NBER WORKING PAPER SERIES

LEVERAGING LOTTERIES FOR SCHOOL VALUE-ADDED:
TESTING AND ESTIMATION

Joshua Angrist
Peter Hull
Parag A. Pathak
Christopher Walters

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

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

We gratefully acknowledge funding from the National Science Foundation, the Laura and John Arnold Foundation, and the Spencer Foundation. 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, the Federal Reserve Bank of New York, the 2015 Becker/Friedman Applied Microeconomics Conference, and the University of Chicago Committee on Education Workshop for suggestions and comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. Angrist's daughter teaches at a Boston charter school.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Joshua Angrist, Peter Hull, Parag A. Pathak, and Christopher Walters. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

ABSTRACT

Conventional value-added models (VAMs) compare average test scores across schools after regression-adjusting for students' demographic characteristics and previous scores. The resulting VAM estimates are biased if the available control variables fail to capture all cross-school differences in student ability. This paper introduces a new test for VAM bias that asks whether VAM estimates accurately predict the achievement consequences of random assignment to specific schools. Test results from admissions lotteries in Boston suggest conventional VAM estimates may be misleading. 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 estimates that optimally combine ordinary least squares and instrumental variables estimates of VAM parameters. Simulations calibrated to the Boston data show that, bias notwithstanding, policy decisions based on conventional VAMs are likely to generate substantial achievement gains. Estimates incorporating lotteries are less biased, however, and yield further gains.

Joshua Angrist
Department of Economics, E17-226
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
angrist@mit.edu

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

Parag A. Pathak
Department of Economics, E17-240
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
ppathak@mit.edu

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

1 Introduction

Public school districts increasingly use value-added models (VAMs) to assess teacher and school effectiveness. Conventional VAM estimates compare test scores across classrooms or schools after regression-adjusting for students' demographic characteristics and earlier scores. Achievement differences remaining after adjustment are attributed to differences in teacher or school quality. Some districts use estimates of teacher value-added to guide personnel decisions, while others use VAMs to generate "report cards" that allow parents to compare schools.[1] Value-added estimation is a high-stakes statistical exercise: low VAM estimates can lead to school 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, 2015 for schools).

Because the stakes are so high, the use of VAM estimates for teacher and school assessment remains controversial. Critics note that VAM estimates 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; Bacher-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, accounting for 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 developed here asks whether conventional VAM estimates correctly predict the effect of randomized admission at every school that has a lottery.

Application of this test to data from Boston suggests conventional VAM estimates are biased and may be misleading. 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. The hybrid VAM 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 in a Monte Carlo simulation of the random coefficients model. Simulation results show that despite the bias in conventional VAM estimates, policy decisions based on estimates that rely on achievement changes from grade to grade or that control for baseline test scores are likely to boost achievement. For example, replacing the lowest-ranked Boston school with an average school is predicted to generate a gain of more than 0.2 standard deviations for affected students, two-thirds of 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 further gains on the order of 0.1 standard deviations. 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 the addition of lottery information improves policy targeting considerably.

The rest of the paper is organized as follows. 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 public school 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.[2] Boston's pilot school sector arose as a union-supported alternative to charter schools, developed jointly by the BPS district and the Boston Teachers Union. Pilot schools are part of the district but typically 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 (students not finishing elementary or middle school 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.,

[2] The 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.

2011; Angrist et al., forthcoming).

2.2 Data and Descriptive Statistics

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

Our analysis combines data from the centralized traditional and pilot match with lottery data from individual charter schools. The BPS lottery instruments code offers at applicants' first choice (highest ranked) middle schools in the match. In particular, BPS lottery offers indicate applicants whose lottery numbers are at least as high as the worst number offered a seat at their first-choice school, among those in the same priority group. Conditional on application year, first-choice school, and an applicant's priority at that school, 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 sixth 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 fifth grade. Table 2 also documents the comparability of students who were and were not offered seats in a lottery. These results, reported in columns 3-6, 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 basic premise of the VAM framework: internally valid treatment effects from earlier years and cohorts are presumed to have predictive value for future cohorts. Student i's potential test score at school j, Y_{ij}, is therefore written as the sum of two non-interacting components, specifically:

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

where μ_j is the mean potential outcome at school j and a_i is student i's "ability," or latent achievement potential. This additively-separable 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 heterogeneity as well).

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

$$\begin{aligned} Y_i &= Y_{i0} + \sum_{j=1}^{J} \left(Y_{ij} - Y_{i0}\right) D_{ij} \\ &= \mu_0 + \sum_{j=1}^{J} \beta_j D_{ij} + a_i. \end{aligned} \tag{2}$$

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

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

$$a_i = X_i'\gamma + \epsilon_i, \tag{3}$$

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

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

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

We'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:

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

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

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

$$E\left[\epsilon_i | D_{ij}\right] = 0; j = 1, ..., J. \tag{6}$$

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

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

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

4 Validating 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 ℓ are randomly assigned conditional on a set of lottery-specific stratifying variables, $C_{i\ell}$. These variables include an indicator for applying to school ℓ and possibly other variables such as application cohort and walk zone status. The vector $C_i = (C_{i1}', .., C_{iL}')'$ collects these variables across all lotteries. 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,

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

for a vector of parameters λ_0 and λ_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

$$E[v_i|C_i, Z_i] = \lambda_0 + C_i'\lambda_c. \tag{8}$$

The restrictions described by (8) generate an overidentification test of the sort widely used with IV estimators. In particular, equation (8) can be tested by checking whether $\phi_z = 0$ in the regression model

$$v_i = \phi_0 + C_i'\phi_c + Z_i'\phi_z + \omega_i. \tag{9}$$

In practice, equation (9) is estimated using sample OLS residuals, $\hat{v}_i$, rather than population residuals, 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-identifed 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 this test procedure help to clarify the nature of the restrictions generated by 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

$$Y_i = \rho_0 + C_i'\rho_c + Z_i'\rho_z + \eta_i, \tag{10}$$

and

$$\hat{Y}_i = \psi_0 + C_i'\psi_c + Z_i'\psi_z + u_i, \tag{11}$$

should therefore produce the same result. In other words, testing the vector equality (again, a set of L restrictions)

$$\rho_z = \psi_z \tag{12}$$

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

A further revealing 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 an endogenous variable, on the right. In what follows, we refer to this IV estimate as the VAM "forecast coefficient," since the resulting estimate gauges the predictive value of VAM estimates. Using all lottery offers as instruments, or using them one at a time, IV estimates of the forecast coefficient should equal 1. This too amounts to a set of L restrictions.

The IV interpretation of the restrictions generated by lotteries links our approach with the tests of "forecast bias" 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 IV procedures similar to the one described here. Previously applied tests have one degree of freedom, however, even though the underlying models generate as many testable restrictions are there are lottery (or other quasi-experimental) instruments in the relevant data set. Our test evaluates all of these overidentifying restrictions jointly, exhausting the lottery information in the data. In practice this means that the test procedure looking only at whether a single over-identified instrumental variables estimate of $^{\rho_z}/_{\psi_z}$ is close to 1 might accept the null hypothesis if $\rho_z = \psi_z$ on average across lotteries, even while deviations from equality might be large and statistically significant for specific lotteries.

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 the "lagged score" model, adds cubic functions of baseline math and ELA test scores. Lagged score specifications of this type are at the heart of the econometric literature on value-added models (Kane et al., 2008; Rothstein, 2010; Chetty et al., 2014a). Finally, we consider a "gains" specification that replaces score levels with grade-to-grade score changes in the demographic specification. This model parallels commonly seen accountability policies that measure test score growth.[3]

Figure 1 summarizes the value-added estimates generated by sixth-grade math scores.[4] Each bar reports an estimated standard deviation of α_j across schools, expressed in test score standard deviation units (σ) and adjusted for estimation error.[5] Adding controls for demographic variables and previous scores reduces the standard deviation of α_j from 0.5σ in the uncontrolled model to about 0.2σ in the lagged score and gains models. This 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 dummies for schools in the charter and pilot school sectors. Controlling for sector effects reduces variation in value-added, suggesting the presence of large differences in value-added across sectors.

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

[4] We focus on math scores because value-added for math appears to be more variable across schools than value-added for ELA. Bias tests for ELA, presented in Appendix Table A1, yield similar results.

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

Table 3 describes the results of tests for bias in conventional VAMs. The first row reports IV estimates of the VAM forecast coefficient, that is, the coefficient generated by instrumenting VAM fitted values with lottery offers as in equations (10) and (11). The estimator used here is an 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.[6] A weak first stage therefore makes the VAM bias test less likely to reject. The F-statistics in columns 1-4 range from 26 to 46, suggesting finite-sample bias is unlikely to be a concern in this application.

The remaining rows of Table 3 report p-values for three VAM validity tests. The first is for the null hypothesis that the forecast coefficient equals one. The second tests the associated set of overidentifying restrictions, which require that IV estimates of the forecast coefficient be the same for all lotteries, though not necessarily equal to one. The third tests overidentifying restrictions plus the restriction of a common value of one for each lottery-generated forecast, implemented by regressing conventional 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 refinement developed in Hall and Horowitz (1996), implemented via the Bayesian bootstrap (see Appendix B.1 for details).

The results of these tests suggest that conventional value-added estimates are biased. As can be seen in columns 1 and 2 of Table 3, the uncontrolled and demographic specifications generate forecast coefficient estimates of about 0.40 and 0.65, and all three specification tests reject the null hypothesis of VAM validity at conventional levels for these models. Not surprisingly, however, the lagged score and gains models do better. The forecast coefficient estimates for the lagged score and gains specifications, reported in columns 3 and 4 of the table, equal 0.86 and 0.95; the latter estimate is not statistically different from one. Importantly, however, the overidentifying restrictions for both of these richly controlled models are rejected, as are the full set of restrictions implied by the conventional VAM framework.[7]

The source of these rejections can be seen in Figure 2, which plots reduced form estimates of the effect of lottery offers on test scores against the corresponding first-stage estimates of the effect of lottery offers on conventional VAM fitted values. Each panel also shows a line through the origin with slope equal to the relevant IV coefficient from Table 3 (plotted as a solid line) along with the 45-degree line (plotted as a dashed line). In other words, Figure 2 gives a visual instrumental variables representation of the VAM forecast coefficient. A valid VAM should generate points along the 45-degree line, with deviations due solely to sampling error. Consistent with the results in Table 3, points for the uncontrolled and demographic specifications are far from the line and many deviations are at least marginally significant (deviations significant

[6] The OLS version of this model is a regression of test scores on VAM fitted values. When estimated in the same sample as the value-added model, with no additional controls, any regression of a dependent variable on the corresponding OLS fitted values necessarily produces a coefficient of one. In practice, the OLS and IV specifications used here differ in that the latter control for lottery strata and exclude some students.

[7] As a point of comparison, we also tested VAM validity in the Charlotte-Mecklenberg lottery data analyzed by Deming (2014). Tests of the full set of conventional VAM restrictions in these data generate a bootstrap-refined p-value of 0.002.

at 10% or better are shaded). Slopes are much closer to one for the lagged score and gains specifications, but points for many individual lotteries remain far from the diagonal, leading to rejection of the overidentifying restrictions implied by the conventional VAM framework.

The results in Figure 2 highlight the difference between the testing strategy developed here and previous efforts to validate VAMs. Many discrepancies between VAM predictions and lottery effects arise from points near the vertical axis and far from the horizontal axis, implying negligible predicted effects but large actual effects of random offers. Such lotteries contribute weak instruments to the IV model and therefore have little influence on the overall forecast 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. Earlier validation strategies focus on the forecast coefficient, ignoring overidentifying restrictions. Our test reveals that forecast bias may be small, even while conventional VAM estimates produce a significantly biased account of score gains from random assignment to some schools.

Figure 2 also reveals that much of the conventional VAM estimates' 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 estimates is weaker in the pilot and traditional public school sectors. This is confirmed in column 5 of Table 3, which reports results for the lagged score specification excluding charter lotteries. At 0.55, the estimated forecast coefficient for this model is much farther from one than the forecast coefficient estimate in the full sample, though the absence of charter lotteries also reduces the precision of the estimate.[8] The corresponding p-value from a test of all restrictions is 0.002, so the conventional VAM specification can also be rejected without charter lotteries.

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 the 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 general problem for the value-added framework. In the presence of heterogeneous school effects, OLS VAM estimators capture 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 use if they fail to reliably predict the effects of

[8] The first stage F-statistic for the specification without charter lotteries is 11.2, suggesting weak instruments might be a problem. It's encouraging, therefore, that the LIML estimate of the forecast coefficient for this specification is virtually the same as the IV estimate.

changing school assignments.[9] Nevertheless, it's worth exploring the roles of bias and effect heterogeneity in driving the rejections in Table 3.

Two analyses shed light on the distinction between heterogeneity and bias. The first is a set of bias tests using OLS VAM specifications that allow school effects to differ across covariate-defined "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 strategy for quantifying heterogeneity. The second analysis tests for bias in OLS VAMs estimated in the lottery sample. This approach asks 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 the test results in Table 3 reflect bias rather than heterogeneity. Panel A of Table 4 reports test results for a version of the lagged score model with school effects that vary across student types. Column 2 shows the results of allowing VAM estimates to differ by year, thereby accommodating "drift" in school effects over time; Chetty et al. (2014a) document such drift in teacher value-added. Columns 3-5 show results for subgroups defined by subsidized lunch eligibility, special education status, and baseline test score terciles. Finally, column 6 reports the test results from models that allow value-added to differ across cells constructed by fully interacting race, sex, subsidized lunch eligibility, special education, English-language learner status, and baseline score tercile. Each variation generates a clear rejection, in spite of the fact that the underlying subsample and cell-specific VAM estimates are relatively imprecise. Similarly, as can be seen in panel B of Table 4, test statistics constructed for the quasi-experimental sample also reject the null hypothesis of conventional VAM validity. These findings suggest that differences between lottery applicants or compliers and other students are not the primary force behind the rejections in Table 3.

5 The Distribution of School Effectiveness

The test results in Table 3 suggest conventional VAM estimates are biased. At the same time, Figure 2 shows that OLS VAM estimates are correlated with lottery reduced forms, while the estimated forecast coefficients for the lagged score and gains specifications are close to one. OLS estimates would therefore seem to be of value even if imperfect. This section develops a hybrid estimation strategy that combines lottery and OLS estimates in an effort to produce the most accurate value-added estimates possible. We also develop a strategy to gauge the consequences of accountability policies that rely on value-added models.

[9] See Condie et al. (2014) for a discussion of the hazards of teacher value-added estimation with heterogeneous effects.

5.1 A Random Coefficients Lottery Model

The hybrid estimation strategy uses a random coefficients model to describe the joint distribution of value-added, bias, and lottery compliance across schools. The model is built on a set of OLS, lottery reduced form, and first stage estimates. Let ρ_z^ℓ denote the element of the reduced form coefficient vector, ρ_z, corresponding to lottery offer dummy $Z_{i\ell}$. Equations (4) and (7) imply that

$$\rho_z^\ell = \sum_{j=1}^{J} \pi_j^\ell \beta_j,$$

where π_j^ℓ 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 ℓ identifies a linear combination of value-added parameters, with coefficients given by the shares of students shifted into or out of each school by the ℓth lottery offer. Estimates of the first-stage coefficients, π_j^ℓ, can be 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 distribution of random coefficients in the population of schools. Specifically, the estimates are written:

$$\begin{aligned} \hat{\alpha}_j &= \beta_j + b_j + e_j^\alpha, \\ \hat{\rho}_z^\ell &= \sum_j \pi_j^\ell \beta_j + e_\ell^\rho, \qquad (13) \\ \hat{\pi}_j^\ell &= \pi_j^\ell + e_{\ell j}^\pi; \end{aligned}$$

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

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

$$\pi_\ell^\ell = \frac{\exp\left(\delta_\ell\right)}{1 + \exp(\delta_\ell)}, \qquad (14)$$

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

[10] Conditional random assignment of admission offers implies that Z_i is conditionally independent of baseline covariates X_i in addition to ϵ_i. Z_i is therefore conditionally mean 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 ρ_z.

school j is modeled as

$$\pi_j^\ell = -\pi_\ell^\ell \cdot \frac{\exp\left(\xi_j + \nu_j^\ell\right)}{1 + \sum_{k \neq \ell} \exp\left(\xi_k + \nu_k^\ell\right)}. \tag{15}$$

In this expression, the quantity $\xi_j + \nu_j^\ell$ is the mean utility for school j in a multinomial logit model predicting alternative school choices among students that comply with offers in lottery ℓ. The parameter ξ_j allows for the possibility that some schools are systematically more or less likely to serve as fallback options for lottery losers. ν_j^ℓ is a random utility shock specific to school j in the lottery at school ℓ. The parametrization in (14) and (15) ensures that lottery offers increase the probability of enrollment at the target school and reduce enrollment probabilities at other schools, and that offer effects on all probabilities are between zero and one in absolute value.[11]

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

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

$$(\beta_j, b_j, \delta_j, \xi_j)|Q_j \sim N\left((\beta_0 + \beta_Q Q_j, b_0, \delta_0, \xi_0), \Sigma\right). \tag{16}$$

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

$$\nu_j^\ell | Q_j \sim N\left(0, \sigma_\nu^2\right). \tag{17}$$

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

[11] This parametrization implies that $0 < \pi_\ell^\ell < 1$, $-1 < \pi_j^\ell < 0$ for $j \neq \ell$, and $\pi_\ell^\ell > -\sum_{j=1, j\neq\ell}^J \pi_j^\ell$. The total probability $\sum_{j=1}^J \pi_j^\ell$ is minus the effect of an offer at the omitted school with index zero, also guaranteed to be between -1 and 0.

5.2 Simulated Minimum Distance Estimation

We estimate hyperparameters by simulated minimum distance (SMD), a variant of the method of simulated moments (McFadden, 1989). SMD focuses on moments that are determined by the parameters of interest, minimizing deviations between sample moments and the corresponding model-based predictions. Our SMD implementation uses means, variances, and covariances of functions of the OLS value-added estimates $\hat{\alpha}_j$, lottery reduced forms, $\hat{\rho}^{\ell}_{z}$, and first stage coefficients, $\hat{\pi}^{\ell}_{j}$. For example, one moment to be fit is the average $\hat{\alpha}_j$ across schools; another is the cross-school variance of the $\hat{\alpha}_j$. Other moments are means and variances of reduced form and first stage estimates across lotteries. Appendix B.3 lists all moments used for SMD estimation.

The fact that the moments in this context are complicated functions of the hyperparameters motivates a simulation approach. For example, the mean reduced form is $E[\rho^{z}_{\ell}] = \sum_j E\left[\pi^{\ell}_{j}\beta_j\right]$. This is the expectation of the product of a normally distributed random variable with a ratio involving correlated log-normals, a moment for which no analytical expression is readily available. Moments are therefore simulated by fixing a value of θ and drawing a vector of school-level parameters using equations (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. The parameter and estimation draws are combined to generate a simulated vector of parameter estimates for the given value of θ. Finally, these are used to construct a set of model-based predicted moments. The SMD estimator minimizes a quadratic form in the difference between predicted moments and the corresponding moments observed in the data. As described in Appendix B.3, the SMD estimates reported here are generated by a two-step procedure with an efficient weighting matrix in the second step.

5.3 Empirical Bayes Posteriors

Studies of teacher and school value-added typically employ EB strategies that generate posterior predictions of value-added by shrinking 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 estimates are presumed unbiased, the posterior mean value-added for school j is

$$E\left[\alpha_j|\hat{\alpha}_j\right] = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + Var(e^\alpha_j)}\hat{\alpha}_j + \left(1 - \frac{\sigma^2_\alpha}{\sigma^2_\alpha + Var(e^\alpha_j)}\right)\alpha_0, \tag{18}$$

where α_0 and σ^2_α are the mean and variance of the conventional OLS VAM estimates. 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 bias. The price for eliminating bias is a loss of precision: because IV uses only the variation generated by random assignment, lottery-based estimates are much less precise than the 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 trades off the advantages and disadvantages of OLS and IV to generate minimum mean squared error (MMSE) estimates of value-added.[12]

To see how this trade-off works in our setting, suppose the first stage parameters, π_j^ℓ, are known rather than estimated (i.e. $e_{\ell j}^\pi = 0 \ \forall \ell, j$). Let Π denote the $L \times J$ matrix of these parameters, and let β, $\hat{\alpha}$ and $\hat{\rho}_z$ denote vectors collecting β_j, $\hat{\alpha}_j$ and $\hat{\rho}_z^\ell$. Appendix B.4 shows that the posterior distribution for β in this case is multivariate normal with mean:

$$E\left[\beta|\hat{\alpha}, \hat{\rho}_z\right] = W_1(\hat{\alpha} - b_0\iota) + W_2\hat{\rho}_z + (I_J - W_1 - W_2\Pi)\beta_0\iota, \tag{19}$$

where ι 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, $(\hat{\alpha} - b_0\iota)$, lottery reduced form estimates, $\hat{\rho}_z$, and mean value-added, $\beta_0\iota$. The weighting matrices, W_1 and W_2, are functions of the first stage parameters and the covariance matrix of estimation error, value-added, and bias. Expressions for these matrices appear in Appendix B.4.

As with conventional EB posteriors, an empirical Bayes version of the posterior mean plugs first-step estimates of b_0, β_0, W_1, and W_2 into equation (19). With known first stage coefficients, hyperparameter estimation is simplified by noting that

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

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

The mean hyperparameters, β_0 and b_0, may therefore be estimated as

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

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

The hyperparameters that determine W_1 and W_2 may likewise 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, Π, 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 then be computed by indirect least squares as $\hat{\beta} = \Pi^{-1}\hat{\rho}_z$. The posterior mean in equation (19) becomes

$$E\left[\beta|\hat{\alpha}, \hat{\rho}_z\right] = W_1(\hat{\alpha} - b_0\iota) + \tilde{W}_2\hat{\beta} + (I_J - W_1 - \tilde{W}_2)\beta_0\iota, \tag{20}$$

[12] This is in the spirit of the combination estimators discussed by Judge and Mittlehammer (2004; 2005; 2007). Chetty and Hendren (2015) apply related techniques in an analysis of neighborhood effects.

for $\tilde{W}_2 = W_2\Pi$. This expression reveals that when a lottery-based value-added model is identified, the posterior mean for value-added is a weighted average of IV estimates, OLS estimates net of mean bias, and mean value-added, with weights that sum to the identity matrix. The weights are chosen so as to minimize (asymptotic) mean-squared error.

In practice, some lotteries are undersubscribed, so IV estimates of value-added can't be computed. 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 on the value-added of all other schools that applicants to this school might attend. If $\pi_j^\ell \neq 0$ the reduced form for lottery ℓ contains information that can be used to improve the posterior prediction of β_j.

Finally, equation (19) reveals 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 β_j gives no weight to $\hat{\rho}_z$ when $\pi_j^\ell = 0$ and $Cov(e_j^\alpha, e_\ell^\rho) = 0$ for all lotteries, ℓ. With the added assumption that $Cov(e_j^\alpha, e_k^\alpha) = 0$ for $k \neq j$, the jth element of equation (19) simplifies to

$$E\left[\beta_j|\hat{\alpha}, \hat{\rho}_z\right] = \frac{\sigma_\beta^2 + \tau\sigma_\beta\sigma_b}{\sigma_\beta^2 + \sigma_b^2 + 2\tau\sigma_\beta\sigma_b + Var\left(e_j^\alpha\right)}(\hat{\alpha}_j - b_0) + \left(1 - \frac{\sigma_\beta^2 + \tau\sigma_\beta\sigma_b}{\sigma_\beta^2 + \sigma_b^2 + 2\tau\sigma_\beta\sigma_b + Var\left(e_j^\alpha\right)}\right)\beta_0, \qquad (21)$$

where σ_β and σ_b are the standard deviations of β_j and b_j and τ is their correlation. Even without a lottery at school j, predictions based on equation (21) improve upon the conventional VAM posterior given by equation (18). The improvement here comes from the fact that the schools with 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.[13]

Equation (19) is a pedagogical formula derived assuming first stage parameters are known. With an estimated first stage, the posterior distribution does not have a closed form. Although the posterior mean for the general case can be approximated using Markov Chain Monte Carlo (MCMC), with a high-dimensional random coefficient vector, MCMC may be sensitive to starting values or other user parameters. Therefore, as in Chamberlain and Imbens (2004), we report EB posterior modes (also known as maximum *a posteriori* estimates; see, e.g., Gelman et al. (2013)). The posterior mode is relatively easily calculated; see Appendix B.4 for details. When posterior value-added is normally distributed as in the fixed first stage case, the posterior mode and posterior mean coincide. As a practical matter, the posterior modes computed here are similar to the weighted averages generated by equation (19) under the fixed first stages assumption, with a correlation across schools of 0.82 in the lagged score model.

[13] Using the fact that $\alpha_j = \beta_j + b_j$, equation (18) can be written to look more like equation (21):

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

6 Parameter Estimates

6.1 Hyperparameters

The SMD procedure for estimating the hyperparameters takes as input a single set of lottery reduced form and first stage estimates, along with conventional VAM estimates from the four models tested in Table 3. The lottery estimates come from regressions of test scores and school attendance indicators (the set of D_{ij}) on lottery offer dummies (Z_i), with controls for randomization strata (C_i) and the baseline covariates from the lagged score VAM specification (strata controls are necessary for instrument validity, while lagged scores and other covariates increase precision). Combining the lottery estimates with the four sets of estimates of α_j generates four sets of hyperparameter estimates, one for each VAM model.

As can be seen in columns 1-4 of Table 5, the hyperparameter estimates reveal substantial heterogeneity in both causal value-added and selection bias across schools. The standard deviation of value-added, σ_β, is similar across specifications, ranging from about 0.20σ in the gains specification to 0.22σ in the lagged score 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 the relatively stable estimates of σ_β, the estimated standard deviation of bias, σ_b, shrinks from 0.49σ in the uncontrolled model to about 0.17σ 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, the estimated standard deviations of bias are statistically significant for all models, implying that controls for demographic variables and baseline achievement are not sufficient to produce unbiased comparisons. The estimates in columns 3 and 4 of Table 5 show that even for the relatively successful lagged score and gains specifications, the estimated variance of bias is almost as large as the estimated variance of causal value-added.

The estimated correlation between β_j and b_j (the hyperparameter τ) 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 of τ are too imprecise to be conclusive. Estimates of β_Q, the lottery school value-added shifter, are mostly small and none are significantly different from zero. This suggests that differences between oversubscribed and undersubscribed schools are modest. The negative estimates of β_Q for models without sector effects, reported in columns 1-4 of the table, imply that, if anything, lottery schools generate slightly smaller gains than schools without over-subscribed lotteries.

Earlier work on school effectiveness explores differences between Boston's charter, pilot, and traditional public sectors (Abdulkadiroğlu et al., 2011; Angrist et al., forthcoming). These estimates show strong charter school treatment effects in Boston, a finding that suggests accounting for sector differences may improve the predictive accuracy of school value-added models. Columns 5 and 6 of Table 5 therefore report estimates of lagged score and gains models in which the means of random coefficients depend on school sector (Appendix

Table A2 reports the complete set of parameter estimates for this model).

Consistent with earlier findings, models with sector effects suggest that average charter school value-added exceeds traditional public school value-added by 0.35σ. Estimated differences in value-added between pilot and traditional public schools are smaller and statistically insignificant. By contrast, bias seems unrelated to sector, implying that conventional VAM models with demographic and lagged achievement controls accurately reproduce lottery-based comparisons of the charter, pilot and traditional sectors. Estimates of β_Q in these specifications are positive, but again small and not significantly different from zero. Finally, the estimates of σ_β and σ_b show that sector effects reduce cross-school variation in value-added and bias by about 20-25 percent. In other words, most of the variation in school quality appears to be within sectors, rather than between.

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 posterior means for each conventional model. Not surprisingly, the correlation is strongest for the lagged score and gains specifications. On the other hand, rank correlations in the lagged score and gains models are around 0.77, so hybrid estimation changes some schools' ranks. This finding suggests that accountability decisions based on estimated value-added might change when based on hybrid as opposed to conventional estimates.

Hybrid estimation generates posterior modes for bias as well as value-added. The resulting estimated bias modes can be used to explore the relationship between bias and over-subscription, providing evidence relevant for the assumption that bias distributions are the same for schools with and without lotteries. A weak or nonexistent relationship between bias and the *degree* of oversubscription is consistent with the hypothesis that bias distributions are similar for schools where lottery information is and is not available. For the purposes of this exploration, the oversubscription rate is defined as the ratio of the annual average number of lottery applicants to the average number of seats for charter schools, and the ratio of the average number of first-choice applicants to the average number of seats for traditional and pilot schools.

As can be seen in Figure 4, these measures of oversubscription are essentially unrelated to bias. Specifically, the figure plots bias posterior modes from the lagged score model with sector effects against oversubscription rates at lottery schools, after regression-adjusting for sector. The slope of the regression line in the figure is 0.01 with a standard error of 0.04.

7 Policy Simulations

Accurate value-added estimates are useful for policymakers and parents making decisions about schools; we use a Monte Carlo simulation to gauge the accuracy and value of VAM estimates for such decisionmaking. The simulation draws values of causal value-added, bias, and lottery parameters from the estimated distributions underlying Table 5. Simulations of the uncontrolled, demographic and gains specifications are based on restricted SMD estimates imposing the joint distribution of (β_j, π_j^{ℓ}) estimated in the lagged score model. Remaining cross-model differences in simulated values are therefore driven solely by differences in sampling error and bias. Estimation errors are drawn from the joint asymptotic distribution of OLS and lottery estimates. The parameter and estimation error draws are combined to construct simulated OLS, reduced form and first stage parameter estimates. Finally, these estimates are used to construct conventional and hybrid EB posterior predictions for each simulation.

7.1 Mean Squared Error and Accountability Targeting

The root-mean-squared error (RMSE) of conventional VAM estimates across simulations falls sharply as controls are added (and when the outcome is a score gain). In contrast, 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 reflects reduced bias in the conventional estimates, while the stability of the hybrid procedure's RMSE is evidence of successful bias mitigation regardless of the conventional starting point. For example, starting from RMSEs of 0.48σ and 0.31σ for conventional estimates in the uncontrolled and demographic specifications, the hybrid posterior mode pulls RMSE below 0.2σ for both models. The hybrid posterior corrects for the fact that most of the variation in $\hat{\alpha}_j$ is due to bias when $\hat{\alpha}_j$ is estimated in models that don't adjust for previous scores.

Conventional VAM estimates generate much lower RMSE values when computed using the lagged score and gains specifications, but the hybrid approach yields improvements for these specifications as well. An RMSE of 0.17σ for conventional posteriors derived from the lagged score and gains specifications falls to 0.14σ for the hybrid. When sector effects are included, hybrid posteriors generate an even larger improvement, reducing RMSE from 0.14σ to about 0.10σ, a reduction of almost 30 percent.

Like many states and school districts, the Massachusetts Department of Elementary and Secondary Education implements an accountability scheme based on standardized tests. Massachusetts' Framework for School Accountability and Assistance places schools into five "levels" based on four-year histories of test score levels and changes. Schools in the bottom quintile of this measure are designated level 3 or higher. A subset of these schools are classified in levels 4 and 5, a designation that puts them at risk of restructuring or closure.[14]

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 traditional BPS and pilot schools above or below various percentiles of the true value-added distribution. The error rate in column

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

1 is the frequency at which schools in the lowest decile of causal value-added are mis-identified as having higher scores. Columns 2-3 report the same sort of misclassification rates for lowest quintile and lowest tercile schools, while columns 4-6 show error rates for highest decile, quintile and tercile schools. Note that because true and estimated classification tiers are the same size, the probability that a school is incorrectly classified as being, say, outside the lowest decile is equal to the fraction of schools classified in the lowest decile that do not belong there.

Uncontrolled value-added estimates produce highly inaccurate school rankings. As can be seen in the second row of Table 6, uncontrolled VAM misclassifies 86 percent of lowest decile schools, 74 percent of lowest quintile schools, and 63 percent of lowest tercile schools. These rates are not much better than the error rates for a policy that simply ranks schools randomly (90, 80 and 67 percent, shown in the first row). This finding implies that school report cards based on unadjusted achievement levels, distributed in many states and districts, are likely to be highly misleading.[15] Hybrid posterior modes that combine uncontrolled OLS and lottery estimates misclassify 67, 52 and 41 percent of lowest decile, quintile and tercile schools. Although still high, these error rates represent a marked improvement on the rates produced by the conventional posterior mean from an uncontrolled model. Compare, for example a hybrid misclassification rate of 41 percent for lowest-tercile schools with the corresponding rate of 63 percent when using conventional posterior means.

Adding controls for demographics and previous achievement reduces misclassification rates based on both conventional and hybrid estimates, but does not eliminate the utility of hybrid estimation. Conventional misclassification rates for lowest decile, quintile and tercile schools are 57, 47 and 39 percent when rankings are based on estimates from the gains specification. In this model, hybrid estimation reduces classification error in the lowest decile from 57 to 50 percent, 12 percent fewer mistakes. The hybrid advantage is larger when classifying lowest quintile and lowest tercile schools, reducing error rates by 21 and 18 percent in the gains specification. The improvements generated by hybrid classification of high-performing schools are even larger: incorporating lotteries cuts mistakes by 24, 21 and 20 percent for highest decile, quintile and tercile schools. The pattern of classification improvement from the lagged score and gains specifications are broadly similar. For both the lagged score and gains models, hybrid estimation cuts mistakes in classifying upper and lower tercile schools to under one third.

The relationship between school rankings based on true and estimated value-added summarizes the predictive value of VAM estimates. Column 7 reports coefficients from regressions of a school's rank in the causal value-added distribution on its rank in each estimated distribution. This rank coefficient increases from 0.14 in the uncontrolled conventional model to 0.61 in the conventional gains specification. Hybrid estimation boosts the rank coefficient for gains to 0.72. In other words, sufficiently controlled VAM estimates strongly predict relative value-added: a one-position increase in a school's VAM rank translates into an average increase of 0.6-0.7 positions in the distribution of true school quality. At the same time, even the largest

[15] 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`).

rank coefficients in the table remain well below one, suggesting there is still considerable scope for mistakes in classification decisions based on estimated VAM. It's of interest, therefore, to consider the consequences of imperfect classification for the students directly affected by policies built on these estimates.

7.2 Effects on Students

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

This analysis ignores any transition effects such as possible disruption due to school closure, peer effects resulting from changes in school composition, or general equilibrium effects that might inhibit replication of successful schools. The results should nevertheless provide a rough guide to the potential consequences of VAM-based policy decisions. Quasi-experimental analyses of charter takeovers and other school reconstitutions in Boston, New Orleans, and Houston have shown large gains when low-performing schools are replaced by schools operating according to pedagogical principles seen to be effective elsewhere (Fryer, 2014; Abdulkadiroğlu et al., 2015). This suggests transitional consequences are dominated by longer-run determinants of school quality, at least for modest policy interventions of the sort considered here.

Consistent with the high misclassification rates generated by uncontrolled VAMs, Table 7 shows that policies based on uncontrolled VAM estimates generate negligible score gains. Using the simplest VAM model as a guide, replacing the lowest-scoring district school with an average school is predicted to increase scores for affected students by only about 0.04σ. Likewise, a policy that replaces the lowest-ranked school with an expansion of an average school ranked in the top quintile generates a gain of rougly 0.06σ. These small effects reflect the large variation in bias evident for the uncontrolled model in Table 5; closure decisions based on uncontrolled estimates target schools with many low achievers rather than low value-added. On the other hand, adjusting naive VAM estimates for selection bias via hybrid estimation increases student gains from closure to 0.16σ when the replacement school is average and to 0.29σ when the replacement school is from the top quintile.

Conventional VAM specifications that adjust for lagged scores are considerably less biased than are the estimates from uncontrolled models and models that adjust only for demographic controls. Closure and replacement decisions based on the lagged score and gains models are therefore predicted to yield substantial achievement gains. For instance, replacing the lowest-ranked school with an average school boosts scores by roughly 0.24σ in the gains specification. This is 66 percent of the corresponding benefit for an infeasible policy that ranks schools by true value-added (0.36σ). Hybrid estimation of the gains model increases the estimated gains from closure to 0.29σ, generating over 80 percent of the maximum possible gain using true value-added.

The effects of VAM-based policies and the incremental benefits of using lotteries to estimate value-added grow when value-added predictions are used to choose expansion schools in addition to closures. In the gains specification, for example, replacing the lowest-ranked school with a typical top-quintile school generates an improvement of 0.38σ when conventional posteriors are used to estimate VAM and an improvement of 0.47σ when rankings are based on hybrid predictions. The latter effect is 82 percent of the theoretical maximum benefit generated by replacing the least effective school in the Boston district (as ranked by causal value-added) with an average school in the top quintile of causal value-added.

The largest gains seen in Table 7 result from a policy that replaces the lowest-ranking school with an average charter school. This mirrors Boston's ongoing in-district charter conversion policy experiment. As a result of the large difference in mean value-added between charter and district schools, charter conversion is predicted to generate large gains regardless of how value-added is estimated. Accurate value-added estimation increases the efficacy of charter conversion, however: selecting a school for conversion based on the hybrid gains specification rather than the naive uncontrolled model boosts the effect of charter expansion from 0.35σ to 0.61σ, close to the maximum possible gain of 0.67σ.

Tables 6 and 7 reveal that, despite substantial misclassification rates, VAM-based policies have the potential to boost student achievement markedly. Even when VAM estimates are far from perfect, they predict causal value-added. For example, causal value-added is more than 0.2σ below average for schools ranked at the bottom by the conventional lagged score and gains specifications. As can be seen in Table 5, this represents roughly a full standard deviation in the distribution of true school quality. Value-added for low-ranked schools is even more negative when rankings are based on hybrid estimates. The misclassification rates in Table 6 show that schools selected for closure based on VAM-based rankings are unlikely to be the very worst schools in the district. At the same time, they are likely to be worse than average, so policies that replace them with schools predicted to do better generate large achievement gains. We see especially large improvements when value-added is estimated using the relatively sophisticated lagged score and gains specifications, augmented with lotteries to mitgate bias.

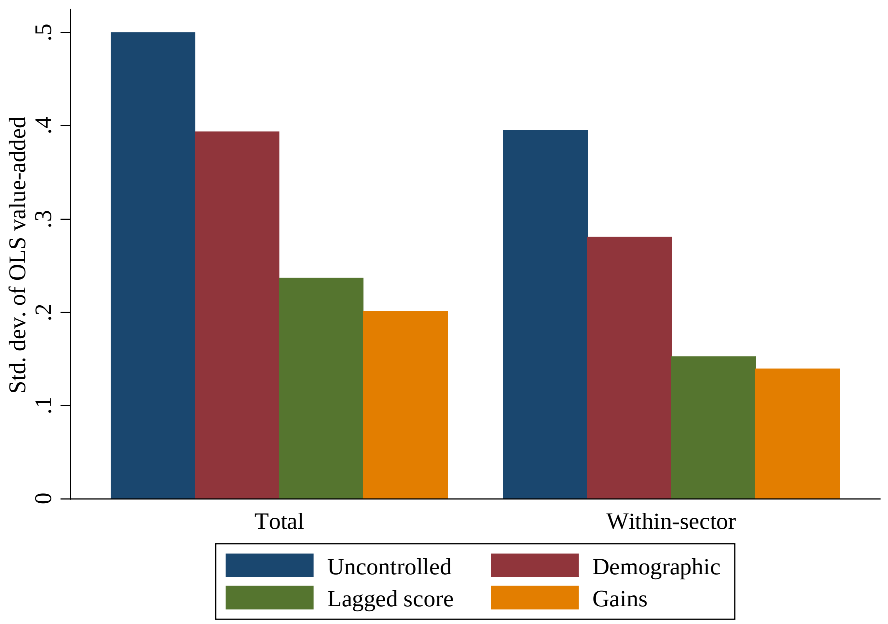
8 Conclusions and Next Steps

School districts increasingly rely on regression-based value-added models to gauge and report on school quality. This paper leverages admissions lotteries to test and improve conventional OLS estimates of school value-added. An application of our approach to data from Boston suggests that conventional value-added estimates for Boston's traditional public schools are biased. Controls for lagged test scores reduce but do not eliminate this bias. Nevertheless, policy simulations show that accountability decisions based on conventional VAM estimates are likely to boost achievement. A hybrid estimation procedure that combines conventional and lottery-based estimates leads to substantially more accurate value-added predictions, improved policy targeting, and larger achievement gains.

Our hybrid approach requires some kind of lottery-based admissions scheme, such as those increasingly used for student assignment in many of America's large urban districts. As our analysis of charter schools shows, however, admissions need not be centralized for lotteries to be useful. Equally important, the strategies outlined here remain useful even for districts, like Boston, where lottery data are missing or irrelevant for a large minority of schools.

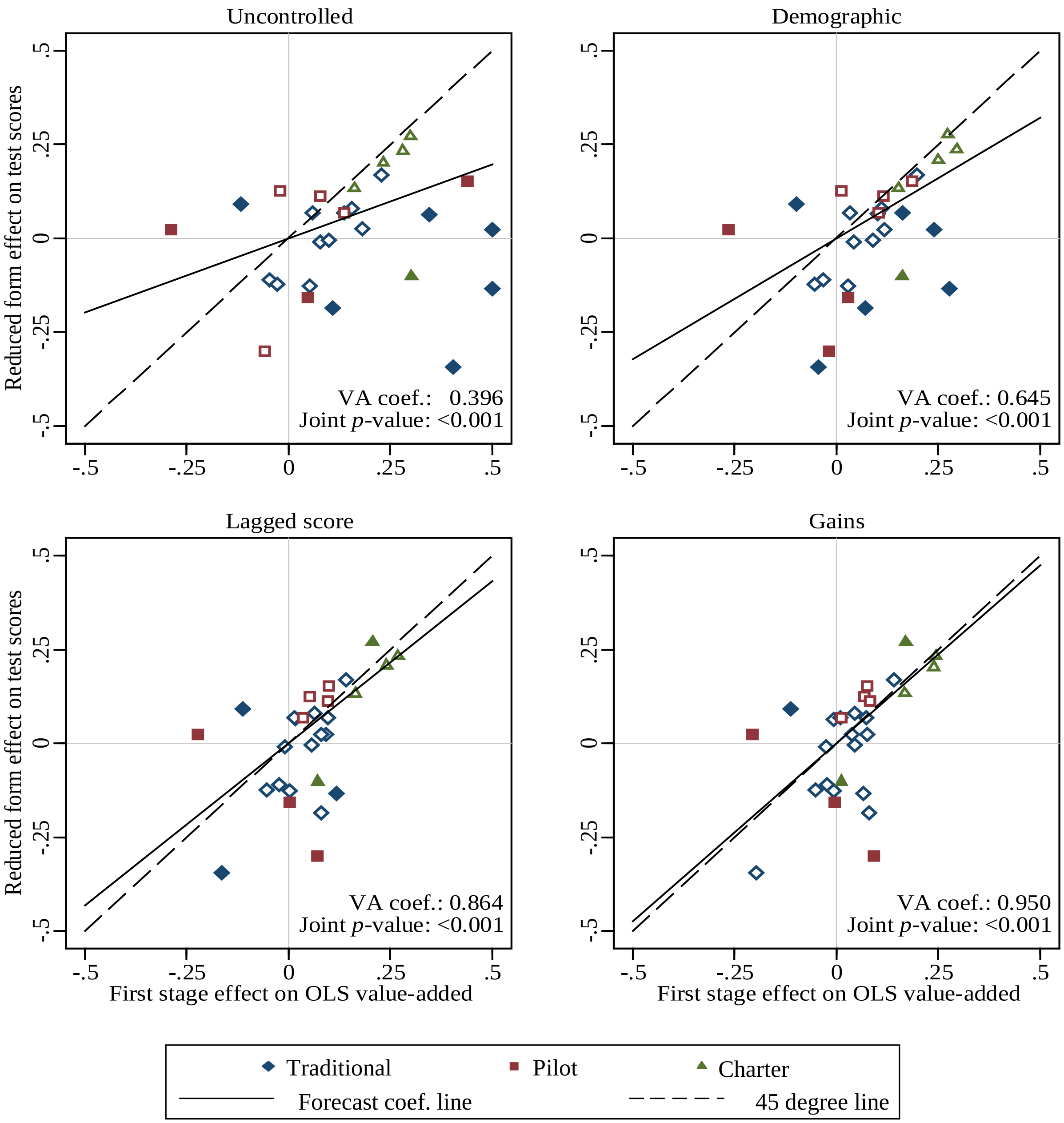
The methods developed here may also be useful for the estimation of teacher value-added. Lotteries for teacher assignment are rare, but the methods outlined here may be extended to exploit other sources of quasi-experimental variation. A complication in the teacher context is the more elaborate hierarchical data structure arising from the fact that teacher assignment has both within- and between-school components. This suggests a somewhat more complicated model for bias in teacher VAMs may be necessary. 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.

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



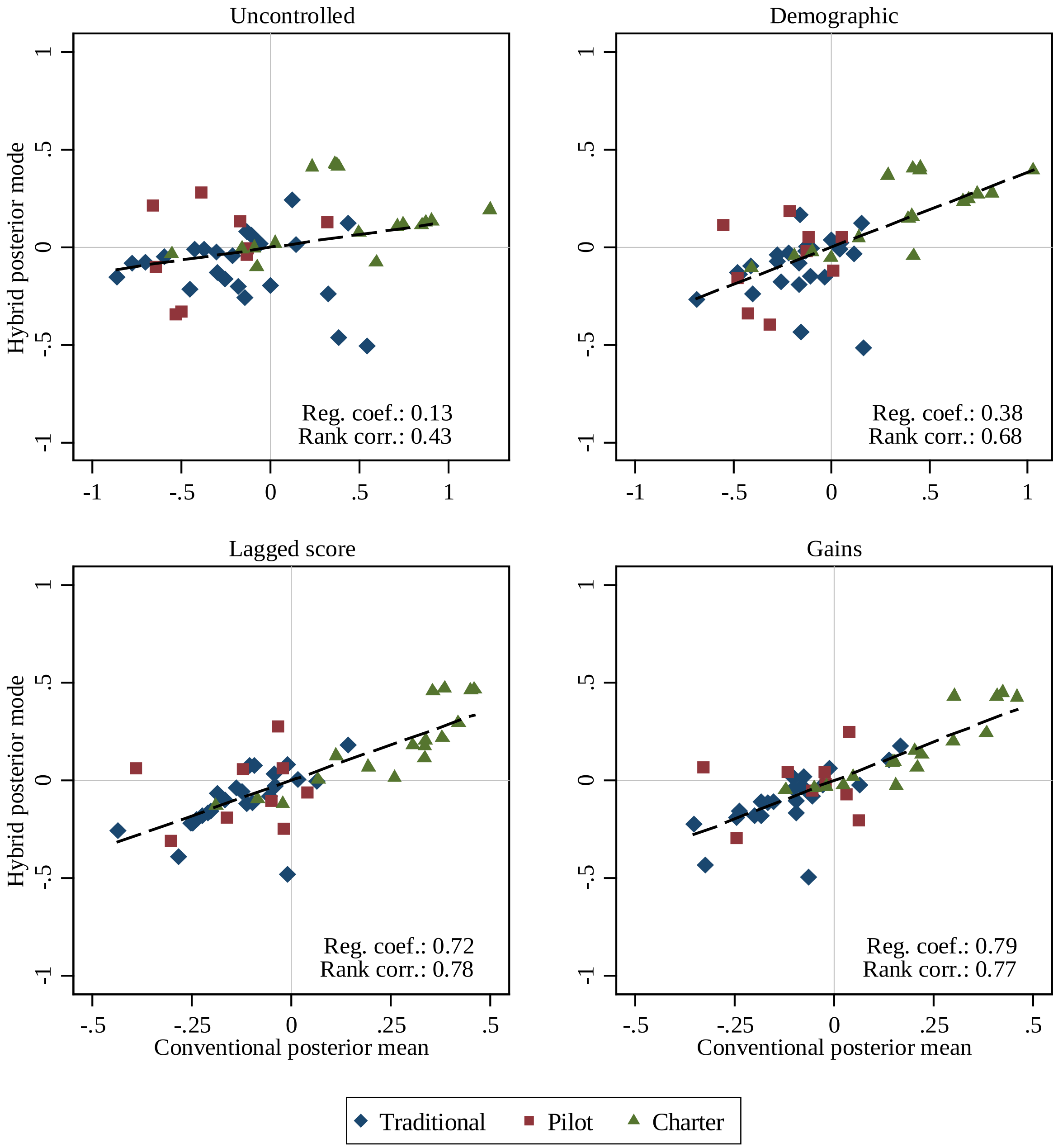
Notes: This figure compares standard deviations of school effects from four math value-added models; see Table 3's notes for a description of the models. For each model, the total variance of school effects is obtained by subtracting the average squared standard error from the sample variance of value-added 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



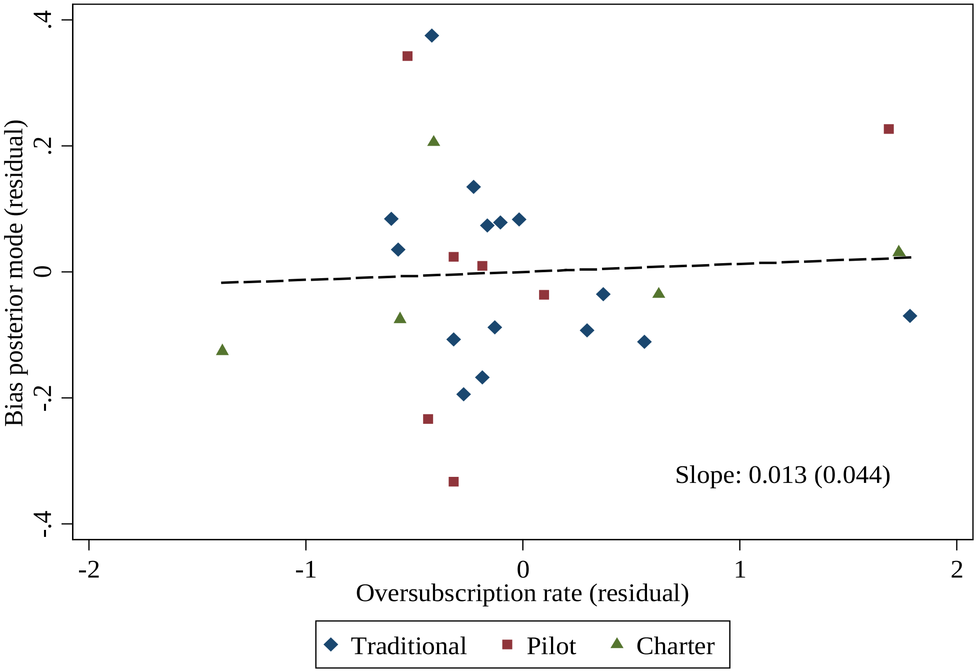
Notes: This figure plots lottery reduced form effects against value-added first stages from each of the 28 school lotteries. See the notes for Table 3 for a description of the value-added models and lottery specification. Filled markers indicate estimates that are significant at the 10% level. Slopes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

Figure 3: Empirical Bayes posterior predictions of school value-added



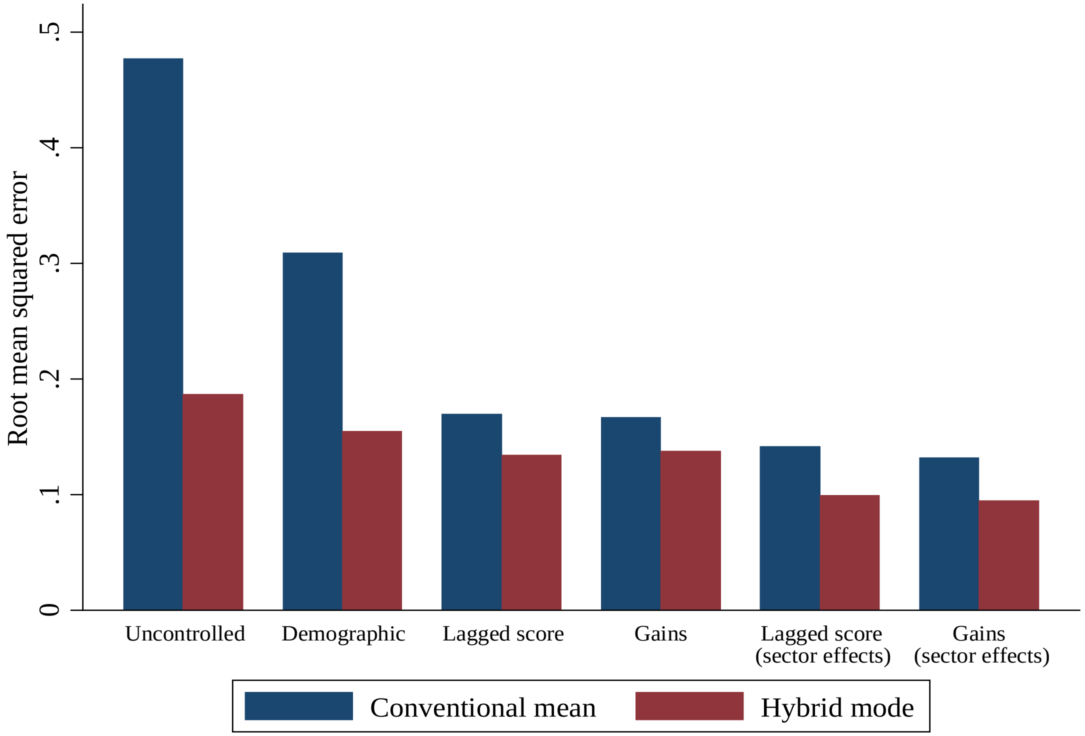
Notes: This figure plots empirical Bayes posterior mode predictions of value-added from the random coefficients model against posterior means based on OLS value-added. Posterior modes are computed by maximizing the sum of the log-likelihood of the OLS, reduced form, and first stage estimates conditional 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 score 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.

Figure 5: Root mean squared error for value-added posterior predictions



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 Table 5. The random coefficients model is re-estimated in each simulated sample.

Table 1: Boston students and schools

	Enrollment				Enrollment		
School	OLS sample	Lottery sample	Lottery school?	School	OLS sample	Lottery sample	Lottery school?
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. Traditional publics				B. Pilots		
1	1,095	79	Y	1	538	310	Y
2	1,025	445	Y	2	1,260	433	Y
3	1,713	1,084	Y	3	585	296	Y
4	547	218	Y	4	78	5	
5	217	46		5	453	46	Y
6	1,354	581	Y	6	380	67	Y
7	263	44		7	242	179	Y
8	1,637	492	Y	8	558	73	Y
9	472	104		9	18	12	
10	1,238	591	Y		C. Charters		
11	537	11		1	738	406	Y
12	331	35	Y	2	361	23	
13	335	82		3	357	215	
14	952	232	Y	4	393	332	Y
15	294	71	Y	5	338	16	
16	333	90		6	511	115	Y
17	766	243	Y	7	71	8	
18	372	47	Y	8	300	23	
19	137	14		9	389	342	Y
20	1,091	225	Y	10	654	34	
21	1,086	127	Y	11	45	3	
22	577	104	Y	12	53	2	
23	622	61		13	415	305	Y
24	906	270	Y	14	70	6	
25 *(Ref.)*	267	19		15	104	23	
				16	701	92	
All schools:	27,864	8,718	28	17	85	37	

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 2: Descriptive statistics

	Means		Offer instrument balance			
	OLS sample	Lottery sample	All lotteries	Traditional	Pilot	Charter
Baseline covariate	(1)	(2)	(3)	(4)	(5)	(6)
Hispanic	0.345	0.354	-0.017	-0.007	0.003	-0.006
			(0.013)	(0.017)	(0.033)	(0.018)
Black	0.410	0.485	-0.011	-0.005	-0.052	-0.009
			(0.014)	(0.018)	(0.034)	(0.020)
White	0.122	0.072	0.010	0.006	0.005	0.009
			(0.007)	(0.008)	(0.015)	(0.010)
Female	0.490	0.504	0.017	0.034*	-0.013	-0.025
			(0.014)	(0.019)	(0.037)	(0.020)
Subsidized lunch	0.806	0.830	0.020*	0.020	0.006	-0.005
			(0.010)	(0.013)	(0.026)	(0.016)
Special education	0.208	0.195	0.006	-0.003	-0.022	0.015
			(0.011)	(0.013)	(0.030)	(0.016)
English-language learner	0.205	0.206	0.006	-0.001	0.018	0.004
			(0.011)	(0.014)	(0.027)	(0.016)
Suspensions	0.093	0.076	-0.025	-0.025	0.009	-0.016
			(0.016)	(0.023)	(0.025)	(0.017)
Absences	1.710	1.534	-0.087	-0.138*	-0.092	0.110
			(0.095)	(0.080)	(0.260)	(0.167)
Math score	0.058	0.004	0.022	-0.026	0.080	0.036
			(0.024)	(0.030)	(0.061)	(0.035)
ELA score	0.030	0.013	0.035	0.045	0.060	0.013
			(0.025)	(0.030)	(0.061)	(0.036)
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 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 3: Tests for bias in conventional value-added models

	Uncontrolled	Demographic	Lagged score	Gains	Lagged score, no charter lotteries
	(1)	(2)	(3)	(4)	(5)
Forecast coefficient	0.396	0.645	0.864	0.950	0.549
	(0.056)	(0.065)	(0.075)	(0.084)	(0.164)
First stage *F*-statistic	45.6	36.1	29.6	26.6	11.2
p-values:					
Forecast coef. equals 1	<0.001	<0.001	0.071	0.554	0.006
Overid. restrictions	<0.001	<0.001	0.003	0.006	0.043
All restrictions	<0.001	<0.001	<0.001	<0.001	0.002
All restrictions (bootstrap refinement)	<0.001	<0.001	<0.001	<0.001	0.002

Notes: This table reports estimates of the VAM forecast coefficient and the results of tests for bias in conventional value-added models for 6th grade math scores. Estimated forecast coefficients are from regressions of 6th grade scores on fitted values from conventional 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 demographic model adds indicators for student sex, race, subsidized lunch, special education, limited-English proficiency, and counts of baseline absences and suspensions. The lagged score model adds cubic polynomials in baseline math and ELA scores. The gains model includes the same controls as the demographic model and uses score gains from baseline as the outcome. Column (5) excludes charter school lotteries from the lottery sample in testing the lagged score model. All IV models control for lottery strata fixed effects, demographic variables, and lagged scores. Standard errors are reported in parentheses. Bootstrap p-values are based on 500 Bayesian bootstrap replications (see Appendix B for details).

Table 4: Robustness of bias tests to effect heterogeneity

		VAM estimated by subgroup				
	Baseline VAM specification	Baseline year	Subsidized lunch	Special education	Baseline score tercile	Interacted groups
	(1)	(2)	(3)	(4)	(5)	(6)
		A. VAM estimated on the OLS sample				
Forecast coefficient	0.864	0.916	0.849	0.863	0.866	0.930
	(0.075)	(0.072)	(0.075)	(0.074)	(0.075)	(0.061)
Bootstrap-refined VAM validity *p*-value	<0.001	<0.001	<0.001	<0.001	<0.001	0.002
		B. VAM estimated on the lottery sample				
Forecast coefficient	0.868	0.962	0.851	0.872	0.873	0.934
	(0.070)	(0.068)	(0.069)	(0.070)	(0.070)	(0.052)
Bootstrap-refined VAM validity *p*-value	<0.001	0.002	<0.001	<0.001	<0.001	0.004

Notes: This table reports lottery-based tests for bias in school value-added models that allow for treatment effect heterogeneity by baseline characteristics. Forecast 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 schol 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 IV models 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 5: Joint distribution of causal value-added and OLS bias

		Models without sector effects				Models with sector effects	
		Uncontrolled	Demographic	Lagged score	Gains	Lagged score	Gains
Parameter	Description	(1)	(2)	(3)	(4)	(5)	(6)
σ_β	Std. dev. of causal	0.210	0.212	0.218	0.199	0.158	0.169
	value-added	(0.063)	(0.062)	(0.061)	(0.060)	(0.070)	(0.066)
σ_b	Std. dev. of bias in	0.487	0.314	0.171	0.168	0.140	0.130
	OLS value-added	(0.068)	(0.055)	(0.074)	(0.060)	(0.068)	(0.078)
τ	Correlation of	-0.132	0.048	-0.293	-0.399	-0.480	-0.595
	value-added and bias	(0.250)	(0.329)	(0.385)	(0.356)	(0.428)	(0.348)
VA shifters	Lottery school	-0.055	-0.103	-0.058	-0.066	0.077	0.059
		(0.141)	(0.110)	(0.068)	(0.056)	(0.048)	(0.046)
	Charter					0.353	0.358
						(0.118)	(0.124)
	Pilot					0.057	0.064
						(0.137)	(0.143)
Bias shifters	Charter					0.052	-0.023
						(0.115)	(0.119)
	Pilot					-0.060	-0.043
						(0.134)	(0.138)
	J-statistic (d.f.):	5.64 (4)	7.41 (4)	3.05 (4)	2.98 (4)	4.09 (4)	2.95 (4)
	Overid. *p*-value:	0.228	0.116	0.549	0.562	0.393	0.566

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 Appendix B. 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 6: Error rates for classification decisions among district schools

		Low-performing schools			High-performing schools			
		Lowest decile	Lowest quintile	Lowest tercile	Highest decile	Highest quintile	Highest tercile	Rank coefficient
Value-added model	Posterior method	(1)	(2)	(3)	(4)	(5)	(6)	(7)
-	Random	0.900	0.800	0.667	0.900	0.800	0.667	0.000
Uncontrolled	Conventional	0.863	0.736	0.628	0.893	0.740	0.606	0.144
	Hybrid	0.670	0.516	0.411	0.573	0.480	0.418	0.538
Demographic	Conventional	0.774	0.632	0.526	0.795	0.639	0.507	0.352
	Hybrid	0.623	0.486	0.402	0.599	0.465	0.404	0.563
Lagged score	Conventional	0.619	0.504	0.412	0.704	0.518	0.416	0.548
	Hybrid	0.517	0.391	0.326	0.486	0.370	0.306	0.702
Gains	Conventional	0.570	0.474	0.390	0.650	0.466	0.374	0.611
	Hybrid	0.503	0.376	0.323	0.493	0.366	0.299	0.724

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. Column (7) reports the coefficient from a regression of a school's rank in the true value-added distribution on its rank in the estimated distribution. 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

		Replacement school:			
			Average above-	Average top-	Average charter
		Average school	median school	quintile school	school
Value-added model	Posterior method	(1)	(2)	(3)	(4)
-	True value-added	0.357	0.488	0.570	0.666
Uncontrolled	Conventional	0.036	0.057	0.064	0.345
	Hybrid	0.163	0.242	0.294	0.472
Demographic	Conventional	0.106	0.154	0.182	0.415
	Hybrid	0.188	0.266	0.321	0.496
Lagged score	Conventional	0.206	0.281	0.332	0.515
	Hybrid	0.266	0.363	0.434	0.575
Gains	Conventional	0.237	0.326	0.383	0.557
	Hybrid	0.290	0.394	0.467	0.610

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

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

Abdulkadiroğlu, A., J. D. Angrist, P. D. Hull, and P. A. Pathak (2015): "Charters without lotteries: Testing takeovers in New Orleans and Boston," IZA discussion paper no. 8985.

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

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

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

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

Angrist, J. D. and I. Fernandez-Val (2013): "ExtrapoLATE-ing: External validity and overidentification in the LATE framework," in *Advances in Economics and Econometrics*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, Cambridge University Press, vol. III of *Econometric Society Monographs, Econometrics*, chap. 11, 401–434.

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

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

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

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

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

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

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

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

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

Condie, S., L. Lefgren, and D. Sims (2014): "Heterogeneous match quality and teacher value-added: theory and empirics," *Economics of Education Review*, 40, 76–92.

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

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

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

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

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

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

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

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013): *Bayesian Data Analysis*, Chapman and Hall/CRC, third ed.

Hall, P. and J. L. Horowitz (1996): "Bootstrap critical values for tests based on generalized-method-of-moments estimators," *Econometrica*, 64, 891–916.

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

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

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

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

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

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

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

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

Kane, T. J. and D. O. Staiger (2008): "Estimating teacher impacts on student achievement: An experimental evaluation," NBER working paper no. 14607.

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

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

Mason, D. M. and M. A. Newton (1992): "A rank statistics approach to the consistency of a general bootstrap," *The Annals of Statistics*, 20, 1611–1624.

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

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

——— (2014): "Revisiting the impacts of teachers," Mimeo, University of California, Berkeley.

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

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

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

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

Appendix A: Data

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

A.1 Student enrollment, demographics, and test scores

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

Enrollment in the SIMS is grade-specific. When a student repeats grades, we retain the first school a student attended in that grade. We then record students attending multiple schools in a given school year as enrolled in the school for which the attendance duration is longest, with duration ties broken randomly. This results in a unique student panel across grades; for the purposes of this study we restrict focus to 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 tie-breaking 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 $\mathbf{X}_i = (1, D_{i1}, \ldots, D_{iJ}, X_i')'$ and $\mathbf{Z}_i = (1, C_i', Z_i')'$. OLS estimates of equation (9) may be written in matrix form as

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

Moreover,

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

while by equation (5),

$$\begin{aligned} \hat{v} - v \quad &= \quad Y - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\beta + v\right) - v \\ &= \quad -\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'v. \end{aligned}$$

We then have that:

$$\begin{aligned} \sqrt{N}(\hat{\phi} - \phi) \quad &= \quad \sqrt{N}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\mathbf{Z}'\omega - \mathbf{Z}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'v\right) \\ &= \quad \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{Z}_i'\right)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{Z}_i\omega_i - \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i\mathbf{X}_i'\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_i'\right)^{-1}\mathbf{X}_iv_i\right). \end{aligned}$$

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

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{Z}_i\omega_i - E\left[\mathbf{Z}_i\mathbf{X}_i'\right]\left(E\left[\mathbf{X}_i\mathbf{X}_i'\right]\right)^{-1}\mathbf{X}_iv_i\right) \quad \Rightarrow \quad N\left(0, \Lambda\right),$$

provided

$$\Lambda \quad = \quad E\left[\left(\mathbf{Z}_i\omega_i - E\left[\mathbf{Z}_i\mathbf{X}_i'\right]\left(E\left[\mathbf{X}_i\mathbf{X}_i'\right]\right)^{-1}\mathbf{X}_iv_i\right)\left(\mathbf{Z}_i\omega_i - E\left[\mathbf{Z}_i\mathbf{X}_i'\right]\left(E\left[\mathbf{X}_i\mathbf{X}_i'\right]\right)^{-1}\mathbf{X}_iv_i\right)'\right]$$

is finite. By Slutsky's theorem,

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

for

$$\Xi \quad = \quad \left(E\left[\mathbf{Z}_i \mathbf{Z}_i'\right]\right)^{-1} \Lambda \left(E\left[\mathbf{Z}_i \mathbf{Z}_i'\right]\right)^{-1}.$$

We base asymptotic inference on ϕ by forming sample analogues of each component of Ξ. Note that this estimate differs from the usual White (1980) heteroskedasticity-consistent covariance matrix estimator by the second term in the inner product of Λ. 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

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

where $\hat{\Xi}_z$ is the sub-matrix of our estimate of Ξ corresponding to $\hat{\phi}_z$. The distribution of F is first-order equivalent to χ^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 re-estimate 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,

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

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

$$p \quad = \quad \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}[F^b > F].$$

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

$$Y_{ij} = \sum_t \mu_{jt} T_{it} + a_i + m_{ij}, \tag{22}$$

where T_{it} is an indicator for belonging to type t, μ_{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

$$Y_i = \sum_t \mu_{0t} T_{it} + \sum_{j=1}^{J} \sum_t \beta_{jt} D_{ij} T_{it} + X_i' \gamma + \epsilon_i, \tag{23}$$

where $\beta_{jt} = \mu_{jt} - \mu_{j0}$ is the value-added of school j for type t and the error term is $\epsilon_i = \sum_{j=1}^{J} \sum_t (m_{ij} - m_{i0}) D_{ij} T_{iw} + \epsilon_{i0}$, with ϵ_{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:

$$E\left[\epsilon_{i0} | D_i\right] = 0, \tag{24}$$

$$E\left[m_{ij} - m_{i0} | D_i, T_i\right] = 0, \tag{25}$$

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 β_{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

$$Y_i = \sum_t \alpha_{0t} T_{it} + \sum_{j=1}^{J} \sum_t \alpha_{jt} D_{ij} T_{it} + X_i' \Gamma + v_i, \tag{26}$$

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 distanced (SMD). The vector of parameters to be estimated is

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

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

$$\hat{\Omega} = \left(\hat{\alpha}_1, ..., \hat{\alpha}_J, \hat{\rho}_z^1, ..., \hat{\rho}_z^L, \hat{\pi}_1^1, ..., \hat{\pi}_1^L, ..., \hat{\pi}_J^L\right)'.$$

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

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

where V_e is a covariance matrix derived from conventional asymptotics. This requires within-school and within-lottery samples to be large enough for asymptotic approximations to be accurate. Under this assumption and the distributional assumptions in equations (14) through (17), values of Ω and $\hat{\Omega}$ can be simulated for any value of θ. We use this procedure to generate simulated data sets, and estimate θ by minimizing the distance between simulated and observed moments.

Our estimation procedure targets the following first moments:

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

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

$$\hat{m}_3 = \tfrac{1}{L}\textstyle\sum_\ell \hat{\rho}_z^\ell,$$

$$\hat{m}_4 = \tfrac{1}{L}\textstyle\sum_\ell \hat{\pi}_\ell^\ell,$$

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

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

$$\hat{m}_7 = \tfrac{1}{L}\textstyle\sum_\ell \left[\frac{\left(\hat{\pi}_\ell^\ell\right)^2}{\sum_k \left(\hat{\pi}_k^k\right)^2}\right] \cdot \left(\frac{\hat{\rho}_z^\ell}{\hat{\lambda}_z^\ell}\right).$$

$\hat{m}_1$ is the mean OLS coefficient, which provides information about $\beta_0 + b_0$, the sum of mean value-added and mean bias. $\hat{m}_2$ is the mean OLS coefficient among lottery schools, which helps to identify β_Q, the difference in value-added between lottery and non-lottery schools. $\hat{m}_3$ is the mean reduced form, which provides information about β_0. $\hat{m}_4$ is the mean first stage across lotteries, which can be used to estimate δ_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 ξ_0, the mean fallback utility for included schools relative to the omitted school. $\hat{m}_7$ is the average ratio of the lottery reduced form to a "pseudo-reduced form" prediction that uses OLS value-added estimates, given by $\hat{\lambda}_z^\ell = \sum_j \hat{\pi}_j^\ell \hat{\alpha}_j$. 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 β_j and b_j.

The next seven moments are variances of parameter estimates:

$$\hat{m}_8 = \tfrac{1}{J}\textstyle\sum_j \left(\hat{\alpha}_j - \bar{\alpha}\right)^2,$$

$$\hat{m}_9 = \tfrac{1}{L}\textstyle\sum_\ell \left(\hat{\rho}_z^\ell - \bar{\rho}\right)^2,$$

$$\hat{m}_{10} = \tfrac{1}{L}\textstyle\sum_\ell (\hat{\lambda}_z^\ell - \bar{\lambda})^2,$$

$$\hat{m}_{11} = \tfrac{1}{L}\textstyle\sum_\ell \left(\hat{\pi}_\ell^\ell - \bar{\pi}_{own}\right)^2,$$

$$\hat{m}_{12} = \tfrac{1}{J}\textstyle\sum_j \left[\left(\tfrac{1}{L-1}\sum_{\ell\neq j} \hat{\pi}_j^\ell\right) - \bar{\pi}_{other}\right]^2,$$

$$\hat{m}_{13} = \tfrac{1}{J}\textstyle\sum_j \left[\left(\tfrac{1}{L-1}\sum_{\ell\neq j} \frac{\hat{\pi}_j^\ell}{\hat{\pi}_\ell^\ell}\right) - \bar{s}_{other}\right]^2,$$

$$\hat{m}_{14} = \tfrac{1}{J(L-1)}\textstyle\sum_j \sum_{\ell\neq j} \left(\hat{\pi}_j^\ell - \bar{\pi}_{j,other}\right)^2.$$

Here $\bar{\alpha}$ indicates the sample average of the α_j, and similarly for other variables. $\hat{m}_8$ 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. $\hat{m}_9$ and $\hat{m}_{10}$ are variances of the lottery reduced form and predicted reduced form, which contain additional information about the joint distribution of value-added and bias. $\hat{m}_{11}$ is the variance of the first stage across lotteries, which helps to identify the variance of δ_j. $\hat{m}_{12}$ 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. $\hat{m}_{13}$ 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 ξ_j and ν_j^ℓ, which govern heterogeneity in fallback probabilities. $\hat{m}_{14}$ 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 ξ_j, the school-specific mean fallback utility, from σ_ν^2, the variance of idiosyncratic school-by-lottery utility shocks.

Finally, we match six covariances:

$$\hat{m}_{15} = \tfrac{1}{L}\textstyle\sum_\ell \left(\hat{\rho}_z^\ell - \bar{\rho}\right)\left(\hat{\lambda}_z^\ell - \bar{\lambda}\right),$$

$$\hat{m}_{16} = \tfrac{1}{L}\textstyle\sum_\ell \left(\hat{\rho}_z^\ell - \bar{\rho}\right)\left(\hat{\pi}_\ell^\ell - \bar{\pi}_{own}\right),$$

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

$$\hat{m}_{18} = \frac{1}{L}\sum_{\ell}\left(\hat{\rho}_{z}^{\ell} - \bar{\rho}\right)\left[\left(\frac{1}{L-1}\sum_{k\neq\ell}\hat{\pi}_{\ell}^{k}\right) - \bar{\pi}_{other}\right],$$

$$\hat{m}_{19} = \frac{1}{L}\sum_{\ell}\left(\hat{\alpha}_{\ell} - \bar{\alpha}\right)\left[\left(\frac{1}{L-1}\sum_{k\neq\ell}\hat{\pi}_{\ell}^{k}\right) - \bar{\pi}_{other}\right],$$

$$\hat{m}_{20} = \frac{1}{L}\sum_{\ell}\left(\hat{\pi}_{\ell}^{\ell} - \bar{\pi}_{own}\right)\left[\left(\frac{1}{L-1}\sum_{k\neq\ell}\hat{\pi}_{\ell}^{k}\right) - \bar{\pi}_{other}\right].$$

$\hat{m}_{15}$ 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 β_j and π_j^j. $\hat{m}_{17}$ is the covariance of conventional value-added and the first stage, which helps to identify the covariance between b_j and δ_j. $\hat{m}_{18}$ is the covariance of the reduced form and average fallback probability, which helps to identify the covariance of β_j and ξ_j. $\hat{m}_{19}$ is the covariance of OLS value-added with the average fallback probability, which depends on the covariance between b_j and ξ_j. $\hat{m}_{20}$ is the covariance of a school's first stage and average fallback probability, which provides information about the covariance of ξ_j and δ_j.

There are 16 elements of θ and 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 $\tilde{m}(\theta)$ be the corresponding vector of simulated predictions. The simulated minimum distance estimator with weighting matrix A is

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

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

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

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

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

Under the null hypothesis that the model's overidentifying restrictions hold and standard regularity conditions, the minimized SMD criterion function follows a χ^2 distribution (Sargan, 1958; Hansen, 1982):

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

where q is the 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, Π, is known. In this case the posterior distribution for β_j and b_j can be derived analytically. In matrix form the model can be written

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

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

$$(e_\alpha',e_\rho')|\beta,b\sim N(0,V_e),$$

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

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

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

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

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

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

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

Rearranging this expression yields

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

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

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

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

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

and

$$\beta^* = W_1(\hat{\alpha} - b_0\iota) + W_2\hat{\rho}_z + (I - W_1 - W_2\Pi)\beta_0\iota$$

with

$$W_1 = B^{-1}((v_{\alpha\alpha} + v_{bb})(v_{\alpha\alpha} + \Pi' v'_{\alpha\rho} + v_{\beta b})^{-1}(v_{\alpha\alpha} + \Pi' v'_{\alpha\rho}) - v_{\alpha\alpha}),$$

$$W_2 = B^{-1}((v_{\alpha\alpha} + v_{bb})(v_{\alpha\alpha} + \Pi' v'_{\alpha\rho} + v_{\beta b})^{-1}(v_{\alpha\rho} + \Pi' v_{\rho\rho}) - v_{\alpha\rho}),$$

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

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

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

with

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

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

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

$$\Theta = \left(\beta_1, b_1, \delta_1, \xi_1, ..., \beta_J, b_J, \delta_J, \xi_J, \nu^1_1, ..., \nu^L_J\right)'.$$

Up to a scaling constant, the posterior density for Θ conditional on the observed estimates $\hat{\Omega}$ and the prior parameters θ can be expressed

$$f_{\Theta|\hat{\Omega}}\left(\Theta|\hat{\Omega};\theta\right) \propto \phi_m\left(\hat{\Omega} - \Omega(\Theta); V\right)\phi_m\left(\Theta - \bar{\Theta}(\theta); \Gamma(\theta)\right), \tag{29}$$

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

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

Note that the probability limit of the vector of observed estimates, Ω, is a function of Θ, so we write $\Omega(\Theta)$.

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

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

Since the first stage parameters π^ℓ_j are nonlinear functions of δ and ξ, the density in (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:

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

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

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

$$\alpha^*_{j.} = \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + Var(e^\alpha_j)}\hat{\alpha}_j + \left(1 - \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + Var(e^\alpha_j)}\right)\hat{\mu}_\alpha, \tag{30}$$

where

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

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

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

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

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

	Uncontrolled (1)	Demographic (2)	Lagged score (3)	Gains (4)	Lagged score, no charter lotteries (5)
Forecast coefficient	0.358	0.660	0.864	0.722	0.423
	(0.087)	(0.130)	(0.167)	(0.172)	(0.310)
First stage *F*-statistic	33.1	27.0	26.8	29.4	14.0
p-values:					
Forecast coef. equals 1	<0.001	0.009	0.416	0.105	0.063
Overid. restrictions	0.011	0.057	0.039	0.007	0.157
All restrictions	<0.001	0.008	0.018	0.001	0.040
All restrictions (bootstrap refinement)	<0.001	<0.001	<0.001	0.002	<0.001

Notes: This table reports estimates of the VAM forecast coefficient and the results of tests for bias in conventional value-added models for 6th grade ELA scores. Estimated forecast coefficients are from regressions of 6th grade scores on fitted values from conventional 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 demographic model adds indicators for student sex, race, subsidized lunch, special education, limited-English proficiency, and counts of baseline absences and suspensions. The lagged score model adds cubic polynomials in baseline math and ELA scores. The gains model includes the same controls as the demographic model and uses score gains from baseline as the outcome. Column (5) excludes charter school lotteries from the lottery sample in testing the lagged score model. All IV models control for lottery strata fixed effects, demographic variables, and lagged scores. Standard errors are reported in parentheses. Bootstrap p-values are based on 500 Bayesian bootstrap replications (see Appendix B for details).

Table A2: Random coefficient distribution

	β_j (1)	b_j (2)	δ_j (3)	ξ_j (4)
Standard deviation	0.158	0.140	0.954	0.951
	(0.070)	(0.068)	(0.112)	(0.227)
Correlation w/β_j	1			
	-			
Correlation w/b_j	-0.480	1		
	(0.427)	-		
Correlation w/δ_j	0.119	0.361	1	
	(0.391)	(0.426)	-	
Correlation w/ξ_j	0.428	-0.522	-0.651	1
	(0.544)	(0.578)	(0.175)	-
Std. dev. of ν_{lj}		1.315		
		(0.183)		

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 score value-added model with sector effects. See notes to Table 5 for a description of the estimation procedure.