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

Barbara Biasi, Chao Fu and John Stromme†

March 12, 2025

Abstract

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

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

Keywords: Wage Rigidity, Equilibrium Sorting, Education Efficiency and Equity, Teachers' Comparative Advantages, Structural Estimation

*We thank Manuel Arellano, Adam Kapor, 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, Cowles Summer Conference, 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, barbara.biasi@yale.edu; Fu: University of Wisconsin and NBER, cfu@ssc.wisc.edu; Stromme: Vanderbilt University, john.stromme@vanderbilt.edu.

1 Introduction

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

Unfortunately, the typical salary structure for U.S. public school teachers fails to incentivize comparative-advantage based sorting. Teacher salaries follow rigid experience-education schedules, often set via collective bargaining. We label this regime the "rigid-pay" regime. Under rigid pay, districts cannot use salary schemes to attract teachers better suited for their students, despite the substantial cross-district variation in students' preparedness.[1] Associated with rigid pay, sorting is largely vertical, with better teachers working in districts with more advantaged students (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 inequalities across children from different backgrounds.

An alternative regime—"flexible pay"— is one where districts have the flexibility to design their own teacher pay schedules.[2] How flexible pay may affect the equity and efficiency of teacher-district allocations is not obvious ex ante: It depends on how districts would choose teacher pay if allowed to do so and how both sides of the teachers' labor market interact in the equilibrium under such a regime. These choices and interactions, in turn, depend on teachers' and districts' preferences. Understanding these essential factors is challenging: Very similar and rigid pay schedules are often imposed on all districts within a state, making inference difficult.

We utilize a real-life exception to gain more insights: 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 study the implications of flexible pay in a market equilibrium setting. We start by extending the traditional value-added (VA) model to allow for two-dimensional teacher VA for low- and high-achieving students. The correlation between the two VA measures is 0.67, which indicates the presence of significant comparative advantages. We then embed this empirical fact into our model, where 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. A district cares about a teacher's contribution

[1] For example, in 2021, the share of a district's 5th graders who meet the state's math proficiency requirement ranged between 10-80% across districts in Wisconsin, 10-85% in California, and 10-90% in Texas (Express, 2021).

[2] 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. We also use the words pay, wage, and salary interchangeably.

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. 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 the offers received, a teacher chooses their most preferred district, net of moving costs. An equilibrium requires districts' beliefs to 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. 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.

To quantify the trade-off mentioned above, we first need to tackle a major identification challenge: The researcher observes only the accepted offers, making it hard to separate teacher preferences (which govern how effectively pay schemes can incentivize teachers to move across jobs) from district preferences (which govern districts' hiring decisions and, if given the flexibility, their choices of teacher pay schedules). Our identification argument is as follows: First, under Act 10, districts have control over teacher pay; the extent to which a district's wage schedule favors different teachers contains information about its preferences. Second, 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. Given teachers' preferences, districts' preference parameters, which govern districts' offer decisions, must rationalize the realized teacher-district matches.

We apply our model to administrative data from the Wisconsin Department of Public Instruction with three linked panel data sets at the student, teacher, and district level. The data allow us to track a teacher's employment history within the state's public school system, including their salaries and job characteristics. We use post-Act 10 data to estimate the 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.

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 0.04% improvement in overall student achievement. However, it enlarges the achievement gap between low- and high-achieving students and reduces student achievement in districts with high fractions of low-achieving students.

These findings indicate that, while flexible pay may improve efficiency, this may come at the cost of increased inequality. To explore the possibility of improving educational equity and efficiency, we use a widely used tool—state-funded teacher bonuses—and design a novel, yet simple formula with three components. The first component rewards a teacher for their total contribution to overall student achievement (TC). This component incentivizes more efficient sorting because a teacher's TC (and thus their bonus) is higher when their comparative advantage better matches a district's student composition. The second component rewards a teacher additionally for their contribution to low-achieving students. This incentivizes teachers who are good at teaching low-achieving students to teach in districts with more students of this type, incentivizing a more equitable allocation. The third component ties the bonus to a district's wage schedule such that districts are incentivized to increase their own rewards for teachers' contribution to student achievement.

At a relatively mild cost compared to programs implemented in other states, we find that it is possible, with carefully-chosen bonus rates, to improve both efficiency and equity. Comparing bonus programs that are equally costly in the equilibrium, we find that programs that favor districts with more low-achieving students can outperform purely TC-based bonus programs in improving overall efficiency. This occurs because most teachers, including some who have a comparative advantage in teaching low-achieving students, prefer to teach high-achieving students. However, a program that rewards teachers only for their contribution to low-achieving students would hurt high-achieving students. Most importantly, we identify a possible and practical route to achieve the difficult goal of increasing both equity and efficiency: With a proper combination of the efficiency incentive and the equity incentive, our bonus program can improve achievement for both low- and high-achieving students, although the impacts are quantitatively small.

Related Literature A proper allocation of public servants across local employers can have important implications for both efficiency and equity. In reality, 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 allocation of workers to employers and its equity-efficiency implications. This is a general problem that exists 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. Indeed, efficiency is part of the reason why many OECD countries have been increasingly introducing performance-related pay for government employees (Cardona, 2006).

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, as we do in our model. Using data from Peru, Bobba et al. (2021) and Ederer (2023) study teacher-school sorting in a centralized application-assignment environment. Assuming that the observed teacher-school matches are stable, Boyd et al. (2013) estimate a two-sided matching model to disentangle teacher and school preferences. 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. Our paper and Tincani (2021) well complement each other. Tincani (2021) focuses on how a given wage function for public school teachers would induce reactions from the private school and affect teachers' and households' choices between public and private sectors. We are interested in efficiency and equity within the public sector, and we study how public school districts use wage and hiring strategies to compete with one another for better teachers.

Our paper also contributes to the literature on the effect of teacher pay on teachers' behavior and student outcomes (see Neal et al., 2011; Jackson et al., 2014, for reviews), and more specifically on teachers' mobility and educational inequality. This literature has produced mixed findings. 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 other studies find little or no effect (e.g., Clotfelter et al., 2011; Russell, 2020). Using a discrete-choice experiment, Johnston (2021) finds that high-performing teachers have stronger preferences for performance pay compared with other teachers. At the same time, Hanushek et al. (2004) find that teacher mobility is more related to student composition than salary, but salary has a modest impact. Evidence on the effect of pay rigidity on teachers' labor market is also mixed. Burgess et al. (2022) find that schools became better at retaining their teachers after a reform that compelled schools in England to replace centralized pay scales with a school-level performance pay regime. A similar reform in Sweden, which replaced centralized pay with individually bargained wages, had no effects on teacher retention, recruitment, or composition (Willén, 2021). Biasi (2021) shows that, after Wisconsin implemented Act 10, higher-quality teachers moved to districts that adopted flexible pay.[3] 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.

An innovation of our study relative to the works mentioned above is that we allow for multi-dimensional teacher effectiveness in teaching different types of students, which leaves open the possibility that changing teacher-district sorting can improve both equity and efficiency. This consideration is supported by previous findings that teacher effectiveness might be specific to student composition. For example, Jackson (2013) demonstrates the importance of match quality between teachers and schools. Aucejo et al. (2019) and Graham et al. (2023) find significant complementarities between teachers and classroom composition and show that reassigning teachers across classrooms could have sizable effects on teachers' contribution to learning. Other studies have considered heterogeneity in teacher effectiveness by race (Dee, 2004; Gershenson et al., 2022) and other student demographics (Lavy, 2016; Bates et al., 2022), and by subjects (Fox, 2016). In a 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

[3] 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.

by directly affecting teachers' preferences over schools. In contrast, 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, in February of 2011 Republican Governor Scott Walker proposed a Budget Repair Bill to the state legislature, aimed at reducing benefits for public-sector employees and reducing their bargaining rights. The bill, which came to be known as Act 10, encountered major opposition both from within the State Senate, where all 14 Democratic senators fled to Illinois to delay the vote, and from the general public, where multiple protests occurred under the lead of the unions and other groups. In spite of the controversy, the bill was approved in the senate on June 29, 2011. However, opposition to Act 10 persisted and led to an (unsuccessful) attempt to recall Walker from office in 2012.

Act 10 led to major reforms to public sector employment and collective bargaining. For public school teachers, who account for approximately 45% of all public employees in Wisconsin, the most dramatic change was to limit salary negotiations between districts and unions up to the base pay (i.e., the starting pay for new employees) while excluding salary schedules from collective bargaining agreements negotiations. The growth rate of base pay was also capped to the rate of inflation. Above and beyond base salaries, Act 10 gave school districts 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." In addition, Act 10 prohibited unions from automatically collecting dues from employees' paychecks and required unions 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. Lastly, Act 10 reduced benefits for all public sector employees, except for law enforcement and fire department personnel, through an increase in employee contributions to pensions and health care premia.[4]

In sum, Act 10 was a highly controversial comprehensive reform that affected public school teachers through multiple provisions. Most of these provisions, such as the reduction in employee benefits, led to changes that were common among school districts. Changes in pay schedules, instead, occurred at the discretion of each district. The model we develop in Section 3 will allow us to isolate the equilibrium effect of changes in pay regimes on the teachers' labor market.

2.1 Motivating Data Facts

Our data, which we describe in detail in Section 5, consist of three linked data sets at the teacher, student, and district level, respectively. We present a set of basic data facts to motivate our model.

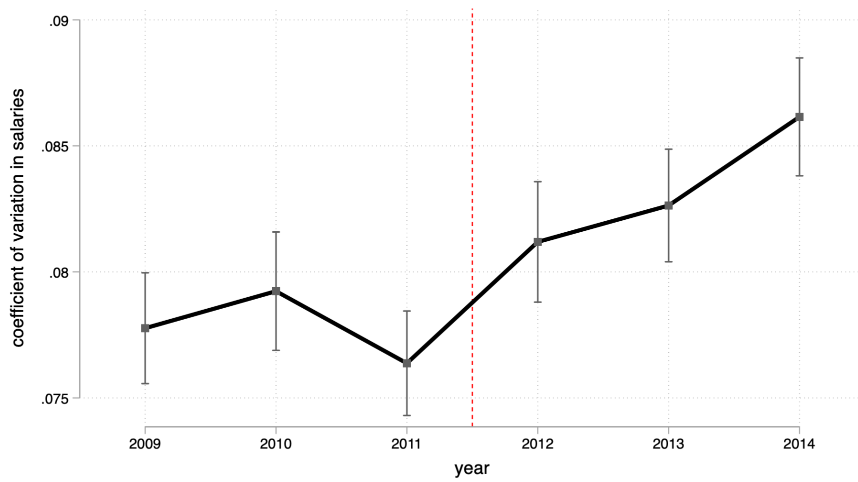
2.1.1 A Glance at the Market Before and After Act 10

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: The left panel of Figure 2 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. **The right panel** shows within-teacher changes in real wages over time, for movers and non-movers. Before Act 10, real wage growth was small for both movers and stayers. After Act 10, wage growth

[4]On July 1, 2011 the state 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).

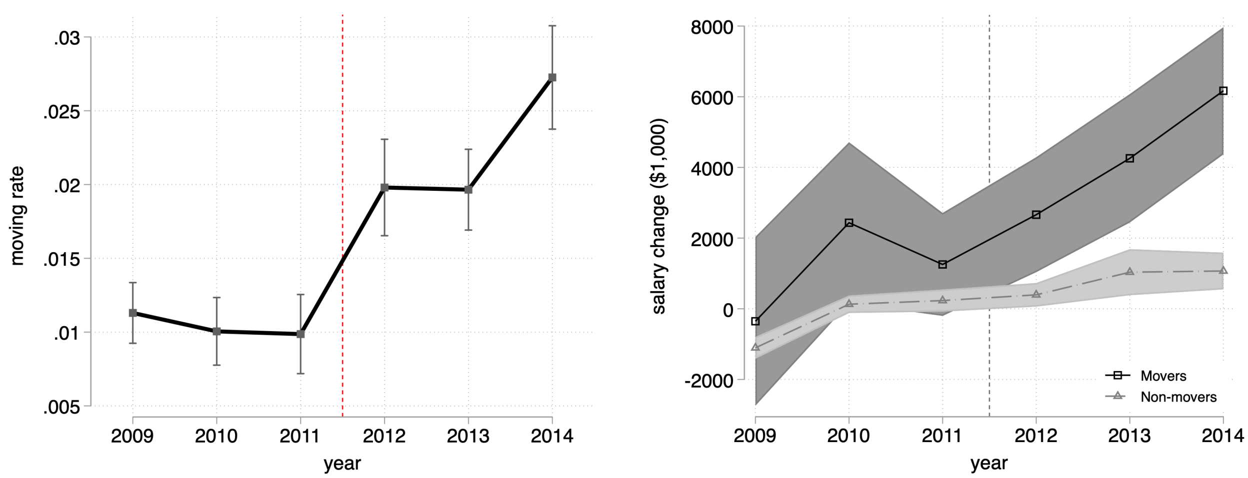
Figure 1: Variation in Teacher Salaries



Notes: Point estimates and robust confidence intervals of the coefficients of a regression of salary residuals on year fixed effects. Salary residuals are obtained from a regression of salaries on district-by-year and experience-by-years of education fixed effects.

remained small for stayers but significantly increased for movers. This pattern is consistent with districts using wage strategies to compete for teachers after Act 10.

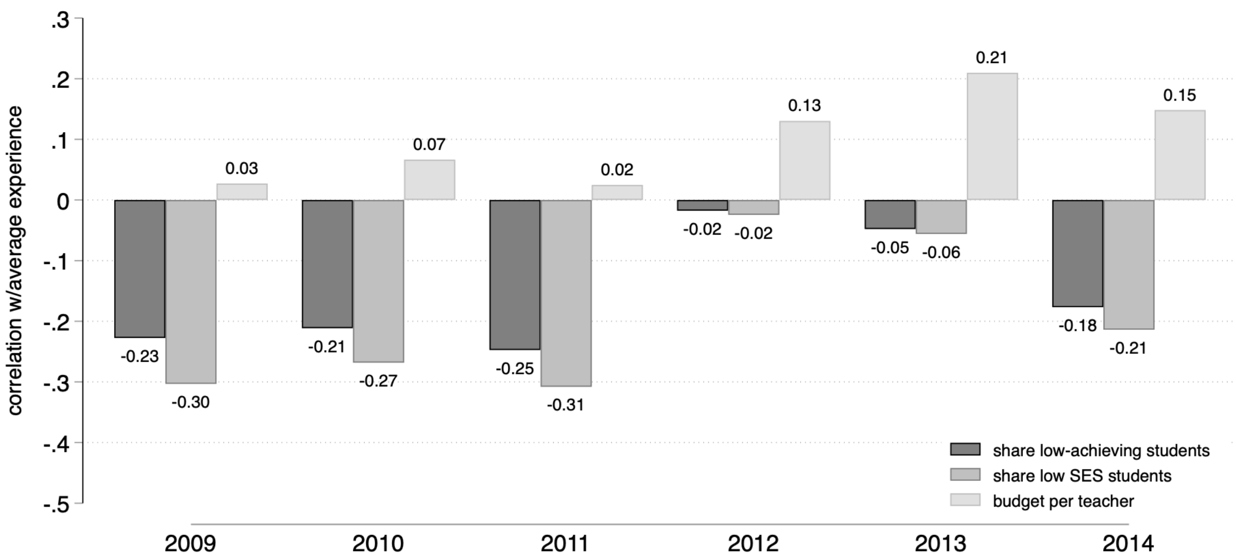
Figure 2: Rates of Teacher Movements Across Districts



Notes: The left panel shows estimates and and confidence intervals of the coefficients of a regression of an indicator for a teacher changing district on year fixed effects, clustering standard errors at the district level. The right panel shows estimates and confidence intervals of year fixed effects in a regression of real salaries, where we control for teacher fixed effects and we cluster standard errors at the district level; we show estimates separately for movers and non-movers in each year.

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 with a larger share of low-achieving or low-SES students. Figure 3 shows that the average teacher experience of each district was negatively correlated with the fraction of low-achieving students (those with math scores below the state median) and the fraction of economically-disadvantaged students in the district. These relationships became much weaker after Act 10. Before Act 10, there was barely any (positive) correlation between average teacher experience of each district and the district's budget; after Act 10, this correlation became stronger. This suggests that, with the flexibility of teacher pay at hand, districts with larger budgets found it even easier to attract teachers after Act 10.

Figure 3: Variation in Teacher Salaries



Notes: Year-specific correlations between district characteristics and the average experience of teachers in the district. Correlations are obtained weighing districts by student enrollment.

Figures 1 to 3 provide some suggestive evidence that, under flexible pay, districts used wage strategies to compete for teachers and teacher-district sorting became less vertical. However, we cannot directly interpret these pre- versus post-Act 10 differences 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, λ_d is the fraction of students in d who are low-achieving (with prior test scores below the state median), κ_d is district d's capacity (number of teaching slots), and M_d is its budget. The sum of 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 λ_d.

Timing: The timing of the model is as follows:

1. Districts simultaneously choose their wage schedules $\{w_d(x, c)\}$ and job offers $\{o_d(x, c, d_0)\}$, where $o_d(x, c, d_0) = 1$ if d makes an offer to teacher (x, c, d_0) and 0 otherwise.
2. Each teacher observes their taste shocks and chooses their most preferred offer.

Notice that wages are assumed be blind to a teacher's origin d_0, which is consistent with real-life practice.[5] 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
$V_d(x, c, d_0) + \epsilon_d \equiv$

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

[5] 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.

where ϵ_d is an i.i.d. Type 1 extreme-value distributed taste shock with a scale parameter σ_ϵ. Wage enters with a normalized coefficient of 1, so that teachers' preferences are measured in \$1,000. Teachers' preferences for district characteristics q_d are governed by the vector θ_0. The next two terms capture teachers' preferences for student composition (λ_d); these preferences may vary across teachers with different effectiveness in teaching low-achieving students (c_1).[6] $\Gamma(\cdot)$ is the cost of moving from d_0 to d, given by

$$\Gamma(d, d_0, x) = \begin{cases} \delta_0(x_1) + \ln(dist_{d,d_0})\,\delta_1 + I(z_d \neq z_{d_0})\,\delta_2 \text{ if } d_0 > 0, \\ \Gamma_0(x) \text{ if } d_0 = 0. \end{cases} \tag{3}$$

For teachers with $d_0 > 0$, $\delta_0(x_1)$ is an experience-group specific moving cost parameter, δ_1 measures how moving costs varies with (log) physical distance between two districts, and δ_2 is the additional cost if the new district is in a different commuting zone than one's current district. For teachers new to the market $(d_0 = 0)$, who are not attached to any district at the beginning of the period, we assume that the moving cost from $d_0 = 0$ to any district d is the same, given by $\Gamma_0(x)$.[7]

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

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

[6] We use e^{λ_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 λ_d might be convex. We have estimated a model with λ_d instead of e^{λ_d}; it does not fit the data well. We include only the interaction $\lambda_d c_1$ in (2) because adding the interaction $c_2\lambda_d$ does not improve the fit.

[7] For teachers with $d_0 = 0$, we do not observe their initial locations. This prevents us from allowing their moving costs to differ by d. Given that $\Gamma_0(x)$ is constant across districts, we set it to zero without loss of generality because it is irrelevant for a teacher's choice over different districts.

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.[8] We assume that $b \geq 0$, i.e., district preferences are weakly increasing in all teacher attributes, and we normalize b_1 to 1. A special case is $b_0 = 0$ and $b_1 = b_2$, in which Equation (5) is equivalent to $TC(c, \lambda_d)$, i.e., a district values a teacher only for their total contribution to its students. More generally, if b_1 and b_2 are large relative to b_0, districts would rank teachers differently depending on their student compositions λ_d; if b_0 is dominant, districts would largely agree on their rankings of teachers.

3.2.2 Choice Space for Wage Schedules

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

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

To avoid unrealistically high or low wages, we bound wages by $[\underline{w}, \overline{w}]$, such that[9]

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

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. Online Appendix B1.4.3 shows that wages predicted by three alternative and more flexible wage rules (e.g., rewarding c_1 and c_2 differently) are very similar to those predicted by (6). Therefore, we choose the more parsimonious specification (6).

[8]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.

[9]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.

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: First, a district chooses a wage schedule $\omega = (\omega_1, \omega_2)$; and second, it makes job offers conditional on ω. We solve a district's problem via backward induction.

Job Offers For a given wage schedule ω, 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 other districts that offer them a job. Teachers' decision rule in Equation (4) implies

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

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

Wage Schedule District d chooses ω to solve the following problem

$$\max_{\omega \in \Omega} \left\{ \frac{\pi_d(\omega)}{\kappa_d} - I(\omega \neq [1,0]) R_d(\omega) + \eta_\omega \right\}, \tag{9}$$

where $\pi_d(\omega)$ (given by (7)) is normalized by district capacity to make the scale comparable across districts with different capacities. Given the highly controversial nature of Act 10, we incorporate a resistance cost $R_d(\cdot)$ that district d faces for deviating from its pre-reform wage schedule $\omega_d = (1, 0)$. Finally, η_ω is an i.i.d. extreme-value distributed shock associated with choosing ω, with a scale parameter σ_η.

To specify $R_d(\cdot)$ empirically, we let the institutional background and our data guide us. As described in Section 2, Act 10 was a reform with political connotations: It was proposed and passed by a largely Republican state legislature with strong opposition from Democrats, and it was perceived by many as an attack to public-sector unions. It is therefore plausible that districts with a larger share of Democratic voters may be more opposed to deviating from the pre-reform rigid pay regime that is preferred by unions. Our data support this hypothesis. As we show in Section 5.2, controlling for factors that enter a district's payoff $\pi_d(\omega)$, the share of Democratic votes among district d's residents (dem_d) is significantly negatively correlated with the deviation from the rigid wage schedule; this correlation remains when we additionally control for the characteristics of a district's residents.[10] We therefore allow the resistance cost to vary with residents' political views as proxied by dem_d and specify $R_d(\cdot)$ as $R(\omega, dem_d)$, detailed in Appendix A2.

3.3 Equilibrium

Definition 1 *An equilibrium is a tuple of decisions* $\{\{d^*(x, c, d_0, \epsilon)\}, \{\omega_d^*, \{o_d^*(x, c, d_0|\omega)\}\}_d\}$ *and belief* $\{\{h_d^*(x, c, d_0, \omega)\}_d\}$ *such that*

1) Given $\{\omega_d^*, \{o_d^*(\cdot|\omega_d^*)\}\}_d$, $d^*(x, c, d_0, \epsilon)$ *solves the teacher's problem, for all* (x, c, d_0, ϵ).

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

3) $\{h_d^*(\cdot)\}_d$ *is consistent with* $\{\{d^*(\cdot)\}, \{\omega_d^*, \{o_d^*(\cdot|\omega_d^*)\}\}_d\}$.

To solve its problem, it is sufficient for a district to know teachers' acceptance probabilities $\{h_d(x, c, d_0, \omega)\}$: 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]

[10] We exclude dem_d from teachers' preferences for parsimony: Including dem_d in our auxiliary regression models does not change the estimates of other coefficients nor does it improve the R^2 of the auxiliary models, which suggests that this variable does not explain teacher-district sorting.

[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

As a feasible alternative, we assume that districts make their decisions based on a simplified parametric belief about teachers' acceptance probabilities,[12] given by

$$\widetilde{h}_d\left(x, c, d_0, \omega\right)=\frac{1}{1+\exp\left(f\left(x, c, d_0, w_d, q_d, \lambda_d\right)\right)}, \tag{10}$$

$$\begin{aligned} f(\cdot)= & x\zeta_1+\zeta_2\frac{c_1+c_2}{2}+\zeta_3\left(\frac{w_d\left(x, c|\omega\right)-\overline{w}\left(x, c\right)}{\sigma_{w(x,c)}}\right)+\zeta_4 q_d+\zeta_5 e^{\lambda_d}+\zeta_6\lambda_d c_1 \\ & +\left(1-I\left(d_0=0\right)\right)\left[I\left(d\neq d_0\right)\left(\zeta_{7x_1}+\zeta_8\ln\left(dist_{dd_0}\right)+\zeta_9 I\left(z_d\neq z_{d_0}\right)\right)\right]. \end{aligned} \tag{11}$$

This simplified belief function captures all the factors governing its counterpart $\{h_d(\cdot)\}$ defined in (8). The first two terms in (11) 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 (11) 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 B2.

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

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).

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 suggest that the effect of performance pay on the selection into and out of the teaching profession is very small (e.g., Rothstein, 2015) and that the decision to leave teaching is not driven by pay (Scafidi et al., 2006). Other studies instead 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). If the latter is true, 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 (Online Appendix B.3.3). That being said, focusing on cross-district movement and competition limits our ability to conduct counterfactual policies that introduce within-district pay variation to induce within-district teacher reallocation. On the one hand, such policies can be fruitful, because teachers' within-district moving costs are presumbaly lower than their cross-district moving costs. On the other hand, districts appear to be averse to within-district pay variation (Online Appendix B3); such policies may face resistance from districts. Introducing within-district competition and teacher reallocation into our framework would allow for a more complete view but would involve substantial complications. We therefore leave it for future research.

Third, we take a district's student composition λ_d as given. In particular, we assume away potential households re-sorting across districts in response to our policy interventions. 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.[14] 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 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

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

[14] 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.

when interpreting our results.

Finally, although we explicitly model how teachers' preferences, effectiveness, opportunity sets, and hence decisions vary with their experience, we do so in a cross-sectional reduced-form way rather than in a life-cycle framework, with teachers internalizing their future moving costs and job prospects. Incorporating life-cycle concerns is important but challenging; we leave this for future work. Relatedly, 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.[15] We complement this literature by focusing on a different channel through which financial incentives may improve education, i.e., by incentivizing 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 Θ) 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 Θ, which consists of teachers' and districts' preference parameters, and an inner loop solving the model for each given Θ (detailed in Online Appendix B2). In our counterfactual policy simulations, we need to find the fixed point for the belief parameter vector ζ *and* wage statistics $\{\overline{w}(\cdot), \sigma_w(\cdot)\}$ that enter the belief function, but we only need to find the fixed point for ζ during the estimation. Assuming that the data were generated from an equilibrium, the realized equilibrium $\{\overline{w}^o(\cdot)\}$

[15] 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).

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 (11) and search only for the fixed point for ζ 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 (see, for example, Menzel, 2015).[16] The observed wage schedules and matches, however, contain rich information that allow us to overcome this obstacle under certain assumptions, 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, it is useful to 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 $R(\cdot)$.

[16] In the setting of two-sided matching markets with non-transferable utility, Menzel (2015) shows that, under pairwise stable matching, the joint surplus is non-parametrically identified and that, without further assumptions, it is impossible to identify preferences on each side of the market separately. Ederer (2023) extends Menzel (2015) to a dynamic setting.

4.1.2 Optimal Job Offers and Observed Matches

The observed teacher-district matches provide the second major source of information for identification. If we observed what options were available for each teacher, we could identify teachers' preferences in a straightforward way: Choices out of (multiple) feasible options reveal preferences. Observing only accepted offers complicates the inference.

However, combining the observed matches with districts' optimal offer decisions allows us to infer a subset of all offers received by each teacher and thus to identify teachers' preferences. 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 ω_d^o) plus the shadow price of a slot. If d hires a teacher i who is not a tenured incumbent in d (which implies that the offer is based on d's preference rather than on the non-dismissal constraint), then for any district preference parameter vector $b \geq 0$, a teacher j is at least as preferable as i and hence must also receive an offer from d if the following (sufficient but not necessary) conditions are met: 1) j has 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). Then, for each teacher i, we can construct (observe) a subset O_i^s of all the offers they received, consisting of the inferred offers, the accepted offer, and, if i is tenured, the guaranteed offer from i's original employer d_{0i}. Given the assumption that $b \geq 0$, all options in O_i^s are offered to teacher i, therefore as long as 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 (the observed) O_i^s identifies their preferences— how much teachers value districts' characteristics relative to wages and how dispersed their preference shocks are.[18]

Observed matches are also informative of district preferences. First, as the number of teachers per district grows large, the lowest (x, c) among teachers working in d is the lowest (x, c) that district d is willing to hire. This is because teachers are subject to preference shocks with an unbounded support and an offer would be accepted by some teacher. The lowest (x, c) within each district identifies districts' preferences over teachers.

In practice, many districts are small, therefore, our identification also relies on the following observation. Given that the distribution of teachers' preferences is revealed from their choices

[17] We assume that teacher experience (x_1) enters district preferences 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, 1977) 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.

within O_i^s, we can predict the probability that a teacher would choose to work in each district if they had offers from *all districts.* As long as at least some districts are selective (i.e., they do not make offers to all teachers), accounting for teacher preference shocks, this predicted distribution of teacher-district matches will be systematically different from the observed matches, because a teacher can choose a district d only if they have an offer from d. That is, given teachers' preferences, districts' offer decisions—which are governed by districts' preferences—must rationalize the realized match distribution.

The identification argument can be illustrated with a simple example (see Online Appendix B4.3 for a graphical illustration). Consider the simpler case where teachers do not have preference shocks and suppose that two teachers i and j both prefer district 1 over district 2. If we observe i working in district 1 and j working in district 2, it must be the case that district 1 prefers i over j. The same argument applies when teachers have preference shocks: If teachers systematically prefer district 1 over district 2, then district 1 must prefer their hires over (most) teachers working in district 2. As long as the distribution of (x, c) in district 1 does not systematically dominate the distribution of (x, c) in district 2 in all dimensions, we can infer how much district 1 cares about x and c_2 relative to c_1 (the coefficient for c_1 is normalized to 1).

Discussion The argument above relies on three maintained assumptions.
A1: (x, c) are observable to all districts.[19] 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 in Online Appendix B4: 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.
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

[19] Jacob and Lefgren (2008) find that principals can generally identify very effective and very ineffective teachers but are less able to distinguish between teachers in the middle of the effectiveness distribution.

[20] In a centralized student-school matching system, Fack et al. (2019) define a student's feasible choice set as the set of 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).

true preferences than those inferred assuming that teachers had offers from all districts. As a robustness check, in Online Appendix B4, 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.

The most important informational assumptions we have made relate to how much districts and the researcher know about teacher effectiveness (or more generally, teacher traits that enter districts' payoff function). We discuss these assumptions in detail in Online Appendix B4.1. In particular, as long as districts and the researcher have the same measure of teacher effectiveness, be it an exact measure or a noisy (but unbiased) measure, our estimation strategy, estimates, and policy implications on expected outcomes remain the same. However, our identification strategy as described in Section 4.1.2 depends critically on the information symmetry between districts and the researcher, which allows us to separate teachers' preferences from districts' preferences.

4.2 Auxiliary Models

The identification argument in Section 4.1.1 suggests targeting statistics on the relationship between districts' wage schedules and their pre-determined conditions; the argument in Section 4.1.2 suggests targeting statistics that summarize 1) teachers' choices within the inferred offer subset O_i^s and 2) the realized distribution of teacher-district matches. 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 information extracted from *all* auxiliary models. Therefore, we target the following auxiliary models *jointly*. 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, measuring the responses of the predicted auxiliary models.

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

Aux 1 Coefficients from two regressions of the form

$$y_{id} = \beta_1^m w_{id} + I\left(\begin{array}{c} d_{0i} > 0, \\ d \neq d_{0i} \end{array}\right)\left[\begin{array}{c} \beta_2^m (x_{i1}) + \beta_3^m \ln (dist_{id}) \\ +\beta_4^m I (z_d \neq z_{d_{0i}}) \end{array}\right] + q_d\beta_4^m + \beta_5^m e^{\lambda_d} + \beta_6^m c_{1i}\lambda_d + \psi_i + \varepsilon_{id}^m,$$

where $y_{id} = 1$ if teacher i is matched with district d, and 0 otherwise. The right-hand-side variables are the same as those entering teachers' preferences, including $w_{id} \equiv w(x_i, c_i|\omega_d)$, the wage i would be paid by district d under wage rule (6). ψ_i is a teacher dummy that relates all $\{(i,d)\}_d$ observations associated with teacher i.[22] The two regressions differ in the number of observations, reflecting the 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.

Aux 1b The second regression includes every possible teacher-district pair, which serves to summarize observed teacher-district matches.

Aux 2 Moments of district-level teacher characteristics (x, c_1, c_2) by district groups (quintiles of λ_d, quintiles of budget per slot, and urban/suburban status), which supplement Aux 1b.

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

$$\begin{aligned}\omega_{dn} = &\beta_{0n}^w + q_d\beta_{1n}^w + \beta_{2n}^w\lambda_d + \beta_{3n}^w\kappa_d + \beta_{4n}^w M_d + X_d\beta_{5n}^w + \beta_{6n}^w\overline{TC}_d + \beta_{7n}^w\overline{TC}_d^{tenure} \\ &+ \beta_{8n}^w\overline{TC}_{z_d} + \beta_{9n}^w\overline{Tenured}_{z_d} + \beta_{10n}^w dem_d + \varepsilon_{dn}^w,\end{aligned}$$

where coefficients β_{1n}^w to β_{4n}^w are associated with district characteristics and constraints, and β_{5n}^w to β_{7n}^w are associated with the composition of district incumbents. In particular, X_d is the average x, $\overline{TC}_d$ is the average 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 $x_{1i} \geq 3$). The next two terms are about teachers originally working in other districts within d's commuting zone (i.e., $d_{0i} \neq d$, but $z_{d_{0i}} = z_d$): their average TC ($\overline{TC}_{z_d}$) and tenured rate ($\overline{Tenured}_{z_d}$). Lastly, dem_d is the share of Democratic votes in the district.

Aux 4 Cross-district wage schedule moments, which supplement Aux 3: $E(\omega_1)$, $E(\omega_2)$, $E(\omega_1^2)$, $E(\omega_2^2)$, $E(\omega_1\omega_2)$, $E(\omega = (1,0))$, $E(\omega_1 dem)$, and $E(\omega_2 dem)$.

[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 fixed effects. These fixed effects (not targeted) serve to capture the idea that the same teacher is choosing one district out of a given set of districts.

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 year 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 λ_d, the fraction of students in district d whose prior math scores were below the grade-specific state median. District characteristics q_d include indicators for urbanicity (urban, suburban or rural) and 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 B1).

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 Act 10 had expired and districts had obtained full autonomy over teacher pay.[23] 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.[24] We exclude secondary-school teachers because they often teach multiple grades, making it hard to

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

[24]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.

identify individual teacher contributions (Kane and Staiger, 2008; Chetty et al., 2014). Among elementary-school teachers, we focus on those for whom we can construct effectiveness measures, i.e., teachers in Grades 4-6. We further restrict attention to teachers of the same subject (math), so that the effectiveness measures are comparable across teachers.[25] 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 this subgroup 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. Online Appendix Figure B8 supports this hypothesis: The share of a district's budget spent on Grade 4-6 math teachers is stable over time.

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:

$$A_{kt} = \gamma Z_{kt}^{s} + \sum_{i:SG_{kt}=SG_{it}^{T}} \begin{pmatrix} I(\tau_k = 1)(\rho_1 x_{it} + v_{i1}) \\ +I(\tau_k = 2)(\rho_2 x_{it} + v_{i2}) \end{pmatrix} + \varepsilon_{kt}, \tag{12}$$

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; ε_{kt} is an i.i.d. idiosyncratic component. In the summation, SG_{kt} (SG_{it}^{T}) denotes the school-grade student k (teacher i) belongs to in year t; τ_k denotes a student's type ($\tau_k = 1$ if k is low-achieving, i.e., if 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 experience in year t and v_{in} is the part unexplained by x_{it}. Assuming that student-teacher sorting is random across school-grades conditional on observables, we estimate γ, ρ_1 and ρ_2 via OLS using data from 2007 to 2016; then, we use the Bayes estimator of Kane and Staiger (2008) to estimate v_{i1} and v_{i2} (Online

[25] The achievement models used to calculate teachers' effectiveness include students' lagged test scores; since students are tested starting from Grade 3, we can only calculate teacher effectiveness starting from Grade 4. 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). Online Appendix Figure B7 shows that the fraction of teachers switching into or out of math was very small in the data and the frequency did not change after Act 10. Therefore, we take the distribution of teachers across math and non-math subjects as given and our counterfactual analysis abstracts away from potential policy effects on teachers' sorting across subjects.

Appendix B1.3.1). Finally, we construct teacher effectiveness (c_{i1}, c_{i2}) in our model as

$$c_{in} \equiv \widehat{\rho}_n x_{it^*} + \widehat{v}_{in},\ n \in \{1, 2\}, \tag{13}$$

where t^* is 2014 for the estimation sample and 2010 for the validation sample.[26]

Two features of our achievement model deserve further discussion. 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 (Online Appendix B1.3.4). 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 (Online Appendix B1.3.5).

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 (12) 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 ε_{kt} and v_{in} are uncorrelated. Notice that this assumption allows for endogeneous district-teacher sorting (as is the case in our model), because we control for Z^s_{kt}, which includes school-grade fixed effects and year fixed effects. As we show in Online Appendix B1.3.3, 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. In addition, the precision of our effectiveness measures is comparable with that of measures from studies using teacher-classroom linked data. That being said, since teacher effectiveness is estimated with noise, our counterfactual policy implications should be interpreted with due caveat.

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.

[26]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.

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.

Choice Set for Wage Schedules **(Ω):** We first construct a grid Ω^o such that wages $w_d(x_i, c_i|\omega)$ under (6) and $\omega \in \Omega^o$ provide a good coverage of the observed wage distribution. We then expand the grid range, such that the model choice set $\Omega \supset \Omega^o$, to allow for the possibility that district choices may go out of the empirical range in counterfactual scenarios. We use the same $\Omega = \{0.9, 0.95, 1, 1.05, 1.1, 1.15\} \times \{0, 10, 30, 50, 75, 100, 200, 225\}$ throughout.

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

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

where $w_d(x_i, c_i|\omega)$ is given by wage rule (6); $(\omega_{d1}^o, \omega_{d2}^o)$ 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.[27]

District Capacity and Budget Constraints: Assuming data are generated from an equilibrium, in which districts' constraints bind, κ_d is then the number of teachers in our sample working in d in year t, and M_d is the sum of wages ($w_d(x_i, c_i|\omega_d^o)$) among these teachers, where $t = 2014$ (2010) for the estimation (validation) sample.

5.2 Summary Statistics

Panel A of Table 1 shows summary statistics for all 6,600 teachers in the estimation sample, for non-tenured teachers ($x_1 < 3$), and for those with over 10 years of experience ($x_1 \geq 10$). Fifty-three percent of all teachers have a graduate degree; this share is 6% among non-tenured teachers and 68% among teachers with over 10 years of experience. On average, non-tenured teachers are less effective than more experienced teachers in terms of both c_1 and c_2. However, the overall correlation between experience (x_1) and either c_1 or c_2, not shown in the table, is 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.

[27]The estimated slope coefficient of a linear model of w_i^o as a function of $w_d(x_i, c_i|\omega_d^o)$ equals 0.88 (with a standard error of 0.004) and an R^2 of 0.85. Focusing on the subsample of movers, the estimated slope coefficient is 1.08, with a standard error of 0.05 and an R2 of 0.75.

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

Table 1: Teacher and District Characteristics (2014)

A. Teacher Characteristics	All	$x_1 < 3$	$x_1 \geq 10$
x_1: Experience	14.6 (9.2)	1.4 (0.5)	19.7 (6.9)
x_2: MA or above	0.53 (0.50)	0.06 (0.24)	0.68 (0.47)
$10c_1$	0.12 (0.29)	0.04 (0.37)	0.12 (0.26)
$10c_2$	0.11 (0.33)	0.02 (0.42)	0.12 (0.31)
Corr (c_1, c_2)	0.67	-	-
# Teachers	6,600	627	4,384
B. District Characteristics	**All**	**λ_d 1st Quartile**	**λ_d 4th Quartile**
Urban	0.04	0.02	0.03
Suburban	0.15	0.34	0.09
λ_d	0.50 (0.12)	0.34 (0.07)	0.64 (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

Notes: Means and standard 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.[28] Column 2 shows 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

[28]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 (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).

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

Teacher's Choice Set	Inferred Offer Set		All Districts	
	coeff.	s.d.	coeff.	s.d.
wage	0.001	(0.0002)	-1.5$\times 10^5$	(2.6$\times 10^6$)
e_d^λ	-0.002	(0.008)	-0.0001	(0.0001)
$c_1 \times \lambda_d$	0.568	(0.283)	-0.020	(0.006)
$d \neq d_0$	-0.826	(0.012)	-0.982	(0.002)
$d \neq d_0 \times \exp \in [1, 2]$	0.476	(0.098)	0.833	(0.039)
$d \neq d_0 \times \exp \in [3, 4]$	0.267	(0.031)	0.236	(0.026)
$d \neq d_0 \times \exp \in [5, 9]$	0.085	(0.013)	0.099	(0.010)
$d \neq d_0 \times \exp \in [10, 14]$	0.020	(0.011)	0.014	(0.005)
$z_d \neq z_{d_0}$	-0.027	(0.005)	-0.0004	(0.0001)
ln(distance)	-0.019	(0.002)	-0.0001	(0.00002)
q_d :urban	0.014	(0.002)	0.004	(0.0002)
q_d :suburban	0.011	(0.002)	0.001	(0.0001)
q_d :large metro	0.096	(0.028)	0.012	(0.002)
# Obs	60,841		2,712,600	

Notes: OLS estimates of equations Aux 1a (left) and Aux 1b (right), obtained controlling for teacher fixed effects. Robust standard errors are in parentheses.

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.

Panel A of Table 3 summarizes districts' wage schedules. Districts' choices of ω_2 (rewards for teacher contribution) are more dispersed than their choices of ω_1. Although given the flexibility, 24% of districts continued to use their pre-reform wage schedules ($\omega = (1, 0)$) and only 50% of districts chose to reward teacher contribution ($\omega_2 > 0$). As we show in Table 4 below, the decision of rewarding teacher effectiveness is largely driven by political reasons. Two other possible explanations, which we abstract from, are 1) districts are ill-informed of teacher effectiveness in a biased way and therefore cannot reward them accordingly, and 2) districts care a lot about teachers' attributes beyond their experience, education, and effectiveness.[29] Panel

[29]The fact that we can fit the observed wages very well with a wage function that depends only on (x, TC) implies that 2) has very limited explanatory power; we thank the editor for pointing this out.

Table 3: District Wage Schedules (2014)

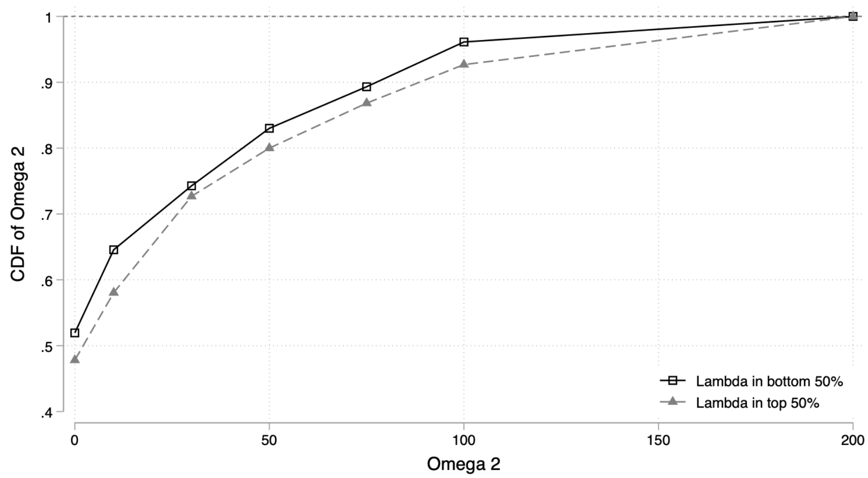
A. Summary stats of (ω_1,ω_2)		B. $w_d\left(x, c \mid \omega_d^o\right)$ in Realized Matches (\$1,000)		
ω_1 mean (std)	0.99 (0.04)	All Teachers: mean (std)		55.1 (11.6)
ω_2 mean (std)	31.3 (50.8)	Experience	< 3	37.2 (4.4)
$Corr$ (ω_1,ω_2)	-0.19		$\in [3, 4]$	41.0 (5.6)
Fr($(\omega_1, \omega_2) = (1, 0)$)	0.24		$\in [5, 9]$	48.0 (6.4)
Fr($\omega_2 > 0$)	0.50		≥ 10	56.5 (7.2)
C. District Characteristics by ω_2		$\omega_2 = 0$		$\omega_2 > 0$
Rural		0.80		0.83
λ_d>median		0.48		0.52
Budget/Capacity (\$1,000)		51.2		50.7
# Districts		205		206

Notes: Summary statistics of districts' wage schedules. Panel A shows means and standard deviations of ω_1 and ω_2, their correlation, and the fractions of districts with $\omega_1 = 1$ and $\omega_2 = 0$ (Fr$(\omega_1, \omega_2) = (1, 0)$) and with $\omega_2 > 0$ (Fr$\omega_2 > 0$). Panel B shows summary statistics of teachers' wages in the realized matches, given districts' wage schedules, for all teachers and separately by teachers' experience intervals. Panel C shows characteristics of districts with $\omega_2 = 0$ (left column) and $\omega_2 > 0$ (right column).

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. Districts with $\omega_2 > 0$ are more likely to be in rural areas and have higher fractions of low-achieving students (λ_d) and smaller per teacher budgets; these differences, though, are not statistically significant. Figure 4 further illustrates the correlation between ω_2 and student composition λ_d by plotting the cumulative distribution function of ω_2 separately for districts with λ_d above and below the median. The distribution of ω_2 in the first group first-order stochastically dominates that in the second group, i.e., districts with more low-achieving students are more likely to reward teacher effectiveness. One possible explanation is that it is difficult for disadvantaged districts to compete for experienced and effective teachers, therefore they set a higher ω_2 (which implies a lower ω_1 to balance the budget) to improve their attractiveness to young but effective teachers.

To further investigate districts' wage schedule choices, in Table 4 we use OLS regressions to relate a district's reward for teacher contribution (ω_{d2}) to the following groups of variables: 1) the share of Democratic votes in the 2012 Presidential election among residents in d's county (dem_d); 2) district d's student body, budget, capacity, and urbanicity; 3) the characteristics of teachers initially working in d and in d's commuting zone; and 4) the characteristics of district d's residents (American Community Survey 2013–17). In Column A, we show a specification

Figure 4: Cumulative Distribution Function of ω_2 by Groups of λ_d



Notes: This Figure plots the CDF of ω_2 for districts with λ_d above and below the median. The distribution of ω_2 in the first group first-order stochastically dominates that in the second group, i.e., districts with more low-achieving students are more likely to reward teacher effectiveness.

that includes the variables in groups 1), 2), and 3). We find that districts with a larger share of Democratic votes and districts with larger capacity tend to offer lower rewards for teacher effectiveness, i.e., they are more reluctant to deviate from the old zero-reward regime. We also find that ω_{d2} is (insignificantly) positively correlated with the share of low-achieving students in d and with d's budget, while it is (insignificantly) negatively correlated with the effectiveness of tenured teachers initially working in d. Although insignificant, the negative point estimate is intuitive: If tenured incumbents, whom the district cannot dismiss, are ineffective, a district can set higher ω_{d_2} (higher penalty for low effectiveness) to push these teachers out. In Column B, we add controls of the income distribution in the district; in Column C, we further control for residents' age and education. Estimates of the coefficient associated with dem_d are robust across specifications. This suggests that a district's wage schedule choice is explained, to a large extent, by the political views of its residents.

6 Estimation Results

6.1 Parameter Estimates

Table 5 shows estimated model parameters, along with their standard errors (in parentheses) derived numerically via the Delta Method. Panel A shows estimated parameters governing teachers' preferences. For most teachers, districts with higher fractions of low-achieving students (λ_d) are less desirable: An average teacher puts a premium of about \$4,100 on a district with

Table 4: Explaining District Reward for Teacher Contribution ω_{d2} (2014)

	A		B		C	
dem_d Share Democratic votes (2012)	-56.46	(26.39)	-57.52	(26.43)	-54.50	(27.12)
Capacity	-0.35	(0.17)	-0.35	(0.17)	-0.30	(0.18)
λ_d	25.24	(24.58)	36.56	(28.20)	36.62	(31.18)
Budget per teacher	0.53	(0.46)	0.47	(0.48)	0.55	(0.48)
Average TC of incumbent teachers	200.4	(2740.2)	193.3	(2719.0)	306.9	(2708.3)
Average TC of tenured incumbent teachers	-508.9	(2749.3)	-501.8	(2731.4)	-614.1	(2720.1)
Experience and education of incumbent teachers	Yes		Yes		Yes	
TC and share tenured of teachers in nearby districts	Yes		Yes		Yes	
District urbanicity	Yes		Yes		Yes	
Distribution of residents' income	No		Yes		Yes	
Distribution of residents' age and education	No		No		Yes	
# obs.			411			

Notes: OLS estimates of regressions with ω_{d2} as the dependent variable. Residents' income is captured by the natural logarithm of median income, the poverty rate, and the fraction of households with incomes higher than \$200,000. The distribution of residents' age and education is captured by the fractions of people younger than 15, those older than 64, and those with a college education. Robust standard errors are in parentheses.

$\lambda_d = 0.3$ over an otherwise identical district with $\lambda_{d'} = 0.7$. However, teachers who are more effective in teaching low-achieving students are more willing to teach in these districts: The coefficient of $c_1 \times \lambda_d$ is positive, although imprecisely estimated. Rural districts are less attractive than their urban and suburban counterparts. The remainder of Panel A shows that teachers face high moving costs, especially among more experienced teachers.[30] Individuals compare the total value of each option when making their choices, including their preference shocks (governed by σ_ϵ). High moving costs help explain the lack of teacher mobility, especially among more experienced teachers, while preference shocks absorb idiosyncratic reasons for mobility. Our findings of large moving cost and dispersion of preference shocks are consistent with previous studies on worker mobility (e.g., Kennan and Walker, 2011).

Panel B shows district preference estimates. Districts significantly value a teacher's contribution to its students' achievement but do not value teacher experience and education per se. In addition, districts value a teacher's contribution to its low-achieving students more than their contribution to its high-achieving students. Table A1 shows the estimated parameters governing the cost a district faces for deviating from the pre-Act 10 wage schedule; consistent with Table 4, the estimates indicate that districts with a larger share of democratic voters face

[30]Moving rates are consistently higher for less experienced teachers throughout our simulations. For example, in the baseline equilibrium, among teachers with $d_0 \neq 0$, the average moving rate by experience group is 0.71, 0.17, 0.05, 0.001, and 0.001.

Table 5: Parameter Estimates

A. Teacher Preference						
wage ($1,000)	1	normalized	$I\left(d \neq d_0\right)$ ×Yrs of experience:	1-2	-10.66	(6.69)
e^{λ_d}	-6.22	(1.38)		3-4	-55.23	(22.03)
$c_1 \times \lambda_d$	22.11	(26.34)		5-9	-78.94	(6.33)
q_d :urban	16.55	(3.36)		10-14	-150.96	(41.10)
q_d :suburban	21.19	(2.13)		≥ 15	-151.29	(50.26)
q_d :large metro	2.96	(8.31)	$I\left(z_d \neq z_{d_0}\right)$		-19.98	(22.86)
σ_ϵ	18.73	(0.94)	ln (distance in miles)		-14.42	(5.24)
B. District Preference						
c_1	1	normalized	Yrs of experience:	1-2	0.012	(0.11)
c_2	0.67	(0.02)		3-4	0.016	(0.02)
MA or above	2.0×10^{-5}	(0.03)		5-9	0.017	(0.37)
				10-14	0.033	(0.02)
				≥ 15	0.040	(0.03)

Notes: Model parameter estimates. Standard errors (in parentheses) are calculated numerically via the Delta Method.

a higher cost. In Section 7.2.1, we explore the implications of both teachers' moving costs and districts' costs for deviating from rigid wage schedules.

6.2 Within-Sample Fit and Model Validation

Within-Sample Fit: Table A2 shows that the model well captures the teacher-school sorting as summarized by the two regressions specified in Section 4.2 (Aux 1a on the left and Aux 1b on the right). The model fits the data well as measured by Aux 1a but less so as measured by Aux 1b. Notably, in Aux 1a, the model replicates the strong positive coefficient on $c_1 \times \lambda_d$ (0.57 in the data vs 0.50 in the model); in Aux 1b, this coefficient is -0.02 in the data but 0.001 in the model. The upper panel of Table A3 shows model fits for the distribution of ω. Overall, the model fits the data well, although it underpredicts the dispersion of ω_2 and the fraction of districts choosing $\omega_2 = 0$. The lower panel shows model fits for district characteristics by whether or not they reward teacher contribution; these statistics are not directly targeted in the estimation. Consistent with the data, the model predicts that districts with $\omega_2 > 0$ are slightly more disadvantaged. Online Appendix Tables B21 and B22 show additional model fit measures (Aux 2 and Aux 4), which are also good.

Model Validation: Using the parameter estimates in Table 5, we apply our model to data from the pre-Act 10 era, when districts were forced to use the rigid wage schedule. We simulate the model under rigid pay and initial conditions from 2010 and contrast model-predicted outcomes with those in the 2010 data (Table A4 and Online Appendix Table B23). Despite the substantial change in the policy environment, our model (estimated using post-Act 10 data) is able to match the pre-Act 10 data well, although the sorting of high-c_1 teachers into high-λ_d districts, captured by the coefficient for $c_1 \times \lambda_d$, is weaker in the model than in the data (Table A4).

7 Counterfactual Experiments

We use our estimated model to examine the equity-efficiency implication of flexible pay and to evaluate a set of counterfactual state bonus programs. Throughout, we measure efficiency and equity in terms of students' access to effective teachers, which maps into students' achievement.[31] Specifically, we pay special attention to the following metrics:

1. Average total contribution $\overline{TC}_{D'}$ among teachers working in a given group of districts $D' \subseteq D$. Since teacher contribution enters student achievement additively, an increase in $\overline{TC}_{D'}$ maps one-to-one into an increase in the average achievement for students in D'. When $D' = D$, $\overline{TC}_D$ measures the overall match *efficiency* in the market. Moreover, a policy will improve cross-district educational equity if it increases $\overline{TC}_{D'}$ more for high-λ_d districts, i.e., districts with higher fractions of low-achieving students, than it does for low-λ_d districts.
2. Average teacher contribution to the state's low-achieving students $\overline{C}_1$ and high-achieving students $\overline{C}_2$. An increase in $\overline{C}_1$ ($\overline{C}_2$) maps one-to-one into an increase in the average achievement for low-achieving (high-achieving) students. A policy will narrow the achievement gap between the two groups if it improves $\overline{C}_1$ more than it improves $\overline{C}_2$.

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.

[31] Although teachers make choices to maximize their welfare in each case, we do not consider teachers' welfare in our efficiency-equity calculations.

Table 6: Flexible Pay vs Rigid Pay

	$\frac{\text{Flexible-Rigid}}{\lvert\text{Rigid}\rvert}$(%)			Frac. Low-achieving
	$\overline{TC}$(all)	$\overline{C}_1$(low-achieving)	$\overline{C}_2$(high-achieving)	λ_d
All Districts	0.04	-0.01	0.10	0.50 (0.12)
By quart. of λ_d: 4th	-0.34	-0.24	-0.57	0.64 (0.06)
3rd	0.37	0.45	0.25	0.54 (0.02)
2nd	0.05	0.03	0.06	0.47 (0.02)
1st	0.05	-0.37	0.27	0.34 (0.07)

Col. 1-3: % difference $\overline{TC}$, $\overline{C}_1$, $\overline{C}_2$ between the flexible-pay and the rigid-pay regime.

Col. 4: mean (cross-district std dev.) of the fraction of low-achieving students in a district group.

The first three columns in Table 6 show percentage changes in $\overline{TC}$, $\overline{C}_1$, and $\overline{C}_2$ associated with a shift from the rigid-pay regime to the flexible-pay regime. These effects are shown in Row 1 for the entire state, and in Row 2 to Row 5 for each of the four groups of districts ranked by λ_d (fraction of low-achieving students) from the highest to the lowest. The last column summarizes the mean and the standard deviation of λ_d among districts in each row. The following four findings stand out:

(1) Flexible pay allows districts to directly reward teacher contribution, which encourages comparative advantage-based sorting and hence improves efficiency: At the state level, teachers' total contribution $\overline{TC}$ increases by 0.04%. However, on average, the gain accrues entirely to high-achieving students, while low-achieving students experience a slight decline in teacher contribution, widening achievement gap between the state's two groups of students.

(2) As districts' competition intensifies under flexible pay, districts with the highest fraction of low-achieving students lose by all three metrics, with both types of students losing effective teachers to other districts. This is because, all else being equal, most teachers prefer districts with fewer low-achieving students.

(3) Perhaps surprisingly, the second-highest-λ_d district group gains the most by almost all three metrics. This is largely explained by teachers' moving costs. To avoid large moving costs, as effective teachers move out of highest-λ_d districts, they are more likely to move to nearby districts, which are less likely to be low-λ_d districts due to the fact that student bodies tend to be similar among nearby districts (Online Appendix Figure B10).

(4) Flexible pay allows districts to compete more effectively for teachers better suited for their students; as a result, a district's majority student group—low- or high-achieving—benefits more or loses less than the minority group (Column 2 versus Column 3). Most evidently, in districts where a large majority of students are high-achieving (Row 5), teacher effectiveness improves by 0.27% for high-achieving students while it decreases by 0.37% for low-achieving students, widening the within-district achievement gap between the two groups of students.

In sum, the impacts shown in Table 6, although small in magnitude, reflect a trade-off between efficiency and equity. Flexible pay improves efficiency, but it also widens achievement gaps across districts and between low- and high-achieving students.

7.2 State-Funded Bonuses

Given the trade-off reflected in Table 6, we now explore the possibility of improving *both* efficiency *and* equity under flexible pay via a commonly used policy tool: state-funded teacher bonuses. To design our bonus formula, we build on the insights provided by our estimated model. On the supply side, although teachers differ in their comparative advantages, they face large moving costs, which act as an obstacle to efficient sorting. In addition, most teachers prefer teaching high-achieving students, which acts as an obstacle to equity. On the demand side, a large fraction of districts choose not to reward teacher effectiveness, which constitutes an additional obstacle to both efficiency and equity.

Our bonus formula is designed to overcome these obstacles. For the supply side, the formula consists of an efficiency incentive and an equity incentive; for the demand side, it gives a direct incentive for districts to reward effectiveness. Specifically, a teacher with effectiveness $c = (c_1, c_2)$ teaching in district d would obtain a state-funded bonus given by

$$B\left(c, \lambda_d, \omega_d\right) = \min\left\{\max\left\{\left[r_0 TC\left(c, \lambda_d\right) + r_1 c_1 \lambda_d\right]\omega_{d2}, 0\right\}, \overline{B}\right\}. \tag{B}$$

To avoid extreme values, we bound bonuses between 0 and $\overline{B}$ (set at twice the standard deviation of the observed wage distribution). Between the two bounds, the formula has three features that make our formula novel and distinct from existing teachers' bonus formulae (see Neal et al., 2011; Pham et al., 2021, for reviews).

(1) The component $r_0 TC\left(c, \lambda_d\right)$ rewards a teacher for their TC at the rate r_0. Since a teacher's TC is higher when their comparative advantage better matches a district's student composition, this component incentivizes more efficient sorting.

(2) The component $r_1 c_1 \lambda_d$ additionally rewards a teacher for their $c_1 \lambda_d$ (contribution to d's low-achieving students) at the rate r_1. By providing an additional incentive for high-c_1 teachers to teach in low-achievement districts, this component directly favors these districts, improving equity. The goal of equity is shared by many existing formulae that reward teachers for working in disadvantaged places (Bobba et al., 2021; Swain et al., 2019); however, designed solely to improve equity, these formulae do not account for efficiency. In contrast, our formula accounts for both equity (via $r_1 c_1 \lambda_d$) and efficiency (via $r_0 TC(\cdot)$) .

(3) The factor ω_{d2} ties a teacher's bonus to the rate at which the district rewards teacher contribution. Since a higher ω_{d2} would enable district d to obtain more state bonuses for

its teachers, this feature incentivizes districts to reward teacher effectiveness in their own wage schedule. This component—missing from existing bonus formulae—proves to be very important for program effectiveness.[32]

The key policy variables in our formula are the bonus rates (r_0, r_1). Different vectors of (r_0, r_1) would induce different reactions from districts and teachers and hence different equilibrium outcomes.[33] For illustration, we present results from three bonus programs under flexible pay (labeled as New1 to New3), each associated with a different (r_0, r_1), and contrast them with the baseline flexible pay equilibrium (labeled as Base); further details and additional simulation results are reported in Online Appendix B6.3. We calibrate the vector of bonus rates such that all programs New1 to New3 are equally costly in the equilibrium, at about $1,620 per teacher. Given this total cost, the equilibrium average state bonus for each recipient ranges from $4,120 to $4,210, depending on program specifics (lower panel of Table 7). These amounts are comparable to relatively mild bonus programs implemented in other states, but, as mentioned above, with very different formulae than ours.[34]

Each of our three programs serves a different policy goal. In **New1**, we seek to improve efficiency, as measured by total $\overline{TC}$ at the state level. Among the simulations we have conducted, the vector $(r_0, r_1) = (2.3, 3.1)$ performs the best and improves efficiency by 0.26%. It is worth noting that, even though our goal is to improve efficiency, in New1, teachers receive an additional bonus for their contribution to low-achieving students ($r_1 > 0$). This is because most teachers, including many who have a comparative advantage in teaching low-achieving students, prefer to teach high-achieving students. Therefore, bonus schemes that favor districts with more low-achieving students outperform a purely TC-based bonus scheme: At the same total cost, the latter leads to a lower efficiency gain of 0.18% (Online Appendix Table B17). However, the gains under New1 accrue only to high-achieving students, who gain by 0.59% on average, while low-achieving students lose by 0.06% on average. Similarly, the gains go to districts serving fewer low-achieving students, at the expense of districts serving more low-achieving students. In sum, New1 succeeds in its goal of improving efficiency, but ends up

[32]Without the third feature, i.e., a teacher's bonus is based on $r_0 TC(c, \lambda_d) + r_1 c_1 \lambda_d$ regardless of ω_{d2}, programs will be much less effective. For example, if we set $r_1 = 0$ and adjust r_0 to exhaust the same total budget of $1,620 per teacher, this alternative formula will lead to an efficiency gain of 0.08%, while our original formula will lead to an efficiency gain of 0.18%.

[33]We assume our model parameters are policy-invariant. In particular, we do not consider the possibility that state bonuses may directly change voters' supports for teacher unions and wage rigidity, thereby changing the resistance cost function $R(\cdot)$. In Section 7.2.1 and Online Appendix B6.3, we examine how policy effects differ in the absence of resistance costs.

[34]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 et al., 2011). Findings from these programs are mixed. Math scores improved by 3% under CLE; the NYC bonus program had no effect on achievement; and POINT had no effect on achievement except for one grade (the effect was positive for one year).

hurting equity.

Table 7: State-Funded Bonuses

(%)	New1-Base / |Base|	New2-Base / |Base|	New3-Base / |Base|
$\overline{TC}$ for all students in the state (efficiency)	0.26	0.04	0.15
$\overline{C}_1$ for low-achieving students in the state	-0.06	0.35	0.16
$\overline{C}_2$ for high-achieving students in the state	0.59	-0.26	0.14
$\overline{TC}$ in 4th quartile λ_d districts	-0.33	1.01	0.47
$\overline{TC}$ in 3rd quartile λ_d districts	-0.34	0.69	0.21
$\overline{TC}$ in 2nd quartile λ_d districts	0.41	0.27	0.32
$\overline{TC}$ in 1st quartile λ_d districts	1.08	-1.37	-0.27
Bonus Rates (r_0, r_1)	(2.3, 3.1)	(0, 7.0)	(1.6, 4.3)
Teachers receiving state bonuses (B>0)	38.5%	39.4%	39.0%
Avg. bonus for recipients E(B|B>0) ($1,000)	4.21	4.12	4.15
Program cost ($1,000 per teacher)		1.62	

Notes: Flexible-pay equilibrium with different bonus programs. The top panel shows changes in the metrics described in the text between the equilibrium with the bonus program and the baseline flexible-pay equilibrium. The bottom panel summarizes the characteristics of each bonus program.

In an attempt to improve equity, in **New2** we reward teachers only based on $c_1\lambda_d$, thus providing strong incentives for high-c_1 teachers to teach in districts with more low-achieving students. New2 leads to a state average gain of 0.35% for low-achieving students but a 0.26% loss for high-achieving students. Cross-district achievement gaps also narrow as districts serving more low-achieving students gain more from this program.

Results from New1 and New2 illustrate the dilemma often faced by policy makers: the trade-off between efficiency and equity. Is it possible to gain in both efficiency and equity in a more balanced manner and benefit both types of students? Our final program **New3** aims at achieving this goal. Intuitively, the vector of bonus rates in this case should be in between those under New1 and New2. Indeed, at bonus rates $(r_0^2, r_1^2) = (1.6, 4.3)$, New3 leads to a 0.15% efficiency gain, with low-achieving students gaining slightly more than high-achieving students. In addition, New3's impacts are also more evenly distributed across districts than both New1 and New2.

Remark 1 *For equity, we focus on the achievement gap between the state's low- and high-achieving students ($\overline{C}_1$ versus $\overline{C}_2$) and the achievement gap across districts ($\overline{TC}$ in high- versus low-λ_d districts). A program that improves the average achievement of a student group or a district group is not guaranteed to benefit all students in that group. For example, although*

New2 is designed to benefit low-achieving students on average, it causes losses for both types of students in the lowest-λ_d district group (Online Appendix Table B17). Similarly, for efficiency, we focus on the state's average student achievement (state level $\overline{TC}$); it is impossible to achieve Pareto improvement at the student level.

7.2.1 Discussion: Magnitudes of Policy Impacts

Our counterfactual simulations demonstrate that it is possible to design bonus programs to improve efficiency and equity. In general, these two goals are hard to achieve simultaneously. For example, Section 7.1 shows that a flexible pay regime improves efficiency but hurts equity. Our contribution is to show that, by exploiting teachers' comparative advantages, *carefully designed* policies can lead the market equilibrium toward more efficient and equitable allocations. However, the magnitudes of our counterfactual policy impacts are small, consistent with previous studies on the limited impacts of monetary incentives on teacher sorting (e.g., Clotfelter et al., 2011; Russell, 2020). We now investigate some possible reasons.

One possible reason is that the current equilibrium teacher-school sorting is close to being efficient and/or equal, so there is not much room for improvement. This is not the case in our data: Online Appendix B6.1 shows that it is possible to achieve much higher efficiency or equity if one could ignore teachers' preferences and dictatorially allocate teachers to districts. An efficiency-seeking dictator can improve efficiency by 31% relative to the baseline (at the cost of equity); a dictator who seeks to help low-achieving students can improve their performance by 70% (while hurting high-achieving students by 56%).

In reality, though, a feasible policy intervention needs to respect teachers' and districts' preferences and constraints, which ultimately govern equilibrium outcomes and policy effects. We quantify the impact of two potentially important factors: teachers' moving costs $\Gamma(\cdot)$ and districts' resistance costs $R(\cdot)$. 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 8. Relative to the baseline, teacher TC is 2.43% higher in the equilibrium without moving costs (Column 1). This comes at the cost of equity: Average TC declines by 6.28% for low-achieving students and increases by 10.78% for high-achieving students. These changes are larger than the impact of any of our bonus programs shown in Table 7. This is unsurprising since teachers' moving costs (Table 5) are much higher than our bonuses. When we additionally remove districts' resistance costs

(Column 2), TC increases further and equity worsens.[35]

Table 8: Moving Costs, Resistance Costs, and Policy Impacts

	Zero Costs vs Base		New2 Impacts	
	1	2	3	4
%	$\frac{\text{Case1-Base}}{\lvert\text{Base}\rvert}$	$\frac{\text{Case2-Base}}{\lvert\text{Base}\rvert}$	Case2 $\frac{\text{New2-Case2}}{\lvert\text{Case2}\rvert}$	$\frac{\text{New2-Base}}{\lvert\text{Base}\rvert}$
$\overline{TC}$	2.43	2.72	0.05	0.04
$\overline{C}_1$	-6.28	-7.01	1.86	0.35
$\overline{C}_2$	10.78	12.05	-1.41	-0.26

Differences in TC, c_1, and c_2, between the baseline flexible-pay equilibrium and equilibria under different scenarios. *Base* refers to the baseline flexible-pay equilibrium. *Case 1* refers to flexible-pay equilibrium with zero moving cost. *Case 2* refers to flexible-pay equilibrium with zero moving cost and zero resistance costs.

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 larger. Districts' resistance costs for deviating from rigid pay mildly enhance these effects. These two costs could also play a major role in mediating the impact of our counterfactual bonus programs. To illustrate this, we introduce our bonus program **New2** to the market under Case 2. Recall that New2 seeks to help low-achieving students by rewarding teachers only based on $c_1\lambda_d$; we re-calibrated bonus rates to keep the total cost the same as before. We find that the impact of New2 is larger on an economy without moving and resistance costs (Column 3) than it is on the baseline economy (Column 4). Without resistance costs, districts are more willing to reward teacher contribution; without moving costs, teachers are more willing to move. As a result, New2 increases low-achieving students by 1.86% without moving and resistance costs, but only by 0.35% 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.

7.3 Additional Simulations

We have conducted additional simulations in Online Appendix B6, the results of which are summarized below.

[35] By setting both the moving cost and the resistance cost to zero, Case 2 can provide a sense of the long-run equilibrium outcomes under flexible pay, when the market has fully adjusted to a new steady state. We thank an anonymous referee for pointing this out.

Extensive Margin Responses and Repeated Static Games As discussed in Section 3.4, our framework is static and it abstracts from teachers' entry and exit decisions. As such, our policy implications are best interpreted as short-run effects. A rigorous analysis of policy impacts in the long run with teachers' extensive-margin responses is beyond the scope of our paper. As an attempt to understand the impact of state bonus programs when teachers' entries and exits are taken into account, we conduct additional simulations in a static setting and in a setting where the static game is played repeatedly over time. To conduct these exercises, we need to make additional assumptions. With these assumptions, we first estimate teachers' entry and exit probabilities in the baseline equilibrium. This procedure estimates these baseline probabilities as functions of teachers' observables including (c_1, c_2), but it does not allow us to estimate how these probabilities would change with respect to pay. Therefore, we borrow extensive-margin elasticities from the literature. With due caveats, the results from these simulations suggest that if teachers' extensive-margin responses are non-trivial and if the programs are in place for multiple years, the equity-efficiency gains from our bonus programs could be significantly larger. For example, assuming that teachers' entry-exit elasticity with respect to pay is one and that the static market equilibrium is played repeatedly for 5 years under a given bonus program, we simulate a teacher's path of movements at both the extensive margin (entry/exit) and the intensive margin (the history of districts they work in during these 5 years). Five years later, our program New1 (New2) would lead to an efficiency gain of 2.99% (2.08%) and benefit both low- and high-achieving students.

Program Generosity To illustrate the role of program generosity, we simulate three bonus programs with bonus rates (r_0, r_1) set at 1.5, 2.0, and 2.5 times $(2.3, 3.1)$, the latter being the bonus rates used in New1. These three programs are increasingly more costly for the state and they all lead to higher efficiency gains than the gain achieved by New1 (0.26%). However, the program effect levels off quickly: The efficiency gain under these three programs are 0.32%, 0.34%, and 0.34% respectively. This result demonstrates that with equilibrium responses from both sides of the market, one should not expect a simple linear relationship between program costs and their impacts. As such, the design of the bonus program is at least as important as the bonus budget.

The Role of $R(\cdot)$ To study the role of districts' resistance cost $R(\cdot)$ by itself, we re-simulated all our three bonus programs by setting $R(\cdot) = 0$ while keeping all other parameters at their estimated values. Perhaps not surprisingly, without resistance costs, districts are more responsive to our bonus programs. For example, at New1's bonus rates $(2.3, 3.1)$, if we set $R(\cdot) = 0$, this program will yield an efficiency gain of 0.34% rather than 0.26%, but it will also be twice as costly as New1 with resistance costs.

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. 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, our static equilibrium model is better suited to study short-run rather than long-run policy impacts; an important but rather difficult extension is to consider the market in a dynamic equilibrium setting. On the supply side, this extension would reflect teachers' life-cycle concerns; on the demand side, districts would consider how their choices would affect both their current return and their future competitiveness. Ultimately, this extension 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. Technical report, National Bureau of Economic Research.

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. Johnston, and I. Sorkin (2022). Teacher labor market equilibrium and the distribution of student achievement. Technical report, National Bureau of Economic Research.

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. Technical report, National Bureau of Economic Research.

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*.

Burgess, S., E. Greaves, and R. Murphy (2022). Deregulating teacher labor markets. *Economics of Education Review 88*, 102253.

Cardona, F. (2006). Performance related pay in the public service in oecd and eu member states. *Programme SIGMA de l'OCDE. Paris*.

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. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of economics and statistics 86*(1), 195–210.

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.

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.

Ederer, T. (2023). Labor market dynamics and teacher spatial sorting.

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.

Express, E. D. (2021). Data library.

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.

Gershenson, S., C. M. Hart, J. Hyman, C. A. Lindsay, and N. W. Papageorge (2022). The long-run impacts of same-race teachers. *American Economic Journal: Economic Policy 14*(4), 300–342.

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. Thiemann, and G. Zamarro (2023). Teacher-to-classroom assignment and student achievement. *Journal of Business & Economic Statistics 41*(4), 1328–1340.

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.

Jacob, B. A. and L. Lefgren (2008). Can principals identify effective teachers? evidence on subjective performance evaluation in education. *Journal of labor Economics 26*(1), 101–136.

Johnston, A. C. (2021). Preferences, selection, and the structure of teacher compensation.

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research.

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. (1977). Modelling the choice of residential location.

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

Menzel, K. (2015). Large matching markets as two-sided demand systems. *Econometrica 83*(3), 897–941.

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

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

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

Pham, L. D., T. D. Nguyen, and M. G. Springer (2021). Teacher merit pay: A meta-analysis. *American Educational Research Journal 58*(3), 527–566.

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.

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 (2006). Do teachers really leave for higher paying jobs in alternative occupations? *Advances in economic analysis & policy 6*(1).

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.

Swain, W. A., L. A. Rodriguez, and M. G. Springer (2019). Selective retention bonuses for highly effective teachers in high poverty schools: Evidence from tennessee. *Economics of Education Review 68*, 148–160.

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

Willén, A. (2021). Decentralization of wage determination: Evidence from a national teacher reform. *Journal of Public Economics 198*, 104388.

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

$$\max_{\{o_d(\cdot)\}} \left\{ \int o_d(x, c, d_0|\omega) h_d(x, c, d_0, \omega) [xb_0 + b_1\lambda_d c_1 + b_2(1 - \lambda_d) c_2] dF(x, c, d_0) \right\} \tag{14}$$

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

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 ν_κ and ν_M are the non-negative multipliers associated with the adjusted capacity and budget constraints. The capacity (budget) is adjusted by netting out the expected slots (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:

$$o_d(x, c, d_0|\omega) \begin{cases} = 1 & \text{if } \varphi(x, c, \lambda_d) - \nu_\kappa - w_d(x, c|\omega) \nu_M > 0 \\ = 0 & \text{if } \varphi(x, c, \lambda_d) - \nu_\kappa - w_d(x, c|\omega) \nu_M < 0 \\ \in [0, 1] & \text{if } \varphi(x, c, \lambda_d) - \nu_\kappa - w_d(x, c|\omega) \nu_M = 0 \end{cases}, \tag{15}$$

$$\int o_d(x, c, d_0|\omega) h_d(x, c, d_0|\omega) dF(x, c, d_0) \leq \kappa_d, \tag{16}$$

$$\text{and} \int o_d(x, c, d_0|\omega) h_d(x, c, d_0|\omega) w_d(x, c|\omega) dF(x, c, d_0) \leq M_d. \tag{17}$$

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 (15).

For a given ω, 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 ν_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., (16) or (17) is binding.
4) Calculate the district's value associated with this ν_M, and optimize over ν_M to find the maximum; $o_d(\cdot|\omega)$ associated with the optimal ν_M are the optimal job offers under ω.
In the outer loop, the district searches over ω to optimize its objective (9). Both (16) and (17) bind in the equilibrium throughout our simulations.

A2. Detailed Function Forms:

We model the resistance cost a district faces $R_d(\omega)$ as the following
$R_d(\omega) = R(\omega, dem_d) =$

$$\exp(\xi_{00} + \xi_{01} dem_d) + \exp(\xi_{11} + \xi_{12} dem_d) |\omega_1 - 1| + \exp(\xi_{21} + \xi_{22} dem_d) \frac{\omega_2}{100}. \qquad (18)$$

The vector $\xi_0 \equiv (\xi_{00}, \xi_{01})$ captures the fixed cost of deviating from the rigid-pay schedule; ξ_1 and ξ_2 capture the incremental costs for larger deviations from the pre-reform ω_1 and ω_2 respectively. Parameters $\{\xi_{k1}\}_{k=0}^{2}$ measure the extent to which these costs vary with the political views of district residents, as measured by dem_d. To impose the restriction that costs are non-negative, we use exponential function to parameterize each of the three parts of the cost but parameters ξ unrestricted.[36] Estimates of these parameters are reported in Table A1.

Table A1: Other Parameter Estimates: Wage Setting Cost $R(\omega, dem_d)$

		constant			dem_d	
Fixed cost for deviation	ξ_{00}	-0.48	(0.20)	ξ_{01}	1.51	(0.45)
Incremental cost wrt $\omega_1$deviation	ξ_{10}	1.71	(0.13)	ξ_{11}	2.48	(0.30)
Incremental cost wrt $\omega_2$deviation	ξ_{20}	-1.00	(0.10)	ξ_{21}	2.07	(0.34)
std dev of ω choice shocks (σ_{η_ω})		0.91 (0.03)				

Notes: Estimates of the parameters in (18). Standard errors (in parentheses) are derived numerically via the Delta Method.

[36]We experiemented with different specifications (e.g., linear rather than log cost with respect to ω_2 deviation), the current specification fits the data pattern the best.

Table A2: Model Fit: OLS of Teacher-District Match (post-Act 10)

Teacher's Choice Set	Inferred Offer Set		All Districts	
	Data	Model	Data	Model
wage	0.001	0.001	-1.5×10^5	-0.5×10^{-5}
e_d^λ	-0.002	-0.002	-0.0001	-0.0002
$c_1 \times \lambda_d$	0.568	0.499	-0.020	0.001
$d \neq d_0$	-0.826	-0.896	-0.982	-0.998
$d \neq d_0 \times \exp \in [1,2]$	0.476	0.339	0.833	0.728
$d \neq d_0 \times \exp \in [3,4]$	0.267	0.203	0.236	0.177
$d \neq d_0 \times \exp \in [5,9]$	0.085	0.078	0.099	0.057
$d \neq d_0 \times \exp \in [10,14]$	0.020	0.010	0.014	-0.0002
$z_d \neq z_{d_0}$	-0.027	-0.029	-0.0004	-0.0003
ln(distance)	-0.019	-0.010	-0.0001	0.00005
q_d :urban	0.014	-0.002	0.004	0.001
q_d :suburban	0.011	0.01	0.001	0.001
q_d :large metro	0.096	0.127	0.012	0.001

Notes: OLS estimates of equations Aux 1a (*Inferred Offer set*) and Aux 1b (*All Districts*), obtained controlling for teacher fixed effects, obtained using the data and the model. Data are from post-Act 10.

Table A3: Model Fit: District Wage Schedules

A. Summary Stats (ω_1,ω_2)	Data	Model		Data	Model
$E(\omega_1)$	0.99	0.99	$E(\omega_2)$	31.3	31.2
$E(\omega_1^2)$	0.98	0.99	$E(\omega_2^2)$	3562.2	3194.1
$E(\omega_1\omega_2)$	30.47	30.87			
$\text{Fr}((\omega_1,\omega_2)=(1,0))$	0.24	0.31	$\text{Fr}(\omega_2=0)$	0.50	0.42
B. District Characteristics by ω_2	$\omega_2 = 0$			$\omega_2 > 0$	
	Data	Model		Data	Model
Rural	0.80	0.80		0.83	0.81
λ_d>median	0.48	0.49		0.52	0.50
Budget/Capacity ($1,000)	51.2	51.0		50.7	50.9

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

Table A4: 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	0.001	-1.6×10^{-6}	-3.4×10^{-6}
e^{λ_d}	-0.017	-0.003	-0.0002	-0.0001
$c_1\times\lambda_d$	1.00	0.42	0.005	-0.0003
$d \neq d_0$	-0.94	-0.96	-0.99	-1.00
$d \neq d_0 \times \exp \in [1,2]$	0.65	0.50	0.79	0.68
$d \neq d_0 \times \exp \in [3,4]$	0.16	0.26	0.14	0.23
$d \neq d_0 \times \exp \in [5,9]$	0.06	0.07	0.05	0.06
$d \neq d_0 \times \exp \in [10,14]$	0.02	0.01	0.01	-0.0002
$I\left(z_d \neq z_{d_0}\right)$	-0.003	-0.01	-0.0001	-0.0003
ln(distance)	-0.009	-0.004	-0.00003	0.00001
q_d : urban	0.008	-0.001	0.002	0.001
q_d : suburban	0.003	0.001	0.001	0.001
q_d : large metro	0.04	-0.003	0.009	0.0004

Notes: OLS estimates of equations Aux 1a (*Inferred Offer set*) and Aux 1b (*All Districts*), obtained controlling for teacher fixed effects. Data vs model estimates, pre-Act 10 .