Online Appendix to "Leveraging Lotteries for School Value-Added: Testing and Estimation"

Joshua D. Angrist, MIT and NBER
Peter D. Hull, MIT
Parag A. Pathak, MIT and NBER
Christopher R. Walters, UC Berkeley and NBER

August 2016

Contents

Appendix A: Data

The administrative data used for this project come from student demographic and attendance information in the Massachusetts Student Information Management System (SIMS), standardized student test scores from the Massachusetts Comprehensive Assessment System (MCAS) database, Boston charter school admission lottery records, and information from the centralized BPS student assignment system. We describe each data source and our cleaning and matching process in detail below; the construction of our main analysis file closely follows that of previous studies, in particular Abdulkadiroğlu et al. (2011).

A.1 Student enrollment, demographics, and test scores

The Massachusetts SIMS contains snapshots of all students in a public school in Massachusetts in October and at the end of each school year. These records contain demographic information on students, their current schools, their residence, and their attendance. We work with SIMS files for the 2005-2006 through the 2013-2014 school years and limit the sample to students enrolled in a Boston school over this period. Schools are classified as charters by the Massachusetts Department of Elementary and Secondary Education website (`http://www.profiles.doe.mass.edu`), and as pilots by the Boston pilot school network website (`http://www.ccebos.org/pilotschools/schools.html`). All remaining Boston schools are considered traditional public schools for the purposes of this study.

Enrollment in the SIMS is grade-specific. When a student repeats grades, we retain the first school a student attended in that grade. We then record students attending multiple schools in a given school year as enrolled in the school for which the attendance duration is longest, with duration ties broken randomly. This results in a unique student panel across grades; for the purposes of this study we restrict focus to sixth grade students enrolled from 2006-2007 to 2013-2014, using their fifth grade information for baseline controls. These controls include indicators for student race (Hispanic, black, white, Asian, and other race), sex, free- or reduced-price lunch eligibility, special education status, and English-language learner status, as well as counts of the number of days a student was suspended or truant over the school year. Suspension data are unavailable in the SIMS starting in the 2012-2013 school year; we include an indicator for students missing this baseline information whenever suspensions are used.

Our primary outcomes for measuring school value-added are sixth, seventh and eighth grade standardized test scores from the Massachusetts Comprehensive Assessment System (MCAS) database. MCAS math and ELA scores are standardized by grade and year to a combined BPS and Boston charter school reference population. MCAS outcome scores are merged to SIMS data via a state-assigned unique student identifier. We also merge baseline (fifth grade) MCAS scores for each student in our sample (fifth grade MCAS information is available starting in the 2005-2006 school year).

A.2 Charter school lotteries

The lottery analysis uses records for five of the six Boston middle school charters with sixth grade admission for the 2006-2007 through the 2013-2014 academic year. These schools are Academy of the Pacific Rim, Boston Preparatory, MATCH Charter Public Middle School, Roxbury Preparatory, and

UP Academy Boston. The remaining school, Smith Leadership Academy, has declined to participate in our studies. For each school and each oversubscribed year we obtain a list of names of students eligible for entry by lottery, as well as information on whether each student was offered a seat on lottery night. Students are marked as ineligible if they submit an incomplete or late application; we also exclude students with a sibling currently enrolled in the school, as they are guaranteed admission. For UP Boston, which is an in-district charter school, students applying from outside of BPS are placed in a lower lottery priority group.

A student is coded as receiving a charter admission offer if he or she is offered a seat on lottery night. These offers are randomly assigned within strata defined by school, application year, and, in the case of UP Boston, BPS priority group. Students are retained the first year they apply to a charter school. We match the set of charter offers and randomization strata to state data by student name, grade, and application year; 97% of charter lottery applicants are successfully matched.

A.3 The BPS mechanism

We obtained a complete record of student-submitted preferences, school priorities, random tie-breaking sequence numbers, and assignments from the BPS deferred-acceptance mechanism, for seats in fall 2006 through fall 2013. For each application year, students are classified by priority (given by whether the applicant has an enrolled sibling and by walk-zone residence, a 1.5 mile radius) at schools that they rank first. Students guaranteed admission by virtue of current enrollment, as well as certain other students with guaranteed or nonstandard priorities, are not subject to random assignment (see Abdulkadiroglu et al. [2006] for a complete description of priorities in BPS). We construct indicators for whether an applying student was offered a seat; these offers are randomly assigned within strata defined by school, application year, and priority. We drop all schools with fewer than 50 students subject to conditionally-random admission, and match offers and randomization strata to state data via a BPS unique student identifier. Students are retained the first year they enter the BPS mechanism for sixth grade entry.

A.4 Sample Selection

We restrict attention to Boston public schools with at least 25 sixth grade students enrolled in each year of operation from 2006-2007 to 2013-2014. In our merged analysis file this leaves 51 schools (see Table I). Students enrolled at these schools are retained if they were enrolled in Boston in both fifth and sixth grade, if their baseline demographic, attendance, and test score information is available, and if we observe their sixth grade MCAS test scores. These restrictions leave a total of 27,864 Boston students, summarized in detail in Table II. Of these, 8,718 students are subject to quasi-experimental variation in sixth grade admission at 28 schools, either from a charter school lottery or from assignment by the BPS mechanism.

Appendix B: Econometric Methods

B.1 Comparison of Compliance Groups

Figure III compares average predicted value-added for lottery compliers, always-takers and never-takers. Predicted value-added comes from a version of equation (5) that interacts the school dummies D_{ij} with race, gender, subsidized lunch, special education, English language learner status, and baseline score terciles. For a student with covariates X_i, this interacted VAM yields an estimate of a covariate-specific value-added parameter $\alpha_j(X_i)$ for each school j.

The arguments in Abadie (2003) imply that $E[\alpha_j(X_i)\kappa^a_{ij}]/E[\kappa^a_{ij}]$ equals the average value-added of school j for always-takers in lottery j, where $\kappa^a_{ij} = D_{ij}(1-Z_{ij})/(1-E[Z_{ij}|C_{ij}])$. Averages of value-added for never-taker and compliers are similarly given by $E[\alpha_j(X_i)\kappa^n_{ij}]/E[\kappa^n_{ij}]$ and $E[\alpha_j(X_i)(1-\kappa^a_{ij}-\kappa^n_{ij})]/E[1-\kappa^a_{ij}-\kappa^n_{ij}]$, for $\kappa^n_{ij} = (1-D_{ij})Z_{ij}/E[Z_{ij}|C_{ij}]$. We construct points in Figure III based on the sample analogues of these quantities, using a saturated model for lottery strata to estimate $E[Z_{ij}|C_{ij}]$.

To adjust for first-step error in the estimation of $\alpha_j(X_i)$ and $E[Z_{ij}|C_{ij}]$, inference in Figure III uses a Bayesian bootstrap procedure (Rubin 1981). The Bayesian bootstrap smooths bootstrap samples by reweighting rather than resampling observations, preventing the omission of small lottery strata that would occasionally be dropped in a standard nonparametric bootstrap. The Bayesian bootstrap used here is implemented by drawing vectors of Dirichlet$(1, ..., 1)$ weights, then re-estimating the interacted VAM and recomputing predicted value-added for compliers, always-takers and never-takers, weighting all moments with the Dirichlet weights. Inference for differences in means between compliance groups are based on the bootstrap covariance matrix of these differences across trials.

B.2 Simulated Minimum Distance

We estimate Bayesian hyperparameters via simulated minimum distance (SMD). The vector of parameters to be estimated is

$$\theta = \left(\alpha_0, \beta_0, \beta_Q, \delta_0, \xi_0, \Sigma, \sigma^2_\nu\right)'.$$

These parameters are estimated by fitting means, variances, and covariances of OLS value-added, lottery reduced form, and first stage estimates. The complete vector of observed estimates is

$$\hat{\Omega} = (\hat{\alpha}_1, ..., \hat{\alpha}_J, \hat{\rho}_1, ..., \hat{\rho}_L, \hat{\pi}_{11}, ..., \hat{\pi}_{L1}, ..., \hat{\pi}_{LJ})'.$$

Let $\Omega = (\alpha_1, ..., \pi_{LJ})'$ denote the probability limits of these estimates. Assume that the sampling distribution of $\hat{\Omega}$ is well approximated by asymptotic theory, so that

$$\hat{\Omega} \sim N\left(\Omega, V_e\right),$$

where V_e is a covariance matrix derived from conventional asymptotics. This requires within-school and within-lottery samples to be large enough for asymptotic approximations to be accurate. Under this assumption and the distributional assumptions in equations (12) through (15), values of Ω and $\hat{\Omega}$ can be simulated for any value of θ. We use this procedure to generate simulated data sets, and estimate θ by minimizing the distance between simulated and observed moments.

Our estimation procedure targets the following first moments:

$$\hat{m}_1 = \frac{1}{J}\sum_j \hat{\alpha}_j,$$

$$\hat{m}_2 = \frac{1}{L}\sum_j Q_j\hat{\alpha}_j,$$

$$\hat{m}_3 = \frac{1}{L}\sum_\ell \hat{\rho}_\ell,$$

$$\hat{m}_4 = \frac{1}{L}\sum_\ell \left(\frac{\hat{\rho}_\ell}{\hat{\pi}_{\ell\ell}}\right)$$

$$\hat{m}_5 = \frac{1}{L}\sum_\ell \hat{\psi}_\ell$$

$$\hat{m}_6 = \frac{1}{L}\sum_\ell \hat{\pi}_{\ell\ell},$$

$$\hat{m}_7 = -\frac{1}{J}\sum_j \frac{1}{L-Q_j}\sum_{\ell\neq j} \hat{\pi}_{\ell j},$$

$$\hat{m}_8 = -\frac{1}{J}\sum_\ell \frac{1}{L-Q_j}\sum_{j\neq \ell} \frac{\hat{\pi}_{\ell j}}{\hat{\pi}_{\ell\ell}},$$

$$\hat{m}_9 = \frac{1}{L}\sum_\ell \left[\frac{(\hat{\pi}_{\ell\ell})^2}{\sum_k(\hat{\pi}_{kk})^2}\right]\cdot\left(\frac{\hat{\rho}_\ell}{\hat{\psi}_\ell}\right).$$

$\hat{m}_1$ is the mean OLS coefficient, which provides information about $\beta_0 + b_0$, the sum of mean value-added and mean bias. $\hat{m}_2$ is the mean OLS coefficient among lottery schools, which helps to identify β_Q, the difference in value-added between lottery and non-lottery schools. $\hat{m}_3$ and $\hat{m}_4$ are the mean reduced form and IV coefficients, which provide information about β_0. $\hat{m}_5$ is the mean of a "pseudo-reduced form" prediction that uses OLS value-added estimates, given by $\hat{\psi}_\ell = \sum_j \hat{\pi}_{\ell j}\hat{\alpha}_j$. $\hat{m}_6$ is the mean first stage across lotteries, which can be used to estimate δ_0. $\hat{m}_7$ is the average fallback probability across lotteries, and $\hat{m}_8$ is the average ratio of this probability to the first stage, which gives the share of compliers drawn from included schools. These two moments help to estimate ξ_0, the mean fallback utility for included schools relative to the omitted school. $\hat{m}_9$ is the average ratio of the lottery reduced form to the pseudo-reduced form (the forecast coefficient). We weight this average by the squared lottery first stage to avoid unstable ratios caused by small first stages. This moment yields information about the variance of b_j, the bias in conventional value-added estimates, along with the correlation between β_j and b_j.

The next seven moments are variances of parameter estimates:

$$\hat{m}_{10} = \frac{1}{J}\sum_j (\hat{\alpha}_j - \bar{\alpha})^2,$$

$$\hat{m}_{11} = \frac{1}{L}\sum_\ell (\hat{\rho}_\ell - \bar{\rho})^2,$$

$$\hat{m}_{12} = \frac{1}{L}\sum_\ell (\hat{\psi}_\ell - \bar{\psi})^2,$$

$$\hat{m}_{13} = \frac{1}{L}\sum_\ell (\hat{\pi}_{\ell\ell} - \bar{\pi}_{own})^2,$$

$$\hat{m}_{14} = \frac{1}{J}\sum_j \left[\left(\frac{1}{L-Q_j}\sum_{\ell\neq j}\hat{\pi}_{\ell j}\right) - \bar{\pi}_{other}\right]^2,$$

$$\hat{m}_{15} = \frac{1}{J}\sum_j \left[\left(\frac{1}{L-Q_j}\sum_{\ell\neq j}\frac{\hat{\pi}_{\ell j}}{\hat{\pi}_{\ell\ell}}\right) - \bar{s}_{other}\right]^2,$$

$$\hat{m}_{16} = \frac{1}{J(L-1)} \sum_j \sum_{\ell \neq j} \left(\hat{\pi}_{\ell j} - \bar{\pi}_j\right)^2.$$

Here $\bar{\alpha}$ indicates the sample average of the α_j, and similarly for other variables. $\hat{m}_{10}$ is the variance of conventional value-added estimates across schools, which depends on the variances of value-added and bias as well as their covariance. $\hat{m}_{11}$ and $\hat{m}_{12}$ are variances of the lottery reduced form and predicted reduced form, which contain additional information about the joint distribution of value-added and bias. $\hat{m}_{13}$ is the variance of the first stage across lotteries, which helps to identify the variance of δ_j. $\hat{m}_{14}$ computes the mean share of students drawn from each school across lotteries, then takes the variance of this mean share across schools. This is the between-school variance in fallback probabilities. $\hat{m}_{15}$ is the variance of the mean share of compliers drawn from a particular school; $\bar{s}_{other}$ is the mean of this variable. These two moments yield information about the variances of ξ_j and $\nu_{\ell j}$, which govern heterogeneity in fallback probabilities. $\hat{m}_{16}$ computes the variance of fallback shares across lotteries at every school, then averages across schools. This is the average within-school variance in fallback probabilities. This moment helps to separate the variance of ξ_j, the school-specific mean fallback utility, from σ_ν^2, the variance of idiosyncratic school-by-lottery utility shocks.

Finally, we match six covariances:

$$\hat{m}_{17} = \frac{1}{L} \sum_\ell \left(\hat{\rho}_\ell - \bar{\rho}\right)\left(\hat{\alpha}_\ell - \bar{\alpha}\right),$$

$$\hat{m}_{18} = \frac{1}{L} \sum_\ell \left(\hat{\rho}_\ell - \bar{\rho}\right)\left(\hat{\psi}_\ell - \bar{\psi}\right),$$

$$\hat{m}_{19} = \frac{1}{L} \sum_\ell \left(\hat{\rho}_\ell - \bar{\rho}\right)\left(\hat{\pi}_{\ell\ell} - \bar{\pi}_{own}\right),$$

$$\hat{m}_{20} = \frac{1}{L} \sum_\ell \left(\hat{\alpha}_\ell - \bar{\alpha}\right)\left(\hat{\pi}_{\ell\ell} - \bar{\pi}_{own}\right),$$

$$\hat{m}_{21} = \frac{1}{L} \sum_\ell \left(\hat{\rho}_\ell - \bar{\rho}\right)\left[\left(\frac{1}{L-1} \sum_{k \neq \ell} \hat{\pi}_{k\ell}\right) - \bar{\pi}_{other}\right],$$

$$\hat{m}_{22} = \frac{1}{J} \sum_j \left(\hat{\alpha}_j - \bar{\alpha}\right)\left[\left(\frac{1}{L-Q_j} \sum_{\ell \neq j} \hat{\pi}_{\ell j}\right) - \bar{\pi}_{other}\right],$$

$$\hat{m}_{23} = \frac{1}{L} \sum_\ell \left(\hat{\pi}_{\ell\ell} - \bar{\pi}_{own}\right)\left[\left(\frac{1}{L-1} \sum_{k \neq \ell} \hat{\pi}_{k\ell}\right) - \bar{\pi}_{other}\right].$$

$\hat{m}_{17}$ and $\hat{m}_{18}$ are covariances of the reduced form with conventional value-added and the pseudo-reduced form, which help to identify variation in bias, as well as the covariance between bias and value-added. $\hat{m}_{19}$ is the covariance between reduced forms and first stages, which is informative about the covariance between β_j and δ_j. $\hat{m}_{20}$ is the covariance of conventional value-added and the first stage, which helps to identify the covariance between b_j and δ_j. $\hat{m}_{21}$ is the covariance of the reduced form and average fallback probability, which helps to identify the covariance of β_j and ξ_j. $\hat{m}_{22}$ is the covariance of OLS value-added with the average fallback probability, which depends on the covariance between b_j and ξ_j. $\hat{m}_{23}$ is the covariance of a school's first stage and average fallback probability, which provides information about the covariance of ξ_j and δ_j.

There are 16 elements of θ and 23 moments, so the model has seven overidentifying restrictions. Models that include charter and pilot school effects add sector-specific values of $\hat{m}_1$, $\hat{m}_3$, $\hat{m}_4$, $\hat{m}_5$, $\hat{m}_6$, $\hat{m}_7$ and $\hat{m}_8$, yielding 24 parameters and 37 moments. Let $\hat{m}$ be the vector of all observed moments, and let $\tilde{m}(\theta)$ be the corresponding vector of simulated predictions. The simulated minimum distance estimator with weighting matrix A is

$$\hat{\theta}_{SMD}(A) = \arg\min_{\theta} \; J\left(\hat{m} - \tilde{m}(\theta)\right)' A \left(\hat{m} - \tilde{m}(\theta)\right).$$

The set of simulation draws used to construct $\tilde{m}(\theta)$ is held constant throughout the optimization. For each evaluation of the objective function the vector θ is used to transform these draws to have the appropriate distributions.

We produce a first-step estimate of θ with an identity weighting matrix, then use this estimate to compute a model-based covariance matrix by simulation. Altonji and Segal (1996) show that estimation error in the weighting matrix can generate finite-sample bias in two-step optimal minimum distance estimates. This bias is caused by correlation between the observations used to compute the moment conditions and those used to construct the weighting matrix. We therefore compute the model-based weighting matrix using a second set of simulation draws independent of the draws used to compute the moments. The weighting matrix is given by

$$\hat{A} = \left[J \cdot \frac{1}{R} \sum\nolimits_r \left(\tilde{m}^r\left(\hat{\theta}_{SMD}(I)\right) - \bar{m}\right)\left(\tilde{m}^r\left(\hat{\theta}_{SMD}(I)\right) - \bar{m}\right)'\right]^{-1},$$

where r indexes a second independent set of $R = 1,000$ simulation draws and $\bar{m}$ is the mean of the simulated moments. An efficient two-step estimate is given by $\hat{\theta}_{SMD}\left(\hat{A}\right)$.

Under standard regularity conditions the minimized SMD criterion function follows a χ^2 distribution (Sargan, 1958; Hansen, 1982):

$$J\left(\hat{m} - \tilde{m}\left(\hat{\theta}_{SMD}(\hat{A})\right)\right)' \hat{A}\left(\hat{m} - \tilde{m}\left(\hat{\theta}_{SMD}(\hat{A})\right)\right) \sim \chi^2_q,$$

where q is the difference between the number of moments and the number of parameters to be estimated. Table VI reports this J-statistic.

B.3 Empirical Bayes Posteriors with a Known First Stage

With a know first stage matrix, Π, the posterior distribution for β_j and b_j can be derived analytically. In matrix form, the model can be written

$$\hat{\alpha} = \beta + b + e_{\alpha},$$

$$\hat{\rho} = \Pi\beta + e_{\rho},$$

$$(e'_{\alpha}, e'_{\rho})|\beta, b \sim N(0, V_e),$$

$$(\beta', b')' \sim N\left((\iota'\beta_0, \iota' b_0)', V_{\Theta}\right),$$

where we have set $\beta_Q = 0$. The posterior density for the random coefficients $\Theta = (\beta, b)$ conditional on the observed estimates $\hat{\Omega} = (\hat{\alpha}, \hat{\rho})$ is given by

$$f_{\Theta|\hat{\Omega}}\left(\Theta|\hat{\Omega};\theta\right) = \frac{f_{\hat{\Omega}|\Theta}\left(\hat{\Omega}|\Theta\right) f_{\Theta}\left(\Theta;\theta\right)}{f_{\hat{\Omega}}\left(\hat{\Omega};\theta\right)}.$$

The estimation errors and random coefficients are normally distributed, so we can write

$$-2\log f_{\Theta|\hat{\Omega}}\left(\Theta|\hat{\Omega};\theta\right) = \left((\hat{\alpha}-\beta-b)', (\hat{\rho}-\Pi\beta)'\right)' \begin{bmatrix} v_{\alpha\alpha} & v_{\alpha\rho} \\ v'_{\alpha\rho} & v_{\rho\rho} \end{bmatrix} \begin{pmatrix} \hat{\alpha}-\beta-b \\ \hat{\rho}-\Pi\beta \end{pmatrix}$$

$$+ ((\beta-\beta_0\iota)', (b-b_0\iota)')' \begin{bmatrix} v_{\beta\beta} & v_{\beta b} \\ v'_{\beta b} & v_{bb} \end{bmatrix} \begin{pmatrix} \beta-\beta_0\iota \\ b-b_0\iota \end{pmatrix} + C_1,$$

where $v_{\alpha\alpha}$, $v_{\alpha\rho}$ and $v_{\rho\rho}$ are blocks of V_e^{-1}; $v_{\beta\beta}$, $v_{\beta b}$ and v_{bb} are blocks of V_Θ^{-1}; and C_1 is a constant that does not depend on Θ.

Rearranging this expression yields

$$-2\log f_{\Theta|\hat{\Omega}}\left(\Theta|\hat{\Omega};\theta\right) = \left((\beta-\beta^*)', (b-b^*)'\right)' \begin{bmatrix} v^*_{\beta\beta} & v^*_{\beta b} \\ v^{*\prime}_{\beta b} & v^*_{bb} \end{bmatrix} \begin{pmatrix} \beta-\beta^* \\ b-b^* \end{pmatrix} + C_2, \quad \text{(A1)}$$

where C_2 is another constant. The parameters of this expression are

$$v^*_{\beta\beta} = v_{\alpha\alpha} + \Pi' v'_{\alpha\rho} + v_{\alpha\rho}\Pi + \Pi' v_{\rho\rho}\Pi + v_{\beta\beta},$$

$$v^*_{\beta b} = v_{\alpha\alpha} + \Pi' v'_{\alpha\rho} + v_{\beta b},$$

$$v^*_{bb} = v_{\alpha\alpha} + v_{bb},$$

and

$$\beta^* = W_\alpha(\hat{\alpha} - b_0\iota) + W_\rho\hat{\rho} + (I - W_\alpha - W_\rho\Pi)\beta_0\iota$$

with

$$W_\alpha = B^{-1}((v_{\alpha\alpha}+v_{bb})(v_{\alpha\alpha}+\Pi' v'_{\alpha\rho}+v_{\beta b})^{-1}(v_{\alpha\alpha}+\Pi' v'_{\alpha\rho}) - v_{\alpha\alpha}),$$

$$W_\rho = B^{-1}((v_{\alpha\alpha}+v_{bb})(v_{\alpha\alpha}+\Pi' v'_{\alpha\rho}+v_{\beta b})^{-1}(v_{\alpha\rho}+\Pi' v_{\rho\rho}) - v_{\alpha\rho}),$$

$$B = (v_{\alpha\alpha}+v_{bb})(v_{\alpha\alpha}+\Pi' v'_{\alpha\rho}+v_{\beta b})^{-1}(v_{\alpha\alpha}+\Pi' v'_{\alpha\rho}+v_{\alpha\rho}\Pi+\Pi' v_{\rho\rho}\Pi+v_{\beta\beta}) - (v_{\alpha\alpha}+v_{\alpha\rho}\Pi+v'_{\beta b}).$$

Equation (A1) implies that the posterior for (β, b) is normal:

$$(\beta', b')'|\hat{\alpha},\hat{\rho} \sim N\left((\beta^{*\prime}, b^{*\prime})', V^*\right),$$

with

$$V^* = \begin{bmatrix} v^*_{\beta\beta} & v^*_{\beta b} \\ v^{*\prime}_{\beta b} & v^*_{bb} \end{bmatrix}^{-1}.$$

An empirical Bayes version of the posterior mean β^* is formed by plugging $\hat{\theta}_{SMD}$ and an estimate of V_e into the expressions for W_α and W_ρ.

Section V.C gives three special cases of the posterior mean. The first is when Π is invertible. Equation (18) is obtained by defining $W_\beta = W_\rho\Pi$ and substituting $W_\beta\Pi^{-1}$ for W_ρ in (17). The second special case adds the conditions that $Var(e_\alpha) = 0$ (α_j is known with certainty) and $Var(e_\beta) = \Pi^{-1}Var(e_\rho)\Pi^{-1\prime}$ is diagonal (sampling errors in IV estimates are independent across schools). In this case the only information in the sample about β_j comes from $(\alpha_j, \hat{\beta}_j)$ since β_j is uncorrelated with $\hat{\beta}_k$ and α_k for $k \neq j$. The vector $(\beta_j, \beta_j + b_j, \beta_j + e_j^\beta)$ is jointly normally distributed, so the posterior mean for β_j is the prediction from a linear regression of β_j on α_j and $\hat{\beta}_j$, given by:

$$\beta_j = \kappa_0 + \kappa_\alpha \alpha_j + \kappa_\beta \hat{\beta}_j + v_j.$$

Standard multivariate regression algebra implies the coefficients in this regression are

$$\kappa_\alpha = \frac{Var(\hat{\beta}_j)Cov(\alpha_j, \beta_j) - Cov(\hat{\beta}_j, \alpha_j)Cov(\beta_j, \hat{\beta}_j)}{Var(\alpha_j)Var(\hat{\beta}_j) - Cov(\alpha_j, \hat{\beta}_j)^2},$$

$$\kappa_\beta = \frac{Var(\alpha_j)Cov(\hat{\beta}_j, \beta_j) - Cov(\hat{\beta}_j, \alpha_j)Cov(\beta_j, \alpha_j)}{Var(\alpha_j)Var(\hat{\beta}_j) - Cov(\alpha_j, \hat{\beta}_j)^2},$$

$$\kappa_0 = E\left[\beta_j\right] - \kappa_\alpha E\left[\alpha_j\right] - \kappa_\beta E\left[\hat{\beta}_j\right].$$

Simplifying these expressions yields

$$\kappa_\alpha = \frac{Cov(\alpha_j, \beta_j)}{Var(\alpha_j)} \times \frac{Var(e_j^\beta)}{Var(e_j^\alpha) + \sigma_\beta^2 - \frac{Cov(\alpha_j, \hat{\beta}_j)^2}{Var(\alpha_j)}} = r_\alpha \times \frac{Var(e_j^\beta)}{Var(e_j^\beta) + \sigma_\beta^2(1 - R^2)},$$

$$\kappa_\beta = \frac{\sigma_\beta^2 - \frac{Cov(\beta_j, \alpha_j)^2}{Var(\alpha_j)}}{Var(e_j^\alpha) + \sigma_\beta^2 - \frac{Cov(\alpha_j, \hat{\beta}_j)^2}{Var(\alpha_j)}} = \frac{\sigma_\beta^2(1 - R^2)}{Var(e_j^\beta) + \sigma_\beta^2(1 - R^2)},$$

$$\kappa_0 = (1 - \kappa_\alpha - \kappa_\beta)\beta_0 - \kappa_\alpha b_0 = (1 - \kappa_\beta)(1 - r_\alpha)\beta_0 - (1 - \kappa_\beta)r_\alpha b_0,$$

where $r_\alpha = Cov(\alpha_j, \beta_j)/Var(\alpha_j)$ and $R^2 = Cov(\alpha_j, \beta_j)^2/(Var(\alpha_j)Var(\beta_j))$. These are the coefficients in equation (19).

The third special case is when lotteries provide no information about β_j (so $Cov(\hat{\rho}_\ell, \beta_j) = 0 \ \forall \ell$) and conventional VAM sampling errors are uncorrelated (so $Cov(e_j^\alpha, e_k^\alpha) = 0 \ \forall k \neq j$). In this case the posterior mean is simply the regression of β_j on $\hat{\alpha}_j$:

$$\beta_j = \tilde{\kappa}_0 + \tilde{\kappa}_\alpha \hat{\alpha}_j + \tilde{v}_j,$$

which has coefficients

$$\tilde{\kappa}_\alpha = \frac{Cov(\beta_j, \hat{\alpha}_j)}{Var(\hat{\alpha}_j)},$$

$$\tilde{\kappa}_0 = E\left[\beta_j\right] - \tilde{\kappa}_\alpha E\left[\alpha_j\right].$$

Simplifying these yields

$$\tilde{\kappa}_\alpha = \frac{Cov(\beta_j, \beta_j + b_j + e_j^\alpha)}{Var(\beta_j + b_j + e_j^\alpha)} = \frac{\sigma_\beta^2 + \sigma_{\beta b}}{\sigma_\beta^2 + \sigma_b^2 + 2\sigma_{\beta b} + Var(e_j^\alpha)},$$

$$\tilde{\kappa}_0 = (1 - \tilde{\kappa}_\alpha)\beta_0 - \tilde{\kappa}_\alpha b_0,$$

which are the coefficients in (20).

B.4 Empirical Bayes Posterior Modes

In practice the first stage matrix Π is unknown and must be estimated. The vector of unknown school-specific parameters is then

$$\Theta = (\beta_1, b_1, \delta_1, \xi_1, ..., \beta_J, b_J, \delta_J, \xi_J, \nu_{11}, ..., \nu_{LJ})'.$$

Up to a scaling constant, the posterior density for Θ conditional on the observed estimates $\hat{\Omega}$ (which now include the estimated $\hat{\pi}_{\ell j}$) and the prior parameters θ can be expressed

$$f_{\Theta|\hat{\Omega}}\left(\Theta|\hat{\Omega};\theta\right) \propto \phi_m\left(\hat{\Omega} - \Omega(\Theta); V\right)\phi_m\left(\Theta - \bar{\Theta}(\theta); \Gamma(\theta)\right),$$

where

$$\bar{\Theta}(\theta) = (\beta_0 + \beta_Q, b_0, \delta_0, \xi_0, ...\beta_0, b_0, \delta_0, \xi_0, 0, ...0)',$$

$\phi_m(x; v)$ is the multivariate normal density function with mean zero and covariance matrix v, and

$$\Gamma(\theta) = \begin{bmatrix} I_J \otimes \Sigma & 0 \\ 0 & \sigma_\nu^2 I_{LJ} \end{bmatrix},$$

where I_J and I_{LJ} are identity matrices of dimension J and $L \times J$. Note that the probability limit of the vector of observed estimates, Ω, is a function of Θ, so we write $\Omega(\Theta)$.

As before we form an empirical Bayes posterior density by plugging $\hat{\theta}_{SMD}$ into (A1) . The empirical Bayes posterior mean is

$$\Theta^*_{mean} = \int \Theta f_{\Theta|\hat{\Omega}}\left(\Theta|\hat{\Omega};\hat{\theta}_{SMD}\right) d\Theta.$$

Since the first stage parameters $\pi_{\ell j}$ are nonlinear functions of δ and ξ, the density in (A1) will not generally be normal. As a result the integral for the posterior mean does not have a closed form and it is not possible to sample directly from the posterior distribution. To avoid integration we instead work with the posterior mode:

$$\Theta^*_{mode} = \arg\max_{\Theta}\ \log\phi_m\left(\hat{\Omega} - \Omega(\Theta); V_e\right) + \log\phi_m\left(\Theta - \bar{\Theta}\left(\hat{\theta}_{SMD}\right); \Gamma\left(\hat{\theta}_{SMD}\right)\right).$$

The posterior mode coincides with the posterior mean in the fixed first stage case where the posterior distribution is normal. The mode is computationally convenient in the estimated first stage case, as it simply requires solving a regularized maximum likelihood problem.

We compare posterior modes for the β_j with conventional empirical Bayes posterior means based on OLS estimates of value-added. The conventional predictions are given by

$$\alpha^*_{j.} = \left(\frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + Var(e^\alpha_j)}\right)\hat{\alpha}_j + \left(1 - \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + Var(e^\alpha_j)}\right)\hat{\mu}_\alpha, \qquad \text{(A2)}$$

where

$$\hat{\mu}_\alpha = \frac{1}{J}\sum_j \hat{\alpha}_j,$$

$$\hat{\sigma}^2_\alpha = \frac{1}{J}\sum_j \left[(\hat{\alpha}_j - \hat{\mu}_\alpha)^2 - Var\left(e^\alpha_j\right)\right].$$

Models with sector effects replace $\hat{\mu}_\alpha$ in equation (A2) with the regression predictions

$$\hat{\mu}_{\alpha j} = S'_j\left[\tfrac{1}{J}\sum_k S_k S'_k\right]^{-1}\left[\tfrac{1}{J}\sum_k S_k\hat{\alpha}_k\right],$$

where S_j is a vector including a constant and charter and pilot school indicators.

B.5 Relationship Between Forecast Coefficient and VAM Reliability

Here we derive the relationship between the probability limit of the IV forecast coefficient, φ, and the VAM reliability ratio, $r_\alpha = Cov(\beta_j, \alpha_j)/Var(\alpha_j)$. The IV model that generates φ is

$$Y_i = \Delta_0 + C_i'\Delta_c + \varphi\hat{Y}_i + \zeta_i.$$

The corresponding reduced form is

$$Y_i = \tau_0 + C_i'\tau_c + Z_i'\rho + u_i,$$

while the first stage is

$$\hat{Y}_i = \tilde{\tau}_0 + C_i'\tilde{\tau}_c + Z_i'\psi + \tilde{u}_i.$$

When C_i is a set of mutually exclusive and exhaustive indicator variables for participation in L lotteries, Theorem 4.5.1 in Angrist and Pischke (2009) implies that 2SLS estimation of this system yields the probability limit

$$\varphi = \sum_{\ell=1}^{L}\left(\frac{\omega_\ell}{\sum_{\ell'}\omega_{\ell'}}\right)\left(\frac{\rho_\ell}{\psi_\ell}\right),$$

where ρ_ℓ and ψ_ℓ are the elements of ρ and ψ corresponding to $Z_{i\ell}$, and

$$\omega_\ell = Pr\left[C_{i\ell} = 1\right] Var(Z_{i\ell}|C_{i\ell} = 1)\left(\psi_\ell\right)^2.$$

This equation shows that the forecast coefficient generated by an overidentified instrumental variables model equals a particular weighted average of lottery-specific forecast coefficients.

The equation for the forecast coefficient can be rewritten

$$\varphi = \frac{\sum_{\ell=1}^{L}\left(\frac{\tilde{\omega}_\ell}{\sum_k \tilde{\omega}_k}\right)\rho_\ell\psi_\ell}{\sum_{\ell=1}^{L}\left(\frac{\tilde{\omega}_\ell}{\sum_k \tilde{\omega}_k}\right)\left(\psi_\ell\right)^2},$$

where $\tilde{\omega}_\ell = Pr\left[C_{i\ell} = 1\right] Var\left(Z_{i\ell}|C_{i\ell} = 1\right)$. This expression shows that φ may be written as the coefficient from a weighted least squares regression through the origin of reduced form effects on test scores, ρ_ℓ, on first-stage effects on predicted value-added, ψ_ℓ.

In the notation of the random coefficients model, these reduced form and first stage effects are given by

$$\rho_\ell = \sum_{j=1}^{J}\pi_{\ell j}\beta_j,$$

$$\psi_\ell = \sum_{j=1}^{J}\pi_{\ell j}\alpha_j.$$

In a scenario with $E\left[\psi_\ell\right] = 0$, we can then write

$$\varphi = \frac{Cov\left(\sqrt{\tilde{\omega}_\ell}\sum_{j=1}^{J}\pi_{\ell j}\beta_j, \sqrt{\tilde{\omega}_\ell}\sum_{j=1}^{J}\pi_{\ell j}\alpha_j\right)}{Var\left(\sqrt{\tilde{\omega}_\ell}\sum_{j=1}^{J}\pi_{\ell j}\alpha_j\right)}.$$

This expression shows that the forecast coefficient φ is a weighted regression of linear combinations of the β_j's on the same linear combinations of the α_j's. The weights depend on the size of each lottery and the offer rate, while the first stage coefficients that form the linear combinations depend on offer takeup rates and the distribution of fallback schools for lottery applicants. Although both φ and r_α measure the correlation of OLS and causal value-added, and coincide when the expectation of β_j given α_j is linear and the first stage parameters $\pi_{\ell j}$ are non-stochastic, in general these summary statistics should be expected to differ.

Appendix C: Supplementary Results

C.1 Results for Seventh and Eighth Grade

As for the bias tests reported in Table III, school effects on seventh and eighth grade test scores are modeled as linear in the number of years spent in each school. The random coefficients framework of Section V.A is adapted to data from seventh grade by modifying the lottery first stage estimand as follows:

$$\pi_{\ell\ell} = 2 \times \frac{\exp(\delta_\ell)}{1+\exp(\delta_\ell)}.$$

The remaining equations describing the random coefficients model are unchanged. This specification guarantees that the effects of lottery offers on time spent in each school are less than two years in absolute value, which is the maximum potential attendance through seventh grade. Likewise, the model for eighth grade uses three years of potential attendance. The value-added and bias parameters, β_j and b_j, may then be interpreted as causal effects and VAM bias associated with one additional year of attendance at school j.

Appendix Table A.IV reports math hyperparameter estimates separately by grade. The results for seventh and eighth grade are qualitatively similar to those for sixth. Standard deviations of causal value-added in each grade are somewhat larger than the corresponding bias standard deviations, and covariances between value-added and bias are uniformly negative. Standard deviations of annual school effects are smaller for the higher grades, which suggests there is some concavity in the relationship between achievement and years of exposure to a particular school. Similarly, differences in value-added between lottery and non-lottery schools and between charter and traditional schools are positive for seventh and eighth grade but smaller than these differences for sixth grade.

Appendix Table A.V shows policy simulation results for school closure decisions based on seventh and eighth grade outcomes. The reported impacts are effects of one extra year spent at the replacement school rather than the closed school. Like the sixth grade results from Table VIII, the simulations for higher grades show large gains associated with using conventional value-added models for accountability decisions. For example, replacing the lowest-performing district school according to the lagged score model with a typical top quintile school is predicted to generate an impact of 0.24σ per year on eighth grade scores, 63 percent of the gain attainable with knowledge of true value-added (0.38σ). Hybrid estimation boosts this effect to 0.29σ, a 22 percent improvement over the conventional model.

C.2 Models with Time-Varying Value-Added

The hyperparameter estimates reported in Table VI are estimated under the assumption that causal school quality and bias are stable over time. Chetty et al. (2014a) document temporal instability in conventional teacher VAM estimates; a model that presumes constant school quality may be inappropriate if school value-added is similarly unstable.

To probe the stability of school value-added, we report estimates from a model allowing school effects to vary by year, fit to year-specific OLS and lottery estimates. This model is based on the specification

$$\beta_{jt} = \beta_j + \tilde{\beta}_{jt},$$

$$b_{jt} = b_j + \tilde{b}_{jt},$$

where (β_j, b_j) are joint normal as in equation (14), and $\tilde{\beta}_{jt}$ and $\tilde{b}_{jt}$ are *iid* uncorrelated normal shocks with mean zero and standard deviations $\sigma_{\tilde{\beta}}$ and $\sigma_{\tilde{b}}$. The first stage mean utility parameters δ_j and ξ_j are assumed stable over time, so changes in the first stage are captured by the idiosyncratic shocks $\nu_{\ell jt}$. The simulated minimum distance procedure uses time averages of the moments listed in Appendix B.2, and is augmented with variances of the year to year changes in $\hat{\alpha}_{jt}$, $\hat{\tau}_{\ell t}$ and $\hat{\rho}_{\ell t}$ in order to estimate the standard deviations of the idiosyncratic value-added and bias shocks.

Minimum distance estimates from the model with time-varying value-added appear in Appendix Table A.VIII. These estimates suggest that the permanent components of value-added and bias are more important than the idiosyncratic components. Estimated standard deviations of the permanent component of value-added are between 0.17σ and 0.22σ across models, roughly similar to the corresponding estimates from Table VI. Estimated standard deviations of the idiosyncratic component are around 0.1σ in each model. Likewise, estimated standard deviations of the permanent component of bias equal 0.41σ, 0.21σ and 0.20σ, compared to 0.05σ, 0.07σ and 0.06σ. We cannot reject the null hypothesis that bias is constant over time at conventional levels. These results suggest that school value-added and bias are reasonably stable across years, so our preferred specifications use the more parsimonious model that abstracts from time variation.

C.3 Misclassification Results

Like many states and school districts, the Massachusetts Department of Elementary and Secondary Education implements an accountability scheme based on standardized tests. Massachusetts' Framework for School Accountability and Assistance places schools into five "levels" based on four-year histories of test score levels and changes. Schools in the bottom quintile of this measure are designated level 3 or higher. A subset of these schools are classified in levels 4 and 5, a designation that puts them at risk of restructuring or closure.[1] Appendix Table A.IX uses the simulations described in Section VII to calculate the frequency of classification errors in accountability schemes of this sort.

Uncontrolled value-added estimates produce highly inaccurate school rankings. As can be seen in the second row of Table A.IX, uncontrolled VAM misclassifies 86 percent of lowest decile schools, 71 percent of lowest quintile schools, and 59 percent of lowest tercile schools. These rates are not much better than the error rates for a policy that simply ranks schools randomly (90, 80 and 67 percent, shown in the first row). Hybrid posterior modes that combine uncontrolled OLS and lottery estimates misclassify 73, 45 and 36 percent of lowest decile, quintile and tercile schools. Although still high, these error rates represent a marked improvement on the rates produced by the conventional posterior mean from an uncontrolled model.

Adding controls for demographics and previous achievement reduces misclassification rates based on both conventional and hybrid estimates. Conventional misclassification rates for lowest decile, quintile and tercile schools are 59, 47 and 38 percent when rankings are based on estimates from

[1]The Massachusetts accountability system also uses information on graduation, dropout rates and from site visits to classify schools; see `http://www.doe.mass.edu/apa/sss/turnaround/level5/schools/FAQ.html` for details.

the gains specification. In this model, hybrid estimation reduces classification error in the lowest decile from 59 to 41 percent, 31 percent fewer mistakes. The hybrid advantages in classifying lowest quintile and lowest tercile schools equal 38 and 39 percent in the gains specification. The pattern of classification improvement from the lagged score and gains specifications are broadly similar. For both the lagged score and gains models, hybrid estimation cuts mistakes in classifying upper and lower tercile schools to under one third.

The relationship between school rankings based on true and estimated value-added summarizes the predictive value of VAM estimates. Column 7 of Table A.IX reports coefficients from regressions of a school's rank in the causal value-added distribution on its rank in each estimated distribution. This rank coefficient increases from 0.15 in the uncontrolled conventional model to 0.61 in the conventional gains specification. Hybrid estimation boosts the rank coefficient for gains to 0.84. In other words, sufficiently controlled VAM estimates strongly predict relative value-added: a one-position increase in a school's VAM rank translates into an average increase of roughly 0.8 positions in the distribution of true school quality.

C.4 Sensitivity Analysis for Bias Assumption

As noted in Section V, a key assumption underlying the hybrid approach is that bias distributions are the same for lottery and non-lottery schools. It is worth documenting the sensitivity of our results to violations of this assumption. To this end, we simulate versions of the model in which the parameters of the bias distribution differ for non-lottery schools. Policymakers in these simulations continue to presume that there is no difference between these groups.

As can be seen in Appendix Table A.X, realistic differences in bias distributions between lottery and non-lottery schools tend to modestly degrade the performance of hybrid posterior estimates. The second and third rows report results from simulations that set the mean bias, b_0, 0.2σ higher or lower for non-lottery schools. As shown in Table VI, these differences are roughly one standard deviation in the distribution of causal school quality, and more than one standard deviation in the distribution of bias. These changes cause root mean squared error for the hybrid estimator to grow from 0.12σ to 0.15σ in the lagged score specification and 0.10σ to 0.14σ in the gains model.

The remaining rows of Table A.X display the effects of changing the standard deviation of bias, σ_b, and the covariance between value-added and bias, $\sigma_{\beta b}$, for non-lottery schools. Doubling σ_b for schools without lotteries increases RMSE to 0.16σ and 0.14σ in the lagged score and gains models. As shown in the fifth row, cutting σ_b in half for non-lottery schools improves the performance of hybrid estimates; in this case, misspecification error is outweighted by the decline in the magnitude of bias for non-lottery schools. The final row shows that reversing the sign of $\sigma_{\beta b}$ for non-lottery schools increases RMSE to 0.13σ and 0.12σ in the lagged score and gains specifications.

Columns 2-5 of Table A.X show that increases in RMSE due to misspecification of the bias distribution are accompanied by decreases in the expected benefits of school closure decisions based on the hybrid estimates. At the same time, these benefits remain substantial in all simulations. When σ_b is doubled for non-lottery schools, for example, a policy that uses hybrid gains estimates to replace the lowest-ranked school with an average school generates an average improvement of 0.24σ. The results in Table A.X indicate that hybrid estimation is likely to be of value as long as differences in bias distributions between lottery and non-lottery schools are modest.

Additional Appendix References

1. Abadie, Alberto, "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113 (2003), 231-263.

2. Altonji, Joseph G., and Lewis M. Segal, "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business and Economic Statistics*, 14 (1996), 353-366.

3. Angrist, Joshua D., and Jörn-Steffen Pischke, *Mostly Harmless Econometrics*, Princeton, NJ: Princeton University Press, 2009.

4. Hansen, Lars Peter, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50 (1982), 1029-1054.

5. Rubin, Donald B., "The Bayesian Bootstrap," *Annals of Statistics*, 9 (1981), 130-134.

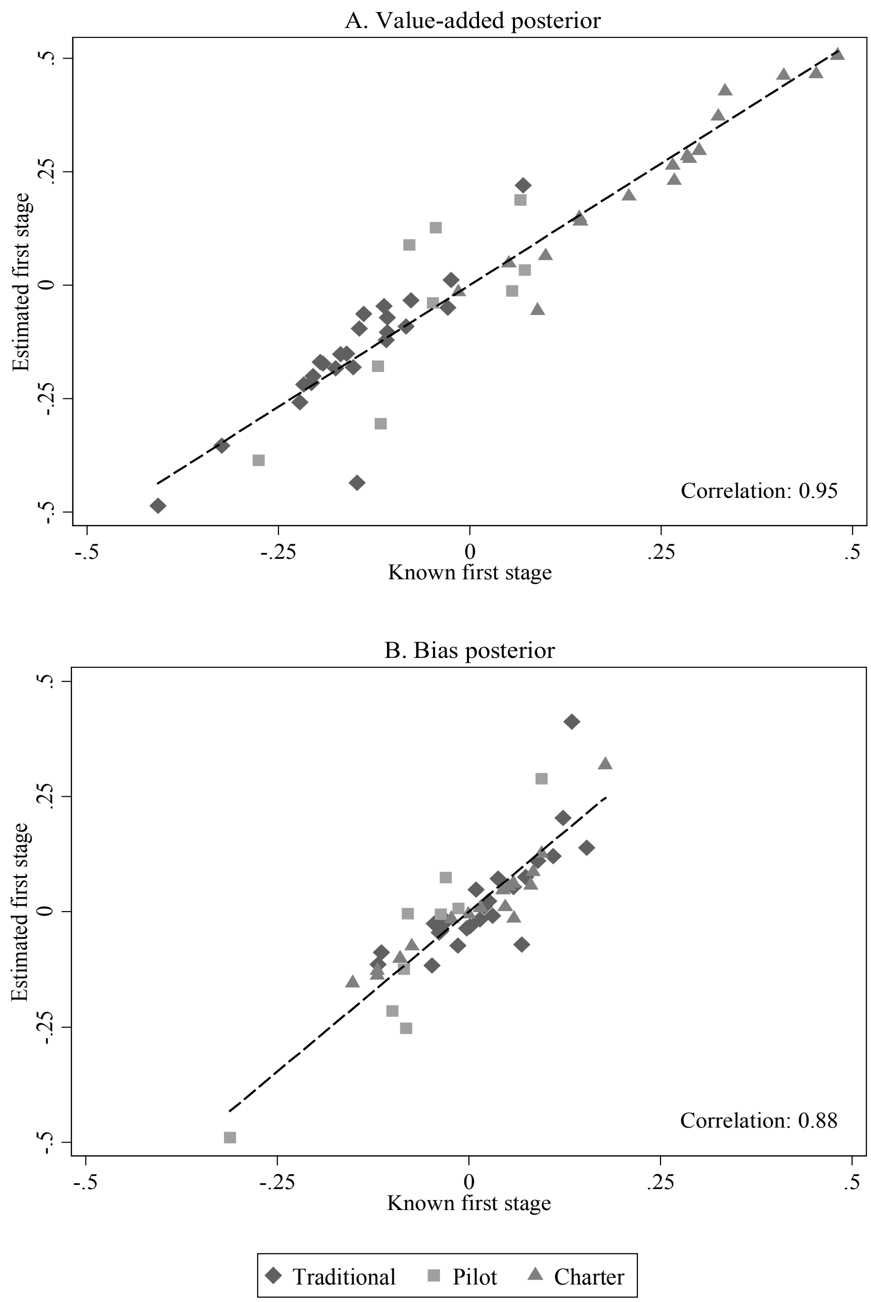

Figure A.I. Posterior Predictions With and Without a Known First Stage

Notes: This figure displays the correlation between posterior predictions of value-added and bias when the lottery first stage is treated as known vs. estimated. Estimates come from the lagged score value-added model with sector effects for sixth grade math scores. The horizontal axis in each panel displays posterior means computed under the assumption that there is no sampling error in the first stage coefficients. The vertical axis in each panel displays posterior modes accounting for estimation error in the first stage. Dashes show OLS lines of best fit.

Table A.I. Lottery Attrition

		Offer balance			
	Mean	All lotteries	Traditional	Pilot	Charter
	(1)	(2)	(3)	(4)	(5)
In lottery sample	0.813	0.028***	0.036***	-0.003	0.010
		(0.010)	(0.011)	(0.023)	(0.015)
N	10,718	10,718	5,589	1,512	4,867

Notes: This table reports the followup rate for the lottery sample and investigates differential attrition by lottery offer status. Column 1 shows the fraction of randomized lottery applicants that appear in the Boston sixth grade sample. Columns 2-5 report coefficients from regressions of an indicator for followup on lottery offers, controlling for lottery strata. Robust standard errors are reported in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table A.II. Tests for Bias in ELA Value-Added Models

	All lotteries		Excluding charter lotteries	
	Lagged score	Gains	Lagged score	Gains
	(1)	(2)	(3)	(4)
	A. Sixth grade			
Forecast coefficient (φ)	0.864	0.722	0.423	0.133
	(0.167)	(0.172)	(0.310)	(0.308)
First stage F-statistic	26.8	29.4	14.0	13.1
p-values:				
Forecast bias	0.416	0.105	0.063	0.005
Overidentification	0.039	0.007	0.157	0.127
Omnibus test χ^2 statistic (d.f.)	46.0 (28)	56.8 (28)	36.1 (23)	43.2 (23)
p-value	0.018	0.001	0.040	0.007
N	8,718		6,162	
	B. All middle school grades			
Forecast coefficient (φ)	0.969	0.924	0.699	0.550
	(0.101)	(0.107)	(0.193)	(0.191)
First stage F-statistic	11.3	12.3	6.7	6.6
p-values:				
Forecast bias	0.759	0.481	0.118	0.019
Overidentification	0.062	0.014	0.122	0.081
Omnibus test χ^2 statistic (d.f.)	119.0 (75)	137.1 (75)	80.7 (60)	92.9 (60)
p-value	<0.001	<0.001	0.039	0.004
N	20,935		15,027	

Notes: This table reports the results of tests for bias in conventional value-added models for sixth through eighth grade ELA scores. See the notes to Table III for a description of the value-added models and test procedure. Standard errors, clustered by student, are reported in parentheses.

Table A.III. Joint Distribution of Value-Added, Bias, and Lottery Compliance

	β_j (1)	b_j (2)	δ_j (3)	ξ_j (4)
Standard deviation	0.171	0.148	0.764	0.864
	(0.028)	(0.029)	(0.131)	(0.584)
Covariance w/b_j	-0.016			
	(0.006)			
Covariance w/δ_j	0.009	0.043		
	(0.018)	(0.032)		
Covariance w/ξ_j	0.077	-0.102	-0.491	
	(0.029)	(0.056)	(0.151)	
Charter effect	0.426	-0.005	0.241	-1.934
	(0.104)	(0.103)	(0.387)	(0.425)
Pilot effect	0.130	-0.121	0.074	-0.479
	(0.129)	(0.124)	(0.312)	(0.434)
Std. dev. of v_{lj}	1.566			
	(0.152)			

Notes: This table reports simulated minimum distance estimates of parameters governing the distribution of value-added, bias, and lottery compliance probabilities for the lagged score value-added model fit to sixth grade school attendance and math scores. See the notes to Table VI for a description of the estimation procedure.

Table A.IV. Minimum Distance Estimates by Grade

		Sixth grade		Seventh grade		Eighth grade	
		Lagged score	Gains	Lagged score	Gains	Lagged score	Gains
		(1)	(2)	(3)	(4)	(5)	(6)
σ_β	Std. dev. of causal VA	0.171	0.170	0.137	0.120	0.109	0.101
		(0.028)	(0.023)	(0.022)	(0.021)	(0.019)	(0.018)
σ_b	Std. dev. of OLS bias	0.148	0.133	0.119	0.094	0.079	0.078
		(0.029)	(0.030)	(0.040)	(0.032)	(0.025)	(0.025)
$\sigma_{\beta b}$	Covariance of VA and bias	-0.016	-0.013	-0.007	-0.006	-0.004	-0.005
		(0.006)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
r_α	Regression of VA on OLS (reliability ratio)	0.694	0.783	0.625	0.747	0.771	0.798
		(0.152)	(0.122)	(0.084)	(0.096)	(0.108)	(0.105)
VA shifters	Charter	0.426	0.396	0.210	0.192	0.145	0.133
		(0.104)	(0.106)	(0.091)	(0.081)	(0.075)	(0.073)
	Pilot	0.130	0.111	-0.039	-0.019	0.004	0.008
		(0.129)	(0.129)	(0.110)	(0.101)	(0.085)	(0.082)
	Lottery school (β_Q)	0.104	0.066	0.003	0.034	0.047	0.056
		(0.042)	(0.041)	(0.042)	(0.033)	(0.027)	(0.027)
Bias shifters	Charter	-0.005	-0.063	0.010	0.030	0.049	0.039
		(0.103)	(0.099)	(0.088)	(0.077)	(0.070)	(0.073)
	Pilot	-0.121	-0.089	0.009	0.062	-0.036	0.011
		(0.124)	(0.121)	(0.107)	(0.097)	(0.078)	(0.077)
	$\chi^2(13)$ statistic:	9.0	6.0	9.2	4.5	6.4	8.8
	Overid. p- value:	0.773	0.946	0.759	0.985	0.932	0.785

Notes: This table reports minimum distance estimates of parameters of the joint distribution of causal school value-added and OLS bias for each middle school grade. School exposure for seventh and eighth grade is measured as the number of years spent in each school. See the notes to Table VI for a description of the estimation procedure.

Table A.V. Per-year Effects of Closing the Lowest-Ranked District School for Affected Children, by Grade

Grade	Model	Posterior method	Replacement school: Average school (1)	Average above-median school (2)	Average top-quintile school (3)	Average charter school (4)
Seventh	-	True value-added	0.284	0.389	0.468	0.500
			[0.059]	[0.067]	[0.073]	[0.076]
	Lagged score	Conventional	0.187	0.253	0.299	0.403
			[0.108]	[0.116]	[0.119]	[0.116]
		Hybrid	0.225	0.301	0.356	0.441
			[0.101]	[0.108]	[0.113]	[0.112]
	Gains	Conventional	0.157	0.215	0.259	0.344
			[0.094]	[0.101]	[0.103]	[0.101]
		Hybrid	0.190	0.257	0.311	0.377
			[0.088]	[0.096]	[0.103]	[0.097]
Eighth	-	True value-added	0.229	0.316	0.384	0.369
			[0.049]	[0.055]	[0.059]	[0.062]
	Lagged score	Conventional	0.162	0.226	0.273	0.302
			[0.083]	[0.088]	[0.092]	[0.092]
		Hybrid	0.193	0.267	0.325	0.333
			[0.071]	[0.078]	[0.083]	[0.079]
	Gains	Conventional	0.142	0.199	0.241	0.268
			[0.080]	[0.084]	[0.088]	[0.087]
		Hybrid	0.171	0.238	0.293	0.297
			[0.079]	[0.088]	[0.093]	[0.087]

Notes: This table reports simulated test score impacts of closing the lowest-ranked BPS district school based on value-added predictions for seventh and eighth grade. The reported effects are average impacts of one year of attendance at the replacement school rather than the closed school. Standard deviations of these effects across simulations appear in brackets. See the notes to Table VIII for a description of the simulation procedure.

Table A.VI. Tests for Bias in Hybrid Value-Added Posteriors

	All lotteries		Excluding charter lotteries	
	Lagged score	Gains	Lagged score	Gains
	(1)	(2)	(3)	(4)
Forecast coefficient (φ)	1.040	1.001	1.143	1.086
	(0.130)	(0.129)	(0.316)	(0.308)
First stage F-statistic	24.0	23.6	7.5	7.3
p-values:				
Forecast bias	0.760	0.994	0.651	0.780
Overidentification	0.849	0.710	0.808	0.783
Omnibus test χ^2 statistic (d.f.)	0.7 (28)	0.8 (28)	0.8 (23)	0.8 (23)
p-value	0.841	0.701	0.795	0.775
N	8,718		6,162	

Notes: This table reports the results of tests for bias in posterior value-added predictions for sixth grade math scores. Empirical Bayes posterior modes come from random coefficient models with sector effects. See the notes to Table III for a description of the value-added models and test procedure. Robust standard errors are reported in parentheses.

Table A.VII. Consequences of Closing the Lowest-Ranked District School Based on Four Years of Data

		Replacement school			
		Average school	Average above-median school	Average top-quintile school	Average charter school
Model	Posterior method	(1)	(2)	(3)	(4)
-	True value-added	0.370	0.507	0.610	0.711
		[0.080]	[0.089]	[0.094]	[0.094]
Uncontrolled	Conventional	0.051	0.071	0.087	0.274
		[0.193]	[0.200]	[0.206]	[0.199]
	Hybrid	0.148	0.217	0.256	0.371
		[0.143]	[0.154]	[0.168]	[0.151]
Lagged score	Conventional	0.223	0.305	0.364	0.575
		[0.152]	[0.162]	[0.172]	[0.158]
	Hybrid	0.302	0.420	0.511	0.651
		[0.134]	[0.146]	[0.154]	[0.146]
Gains	Conventional	0.229	0.315	0.380	0.570
		[0.144]	[0.153]	[0.162]	[0.149]
	Hybrid	0.302	0.417	0.506	0.641
		[0.122]	[0.133]	[0.141]	[0.133]

Notes: This table reports simulated test score impacts of closing the lowest-ranked BPS district school based on value-added predictions computed from four years of data. These effects are computed by scaling up the covariance matrix of sampling errors underlying the simulations in Table VIII by a factor of two. See the notes to Table VIII for a description of the simulation procedure.

Table A.VIII. Models with Time-Varying Value-Added and Bias

		Uncontrolled	Lagged score	Gains
		(1)	(2)	(3)
σ_β	Std. dev. of causal VA (permanent)	0.168	0.215	0.193
		(0.034)	(0.027)	(0.024)
	Std. dev. of causal VA (transitory)	0.103	0.084	0.091
		(0.050)	(0.043)	(0.040)
σ_b	Std. dev. of OLS bias (permanent)	0.414	0.214	0.199
		(0.018)	(0.025)	(0.023)
	Std. dev. of OLS bias (transitory)	0.046	0.066	0.066
		(0.164)	(0.062)	(0.064)
$\sigma_{\beta b}$	Covariance of VA and bias	-0.034	-0.037	-0.029
		(0.008)	(0.007)	(0.005)
r_α	Regression of VA on OLS (reliability ratio)	-0.041	0.510	0.431
		(0.189)	(0.182)	(0.147)
VA shifters	Charter	0.263	0.487	0.354
		(0.129)	(0.153)	(0.134)
	Pilot	-0.024	0.078	0.091
		(0.145)	(0.170)	(0.140)
	Lottery school (β_Q)	0.178	0.133	0.127
		(0.136)	(0.061)	(0.053)
Bias shifters	Charter	0.324	-0.043	0.020
		(0.179)	(0.151)	(0.138)
	Pilot	-0.151	-0.157	-0.117
		(0.204)	(0.170)	(0.139)
	$\chi^2(14)$ statistic:	15.5	12.1	15.5
	Overid. p- value:	0.347	0.597	0.343

Notes: This table reports simulated minimum distance estimates of parameters of the joint distribution of causal school value-added and OLS bias for sixth grade math scores from a model that allows school effects to vary by year. School value-added and bias are assumed to consist of a permanent component plus an independent and identically distributed transitory shock each year. See the notes to Table VI for a description of the estimation procedure.

Table A.IX. Error Rates for Classification Decisions Among District Schools

Value-added model	Posterior method	Low-performing schools: Lowest decile (1)	Lowest quintile (2)	Lowest tercile (3)	High-performing schools: Highest decile (4)	Highest quintile (5)	Highest tercile (6)	Rank coefficient (7)
-	Random	0.900	0.800	0.667	0.900	0.800	0.667	0.000
		[0.161]	[0.139]	[0.118]	[0.161]	[0.139]	[0.118]	[0.177]
Uncontrolled	Conventional	0.857	0.710	0.593	0.863	0.733	0.608	0.146
		[0.190]	[0.150]	[0.122]	[0.182]	[0.151]	[0.122]	[0.171]
	Hybrid	0.726	0.449	0.359	0.781	0.606	0.491	0.502
		[0.246]	[0.176]	[0.126]	[0.226]	[0.161]	[0.126]	[0.171]
Lagged score	Conventional	0.639	0.501	0.411	0.670	0.523	0.422	0.542
		[0.256]	[0.155]	[0.121]	[0.246]	[0.154]	[0.113]	[0.134]
	Hybrid	0.438	0.316	0.249	0.382	0.305	0.234	0.825
		[0.252]	[0.141]	[0.106]	[0.231]	[0.146]	[0.105]	[0.087]
Gains	Conventional	0.594	0.469	0.379	0.611	0.483	0.393	0.606
		[0.260]	[0.152]	[0.115]	[0.483]	[0.152]	[0.112]	[0.123]
	Hybrid	0.411	0.293	0.232	0.350	0.286	0.225	0.841
		[0.237]	[0.137]	[0.103]	[0.243]	[0.147]	[0.108]	[0.096]

Notes: This table reports simulated misclassification rates for policies based on empirical Bayes posterior predictions of value-added. The first row shows results for a system that ranks schools at random. Column 1 shows the fraction of district schools in the lowest decile of true sixth grade math value-added that are not classified in the lowest decile of estimated value-added for each model. Columns 2 and 3 report corresponding misclassification rates for the lowest quintile and tercile. Columns 4-6 report misclassification rates for schools in the highest decile, quintile and tercile of true value-added. Column 7 reports the coefficient from a regression of a school's rank in the true value-added distribution on its rank in the estimated distribution. Standard deviations of misclassification rates and rank coefficients across simulations appear in brackets.

Table A.X. Sensitivity of Hybrid Posteriors to Differences in Bias Distributions Between Lottery and Non-Lottery Schools

			Closure effects			
Difference in bias distributions	Model	RMSE (1)	Average school (2)	Average above-median school (3)	Average top-quintile school (4)	Average charter school (5)
None	Lagged score	0.116	0.315	0.437	0.529	0.665
			[0.131]	[0.141]	[0.147]	[0.145]
	Gains	0.100	0.316	0.434	0.525	0.657
			[0.115]	[0.126]	[0.136]	[0.128]
Non-lottery b_0	Lagged score	0.146	0.295	0.408	0.511	0.647
0.2σ lower			[0.135]	[0.151]	[0.157]	[0.143]
	Gains	0.140	0.283	0.391	0.493	0.626
			[0.124]	[0.136]	[0.138]	[0.133]
Non-lottery b_0	Lagged score	0.152	0.263	0.365	0.454	0.615
0.2σ higher			[0.131]	[0.144]	[0.152]	[0.144]
	Gains	0.143	0.277	0.373	0.446	0.619
			[0.122]	[0.138]	[0.155]	[0.134]
Non-lottery σ_b	Lagged score	0.157	0.214	0.318	0.412	0.568
twice as large			[0.163]	[0.172]	[0.181]	[0.168]
	Gains	0.144	0.238	0.341	0.428	0.582
			[0.149]	[0.157]	[0.165]	[0.155]
Non-lottery σ_b	Lagged score	0.100	0.337	0.462	0.553	0.692
half as large			[0.125]	[0.137]	[0.144]	[0.132]
	Gains	0.092	0.334	0.454	0.543	0.679
			[0.111]	[0.123]	[0.127]	[0.116]
Non-lottery $\sigma_{\beta b}$	Lagged score	0.129	0.306	0.423	0.508	0.660
has opposite sign			[0.138]	[0.151]	[0.159]	[0.143]
	Gains	0.118	0.290	0.407	0.492	0.634
			[0.127]	[0.135]	[0.140]	[0.134]

Notes: This table explores the sensitivity of simulated effects of closing the lowest-ranked district school to violations of the assumption that bias distributions are the same for lottery and non-lottery schools. Policymakers are assumed to make closure decisions based on hybrid posterior modes constructed under the incorrect assumption that there is no difference between these groups. See the notes to Table VIII for a description of the simulation procedure.