaluating the policy from an equilibrium perspective.

differs across districts and that teachers differ in their comparative advantages, teacher-district matching is not necessarily a zero-sum game, and such interventions may improve both equity *and* efficiency. In the following, we explore this possibility under flexible pay via a commonly used policy tool: state-funded teacher bonuses.[60]

We focus on the design of the bonus structure and develop two bonus formulae. Under our first formula, a teacher with effectiveness $c = [c_1, c_2]$ teaching in district d would obtain a state-funded bonus given by

$$B_d^1(c) = \min\left\{\max\left\{r_d^1 TC(c, \lambda_d), 0\right\}, \overline{B}\right\}. \tag{B1}$$

Between the lower bound 0 and the upper bound $\overline{B}$,[61] a teacher's bonus is their total contribution $TC(c, \lambda_d)$ multiplied by a district-specific bonus rate r_d^1. Because $TC(c, \lambda_d)$ is higher if a teacher's c better matches the district's student composition (λ_d), B1 incentivizes comparative advantage-based sorting and therefore can improve efficiency. B1 also accounts for equity because bonus rates $\{r_d^1\}_d$ are policy parameters that can be adjusted to provide stronger incentives for effective teachers to teach in disadvantaged districts. Different bonus rate vectors $\{r_d^1\}_d$ would induce different reactions from districts and teachers and hence different equilibrium outcomes.

Our second formula is similar to B1, but with an additional feature:

$$B_d^2(c, \omega_d) = \min\left\{\max\left\{r_d^2 \omega_{d2} TC(c, \lambda_d), 0\right\}, \overline{B}\right\}. \tag{B2}$$

That is, B2 ties bonuses for teachers working in district d to the district's own reward rate for teacher contribution ω_{d2}. District d would obtain more "free money" in the form of state-funded bonuses for its teachers if it chooses a higher ω_{d2}. Therefore, B2 directly incentivizes districts to reward teacher contribution in their own wage schemes.

For illustration, we present equilibrium results from three bonus programs under flexible pay. We calibrate the vector of bonus rates in each program such that all programs are equally costly in the equilibrium, at about \$1,560 per teacher or 10.3 million dollars in total. Given this total cost, the equilibrium average state bonus for each recipient is about \$2,360 or \$3,940, depending on program specifics. These amounts are comparable to relatively mild bonus programs implemented in other states but with very different formulae than ours.[62]

[60] State-funded bonus programs have been used in some states (e.g., North Carolina and California) to reward teachers for teaching in schools with more low-income or low-achieving students.

[61] We consider non-negative state bonuses and set the lower bound at 0 (recall that student test scores are normalized with mean 0 and that $TC(\cdot)$ can be negative). To avoid extreme bonuses, we also impose an upper bound: $\overline{B}$ is twice the standard deviation of the overall wage distribution.

[62] For example, in 2014 dollars, the per recipient bonus was between \$1,910 and \$13,370 in the 1989 Tennessee Career Ladder Evaluation (CLE) program, between \$1,719 and \$3,420 in the 2007 NYC bonus program, and

We start with two flat-rate programs: **B1(flat)** under Formula B1, with $r_d^1 = \$87,550$ for all d;[63] and **B2(flat)** under Formula B2, with $r_d^2 = 3.4$ for all d. The bonus rates are calibrated to exhaust the same pre-specified total cost. Effects of the two programs are presented in the first two columns in Panel A of Table 11. Compared to the baseline flexible-pay equilibrium, B1(flat) leads to a 0.08% improvement in the overall teacher total contribution or efficiency. The gains are similarly shared between low- and high-achieving students. In contrast, B2(flat) leads to a higher efficiency gain at 0.13%. However, most of the gains are enjoyed by high-achieving students, and districts with higher fractions of low-achieving students experience a *decline* in total teacher contribution.

Motivated by the fact that B2(flat) leads to more efficient but more unequal allocation than B1(flat), we conduct a series of experiments under B2 with different vectors of progressive bonus rates, in order to explore possible gains in both equity and efficiency. To be specific, we divide districts into quintiles based on their λ_d (the fraction of low-achieving students), and we experiment with group-specific bonus rates such that r_d^2 (weakly) increases as we move from the lowest-λ_d group to the highest-λ_d group.[64] Among the set of bonus vectors we have tried that satisfy the pre-specified bonus budget, the following delivers the most promising results: **B2(Pro)** under Formula B2, with bonus rates r_d^2 set at 3, 3.25, 3.25, 3.75 and 4.5 for districts in the 1st to 5th quintiles of the λ_d distribution, respectively.

The effect of B2(pro) is shown in the last column in Panel A of Table 11. B2(pro) leads to 0.13% gain in the overall efficiency (the same as B2(flat)). Both types of students gain, with a 0.16% gain for low-achieving students and a 0.10% gain for high-achieving students, implying a narrowed achievement gap. Moreover, districts with higher fractions of low-achieving students enjoy larger gains than an average district in terms of total teacher contribution. Admittedly, these effects are not large, which is in line with findings in other studies that monetary incentives have rather limited effects on attracting and retaining teachers.[65] Nevertheless, counterfactual results under B2(pro) demonstrate that carefully-designed bonus programs can improve both efficiency and equity.

between \$5,500 and \$16,500 in the 2008 Tennessee POINT program (Neal et al., 2011). Findings from these programs are mixed: Math scores improved by 3% under CLE; the NYC bonus program had no effect on achievement; and POINT had no effect on achievement except for one grade, where the effect was positive but only for one year.

[63] In the baseline equilibrium, $TC(c, \lambda_d)$ in realized teacher-district matches has a mean of 0.011 and a standard deviation of 0.027.

[64] An optimal search for district-specific bonus rates would be computationally too burdensome to do, as we would have to solve for the market equilibrium for each vector of bonus rates and also guarantee the total cost of bonuses be the same in the equilibrium.

[65] See, for example, Clotfelter et al. (2008, 2011); Russell (2020).

Table 11: State-Funded Teacher Bonuses

A. Effects on Teacher Contribution[a]				
(%)		$\frac{\text{B1(flat)-Base}}{\text{Base}}$	$\frac{\text{B2(flat)-Base}}{\text{Base}}$	$\frac{\text{B2(pro)-Base}}{\text{Base}}$
M1: TC for all students in the state (efficiency)		0.08	0.13	0.13
M2.1 c_1 for all low-achieving students		0.07	0.02	0.16
M2.2 c_2 for all high-achieving students		0.08	0.25	0.10
M1: TC in top quintile λ_d districts		0.02	-0.89	0.24
M1: TC in above median λ_d districts		0.10	-0.13	0.42
B. Teacher Reward and Program Cost				
	Baseline	B1(flat)	B2(flat)	B2(pro)
Districts choosing $\omega_{2d} > 0$	59%	59%	60%	60%
Teachers rewarded by districts ($\omega_{2d}TC > 0$)[b]	39%	39%	40%	40%
Avg. reward $E(\omega_{2d}TC \mid \omega_{2d}TC > 0)$ ($1,000)	1.29	1.30	1.36	1.36
Teachers receiving state bonuses (B>0)	-	66%	40%	40%
Avg. state bonus E(B|B>0) ($1,000)	-	2.36	3.94	3.94
Program cost ($1,000 per teacher)	-		1.56	

[a]Flexible-pay equilibrium with a given bonus scheme vs baseline flexible-pay equilibrium.

[b]Teacher wage is given by $\max\left\{\min\left\{\omega_1 W_d^0(x) + \omega_2 TC(c, \lambda_d), \overline{w}\right\}, \underline{w}\right\}$

Finally, Panel B of Table 11 shows the equilibrium reward for teacher contribution and program costs. In the baseline, 59% of districts reward teachers for teacher contribution (TC) by setting $\omega_{2d} > 0$; 39% of teachers are rewarded ($\omega_{2d}TC > 0$), with a mean reward of $1,290. There is almost no change in any of these figures under B1(flat). By tying state bonuses to ω_{2d}, both B2(flat) and B2(pro) have some very limited effects on districts' wage choices with 60% of districts setting $\omega_{2d} > 0$; 40% of teachers receive district reward for TC, with a mean of $1,360. The lack of effects on ω_{2d} arises mainly from two costs faced by districts, which may outweigh the small state bonuses we introduce. First, although by increasing its own ω_{2d}, a district can obtain more state bonuses for its effective teachers, it has to reallocate its total wage budget across its teachers with different TC, experience, and education. This distortion can be very costly: A district cares about attracting and retaining teachers of higher values to fill its capacity, where the value is based not only on effectiveness, but also on experience and education. Second, districts also face an resistance cost, which increases with its deviation from $\omega_{2d} = 0$.[66] It should be noted that our findings are better interpreted as short-run policy effects. For example, in the long run, the resistance against deviating from rigid pay may fade off, and state bonus programs may induce larger policy impacts.

[66]As a purely illustrative exercise, in Appendix A3 we show the effect of cutting $R(\cdot)$ to $0.5R(\cdot)$.

8 Conclusion

We have developed an equilibrium model of the teachers' labor market, where teachers differ in their comparative advantages in teaching low- and high-achieving students, and districts compete for teachers using both wage and hiring strategies. We have estimated the model using data from Wisconsin after a reform that gave districts control over teacher pay, and we have validated the model using the pre-reform data under rigid pay.

The estimated model implies that, ceteris paribus, giving districts control over teacher pay would lead to more efficient but also more unequal teacher-district sorting. Efficiency is improved because districts are allowed to directly reward teacher contribution, which encourages comparative advantage-based sorting. Inequality is enlarged because all else equal, (most) teachers prefer working in districts with more high-achieving students, and flexible pay makes it even easier for these districts to attract teachers.

We have further demonstrated that under flexible pay, carefully-designed interventions can improve both equity and efficiency. In particular, progressive state-funded bonus schemes that incentivize comparative advantage-based teacher-district sorting could both improve overall student achievement and narrow the achievement gap between low- and high-achieving students.

We have abstracted from several important aspects of the teachers' market; extending our framework along these lines is worth pursuing. With additional data, the first extension is to incorporate decisions by the private education sector and to consider the competition not only among public school districts, but also between public and private sectors. The second extension is to incorporate household sorting (e.g., Epple and Ferreyra, 2008). Third, the efficiency gains we have found are likely to under-state the total effect of our counterfactual policy intervention, since our model takes teacher effectiveness as pre-determined. An interesting extension is to add teachers' effort choices into our framework. Finally, our static equilibrium model is better suited to study short-run policy effects. An important but very difficult extension is to consider the market in a dynamic equilibrium setting, which would allow for investigation of long-run policy impacts.

References

Atkinson, A., S. Burgess, B. Croxson, P. Gregg, C. Propper, H. Slater, and D. Wilson (2009). Evaluating the impact of performance-related pay for teachers in England. *Labour Economics 16*(3), 251–261.

Aucejo, E. M., P. Coate, J. C. Fruehwirth, S. Kelly, and Z. Mozenter (2019). Match effects in the teacher labor market: Teacher effectiveness and classroom composition.

Behrman, J. R., M. M. Tincani, P. E. Todd, and K. I. Wolpin (2016). Teacher quality in public and private schools under a voucher system: The case of Chile. *Journal of Labor Economics 34*(2), 319–362.

Biasi, B. (2020). The labor market for teachers under different pay schemes. *American Economic Journal: Economic Policy* (forthcoming).

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2005). Explaining the short careers of high-achieving teachers in schools with low-performing students. *American Economic Review 95*(2), 166–171.

Boyd, D., H. Lankford, S. Loeb, and J. Wyckoff (2013). Analyzing the determinants of the matching of public school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics 31*(1), 83–117.

Brehm, M., S. A. Imberman, and M. F. Lovenheim (2017). Achievement effects of individual performance incentives in a teacher merit pay tournament. *Labour Economics 44*, 133–150.

Brown, C. and T. Andrabi (2020). Inducing positive sorting through performance pay: Experimental evidence from Pakistani schools. *University of California at Berkeley Working Paper*.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

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

Clotfelter, C. T., H. F. Ladd, and J. Vigdor (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education review 24*(4), 377–392.

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

Dee, T. S. and J. Wyckoff (2015). Incentives, selection, and teacher performance: Evidence from impact. *Journal of Policy Analysis and Management 34*(2), 267–297.

Dinerstein, M. and T. D. Smith (2016). Quantifying the supply response of private schools to public policies. *University of Chicago Working Paper*.

Dolton, P. and W. v. d. Klaauw (1999). The turnover of teachers: A competing risks explanation. *Review of Economics and Statistics 81*(3), 543–550.

Duflo, E., R. Hanna, and S. P. Rya (2012). Incentives work: Getting teachers to come to school. *American Economic Review 102*(4), 1241–1278.

Epple, D. and M. M. Ferreyra (2008). School finance reform: Assessing general equilibrium effects. *Journal of Public Economics 92*(5-6), 1326–1351.

Epple, D. and H. Sieg (1999). Estimating equilibrium models of local jurisdictions. *Journal of Political Economy 107*(4), 645–681.

Epple, D. N. and R. Romano (2003). Neighborhood schools, choice, and the distribution of educational benefits. In *The economics of school choice*, pp. 227–286. University of Chicago Press.

Feng, L. and T. R. Sass (2018). The impact of incentives to recruit and retain teachers in "hard-to-staff" subjects. *Journal of Policy Analysis and Management 37*(1), 112–135.

Ferreyra, M. M. (2007). Estimating the effects of private school vouchers in multidistrict economies. *American Economic Review 97*(3), 789–817.

Fox, J. T. (2007). Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *The RAND Journal of Economics 38*(4), 1002–1019.

Fox, L. (2016). Playing to teachers' strengths: Using multiple measures of teacher effectiveness to improve teacher assignments. *Education Finance and Policy 11*(1), 70–96.

Fryer, R. G. (2013). Teacher incentives and student achievement: Evidence from New York City Public Schools. *Journal of Labor Economics 31*(2), 373–407.

Gemici, A. (2011). Family migration and labor market outcomes. *Royal Holloway, University of London Working Paper*.

Glewwe, P., N. Ilias, and M. Kremer (2010). Teacher incentives. *American Economic Journal: Applied Economics 2*(3), 205–227.

Graham, B. S., G. Ridder, P. M. Thiemann, and G. Zamarro (2020). Teacher-to-classroom assignment and student achievement. *NBER Working Paper*.

Guarino, C. M., A. B. Brown, and A. E. Wyse (2011). Can districts keep good teachers in the schools that need them most? *Economics of Education Review 30*(5), 962–979.

Hanushek, E. A., J. F. Kain, and S. G. Rivkin (2004). Why public schools lose teachers. *Journal of Human Resources 39*(2), 326–354.

Imberman, S. A. and M. F. Lovenheim (2015). Incentive strength and teacher productivity: Evidence from a group-based teacher incentive pay system. *Review of Economics and Statistics 97*(2), 364–386.

Ingersoll, R. M. (2004). *Why do high-poverty schools have difficulty staffing their classrooms with qualified teachers?*

Jackson, C. K. (2013). Match quality, worker productivity, and worker mobility: Direct evidence from teachers. *Review of Economics and Statistics 95*(4), 1096–1116.

Jackson, C. K. (2018). What do test scores miss? the importance of teacher effects on non–test score outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Jackson, C. K., J. E. Rockoff, and D. O. Staiger (2014). Teacher effects and teacher-related policies. *Annual Review of Economics 6*(1), 801–825.

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *The Future of Children*, 129–153.

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. *NBER Working Paper*.

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

Lang, K. and M. D. Palacios (2018). The determinants of teachers' occupational choice. *NBER Working Paper*.

Lankford, H., S. Loeb, and J. Wyckoff (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational evaluation and policy analysis 24*(1), 37–62.

Lavy, V. (2016). What makes an effective teacher? Quasi-experimental evidence. *CESifo Economic Studies 62*(1), 88–125.

Lavy, V. C. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy 110*(6), 1286–1317.

Lee, D. and K. I. Wolpin (2006). Intersectoral labor mobility and the growth of the service sector. *Econometrica 74*(1), 1–46.

Mansfield, R. K. (2015). Teacher quality and student inequality. *Journal of Labor Economics 33*(3), 751–788.

Meghir, C., R. Narita, and J.-M. Robin (2015). Wages and informality in developing countries. *American Economic Review 105*(4), 1509–46.

Mehta, N. (2017). Competition in public school districts: charter school entry, student sorting, and school input determination. *International Economic Review 58*(4), 1089–1116.

Moe, T. M. (2013). *A primer on America's schools*, Volume 486. Hoover Institution Press.

Muralidharan, K. and V. Sundararaman (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy 119*(1), 39–77.

Neal, D. et al. (2011). The design of performance pay in education. *Handbook of the Economics of Education 4*, 495–550.

Podgursky, M. (2006). Teams versus bureaucracies: Personnel policy, wage-setting, and teacher quality in traditional public, charter, and private schools. *Education Working Paper Archive*.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005). Teachers, schools, and academic achievement. *Econometrica 73*(2), 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review 94*(2), 247–252.

Rothstein, J. (2015). Teacher quality policy when supply matters. *American Economic Review 105*(1), 100–130.

Russell, L. (2020). Effects of the federal Teacher Loan Forgiveness Program on school-level outcomes. *Dartmouth College Working Paper*.

Scafidi, B., D. L. Sjoquist, and T. R. Stinebrickner (2007). Race, poverty, and teacher mobility. *Economics of Education Review 26*(2), 145–159.

Steele, J. L., R. J. Murnane, and J. B. Willett (2010). Do financial incentives help low-performing schools attract and keep academically talented teachers? Evidence from California. *Journal of Policy Analysis and Management 29*(3), 451–478.

Stinebrickner, T. R. (2001a). Compensation policies and teacher decisions. *International Economic Review 42*(3), 751–780.

Stinebrickner, T. R. (2001b). A dynamic model of teacher labor supply. *Journal of Labor Economics 19*(1), 196–230.

Tincani, M. M. (2021). Teacher labor markets, school vouchers and student cognitive achievement: Evidence from Chile. *Quantitative Economics 12*, 173–2016.

Wiswall, M. (2007). Licensing and occupational sorting in the market for teachers. *Unpublished manuscript, Department of Economics, New York University*.

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics 100*, 61–78.

Appendix

A1. Optimal Job Offer Decisions

For a given wage schedule ω, district d's job offers $o_d(x, c, d_0|\omega) \in \{0, 1\}$ solve the following problem:
$\pi_d(\omega) =$

$$\max_{\{o_d(\cdot)\}} \left\{ \int o_d(x, c, d_0|\omega)\, h_d(x, c, d_0|\omega)\left[xb_0 + b_1\lambda_d c_1 + b_2(1-\lambda_d)c_2\right] dF(x, c, d_0) \right\} \tag{15}$$

$$s.t. \int o_d(x, c, d_0|\omega)\, h_d(x, c, d_0|\omega)\, dF(x, c, d_0) \leq \kappa_d,$$

$$\int o_d(x, c, d_0|\omega)\, h_d(x, c, d_0|\omega)\, w_d(x, c|\omega)\, dF(x, c, d_0) \leq M_d$$

$$o_d(x, c, d_0|\omega) = 1 \text{ if } x_1 \geq 3 \text{ and } d_0 = d.$$

Let $\varphi(x,c,\lambda_d) \equiv [xb_0 + b_1\lambda_d c_1 + b_2(1-\lambda_d)c_2]$ be the teacher's marginal contribution to the district's payoff. For non-incumbent teachers and untenured teachers, the first-order condition is

$$\varphi(x,c,\lambda_d) - \nu_\kappa - w_d(x,c|\omega)\,\nu_M = 0,$$

where ν_κ and ν_M are the non-negative multipliers associated with the adjusted capacity and budget constraints. The capacity (budget) is adjusted by netting out the expected slots (wage bill) filled by tenured incumbent teachers ($x_1 \geq 3$ and $d_0 = d$), for whom $o_d(x,c,d_0)$ has to be 1.

If the district makes an offer to (x,c) and if it is accepted, the district must surrender a slot from its limited capacity and pay the wage $w_d(x,c|\omega)$, thus inducing the marginal cost $\nu_\kappa + w_d(x,c|\omega)\,\nu_M$. Balancing between the marginal benefit and the marginal cost, the solution is characterized by:

$$o_d(x,c,d_0|\omega)\begin{cases} = 1 & \text{if } \varphi(x,c,\lambda_d) - \nu_\kappa - w_d(x,c|\omega)\,\nu_M > 0 \\ = 0 & \text{if } \varphi(x,c,\lambda_d) - \nu_\kappa - w_d(x,c|\omega)\,\nu_M < 0 \\ \in [0,1] & \text{if } \varphi(x,c,\lambda_d) - \nu_\kappa - w_d(x,c|\omega)\,\nu_M = 0 \end{cases}, \tag{16}$$

and

$$\int o_d(x,c,d_0|\omega)\,h_d(x,c,d_0|\omega)\,dF(x,c,d_0) \leq \kappa_d, \tag{17}$$

$$\int o_d(x,c,d_0|\omega)\,h_d(x,c,d_0|\omega)\,w_d(x,c|\omega)\,dF(x,c,d_0) \leq M_d. \tag{18}$$

Notice that d_0 affects the optimal job offer decision $o_d(x,c,d_0|\omega)$ only up to tenured incumbent teachers; for other teachers, $o_d(x,c,d_0|\omega)$ is independent of d_0, as seen in (16).

For a given ω, a district's job offer decision can be derived by the following procedure.

0) Set $o_d(x,c,d_0|\omega) = 1$ for teachers with $x_1 \geq 3$ and $d_0 = d$.

1) Guess ν_M, rank other teachers by $\varphi(x,c,\lambda_d) - w_d(x,c|\omega)\,\nu_M$.

2) Give offers to teachers from the top-ranked downwards, until the expected capacity or budget is filled, i.e., (17) or (18) is binding.

3) Calculate the district's objective value associated with this ν_M, and optimize over ν_M to find the maximum; $o_d(\cdot|\omega)$ associated with the optimal ν_M are the optimal job offer decisions under ω.

A2. Model Fit

Table A1 presents the model fit for Auxiliary Model 3 as specified in Section 4.2. The left (right) panel shows the coefficients from the OLS of a district's wage policy ω_{d1} (ω_{d2}) on the composition of its incumbent teachers, district characteristics, and the average TC (as defined by Equation (1)) of teachers in other districts within the same commuting zone. Overall, model predictions are well within the 95% confidence intervals (CI's) of the data estimates; model predictions that are outside of these CI's are marked with asterisks.

Table A1: Model Fit: OLS of District Wage Schedule

Auxiliary Model 3	ω_{d1}		ω_{d2}	
	Data	Model	Data	Model
Composition of incumbent teachers ($d_0 = d$)				
Fr(experience 3-4)[a]	0.02	0.01	24.07	6.15
Fr(experience 5-9)	0.01	0.01	7.54	-0.87
Fr(experience 10-14)	-0.001	0.004	17.28	1.94
Fr(experience $\geq$ 15)	0.03	-0.0001	-13.48	1.70
Fr(MA or above)	-0.03	-0.005*	-11.41	-2.16
Average TC	-0.62	0.84	532.71	90.31
Std dev. TC	-0.19	-0.06	-468.43	-33.93*
Average TC among the Tenured	0.49	0.76	-770.68	-255.40
District Characteristics				
λ_d	0.01	0.03	17.52	4.16
budget per teacher	0.002	0.001*	0.57	0.002
capacity	-0.00001	0.0002	-0.36	-0.03*
urban	-0.03	-0.01	19.64	2.84
suburban	-0.02	-0.002*	-1.26	1.98
Teachers in nearby districts ($z_{d_0} = z_d$, $d_0 \neq d$)				
Average TC	-1.04	0.01	1290.30	237.39
# obs.	411		411	

OLS as specified in Aux 3: data vs model, post Act 10.

[a]The lowest experience of incumbent teachers is 1 to 2 years. Experience=0 teachers have d_0=0.

* Outside of the 95% confidence interval of the estimates from the data.

A3. Illustrating the Effect of Resistance Costs $R(\cdot)$

We examine the effect of reducing the resistance cost, which is for illustration purposes only because our model is silent on what underlies $R(\cdot)$. Column 1 of Table A2 shows the effect (relative to the baseline flexible-pay equilibrium) of reducing $R(\cdot)$ to $0.5R(\cdot)$: The overall efficiency would improve by 0.02%; the fraction of districts rewarding teacher contribution would increase from 59% to 75%. Column 2 shows the effect (relative to the baseline) of B1(flat) combined with reducing $R(\cdot)$ to $0.5R(\cdot)$. For comparison, Column 3 repeats the effect of B1(flat) under the original $R(\cdot)$ (i.e., it is the same as Column 1 in Table 11). Relative to B1(flat) under $R(\cdot)$, Column 2 shows that combining B1(flat) with a 50% cut in the resistance cost would lead to a larger gain in efficiency (0.11% vs. 0.08%), but it would *reduce* teacher total contribution in high-λ_d districts.

Table A2: The Effect of Reducing $R(\cdot)$: An Illustration

A. Effects on Teacher Contribution				
(%)		$\frac{0.5R(\cdot)\text{-Base}}{\text{Base}}$[a]	$\frac{\text{B1(flat)\&}0.5R(\cdot)\text{-Base}}{\text{Base}}$[b]	$\frac{\text{B1(flat)-Base}}{\text{Base}}$[c]
M1: TC for all students in the state (efficiency)		0.02	0.11	0.08
M2.1 c_1 for all low-achieving students		-0.03	0.02	0.07
M2.2 c_2 for all high-achieving students		0.07	0.20	0.08
M1: TC in top quintile λ_d districts		-0.28	-0.39	0.02
M1: TC in above median λ_d districts		-0.21	-0.23	0.10
B. Teacher Reward and Program Cost				
	Base	$0.5R(\cdot)^a$	B1(flat)&$0.5R(\cdot)^b$	B1(flat)^c
Districts choosing $\omega_{2d} > 0$	59%	75%	76%	59%
Teachers rewarded by districts ($\omega_{2d}TC > 0$)	39%	50%	51%	39%
Avg. reward E($\omega_{2d}TC \mid \omega_{2d}TC > 0$) ($1,000)	1.29	1.67	1.92	1.30
Teachers receiving state bonuses (B>0)	-	-	66%	66%
Avg. state bonus E(B|B>0) ($1,000)	-	-	2.36	2.36
Program cost ($1,000 per teacher)	-	-	1.56	

[a] Flexible-pay equilibrium under $0.5R(\cdot)$ vs. the baseline equilibrium.

[b] Flexible-pay equilibrium under B1(flat) and $0.5R(\cdot)$ vs. the baseline equilibrium.

[c] Flexible-pay equilibrium under B1(flat) and $R(\cdot)$ vs. the baseline equilibrium.(Column 1 in Table 11)