NBER WORKING PAPER SERIES

LEVERAGING LOTTERIES FOR SCHOOL VALUE-ADDED: TESTING AND ESTIMATION

Joshua Angrist
Peter Hull
Parag A. Pathak
Christopher Walters

Working Paper 21748
http://www.nber.org/papers/w21748

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2015

We gratefully acknowledge funding from the National Science Foundation, the Laura and John Arnold Foundation, and the Spencer Foundation. We are indebted to SEII research managers Annice Correia and Eryn Heying for invaluable help and support. Thanks also go to Isaiah Andrews, Pat Kline, Guido Imbens, Rick Mansfield, Chris Nielson, Stephen Raudenbush, Jesse Rothstein, Doug Staiger, and seminar participants at the 2014 All California Labor Economics Conference, the APPAM Fall 2014 research conference, the 2014 AEFP meeting, the 2015 ASSA annual meeting, the 2015 SOLE/EALE annual meeting, the 2015 NBER Summer Institute, the Federal Reserve Bank of New York, the 2015 Becker/Friedman Applied Microeconomics Conference, the University of Chicago Committee on Education Workshop, Brown University, and the University of Chicago Workshop on Quantitative Research Methods for suggestions and comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. Angrist's daughter teaches at a Boston charter school.

Leveraging Lotteries for School Value-Added: Testing and Estimation
Joshua Angrist, Peter Hull, Parag A. Pathak, and Christopher Walters
NBER Working Paper No. 21748
November 2015, Revised July 2016
JEL No. I20,J24

ABSTRACT

Conventional value-added models (VAMs) compare average test scores across schools after regression-adjusting for students' demographic characteristics and previous scores. This paper tests for VAM bias using a procedure that asks whether VAM estimates accurately predict the achievement consequences of random assignment to specific schools. Test results from admissions lotteries in Boston suggest conventional VAM estimates are biased, which motivates the development of a hierarchical model describing the joint distribution of school value-added, bias, and lottery compliance. We use this model to assess the substantive importance of bias in conventional VAM estimates and to construct hybrid value-added estimates that optimally combine ordinary least squares and lottery-based instrumental variables estimates of VAM parameters. The hybrid estimation strategy provides a general recipe for combining non-experimental and quasi-experimental estimates. While still biased, hybrid school value-added estimates have lower mean squared error than conventional VAM estimates. Simulations calibrated to the Boston data show that, bias notwithstanding, policy decisions based on conventional VAMs are likely to generate substantial achievement gains. Hybrid estimates that incorporate lotteries yield further gains.

Joshua Angrist
Department of Economics, E52-436
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
angrist@mit.edu

Parag A. Pathak
Department of Economics, E52-426
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
ppathak@mit.edu

Peter Hull
Department of Economics,
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
hull@mit.edu

Christopher Walters
Department of Economics
University of California at Berkeley
530 Evans Hall #3880
Berkeley, CA 94720-3880
and NBER
crwalters@econ.berkeley.edu

1 Introduction

Public school districts increasingly use value-added models (VAMs) to assess teacher and school effectiveness. Conventional VAM estimates compare test scores across classrooms or schools after regression-adjusting for students' demographic characteristics and earlier scores. Achievement differences remaining after adjustment are attributed to differences in teacher or school quality. Some districts use estimates of teacher value-added to guide personnel decisions, while others use VAMs to generate "report cards" that allow parents to compare schools.[1] Value-added estimation is a high-stakes statistical exercise: low VAM estimates can lead to school closures and teacher dismissals, while a growing body of evidence suggests the near-term achievement gains produced by effective teachers and schools translate into improved outcomes in adulthood (see, e.g., Chetty et al., 2011 and Chetty et al., 2014b for teachers and Angrist et al., 2016a and Dobbie and Fryer, 2015 for schools).

Because the stakes are so high, the use of VAM estimates for teacher and school assessment remains controversial. Critics note that VAM estimates may be misleading if the available control variables are inadequate to ensure *ceteris paribus* comparisons. VAM estimates may also reflect considerable sampling error. The accuracy of teacher value-added models is the focus of a large and expanding body of research. This work demonstrates that teacher VAM estimates have predictive value, but has yet to generate a consensus on the substantive importance of bias or guidelines for "best practice" VAM estimation (see, for example, Kane and Staiger, 2008; Rothstein, 2010; Koedel and Betts, 2011; Kinsler, 2012; Kane et al., 2013; Chetty et al., 2014a; Bacher-Hicks et al., 2014; Rothstein, 2015; and Chetty et al., 2016). While the social significance of school-level VAMs is similar to that of teacher VAMs, validation of VAMs for schools has received less attention.

The proliferation of partially-randomized urban school assignment systems provides a new tool for measuring school value-added. Centralized assignment mechanisms based on the theory of market design, including those used in Boston, Chicago, Denver, New Orleans, and New York, use information on parents' preferences over schools and schools' priorities over students to allocate scarce admission offers. These matching algorithms typically use random sequence numbers to distinguish between students with the same priorities, thereby creating stratified student assignment lotteries. Similarly, independently-run charter schools often use admissions lotteries when oversubscribed. Scholars increasingly use these lotteries to identify causal effects of enrollment in various school sectors, including charter schools, pilot schools, small high schools, and magnet schools (Cullen et al., 2006; Hastings and Weinstein, 2008; Abdulkadiroğlu et al., 2011; Angrist et al., 2013; Dobbie and Fryer, 2013; Bloom and Unterman, 2014; Deming et al., 2014). Lottery-based estimation of individual school value-added is less common, however, reflecting the fact that lottery samples for many schools are small, while other schools are undersubscribed.

[1]The Education Commission of the States notes that Alabama, Arizona, California, Florida, Indiana, Louisiana, Maine, Mississippi, New Mexico, North Carolina, Texas, Utah, and Virginia issue letter-grade report cards with grades determined at least in part by adjusted standardized test scores (`http://www.ecs.org/html/educationissues/accountability/stacc_intro.asp`).

This paper develops econometric methods that leverage school admissions lotteries for VAM testing and estimation, accounting for the partial coverage of lottery data. Our first contribution is the formulation of a new lottery-based test of conventional VAMs. This test builds on recent experimental and quasi-experimental VAM validation strategies, including the work of Kane and Staiger (2008), Deutsch (2012), Kane et al. (2013), Chetty et al. (2014a) and Deming (2014). In contrast with earlier studies, which implicitly look at average-across-schools validity in a test with one degree of freedom, ours is an over-identification test that looks at each of the orthogonality restrictions generated by a set of lottery instruments. Intuitively, the test developed here asks whether conventional VAM estimates correctly predict the effect of randomized admission at every school that has a lottery, as well as the effects of VAM on average. Our test of VAM validity parallels a classical over-identification test, since the latter can be described either as testing instrument-error orthogonality or as a comparison of alternative just-identified IV estimates that should be the same under the null hypothesis.[2]

Application of this test to data from Boston reveals moderate but statistically significant bias in conventional VAM estimates. This finding notwithstanding, conventional VAM estimates may nevertheless provide a useful guide to school quality if the degree of bias is modest. To assess the practical value of VAM estimates, we develop and estimate a hierarchical random coefficients model that describes the joint distribution of value-added, VAM bias, and lottery compliance across schools. The model is estimated via a simulated minimum distance (SMD) procedure that matches moments of the distribution of conventional VAM estimates, lottery reduced forms, and first stages to those predicted by the random coefficients structure. Estimates of the model indicate substantial variation in both causal value-added and selection bias across schools. The estimated joint distribution of these parameters implies that conventional VAM estimates are highly correlated with school effectiveness.

A second contribution of our study is to use the random coefficients framework and lottery variation to improve conventional VAM estimates. Our approach builds on previous estimation strategies that trade reduced bias for increased variance (Morris, 1983; Judge and Mittlehammer, 2004, 2005, 2007). Specifically, we compute empirical Bayes (EB) hybrid posterior predictions that optimally combine relatively imprecise but unbiased lottery-based estimates with biased but relatively precise conventional VAM estimates. Importantly, our approach makes efficient use of the available lottery information without requiring a lottery for every school. Hybrid estimates for undersubscribed schools are improved by information on the *distribution* of bias contributed by schools with oversubscribed lotteries. The hybrid estimation procedure generates estimates that, while still biased, have lower mean squared error than conventional VAM estimates. Our framework provides a general recipe for combining non-experimental and quasi-experimental estimators and may therefore be useful in other settings.[3]

Finally, we quantify the consequences of bias in conventional VAM estimates and the payoff to hybrid

[2]The theory behind VAM over-identification testing is sketched in Angrist et al. (2016b).

[3]These settings include the analysis of teacher, hospital, doctor, firm, and neighborhood effects, as in Chetty et al. (2014a,b), Finkelstein et al. (2013), Fletcher et al. (2014), Card et al. (2013), and Chetty and Hendren (2015). Chetty and Hendren combine observational and quasi-experimental estimates of neighborhood effects, a connection discussed in Section 5.

estimation using a Monte Carlo simulation calibrated to our Boston estimates. Simulation results show that policy decisions based on conventional estimates that control for baseline test scores or measure score growth are likely to boost achievement. For example, replacing the lowest-ranked Boston school with an average school is predicted to generate a gain of 0.24 test score standard deviations (σ) for affected students, roughly two-thirds of the benefit obtained when true value-added is used to rank schools (0.37σ). Hybrid estimates are highly correlated with conventional estimates (the rank correlation is 0.74), and hybrid estimation generates modest additional gains, reducing mean squared error by 30 percent and increasing the benefits of school closure policies by about 0.08σ (33 percent). Conventional school VAMs would therefore appear to provide a useful guide for policy-makers, while hybrid estimators generate worthwhile improvements in policy targeting.

The next section describes the Boston data used for VAM testing and estimation, and Section 3 describes the conventional value-added framework as applied to these data. Section 4 derives our VAM validation test and discusses test implementation and results. Section 5 outlines the random coefficients model and empirical Bayes approach to hybrid estimation, while Section 6 reports estimates of the model's hyperparameters and the resulting posterior predictions of value-added. Section 7 discusses policy simulations. Finally, Section 8 concludes with remarks on how the framework developed here might be used in other settings.

2 Setting and Data

2.1 Boston Public Schools

Boston public school students can choose from a diverse set of enrollment options, including traditional Boston Public School (BPS) district schools, charter schools, and pilot schools. As in most districts, Boston's charter schools are publicly funded but free to operate within the confines of their charters. For the most part, charter staff are not covered by collective bargaining agreements and other BPS regulations.[4] Boston's pilot school sector arose as a union-supported alternative to charter schools, developed jointly by the BPS district and the Boston Teachers Union. Pilot schools are part of the district but typically have more control over their budgets, scheduling, and curriculum than do traditional public schools. On the other hand, pilot school teachers work under collective bargaining provisions similar to those in force at traditional public schools.

Applicants to traditional public and pilot schools rank between three and ten schools as the first step in a centralized match (students not finishing elementary or middle school who are happy to stay where they are need not participate in the match). Applicants are then assigned to schools via a student-proposing deferred acceptance mechanism, as described in Abdulkadiroğlu et al. (2006). This mechanism combines student preferences with a strict priority ranking over students for each school. Priorities are determined by whether an applicant is already enrolled at the school and therefore guaranteed admission, has a sibling enrolled at the school, or lives in the school's walk-zone. Ties within these coarse priority groups are broken

[4] Boston's charter sector includes both "Commonwealth" charters, which are authorized by the state and run as independent school districts, and "in-district" charters, which are authorized and overseen by the Boston School Committee.

by random sequence numbers, which we refer to as lottery numbers. In an evaluation of the pilot sector exploiting this centralized random assignment scheme, Abdulkadiroğlu et al. (2011) find mostly small and statistically insignificant effects of pilot school attendance relative to the traditional public school sector.

In contrast with the centralized match that assigns seats at traditional and pilot schools, charter applicants apply to individual charter schools separately in the spring of the year they hope to enter. By Massachusetts law, oversubscribed charter schools must select students in public admissions lotteries, with the exception of applicants with siblings already enrolled in the charter, who are guaranteed seats. Charter offers and centralized assignment offers are made independently; students applying to the charter sector can receive multiple offers. In practice, some Boston charter schools offer all of their applicants seats, while others fail to retain usable information on admissions lotteries. Studies based on charter lotteries show that Boston charter schools boost test scores and increase four-year college attendance (see, for example, Abdulkadiroğlu et al. (2011) and Angrist et al. (2016a)).

2.2 Data and Descriptive Statistics

The data analyzed here consist of a sample of roughly 28,000 sixth-grade students attending 51 Boston traditional, pilot, and charter schools in the 2006-2007 through 2013-2014 school years. In Boston, sixth grade marks the first grade of middle school, so most rising sixth graders participate in the centralized match. For our purposes, baseline test scores come from fifth grade Massachusetts Comprehensive Assessment System (MCAS) tests in math and English Language Arts (ELA), while outcomes are measured at the end of sixth grade. Test scores are standardized to have mean zero and unit variance in the population of Boston charter, pilot, and traditional public schools, separately by subject, grade, and year. Other variables used in the empirical analysis are school enrollment, race, sex, subsidized lunch eligibility, special education status, English-language learner status, and suspensions and absences. Appendix A describes the administrative files and data processing conventions used to construct the working extract.

Our analysis combines data from the centralized traditional and pilot match with lottery data from individual charter schools. The BPS lottery instruments code offers at applicants' first choice (highest ranked) middle schools in the match. In particular, BPS lottery offers indicate applicants whose lottery numbers are at least as high as the worst number offered a seat at their first-choice school, among those in the same priority group. Conditional on application year, first-choice school, and an applicant's priority at that school (what we call the assignment strata), offers of seats at a first choice are randomly assigned. Charter lottery instruments indicate offers made on the night of the admissions lottery at each charter school. These offers are randomly assigned for non-siblings conditional on the target school and application year.[5]

The schools and students analyzed here are described in Table 1. We exclude schools serving fewer than 25 sixth graders in each year, leaving a total of 25 traditional public schools, 9 pilot schools, and 17 charter

[5] For a much smaller group of applicants, the centralized BPS mechanism induces random tie-breaking for lower-ranked school choices. The use of tie-breaking from these choices generates complications beyond the scope of this paper; see Abdulkadiroğlu et al. (2015) for a comprehensive analysis of empirical strategies that exploit centralized assignment.

schools. Of these, 37 schools have sixth grade as a primary entry point and 28 (16 traditional, 7 pilot, and 5 charter) had at least 50 students subject to random sixth grade assignment. Applicants to these 28 schools constitute our lottery sample. Conventional ordinary least squares (OLS) value-added models are estimated in a sample of 27,864 Boston sixth graders with complete baseline, demographic, and outcome information; 8,718 of these students are also in the lottery sample.

About 77 percent of Boston sixth graders enroll at schools with usable lotteries, and, as can be seen in the descriptive statistics reported in Table 2, demographic characteristics for this group are comparable to those of the full BPS population. Columns 3 and 4 of Table 2 report characteristics of the subset of students subject to randomized lottery assignment. Lotteried students are slightly more likely to be African American and to qualify for a subsidized lunch, and somewhat less likely to be white or to have been suspended or recorded as absent in fifth grade. Table 2 also documents the comparability of students who were and were not offered seats in a lottery. These results, reported in columns 5-7, compare the baseline characteristics of lottery winners and losers, controlling for assignment strata. Consistent with conditional random assignment of offers, estimated differences by offer status are small and not significantly different from zero, both overall and within school sectors.[6]

3 Value-added Framework

As in earlier investigations of school value-added, the analysis here builds on a constant-effects causal model. This reflects a basic premise of the VAM framework: internally valid treatment effects from earlier years and cohorts are presumed to have predictive value for future cohorts. Student i's potential test score at school j, Y_{ij}, is therefore written as the sum of two non-interacting components, specifically:

$$Y_{ij} = \mu_j + a_i, \tag{1}$$

where μ_j is the mean potential outcome at school j and a_i is student i's "ability," or latent achievement potential. This additively-separable model implies that causal effects are the same for all students. The constant effects framework focuses attention on the possibility of selection bias in VAM estimates rather than treatment effect heterogeneity (though we explore heterogeneity as well).

A dummy variable, D_{ij}, is used to indicate whether student i attended school j in sixth grade. The observed sixth-grade outcome for student i can therefore be written

[6]Lottery estimates may be biased by selective sample attrition. As shown in Appendix Table A1, follow-up data are available for 81 percent of lottery applicants, while sample retention is 2.8 percentage points higher for lottery winners than for losers, a difference driven by traditional public school lotteries. Table 2 shows that that baseline characteristics are balanced in the sample with follow-up scores, so the modest differential attrition documented in Table A1 seems unlikely to affect the results reported here.

$$Y_i = Y_{i0} + \sum_{j=1}^{J} (Y_{ij} - Y_{i0}) D_{ij}$$

$$= \mu_0 + \sum_{j=1}^{J} \beta_j D_{ij} + a_i. \quad (2)$$

The parameter $\beta_j \equiv \mu_j - \mu_0$ measures the causal effect of school j relative to an omitted reference school with index value 0. In other words, β_j is school j's value-added.

Conventional value-added models use regression methods to mitigate selection bias. Write

$$a_i = X_i'\gamma + \epsilon_i, \quad (3)$$

for the regression of a_i on a vector of controls, X_i, which includes lagged test scores. Note that $E[X_i \epsilon_i] = 0$ by definition of γ. This decomposition implies that observed outcomes can be written

$$Y_i = \mu_0 + \sum_{j=1}^{J} \beta_j D_{ij} + X_i'\gamma + \epsilon_i. \quad (4)$$

It bears emphasizing that equation (4) is a causal model: ϵ_i is defined so as to be orthogonal to X_i, but need not be uncorrelated with the school attendance indicators, D_{ij}.

We are interested in how OLS regression estimates compare with the causal parameters in equation (4). We therefore define population regression coefficients in a model with the same conditioning variables:

$$Y_i = \alpha_0 + \sum_{j=1}^{J} \alpha_j D_{ij} + X_i'\Gamma + v_i. \quad (5)$$

This is a population projection, so the residuals in this model, v_i, are necessarily orthogonal to all right-hand-side variables, including school attendance dummies.

Regression model (5) has a causal interpretation when the parameters in this equation coincide with those in the causal model, equation (4). This in turn requires that school choices be unrelated to the unobserved component of student ability, an assumption that can be expressed as:

$$E[\epsilon_i | D_{ij}] = 0; \; j = 1, ..., J. \quad (6)$$

Restriction (6), sometimes called "selection-on-observables," means that $\alpha_j = \beta_j$ for each school. In practice, of course, regression estimates need not have a causal interpretation; rather, they may be biased. This possibility is represented by writing

$$\alpha_j = \beta_j + b_j,$$

where the bias parameter b_j is the difference between the regression and causal parameters for school j.

4 Validating Conventional VAMs

4.1 Test Procedure

The variation in school attendance generated by oversubscribed admission lotteries allows us to assess the causal interpretation of conventional VAM estimates. A vector of dummy variables, $Z_i = (Z_{i1}, .., Z_{iL})'$, indicates lottery offers to student i for seats at L oversubscribed schools. Offers at school ℓ are randomly assigned conditional on a set of lottery-specific stratifying variables, $C_{i\ell}$. These variables include an indicator for applicants to school ℓ and possibly other variables such as application cohort and walk zone status. The vector $C_i = (C'_{i1}, .., C'_{iL})'$ collects these variables across all lotteries. The models used here also add the OLS VAM controls (X_i in equation (5)) to the vector C_i to increase precision.

We assume that lottery offers are (conditionally) mean-independent of student ability. In other words,

$$E[\epsilon_i|C_i, Z_i] = \lambda_0 + C'_i\lambda_c, \tag{7}$$

for a set of parameters λ_0 and λ_c. This implies that admission offers are valid instruments for school attendance after controlling for lottery assignment strata, an assumption that underlies recent lottery-based analyses of school effectiveness (Cullen et al., 2006; Abdulkadiroğlu et al., 2011; Deming et al., 2014).

With fewer lotteries than schools (that is, when $L < J$), the restrictions in (7) are insufficient to identify the parameters of the causal model, equation (4). Even so, these restrictions can be used to test the selection-on-observables assumption. Equations (6) and (7) imply that $L + J$ orthogonality conditions are available to identify J school effects β_j. The resulting L overidentifying restrictions generate an over-identification test of the sort widely used with instrumental variables (IV) estimators.

To describe the over-identification test statistic, let Z denote the $N \times L$ matrix of lottery offers for a sample of N students, and let C denote the corresponding matrix of stratifying variables, with associated projection matrix $P_C = C(C'C)^{-1}C$ and annihilator matrix $M_C = I - P_C$. A Lagrange multiplier (LM) over-identification test statistic, associated with two-stage least squares (2SLS) models estimated assuming homoskedasticity, can be written:

$$\hat{T} \quad = \quad \frac{\hat{\epsilon}' P_{\tilde{Z}} \hat{\epsilon}}{\hat{\sigma}^2_{\hat{\epsilon}}}, \tag{8}$$

where $P_{\tilde{Z}} = M_C Z(Z' M_C Z)^{-1} Z' M_C$ is the lottery offer projection matrix after partialling out randomization strata, $\hat{\epsilon}$ is an $N \times 1$ vector of OLS VAM residuals (since OLS and 2SLS coincide when D_{ij} itself is in the instrument list), and $\hat{\sigma}^2_{\hat{\epsilon}} = \hat{\epsilon}' M_C \hat{\epsilon}/N$ is an estimate of the residual variance of ϵ_i partialling out strata effects. Under the joint null hypothesis described by selection-on-observables and lottery exclusion (equations (6) and (7)), the statistic $\hat{T}$ has an asymptotic χ^2_L distribution.[7]

A simple decomposition of $\hat{T}$ reveals an important connection with previously used VAM validity tests.

[7] The test statistic in (8) is derived assuming homoskedastic errors. An analogous test allowing heteroskedasticity uses a White (1980) robust covariance matrix to test that coefficients on lottery offers equal zero in a regression of $\hat{\epsilon}_i$ on Z_i and C_i.

Let $\hat{Y}_i$ denote the fitted values generated by OLS VAM estimation (computed from regression model (5)), and let Y and $\hat{Y}$ denote $N \times 1$ vectors collecting individual Y_i and $\hat{Y}_i$. The LM statistic can be written

$$\begin{aligned} \hat{T} &= \frac{((Y - \hat{\varphi}\hat{Y}) + (\hat{\varphi} - 1)\hat{Y})' P_{\tilde{Z}} ((Y - \hat{\varphi}\hat{Y}) + (\hat{\varphi} - 1)\hat{Y})}{\hat{\sigma}^2_{\tilde{\epsilon}}} \\ &= \frac{(\hat{\varphi} - 1)^2}{\hat{\sigma}^2_{\tilde{\epsilon}} (\hat{Y}' P_{\tilde{Z}} \hat{Y})^{-1}} + \frac{(Y - \hat{\varphi}\hat{Y})' P_{\tilde{Z}} (Y - \hat{\varphi}\hat{Y})}{\hat{\sigma}^2_{\tilde{\epsilon}}}. \end{aligned} \tag{9}$$

Here, the scalar $\hat{\varphi} = (\hat{Y}' P_{\tilde{Z}} \hat{Y})^{-1} \hat{Y}' P_{\tilde{Z}} Y$ is the 2SLS estimate from a model that uses lottery offers as instruments in an equation with Y_i on the left and $\hat{Y}_i$, treated as endogenous, on the right. Equation (9) shows that the omnibus test statistic $\hat{T}$ combines two terms. The first is a one-degree-of-freedom Wald-type test statistic for $\hat{\varphi} = 1$ (note that the denominator of this term estimates the asymptotic variance of $\hat{\varphi}$). The second is the Sargan (1958) statistic for testing the $L - 1$ overidentifying restrictions generated by the availability of L instruments to estimate this single parameter.[8]

In what follows, the estimate $\hat{\varphi}$ is called a "forecast coefficient." This connects $\hat{T}$ with tests of "forecast bias" implemented in previous VAM validation efforts (Kane and Staiger, 2008; Kane et al., 2013; Deming, 2014; Chetty et al., 2014a). These earlier tests similarly ask whether the coefficient on predicted value-added equals one in IV procedures relating outcomes to VAM fitted values (though the details sometimes differ). Forecast bias arises when VAM estimates for a group of schools are off the mark, a failure of average predictive validity. Importantly, the omnibus test statistic, $\hat{T}$, checks more than forecast bias: this statistic asks whether *each* over-subscribed lottery generates score gains commensurate with the gains predicted by OLS VAM.

4.2 Test Results

The conventional VAM setup assessed here includes two value-added specifications. The first, referred to as the "lagged score" model, includes indicators for sex, race, subsidized lunch eligibility, special education status, English-language learner status, and counts of baseline absences and suspensions, along with cubic functions of baseline math and ELA test scores. Specifications of this type are at the heart of the econometric literature on value-added models (Kane et al., 2008; Rothstein, 2010; Chetty et al., 2014a). The second, a "gains" specification, uses grade-to-grade score changes as the outcome variable and includes all controls from the lagged score model except baseline test scores. This model parallels widely used accountability policies that measure test score growth.[9] As in Rothstein (2009), we benchmark the extent of cross-school

[8] Angrist et al. (2016b) interpret VAM validity tests using the general theory of specification testing developed by Newey (1985) and Newey and West (1987). In practice, Wald and LM test statistics typically use different variance estimators in the denominator.

[9] The gains specification can be motivated as follows: suppose that human capital in grade g, denoted A_{ig}, equals lagged human capital plus school quality, so that $A_{ig} = A_{ig-1} + q_{ig}$ where $q_{ig} = \sum_j \beta_j D_{ij} + \eta_{ig}$ and η_{ig} is a random component independent of school choice. Suppose further that test scores are noisy proxies for human capital, so that $Y_{ig} = A_{ig} + \nu_{ig}$ where ν_{ig} is classical measurement error. Finally, suppose that school choice in grade g is determined solely by A_{ig-1} and variables unrelated to achievement. Then a lagged score model that controls for Y_{ig-1} generates biased estimates, but a gains model with $Y_{ig} - Y_{ig-1}$ as the outcome variable measures value-added correctly.

ability differences by also estimating an "uncontrolled" model that adjusts only for year effects. Although the uncontrolled model almost certainly provides a poor measure of school value added, many districts distribute school report cards based on unadjusted test score levels.[10]

Figure 1 summarizes the value-added estimates generated by sixth-grade math scores. We focus on math scores because value-added for math appears to be more variable across schools than value-added for ELA (bias tests for ELA, presented in Appendix Table A2, yield similar results). Each bar in Figure 1 reports an estimated standard deviation of α_j across schools, expressed in test score standard deviation units and adjusted for estimation error.[11] Adding controls for demographic variables and previous scores reduces the standard deviation of α_j from 0.5σ in the uncontrolled model to about 0.2σ in the lagged score and gains models. This shows that observed student characteristics explain a substantial portion of the variation in school averages. The last three bars in Figure 1 report estimates of within-sector value-added standard deviations, constructed using residuals from regressions of $\hat{\alpha}_j$ on dummies for schools in the charter and pilot sectors. Controlling for sector effects reduces variation in α_j, reflecting sizable differences in average conventional value-added across sectors.

Panel A of Table 3 summarizes test results for sixth grade math VAMs. The first row shows the forecast coefficient, $\hat{\varphi}$. The estimator used here is a version of the optimal IV procedure for heteroskedastic models described by White (1982). The second row reports first stage F-statistics measuring the strength of the relationship between lottery offers and predicted value-added. With a weak first stage, forecast coefficient estimates may be biased towards the corresponding OLS estimand, that is, the coefficient from a regression of test scores on VAM fitted values. In simple models this regression coefficient must equal one, so a weak first stage makes a test of $H_0 : \varphi = 1$ less likely to reject.[12] First-stage F-statistics for the sixth grade lagged score and gains models are close to 30, suggesting finite-sample bias is not an issue in the full lottery sample. First-stage strength is more marginal, however, when charter lotteries are omitted.

Table 3 reports p-values for three VAM validity tests. The first is for forecast bias, that is, the null hypothesis that the forecast coefficient equals one. The second tests the associated set of overidentifying restrictions, which require that just-identified IV estimates of the forecast coefficient be the same for each lottery instrument, though not necessarily equal to one. The third omnibus test combines these restrictions.

On average, VAM fitted values predict the score gains generated by random assignment remarkably well. This can be seen in columns 1 and 2 of Panel A in Table 3, which show that the lagged score and gains specifications generate forecast coefficients equal to 0.86 and 0.95; the former is only marginally statistically different from one ($p = 0.07$), while the second has $p = 0.55$. At the same time, the over-identification and

[10] California's School Accountability Report Cards list school proficiency levels (see `http://www.sarconline.org`). Massachusetts' school and district profiles provide information on proficiency levels and test score growth (see `http://profiles.doe.mass.edu`).

[11] The estimated standard deviations plotted in the figure are given by $\hat{\sigma}_\alpha = (\frac{1}{J}\sum_j[(\hat{\alpha}_j - \hat{\mu}_\alpha)^2 - SE(\hat{\alpha}_j)^2])^{1/2}$, where $\hat{\mu}_\alpha$ is mean value-added and $SE(\hat{\alpha}_j)$ is the standard error of $\hat{\alpha}_j$.

[12] When estimated in the same sample with no additional controls, OLS regressions on OLS fitted values necessarily produce a coefficient of one. In practice, the specification used here to test VAM differs from the model producing fitted values in that it also controls for lottery strata and excludes non-lotteried students.

omnibus tests reject for both models.[13]

The source of these rejections can be seen in Figure 2, which plots reduced form estimates of the effects of lottery offers on test scores against corresponding first-stage effects of lottery offers on conventional VAM fitted values for sixth grade math scores. Each panel also shows a line through the origin with slope equal to the forecast coefficient reported in Table 3 (plotted as a solid line) along with a dashed 45-degree line. In other words, Figure 2 gives a visual instrumental variables representation of the forecast coefficient: VAM models that satisfy equation (6) should generate points along the 45-degree line, with deviations due solely to sampling error. Though the lines of best fit have slopes close to one, points for many lotteries are farther from the diagonal than sampling variance alone would lead us to expect. Earlier validation strategies focus on forecast coefficients, ignoring overidentifying restrictions. Figure 2 shows that such strategies may fail to detect substantial deviations between conventional VAM predictions and reduced form lottery effects for some lotteries.

Figure 2 also suggests that a good portion of conventional VAM estimates' predictive power for Boston schools comes from charter school lotteries, which contribute large first stage and reduced form effects. The relationship between OLS value-added and lottery estimates is weaker in the traditional public and pilot school sectors. This is confirmed in columns 3 and 4 of Table 3, which report results of VAM bias tests for sets of instruments that exclude charter lotteries. At 0.55 and 0.68, estimated forecast coefficients from traditional public and pilot lotteries are further from one than the coefficients computing using all lotteries. Although removal of charter lotteries reduces precision, omnibus tests continue to reject at the 1-percent level.[14]

Finally, Panel B of Table 3 reports test results combining data from sixth through eighth grade. As in Abdulkadiroğlu et al. (2011) and Dobbie and Fryer (2013), school effects on seventh and eighth grade scores are modeled as linear in the number of years spent in each school. Given constant linear school enrollment effects, regressions of later-grade outcomes on baseline controls and years of enrollment in each school recover causal school effects in the absence of sorting on unobserved ability. The omnibus VAM validity test in this case regresses residuals from the multi-grade (stacked) model on sixth grade lottery offers while the forecast coefficient is generated by using lottery offers to instrument OLS VAM fitted values from the multi-grade model. The omnibus tests show clear rejections in the multi-grade set-up as well as for sixth grade only, in spite of the fact that the first-stage F-statistics are noticeably lower.

[13] As a point of comparison, Angrist et al. (2016b) report tests of VAM validity in the Charlotte-Mecklenberg lottery data analyzed by Deming (2014). There as well the forecast coefficient is close to one, while the omnibus test generates a p-value of 0.02.

[14] The first stage F-statistics for the specifications without charter lotteries are 11.2 and 9.3, suggesting weak instruments might be a problem in this model. It is encouraging, therefore, that limited information maximum likelihood (LIML) forecast coefficient estimates are virtually the same as the estimates reported in Table 3. A related concern is whether the heteroskedastic-robust standard errors and test statistics used in Table 3 are misleading due to common school-year shocks (as suggested by Kane and Staiger (2002) for teachers). Reassuringly, cluster-robust test results are also similar to those in Table 3.

4.3 Heterogeneity vs. Bias

The omnibus test results reported in Table 3 suggest conventional VAM estimates fail to predict the effects of lottery offers perfectly. This is consistent with bias in OLS VAMs. In a world of heterogeneous causal effects, however, these rejections need not reflect selection bias. Rather, they might signal divergence between the local average treatment effects (LATEs) identified by lottery instruments and possibly more representative effects captured by OLS, at least for some lotteries (Imbens and Angrist, 1994; Angrist et al., 1996). Moreover, with unrestricted potential outcomes, even internally valid OLS VAM estimates (that is, those satisfying selection-on-observables) capture weighted average causal effects that need not match average effects for the entire sample of students attending particular schools (Angrist, 1998).

Three analyses shed light on the distinction between heterogeneity and bias. The first is a set of bias tests using OLS VAM specifications that allow school effects to differ across covariate-defined subsamples (e.g. special education students or those with low levels of baseline achievement). This approach accounts for variation in school effects across covariate cells that may be weighted differently by IV and OLS. The second analysis tests for bias in OLS VAMs estimated in the lottery sample. This asks whether differences between IV and OLS are caused by differences between students subject to lottery assignment and the general student population. The final analysis estimates OLS VAM separately for applicants who respond to lottery offers ("compliers") and for other groups, within the sample of lottery applicants.

Estimates by subgroup, reported in Panel A of Table 4, consistently generate rejections in omnibus tests of VAM validity. Column 2 reports test results allowing VAM estimates to differ by year, thereby accommodating "drift" in school effects over time (Chetty et al. (2014a) document such drift in teacher value-added); columns 3-5 show results for subgroups defined by subsidized lunch eligibility, special education status, and baseline test score terciles; and column 6 reports results from models that allow value-added to differ across cells constructed by fully interacting race, sex, subsidized lunch eligibility, special education, English-language learner status, and baseline score tercile. The forecast coefficients and omnibus test statistics generated by each of these subgroup schemes are similar to those for the full sample. As can be seen in panel B of Table 4, test results for models that use only the quasi-experimental sample for OLS VAM estimation are also similar to the full sample results. This suggests that rejection of the omnibus test is not driven by differences in OLS VAM between students subject to random assignment and the general population.[15]

Lottery-based IV estimates identify average causal effects for compliers, that is, lottery applicants whose attendance choices shift in response to random offers, rather than the full population of students that enroll in a particular school. To investigate the link between lottery compliance and treatment effects, we predict

[15] In a subset of the data used here, Walters (2014) documents a link between the propensity to apply to Boston charter schools and the causal effect of charter school attendance. This finding is not at odds with our constant effects assumption because Walters studies the effects of charter schools relative to a heterogeneous mix of traditional public schools, while we allow a distinct effect for every traditional public school. The effect heterogeneity uncovered by Walters may reflect variation in the quality of fallback public school options across charter applicants. Consistent with this possibility, Walters' Figure 2 demonstrates that the relationship between charter application choices and causal effects is driven primarily by heterogeneity in outcomes at fallback traditional public schools.

value-added at the target school for every lottery applicant based on covariate-specific OLS estimates from the model in column 6 of Table 4 (estimated in the lottery sample). Maintaining the hypothesis of OLS VAM validity, we allow for the possibility that heterogeneous effects are reflected in a set of covariate-specific estimates. These predictions are then used to compare an imputed average value-added for compliers to imputed average value-added for "never takers" (those who decline lottery offers) and "always takers" (those who enroll in the target school even when denied an offer) in each lottery. Averages for the three lottery compliance groups are estimated using methods described in Abadie (2003) (see Appendix B.1).

Figure 3 shows that imputed OLS value-added estimates for compliers, always takers, and never takers are similar. Formal tests for equality fail to reject the hypotheses that predicted effects for compliers equal predicted effects for always takers ($p = 0.72$) or never takers ($p = 0.39$). This suggests that lottery compliance is not a major source of treatment effect heterogeneity linked to observable characteristics, though we cannot rule out unobserved differences between compliers and other groups.

5 The Distribution of School Effectiveness

The test results in Table 3 suggest conventional VAM estimates are biased. At the same time, OLS VAM estimates tend to predict lottery impacts on average, with estimated forecast coefficients close to one. OLS estimates would therefore seem to be useful even if imperfect. This section develops a hybrid estimation strategy that combines lottery and OLS estimates in an effort to quantify the bias in conventional VAMs and produce more accurate value-added estimates.

5.1 A Random Coefficients Lottery Model

The hybrid estimation strategy uses a random coefficients model to describe the joint distribution of value-added, bias, and lottery compliance across schools. The model is built on a set of OLS, lottery reduced form, and first stage estimates. The OLS estimates come from equation (5), while the lottery reduced form and first stage equations are:

$$Y_i = \tau_0 + C_i'\tau_c + Z_i'\rho + u_i, \tag{10}$$

$$D_{ij} = \phi_{0j} + C_i'\phi_{cj} + Z_i'\pi_j + \eta_{ij};\ j = 1, ..., J.$$

Assumption (7) implies that the reduced form effect of admission in lottery ℓ is given by

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

where ρ_ℓ and $\pi_{\ell j}$ are the elements of ρ and π_j corresponding to $Z_{i\ell}$. This expression shows that the lottery at school ℓ identifies a linear combination of value-added parameters, with coefficients $\pi_{j\ell}$ equal to the shares of students shifted into or out of each school by the ℓth lottery offer.

OLS, reduced form, and first stage estimates are modeled as noisy measures of school-specific parameters, which are in turn modeled as draws from a distribution of random coefficients in the population of schools. Specifically, we have:

$$\hat{\alpha}_j = \beta_j + b_j + e_j^{\alpha},$$

$$\hat{\rho}_\ell = \sum_j \pi_{\ell j}\beta_j + e_\ell^{\rho}, \tag{11}$$

$$\hat{\pi}_{\ell j} = \pi_{\ell j} + e_{\ell j}^{\pi};$$

where $e_j^{\alpha}, e_\ell^{\rho}$ and $e_{\ell j}^{\pi}$ are mean-zero estimation errors that vanish as within-school and within-lottery samples tend to infinity. Subject to the usual asymptotic approximations, these errors can be modeled as normally distributed with a known covariance structure. Table 1 shows that the OLS and lottery estimation samples used here typically include hundreds of students per school, so the use of asymptotic results seems justified.

The second level of the model treats the school-specific parameters β_j, b_j, and $\{\pi_{\ell j}\}_{\ell=1}^{L}$ as draws from a joint distribution of causal effects, bias, and lottery compliance patterns. The effect of admission at school ℓ on the probability of attending this school is parameterized as

$$\pi_{\ell\ell} = \frac{\exp(\delta_\ell)}{1+\exp(\delta_\ell)}, \tag{12}$$

where the parameter δ_ℓ can be viewed as the mean utility in a binary logit model predicting compliance with a random offer of a seat at school ℓ. Likewise, the effect of an offer to attend school $\ell \neq j$ on attendance at school j is modeled as

$$\pi_{\ell j} = -\pi_{\ell\ell} \times \frac{\exp(\xi_j + \nu_{\ell j})}{1+\sum_{k\neq \ell}\exp(\xi_k + \nu_{\ell k})}. \tag{13}$$

In this expression, the quantity $\xi_j + \nu_{\ell j}$ is the mean utility for school j in a multinomial logit model predicting alternative school choices among students that comply with offers in lottery ℓ. The parameter ξ_j allows for the possibility that some schools are systematically more or less likely to serve as fallback options for lottery losers, while $\nu_{\ell j}$ is a random utility shock specific to school j in the lottery at school ℓ. The parametrization in (12) and (13) ensures that lottery offers increase the probability of enrollment at the target school and reduce enrollment probabilities at other schools, and that effects on all probabilities are between zero and one in absolute value.

Each school is characterized by a vector of four parameters: a value-added coefficient, β_j; a selection bias term, b_j; an offer compliance utility, δ_j; and a mean fallback utility, ξ_j. These are modeled as draws from a prior distribution in a hierarchical Bayesian framework. A key assumption in this framework is that the distribution of VAM bias is the same for schools with and without oversubscribed lotteries. This assumption allows the model to "borrow" information from schools with lotteries and to generate posterior distributions for non-lottery schools that account for bias in conventional VAM estimates. Importantly, however, we allow for the possibility that average value-added may differ between schools with and without lotteries. Section 6.2 investigates the empirical relationship between oversubscription and bias.

Let Q_j denote an indicator for whether quasi-experimental lottery data are available for school j. School-specific parameters are modeled as draws from the following conditional multivariate normal distribution:

$$(\beta_j, b_j, \delta_j, \xi_j)|Q_j \sim N\left((\beta_0 + \beta_Q Q_j, b_0, \delta_0, \xi_0), \Sigma\right). \tag{14}$$

The parameter β_Q captures the possibility that average value-added differs for schools with lotteries. The matrix Σ describes the variances and covariances of value-added, bias, and first stage utility parameters, and is assumed to be the same for lottery and non-lottery schools. Finally, lottery and school-specific utility shocks are also modeled as conditionally normal:

$$\nu_{\ell j}|Q_j \sim N\left(0, \sigma_\nu^2\right). \tag{15}$$

The vector $\theta \equiv (\beta_0, \beta_Q, b_0, \delta_0, \xi_0, \Sigma, \sigma_\nu^2)$ collects the hyperparameters governing the prior distribution of school-specific parameters. Our empirical Bayes (EB) framework first estimates these hyperparameters and then uses the estimated prior distribution to compute posterior value-added predictions for individual schools. Some of the specifications considered below extend the setup outlined here to allow the prior mean vector $(\beta_0, b_0, \delta_0, \xi_0)$ to vary across school sectors (traditional, charter, and pilot).

5.2 Simulated Minimum Distance Estimation

We estimate hyperparameters by simulated minimum distance (SMD), a variant of the method of simulated moments (McFadden, 1989). SMD focuses on moments that are determined by the parameters of interest, choosing hyperparameters to minimize deviations between sample moments and the corresponding model-based predictions. Our SMD implementation uses means, variances, and covariances of functions of the OLS value-added estimates, $\hat{\alpha}_j$, lottery reduced forms, $\hat{\rho}_\ell$, and first stage coefficients, $\hat{\pi}_{\ell j}$. For example, one moment to be fit is the average $\hat{\alpha}_j$ across schools; another is the cross-school variance of the $\hat{\alpha}_j$. Other moments are means and variances of reduced form and first stage estimates across lotteries. Appendix B.2 lists the moments used for SMD estimation.

The fact that the moments in this context are complicated nonlinear functions of the hyperparameters motivates a simulation approach. For example, the mean reduced form is $E[\rho_\ell] = \sum_j E\left[\pi_{\ell j}\beta_j\right]$. This is the expectation of the product of a normally distributed random variable with a ratio involving correlated log-normals (described by (12) and (13)), a moment for which no analytical expression is readily available. Moments are therefore simulated by fixing a value of θ and drawing a vector of school-level parameters using equations (14) and (15). Likewise, the simulation draws a vector of the estimation errors in (11) from the joint asymptotic distribution of the OLS, reduced form and first stage estimates. The parameter and estimation draws are combined to generate a simulated vector of parameter estimates for the given value of θ. Finally, these are used to construct a set of model-based predicted moments. The SMD estimator minimizes a quadratic form in the difference between predicted moments and the corresponding moments observed in the data. As described in Appendix B.2, the SMD estimates reported here are generated by a two-step procedure with an efficient weighting matrix in the second step.

5.3 Empirical Bayes Posteriors

Studies of teacher and school value-added typically employ EB strategies that shrink noisy teacher- and school-specific value-added estimates towards the grand mean, reducing mean squared error (see, e.g., Kane et al., 2008 and Jacob and Lefgren, 2008). In a conventional VAM model where OLS estimates are presumed unbiased, the posterior mean value-added for school j is

$$E\left[\alpha_j|\hat{\alpha}_j\right] = \left(\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + Var(e_j^\alpha)}\right)\hat{\alpha}_j + \left(1 - \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + Var(e_j^\alpha)}\right)\alpha_0, \tag{16}$$

where α_0 and σ_α^2 are the mean and variance of the conventional OLS VAM parameters α_j. An EB posterior mean plugs estimates of these hyperparameters into (16).

Our setup extends this idea to a scenario where the estimated $\hat{\alpha}_j$ may be biased but lotteries are available to reduce this bias. The price for bias reduction is a loss of precision: because IV uses only the variation generated by random assignment, lottery-based estimates are less precise than the corresponding OLS estimates. Because some schools are undersubscribed, there are also fewer lottery instruments than schools and a VAM is not identified using lotteries alone. Even so, in the spirit of the combination estimators discussed by Judge and Mittlehammer (2004; 2005; 2007), our empirical Bayes approach trades off the advantages and disadvantages of OLS and IV to construct minimum mean squared error (MMSE) estimates of value-added.

To see how this trade-off works, suppose the first stage parameters, $\pi_{\ell j}$, are known rather than estimated (equivalently, $e_{\ell j}^\pi = 0\ \forall \ell, j$). Let Π denote the $L \times J$ matrix of these parameters, and let β, $\hat{\alpha}$ and $\hat{\rho}$ denote vectors collecting β_j, $\hat{\alpha}_j$ and $\hat{\rho}_\ell$. Appendix B.3 shows that the posterior distribution for β in this case is multivariate normal with mean:

$$E\left[\beta|\hat{\alpha}, \hat{\rho}\right] = W_\alpha(\hat{\alpha} - b_0\iota) + W_\rho\hat{\rho} + (I - W_\alpha - W_\rho\Pi)\beta_0\iota, \tag{17}$$

where ι is a $J \times 1$ vector of ones. Posterior mean value-added is a linear combination of OLS estimates net of mean bias, $(\hat{\alpha} - b_0\iota)$, lottery reduced form estimates, $\hat{\rho}$, and mean value-added, $\beta_0\iota$. The weighting matrices, W_α and W_ρ, are functions of the first stage parameters and the covariance matrix of estimation error, value-added, and bias. Expressions for these matrices appear in Appendix B.3. As with conventional EB posteriors, an empirical Bayes version of the posterior mean plugs first-step estimates of b_0, β_0, W_α, and W_ρ into equation (17).

Suppose that all schools are oversubscribed, so that $L = J$. In this case, the first stage matrix Π is square; if it is also full rank, the parameters of equation (4) are identified using lotteries alone. Lottery-based value-added estimates may then be computed by indirect least squares as $\hat{\beta} = \Pi^{-1}\hat{\rho}$, and the posterior mean in equation (17) becomes

$$E\left[\beta|\hat{\alpha},\hat{\beta}\right] = W_\alpha(\hat{\alpha} - b_0\iota) + W_\beta\hat{\beta} + (I - W_\alpha - W_\beta)\beta_0\iota, \tag{18}$$

for $W_\beta = W_\rho\Pi$. This expression reveals that when a lottery-based value-added model is identified, the posterior mean for value-added is a matrix-weighted average of three quantities: quasi-experimental IV estimates, conventional OLS estimates net of mean bias, and prior mean value-added, with weights (that sum to the identity matrix) optimally chosen to minimize mean-squared error.

In the same spirit as our hybrid strategy, Chetty and Hendren (2015) combine noisy quasi-experimental estimates of neighborhood effects based on movers with precise averages of permanent resident outcomes to generate optimal forecasts of neighborhood causal effects. A further special case of equation (18) illuminates the link between this approach and ours. Suppose the estimation error in OLS estimates is negligible ($Var(e_j^\alpha) = 0$), and that IV estimation error e_j^β is uncorrelated across schools. Appendix B.3 shows that under these simplifying assumptions, the jth element of equation (18) becomes

$$E\left[\beta_j|\hat{\alpha},\hat{\beta}\right] = \left(\frac{\sigma_\beta^2(1-R^2)}{Var(e_j^\beta)+\sigma_\beta^2(1-R^2)}\right)\hat{\beta}_j + \left(1 - \frac{\sigma_\beta^2(1-R^2)}{Var(e_j^\beta)+\sigma_\beta^2(1-R^2)}\right)\left(r_\alpha(\hat{\alpha}_j - b_0) + (1-r_\alpha)\beta_0\right), \tag{19}$$

where σ_β^2 is the variance of β_j, $r_\alpha = Cov(\beta_j,\alpha_j)/Var(\alpha_j)$ is the reliability ratio from a regression of causal value-added on OLS value-added, and R^2 is the R-squared from this regression. This expression coincides with equation (21) in Chetty and Hendren (2015) and can also be seen to be the same as the canonical empirical Bayes shrinkage formula in equation (1.5) of Morris (1983).[16]

In practice, some schools are undersubscribed, so IV estimates of individual school value-added cannot be computed. Nevertheless, equation (17) shows that predictions at schools without lotteries can be improved using lottery information from other schools. Lottery reduced form parameters embed information for all fallback schools, including those without lotteries. This is a consequence of the relationship described by equation (11), which shows that the reduced form for any school with a lottery depends on the value-added of all other schools that applicants to this school might attend. Specifically, as long as $\pi_{\ell j} \neq 0$, the reduced form for lottery ℓ contains information that can be used to improve the posterior prediction of β_j. The test results in columns 2 and 3 of Table 5 shows that estimates of $\pi_{\ell j}$ are significantly different from zero (at 5 percent) for 12 of the 22 undersubscribed schools in our sample. The ten schools not on this list have primary entry grades other than sixth. In other words, oversubscribed sixth grade lotteries contribute information on all schools with sixth grade entry.

[16] The connection with Morris can be made by observing that when $\hat{\alpha}_j = \alpha_j$, the term $r_\alpha(\hat{\alpha}_j - b_0) + (1 - r_\alpha)\beta_0$ is the fitted value from the regression of β_j on α_j. Chetty and Hendren (2015) normalize mean value-added and bias to $\beta_0 = b_0 = 0$, rearranging (19) to read:

$$E\left[\beta_j|\hat{\alpha}_j,\hat{\beta}_j\right] = r_\alpha\hat{\alpha}_j + \left(\frac{\sigma_\beta^2(1-R^2)}{Var(e_j^\beta)+\sigma_\beta^2(1-R^2)}\right)\left(\hat{\beta}_j - r_\alpha\hat{\alpha}_j\right).$$

Finally, equation (17) also reveals how knowledge of conventional VAM bias can be used to improve posterior predictions even for schools that are never lottery fallbacks. Appendix B.3 shows that the posterior mean for β_j gives no weight to $\hat{\rho}$ when $\pi_{\ell j} = 0$ and $Cov(e_j^{\alpha}, e_{\ell}^{\rho}) = 0$ across all lotteries, ℓ. In this case the posterior mean for β_j simplifies to

$$E\left[\beta_j|\hat{\alpha}, \hat{\rho}\right] = r_{\alpha}(\hat{\alpha}_j - b_0) + (1 - r_{\alpha})\,\beta_0. \tag{20}$$

Even without a lottery at school j, predictions based on equation (20) improve upon the conventional VAM posterior given by equation (16). The improvement here comes from the fact that the schools with lotteries provide information that can be used to determine the reliability of conventional VAM estimates.[17]

Equations (17) through (20) are pedagogical formulas derived assuming first stage parameters are known. With an estimated first stage, the posterior distribution for value-added does not have a closed form. Although the posterior mean for the general case can be approximated using Markov Chain Monte Carlo (MCMC) methods, with a high-dimensional random coefficient vector, MCMC may be sensitive to starting values or other tuning parameters. We therefore report EB posterior modes (as in Chamberlain and Imbens (2004); these are also known as maximum *a posteriori* estimates). The posterior mode is relatively easily calculated, and coincides with the posterior mean when value-added is normally distributed as in the fixed first stage case (see Appendix B.4 for details). As a practical matter, the posterior modes for value-added turn out to be similar to the weighted averages generated by equation (17) under the fixed first stage assumption, with a correlation across schools of 0.95 in the lagged score model (see Appendix Figure A1).

6 Parameter Estimates

6.1 Hyperparameters

The SMD procedure for estimating hyperparameters takes as input a set of lottery reduced form and first stage estimates, along with conventional VAM estimates for each value-added model. The lottery estimates come from regressions of test scores and school attendance indicators (the set of D_{ij}) on lottery offer dummies (Z_i), with controls C_i for randomization strata and the baseline covariates from the lagged score VAM specification (strata controls are necessary for instrument validity, while baseline covariates increase precision). Combining the lottery estimates with OLS estimates of the α_j generates hyperparameter estimates for a particular value-added model.

As can be seen in columns 1-3 of Table 6, the hyperparameter estimates reveal substantial variation in both causal value-added and selection bias across schools. The standard deviation of value-added, σ_β,

[17] Using the fact that $\alpha_j = \beta_j + b_j$, equation (16) can be written to look more like equation (20):

$$E\left[\alpha_j|\hat{\alpha}_j\right] = r_{\alpha}\left(\frac{\sigma_\beta^2+\sigma_b^2+2\sigma_{\beta b}}{\sigma_\beta^2+\sigma_{\beta b}}\right)(\hat{\alpha}_j - b_0) + \left(1 - r_{\alpha}\left(\frac{\sigma_\beta^2+\sigma_b^2+2\sigma_{\beta b}}{\sigma_\beta^2+\sigma_{\beta b}}\right)\right)\beta_0 + b_0,$$

This adds bias, b_0, to a weighted average of bias-corrected OLS and global mean value-added.

is similar across specifications, ranging from about 0.20σ in the uncontrolled specification to 0.22σ in the lagged score and gains models. This stability is reassuring: the control variables that distinguish these models should not change the underlying distribution of causal school effectiveness if our estimation procedure works as we hope.

In contrast with the relatively stable estimates of σ_β, the estimated standard deviation of bias, σ_b, shrinks from 0.50σ with no controls to under 0.2σ in the lagged score and gains specifications. In other words, controlling for observed student characteristics and past scores reduces bias in conventional value-added estimates markedly. On the other hand, the estimated standard deviations of bias are statistically significant for all models, implying that controls for demographic variables and baseline achievement are not sufficient to produce unbiased comparisons. Columns 2 and 3 of Table 6 show that the standard deviations of bias in the lagged score and gains models equal 0.18σ and 0.17σ, slightly smaller than the standard deviation of causal value-added.[18]

Earlier work on school effectiveness explores differences between Boston's charter, pilot, and traditional public sectors (Abdulkadiroğlu et al., 2011; Angrist et al., 2016a). These estimates show strong charter school treatment effects in Boston, a finding that suggests accounting for sector differences may improve the predictive accuracy of school value-added models. Columns 4 and 5 of Table 6 therefore report estimates of lagged score and gains models in which the means of the random coefficients depend on school sector (Appendix Table A3 reports the complete set of parameter estimates for the lagged score model). Consistent with earlier findings, models with sector effects show that average charter school value-added exceeds traditional public school value-added by roughly 0.4σ. Estimated differences in value-added between pilot and traditional public schools are smaller and statistically insignificant. By contrast, bias seems unrelated to sector, implying that conventional VAM models with demographic and lagged achievement controls accurately reproduce lottery-based comparisons of the charter, pilot and traditional sectors (also consistent with the findings of Abdulkadiroğlu et al. (2011)). The estimates of σ_β and σ_b show that sector effects reduce cross-school variation in both value-added and bias by about 20-25 percent. The large charter effect on value added notwithstanding, most of the variation in middle school quality in Boston is within sectors rather than between.

Estimated covariances between β_j and b_j, denoted $\sigma_{\beta b}$, are negative and mostly statistically significant, a result that can be seen in the third row of Table 6. A negative covariance between value-added and bias suggests that conditional on demographics and past achievement, students with higher ability tend to enroll in schools with lower value-added. Conventional VAMs therefore overestimate the effectiveness of low-quality schools and underestimate the effectiveness of high-quality schools. Estimates of β_Q, the lottery school value-added shifter, are close to zero in models without sector effects, and positive but small when sector effects are included. The estimate of β_Q for the lagged score model is statistically significant, implying that schools with lotteries are slightly more effective than undersubscribed schools in the same sector.

[18] Rothstein (2009) tests for bias in teacher VAMs using granger-type causality tests that regress lagged test scores on future teacher dummies. Like our random coefficients model, these tests generate estimates of the standard deviation of bias in VAM estimates.

Studies of teacher value-added emphasize the reliability ratio $r_\alpha = Cov(\alpha_j, \beta_j)/Var(\alpha_j)$ as a summary measure of the predictive value of VAM estimates (Chetty et al., 2014a; Rothstein, 2015).[19] The fourth row of Table 6 reports model-based estimates of this parameter. The estimated reliability of the uncontrolled specification equals 0.08 with a standard error of 0.20, implying that school average test scores are only weakly related to school effectiveness. Reliability ratios in the lagged score and gains models equal 0.64 and 0.75 in models without sector effects, and 0.69 and 0.78 in models with sector effects. Consistent with the test results in Section 4, these estimates show that conventional VAM estimates are strongly, but not perfectly, linked to causal school quality.

6.2 School Characteristics, Value-added, and Bias

The individual school value-added posterior modes generated by our hybrid estimation strategy are positively correlated with conventional posterior means that ignore bias in OLS value-added estimates. This is evident in Figure 4, which plots hybrid modes against posterior means from conventional value-added models. Rank correlations in the lagged score and gains models are 0.79 and 0.74. Although hybrid and conventional posteriors are strongly correlated, hybrid estimation changes some schools' ranks, so accountability decisions may be improved using the hybrid estimates.

Hybrid estimation generates posterior modes for bias as well as value-added. The value-added and bias posteriors therefore permit an exploration of the association between school characteristics, causal value-added and bias. Table 7 reports coefficients from regressions of posterior modes for bias and value-added on school characteristics, with and without controls for sector. As can be seen in columns 1 and 3, students that appear more advantaged (based on baseline scores and special education status, for example) tend to enroll in schools with higher value-added, but this pattern is largely explained by the higher likelihood that these students enroll in charter schools. By contrast, column 4 shows that VAM bias is more positive for schools with more advantaged students, including those with higher average baseline test scores, fewer black students, fewer special education students, and fewer students eligible for subsidized lunches. The correlation of bias with baseline scores is especially noteworthy: although we see positive selection into the Boston charter sector, the popular impression that good schools have good peers is driven mostly by selection bias.

A key assumption underlying the hybrid approach is that the distribution of bias in conventional VAM estimates is unrelated to lottery over-subscription. This assumption restricts the relationship between student ability and school enrollment patterns. For example, it requires that students who enroll in more and less popular schools have similar ability conditional on demographic variables and lagged scores scores. Evidence in support of this assumption comes from the relationships between oversubscription rates, posterior bias estimates, and baseline scores.

[19] Chetty et al. (2014a) use this parameter to define "forecast bias," equal to $1 - r_\alpha$. We use "reliability" here to distinguish between r_α and the forecast coefficient $\hat{\varphi}$, which captures a different weighted average across schools. Appendix B.5 discusses the relationship between $\hat{\varphi}$ and r_α.

As can be seen in Panel A of Figure 5, posterior bias estimates are uncorrelated with the extent of oversubscription among lottery schools. Specifically, a regression of predicted bias from the lagged score model on the log of the oversubscription rate yields a slope coefficient of -0.02 with a standard error of 0.06.[20] The weak relationship between bias and the *degree* of oversubscription apparent in the figure is consistent with the hypothesis that bias distributions are similar for schools where lottery information is and is not available. Note also that this finding is not a mechanical consequence of assumptions imposed by the hybrid model, since the model ignores the degree of oversubscription within the lottery sample.

Recall that Table 2 shows that baseline scores and other observed characteristics are similar for students enrolled at schools with and without lotteries. Panel B of Figure 5 explores this pattern further by showing that oversubscription rates are uncorrelated with average baseline scores at oversubscribed schools. A regression of average baseline scores on log oversubscription produces a coefficient of -0.03 with a standard error of 0.10. This finding, which does not rely on estimates from the model, shows that observed ability of enrolled students is unrelated to lottery oversubscription within the lottery sample. We might therefore expect unobserved ability to be unrelated to oversubscription as well. Both panels of Figure 5 support the assumption postulating similar bias distributions for schools that are more and less heavily over-subscribed.[21]

7 Policy Simulations

We use a Monte Carlo simulation to gauge the accuracy and value of VAM estimates for decision-making. The simulation draws values of causal value-added, bias, and lottery first stage parameters from the estimated hyperparameter distributions underlying Table 6.[22] Estimation errors are also drawn from their joint asymptotic distribution and are combined with parameter draws to construct simulated OLS, reduced form and first stage estimates. These simulated estimates are then used to re-estimate the random coefficients model and construct conventional and hybrid EB posterior predictions. Each simulation therefore replicates the information available to a policymaker or parent, armed with both conventional and hybrid estimates, in a world calibrated to our model.

7.1 Mean Squared Error

Our first statistic for model assessment is root-mean-squared error (RMSE). Conventional VAMs generate value added estimates of school quality with an RMSE far below that of a naive uncontrolled benchmark. This can be seen in Figure 6, which compares RMSE across specifications and estimation procedures. RMSE in the uncontrolled model is about 0.5σ, falling to around 0.18σ and 0.17σ in the lagged score and gains

[20]The oversubscription rate is defined as the ratio of the annual average number of lottery applicants to the average number of seats for charter schools, and the ratio of the average number of first-choice applicants to the average number of seats for traditional and pilot schools.

[21]Appendix C investigates the sensitivity of policy simulation results to violations of this assumption. These results show that hybrid estimation generates substantial gains even when the difference in mean bias between lottery and non-lottery schools is on the order of 0.2σ.

[22]Simulation results for seventh and eighth grade, reported in Appendix Tables A4 and A5, yield conclusions similar to those for sixth grade. These and other supplementary simulation results are discussed in Appendix C.

VAMs. Adjustments for past scores and other student demographics eliminate a good portion of the bias in uncontrolled estimates.

The RMSE of hybrid estimates is impressively stable across specifications, starting at 0.17σ in an uncontrolled benchmark model and falling to 0.14σ in the lagged score and gains models. With sector effects included, hybrid estimation reduces RMSE from 0.15σ to about 0.12σ in the lagged score model and from 0.14σ to about 0.10σ in the gains model. The relatively stable hybrid RMSE shows how the hybrid estimator manages to reduce bias even when non-lottery estimates are badly biased. Although the largest bias mitigation seen in the figure comes from controlling for covariates, hybrid estimation reduces RMSE by a further 20-30 percent.

Not surprisingly, the RMSE reduction yielded by the hybrid estimator reflects reduced bias at the cost of increased sampling variance. This can be seen by writing the mean squared error of an estimator, β_j^*, as

$$E\left[\left(\beta_j^* - \beta_j\right)^2\right] = E\left[Var\left(\beta_j^* | \beta_j\right)\right] + \sigma_{b^*}^2,$$

where $\sigma_{b^*}^2 = E\left[\left(E\left[\beta_j^* | \beta_j\right] - \beta_j\right)^2\right]$ is average bias squared and the expectation treats the value-added parameters, β_j, as random. Blue and red shading in Figure 6 shows the proportions of MSE due to bias and variance. OLS VAMs are precisely estimated: sampling variance contributes only a small part of their overall MSE. Hybrid estimation reduces MSE, while also increasing the proportion of error due to sampling variance to around 30 percent. This reflects the core tradeoff motivating the hybrid approach: hybrid posteriors leverage lottery estimates to reduce bias in exchange for increased sampling variance relative to conventional VAMs.[23]

7.2 Consequences of School Closure

Massachusetts' school accountability framework uses value-added measures to guide decisions about school closures, school restructuring and turnarounds, and charter school expansion. A stylized description of these decisions is that they replace weak schools with those judged to be stronger on the basis of value-added estimates. We therefore simulate the achievement consequences of closing the lowest-ranked district school (traditional or pilot) and sending its students to schools with average or better estimated value-added.

This analysis ignores possible transition effects such as disruption due to school closure, peer effects from changes in school composition, and other factors that might inhibit replication of successful schools. The results should nevertheless provide a rough guide to the potential consequences of VAM-based policy decisions. Quasi-experimental analyses of charter takeovers and other school reconstitution efforts in Boston, New Orleans, and Houston have shown large gains when low-performing schools are replaced by schools operating according to pedagogical principles seen to be effective elsewhere (Fryer, 2014; Abdulkadiroğlu

[23] Appendix Table A6 shows hybrid estimates generate forecast coefficients close to one in both the lagged score and gains specifications, with or without charter lotteries. The hybrid estimates also pass the overidentification and omnibus specification tests.

et al., 2016). This suggests transitional consequences are dominated by longer-run determinants of school quality, at least for modest policy interventions of the sort considered here.

The potential for VAMs to guide decision-making is highlighted by the first row of Table 8, which shows the score gains produced by decisions based on true value added. Closing the worst school and replacing it with an average school boosts achievement by 0.37σ, while more targeted replacement policies generate even larger gains. Consistent with the high RMSE of uncontrolled estimates, however, Table 8 also shows that policies based on uncontrolled test score levels generate only small gains. For example, replacing the lowest-scoring district school with an average school is predicted to increase scores for affected students by 0.06σ on average. Likewise, a policy that replaces the lowest-ranked school with an average top quintile school generates a gain of 0.10σ. These small effects reflect the large variation in bias evident for the uncontrolled model in Table 6: closure decisions based on average test scores target schools with many low achievers rather than low value-added. The bias in uncontrolled VAM estimates also leads to a wide dispersion of simulated closure effects, with a cross-simulation standard deviation (reported in brackets) of around 0.2σ.

In contrast, closure and replacement decisions based on conventional lagged score and gains models yield substantial achievement gains. For instance, replacing the lowest-ranked school with an average school boosts scores by an average of 0.24σ when rankings are based on the gains specification. This is 65 percent of the corresponding benefit generated by a policy that ranks schools by true value-added. Hybrid estimation increases these gains to 0.32σ, an improvement of over 30 percent relative to the conventional model. This incremental effect closes roughly half the gap between conventional estimates and the maximum possible impact.

The effects of VAM-based policies and the incremental benefits of using lotteries grow when value-added predictions are used to choose expansion schools in addition to closures. In the gains specification, for example, replacing the lowest-ranked school with a typical top-quintile school generates an average improvement of 0.39σ when conventional posteriors are used to estimate VAM and an improvement of 0.53σ when rankings are based on hybrid predictions. The hybrid approach also modestly reduces the uncertainty associated with VAM-based policies by doing a better job of finding reliably good replacement schools.

The largest gains seen in Table 8 result from a policy that replaces the lowest-ranking traditional or pilot school with a charter school. This mirrors Boston's ongoing in-district charter conversion policy experiment (Abdulkadiroğlu et al., 2016). Reflecting the large difference in mean value-added between charter and district schools, charter conversion is predicted to generate significant gains regardless of how value-added is estimated. Accurate value-added estimation increases the efficacy of charter conversion, however: selecting schools for conversion based on the lagged score value-added model rather than the uncontrolled model boosts the effect of charter expansion from 0.28σ to 0.58σ, while use of the hyrid estimator pulls this up to 0.67σ, close to the maximum possible gain of 0.71σ.

The results in Table 8 show that, even when VAM estimates are imperfect, they predict causal value-added well enough to be useful for policy. For example, causal value-added is more than 0.2σ below-average for schools ranked at the bottom by the conventional lagged score and gains specifications. As can be seen in Table 6, this represents roughly a full standard deviation in the distribution of true school quality. Value-added for low-ranked schools is even more negative when rankings are based on hybrid estimates. Schools selected for replacement may not be the very worst schools in the district. At the same time, these schools are likely to be much worse than average, so policies that replace them with schools predicted to do better generate large gains.[24]

8 Conclusions and Next Steps

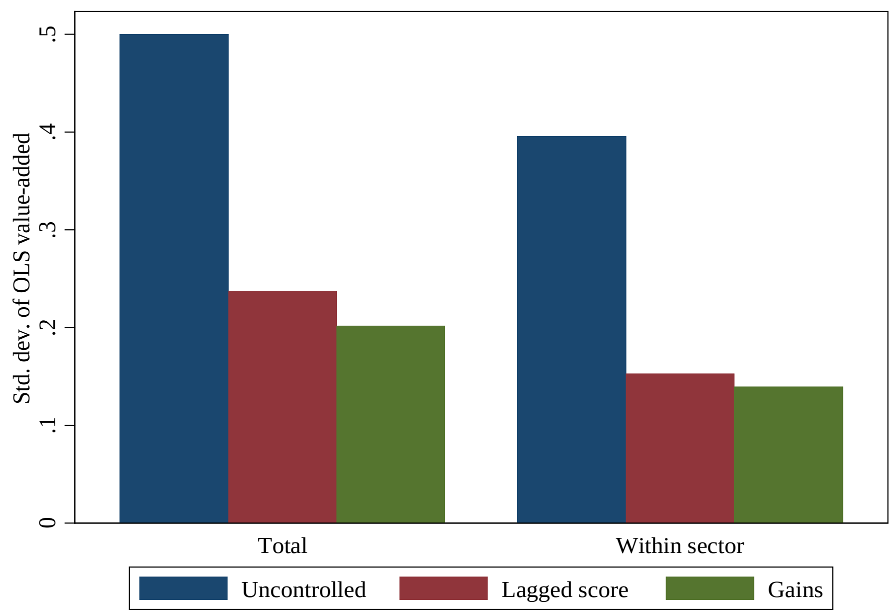
School districts increasingly rely on regression-based value-added models to gauge and report on school quality. This paper leverages admissions lotteries to test and improve conventional VAM estimates of school value-added. An application of our approach to data from Boston suggests that conventional value-added estimates for Boston's schools are biased. Nevertheless, policy simulations show that accountability decisions based on estimated VAM are likely to boost achievement. A hybrid estimation procedure that combines conventional and lottery-based estimates generates predictions that, while still biased, achieve lower mean-squared error and improved policy targeting relative to conventional VAMs.

Hybrid school value-added estimation requires some kind of lottery-based admissions scheme, such as those increasingly used for student assignment in many of America's large urban districts. As our analysis of charter schools shows, however, admissions need not be centralized for lotteries to be of value. The utility of hybrid estimation in other cities will vary with the extent of lottery coverage, but results for Boston show hybrid estimation remains useful even when lottery data are missing for many schools. Our approach also rules out effect heterogeneity linked to school choices, which may be less appropriate in settings with more specialized schools and very heterogeneous student populations.

The methods developed here may be useful for combining quasi-experimental and non-experimental estimators in other contexts. Candidates for this extension include the quantification of teacher, doctor, hospital, firm, or neighborhood effects. Assignment lotteries in these settings are rare, but our hybrid estimation strategy may be extended to exploit other sources of quasi-experimental variation. A hybrid approach to testing and estimation is likely to be fruitful in any context where a set of credible quasi-experiments is available to discipline a larger set of non-experimental comparisons.

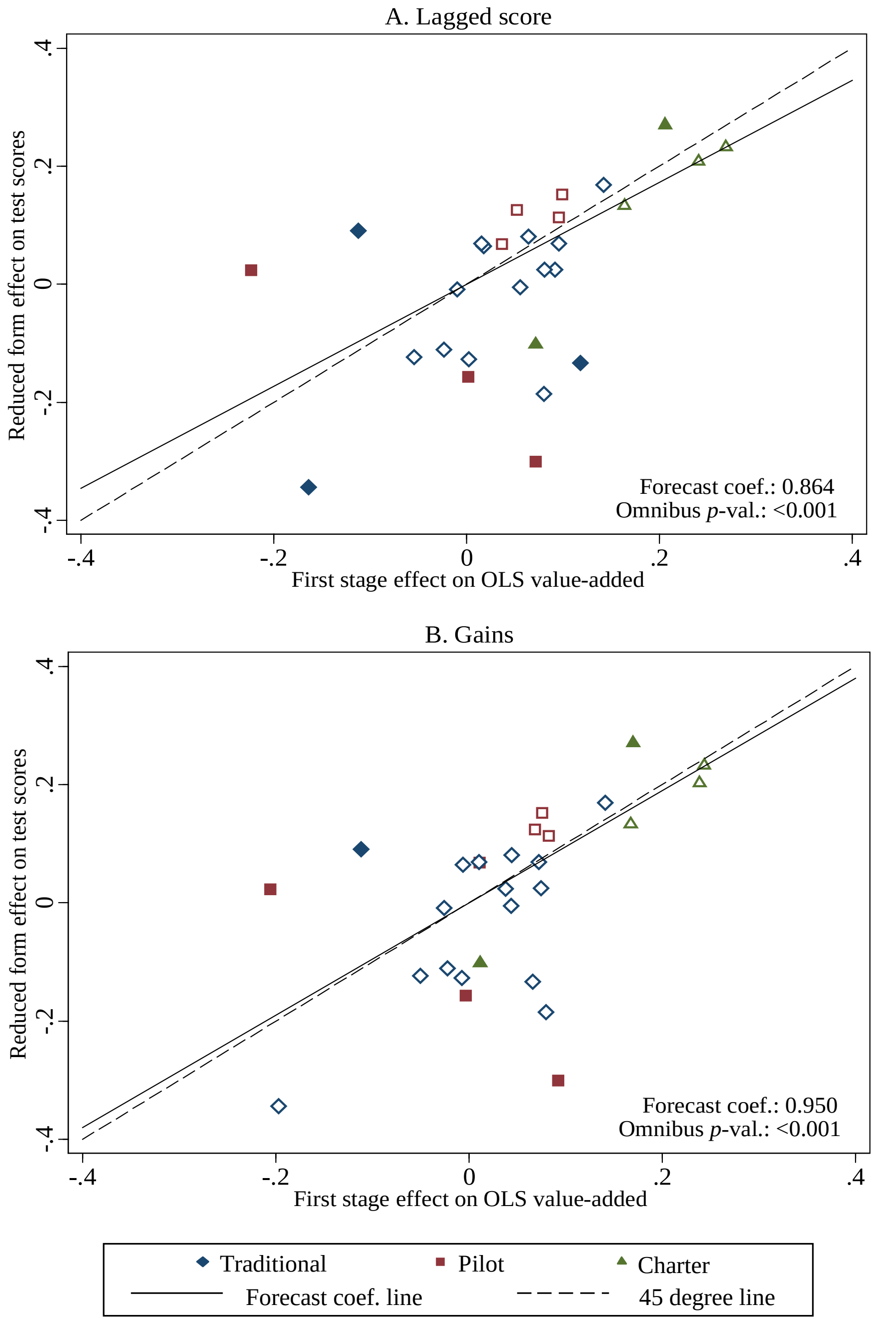
[24] The simulations in Table 8 predict the consequences of decisions based on the eight years of data in our sample. Districts often estimate value-added over shorter time periods. To gauge the effects of using four years of data, Appendix Table A7 reports simulation results that double sampling variance. This produces results which are qualitatively similar to those from the full sample, with slightly smaller closure effects. Appendix Table A8 reports estimates from a model (described in Appendix C) that allows value-added and bias to vary by year. These estimates suggest a limited role for idiosyncratic temporal variation in VAM hyperparameters.

Figure 1: Standard deviations of school effects from OLS value-added models



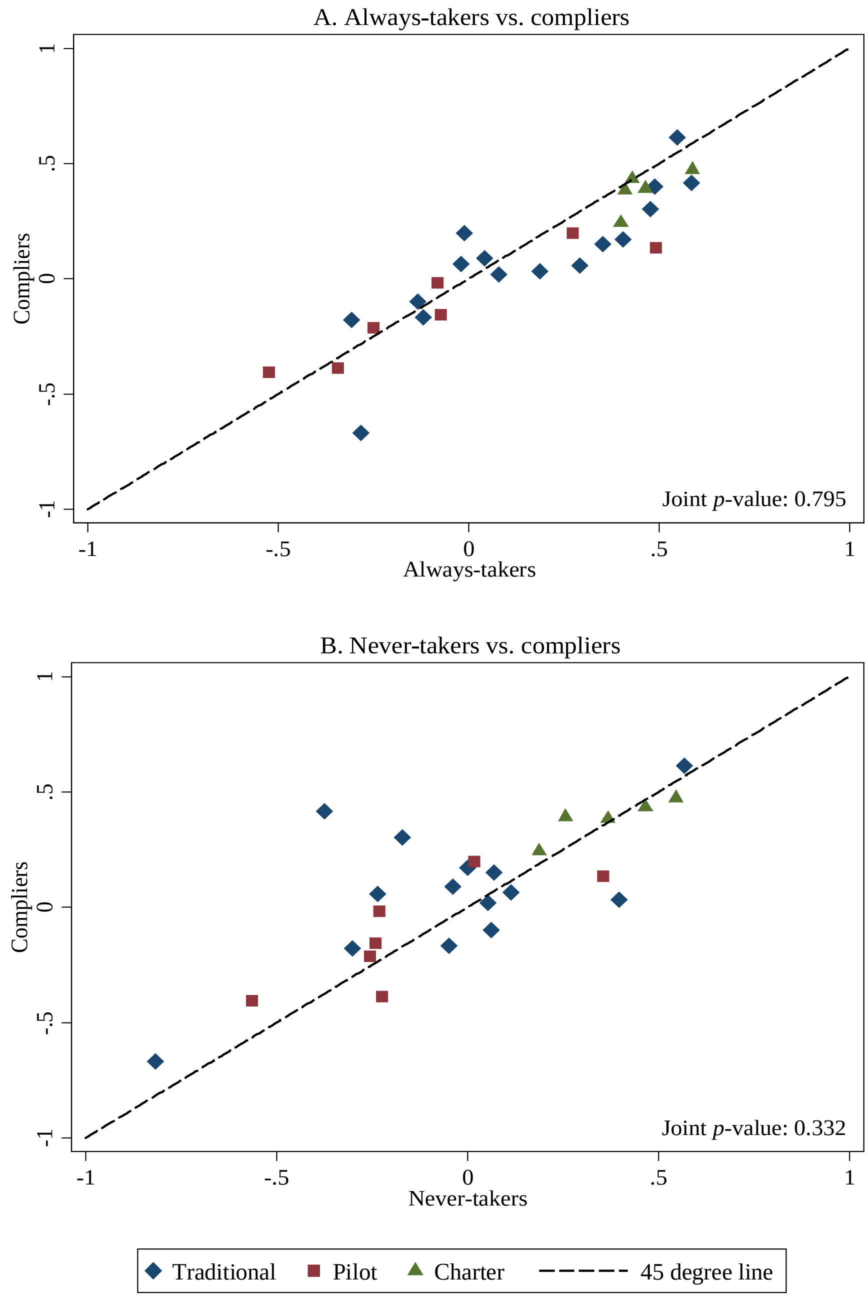
Notes: This figure displays standard deviations of school effects from OLS value-added models. See notes to Table 3 for a description of the controls included in the lagged score and gains models; the uncontrolled model includes only year effects. The variance of OLS value-added is obtained by subtracting the average squared standard error from the sample variance of value-added estimates, then taking the square root. Within-sector variances are obtained by first regressing value-added estimates on charter and pilot dummies, then subtracting the average squared standard error from the sample variance of residuals and taking the square root.

Figure 2: Visual instrumental variables tests for bias



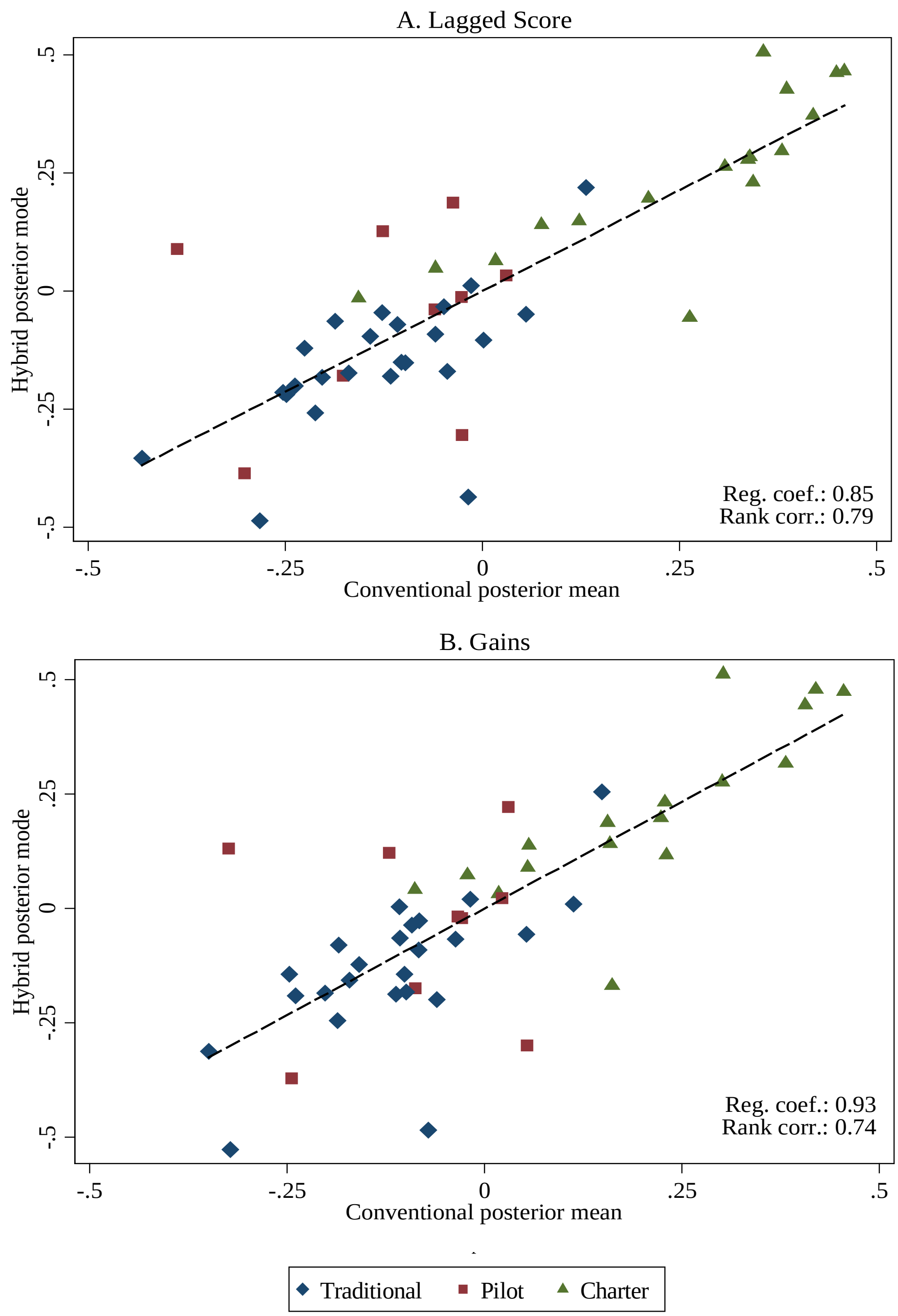
Notes: This figure plots lottery reduced form effects against value-added first stages from each of 28 school admission lotteries. See the notes for Table 3 for a description of the value-added models and lottery specification. Filled markers indicate estimates that are significant at the 10% level. Slopes of solid lines correspond to the forecast coefficients from Table 3, while dashed lines indicate the 45-degree line.

Figure 3: Comparisons of conventional value-added by lottery compliance



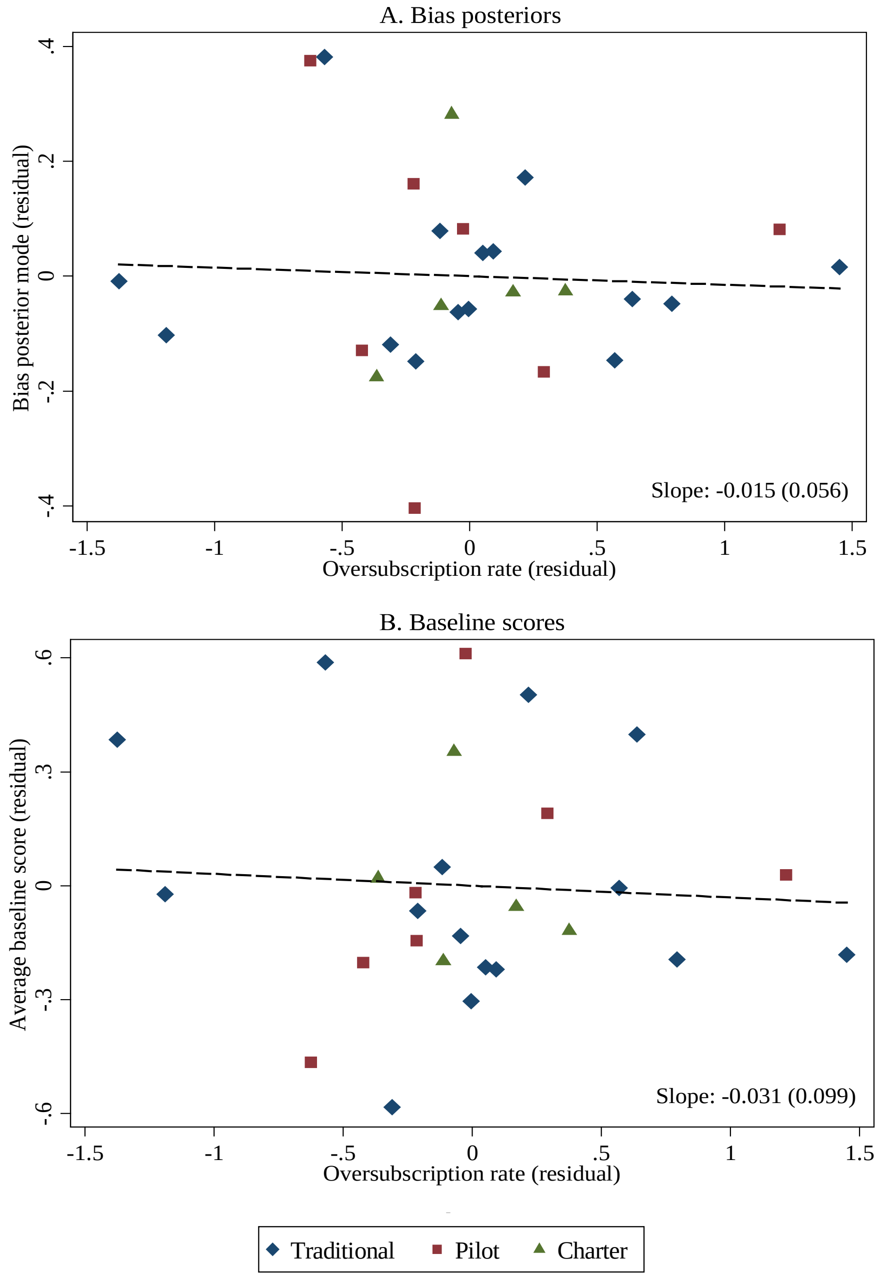
Notes: This figure compares OLS estimates of average value-added for admission lottery compliers to estimates for always- and never-takers in each of 28 school lotteries. OLS estimates comes from a lagged-score VAM that allows school effects to differ across the subgroups used in column 6 of Table 4, estimated in the lottery sample. Complier, always-taker, and never-taker means are estimated by the method proposed by Abadie (2003) using a saturated model of lottery strata indicators for E[Z|C]. *P*-values are for the joint tests that complier and always-taker or never-taker means are equal in every lottery and are based on 500 Bayesian bootstrap replications. The joint p-value for both panels is 0.289.

Figure 4: Empirical Bayes posterior predictions of school value-added



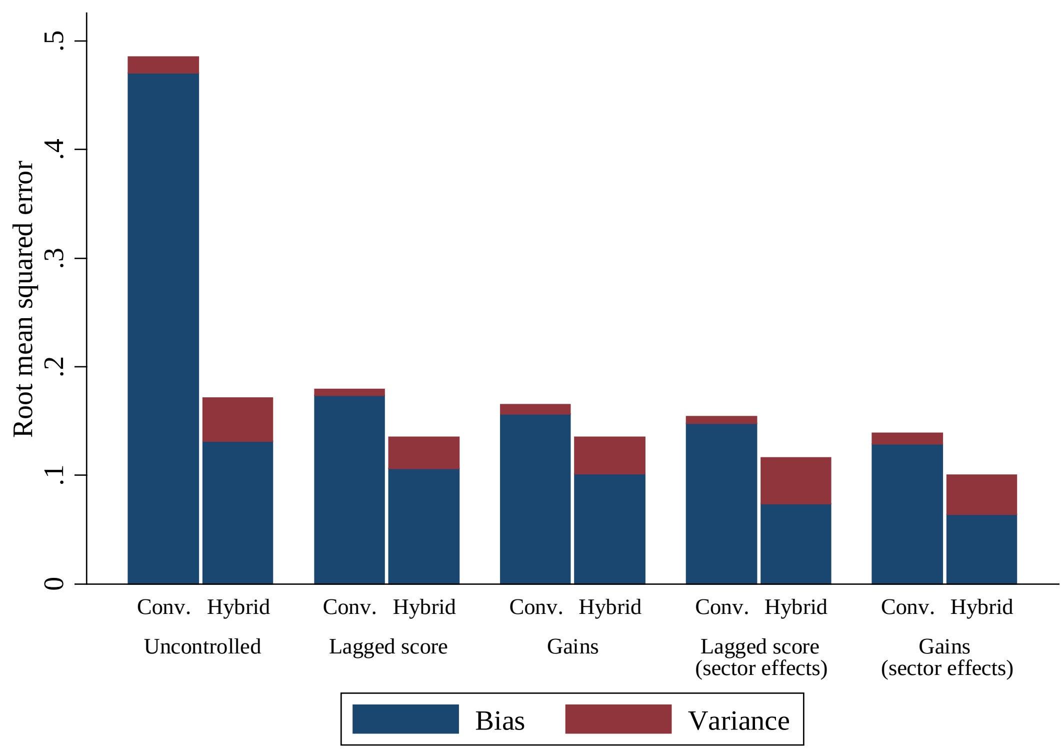
Notes: This figure plots empirical Bayes posterior mode predictions of value-added from the random coefficients model against posterior means based on OLS value-added. Posterior modes are computed by maximizing the sum of the log-likelihood of the OLS, reduced form, and first stage estimates conditional on all school-specific parameters plus the log-likelihood of these parameters given the estimated random coefficient distribution. Conventional posteriors shrink OLS estimates towards the mean in proportion to one minus the signal-to-noise ratio. Dashes indicate OLS regression lines.

Figure 5: Relationship between oversubscription and bias measures for lottery schoools



Notes: Panel A of this figure plots posterior mode predictions of bias in sixth grade math VAMs against oversubscription rates for schools with admission lotteries. The oversubscription rate is defined as the log of the ratio of the average number of first-choice applicants (for traditional and pilot schools) or the average number of total applicants (for charters) to the average number of available seats for each admission grade. Bias modes come from the lagged score model with sector effects. Panel B plots school average baseline math and ELA scores against oversubscription rates. Points in the figure are constructed by first regressing bias modes, mean baseline scores and oversubscription rates on pilot and charter indicators, then computing residuals from these regressions. Dashes indicate OLS regression lines.

Figure 6: Root-mean-squared error for value-added posterior predictions



Notes: This figure plots root-mean-squared error for posterior predictions of sixth grade math value-added. Conventional predictions are posterior means constructed from OLS value-added estimates. Hybrid predictions are posterior modes constructed from OLS and lottery estimates. The total height of each bar indicates root mean squared error (RMSE). Blue bars display shares of mean squared error due to bias, and red bars display shares due to variance. RMSE is calculated from 500 simulated samples drawn from the data generating processes implied by the estimates in Table 6. The random coefficients model is re-estimated in each simulated sample.

Table 1: Boston students and schools

Total enrollment				Total enrollment			
OLS sample	Lottery sample	6th grade entry?	Lottery school?	OLS sample	Lottery sample	6th grade entry?	Lottery school?
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Traditional publics (25)				B. Pilots (9)			
1,095	79	Y	Y	538	310	Y	Y
1,025	445	Y	Y	1,260	433	Y	Y
1,713	1,084	Y	Y	585	296	Y	Y
547	218	Y	Y	78	5	Y	
217	46	Y		453	46	Y	Y
1,354	581	Y	Y	380	67	Y	Y
263	44	Y		242	179	Y	Y
1,637	492	Y	Y	558	73	Y	Y
472	104	Y		18	12		
1,238	591	Y	Y	C. Charters (17)			
537	11			738	406	Y	Y
331	35	Y	Y	361	23		
335	82	Y		357	215	Y	
952	232	Y	Y	393	332	Y	Y
294	71	Y	Y	338	16		
333	90	Y		511	115	Y	Y
766	243	Y	Y	71	8		
372	47	Y	Y	300	23		
137	14	Y		389	342	Y	Y
1,091	225	Y	Y	654	34		
1,086	127	Y	Y	45	3		
577	104	Y	Y	53	2		
622	61	Y		415	305	Y	Y
906	270	Y	Y	70	6		
267	19			104	23		
				701	92		
				85	37		

Notes: This table counts the students included in each school in the OLS value-added and lottery samples. The sample covers cohorts attending sixth grade in Boston between the 2006-2007 and 2013-2014 school years. Columns 3 and 7 indicate schools for which sixth grade is the primary entry grade; columns 4 and 8 indicate whether the school has enough students subject to conditionally-random admission variation to be included in the lottery sample. Total numbers of schools in each sector are included in parentheses in the column headings.

Table 2: Descriptive statistics

	Means				Lottery offer balance			
	OLS sample		Lottery sample					
Baseline covariate	All students (1)	Lottery school students (2)	All students (3)	Lottery school students (4)	All lotteries (5)	Traditional (6)	Pilot (7)	Charter (8)
Hispanic	0.345	0.342	0.354	0.361	-0.017	-0.007	0.003	-0.006
					(0.013)	(0.017)	(0.033)	(0.018)
Black	0.410	0.394	0.485	0.468	-0.011	-0.005	-0.052	-0.009
					(0.014)	(0.018)	(0.034)	(0.020)
White	0.122	0.125	0.072	0.078	0.010	0.006	0.005	0.009
					(0.007)	(0.008)	(0.015)	(0.010)
Female	0.490	0.487	0.504	0.502	0.017	0.034*	-0.013	-0.025
					(0.014)	(0.019)	(0.037)	(0.020)
Subsidized lunch	0.806	0.811	0.830	0.831	0.020*	0.020	0.006	-0.005
					(0.010)	(0.013)	(0.026)	(0.016)
Special education	0.208	0.214	0.195	0.196	0.006	-0.003	-0.022	0.015
					(0.011)	(0.013)	(0.030)	(0.016)
English-language learner	0.205	0.224	0.206	0.214	0.006	-0.001	0.018	0.004
					(0.011)	(0.014)	(0.027)	(0.016)
Suspensions	0.093	0.073	0.076	0.070	-0.025	-0.025	0.009	-0.016
					(0.016)	(0.023)	(0.025)	(0.017)
Absences	1.710	1.567	1.534	1.466	-0.087	-0.138*	-0.092	0.110
					(0.095)	(0.080)	(0.260)	(0.167)
Math score	0.058	0.053	0.004	0.016	0.022	-0.026	0.080	0.036
					(0.024)	(0.030)	(0.061)	(0.035)
ELA score	0.030	0.006	0.013	0.016	0.035	0.045	0.060	0.013
					(0.025)	(0.030)	(0.061)	(0.036)
N	27,864	21,446	8,718	7,748	8,718	4,849	1,303	3,655

Notes: This table reports sample mean characteristics and investigates balance of random lottery offers. Column 1 shows mean characteristics for all Boston sixth graders enrolled between the 2006-2007 and 2013-2014 school years, and column 2 shows means for students enrolled at schools that have randomized entrance lotteries in at least one year. Columns 3 and 4 report mean characteristics for students subject to random lottery assignment. Columns 5-8 report coefficients from regressions of baseline characteristics on lottery offers, controlling for lottery strata. Robust standard errors are reported in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 3: Tests for bias in conventional 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	0.950	0.549	0.677
	(0.075)	(0.084)	(0.164)	(0.193)
First stage *F* -statistic	29.6	26.6	11.2	9.3
p -values:				
Forecast bias	0.071	0.554	0.006	0.095
Overidentification	0.003	0.006	0.043	0.052
Omnibus test χ^2 statistic (d.f.)	77.7 (28)	72.1 (28)	48.0 (23)	41.7 (23)
p -value	<0.001	<0.001	<0.001	0.010
N	8,718		6,162	
		B. All middle school grades		
Forecast coefficient (φ)	0.880	0.924	0.683	0.726
	(0.055)	(0.060)	(0.124)	(0.133)
First stage *F* -statistic	14.7	15.0	7.6	7.8
p -values:				
Forecast bias	0.028	0.204	0.011	0.039
Overidentification	0.011	0.011	0.062	0.065
Omnibus test χ^2 statistic (d.f.)	172.8 (75)	167.0 (75)	111.6 (60)	107.9 (60)
p -value	<0.001	<0.001	<0.001	<0.001
N	20,935		15,027	

Notes: This table reports the results of tests for bias in conventional value-added models (VAMs) for sixth through eighth grade math scores. The lagged score VAM includes cubic polynomials in baseline math and ELA scores, along with indicators for application year, sex, race, subsidized lunch, special education, limited-English proficiency, and counts of baseline absences and suspensions. The gains VAM drops the lagged score controls and uses score growth from baseline as the outcome. Seventh and eighth grade VAMs measure exposure to each school using total years of enrollment since the lottery. Forecast coefficients are from instrumental variables regressions of test scores on fitted values from conventional VAMs, instrumenting fitted values with lottery offer indicators. IV models are estimated via an asymptotically efficient GMM procedure and control for lottery strata fixed effects, demographic variables, and lagged scores. The forecast bias test checks whether the coefficient from this model equals one, and the overidentificiation test checks the model's overidentifying restrictions. The omnibus test combines forecast bias and overidentifying restrictions. Panel A uses sixth grade math scores, while Panel B stacks math score outcomes and VAM fitted values from sixth through eighth grade. Standard errors and test statistics in Panel B cluster on student. Columns 3 and 4 exclude charter school lotteries.

Table 4: Robustness of sixth grade bias tests to effect heterogeneity

Value-added model		Baseline VAM specification (1)	VAM estimated by subgroup: Baseline year (2)	Subsidized lunch (3)	Special education (4)	Baseline score tercile (5)	Interacted groups (6)
			A. VAM estimated on the OLS sample				
Lagged score	Forecast coefficient (φ)	0.864	0.916	0.849	0.863	0.866	0.930
		(0.075)	(0.072)	(0.075)	(0.074)	(0.075)	(0.061)
	Omnibus test $\chi^2(28)$ statistic	77.7	68.2	82.8	79.0	83.3	73.4
	p -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Gains	Forecast coefficient (φ)	0.950	1.016	0.944	0.955	0.891	0.953
		(0.084)	(0.082)	(0.083)	(0.083)	(0.079)	(0.065)
	Omnibus test $\chi^2(28)$ statistic	72.1	65.7	74.4	72.4	80.9	66.6
	p -value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
			B. VAM estimated on the lottery sample				
Lagged score	Forecast coefficient (φ)	0.868	0.962	0.851	0.872	0.873	0.934
		(0.070)	(0.068)	(0.069)	(0.070)	(0.070)	(0.052)
	Omnibus test $\chi^2(28)$ statistic	62.3	51.8	67.9	63.5	65.9	56.3
	p -value	<0.001	0.004	<0.001	<0.001	<0.001	0.001
Gains	Forecast coefficient (φ)	0.926	1.035	0.912	0.937	0.890	0.941
		(0.077)	(0.077)	(0.076)	(0.077)	(0.073)	(0.055)
	Omnibus test $\chi^2(28)$ statistic	57.8	50.2	60.1	58.4	61.1	42.2
	p -value	<0.001	0.006	<0.001	<0.001	<0.001	0.041

Notes: This table reports lottery-based tests for bias in school value-added models that allow for effect heterogeneity by baseline characteristics. See the notes to Table 3 for a description of the lagged score and gains models and test procedure. Panel A estimates value-added in the full OLS sample, while Panel B restricts estimation to the lottery subsample. Column 1 repeats estimates that do not allow effect heterogeneity, while columns 2-6 allow value-added to differ across groups defined by the covariates in the column headings. The covariates used to define groups in column 6 are race, gender, subsidized lunch, special education, English language learner status, and baseline score terciles based on average fifth grade math and ELA test scores in the OLS sample.

Table 5: Fallback status of schools without sixth grade lotteries

Lottery students with fallback enrollment	*p*-value: not a lottery fallback	Sixth grade entry?	Lottery students with fallback enrollment	*p*-value: not a lottery fallback	Sixth grade entry?
(1)	(2)	(3)	(4)	(5)	(6)
A. Traditional publics			C. Charters		
39	0.013	Y	320	<0.001	Y
36	0.018	Y	11	0.080	
113	<0.001	Y	16	0.427	
79	<0.001	Y	16	0.204	
94	<0.001	Y	24	0.724	
21	0.045	Y	42	0.145	
60	0.006	Y	3	0.111	
12	0.016		2	0.390	
B. Pilots			10	0.257	
5	0.033	Y	33	0.010	
15	0.169		112	<0.001	
			34	0.378	

Notes: This table reports *p*-values for tests of whether each non-lottery school in the OLS sample serves as a fallback for one of the 28 lottery schools. Columns 1 and 4 count the number of students in the lottery sample who are observed enrolling in the undersubscribed school when not given a offer. Columns 2 and 5 test jointly whether the undersubscribed school's first stage coefficients are zero in all lotteries with such students. Columns 3 and 6 indicate whether sixth grade is a school's primary entry point. First stage regressions control for lottery strata indicators, demographic variables, and lagged test scores.

Table 6: Joint distribution of causal value-added and VAM bias for sixth grade math scores

		Models without sector effects			Models with sector effects	
		Uncontrolled	Lagged score	Gains	Lagged score	Gains
		(1)	(2)	(3)	(4)	(5)
σ_β	Std. dev. of causal VA	0.195	0.220	0.222	0.171	0.170
		(0.024)	(0.021)	(0.023)	(0.028)	(0.023)
σ_b	Std. dev. of VAM bias	0.501	0.182	0.166	0.148	0.133
		(0.061)	(0.048)	(0.048)	(0.029)	(0.030)
$\sigma_{\beta b}$	Covariance of	-0.018	-0.014	-0.017	-0.016	-0.013
	VA and bias	(0.010)	(0.003)	(0.004)	(0.006)	(0.003)
r_α	Regression of VA	0.078	0.644	0.753	0.694	0.783
	on OLS (reliability ratio)	(0.204)	(0.066)	(0.072)	(0.152)	(0.122)
VA shifters	Charter				0.426	0.396
					(0.104)	(0.106)
	Pilot				0.130	0.111
					(0.129)	(0.129)
	Lottery school (β_Q)	0.040	-0.024	-0.033	0.104	0.066
		(0.127)	(0.061)	(0.054)	(0.042)	(0.041)
Bias shifters	Charter				-0.005	-0.063
					(0.103)	(0.099)
	Pilot				-0.121	-0.089
					(0.124)	(0.121)
	χ^2 statistic (d.f.):	10.9 (7)	10.8 (7)	9.1 (7)	9.0 (13)	6.0 (13)
	Overid. *p*- value:	0.145	0.147	0.247	0.773	0.946

Notes: This table reports simulated minimum distance estimates of parameters of the joint distribution of causal school value-added and OLS bias. The moments used in estimation are functions of OLS value-added, lottery reduced form, and first stage estimates, as described in Appendix B. Uncontrolled estimates come from an OLS regression that controls only for year effects. See notes to Table 3 for a description of the control variables included in the lagged score and gains value-added models. Simulated moments are computed from 500 samples constructed by drawing school-specific parameters from the random coefficient distribution along with estimation errors based on the asymptotic covariance matrix of the estimates. Columns 4 and 5 allow the means of the random coefficients distribution to depend on school sector. Moments are weighted by an estimate of the inverse covariance matrix of the moment conditions, calculated from a first-step estimate using an identity weighting matrix. The weighting matrix is produced using 1,000 simulations, drawn independently from the samples used to simulate the moments.

Table 7: Correlates of posterior value-added and VAM bias

	Overall		Within-sector	
	Value-added	Bias	Value-added	Bias
School characteristic	(1)	(2)	(3)	(4)
Fraction black	0.158	-0.208***	-0.050	-0.217***
	(0.143)	(0.075)	(0.083)	(0.073)
Fraction hispanic	0.065	0.031	0.268**	0.048
	(0.201)	(0.105)	(0.127)	(0.112)
Fraction subsidized lunch	-0.132	-0.452**	0.085	-0.474***
	(0.306)	(0.181)	(0.203)	(0.164)
Fraction special education	-0.977***	-0.501***	0.009	-0.508**
	(0.330)	(0.157)	(0.316)	(0.217)
Fraction English-language learne	-0.542**	-0.135	0.297	-0.092
	(0.247)	(0.221)	(0.243)	(0.254)
Mean baseline math score	0.157*	0.143***	0.012	0.145***
	(0.088)	(0.051)	(0.070)	(0.047)
Mean baseline ELA score	0.201**	0.135**	0.039	0.138**
	(0.085)	(0.060)	(0.074)	(0.060)
Charter and pilot controls?			Y	Y

Notes: This table reports coefficients from regressions of empirical Bayes posterior modes for causal value-added and VAM bias on school characteristics. Columns 1 and 2 show coefficients from bivariate regressions, while columns 3 and 4 show coefficients from regressions controlling for charter and pilot indicators. Posterior modes come from the lagged score model with sector effects for sixth grade math scores. Robust standard errors are reported in parentheses.
*significant at 10%; **significant at 5%; ***significant at 1%

Table 8: Consequences of closing the lowest-ranked district school for affected children

Model	Posterior method	Replacement school: Average district school (1)	Average above-median school (2)	Average top-quintile school (3)	Average charter school (4)
-	True value-added	0.370	0.507	0.610	0.711
		[0.080]	[0.089]	[0.094]	[0.094]
Uncontrolled	Conventional	0.056	0.078	0.095	0.280
		[0.191]	[0.197]	[0.204]	[0.198]
	Hybrid	0.153	0.223	0.259	0.377
		[0.143]	[0.156]	[0.169]	[0.151]
Lagged score	Conventional	0.226	0.307	0.367	0.577
		[0.159]	[0.168]	[0.176]	[0.165]
	Hybrid	0.315	0.437	0.529	0.665
		[0.131]	[0.141]	[0.147]	[0.145]
Gains	Conventional	0.240	0.327	0.391	0.580
		[0.148]	[0.156]	[0.163]	[0.153]
	Hybrid	0.316	0.434	0.525	0.657
		[0.115]	[0.126]	[0.136]	[0.128]

Notes: This table reports simulated test score impacts of closing the lowest-ranked BPS district school based on value-added predictions. The reported impacts are average effects on test scores for students at the closed school. Standard deviations of these effects across simulations appear in brackets. Column 1 replaces the lowest-ranked district school with an average district school. Column 2 replaces the lowest-ranked school with an average above-median district school, and column 3 uses an average top-quintile district school. Column 4 replaces the lowest-ranked district school with an average charter school. See notes to Table 3 for a description of the controls included in each value-added model. Conventional empirical Bayes posteriors are means conditional on OLS estimates only, while hybrid posteriors are modes conditional on OLS and lottery estimates. All models include sector effects. Statistics are based on 500 simulated samples, and the random coefficients model is re-estimated in each sample.

References

Abadie, A. (2003): "Semiparametric instrumental variable estimation of treatment response models," *Journal of Econometrics*, 113, 231–263.

Abdulkadiroğlu, A., J. D. Angrist, S. Dynarski, T. J. Kane, and P. A. Pathak (2011): "Accountability and flexibility in public schools: Evidence from Boston's charters and pilots," *Quarterly Journal of Economics*, 126(2), 699–748.

Abdulkadiroğlu, A., J. D. Angrist, P. D. Hull, and P. A. Pathak (2016): "Charters without lotteries: Testing takeovers in New Orleans and Boston," *American Economic Review*, 106(7), 1878–1920.

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2015): "Market design meets research design," NBER working paper no. 21705.

Abdulkadiroğlu, A., P. A. Pathak, A. E. Roth, and T. Sönmez (2006): "Changing the Boston school choice mechanism," NBER working paper no. 11965.

Altonji, J. G. and L. M. Segal (1996): "Small-sample bias in GMM estimation of covariance structures." *Journal of Business and Economic Statistics*, 14, 353–366.

Angrist, J. D. (1998): "Estimating the labor market impact of voluntary military service using Social Security data on military applicants," *Econometrica*, 66, 249–288.

Angrist, J. D., S. R. Cohodes, S. M. Dynarski, P. A. Pathak, and C. R. Walters (2016a): "Stand and deliver: Effects of Boston's charter high schools on college preparation, entry and choice," *Journal of Labor Economics*, 34, 275–318.

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters (2016b): "Interpreting tests of school VAM validity," *American Economic Review: Papers & Proceedings*, 106, 388–392.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of causal effects using instrumental variables," *Journal of the American Statistical Association*, 91, 444–455.

Angrist, J. D., P. A. Pathak., and C. R. Walters (2013): "Explaining charter school effectiveness," *American Economic Journal: Applied Economics*, 5, 1–27.

Angrist, J. D. and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

Bacher-Hicks, A., T. J. Kane, and D. O. Staiger (2014): "Validating teacher effect estimates using changes in teacher assignments in Los Angeles," NBER working paper no. 20657.

Bloom, H. S. and R. Unterman (2014): "Can small high schools of choice improve educational prospects for disadvantaged students?" *Journal of Policy Analysis and Management*, 33, 290–319.

Card, D., J. Heining, and P. Kline (2013): "Workplace heterogeneity and the rise of West German wage inequality," *Quarterly Journal of Economics*, 128, 967–1015.

Chamberlain, G. and G. W. Imbens (2004): "Random effects estimators with many instrumental variables," *Econometrica*, 72, 295–306.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011): "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *Quarterly Journal of Economics*, 126, 1593–1660.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a): "Measuring the impact of teachers I: Evaluating bias in teacher value-added estimates," *American Economic Review*, 104, 2593–2563.

——— (2014b): "Measuring the impact of teachers II: Teacher value-added and student outcomes in adulthood," *American Economic Review*, 104, 2633–2679.

——— (2016): "Using lagged outcomes to evaluate bias in value-added models," *American Economic Review: Papers & Proceedings*, 106, 393–399.

Chetty, R. and N. Hendren (2015): "The impacts of neighborhoods on intergenerational mobility: childhood exposure effects and county-level estimates," Mimeo, Harvard University.

Cullen, J. B., B. A. Jacob, and S. D. Levitt (2006): "The effect of school choice on participants: evidence from randomized lotteries," *Econometrica*, 74, 1191–1230.

Deming, D. (2014): "Using school choice lotteries to test measures of school effectiveness," *American Economic Review: Papers & Proceedings*, 104, 406–411.

Deming, D. J., J. S. Hastings, T. J. Kane, and D. O. Staiger (2014): "School choice, school quality, and postsecondary attainment," *American Economic Review*, 104, 991–1013.

Deutsch, J. (2012): "Using school lotteries to evaluate the value-added model," Mimeo, University of Chicago.

Dobbie, W. and R. G. Fryer (2013): "Getting beneath the veil of effective schools: evidence from New York City," *American Economic Journal: Applied Economics*, 5, 28–60.

——— (2015): "The medium-term impacts of high-achieving charter schools," *Journal of Political Economy*, 123, 985–1037.

Finkelstein, A., A. Sacarny, and C. Syverson (2013): "Healthcare exceptionalism? Productivity and allocation in the US healthcare sector," NBER working paper no. 19200.

Fletcher, J. M., L. I. Horwitz, and E. Bradley (2014): "Estimating the value added of attending physicians on patient outcomes," NBER working paper no. 20534.

Fryer, R. G. (2014): "Injecting charter school best practices into traditional public schools: Evidence from field experiments," *Quarterly Journal of Economics*, 129, 1355–1407.

Hansen, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica*, 50, 1029–1054.

Hastings, J. S. and J. M. Weinstein (2008): "Information, school choice, and academic achievement: Evidence from two experiments," *Quarterly Journal of Economics*, 123, 1373–1414.

Imbens, G. W. and J. D. Angrist (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

Jacob, B. A. and L. Lefgren (2008): "Principals as agents: subjective performance assessment in education," *Journal of Labor Economics*, 26, 101–136.

Judge, G. G. and R. C. Mittlehammer (2004): "A semiparametric basis for combining estimation problems under quadratic loss," *Journal of the American Statistical Association*, 99, 479–487.

——— (2005): "Combing estimators to improve structural model estimation and inference under quadratic loss," *Journal of Econometrics*, 128, 1–29.

——— (2007): "Estimation and inference in the case of competing sets of estimating equations," *Journal of Econometrics*, 138, 513–531.

Kane, T. J., D. F. McCaffrey, and D. O. Staiger (2013): "Have we identified effective teachers? Validating measures of effective teaching using random assignment," *Gates Foundation Report*.

Kane, T. J., J. E. Rockoff, and D. O. Staiger (2008): "What does certification tell us about teacher effectiveness? Evidence from New York City," *Economics of Education Review*, 27, 615–631.

Kane, T. J. and D. O. Staiger (2002): "The Promise and PItfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives*.

——— (2008): "Estimating teacher impacts on student achievement: An experimental evaluation," NBER working paper no. 14607.

Kinsler, J. (2012): "Assessing Rothstein's critique of teacher value-added models," *Quantitative Economics*, 3, 333–362.

Koedel, C. and J. R. Betts (2011): "Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique," *Education Finance and Policy*, 6, 18–42.

McFadden, D. (1989): "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 57, 995–1026.

Morris, C. N. (1983): "Parametric empirical Bayes inference: Theory and applications," *Journal of the American Statistical Association*, 78, 47–55.

Newey, W. K. (1985): "Generalized method of moments specification testing," *Journal of Econometrics*, 29, 229–256.

Newey, W. K. and K. D. West (1987): "Hypothesis testing with efficient method of moments estimation," *International Economic Review*, 28, 777–787.

Rothstein, J. (2009): "Student sorting and bias in value-added estimation: Selection on observables and unobservables," *Education Finance and Policy*, 4, 537–571.

——— (2010): "Teacher quality in educational production: Tracking, decay, and student achievement," *Quarterly Journal of Economics*, 125, 175–214.

——— (2015): "Revisiting the impacts of teachers," Mimeo, University of California, Berkeley.

Rubin, D. B. (1981): "The Bayesian bootstrap," *Annals of Statistics*, 9, 130–134.

Sargan, J. (1958): "The estimation of economic relationships using instrumental variables," *Econometrica*, 26, 393–415.

Walters, C. R. (2014): "The demand for effective charter schools," NBER Working Paper no. 20640.

White, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 48, 817–838.

——— (1982): "Instrumental variables regression with independent observations," *Econometrica*, 50, 483–499.

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 outcome for measuring school value-added are sixth grade standardized test scores from the Massachusetts Comprehensive Assessment System (MCAS) database. We normalize MCAS math and ELA scores by grade and year to be mean-zero and have standard deviation one within a combined BPS and Boston charter school reference population. MCAS scores are merged to SIMS data via a state-assigned unique student identifier. We also merge baseline (fifth grade) math and ELA test 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

We use annual lottery 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 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 obtain a complete record of student-submitted preferences, school priorities, random tie-breaking sequence numbers, and assignments from the BPS deferred-acceptance mechanism, 2006-2013. For each year, we identify groups of students subject to the same priorities (given by whether a student has an enrolled sibling and whether she resides in a school's walk-zone, a 1.5 mile radius) at schools that they rank first. In forming these groups we exclude students that are guaranteed admission by virtue of current enrollment, as well as certain other students with guaranteed or nonstandard priorities (see Abdulkadiroğlu et al. (2006) for a complete description of priorities in BPS). We construct indicators for whether an applying student was offered a seat, with such offers are randomly assigned within strata defined by school, application year, and priority group. 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 1). 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 2. 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 3 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 $E[\alpha_j(X_i)\kappa^n_{ij}]/E[\kappa^a_{ij}]$ 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 3 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 $\alpha_j(X_i)$, inference in Figure 3 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, alway-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

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

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:

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

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

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

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

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

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

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

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 ν_j^ℓ, 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

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

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 6 reports this J-statistic.

B.3 Empirical Bayes Posteriors with a Known First Stage

We next derive expressions for hybrid empirical Bayes posterior predictions of school value-added that condition on lottery and OLS estimates. Begin by assuming that the first stage matrix, Π, is known. In this case 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 $\beta_Q = 0$ for simplicity. The posterior density for the random coefficients $\Theta = (\beta, b)$ conditional on the observed estimates $\hat{\Omega} = (\hat{\alpha}, \hat{\rho}_z)$ is given by

$$f_{\Theta|\hat{\Omega}}\left(\Theta|\hat{\Omega};\theta\right) = \frac{f_{\hat{\Omega}|\Theta}\left(\hat{\Omega}|\Theta\right) f_\Theta\left(\Theta;\theta\right)}{f_{\hat{\Omega}}\left(\hat{\Omega};\theta\right)}. \tag{21}$$

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

$$-2\log f_{\Theta|\hat{\Omega}}\left(\Theta|\hat{\Omega};\theta\right) = ((\beta - \beta^*)', (b - b^*)' \begin{bmatrix} v^*_{\beta\beta} & v^*_{\beta b} \\ v^{*\prime}_{\beta b} & v^*_{bb} \end{bmatrix} \begin{pmatrix} \beta - \beta^* \\ b - b^* \end{pmatrix} + C_2, \tag{22}$$

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 (22) 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 5.3 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 $(\beta_j, \beta_j + b_j, \beta_j + e^\beta_j)$ 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

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

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

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

Simplifying these expressions yields

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

$$\kappa_\beta = \frac{\sigma^2_\beta - \frac{Cov(\beta_j, \alpha_j)^2}{Var(\alpha_j)}}{Var(e^\alpha_j) + \sigma^2_\beta - \frac{Cov(\alpha_j, \hat{\beta}_j)^2}{Var(\alpha_j)}} = \frac{\sigma^2_\beta(1-R^2)}{Var(e^\beta_j) + \sigma^2_\beta(1-R^2)},$$

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

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

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

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

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 $\hat{\pi}_j^\ell$) and the prior parameters θ can be expressed

$$f_{\Theta|\hat{\Omega}}\left(\Theta|\hat{\Omega};\theta\right) \propto \phi_m\left(\hat{\Omega} - \Omega(\Theta); V\right)\phi_m\left(\Theta - \bar{\Theta}(\theta); \Gamma(\theta)\right), \tag{23}$$

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 (23). 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 π^ℓ_j are nonlinear functions of δ and ξ, the density in (23) 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

$$\alpha^*_{j.} = \left(\frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + Var(e^\alpha_j)}\right)\hat{\alpha}_j + \left(1 - \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + Var(e^\alpha_j)}\right)\hat{\mu}_\alpha, \tag{24}$$

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 (24) 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

This Appendix derives 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

Consistent with the bias tests reported in Table 3, we model school effects on seventh and eighth grade test scores as linear in the number of years spent in each school. The random coefficients framework of Section 5.1 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 A4 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 A5 displays 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 8, 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 6 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 ν^{ℓ}_{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 A8. These estimates suggest that the permanent component 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 6. 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 away 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.[25] Appendix Table A9 uses the simulations described in Section 7 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 A9, 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.

[25] 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.

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 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.

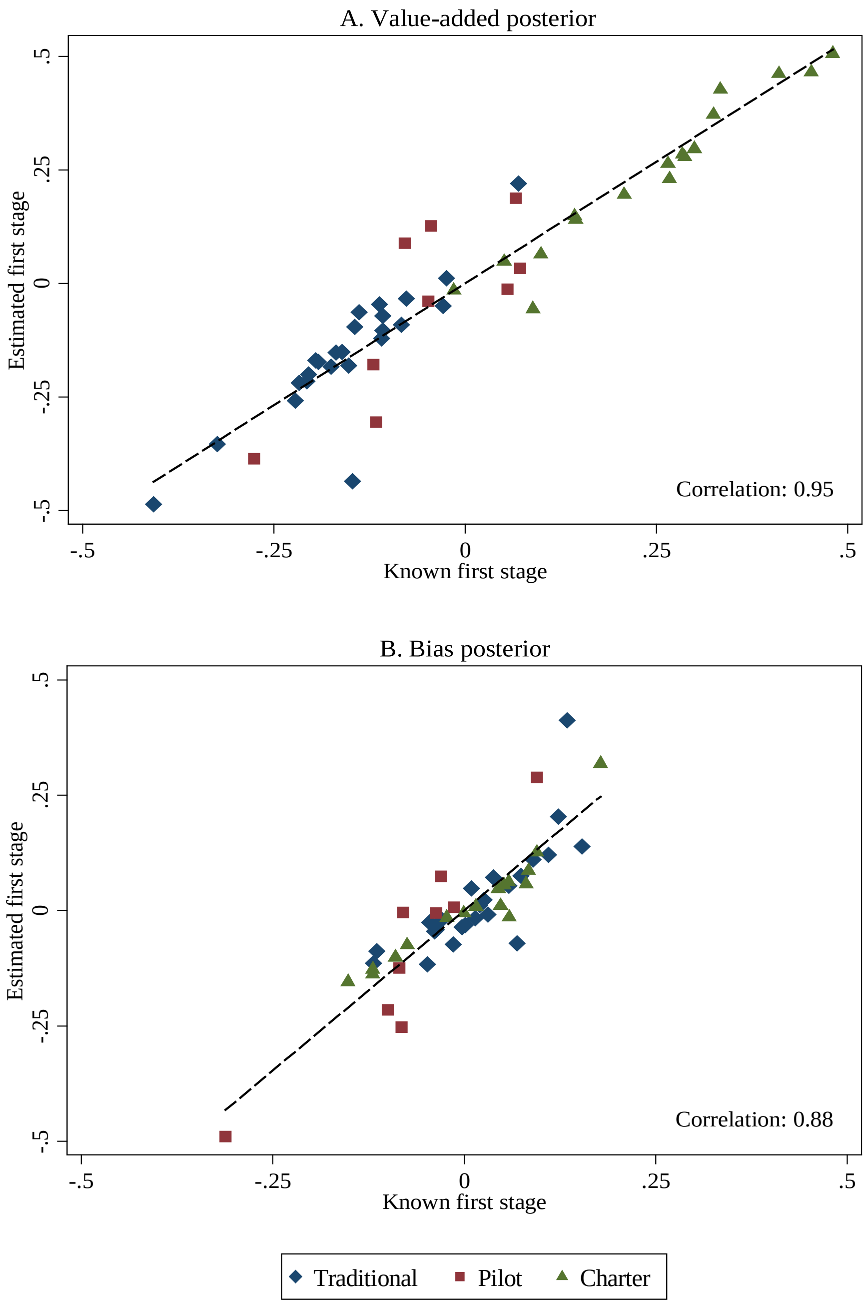
The relationship between school rankings based on true and estimated value-added summarizes the predictive value of VAM estimates. Column 7 of Table A9 reports coefficients from regressions of a school's rank in the causal value-added distribution on its rank in each estimated distribution. This rank 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 5, 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 mean bias for lottery schools is either 0.2σ above or 0.2σ below mean bias for non-lottery schools. As shown in Table 6, 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. Policymakers in these simulations continue to presume that there is no difference in bias between these groups.

As can be seen in Appendix Table A10, differences in bias between lottery and non-lottery schools slightly reduce the accuracy of hybrid estimates and the effects of VAM-based policies. Root mean squared error for hybrid gains estimates grows from 0.10σ to 0.14σ when bias for lottery schools is 0.2σ above or below bias for undersubscribed schools. A closure decision that replaces the lowest-ranked school with an average school generates a benefit of 0.31σ when the bias assumption is satisfied, and 0.28σ when mean bias for lottery schools is 0.2σ above or below average. This analysis shows that hybrid estimation is likely to be of value as long as differences in mean bias between lottery and non-lottery schools are modest.

Figure A1: Posterior predictions from treating the first stage as estimated vs. known



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 A1: 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 A2: 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 3 for a description of the value-added models and test procedure. Standard errors, clustered by student, are reported in parentheses.

Table A3: 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 ν_{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 6 for a description of the estimation procedure.

Table A4: 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	-0.016	-0.013	-0.007	-0.006	-0.004	-0.005
	VA and bias	(0.006)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
r_α	Regression of VA	0.694	0.783	0.625	0.747	0.771	0.798
	on OLS (reliability ratio)	(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 6 for a description of the estimation procedure.

Table A5: Per-year effects of closing the lowest-ranked district school for affected children, by grade

			Replacement school:			
Grade	Model	Posterior method	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 8 for a description of the simulation procedure.

Table A6: 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 3 for a description of the value-added models and test procedure. Robust standard errors are reported in parentheses.

Table A7: 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 8 by a factor of two. See the notes to Table 8 for a description of the simulation procedure.

Table A8: 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)
	χ^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 6 for a description of the estimation procedure.

Table A9: 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 A10: Sensitivity of hybrid posteriors to difference in bias between lottery and non-lottery schools

			Closure effects			
Difference in bias	Model	RMSE (1)	Average school (2)	Average above-median school (3)	Average top-quintile school (4)	Average charter school (5)
0σ	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]
0.2σ	Lagged score	0.146	0.295	0.408	0.511	0.647
			[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]
-0.2σ	Lagged score	0.152	0.263	0.365	0.454	0.615
			[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]

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. The difference in bias gives the mean difference in conventional VAM bias between schools with and without lotteries. 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 8 for a description of the simulation procedure.