Equilibrium in the Market for Public School Teachers: District Wage Strategies and Teacher Comparative Advantage
(Online Appendix)

Barbara Biasi, Chao Fu and John Stromme*

April, 2022

B1 Data Details

B1.1 Sample Construction

We construct our samples as follows. For estimation and empirical analysis, we focus on full-time Grades 4-6 math teachers employed in Wisconsin school districts in 2014 (411 districts and 6,625 individuals).[1] We exclude 3 teachers from the sample, whose schools did not report test scores. We also exclude 22 teachers with missing information on years of experience. This leaves us with 6,600 teachers and 411 districts in the final estimation sample.
For the validation sample, we focus on 6,751 full-time Grades 4-6 math teachers employed in 411 districts in 2010. We exclude 10 teachers with missing information on years of experience. This leaves us with 6,741 teachers and 411 districts in the final validation sample.

B1.2 Teacher's Previous District

Our model requires identifying the district where each teacher was working at the beginning of the model period (d_{i0}). For the estimation sample, which is based on 2014 data, we define d_{i0} as follows. If the teacher never moved or moved only once between 2011 and 2014, d_{i0}

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

[1]Wisconsin had 424 school districts in 2014, 11 of which did not have any elementary school, and 2 of which did not have any full-time Grades 4-6 math teachers.

is the district where she was employed in 2011. If a teacher moved more than once between 2011 and 2014, we set d_{i0} to be the last employer she worked for before 2014. For example, if teacher i worked in District A in 2011 and 2012, and District C in 2013 and 2014, then $d_{i0} = C$. If teacher i worked in District A in 2011 and 2012, in District B in 2013, and in District C in 2014, then $d_{i0} = B$.
For the validation sample, based on data from 2010, we obtain teachers' d_{i0} following the same procedure as above, using a teacher's employment history between 2007 and 2010.

B1.3 Teacher Effectiveness

Students were tested on math and language in the Wisconsin Knowledge and Concepts Examination (WKCE, 2007-2014) and Badger test (2015-2016); we focus on their math scores. The WKCE was administered in November of each school year, whereas the Badger test was administered in March. To account for this change, for the years 2007–2014 we assign each student a score equal to the average of the standardized scores for the current and the following year. The test score data also include individual characteristics of test takers, such as gender, race and ethnicity, socioeconomic (SES) status, migration status, English-learner status, and disability status.

Our data allow us to link students and teachers up to the school-grade level, rather than the classroom level. To account for this data structure, we estimate two student achievement models and derive teacher effectiveness measures from each of them. In the following, we first describe the achievement model used in our empirical analysis, and its estimation and identification. The distribution of effectiveness measures estimated with this achievement model is summarized in Tables B1 and B2 and Figures B1 and B2. Next, we describe the alternative model, and its estimation and identification. Finally, we show that the effectiveness measures we obtain from both models are strongly correlated and that our auxiliary models used in our structural estimation are robust to the choice of effectiveness measures.

B1.3.1 Achievement Model 1 (Main)

The effectiveness measures used in our empirical analysis are estimated using the following achievement model:

$$A_{kt} = \gamma Z_{kt}^{s} + \sum_{i:SG_{kt}=SG_{it}^{T}} \sum_{n=1}^{2} I\left(\tau_k = n\right)\left(\rho_n x_{it} + v_{in}\right) + \varepsilon_{kt} \tag{1}$$

$$= \gamma Z_{kt}^{s} + \sum_{i:SG_{kt}=SG_{it}^{T}} \sum_{n=1}^{2} I\left(\tau_k = n\right)\rho_n x_{it} + \varphi_{kt} \tag{2}$$

where A_{kt} is achievement (measured as the standardized Math test score) of student k in year t. The vector Z_{kt}^{s} contains the following: a cubic polynomial of previous year's test scores, interacted with grade fixed effects; a cubic polynomial of previous year's average test scores for k's cohort in the school, interacted with grade fixed effects; a set of student characteristics, including gender, race and ethnicity, disability status, English-language status, and socioeconomic status; the same average characteristics for student k's cohort; cohort size; grade-by-school fixed effects; and year fixed effects. The variable ε_{kt} is an i.i.d. unobservable component of achievement, idiosyncratic to each student and year. SG_{kt} (SG_{it}^{T}) denotes the school-grade of student k (teacher i) in year t. The variable τ_k equals 1 for low-achieving students and 2 for high-achieving ones; we consider a student to be low-achieving if their test score in the previous year was below the grade-specific median in the state, and high-achieving otherwise. The contribution of teacher i to the achievement of a student of type $n \in \{1, 2\}$ is $\rho_n x_{it} + v_{in}$, where x_{it} denotes i's education and experience in year t and v_{in} is the part unexplained by x_{it}.

The achievement model in (1) assumes that all teachers in a given school-grade contribute to the achievement of all students in the same school-grade. We make this choice to be able to allow x_{it} to directly enter teacher effectiveness (since experience has been shown to affect teacher effectiveness (Wiswall 2013), especially in the first years of a teacher's career (Rockoff 2004)), even if we do not observe all the teacher-student classroom links in the data. Model (1) allows us to identify the component of teacher effectiveness that depends on a teacher's experience and education.

Constructing our measures of effectiveness (c_{i1}, c_{i2}) requires estimating v_{in} and ρ_n for $n \in \{1, 2\}$. We make the following two assumptions:

A1. ε_{kt} is i.i.d. with mean 0 and variance σ_{ε}^{2}.

A2. $Cov(\varepsilon_{kt}, v_{in}) = 0 \; \forall k, i, t, n : SG_{it}^{T} = SG_{kt}$. This implies that there is no sorting on unobservables of teachers across school-grades within a district. Although there is no direct test of this assumption, in Section B3.3 we combine the approaches of Chetty et al. (2014) and Rothstein (2010) and we do not find evidence of non-random sorting.

Estimation Procedure: Model 1

1. Given A1 and A2, we estimate γ and ρ_n via OLS on equation (1), to obtain $\hat{\gamma}$ and $\hat{\rho}_n$.

2. With the estimated $\hat{\gamma}$ and $\hat{\rho}_n$, we can then estimate v_{in} using an empirical Bayes estimator similar to the one of Kane and Staiger (2008) which we adapt to take into account the structure of our data.

   (a) Let

$$\widehat{\varphi}_{kt} = A_{kt} - \hat{\gamma} Z^s_{kt} - \sum_{i:SG_{kt}=SG^T_{it}} \sum_{n=1}^{2} \hat{\rho}_n x_{it} I(\tau_k = n). \tag{3}$$

   The quantity $\widehat{\varphi}_{kt}$ is an estimate for φ_{kt}, i.e.,

$$\varphi_{kt} \equiv \sum_{i':SG_{kt}=SG^T_{i't}} \sum_{n=1}^{2} v_{i'n} I(\tau_k = n) + \varepsilon_{kt}.$$

   Let $K_{SG^T_{it}n}$ be the number of achievement type-n students in the school-grade that i belongs to. For each teacher i we define, for $n \in \{1, 2\}$

$$\widehat{v}_{int} = \frac{1}{K_{SG^T_{it}n}} \sum_{k:SG_{kt}=SG^T_{it}} \widehat{\varphi}_{kt} I\left(\tau_k = n\right) \tag{4}$$

   which is an estimate of

$$\sum_{i':SG^T_{i't}=SG^T_{it}} v_{i'n} + \frac{1}{K_{SG^T_{it}n}} \sum_{k:SG_{kt}=SG^T_{it}} \varepsilon_{kt}.$$

   This quantity corresponds to the average test score residuals of type-n students in teacher i's school-grade in year t, conditional on observables Z^s_{kt} and the characteristics x of all teachers in the same school-grade in t.

   (b) We form a weighted average of the residuals $\{\widehat{v}_{int}\}_t$ by weighting each $\widehat{v}_{int}$ by $\varpi_{int} = \frac{K_{SG^T_{it}n}}{\sum_t K_{SG^T_{it}n}}$, so that residuals corresponding to more observations receive more weight:

$$\bar{v}_{in} = \sum_t \varpi_{int} \widehat{v}_{int} \tag{5}$$

   Note that assumption A1 implies

$$E(\bar{v}_{in}) = v_{in} + \sum_t \varpi_{int} \sum_{i' \neq i:SG^T_{i't}=SG^T_{it}} v_{i'n}$$

Taking the limit of this expectation as t approaches infinity yields

$$\lim_{t\to\infty} E(\bar{v}_{in}) = v_{in} + \lim_{t\to\infty} \sum_t \varpi_{int} \sum_{i'\neq i:SG^T_{i't}=SG^T_{it}} v_{i'n}$$

It follows that a requirement for the estimator $\bar{v}'_{in}$ to be asymptotically unbiased is that $\lim_{t\to\infty} \sum_t \varpi_{int} \sum_{i'\neq i:SG^T_{i't}=SG^T_{it}} v_{i'n} = 0$. In words, the weighted sum of the effects of all teachers in i's school-grade over time has to approach 0 as the number of periods grows large. This requirement is met because 1) the teacher effect v_{in} is defined as a residual component of standardized test scores conditioning on grade-by-school fixed effects (which implies that, across time, the mean of v_{in} is zero within each school-grade) and 2) Assumption A2 guarantees that there is no sorting of teachers on unobservables across school-grades over time.

(c) Armed with $\bar{v}_{in}$, we can construct the empirical Bayes estimator of v_{in} by multiplying $\bar{v}_{in}$ by the shrinkage factor, a measure of the reliability of the estimator defined as the ratio between the estimated variance of the quantity to be estimated, $\hat{\sigma}_{vn} = Var(v_{in})$, and the variance of the estimator:

$$\hat{v}_{in} = \bar{v}_{in} \left(\frac{\widehat{\sigma}^2_{vn}}{Var(\bar{v}_{in})} \right),$$

where, given assumptions A1 and A2, we can estimate $\hat{\sigma}^2_{vn}$ as

$$\widehat{\sigma}^2_{vn} = \frac{Cov(\widehat{v}_{int}, \widehat{v}_{int-1})}{J_{SG^T_{it,t-1}}}$$

and $J_{SG^T_{it,t-1}} = \sum_{i'} I(SG^T_{i't} = SG^T_{it}) I(SG^T_{i't-1} = SG^T_{it-1})$ is the number of teachers who are in the same school-grade as i in both t and $t-1$.

Identification: Model 1 The identification of teacher effects v_{in} leverages teacher turnover across school-grades over time. Our identifying assumption is that turnover of teachers across school-grades, within a district, is unrelated to v_{in} (Assumption A2). Importantly, this assumption allows for the endogenous sorting of teachers across districts based on v_{i1} and v_{i2}, as is the case in our model. In the estimation of v_{in}, this type of sorting is accounted for by the school-grade fixed effects included in Z^s_{kt}.

Teacher turnover across school-grades allows us to identify v_{in} from $\bar{v}_{in}$ for all i and n. In particular, we can stack all the equations (5) for all I teachers and $n = 1, 2$, forming a system of $2I$ equations (where I is the total number of teachers) in $2I$ unknowns ($\{v_{in}\}_{i,n\in\{1,2\}}$).

Identification is achieved if the rank condition of the system is satisfied, i.e., if the coefficient matrix of the system is full-rank.

In practice, this requires that the set $\{i' : SG^T_{i't} = SG^T_{it} \forall t\}$ is empty for all i, which means that there are no two teachers who teach the same school-grade in all t. When this is the case, the system (and the v_{in} for all i and n) is perfectly identified. In our data, $\{i' : SG^T_{i't} = SG^T_{it} \forall t\}$ is empty for 75% of teachers, for whom we can precisely estimate (v_{i1}, v_{i2}). For the remaining 25% of teachers, $\{i' := SG^T_{i't} = SG^T_{it} \forall t\}$ is non-empty, and our estimated v_{in} is the average of $v_{i'n}$ for $i' : SG^T_{i't} = SG^T_{it} \forall t$.

B1.3.2 Achievement Model 2 (Alternative)

An alternative model would feature the assumption that each teacher contributes only to the achievement of the students in her classroom, while also assuming that teacher effectiveness is fixed over time. These assumptions have been used extensively in the value-added literature (e.g. Rockoff, 2004; Aaronson et al., 2007; Kane and Staiger, 2008).[2] The achievement model in this case would be:

$$A_{kt} = \gamma Z^s_{kt} + \sum_{n=1}^{2} I\left(\tau_k = n\right) v_{i(kt)n} + \varepsilon_{kt} \tag{6}$$

$$= \gamma Z^s_{kt} + \varphi_{kt} \tag{7}$$

where $i(kt)$ denotes student k's teacher in year t, i.e., k is in teacher i's classroom in year t. The contribution of teacher i to the achievement of a student of type $n \in \{1, 2\}$ is simply v_{in}. To estimate this quantity, we add the following assumption to A1 and A2:
A3. The variable $j_{int} = K_{int}/K_{SG^T_{it}n}$ is i.i.d. with mean $1/J_{SG^T_{it}n}$, where K_{int} is the number of students of type n in the classroom of teacher i in year t and $J_{SG^T_{it}}$ is the number of teachers in school-grade SG^T_{it} in t. Furthermore, $Cov(j_{int}, v_{i'n}) = 0\ \forall i, i', t$. That is, class size is unrelated to teacher effectiveness within each school-grade.

Estimation: Model 2 With A1-A3, we can adapt the estimation procedure as follows.

1. We estimate γ via OLS on equation (6) to obtain $\widehat{\gamma}$.

2. We construct

$$\widehat{\varphi}'_{kt} = A_{kt} - \hat{\gamma} Z^s_{kt} \tag{8}$$

[2] Besides assuming that teacher effectiveness is fixed over time, these studies assume that teacher effectiveness is one-dimensional, rather than student-type-specific.

which is an estimate for $\sum_{n=1}^{2} v_{i(kt)n} I(\tau_k = n) + \varepsilon_{kt}$. For each teacher i, we define, for $n \in \{1, 2\}$

$$\widehat{v}'_{int} = \frac{1}{K_{SG_{it}^T n}} \sum_{k: SG_{kt} = SG_{it}^T} \widehat{\varphi}'_{kt} I\left(\tau_k = n\right) \tag{9}$$

which is an estimate of $$\sum_{i': SG_{it}^T = SG_{i't}^T} j_{i'nt} v_{i'n} + \frac{1}{K_{SG_{it}^T n}} \sum_{k: SG_{kt} = SG_{it}^T} \varepsilon_{kt} \tag{10}$$

3. We form a weighted average of $\{\widehat{v}'_{int}\}_t$, with the same weights ϖ_{int} as before:

$$\bar{v}'_{in} = \sum_t \varpi_{int} \widehat{v}'_{int}$$

Assumption A1. implies

$$E(\bar{v}'_{in}) = v_{in} \sum_t \frac{\varpi_{int}}{J_{SG_{it}^T}} + \sum_t \frac{\varpi_{int}}{J_{SG_{it}^T}} \sum_{i': SG_{it}^T = SG_{i't}^T} v_{i'n}$$

Taking the limit of this expectation as t approaches infinity implies

$$\lim_{t \to \infty} E(\bar{v}'_{in}) = v_{in} \sum_t \frac{\varpi_{int}}{J_{SG_{it}^T}} + \lim_{t \to \infty} \sum_t \frac{\varpi_{int}}{J_{SG_{it}^T}} \sum_{i': SG_{it}^T = SG_{i't}^T} v_{i'n}$$

It follows that the estimator

$$\bar{\bar{v}}'_{in} = \frac{1}{\sum_t \frac{\varpi_{int}}{J_{SG_{it}^T}}} \bar{v}'_{in} \tag{11}$$

is asymptotically unbiased if $\lim_{t \to \infty} \sum_t \frac{\varpi_{int}}{J_{SG_{it}^T}} \sum_{i': SG_{it}^T = SG_{i't}^T} v_{i'n} = 0$. As before, this requirement implies that the weighted average of the effects of all teachers in i's school-grade over time has to approach 0 as the number of periods grows large. Assumption A2 and the fact that we are conditioning on school-grade fixed effects guarantees that this is the case asymptotically.

4. Finally, we construct the empirical Bayes estimator for v_{in} as

$$\hat{v}'_{in} = \bar{\bar{v}}'_{in} \left(\frac{\widehat{\sigma}^{2\prime}_{vn}}{Var(\bar{\bar{v}}'_{in})} \right)$$

and we can estimate the variance of v_{in}, $\widehat{\sigma}^{2\prime}_{vn}$, as

$$\widehat{\sigma}^{2\prime}_{vn} = J_{SG^T_{it}} J_{SG^T_{it-1}} \frac{Cov(\widehat{v}'_{int}, \widehat{v}'_{int-1})}{J_{SG^T_{it,t-1}}}$$

Identification: Model 2 The identification of this alternative model also relies on within-district school-grade turnover as in Model 1. Equation (11) represents a system of $2I$ equations (where I is the total number of teachers) in $2I$ unknowns, where the unknowns are $\{v_{in}\}_{i,n\in\{1,2\}}$. Teacher effectiveness v_{in} is perfectly identified for teachers for whom there are at least two periods t and t' with $SG^T_{it} \neq SG^T_{it'}$.

B1.3.3 Teacher Effectiveness: Model 1 vs Model 2

Correlation of Teacher Effectiveness Measures Table B3 displays the correlations between (c_{i1}, c_{i2}), the measures of teacher effectiveness we use in our preferred model (Model 1), and $(\hat{v}'_{i1}, \hat{v}'_{i2})$, estimates of teacher effectiveness obtained with the alternative model (Model 2). We report these for both the estimation sample (2014) and the validation sample (2010). Teacher effectiveness measures estimated from the two models are highly correlated.

Inferred Offer Sets As discussed in the identification section of the paper, an important step of our estimation is to infer subsets of the offers received by each teacher from the observed teacher-district matches (we denote these as O^s_i). To show that the model estimates are robust to using $(\hat{v}'_{i1}, \hat{v}'_{i2})$ in place of (c_{i1}, c_{i2}), we re-constructed the inferred offer (sub)sets using $(\hat{v}'_{i1}, \hat{v}'_{i2})$, denoted by $\widetilde{O}^s_i$. Comparing O^s_i with $\widetilde{O}^s_i$ for each of the 6,600 teachers in our estimation sample, we find that 1) $O^s_i = \widetilde{O}^s_i$ for 27% of teachers, 2) $O^s_i \supset \widetilde{O}^s_i$ for 23% of teachers, 3) $O^s_i \subset \widetilde{O}^s_i$ for 21% of teachers, and 4) for the rest 28% of teachers, there are some districts in O^s_i but not in $\widetilde{O}^s_i$ and some districts in $\widetilde{O}^s_i$ but not in O^s_i. For the robustness of teacher preferences under O^s_i in place of $\widetilde{O}^s_i$, case 1) is ideal, and cases 2) and 3) are not concerning, because we only need subsets of offers to infer teacher preferences (Fox, 2007). These three cases account for 72% of teachers.

Auxiliary Models A key source of identification comes from our auxiliary models Aux 1a and Aux 1b that characterize teacher-district matches via regressions,

$$y_{id} = \beta^m_1 w_{id} + I\begin{pmatrix} d_{0i} > 0, \\ d \neq d_{0i} \end{pmatrix}\begin{bmatrix} \beta^m_2(x_{i1}) + \beta^m_3 \ln(dist_{id}) \\ +\beta^m_4 I(z_d \neq z_{d_{0i}}) \end{bmatrix} + q_d\beta^m_4 + \beta^m_5 e^{\lambda_d} + \beta^m_6 c_{1i}\lambda_d + \psi_i + \varepsilon^m_{id},$$

In Aux 1a, i's are all the teachers whose inferred subsets of offers O_i^s contain more than one district, and an observation (i, d) is a teacher-district pair in these inferred subsets. In Aux 1b, an observation is any teacher-district pair, with $I \times D$ total observations.

In Table B4, we compare Aux 1a and Aux 1b when a teacher is characterized by (x, c) (Model 1) against their counterparts when a teacher is characterized by $(x, \hat{v}')$ (Model 2). Between the two cases, regression coefficients in Aux 1a are very similar, and those in Aux 1b are almost identical.

Precision of Teacher Effectiveness Estimates We compare the precision of our estimates with that in previous studies. In particular, it is useful to compare the signal-to-noise ratio of our measure with that reported by other papers that estimate teacher value-added using data with classroom links. We perform this comparison in Table B5. Since all previous papers use one-dimensional value-added (VA) measures, we begin by comparing a one-dimensional measure of VA constructed with our data. Row 1 compares the signal-to-noise ratio of the estimated one-dimensional VA in our data and those found in previous papers. The precision of our estimate is comparable to that in previous studies; we believe this assuages concerns about the noise in our estimates due to the absence of classroom linkages and classroom effects from our model.

For completeness, in rows 2 and 3 we also report the signal-to-noise ratios of $c_1$and c_2, the effectiveness measures used in our model. These are 0.55 and 0.61, respectively. Since previous papers do not estimate multi-dimensional VA, we do not have a benchmark for these metrics. However, we believe these values to be reasonably smaller than that in row 1, since they involve estimating two effectiveness measures per teacher using the same data. Together with the estimate in row 1, they suggest that our estimators perform reasonably well.

B1.3.4 Teacher Effectiveness: Two-Dimensional vs One-Dimensional

To check whether allowing teacher effectiveness to vary by student type provides gains in terms of explaining the overall variation in test scores, we estimate a counterpart of Model (1) with one-dimensional rather than two-dimensional teacher effectiveness and compare it with Model (1). Table B6 compares the average sum of squared test score residuals $\hat{\varphi}_{kt}$, by student type, obtained from each model. Our two-dimensional teacher effectiveness model explains approximately 20% more variation in test scores compared to its one-dimensional effectiveness counterpart.

B1.3.5 Teacher Effectiveness: Race

Previous studies suggest that the match between the teacher's race and the student's race can matter for achievement. In comparison, we focus on teachers' comparative advantages in teaching students with different prior achievement types. We make this choice for two reasons. First, as shown in Table B7, if we add teacher race and the interaction of teacher and student race to our achievement model (student race is already included in our achievement model), almost none of the added terms are significant. Second, if we add a teacher's race and gender and their interactions with the district's racial and gender composition of students to our Aux 1a (Column 1 of Table 2 in the main text), the R^2 is barely improved (from 0.68 to 0.681).

B1.4 Wage Schedules

B1.4.1 Pre-Reform Wage Schedules

We obtain $W_d^0(x_i)$ as the predicted values from the following regression, estimated using data from 2007 to 2011:

$$w_{it}^0 = \delta_d^0 + Exp_{it}\delta_{g(i)}^e + MA_{it}\delta_{g(i)}^m + \varepsilon_{it}, \tag{12}$$

where i and t refer to teacher and year, respectively; w_{it}^0 is the wage of teacher i in year t; Exp_{it} is a vector of indicators for six classes of years of experience: 0, $[1,2]$, $[3,4]$, $[5,9]$, $[10,14]$, and $[15,+\infty)$; and MA_{it} is an indicator for having a Master's degree (MA) or a higher degree. The parameter δ_0 can be interpreted as the average wages for teachers with zero experience and without a MA; with $\delta_{g(i)}^e$ normalized to 0 for those with zero experience, $\delta_{g(i)}^e$ is the average wage premium for teachers in each of the higher experience category, relative to those with zero experience with the same education; and δ^m is the wage premium for teachers who have a MA.

We estimate the intercept δ_d^0 separately for each district. Trading off the accuracy of our wage schedules with power, we estimate the coefficients δ^e and δ^m by groups of districts, defined as follows:

1. For the 35 large districts (i.e., those with at least 10 teachers in each experience and education category), each group corresponds to a district.

2. For the remaining 356 districts, we construct groups based on the similarity in their salary schedules. To do so, we proceed as follows.

(a) For each district, we calculate the following summary statistics for their salary schedules: (i) wages for teachers with 0 years of experience and $MA_{it} = 0$ (i.e., the lowest possible wage category); (ii) wages for teachers with over 15 years of experience and $MA_{it} = 0$ (i.e., the highest possible wage category for those without MA); (iii) average salary difference between a teacher with more than 15 years of experience and a MA, and one with the same experience and no MA.

(b) We check whether each district is above or below the median of the cross-districts distribution for each of the three statistics.

(c) We form eight groups based on the statistics (i), and (ii), and (iii), and assign each district to a group as follows:

Group	(i)	(ii)	(iii)
1	≥median	≥median	≥median
2	≥median	≥median	<median
3	≥median	<median	≥median
4	<median	≥median	≥median
5	<median	<median	≥median
6	<median	≥median	<median
7	≥median	<median	<median
8	<median	<median	<median

Table B8 summaries the point estimates from Equation (12). In particular, it reports the cross-district means and standard deviations of the estimated vectors δ. Figure B1 shows a binned scatter plot of $W_d^0(x_i)$ and data wage w_{it}^0 in 2010. The former predicts the latter remarkably well, with a correlation coefficient of 0.93 (significant at 1 percent).

B1.4.2 Districts' Choice Set of Wage Schedules

A district's wage rule is given by

$$w_d(x, c|\omega) = \max\left\{\min\left\{\omega_1 W_d^0(x) + \omega_2(\lambda_d c_1 + (1-\lambda_d)c_2), \overline{w}\right\}, \underline{w}\right\}. \tag{13}$$

A district chooses (ω_1, ω_2) from a discrete set Ω, the grid points of which are chosen as follows.

1. We start by estimating the parameters $(\widetilde{\omega}_{d1}, \widetilde{\omega}_{d2}) \geq 0$ separately for each district from

$$w_i = \widetilde{\omega}_{d1} W_d^0(x_i) + \widetilde{\omega}_{d2} TC(c_i, \lambda_d) + \varepsilon_i^w, \text{ for } i : d(i) = d$$

where w_i is the observed 2014 wage for teacher i working in district d $(i : d(i) = d)$, $W_d^0(x_i)$ is defined as in Section B1.4.1, and teacher contribution $TC(c_i, \lambda_d)$ is given by

$$TC(c_i, \lambda_d) = \lambda_d c_{i1} + (1 - \lambda_d) c_{i2}.$$

2. Based on the estimated $\{(\widetilde{\omega}_{d1}, \widetilde{\omega}_{d2})\}_d$, we choose a set of equally spaced grid points that provides a good coverage of the empirical distribution in the data:

$$\Omega^o = \{0.9, 0.95, 1, 1.05, 1.1\} \times \{0, 10, 30, 50, 75, 100, 200\}.$$

3. We assign each district the wage schedule $(\omega_{d1}^o, \omega_{d2}^o) \in \Omega^o$ that best summarizes the distribution of teacher wages in that district $\{i : d(i) = d\}$, i.e.,

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

$$s.t.\ w_d(x_i, c_i; \omega) = \begin{cases} \underline{w} \text{ if } \omega_1 W_d^0(x_i) + \omega_2 TC(c_i, \lambda_d) < \underline{w} \\ \overline{w} \text{ if } \omega_1 W_d^0(x_i) + \omega_2 TC(c_i, \lambda_d) > \overline{w} \\ \omega_1 W_d^0(x_i) + \omega_2 TC(c_i, \lambda_d) \text{ otherwise} \end{cases},$$

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

- The $(\omega_{d1}^o, \omega_{d2}^o)$ selected with this procedure predicts teachers' actual salaries quite well: 1) the absolute percentage deviation of predicted wages from actual wages in 2014, i.e., $\left|1 - \frac{w_d(x_i, c_i; \omega)}{w_i}\right|$, is less than 10% for 95% of teachers in our sample; and 2) regressing w_i on $w_d(x_i, c_i; \omega)$ yields a slope coefficient of 0.98 (with a standard error of 0.001) and an R^2 of 0.99.

4. Finally, we expand the grid range to allow for the possibility that district choices may go out of the empirical range in counterfactual scenarios. The choice set in the model is given by

$$\Omega = \{0.9, 0.95, 1, 1.05, 1.1, 1.15\} \times \{0, 10, 30, 50, 75, 100, 200, 225\}.$$

where both $\omega_1 = 1.15$ and $\omega_2 = 225$ are outside of Ω^o.

B1.4.3 Alternative Wage Rules

Three ω's We have also tried to allow for a more flexible alternative wage schedule as follows

$$w_d(x, c|\omega) = \max\left\{\min\left\{\omega_1 W_d^0(x) + \omega_2 \lambda_d c_1 + \omega_3 (1-\lambda_d) c_2, \overline{w}\right\}, \underline{w}\right\}. \tag{14}$$

Wage rule (13) we use in the paper is a special case of (14) with $\omega_2 = \omega_3$. We repeat the exercise as in Section B1.4.2, but under the three-ω specification (14). This procedure yields the triplet $(\omega'_{d1}, \omega'_{d2}, \omega'_{d3})$ that best summarizes the observed distribution of teacher wages in each district d. Figure B4 compares the predicted wage under rule (13) and that under rule (14). The two predicted wages are nearly indistinguishable from each other, indicating the absence of large predictive gains associated with the use of (14) instead of (13).

Tenured vs untenured We have also tested for the possibility that the relationship between teacher contribution and wages depends on whether the teacher is tenured or not. To do so, we modify the wage function in the paper to be

$$w_d(x, c|\omega) = \max\left\{\min\left\{\omega_1 W_d^0(x) + \omega_{21} TC(c, \lambda_d) x_1 + \omega_{22} TC(c, \lambda_d)(1 - x_1), \overline{w}\right\}, \underline{w}\right\} \tag{15}$$

where $x_1 = 1$ if the teacher is untenured, and zero otherwise. We then estimated ω for each district and constructed each teacher's wage using this new schedule.

Figure B5 shows a scatter plot of wage residuals (i.e., the difference between actual wages and wages predicted using the wage schedule) obtained using the schedule originally contained in the paper (y-axis) and the alternative schedule shown above (x-axis). The relationship is close to a 45-degree line, indicating that the predicted wages are very similar using the two schedules.

Experience = 0 vs experience > 0 teachers Next, we test whether the relationship between teacher contribution and wages depends on experience. We do so by defining, in equation (15), $x_1 = 1$ for teachers with no experience. Figure B6 below shows a scatter plot of wage residuals obtained using the schedule originally contained in the paper (y-axis) and the alternative schedule shown above (x-axis). As before, the relationship is close to a 45-degree line, indicating that the predicted wages are very similar using the two schedules.

B2 Algorithms

Teachers' decision rule implies that if District d makes an offer to the teacher, the teacher's acceptance probability is given by

$$h_d(x,c,d_0)=\frac{\exp\left(\frac{V_d(x,c,d_0)}{\sigma\epsilon}\right)}{\exp\left(\frac{V_d(x,c,d_0)}{\sigma_\epsilon}\right)+\sum_{d'\in D\backslash d}o_{d'}(x,c,d_0)\exp\left(\frac{V_{d'}(x,c,d_0)}{\sigma\epsilon}\right)}. \tag{16}$$

We assume that districts make decisions based on a simplified belief, given by

$$\widetilde{h}_d\left(x,c,d_0|\overline{w}(x,c),\sigma_w(x,c)\right)=\frac{1}{1+\exp\left(f\left(x,c,d_0,w_d,q_d,\lambda_d\right)\right)}, \tag{17}$$

$$\text{with } f(\cdot)=x\zeta_1+\zeta_2\frac{c_1+c_2}{2}+\zeta_3\left(\frac{w_d-\overline{w}(x,c)}{\sigma_{w(x,c)}}\right)+\zeta_4q_d+\zeta_5e^{\lambda_d}+\zeta_6\lambda_dc_1$$
$$+(1-I(d_0=0))\left[I(d\neq d_0)(\zeta_7+\zeta_8x_1)+\zeta_9I(z_d\neq z_{d_0})\right],$$

where $\overline{w}(x,c)$ and $\sigma_w(x,c)$ are the mean and standard deviation of wages across all districts for a teacher with (x,c), i.e.,

$$\overline{w}(x,c)\equiv\frac{1}{D}\sum_d w_d(x,c;\omega_d) \tag{18}$$

$$\sigma_{w(x,c)}\equiv\sqrt{\frac{1}{D-1}\sum_d\left(w_d(x,c;\omega_d)-\overline{w}(x,c)\right)^2}. \tag{19}$$

An equilibrium requires beliefs $\widetilde{h}_d(x,c,d_0)$, and in particular the vector ζ and the wage statistics $\left\{\overline{w}(x,c),\sigma_{w(x,c)}\right\}_{x,c}$, to be consistent with decisions made by teachers and districts.

B2.1 Estimation Algorithm

The estimation algorithm involves an outer loop searching for the parameter vector Θ and an inner loop solving the model for each given Θ. This inner loop does not require finding the fixed point for all components in $\{\zeta,\overline{w}(\cdot),\sigma_w(\cdot)\}$: Assuming that data were generated from an equilibrium, $\{\overline{w}(\cdot)\}$ and $\{\sigma_w(\cdot)\}$ can be derived directly from the observed district wage schedules $\{\omega_d^o\}_d$, where the superscript o denotes "observed." For estimation, one only needs to find the fixed point for ζ; the observed equilibrium wage statistics $\{\overline{w}^o(\cdot),\sigma_w^o(\cdot)\}$ can be plugged directly into the belief function (17). Given a parameter vector Θ, the inner loop of the estimation algorithm involves the following steps.

1. Search for $\zeta^*(\Theta)$

(a) Guess ζ, which, together with $\overline{w}^o(\cdot)$ and $\sigma_w^o(\cdot)$, implies a belief $\left\{\widetilde{h}_d\left(\cdot|\zeta,\overline{w}^o(\cdot),\sigma_w^o(\cdot)\right)\right\}$ as defined in (17).

(b) Given $\widetilde{h}_d\left(\cdot|\zeta,\overline{w}^o(\cdot),\sigma_w^o(\cdot)\right)$, solve for the optimal job offers $o_d^*\left(\cdot;\omega_d^o\right)$ under the observed ω_d^o for each district d.

(c) Given the job offers and the wages implied by $\left\{o_d^*\left(\cdot;\omega_d^o\right),\omega_d^o\right\}_d$, calculate each teacher's acceptance probabilities $h_d(\cdot)$ for each d, as in (16), and the distance $\left\|h(\cdot)-\widetilde{h}\left(\cdot|\zeta,\overline{w}^o(\cdot),\sigma_w^o(\cdot)\right)\right\|$.

(d) Repeat Steps 1a-1c until $\left\|h(\cdot)-\widetilde{h}\left(\cdot|\zeta,\overline{w}^o(\cdot),\sigma_w^o(\cdot)\right)\right\|$ is below a tolerance level; the associated ζ is the consistent belief parameter vector $\zeta^*(\Theta)$.

2. Given job offers $\left\{o_d^*\left(\cdot;\omega_d^o\right)\right\}_d$ under $\widetilde{h}_d\left(\cdot|\zeta^*(\Theta),\overline{w}^o(\cdot),\sigma_w^o(\cdot)\right)$ and wages implied by $\left\{\omega_d^o\right\}$, each teacher chooses the most preferred among their received offers. The implied teacher-district matches will be compared with the observed matches in the outer loop.

3. Given $\widetilde{h}_d\left(\cdot|\zeta^*(\Theta),\overline{w}^o(\cdot),\sigma_w^o(\cdot)\right)$, each district makes optimal decisions on its wage schedule $\omega_d^*(\Theta)$.[3] The resulting $\left\{\omega_d^*(\Theta)\right\}_d$ will be compared with the observed $\left\{\omega_d^o\right\}_d$ in the outer loop.

B2.2 Solving for the Equilibrium

Both the teacher-specific wage statistics $\left\{\left(\overline{w}(x,c),\sigma_{w(x,c)}\right)\right\}_{x,c}$ and the wage rules $\left\{\left(\omega_{d1},\omega_{d2}\right)\right\}_d$ that govern these statistics are high-dimensional objects. However, notice that districts' wages are given by

$$w_d(x,c;\omega)=\begin{cases}\underline{w} \text{ if } \omega_1 W_d^0(x)+\omega_2\left[\lambda_d c_1+\left(1-\lambda_d\right)c_2\right]<\underline{w}\\ \overline{w} \text{ if } \omega_1 W_d^0(x)+\omega_2\left[\lambda_d c_1+\left(1-\lambda_d\right)c_2\right]>\overline{w}\\ \omega_1 W_d^0(x)+\omega_2\left[\lambda_d c_1+\left(1-\lambda_d\right)c_2\right] \text{ otherwise}\end{cases}, \tag{20}$$

where the pre-reform wage schedule $W_d^0(x)$ is a linear function of experience categories (x_1) and the MA dummy (x_2). It follows that the mean wage is a linear function of the following form governed by some parameter vector θ^1

$$\widetilde{\overline{w}}(x,c)=\begin{cases}\underline{w} \text{ if } \sum_n \theta_{1n}^1 I\left(x_1=n\right)+\theta_2^1 x_2+\theta_3^1 c_1+\theta_4^1 c_2<\underline{w}\\ \overline{w} \text{ if } \sum_n \theta_{1n}^1 I\left(x_1=n\right)+\theta_2^1 x_2+\theta_3^1 c_1+\theta_4^1 c_2>\overline{w}\\ \sum_n \theta_{1n}^1 I\left(x_1=n\right)+\theta_2^1 x_2+\theta_3^1 c_1+\theta_4^1 c_2 \text{ otherwise.}\end{cases} \tag{21}$$

[3] We assume that changing a single district's wage for Teacher i has a negligible effect on wage statistics $\left(\overline{w}^o\left(x_i,c_i\right),\sigma_w^o\left(x_i,c_i\right)\right)$, i.e., the mean and standard deviation of Teacher i's wage across the 411 districts in our sample.

Similarly, the cross-district wage standard deviation for a teacher will be the square root of a quadratic function (Q), governed by some parameter vector θ^2, and bounded from above by the largest possible wage spread, i.e.,

$$\widetilde{\sigma}_{w(x,c)} = \min\left\{\sqrt{\max\left\{Q\left(x_1, x_2, c_1, c_2; \theta^2\right), 0\right\}}, \overline{w} - \underline{w}\right\}. \tag{22}$$

Instead of searching for fixed points of $\left\{\{h_d(x, c, d_0)\}_{x,c}, (\omega_{d1}, \omega_{d2})\right\}_d$, one can search for parameter vectors ζ, θ^1, and θ^2 in (17), (21) and (22) to guarantee equilibrium consistency. Note that ζ, θ^1, and θ^2 are not structural parameters; rather, they serve to summarize the equilibrium under a given policy scenario and are policy dependent. We now describe the algorithm we use to simulate the equilibrium outcomes, for a given policy environment.

B2.2.1 Equilibrium Algorithm

We draw M economies, each with D districts and N teachers. All economies share the same observable teacher and district characteristics as those in the data, but each economy is assigned a different realization of wage-choice-specific shocks $\{\{\eta_{d\omega}\}_{\omega}\}_d$, drawn from the i.i.d. extreme value distribution, with the scaling parameter σ_η. The expected equilibrium outcomes are calculated as the average outcomes across M economies. For each economy m, we apply the following procedure.

1. Guess parameters ζ, θ^1, and θ^2, which imply $\left\{\widetilde{\overline{w}}(x, c), \widetilde{\sigma}_{w(x,c)}, \widetilde{h}_d\left(x, c, d_0 | \widetilde{\overline{w}}(x, c), \widetilde{\sigma}_{w(x,c)}\right)\right\}$ from (17), (21) and (22).

2. Given $\left\{\widetilde{h}_d\left(x, c, d_0 | \widetilde{\overline{w}}(x, c), \widetilde{\sigma}_{w(x,c)}\right)\right\}$, each district d chooses its optimal wage and offer policies $\{\omega_d, O(\omega_d)\}$.

3. Given $\{\omega_d, O(\omega_d)\}_d$, compute teacher acceptance probabilities $h_d(\cdot)$ from their decision rules (16), the mean wage $\overline{w}(x, c)$ based on (18), and standard deviation $\sigma_{w(x,c)}$ based on (19).

4. Calculate the distance between $\left\{\widetilde{\overline{w}}(x, c), \widetilde{\sigma}_{w(x,c)}, \widetilde{h}_d\left(x, c, d_0 | \widetilde{\overline{w}}(x, c), \widetilde{\sigma}_{w(x,c)}\right)\right\}$ and $\left\{\overline{w}(x, c), \sigma_{w(x,c)}, h_d\left(x, c, d_0 | \left(\overline{w}(x, c), \sigma_{w(x,c)}\right)\right)\right\}$.

5. Repeat Step 1 to Step 4 and search for $\{\zeta^*, \theta^{1*}, \theta^{2*}\}$ that bring the distance in Step 4 below a tolerance level. The vector $\{\zeta^*, \theta^{1*}, \theta^{2*}\}$ renders the consistent belief (17). Equilibrium outcomes in economy m consist of the decisions made by districts and teachers under this consistent belief.

B3 Across-District vs Within-District Variation

In our model we abstract from within-district competition for teachers, focusing on competition across districts. Here we show that cross-district variation clearly dominates within-district, cross-school variation in terms of both teacher wages and the share of low-achieving students.

B3.1 Wages

Table B9 shows the adjusted R^2 and the root mean-squared error (MSE) of a regression of post-Act 10 salaries on c_1, c_2, experience and education (first row). It then shows how the R^2 and MSE change as we sequentially add district fixed effects (second row) and school fixed effects (third row). Adding district fixed effects reduces the root MSE by 31.3%; this implies that differences across districts explain 31.3% of the residual variation in salaries, conditional on teacher characteristics. Adding school fixed effects instead only explains an additional 2.7% of the root MSE. We can conclude that the main source of variation in wages is across districts, not across schools within districts.

B3.2 Student Composition

The cross-district variation in the share of low-achieving students (λ in our model) largely dominates the within-district, cross-school variation. We provide evidence of this in three different ways.

1. Estimates from an OLS student-level regression of an indicator for being low-achieving, to which we progressively add district and school fixed effects, indicates that districts explain 8.7% of the variation in this probability whereas schools only explain an additional 2.7%.
2. The estimated R^2 of an OLS regression of the school-level share of low-achieving students on district fixed effects, weighted by enrollment, indicates that 74% of the variation in the school-level share is explained by the district.
3. For each school, we calculate the absolute difference between the school-level and the district-level shares of low-achieving students. This absolute difference has a mean of 0.05 and a standard deviation of 0.06. The 25th, 50th and 75th percentile of this absolute difference are 0.01, 0.03 and 0.07 respectively.

B3.3 Teacher Assignment Across School-Grades Within a District

The identification of c_1 and c_2 in our achievement model relies on the assumption of random sorting of teachers across school-grades within each district, conditional on all the covariates described in Appendix B1.3. To test for the presence of non-random sorting, in Table B10 we combine the approaches of Chetty et al. (2014) and Rothstein (2010). In columns 1 and 2 we follow Chetty et al. (2014) and estimate the slope of the relationship between changes in students' test score residuals (obtained from a regression of test scores on all the covariates in equation (13)) and changes in c_1 and c_2. As in Chetty et al. (2014), we control for school-by-grade and school-by-year fixed effects. These tests, shown in columns 1 and 2, reveal a slope coefficient that is statistically indistinguishable from one, indicating that our estimates of (c_1, c_2) are forecast unbiased for (c_1, c_2).

In columns 3 and 4 of Table B10 we combine the above empirical design with the test proposed by Rothstein (2010) and estimate the relationship between changes in (c_1, c_2) and changes in *lagged* test score residuals. If the estimates in this specification were significant, they would indicate non-random sorting of teachers across grade-schools. Reassuringly, the slope coefficients in columns 3 and 4 are smaller than those in columns 1 and 2 and statistically indistinguishable from zero.

B4 Robustness Checks: Identification Assumptions

In this section, we conduct two sets of robustness checks with respect to the two maintained assumptions underlying our identification strategy:
A1: (x, c) are observable to all districts. With our data, it is difficult to separate preferences from information friction; we abstract away from the latter.
A2: Districts cannot discriminate among teachers by factors other than (x, c).

As a partial test for the robustness of our results with respect to A1, we conduct the following exercise: Instead of (c_1, c_2), districts observe $(c_1 + err_1, c_2 + err_2)$ and make wage and job offer decisions based on these noisy measures. Assuming that $err_k \sim N\left(0, \sigma^2_{err_k}\right)$ are i.i.d. random noises and considering values of σ_{err_k} equal to one, two, or four times the standard deviation of c_k, for $k = 1, 2$, we repeat the procedure described in Section 4.1.2 of the main text to construct sub-offer sets using the observed matches. Column 1 of Table B11 reports the baseline estimates of Aux 1a, which are key for the identification of teachers' preferences. Columns 2-4 show estimates obtained assuming that both teachers' and districts' decisions are based on $(c_1 + err_1, c_2 + err_2)$, while the researcher observes (c_1, c_2). Columns 5-7 show the corresponding estimates assuming that districts' decisions are based on $(c_1 + err_1, c_2 + err_2)$, while teachers' decisions are based on (c_1, c_2). Throughout these exercises, the estimates of Aux 1a are robust.

To investigate robustness to a violation of A2, we consider the possibility that some ineffective teachers may have been hired for reasons other than (x, c). Table B12 compares our auxiliary model Aux 1a with its counterpart that does not use observed teacher-district (i, d) matches to infer offers for other teachers if i's effectiveness with either low- or high-achieving students is below the 10th percentile among all teachers. Doing so has a significant impact on the number of inferred offers for other teachers; yet Aux 1a remains robust.

It should be noted that although our robustness checks give some comfort that simple violations of A1 and A2 may not seriously affect our inference, they are no proof that these assumptions (maintained throughout) are innocuous.

B5 The Impact of Changes in Parameter Values on Auxiliary Models

Following Einav et al. (2018), we provide more evidence on the mapping between data and parameters via a perturbation exercise. We adjust each parameter one at a time and measure responses of the predicted auxiliary models we use for estimation.

To be specific, letting $\left\{\widehat{\theta}_n\right\}_{n=1}^{N}$ be the vector of estimated structural parameters and $\{\widehat{\sigma}_{\theta_n}\}_{n=1}^{N}$ be the vector of their standard errors, we re-simulate our model N times. In the n^{th} simulation, we use the parameter vector $\left\{\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_{n-1}, \widehat{\theta}_n + \widehat{\sigma}_{\theta_n}, \widehat{\theta}_{n+1}, ..., \widehat{\theta}_N\right\}$, where the n^{th} parameter is perturbed by its standard error, and obtain new estimates of the auxiliary models. We then compute the percent change in absolute terms for each auxiliary model (regression coefficient or moment). This exercise produces a matrix of dimension (number of auxiliary models $\times$ number of parameters). To ease exhibition, we take simple averages within sub-blocks of this matrix. Specifically, we split the auxiliary models into five groups as specified in the paper (Aux 1a, Aux 1b, Aux 2, Aux 3, and Aux 4) and split parameters into three groups (teacher preference parameters, district preference parameters, and wage-setting resistance cost parameters). This results in the 5 x 3 summary matrix shown in Table B13. Each cell in Table B13 shows the average percent change across auxiliary models and parameter permutations within a given sub-block.

Column 1 of Table B13 shows that teacher preference parameters primarily affect the sub-offer and all-offer regression models (Aux 1a and Aux1b), as well as the regression coefficients that link districts' wage choices to their pre-determined conditions (Aux 3). It is unsurprising that Aux 1a and Aux 1b are closely related to teachers' preferences, as these regressions are designed to mimic a conditional logit model of teachers' choices. Additionally, as teachers' preferences change, districts change their wage schedules in order to attract their preferred teachers; such responses are captured by changes in Aux 3.

Column 2 shows that district preference parameters mostly affect the regression coefficients that link wages to districts' pre-determined conditions (Aux 3) and the offer regression models (Aux 1a and Aux1b). As we argued in our identification section, Aux 3 should be informative of districts' preferences as districts can use wage choice to push or pull teachers; moreover, the *difference* between Aux 1a and 1b are also informative of districts' preferences.

Finally, Column 3 shows that the wage-setting resistance cost parameters affect the wage regressions and cross-district wage moments (Aux 3 and Aux 4). This is unsurprising as these two auxiliary models directly summarize wage choices. Notice that, by design, resistance cost parameters should have zero impact on Aux 1a, Aux 1b, and Aux 2, because these auxiliary models are obtained while holding wage schedules at the observed equilibrium levels.

B6 Counterfactual Experiments

Throughout the analyses in the main paper and the online appendix, we use the following efficiency-equity metrics: Letting $\Pr(i \text{ in } d|\Upsilon)$ be the equilibrium probability that teacher i works in district d in a given policy environment Υ, our metrics include:[4]

M1. Average total contribution among teachers working in a given group of districts $d \in D'$:

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

where $TC(c_i, \lambda_d) = c_{i1}\lambda_d + c_{i2}(1-\lambda_d)$ is teacher i's total contribution to students in d (if i works in d) and the numerator is the expected total contribution among teachers working in D'. The denominator is the expected total number of teachers working in D'.[5]

M2. Average teacher contribution to low-achieving students in the state

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

and to high-achieving students in the state

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

where $c_{i1}\lambda_d$ and $c_{i2}(1-\lambda_d)$ are teacher i's contributions to low- and high-achieving students in district d (if i works in d), respectively.

[4]Teacher-district matching is probabilistic because of shocks $\{\epsilon_d\}$ to teachers' choices, and shocks $\{\eta_\omega\}$ to districts' wage choices. For a given policy, we calculate the expected equilibrium outcomes by numerically integrating over $\{\eta_\omega\}$ and deriving teachers' choice probabilities analytically, as detailed in Online Appendix B1.2.1. We use the same set of random shocks throughout our analysis.

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

In the following, we examine the efficiency and equity of several alternative allocations of teachers to districts. First, we attempt to improve efficiency or equity by allocating teachers at will, regardless of their preferences. Second, we re-examine the effects of teacher bonus programs under additional assumptions about teachers' entry/exit decisions and the evolution of the market over multiple years.

B6.1 Allocating Teachers at Will

To gauge the potential gain in efficiency and that in equity if one can assign teachers at one's will (e.g., under a dictatorship), we conduct two exercises that we label as Dictator1 and Dictator2, respectively.

In Dictator1, our goal is to maximize efficiency, i.e., to increase the total TC in the market. To do so, we allocate teachers with the largest comparative advantage in teaching low-achieving students (measured as $(c_1 - c_2)$) to districts with the highest share of low achieving students. To implement this allocation, we sort districts by λ_d (fraction of low-achieving students) and teachers by $(c_1 - c_2)$. We first fill the highest-λ_d district with the highest-$(c_1 - c_2)$ teachers until the district's capacity is filled. We then move to the next district and fill its capacity with the highest-$(c_1 - c_2)$ teachers among those yet to be assigned. We proceed by filling the capacity of all districts according to this rule, in decreasing order according to λ_d.

In Dictator2, we aim at improving the performance of low-achieving students. To do so, we sort teachers based on their absolute advantage towards teaching low achieving students (c_1). We first fill the highest-λ_d district with the highest-c_1 teachers until the district's capacity if filled; then, we move to the next district to fill its capacity with the highest-c_1 teachers among those yet to be assigned. We proceed by filling the capacity of all districts according to this rule, in decreasing order according to λ_d.

Table B14 shows the results from these two exercises. Under Dictator1, total efficiency (TC) improves by 31.0%. These gains are unequally distributed: High-achieving students gain 43% and low-achieving students gain 19%. Under Dictator2, low-achieving students gain 70.5% and high-achieving student lose 55.8%. These estimates indicate that there exist reallocations of teachers that yield large increases in teacher contributions, which implies that the baseline market equilibrium under flexible pay leaves room for improvement.

B6.2 Policy Impacts: Entry/Exit Margin and Repeated Game

In our model, we model the teachers' labor market in a static equilibrium setting and we abstract from teachers' decisions to enter or exit the market. A rigorous analysis of policy impacts in the long run with teachers' extensive-margin responses is beyond the scope of our paper. To begin understanding how the impact of our policies could differ in the medium run, in this section we conduct further counterfactual policy simulations under additional assumptions about teachers' entry/exit decisions and about how the market evolves over multiple years.

B6.2.1 Setting 1: One-Shot Game

To incorporate teachers' entry/exit decisions into our framework, we extend our model to include a Stage 0, where teachers make these decisions. We refer to the potential entrants who decide to enter and the incumbents who decide to stay as participants to the market. After Stage 0, districts and participant teachers make decisions as specified in our model in the main text.

More specifically, our exercise involves the following assumptions and steps.

1. Baseline Entry Probability: Ideally, to estimate entry probabilities we would like to observe the choices of all potential entrants into the market. In the absence of such data, we assume that the distribution of the characteristics (x, c_1, c_2) of *potential* entrants is the same as the distribution of all Wisconsin teachers in our 2013-2015 pooled sample, including new teachers but excluding those beyond age 62 (a standard retirement age; our results are robust to this choice). This assumption essentially treats the Wisconsin teachers' labor market as being surrounded by a larger market of teachers and implies that teachers who work in Wisconsin are a representative sample of all teachers (i.e., potential entrants) in this larger market. Of course, the distribution of entrants will be different from the distribution of potential teachers. We model the entry probability as follows: Using this potential pool of entrants, we estimate a logistic model to predict the baseline probabilities for all potential entrants, given their observable characteristics. The explanatory variables of the model include age group fixed effects; experience group fixed effects; an indicator for holding a Master's degree; the interaction of these three sets of fixed effects; teacher contributions c_1, c_2; the interaction between contributions and experience-education fixed effects; and year fixed effects.

2. Baseline Exit Probability: Using a similar logistic model with the same explanatory variables, we also estimate each teacher's exit probability in the baseline equilibrium. To improve the precision of our estimates, we pool teachers in the 2013-2015 sample and control for year fixed effects. Consistent with the Stage-0 framework, here we aim at estimating baseline probabilities of voluntary exits rather than layoffs (layoffs almost exclusively affect untenured teachers). To achieve this, we exclude untenured

teachers (i.e., those with fewer than 3 years of experience) from the estimation sample; we then extrapolate untenured teachers' voluntary exit probability using the estimated coefficients for teachers with 3-4 years of experience.[6]

3. Given 1 and 2, we can simulate the baseline equilibrium with entry and exit. In Stage 0, potential entrants decide whether to enter and incumbents decide whether to exit. Then, districts and participating teachers make decisions as specified in our main model.

4. We then additionally simulate the model under a given counterfactual bonus program as follows.

   (a) Using the baseline equilibrium, we calculate each teacher's expected pay under the baseline and under a given state bonus program. This gives us an (expected) percentage change in a teacher's pay.

   (b) Assuming a given entry/exit elasticity with respect to pay that is consistent with the literature, we can calculate the change in a teacher's entry/exit probability in response to the pay change calculated in 4 (a). This further allows us to construct the new pool of market participants in the counterfactual. In our application, we follow Rothstein (2015) and use three elasticities: 0.5, 1.0, and 1.5.

   (c) Given this new pool of participants, we simulate the equilibrium interaction between districts and participating teachers as specified in our main model.

Given the additional assumptions we have made (e.g., entry/exit elasticities), results from this exercise allow us to assess the impact of our counterfactual policies when we take exit/entry into account in a static setting.

B6.2.2 Setting 2: Repeated Static Game

Lastly, we attempt to understand how our bonus programs may affect the market after having been in effect for T years. To do so, we assume that teachers and districts play the static game (as described in Section B6.2.1) repeatedly over T years.[7] For each year t, we assume that the distribution of potential entrants remains the same; for each participating teacher, we update their age and experience from year to year and calculate their exit probability accordingly. To simplify the simulation exercise, we assume that the state uses

[6] Observed exit rates are very similar across experience levels from 0 to 4, suggesting that this extrapolation may not cause large biases.

[7] In particular, when simulating entry/exit decisions, we assume that an individual teacher calculates their expected pay (wage plus bonus) in t based on the equilibrium wage schedules in $t-1$.

the baseline district characteristics, including λ_d, to calculate teacher bonuses. As a caveat, this assumption is reasonable when we consider shorter T, but is not well suited for very long T.[8] Therefore, we consider a relatively short $T = 5$ in our following simulation.

B6.2.3 Results: State-Funded Teacher Bonus Programs

For illustration, we simulate the results for our counterfactual bonus programs New1 and New2. As described in the main text, both programs use the following formula

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

In **New1**, we seek to improve efficiency, with bonus rates $(r_0, r_1) = (2.3, 3.1)$. In **New2,** we seek to improve equity by rewarding teachers only based on $c_1\lambda_d$, with bonus rates $(r_0, r_1) = (0, 7.0)$.

We begin by studying the impact of alternative programs in the short run, i.e., when teachers and districts play a static game (as in our main model), and accounting for entry and exit. For each program, Table B15 contrasts its impacts in four cases. Within each block of columns, the first column refers to the setting in our main text (a static game without teachers' extensive-margin responses); the next three columns refer to the setting described in Section B6.2.1 with three different entry/exit elasticities (0.5, 1.0, and 1.5). Under the assumptions about entry/exits stated in Section B6.2.2, we find that both programs would lead to larger gains in student achievement than our baseline setting and the impacts are larger with higher entry/exit elasticities.

Next, we study the impact of the same programs in the medium run, i.e., allowing teachers and districts to repeatedly play the static game. For each program, Table B16 contrasts its impacts in two cases. Within each block of columns, the first column refers to the setting in our main text (a static game without teachers' extensive-margin responses); the second column refers to the setting in Section B6.2.2 (the game played repeatedly for 5 years) with an entry/exit elasticity of 1.0.[9] Under the assumptions stated in Section B6.2.2, we find that both programs would lead to much larger gains in efficiency (total TC in the state) after 5 years of implementation: 2.99% under New1 and 2.08% under New2. Moreover, these programs benefit both low- and high-achieving students, although New1 benefits high-achievers more and New2 benefits low-achievers more.

Taken together, these findings suggest that if teachers' extensive-margin responses are

[8] State transfers to school districts tend to be stable over time as functions of district-level observables (e.g., property values) that get reassessed very infrequently (Biasi, 2023)

[9] The second column compares the year-5 outcomes from the repeated static game with versus without the bonus program.

non-trivial and if the programs last longer, the equity-efficiency gains from our bonus programs can be significantly larger than those reported in the main text. However, we would like to emphasize that findings in Tables B15 and B16 are obtained with additional assumptions, including externally set entry/exit elasticity parameters; readers should interpret these tables with the due caveats.

B7 Tables

B7.1 Tables for Section B1

Table B1: Estimated parameters of teacher effectiveness

	$\hat{\rho}_1$	$\hat{\rho}_2$
exp $= 0$	0	0
exp $\in [1, 2]$	0.0068	0.0009
exp $\in [3, 4]$	0.0154	0.0057
exp $\in [5, 9]$	0.0117	0.0028
exp $\in [10, 14]$	0.0117	0.0049
exp $\in [15, +\infty)$	0.0112	0.0038
R^2	0.677	0.625

Notes: The table shows the parameters on indicators for teacher experience categories in achievement Model 1.

Table B2: Distribution of teacher effectiveness

	c_1	c_2
min	-0.1398	-0.1988
p1	-0.0630	-0.0779
p5	-0.0345	-0.0417
p10	-0.0225	-0.0278
p25	-0.0049	-0.0075
median	0.0115	-0.0108
mean	0.0116	0.0110
p75	0.0282	0.0300
p90	0.0454	0.0503
p95	0.0582	0.0664
p99	0.0894	0.0978
max	0.1532	0.2362

Table B3: Correlation of Teacher Effectiveness between Model 1 and Model 2

	Estimation Sample (2014)		Validation Sample (2010)	
experience	$corr(c_{i1}, \hat{v}'_{i1})$	$corr(c_{i2}, \hat{v}'_{i2})$	$corr(c_{i1}, \hat{v}'_{i1})$	$corr(c_{i2}, \hat{v}'_{i2})$
$= 0$	0.91	0.98	0.86	0.90
$\in [1, 2]$	0.85	0.87	0.86	0.90
$\in [3, 4]$	0.88	0.93	0.88	0.91
$\in [5, 9]$	0.85	0.91	0.85	0.87
$\in [10, 14]$	0.85	0.86	0.86	0.88
≥ 15	0.86	0.87	0.84	0.86

Table B4: Auxiliary Models Aux 1a and Aux 1b, Under Achievement Models 1 and 2

	Aux 1a		Aux 1b	
Achievement	Model 1	Model 2	Model 1	Model 2
wage	0.001 (0.0002)	0.002 (0.0003)	-0.00002 (0.000002))	-0.00002 (0.000003)
e^{λ_d}	-0.002 (0.008)	0.011 (0.010)	-0.0001 (0.0001)	-0.0002 (0.0001)
$c_1 \times \lambda_d$	0.57 (0.29)	0.178 (0.496)	-0.02 (0.006)	-0.02 (0.015)
$I\,(d \neq d_0)$	-0.83 (0.01)	-0.83 (0.02)	-0.98 (0.002)	-0.98 (0.002)
$I\,(d \neq d_0)$ ×untenured	0.48 (0.10)	0.39 (0.13)	0.83 (0.04)	0.84 (0.04)
$I\,(d \neq d_0)$ ×exp $\in [4, 5]$	0.267 (0.031)	0.317 (0.039)	0.236 (0.026)	0.237 (0.027)
$I\,(d \neq d_0)$ ×exp $\in [6, 10]$	0.085 (0.013)	0.099 (0.016)	0.099 (0.010)	0.095 (0.010)
$I\,(d \neq d_0)$ ×exp $\in [11, 15]$	0.020 (0.011)	0.009 (0.011)	0.014 (0.005	0.012 (0.005)
$I\,(z_d \neq z_{d_0})$	-0.0269 (0.005)	-0.0357 (0.007)	-0.0004 (0.0001)	-0.0001 (0.00002)
ln(distance)	-0.019 (0.0019)	-0.019 (0.0026)	-0.0001 (0.00002)	-0.0001 (0.00002)
q_d : urban	0.01 (0.002)	0.003 (0.003)	0.004 (0.0002)	0.003 (0.0002)
q_d : suburban	0.01 (0.002)	0.011 (0.002)	0.001 (0.0001)	0.001 (0.0001)
q_d : large metro	0.10 (0.03)	0.02 (0.03)	0.01 (0.002)	0.01 (0.002)
# Obs	60,841	36,566	2,712,600	

Notes: Achievement Models 1 and 2 as described in Section B1.3.3. Standard errors are in parentheses.

Table B5: Comparison of Signal-to-Noise Ratios with Estimates of Math Teacher Value-Added in The Literature

Estimate	signal-to-noise ratio	calculated as
c (no comparative advantage)	0.69	see Appendix B1.3.4
c_1	0.55	$\hat{\sigma}_{i1}$ (adjusted) $/\bar{\sigma}_{i1}$ (unadjusted) , as defined in our Online Appendix B1.3.1
c_2	0.61	$\hat{\sigma}_{i2}$ (adjusted) $/\bar{\sigma}_{i2}$ (unadjusted) as defined in our Online Appendix B1.3.1
Chetty et al. (2014) elem VA	0.70	ratio between sds in Appendix Figure 1 and in Table 2
Aaronson et al. (2007) high-school VA	0.70	ratio between adjusted and unadjusted sds in Table 6, column 5
Kane and Staiger (2008) elem VA	0.85	ratio between sds in Table 1 and in Table 2 (specification w/student controls)

Table B6: Sum of Squared Test Score Residuals Under c and Under (c_1, c_2)

Effectiveness measure	c	(c_1, c_2)	% difference
Student type:			
All students	0.1680	0.1370	22.61%
$\tau_k = 1$	0.1922	0.1552	23.87%
$\tau_k = 2$	0.1438	0.1189	20.97%

Table B7: Achievement Production Function: Controlling for Teachers and Students' Race/Ethnicity

	$\tau = 1$ (1)	$\tau = 2$ (2)
Black S	-0.056***	-0.067***
	(0.003)	(0.003)
Hisp S	-0.007**	-0.022***
	(0.003)	(0.003)
Asian S	0.053***	0.081***
	(0.004)	(0.004)
Black T	-0.001	0.0001
	(0.005)	(0.005)
Black T * Black S	-0.008	-0.019*
	(0.006)	(0.010)
Hisp T	-0.010*	-0.006
	(0.005)	(0.005)
Hisp T * Hisp S	0.007	0.008
	(0.007)	(0.009)
Asian T	0.003	0.004
	(0.007)	(0.008)
Asian T * Asian S	0.015	0.022
	(0.017)	(0.016)
Observations	3,360,517	3,635,942

Notes: Estimates of achievement model in equation (13), obtained controlling for teachers' (T) and students' (S) race/ethnicity indicators and their interactions.

Table B8: Cross-district Summary of Pre-Reform Wage Schedules

		Cross-district Mean	Cross-district Std Dev.
δ^0		34,686.8	3,286.1
δ^e:	$[1,2]$	1,719.2	598.3
	$[3,4]$	3,939.1	1,103.3
	$[5,9]$	8,227.8	1,536.6
	$[10,14]$	14,644.0	2,348.5
	$\geq$15	21,235.4	3,063.4
δ^m(MA)		7,008.5	2,456.6

B7.2 Tables for Section B3

Table B9: Variation in Salaries Across and Within Districts, 2013-2016

Specification	sqrt(MSE)	R^2	Δsqrt(MSE) from Baseline
Baseline: Experience, Education, c_1,c_2	6,856	0.69	–
+ District FE	4,711	0.86	31.3%
+ School FE	4,523	0.87	34.0%

Table B10: Test for Non-Random Teacher Sorting Across Grade-Schools (Rothstein 2010)

	Residuals		Lagged residuals	
	(1)	(2)	(3)	(4)
Δc_0	1.204***		0.365	
	(0.072)		(0.250)	
Δc_1		0.905***		0.394
		(0.164)		(0.286)
School-by-year FE	Yes	Yes	Yes	Yes
N	6448	1269	1518	298
# school-grades	1950	694	582	174

B7.3 Tables for Section B4

Table B11: Estimates of Aux 1a Assuming Noisy Measures of (c_1, c_2)

	Baseline[a]	For teachers and districts			For districts only		
		σ_{err_k}	2* σ_{err_k}	4* σ_{err_k}	σ_{err_k}	2* σ_{err_k}	4* σ_{err_k}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
wage	0.0012***	0.0017***	0.0018***	0.0028***	0.00172***	0.00177***	0.00280***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
e_d^{λ}	-0.0024	-0.0230**	-0.0177	-0.0416***	-0.0137	-0.00835	-0.00671
	(0.0084)	(0.0114)	(0.0148)	(0.0154)	(0.0093)	(0.0095)	(0.0076)
$c_1 \times \lambda_d$	0.5680**	1.0365***	0.6565**	0.8840***	1.025***	0.792**	0.826***
	(0.2828)	(0.2964)	(0.3086)	(0.2386)	(0.3029)	(0.3122)	(0.2830)
$d \neq d_0$	-0.8259***	-0.7984***	-0.7969***	-0.7843***	-0.799***	-0.797***	-0.786***
	(0.0122)	(0.0138)	(0.0135)	(0.0148)	(0.0138)	(0.0135)	(0.0148)
$d \neq d_0 \times$ untenured	0.4762***	0.3233***	0.3819***	0.3117***	0.324***	0.382***	0.314***
	(0.0981)	(0.1194)	(0.1195)	(0.1171)	(0.1193)	(0.1194)	(0.1169)
$d \neq d_0 \times$ exp	0.2675***	0.2725***	0.2953***	0.2853***	0.273***	0.295***	0.286***
	(0.0314)	(0.0322)	(0.0336)	(0.0337)	(0.0322)	(0.0336)	(0.0337)
$d \neq d_0 \times$exp$\in [6, 10]$	0.0847***	0.0791***	0.0875***	0.0849***	0.0793***	0.0874***	0.0852***
	(0.0126)	(0.0130)	(0.0127)	(0.0134)	(0.0130)	(0.0127)	(0.0134)
$d \neq d_0 \times$exp$\in [11, 15]$	0.0204*	0.0173	0.0308**	0.0084	0.0175	0.0309**	0.00929
	(0.0114)	(0.0118)	(0.0123)	(0.0109)	(0.0118)	(0.0123)	(0.0109)
$z_d \neq z_{d_0}$	-0.0269***	-0.0307***	-0.0310***	-0.0345***	-0.0307***	-0.0311***	-0.0347***
	(0.0048)	(0.0059)	(0.0059)	(0.0068)	(0.0059)	(0.0059)	(0.0068)
urban	0.0138***	0.0243***	0.0225***	0.0205***	0.0243***	0.0225***	0.0209***
	(0.0021)	(0.0027)	(0.0026)	(0.0031)	(0.0027)	(0.0026)	(0.0031)
suburban	0.0115***	0.0103***	0.0123***	0.0021	0.0102***	0.0122***	0.0021
	(0.0021)	(0.0022)	(0.0022)	(0.0025)	(0.0022)	(0.0022)	(0.0025)
ln(distance)	-0.0194***	-0.0227***	-0.0241***	-0.0237***	-0.0227***	-0.0241***	-0.0235***
	(0.0019)	(0.0021)	(0.0021)	(0.0024)	(0.0021)	(0.0021)	(0.0024)
large metro	0.0962***	0.0855***	0.0866***	0.0798***	0.0844***	0.0858***	0.0751**
	(0.0278)	(0.0280)	(0.0273)	(0.0303)	(0.0280)	(0.0273)	(0.0303)
N	60841	52439	53310	46906	52439	53310	46906

Notes: Estimates of Aux 1; Column 1 shows the estimates also shown in column 1 of Table 2 of the paper. Columns 2-4 assume noise in the measures of teacher effectiveness for both teachers and districts, with various variances; and columns 5-7 assume noise in the measures of teacher effectiveness only for districts, with various variances. Robust standard errors in parentheses.

Table B12: OLS of Teacher-District Matches (Aux 1a): Baseline and Excluding Matches for Teachers with c_{1i} or c_{2i} Below the 10th Percentile

Teacher's Choice Set	Baseline Inferred Offer Set[a]	Robustness Inferred Offer Set[b]
wage	0.001 (0.0002)	0.002 (0.0003)
e^{λ_d}	-0.002 (0.008)	-0.007 (0.0142
$c_1 \times \lambda_d$	0.57 (0.28)	0.83 (0.38)
$I(d \neq d_0)$	-0.83 (0.01)	-0.81 (0.02)
$I(d \neq d_0) \times$ untenured	0.48 (0.10)	0.37 (0.11)
$I(d \neq d_0) \times \exp \in [4, 5]$	0.27 (0.03)	0.31 (0.04)
$I(d \neq d_0) \times \exp \in [6, 10]$	0.09 (0.01)	0.10 (0.02)
$I(d \neq d_0) \times \exp \in [11, 15]$	0.02 (0.01)	0.01 (0.01)
$I(z_d \neq z_{d_0})$	-0.03 (0.005)	-0.04 (0.007)
ln(distance)	-0.019 (0.002)	-0.20 (0.002)
q_d : urban	0.01 (0.002)	0.003 (0.003)
q_d : suburban	0.01 (0.002)	0.01 (0.003)
q_d : large metro	0.10 (0.03)	0.08 (0.03)
# Obs	60,841	37,842

Notes:Estimates of Aux 1a on 2014 data. Column 1 baseline estimates; column 2 shows estimates obtained ignoring teacher-district matches (i, d) for teachers with c_{1i} or c_{2i} below the 10th percentile of their respective distribution when inferring matches. Robust standard errors are in parentheses.

B7.4 Tables for Section B5

Table B13: Parameter Permutation Exercise: Change in Estimates of Auxiliary Models from Parameter Perturbation

	Parameter Group		
Auxiliary Model:	Teacher Preferences	District Preferences	Wage-Setting Resistance Costs
Aux 1a	29.10%	2.43%	0.00%
Aux 1b	32.95%	2.46%	0.00%
Aux 2	0.51%	0.04%	0.00%
Aux 3	51.30%	14.03%	27.06%
Aux 4	1.35%	0.03%	5.33%

Notes: Estimates of changes in auxiliary model estimates when we perturb the true preference parameters.

B7.5 Tables for Section B6

Table B14: Allocating Teachers at Will

%	$\frac{\text{Dictator1-Base}}{\lvert\text{Base}\rvert}$	$\frac{\text{Dictator2-Base}}{\lvert\text{Base}\rvert}$
TC	30.97	7.84
c_1	19.37	70.55
c_2	42.75	-55.85

Notes: Policy impacts when teachers are allocated to districts at will.

Table B15: State-Funded Teacher Bonuses: Extensive Margin

(%)	$\frac{\text{New1-Base}}{\lvert\text{Base}\rvert}$				$\frac{\text{New2-Base}}{\lvert\text{Base}\rvert}$			
Entry/Exit Elasticity	-	0.5	1.0	1.5	-	0.5	1.0	1.5
TC for all students	0.26	0.55	0.56	0.86	0.04	0.21	0.53	0.81
$c_1$for low-achieving students	-0.06	0.14	0.26	0.50	0.35	0.22	0.64	0.88
$c_2$for high-achieving students	0.59	0.97	0.87	1.22	-0.26	0.20	0.41	0.75

Notes: Policy impacts when allowing for exit and entry into the market, with the elasticities reported in the column headers.

Table B16: State-Funded Teacher Bonuses: Extensive Margin and Repeated Game

(%)	$\frac{\text{New1-Base}}{\|\text{Base}\|}$		$\frac{\text{New2-Base}}{\|\text{Base}\|}$	
Entry/Exit Elasiticity	-	1.0	-	1.0
Repeated Game (5 Yrs)	No	Yes	No	Yes
TC for all students (efficiency)	0.26	2.99	0.04	2.08
c_1 for low-achieving students	-0.06	2.87	0.35	2.49
c_2 for high-achieving students	0.59	3.13	-0.26	1.60

Notes: Policy impacts when allowing for exit and entry into the market, with the elasticities reported in the column headers, and assuming districts play a repeated static game.

B7.6 Additional Tables: Data and Model Fit

Table B17: Teacher and District Characteristics (2010)

A. Teacher Characteristics	All	$x_1 < 3$	$x_1 \geq 10$
x_1: Experience	15.6 (9.6)	1.6 (0.5)	20.2 (7.7)
x_2: MA or above	0.55 (0.50)	0.05 (0.22)	0.66 (0.48)
$10c_1$	0.11 (0.25)	0.07 (0.27)	0.11 (0.25)
$10c_2$	0.12 (0.30)	0.06 (0.32)	0.12 (0.29)
Corr (c_1, c_2)	0.65	-	-
# Teachers	6,741	391	4,675
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	16.4 (30.7)	18.4 (16.2)	15.1 (46.2)
Budget/Capacity ($1,000)	52.4 (6.1)	54.3 (6.7)	51.2 (5.7)
Characteristics of District Incumbent Teachers ($d_0 = d$)			
Average experience	17.5 (5.1)	16.6 (4.6)	18.0 (5.6)
Share w/MA or above	0.52 (0.26)	0.57 (0.26)	0.48 (0.28)
Average $10c_1$	0.10 (0.10)	0.10 (0.09)	0.09 (0.13)
Average $10c_2$	0.11 (0.13)	0.11 (0.11)	0.09 (0.15)
# Districts	411	103	103

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

Table B18: Model Fit: Average District Employee Characteristics ($d^*(\cdot) = d$)

District Group		Experience		Share MA or above		$10c_1$		$10c_2$	
		Data	Model	Data	Model	Data	Model	Data	Model
λ_d	Quintile 1	14.7	13.5	0.53	0.48	0.13	0.14	0.11	0.13
	Quintile 2	15.5	14.7	0.51	0.49	0.12	0.14	0.13	0.15
	Quintile 3	15.6	14.7	0.48	0.46	0.14	0.14	0.12	0.13
	Quintile 4	16.3	15.6	0.48	0.48	0.14	0.14	0.16	0.15
$\frac{\text{Budget}}{\text{Capacity}}$	Quintile 1	11.5	11.8	0.29	0.33	0.14	0.15	0.12	0.14
	Quintile 2	14.8	14.1	0.38	0.38	0.11	0.14	0.12	0.14
	Quintile 3	15.9	15.0	0.48	0.46	0.13	0.13	0.12	0.13
	Quintile 4	17.7	16.2	0.59	0.56	0.13	0.13	0.13	0.14
Urban		14.2	15.2	0.57	0.59	0.10	0.11	0.09	0.09
Suburban		14.7	12.7	0.60	0.52	0.14	0.13	0.13	0.12

Notes: Moments as specified in Aux 2. All estimates use data post-Act 10.

Table B19: Model Fit: OLS of District Wage Schedule

Auxiliary Model 3	ω_{d1}		ω_{d2}	
	Data	Model	Data	Model
Composition of incumbent teachers ($d_0 = d$)				
Fr(experience 3-4)	0.01	0.001	1.57	5.02
Fr(experience 5-9)	0.01	0.01	1.50	-1.04
Fr(experience 10-14)	-0.004	0.008	11.36	-0.44
Fr(experience $\geq$ 15)	0.03	-0.0001	-21.97	-0.49
Fr(MA or above)	-0.03	-0.004	-11.88	-1.41
Average TC	-0.52	0.61	200.40	2.16
Average TC among Tenured	0.38	-0.54	-508.90	-264.9
District Characteristics				
λ_d	0.001	0.01	25.24	2.16
budget per teacher	0.002	0.001*	0.53	0.02
capacity	-0.00002	0.0001	-0.35	-0.01*
urban	-0.02	-0.01	19.77	1.84
suburban	-0.02	-0.004*	2.59	1.69
large metro	0.02	-0.056	97.94	3.45
share Democratic votes (2012)	0.03	0.04	-56.46	-45.42
Teachers in nearby districts ($z_{d_0} = z_d$, $d_0 \neq d$)				
Average TC	-0.65	0.10	1206.19	-70.72
Share of Tenured	-0.04	0.02	143.830	5.19
# obs.	411		411	

Notes: OLS estimates of Aux 3. * denotes model estimates outside of the 95% CI of the estimates from the data.

Table B20: Model Validation: Average District Employee Characteristics (pre-Act 10)

District Group		Experience		Share MA or above		$10c_1$		$10c_2$	
		Data	Model	Data	Model	Data	Model	Data	Model
λ_d	Quintile 1	16.1	15.8	0.56	0.52	0.10	0.11	0.11	0.12
	Quintile 2	16.4	16.0	0.51	0.50	0.10	0.10	0.11	0.11
	Quintile 3	17.6	16.8	0.46	0.46	0.10	0.11	0.13	0.13
	Quintile 4	17.5	16.9	0.52	0.50	0.09	0.10	0.10	0.11
$\frac{\text{Budget}}{\text{Capacity}}$	Quintile 1	13.5	13.5	0.27	0.29	0.10	0.11	0.12	0.13
	Quintile 2	17.7	16.9	0.42	0.42	0.11	0.12	0.12	0.12
	Quintile 3	17.2	16.6	0.52	0.51	0.09	0.09	0.10	0.10
	Quintile 4	18.7	17.5	0.60	0.56	0.08	0.09	0.10	0.10
Urban		15.2	15.5	0.56	0.56	0.14	0.14	0.13	0.13
Suburban		15.6	14.0	0.62	0.58	0.08	0.09	0.10	0.11

Notes: Moments as specified in Aux 2. All estimates use data pre-Act 10.

B8 Figures

Figure B1: Distribution of teacher effectiveness

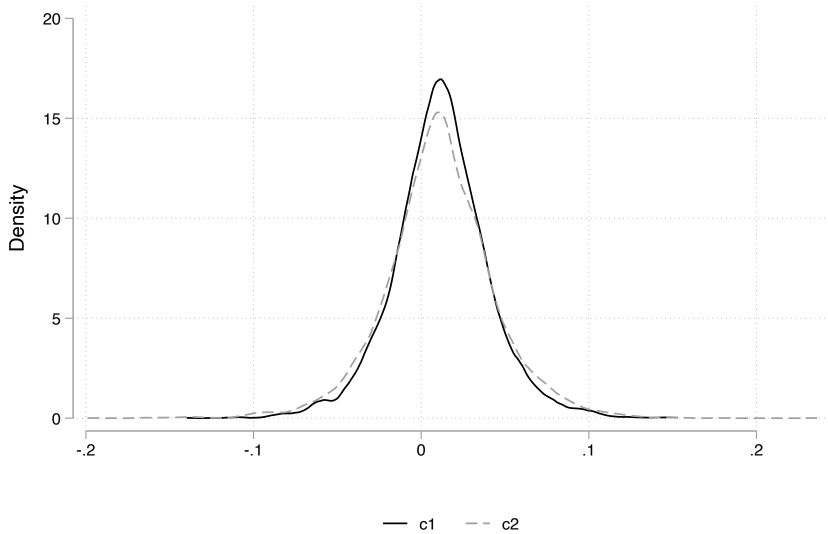


Figure B2: Relationship between c_1 and c_2

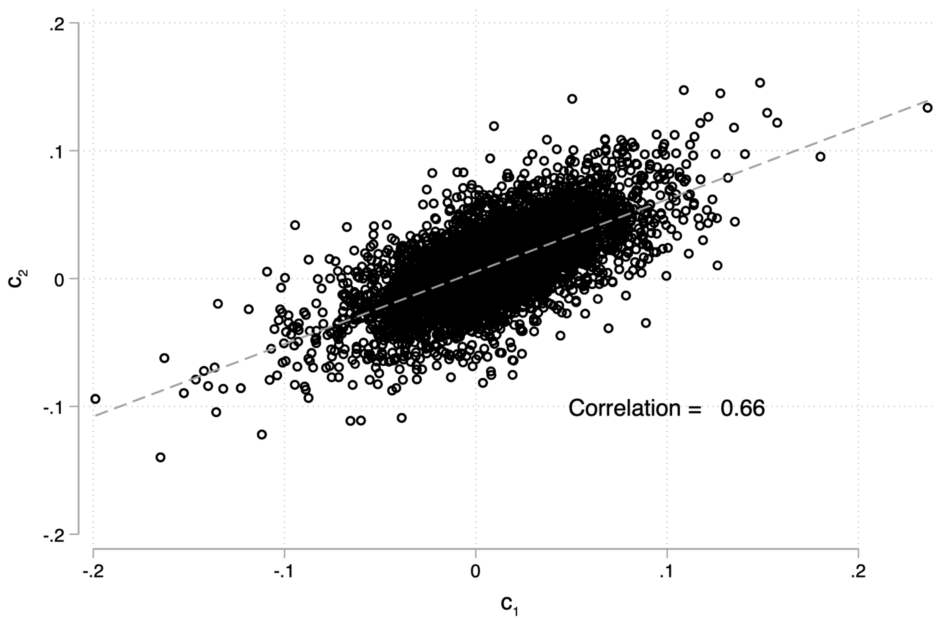
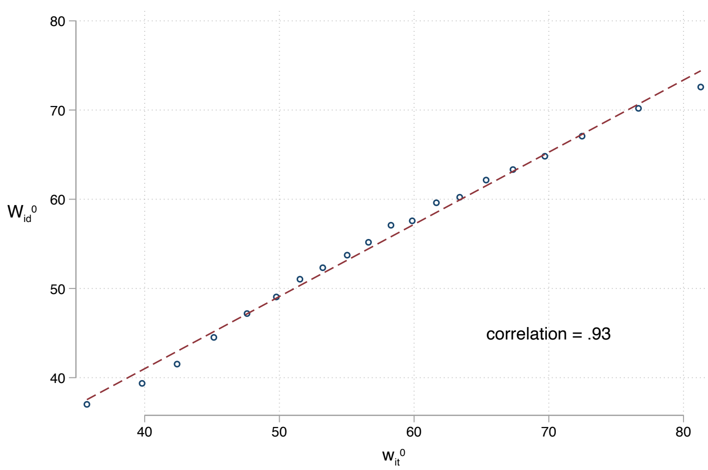


Figure B3: Relationship between $W_d^0(x_i)$ and w_{it}^0

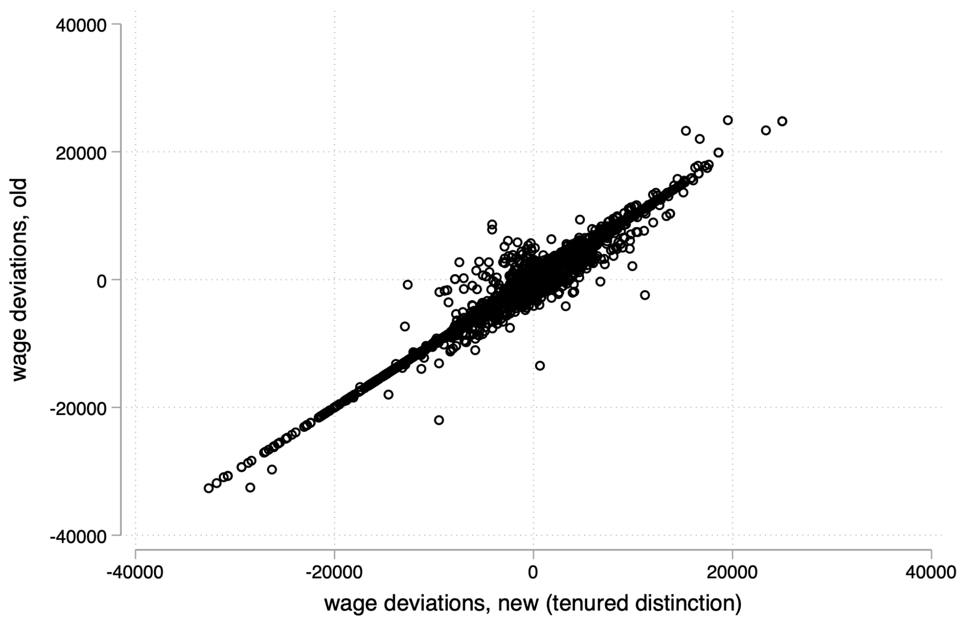


Note: Binned scatterplot of $W_d^0(x_i)$ and w_{it}^0 using wage data from 2010.

Figure B4: Relationship between deviations of true wages from $w_d(x, c|\omega)$, obtained using rules (13) and (14)

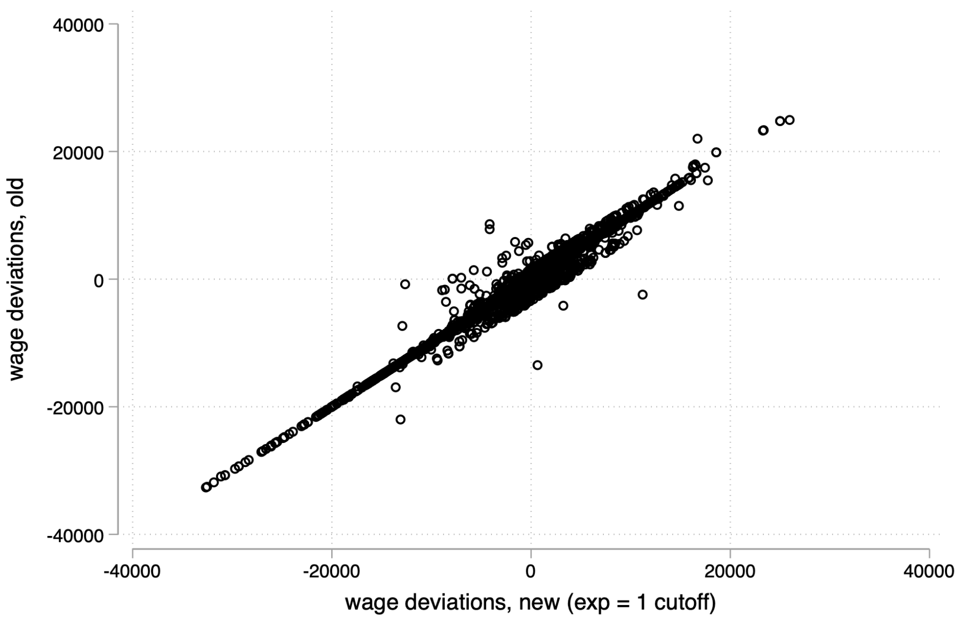
Note: Binned scatterplot of the difference between true 2014 teacher wages and $w_d(x, c|\omega)$, calculated using (13) (vertical axis) and (14) (horizontal axis).

Figure B5: Relationship between deviations of true wages from $w_d\,(x, c|\omega)$, obtained using rules (13) and (15), where $x_1 = 1$ for untenured teachers



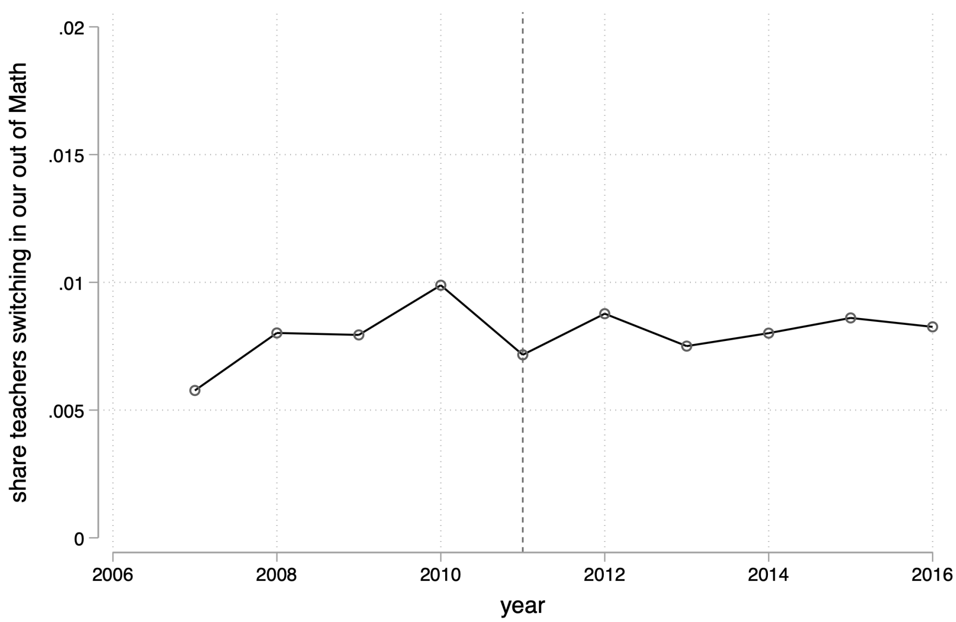
Note: Binned scatterplot of the difference between true 2014 teacher wages and $w_d\,(x, c|\omega)$, calculated using (13) (vertical axis) and (15), where $x_1 = 1$ for untenured teachers (horizontal axis).

Figure B6: Relationship between deviations of true wages from $w_d\,(x, c|\omega)$, obtained using rules (13) and (15), where $x_1 = 1$ for teachers with no experience



Note: Binned scatterplot of the difference between true 2014 teacher wages and $w_d\,(x, c|\omega)$, calculated using (13) (vertical axis) and (15), where $x_1 = 1$ for teachers with no experience (horizontal axis).

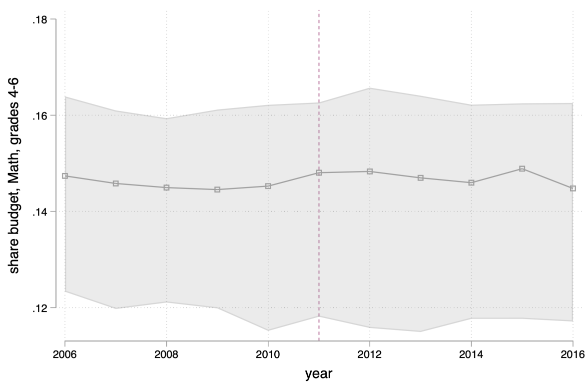
Figure B7: Share of Teachers Who Switch In and Out of Math Teaching, By Year



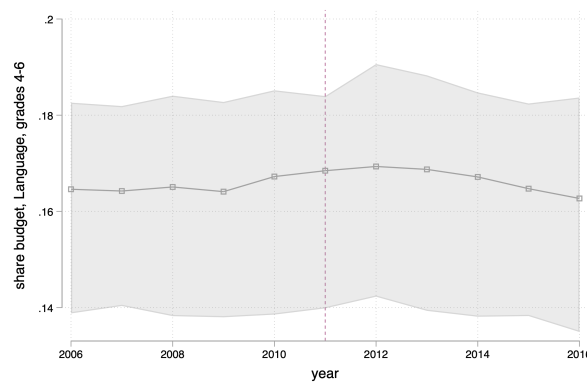
Note: Share of teachers who switch into or out of math teaching, in each year and out of the total number of teachers in Wisconsin.

Figure B8: Share of districts' budgets spent on teacher salaries, by grade and subject

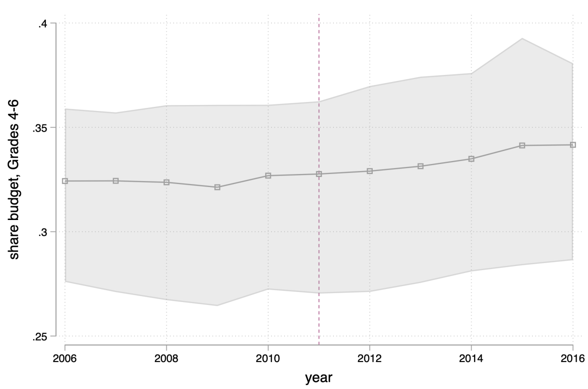
(a) Math, Grades 4-6



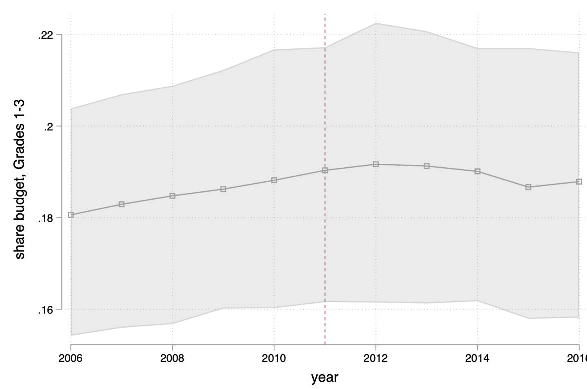
(b) Reading/ELA, Grades 4-6



(c) All Teachers, Grades 4-6



(d) All Teachers, Grades 1-3



Notes: The figure shows the mean and the interquartile range of the share of total wages paid to teachers in each subgroup, calculated for each district.

References

Aaronson, D., L. Barrow, and W. Sander (2007). Teachers and student achievement in the chicago public high schools. *Journal of labor Economics 25*(1), 95–135.

Biasi, B. (2023). School finance equalization increases intergenerational mobility. *Journal of Labor Economics 41*(1), 1–38.

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.

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

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.

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

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.