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, 2021

B1 Algorithms

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

$$h_d\left(x,c,d_0\right)=\frac{\exp\left(\frac{V_d(x,c,d_0)}{\sigma\epsilon}\right)}{\exp\left(\frac{V_d(x,c,d_0)}{\sigma_\epsilon}\right)+\sum_{d'\in D\backslash d}o_{d'}\left(x,c,d_0\right)\exp\left(\frac{V_{d'}(x,c,d_0)}{\sigma\epsilon}\right)}. \tag{1}$$

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

$$\widetilde{h}_d\left(x,c,d_0|\overline{w}\left(x,c\right),\sigma_w\left(x,c\right)\right)=\frac{1}{1+\exp\left(f\left(x,c,d_0,w_d,q_d,\lambda_d\right)\right)}, \tag{2}$$

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

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

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.,

$$\overline{w}(x,c) \equiv \frac{1}{D}\sum_d w_d(x,c;\omega_d) \tag{3}$$

$$\sigma_{w(x,c)} \equiv \sqrt{\frac{1}{D-1}\sum_d \left(w_d(x,c;\omega_d) - \overline{w}(x,c)\right)^2}. \tag{4}$$

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.

B1.1 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 (2). 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 $\left\{\widetilde{h}_d(\cdot|\zeta, \overline{w}^o(\cdot), \sigma_w^o(\cdot))\right\}$ as defined in (2).

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

    (c) Given the job offers and the wages implied by $\{o_d^*(\cdot;\omega_d^o), \omega_d^o\}_d$, calculate each teacher's acceptance probabilities $h_d(\cdot)$ for each d, as in (1), and the distance $\left\|h(\cdot) - \widetilde{h}(\cdot|\zeta, \overline{w}^o(\cdot), \sigma_w^o(\cdot))\right\|$.

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

2. Given job offers $\{o_d^*(\cdot;\omega_d^o)\}_d$ under $\widetilde{h}_d(\cdot|\zeta^*(\Theta), \overline{w}^o(\cdot), \sigma_w^o(\cdot))$ and wages implied by $\{\omega_d^o\}$, 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 $\omega_d^*(\Theta)$.[1] The resulting $\{\omega_d^*(\Theta)\}_d$ will be compared with the observed $\{\omega_d^o\}_d$ in the outer loop.

B1.2 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 $\{(\omega_{d1}, \omega_{d2})\}_d$ that govern these statistics are high-dimensional objects. However, notice that districts' wages are given by

$$w_d(x, c; \omega) = \begin{cases} \underline{w} \text{ if } \omega_1 W_d^0(x) + \omega_2\left[\lambda_d c_1 + (1-\lambda_d) c_2\right] < \underline{w} \\ \overline{w} \text{ if } \omega_1 W_d^0(x) + \omega_2\left[\lambda_d c_1 + (1-\lambda_d) c_2\right] > \overline{w} \\ \omega_1 W_d^0(x) + \omega_2\left[\lambda_d c_1 + (1-\lambda_d) c_2\right] \text{ otherwise} \end{cases}, \tag{5}$$

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

$$\widetilde{\overline{w}}(x, c) = \begin{cases} \underline{w} \text{ if } \sum_n \theta_{1n}^1 I(x_1 = n) + \theta_2^1 x_2 + \theta_3^1 c_1 + \theta_4^1 c_2 < \underline{w} \\ \overline{w} \text{ if } \sum_n \theta_{1n}^1 I(x_1 = n) + \theta_2^1 x_2 + \theta_3^1 c_1 + \theta_4^1 c_2 > \overline{w} \\ \sum_n \theta_{1n}^1 I(x_1 = n) + \theta_2^1 x_2 + \theta_3^1 c_1 + \theta_4^1 c_2 \text{ otherwise.} \end{cases} \tag{6}$$

Similary, 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.,

$$\widetilde{\sigma}_{w(x,c)} = \min\left\{\sqrt{\max\left\{Q\left(x_1, x_2, c_1, c_2; \theta^2\right), 0\right\}}, \overline{w} - \underline{w}\right\}. \tag{7}$$

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 (2), (6) and (7) to guarantee equilibrium consistency. Note that none of 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.

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

B1.2.1 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 conduct the following procedure.

1. Guess parameters ζ, θ^1, and θ^2, which imply $\left\{\widetilde{\overline{w}}\left(x,c\right), \widetilde{\sigma}_{w(x,c)}, \widetilde{h}_d\left(x,c,d_0|\widetilde{\overline{w}}\left(x,c\right), \widetilde{\sigma}_{w(x,c)}\right)\right\}$ from (2), (6) and (7).
2. Given $\left\{\widetilde{h}_d\left(x,c,d_0|\widetilde{\overline{w}}\left(x,c\right), \widetilde{\sigma}_{w(x,c)}\right)\right\}$, each district d chooses its optimal wage and offer policies $\{\omega_d, O\left(\omega_d\right)\}$.
3. Given $\{\omega_d, O\left(\omega_d\right)\}_d$, compute teacher acceptance probabilities $h_d\left(\cdot\right)$ from their decision rules (1), the mean wage $\overline{w}\left(x,c\right)$ based on (3), and standard deviation $\sigma_{w(x,c)}$ based on (4).
4. Calculate the distance between $\left\{\widetilde{\overline{w}}\left(x,c\right), \widetilde{\sigma}_{w(x,c)}, \widetilde{h}_d\left(x,c,d_0|\widetilde{\overline{w}}\left(x,c\right), \widetilde{\sigma}_{w(x,c)}\right)\right\}$ and $\left\{\overline{w}\left(x,c\right), \sigma_{w(x,c)}, h_d\left(x,c,d_0|\left(\overline{w}\left(x,c\right), \sigma_{w(x,c)}\right)\right)\right\}$.
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 (2). Equilibrium outcomes in economy m consist of the decisions made by districts and teachers under this consistent belief.

B2 Data Details

B2.1 Sample Construction

We construct our samples as follows. For estimation and empirical analysis, we focus on full-time Grade 4-6 math teachers employed in Wisconsin school districts in 2014 (411 districts and 6,625 individuals).[2] 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 Grade 4-6 math teachers employed in

[2]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.

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.

B2.2 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} 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.

B2.3 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, socio-economic (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 B6 and B7 and Figure 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.

B2.3.1 Achievement Model 1 (Main)

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

$$A_{kt} = \gamma Z_{kt}^{s} + \sum_{i:SG_{kt}=SG_{it}^{T}} \sum_{n=1}^{2} I\left(\tau_k = n\right)\left(\rho_n x_{it} + v_{in}\right) + \varepsilon_{kt} \tag{8}$$

$$= \gamma Z_{kt}^{s} + \sum_{i:SG_{kt}=SG_{it}^{T}} \sum_{n=1}^{2} I\left(\tau_k = n\right)\rho_n x_{it} + \varphi_{kt} \tag{9}$$

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 is 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 (8) 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 (8) 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 iid 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.

Estimation Procedure: Model 1

1. Given A1 and A2, we estimate γ and ρ_n via OLS on equation (8), 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

   $$\widehat{\varphi}_{kt} = A_{kt} - \hat{\gamma} Z_{kt}^s - \sum_{i:SG_{kt}=SG_{it}^T} \sum_{n=1}^{2} \hat{\rho}_n x_{it} I(\tau_k = n). \tag{10}$$

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

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

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

   $$\widehat{v}_{int} = \frac{1}{K_{SG_{it}^T n}} \sum_{k:SG_{kt}=SG_{it}^T} \widehat{\varphi}_{kt} I\left(\tau_k = n\right) \tag{11}$$

   which is an estimate of

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

   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_{kt}^s 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 $\varpi_{int} = \frac{K_{SG_{it}^T n}}{\sum_t K_{SG_{it}^T n}}$, so that residuals corresponding to more observations receive more weight:

   $$\bar{v}_{in} = \sum_t \varpi_{int} \widehat{v}_{int} \tag{12}$$

   Note that assumption A1 implies

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

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 (12) 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 $i' : SG^T_{i't} = SG^T_{it} \forall t$.

B2.3.2 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).[3] The achievement model in this case would be:

$$A_{kt} = \gamma Z^s_{kt} + \sum_{n=1}^{2} I\left(\tau_k = n\right) v_{i(kt)n} + \varepsilon_{kt} \tag{13}$$

$$= \gamma Z^s_{kt} + \varphi_{kt} \tag{14}$$

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 (13) to obtain $\widehat{\gamma}$.

2. We construct
$$\widehat{\varphi}'_{kt} = A_{kt} - \hat{\gamma} Z^s_{kt} \tag{15}$$

[3] 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 $n \in \{1, 2\}$

$$\widehat{v}'_{int} = \frac{1}{K_{SG^T_{it} n}} \sum_{k:SG_{kt}=SG^T_{it}} \widehat{\varphi}'_{kt} I\left(\tau_k = n\right) \tag{16}$$

$$\text{which is an estimate of} \quad \sum_{i':SG^T_{it}=SG^T_{i't}} j_{i'nt} v_{i'n} + \frac{1}{K_{SG^T_{it} n}} \sum_{k:SG_{kt}=SG^T_{it}} \varepsilon_{kt} \tag{17}$$

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

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

Taking the limit of this expectation as t approaches infinity implies

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

It follows that the estimator

$$\bar{\bar{v}}'_{in} = \frac{1}{\sum_t \frac{\varpi_{int}}{J_{SG^T_{it}}}} \bar{v}'_{in} \tag{18}$$

is asymptotically unbiased if $\lim_{t\to\infty} \sum_t \frac{\varpi_{int}}{J_{SG^T_{it}}} \sum_{i':SG^T_{it}=SG^T_{i't}} v_{i'n} = 0$. 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

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

Table B1: 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

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 (18) 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 $SG^T_{it} \neq SG^T_{it'}$.

B2.3.3 Teacher Effectiveness: Model 1 vs Model 2

Correlation of Teacher Effectiveness Measures Table B1 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.

Table B2: OLS of Teacher-District Match

	Aux 1a		Aux 1b	
Achievement	Model 1	Model 2	Model 1	Model 2
wage	0.002 (0.0002)	0.001 (0.0001)	-0.000005 (0.000002)	-0.000005 (0.000002)
e^{λ_d}	-0.004 (0.009)	-0.003 (0.005)	-0.0001 (0.0001)	-0.0002 (0.0001)
$c_1 \times \lambda_d$	0.52 (0.29)	0.32 (0.19)	-0.02 (0.006)	-0.02 (0.015)
$I(d \neq d_0)$	-0.72 (0.02)	-0.73 (0.02)	-0.80 (0.01)	-0.80 (0.01)
$I(d \neq d_0) \times x_1$	-0.008 (0.001)	-0.008 (0.001)	-0.008 (0.0005)	-0.008 (0.0005)
$I(z_d \neq z_{d_0})$	-0.06 (0.006)	-0.06 (0.005)	-0.0006 (0.0001)	-0.0006 (0.0001)
q_d : urban	0.01 (0.002)	0.01 (0.002)	0.003 (0.0002)	0.003 (0.0002)
q_d : suburban	0.01 (0.002)	0.01 (0.002)	0.001 (0.0001)	0.001 (0.0001)
q_d : large metro	0.11 (0.03)	0.09 (0.02)	0.01 (0.002)	0.01 (0.002)
# Obs	57,068	60630	2,712,600	

Standard errors are in parentheses.

For the robustness of teacher preferences under O_i^s in place of $\widetilde{O}_i^s$, case 1) is ideal; 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,

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

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 B2, 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.

B2.3.4 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

Table B3: Sum of squared test score residuals under (c_1, c_2) and under c

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%

(8) with one-dimensional rather than two-dimensional teacher effectiveness and compare it with Model (8). Table B3 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.

B2.4 Wage Schedules

B2.4.1 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:

$$w_{it}^0 = \delta_d^0 + Exp_{it}\delta_{g(i)}^e + MA_{it}\delta_{g(i)}^m + \varepsilon_{it}, \tag{19}$$

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	$\geq$median	$\geq$median	$\geq$median
2	$\geq$median	$\geq$median	$<$median
3	$\geq$median	$<$median	$\geq$median
4	$<$median	$\geq$median	$\geq$median
5	$<$median	$<$median	$\geq$median
6	$<$median	$\geq$median	$<$median
7	$\geq$median	$<$median	$<$median
8	$<$median	$<$median	$<$median

Table B4 summaries the point estimates from Equation (19). In particular, it reports the cross-district means and standard deviations of the estimated vectors δ. Figure B1 shows a binned scatterplot 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).

B2.4.2 Districts' Choice Set of Wage Schedules

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

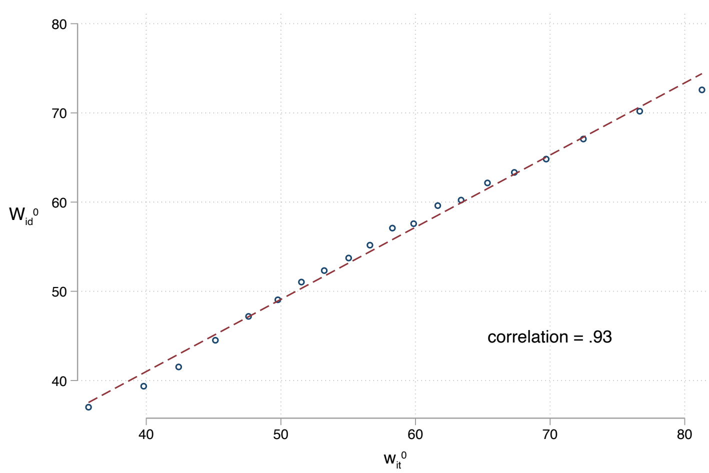
1. We start by estimating the parameters $(\widetilde{\omega}_{d1}, \widetilde{\omega}_{d2}) \geq 0$ separately for each districts from

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

Table B4: 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
	≥ 15	21,235.4	3,063.4
δ^m(MA)		7,008.5	2,456.6

Figure B1: 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.

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 B2.4.1, 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 allow for a wider support of wage schedule choices in our model than the observed support Ω^o because in counterfactual policy experiments. 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.

Table B5: 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%

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 B5 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 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.

B4 Additional Tables and Figures

Table B6: 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

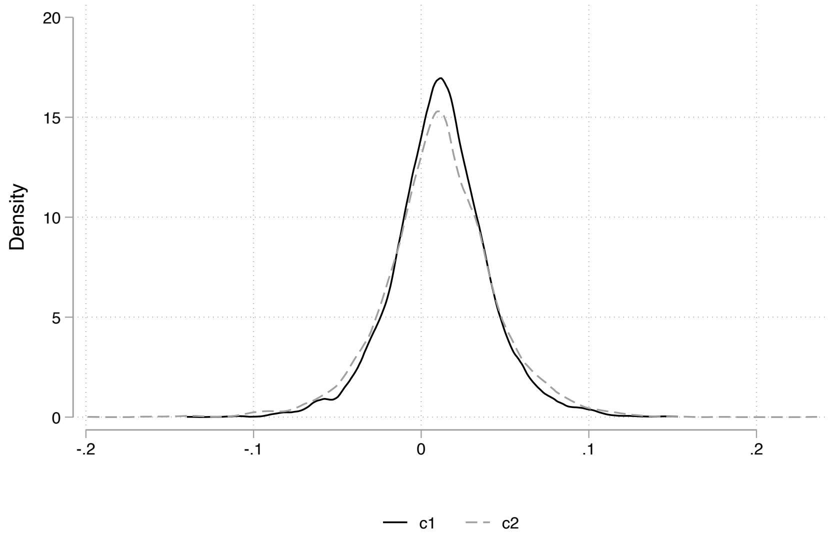
Table B7: Distribution of teacher effectiveness

	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

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.

Figure B2: Distribution of teacher effectiveness



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.