

Estimating Heterogeneous Teacher Effects with Grouped Three-Way Interactive Fixed Effects

Sangmyung Ha

Department of Economics, Indiana University

2026 Midwest Econometrics Group Meeting

University of Cincinnati

October 10, 2026

Research Questions

1. How does teacher effectiveness vary across student characteristics and academic years?
2. How much of the variation in teacher value added is attributable to student–teacher matching versus general level effects?
3. For which student groups does teacher assignment matter most, and how sensitive are teacher rankings and student outcomes to classroom assignment?

Modeling challenge

An unrestricted group–teacher–year effect,

$$\{\tau_{gjt}\}_{g,j,t},$$

requires *GMT* parameters.

Our approach

We impose a parsimonious grouped three-way interactive structure:

$$\tau_{gjt} = \nu_j + \sum_{k=1}^2 u_{gk} v_{jk} w_{tk}, \quad G = 5.$$

Five prior-achievement student groups are linked through two latent dimensions of teacher comparative advantage.

Takeaways

Theoretical

Imposing more structure in the interactive fixed effect

- ▶ helps the identification of the slope parameter,
- ▶ and reduces the order of incidental-parameter bias.

Empirical

- ▶ Student–teacher matching accounts for 27% of the fitted teacher-value-added variation.
- ▶ Teacher assignment is most consequential for students in the bottom prior-achievement quintiles.
- ▶ Counterfactual analysis shows that up to 27.59% of the teachers ranked at the bottom 5% will escape the bottom 5% if assigned to a different classroom within the same school.
- ▶ Optimally allocating teachers to different classrooms within schools can increase students' test scores by an average of up to 0.054 standard deviations. (equivalent to 31.7 additional school days, Aucejo and Romano (2016))
- ▶ Comparison with randomized roster benchmark shows strong agreement.

Panel model with grouped three-way interactive effects

For $i = 1, \dots, N$, $j = 1, \dots, M$, and $t = 1, \dots, T$, the observed outcome satisfies

$$y_{ijt} = x'_{ijt}\delta + \sum_{k=1}^{\ell} u_{g(i),k} v_{jk} w_{tk} + \varepsilon_{ijt},$$

where $x_{ijt} \in \mathbb{R}^p$ is observed, $\sum_{k=1}^{\ell} u_{g(i),k} v_{jk} w_{tk}$ is unobserved fixed effects, and ε_{ijt} is an idiosyncratic error.

- ▶ (j, t) a classroom taught by teacher j in period t
 i local student position within (j, t) .
- ▶ y_{ijt} : The test score of i -th student in classroom (j, t) .
- ▶ x_{ijt} : The characteristics of i -th student in classroom (j, t) .
- ▶ The group mapping

$$g : \{1, \dots, N\} \longrightarrow \{1, \dots, G\}$$

is observed and treated as fixed. Units in the same group share the first-mode loading vector u_g .

Model in Tensor Form

Given a balanced sample, the model can be written in tensor form as

$$\mathbf{Y} = \mathbf{X} \cdot \delta + \sum_{k=1}^{\ell} (Du_k \circ v_k \circ w_k) + \mathbf{E},$$

where $\mathbf{Y}, \mathbf{E} \in \mathbb{R}^{N \times M \times T}$ and $\mathbf{X} \in \mathbb{R}^{N \times M \times T \times p}$ with

$$\mathbf{X} \cdot \delta = (x'_{ijt} \delta)_{i,j,t} \in \mathbb{R}^{N \times M \times T}.$$

The grouping matrix $D = (1\{g(i) = h\})_{i,h=1}^{N,G} \in \mathbb{R}^{N \times G}$ assigns students to groups and is known. We collect the component-specific vectors into the factor matrices

$$U = (u_1, \dots, u_{\ell}) \in \mathbb{R}^{G \times \ell}, \quad V = (v_1, \dots, v_{\ell}) \in \mathbb{R}^{M \times \ell}, \quad W = (w_1, \dots, w_{\ell}) \in \mathbb{R}^{T \times \ell}.$$

Model in Vector Form

The model admits the vectorized representation

$$y = X\delta + \sum_{k=1}^{\ell} (Du_k \otimes v_k \otimes w_k) + e,$$

where $y = \text{vec}(\mathbf{Y})$, $e = \text{vec}(\mathbf{E}) \in \mathbb{R}^{NMT}$ and $\text{vec}(\mathbf{X} \cdot \delta) = X\delta$ with X defined as lexicographic ordering of (x_{ijt}) given by

$$X = \begin{pmatrix} x'_{111} \\ x'_{112} \\ \vdots \\ x'_{NM,T-1} \\ x'_{NMT} \end{pmatrix} \in \mathbb{R}^{NMT \times p}, \quad X'X = \sum_{i=1}^N \sum_{j=1}^M \sum_{t=1}^T x_{ijt} x'_{ijt}.$$

Estimation

For a given number ℓ of interactive components, we estimate the model by minimizing the average squared residual

$$S_{NMT}(\delta, U, V, W) = \frac{1}{NMT} \sum_{i=1}^N \sum_{j=1}^M \sum_{t=1}^T \left(y_{ijt} - x'_{ijt} \delta - \sum_{k=1}^{\ell} u_{g(i),k} v_{jk} w_{tk} \right)^2 \quad (1)$$

with respect to (δ, U, V, W) . The minimizer may be obtained using alternating least squares.

Three FWL representations of $\hat{\delta}$

Depending on whether $(v_k \otimes w_k)$, $(Du_k \otimes w_k)$, or $(Du_k \otimes v_k)$ are interpreted as regressors in

$$y = X\delta + \sum_{k=1}^{\ell} (Du_k \otimes v_k \otimes w_k) + e,$$

we have three Frisch–Waugh–Lovell-type representation of $\hat{\delta}$

$$\hat{\delta} = (X'(I_{NMT} - P_D \otimes P_{\hat{V}*\hat{W}})X)^{-1} X'(I_{NMT} - P_D \otimes P_{\hat{V}*\hat{W}})y$$

$$\hat{\delta} = (X'(K_{1,2} (I_M \otimes M_{D\hat{U}*\hat{W}}) K_{2,1})X)^{-1} X'(K_{1,2} (I_M \otimes M_{D\hat{U}*\hat{W}}) K_{2,1})y$$

$$\hat{\delta} = (X'(M_{D\hat{U}*\hat{V}} \otimes I_T)X)^{-1} X'(M_{D\hat{U}*\hat{V}} \otimes I_T)y$$

Here $K_{2,1} = K'_{1,2}$, and

$$K_{1,2}(b \otimes a \otimes c) = a \otimes b \otimes c$$

for $a \in \mathbb{R}^N$, $b \in \mathbb{R}^M$, and $c \in \mathbb{R}^T$.

Least squares estimator: tangent-space representation

Define the space of generated regressors

$$\hat{\mathcal{T}} = \left\{ \sum_{k=1}^{\ell} \begin{pmatrix} Da_k \otimes \hat{v}_k \otimes \hat{w}_k \\ + D\hat{u}_k \otimes b_k \otimes \hat{w}_k \\ + D\hat{u}_k \otimes \hat{v}_k \otimes c_k \end{pmatrix} : \begin{array}{l} a_k \in \mathbb{R}^G, \\ b_k \in \mathbb{R}^M, \\ c_k \in \mathbb{R}^T \end{array} \right\}.$$

The three Frisch–Waugh–Lovell-type representations imply

$$\hat{\delta} = (X' M_{\hat{\mathcal{T}}} X)^{-1} X' M_{\hat{\mathcal{T}}} y.$$

Assumption 1 (Strong Factors and Nonpervasive Errors)

There exists finite positive-definite matrices $\Sigma_U, \Sigma_V, \Sigma_W$

$$\frac{1}{N} \sum_{i=1}^N u_{g(i)} u'_{g(i)} \xrightarrow{p} \Sigma_U, \quad \frac{1}{M} \sum_{j=1}^M v_j v'_j \xrightarrow{p} \Sigma_V, \quad \frac{1}{T} \sum_{t=1}^T w_t w'_t \xrightarrow{p} \Sigma_W,$$

as $\min\{NMT/G, NT, NM\} \rightarrow \infty$.

Let

$$\mathbf{E}^G = \left(\frac{1}{\sqrt{N_g}} \sum_{\{i:g(i)=g\}} \varepsilon_{ijt} \right)_{g,j,t=1}^{G,M,T} \in \mathbb{R}^{G \times M \times T}.$$

The idiosyncratic-error tensor satisfies

$$\|\mathbf{E}^G\| = O_p(\sqrt{G + M + T}).$$

where $\|\cdot\|$ denotes the tensor spectral norm.

Assumption 2 (Conditional Covariance Summability)

Let

$$\mathcal{F} = \sigma(X, D, U, V, W), \quad \mathcal{I} = \{1, \dots, N\} \times \{1, \dots, M\} \times \{1, \dots, T\}.$$

For $a = (i_a, j_a, t_a) \in \mathcal{I}$, write $\varepsilon_a = \varepsilon_{ijt}$. We have

$$\mathbb{E}[\varepsilon_a \mid \mathcal{F}] = 0$$

for every $a \in \mathcal{I}$. Furthermore, there exists some finite constant $C_\varepsilon > 0$ such that with probability approaching one,

$$\sup_{a \in \mathcal{I}} \sum_{b \in \mathcal{I}} |\mathbb{E}[\varepsilon_a \varepsilon_b \mid \mathcal{F}]| \leq C_\varepsilon.$$

This implies $\|\mathbb{E}[ee' \mid \mathcal{F}]\| < \infty$.

Assumption 3 (Identification of δ)

There exists a constant $c_\delta > 0$ such that, with probability approaching one,

$$\inf_{\tilde{U} \in \mathbb{R}^{G \times \ell}, \tilde{V} \in \mathbb{R}^{M \times \ell}} \lambda_{\min} \left[\frac{1}{NMT} X' (M_{D\tilde{U}*} \tilde{V} \otimes M_W) X \right] \geq c_\delta,$$

where W are the true factors.

Assumption 3 ensures that the common slope coefficient, δ , can be identified separately from the interactive effects. This condition ensures some regressor variation that cannot be absorbed by changing the factors.

Comparison with Bai (2009)

Suppose instead that the indices (i, j) are enumerated into a single cross-sectional index and the model is estimated as the unrestricted two-way interactive fixed-effects model of Bai (2009). The corresponding identification condition is given by:

$$\inf_{H \in \mathbb{R}^{NM \times \ell}} \lambda_{\min} \left[\frac{1}{NMT} X' (M_H \otimes M_W) X \right] \geq c_{\text{flat}} > 0, \quad (2)$$

where W are the true factors.

A scalar regressor of the form

$$x_{ijt} = q_{ij} \tau_t$$

does not satisfy the identification condition (2).

The grouped condition can still hold if the cross-sectional component q is uniformly separated from all rank- ℓ separable loading spaces and the time component τ is uniformly separated from $\text{col}(W)$.

Stochastic Expansion for $\hat{\delta}$

Let

$$\Gamma_{NMT} = \frac{1}{NMT} X' M_{\mathcal{T}} X, \quad \mathcal{Z}_{NMT} = \frac{1}{\sqrt{NMT}} X' M_{\mathcal{T}} e.$$

The least squares estimator $\hat{\delta}$ satisfies

$$\Gamma_{NMT} \sqrt{NMT} (\hat{\delta} - \delta) = \mathcal{Z}_{NMT} - \mathcal{B}_1 - \mathcal{B}_2 - \mathcal{B}_3 + O_p \left(\sqrt{\frac{G + M + T}{NMT}} + \frac{(G + M + T)^{3/2}}{NMT} \right),$$

where

$$\mathcal{B}_1 = \frac{1}{\sqrt{NMT}} X' P_{\mathcal{T}} Q \text{vec} \left(\bar{U} (\bar{U}' \bar{U})^{-1} ((V * W)' (V * W))^{-1} (V * W)' \Omega_1^G M_{V * W} \right)$$

$$\mathcal{B}_2 = \frac{1}{\sqrt{NMT}} X' P_{\mathcal{T}} Q K_{1,2}^G \text{vec} \left(V (V' V)^{-1} ((\bar{U} * W)' (\bar{U} * W))^{-1} (\bar{U} * W)' \Omega_2^G M_{\bar{U} * W} \right)$$

$$\mathcal{B}_3 = \frac{1}{\sqrt{NMT}} X' P_{\mathcal{T}} Q K_{1,3}^G \text{vec} \left(W (W' W)^{-1} ((\bar{U} * V)' (\bar{U} * V))^{-1} (\bar{U} * V)' \Omega_3^G M_{\bar{U} * V} \right).$$

Here, $\bar{U} = (D' D)^{1/2} U$ and $\Omega_m^G = \mathbb{E}[(E_{(m)}^G)' E_{(m)}^G \mid \mathcal{F}]$ for $m = 1, 2, 3$.

Stochastic Expansion for $\hat{\delta}$

The remainder terms in the stochastic expansion has the bound

$$R_{NMT} = O_p \left(\sqrt{\frac{G + M + T}{NMT}} + \frac{(G + M + T)^{3/2}}{NMT} \right).$$

This allows one dimension to remain fixed, as long as the other two dimensions diverge at the same rate. Let any two of N, M, T have the same order as $L \rightarrow \infty$ while the remaining dimension is $O(L)$. Since $G \leq N$,

$$G + M + T \leq N + M + T = O(L), \quad NMT \gtrsim L^2.$$

Consequently,

$$\frac{G + M + T}{NMT} = O(L^{-1}) = o_p(1), \quad \frac{(G + M + T)^{3/2}}{NMT} = O(L^{-1/2}) = o_p(1).$$

This is regardless of which two dimensions diverge.

Comparison with Bai (2009)

Mode-specific bias terms are bounded in norm by

$$\text{Grouped Three-Way Separable} = O_p \left(\frac{G + M}{\sqrt{NMT}} \right) + O_p \left(\frac{T}{\sqrt{NMT}} \right)$$

$$\text{Flattened Two-Way Separable} = O_p \left(\frac{NM}{\sqrt{NMT}} \right) + O_p \left(\frac{T}{\sqrt{NMT}} \right)$$

- ▶ In the three-way separable model, bias correction as in Hahn and Kuersteiner (2002) or Bai (2009) is unnecessary when $N \asymp M \asymp T \rightarrow \infty$ in as long as $G \leq N$.
- ▶ In the two-way separable model, rate conditions cannot make both $O_p \left(\frac{NM}{\sqrt{NMT}} \right) = O_p \left(\sqrt{\frac{NM}{T}} \right)$ and $O_p \left(\frac{T}{\sqrt{NMT}} \right) = O_p \left(\sqrt{\frac{T}{NM}} \right)$ asymptotically negligible. Bai (2009) uses the rate condition $\frac{NM}{T} \rightarrow c \in (0, \infty)$.

Simulations

Model:

$$y_{ijt} = x'_{ijt}\delta + \kappa + \mu_{g(i)} + \nu_j + \xi_t + \pi_{ijt} + \varepsilon_{ijt}, \quad \pi_{ijt} = \sum_{k=1}^{\ell} u_{g(i),k} v_{jk} w_{tk}.$$

- ▶ $p = 2, \ell = 2, \delta = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \kappa = 1$
- ▶ The fixed effects (μ_g, ν_j, ξ_t) and $(u_{g,k}, v_{jk}, w_{tk})$ all drawn from $\mathcal{N}(0, 1)$.
- ▶ For $r = 1, 2$,

$$x_{ijt,r} = 1 + \pi_{ijt} + \eta_{ijt,r}, \quad \eta_{ijt,r} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

▶

$$\varepsilon_{ijt} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

Estimation Models

Given the estimation model,

$$y_{ijt} = x'_{ijt}\delta + \kappa + \alpha_{ij} + \xi_t + \sum_{k=1}^{\ell} \lambda_{ij,k} f_{tk} + e_{ijt},$$

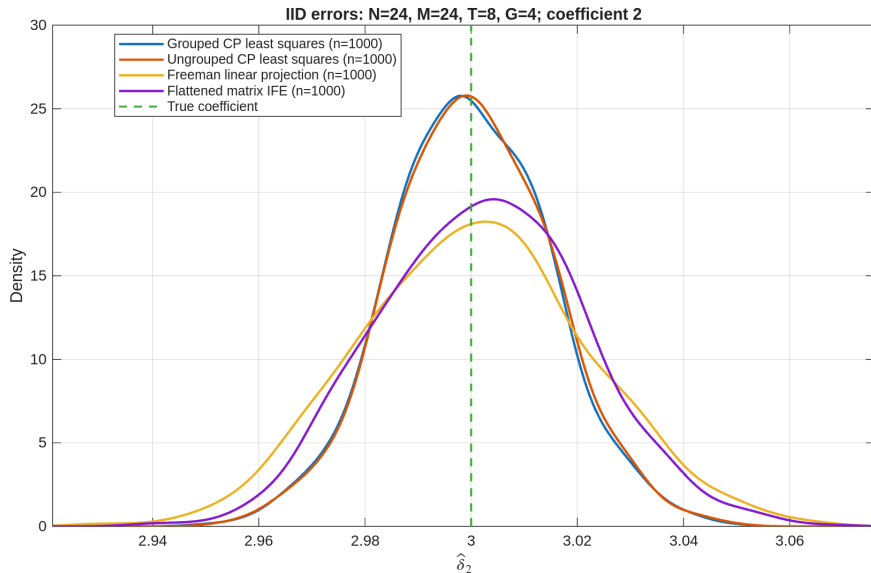
we consider two restrictions

Separability, (R_1) : $\alpha_{ij} = \mu_i + \nu_j$, $\lambda_{ij,k} = a_{ik}b_{jk}$, Grouping, (R_2) : $\mu_i = \mu_{g(i)}$, $a_{ik} = a_{g(i),k}$.

Methods	FE structure	Interactive parameters	Estimation
1	Grouped three-way (R_1, R_2)	$(G + M + T)\ell$	Joint LS
2	Ungrouped three-way (R_1)	$(N + M + T)\ell$	Joint LS
3	Ungrouped three-way (R_1)	$(N + M + T)\ell$	Freeman (2026)
4	Flattened two-way	$(NM + T)\ell$	Joint LS

All four methods are correctly specified; differences arise from fixed effects parameter parsimony and estimation method.

Distribution of $\hat{\delta}_2$



Simulations

Model:

$$y_{ijt} = x'_{ijt}\delta + \kappa + \mu_{g(i)} + \nu_j + \xi_t + \pi_{ijt} + \varepsilon_{ijt}, \quad \pi_{ijt} = \sum_{k=1}^{\ell} u_{g(i),k} v_{jk} w_{tk}.$$

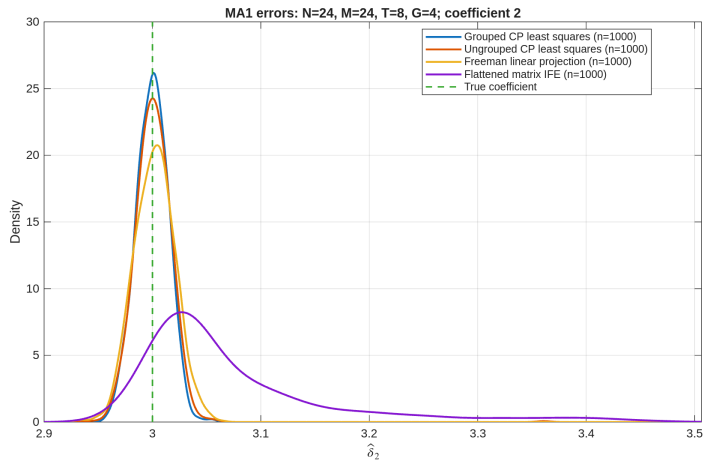
- ▶ $p = 2, \ell = 2, \delta = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \kappa = 1$
- ▶ The fixed effects (μ_g, ν_j, ξ_t) and $(u_{g,k}, v_{jk}, w_{tk})$ all drawn from $\mathcal{N}(0, 1)$.
- ▶ For $r = 1, 2$,

$$x_{ijt,r} = 1 + \pi_{ijt} + \eta_{ijt,r}, \quad \eta_{ijt,r} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

▶

$$\varepsilon_{ijt} = \frac{\tilde{\varepsilon}_{ijt} + \tilde{\varepsilon}_{ij,t-1}}{\sqrt{2}}, \quad \tilde{\varepsilon}_{ijt} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

Distribution of $\hat{\delta}_2$



Data: National Center for Teacher Effectiveness

We use matched student–teacher records from the National Center for Teacher Effectiveness (NCTE) Main Study, distributed as ICPSR 36095. Our analytic sample contains:

- ▶ 11,390 student–year observations in grades 4 and 5, with approximately 20 students per classroom.
- ▶ 316 fourth- and fifth-grade mathematics teachers from four public school districts on the U.S. East Coast.
- ▶ Three academic years, from 2010–11 through 2012–13.
- ▶ In 2012–13, a subset of participating teachers was randomly assigned to preformed classroom rosters within schools and grades.

The records contain standardized mathematics test scores, prior achievement, and student demographic characteristics.

Empirical Specification

We estimate

$$y_{ijt} = x'_{ijt}\delta + \nu_j + \sum_{k=1}^2 u_{g(i),k} v_{jk} w_{tk} + \varepsilon_{ijt}, \quad g(i) \in \{1, \dots, 5\},$$

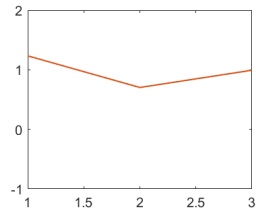
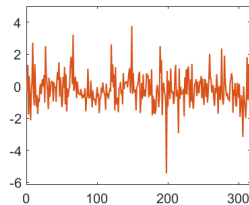
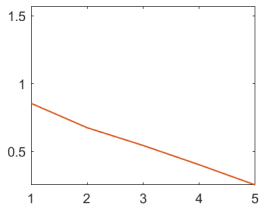
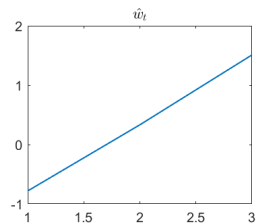
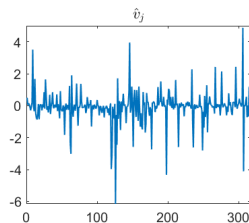
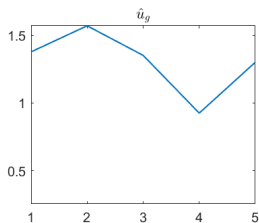
where $g(i)$ is the student's prior-achievement quintile and x_{ijt} includes prior-year test score and dummies for

- ▶ grade
- ▶ gender
- ▶ race/ethnicity
- ▶ special education
- ▶ English learner
- ▶ free- or reduced-price lunch status.

A causal value-added interpretation requires teacher assignment to be conditionally independent of unobserved determinants of achievement.

Estimated Teacher Components

The estimated teacher components are given by



Averages of the Fitted Teacher Component

For achievement group g , teacher j , and period t , let $\tau_{gjt} = \nu_j + \sum_{k=1}^{\ell} u_{gk} v_{jk} w_{tk}$ denote the fitted teacher component. Let

$$\begin{aligned}\mathcal{O} &= \{(j, t) : \text{teacher } j \text{ is observed in period } t\}, \\ \mathcal{J}_t &= \{j : (j, t) \in \mathcal{O}\}, \quad J_t = |\mathcal{J}_t|, \\ \mathcal{T}_j &= \{t : (j, t) \in \mathcal{O}\}, \quad T_j = |\mathcal{T}_j|.\end{aligned}$$

Let $\omega_g \geq 0$ denote the reference weight for group g , with $\sum_{g=1}^G \omega_g = 1$. Define

$$\begin{aligned}\bar{\tau}_{jt} &= \sum_{g=1}^G \omega_g \tau_{gjt}, & \bar{\tau}_{gt} &= \frac{1}{J_t} \sum_{j \in \mathcal{J}_t} \tau_{gjt}, \\ \bar{\tau}_t &= \frac{1}{J_t} \sum_{j \in \mathcal{J}_t} \bar{\tau}_{jt} = \sum_{g=1}^G \omega_g \bar{\tau}_{gt}, & \bar{\tau} &= \frac{1}{|\mathcal{O}|} \sum_{t=1}^T J_t \bar{\tau}_t.\end{aligned}$$

For every active teacher–year cell, τ_{gjt} is evaluated for all $g = 1, \dots, G$, whether or not group g is observed in that cell.

Relative Teacher Value Added: Level and Matching Effects

Relative teacher value added of the j -th teacher for group g in period t is

$$VA_{gjt} = \tau_{gjt} - \bar{\tau}_{gt} = L_{jt} + M_{gjt},$$

where

$$L_{jt} = \bar{\tau}_{jt} - \bar{\tau}_t,$$

$$M_{gjt} = \tau_{gjt} - \bar{\tau}_{jt} - \bar{\tau}_{gt} + \bar{\tau}_t.$$

- ▶ L_{jt} measures teacher j 's general effectiveness relative to the average active teacher in period t .
- ▶ M_{gjt} measures the teacher's comparative advantage for achievement group g .

Relative Teacher Value Added: Level and Matching Effects

By construction,

$$\sum_{g=1}^G \omega_g M_{gjt} = 0 \quad \text{for every } (j, t) \in \mathcal{O},$$
$$\frac{1}{J_t} \sum_{j \in \mathcal{J}_t} M_{gjt} = 0 \quad \text{for every } (g, t).$$

These conditions capture the economic meaning of matching:

- ▶ Teacher effect common to student-group is not matching. (For example, if a teacher is equally effective in teaching different student-groups, this is captured in the level effect.)
- ▶ Group effect common to teacher is not matching. (For example, if all teachers are equally more effective in teaching high-prior achievement groups, this is not a matching effect.)
- ▶ Only the teacher-by-group interaction remains.

Complete Decomposition of the Fitted Teacher Component

The fitted teacher component admits the orthogonal decomposition

$$\begin{aligned}
 \tau_{gjt} - \bar{\tau} &= \underbrace{\bar{\tau}_t - \bar{\tau}}_{\text{common year component}} + \underbrace{\bar{\tau}_{gt} - \bar{\tau}_t}_{\text{common group level component}} \\
 &+ \underbrace{\bar{\tau}_{jt} - \bar{\tau}_t}_{\text{teacher level effect}} + \underbrace{\tau_{gjt} - \bar{\tau}_{jt} - \bar{\tau}_{gt} + \bar{\tau}_t}_{\text{teacher-group matching effect}} \\
 &= T_t + G_{gt} + L_{jt} + M_{gjt}.
 \end{aligned}$$

Under the common teacher-year and group weights, the sample variance can be decomposed as

$$\hat{V}_\tau = \hat{V}_T + \hat{V}_G + \hat{V}_L + \hat{V}_M.$$

The estimated variance components are

$$(\hat{V}_T, \hat{V}_G, \hat{V}_L, \hat{V}_M) = (0.000, 0.003, 0.097, 0.036).$$

Consequently,

$$\hat{V}_\tau = 0.136, \quad \hat{V}_{VA} = \hat{V}_L + \hat{V}_M = 0.133.$$

The common year and group-level components do not vary across teachers and are therefore excluded from relative teacher value added.

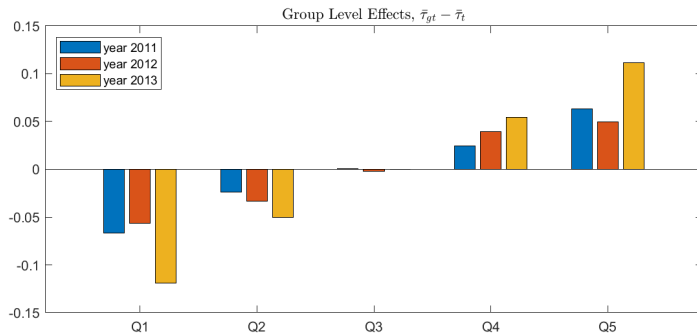
Level and Matching Shares of Relative Teacher Value Added

	Grouped Three-Way		Ahn, Aucejo, and James (2025): Math	
	Variance	Share	Variance	Share
Relative teacher value added	0.133	100%	0.056	100%
Level	0.097	73%	0.049	88%
Matching	0.036	27%	0.007	12%

The matching component accounts for approximately 27% of relative teacher-value-added variation in the grouped three-way model, compared with 12% in the math estimates of Ahn, Aucejo, and James (2025).

Notes: The comparison is descriptive. Ahn, Aucejo, and James estimate the population variance of latent teacher coefficients using an empirical-Bayes Gaussian-mixture model, whereas the grouped three-way entries are plug-in second moments of the estimated teacher-effect surface.

Group Level Effect

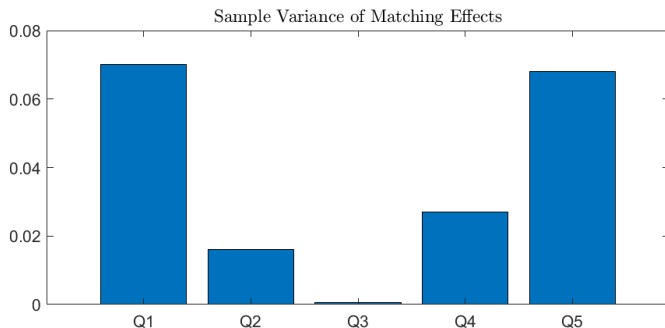


The group-level effect $G_{gt} = \bar{\tau}_{gt} - \bar{\tau}_t$ suggests that the average teacher effects tend to be lower for the prior low achievement groups compared to the high achievement groups. However, variation in the match effect $\hat{V}_M = 0.036$ is much larger than that of the group-level effect $\hat{V}_G = 0.003$.

For Which Groups Does Teacher Assignment Matter Most?

Define the variance of matching effects for achievement group h

$$\hat{V}_{M,h} = \frac{1}{|\mathcal{O}|} \sum_{t=1}^T \sum_{j \in \mathcal{J}_t} \hat{M}_{hjt}^2.$$



The estimates suggest that the matching dispersion is largest for students in the bottom and top prior-achievement quintiles and smallest for students in the middle quintile.

How do teacher rankings change when assigned to the best classroom?

Table 1: Counterfactual gains diagnostics

	Within school	Within district
<i>Panel A: Individually best classroom</i>		
Mean gain	0.0644	0.1499
Median gain	0.0138	0.0885
<i>Panel B: Individual-best matching diagnostics</i>		
Already in a best classroom	206 (36.40%)	26 (4.59%)
Exceed original 5th-percentile cutoff	8 (27.59%)	16 (55.17%)
Escape after re-ranking	5 (17.24%)	8 (27.59%)

Each teacher–year observation is evaluated in its individually best eligible classroom.

How does teacher effectiveness change under optimal feasible allocation?

Table 2: Counterfactual matching diagnostics

	Within school	Within district
<i>Panel C: Feasible total-value-added optimum</i>		
Total value-added gain	0.0538	0.1352
Matching contribution	0.0429	0.0916
Level contribution	0.0109	0.0436
<i>Panel D: Feasible matching-only optimum</i>		
Total value-added gain	0.0422	0.0939
Matching contribution	0.0476	0.0989
Level contribution	-0.0054	-0.0050

Eligible teachers are reassigned one-to-one to existing classrooms within the same grade, year, and school or district, holding classroom rosters and estimated teacher-effect functions fixed. An increase of 0.0538 SD is equivalent to 31.7 additional school days, Aucejo and Romano (2016).

For Which Groups Does Teacher Assignment Matter Most?

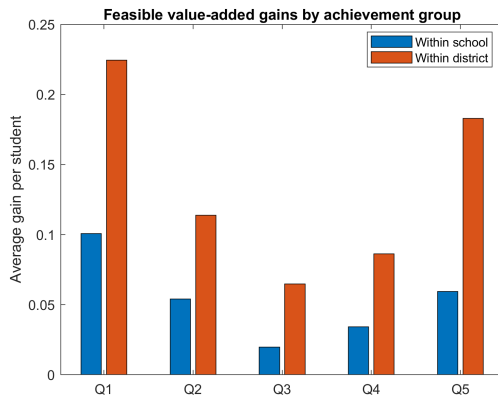


Figure 1: Under feasible allocation, gains are largest in the lowest-achievement group, followed by the highest-achievement group, and smallest in the middle group.

Comparison with Randomized-Roster Benchmark

Let T correspond to 2012–2013, and let $\mathcal{R}$ contain the student observations belonging to the randomized classroom rosters. Using observations $(i, j, T) \in \mathcal{R}$, we estimate

$$y_{ijT} = x'_{ijT}\delta + \tau_{g(i),j} + \varepsilon_{ijT}$$

without imposing the separable fixed effect structure.

Table 3: All-teacher correlations with the randomized-roster benchmark

<i>Panel A: Pooled correlations and sample sizes</i>			
Component	$m = 1$	$m = 2$	$m = 3$
Group level	0.923	0.938	0.963
Teacher level	0.755	0.760	0.809
Matching	0.234	0.415	0.585
Relative value added ($L + M$)	0.508	0.647	0.718
Paired teachers	71	70	60
Paired cells	284	215	162
Observations in paired cells	968	899	793

Conclusion

- ▶ Grouped three-way interactive fixed effects provide a parsimonious model of heterogeneous teacher effectiveness and reduce the order of incidental-parameter bias.
- ▶ Student-teacher matching accounts for **27% of fitted relative teacher-value-added variation** under the reference weights.
- ▶ Feasible within-school teacher reallocation yields a **model-implied average gain of 0.054 standard deviations per student**, holding classroom rosters and estimated teacher-effect functions fixed.

Additional results on factor estimation, additive fixed effects, and unbalanced panels are available at:

Personal Webpage: <https://sangmyung3.github.io>

APPENDIX

Main References

Estimating Teacher Effects

- ▶ Ahn, Aucejo, and James (2025). The importance of student-teacher matching: A multidimensional value-added approach. *The Review of Economics and Statistics*
- ▶ Bates, Dinerstein, Johnston, and Sorkin (2025). Teacher labor market policy and the theory of the second best. *The Quarterly Journal of Economics*
- ▶ Delgado (2025). Disparate teacher effects, comparative advantage, and match quality. *Economics of Education Review*
- ▶ Biasi, Fu, and Stromme (2026). Equilibrium in the market for public school teachers: District wage strategies and teacher comparative advantage.
- ▶ Chetty, Friedman, and Rockoff (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*
- ▶ Aucejo and Romano (2016) Assessing the effect of school days and absences on test score performance, *Economics of Education Review*

Main References

Panel Regression with Interactive Fixed Effects

- ▶ [Bai](#) (2009). Panel data models with interactive fixed effects. *Econometrica*
- ▶ [Moon and Weidner](#) (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*
- ▶ [Freeman](#) (2026). Linear multidimensional regression with interactive fixed-effects.

Tensor Factor Models with Canonical Polyadic Decomposition

- ▶ [Chen, Han, and Yu](#) (2026). Estimation and inference for CP tensor factor models. *Journal of Econometrics*
- ▶ [Chang, Ha, and Park](#) (2026). A general factor model with common components identified by tensor structure.

Comparison with Bai (2009)

By considering (i, j) as a single cross-sectional index, we may model

$$y_{ijt} = x'_{ijt}\delta + \sum_{k=1}^{\ell} \lambda_{ij,k} f_{tk} + \varepsilon_{ijt}.$$

We therefore consider a restricted version of Bai (2009) where

$$\lambda_{ij,k} = a_{ik}v_{jk}, \quad a_{ik} = u_{g(i),k}.$$

The separability in i and j reduces the number of fixed effects from $(NM + T)\ell$ to $(N + M + T)\ell$. The grouping restriction further reduces this to $(G + M + T)\ell$.

Comparison with Freeman (2026)

Freeman's linear-projection specification uses preliminary matrix interactive fixed-effect regressions to estimate the three loading spaces. Let $\hat{Q}_F$ be the projection matrix given by this preliminary regression, defined as

$$\hat{Q}_F = M_{D\hat{U}} \otimes M_{\hat{V}} \otimes M_{\hat{W}}.$$

The two slope estimators have the representations

$$\begin{aligned}\hat{\delta}_F &= \left(X' \hat{Q}_F X\right)^{-1} X' \hat{Q}_F y, \\ \hat{\delta}_{LS} &= \left(X' M_{\hat{\mathcal{T}}} X\right)^{-1} X' M_{\hat{\mathcal{T}}} y.\end{aligned}$$

At common true factors, let $Q = M_{DU} \otimes M_V \otimes M_W$. Then

$$Q \preceq M_{\mathcal{T}} \implies X' Q X \preceq X' M_{\mathcal{T}} X.$$

Joint least squares can retain identifying regressor variation that the product projection discards.