Leveraging Lotteries for School Value-Added:
Testing and Estimation

Joshua Angrist  Peter Hull  Parag Pathak  Christopher Walters*
MIT and NBER  MIT  MIT and NBER  UC Berkeley and NBER

October 26, 2015

Abstract

We test and improve school value-added models (VAMs) using randomized admissions lotteries from Boston. Conventional VAMs compare average test scores across schools after regression-adjusting for demographic characteristics and previous scores. Estimates from such models are biased if the available controls are insufficient to capture differences in student ability. A new specification test developed here asks whether VAMs accurately predict the achievement consequences of random assignment to specific schools. The results indicate bias in conventional VAM estimates. This finding motivates the development of a hierarchical model describing the joint distribution of school value-added, VAM 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 predictions that optimally combine ordinary least squares and instrumental variables estimates of VAM parameters. Simulations calibrated to the Boston data show that, bias notwithstanding, accountability decisions based on conventional VAMs are likely to generate substantial achievement gains. Estimates incorporating lotteries are less biased and yield significant further gains.

*We gratefully acknowledge funding from the National Science Foundation, the Laura and John Arnold Foundation, and the Spencer Foundation. As always, we’re indebted to SEII research manager Annice Correia for invaluable help and support. Thanks also go to Isaiah Andrews, Pat Kline, Rick Mansfield, Chris Nielson, 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, and the Federal Reserve Bank of New York for suggestions and comments.
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 closure 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., forthcoming and Dobbie and Fryer, forthcoming 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 are misleading if the available control variables are inadequate to ensure *ceteris paribus* comparisons. VAM estimates are also likely to reflect considerable sampling error. The accuracy of teacher value-added models is the focus of a large and expanding body of research, but this work has yet to generate a consensus on the predictive value of VAM estimates 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; Baicher-Hicks et al., 2014; and Rothstein, 2014). The VAM research agenda has also been tilted towards teachers; in particular, while the social significance of school-level VAMs is similar to that of teacher VAMs, validation of VAMs for schools has received comparatively less attention than tests of VAMs for teachers.

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; 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, taking account of the partial coverage of lottery data. Our first contribution is the formulation of a new lottery-based test of VAM bias. 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 these earlier studies, however, we test the complete set of overidentifying restrictions implicit in an empirical VAM framework augmented with admissions lotteries. The test asks whether conventional VAM estimates correctly predict the effect of randomized admission at every school with a lottery.

Application of this test to data from Boston strongly suggests bias in conventional VAM estimates. Motivated by this finding, we develop and estimate a hierarchical random coefficients model that describes the joint distribution of value-added, selection 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. The SMD estimates are then used to compute empirical Bayes posterior predictions of individual school value-added. As shown below, the hybrid estimates that emerge from this procedure optimally combine relatively imprecise but unbiased instrumental variables (IV) estimates derived from lotteries with biased but relatively precise ordinary least squares (OLS) estimates. Importantly, the hybrid estimates make 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.

We assess the practical consequences of bias in conventional VAM estimates and the payoff to hybrid estimation with a Monte Carlo experiment based on data simulated from our estimated model. Simulation results show that despite the bias in conventional VAM estimates, policies based on estimates that rely on achievement changes from grade to grade or control for baseline scores appear worthwhile. For example, replacing the lowest-ranked Boston school with an average school is predicted to generate a gain of 0.2 standard deviations for affected students. This represents more than half the benefit that could be attained with knowledge of true value-added. Hybrid estimation reduces the root mean squared error of VAM estimates by about 30 percent, and closure decisions using hybrid estimates yield a further 0.1 standard deviations of achievement gain. Policies that replace low value-added schools with expansions of schools estimated to be high-performing are predicted to generate even larger impacts. These findings suggest that bias in conventional VAMs is not severe enough to fully offset the benefits of replacing schools that appear to be low-performing, while incorporating lotteries has the potential to improve policy targeting considerably.

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 the policy simulations. Finally, Section 8 concludes with remarks on how the framework outlined here might be used in other settings.

2 Setting and Data

2.1 Boston Public Schools

Boston students can choose from a diverse set of 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. 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 BPS district but typically operate with more flexibility over school 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 (grade-appropriate students who are happy to stay where they are need not participate in the match). With student preferences in hand, applicants are assigned to schools via a student-proposing deferred acceptance mechanism (described in Abdulkadiroğlu et al., 2006). This mechanism combines student preference rankings 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 by random sequence numbers, which we refer to as lottery numbers. In an evaluation of the pilot sector exploiting centralized random assignment, 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 college attendance (see, for example, Abdulkadiroğlu et al., 2011; Angrist et al., forthcoming).

---

2The 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.
2.2 Data and Descriptive Statistics

The data analyzed here consist of a sample of roughly 28,000 6th 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, offers are randomly assigned. Charter lottery offer 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.

The schools and students analyzed here are described in Table 1. We exclude schools serving fewer than 25 6th graders in each year, leaving a total of 25 traditional public schools, 9 pilot schools, and 17 charter schools. Of these, 28 schools (16 traditional, 7 pilot, and 5 charter) had at least 50 students subject to random 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.

Overall, lottery applicants look broadly similar to the larger BPS population. As shown in Table 2, 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 5th grade. Table 2 also documents the comparability of students who were and were not offered seats in a lottery. These results, reported in columns (3) through (6), compare the baseline characteristics of lottery winners and losers, controlling for assignment strata. Consistent with conditional random assignment of offers, the estimated differences by offer status are small and not significantly different from zero, both overall and within school sectors.
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 fundamental 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:

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

where $\mu_j$ is the mean potential outcome at school $j$ and $a_i$ is student $i$’s “ability,” or latent achievement potential. This additively-separable form 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 briefly explore this 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

$$Y_i = Y_{i0} + \sum_{j=1}^{J} (Y_{ij} - Y_{i0}) D_{ij} = \mu_0 + \sum_{j=1}^{J} \beta_j D_{ij} + a_i.$$

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

Conventional value-added models use regression methods in an attempt to eliminate selection bias. Write

$$a_i = X_i' \gamma + \epsilon_i,$$

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 $\gamma$. This decomposition implies that observed outcomes can be written

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

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

We’re 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:

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

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

The 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, a requirement that can be expressed as:

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

Restriction (6), sometimes called “selection-on-observables,” means that $\alpha_j = \beta_j$ for each school. In practice, of course, regression estimates need need 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 defined as the difference between regression and causal parameters for school $j$.

4 Testing Conventional VAM

4.1 Test Procedure

The variation in school attendance generated by admission lotteries at oversubscribed schools 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 $\ell$ are randomly assigned conditional on a set of lottery-specific stratifying variables, $C_{i\ell}$. These include an indicator for applying to school $\ell$ 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. In practice $C_i$ may include the vector of value-added controls $X_i$ as well.

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

$$E[\epsilon_i | C_i, Z_i] = \lambda_0 + C_i' \lambda_c,$$  \hspace{1cm} (7)

for a vector of parameters $\lambda_0$ and $\lambda_c$. This implies that admission offers are valid instruments for school attendance after controlling for lottery assignment strata.

With fewer lotteries than schools (that is, $L < J$), 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 implications of the conventional value-added framework. Selection-on-observables implies that $v_i = \epsilon_i$, so under assumption (6) the moment conditions (7) are equivalent to

$$E[v_i | C_i, Z_i] = \lambda_0 + C_i\lambda_c. \quad (8)$$

These restrictions generate an overidentification test of the sort widely used with IV estimators. Equation (8) can be tested by checking whether $\phi_z = 0$ in the regression model

$$v_i = \phi_0 + C_i\phi_c + Z_i\phi_z + \omega_i. \quad (9)$$

In practice equation (9) must be estimated using sample OLS residuals, $\hat{v}_i$, rather than population residuals $v_i$, adjusting inference to account for first-step estimation of the residuals (as described in Appendix B.1). A conventional instrumental variables (IV) overidentification test statistic has degrees of freedom given by the degree of overidentification; the orthogonality restrictions motivating a just-identified IV model can’t be tested. Here, however, instruments are unnecessary under the null hypothesis of VAM validity, so the relevant test procedure has $L$ degrees of freedom and even a single lottery generates a testable restriction.

Two variations on the test procedure based on equation (9) help to clarify the nature of the restrictions provided by randomized admissions lotteries. First, note that the VAM regression residual, $v_i$, is necessarily the difference between observed achievement and the fitted values generated by conventional OLS VAM estimation, denoted $\hat{Y}_i$. The two regressions

$$Y_i = \rho_0 + C_i\rho_c + Z_i\rho_z + \eta_i, \quad (10)$$

and

$$\hat{Y}_i = \psi_0 + C_i\psi_c + Z_i\psi_z + u_i, \quad (11)$$

should therefore produce the same result. In other words, testing the vector equality (with $L$ restrictions)

$$\rho_z = \psi_z \quad (12)$$

is the same as testing $\phi_z = 0$ in (9). This version of the test captures the intuition that effects of lottery offers on test scores should equal their effects on the predictions generated by an unbiased value-added model.

A further useful implication of the selection-on-observables restriction builds on (12). Specifically, (10) and (11) can be interpreted as the reduced form and first stage equations associated with a two-stage least squares procedure that uses lottery offers to instrument a model with $Y_i$ on the left-hand side and $\hat{Y}_i$, treated as endogenous, on the right. Using all lottery offers as instruments for this model, or using them one at a time, each of the resulting IV estimates should equal 1. This too amounts to a test of $L$ restrictions.
This interpretation links our approach with the “forecast bias” tests implemented in previous efforts to validate VAMs (Kane and Staiger, 2008; Kane et al., 2013; Deming, 2014; Chetty et al., 2014a). These tests ask whether the coefficient on predicted value-added equals one in procedures similar to the IV model described here. Such tests have one degree of freedom, however, even though the underlying models generate an additional testable restriction for every available quasi-experiment. Our test leverages all of these overidentifying restrictions rather than just one.

4.2 Test Results

The conventional VAM setup assessed here includes four models. The first, referred to as “uncontrolled”, adjusts only for year effects; estimates from this model are essentially school average test score levels. The second, a “demographic” specification, includes indicators for sex, race, subsidized lunch eligibility, special education, English-language learner status, and counts of baseline absences and suspensions. The third, labeled a “lagged score” specification, adds cubic functions of baseline math and ELA test scores. Lagged score models 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). Finally, we consider a “gains” specification that replaces score levels with grade-to-grade score changes in the demographic specification. This model, which parallels common accountability policies that measure test score growth, identifies causal effects under the alternative assumption that test scores imperfectly proxy for a latent non-depreciating human capital factor that drives school choices.3

Figure 1 summarizes value-added estimates from the four models applied to math scores.4 Each bar reports an estimated standard deviation of \( \alpha_j \) across schools, expressed in test score standard deviation units (\( \sigma \)) and adjusted for estimation error.5 Adding controls for demographic variables and previous scores reduces the standard deviation of \( \alpha_j \) from 0.5\( \sigma \) in the uncontrolled model to about 0.2\( \sigma \) in the lagged score and gains models. This implies that observed student characteristics explain a substantial portion of the variation in school test score levels. The last four bars in Figure 1 report estimates of within-sector value-added standard deviations, constructed using residuals from regressions of \( \hat{\alpha}_j \) on charter and pilot school indicators. Including sector effects reduces variation in value-added, suggesting large differences in school effects across sectors.

Table 3 describes the results of tests for bias in conventional VAMs. The first row shows value-added coefficients from overidentified IV models instrumenting VAM fitted values with lottery offers as in equations

3The gains specification can be motivated as follows: suppose that human capital in grade \( g \) equals lagged human capital plus educational quality, so that \( A_{ig} = A_{ig-1} + q_{ig} \) where \( q_{ig} = \sum_j \beta_j D_{ij} + \eta_{ig} \) and \( \eta_{ig} \) is a random error independent of school choices. Suppose further that test scores are noisy proxies for human capital, so that \( Y_{ig} = A_{ig} + \nu_{ig} \) where \( \nu_{ig} \) is classical measurement error. Finally, suppose that students sort into schools in grade \( g \) on the basis of \( A_{ig-1} \). Then a lagged scores model that controls for \( Y_{ig-1} \) generates biased estimates, but a gains model that uses \( Y_{ig} - Y_{ig-1} \) as the outcome variable measures value-added correctly.

4We 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 A1, yield similar results.

5The estimated standard deviations 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 \( \hat{\alpha}_j \).
These models are estimated via a two-stage optimal IV procedure that is asymptotically efficient under heteroskedasticity (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. A strong first stage is an important requirement for the IV version of the test: with a weak first stage, IV estimates are biased towards the corresponding OLS estimates, which in this case equal one by construction. A weak first stage therefore makes the IV version of the VAM bias test less likely to reject. The $F$-statistics in columns (1) through (4) range from 26 to 46, suggesting finite-sample bias is unlikely to be a concern.

The remaining rows of Table 3 report $p$-values for three tests. The first of these tests the null hypothesis that the IV value-added coefficient equals one. The second tests the IV model’s overidentifying restrictions, which require that value-added coefficients are the same in all lotteries but not necessarily equal to one. The third tests the full set of restrictions by regressing VAM residuals on lottery offers as in equation (9). Since asymptotic critical values for overidentification tests can be inaccurate, the table also reports $p$-values for the third test based on a version of the bootstrap procedure developed in Hall and Horowitz (1996), which has been shown to yield higher-order refinements. This refinement uses a Bayesian bootstrap (as described by Rubin, 1981; see Appendix B.1 for details).

The results of these tests suggest that conventional value-added estimates are biased. The uncontrolled and demographic specifications generate IV coefficients of 0.40 and 0.65, and all three tests reject at conventional levels for these models ($p < 0.001$). The lagged score and gains specifications account for past test scores, powerful control variables in models measuring test score value-added. IV coefficients for these models equal 0.86 and 0.95, and the latter is not statistically different from one at conventional levels. Nonetheless, the overidentifying restrictions for these models are rejected ($p < 0.01$), and the null hypothesis of joint predictive accuracy across all lotteries is decisively rejected for both specifications ($p < 0.001$). These findings show that conventional value-added estimates fail to predict lottery effects even after controlling for a rich set of baseline covariates.

The source of this failure can be seen in Figure 2, which plots lottery-specific reduced form effects on test scores against first stage effects on OLS fitted values. Each panel also displays a line through the origin with slope equal to the relevant IV coefficient from Table 3 (solid lines) along with the 45-degree line (dashed lines). If VAM estimates perfectly predict lottery effects, the 45-degree line should fit all points up to sampling error. Shaded points indicate estimates that differ from the 45-degree line at the 10-percent confidence level. Consistent with the results in Table 3, the clouds of points for the uncontrolled and demographic specifications do not seem to follow the prescribed trend, with most quasi-experiments rejecting VAM validity at the 10-percent level. Slopes are much closer to one for the lagged score and gains specifications, but points for many individual lotteries remain far from the trend lines, leading to rejection.

---

6The OLS version of this model is a regression of test scores on VAM fitted values. If estimated in the same sample as the value-added model with no additional controls, this regression produces a coefficient that is algebraically one. In practice, the OLS and IV specification differ in that the latter control for lottery strata and exclude some students.

7To check whether this finding generalizes beyond Boston we also applied our test procedure to the Charlotte-Mecklenberg lottery data analyzed by Deming (2014). For a VAM including controls for past achievement, our test generated a bootstrap-refined joint $p$-value of 0.002.
of the overidentifying restrictions.

The results in Figure 2 highlight the advantage of our approach over previous efforts to validate VAMs. It is visually evident that much of the power to detect bias in the lagged score and gains models comes from lotteries that generate points near the vertical axis, implying small value-added first stages and large impacts on test scores. Such lotteries contribute weak instruments to the IV model and therefore have little influence on the value-added coefficient. These points clearly indicate that the predictions of the value-added model are violated, however, a finding that is captured by the overidentification test results. Previous efforts to validate VAMs focus only on the value-added coefficient in similar models, ignoring overidentifying restrictions. In our Boston data, testing these extra restrictions results in a more decisive rejection of value-added than a test based on the IV coefficient alone.

Figure 2 also reveals that much of the value-added models’ predictive power is generated by charter school lotteries, which contribute large first stage and reduced form effects. The relationship between OLS value-added and lottery effects is less clear in the pilot and traditional public school sectors. This impression is confirmed in column (5) of Table 3, which reports results for the lagged score specification excluding charter lotteries. The value-added coefficient for this model is further from one (0.55), though the absence of charter lotteries also reduces its precision. The corresponding p-value from a joint test of all restrictions equals 0.002, indicating that the inclusion of charter lotteries is not necessary to detect bias in OLS value-added models.

4.3 Heterogeneity vs. Bias

The test results in Table 3 show that conventional VAM estimates fail to predict the changes in achievement generated by randomly assigned offers of admission. In a constant effects model this implies that conventional estimates are biased. In a world of heterogeneous causal effects, however, these test results might instead signal divergence between the local average treatment effects (LATEs) identified by lottery instruments and possibly more representative effects captured by OLS (Imbens and Angrist, 1994; Angrist et al., 1996).

It bears emphasizing that rejections driven by effect heterogeneity pose a problem for the value-added framework more generally. In the presence of heterogeneous effects OLS identifies a variance-of-treatment weighted average causal effect that need not have predictive value for specific individuals (Angrist, 1998). The value-added enterprise is built on a foundation of limited variation in causal effects: the goal here is prediction, and value-added estimates are of little policy relevance if they fail to reliably predict the effects of changing school assignments. Nevertheless, it is useful to know whether our rejection of value-added is due to bias or effect heterogeneity. We therefore present a brief exploration of heterogeneity in value-added estimates.

8The first stage F-statistic for the specification without charter lotteries is 11.2, on the border of the range conventionally thought to indicate weak instruments (Staiger and Stock, 1997). To explore this possibility we also estimated limited information maximum likelihood (LIML) models which are more robust to weak instruments (Stock and Yogo, 2005). The LIML estimate corresponding to column (5) equals 0.54 with a standard error of 0.19, very similar to the IV estimate.

9See Condie et al. (2014) for a discussion of the hazards of teacher value-added estimation with heterogeneous effects.
This investigation tests whether differences between IV and OLS estimates are explained by differences in observed characteristics between lottery compliers and other students, or by differences in the samples used to construct both sets of estimates. Two analyses shed light on these possibilities. The first is a set of bias tests using OLS VAM specifications that allow school effects to differ across covariate-defined “types” of students (e.g. special education students or those with low levels of baseline achievement). Intuitively, this approach accounts for variation in school effects across covariate cells that may be weighted differently by IV and OLS; see Appendix B.2 for a formal justification of this technique. The second tests for bias in OLS VAMs estimated in the lottery sample. This approach checks whether differences between IV and OLS are caused by differences between students subject to lottery assignment and the general student population.

The results of these analyses suggest that our rejection of value-added models is due to bias rather than effect heterogeneity. Panel A of Table 4 reports test results for a version of the lagged scores model with school effects that vary across student types. Column (2) allows value-added to differ by school year to address the concern that school effects may “drift” over time (Chetty et al., 2014a). Although the IV coefficient rises relative to the unadjusted specification repeated in column (1), the joint test for VAM bias continues to reject with a bootstrap-refined $p$-value less than 0.001. Columns (3) through (5) define types based on subsidized lunch eligibility, special education status, and baseline test score terciles. Results from these specifications are similar to the pooled results in column (1). Column (6) allows value-added to differ across cells constructed by fully interacting six covariates: race, sex, subsidized lunch eligibility, special education, English-language learner status, and baseline score tercile. The test for VAM validity continues to reject at the 1-percent level in this specification ($p = 0.002$). Finally, panel B of Table 4 shows that tests based on OLS VAMs estimated within the quasi-experimental sample yield similar $p$-values. These findings indicate that differences between lottery applicants and other students are not the source of differences between OLS and lottery estimates.

5 The Distribution of School Effectiveness

While our results indicate the presence of bias in conventional value-added models, the highly significant IV coefficients for the lagged score and gains specifications also show that VAM estimates have substantial predictive power. It’s worth asking whether bias in OLS VAM estimates is severe enough to meaningfully affect accountability policies that rely on value-added models. The answer to this question depends on the joint distribution of value-added and bias across schools, which cannot be determined from the estimates in Table 3 alone. This section develops methods to quantify and reduce bias in value-added estimates.

5.1 Model

Our analysis 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 observed OLS, lottery reduced form, and
first stage coefficient estimates. Let $\rho^\ell_z$ denote the element of $\rho_z$, the reduced form coefficient vector from equation (10), corresponding to $Z_{i\ell}$. Equations (4) and (7) imply that

$$
\rho^\ell_z = \sum_{j=1}^{J} \pi^\ell_j \beta_j,
$$

where $\pi^\ell_j$ is the first-stage coefficient on $Z_{i\ell}$ from a regression of $D_{ij}$ on $Z_i$ and $C_i$.\(^{10}\) This expression shows that the lottery at school $\ell$ identifies a linear combination of value-added parameters, with coefficients equal to shares of students shifted into (or out of) each school by the lottery offer. Estimates of the first-stage coefficients, $\pi^\ell_j$, are obtained by substituting $D_{ij}$ for $Y_i$ on the left-hand side of (10).

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 random coefficient distribution in the population of schools. The observed estimates can be written:

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

$$
\hat{\rho}^\ell_z = \sum_{j} \pi^\ell_j \beta_j + e^\rho_{\ell},
$$

$$
\hat{\pi}^\ell_j = \pi^\ell_j + e^\pi_{\ell j},
$$

(13)

The $e_j$ terms in these equations denote 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 these asymptotic approximations seem reasonable.

The second level of the model treats the school-specific parameters $\beta_j$, $b_j$ 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 $\ell$ on the probability of attending this school is parameterized as

$$
\pi^\ell = \frac{\exp (\delta^\ell)}{1 + \exp (\delta^\ell)},
$$

(14)

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

$$
\pi_j = -\pi^\ell_j \cdot \frac{\exp (\xi_j + \nu^\ell_j)}{1 + \sum_{k \neq \ell} \exp (\xi_k + \nu^\ell_k)},
$$

(15)

---

\(^{10}\)Conditional random assignment of admission offers implies that $Z_i$ is conditionally independent of baseline covariates $X_i$ in addition to $\epsilon_i$. $Z_i$ is therefore conditionally independent of all terms in equation (4) except $\sum_j \beta_j D_{ij}$, so a regression of this quantity on $Z_i$ controlling for $C_i$ produces $\rho_z$. 

13
The quantity $\xi_j + \nu^\ell_j$ is the mean utility for school $j$ in a multinomial logit model predicting fallback school choices among students that comply with offers in lottery $\ell$. The parameter $\xi_j$ is constant across lotteries, which allows for the possibility that some schools are consistently likely to serve as fallback options for lottery losers. $\nu^\ell_j$ is a random utility shock specific to school $j$ in the lottery at school $\ell$. The parametrization in (14) and (15) guarantees that a lottery offer increases the probability of enrollment at the target school and reduces enrollment probabilities at all other schools, and that effects on all probabilities are between zero and one in absolute value.\footnote{This parametrization implies that $0 < \pi^\ell_\ell < 1$, $-1 < \pi^j_\ell < 0$ for $j \neq \ell$, and $\pi^j_\ell > -\sum_{j \neq \ell} \pi^j_\ell$. The residual probability $\pi^\ell_\ell + \sum_{j \neq \ell} \pi^j_\ell$ is minus the effect of an offer at the omitted school with index zero.}

Each school is characterized by a vector of four parameters: a value-added coefficient $\beta_j$, a selection bias term $b_j$, an offer compliance utility $\delta_j$, and a mean fallback utility $\xi_j$. We treat these vectors as draws from a prior distribution in a hierarchical Bayesian framework. A key assumption in this analysis 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 generate posteriors for non-lottery schools that account for bias in OLS 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 this empirically).

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 multivariate normal distribution:

$$
(\beta_j, b_j, \delta_j, \xi_j)|Q_j \sim N((\beta_0 + \beta_Q Q_j, b_0, \delta_0, \xi_0), \Sigma).
$$

(16)

The parameter $\beta_Q$ capture the possibility that average value-added differs for schools with lotteries. The covariance matrix $\Sigma$ measures variances and covariances of value-added, bias, and first stage 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 normal:

$$
\nu^\ell_j|Q_j \sim N(0, \sigma^2_\nu).
$$

(17)

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 informative about the parameters of
interest, minimizing deviations between sample moments and corresponding model 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 \( \hat{\pi}_\ell \). One moment, for example, is the average \( \hat{\alpha}_j \) across schools; another is the variance of the \( \hat{\alpha}_j \). Other moments are means and variances of reduced form and first stage estimates across lotteries. Appendix B.3 lists all moments used for SMD estimation.

The moments are complicated functions of the hyperparameters, which motivates a simulation approach.\(^{12}\) Moments are simulated by fixing a value of \( \theta \) and drawing a vector of school-level parameters using equations (16) and (17). Likewise, the simulation draws a vector of the estimation errors in (13) from the joint asymptotic distribution of the OLS, reduced form and first stage estimates. Putting the parameter and estimation draws together generates a simulated vector of parameter estimates for the given value of \( \theta \), which are then transformed into a set of 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.3, we report two-step SMD estimates with an efficient weighting matrix in the second step.

### 5.3 Empirical Bayes Posteriors

SMD estimates of \( \theta \) can be used to form posterior predictions of school quality. Studies of teacher and school value-added typically employ empirical Bayes strategies that shrink noisy teacher- and school-specific 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 is presumed unbiased and estimation error is uncorrelated, the posterior mean value-added for school \( j \) is

\[
E[\alpha_j|\hat{\alpha}_j] = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + Var(\epsilon^\alpha_j)} \hat{\alpha}_j + \left(1 - \frac{\sigma^2_\alpha}{\sigma^2_\alpha + Var(\epsilon^\alpha_j)}\right) \alpha_0,
\]

where \( \alpha_0 \) and \( \sigma^2_\alpha \) are the mean and variance of OLS VAM parameters. An EB posterior mean plugs estimates of these hyperparameters into (18).

Our setup extends this idea to a scenario where the estimated \( \hat{\alpha}_j \) may be biased but lottery estimates are available to reduce this bias. Although lottery-based IV estimates are consistent estimates of causal effects, OLS estimates may still be valuable. The price for eliminating bias is a loss of precision: because IV uses only the variation generated by random assignment, lottery-based estimates are likely to be considerably less precise than corresponding OLS estimates. Some schools are also undersubscribed, so there are fewer instruments than schools and a lottery-based IV model is underidentified. The empirical Bayes approach optimally trades off the advantages and disadvantages of each approach to generate minimum mean squared error (MMSE) estimates of value-added (Judge and Mittlehammer, 2004, 2005, 2007).

\(^{12}\)For example, the mean reduced form is \( E[\rho^\ell] = \sum_j E[\pi^\ell_j \beta_j] \). The right-hand side is the expected product of a normally distributed random variable with a ratio involving correlated log-normals, a transformation for which no analytical expression is readily available.
To see how this trade-off works, consider a simplified setting in which the first stage parameters $\pi^\ell_j$ are known rather than estimated (i.e. $e_{ij}^\ell = 0 \forall \ell, j$). Let $\Pi$ denote the $L \times J$ matrix of these parameters, and let $\beta, \hat{\alpha}$ and $\hat{\rho}_z$ denote vectors collecting $\beta_j, \hat{\alpha}_j$ and $\hat{\rho}_z^\ell$. Appendix B.4 shows that the posterior distribution for $\beta$ is multivariate normal with mean:

$$E[\beta | \hat{\alpha}, \hat{\rho}_z] = W_1(\hat{\alpha} - b_0 \iota) + W_2 \hat{\rho}_z + (I_J - W_1 - W_2 \Pi) \beta_0 \iota,$$

(19)

where $\iota$ is a $J \times 1$ vector of ones and $I_J$ is the $J$-dimensional identity matrix. Posterior mean value-added is a linear combination of OLS estimates net of the mean bias $b_0$, lottery reduced form estimates, and the mean value-added $\beta_0$. The weighting matrices, $W_1$ and $W_2$, are functions of the first stage parameters and second moments of estimation error, value-added, and bias. Closed-form expressions for these matrices are provided in Appendix B.4.

As with conventional EB posteriors, an empirical Bayes version of the posterior mean plugs first-step estimates of $b_0, \beta_0, W_1, \text{and} W_2$ into equation (19). With known first stage coefficients this hyperparameter estimation is simplified by noting that

$$E[\hat{\rho}_z^\ell] = \sum_j \pi^\ell_j \beta_0,$$

$$E[\hat{\alpha}_j] = \beta_0 + b_0.$$

The mean hyperparameters $\beta_0$ and $b_0$ may therefore be estimated as

$$\hat{\beta}_0 = \frac{1}{L} \sum_{\ell} \sum_j \pi^\ell_j \hat{\rho}_z^\ell,$$

$$\hat{b}_0 = \frac{1}{J} \sum_j \hat{\alpha}_j - \hat{\beta}_0.$$

Similarly, the variance hyperparameters that constitute $W_1$ and $W_2$ may be estimated from second moments of $\hat{\alpha}_j$ and $\hat{\rho}_z^\ell$.

Suppose that all schools are oversubscribed, so $L = J$. In this case the first stage matrix $\Pi$ is square; if it is also full rank, the parameters of equation (4) are identified using lotteries alone. A vector of IV value-added estimates may be computed by indirect least squares as $\hat{\beta} = \Pi^{-1} \hat{\rho}_z$. In this case the posterior mean in equation (19) becomes

$$E[\beta | \hat{\alpha}, \hat{\rho}_z] = W_1(\hat{\alpha} - b_0 \iota) + \tilde{W}_2 \hat{\beta} + (I_J - W_1 - \tilde{W}_2) \beta_0 \iota,$$

(20)

for $\tilde{W}_2 = W_2 \Pi$. When there are enough instruments to estimate an IV value-added model, the posterior mean is a weighted average of unbiased IV estimates, OLS estimates net of mean bias, and mean value-added,
with weights that sum to the identity matrix.\textsuperscript{13} This estimate jointly shrinks the IV and OLS estimates toward the mean to account for sampling error in both estimates and the bias of OLS. The shrinkage factors $W_1$ and $\tilde{W}_2$ minimize the mean squared error of the resulting value-added predictions.

In practice some schools are undersubscribed, so $\hat{\beta}_j$ is not estimable at every school. Nevertheless, equation (19) shows that predictions at schools without lotteries may still be improved by lottery information from other schools. Lottery reduced form parameters contain information for all fallback schools, including for those without their own lotteries. This is a consequence of the relationship described by equation (13), which shows that the reduced form for any school with a lottery depends in part on the value-added of the other schools that the applicants to this school might attend. If $\pi_j^\ell \neq 0$ the reduced form for school $\ell$ puts weight on $\beta_j$ and therefore contains a signal that can be used to improve the posterior prediction of this parameter.

Finally, equation (19) shows how knowledge of conventional VAM bias can be used to improve posterior predictions even for schools that are never lottery fallbacks. Appendix B.4 shows that the posterior mean for this case can be approximated via Markov Chain Monte Carlo (MCMC). In practice, the posterior mean does not have a closed form. The improvement here comes from the fact that the schools for which we do have lotteries provide information that can be used to estimate the distribution of bias, allowing at least a partial correction for bias in OLS estimates at schools without lotteries.

Equation (19) is a pedagogical formula derived assuming a known first stage matrix. With an estimated first stage the posterior distribution for $\beta$ is not normal and the posterior mean does not have a closed form. The posterior mean for this case can be approximated via Markov Chain Monte Carlo (MCMC). In practice, however, with a high-dimensional random coefficient vector such the one featured in our model, MCMC may be sensitive to starting values or other user parameters. As in Chamberlain and Imbens (2004) we instead work with EB posterior modes (also known as maximum \textit{a posteriori} estimates; see Gelman et al., 2013),

\textsuperscript{13}We use the term “unbiased” here somewhat loosely. Just-identified IV estimates have no moments, so the mean of $e^\beta_j$ is undefined. Still, as with OLS, IV estimation error disappears as the sample size grows. Note also that because the IV procedure sketched here is just-identified, the resulting estimates are approximately median-unbiased in the sense that the estimation error has a median of approximately zero (Angrist and Pischke, 2009).

\textsuperscript{14}Recalling that $\alpha_j = \beta_j + b_j$, equation (18) may be written

\[ E [\alpha_j | \hat{\alpha}_j] = \frac{\sigma_\beta^2 + \tau \sigma_\beta \sigma_b + 2 \tau \sigma_\beta \sigma_b + Var (e_j^\alpha)}{\sigma_\beta^2 + \sigma_b^2 + 2 \tau \sigma_\beta \sigma_b + Var (e_j^\alpha)} \hat{\alpha}_j + \left(1 - \frac{\sigma_\beta^2 + \tau \sigma_\beta \sigma_b + 2 \tau \sigma_\beta \sigma_b + Var (e_j^\alpha)}{\sigma_\beta^2 + \sigma_b^2 + 2 \tau \sigma_\beta \sigma_b + Var (e_j^\alpha)} \right) (\beta_0 + b_0). \]

A comparison with equation (21) reveals that the conventional posterior is centered at a difference place than the hybrid posterior and places a different weight on the OLS estimate $\hat{\alpha}_j$. The conventional posterior puts more weight on OLS if $\sigma_b > -\tau \sigma_\beta$. 


which maximize the log likelihood of the observed estimates conditional on the random coefficients plus the log likelihood of the random coefficients conditional on the estimated hyperparameters. In our model calculating the posterior mode turns out to be relatively quick and straightforward; see Appendix B.4 for details. When posterior value-added is normally distributed as in the fixed first stage case, the posterior mode and mean coincide – in practice the posterior modes are similar to the weighted averages generated by equation (19), with a correlation coefficient of 0.82.

6 Parameter Estimates

6.1 Hyperparameters

We estimate hyperparameters starting with each of the four math value-added specifications tested in Table 3, along with sets of lottery reduced form and first stage estimates.\footnote{These come from regressions of test scores and school attendance on lottery offers with controls for randomization strata and the baseline covariates from the lagged score VAM specification.} As can be seen in columns (1) through (4) of Table 5, the results reveal substantial heterogeneity in both causal value-added and selection bias across schools. The standard deviation of value-added, $\sigma_\beta$, is similar across specifications, ranging from about 0.20$\sigma$ in the gains specification to 0.22$\sigma$ in the lagged scores model. This stability is reassuring: the control variables that distinguish these models should not change the underlying distribution of school effectiveness if our estimation procedure works as we hope.

In contrast with relatively stable estimates of $\sigma_\beta$, the standard deviation of bias, $\sigma_\theta$, shrinks from 0.49$\sigma$ in the uncontrolled model to 0.17$\sigma$ in the lagged score and gains specifications. Evidently, controlling for observed student characteristics dramatically reduces the degree of bias in conventional value-added estimates. On the other hand, 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. Our estimates suggest variation in causal value-added is slightly larger than variation in bias due to selective student enrollment.

The estimated correlation between $\beta_j$ and $b_j$ (the hyperparameter $\tau$) is negative for the lagged score and gains specifications, a result that can be seen in the third row of Table 5. This suggests that conventional models may overstate the effectiveness of low-quality schools and understate the effectiveness of high-quality schools, though the estimates are too imprecise to be conclusive. Estimates of $\beta_Q$, the lottery school value-added shifter, are negative but statistically insignificant in columns (1) through (4). This implies that differences between oversubscribed and undersubscribed schools are modest; if anything, lottery schools appear to generate slightly smaller gains.

In previous work we’ve demonstrated differences in effectiveness between Boston’s charter, pilot and traditional public sectors (Abdulkadıroğlu et al., 2011; Angrist et al., forthcoming). This suggests that accounting for sector effects may improve the predictive accuracy of school value-added models. Columns
(5) and (6) of Table 5 report estimates from versions of the lagged score and gains models that allow the means of the four random coefficients to depend on school sector. Appendix Table A2 shows the full joint distribution of random coefficients from the lagged score model.

Consistent with earlier findings, the results imply a large charter advantage: charter school value-added exceeds traditional public school value-added by 0.35σ on average. Differences in value-added between pilot and traditional public schools are smaller and statistically insignificant. The estimated bias parameters indicate no systematic difference in bias across sectors in either model. This also corroborates past work indicating that observational models with demographic and lagged achievement controls accurately reproduce lottery-based comparisons of the charter, pilot and traditional sectors (Abdulkadiroğlu et al., 2011; Angrist et al., 2013). With sector effects included estimates of β_Q are positive, implying that the negative estimates in models without sector effects are driven by sectoral differences in oversubscription and value-added. Within sectors, schools with lotteries appear to perform slightly better than schools without. Finally, the estimates of σ_β and σ_b show that sector effects do not eliminate variation in value-added and bias. Within-sector standard deviations of value-added and bias are smaller than the corresponding estimates from models without sector effects, but these estimates are statistically significant in columns (5) and (6).

6.2 Empirical Bayes Posteriors for Value-added and Bias

The posterior modes generated by our hybrid estimation strategy are positively correlated with conventional posterior means that presume no bias in OLS value-added estimates. This is evident in Figure 3, which plots hybrid modes against conventional means for each of the four models. Hybrid modes are highly correlated with conventional means in the demographic, lagged score and gains models, implying that the posterior modes place significant weight on the OLS estimates. On the other hand, there are some rank reversals: rank correlations in the lagged score and gains models are 0.78 and 0.77, respectively. This suggests that incorporating information from lotteries may affect accountability decisions based on estimated value-added. The next section compares the effects of policies based on conventional and hybrid measures of school quality.

Hybrid estimation generates posterior modes for selection bias terms b_j in addition to value-added parameters β_j. We use the bias modes to assess the plausibility of our assumption that bias distributions are the same for schools with and without lotteries. While this assumption is not directly testable, we can look for a relationship between the estimated bias distribution and the extent of oversubscription for schools that have lotteries. Lack of a connection between the degree of oversubscription and bias is consistent with the hypothesis that bias distributions are similar for schools where lottery information is entirely absent.

The relationship between bias and oversubscription supports this assumption. This is documented in Figure 4, which plots bias posterior modes from the lagged score model with sector effects against oversubscription rates at lottery schools. The oversubscription rate equals 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. Points in the figure
are residuals from regressions of bias posterior modes and oversubscription rates on sector indicators. Figure 4 shows an essentially flat scatter plot: a regression of bias residuals on oversubscription residuals produces a slope coefficient of 0.01 with a standard error of 0.04.

7 Policy Simulations

Accurate value-added estimates are useful for parents looking for high-quality schools, and for policymakers looking to expand or reward successful schools while scaling back or sanctioning laggards. We next assess the accuracy of conventional and hybrid value-added estimates with these purposes in mind. This assessment relies on a set of Monte Carlo simulations applying VAM-based policy rules to data from our estimated hierarchical model. Each simulation draws vectors of causal value-added, bias, and lottery parameters from the estimated distributions underlying Table 5, along with vectors of estimation error from the joint asymptotic distribution of OLS and lottery estimates. These draws are used to construct simulated OLS, reduced form and first stage parameter estimates. After re-estimating the model on these simulated data we construct conventional and hybrid EB posterior predictions. Finally, these predictions are compared to the underlying causal value-added parameters. Key metrics for evaluating VAM performance include root mean squared error (RMSE), misclassification rates for schemes aimed at identifying high- and low-performing schools, and student achievement gains resulting from closure and replacement of low-ranking schools.

7.1 Statistical Error

RMSE in conventional VAM estimates falls sharply as controls for observable characteristics are added, while the RMSE of hybrid VAM estimates is much more stable. This can be seen in Figure 5, which compares RMSE across specifications and estimation procedures. The sharp decline in the RMSE of conventional VAM estimates as the set of controls grows richer reflects reduced bias in the conventional estimates; the stability of the hybrid procedure’s RMSE is evidence of successful bias mitigation regardless of the conventional starting point. For example, the hybrid posterior mode reduces RMSE to just under $0.2\sigma$ for the uncontrolled and demographic specifications from starting points of $0.48\sigma$ and $0.31\sigma$, an impressive gain. In other words, the hybrid estimates correctly reflect the fact that most of the variation in $\hat{\alpha}_j$ is due to bias in models that don’t adjust for previous scores.

Although conventional VAM estimates generate much lower RMSE values using the lagged score and gains specifications, the hybrid approach yields improvements for these specifications as well. An RMSE of $0.17\sigma$ for conventional posteriors derived from the lagged score and gains specifications falls to $0.14\sigma$ when sector effects are included in the lagged score model. Hybrid posteriors pull this down to $0.10\sigma$ when sector effects are included, a gain of almost 30 percent.
7.2 Accountability Targeting

Many states and districts use accountability schemes based on standardized tests. Massachusetts’ Framework for School Accountability and Assistance, for example, places schools into five tiers based on four-year histories of test score levels and changes. Schools in the bottom quintile of this measure are automatically placed in level 3 or above, and a subset are then classified in levels 4 and 5, putting them at risk of restructuring or closure.\[16\]

Table 6 looks at the accuracy of VAM-based accountability classification schemes of this sort. Specifically, this table reports simulated misclassification rates for policies aimed at identifying BPS district (traditional and pilot) schools above or below particular percentiles of the true value-added distribution. Column (1) displays the frequency at which schools in the lowest decile of causal value-added are not ranked in the lowest decile by various value-added prediction methods. This is effectively the type II error rate for a test that rejects when a school is ranked in the lowest decile. Columns (2) and (3) report corresponding misclassification rates for lowest quintile and lowest tercile schools, while columns (4) through (6) show error rates for highest decile, quintile and tercile schools.

Uncontrolled value-added estimates produce very inaccurate school rankings. As shown in the second row of Table 6, an uncontrolled VAM misclassifies 84 percent of lowest decile schools, 68 percent of lowest quintile schools, and 54 percent of lowest tercile schools. These rates are not much better than the error rates for a policy that ranks schools at random (90, 80 and 67 percent, shown in the first row). This finding implies that school report cards based on unadjusted achievement levels, distributed in many states and districts, are likely to be highly misleading.\[17\] Hybrid posterior modes that combine uncontrolled OLS and lottery estimates misclassify 65, 51 and 44 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.

Adding controls for observed characteristics reduces misclassification rates for both value-added prediction methods but does not eliminate the utility of hybrid estimates. Conventional misclassification rates for lowest decile, quintile and tercile schools are 65, 47 and 39 percent when rankings are based on estimates from the gains specification. Hybrid estimation reduces the error rate in the lowest decile to 55 percent, implying a 15 percent decline in mistakes. Similarly, the hybrid procedure reduces misclassification for lowest quintile and tercile schools by 16 and 10 percent in the gains specification. Improvements for high-performing schools are even larger: incorporating lotteries cuts mistakes by 31, 26 and 24 percent for highest decile, quintile and tercile schools. The lagged score specification typically produces slightly higher error rates than the gains model, but patterns of results for these two specifications are broadly similar.

\[16\] The Massachusetts accountability system also uses information on graduation, dropout rates and site visits to classify schools; see http://www.doe.mass.edu/aps/sss/turnaround/level5/schools/FAQ.html for more information.

\[17\] 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).

21
7.3 Effects on Students

Figure 5 and Table 6 indicate that the combination of lagged score adjustments and lottery-based bias correction reduces statistical error and misclassification rates. To better interpret the magnitudes of these improvements we next consider the effects on student achievement that would result from using VAM estimates for school closure and expansion decisions. The analysis here ignores any direct disruptive effects of school closure, peer effects resulting from changes in school composition, and general equilibrium effects that might prevent replication of successful schools. While these assumptions are strong, the results provide a rough guide to the likely effects of VAM-based policy decisions.

Table 7 predicts impacts of closing the lowest-ranked district school for a given value-added model and sending the displaced students to another school selected by the same model. Consistent with the high misclassification rates in Table 6, policies based on uncontrolled test score levels generate negligible benefits for students. Replacing the lowest-scoring BPS school with an average school is predicted to increase scores for affected students by $0.03\sigma$. Likewise, a policy that replaces the lowest-ranked school with an expansion of the highest-ranked school generates a gain of only $0.06\sigma$. These small effects are due to the large variation in bias evident for the uncontrolled model in Table 5; closure decisions based on uncontrolled estimates are likely to select schools with very negative selection bias rather than low value-added. Adjusting for bias via hybrid estimation increases these gains to $0.15\sigma$ and $0.28\sigma$, a substantial improvement.

Conventional VAM specifications that adjust for lagged scores are less severely biased than uncontrolled models, and causal value-added is low for schools ranked at the bottom by these specifications. As a result, closure and expansion decisions based on the lagged score and gains models are predicted to yield substantial achievement gains for students. Replacing the lowest-ranked school with an average school boosts scores by roughly $0.2\sigma$ in the gains specification. This is 54 percent of the corresponding benefit for an infeasible policy that ranks schools by true value-added ($0.37\sigma$). Hybrid estimation increases the effect of closure to $0.3\sigma$, thereby eliminating more than half the gap between the conventional model and the maximum possible gain. As in Table 6, use of the lagged scores specification results in slightly less effective policies than the gains model, but effects for these two models are typically similar in magnitude.

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, replacing the lowest-ranked school with a typical top-quintile school generates an improvement of $0.34\sigma$ when conventional posteriors are used and $0.48\sigma$ when rankings are based on hybrid predictions. Expanding the highest-ranked school produces corresponding impacts of $0.39\sigma$ and $0.57\sigma$. The latter effect is 80 percent of the benefit that would result from replacing the least effective school in the Boston district with the most effective ($0.57/0.71$).

Finally, column (5) of Table 7 simulates the effects of replacing low-ranked schools with an average charter school, a policy similar to Boston’s in-district charter conversions (Abdulkadiroğlu et al., 2015). As a result of the substantial difference in mean value-added between charter and district schools, charter...
conversion is predicted to generate large gains for all value-added specifications and estimation methods. Accurate value-added estimation improves the efficacy of charter conversion, however: selecting a school for conversion based on a hybrid gains specification rather than a naive uncontrolled model boosts the effect of charter expansion from $0.41\sigma$ to $0.62\sigma$.

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 on conventional estimates of school value-added. An application of our approach to data from Boston suggests that conventional OLS value-added estimates for Boston’s traditional public schools are systematically biased. Estimates of the joint distribution of value-added and bias imply highly misleading level comparisons of test scores across schools. Controls for lagged test scores reduce but do not eliminate bias. Nevertheless, policy simulations show accountability decisions based on conventional VAMs are likely to boost achievement. A hybrid estimation procedure that combines conventional and lottery-based estimates to reduce bias leads to more accurate value-added predictions, fewer policy errors, and even larger gains.

A necessary ingredient for our approach is some kind of lottery-based admissions scheme, such as are increasingly found in America’s large urban districts. As our analysis of charter schools shows, however, admissions need not be centralized for lotteries to be useful. Equally important, our analysis shows that the strategies outlined here remain useful even for districts (like Boston) where lottery data are missing or irrelevant for a large minority of schools.

Application of our framework to the study of teacher value-added is a natural direction for further research. Lotteries for teacher assignment are rare, but the methods outlined here may be extended to exploit other sources of quasi-experimental variation. A complication in the teacher context is the more elaborate hierarchical data structure arising from the fact that teacher assignment has both within and between-school components. Our results are silent on the nature of sorting into classrooms within schools. On the other hand, evidence that school-to-school comparisons are biased would seem to raise concerns regarding the causal interpretation of conventional teacher VAM estimates, which typically use both sources of variation (as in Chetty et al., 2014a). Finally, the framework outlined here seems likely to be useful for testing and improving VAM estimates in settings outside schools. Candidates for this extension include the quantification of doctor, hospital, and neighborhood effects (Finkelstein et al., 2013; Chetty and Hendren, 2015).
Notes: This figure compares standard deviations of school effects from four math value-added models; see Table 3's notes for a description of the models. For each model, the total variance of school effects 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 the 28 school 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 IV value-added coefficients from Table 3, while dashed lines indicate the 45-degree line.

Uncontrolled

Demographics

VA coeff.: 0.396  
Joint p-value: <0.001

VA coeff.: 0.645  
Joint p-value: <0.001

Lagged scores

Gains

VA coeff.: 0.864  
Joint p-value: <0.001

VA coeff.: 0.950  
Joint p-value: <0.001

Notes: This figure plots lottery reduced form effects against value-added first stages from each of the 28 school 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 IV value-added coefficients from Table 3, while dashed lines indicate the 45-degree line.
Figure 3: 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.
Figure 4: Relationship between bias and oversubscription among lottery schools

Notes: This figure plots posterior mode predictions of bias against oversubscription rates for schools with lotteries. The oversubscription rate is defined as 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. Bias modes come from the lagged scores model with sector effects. Points in the figure are constructed by first regressing bias modes and oversubscription rates on pilot and charter indicators, then computing residuals from these regressions.
Notes: This figure plots root mean squared error for posterior predictions of school value-added. Conventional predictions are posterior means constructed from OLS value-added estimates. Hybrid predictions are posterior modes constructed from OLS and lottery estimates. Root mean squared error is calculated from 100 simulated samples drawn from the data generating processes implied by the estimates in tables 5 and 6. The random coefficients model is re-estimated in each simulated sample.
Table 1: Boston students and schools

<table>
<thead>
<tr>
<th>School</th>
<th>Enrollment</th>
<th>OLS sample</th>
<th>Lottery sample</th>
<th>Lottery school?</th>
<th>School</th>
<th>OLS sample</th>
<th>Lottery sample</th>
<th>Lottery school?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td></td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>A. Traditional publics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B. Pilots</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1,095</td>
<td>79</td>
<td>Y</td>
<td></td>
<td>1</td>
<td>538</td>
<td>310</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>1,025</td>
<td>445</td>
<td>Y</td>
<td></td>
<td>2</td>
<td>1,260</td>
<td>433</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>1,713</td>
<td>1,084</td>
<td>Y</td>
<td></td>
<td>3</td>
<td>585</td>
<td>296</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>547</td>
<td>218</td>
<td>Y</td>
<td></td>
<td>4</td>
<td>78</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>217</td>
<td>46</td>
<td>Y</td>
<td></td>
<td>5</td>
<td>453</td>
<td>46</td>
<td>Y</td>
</tr>
<tr>
<td>6</td>
<td>1,354</td>
<td>581</td>
<td>Y</td>
<td></td>
<td>6</td>
<td>380</td>
<td>67</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>263</td>
<td>44</td>
<td>Y</td>
<td></td>
<td>7</td>
<td>242</td>
<td>179</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>1,637</td>
<td>492</td>
<td>Y</td>
<td></td>
<td>8</td>
<td>558</td>
<td>73</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>472</td>
<td>104</td>
<td>Y</td>
<td></td>
<td>9</td>
<td>18</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1,238</td>
<td>591</td>
<td>Y</td>
<td></td>
<td>10</td>
<td>738</td>
<td>406</td>
<td>Y</td>
</tr>
<tr>
<td>11</td>
<td>537</td>
<td>11</td>
<td></td>
<td></td>
<td>11</td>
<td>361</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>331</td>
<td>35</td>
<td>Y</td>
<td></td>
<td>12</td>
<td>357</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>335</td>
<td>82</td>
<td>Y</td>
<td></td>
<td>13</td>
<td>393</td>
<td>332</td>
<td>Y</td>
</tr>
<tr>
<td>14</td>
<td>952</td>
<td>232</td>
<td>Y</td>
<td></td>
<td>14</td>
<td>338</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>294</td>
<td>71</td>
<td>Y</td>
<td></td>
<td>15</td>
<td>511</td>
<td>115</td>
<td>Y</td>
</tr>
<tr>
<td>16</td>
<td>333</td>
<td>90</td>
<td></td>
<td></td>
<td>16</td>
<td>71</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>766</td>
<td>243</td>
<td>Y</td>
<td></td>
<td>17</td>
<td>300</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>372</td>
<td>47</td>
<td>Y</td>
<td></td>
<td>18</td>
<td>53</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>137</td>
<td>14</td>
<td></td>
<td></td>
<td>19</td>
<td>415</td>
<td>305</td>
<td>Y</td>
</tr>
<tr>
<td>20</td>
<td>1,091</td>
<td>225</td>
<td>Y</td>
<td></td>
<td>20</td>
<td>70</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>1,086</td>
<td>127</td>
<td>Y</td>
<td></td>
<td>21</td>
<td>104</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>577</td>
<td>104</td>
<td>Y</td>
<td></td>
<td>22</td>
<td>701</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>622</td>
<td>61</td>
<td></td>
<td></td>
<td>23</td>
<td>85</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>906</td>
<td>270</td>
<td>Y</td>
<td></td>
<td>24</td>
<td>85</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>25 (Ref.)</td>
<td>267</td>
<td>19</td>
<td></td>
<td></td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All schools:</td>
<td>27,864</td>
<td>8,718</td>
<td>28</td>
<td></td>
<td></td>
<td>17</td>
<td>85</td>
<td>37</td>
</tr>
</tbody>
</table>

Notes: This table counts the students and schools included in the observational (OLS) and lottery samples. The sample covers cohorts attending 6th grade in Boston between the 2006-2007 and 2013-2014 school years. Traditional public school #25 is the designated omitted enrollment category for value-added estimation. Columns (4) and (8) indicate whether the school has enough students subject to conditionally-random offer variation to be included in the lottery sample.
<table>
<thead>
<tr>
<th>Baseline covariate</th>
<th>OLS sample (1)</th>
<th>Lottery sample (2)</th>
<th>All lotteries (3)</th>
<th>Traditional (4)</th>
<th>Pilot (5)</th>
<th>Charter (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>0.345</td>
<td>0.354</td>
<td>-0.017</td>
<td>-0.007</td>
<td>0.003</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.017)</td>
<td>(0.033)</td>
<td>(0.018)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>0.410</td>
<td>0.485</td>
<td>-0.011</td>
<td>-0.005</td>
<td>-0.052</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.018)</td>
<td>(0.034)</td>
<td>(0.020)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>0.122</td>
<td>0.072</td>
<td>0.010</td>
<td>0.006</td>
<td>0.005</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.008)</td>
<td>(0.015)</td>
<td>(0.010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.490</td>
<td>0.504</td>
<td>0.017</td>
<td>0.034*</td>
<td>-0.013</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.019)</td>
<td>(0.037)</td>
<td>(0.020)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subsidized lunch</td>
<td>0.806</td>
<td>0.830</td>
<td>0.020*</td>
<td>0.020</td>
<td>0.006</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.013)</td>
<td>(0.026)</td>
<td>(0.016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special education</td>
<td>0.208</td>
<td>0.195</td>
<td>0.006</td>
<td>-0.003</td>
<td>-0.022</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.013)</td>
<td>(0.030)</td>
<td>(0.016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English-language learner</td>
<td>0.205</td>
<td>0.206</td>
<td>0.006</td>
<td>-0.001</td>
<td>0.018</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.014)</td>
<td>(0.027)</td>
<td>(0.016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspensions</td>
<td>0.093</td>
<td>0.076</td>
<td>-0.025</td>
<td>-0.025</td>
<td>0.009</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.023)</td>
<td>(0.025)</td>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absences</td>
<td>1.710</td>
<td>1.534</td>
<td>-0.087</td>
<td>-0.138*</td>
<td>-0.092</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>(0.095)</td>
<td>(0.080)</td>
<td>(0.260)</td>
<td>(0.167)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math score</td>
<td>0.058</td>
<td>0.004</td>
<td>0.022</td>
<td>-0.026</td>
<td>0.080</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.030)</td>
<td>(0.061)</td>
<td>(0.035)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA score</td>
<td>0.030</td>
<td>0.013</td>
<td>0.035</td>
<td>0.045</td>
<td>0.060</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.030)</td>
<td>(0.061)</td>
<td>(0.036)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N 27,864 8,718 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 students 6th graders enrolled between the 2006-2007 and 2013-2014 school years, and column (2) shows mean characteristics for randomized lottery applicants. Columns (3)-(6) 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>
<thead>
<tr>
<th></th>
<th>Uncontrolled</th>
<th>Demographics</th>
<th>Lagged scores</th>
<th>Gains</th>
<th>Lagged scores, no charter lotteries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value-added coefficient</td>
<td>0.396</td>
<td>0.645</td>
<td>0.864</td>
<td>0.950</td>
<td>0.549</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.065)</td>
<td>(0.075)</td>
<td>(0.084)</td>
<td>(0.164)</td>
</tr>
<tr>
<td>First stage F-statistic</td>
<td>45.6</td>
<td>36.1</td>
<td>29.6</td>
<td>26.6</td>
<td>11.2</td>
</tr>
</tbody>
</table>

*p-values:*

<table>
<thead>
<tr>
<th></th>
<th>Uncontrolled</th>
<th>Demographics</th>
<th>Lagged scores</th>
<th>Gains</th>
<th>Lagged scores, no charter lotteries</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA coef. equals 1</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.071</td>
<td>0.554</td>
<td>0.006</td>
</tr>
<tr>
<td>Overid. restrictions</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.003</td>
<td>0.006</td>
<td>0.043</td>
</tr>
<tr>
<td>All restrictions</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>All restrictions (bootstrap refinement)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Notes: This table reports lottery-based tests for bias in school value-added models for math scores. Value-added coefficients come from IV regressions of 6th grade scores on fitted values from value-added models, instrumented by the set of offer dummies for all school lotteries. Models are estimated via a two-step optimal GMM procedure that is efficient with arbitrary heteroskedasticity. Joint *p*-values come from OLS regressions of value-added residuals on offer dummies. The uncontrolled model includes only year-of-test indicators as controls. The demographics model adds indicators for student sex, race, subsidized lunch, special education, limited-English proficiency, and counts of baseline absences and suspensions. The lagged scores model adds cubic polynomials in baseline math and ELA scores. The gains model includes the same controls as the demographics model and uses score gains from baseline as the outcome. Column (5) excludes charter school lotteries from the lottery sample in testing the lagged scores model. All IV models control for lottery strata fixed effects, demographics, and lagged scores. Inference is robust to heteroskedasticity and accounts for first-step estimation error in school value-added. Bootstrap refinements to first-order asymptotics are based on 500 Bayesian bootstrap replications (see Appendix B).
### Table 4: Robustness of bias tests to effect heterogeneity

<table>
<thead>
<tr>
<th>Baseline VAM specification (1)</th>
<th>Baseline year (2)</th>
<th>Subsidized lunch (3)</th>
<th>Special education (4)</th>
<th>Baseline score tercile (5)</th>
<th>Interacted groups (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value-added coefficient</td>
<td>0.864</td>
<td>0.916</td>
<td>0.849</td>
<td>0.863</td>
<td>0.866</td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td>(0.072)</td>
<td>(0.075)</td>
<td>(0.074)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>Bootstrap-refined VAM validity p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Value-added coefficient</td>
<td>0.868</td>
<td>0.962</td>
<td>0.851</td>
<td>0.872</td>
<td>0.873</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
<td>(0.068)</td>
<td>(0.069)</td>
<td>(0.070)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>Bootstrap-refined VAM validity p-value</td>
<td>&lt;0.001</td>
<td>0.002</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Notes: This table reports lottery-based tests for bias in school value-added models that allow for treatment effect heterogeneity by baseline characteristics. Value-added coefficients come from IV regressions of 6th grade math scores on fitted values from value-added models, instrumented by the set of offer dummies for all school lotteries. Models are estimated via a two-step optimal GMM procedure that is efficient with arbitrary heteroskedasticity. Joint p-values come from OLS regressions of value-added residuals on offer dummies. The OLS value-added specification includes demographics and lagged scores. Panel A estimates value-added in the full observational sample, while Panel B restricts estimation to the lottery subsample. Column (1) repeats estimates from Table 3, while columns (2)-(6) allow value-added to differ across cells defined by the covariates in the column headings. The covariates used to define subgroups in column (6) are Hispanic, black, and female indicators, dummies for subsidized lunch, special education, and English language learner status, and indicators for baseline score terciles, based on average 5th grade math and ELA test scores in the observational sample. All regressions control for lottery strata fixed effects, demographics, and lagged scores. Inference is robust to heteroskedasticity and accounts for first-step VAM estimation error. Bootstrap refinements to first-order asymptotics are based on 500 Bayesian bootstrap replications (see Appendix B).
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Models without sector effects</th>
<th>Models with sector effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$\sigma_Y$</td>
<td>Std. dev. of causal value-added</td>
<td>0.210</td>
<td>0.212</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.063)</td>
<td>(0.062)</td>
</tr>
<tr>
<td>$\sigma_b$</td>
<td>Std. dev. of bias in OLS value-added</td>
<td>0.487</td>
<td>0.314</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.068)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Correlation of value-added and bias</td>
<td>-0.132</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.250)</td>
<td>(0.329)</td>
</tr>
<tr>
<td>VA shifters</td>
<td>Lottery school</td>
<td>-0.055</td>
<td>-0.103</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.141)</td>
<td>(0.110)</td>
</tr>
<tr>
<td>Charter</td>
<td></td>
<td>0.353</td>
<td>0.358</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.118)</td>
<td>(0.124)</td>
</tr>
<tr>
<td>Pilot</td>
<td></td>
<td>0.057</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.137)</td>
<td>(0.143)</td>
</tr>
<tr>
<td>Bias shifters</td>
<td>Charter</td>
<td>0.052</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.115)</td>
<td>(0.119)</td>
</tr>
<tr>
<td>Pilot</td>
<td></td>
<td>-0.060</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.134)</td>
<td>(0.138)</td>
</tr>
<tr>
<td>J-statistic (d.f.):</td>
<td>5.64 (4)</td>
<td>7.41 (4)</td>
<td>3.05 (4)</td>
</tr>
<tr>
<td>Overid. p-value:</td>
<td>0.228</td>
<td>0.116</td>
<td>0.549</td>
</tr>
</tbody>
</table>

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 the observed OLS, reduced form, and first stage estimates, as described in the Appendix. Simulated moments are computed from 500 samples constructed by drawing estimation errors from the asymptotic covariance matrix of the observed estimates, along with school-specific parameters drawn from the random coefficient distribution. 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 10,000 simulations, drawn independently from the samples used to compute the estimator. See notes to Table 3 for a description of the control variables included in each OLS value-added model.
<table>
<thead>
<tr>
<th>Value-added model</th>
<th>Posterior method</th>
<th>Low-performing schools</th>
<th>High-performing schools</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Lowest decile (1)</td>
<td>Lowest quintile (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Highest decile (4)</td>
<td>Highest quintile (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lowest tercile (3)</td>
<td>Highest tercile (6)</td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td>0.900</td>
<td>0.800</td>
</tr>
<tr>
<td>Uncontrolled</td>
<td>Conventional</td>
<td>0.838</td>
<td>0.680</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.646</td>
<td>0.510</td>
</tr>
<tr>
<td>Demographics</td>
<td>Conventional</td>
<td>0.729</td>
<td>0.603</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.594</td>
<td>0.455</td>
</tr>
<tr>
<td>Lagged scores</td>
<td>Conventional</td>
<td>0.622</td>
<td>0.491</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.544</td>
<td>0.438</td>
</tr>
<tr>
<td>Gains</td>
<td>Conventional</td>
<td>0.646</td>
<td>0.468</td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.546</td>
<td>0.394</td>
</tr>
</tbody>
</table>

Notes: This table reports 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 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. 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 100 simulated samples, and the random coefficients model is re-estimated in each sample.
Table 7: Consequences of closing the lowest-ranked district school for affected children

<table>
<thead>
<tr>
<th>Value-added model</th>
<th>Posterior method</th>
<th>Replacement school:</th>
<th>Average school</th>
<th>Average above-median school</th>
<th>Average top-quintile school</th>
<th>Highest-ranked school</th>
<th>Average charter school</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>True value-added</td>
<td></td>
<td></td>
<td>0.368</td>
<td>0.504</td>
<td>0.588</td>
<td>0.705</td>
<td>0.685</td>
</tr>
<tr>
<td>Uncontrolled</td>
<td>Conventional</td>
<td>0.030</td>
<td>0.098</td>
<td>0.164</td>
<td>0.223</td>
<td>0.409</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.146</td>
<td>0.219</td>
<td>0.294</td>
<td>0.363</td>
<td>0.525</td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td>Conventional</td>
<td>0.130</td>
<td>0.195</td>
<td>0.234</td>
<td>0.282</td>
<td>0.455</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.234</td>
<td>0.329</td>
<td>0.403</td>
<td>0.507</td>
<td>0.559</td>
<td></td>
</tr>
<tr>
<td>Lagged scores</td>
<td>Conventional</td>
<td>0.166</td>
<td>0.234</td>
<td>0.282</td>
<td>0.316</td>
<td>0.475</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.260</td>
<td>0.353</td>
<td>0.424</td>
<td>0.532</td>
<td>0.569</td>
<td></td>
</tr>
<tr>
<td>Gains</td>
<td>Conventional</td>
<td>0.198</td>
<td>0.283</td>
<td>0.343</td>
<td>0.386</td>
<td>0.515</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hybrid</td>
<td>0.303</td>
<td>0.403</td>
<td>0.479</td>
<td>0.571</td>
<td>0.620</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table reports simulated test score impacts of closing the lowest-ranked district school based on value-added predictions. The reported impacts are effects on test scores for students at the closed school. Column (1) replaces the lowest-ranked district school with an average district school. Columns (2), (3) and (4) replace the lowest-ranked school with an average above-median district school, an average top-quintile district school, or the highest-ranked district school. Column (5) 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 100 simulated samples, and the random coefficients model is re-estimated in each sample.
References


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 6th grade students enrolled from 2006-2007 to 2013-2014, using their 5th 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 6th 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 (5th grade) math and ELA test scores for each student in our sample (5th 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 6th 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 Charter School of 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 tiebreaking sequence numbers, and assignments from the BPS deferred-acceptance mechanism, 2006-2007 though 2013-2014. 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 being currently enrolled in the school, as well as certain other students with guaranteed or nonstandard priorities (see Abdulkadiroğlu et al. (2006) for a complete description of priorities in BPS). Within groups we construct indicators for whether an applying student has a random sequence number that is better than the worst number belonging to a student in the group that is assigned to each school. A student qualified in such a way is assigned to it by the mechanism, and 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 6th grade entry.

A.4 Sample Selection

We restrict attention to Boston public schools with at least 25 6th 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 5th and 6th grade, if
their baseline demographic, attendance, and test score information is available, and if we observe their 6th 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 6th grade admission at 28 schools, either from a charter school lottery or from assignment by the BPS mechanism.
Appendix B: Econometric Methods

B.1 VAM Bias Tests

We test for bias in conventional VAMs by regressing VAM residuals on a vector of lottery offers $Z_i$ and lottery strata controls $C_i$ as in equation (9). In practice this regression uses sample residuals $\hat{v}_i$, which are noisy estimates of the population residuals $v_i$ that would be observable if the coefficients in the OLS value-added model (5) were known rather than estimated. We therefore adjust inference to account for the resulting estimation error.

Let $X_i = (1, D_{i1}, \ldots, D_{iJ}, X_i')'$ and $Z_i = (1, C_i, Z_i')'$. OLS estimates of equation (9) may be written in matrix form as

$$\hat{\phi} = (Z'Z)^{-1}Z'\hat{v}.$$ 

Moreover,

$$\hat{\phi} - \phi = (Z'Z)^{-1}Z'(\omega + (\hat{v} - v)),$$ 

while by equation (5),

$$\hat{v} - v = Y - X(X'X)^{-1}X'(X\beta + v) - v$$

$$= -X(X'X)^{-1}X'v.$$ 

We then have that:

$$\sqrt{N}(\hat{\phi} - \phi) = \sqrt{N}(Z'Z)^{-1}\left(Z'\omega - Z'X(X'X)^{-1}X'v\right)$$

$$= \left(\frac{1}{N} \sum_{i=1}^N Z_i Z_i'\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\frac{1}{N} \sum_{i=1}^N Z_i X_i' \left(\frac{1}{N} \sum_{i=1}^N X_i X_i'\right)^{-1} X_i v_i\right).$$ 

Under iid sampling the weak law of large numbers ensures $\frac{1}{N} \sum_{i=1}^N Z_i Z_i' \xrightarrow{p} E[Z_i Z_i']$, $\frac{1}{N} \sum_{i=1}^N Z_i X_i' \xrightarrow{p} E[Z_i X_i']$, and $\frac{1}{N} \sum_{i=1}^N X_i X_i' \xrightarrow{p} E[X_i X_i']$, provided such moments exist. Furthermore the Lindeberg–Lévy central limit theorem implies:

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \left(Z_i \omega_i - E[Z_i X_i'] (E[X_i X_i'])^{-1} X_i v_i\right) \xrightarrow{d} N(0, \Lambda),$$

provided

$$\Lambda = E\left[\left(Z_i \omega_i - E[Z_i X_i'] (E[X_i X_i'])^{-1} X_i v_i\right) \left(Z_i \omega_i - E[Z_i X_i'] (E[X_i X_i'])^{-1} X_i v_i\right)^\prime\right].$$
is finite. By Slutsky’s theorem,

$$\sqrt{N}(\hat{\phi} - \phi) \Rightarrow N(0, \Xi)$$

for

$$\Xi = (E[Z_i Z_i'])^{-1} \Lambda (E[Z_i Z_i'])^{-1}.$$  

We base asymptotic inference on $\phi$ by forming sample analogues of each component of $\Xi$. Note that this estimate differs from the usual White (1980) heteroskedasticity-consistent covariance matrix estimator by the second term in the inner product of $\Lambda$. This term accounts for estimation error in the first-step residuals.

The Wald test statistic for the null hypothesis of VAM validity, $\phi_z = 0$, is then

$$F = \hat{\phi}_z' \left( \hat{\Xi}_z / N \right)^{-1} \hat{\phi}_z,$$

where $\hat{\Xi}_z$ is the submatrix of our estimate of $\Xi$ corresponding to $\hat{\phi}_z$. The distribution of $F$ is first-order equivalent to $\chi^2_L$. Analogous steps are used to derive asymptotic variances for instrumental variables estimators based on equations (10) and (11).

Bootstrapping asymptotically-pivotal test statistics often yields critical values that are more accurate than those derived from first-order asymptotics (Hall and Horowitz, 1996). We report bootstrap-refined $p$-values for tests based on equation (9). Our implementation uses the Bayesian bootstrap procedure of Rubin (1981), which smooths out bootstrap samples by reweighting rather than resampling observations. This prevents the omission of small lottery strata in our data that would occasionally be dropped in standard nonparametric bootstrap resampling. The Bayesian and nonparametric bootstraps are special cases of the generalized bootstrap and both are consistent under weak conditions (Mason and Newton, 1992).

To implement the Bayesian bootstrap we draw random vectors of Dirichlet(1, ...., 1) weights, then reestimate the value-added model (5) and residual regression (9) by weighted least squares. We then use the results to construct a set of re-centered test statistics,

$$F^b = \left( \hat{\phi}_z^b - \hat{\phi}_z \right)' \left( \hat{\Xi}_z^b / N \right)^{-1} \left( \hat{\phi}_z^b - \hat{\phi}_z \right),$$

where the variance matrix $\hat{\Xi}_z^b$ is estimated in each bootstrap trial as described above, weighting all sample moments with the Dirichlet weights. The resulting bootstrap-refined $p$-value is

$$p = \frac{1}{B} \sum_{b=1}^B 1[F^b > F].$$
B.2 Heterogeneous School Effects

We next generalize our lottery-based tests for bias to a model that allows school effects to vary across students. For a set of mutually-exclusive and exhaustive “types” \( t \), we write potential achievement as

\[
Y_{ij} = \sum_t \mu_{jt} T_{it} + a_i + m_{ij},
\]

where \( T_{it} \) is an indicator for belonging to type \( t \), \( \mu_{jt} \) is average potential achievement at school \( j \) for individuals of type \( t \), and \( a_i \) is student ability, which is orthogonal to student type by definition. The decomposition in equation (22) also allows for an unrestricted match component in achievement, \( m_{ij} \), and is therefore fully general.

Equation (22) implies that the observed outcome for student \( i \) can be written

\[
Y_i = \sum_t \mu_{0t} T_{it} + \sum_j \sum_t \beta_{jt} D_{ij} T_{it} + X'_i \gamma + \epsilon_i,
\]

where \( \beta_{jt} = \mu_{jt} - \mu_{j0} \) is the value-added of school \( j \) for type \( t \) and the error term is

\[
\epsilon_i = \sum_j \sum_t (m_{ij} - m_{i0}) D_{ij} T_{iw} + \epsilon_{i0},
\]

with \( \epsilon_{i0} \) the residual from a projection of \( a_i + m_{i0} \) on \( X_i \).

The following assumptions extends selection-on-observables to the case with effect heterogeneity:

\[
E[\epsilon_{i0} | D_i] = 0,
\]

where \( T_i \) is the vector of all type dummies. Assumption (24) requires general student ability to be unrelated to school choices after controlling for \( X_i \), similar to (6). The additional assumption (25) requires idiosyncratic match effects to also be independent of school choices conditional on \( T_i \). In other words, any relationship between school effects and school choices occurs through the type-specific effects \( \beta_{jt} \), not through sorting on gains within type. Assumption (25) relates to the “conditional effect ignorability” assumption described by Angrist and Fernandez-Val (2013).

The OLS regression corresponding to the causal model (23) is

\[
Y_i = \sum_t \alpha_{0t} T_{it} + \sum_j \sum_t \alpha_{jt} D_{ij} T_{it} + X'_i \Gamma + v_i,
\]

with \( v_i \) orthogonal to \( T_{it}, D_{ij} \) and \( X_i \) by definition. Together with the maintained exclusion restriction (7), assumptions (24) and (25) imply that residuals from model (26) should be orthogonal to lottery offers after controlling for randomization strata. If rejection of the more restrictive model (4) is caused by heterogeneity in school effects across student types rather than bias, tests based on (26) will not reject. In section 4.4 we conduct tests with student types defined in a variety of ways.
B.3 Simulated Minimum Distance

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

$$\theta = (\alpha_0, \beta_0, \beta_Q, \delta_0, \xi_0, \Sigma, \sigma^2_0)'$$.

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

$$\hat{\Omega} = (\hat{\alpha}_1, ..., \hat{\alpha}_J, \hat{\rho}_1^z, ..., \hat{\rho}_L^z, \hat{\pi}_1^1, ..., \hat{\pi}_L^1, ..., \hat{\pi}_L^J)'$$.

Let $$\Omega = (\alpha_1, ..., \pi_L^J)'$$ 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 (14) through (17), values of $$\Omega$$ and $$\hat{\Omega}$$ can be simulated for any value of $$\theta$$. We use this procedure to generate simulated data sets, and estimate $$\theta$$ by minimizing the distance between simulated and observed moments.

Our estimation procedure targets the following first moments:

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

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

$$\hat{m}_3 = \frac{1}{L} \sum_\ell \hat{\beta}_z^\ell$$,

$$\hat{m}_4 = \frac{1}{L} \sum_\ell \hat{\pi}_\ell$$,

$$\hat{m}_5 = -\frac{1}{L} \sum_\ell \sum_{j \neq \ell} \hat{\pi}_j^f$$,

$$\hat{m}_6 = -\frac{1}{L(L-1)} \sum_\ell \sum_{j \neq \ell} \hat{\pi}_j^f$$,

$$\hat{m}_7 = \frac{1}{L} \sum_\ell \left[ \frac{(\hat{\pi}_\ell^f)^2}{\sum_k (\hat{\pi}_k^f)^2} \right] \cdot \left( \frac{\hat{\beta}_z^\ell}{\hat{\Sigma}^\ell} \right)$$.

$$\hat{m}_1$$ 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 $$\beta_Q$$, the difference in value-added between lottery and non-lottery schools. $$\hat{m}_3$$ is the mean reduced form, which provides information about $$\beta_0$$. $$\hat{m}_4$$ is the mean first stage across lotteries, which can be used to estimate $$\delta_0$$. $$\hat{m}_5$$ is the average sum of fallback probabilities for included schools across lotteries, and $$\hat{m}_6$$ is the average
ratio of this sum to the first stage, which gives the share of compliers drawn from included schools. These
two moments help to estimate $\xi_0$, the mean fallback utility for included schools relative to the omitted school.
$m_7$ is the average ratio of the lottery reduced form to a “pseudo-reduced form” prediction that uses OLS
value-added estimates, given by $\hat{\lambda}_z^f = \sum_j \hat{\pi}_j^f \hat{\alpha}_j$. 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 $\beta_j$ and $b_j$.

The next seven moments are variances of parameter estimates:

$$m_8 = \frac{1}{L} \sum_j (\hat{\alpha}_j - \bar{\alpha})^2,$$

$$m_9 = \frac{1}{L} \sum\ell (\hat{\rho}_z^\ell - \bar{\rho})^2,$$

$$m_{10} = \frac{1}{L} \sum\ell (\hat{\lambda}_z^\ell - \bar{\lambda})^2,$$

$$m_{11} = \frac{1}{L} \sum\ell (\hat{\pi}_z^\ell - \bar{\pi}_{own})^2,$$

$$m_{12} = \frac{1}{L} \sum_j \left( \frac{1}{L} \sum\ell \hat{\pi}_j^\ell - \bar{\pi}_{other} \right)^2,$$

$$m_{13} = \frac{1}{L} \sum_j \left( \frac{1}{L} \sum\ell \hat{\pi}_j^\ell - \bar{\pi}_{other} \right)^2,$$

$$m_{14} = \frac{1}{L(L-1)} \sum_j \sum\ell \hat{\pi}_j^\ell (\hat{\pi}_j^\ell - \bar{\pi}_{other})^2.$$

Here $\bar{\alpha}$ indicates the sample average of the $\alpha_j$, and similarly for other variables. $m_8$ 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. $m_9$ and $m_{10}$ are variances of the lottery reduced form and predicted reduced
form, which contain additional information about the joint distribution of value-added and bias. $m_{11}$ is
the variance of the first stage across lotteries, which helps to identify the variance of $\delta_j$. $m_{12}$ 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. $m_{13}$ 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 $\xi_j$ and $\nu_j^f$, which govern heterogeneity in fallback probabilities. $m_{14}$
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 $\xi_j$, the school-specific mean fallback utility, from $\sigma_n^2$, the variance of idiosyncratic school-by-lottery utility
shocks.

Finally, we match six covariances:

$$m_{15} = \frac{1}{L} \sum\ell (\hat{\rho}_z^\ell - \bar{\rho}) (\hat{\lambda}_z^\ell - \bar{\lambda}),$$

$$m_{16} = \frac{1}{L} \sum\ell (\hat{\rho}_z^\ell - \bar{\rho}) (\hat{\pi}_z^\ell - \bar{\pi}_{own}).$$

47
\[ \hat{m}_{17} = \frac{1}{L} \sum_{t} (\hat{\alpha}_t - \bar{\alpha}) (\hat{x}_{t} - \bar{\pi}_{own}), \]
\[ \hat{m}_{18} = \frac{1}{L} \sum_{t} (\hat{\beta}_{zt} - \bar{\beta}) \left( \left( \frac{1}{L-1} \sum_{k \neq t} \hat{\pi}^k_{t} \right) - \bar{\pi}_{other} \right), \]
\[ \hat{m}_{19} = \frac{1}{L} \sum_{t} (\hat{\alpha}_t - \bar{\alpha}) \left( \left( \frac{1}{L-1} \sum_{k \neq t} \hat{\pi}^k_{t} \right) - \bar{\pi}_{other} \right), \]
\[ \hat{m}_{20} = \frac{1}{L} \sum_{t} (\hat{\pi}_{own} - \bar{\pi}_{own}) \left( \left( \frac{1}{L-1} \sum_{k \neq t} \hat{\pi}^k_{t} \right) - \bar{\pi}_{other} \right). \]

\( \hat{m}_{15} \) is the covariance of the reduced form and pseudo-reduced form, which helps to identify variation in bias, as well as the covariance between bias and value-added. \( \hat{m}_{16} \) is the covariance between reduced forms and first stages, which is informative about the covariance between \( \beta_j \) and \( \pi^j_p \). \( \hat{m}_{17} \) is the covariance of conventional value-added and the first stage, which helps to identify the covariance between \( b_j \) and \( \delta_j \). \( \hat{m}_{18} \) is the covariance of the reduced form and average fallback probability, which helps to identify the covariance of \( \beta_j \) and \( \xi_j \). \( \hat{m}_{19} \) is the covariance of OLS value-added with the average fallback probability, which depends on the covariance between \( b_j \) and \( \xi_j \). \( \hat{m}_{20} \) is the covariance of a school’s first stage and average fallback probability, which provides information about the covariance of \( \xi_j \) and \( \delta_j \).

There are 16 elements of \( \theta \) and 20 moments, so the model has four overidentifying restrictions. Models that include charter and pilot school effects add sector-specific values of \( \hat{m}_1, \hat{m}_2, \hat{m}_3 \) and \( \hat{m}_4 \), yielding 20 parameters and 24 moments. Let \( \hat{m} = (\hat{m}_1, ..., \hat{m}_{24})' \) be the vector of observed moments, and let \( \hat{m}(\theta) \) be the corresponding vector of simulated predictions. The simulated minimum distance estimator with weighting matrix \( A \) is

\[ \hat{\theta}_{SMD}(A) = \arg \min_{\theta} J (\hat{m} - \hat{m}(\theta))' A (\hat{m} - \hat{m}(\theta)). \]

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

We produce a first-step estimate of \( \theta \) 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

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

where \( r \) indexes a second independent set of \( R = 10,000 \) simulation draws and \( \bar{m} \) is the mean of the simulated moments. An efficient two-step estimate is given by \( \hat{\theta}_{SMD} (\hat{A}) \).

Under the null hypothesis that the model’s overidentifying restrictions hold and standard regularity conditions, the minimized SMD criterion function follows a \( \chi^2 \) distribution (Sargan, 1958; Hansen, 1982):
\[ J \left( \hat{m} - \hat{m} \left( \hat{\theta}_{SMD}(\hat{A}) \right) \right)' \hat{A} \left( \hat{m} - \hat{m} \left( \hat{\theta}_{SMD}(\hat{A}) \right) \right) \sim \chi^2_q, \]

where \( q \) is the number of overidentifying restrictions. Table 5 reports the results of overidentification tests based on this \( J \)-statistic.

**B.4 Empirical Bayes Posteriors**

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, \( \Pi \), is known. In this case the posterior distribution for \( \beta_j \) and \( b_j \) can be derived analytically. In matrix form the model can be written

\[
\hat{\alpha} = \beta + b + e_{\alpha},
\]

\[
\hat{\rho}_z = \Pi\beta + e_{\rho},
\]

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

\[
(\beta', b')|\theta \sim N((\iota'\beta_0, \iota'b_0)', V_\theta),
\]

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

\[
 f_{\Theta|\hat{\Omega}} \left( \Theta|\hat{\Omega}; \theta \right) = \frac{f_{\hat{\Omega}|\Theta} \left( \hat{\Omega}|\Theta; \theta \right) f_{\Theta} (\Theta; \theta)}{f_{\hat{\Omega}} (\hat{\Omega}; \theta)}. \tag{27}
\]

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

\[
-2 \log f_{\Theta|\hat{\Omega}} \left( \Theta|\hat{\Omega}; \theta \right) = ((\hat{\alpha} - \beta - b)', (\hat{\rho}_z - \Pi\beta)')' \begin{bmatrix} v_{\alpha\alpha} & v_{\alpha\rho} \\ v'_{\alpha\rho} & v_{\rho\rho} \end{bmatrix} \begin{bmatrix} \hat{\alpha} - \beta - b \\ \hat{\rho}_z - \Pi\beta \end{bmatrix} + \begin{bmatrix} v_{\beta\beta} & v_{\beta b} \\ v'_{\beta b} & v_{bb} \end{bmatrix} \begin{bmatrix} \beta - \beta_0 \iota \\ b - b_0 \iota \end{bmatrix} + C_1,
\]

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 \( V_\theta^{-1} \); and \( C_1 \) is a constant that does not depend on \( \Theta \).

Rearranging this expression yields

\[
-2 \log f_{\Theta|\hat{\Omega}} \left( \Theta|\hat{\Omega}; \theta \right) = ((\beta - \beta^*)', (b - b^*)')' \begin{bmatrix} v_{\beta\beta} & v_{\beta b} \\ v'_{\beta b} & v_{bb} \end{bmatrix} \begin{bmatrix} \beta - \beta^* \\ b - b^* \end{bmatrix} + C_2, \tag{28}
\]

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

\[
v_{\beta\beta}^* = v_{\alpha\alpha} + \Pi'v_{\alpha\rho} + v_{\alpha\rho}\Pi + \Pi'v_{\rho\rho}\Pi + v_{\beta\beta},
\]

49
\[ v_{bb}^* = v_{aa} + v_{bb}, \]

and

\[
\beta^* = W_1(\hat{\alpha} - b_0\ell) + W_2\hat{\rho} + (I - W_1 - W_2\Pi)\beta_0\ell
\]

with

\[
W_1 = B^{-1}((v_{aa} + v_{bb})(v_{aa} + \Pi'v_{\alpha'\rho} + v_{\beta b})^{-1}(v_{aa} + \Pi'v_{\alpha'\rho}) - v_{aa}),
\]

\[
W_2 = B^{-1}((v_{aa} + v_{bb})(v_{aa} + \Pi'v_{\alpha'\rho} + v_{\beta b})^{-1}(v_{\alpha'\rho} + \Pi'v_{\rho\rho} - v_{\alpha'\rho}),
\]

\[
B = (v_{aa} + v_{bb})(v_{aa} + \Pi'v_{\alpha'\rho} + v_{\beta b})^{-1}(v_{aa} + \Pi'v_{\alpha'\rho} + v_{\alpha'\rho}\Pi + \Pi'v_{\rho\rho}\Pi + v_{\beta b}) - (v_{aa} + v_{\alpha'\rho}\Pi + v_{\beta b}).
\]

Equation (28) implies that the posterior for \((\beta, b)\) is normal:

\[(\beta', b')|\hat{\alpha}, \hat{\rho} \sim N((\beta^*|, b^*)', V^*),\]

with

\[
V^* = \begin{bmatrix} v_{bb}^* & v_{bb}^* \\ v_{bb}^* & v_{bb}^* \end{bmatrix}^{-1}.
\]

An empirical Bayes version of the posterior mean \(\beta^*\) is formed by plugging \(\hat{\theta}_{SMD}\) and an estimate of \(V_e\) into the expressions for \(W_1\) and \(W_2\).

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

\[
\Theta = (\beta_1, \beta_1, \delta_1, \xi_1, ..., \beta_J, \beta_J, \delta_J, \xi_J, v_1, ..., v_J)'.
\]

Up to a scaling constant, the posterior density for \(\Theta\) conditional on the observed estimates \(\hat{\Omega}\) and the prior parameters \(\theta\) can be expressed

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

where

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

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

\[
\Gamma(\theta) = \begin{bmatrix} I_J \otimes \Sigma & 0 \\ 0 & \sigma_v^2 I_J \end{bmatrix}.
\]
Note that the probability limit of the vector of observed estimates, $\Omega$, is a function of $\Theta$, so we write $\Omega(\Theta)$.

As before we form an empirical Bayes posterior density by plugging $\hat{\theta}_{SMD}$ into (29). The empirical Bayes posterior mean is

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

Since the first stage parameters $\pi^j_\ell$ are nonlinear functions of $\delta$ and $\xi$, the density in (29) 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:

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

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 $\beta_j$ with conventional empirical Bayes posterior means based on OLS estimates of value-added. The conventional predictions are given by

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

where

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

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

Models with sector effects replace $\hat{\mu}_\alpha$ in equation (30) with the regression predictions

$$
\hat{\mu}_{\alpha j} = P'_j \left[ \frac{1}{J} \sum_k P_k P_k' \right]^{-1} \left[ \frac{1}{J} \sum_k P_k \hat{\alpha}_k \right],
$$

where $P_j$ is a vector including a constant and charter and pilot school indicators.
Table A1: Tests for bias in ELA school value-added models

<table>
<thead>
<tr>
<th></th>
<th>Uncontrolled (1)</th>
<th>Demographics (2)</th>
<th>Lagged scores (3)</th>
<th>Gains (4)</th>
<th>Lagged scores, no charter lotteries (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value-added coefficient</td>
<td>0.358</td>
<td>0.660</td>
<td>0.864</td>
<td>0.722</td>
<td>0.423</td>
</tr>
<tr>
<td></td>
<td>(0.087)</td>
<td>(0.130)</td>
<td>(0.167)</td>
<td>(0.172)</td>
<td>(0.310)</td>
</tr>
<tr>
<td>First stage $F$-statistic</td>
<td>33.1</td>
<td>27.0</td>
<td>26.8</td>
<td>29.4</td>
<td>14.0</td>
</tr>
<tr>
<td>$p$-values:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VA coef. equals 1</td>
<td>$&lt;0.001$</td>
<td>0.009</td>
<td>0.416</td>
<td>0.105</td>
<td>0.063</td>
</tr>
<tr>
<td>Overid. restrictions</td>
<td>0.011</td>
<td>0.057</td>
<td>0.039</td>
<td>0.007</td>
<td>0.157</td>
</tr>
<tr>
<td>All restrictions</td>
<td>$&lt;0.001$</td>
<td>0.008</td>
<td>0.018</td>
<td>0.001</td>
<td>0.040</td>
</tr>
<tr>
<td>All restrictions (bootstrap refinement)</td>
<td>$&lt;0.001$</td>
<td>$&lt;0.001$</td>
<td>$&lt;0.001$</td>
<td>0.002</td>
<td>$&lt;0.001$</td>
</tr>
</tbody>
</table>

Notes: This table reports lottery-based tests for bias in school value-added models for ELA scores. Value-added coefficients come from IV regressions of 6th grade scores on fitted values from value-added models, instrumented by the set of offer dummies for all school lotteries. Models are estimated via a two-step optimal GMM procedure that is efficient with arbitrary heteroskedasticity. Joint $p$-values come from OLS regressions of value-added residuals on offer dummies. The uncontrolled model includes only year-of-test indicators as controls. The demographics model adds indicators for student sex, race, subsidized lunch, special education, limited-English proficiency, and counts of baseline absences and suspensions. The lagged scores model adds cubic polynomials in baseline math and ELA scores. The gains model includes the same controls as the demographics model and uses score gains from baseline as the outcome. Column (5) excludes charter school lotteries from the lottery sample in testing the lagged scores model. All IV models control for lottery strata fixed effects, demographics, and lagged scores. Inference is robust to heteroskedasticity and accounts for first-step estimation error in school value-added. Bootstrap refinements to first-order asymptotics are based on 500 Bayesian bootstrap replications (see Appendix B).
<table>
<thead>
<tr>
<th></th>
<th>$\beta_j$ (1)</th>
<th>$b_j$ (2)</th>
<th>$\delta_j$ (3)</th>
<th>$\xi_j$ (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>0.158 (0.070)</td>
<td>0.140 (0.068)</td>
<td>0.954 (0.112)</td>
<td>0.951 (0.227)</td>
</tr>
<tr>
<td>Correlation w/$\beta_j$</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Correlation w/$b_j$</td>
<td>-0.480 (0.427)</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Correlation w/$\delta_j$</td>
<td>0.119 (0.391)</td>
<td>-0.522 (0.544)</td>
<td>-0.651 (0.578)</td>
<td>1</td>
</tr>
<tr>
<td>Std. dev. of $\nu_l$</td>
<td>1.315 (0.183)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Notes: This table reports simulated minimum distance estimates of parameters governing the distribution of value-added, bias, and first-stage compliance across schools for a lagged scores value-added model with sector effects. See notes to Table 5 for a description of the estimation procedure.