The Review *of* Economics *and* Statistics

VOL. CVI JANUARY 2024 NUMBER 1

CREDIBLE SCHOOL VALUE-ADDED WITH UNDERSUBSCRIBED SCHOOL LOTTERIES

Joshua Angrist, Peter Hull, Parag A. Pathak, and Christopher Walters*

Abstract—We introduce two empirical strategies harnessing the randomness in school assignment mechanisms to measure school value-added. The first estimator controls for the probability of school assignment, treating take-up as ignorable. We test this assumption using randomness in assignments. The second approach uses assignments as instrumental variables (IVs) for low-dimensional models of value-added and forms empirical Bayes posteriors from these IV estimates. Both strategies solve the underidentification challenge arising from school undersubscription. Models controlling for assignment risk and lagged achievement in Denver and New York City yield reliable value-added estimates. Estimates from models with lower-quality achievement controls are improved by IV.

I. Introduction

POLICYMAKERS and families increasingly rely on achievement-based measures of school quality to make high-stakes decisions. Families use school quality information to decide where to enroll and—in some cases—where to live. School leaders and policymakers use measured quality when deciding whether to close, restructure, or expand schools.[1] A common concern with such quality rankings is selection bias due to the nonrandom sorting of students to schools. In a parallel development, a growing number of school districts use centralized, algorithmic assignment schemes to match students and schools. Boston, Denver, and New York City (NYC), for instance, use deferred acceptance (DA) algorithms to assign students to seats.[2] Many of these centralized assignment systems incorporate random lottery numbers to break ties between otherwise similar students. A growing econometric literature shows how the resulting randomness in seat assignment can be used to address selection bias.

This paper introduces two new empirical strategies that exploit randomness in algorithmic school assignment to measure individual school quality. The key to both approaches is a vector of school assignment propensity scores that characterize each student's probability of assignment to each school. In general, the propensity score for treatment assignment is the probability of assignment conditional on a vector of confounding variables; Rosenbaum & Rubin (1983) show that treatments that are independent of potential outcomes conditional on potential confounders are also independent of potential outcomes conditional on the propensity score. Abdulkadiroğlu et al. (2017, 2022) extend this result to matching markets for schools, deriving formulas that quantify assignment risk in centrally assigning districts.

Empirical work exploiting school assignment propensity scores has so far aimed to capture causal effects of attendance in particular sectors, such as charter schools, rather than individual schools. Estimates of such effects are useful for understanding average sectoral effectiveness, but high-stakes decisions for households and policymakers typically hinge on measures of individual school quality. Our aim here is to use school assignment propensity scores to estimate individual school value-added. An important econometric challenge in this context arises because many schools are "undersubscribed," in the sense of having no applicants for whom assignment risk is strictly between 0 and 1. A conventional two-stage least squares (2SLS) model that uses offers to instrument individual school attendance is therefore underidentified. Moreover, many oversubscribed schools face weak demand and, consequently, yield a weak first stage. This paper tackles the econometric issues arising in such common school assignment scenarios.

Received for publication January 13, 2021. Revision accepted for publication October 6, 2021. Editor: Brian A. Jacob.

*Angrist: MIT and NBER; Hull: Brown University and NBER; Pathak: MIT and NBER; Walters: University of California Berkeley and NBER.

Our thanks to Jimmy Chin and Raymond Han for outstanding research assistance and to MIT SEII program managers Eryn Heying and Anna Vallee for invaluable administrative support. We thank Jesse Rothstein and seminar participants at Berkeley and the Online Causal Inference Seminar for helpful comments. Financial support from Arnold Ventures/LJAF and the National Science Foundation is gratefully acknowledged. P.P. also thanks the W. T. Grant Foundation for research support. The research described here was carried out under data use agreements between the New York City and Denver public school districts and MIT. We are grateful to these districts for sharing data.

A supplemental appendix is available online at https://doi.org/10.1162/rest_a_01149.

[1]The Education Commission of the States noted that as of 2021, 49 U.S. states and Washington, DC, included or plan to include a measure of average student achievement growth in their school accountability systems for elementary and middle school grades; 27 states further use or plan to use such measures in high school accountability systems. These growth measures are analogous to the conventional value-added models we study in this paper, which adjust student test scores by lagged achievement.

[2]Other cities with centralized school assignment systems include Baltimore, Cambridge (Massachusetts), Camden (New Jersey), Chicago, Indianapolis, Minneapolis, Newark, New Orleans, Oakland, San Francisco, Seattle, Tulsa, and Washington, DC. Centralized assignment is also widespread and growing globally, with 36 countries using it at either the primary or secondary level as of 2020 (see https://www.ccas-project.org/).

The Review of Economics and Statistics, January 2024, 106(1): 1–19

https://doi.org/10.1162/rest_a_01149

Our first empirical strategy presumes that the only sources of selection bias in value-added estimates are the applicant characteristics integral to school matching, such as where to apply and the priority status that a school assigns its applicants. This amounts to the assumption that compliance with conditionally randomized offers is independent of potential outcomes. We refer to estimation based on this empirical strategy as a *risk-controlled value-added model* (RC VAM). The conditional independence assumption underlying RC VAM echoes that invoked in the Dale & Krueger (2002, 2014) and Mountjoy & Hickman (2021) studies of the earnings consequences of elite college attendance. Importantly, however, centralized school assignment facilitates tests of such conditional independence assumptions, a feature new to applications of this sort of identification strategy.

Our second estimator avoids the conditional independence assumption motivating RC VAM by using randomized school offers as instrumental variables (IVs) for school attended, adjusting for the school assignment propensity scores. Score conditioning makes offer dummies credible instruments, but undersubscription makes a conventional IV approach intractable. Our *instrumental variables value-added model* (IV VAM) procedure solves the underidentification problem by modeling school value-added as a function of a few mediating school characteristics. IV estimates of mediator effects are then used to construct empirical Bayes posterior predictions of value-added for individual schools. This approach builds on and simplifies earlier empirical Bayes strategies that combine observational and quasi-experimental estimates (Angrist et al., 2016, 2017; Chetty & Hendren, 2018; Hull, 2018).

The RC VAM and IV VAM procedures are illustrated by estimating school value-added in Denver and NYC, two large urban districts with centralized public school assignment. Denver matches students to schools in a unified enrollment scheme that employs a single random lottery number as a tie-breaker across students with the same admission priorities. We estimate the Denver school assignment propensity score using recent theoretical results on school assignment risk in such systems (Abdulkadiroğlu et al., 2017). In NYC, admission offers are determined by a match that combines lottery and nonlottery tie-breakers, with the latter relevant for applicants to New York's "screened schools." NYC school assignment scores are therefore estimated using the theoretical results on assignment risk in matching markets with mixed multiple tie-breaking developed in Abdulkadiroğlu et al. (2022).

VAM estimates from both cities suggest that controlling for school assignment risk and lagged student achievement does a remarkably good job of eliminating selection bias. Statistical tests that exploit quasi-experimental offer variation fail to reject the key conditional independence assumption underlying RC VAM. Perhaps surprisingly, RC VAM estimates appear reliable even for the effects of NYC high schools on SAT scores, where lagged score controls come from very different assessments. RC VAM estimates for NYC also beat conventional VAM estimates on mean-squared error (MSE) grounds, with bias reductions more than offsetting the variance penalty for use of narrower identifying variation.

Because RC VAM appears virtually unbiased, there's little scope for IV VAM to improve on it in Denver and NYC. We show, however, that IV VAM can reduce bias and MSE when value-added is estimated with more limited controls than those specified by the RC VAM procedure. IV VAM also boosts the accuracy of VAMs that rely on older measures of lagged achievement for estimation. This scenario is inspired by the COVID-19 pandemic, when few school districts were able to test as originally scheduled. Our encouraging results in this context suggest that IV VAM is likely to be useful in situations where information on lagged outcomes is dated or unavailable.

Our analysis adds to a large and growing literature on value-added measurement in education and other fields. Statistical analyses of U.S. school quality date back to Coleman (1966), which famously concluded that schools "bring little influence to bear on a child's achievement." More recent studies leverage advanced statistical methods to separate the causal effects of schools (or teachers) from the confounding factors of student background and ability (Ladd & Walsh, 2002; Tekwe et al., 2004; Kane & Staiger, 2008; Reardon & Raudenbush, 2009). Key extensions and applications exploiting quasi-experimental variation include Chetty et al. (2014a) for teachers and Deming (2014) and Angrist et al. (2017) for schools. Applications of quasi-experimental value-added estimation to other settings include Chetty and Hendren (2018), Hull (2018), and Abaluck et al. (2021).[3]

The remainder of the paper is organized as follows. Section II outlines a conceptual framework for school value-added estimation and testing in districts that centralize school assignment. Sections III and IV describe our two new estimation strategies. Empirical results are discussed in section V. Section VI summarizes and points to directions for further work.

II. Econometric Framework

A. Setting

Consider a population of N students, each attending one of J schools in a district. Let Y_{ij} denote the potential test score outcome of student i when enrolled at school j. These potential outcomes are described by a constant-effects model.

$$Y_{ij} = \gamma_j + \varepsilon_i, \quad j \in \{1, \dots, J\}, \tag{1}$$

where $\gamma_j = E[Y_{ij}]$ and $\varepsilon_i \equiv Y_{ij} - E[Y_{ij}]$. For any two schools, j and k, the difference $\gamma_j - \gamma_k$ gives the causal effect of attending j rather than k, that is, a comparison of school value-added. Equation (1) is a constant-effects VAM because the common residual, ε_i, implies the same $Y_{ij} - Y_{ik}$ for all students. The constant-effects framework focuses attention on

[3]See also Andrabi et al. (2020) and Ainsworth et al. (2023), who apply the methods of Angrist et al. (2017) to school systems outside the United States.

the possibility of selection bias in VAM estimates rather than treatment effect heterogeneity (though we explore the potential for such heterogeneity below). We refer to ε_i as student i's *ability*, since this reflects an individual student's contribution to her academic achievement.

It's convenient to adopt a parameterization that measures value-added relative to the average for the district. Let D_{ij} denote an indicator equal to 1 if student i attends school j. Observed outcomes Y_i can then be written as

$$Y_i = \beta_0 + \sum_{j=1}^{J} \beta_j D_{ij} + \varepsilon_i, \tag{2}$$

where $\beta_0 = \frac{1}{J}\sum_{j=1}^{J} \gamma_j$ is the average potential achievement in the district and $\beta_j = \gamma_j - \beta_0$.

The value-added parameters β_j describe the effect of randomly generated changes in school enrollment on test score outcomes. Since school enrollment is not randomly assigned, these causal parameters need not coincide with differences in average student outcomes across schools. Schools that attract higher-ability students tend to have better average outcomes, regardless of value-added. In the context of equation (2), this selection bias manifests as correlation between the D_{ij} and ε_i. Consequently, a regression of Y_i on the D_{ij} indicators need not identify the parameters of the causal model equation (2).

B. *Centralized Assignment*

Centralized school assignment schemes ("matches," for short) provide a source of identifying information that we use to overcome selection bias. Centralized matches ask students to submit rank-ordered preferences over schools, while also granting potential applicants a priority at each school; for example, siblings of enrolled students typically receive a higher priority than nonsiblings. Students are matched to schools by a deferred acceptance (DA) algorithm that takes preferences and priorities as inputs. Match-generated offers are also determined by a tie-breaking variable, often randomly assigned, that distinguishes between students with the same preferences and priorities. The DA algorithm outputs a single school assignment for each student. Students may choose to enroll where matched or to enter a later, less systematic round producing negotiated assignments. Random tie-breaking ensures that admission offers are randomly assigned conditional on student preferences and priorities. Because many students enroll where offered, dummies indicating school offers are highly correlated with the school enrollment dummies in equation (2).

The stochastic features of a match are formalized with the aid of notation for preferences and priorities. Let $\succ_i$ denote student i's list of preferences over schools and write ρ_{ij} for student i's priority at school j. The vector $\rho_i = (\rho_{i1}, \ldots, \rho_{iJ})'$ collects a student's priorities at all schools. A student's *type*, denoted θ_i, is the combination of preferences and priorities, that is, $\theta_i \equiv (\succ_i, \rho_i)$. The centralized assignment algorithm takes as inputs the set of θ_i and random tie-breaking numbers; it outputs a set of indicators, Z_{ij}, equal to 1 when the match offers student i a seat at school j. Assignment indicators are collected in the vector $Z_i = (Z_{i1}, \ldots, Z_{iJ})'$. The conditional random assignment of Z_i is then summarized as follows:

Assumption CRA. Student ability is independent of school assignments conditional on student type: $\varepsilon_i \perp\!\!\!\perp Z_i \mid \theta_i$.

Assumption CRA, maintained throughout this paper, suggests that we can estimate the causal effects of school assignment by comparing the outcomes of students receiving different assignments within strata defined by θ_i. In practice, however, we see nearly as many preference and priority combinations as there are students in a match. Consequently, full type conditioning leaves few degrees of freedom for empirical analysis.

To reduce the dimension of the conditioning set needed to control omitted variables bias, students of different types are pooled in a manner that preserves conditional independence of offers and potential outcomes. Pooling relies on the *school assignment propensity score*, computation of which uses theoretical results in Abdulkadiroğlu et al. (2017, 2022). Student i's propensity score for assignment to school j is defined as

$$p_{ij} \equiv \Pr(Z_{ij} = 1 | \theta_i).$$

The p_{ij} are school assignment rates implicitly determined by repeatedly running the assignment algorithm, redrawing tie-breakers each time, while holding preferences and priorities fixed. Abdulkadiroğlu et al. (2017, 2022) show how to compute p_{ij} analytically, an approach we follow here. The vector $p_i = (p_{i1}, \ldots, p_{iJ})'$ collects student i's propensity scores for all schools.[4]

As first shown by Rosenbaum & Rubin (1983), random assignment conditional on a vector of controls implies conditional random assignment given the propensity score obtained from these controls. In our setup, this result can be stated as follows:

Lemma 1. *Under assumption CRA, student ability is independent of school assignments conditional on assignment risk:* $\varepsilon_i \perp\!\!\!\perp Z_i \mid p_i$.

In other words, since school assignment is ignorable conditional on type, it's also ignorable conditional on the school assignment propensity score. Moreover, assignment scores are determined by a few key match parameters. Conditioning on low-dimensional propensity scores thus leaves us with far more degrees of freedom than full type conditioning.

In a linear constant-effects model, score control can be implemented using the following corollary to lemma 1:

[4]The NYC match is complicated by the participation of schools using nonlottery tie-breaking variables like test scores. Propensity scores in this case, derived in Abdulkadiroğlu et al. (2022), turn on a set of tie-breaker classification variables as well as on type. Our notation ignores this complication, which affects estimation but is immaterial to the conceptual framework that allows us to identify VAM.

Corollary 1. *Under assumption CRA, student ability is orthogonal to risk-adjusted school assignments:* $E[\varepsilon_i(Z_{ij} - p_{ij})] = 0$ *for each* j.

This suggests that the set of risk-adjusted offer dummies, $Z_{ij} - p_{ij}$, can be used as instruments for the set of enrollment dummies, D_{ij}, in equation (2). The argument behind this result parallels that in Robinson's (1988) partially linear framework, applied here to IV. A two-stage least squares (2SLS) procedure that uses offers as instruments for school enrollment while controlling for the corresponding propensity scores is equivalent to a 2SLS estimator that uses residuals from a regression of offers on school assignment scores as instruments. Since $p_{ij} = E[Z_{ij} \mid \theta_i]$, the auxiliary regression residual in this case is $Z_{ij} - p_{ij}$.[5]

C. The Undersubscription Challenge

Abdulkadiroğlu et al. (2017, 2022) use orthogonality conditions analogous to those described by corollary 1 to estimate the effects of attending groups of schools that belong to a particular sector, such as charter schools. This work treats assignment to a given sector as an instrument for sector enrollment, controlling for the relevant sector-assignment propensity score. When the sector of interest consists of charter schools, for example, risk adjustment is based on the probability of being assigned a seat at any charter school in the match.

Using centralized assignment to estimate individual school value-added is more challenging. The challenge arises from the fact that many schools are undersubscribed, in the sense of having no applicants for whom assignment risk is strictly between 0 and 1. When assignment risk is 0 or 1, risk-adjusted school-assignment offers are constant at 0, and so uninformative about a school's causal effects. Undersubscription of individual schools is seen more often than undersubscription of an entire sector. Thus, unlike models of, say, charter school effects (a sector of interest), equation (2) is not identified solely by IV-type exclusion restrictions since the number of endogenous variables exceeds the number of useful instruments. The next two sections introduce strategies to address this challenge.[6]

[5]Proof of this and other key identification results appears in the appendix. The supplemental appendix contains technical material related to parameter estimation and additional empirical results. Borusyak and Hull (forthcoming) apply the risk adjustment idea to a broad class of research designs (and outside linear constant-effects models) that use treatments or instruments computed from multiple sources of variation.

[6]We use "undersubscribed" to refer to schools for which all applicants have risk of either 0 or 1. This label comes from the fact that much of this degenerate risk is explained by nonapplication: in the sample of undersubscribed NYC high schools; for example, 85% of school-applicant pairs have $p_{ij} = 0$. Undersubscription of this sort often arises when a school is ranked first by fewer applicants than the school has seats. In this case, those ranking the school have $Z_{ij} = p_{ij} = 1$, while nonapplicants have $Z_{ij} = p_{ij} = 0$. With centralized assignment, the $Z_{ij} = p_{ij} = 0$ scenario may also occur when all students ranking j are guaranteed to find seats at schools they prefer to j.

III. Risk-Controlled VAM

Our first approach to the estimation of causal value-added eschews the use of instruments and instead controls for assignment risk. This strategy is predicated on the assumption that conditional on applicant type and a vector of other preassignment covariates X_i, the identity of a student's enrolled school is independent of their potential outcomes and is therefore as good as randomized. The X_i include conventional VAM controls such as lagged test scores and student demographic characteristics. To formalize this identifying assumption, let $D_i = (D_{i1}, \ldots, D_{iJ})'$ be a vector collecting the J school enrollment indicators. This notation is used to state a key conditional independence assumption as follows:

Assumption CIA. Student ability is independent of school enrollment conditional on student characteristics and assignment risk: $\varepsilon_i \perp\!\!\!\perp D_i \mid (p_i, X_i)$.

The risk-controlled (RC) VAM estimator, presented below, leverages this assumption via a regression of outcomes on school attendance dummies with controls for p_i and X_i.

To motivate the RC VAM approach, note first that because school assignment offers are randomly assigned conditional on type, it must be true that $(\varepsilon_i, X_i) \perp\!\!\!\perp Z_i \mid \theta_i$. Hence, lemma 1 can be modified to say

$$\varepsilon_i \perp\!\!\!\perp Z_i \mid (p_i, X_i). \tag{3}$$

Assumption CIA therefore follows from lemma 1 in a scenario in which all match participants accept any offer yielded by the match, so that $D_i = Z_i$. More generally, assumption CIA overcomes selection bias by requiring compliance with conditionally randomized school assignment offers (i.e., the relationship between D_i and Z_i) to be as good as random conditional on applicant type and covariates. Under this assumption, value-added can be measured for all schools in the district, without regard to undersubscription. Conventional VAM estimates invoke a stronger version of assumption CIA by requiring conditional independence to hold conditional on X_i alone. The identifying assumption behind RC VAM therefore nests that justifying conventional VAM.

Assumption CIA is a central-assignment analog of the Dale & Krueger (2002, 2014) identification strategy for the returns to college selectivity. The Dale and Krueger approach assumes that college enrollment decisions are made independent of potential outcomes conditional on student application choices and college admission offers. In particular, these studies argue that application choices and admission results are likely to capture stable and systematic features of applicants' preferences and qualifications, while subsequent offer take-up decisions reflect idiosyncratic variation unrelated to potential outcomes.[7] College application choices are a

[7]Mountjoy and Hickman (2020) apply this strategy to estimate effects of college selectivity using administrative data from Texas. Abdulkadiroğlu et al. (2020) use a similar strategy to identify the determinants of parental preferences for NYC schools.

decentralized analog of the preference component of applicant type in a centralized match, while admission offer take-up is analogous to offer compliance.

This analogy suggests a close connection between the RC VAM and Dale and Krueger identification strategies, a relationship formalized in the following result:

Lemma 2. *Given* $(\varepsilon_i, X_i) \perp\!\!\!\perp Z_i \mid \theta_i$, *assumption CIA is implied by conditional independence of student ability and enrollment given student characteristics, assignment risk, and school assignments:* $\varepsilon_i \perp\!\!\!\perp D_i \mid (p_i, X_i, Z_i) \Longrightarrow \varepsilon_i \perp\!\!\!\perp D_i \mid (p_i, X_i)$.

At first blush, the Dale and Krueger strategy adds an extra set of conditioning variables: the admissions offers, Z_i. Lemma 2 shows, however, that our CIA assumption is implied by a Dale-and-Krueger-type conditional independence assumption that adds school assignment offers to the conditioning set. Offer conditioning is therefore unnecessary.

We implement RC VAM using ordinary least squares (OLS) regressions of the form

$$Y_i = \alpha_0 + \sum_j \alpha_j D_{ij} + X_i'\Gamma + g(p_i) + \eta_i, \tag{4}$$

where X_i is a conventional VAM control vector and the function $g(p_i)$ parameterizes control for assignment risk. This function includes linear terms in the elements of p_i and a set of dummy variables indicating when each p_{ij} equals 0. The resulting parsimonious specification of $g(p_i)$ has $2J$ terms in a district with J schools.

In a model with offer vector Z_i on the right-hand side, control for a linear function of p_i is equivalent to estimation with dummies for all points of support in the assignment score distribution (this follows from multivariate regression algebra and the definition of p_i). By the propensity score theorem, therefore, linear score control eliminates omitted variables bias arising from the relationship between offers and applicant type. We're interested in the effects of school enrollment rather than offer effects, however. Dummies for zero risk add another nonlinear control for applicant type that may mitigate bias in enrollment effects. Control for zero-risk dummies is motivated by the fact that over 99% of those with zero risk of an offer did not apply. As in Dale & Krueger (2002), application behavior seems likely to be a potential source of omitted variables bias in comparisons by school enrollment.

Given a set of RC VAM coefficient estimates $\hat{\alpha}_j$ (normalized to be mean-zero, as with the β_j), we construct linear shrinkage estimates of individual school quality using the formula

$$\alpha_j^* = \lambda_j \hat{\alpha}_j, \ \lambda_j = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + s_j^2}, \tag{5}$$

where σ_α^2 is the variance of the RC VAM coefficients and s_j^2 is the sampling variance of $\hat{\alpha}_j$. Relative to the OLS estimates $\hat{\alpha}_j$, the α_j^* have lower mean squared error (MSE) as predictors of the population coefficients, α_j. In particular, α_j^* coincides with the posterior mean for α_j under the assumption that the α_j are normally distributed across schools and independent of s_j and gives an MSE-minimizing linear approximation to the posterior mean outside of normality (Morris, 1983). An empirical Bayes (EB) posterior prediction substitutes an estimate of λ_j into equation (5). We estimate s_j^2 with the squared standard error of $\hat{\alpha}_j$ and estimate σ_α^2 by subtracting the average s_j^2 estimate from the sample variance of $\hat{\alpha}_j$.

Testing the CIA with centralized assignment. The α_j in equation (4) coincide with causal value-added parameters β_j when the controls in equation (4) are sufficient to eliminate selection bias. This leads to the following null hypothesis for the validity of RC VAM and other OLS value-added estimators:

$$H_0 : \alpha_j = \beta_j; \quad j \in \{1, \ldots, J\}.$$

A simple example shows how randomized offers can be used to test H_0. Suppose $J = 2$ and we renormalize intercepts to make β_1 and α_1 measure causal and OLS value-added of school 1 relative to school 2. Suppose also that every student faces the same risk of assignment to school 1, so that p_{i1} is constant, and omit the additional VAM controls, X_i. Under H_0, the OLS residual, η_i, is then equal to student ability, ε_i. Corollary 1 therefore implies that the offer Z_{i1} is mean-independent of η_i:

$$E[\eta_i|Z_{i1} = 1] - E[\eta_i|Z_{i1} = 0] = 0.$$

Substituting ε_i for η_i in this expression using a simplified equation (4) yields

$$(E[Y_i|Z_{i1} = 1] - \alpha_1 E[D_{i1}|Z_{i1} = 1]) \\ - (E[Y_i|Z_{i1} = 0] - \alpha_1 E[D_{i1}|Z_{i1} = 0]) = 0.$$

Rearranging and using the fact that $\alpha_1 = E[Y_i|D_{i1} = 1] - E[Y_i|D_{i1} = 0]$, we have

$$\frac{E[Y_i|Z_{i1} = 1] - E[Y_i|Z_{i1} = 0]}{E[D_{i1}|Z_{i1} = 1] - E[D_{i1}|Z_{i1} = 0]} = E[Y_i|D_{i1} = 1] \\ - E[Y_i|D_{i1} = 0].$$

The left side of this expression is an IV estimand using Z_{i1} to instrument D_{i1}. A test of H_0 that checks orthogonality of η_i and Z_{i1} is therefore a Hausman (1978) test for equality of school 1 enrollment effects estimated by IV and OLS.

To extend this test to the general version of equation (4), let L be the number of schools with nondegenerate propensity scores. Note that even when $L < J$, null hypothesis H_0 implies the orthogonality of η_i and all risk-adjusted school offers. This orthogonality yields L testable restrictions of the form

$$E[\eta_i(Z_{i\ell} - p_{i\ell})] = 0; \quad \ell \in \{1, \ldots, L\}, \tag{6}$$

where η_i denotes the RC VAM residual in equation (4).

The restrictions in equation (6) are tested by asking whether $\tau_1 = \cdots = \tau_L = 0$ in the residual regression equation

$$\hat{\eta}_i = \tau_0 + \sum_{\ell=1}^{L} \tau_\ell Z_{i\ell} + \sum_{\ell=1}^{L} \mu_\ell p_{i\ell} + X_i'\Delta + \xi_i, \tag{7}$$

where ℓ indexes dummies, $Z_{i\ell}$, indicating offers at $L < J$ oversubscribed schools. This is a regression version of the Sargan test for instrument-error orthogonality (detailed in Hausman, 1983). The error term here, however, is generated by OLS estimates of the RC VAM model rather than by 2SLS, as for the original Sargan test. Propensity score controls on the right-hand side of equation (7) ensure that this procedure tests risk-adjusted offer-residual orthogonality.[8] Covariates are not needed for test validity (since the X_i are conditionally uncorrelated with the $Z_{i\ell}$), but their inclusion may reduce the residual variance in equation (7), thereby increasing test power.

It's worth noting that even with $L < J$, test statistics based on equation (7) may detect bias in any of the J OLS VAM coefficients, α_j. To see this, note that a specification test based on equation (6) implicitly asks whether effects of randomized offers on test scores equal effects of offers on OLS-predicted value-added (an equivalence detailed in Angrist et al., 2016, and illustrated in the empirical analysis that follows).[9] Because offers of a seat at an oversubscribed school may shift enrollment at any school, whether oversubscribed or not, the two reduced forms in question are likely to be aligned only when all OLS VAM estimates are unbiased.[10]

The omnibus specification test based on equation (7) admits a useful decomposition that distinguishes VAM bias on average from the bias of VAM estimates for specific schools. This decomposition, introduced by Angrist et al. (2016), builds on a regression linking OLS and causal value-added coefficients,

$$\beta_j = \varphi\alpha_j + \nu_j, \tag{8}$$

where ν_j is defined so as to be uncorrelated with α_j in the population of schools. The parameter φ is a forecast coefficient summarizing the degree to which RC VAM parameters predict causal value-added. The corresponding forecast residuals, ν_j, have variance denoted σ_ν^2. In the absence of selection bias, $\alpha_j = \beta_j$, so $\varphi = 1$ and $\sigma_\nu^2 = 0$. More generally, φ summarizes the reliability of RC VAM predictions. At the same time, σ_ν^2 characterizes idiosyncratic school-specific biases that average to 0 over schools.

Assuming the residual ξ_i is conditionally homoskedastic, the omnibus test statistic evaluating $\tau_1 = \cdots = \tau_L = 0$ in equation (7) can be written as the sum of a test of forecast bias and a test of idiosyncratic bias. The omnibus test statistic is

$$\widehat{T} = \frac{(Y - D\hat{\alpha})' P_{Z_\perp}(Y - D\hat{\alpha})}{\hat{\sigma}_\xi^2},$$

where Y is a vector collecting observations of Y_i, D is a matrix collecting observations of D_i, $\hat{\alpha}$ is a vector collecting the OLS estimates of α_j, $\hat{\sigma}_\xi^2 = \frac{1}{N}\sum_i \hat{\xi}_i^2$ estimates the variance of ξ_i, and $P_{Z_\perp}$ is the projection matrix for the set of oversubscribed assignments Z after partialing out controls (that is, $P_{Z_\perp} = Z_\perp(Z_\perp' Z_\perp)^{-1} Z_\perp'$, where $Z_\perp = Z - C(C'C)^{-1}C'Z$, with matrix C collecting observations on risk controls and covariates on the right-hand side of equation [7]). Appendix A.3 shows that this test statistic can be decomposed as

$$\widehat{T} = \frac{(\hat{\varphi} - 1)^2}{\hat{\sigma}_\xi^2(\hat{\alpha}' D' P_{Z_\perp} D\hat{\alpha})^{-1}} + \frac{(Y - D\hat{\alpha}\hat{\varphi})' P_{Z_\perp}(Y - D\hat{\alpha}\hat{\varphi})}{\hat{\sigma}_\xi^2}, \tag{9}$$

where $\hat{\varphi} = (\hat{\alpha}' D' P_{Z_\perp} D\hat{\alpha})^{-1}(\hat{\alpha}' D' P_{Z_\perp} Y)$ is the 2SLS estimate of the forecast coefficient computed by using oversubscribed offer dummies to instrument RC VAM estimates of value-added at the school attended by student i.

The first term in the decomposition of $\widehat{T}$ is a Wald statistic testing $\varphi = 1$ (the denominator of this term estimates the variance of $\hat{\varphi}$). The second term is the Sargan (1958) statistic testing overidentifying restrictions induced by the 2SLS procedure generating $\hat{\varphi}$. The distinction between these tests is illuminated by substituting the forecast regression, equation (8), into the causal model, equation (2), to obtain

$$Y_i = \beta_0 + \varphi\alpha_{j(i)} + \varepsilon_i + \nu_{j(i)}, \tag{10}$$

where $\alpha_{j(i)} = \sum_j D_{ij}\alpha_j$ is the OLS VAM coefficient for student i's school and $\nu_{j(i)} = \sum_j D_{ij}\nu_j$ is the corresponding forecast residual.

By corollary 1, risk-adjusted school offers, $Z_{ij} - p_{ij}$, are orthogonal to ε_i. Under H_0, we also have $\nu_{j(i)} = 0$. Together, these restrictions imply that the $Z_{i\ell} - p_{i\ell}$ are valid instruments for $\alpha_{j(i)}$ in equation (10). The first term of equation (9) asks whether the estimated forecast coefficient yielded by these instruments is indeed statistically indistinguishable from 1. Paralleling a Sargan test of instrument-error orthogonality, the second term asks whether the individual IV estimates computed using one instrument at a time are statistically indistinguishable from one another (whether they're

[8]This residual is formed using the OLS estimates $\hat{\alpha}_j$ with no shrinkage. Shrinkage is relevant for obtaining minimum-MSE (and possibly biased) predictions of individual school quality rather than testing for bias.

[9]This interpretation follows by observing that $E[\eta_i(Z_{i\ell} - p_{i\ell})] = E[Y_i(Z_{i\ell} - p_{i\ell})] - E[\sum_j \alpha_j D_{ij}(Z_{i\ell} - p_{i\ell})]$, since risk-adjusted offers are orthogonal to X_i. Testing $E[\eta_i(Z_{i\ell} - p_{i\ell})] = 0$ is therefore equivalent to testing whether $E[Y_i(Z_{i\ell} - p_{i\ell})] = E[\sum_j \alpha_j D_{ij}(Z_{i\ell} - p_{i\ell})]$.

[10]Suppose, for example, there are three schools, A, B, and C. Let A be the reference school and suppose that only B is oversubscribed. OLS predicted value-added is $\hat{Y}_i = \alpha + \alpha_B D_{iB} + \alpha_C D_{iC}$. Offers of a seat at B increase enrollment at B and reduce enrollment at C. Bias in α_C therefore induces divergence between the reduced-form effect of an offer at B on Y_i and the corresponding reduced-form effect on $\hat{Y}_i$, even when α_B is unbiased.

equal to 1 or not). Maintaining orthogonality of the instruments with ε_i, this amounts to a test of $\nu_{j(i)} = 0$ since $\nu_{j(i)} \neq 0$ typically makes one-at-a-time IV estimates diverge. Note also that the omnibus test and associated decomposition detailed here apply to OLS estimates other than RC VAM, including conventional VAM estimates that control for X_i alone.

IV. IV VAM in Underidentified Models

The randomness induced by centralized assignment identifies key features of the value-added distribution even when OLS VAM estimates are biased. Our IV VAM estimator exploits this variation by including RC VAM or other OLS estimates in a low-dimensional model of school quality. Because systematic variation in quality is presumed to flow through only a few school characteristics, this model is identified even in the face of widespread undersubscription. The resulting parameter estimates can then be used to predict value-added for individual schools. As in Angrist et al. (2017), these predictions can be seen as optimally weighted combinations of OLS estimates and IV reduced forms. We improve on earlier efforts in this spirit by combining RC VAM and IV estimates in a computationally attractive linear framework.

IV VAM is derived from first-stage and reduced-form regressions of school attendance and outcomes on risk-adjusted offers for admission to oversubscribed schools:

$$D_{ij} = \phi_j + \sum_{\ell=1}^{L} \pi_{\ell j}(Z_{i\ell} - p_{i\ell}) + u_{ij}, \quad j \in \{1, \ldots, J\}, \quad (11)$$

and

$$Y_i = \kappa + \sum_{\ell=1}^{L} \rho_\ell (Z_{i\ell} - p_{i\ell}) + \omega_i. \quad (12)$$

Reduced-form equation (12) is obtained by substituting the first stages described by equation (11) into the causal model, equation (2).

Matrix notation highlights the identification problem raised by undersubscription. Array the $\pi_{\ell j}$ coefficients in equation (11) in the $L \times J$ matrix Π, and collect the ρ_ℓ coefficients in equation (12) in the $L \times 1$ vector ρ. We then have $\rho = \Pi\beta$, where $\beta = (\beta_1, \ldots, \beta_J)'$. When $L = J$, school value-added is identified by solving $\beta = \Pi^{-1}\rho$. When some schools are undersubscribed, $L < J$ and the first stage and reduced form alone are insufficient to identify the causal effects of interest.

This identification problem is tackled here by modeling the relationship between school quality and a lower-dimensional set of mediating variables. We show below how this model, once estimated, can be used to produce individual school quality estimates. Specifically, we extend the simple school-level forecast regression, equation (8), to a more general model for school value-added,

$$\beta_j = M_j'\varphi + \nu_j, \quad (13)$$

where $M_j = (M_{j1}, \ldots, M_{jK})'$ is a $K \times 1$ vector of school characteristics, normalized to have mean zero across schools, with K assumed to be much smaller than L. These value-added mediators may include RC VAM or conventional OLS VAM parameters, as well as other school characteristics like indicators for school sector. The forecast coefficient, φ (now a vector), captures the relationship between mediators and causal value-added, β_j. The forecast residual, ν_j, is defined to be mean zero and uncorrelated with M_j across schools. Substituting equation (13) into equation (2) yields a generalization of equation (10),

$$Y_i = \beta_0 + M_{j(i)}'\varphi + \varepsilon_i + \nu_{j(i)}, \quad (14)$$

where $M_{j(i)} = \sum_j D_{ij} M_j$ is the vector of mediators associated with student i's enrolled school and $\nu_{j(i)} = \sum_j D_{ij}\nu_j$ is the associated forecast residual.

Our IV VAM procedure estimates φ using risk-adjusted school offers as instruments for $M_{j(i)}$ in equation (14). In contrast with the testing procedure detailed in section III, IV VAM allows for multiple mediators in $M_{j(i)}$ rather than RC VAM coefficients alone as in equation (8). Moreover, the forecast regression is no longer assumed to fit causal value-added perfectly, that is, IV VAM allows $\nu_j \neq 0$. This scenario, which arises when M_j fails to explain all of the variation in school value-added, leads to a violation of the relevant exclusion restrictions even when assumption CRA holds. To see this, substitute the first-stage equation, (11), into the definition of $\nu_{j(i)}$ in equation (14) to obtain

$$Y_i = \tilde{\beta}_0 + M_{j(i)}'\varphi + \sum_{\ell=1}^{L} \delta_\ell (Z_{i\ell} - p_{i\ell}) + \tilde{\varepsilon}_i, \quad (15)$$

where $\tilde{\beta}_0 = \beta_0 + \sum_j \nu_j \phi_j$, $\tilde{\varepsilon}_i = \varepsilon_i + \sum_j \nu_j u_{ij}$, and $\delta_\ell = \sum_j \pi_{\ell j}\nu_j$. Parameters δ_ℓ in this expression capture effects of the instruments operating through channels other than $M_{j(i)}$. The resulting exclusion restriction violations emerge whenever offers shift students across schools with different forecast residuals.

We account for potential exclusion violations by adapting the Kolesár et al. (2015) framework for models with many invalid instruments. This framework allows exclusion violations for individual instruments, but requires these violations to average out in an asymptotic sequence that increases the number of instruments in proportion to the sample size. To see how this strategy applies here, it's useful to write the first stages for the mediators as

$$M_{j(i)k} = \psi_{0k} + \sum_{\ell=1}^{L} \psi_{\ell k}(Z_{i\ell} - p_{i\ell}) + \upsilon_{ik}, \quad k \in \{1, \ldots, K\}, \quad (16)$$

with $\psi_{0k} = \sum_j M_{jk}\phi_j$, $\upsilon_{ik} = \sum_j M_{jk} u_{ij}$, and $\psi_{\ell k} = \sum_j M_{jk}\pi_{\ell j}$. Parameters $\psi_{\ell k}$ characterize school offer effects on school characteristics, filtered through the effects of offers on enrollment as parameterized by $\pi_{\ell j}$.

With this notation in hand, the second key identifying assumption for IV VAM (after assumption CRA) can be stated as follows

Assumption MIV. Exclusion violations are orthogonal to first-stage fitted values:

$$E\left[\left(\sum_{\ell=1}^{L}\delta_\ell(Z_{i\ell}-p_{i\ell})\right)\left(\sum_{\ell=1}^{L}\psi_{\ell k}(Z_{i\ell}-p_{i\ell})\right)\right]=0,$$
$$k\in\{1,\ldots,K\}.$$

This assumption requires the first-stage predicted values for mediators generated by equation (16) to be uncorrelated with the terms generating exclusion violations in equation (15).

When the first-stage and reduced-form parameters are viewed as fixed, assumption MIV holds only in a scenario in which the δ_ℓ and $\psi_{\ell k}$ coefficients are arranged so that exclusion violations fortuitously average to 0 in the data at hand. In this case, the only random variables relevant to assumption MIV are the instruments. In a random coefficients framework, however, school-specific features M_j, ν_j, and $\{\pi_{\ell j}\}_{\ell=1}^{L}$ are seen as draws from a joint distribution of school characteristics, forecast residuals, and offer compliance behavior. In this framework (similar to that used in Angrist et al., 2017), a sufficient condition for assumption MIV can be formalized as:

Lemma 3. *Let ν denote the $J\times 1$-vector of forecast residuals, ν_j; let M denote the $J\times K$ matrix of school characteristics, M_j; let $\tilde{Z}$ denote the $N\times L$ matrix of risk-adjusted offers, $Z_{i\ell}-p_{i\ell}$. Suppose that M_j, ν_j, and $\{\pi_{\ell j}\}_{\ell=1}^{L}$ are drawn from a joint distribution of school features, so that conditional expectations involving these variables are well defined and that $E[\nu\mid\Pi,M,\tilde{Z}]=0$. Then assumption MIV holds.*

This result requires that the unexplained component of school value-added be unrelated to offer compliance rates. When M_j includes characteristics strongly predictive of school quality, ν_j can be thought of as capturing the bias in these predictions. MIV is then satisfied when bias and offer compliance rates are uncorrelated but fails when students are more likely to accept offers at higher value-added schools, conditional on mediators and offers. It's also worth noting that when M_j includes the RC VAM coefficient α_j, assumption MIV is strictly weaker than assumption CIA: the latter requires $\nu_j=0$ for all schools, while the former is compatible with nonzero ν_j, provided these idiosyncratic bias components are conditionally mean-independent of Π.

Increasing the number of instruments increases the plausibility of assumption MIV, though at the risk of increased finite-sample bias in 2SLS estimates of φ. In view of possible bias in heavily overidentified models, Kolesár et al. (2015) propose a bias-corrected 2SLS estimator (B2SLS) that is consistent in a many-instrument asymptotic sequence similar to that in Bekker (1994). The supplemental appendix adapts the assumptions of Kolesár et al. (2015) to our setting and shows that the B2SLS estimator is consistent for φ under assumptions CRA and MIV in a many-instrument asymptotic sequence (here, this means increasing L in proportion to N). This appendix also derives a consistent estimator of the forecast residual variance, σ_ν^2, under a homoskedasticity assumption. In practice, B2SLS and 2SLS estimates of value added are virtually indistinguishable in our application.

Empirical Bayes posterior predictions. In an empirical Bayes framework, IV VAM estimates of φ and σ_ν^2 can be seen as characterizing a prior distribution of school quality. These estimates, in combination with estimates of ρ and Π, yield posterior predictions of school value-added. The rest of this section sketches an empirical Bayes procedure using IV VAM estimates for posterior prediction (proofs are in the supplemental appendix).

Consider the minimum mean squared error (MSE) predictor of β as a function of OLS estimates of reduced form offer effects $\hat{\rho}$ in equation (12), conditional on M, Π, and $\tilde{Z}$:

$$\beta^*=\arg\min_{b(\cdot)}E\left[(b(\hat{\rho})-\beta)'(b(\hat{\rho})-\beta)|\Pi,M,\tilde{Z}\right]. \tag{17}$$

As always, MSE is minimized by a conditional expectation, here $\beta^*=E[\beta\mid\hat{\rho},\Pi,M,\tilde{Z}]$. The following result characterizes this function when reduced-form estimates and forecast residuals are both normally distributed:

Proposition 1. *Suppose $\hat{\rho}|(\rho,\Pi,M,\tilde{Z})\sim N(\rho,\Sigma)$ and $\nu|(\Pi,M,\tilde{Z})\sim N(0,\sigma_\nu^2 I)$. Then:*

$$\beta^*=\Omega\hat{\rho}+(I-\Omega\Pi)M\varphi, \tag{18}$$

where $\Omega=\Pi'(\Pi\Pi'+\Sigma/\sigma_\nu^2)^{-1}$.

Equation (18) defines a set of hybrid value-added predictions determined by a linear combination of reduced-form offer effects, $\hat{\rho}$, and value-added as predicted by the forecast regression, $M\varphi$. Given the assumptions invoked in the proposition, the hybrid value-add vector β^* can be interpreted as the posterior mean of β given a prior based on $(M,\Pi,\tilde{Z})$, updated with estimates of $\hat{\rho}$. Plugging estimates of φ, σ_ν^2, Π, Σ, and M into equation (18) yields an EB posterior mean $\hat{\beta}^*$. These IV VAM predictions generalize the EB shrinkage estimators commonly used to reduce the MSE of noisy OLS estimates (as in, e.g., Kane et al., 2008, and Chetty et al., 2014a) for teachers, and as described for RC VAM in section III).

Proposition 1 generates EB posterior VAM estimates for schools under weaker assumptions than deployed for this purpose in Angrist et al. (2017), which requires all school-specific parameters in the underlying random coefficients model to be normally distributed. The first assumption in proposition 1, that the reduced-form estimates are conditionally normally distributed, can be justified by an asymptotic approximation to the distribution of $\hat{\rho}$ or by the normality of the error term ω_i in equation (12). Note also that β^* is the best linear predictor of β even when the ν_j are non-normal:

Corollary 2. *Suppose* $\hat{\rho}|(\rho, \Pi, M, \tilde{Z}) \sim N(\rho, \Sigma)$, $E[\nu \mid \Pi, M, \tilde{Z}] = 0$, *and* $Var(\nu|\Pi, M, \tilde{Z}) = \sigma_\nu^2 I$. *Then the* β^* *in equation (18) solves equation (17) in the class of linear predictors of the form* $b_0 + B_1\hat{\rho}$, *where* b_0 *is a* $J \times 1$ *vector of constants and* B_1 *is a* $J \times L$ *coefficient matrix.*

This follows from the fact that the normality of the reduced-form estimates and conditional homoskedasticity of forecast residuals imply that β^* and the regression of β on $\hat{\rho}$ coincide.

The formula for Ω reveals that when M is a vector of unbiased OLS VAM coefficients α, such that $\varphi = 1$ and $\sigma_\nu^2 = 0$, the posterior β^* puts no weight on the reduced-form estimates; in this case, $\beta^* = M\varphi = \alpha$. The supplemental appendix extends this formula to allow for sampling variance in the OLS estimates, that is, for M to be the vector of $\hat{\alpha}_j$, a generalization used in the empirical work. When $\hat{\alpha}_j$ is unbiased and the underlying errors are homoskedastic, β^* again puts no weight on reduced-form estimates, simplifying to a conventional EB shrinkage formula of the form $\beta_j^* = \lambda_j \hat{\alpha}_j$, as in section III. The supplemental appendix also shows that when $L = J$, so that all schools are oversubscribed, the estimated posterior $\hat{\beta}^*$ is a weighted average of IV estimates $\hat{\Pi}^{-1}\hat{\rho}$ and the forecast regression fitted value $M\hat{\varphi}$. In general, the EB posterior combines the quasi-experimental information generated by randomized school assignment with a value-added forecast based on OLS estimates and other school characteristics, weighted to account for estimation and forecast errors.

V. Quantifying the Quality of Public Schools

We demonstrate the utility of RC VAM and IV VAM using data from the Denver and NYC public school districts. The Denver sample updates the extract analyzed by Abdulkadiroğlu et al. (2017), adding five new applicant cohorts. This sample includes students applying for sixth-grade seats at any Denver Public Schools (DPS) middle school between the 2012–2013 and 2018–2019 school years. Match data include applicant preferences, priorities, and the assignments generated by the match. We also have data on school enrollment, student demographic characteristics, and scores on the Colorado Student Assessment Program (CSAP) and Colorado Measures of Academic Success (CMAS) state achievement tests. The data appendix for Abdulkadiroğlu et al. (2017) explains how these files are processed.

Our NYC analysis sample covers sixth-grade applicants to NYC middle schools, applying for the 2016–2017 through 2018–2019 school years, and ninth-grade applicants to NYC high schools, applying for 2012–2013 through 2014–2015. This sample is an update of the extract analyzed by Abdulkadiroğlu et al. (2019), adding middle school applicants. As with Denver, the NYC analysis sample includes preferences, priorities, assignments, demographic information, and school enrollment. Middle school outcomes come from New York State achievement tests, while the high school analysis uses SAT score outcomes. The processing of NYC student records is described in the Abdulkadiroğlu et al. (2022) data appendix. In both Denver and NYC, we standardize all achievement tests and SAT scores to have mean 0 and standard deviation 1 separately by year.

Students in Denver rank up to five schools participating in the DPS unified enrollment match, which covers public schools of all types including traditional district schools and charter schools. Priorities are assigned based on criteria like sibling status and applicant neighborhood. A DA algorithm implemented with a single lottery tie-breaker assigns students to schools. We calculate assignment risk for DPS applicants using the propensity score formula derived by Abdulkadiroğlu et al. (2017). This formula is an analytical large-market approximation to the school assignment propensity score for DA with a random tie-breaker.[11]

NYC applicants rank up to 12 academic programs in middle or high school. For purposes of the analysis that follows, multiple programs are aggregated to the school level. The NYC match features a variety of tie-breakers, with "unscreened" schools using a random lottery number and "screened" schools using nonrandom tie-breakers such as sixth-grade test scores and grades. Propensity scores for NYC school assignment are computed as described in Abdulkadiroğlu et al. (2022). These scores depend in part on bandwidths for screened school tie-breakers, similar to those used in standard regression discontinuity designs.[12] As in Abdulkadiroğlu et al. (2022), regression and 2SLS estimates that control for propensity scores also control for local linear functions of the relevant screened-school tie-breakers for applicants inside the relevant bandwidth.

Table 1 describes students and schools in the DPS and NYC samples. The first column shows statistics for the full sample of enrolled DPS middle school students, while column 2 shows statistics for DPS applicants with nondegenerate assignment risk (these students have $p_{ij} \in (0, 1)$ for at least one j). Columns 3 and 4 of the table report corresponding statistics for NYC middle school students, and columns 5 and 6 describe NYC high school students. RC VAM and other OLS

[11] The DPS score is computed using the formula score described in Abdulkadiroğlu et al. (2017). This is computed by identifying sets of applicant types that are never seated, always seated, and conditionally seated at every school in the match, a classification that depends on priorities. The formula score is a function of this classification and the school-specific tie-breaker cutoffs determined by the match. Cutoff dependence is determined by an applicant's preferences. See Abdulkadiroğlu et al. (2017) for details.

[12] The NYC score is the local DA score described in section 4.2 of Abdulkadiroğlu et al. (2022). As with the formula score, it is a function of a classification of types into always, never, and conditionally seated applicants to each school, as well as the tie-breaker cutoffs generated by the match. Applicant classification for the local score depends also on nonlottery tie-breakers and bandwidths around screened-school admissions cutoffs. Bandwidths used here are computed as suggested by Calonico et al. (2019).

TABLE 1.—DESCRIPTIVE STATISTICS

	Denver middle schools		NYC middle schools		NYC high schools	
	All (1)	With risk (2)	All (3)	With risk (4)	All (5)	With risk (6)
Demographics						
Hispanic	0.592	0.581	0.413	0.445	0.380	0.430
Black	0.125	0.140	0.231	0.254	0.281	0.278
White	0.210	0.201	0.154	0.110	0.137	0.106
Female	0.493	0.494	0.494	0.484	0.525	0.520
Free/reduced price lunch	0.723	0.703	0.731	0.763	0.774	0.792
Special education	0.102	0.087	0.201	0.215	0.126	0.066
English language learner	0.393	0.416	0.113	0.113	0.093	0.087
Baseline scores						
Math (standardized)	0.000	0.077	0.000	−0.063	0.000	−0.043
ELA (standardized)	0.000	0.070	0.000	−0.055	0.000	−0.040
Enrollment						
Screened	0.000	0.000	0.067	0.044	0.205	0.118
Lottery	1.000	1.000	0.933	0.956	0.795	0.882
Share noncompliant	0.300	0.291	0.268	0.324	0.365	0.336
Share not offered	0.182	0.048	0.149	0.134	0.188	0.117
Students	37,089	8,100	184,760	46,095	122,214	32,479
Schools	80	75	624	594	486	484
Lotteries (schools with risk)		67		448		382

This table describes Denver and NYC student samples. Column 1 reports descriptive statistics for Denver students enrolled in sixth grade in the 2012–13 through 2018–19 school years. Column 3 reports statistics for NYC middle school students enrolled in sixth grade in the 2016–17 through 2018–19 school years. Column 5 reports statistics for NYC high school students enrolled in ninth grade in the 2012–13 through 2014–15 school years. Columns 2, 4, and 6 report on the corresponding samples of applicants with assignment risk at at least one school. Baseline characteristics and lagged scores are from fifth grade for middle school samples and eighth grade for high school samples. Baseline scores are standardized to be mean 0 and standard deviation 1 in the student-level test score distribution, separately by year. Screened schools are defined as schools without any lottery programs. The share noncompliant is defined as the proportion of students who enroll other than where offered a seat; this includes students receiving no offers.

VAM models are estimated using the full sample of enrolled students, while bias tests and IV VAM estimators use the nondegenerate risk subsamples.

As is typical of large urban districts, most DPS and NYC students are disadvantaged, with over 70% eligible for a subsidized lunch. Roughly a quarter of the students in each sample face some assignment risk, with such applicants appearing broadly representative of enrolled student populations. Table A1 in the supplemental appendix compares the characteristics of students offered seats at higher- and lower-value-added schools (as measured by the conventional VAM estimates discussed below) within the at-risk sample. We see large differences in student characteristics between those offered high- and low-value-added seats. Controlling for assignment risk, however, makes these differences vanish. The fact that risk control makes centralized assignment offers independent of observed characteristics suggests this is likely to be the case for unobserved characteristics as well.[13]

Table 1 shows that most NYC students in our samples attend lottery rather than screened schools (all DPS schools use lottery tie-breaking). Roughly 30% of middle school students in both cities are noncompliant in the sense that they enroll someone other than where offered a seat, while noncompliance is a little higher among NYC high school students. RC VAM implicitly treats the gap between enrollment and assignment as conditionally randomly assigned. The bottom rows of the table show that most Denver and New York schools are oversubscribed. Specifically, at least one student has nondegenerate risk at 67 out of 80 Denver middle schools, at 448 out of 624 NYC middle schools, and at 382 out of 486 NYC high schools. This reflects the interdependence of school assignments in a centralized match: oversubscription at in-demand schools generates assignment risk even for unpopular schools with fewer applicants than seats. At the same time, some schools are undersubscribed and some oversubscribed schools have small at-risk samples or low offer take-up rates. Our RC VAM and IV VAM procedures yield estimates of causal effects for such schools even so.

A. *Evaluating RC VAM*

The testing framework in section III is used to compare the predictive validity of RC VAM estimates with results from OLS VAM estimators relying on fewer controls. The first of these is a benchmark model, labeled *uncontrolled*, that includes only application year dummies in X_i and omits the propensity score controls, p_i. The second is a conventional value-added model adding dummies for sex, race, subsidized lunch status, special education, and limited English proficiency, along with cubic functions of baseline math and English language arts (ELA, or reading) scores, to the control vector, X_i. This model parallels widely used VAM specifications for the measurement of teacher and school effectiveness

[13]Balance checks regress student characteristics on the conventional value-added of the school where applicants are offered a seat, along with a dummy indicating whether the applicant was offered a seat anywhere. Risk controls consist of expected value-added and the probability of receiving any offer. The former is computed as a score-weighted average of school value-added. Table A1 in the supplemental appendix shows that control for risk eliminates imbalances between applicants offered high- and low-value-added seats and between those who do and do not get offers. Differential attrition can create selection bias even with random assignment. Supplemental appendix table A2 shows that follow-up rates for key outcomes are largely unrelated to assigned school value-added, conditional on assignment risk. This makes it unlikely that selective attrition biases estimates of reduced-form school-offer effects.

TABLE 2.—VAM BIAS TESTS FOR MIDDLE SCHOOL MATH SCORES

	Uncontrolled (1)	Conventional (2)	Risk only (3)	RC VAM (4)
A. Denver middle schools				
Forecast coefficient	0.427	1.12	0.813	1.12
	(0.059)	(0.106)	(0.091)	(0.103)
First-stage F-statistic	48.4	104	35.6	94.8
Bias tests				
Forecast bias	95.0	1.19	4.20	1.28
	[0.000]	[0.275]	[0.040]	[0.258]
Overidentification (19 d.f.)	79.0	28.7	49.3	21.9
	[0.000]	[0.070]	[0.000]	[0.292]
Omnibus (20 d.f.)	174	29.9	53.5	23.1
	[0.000]	[0.071]	[0.000]	[0.282]
N (testing)	7,660			
N (VAM estimation)	37,089			
B. NYC middle schools				
Forecast coefficient	0.600	0.933	0.817	0.994
	(0.030)	(0.041)	(0.039)	(0.044)
First-stage F-statistic	154	649	188	530
Bias tests				
Forecast bias	177	2.61	21.8	0.020
	[0.000]	[0.106]	[0.000]	[0.886]
Overidentification (19 d.f.)	86.1	24.3	67.0	24.7
	[0.000]	[0.187]	[0.000]	[0.170]
Omnibus (20 d.f.)	263	26.9	88.8	24.7
	[0.000]	[0.139]	[0.000]	[0.212]
N (testing)	44,494			
N (VAM estimation)	184,760			

This table reports tests for bias in OLS value-added models (VAMs). The uncontrolled VAM includes indicators for application year. The conventional VAM adds cubic functions of baseline math and ELA scores and indicators for sex, race, subsidized lunch, special education, and limited English proficiency, each interacted with application year. Risk-only VAM adds propensity score and running variable controls to the uncontrolled specification. RC VAM combines the controls in the conventional and risk-only VAMs. Forecast coefficients are from instrumental variables regressions of test scores on VAM fitted values, instrumenting fitted values with binned assignment indicators. Assignments are binned by ventile of the estimated conventional VAM. IV models control for propensity scores, running variable controls, and baseline demographics and achievement. Test scores for outcomes and VAMs are standardized to be mean 0 and standard deviation 1 in the student-level test score distribution, separately by year. The forecast bias test checks whether the forecast coefficient equals 1; the overidentification test checks overidentifying restrictions implicit in the procedure used to estimate the forecast coefficient. The omnibus test combines tests for forecast bias and overidentification. Standard errors are reported in parentheses; test p-values are reported in brackets.

(Chetty et al., 2014a). A third model, labeled *risk only*, omits conventional VAM controls but includes assignment propensity scores, p_i. Finally, RC VAM estimates come from regressions that combine conventional VAM controls with controls for school assignment risk.[14]

Uncontrolled middle school VAM estimates are clearly contaminated by selection bias. This can be seen in table 2, which reports the components of the test statistic equation (9), for sixth-grade math scores in each city. For the purpose of VAM testing, schools are classified into twenty bins defined by ventiles of the distribution of estimated conventional value-added. The testing equation, (7), is estimating using bin-level (rather than single-school) offers and propensity scores. This aggregation may increase test power relative to tests derived by using all school-specific offers as instruments for value-added in equation (10).[15] As shown in column 1 of table 2, the uncontrolled VAM specification generates forecast coefficients of 0.43 in DPS (shown in panel A) and 0.60 in NYC (shown in panel B). These estimates are statistically different from 1, while the overidentification and omnibus tests clearly reject the null hypothesis of zero bias in the uncontrolled model.

VAM research to date suggests that controlling for lagged test scores and student demographic characteristics eliminates much of the selection bias in naive comparisons of achievement across teachers and schools (Chetty et al., 2014a; Bacher-Hicks et al., 2014; Deming, 2014; Angrist et al., 2017). Consistent with this finding, the second column of table 2 shows that middle school VAMs estimated with conventional controls boost the forecast coefficient markedly, yielding forecast estimates of 1.12 in DPS and 0.93 in NYC, respectively. The former is statistically indistinguishable from 1, while the latter is only marginally significantly different from 1, reflecting the fact that NYC estimates are considerably more precise. Omnibus test results are marginal for conventional VAM in both cities. The DPS overidentification test also results in a marginal rejection of the null hypothesis.

As can be seen in column 3 of table 2, middle school VAMs estimated with risk controls alone also improve greatly on the uncontrolled estimates. The risk-only model generates forecast coefficients of 0.81 and 0.82 in DPS and NYC. In contrast

[14]Schools out of the match have constant risk equal to $p_{ij} = 0$ for all students. RC VAM for such schools adds no controls relative to conventional VAM. But for undersubscribed schools with a mix of applicants such that $p_{ij} = 0$ and $p_{ij} = 1$, risk control becomes a dummy indicating any risk. The zero-risk group in such cases consists mostly of nonapplicants.

[15]Roodman (2009) discusses the connection between the number of overidentifying restrictions and the power of overidentification tests. Supplemental appendix table A3 reports results using alternative bin schemes; the choice of bin size matters little for test results.

TABLE 3.—VAM BIAS TESTS FOR SAT MATH SCORES, NYC HIGH SCHOOLS

	Uncontrolled (1)	Conventional (2)	Risk only (3)	RC VAM (4)
Forecast coefficient	0.344	0.783	0.674	0.956
	(0.030)	(0.064)	(0.056)	(0.077)
First-stage F-statistic	161	240	87.5	181
Bias tests				
Forecast bias	466	11.3	33.7	0.327
	[0.000]	[0.001]	[0.000]	[0.567]
Overidentification (19 d.f.)	38.7	20.4	20.7	14.5
	[0.005]	[0.369]	[0.352]	[0.756]
Omnibus (20 d.f.)	505	31.8	54.4	14.8
	[0.000]	[0.046]	[0.000]	[0.788]
N (testing)		30,066		
N (VAM estimation)		122,214		

This table reports tests for bias in OLS value-added models (VAMs). See the notes to table 2 for a description of models and test procedures. Standard errors are reported in parentheses; test p-values are reported in brackets.

with the test results for conventional VAM, however, omnibus and overidentification test results clearly reject the null of unbiased risk-only estimates. On the other hand, RC VAM estimates, evaluated in the fourth column of the table, yield remarkably accurate and internally consistent predictions of school quality. For both DPS and NYC middle schools, estimated RC VAM forecast coefficients are close to 1, while the associated omnibus tests offer little evidence against the claim that RC VAM estimates can be interpreted as causal.

Test results for NYC high schools, reported in table 3, likewise show that RC VAM estimates for SAT math are virtually unbiased. In contrast, the other VAM estimators evaluated in the table are almost certainly biased. This is unsurprising for uncontrolled estimates, which yield a forecast coefficient of only 0.34. Conventional and risk-only VAMs do much better, with forecast coefficients of 0.78 and 0.67, respectively. Even so, test results for both models suggest substantial remaining bias. In these cases, forecast bias rather than a failure of overidentifying restrictions is the source of the omnibus test rejection. The fact that the lagged score controls in the conventional VAM estimates come from assessments other than the SAT seems likely to contribute to the relatively poor performance of conventional VAM estimates for high schools (a point made in a different context by Chetty et al., 2014b). The good performance of RC VAM for SAT scores is therefore especially impressive.[16]

As noted in section III, the tests reported in tables 2 and 3 can be interpreted as asking whether OLS-predicted VAM estimates predict the causal effects of randomized school offers. Figure 1 visualizes these predictions. Specifically, the figure plots reduced-form offer effects for each value-added bin (the ρ_ℓ coefficients in equation [12]) against first-stage effects of bin offers on predicted value-added of an applicant's enrolled school (that is, OLS value-added at the school attended, $\hat{\alpha}_{j(i)}$). The corresponding 2SLS estimate of φ is given by the slope of a weighted least squares line of best fit through the origin, while the overidentification test checks whether this line fits all points up to sampling error (Angrist, 1991). Consistent with the estimates in tables 2 and 3, each panel shows that adding either conventional or risk controls pushes the forecast slopes toward 1 and reduces dispersion around the best-fitting lines. Including both sets of controls yields tightly estimated relationships that are statistically indistinguishable from the 45-degree line. Figure A1 and table A4 in the supplemental appendix show similar patterns for reading test score VAMs.

In a constant-effects framework, rejection of an overidentification test is evidence of selection bias. In a world of heterogeneous school effects, however, rejection might be generated by differences in school quality across students. To explore this possibility, we follow Angrist et al. (2017) in testing models that allow value-added to vary with student characteristics. If heterogeneity (mediated by covariates) is the source of VAM misalignment with reduced-form offer effects, the resulting test results should be more forgiving. As can be seen in appendix table A5, however, forecast coefficients and test results are similar when VAM is allowed to vary by baseline year, subsidized lunch status, special education status, and baseline score tercile. This table also shows similar test results when OLS VAMs are estimated using only data from the subsamples with nondegenerate assignment risk.[17]

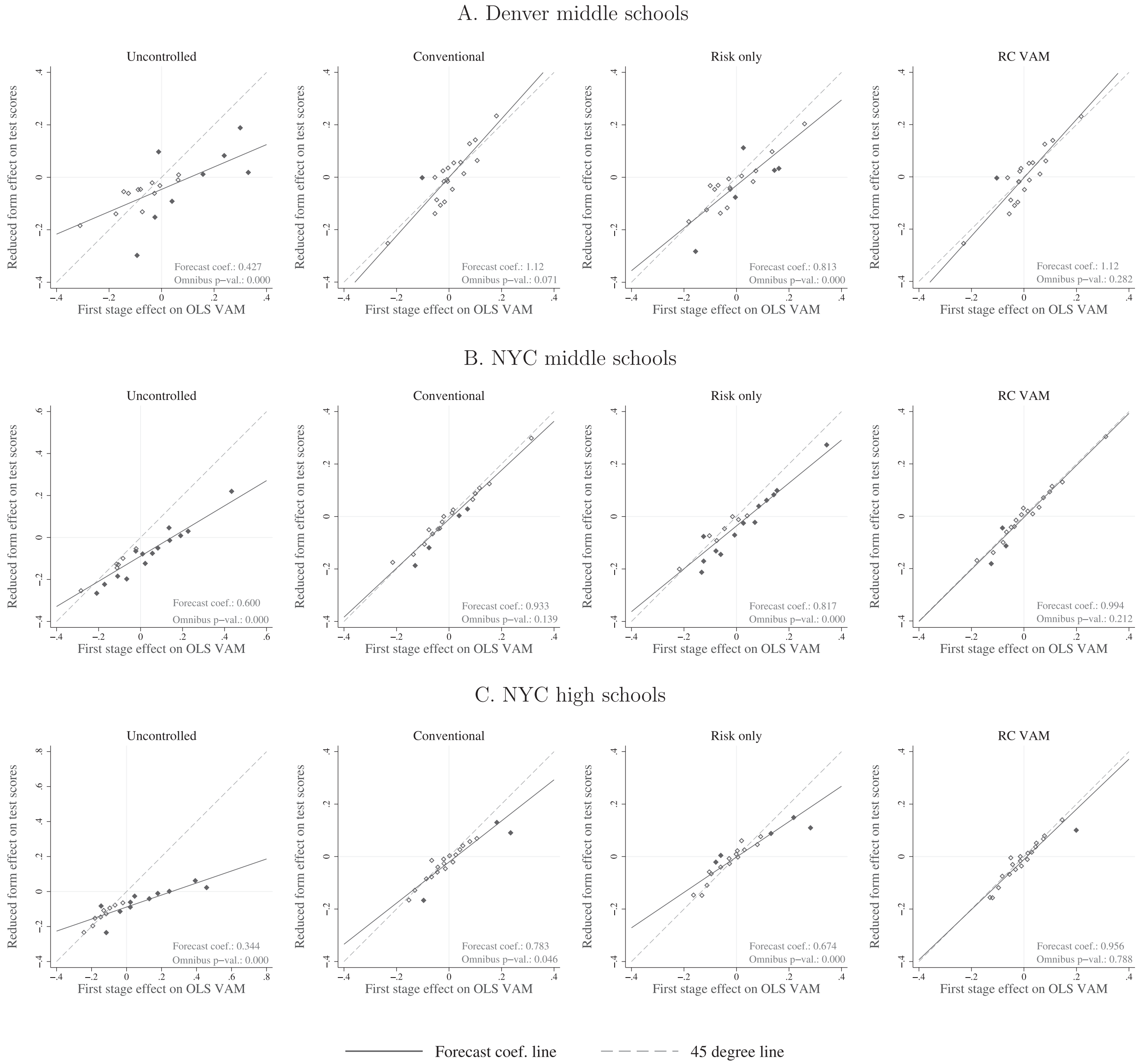
B. IV VAM Estimates

Our IV VAM application focuses on NYC middle and high schools, since these samples have the largest number

[16]Lagged score controls used to compute the estimates in table 2 come from fifth grade, taken the year before the start of middle school; lagged score controls used to compute the estimates in table 3 come from eighth grade, taken the year before the start of high school.

[17]The question of effect heterogeneity is particularly salient for NYC screened schools, which admit students partly on the basis of tie-breaking variables that gauge academic ability or artistic talent. The cutoffs generated by centralized assignment to screened schools act something like a regression discontinuity (RD) design. With centralized assignment, however, students with tie-breakers far from admissions cutoffs may nevertheless have nondegenerate screened-school assignment risk (Abdulkadiroğlu et al., 2021). RC VAM estimates for screened schools therefore reflect outcomes for a wider range of students than the group of compliers usually found inside an RD bandwidth.

FIGURE 1.—VISUAL INSTRUMENTAL VARIABLES TESTS FOR BIAS (MATH)



This figure plots reduced-form estimates against value-added first stages from each of twenty school assignment bins. Outcomes are sixth-grade math CSAP and CMAS scores for Denver, sixth-grade math New York State Assessment scores for NYC middle schools, and SAT math scores for NYC high schools. Scores are standardized to be mean 0 and standard deviation 1 in the student-level test score distribution, separately by year. Assignments are binned by ventile of the estimated conventional VAM. See notes to table 2 for a description of the value-added models and test procedure. 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 2, and dashed lines indicate the 45-degree line.

of schools and students with assignment risk. The vector of mediators used to predict causal value-added in includes a screened school indicator and the OLS VAM estimates evaluated in tables 2 and 3. Forecast regression parameters and the forecast residual variance are estimated by instrumenting $M_{j(i)}$ with a full set of school offer dummies, controlling for school assignment propensity scores and the baseline covariates used to compute the test statistics reported in these tables. We use individual school offers rather than binned offers for IV VAM because the extra instruments make assumption MIV more plausible. As detailed in the supplemental appendix, IV VAM is implemented using a bias-corrected 2SLS estimator consistent in the asymptotic sequence in which the number of instruments grows in proportion to sample size.

Using uncontrolled VAM estimates to predict causal value-added yields an estimated IV VAM forecast coefficient of only 0.26 for middle schools and 0.24 for high schools. These results, reported in the first column of table 4, are qualitatively consistent with the test statistics reported in tables 2 and 3, which show strong evidence of forecast bias and the failure of uncontrolled VAM to satisfy the corresponding overidentification test.

The IV VAM procedure generates an estimated screened-school effect on causal value-added. This is negative and

Table 4.—IV VAM Estimates for NYC Middle and High School Math Scores

	(1)	(2)	(3)	(4)	(5)	(6)
A. NYC Middle Schools						
Mediators						
Uncontrolled VAM	0.255					
	(0.029)					
Conventional VAM		0.919			−0.151	
		(0.044)			(0.136)	
Risk-only VAM			0.613			
			(0.040)			
RC VAM				0.990	1.14	
				(0.041)	(0.138)	
Conventional VAM (older lagged scores)						0.809
						(0.036)
Screened sector dummy	−0.156	−0.032	−0.083	−0.004	−0.001	−0.070
	(0.056)	(0.042)	(0.048)	(0.039)	(0.040)	(0.037)
Standard deviations						
Value-added	0.190	0.194	0.261	0.204	0.205	0.187
Max(forecast residual, 0)	0.130	0.058	0.095	0.028	0.027	0
First-stage *F*-statistic	26.2	26.0	20.5	26.0	26.0	26.1
N	46,095					
B. NYC High Schools						
Mediators						
Uncontrolled VAM	0.244					
	(0.030)					
Conventional VAM		0.801			−0.242	
		(0.059)			(0.160)	
Risk-only VAM			0.600			
			(0.047)			
RC VAM				0.999	1.25	
				(0.065)	(0.178)	
Conventional VAM (older lagged scores)						0.644
						(0.049)
Screened sector dummy	−0.007	0.007	−0.034	−0.025	−0.032	−0.012
	(0.040)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)
Standard deviations						
Value-added	0.135	0.151	0.209	0.169	0.169	0.147
Max(forecast residual, 0)	0.060	0	0	0	0	0
First-stage *F*-statistic	10.0	10.3	6.76	10.5	10.2	10.4
N	32,479					

This table reports IV VAM parameter estimates for math scores. The rows listing mediators report forecast coefficients and sector effects from instrumental variable regressions of test scores on the OLS VAM estimates listed as mediator, along with an indicator for screened school attendance. Mediators are instrumented with individual school assignment offer dummies. Conventional and RC VAM mediators for the estimates reported in columns 1–5 use fifth-grade lagged score controls for middle schools and eighth-grade lagged score controls for high schools. The conventional VAM mediator for the estimates in column 6 relies on third-grade lagged score controls for middle schools and sixth-grade lagged score controls for high schools. Standard errors are reported in parentheses.

significantly different from 0 for middle schools but 0 for high schools. The bottom rows of table 4 report IV VAM estimates of the forecast residual standard deviation, σ_ν, along with the overall standard deviation of causal value-added, σ_β (obtained as $\sigma_\beta^2 = \text{Var}(M_j'\varphi) + \sigma_\nu^2$ from the forecast regression). These estimates are scaled in standard deviation units of the student-level test score distribution. The estimates in column 1 reveal substantial residual variation unexplained by uncontrolled VAM or by screened status (with estimated $\sigma_\nu = 0.13$ for middle schools and $\sigma_\nu = 0.06$ for high schools) and a total value-added standard deviation estimated at $\sigma_\beta = 0.19$ for middle schools and $\sigma_\beta = 0.14$ for high schools. The latter pair of estimates highlights the substantial variation in quality across NYC schools.

As shown in the second and third columns of table 4, replacing uncontrolled VAM estimates with conventional or risk-only value-added estimates increases IV VAM forecast coefficients (to around 0.8 to 0.9 and 0.6 respectively), while also decreasing the standard deviation of the forecast residual. The screened school effects in these columns are small and not significantly different from 0, suggesting that the larger negative estimate for middle schools in the first column is driven by the diminished predictive power of uncontrolled VAM.

IV VAM models generate comparable estimates of σ_β in columns 1, 2, and 4 (around 0.2 for middle schools and 0.15 for high schools), though the estimates in column 3 are notably larger, perhaps reflecting failure of assumption MIV in this case. It is noteworthy that the middle school forecast coefficient estimates for uncontrolled and risk-only VAM in the testing and IV VAM tables differ markedly, though not so for conventional VAM. This reflects the near-unbiasedness of conventional VAM estimates. When the VAM estimates used as mediators generate unreliable predictions, forecast coefficient estimates computed using binned instruments and individual school offers should differ.

The estimates in columns 4 and 5 of table 4 align with those in tables 2 and 3 in highlighting the virtual unbiasedness of

RC VAM. Including the RC VAM estimates as an IV VAM mediator yields a precisely estimated forecast coefficient of 0.99 in both middle and high school, with an estimated residual standard deviation close to 0.[18] At the same time, column 5 shows that models including both conventional and RC VAM as mediators generate an insignificant negative forecast coefficient for conventional VAM, while the RC VAM coefficient remains close to 1 (including these two highly correlated mediators reduces the precision of estimated forecast coefficients). This is notable since conventional VAM predicts causal value-added well on its own. The screened school coefficient estimates in these columns are precisely estimated 0s, consistent with the view that RC VAM tells us everything we need to know about school quality.

On balance, there would seem to be little scope for IV VAM posterior predictions to improve on RC VAM. Conventional VAMs also generate remarkably accurate predictions of causal school effects. But it is worth highlighting the fact that both of these models rely on controls for lagged test scores measured one year prior to the outcome. In practice, this sort of proximate lagged score control may be unavailable. The COVID-19 pandemic, for example, led many districts to temporarily suspend standardized testing programs. Many districts also rely on "skip-year growth" metrics that omit a year of test score data when transitioning to new assessment systems. These scenarios necessitate VAM control strategies that rely on test scores lagged by more than one year.

By way of evidence on the consequences of older lagged score controls, column 6 of table 4 reports IV VAM estimates from a procedure where the mediator is a set of conventional VAM estimates that rely on longer lags. The estimates here replace fifth-grade scores with third-grade scores in a conventional VAM for sixth-grade outcomes and sixth-grade scores for ninth-grade outcomes. Using earlier lagged score controls reduces the forecast coefficient to 0.81 for middle schools and to 0.64 for high schools. This indicates a substantial deterioration in predictive power relative to the estimates in column 2. It remains, however, to gauge the extent to which IV VAM can ameliorate this.

C. *Estimating VAM Mean Squared Error*

Mean squared error provides a natural standard of comparison for IV VAM, RC VAM, and conventional VAM estimators computed using different sets of controls. The MSE of an OLS-based linear shrinkage estimate, α_j^*, for value-added, β_j, is given by

$$E\left[(\alpha_j^* - \beta_j)^2\right] = E\left[\lambda_j^2 s_j^2\right] + \sigma_\alpha^2 E\left[(\varphi - \lambda_j)^2\right] + \sigma_\nu^2. \quad (19)$$

The first term in this formula reflects sampling variance in the OLS estimates, while the last two terms are attributable to bias. A similar (though more involved) calculation in the supplemental appendix gives the MSE of IV VAM posterior predictions, with a parallel decomposition into bias and sampling variance. In practice, we calculate MSE for each estimator by simulation, drawing school value-added from a normal distribution calibrated to match a set of IV VAM forecast coefficient and residual variance estimates.[19]

Figure 2 compares the root mean squared error (RMSE) of OLS VAM and IV VAM posterior predictions of causal value-added. RMSE is minimized by the RC VAM posterior, at around 0.06 standard deviations in the distribution of both middle and high school math scores. Sampling variance contributes far more to the MSE of RC VAM than bias. What bias there is arises from shrinkage (i.e., $\lambda_j \neq 1$). The RMSE of the conventional model exceeds that of RC VAM despite the relative imprecision of the latter. The precision penalty with RC VAM reflects the fact that risk controls absorb a substantial portion of the variation in school enrollment, but this loss is more than outweighed by the reduced bias of RC VAM.

Perhaps surprisingly, IV VAM posteriors computed using RC and conventional estimates as mediators have somewhat higher total RMSE than the corresponding shrunk OLS estimates. This reflects the fact that the forecast coefficient and residual variance are treated as known when deriving the MSE-minimizing prediction in equation (18), while estimation of these parameters generates additional sampling variance in practice. Additional simulations (not reported) confirm that were these hyperparamters known, the IV VAM RMSE necessarily falls below that of the corresponding OLS estimates. Evidently, models that control for lagged scores have such low bias that the IV VAM bias correction is not worth the cost introduced by estimation of the extra parameters that IV VAM requires.

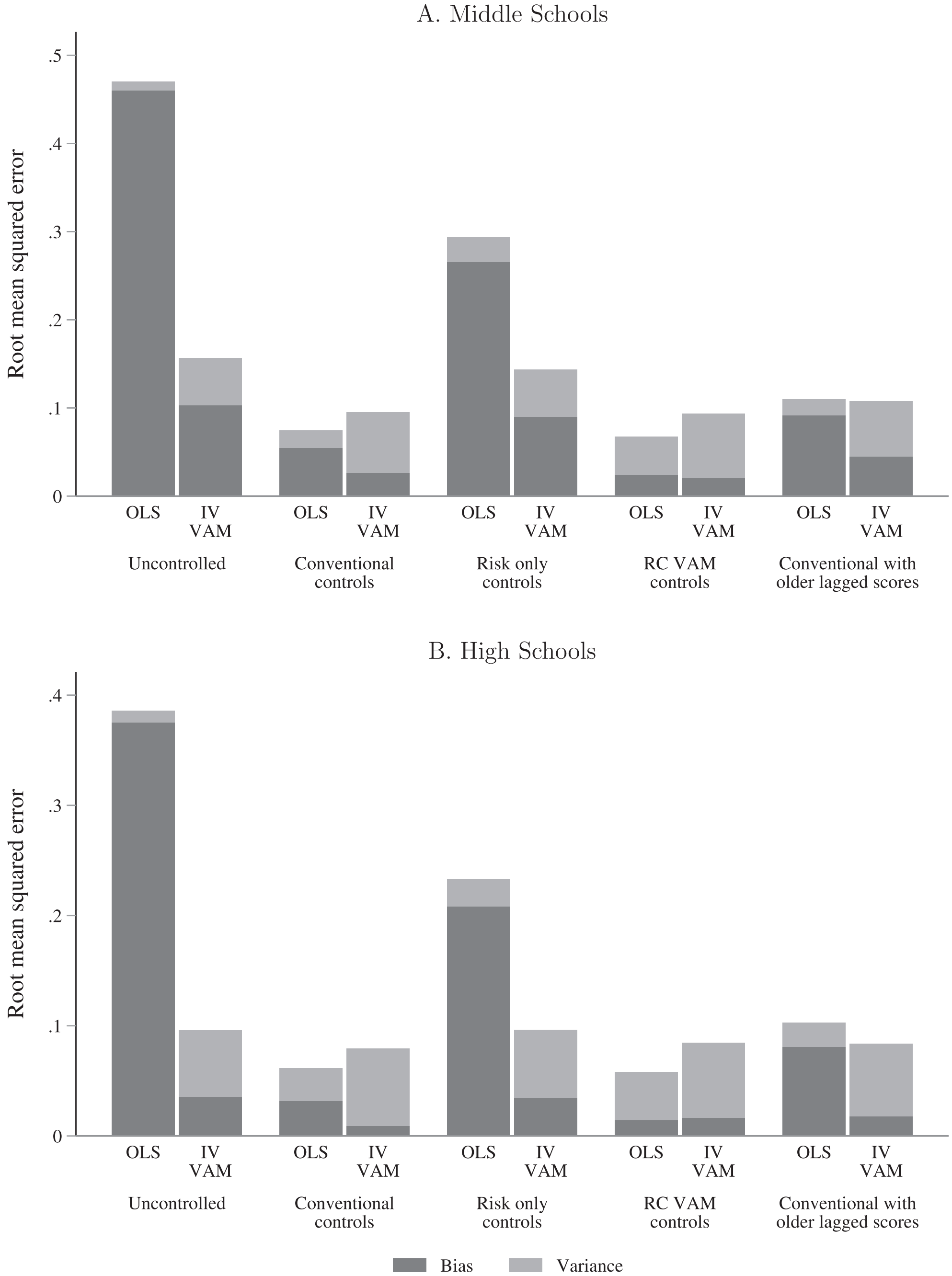
RMSE with risk control alone improves greatly on the RMSE of uncontrolled VAM. Moreover, hyperparameter estimation notwithstanding, IV VAM reduces the RMSE of both uncontrolled and risk-controlled VAM considerably. These gains are perhaps most noteworthy for the latter model, where RMSE is cut by more than half. The improvements generated by IV VAM in the uncontrolled and risk-only models suggest that IV VAM is a good substitute for lagged score controls, a point of some practical relevance.

The last two columns in figure 2 compare RMSE when conventional VAM is estimated with older lagged score controls. Consistent with the estimates in the last column of table 4, use of older lagged score controls increases RMSE substantially. At the same time, IV VAM cuts the bias of these estimates markedly, while also reducing total RMSE. The improvement here is especially large for high schools, where IV VAM cuts the RMSE of conventional models by about 20% relative to OLS (roughly half the gap with RC VAM). This highlights the value of IV VAM estimation in applications with missing or degraded lagged achievement controls.

[18]The estimator of σ_ν^2 can be negative; the table reports 0 in such cases.

[19]The IV VAM estimates used to calibrate the simulation appear in appendix table A6. These estimates come from a model that includes the uncontrolled, conventional, risk only, and RC VAM estimates as mediators along with a screened sector dummy, allowing us to evaluate each estimator maintaining a consistent distribution of school quality.

FIGURE 2.—RMSE OF VALUE-ADDED ESTIMATES FOR NYC MIDDLE AND HIGH SCHOOLS



This figure plots root mean squared error (RMSE) for posterior predictions of value-added generated by OLS and IV VAM. OLS VAM predictions are posterior means constructed from OLS value-added estimates. IV VAM predictions are posterior means constructed from OLS and reduced-form estimates. Bars indicate RMSE. The darker and lighter shading mark the shares of RMSE due to bias and variance. Conventional and RC VAM posterior predictions are from models using fifth- and eighth-grade tests as lagged score controls for middle and high school students, as in table 2. The conventional VAM posterior with older lagged scores relies on third grade lagged score controls for middle and schools, and sixth-grade lagged score controls for high schools.

VI. Summary and Conclusion

VAM estimates may help families choose schools wisely, perhaps with life-changing consequences. Policymakers and educators likewise base high-stakes decisions related to school access, expansion, and closure on VAMs; the federal government and many states require this. Given the stakes, how should the consequences of attendance at individual schools be estimated? Many school accountability frameworks rank schools by performance measures computed with few or no controls. Such poorly controlled VAM estimates confuse school quality with the ability of student bodies. In large urban districts like those examined here, schools sporting the highest test scores and graduation rates tend to enroll an outsized share of nonminority students. These schools are also found in wealthier neighborhoods.

The primary econometric challenge in this context is how to eliminate or moderate this sort of selection bias. We show here that centralized assignment provides an invaluable and easily exploited tool in service of this goal. By matching students to schools as a function of observed characteristics and partially or fully randomized tie-breaking variables, centralized assignment takes much of the mystery out of who goes where. This by-product of centralized matching can be key to VAM estimation strategies with minimal selection bias.

Our RC VAM estimator exploits centralized assignment by controlling for the many student preference and priority variables that govern match outcomes. The problem of high-dimensional controls is solved by conditioning on the relatively coarse school assignment propensity score induced by DA matching algorithms. Importantly, the RC VAM procedure generates school-specific VAM estimates for all schools in a match, regardless of undersubscription. Moreover, the assumptions justifying RC VAM are easily validated by testing whether RC VAM residuals are orthogonal to offers of seats at oversubscribed schools. Application of this test to schools in Denver and NYC suggests that RC VAM estimates provide a remarkably accurate account of school quality. RC VAM estimates exhibit little bias and outperform conventional VAM strategies on mean squared error grounds.

We have also introduced an IV VAM estimator that exploits reduced forms for causal value-added estimation in districts with fewer randomized offers than schools. The IV VAM procedure outlined here, which builds on and simplifies earlier efforts in this direction, amounts to 2SLS estimation of mediating-variable effects followed by a simple weighted-average calculation of empirical Bayes posterior means. IV VAM posterior predictions have an attractive best linear predictor property and require weaker distributional assumptions than the fully parametric hybrid IV-and-OLS VAM estimator developed in Angrist et al. (2017). Our NYC application shows that IV VAM can improve on poorly controlled estimates of school quality and on estimates that rely on older and perhaps less relevant controls.

Further afield, we expect the estimation strategies developed here to find application in other markets with elements of systematic and chance assignment. Possible applications include job assignment systems, such as those used by Teach for America to place interns in school, the measurement of physician and hospital quality, and the consequences of receiving rationed medical resources like new drugs and mechanical ventilation during the recent pandemic. In these contexts, RC VAM and IV VAM can be deployed to answer causal questions about the consequences of receiving a particular assignment or scarce resource. Finally, on the theoretical side, there is work to be done on integrating the large market asymptotic sequence used to derive school assignment propensity scores with the many-instrument asymptotic sequences used to study the behavior of econometric estimators like 2SLS. We plan to explore these applications and questions in future work.

Appendix Proofs

A. *Lemmas 1 and 2*

As in Rosenbaum and Rubin (1983), we establish lemma 1 by showing $\Pr(Z_{ij} = 1 \mid \varepsilon_i, p_i) = \Pr(Z_{ij} = 1 \mid p_i) = p_{ij}$. By the law of iterated expectations,

$$\begin{aligned}\Pr(Z_{ij} = 1 \mid \varepsilon_i, p_i) &= E[E[Z_{ij} \mid \theta_i, \varepsilon_i, p_i] \mid \varepsilon_i, p_i] \\ &= E[E[Z_{ij} \mid \theta_i, p_i] \mid \varepsilon_i, p_i] = p_{ij}.\end{aligned}$$

The second line uses assumption CRA; the last line uses the fact that $E[Z_{ij} \mid \theta_i, p_i] = E[Z_{ij} \mid \theta_i] = p_{ij}$. We establish equation (3) by repeating the same argument, starting with the fact that $(\varepsilon_i, X_i) \perp\!\!\!\perp Z_i \mid \theta_i$ and adding X_i to the conditioning set in the first line above.

To establish lemma 2, start with $\varepsilon_i \perp\!\!\!\perp Z_i \mid (p_i, X_i)$, and assume $\varepsilon_i \perp\!\!\!\perp D_i \mid (p_i, X_i, Z_i)$. Then by the law of iterated expectations,

$$\begin{aligned}\Pr(D_{ij} = 1 \mid \varepsilon_i, p_i, X_i) &= E[E[D_{ij} \mid \varepsilon_i, p_i, X_i, Z_i] \mid \varepsilon_i, p_i, X_i] \\ &= E[E[D_{ij} \mid p_i, X_i, Z_i] \mid \varepsilon_i, p_i, X_i] \\ &= E[E[D_{ij} \mid p_i, X_i, Z_i] \mid p_i, X_i] \\ &= \Pr(D_{ij} = 1 \mid p_i, X_i).\end{aligned}$$

The second line follows from the independence of ε_i and D_i conditional on (p_i, X_i, Z_i), the third follows from independence of Z_i and ε_i conditional on (p_i, X_i), and the fourth follows from another application of the law of iterated expectations.

B. *Corollary 1*

The corollary is a consequence of the law of iterated expectations:

$$\begin{aligned}E[\varepsilon_i(Z_{ij} - p_{ij})] &= E[\varepsilon_i(E[Z_{ij} \mid \varepsilon_i, p_{ij}] - p_{ij})] \\ &= E[\varepsilon_i(E[Z_{ij} \mid p_{ij}] - p_{ij})] \\ &= E[\varepsilon_i(p_{ij} - p_{ij})] = 0.\end{aligned}$$

The second equality uses lemma 1, while the third equality uses the definition of p_{ij}.

Risk adjustment is equivalent to control for the propensity score. To see this, note that

$$E[Z_{ij} \mid p_{i1}, \ldots, p_{iJ}] = p_{ij},$$

so the population regression of Z_{ij} on $p_{i1}, \ldots, p_{iJ}$ is p_{ij}. The auxiliary regression that partials out propensity scores therefore has residual $Z_{ij} - p_{ij}$. Equivalence then follows by standard multivariate regression algebra.

C. Equation (9)

Note that $Y - D\hat{\alpha} = Y - D\hat{\alpha}\hat{\varphi} + D\hat{\alpha}(\hat{\varphi} - 1)$ and that $(Y - D\hat{\alpha}\hat{\varphi})'\tilde{P}_{Z_\perp}D\hat{\alpha}(\hat{\varphi} - 1) = 0$, since the first-stage fitted values, $\tilde{P}_{Z_\perp}D\hat{\alpha}$, are necessarily orthogonal to the 2SLS residuals $Y - D\hat{\alpha}\hat{\varphi}$. Therefore,

$$
\begin{aligned}
&\frac{(Y - D\hat{\alpha})'\tilde{P}_{Z_\perp}(Y - D\hat{\alpha})}{\hat{\sigma}_\xi^2} \\
&= \frac{(Y - D\hat{\alpha}\hat{\varphi} + D\hat{\alpha}(\hat{\varphi} - 1))'\tilde{P}_{Z_\perp}(Y - D\hat{\alpha}\hat{\varphi} + D\hat{\alpha}(\hat{\varphi} - 1))}{\hat{\sigma}_\xi^2} \\
&= \frac{(D\hat{\alpha}(\hat{\varphi} - 1))'\tilde{P}_{Z_\perp}(D\hat{\alpha}(\hat{\varphi} - 1))}{\hat{\sigma}_\xi^2} \\
&\quad + \frac{(Y - D\hat{\alpha}\hat{\varphi})'\tilde{P}_{Z_\perp}(Y - D\hat{\alpha}\hat{\varphi})}{\hat{\sigma}_\xi^2} + 0 \\
&= \frac{(\hat{\varphi} - 1)^2}{\hat{\sigma}_\xi^2(\hat{\alpha}'D'\tilde{P}_{Z_\perp}D\hat{\alpha})^{-1}} + \frac{(Y - D\hat{\alpha}\hat{\varphi})'\tilde{P}_{Z_\perp}(Y - D\hat{\alpha}\hat{\varphi})}{\hat{\sigma}_\xi^2}.
\end{aligned}
$$

D. Lemma 3

Using the law of iterated expectations:

$$
\begin{aligned}
&E\left[\left(\sum_{\ell=1}^{L}\delta_\ell(Z_{i\ell} - p_{i\ell})\right)\left(\sum_{\ell=1}^{L}\psi_{\ell k}(Z_{i\ell} - p_{i\ell})\right)\right] \\
&= E\left[E\left[\left(\sum_{\ell=1}^{L}\delta_\ell(Z_{i\ell} - p_{i\ell})\right)\right.\right. \\
&\quad \left.\left.\times\left(\sum_{\ell=1}^{L}\psi_{\ell k}(Z_{i\ell} - p_{i\ell})\right) \mid \Pi, M, \tilde{Z}\right]\right] \\
&= E\left[E\left[\left(\sum_{\ell=1}^{L}\left(\sum_{j=1}^{J}\pi_{\ell j}E\left[\nu_j \mid \Pi, M, \tilde{Z}\right]\right)(Z_{i\ell} - p_{i\ell})\right)\right.\right. \\
&\quad \left.\left.\times\left(\sum_{\ell=1}^{L}\left(\sum_{j=1}^{J}M_{jk}\pi_{\ell j}\right)(Z_{i\ell} - p_{i\ell})\right) \mid \Pi, M, \tilde{Z}\right]\right] \\
&= 0,
\end{aligned}
$$

since $E\left[\nu_j \mid \Pi, M, \tilde{Z}\right] = 0$.

REFERENCES

Abaluck, J., M. M. C. Bravo, P. Hull, and A. Starc, "Mortality Effects and Choice across Private Health Insurance Plans," *Quarterly Journal of Economics* 136 (2021), 1557–1610. 10.1093/qje/qjab017

Abdulkadiroğlu, A., J. D. Angrist, Y. Narita, and P. A. Pathak, "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica* 85 (2017), 1373–1432.

——— "Breaking Ties: Regression Discontinuity Design Meets Market Design," *Econometrica* 90 (2022), 117–151.

Abdulkadiroğlu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters, "Do Parents Value School Effectiveness?" *American Economic Review* 110 (2020), 1502–1539.

Ainsworth, R., R. Dehijia, C. Pop-Eleches, and M. Urquiola, "Why Do Households Leave School Value Added on the Table? The Roles of Information and Preferences," *American Economic Review* 113 (2023), 1049–1082.

Andrabi, T., N. Bau, J. Das, and A. Khwaja, "Private Schooling, Learning, and Civic Values in a Low-Income Country," unpublished working paper (2020).

Angrist, J. D., "Grouped-Data Estimation and Testing in Simple Labor-Supply Models," *Journal of Econometrics* 47 (1991), 243–266. 10.1016/0304-4076(91)90101-I

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walters, "Interpreting Tests of School VAM Validity," *American Economic Review: Papers and Proceedings* 106 (2016), 388–392. 10.1257/aer.p20161080

——— "Leveraging Lotteries for School Value-Added: Testing and Estimation," *Quarterly Journal of Economics* 132 (2017), 871–919. 10.1093/qje/qjx001

Bacher-Hicks, A., T. J. Kane, and D. O. Staiger, "Validating Teacher Effect Estimates Using Changes in Teacher Assignments in Los Angeles," NBER working paper 20657 (2014).

Bekker, P., "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica* 62 (1994), 657–681. 10.2307/2951662

Borusyak, K., and P. Hull, "Non-Random Exposure to Exogenous Shocks," *Econometrica* (forthcoming).

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik, "Regression Discontinuity Designs Using Covariates," this REVIEW 101 (2019), 442–451.

Chetty, R., J. N. Friedman, and J. E. Rockoff, "Measuring the Impact of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review* 104 (2014a), 2593–2563. 10.1257/aer.104.9.2593

——— "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review* 104 (2014b), 2633–2679. 10.1257/aer.104.9.2633

Chetty, R., and N. Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates," *Quarterly Journal of Economics* 133 (2018), 1163–1228. 10.1093/qje/qjy006

Coleman, J. S., *Equality of Educational Opportunity (Summary Report)*, (Washington, DC: U.S. Department of of Health, Education, and Welfare, Office of Education, 1966).

Dale, S. B., and A. B. Krueger, "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables," *Quarterly Journal of Economics* 117 (2002), 1491–1527. 10.1162/003355302320935089

——— "Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data," *Journal of Human Resources* 49 (2014), 323–358. 10.1353/jhr.2014.0015

Deming, D., "Using School Choice Lotteries to Test Measures of School Effectiveness," *American Economic Review: Papers and Proceedings* 104 (2014), 406–411. 10.1257/aer.104.5.406

Hausman, J. A., "Specification Tests in Econometrics," *Econometrica* 46 (1978), 1251–1271. 10.2307/1913827

——— "Specification and Estimation of Simultaneous Equation Models" (pp. 391–448), in Z. Griliches and M. D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 1 (Amsterdam: North-Holland, 1983). 10.1016/S1573-4412(83)01011-9

Hull, P. D., "Estimating Hospital Quality with Quasi-Experimental Data," unpublished working paper (2018).

Kane, T. J., J. E. Rockoff, and D. O. Staiger, "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City," *Economics of Education Review* 27 (2008), 615–631. 10.1016/j.econedurev.2007.05.005

Kane, T. J., and D. O. Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," NBER working paper 14607 (2008).

Kolesár, M., R. Chetty, J. Friedman, E. Glaeser, and G. W. Imbens, "Identification and Inference with Many Invalid Instruments," *Journal of Business and Economic Statistics* 33 (2015), 474–484.

Ladd, H. F., and R. P. Walsh, "Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right," *Economics* *of Education Review* 21 (2002), 1–17. 10.1016/S0272-7757(00)00039-X

Morris, C. N., "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association* 78 (1983), 47–55. 10.1080/01621459.1983.10477920

Mountjoy, J., and B. Hickman, "The Returns to College(s): Estimating Value-Added and Match Effects in Higher Education," NBER working paper 29276 (2021).

Reardon, S. F., and S. W. Raudenbush, "Assumptions of Value-Added Models for Estimating School Effects," *Education Finance and Policy* 4 (2009), 492–519. 10.1162/edfp.2009.4.4.492

Robinson, P. M., "Root-N-Consistent Semiparametric Regression," *Econometrica* 56 (1988), 931–954. 10.2307/1912705

Roodman, D., "A Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics and Statistics* 71 (2009), 135–158. 10.1111/j.1468-0084.2008.00542.x

Rosenbaum, P. R., and D. B. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70 (1983), 41–55. 10.1093/biomet/70.1.41

Sargan, J., "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica* 26 (1958), 393–415. 10.2307/1907619

Tekwe, C. D., R. L. Carter, C.-X. Ma, J. Algina, M. E. Lucas, J. Roth, M. Ariet, T. Fisher, and M. B. Resnick, "An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance," *Journal of Educational and Behavioral Statistics* 29 (2004), 11–36. 10.3102/10769986029001011