LEVERAGING LOTTERIES FOR SCHOOL VALUE-ADDED: TESTING AND ESTIMATION*

JOSHUA D. ANGRIST
PETER D. HULL
PARAG A. PATHAK
CHRISTOPHER R. WALTERS

Conventional value-added models (VAMs) compare average test scores across schools after regression-adjusting for students' demographic characteristics and previous scores. This article tests for VAM bias using a procedure that asks whether VAM estimates accurately predict the achievement consequences of random assignment to specific schools. Test results from admissions lotteries in Boston suggest conventional VAM estimates are biased, a finding that motivates the development of a hierarchical model describing the joint distribution of school value-added, bias, and lottery compliance. We use this model to assess the substantive importance of bias in conventional VAM estimates and to construct hybrid value-added estimates that optimally combine ordinary least squares and lottery-based estimates of VAM parameters. The hybrid estimation strategy provides a general recipe for combining nonexperimental and quasi-experimental estimates. While still biased, hybrid school value-added estimates have lower mean squared error than conventional VAM estimates. Simulations calibrated to the Boston data show that, bias notwithstanding, policy decisions based on conventional VAMs that control for lagged achievement are likely to generate substantial achievement gains. Hybrid estimates that incorporate lotteries yield further gains. *JEL Codes*: I20, J24, C52.

*We gratefully acknowledge funding from the National Science Foundation, the Laura and John Arnold Foundation, and the Spencer Foundation. We are indebted to SEII research managers Annice Correia and Eryn Heying for invaluable help and support. Thanks also go to Isaiah Andrews, Guido Imbens, Pat Kline, Rick Mansfield, Chris Nielson, Stephen Raudenbush, Jesse Rothstein, Doug Staiger, and seminar participants at the 2014 All California Labor Economics Conference, the APPAM Fall 2014 research conference, the 2014 and 2017 AEFP meetings, the 2015, 2016, and 2017 ASSA annual meetings, the 2015 SOLE/EALE annual meeting, the 2015 NBER Summer Institute, the Federal Reserve Bank of New York, the 2015 Becker/Friedman Applied Microeconomics Conference, the University of Chicago Committee on Education Workshop, Brown University, the University of Chicago Workshop on Quantitative Research Methods, UC Merced, the University of Arizona, Northwestern University, and the SREE Spring 2017 Conference for suggestions and comments.

The Quarterly Journal of Economics (2017), 871–919. doi:10.1093/qje/qjx001.
Advance Access publication on February 1, 2017.

I. Introduction

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

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

The proliferation of partially randomized urban school assignment systems provides a new tool for measuring school value-added. Centralized assignment mechanisms based on the theory

1. The Education Commission of the States noted that, as of 2016, fourteen states – Alabama, Arizona, Florida, Indiana, Louisiana, Maine, Mississippi, New Mexico, North Carolina, Ohio, Oklahoma, Texas, Utah, and Virginia – were issuing 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).

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 employ 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, Jacob, and Levitt 2006; Hastings and Weinstein 2008; Abdulkadiroğlu et al. 2011; Angrist, Pathak, and Walters 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.

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

Application of this test to data from Boston reveals moderate but statistically significant bias in conventional VAM estimates.

2. The theory behind VAM overidentification testing is sketched in Angrist et al. (2016b).

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

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

Finally, we quantify the consequences of bias in conventional VAM estimates and the payoff to hybrid estimation using a Monte Carlo simulation calibrated to our Boston estimates. Simulation results show that policy decisions based on conventional estimates

3. These settings include the analysis of teacher, hospital, doctor, firm, and neighborhood effects, as in Chetty, Friedman, and Rockoff (2014a, 2014b), Chandra et al. (2016), Fletcher, Horwitz, and Bradley (2014), Card, Heining, and Kline (2013), and Chetty and Hendren (2016). Chetty and Hendren combine observational and quasi-experimental estimates of neighborhood effects using a procedure discussed in Section V.

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

The next section describes the Boston data used for VAM testing and estimation, and Section III describes the conventional value-added framework as applied to these data. Section IV derives our VAM validation test and discusses test implementation and results. Section V outlines the random coefficients model and empirical Bayes approach to hybrid estimation, while Section VI reports estimates of the model's hyperparameters and the resulting posterior predictions of value-added. Section VII discusses policy simulations. Finally, Section VIII concludes with remarks on how the framework developed here might be used in other settings. All appendix material appears in the Online Appendix.

II. Setting and Data

II.A. Boston Public Schools

Boston public school students can choose from a diverse set of enrollment options, including traditional Boston Public School (BPS) district schools, charter schools, and pilot schools. As in most districts, Boston's charter schools are publicly funded but free to operate within the confines of their charters. For the most part, charter staff are not covered by collective bargaining agreements or other BPS regulations.[4] Boston's pilot school sector arose as a union-supported alternative to charter schools, developed jointly by the BPS district and the Boston Teachers Union. Pilot schools

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

are part of the district but typically have more control over their budgets, scheduling, and curriculum than do traditional public schools. On the other hand, pilot school teachers work under collective bargaining provisions similar to those in force at traditional public schools.

Applicants to traditional public and pilot schools rank between 3 and 10 schools as the first step in a centralized match (students not finishing elementary or middle school who are happy to stay where they are need not participate in the match). Applicants are then assigned to schools via a student-proposing deferred acceptance mechanism, as described in Abdulkadiroğlu et al. (2006). This mechanism combines student preferences with a strict priority ranking over students for each school. Priorities are determined by whether an applicant is already enrolled at the school and therefore guaranteed admission, has a sibling enrolled at the school, or lives in the school's walk zone. Ties within these coarse priority groups are broken by random sequence numbers, which we refer to as lottery numbers. In an evaluation of the pilot sector exploiting this centralized random assignment scheme, Abdulkadiroğlu et al. (2011) find mostly small and statistically insignificant effects of pilot school attendance relative to the traditional public school sector.

In contrast with the centralized match that assigns seats at traditional and pilot schools, charter applicants apply to individual charter schools separately before the fall of the school year in which they hope to enter. By Massachusetts law, oversubscribed charter schools must select students through 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 complete information on past admissions lotteries. Studies based on charter lotteries show that Boston charter schools boost test scores and increase four-year college attendance (see, for example, Abdulkadiroğlu et al. 2011; Angrist et al. 2016a).

II.B. 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.

Baseline test scores come from fifth-grade Massachusetts Comprehensive Assessment System (MCAS) tests in math and English language arts (ELA), while outcomes are measured in sixth, seventh, and eighth grades. 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 include school enrollment, race, sex, subsidized lunch eligibility, special education status, English-language learner status, and suspensions and absences. Online Appendix A describes the administrative files 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 below the highest (worst) number offered a seat at their first-choice school, among those in the same priority group. Conditional on application year, first-choice school, and an applicant's priority at that school (what we call the assignment strata), offers of seats at a first choice are randomly assigned. Charter lottery instruments indicate offers made on the night of the admissions lottery at each charter school. These offers are randomly assigned for nonsiblings conditional on the target school and application year.[5]

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

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

TABLE I

BOSTON STUDENTS AND SCHOOLS

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

Notes. This table counts the students included in each school in the OLS value-added and lottery samples. The samples cover cohorts attending sixth grade in Boston between the 2006–2007 and 2013–2014 school years. Columns (3) and (7) indicate schools for which sixth grade is the primary entry grade, while columns (4) and (8) indicate whether a school has enough students subject to random admission variation to be included in the lottery sample. Total numbers of schools in each sector appear in parentheses in the school type headings.

About 77% of Boston sixth graders enroll at schools with usable lotteries, and, as can be seen in the descriptive statistics reported in Table II, demographic characteristics for this group are comparable to those of the full BPS population. Columns (3) and (4) of Table II report characteristics of the subset of students subject to randomized lottery assignment. Lotteried students are slightly more likely to be African American and to qualify for a subsidized lunch, and somewhat less likely to be white or to have been suspended or absent in fifth grade. Table II also documents the comparability of students offered and not offered seats in a lottery. These results, reported in

TABLE II

DESCRIPTIVE STATISTICS

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

TABLE II
(CONTINUED)

	Means				Lottery offer balance			
	OLS sample		Lottery sample					
Baseline covariate	All students (1)	Lottery school students (2)	All students (3)	Lottery school students (4)	All lotteries (5)	Traditional (6)	Pilot (7)	Charter (8)
Math score	0.058	0.053	0.004	0.016	0.022	−0.026	0.080	0.036
					(0.024)	(0.030)	(0.061)	(0.035)
ELA score	0.030	0.006	0.013	0.016	0.035	0.045	0.060	0.013
					(0.025)	(0.030)	(0.061)	(0.036)
N	27,864	21,446	8,718	7,748	8,718	4,849	1,303	3,655

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

columns (5)–(7), compare the baseline characteristics of lottery winners and losers, controlling for assignment strata. Consistent with conditional random assignment of offers, estimated differences by offer status are small and not significantly different from zero, both overall and within school sectors.[6]

III. 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, denoted Y_{ij}, is therefore written as the sum of two noninteracting components, specifically:

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

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

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

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

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

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

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

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

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

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

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

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

This is a population projection, so the residuals in this model, 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 parameters for school j.

IV. Validating Conventional VAMs

IV.A. Test Procedure

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

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

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

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

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

To describe the overidentification test statistic, let Z denote the $N \times L$ matrix of lottery offers for a sample of N students, and let C denote the corresponding matrix of stratifying variables,

with associated projection matrix $P_C = C(C'C)^{-1}C'$ and annihilator matrix $M_C = I - P_C$. The Lagrange multiplier (LM) overidentification test statistic associated with two-stage least squares (2SLS) models estimated assuming homoskedasticity can be written:

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

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

A simple decomposition of $\hat{T}$ reveals an important connection with classical overidentification tests and previously used VAM validity tests. Let $\hat{Y}_i$ denote the fitted values generated by OLS VAM estimation (computed from regression model (5)), and let Y and $\hat{Y}$ denote $N \times 1$ vectors collecting individual Y_i and $\hat{Y}_i$. Our LM statistic can then be written

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

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

7. The test statistic in equation (8) is derived assuming homoskedastic errors. An analogous test allowing heteroskedasticity uses a White (1980) robust covariance matrix to test the hypothesis that coefficients on lottery offers equal 0 in a regression of ϵ_i on Z_i and C_i.

for testing the $L - 1$ overidentifying restrictions generated by the availability of L instruments to estimate $\hat{\varphi}$.[8]

In what follows, the estimate $\hat{\varphi}$ is called a "forecast coefficient." This connects $\hat{T}$ with tests of "forecast bias" implemented in previous VAM validation efforts (Kane and Staiger 2008; Chetty, Friedman, and Rockoff 2014a). These earlier tests similarly ask whether the coefficient on predicted value-added equals 1 in IV procedures relating outcomes to VAM fitted values (though the details sometimes differ). Forecast bias arises when VAM estimates for a group of schools are off the mark, a failure of average predictive validity. Importantly, the omnibus test statistic, $\hat{T}$, checks more than forecast bias: this statistic asks whether each oversubscribed lottery generates score gains commensurate with the gains predicted by an OLS VAM.

IV.B. Test Results

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

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

9. The gains specification can be given a theoretical foundation 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.

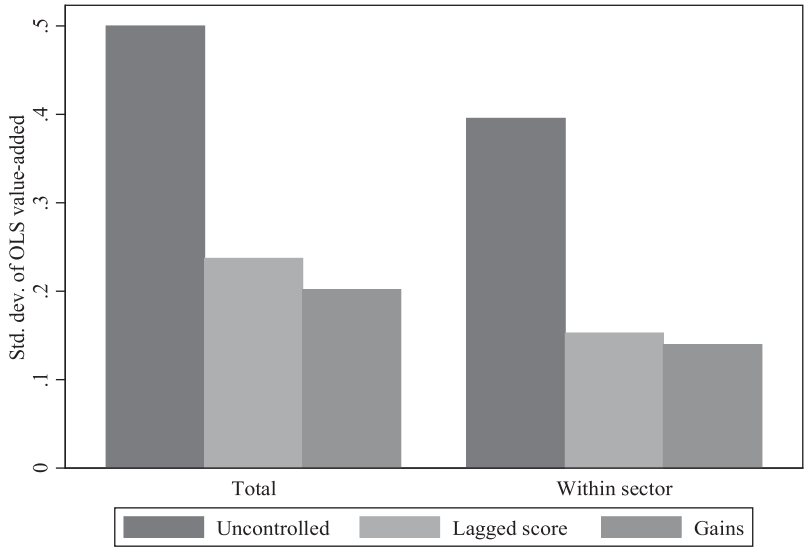

FIGURE I

Standard Deviations of School Effects from OLS Value-Added Models

This figure compares standard deviations of school effects from alternative OLS value-added models. The notes to Table III describe the controls included in the lagged score and gains models; the uncontrolled model includes only year effects. The variance of OLS value-added is obtained by subtracting the average squared standard error from the sample variance of value-added estimates. 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.

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

Figure I summarizes the value-added estimates generated by sixth-grade math scores. We focus on math scores because value-added for math appears to be more variable across schools than value-added for ELA (bias tests for ELA, presented in Online Appendix Table A.II, yield similar results). Each bar in Figure I reports an estimated standard deviation of α_j across schools,

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

expressed in test score standard deviation units and adjusted for estimation error.[11] Adding controls for demographic variables and previous scores reduces the standard deviation of α_j from 0.5σ in the uncontrolled model to about 0.2σ in the lagged score and gains models. This shows that observed student characteristics explain a substantial portion of the variation in school averages. The last three bars in Figure I report standard deviations of within-sector value-added constructed using residuals from regressions of $\hat{\alpha}_j$ on dummies for schools in the charter and pilot sectors. Controlling for sector effects reduces variation in α_j, reflecting sizable differences in average conventional value-added across sectors.

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

Table III also reports p-values for three VAM validity tests. The first is for forecast bias, that is, the null hypothesis that the forecast coefficient equals 1. The second tests the associated set of overidentifying restrictions, which require that just-identified IV estimates of the forecast coefficient be the same for each lottery instrument, though not necessarily equal to 1. The third "omnibus test" combines these restrictions.

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

12. When estimated in the same sample with no additional controls, OLS regressions on OLS fitted values necessarily produce coefficients of 1. In practice, the specification used here to test VAM differs from the model producing fitted values in that it also controls for lottery strata and excludes nonlotteried students.

TABLE III
TESTS FOR BIAS IN CONVENTIONAL VALUE-ADDED MODELS

	All lotteries		Excluding charter lotteries	
	Lagged score (1)	Gains (2)	Lagged score (3)	Gains (4)
Panel A: Sixth grade				
Forecast coefficient (φ)	0.864	0.950	0.549	0.677
	(0.075)	(0.084)	(0.164)	(0.193)
First stage F-statistic	29.6	26.6	11.2	9.3
p-values:				
Forecast bias	0.071	0.554	0.006	0.095
Overidentification	0.003	0.006	0.043	0.052
Omnibus test χ^2 statistic (d.f.)	77.7 (28)	72.1 (28)	48.0 (23)	41.7 (23)
p-value	<0.001	<0.001	<0.001	0.010
N	8,718		6,162	
Panel B: All middle school grades				
Forecast coefficient (φ)	0.880	0.924	0.683	0.726
	(0.055)	(0.060)	(0.124)	(0.133)
First stage F-statistic	14.7	15.0	7.6	7.8
p-values:				
Forecast bias	0.028	0.204	0.011	0.039
Overidentification	0.011	0.011	0.062	0.065
Omnibus test χ^2statistic (d.f.)	172.8 (75)	167.0 (75)	111.6 (60)	107.9 (60)
p-value	<0.001	<0.001	<0.001	<0.001
N	20,935		15,027	

Notes. This table reports the results of tests for bias in conventional value-added models (VAMs) for sixth- through eighth-grade math scores. The lagged score VAM includes cubic polynomials in baseline math and ELA scores, along with indicators for application year, sex, race, subsidized lunch, special education, limited English proficiency, and counts of baseline absences and suspensions. The gains VAM drops the lagged score controls and uses score growth from baseline as the outcome. Seventh- and eighth-grade VAMs measure exposure to each school using total years of enrollment since the lottery. Forecast coefficients are from instrumental variables regressions of test scores on fitted values from conventional VAMs, instrumenting fitted values with lottery offer indicators. IV models are estimated via an asymptotically efficient GMM procedure and control for assignment strata fixed effects, demographic variables, and lagged scores. The forecast bias test checks whether the forecast coefficient equals 1, and the overidentificiation test checks the IV model's overidentifying restrictions. The omnibus test combines forecast bias and overidentifying restrictions. Panel A uses sixth grade math scores, while Panel B stacks outcomes from sixth through eighth grade. Standard errors and test statistics in Panel B cluster by student. Columns (3) and (4) exclude charter school lotteries.

On average, VAM fitted values predict the score gains generated by random assignment remarkably well. This can be seen in columns (1) and (2) of Table III, which show that the lagged score and gains specifications generate forecast coefficients for 6th graders equal to 0.86 and 0.95; the former is only marginally statistically different from 1 ($p = .07$), while the second has $p = .55$. At the same time, the overidentification and omnibus tests reject for both models.[13]

13. As a point of comparison, Angrist et al. (2016b) report tests of VAM validity in the Charlotte-Mecklenburg lottery data analyzed by Deming (2014). There as well the forecast coefficient is close to 1, while the omnibus test generates a p-value of .02.

The source of these rejections can be seen in Figure II, which plots reduced-form estimates of the effects of lottery offers on test scores against corresponding first-stage effects of lottery offers on conventional VAM fitted values for sixth-grade math. Each panel also shows a line through the origin with slope equal to the forecast coefficient reported in Table III (plotted as a solid line) along with a dashed 45-degree line. In other words, Figure II gives a visual representation of the forecast coefficient: VAM models that satisfy equation (6) should generate points along the 45-degree line, with deviations due solely to sampling error. Though the lines of best fit have slopes close to 1, points for many lotteries are farther from the diagonal than sampling variance alone would lead us to expect. Earlier validation strategies focus on forecast coefficients, ignoring overidentifying restrictions. Figure II shows that such strategies may fail to detect substantial deviations between conventional VAM predictions and reduced-form lottery effects for individual lotteries.

Figure II also suggests that a good portion of conventional VAM estimates' predictive power for Boston schools comes from charter school lotteries, which contribute large first-stage and reduced-form effects. The relationship between OLS value-added and lottery estimates is weaker in the traditional public and pilot school sectors. This is confirmed in columns (3) and (4) of Table III, which report results of VAM bias tests for sets of instruments that exclude charter lotteries. At 0.55 and 0.68, estimated forecast coefficients from traditional public and pilot lotteries are farther from 1 than the coefficients computed using all lotteries. Although removal of charter lotteries reduces precision, omnibus tests computed without them also reject at the 1% level.[14]

Finally, Panel B of Table III reports test results combining data from sixth through eighth grade. As in Abdulkadiroğlu et al. (2011) and Dobbie and Fryer (2013), school effects on seventh- and eighth-grade scores are modeled as linear in the number of years spent in each school. In a linear constant-effects framework,

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

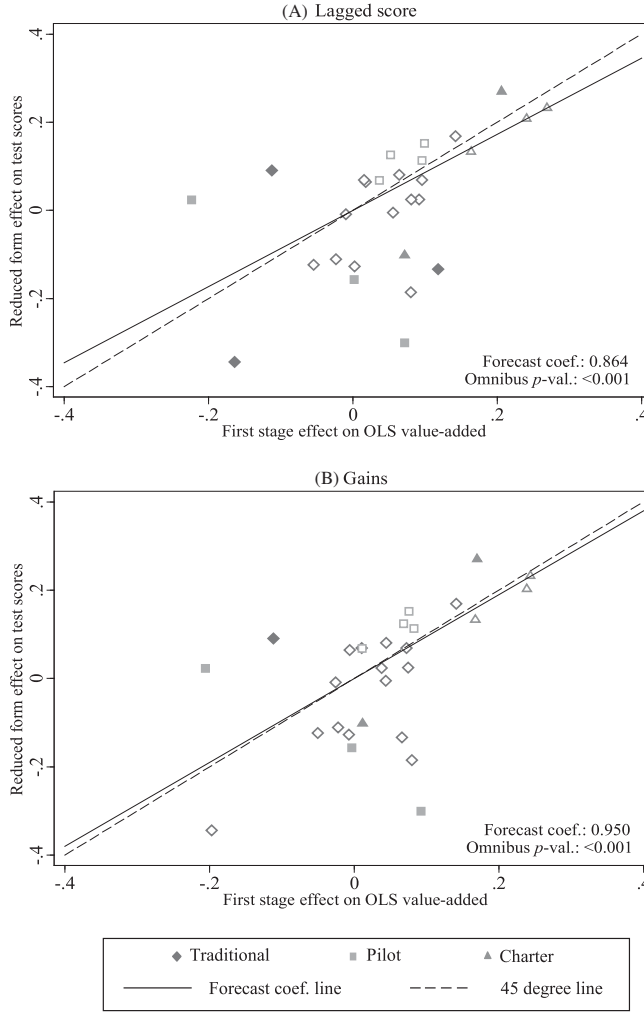

FIGURE II

Visual Instrumental Variables Tests for Bias

This figure plots lottery reduced-form estimates against value-added first stages from each of 28 school admission lotteries. The notes to Table III describe the underlying models. Filled markers indicate reduced form and first stage estimates that are significantly different from each other at the 10% level. The solid lines have slopes equal to the forecast coefficients in Table III, while dashed lines indicate the 45-degree line. Omnibus p-values are for the overidentification test statistic described in Section IV.A.

regressions of test score outcomes on baseline controls and years of enrollment in each school recover causal school effects in the absence of sorting on unobserved ability. The omnibus VAM validity test in this case regresses residuals from the multigrade (stacked) model on sixth-grade lottery offers, while the forecast coefficient is generated by using lottery offers to instrument OLS VAM fitted values from the multigrade model. The omnibus test results show clear rejections in the multigrade setup as well as for sixth grade only, in spite of the fact that the first-stage F-statistics here are noticeably lower.

IV.C. Heterogeneity versus Bias

The omnibus test results reported in Table III suggest conventional VAM estimates fail to predict the effects of lottery offers perfectly. This is consistent with bias in OLS VAMs. In a world of heterogeneous causal effects, however, these rejections need not reflect selection bias. Rather, these results might 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, Imbens, and Rubin 1996). Moreover, with unrestricted potential outcomes, even internally valid OLS VAM estimates (that is, those satisfying selection-on-observables) capture weighted average causal effects that need not match average effects for the entire sample of students attending particular schools (Angrist 1998).

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

Estimates by subgroup, reported in Panel A of Table IV for the OLS VAM sample, consistently generate rejections in omnibus tests of VAM validity. Column (2) shows test results computed

TABLE IV

ROBUSTNESS OF SIXTH-GRADE BIAS TESTS TO EFFECT HETEROGENEITY

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

Notes. This table reports lottery-based tests for bias in school value-added models that allow for effect heterogeneity by baseline characteristics. The notes to Table III describe the underlying models. Panel A shows results for the full OLS sample, while Panel B shows results for the lottery subsample. Column (1) reports estimates that do not allow effect heterogeneity, while columns (2)–(6) are from models allowing value-added to differ across groups defined by the covariates in the column headings. The covariates used to define groups in column (6) are race, gender, subsidized lunch, special education, English language learner status, and baseline score terciles based on average fifth-grade math and ELA test scores in the OLS sample.

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

Lottery-based IV estimates identify average causal effects for compliers, that is, for lottery applicants whose attendance choices shift in response to random offers, rather than for the full population of students that enroll in a particular school. To investigate the link between lottery compliance and treatment effects, we predict value-added at the target school for individual lottery applicants using covariate-specific OLS estimates from the model in column (6) of Table IV (estimated in the lottery sample). Maintaining the hypothesis of OLS VAM validity, we allow for the possibility that heterogeneous effects are reflected in a set of covariate-specific estimates. These predictions are then used to compare imputed average value-added for compliers to imputed average value-added for "never-takers" (those who decline lottery offers) and "always-takers" (those who enroll in the target school even when denied an offer) in each lottery. Averages for the three

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

lottery compliance groups are estimated using methods described in Online Appendix B.1.

Figure III shows that imputed OLS value-added estimates for compliers, always-takers, and never-takers are similar. Formal tests for equality fail to reject the hypotheses that predicted effects for compliers equal predicted effects for always-takers ($p = .80$) or never-takers ($p = .33$). This suggests that lottery compliance is not a major source of treatment effect heterogeneity, though we cannot rule out unobserved differences between compliers and other groups.

V. The Distribution of School Effectiveness

The test results in Table III suggest conventional VAM estimates are biased. At the same time, OLS VAM estimates tend to predict lottery effects on average, with estimated forecast coefficients close to 1. OLS estimates would therefore seem to be useful even if imperfect. This section develops a hybrid estimation strategy that combines lottery and OLS estimates in an effort to quantify the bias in conventional VAMs and produce more accurate value-added estimates.

V.A. A Random Coefficients Lottery Model

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

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

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

Note that Z_i is the vector of all lottery admissions offers, $Z_{i\ell}$ for $\ell = 1, \ldots, L$. Assumption (7) implies that the reduced-form effect of admission in lottery ℓ is given by

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

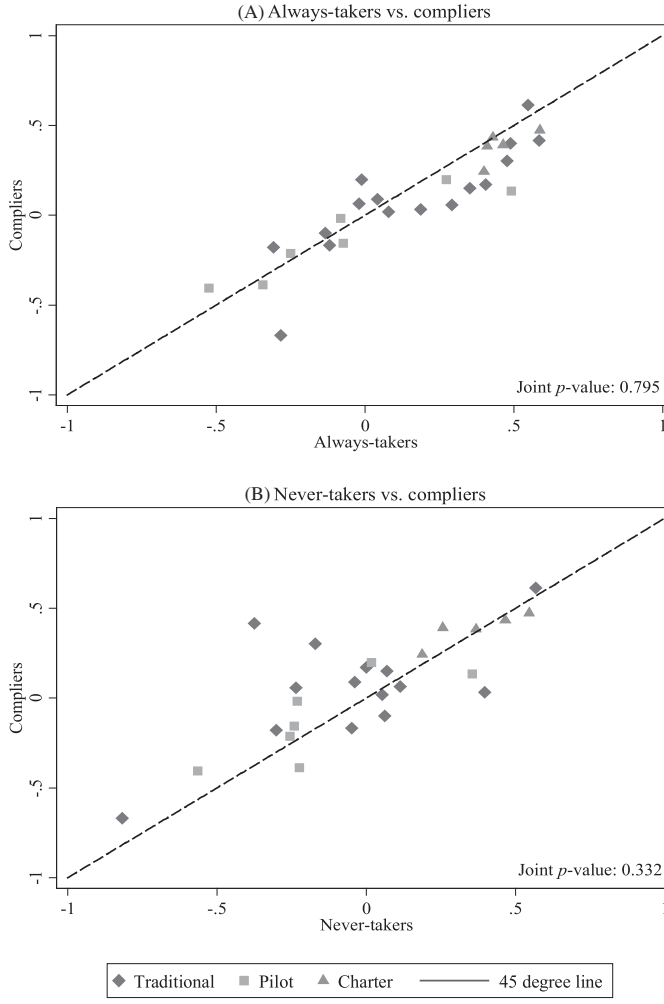

FIGURE III

Comparisons of Conventional Value-Added by Lottery Compliance

This figure compares OLS estimates of average value-added for admission lottery compliers to estimates for always- and never-takers in each of 28 school lotteries. OLS estimates come from a lagged-score VAM that allows school effects to differ across the subgroups used in column (6) of Table IV, estimated in the lottery sample. Complier, always-taker, and never-taker means are estimated using methods described in Online Appendix B. *p*-values are for joint tests of complier and always/never-taker equality across all schools. The *p*-value for a test that pools the estimates in both panels is .289.

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

OLS VAM, lottery reduced-form, and lottery 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 a larger population of schools. Specifically, we have:

$$\begin{aligned} \hat{\alpha}_j &= \beta_j + b_j + e_j^{\alpha}, \\ \hat{\rho}_\ell &= \sum_j \pi_{\ell j}\beta_j + e_\ell^{\rho}, \\ \hat{\pi}_{\ell j} &= \pi_{\ell j} + e_{\ell j}^{\pi}, \end{aligned} \tag{11}$$

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

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

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

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

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

In this expression, the quantity $\xi_j + \nu_{\ell j}$ is the mean utility for school j in a multinomial logit model predicting alternative school choices among students that comply with offers made in lottery ℓ. The parameter ξ_j allows for the possibility that some schools are systematically more or less likely to serve as fallback options for lottery losers, while $\nu_{\ell j}$ is a random utility shock specific to school j in the lottery at school ℓ. The parameterization in equations (12) and (13) ensures that lottery offers increase the probability of enrollment at the target school and reduce enrollment probabilities at other schools, and that effects on all probabilities are between 0 and 1 in absolute value.

Each school is characterized by a vector of four parameters: a value-added coefficient, β_j; a selection bias term, b_j; an offer compliance utility, δ_j; and a mean fallback utility, ξ_j. These are modeled as draws from a prior distribution in a hierarchical Bayesian framework. A key assumption in this framework is that the distribution of VAM bias is the same for schools with and without oversubscribed lotteries. This assumption allows the model to "borrow" information from schools with lotteries and to generate posterior predictions for nonlottery 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 VI.B investigates the empirical relationship between oversubscription and bias.

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

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

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

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

The vector $\theta = (\beta_0, \beta_Q, b_0, \delta_0, \xi_0, \text{vec}(\Sigma), \sigma_\nu^2)'$ collects the hyperparameters governing the prior distribution of school-specific

parameters. Our empirical Bayes framework first estimates these hyperparameters and then uses the estimated prior distribution to compute posterior value-added predictions for individual schools. Some of the specifications considered below extend the setup outlined here to allow the prior mean vector $(\beta_0, b_0, \delta_0, \xi_0)$ to vary across Boston's school sectors (traditional, charter, and pilot).

V.B. Simulated Minimum Distance Estimation

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

The fact that the moments in this context are complicated nonlinear functions of the hyperparameters motivates a simulation approach. For example, the mean reduced form is $E[\rho_\ell] = \sum_j E[\pi_{\ell j}\beta_j]$. This is the expectation of the product of normally distributed random variables (the β_j) with ratios (the elements of π_j) 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 (14) and (15). Likewise, the simulation draws a vector of the estimation errors in equation (11) from the joint asymptotic distribution of the OLS, reduced-form, and first-stage estimates. The parameter and estimation draws are combined to generate a simulated vector of parameter estimates for the given value of θ. Finally, these are used to construct a set of model-based predicted moments. The SMD estimator minimizes a quadratic form that weights differences between predicted moments and the corresponding moments observed in the data. As described in Online Appendix B.2, the SMD estimates reported here are generated by a two-step procedure with an efficient weighting matrix in the second step.

V.C. Empirical Bayes Posteriors

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

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

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

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

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

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

where ι is a $J \times 1$ vector of ones. Posterior mean value-added is a linear combination of OLS estimates net of mean bias, $(\hat{\alpha} - b_0\iota)$, lottery reduced-form estimates, $\hat{\rho}$, and mean value-added, $\beta_0\iota$. The weighting matrices, W_α and W_ρ, are functions of the first-stage parameters and the covariance matrix of estimation error,

value-added, and bias. Expressions for these matrices appear in Online Appendix B.3. As with conventional EB posteriors, an empirical Bayes version of the posterior mean plugs first-step estimates of b_0, β_0, W_α, and W_ρ into equation (17).

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

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

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

In related work, Chetty and Hendren (2016) combine noisy quasi-experimental estimates of neighborhood effects based on movers with relatively precise averages of permanent resident outcomes to generate optimal forecasts of neighborhood causal effects. A further special case of equation (18) illuminates the link between this approach and ours. Suppose the estimation error in OLS estimates is negligible ($Var(e_j^\alpha) = 0$), and that IV estimation error, e_j^β, is uncorrelated across schools. Online Appendix B.3 shows that under these simplifying assumptions, the jth element of equation (18) becomes

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

$$\times \left(r_\alpha(\hat{\alpha}_j - b_0) + (1 - r_\alpha)\beta_0\right), \tag{19}$$

where σ_β^2 is the variance of β_j, $r_\alpha = \frac{Cov(\beta_j,\alpha_j)}{Var(\alpha_j)}$ is the slope (also known as the reliability ratio) from a regression of causal value-added on OLS value-added, and R^2 is the R-squared from this regression. This expression coincides with equation (9) in Chetty and Hendren (2016) and can also be seen to be the same as the

canonical empirical Bayes shrinkage formula expressed by equation (1.5) of Morris (1983).[16]

In practice, some schools are undersubscribed, so IV estimates of individual school value-added cannot be computed. Nevertheless, equation (17) shows that predictions at schools without lotteries can be improved using lottery information from other schools. Lottery reduced-form parameters embed information for all fallback schools, including those without lotteries. This is a consequence of the relationship described by equation (11), which shows that the reduced form for any school with a lottery depends on the value-added of all other schools that applicants to this school might attend. Specifically, as long as $\pi_{\ell j} \neq 0$, the reduced form for lottery ℓ contains information that can be used to improve the posterior prediction of β_j. The test results in columns (2) and (5) of Table V show that estimates of $\pi_{\ell j}$ are significantly different from zero (at the 5% level) for 12 of the 22 undersubscribed schools in our sample. The 10 schools not on this list have primary entry grades other than sixth. In other words, oversubscribed sixth-grade lotteries contribute information on all schools with sixth-grade entry.

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

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

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

16. The connection with Morris can be made by observing that when $\hat{\alpha}_j = \alpha_j$, the term $r_{\alpha}(\hat{\alpha}_j - b_0) + (1 - r_{\alpha})\beta_0$ is the fitted value from a regression of β_j on α_j.

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

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

This formulation of the conventional EB estimand adds bias, b_0, to a weighted average of bias-corrected OLS and global mean value-added.

TABLE V

FALLBACK STATUS OF SCHOOLS WITHOUT SIXTH-GRADE LOTTERIES

Number of lottery losers enrolled (1)	p-value: not a lottery fallback (2)	Sixth-grade entry (3)	Number of lottery losers enrolled (4)	p-value: not a lottery fallback (5)	Sixth-grade entry (6)
Panel A: Traditional public schools			Panel C: Charter schools		
39	0.013	Y	320	<0.001	Y
36	0.018	Y	11	0.080	
113	<0.001	Y	16	0.427	
79	<0.001	Y	16	0.204	
94	<0.001	Y	24	0.724	
21	0.045	Y	42	0.145	
60	0.006	Y	3	0.111	
12	0.016		2	0.390	
Panel B: Pilot schools			10	0.257	
5	0.033	Y	33	0.010	
15	0.169		112	<0.001	
			34	0.378	

Notes. This table reports p-values for tests of whether each nonlottery school in the OLS sample serves as a fallback for one of the 28 lottery schools. Columns (1) and (4) count the number of students in the lottery sample who are observed enrolling in the undersubscribed school when not given an offer. Columns (2) and (5) report p-values from tests of the hypothesis that the undersubscribed school's first-stage coefficients are 0 in all lotteries with such students. Columns (3) and (6) indicate whether sixth grade is a school's primary entry point. First-stage regressions control for assignment strata indicators, demographic variables, and lagged test scores.

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

VI. Parameter Estimates

VI.A. Hyperparameters

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

As can be seen in columns (1)–(3) of Table VI, the hyperparameter estimates reveal substantial variation 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 uncontrolled specification to 0.22σ in the lagged score and gains models. This stability is reassuring: the control variables that distinguish these models should not change the underlying distribution of causal school effectiveness if our estimation procedure works as we hope.

TABLE VI

ESTIMATES OF THE JOINT DISTRIBUTION OF CAUSAL VALUE-ADDED AND VAM BIAS FOR SIXTH-GRADE MATH SCORES

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

Notes. This table reports simulated minimum distance estimates of parameters of the joint distribution of causal school value-added and OLS bias for sixth-grade math scores. The moments used in estimation are functions of OLS value-added, lottery reduced-form, and first-stage estimates, as described in Online Appendix B. Uncontrolled estimates come from an OLS regression that includes year effects. The notes to Table III describe the other value-added models. Simulated moments are computed from 500 samples constructed by drawing school-specific parameters from the random coefficient distribution along with estimation errors based on the asymptotic covariance matrix of the estimates. The estimates in columns (4) and (5) are from models allowing the means of the random coefficients distribution to depend on school sector. Moments are weighted by an estimate of the inverse covariance matrix of the moment conditions, calculated from a first-step estimate using an identity weighting matrix. The weighting matrix is produced using 1,000 simulations, drawn independently from the samples used to simulate the moments.

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

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

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

Estimated covariances between β_j and b_j, denoted $\sigma_{\beta b}$, are negative and mostly statistically significant, a result that can be seen in the third row of Table VI. A negative covariance between value-added and bias suggests that conditional on demographics and past achievement, students with higher ability tend to enroll in schools with lower value-added. Conventional VAMs therefore overestimate the effectiveness of low-quality schools and underestimate the effectiveness of high-quality schools. Estimates of β_Q, the lottery school value-added shifter, are close to 0 in models without sector effects, and positive but small when sector effects are included. The estimate of β_Q for the lagged score model is statistically significant, implying that schools with lotteries are slightly more effective than undersubscribed schools in the same sector.

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

VI.B. School Characteristics, Value-Added, and Bias

The individual school value-added posterior modes generated by our hybrid estimation strategy are positively correlated with conventional posterior means that ignore bias in OLS value-added estimates. This is evident in Figure IV, which plots hybrid modes against posterior means from conventional value-added models. Rank correlations in the lagged score and gains models are 0.79 and 0.74. The relationship between conventional and hybrid posteriors is weaker for lottery schools (indicated by filled markers)

19. Chetty, Friedman, and Rockoff (2014a) use this parameter to define "forecast bias," equal to $1 - r_\alpha$. We use the term "reliability" here to distinguish between r_α and the forecast coefficient, $\hat{\varphi}$. Online Appendix B.5 discusses the relationship between $\hat{\varphi}$ and r_α.

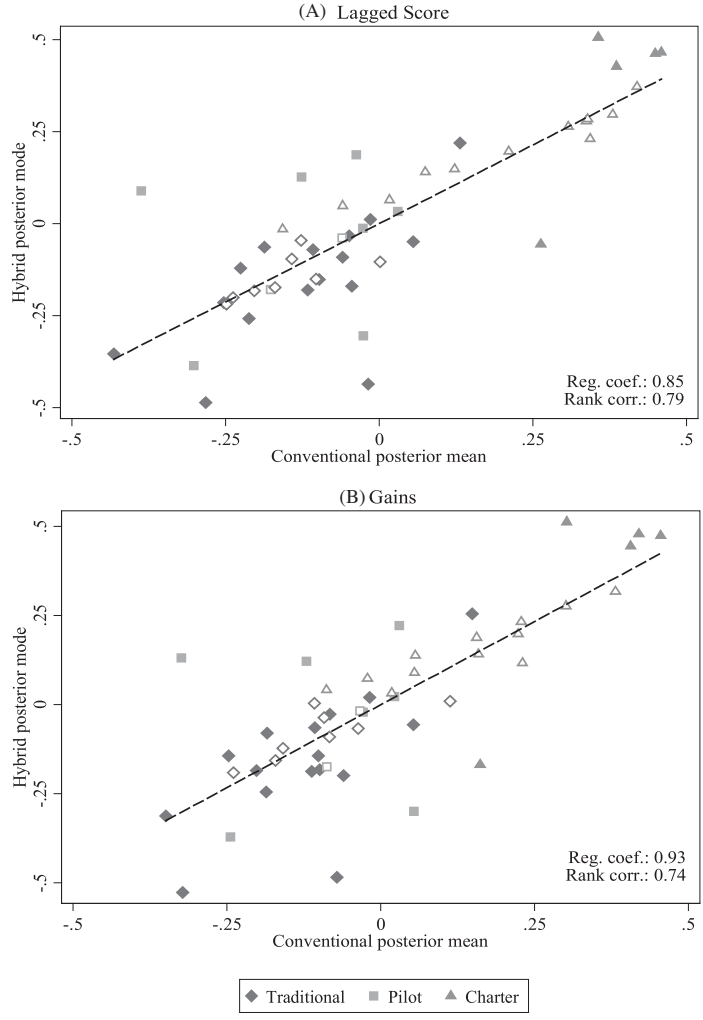

FIGURE IV

Empirical Bayes Posterior Predictions of School Value-Added

This figure plots empirical Bayes posterior modes of value-added from the hybrid 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 toward the mean in proportion to 1 minus the signal-to-noise ratio. Filled markers indicate lottery schools. Dashes indicate OLS regression lines linking the two sets of estimates.

TABLE VII
CORRELATES OF POSTERIOR VALUE-ADDED AND VAM BIAS

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

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

than for schools without lotteries: rank correlations for these two groups equal 0.60 and 0.90 in the gains model. This reflects the fact that lotteries are more informative about causal effects for schools with randomized admission. Importantly, although hybrid and conventional posteriors are strongly correlated, hybrid estimation changes some schools' ranks, so accountability decisions may be improved using the hybrid estimates.

Hybrid estimation generates posterior modes for bias as well as value-added. The value-added and bias posteriors therefore permit an exploration of the associations between school characteristics, causal value-added, and bias. Table VII reports coefficients from regressions of posterior modes for bias and value-added on school characteristics, with and without controls for sector. As can be seen in columns (1) and (3), students who appear more advantaged (as measured by baseline scores and special education status, for example) tend to enroll in schools with higher value-added, but this pattern is largely explained by the higher likelihood that these students enroll in charter schools. By

contrast, column (4) shows that VAM bias within sectors is more positive for schools with more advantaged students, including those with higher average baseline test scores, fewer black students, fewer special education students, and fewer students eligible for subsidized lunches. The correlation of bias with baseline scores is especially noteworthy: although we see positive selection into the Boston charter sector, the popular impression that good schools have good peers is driven mostly by selection bias.

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

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

Recall that Table II shows that baseline scores and other observed characteristics are similar for students enrolled at schools with and without lotteries. The bottom of Figure V explores this pattern further by showing that oversubscription rates are uncorrelated with average baseline scores at oversubscribed schools. A regression of average baseline scores on log oversubscription produces a coefficient of −0.03 with a standard error of 0.10. This finding, which does not rely on estimates from the model, shows

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

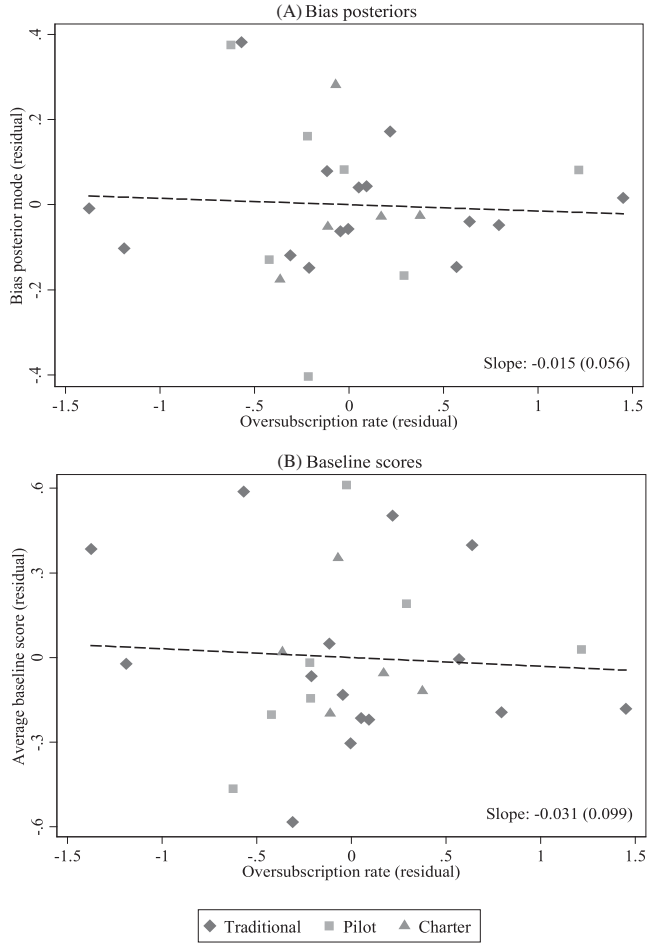

FIGURE V

Relationship between Oversubscription and Bias for Lottery Schools

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

that the observed ability of enrolled students is unrelated to lottery oversubscription within the lottery sample. We might therefore expect unobserved ability to be unrelated to oversubscription as well. Both panels of Figure V support the assumption postulating similar bias distributions for schools that are more and less heavily oversubscribed.[21]

VII. Policy Simulations

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

VII.A. Mean Squared Error

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

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

22. Simulation results for seventh and eighth grade, reported in Online Appendix Tables A.IV and A.V, yield conclusions similar to those for sixth grade. These and other supplementary simulation results are discussed in Online Appendix C.

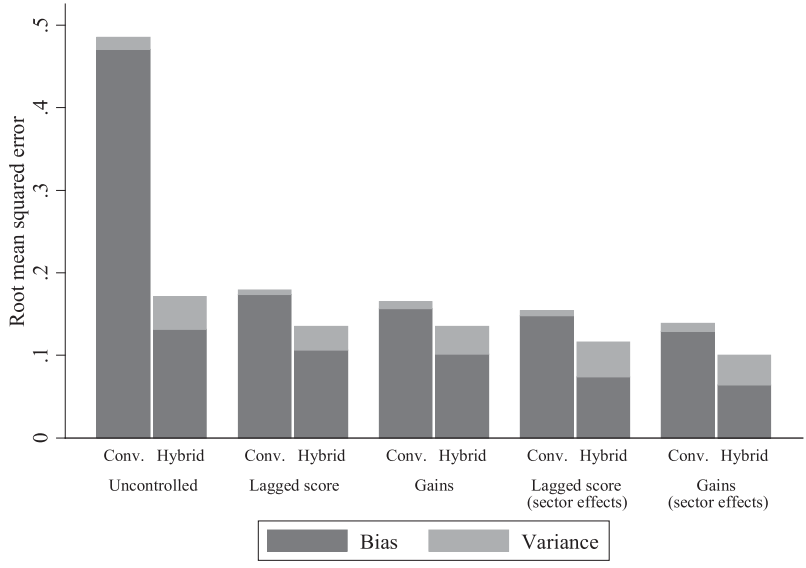

FIGURE VI

Root Mean Squared Error for Value-Added Posterior Predictions

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

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

Not surprisingly, the RMSE reduction yielded by the hybrid estimator reflects reduced bias at the cost of increased sampling

variance. This can be seen by writing the mean squared error of an estimator, β_j^*, as

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

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

VII.B. Consequences of School Closure

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

This analysis ignores possible transition effects such as disruption due to school closure, peer effects from changes in school composition, and other factors that might inhibit replication of successful schools. The results should nevertheless provide a rough guide to the potential consequences of VAM-based policy decisions. Quasi-experimental analyses of charter takeovers and other school reconstruction efforts in Boston, New Orleans, and Houston have shown large gains when low-performing schools are replaced by schools operating according to pedagogical principles seen to be effective elsewhere (Fryer 2014; Abdulkadiroğlu et al. 2016). This suggests transitional consequences are dominated by

23. Online Appendix Table A.VI shows that hybrid estimates generate forecast coefficients close to 1 in both the lagged score and gains specifications, with or without charter lotteries. The hybrid estimates also pass the overidentification and omnibus specification tests.

TABLE VIII

CONSEQUENCES OF CLOSING THE LOWEST-RANKED DISTRICT SCHOOL FOR AFFECTED STUDENTS

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

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

longer-run determinants of school quality, at least for modest policy interventions of the sort considered here.

The potential for VAMs to guide decision making is highlighted by the first row of Table VIII, which shows the score gains produced by decisions based on true value-added. Closing the worst school and replacing it with an average school boosts achievement by 0.37σ, while more targeted replacement policies generate even larger gains. Consistent with the high RMSE of uncontrolled estimates, however, Table VIII also shows that policies based on uncontrolled test score levels generate only small gains. For example, replacing the lowest-scoring district school with an average school is predicted to increase scores for affected students by 0.06σ on average. Likewise, a policy that replaces the lowest-ranked school with an average top quintile school

generates a gain of 0.10σ. These small effects reflect the large variation in bias evident for the uncontrolled model in Table VI: closure decisions based on average test scores target schools with many low achievers rather than low value-added. The bias in uncontrolled VAM estimates also leads to a wide dispersion of simulated closure effects, with a cross-simulation standard deviation (reported in brackets) of around 0.2σ.

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

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

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

The results in Table VIII show that even when VAM estimates are imperfect, they predict causal value-added well enough

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

VIII. Conclusions and Next Steps

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

Hybrid school value-added estimation requires some kind of lottery-based admissions scheme, such as those increasingly used for student assignment in many large urban districts in the United States. As our analysis of Boston's multiple-offer charter sector shows, however, admissions need not be centralized for lotteries to be of value. The utility of hybrid estimation in other cities will vary with the extent of lottery coverage, but results for Boston

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

show hybrid estimation remains useful even when lottery data are missing for many schools. Our approach also ignores effect heterogeneity linked to school choices, a limitation that may matter in settings with more specialized schools and very heterogeneous student populations.

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

MIT AND NBER
MIT
MIT AND NBER
UC BERKELEY AND NBER

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

REFERENCES

Abdulkadiroğlu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, and Parag A. Pathak, "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots," *Quarterly Journal of Economics*, 126 (2011), 699–748.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Peter D. Hull, and Parag A. Pathak, "Charters without Lotteries: Testing Takeovers in New Orleans and Boston," *American Economic Review*, 106 (2016), 1878–1920.

Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak, "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, forthcoming.

Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez, "Changing the Boston School Choice Mechanism," NBER Working Paper 11965, 2006.

Angrist, Joshua D., "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66 (1998), 249–288.

Angrist, Joshua D., Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters, "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice," *Journal of Labor Economics*, 34 (2016a), 275–318.

Angrist, Joshua, Peter Hull, Parag Pathak, and Christopher Walters, "Interpreting Tests of School VAM Validity," *American Economic Review: Papers & Proceedings*, 106 (2016b), 388–392.
Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91 (1996), 444–455.
Angrist, Joshua D., Parag A. Pathak, and Christopher R. Walters, "Explaining Charter School Effectiveness," *American Economic Journal: Applied Economics*, 5 (2013), 1–27.
Bloom, Howard S., and Rebecca Unterman, "Can Small High Schools of Choice Improve Educational Prospects for Disadvantaged Students?," *Journal of Policy Analysis and Management*, 33 (2014), 290–319.
Card, David, Jörg Heining, and Patrick Kline, "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128 (2013), 967–1015.
Chamberlain, Gary, and Guido Imbens, "Random Effects Estimators with Many Instrumental Variables," *Econometrica*, 72 (2004), 295–306.
Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector," *American Economic Review*, 106 (2016), 2110–2144.
Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan, "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR," *Quarterly Journal of Economics*, 126 (2011), 1593–1660.
Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104 (2014a), 2593–2632.
——, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, 104 (2014b), 2633–2679.
——, "Using Lagged Outcomes to Evaluate Bias in Value-Added Models," *American Economic Review: Papers & Proceedings*, 106 (2016), 393–399.
——, "Measuring the Impacts of Teachers: Reply to Rothstein (2016)" *American Economic Review*, forthcoming.
Chetty, Raj, and Nathaniel Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates," NBER Working Paper 23002, 2016.
Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt, "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica*, 74 (2006), 1191–1230.
Deming, David J., "Using School Choice Lotteries to Test Measures of School Effectiveness," *American Economic Review: Papers & Proceedings*, 104 (2014), 406–411.
Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger, "School Choice, School Quality, and Postsecondary Attainment," *American Economic Review*, 104 (2014), 991–1013.
Deutsch, Jonah, "Proposing A Test of the Value-Added Model Using School Lotteries," Working paper, 2013.
Dobbie, Will, and Roland G. Fryer Jr., "Getting Beneath the Veil of Effective Schools: Evidence from New York City," *American Economic Journal: Applied Economics*, 5 (2013), 28–60.
——, "The Medium-Term Impacts of High-Achieving Charter Schools," *Journal of Political Economy*, 123 (2015), 985–1037.
Fletcher, Jason M., Leora I. Horwitz, and Elizabeth Bradley, "Estimating the Value Added of Attending Physicians on Patient Outcomes," NBER Working Paper 20534, 2014.
Fryer, Roland G. Jr."Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments," *Quarterly Journal of Economics*, 129 (2014), 1355–1407.

Hastings, Justine S., and Jeffrey M. Weinstein, "Information, School Choice, and Academic Achievement: Evidence from Two Experiments," *Quarterly Journal of Economics*, 123 (2008), 1373–1414.
Imbens, Guido W., and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62 (1994), 467–475.
Jacob, Brian A., and Lars Lefgren, "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education," *Journal of Labor Economics*, 26 (2008), 101–136.
Judge, George G., and Ron C. Mittelhammer, "A Semiparametric Basis for Combining Estimation Problems under Quadratic Loss," *Journal of the American Statistical Association*, 99 (2004), 479–487.
——, "Estimation and Inference in the Case of Competing Sets of Estimating Equations," *Journal of Econometrics*, 138 (2007), 513–531.
Kane, Thomas J., Daniel F. McCaffrey, Trey Miller, and Douglas O. Staiger, "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment," Gates Foundation Report, 2013.
Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger, "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City," *Economics of Education Review*, 27 (2008), 615–631.
Kane, Thomas J., and Douglas O. Staiger, "The Promise and Pitfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives*, 16 (2002), 91–114.
——, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," NBER Working Paper 14607, 2008.
Kinsler, Josh, "Assessing Rothstein's Critique of Teacher Value-Added Models," *Quantitative Economics*, 3 (2012), 333–362.
Koedel, Cory, and Julian R. Betts, "Does Student Soring Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique," *Education Finance and Policy*, 6 (2011), 18–42.
McFadden, Daniel, "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57 (1989), 995–1026.
Mittelhammer, Ron C., and George G. Judge, "Combining Estimators to Improve Structural Model Estimation and Inference under Quadratic Loss," *Journal of Econometrics*, 128 (2005), 1–29.
Morris, Carl N., "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78 (1983), 47–55.
Newey, Whitney K., "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29 (1985), 229–256.
Newey, Whitney K., and Kenneth D. West, "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28 (1987), 777–787.
Rothstein, Jesse, "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables," *Education Finance and Policy*, 4 (2009), 537–571.
——, "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *Quarterly Journal of Economics*, 125 (2010), 175–214.
——, "Revisiting the Impact of Teachers," *American Economic Review*, forthcoming.
Sargan, J. D., "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26 (1958), 393–415.
Walters, Christopher R., "The Demand for Effective Charter Schools," NBER Working Paper 20640, 2014.
White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48 (1980), 817–838.
——, "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50 (1982), 483–499.