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

Barbara Biasi, Chao Fu and John Stromme†

April 18, 2022

**Abstract**

Proper allocation of public servants across local employers is often hampered by a major institutional friction: wage rigidity. Through the lens of the market for public-school teachers, we study the equilibrium equity-efficiency implication of this friction. In our model, 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**: Wage Rigidity, Equilibrium Sorting, Education Efficiency and Equity, Teachers' Comparative Advantages, Structural Estimation

*We thank Manuel Arellano, Xiaoxia Shi, Chris Taber, Matt Wiswall, seminar participants at Bocconi, Cornell, Johns Hopkins, Penn State, Toulouse, Tsinghua, U Chicago, UPenn and Yale, and conference participants at AEA, SEA, Barcelona GSE Summer Forum, and International Symposium on Labor Economics (China) 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

Labor market efficiency relies on the proper allocation of workers across employers. Toward this end, labor market prices (i.e., wages or compensation in general) play the role of the invisible hand and incentivize workers to sort into firms where they are most productive. However, a large share of the workforce is employed in markets where the link between a worker's productivity and their wage is rather weak, often due to the influence of labor unions that compress wage dispersion. This is most evident in markets for public servants, such as police officers, transport workers, and teachers. In these markets, institutional rules (e.g., collective bargaining) compel employers—such as local enforcement agencies, local transport authorities, and school districts—to compensate workers according to rigid (typically seniority-based) schedules.

This occurs despite the fact that workers might differ in their comparative advantages working for different employers. Teachers are a salient example: Some of them may be better at stimulating high-achieving students, while others at helping low-achieving students. Therefore, it would be most efficient if teachers sorted into teaching students according to teachers' comparative advantages (Roy, 1951). Unfortunately, pay for most U.S. public school teachers fails to incentivize such sorting: It follows rigid experience-education schedules, a regime we call "rigid pay." Although districts serve different student bodies, under rigid pay they cannot use salary schemes to attract teachers better suited for their students. 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 (e.g., Lankford et al., 2002; Ingersoll, 2004; Clotfelter et al., 2005; Mansfield, 2015; Jacob, 2007). Such vertical sorting can lead to both efficiency losses and large educational inequalities across children from different backgrounds.

An alternative arrangement would be one where districts have the flexibility to design their own teacher pay schedules, a regime we label "flexible pay."[1] This paper investigates the implication of flexible pay 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 pay schemes can incentivize teachers to move across jobs. The second is school districts' preferences over various attributes of a teacher, which govern districts' hiring decisions and, if given the flexibility, their choices of teacher pay schedules. The third factor is competition among districts for teachers, which needs to be

[1] 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.

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 observed teacher pay schedules. Due to pattern bargaining by a state's teachers' union, very similar and rigid pay schedules are often imposed on all districts in the state. This has made it difficult to infer districts' preferences, let alone how districts 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.[2] A district cares about a teacher's contribution to its students' achievement, and it 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 achievement of the district's students, 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 bodies differ across districts and teachers differ in their comparative advantages in teaching certain types of students, teacher-district sorting 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 even easier to attract teachers with absolute advantages in teaching both types of students. This 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.

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

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. Our identification argument, which guides our choice of auxiliary models used in our indirect-inference estimation, is as follows. First, under Act 10, districts have control over teacher pay. Therefore, we can learn about districts' preferences over teachers from the degree to which districts' observed pay schedules favor or disfavor certain groups of teachers with different attributes (experience, education, and effectiveness) and how these schedules vary with districts' characteristics. Second, the observed teacher-district matches are informative of both teachers' and districts' preferences. With the mild assumption that district preferences for teachers are weakly increasing in teacher attributes, we can infer from an observed district-teacher match $(d, i)$ 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, and it informs us of district preferences.

We apply our model to administrative data from the Wisconsin Department of Public Instruction, which consists of three linked panel data sets 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 students with low and high prior scores. 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 *the same initial conditions*, 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 two counterfactual bonus formulae (B1 and B2) of state-funded teacher bonus programs 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. This can incentivize more efficient sorting because a teacher's contribution (and thus bonus) is higher when their comparative advantage better matches a district's student composition. To account for equity, we adjust bonus rates based on a district's student composition. Relative to formula B1, formula B2 additionally ties the state bonus to a district's wage schedule, such that districts are incentivized to increase their own reward for teachers' contribution to student achievement. We find that, at the same total cost, B2 programs with progressive bonus rates that favor 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 would improve more in districts with higher fractions of low-achieving students than it would in an average district. In summary, under flexible pay, carefully designed interventions can induce more efficient and more equal teacher-district sorting. Quantitatively, the impacts of our bonus programs are small. Additional counterfactual simulations suggest that the effectiveness of these policy interventions hinges on teachers' willingness to move and districts' willingness to change their wage schedules.

**Related Literature** Proper allocation of public servants across local employers can have important implications for both efficiency and equity. Unfortunately, a socially optimal allocation is hampered by various institutional frictions. Through the lens of the labor market for public school teachers, our paper contributes toward a better understanding of this issue by showing how wage rigidity—a major institutional friction—impacts the efficiency of the allocation of workers to employers and equity. This is a general problem that arises in many settings besides education, including law enforcement, healthcare, and other forms of public service. For example, Ba et al. (2021) show that although police officers' effectiveness in reducing crimes increases with experience, more experienced officers tend to work in low-crime areas in the presence of wage rigidity and seniority-based priority in the centralized assignment process. This hurts not only the equity between high-crime and low-crime areas, but also the efficiency in aggregate crime reduction.

More specifically, our paper contributes to an extensive body of work on the labor market for teachers. Given its goal of evaluating counterfactual policies, our paper is closest to those studying this market through the lens of a structural model. A large subset of these studies focus on the supply side. For example, Stinebrickner (2001a), Stinebrickner (2001b), Wiswall (2007), and 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. Using competing risks models, Dolton and Klaauw (1999) study teachers' decision to leave the profession. Boyd et al. (2005) and Scafidi et al. (2007) study teachers' preferences for schools and find that teachers prefer schools with fewer low-achieving and minority students.

A smaller subset of studies consider both sides of the market. Using data from Peru, Bobba et al. (2021) study how policy measures such as wage bonuses affect teacher sorting. Differently from our context, their setting features centralized applications and teacher assignments; schools have capacity constraints but are not active decision makers. 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 with rigid pay, 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.[3] 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, 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 disadvantaged 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., 2011; Russell, 2020). Biasi (2021) shows that, under Wisconsin Act 10, higher-quality teachers tend to move to districts that adopted flexible pay.[4] Building on this literature, we develop and estimate an equilibrium model to understand districts' and teachers' preferences that underlie the observed outcomes and 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

[3]With focuses different than ours, Mehta (2017) estimates an equilibrium model of charter school entry, school inputs, and students' school choices; Dinerstein and Smith (2016) study private schools' responses to public school funding policies.

[4]Using field experiments in non-US settings, Brown and Andrabi (2020) find that performance pay induced positive teacher sorting, while Leaver et al. (2021) find that it improved teacher effort without significant effects on selection.

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 could have sizable effects on teachers' contribution to learning.[5] In a very recent paper, Bates et al. (2022) study teacher-school allocation within a district, allowing teacher valued added to differ for advantaged and disadvantaged students. They estimate teachers' preferences over various non-wage aspects of a school (given the lack of wage variation) and schools' preferences over teachers. Assuming pair-wise stable teacher-school matching, they find that a meaningfully more efficient allocation can be achieved by directly affecting teachers' preferences over schools. In this paper, we are interested in exploring how policy tools such as teacher bonuses can induce more efficient teacher-district sorting in a market equilibrium setting where districts compete for teachers using both wage and job offer strategies.

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 without collective bargaining (CB), these schedules are typically determined at the state level (e.g. Georgia). Most states use CB, where these schedules are negotiated between school districts and teachers' unions. 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 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. Act 10 led to major reforms

[5] Recent studies have also considered heterogeneity in teacher effectiveness by student demographics (Lavy, 2016; Bates et al., 2022) and by subjects (Fox, 2016).

to public-sector employment in the state. 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. 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."

Act 10 also introduced a series of other provisions, applied uniformly to all school districts in the state. First, Act 10 reduced employees' benefits via an increase in employee contributions to pensions 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 from 83% in 2011 to 45% in 2016. 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).

## 2.1 A Glance at the Market Before and After Act 10

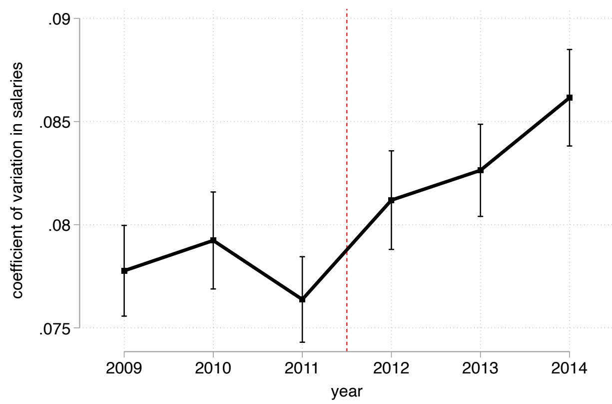
We provide a first glance at the labor market for public school teachers in Wisconsin before and after Act 10, using data from the Wisconsin Department of Public Instruction. The data, which we describe in detail in Section 5, consist of three linked data sets at the teacher, student, and district level,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

**Variation in Teacher Salaries:** As a measure of teacher wage variation, we use the coefficient of variation (CV) obtained from a regression of wages on district-by-year and seniority-by-education fixed effects. Figure 1 shows that, prior to Act 10, teacher wage variation was almost nonexistent within each district among teachers with similar experience and education. 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 year fixed effects.[6] Before Act 10, real wage growth was small and negative for both movers and stayers. After Act 10, wage growth remained small and negative for stayers but significantly

[6]Specifically, letting $w_{it}$ 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 $\varpi_{it}$ from this regression. Letting $\Delta_{it} = \varpi_{it} - \varpi_{it-1}$, Figure 2b shows the median $\Delta_{it}$ among those who moved across districts in $t$ and the median $\Delta_{it}$ among those who stayed in the same district between $t-1$ and $t$, $t = 2010$ and 2014 are shown as examples.

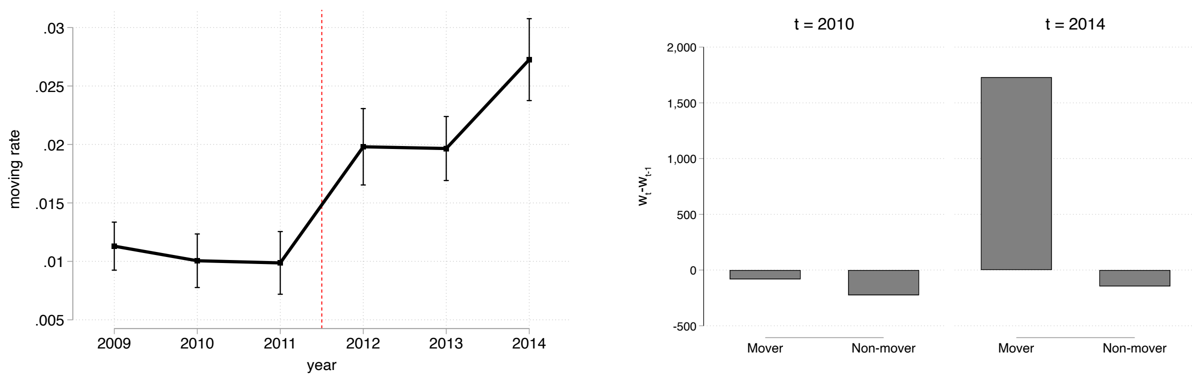
Figure 1: Variation in Teacher Salaries




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

increased for movers ($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

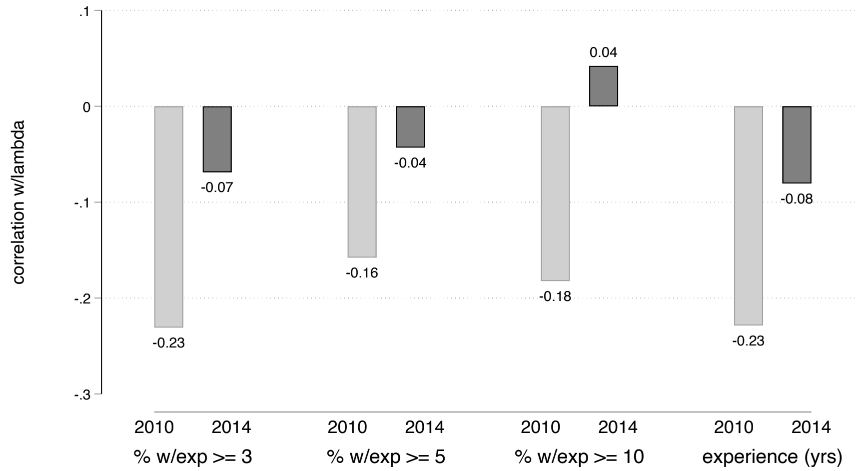
(b) Relative Wage Growth

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

**Teacher-District Sorting:** Prior to Act 10, teachers with higher experience, who tend to be more effective (Wiswall, 2013), were significantly less likely to work in districts where a larger fraction of students are low achieving. Figure 3 shows that the fractions of teachers with experience higher than 3, 5, and 10 years and the average teacher experience of each district were negatively correlated with the fraction of low-achieving students (those with math scores below the state median) in the district. These relationships became much weaker after Act 10.

Figures 1 to 3 provide some suggestive evidence that, under flexible pay, districts used wage

Figure 3: Vertical Sorting: Teacher Experience and Share of Low Students




Correlations between each variable and the share of each district's students who are low-achieving.

strategies to compete for teachers and teacher-district sorting became less vertical. However, 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. To isolate the equilibrium impact of replacing rigid pay with flexible pay and, more importantly, to conduct counterfactual policy analysis, we build 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 $(x, c, d_0)$. The vector $x = [x_1, x_2]$ includes experience and education; $c = [c_1, c_2]$ is one's effectiveness in teaching low- and high-achieving students; $d_0$ is the district one works in at the beginning of the model, where $d_0 \in \{1, ..., D\}$ for incumbent teachers and $d_0 = 0$ for those who are yet to find a job on this market (e.g., new teachers).

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

**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 student composition $\lambda_d$.
**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.[7] 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 $\epsilon_d$ is an i.i.d. Type 1 extreme-value distributed taste shock with a scale parameter $\sigma_\epsilon$. Wage enters with a normalized coefficient of 1, so that teachers' preferences are measured in \$1,000. Teachers' preferences for district characteristics $q_d$ are governed by the vector $\theta_0$. The next two terms capture teachers' preferences for student composition $(\lambda_d)$; these preferences may vary across teachers with different effectiveness in teaching low-achieving students $(c_1)$.[8] $\Gamma(\cdot)$ 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 yet 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,

[7] 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.

[8] We use $e^{\lambda_d}$ in (2) because it is disproportionately rare to see teachers move into districts with a high fraction of low-achieving students, suggesting that teachers' preference over $\lambda_d$ might be convex; indeed, we have estimated a model with $\lambda_d$ instead of $e^{\lambda_d}$, which does not fit the data well. We only include the interaction $\lambda_d c_1$ in (2) because adding the interaction $c_2\lambda_d$ does not improve the fit.

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

#### 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 (gross) 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 $b_0$ allows for the possibility that districts may directly care about teacher experience and education, and $b_1$ and $b_2$ capture how much a district cares about a teacher's contribution to its low- and high-achieving students, respectively.[9] 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 $\lambda_d$; 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 $TC(\cdot)$:

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

[9] 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 assumed the latter away.

To avoid unrealistically high or low wages, we bound wages by $[\underline{w}, \overline{w}]$, such that[10]

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

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, it 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 the case in most U.S. school districts; if $\omega_2 > 0$, teachers are rewarded for their contribution, echoing the idea of performance pay. As we show in Section 5.1.3, wages calculated under (6) match the observed wages very well. We have also tried a more flexible wage rule that rewards $c_1$ and $c_2$ differently. However, wages predicted by this more flexible wage rule are very similar to those predicted by (6) (Online Appendix B2.4.3). Therefore, we choose the more parsimonious specification (6).

### 3.2.3 District's Optimal Decisions

Taking all the other districts' policies and teachers' decision rules as given, a district fills 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: A district first chooses a wage schedule $\omega = (\omega_1, \omega_2)$, and then makes job offers conditional on $\omega$. We solve a district's problem via backward induction.

**Job Offers** For a given wage schedule $\omega$, district $d$'s job offers $\{o_d(x, c, d_0|\omega)\}_{(x,c,d_0)}$ maximize the following total value from teachers it expects to hire:

$$\pi_d(\omega) = \max_{\{o_d(\cdot)\}} \left\{ \int o_d(x, c, d_0|\omega) h_d(x, c, d_0, \omega) [xb_0 + b_1 \lambda_d c_1 + b_2(1 - \lambda_d) c_2] dF(x, c, d_0) \right\} \tag{7}$$

$$\begin{aligned} s.t. \int o_d(x, c, d_0|\omega) h_d(x, c, d_0, \omega) dF(x, c, d_0) &\leq \kappa_d, \\ \int o_d(x, c, d_0|\omega) h_d(x, c, d_0, \omega) w_d(x, c|\omega) dF(x, c, d_0) &\leq M_d \\ o_d(x, c, d_0|\omega) = 1 \text{ if } x_1 \geq 3 \text{ and } d_0 = d&, \end{aligned}$$

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

[10]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 $\omega$. 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 $\omega$ 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, and $r_1$ and $r_2$ capture the incremental costs for larger deviations. Finally, $\eta_\omega$ is an i.i.d. extreme-value distributed shock associated with choosing $\omega$, with a scale parameter $\sigma_\eta$.

## 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* $(x, c, d_0, \epsilon)$.

*2) For all d, given* $\{h_d^*(\cdot)\}$, $\omega_d^*$ *is an optimal wage decision and* $\{o_d^*(\cdot|\omega)\}$ *are optimal job offer decisions under* $\omega$.

*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 $\{h_d(x, c, d_0, \omega)\}$: 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.[11] As a feasible alternative, we assume that districts make their decisions based on a simplified parametric belief about teachers' acceptance probabilities,[12] given by

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

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

This simplified belief function captures all the factors governing its counterpart $\{h_d(\cdot)\}$ defined in (8). The first two terms in (12) relate to the overall desirability of the teacher; a district should expect more competitors for a better teacher and therefore a lower acceptance probability. 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 districts' characteristics as in (2) and 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.

[11] The dimensionality of $\{h_d(\cdot)\}$ is $I \times D \times N_w$ ($I$, $D$ and $N_w$ are the numbers of teachers, districts and potential wage levels, respectively). Alternatively, a district can derive $\{h_d(\cdot)\}$ from Equation (8) with its belief about other districts' strategies. With 411 districts in the market, forming the exact belief about other districts' wage strategies $\{(\omega_{d1}, \omega_{d2})\}$ and offer strategies is also a daunting task.

[12] Similar simplification 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).

### 3.4 Model 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 competition among districts and abstract from their competition against teachers' outside options (e.g., private schools, public schools in other states, and other occupations). For the same reason, although new teachers who ended up working in Wisconsin public schools are included in our sample, we do not model teachers' decisions to enter or exit the market and we take the initial distribution of teachers in 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. Some studies (e.g., Rothstein, 2015) suggest that the effect of performance pay on selection is very small, while some other studies suggest that performance pay in public schools may improve the quality of the overall supply of teachers in both public and private schools (e.g., Tincani, 2021). Although we cannot be certain about how incorporating teacher entry/exit may affect our findings, the efficiency gain we find in our counterfactual policy experiments could be understated.

Second, because wage schedules are set at the district level, we focus on the competition across districts and abstract from the allocation of teachers across schools within a district. Online Appendix B3 shows that the cross-district variation in teacher wages and student bodies clearly dominate the within-district variation.[13] Moreover, implementing the tests proposed by Chetty et al. (2014) and Rothstein (2010), we find no evidence of non-random sorting of teachers across grade-schools within a district.[14] 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 $\lambda_d$ as given. In particular, we assume away potential households re-sorting across districts in response to our policy interventions. In our data, the fraction of students moving across districts was very small and similarly so before and after Act 10; this is true for moves between any two districts and for moves between a district that rewarded teacher effectiveness under Act 10 and one that did not.[15] Our counterfactual policies would change the baseline environment (post-Act 10 Wisconsin) only by the addition of teacher bonuses. This intervention is milder than the introduction of

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

[14] We test for the presence of non-random sorting of teachers across grade-schools by correlating changes in grade-school average $c_1$ and $c_2$ with changes in students' *lagged* test scores (conditional on observables). These correlations are statistically indistinguishable from zero (Online Appendix B3.3).

[15] Between 2007 and 2016, 4.4% of Grades 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. Labeling districts as adopting and non-adopting by whether or not they chose to reward teacher effectiveness ($\omega_2 > 0$ vs $\omega_2 = 0$) after Act 10, the fraction of students moving from non-adopting districts to adopting districts was also stable at 0.8% in 2010, 0.8% in 2011, 0.8% in 2012 and 0.9% in 2013.

Act 10 to the market. Therefore, we do not expect our counterfactual policies to significantly affect households' location choices. However, readers should still be aware of this limitation when 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 with mixed findings.[16] We complement this literature by focusing on a different channel through which financial incentives may improve education, i.e., 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 could 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 the data and $\widehat{\beta}(\Theta)$ denote the corresponding auxiliary model parameters obtained from simulating a large dataset from the model (parameterized by $\Theta$) and computing the same estimators. 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 $\Theta$, which consists of teachers' and districts' preference parameters, and an inner loop solving the model for each given $\Theta$ (detailed in Online Appendix B1). In our counterfactual policy simulations, we need to find the fixed point for the belief parameter vector $\zeta$ *and* wage statistics $\{\overline{w}(\cdot), \sigma_w(\cdot)\}$ that enter the belief function, but we only need to find the fixed point for $\zeta$ during the estimation. Assuming that the data were generated from an equilibrium, the realized equilibrium $\{\overline{w}^o(\cdot)\}$ and $\{\sigma_w^o(\cdot)\}$ can be derived directly from the *observed* wage schedules $\{\omega_d^o\}_d$ (the superscript $o$ denotes "observed"). Therefore, we can plug $\{\overline{w}^o(\cdot), \sigma_w^o(\cdot)\}$ into Equation (12) and search

[16] 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., Fryer (2013); Imberman and Lovenheim (2015); Dee and Wyckoff (2015); Brehm et al. (2017).

only for the fixed point for $\zeta$ during estimation.

## 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. The observed wage schedules and matches, however, contain rich information that allow us to overcome this obstacle, as we argue below. These arguments guide our choice of auxiliary models.

### 4.1.1 Wage Schedule and District Pre-Determined Conditions

Under Act 10, districts can choose how to reward teachers. Therefore, the observed wage schedules provide the first major source of information for identification: One can learn about districts' preferences from the extent to which their wage schedules favor teachers with different attributes $(x, c)$ and how wage schedules relate to districts' pre-determined conditions. To see the intuition, notice that wage schedules can serve to both pull and 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 stronger when these teachers are not district incumbents, because moving is costly for teachers. 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 district $d$ can avoid teachers with $(x', c')$ who are not $d$'s tenured incumbents 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 $(x', c')$.

However, district preferences over teachers may not be sufficient to explain the observed wage schedule choices. For example, in our post-Act 10 data, 24% of districts kept their pre-Act 10 wage schedules and 50% of districts chose not to reward teacher effectiveness. It is hard to rationalize these mass points as optimal wage schedules chosen by districts purely to hire their preferred teachers. Districts' choices that are not explained by their preferences for teachers are attributed to the resistance cost $R(\cdot)$.

### 4.1.2 Optimal Job Offers and Observed Matches

The observed teacher-district matches provide the second major source of information for identification. On the teacher side, identification of preferences would be straightforward if we observed what options were available for each teacher, because choices out of (multiple) feasible options reveal preferences; observing only accepted offers complicates the inference.

However, we can use observed matches to infer a subset of offers each teacher received by exploiting districts' optimal job offer decisions. Specifically, for district $d$, the marginal benefit of hiring a teacher consists of the teacher's 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 experience and education $x$. The marginal cost consists of teacher wage $w_d\,(x, c|\omega_d^o)$ (calculated using wage rule (6) at the observed schedule $\omega_d^o$) plus the shadow price of a slot. If $d$ 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 but not necessary) 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 $i$,[17] 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 $O_i^s$, consisting of the 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 (which is true for 5,170 out of 6,600 teachers in our sample), a teacher's choice within $O_i^s$ informs us of teacher preferences, since all options in $O_i^s$ were feasible.[18]

We can also leverage observed offers in a different way to learn about district preferences. For each teacher $i$, the entire set of districts $D$ is the union of $i$'s full offer set $O_i \equiv \{d : o_d\,(x_i, c_i, d_{0i}) = 1\}$, of which $O_i^s$ is a subset, and non-offer set $D \backslash O_i$.[19] If one were to infer teacher preferences under the (false) assumption that every teacher had offers from *all* districts, the inferred "preferences" would be contaminated by the existence of infeasible choices ($D \backslash O_i$) in a teacher's "choice set", and thus would be different from preferences inferred from choices within $O_i^s$. The discrepancy between the two sets of inferred teacher preferences depends on the composition of the non-offer set $D \backslash O_i$ for each teacher $i$. The non-offer set $D \backslash O_i$ in turn is governed by districts' preferences over teachers. For example, 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 $(\lambda_d)$. Therefore, we can learn about districts' preferences from the aforementioned discrepancy: Districts' preference parameters have to generate not only the observed offers, but also the lack of offers from certain districts to certain teachers ($D \backslash O_i$ for each $i$) that would reconcile this discrepancy.

[17] 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 $(x_1)$ is based on these categories.

[18] Multinomial discrete-choice models can be point-identified using a subset of choices, parametrically (e.g., McFadden, 1978) and semiparametrically (e.g., Fox, 2007). In a framework much more flexible than ours, Barseghyan et al. (2021) allow for *unrestricted correlation* between choice sets and preferences and characterize the sharp identification region of model parameters. We build on insights from these studies to design our auxiliary model Aux 1a (Section 4.2), which is used to extract information useful for identification.

[19] The inference procedure above identifies $O_i^s$, but not $O_i$ or $D \backslash O_i$.

**Discussion** 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 abstract away from the latter.[20] As a robustness check, we conduct the following exercise: Instead of $(c_1, c_2)$, districts observe $(c_1 + err_1, c_2 + err_2)$ and make wage and job offer decisions based on these noisy measures. Assuming that $err_k$ are i.i.d., normally distributed noise terms for $k = 1, 2$, we repeat the procedure described in Section 4.1.2 to construct subsets of offers for each teacher and re-estimate of our key auxiliary models that summarize teachers' choices within these subsets. These auxiliary models are robust to this simple form of information friction (Online Appendix B4).
A2: Districts cannot discriminate among teachers by factors other than $(x, c)$. 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 bias in the inferred teacher preferences based on $O_i^s$. However, as long as most job offers are based on $(x, c)$, the essence of our identification strategy still holds: 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. As a robustness check, we re-estimate our key auxiliary models but do not use observed teacher-district $(i, d)$ matches to infer offers for other teachers if $i$'s effectiveness (either $c_1$ or $c_2$) is below the 10th percentile among all teachers, since these ineffective teachers may indeed have been hired for other reasons. Doing so significantly affects the number of inferred offers for other teachers; yet our auxiliary models remain robust.
A3: We assume away job posting costs. This assumption is plausible because in reality districts post openings publicly on online platforms.[21] We also assume that teachers get offers without having to apply. This assumption does not affect our inference of teacher preferences because 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 it; 2) $j$ was eligible for a job in $d$ but did not apply. 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.

## 4.2 Auxiliary Models

Following the identification argument, we target the following auxiliary models *jointly*. Notice that, although certain auxiliary models are intuitively more informative about certain structural parameters than others (as we explained above), the identification of the model relies on using

[20]In a centralized student-school matching system, Fack et al. (2019) define a student's feasible choice set as schools whose observed ex post admission cutoffs are below the student's priority index; they estimate students' preferences assuming stability and, like we do, assuming complete information (i.e., students can perfectly forecast school-specific admission cutoffs).

[21]See, e.g., https://wecan.education.wisc.edu (Wisconsin Education Career Access Network).

information extracted from all auxiliary models. To provide more evidence on the mapping between data and parameters, in Online Appendix B5 we follow Einav et al. (2018) and perturb structural parameters one by one and measure the responses of the predicted auxiliary models.

Aux 1 Coefficients from two regressions of the following form

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

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). $\psi_i$ is a teacher dummy that relates all $\left\{(i,d)\right\}_d$ observations associated with teacher $i$.[22] The two regressions differ in the number of observations, reflecting the identification argument in Section 4.1.2.

Aux 1a The first regression includes all teachers whose inferred subsets $O_i^s$ contain more than one offer; an observation $(i, d)$ is a teacher-district pair in these inferred subsets. Importantly, we use the *same procedure* to construct $O_i^s$ in the actual data and in the simulated data.

Aux 1b The second regression includes every possible teacher-district pair. Aux 1a is informative of teacher preferences, whereas the *difference* between Aux 1b and Aux 1a (rather than Aux 1b in itself) is informative of districts' preferences. Since Aux 1a only includes teachers whose inferred subsets $O_i^s$ contain more than one offer, Aux 1a and Aux 1b are estimated using slightly different samples of teachers; however, estimates of Aux 1b are very similar when we restrict the sample to teachers included in Aux 1a.

Aux 2 Moments of district-level teacher characteristics $(x, c_1, c_2)$ by district groups (quintiles of $\lambda_d$, quintiles of budget per slot, and urban/suburban status).

Aux 3 Coefficients from regressions of wage schedule $\omega_{dn}$, $n = 1, 2$, on district's pre-determined conditions, reflecting the identification argument in Section 4.1.1:

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

where coefficients $\beta_{1n}^w$ to $\beta_{4n}^w$ are associated with district characteristics and constraints, and $\beta_{5n}^w$ to $\beta_{8n}^w$ are associated with the composition of district incumbents. In particular,

[22] 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 instead use a linear regression with teacher dummies. These dummies (not targeted) serve to capture the idea that the same teacher is choosing one district out of a given set of districts.

$X_d$ is the average $x$, $\overline{TC}_d$ and $\sigma_{TC_d}$ are the average and 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 $z_{d_{0i}} = z_d$).[23]

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

***Teacher-Level Data*** (PI-1202 Fall Staff Report) cover all individuals employed by the WDPI between 2006 and 2016. 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 demographics and state standardized test scores for all public school students in Grades 3 to 8 between 2007 and 2016.

***District-level Information:*** Using student test score data, we calculate $\lambda_d$, 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 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 from 2014, i.e., 3 years after Act 10; by then, all the CB agreements pre-dating

[23] 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 nearby districts in our final specification of Aux 3, because they are insignificant when included.

Act 10 had expired and districts had obtained full autonomy over teacher pay.[24] 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 teachers because they face different types of contracts than regular, full-time teachers.[25] We exclude secondary-school teachers because they often teach multiple grades, making it hard to identify individual teacher contributions (Kane and Staiger, 2008; Chetty et al., 2014). Among elementary-school teachers, effectiveness measures are obtainable for teachers in Grades 4-6; we restrict attention to those teaching the same subject (math), so that the effectiveness measures are comparable across teachers.[26] The estimation sample contains 411 districts and 6,600 teachers; 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 $i$'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:

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

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 $t$; $\tau_k$ 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; $\tau_k = 2$ 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

[24] Biasi (2021) shows that teacher exits surged in 2012 but had stabilized by 2014.

[25] 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.

[26] We need student test scores from the previous year to calculate teachers' effectiveness and students are tested starting 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).

experience in year $t$ and $v_{in}$ is the part unexplained by $x_{it}$. Assuming $\varepsilon_{kt}$ is an i.i.d. idiosyncratic component, we estimate $\gamma$, $\rho_1$ and $\rho_2$ via OLS using data from 2007 to 2016; then, we use the Bayes estimator of Kane and Staiger (2008) to estimate $v_{i1}$ and $v_{i2}$. 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.[27]

Two features of our achievement model deserve further discussion (see Online Appendix B2.3 for details). First, we focus on teachers' comparative advantages in terms of $(c_1, c_2)$ because our two-dimensional effectiveness model explains approximately 20% more variation in test scores compared to the one-dimensional effectiveness model. In contrast, if we add, for example, a teacher's race and its interaction with student race to the achievement model, the interaction terms are indistinguishable from zero.

Second, besides modeling $c$ as being two-dimensional, we also allow $c$ to vary directly with $x$, because experience has been shown to affect teacher effectiveness (e.g., Rockoff, 2004; Wiswall, 2013). To estimate effectiveness with this feature, we have to assume that a teacher contributes to all students in their school-grade in (13) because we can link students and teachers only up to the school-grade level. In an alternative 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 $\varepsilon_{kt}$ 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.[28] 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 the identification of 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. We follow the same procedure for the validation sample, using a teacher's employment history between 2007 and 2010.

[27] 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 have a (weakly) positive contribution 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.

[28] Implementing the tests proposed by Chetty et al. (2014) and Rothstein (2010), we do not find evidence of non-random sorting of teachers across grade-schools (Online Appendix B3.3).

### 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.[29]

***Choice Set for Wage Schedules*** **($\Omega$):** We first construct a grid $\Omega^o$ 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 $\Omega = \{0.9, 0.95, 1, 1.05, 1.1, 1.15\} \times \{0, 10, 30, 50, 75, 100, 200, 225\}$ throughout our analysis.

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

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

where $w_d(x_i, c_i|\omega)$ is given by wage rule (6); $(\omega_{d1}^o, \omega_{d2}^o)$ is treated as district $d$'s wage schedule in the realized equilibrium. The implied $\{w_d(x_i, c_i|\omega_d^o)\}$ matches the data $\{w_i^o\}_i$ very well.[30]

***District Capacity and Budget Constraints:*** Assuming data are generated from an equilibrium, in which districts' constraints bind, $\kappa_d$ 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 ($x_1 \geq 10$).[31] 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, the overall correlation between experience ($x_1$) and either $c_1$ or $c_2$, not shown in the table, is

[29] Among the specifications we have tried, we found that 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.

[30] 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.

[31] Online Appendix Table B10 shows summary statistics for the validation sample (2010).

low at 0.04. This is consistent with previous work (e.g., Rockoff, 2004). 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 $\lambda_d$ (the fraction of low-achieving students). Districts with fewer low-achieving students are more likely to be located in suburban areas and 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** | **$\lambda_d$ 1st Quartile** | **$\lambda_d$ 4th Quartile** |
| Urban | 0.04 | 0.02 | 0.03 |
| Suburban | 0.15 | 0.34 | 0.09 |
| $\lambda_d$ | 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 $O_i^s$.[32] Column 2 shows

[32] Controlling for district-level shares of students who are Black, Hispanic, Asian, or female, and their interactions with the corresponding indicators for teachers' race and ethnicity barely improves the fit of Aux 1a

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 value higher 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. In particular, the wage coefficient in Column 2 is negative. This arises not because teachers dislike being paid more, but because many teachers did not receive offers from high-wage districts and Column 2 falsely assumes that they do. As a result, it appears that many teachers chose low-wage districts over high-wage districts. This example illustrates our identification argument. In general, districts' preference parameters need to rationalize, in addition to the observed offers, the lack of offers that reconcile the discrepancies between Columns 1 and 2.

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

| Teacher's Choice Set | Inferred Offer Set[a] | All Districts[b] |
|---|---|---|
| wage | 0.002 (0.0002) | $-5\times10^{-6}(2\times10^{-6})$ |
| $e^{\lambda_d}$ | -0.004 (0.009) | -0.0001 (0.0001) |
| $c_1\times\lambda_d$ | 0.52 (0.29) | -0.02 (0.006) |
| $I\left(d\neq d_0\right)$ | -0.72 (0.02) | -0.80 (0.01) |
| $I\left(d\neq d_0\right)\times$ experience | -0.008 (0.001) | -0.008 (0.0005) |
| $I\left(z_d\neq z_{d_0}\right)$ | -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 std errors are in parentheses.

Panel A of Table 3 summarizes districts' wage schedules. Districts' choices of $\omega_2$ (rewards for teacher contribution) are more dispersed than their choices of $\omega_1$. 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

(with an increase in $R^2$ from 0.680 to 0.681). We therefore choose a more parsimonious specification of teacher preferences, as in Equation (2).

areas and have higher fractions of low-achieving students, smaller per teacher budgets, and a slightly weaker composition of incumbent teachers. 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 $\omega_2$ (which implies a lower $\omega_1$ to balance the budget), these districts can attracts young but effective teachers.

Table 3: District Wage Schedules (2014)

| A. Summary stats of ($\omega_1$,$\omega_2$) | | B. $w_d\left(x, c \mid \omega_d^o\right)$ in Realized Matches (\$1,000) | | |
|---|---|---|---|---|
| $\omega_1$ mean (std) | 0.99 (0.04) | All Teachers: mean (std) | | 55.1 (11.6) |
| $\omega_2$ 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) |
| $\mathrm{Fr}((\omega_1, \omega_2) = (1, 0))$ | 0.24 | | $\in [5, 9]$ | 48.0 (6.4) |
| $\mathrm{Fr}(\omega_2 > 0)$ | 0.50 | | $\geq 10$ | 56.5 (7.2) |

| C. District Characteristics by $\omega_2$ | $\omega_2 = 0$ | $\omega_2 > 0$ |
|---|---|---|
| Rural | 0.80 | 0.83 |
| $\lambda_d$>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 |

Std deviations are in parentheses.

# 6 Estimation Results

## 6.1 Parameter Estimates

Table 4 shows estimated model parameters with their standard errors in parentheses, which are derived numerically via the Delta Method. Panel A shows estimated parameters governing teachers' preferences. For an average teacher, districts with higher fractions of low-achieving students ($\lambda_d$) are less desirable. However, teachers who are more effective in teaching low-achieving students are more willing to teach in these districts. For example, a teacher whose $c_1$ is at the 10th percentile 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 $90^{th}$ percentile, this premium is only \$1,817. We also find that rural districts are less attractive than their urban counterparts. The rest of Panel A shows 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 $\sigma_\epsilon$). High average moving costs help explain the lack of teacher mobility in general, while preference shocks absorb idiosyncratic reasons for mobility. Our findings of large average moving cost and dispersion of preference shocks are consistent with those in previous studies on worker mobility (e.g., Kennan and Walker, 2011).[33]

Table 4: Parameter Estimates

| A. Teacher Preference | | | B. District Preference | | |
|---|---|---|---|---|---|
| wage (\$1,000) | 1 | normalized | $c_1$ | 1 | normalized |
| $e^{\lambda_d}$ | -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) | $\geq 15$ | 0.058 | (0.07) |
| $I(d \neq d_0)\, x_1$ | -2.33 | (0.08) | MA or above | $1.3\times10^{-5}$ | $(5.8\times10^{-6})$ |
| $\sigma_\epsilon$ | 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) | $\sigma_\eta$ | 0.75 | (0.33) |

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

Panels B and C show district-side parameter estimates, which tend to have larger standard errors than those in Panel A because we have many more teachers (6,600) than districts (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. Our model is silent on what causes these costs, which may arise, for example, from the resistance of teachers or school boards. In Section 7.2.1, we explore the implications of both teachers' moving costs and districts' costs for deviating from rigid wage schedules.

[33] One possible explanation, which we abstract from, is the family joint location problem: The tied stayer (mover) would appear to have very high (low or negative) moving costs (e.g., Gemici, 2011).

## 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 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$^{-6}$ |
| $e^{\lambda_d}$ | -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 |
| $\lambda_d$ : | 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 $\omega$. Overall, the model fits the data well, although it underpredicts the dispersion of $\omega_2$ 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 in the estimation. Consistent with the data, the model predicts that districts with $\omega_2 > 0$ are slightly more disadvantaged. Appendix Table A1 shows that the model captures the correlation between $\omega$ and district pre-determined conditions as summarized by Aux 3.

Table 7: Model Fit: District Wage Schedules

| A. Summary Stats ($\omega_1$,$\omega_2$) | 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 | | | |
| $\mathrm{Fr}((\omega_1, \omega_2) = (1, 0))$ | 0.24 | 0.30 | $\mathrm{Fr}(\omega_2 = 0)$ | 0.50 | 0.41 |
| B. District Characteristics by $\omega_2$ | $\omega_2 = 0$ | | | $\omega_2 > 0$ | |
| | Data | Model | | Data | Model |
| Rural | 0.80 | 0.81 | | 0.83 | 0.81 |
| $\lambda_d$>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 $(\omega_1, \omega_2)$ and district characteristics by $\omega_2$: 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.

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$\times 10^{-6}$ | -3$\times 10^{-6}$ |
| $e^{\lambda_d}$ | -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)

| District Group | | Experience | | Share MA or above | | $10c_1$ | | $10c_2$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | Data | Model | Data | Model | Data | Model | Data | Model |
| $\lambda_d$ : | 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. To quantify the impacts of all our counterfactual policies, we pay special attention to the following metrics. Letting $\Pr(i \text{ in } d|\Upsilon)$ be the equilibrium probability that teacher $i$ works in district $d$ in a given policy environment $\Upsilon$, our metrics include:[34]

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 $D'$.[35] 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-$\lambda_d$ districts, i.e., districts with higher fractions of low-achieving students, than it does for low-$\lambda_d$ 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 contributions 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 into 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 improves 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.

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

[35] Throughout our simulations, in equilibrium, $\sum_i \Pr\left(i \text{ in } d|\Upsilon\right)$ equals $d$'s capacity $\kappa_d$.

Table 10: Flexible Pay vs Rigid Pay

| | Flexible* | $\frac{\text{Flexible-Rigid}}{\lvert\text{Rigid}\rvert}$(%) |
|---|---|---|
| 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 $\lambda_d$ districts | 0.108 | -1.04 |
| M1: TC in above median $\lambda_d$ districts | 0.107 | -0.57 |

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

Column 1 of Table 10 presents outcomes in the flexible-pay equilibrium. The first three 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 trade-off between efficiency and equity.[36] 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. As a result, achievement gaps are enlarged across districts and between low- and high-achieving students.

## 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 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

[36]The small magnitudes and the equity-efficiency trade-off 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 the distribution of teacher qualification.

*and* efficiency. In the following, we explore this possibility under flexible pay via a commonly used policy tool: state-funded teacher bonuses.

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

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

To avoid extreme values, we bound bonuses between 0 and $\overline{B}$ (set at twice the standard deviation of the overall wage distribution). Between 0 and $\overline{B}$, 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 ($\lambda_d$), B1 incentivizes comparative advantage-based sorting and therefore can improve efficiency. B1 also accounts for equity because bonus rates $\{r_d^1\}_d$ 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 $\omega_{d2}$. District $d$ would obtain more "free money" in the form of state-funded bonuses for its teachers if it chooses a higher $\omega_{d2}$. 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 (details are in Appendix A3). These amounts are comparable to relatively mild bonus programs implemented in other states, but with very different formulae than ours.[37]

We start with two flat-rate programs: **B1(flat)** under Formula B1, with $r_d^1 = \$87,550$ for all $d$; and **B2(flat)** under Formula B2, with $r_d^2 = 3.4$ for all $d$. The bonus rates are calibrated to

[37] 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 between \$5,500 and \$16,500 in the 2008 Tennessee POINT program (Neal, 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 (the effect was positive for one year).

exhaust the same pre-specified total cost. The 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 $\lambda_d$ (the fraction of low-achieving students) and experiment with group-specific bonus rates, such that $r_d^2$ (weakly) increases as we move from the lowest-$\lambda_d$ group to the highest-$\lambda_d$ group.[38] 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 $\lambda_d$ distribution, respectively.

Table 11: State-Funded Teacher Bonuses

| (%) | $\frac{\text{B1(flat)-Base}}{\lvert\text{Base}\rvert}$ | $\frac{\text{B2(flat)-Base}}{\lvert\text{Base}\rvert}$ | $\frac{\text{B2(pro)-Base}}{\lvert\text{Base}\rvert}$ |
|---|---|---|---|
| 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 $\lambda_d$ districts | 0.02 | -0.89 | 0.24 |
| M1: TC in above median $\lambda_d$ districts | 0.10 | -0.13 | 0.42 |

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

The effect of B2(pro) is shown in the last column of Table 11. B2(pro) leads to a 0.13% gain in the overall efficiency (the same as B2(flat)). Low-achieving students gain by 0.16% and high-achieving students gain by 0.10%, 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. Albeit small, results under B2(pro) demonstrate that carefully designed bonus programs can improve both efficiency and equity.

[38] An optimal search for district-specific bonus rates would be computationally too burdensome to perform, 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.

#### 7.2.1 Discussion: Magnitudes of Policy Impacts

Via counterfactual simulations, we have shown that there is a meaningful equity-efficiency trade-off associated with flexible pay, and that carefully designed state bonus programs can improve both equity and efficiency. However, the magnitudes of these policy impacts are small; this is in line with findings from previous studies that monetary incentives have very limited effects on attracting and retaining teachers (e.g., Clotfelter et al., 2011; Russell, 2020).

The equilibrium impacts of our counterfactual policies are shaped by teachers' and districts' preferences and constraints. We examine the importance of two of these factors: moving costs $\Gamma(\cdot)$ faced by teachers and resistance costs $R(\cdot)$ faced by districts upon deviating from their rigid wage schedules. To do so, we consider the following two counterfactual cases.
Case 1: Teachers have zero moving costs, i.e., $\Gamma(\cdot) = 0$.
Case 2: In addition to $\Gamma(\cdot) = 0$, districts face zero resistance costs, i.e., $R(\cdot) = 0$.
For each of these cases, we simulate the new flexible-pay equilibrium and compare it with the baseline flexible-pay equilibrium.

The results are shown in the first two columns of Table 12. Relative to the baseline, teacher total contribution is 2.7% higher in the equilibrium without moving costs (Column 1). This higher efficiency comes at the cost of equity: Average teacher contribution declines by 3.9% for low-achieving students and increases by 9.1% for high-achieving students. These changes are larger than the impact of any of our bonus programs shown in Table 11. This is unsurprising since teachers' moving costs (Table 4) are much higher than our bonuses. When we additionally remove districts' resistance costs (Column 2), teachers' total contribution increases further and equity worsens, but only mildly so.

Table 12: Moving Costs, Resistance Costs, and Policy Impacts

| | Zero Costs vs Base | | B1(flat) Impacts | | B2(flat) Impacts | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| % | $\frac{\text{Case1-Base}}{\lvert\text{Base}\rvert}$ | $\frac{\text{Case2-Base}}{\lvert\text{Base}\rvert}$ | $\frac{\text{Case1 B1(flat)-Case1}}{\lvert\text{Case1}\rvert}$ | $\frac{\text{B1(flat)-Base}}{\lvert\text{Base}\rvert}$ | $\frac{\text{Case2 B2(flat)-Case2}}{\lvert\text{Case2}\rvert}$ | $\frac{\text{B2(flat)-Base}}{\lvert\text{Base}\rvert}$ |
| TC | 2.67 | 3.08 | 0.92 | 0.08 | 2.14 | 0.13 |
| $c_1$ | -3.94 | -3.97 | -0.91 | 0.07 | 0.79 | 0.02 |
| $c_2$ | 9.08 | 9.93 | 2.48 | 0.08 | 3.27 | 0.25 |

Base: baseline flexible-pay equilibrium; B1(flat)/B2(flat): flat bonus B1/B2 + base;

Case 1: flexible-pay equilibrium with zero moving cost; Case 1 B1(flat): flat bonus B1 + Case 1;

Case 2: Case 1 + zero resistance costs; Case 2 B2(flat): flat bonus B2 + Case 2.

The results above suggest that teachers' moving costs are a major obstacle for efficiency, but in their absence the achievement gap between low- and high-achieving students would be much larger. Districts' resistance costs for deviating from rigid wage schedules mildly enhance these effects. These two costs could also play a major role in mediating the impact of our counterfactual bonus programs. To quantify this role, we run two additional simulations. In the first simulation, we introduce bonus program B1(flat) to the market under Case 1 (zero moving costs). In the second simulation, we introduce bonus program B2(flat) to the market under counterfactual Case 2 (zero moving costs and zero resistance costs). We re-calibrate bonus rates such that the total cost of each program is the same as before ($1,560 per teacher).

Column 3 of Table 12 shows how B1(flat) impacts equity and efficiency in a market without moving costs (Case 1). For comparison, Column 4 presents again the impact of B1(flat) on the baseline economy. Without moving costs, B1(flat) leads to larger efficiency gains (with a 0.9% increase in total contribution compared with 0.08% under the baseline). This is because B1(flat) incentivizes teachers to move based on their comparative advantages, which should be easier to achieve when teachers face zero moving costs. The absence of moving costs also exacerbates B1(flat)'s negative effect on educational equity. Without moving costs, B1(flat) decreases teachers' average contribution by 0.9% for low-achieving students and increases it by 2.5% for high-achieving students. With moving costs, B1(flat) increases teachers' average contribution for both types of students, although more for high-achieving students.

Columns 5 and 6 show that the impact of B2(flat) is much larger on an economy without moving and resistance costs (Column 5) than it is on the baseline economy (Column 6). This is because B2(flat) provides incentives not only for teachers to sort efficiently, but also for districts to reward teacher contribution. Without resistance costs, districts are more willing to reward teacher contribution; without moving costs, teachers are more willing to move. As a result, B2(flat) increases total teacher contribution by 2.14% without moving and resistance costs, but only by 0.13% when these costs are present. Taken together, our results imply that the effectiveness of teacher bonus schemes hinges on teachers' willingness to move, districts' willingness to change their wage schedules, and the interaction of these two factors.

# 8 Conclusion

Proper allocation of public servants across local employers can have important implications for both efficiency and equity, but it is difficult to achieve due to various institutional frictions such as wage rigidity. We study the equity-efficiency implication of wage rigidity through the lens of the labor market for public school teachers. To that end, 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 following a reform that gave districts control over teacher pay. We have validated the model using the pre-reform data under rigid pay.

Our estimated model implies that, ceteris paribus, giving districts control over teacher pay would lead to more efficient but also more unequal sorting of teachers across districts. Efficiency improves 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. However, the effectiveness of these policy interventions hinges on teachers' willingness to move and districts' willingness to change their wage schedules.

Our analysis abstracts from several important aspects of the teachers' market; extending our framework along these lines is worth pursuing. The first extension, which requires additional data, 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 Sieg, 1999; Epple and Romano, 2003; Ferreyra, 2007; Epple and Ferreyra, 2008). A third extension is to add teachers' effort choices into our framework. Since our model takes teacher effectiveness as pre-determined, the efficiency gains we have found are likely to understate the total effect of our counterfactual policy intervention. For example, Barlevy and Neal (2012) show that "pay for percentile" can induce teachers to allocate socially optimal levels of effort. Finally, while our static equilibrium model is better suited to study short-run policy effects, an important but rather difficult extension is to consider the market in a dynamic equilibrium setting, which would allow for the 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.

Ba, B., P. Bayer, N. Rim, R. Rivera, and M. Sidibé (2021). Police officer assignment and neighborhood crime. *NBER working paper n. 29243*.

Barlevy, G. and D. Neal (2012). Pay for percentile. *American Economic Review 102*(5), 1805–31.

Barseghyan, L., M. Coughlin, F. Molinari, and J. C. Teitelbaum (2021). Heterogeneous choice sets and preferences. *Econometrica 89*(5), 2015–2048.

Bates, M. D., M. Dinerstein, A. C. Johnson, and I. Sorkin (2022). Teacher labor market equilibrium and student achievement. *NBER working paper n. 29728*.

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. (2021). The labor market for teachers under different pay schemes. *American Economic Journal: Economic Policy 13*(3), 63–102.

Bobba, M., T. Ederer, G. Leon-Ciliotta, C. Neilson, and M. G. Nieddu (2021). Teacher compensation and structural inequality: Evidence from centralized teacher school choice in perú. *NBER working paper n. 29068*.

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.

Einav, L., A. Finkelstein, and N. Mahoney (2018). Provider incentives and healthcare costs: Evidence from long-term care hospitals. *Econometrica 86*(6), 2161–2219.

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.

Fack, G., J. Grenet, and Y. He (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review 109*(4), 1486–1529.

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., 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.

Leaver, C., O. Ozier, P. Serneels, and A. Zeitlin (2021). Recruitment, effort, and retention effects of performance contracts for civil servants: Experimental evidence from rwandan primary schools. *American Economic Review*.

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.

McFadden, D. (1978). Modeling the choice of residential location. In A. e. a. Karlqvist (Ed.), *Spatial Interaction Theory and Residential Location*. Amsterdam: North Holland.

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. (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. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics 125*(1), 175–214.

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

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers 3*(2), 135–146.

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. Districts' Optimal Decisions

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

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

$$\begin{aligned} s.t. &\int o_d(x, c, d_0|\omega)\, h_d(x, c, d_0, \omega)\, dF(x, c, d_0) \leq \kappa_d, \\ &\int o_d(x, c, d_0|\omega)\, h_d(x, c, d_0, \omega)\, w_d(x, c|\omega)\, dF(x, c, d_0) \leq M_d \\ &o_d(x, c, d_0|\omega) = 1 \text{ if } x_1 \geq 3 \text{ and } d_0 = d. \end{aligned}$$

Letting $\varphi(x, c, \lambda_d) \equiv [xb_0 + b_1\lambda_d c_1 + b_2(1-\lambda_d)c_2]$, the first-order condition is

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

where $\nu_\kappa$ and $\nu_M$ are the non-negative multipliers associated with the adjusted capacity and budget constraints. The capacity (budget) is adjusted by netting out the expected slots (wages) 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 the offer is accepted, the district must surrender a slot from its limited capacity and pay the wage $w_d(x, c|\omega)$, 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:

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

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

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

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 from $d_0$, as seen in (16).

For a given $\omega$, a district's job offer decision can be derived by the following procedure.
1) Set $o_d(x, c, d_0|\omega) = 1$ for teachers with $x_1 \geq 3$ and $d_0 = d$.
2) Guess $\nu_M$, rank other teachers by $\varphi(x, c, \lambda_d) - w_d(x, c|\omega)\nu_M$.
3) Give offers to teachers from the top-ranked downwards, until the expected capacity or budget is filled, i.e., (17) or (18) is binding.
4) Calculate the district's value associated with this $\nu_M$, and optimize over $\nu_M$ to find the maximum; $o_d(\cdot|\omega)$ associated with the optimal $\nu_M$ are the optimal job offers under $\omega$.
In the outer loop, the district searches over $\omega$ to optimize its objective (9). Both (17) and (18) bind in the equilibrium throughout our simulations.

### 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 $\omega_{d1}$ ($\omega_{d2}$) 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.

### A3. State Bonuses: Reward for Teacher Contribution and Program Costs

Table A2 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 $\omega_{2d}$, 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 $\omega_{2d}$ arises mainly from two costs faced by districts, which may outweigh the small state bonuses we introduce. First, although a district can obtain more state bonuses for its effective teachers by increasing its own $\omega_{2d}$, 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 a resistance cost, which increases with its deviation from $\omega_{2d} = 0$. 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.

Table A1: Model Fit: OLS of District Wage Schedule

| Auxiliary Model 3 | $\omega_{d1}$ | | $\omega_{d2}$ | |
|---|---|---|---|---|
| | Data | Model | Data | Model |
| *Composition of incumbent teachers* ($d_0 = d$) | | | | |
| Fr(experience 3-4) | 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* | | | | |
| $\lambda_d$ | 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. * Outside of the 95% CI of the estimates from the data.

Table A2: State-Funded Teacher Bonuses: 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 ($\omega_{2d}TC > 0$)* | 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 | |

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