

WORKING PAPER · NO. 2020-186

Simple and Credible Value-Added Estimation Using Centralized School Assignment

Joshua Angrist, Peter Hull, Parag Pathak, and Christopher Walters

DECEMBER 2020

Simple and Credible Value-Added Estimation

Using Centralized School Assignment*

Joshua Angrist[†] Peter Hull[‡] Parag Pathak[§] Christopher Walters[¶]

January 4, 2021

Abstract

Many large urban school districts match students to schools using algorithms that incorporate an element of random assignment. We introduce two simple empirical strategies to harness this randomization for value-added models (VAMs) measuring the causal effects of individual schools. The first estimator controls for the probability of being offered admission to different schools, treating the take-up decision as independent of potential outcomes. Randomness in school assignments is used to test this key conditional independence assumption. The second estimator exploits randomness in offers to generate instrumental variables (IVs) for school enrollment. This procedure uses a low-dimensional model of school quality mediators to solve the under-identification challenge arising from the fact that some schools are under-subscribed. Both approaches relax the assumptions of conventional value-added models while obviating the need for elaborate nonlinear estimators. In applications to data from Denver and New York City, we find that models controlling for both assignment risk and lagged achievement yield highly reliable VAM estimates. Estimates from models with fewer controls and older lagged score controls are improved markedly by IV.

*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 seminar participants at Berkeley for helpful input. Financial support from Arnold Ventures/LJAF and the National Science Foundation is gratefully acknowledged. Pathak 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're grateful to these districts for sharing data.

[†]MIT and NBER. Email: angrist@mit.edu

[‡]Corresponding author; University of Chicago and NBER. Email: hull@uchicago.edu

[§]MIT and NBER. Email: ppathak@mit.edu

[¶]University of California Berkeley and NBER. Email: crwalters@econ.berkeley.edu

1 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 policy-makers use the same sort of information when deciding whether to close, restructure, or expand 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. 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 identifies causal effects of school attendance.

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* which 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 and 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, 2019) 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 differences, 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 under-subscribed: that is, they have fewer applicants than seats. A conventional two-stage least squares (2SLS) model that uses school offers to instrument individual school attendance is therefore under-identified. Moreover, many over-subscribed schools face weak demand and, consequently, yield a weak first stage. This paper tackles the econometric issues arising in such common school assignment scenarios.

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 presumption that compliance with conditionally randomized offers is independent of

potential outcomes. We refer to 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 and Krueger (2002, 2014) and Mountjoy and Hickman (2020) studies of the earnings consequences of elite college attendance. Importantly, however, centralized school assignment facilitates validation of such conditional independence assumptions, a feature not seen in previous applications of this sort of identification strategy.

Our second estimator uses randomized school offers as instrumental variables (IVs) for school attended, conditional on the school assignment propensity score. Score conditioning ensures the validity of offer dummies as instruments, but under-subscription makes a conventional IV approach intractable. Our IV VAM approach solves the under-identification problem by modeling school value-added as a function of a few mediating school characteristics, estimated via the “many invalid instruments” framework of Kolesár et al. (2015). The IV VAM estimates are then used to construct empirical Bayes posterior predictions of value-added for individual schools. This approach builds on previous empirical Bayes strategies that combine observational and quasi-experimental estimates (Angrist et al., 2016, 2017; Chetty and Hendren, 2018; Hull, 2018) but which can be handicapped by computational complexity and the need for a host of auxiliary modeling assumptions. We show here how information on school characteristics and assignment risk can be used to simplify IV VAM estimation in markets where schools are under-subscribed.

We illustrate our two approaches 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. The Denver school assignment propensity score is estimated 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 non-lottery 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. (2019).

Estimates from both cities suggest that controlling for school assignment risk does a remarkably good job of controlling selection bias. Statistical tests that exploit quasi-experimental offer variation fail to reject the key conditional independence assumption underlying this first RC VAM approach. Perhaps surprisingly, the approach appears to be reliable even for NYC high schools, where it is not possible to control for lagged longer-run outcomes (here, SAT scores). We find that RC VAM estimates also beat conventional VAM estimates on mean-squared error (MSE) grounds, with bias reductions more than offsetting the modestly

increased variance from using narrower identifying variation.

Because RC VAMs appear to be virtually unbiased, there’s little scope for IV VAM to improve upon them in Denver and NYC. We show, however, that IV VAM can reduce bias and MSE when such VAMs are estimated with more limited controls than those specified by the RC 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 2020 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 other situations where information on lagged outcomes is dated or unavailable.

The remainder of the paper is organized as follows. Section 2 outlines a conceptual framework for school value-added estimation and testing in districts that centralize school assignment. Sections 3 and 4 describe our two new estimation strategies. Empirical results are discussed in Section 5. Section 6 summarizes and points to directions for further work.

2 Econometric Framework

Our theoretical analysis adds to a burgeoning literature on value-added measurement in education and other fields. Much of this work is inspired by the pioneering Kane and Staiger (2008) study of teachers, which applies empirical Bayes estimation to the study of teacher value-added. Key extensions and VAM applications exploiting quasi-experimental variation include Chetty et al. (2014a) for teachers and Deming (2014) and Angrist et al. (2017) for schools. Applications of an empirical Bayes framework for value-added estimation to other settings include Chetty and Hendren (2018), Hull (2018), and Abaluck et al. (2020). As in these previous studies, the foundation of our approach is a causal model describing the effects of attendance at individual schools on student outcomes. We next describe this setup.

2.1 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:

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

where $\gamma_j = E[Y_{ij}]$ and $\varepsilon_i \equiv Y_{ij} - E[Y_{ij}]$. For any two schools, j and k , the difference $\gamma_j - \gamma_k$ is 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 that $Y_{ij} - Y_{ik}$ is the same for all students. We refer to ε_i as student i 's *ability*.

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

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

where $\beta_0 = \frac{1}{J} \sum_{j=1}^J \gamma_j$ and $\beta_j = \gamma_j - \beta_0$.

The value-added parameters in equation (2) describe the effect of randomly generated changes in school enrollment on test score outcomes. Since school enrollment is not randomly assigned, however, 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 D_{ij} and ε_i . Consequently, a regression of Y_i on D_{ij} need not coincide with the causal model (2).

2.2 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; a typical priority allocates seats to the siblings of already-enrolled students first. Assignments are often made 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 who have the same preferences and priorities. The DA algorithms output a single school assignment for each student. Students may choose to enroll where matched, or enter a later less systematic round producing negotiated assignments. The randomness in 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 can be used as instrumental variables for the school enrollment dummies in equation (2).

We formalize this setting by denoting the partial ordering of schools on student i 's preference list by $\succ_i$ and by writing ρ_{ij} for student i 's priority at school j . The vector $\rho_i = (\rho_{i1}, \dots, \rho_{iJ})'$ 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 and outputs a set of indicators Z_{ij} , equal to one when the match offers student i a seat at school j . We collect these assignment indicators in the vector $Z_i = (Z_{i1}, \dots, Z_{iJ})'$. 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: $\varepsilon_i \perp\!\!\!\perp Z_i \mid \theta_i$.*

The maintained Assumption CRA suggests that we can estimate the causal effects of school assignment by comparing the outcomes of students within strata defined by their type. In practice, however, we see nearly as many preference and priority combinations as there are students in a match. Consequently, full type conditioning is impractical.

To solve the problem of a high-dimensional conditioning set, we pool students of different types 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, 2019). Student i 's propensity score for assignment to school j is defined as:

$$p_{ij} \equiv \Pr(Z_{ij} = 1 \mid \theta_i). \quad (3)$$

The p_{ij} are school assignment rates determined by repeatedly running the assignment algorithm, redrawing tie-breakers each time, while holding preferences and priorities fixed. The vector $p_i = (p_{i1}, \dots, p_{iJ})'$ collects student i 's propensity scores for all schools.

As first shown by Rosenbaum and 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:

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, where school assignment is random conditional on type, assignment is also random conditional on the school assignment propensity score.

Why is this conditional independence property useful? As in a complex stratified randomized trial, a signal feature of the school assignment setting is that the propensity score is a coarse function of observed conditioning variables. In particular, while type varies widely, assignment scores are determined by a few school-level cutoffs. We can therefore estimate causal effects of school offers by conditioning on assignment propensity scores rather than student type. Score control is especially straightforward in a linear constant-effects model

like equation (2). In this case, it's enough to subtract scores from offer dummies. Specifically, we have the following corollary to Lemma 1:

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 .*

By virtue of the Corollary, 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 reasoning that yields this result parallels that employed in the partially linear framework pioneered by Robinson (1988), 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 $Z_{ij} - p_{ij}$.¹

2.3 The Under-subscription Challenge

Abdulkadiroğlu et al. (2017, 2019) use orthogonality conditions analogous to those described by Corollary 1 to estimate effects of attending groups of schools seen as belonging 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 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 participating in the match.

Use of centralized assignment to estimate individual school value-added is more challenging because many schools are under-subscribed, in the sense of having more available seats than applicants who rank the school highly. Under-subscription of an entire sector is much less likely than individual school under-subscription. The propensity score for an under-subscribed school equals either zero or one for all applicants, depending on how each applicant ranks the school. Consequently, $Z_{ij} = p_{ij}$ for such schools, making risk-adjusted offers degenerate. With under-subscription, the causal model (2) is under-identified by risk-adjusted offers because the number of endogenous variables exceeds the number of non-degenerate instruments. The next two sections introduce strategies for addressing this challenge.

¹Proof of this and other key identification results appear in the Appendix. A Supplemental Appendix contains technical material related to parameter estimation and additional empirical results. Borusyak and Hull (2020) apply the risk adjustment idea to a broad class of research designs which use treatments or instruments computed from multiple sources of variation, only some of which are exogenous.

3 Risk-Controlled VAM

Our first approach to estimating value-added with centralized assignment data uses assignment risk controls to eliminate selection bias. This strategy is predicated on the assumption that conditional on applicant type, and potentially a vector of other observed pre-assignment covariates X_i , school choice is independent of potential outcomes and is therefore as good as randomized. We think of the X_i as conventional VAM controls; these include, for example, lagged test scores or student demographics. To formalize this identifying assumption, let $D_i = (D_{i1}, \dots, D_{iJ})'$ be a vector collecting the J school enrollment indicators. We then have a conditional independence assumption:

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

Our risk-controlled (RC) VAM estimator leverages this assumptions 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:

$$\varepsilon_i \perp\!\!\!\perp Z_i \mid (p_i, X_i). \quad (4)$$

Assumption CIA therefore follows from Lemma 1 in a scenario in which all match participants accept any offer yielded by the match. More generally, Assumption CIA overcomes the identification challenge raised by under-subscription by requiring that *compliance* with conditionally randomized school assignment offers be as good as randomized conditional on applicant type and covariates. Conventional VAM estimates invoke a stronger version of CIA by requiring conditional independence to hold conditional on covariates alone. Our desire to nest the conventional setup motivates inclusion of X_i in the RC VAM conditioning set.

The nature of Assumption CIA is illuminated by showing it to be a central-assignment analog of the Dale and Krueger (2002, 2014) identification strategy for the returns to college selectivity. The Dale and Krueger approach assumes that college enrollment decisions are made independently 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.² College application choices are a decentralized analog of

²Mountjoy and Hickman (2020) apply this strategy to estimate effects of college selectivity using admin-

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: $\varepsilon_i \perp\!\!\!\perp D_i \mid (p_i, X_i, Z_i) \implies \varepsilon_i \perp\!\!\!\perp D_i \mid (p_i, X_i)$.*

At first blush, the Dale and Krueger strategy includes an extra set of conditioning variables: the admissions offers, Z_i . Lemma 2 shows, however, that our CIA assumption follows from an extended conditional independence assumption that adds school assignment offers to the conditioning set, so that realized-offer conditioning is unnecessary. Lemma 2 therefore translates Dale-and-Krueger reasoning from the decentralized multi-offer college admissions world to a centralized single-offer setting. As we show below, the additional variation in Z_i conditional on (p_i, X_i) can be used to validate the key conditional independence assumption: an attractive feature unique to the centralized assignment setting.

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

$$Y_i = \alpha_0 + \sum_j \alpha_j D_{ij} + X_i' \Gamma + g(p_i) + \eta_i, \quad (5)$$

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 zero. In a model with 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 .) We're interested in the effects of school enrollment rather than offer effects, however. Dummies for zero risk therefore amount to an additional nonlinear control for applicant type. Control for zero-risk dummies is motivated by the fact that those with zero risk of an offer at j are almost all non-applicants to j . In the spirit of Dale and Krueger (2002), application behavior seems likely to be a potential source of omitted variables bias in cross-school comparisons. This parsimonious specification of $g(p_i)$ has $2J$ terms in a district with J schools.

istrative data from Texas. Abdulkadiroğlu et al. (2020) use a similar strategy to identify the determinants of parental preferences for NYC schools.

Testing the CIA with Centralized Assignment

The α_j in (5) coincide with causal value-added parameters β_j when the controls in equation (5) are sufficient to eliminate selection bias. This leads to the following null hypothesis for the validity of RC VAM and other OLS VAM estimators:

$$H_0 : \alpha_j = \beta_j; \quad j \in \{1, \dots, J\}. \quad (6)$$

A simple example shows how randomized offers can be used to test H_0 . Suppose $J = 2$ and we re-normalize 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 :

$$E[\eta_i | Z_{i1} = 1] - E[\eta_i | Z_{i1} = 0] = 0. \quad (7)$$

Substituting ε_i for η_i in this expression using a simplified (5) yields:

$$(E[Y_i | Z_{i1} = 1] - \alpha_1 E[D_{i1} | Z_{i1} = 1]) - (E[Y_i | Z_{i1} = 0] - \alpha_1 E[D_{i1} | Z_{i1} = 0]) = 0. \quad (8)$$

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

$$\frac{E[Y_i | Z_{i1} = 1] - E[Y_i | Z_{i1} = 0]}{E[D_{i1} | Z_{i1} = 1] - E[D_{i1} | Z_{i1} = 0]} = E[Y_i | D_{i1} = 1] - E[Y_i | D_{i1} = 0]. \quad (9)$$

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 (5), with under-subscription, note that even when $L < J$ the null hypothesis H_0 implies η_i is orthogonal to all risk-adjusted school offers. This orthogonality yields L testable restrictions (one for each over-subscribed school) of the form:

$$E[\eta_i (Z_{i\ell} - p_{i\ell})] = 0; \quad \ell \in \{1, \dots, L\}. \quad (10)$$

The restrictions in (10) are evaluated using the regression version of the Sargan test for instrument-error orthogonality (detailed in, e.g., Hausman (1983)). Let $\hat{\eta}_i$ denote the RC VAM residuals generated by OLS estimation of (5). The restrictions in (10) are tested by

asking whether $\tau_1 = \dots = \tau_L = 0$ in the regression equation:

$$\hat{\eta}_i = \tau_0 + \sum_{\ell=1}^L \tau_\ell Z_{i\ell} + \sum_{\ell=1}^L \mu_\ell p_{i\ell} + X_i' \Delta + \xi_i, \quad (11)$$

where ℓ indexes dummies $Z_{i\ell}$ indicating offers at $L < J$ over-subscribed schools. The propensity score controls on the right hand side ensure that this procedure tests risk-adjusted offer-residual orthogonality. The covariates X_i are included to increase statistical precision.

The omnibus specification test based on (11) admits a useful decomposition that distinguishes VAM bias on average from the bias of estimates for specific schools. First introduced by Angrist et al. (2016), this decomposition builds on the following forecast regression:

$$\beta_j = \varphi \alpha_j + \nu_j, \quad (12)$$

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. 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 zero over schools.

Assuming the residual ξ_i is conditionally homoskedastic, the omnibus test statistic evaluating $\tau_1 = \dots = \tau_L = 0$ in equation (11) can be written as the sum of a test of forecast bias and a test of idiosyncratic bias. The omnibus test statistic is:

$$\hat{T} = \frac{(Y - D\hat{\alpha})' P_{Z_\perp} (Y - D\hat{\alpha})}{\hat{\sigma}_\xi^2}, \quad (13)$$

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 over-subscribed assignments Z after partialling 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 (11)). Appendix A.3 shows that this test statistic can be decomposed as

$$\hat{T} = \frac{(\hat{\varphi} - 1)^2}{\hat{\sigma}_\xi^2 (\hat{\alpha}' D' P_{Z_\perp} D \hat{\alpha})^{-1}} + \frac{(Y - D\hat{\alpha}\hat{\varphi})' P_{Z_\perp} (Y - D\hat{\alpha}\hat{\varphi})}{\hat{\sigma}_\xi^2}, \quad (14)$$

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 over-subscribed 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, (12), into the causal model, (2), to obtain

$$Y_i = \beta_0 + \varphi \alpha_{j(i)} + \varepsilon_i + \nu_{j(i)}, \quad (15)$$

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 (15). The first term of equation (14) asks whether the estimated forecast coefficient yielded by these instruments is indeed statistically indistinguishable from one. 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 equal to one 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.

4 IV VAM in Under-identified Models

Our second approach leverages the fact that the randomness induced by centralized assignment can identify key features of the value-added distribution even when OLS VAM estimates are biased. Our IV VAM estimator exploits this variation by integrating a low-dimensional model of the determinants of school quality with RC VAM or other OLS estimates. 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 under-subscription. 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.

The foundations of this framework are first-stage and reduced form regressions of school

attendance and outcomes on risk-adjusted offers for admission to over-subscribed schools:

$$D_{ij} = \phi_j + \sum_{\ell=1}^L \pi_{\ell j}(Z_{i\ell} - p_{i\ell}) + u_{ij}, \quad j \in \{1, \dots, J\}, \quad (16)$$

and

$$Y_i = \kappa + \sum_{\ell=1}^L \rho_{\ell}(Z_{i\ell} - p_{i\ell}) + \omega_i. \quad (17)$$

The reduced form is obtained by substituting the first stages (16) into the causal model (2). It is convenient to array the $\pi_{\ell j}$ coefficients in equation (16) in the $L \times J$ matrix Π , and to collect the ρ_{ℓ} coefficients in equation (17) in the $L \times 1$ vector ρ . We then have $\rho = \Pi\beta$, where $\beta = (\beta_1, \dots, \beta_J)'$. When $L = J$, school value-added is identified by solving for $\beta = \Pi^{-1}\rho$. When $L < J$, because some schools are under-subscribed, 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. Specifically, we extend the simple school-level forecast regression (12) to a more general model for school value-added:

$$\beta_j = M_j' \varphi + \nu_j, \quad (18)$$

where $M_j = (M_{j1}, \dots, M_{jK})'$ 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) again captures the relationship between these school covariates and causal value-added, β_j . The forecast residual, ν_j , is again defined to be mean zero and uncorrelated with M_j across schools. Substituting (18) into (2) yields a generalization of equation (15):

$$Y_i = \beta_0 + M_{j(i)}' \varphi + \varepsilon_i + \nu_{j(i)}, \quad (19)$$

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 (19). This extends the testing procedure in Section 3 to allow for multiple mediators in $M_{j(i)}$ rather than RC VAM coefficients alone. Moreover, the forecast regression is no longer assumed to fit causal value-added perfectly, that is, $\nu_j \neq 0$. 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, (16), into the definition of $\nu_{j(i)}$ in (19) to obtain:

$$Y_i = \tilde{\beta}_0 + M'_{j(i)}\varphi + \sum_{\ell=1}^L \delta_\ell(Z_{i\ell} - p_{i\ell}) + \tilde{\varepsilon}_i, \quad (20)$$

where $\tilde{\beta}_0 = \beta_0 + \sum_j \nu_j \phi_j$, $\tilde{\varepsilon}_i = \varepsilon_i + \sum_j \nu_j u_{ij}$, and $\delta_\ell = \sum_j \pi_{\ell j} \nu_j$. The δ_ℓ parameters 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 violations of the exclusion restriction, 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 mediating variables as:

$$M_{j(i)k} = \psi_{0k} + \sum_{\ell=1}^L \psi_{\ell k}(Z_{i\ell} - p_{i\ell}) + v_{ik}, \quad k \in \{1, \dots, K\}, \quad (21)$$

with $\psi_{0k} = \sum_j M_{jk} \phi_j$, $v_{ik} = \sum_j M_{jk} u_{ij}$, and $\psi_{\ell k} = \sum_j M_{jk} \pi_{\ell j}$. The $\psi_{\ell k}$ 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:*

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

This assumption requires the first stage predicted values for mediators generated by (21) to be uncorrelated with the terms generating exclusion violations in (20).

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 zero 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 patterns across schools. 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 M_j and $\tilde{Z}$). 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 non-zero 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 over-identified models, Kolesár et al. (2015) propose a bias-corrected 2SLS estimator (B2SLS) that is consistent in a many-instrument asymptotic sequence similar to that used 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 an additional homoskedasticity assumption. In practice, B2SLS and 2SLS estimates of the parameters of interest here are virtually indistinguishable.

Empirical Bayes Posterior Predictions

In addition to characterizing the distribution of school quality, IV VAM estimates of φ and σ_ν^2 can be used to produce value-added predictions for individual schools. In contrast with RC VAM, however, which generates individual school VAM estimates directly, IV VAM requires an extra step. This step, sketched here (with proofs in the Supplemental Appendix), computes empirical Bayes posterior means from estimates of ρ , Π , φ , and σ_ν^2 .

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 (17), conditional on M , Π , and $\tilde{Z}$:

$$\beta^* = \arg \min_{b(\cdot)} E \left[(b(\hat{\rho}) - \beta)'(b(\hat{\rho}) - \beta) | \Pi, M, \tilde{Z} \right]. \quad (22)$$

As always, MSE is minimized by a conditional expectation, here $\beta^* = E[\beta | \hat{\rho}, \Pi, M, \tilde{Z}]$. 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:*

$$\beta^* = \Omega \hat{\rho} + (I - \Omega \Pi) M \varphi, \quad (23)$$

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

Equation (23) defines a set of hybrid value-added predictions determined by a linear combination of reduced-form offer effects and value-added as predicted by the forecast regression. 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 (23) yields an empirical Bayes posterior mean $\hat{\beta}^*$. These IV VAM predictions generalize the empirical Bayes shrinkage estimators commonly used to reduce MSE in noisy OLS estimates of VAM (as in, e.g., Kane et al. (2008) and Chetty et al. (2014a) for teachers).

Proposition 1 generates empirical Bayes 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 Normality of the error term ω_i in equation (17). Note also that β^* is the best linear predictor of β even when the ν_j are non-Normal:

Corollary 2. *Suppose $\hat{\rho} | (\rho, \Pi, M, \tilde{Z}) \sim N(\rho, \Sigma)$, $E[\nu | \Pi, M, \tilde{Z}] = 0$, and $Var(\nu | \Pi, M, \tilde{Z}) = \sigma_\nu^2 I$. Then the β^* in equation (23) solves (22) 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 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:

$\beta^* = M\varphi = \alpha$. The Supplemental Appendix extends this formula to allow for sampling variance in the OLS estimates, $\hat{\alpha}_j$, a generalization that is used in the empirical work. When $\hat{\alpha}_j$ is unbiased and the underlying errors are homoskedastic, the posterior again puts no weight on reduced-form estimates, simplifying to a conventional empirical Bayes shrinkage estimator of the form $\hat{\beta}_j^* = \lambda_j \hat{\alpha}_j$. The Supplemental Appendix also shows that when $L = J$, so that all schools are over-subscribed, the 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 hybrid 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.

5 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. (2019) data appendix. In both Denver and NYC, we standardize all achievement tests and SAT scores to have mean zero and standard deviation one 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-breaking number

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.

NYC applicants rank up to 12 academic programs in middle or high school. We aggregate schools with multiple programs at the school level. The NYC match features a variety of tie-breakers, with “unscreened” schools using a random lottery number and other “screened” schools using non-random tie-breakers such as past test scores. Propensity scores for NYC school assignment are computed as described in Abdulkadiroğlu et al. (2019). These scores depend in part on a bandwidth for screened school tie-breakers, similar to those used in standard regression discontinuity designs, which we calculate using a procedure suggested by Calonico et al. (2019). As in Abdulkadiroğlu et al. (2019), 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 bandwidth of each tie-breaking variable.

Table 1 describes the students and schools in the DPS and NYC samples. The first column shows statistics for the full sample of DPS middle school students, while column 2 shows statistics for DPS applicants with non-degenerate assignment risk (these students have $p_{ij} \in (0, 1)$ for at least one j). Columns 3-4 of the table report corresponding statistics for NYC middle school students, and columns 5-6 describe NYC high school students. RC VAM and other OLS VAM models are estimated using full samples, while bias tests and IV VAM estimation use the non-generate risk subsamples.

As is typical of large urban districts, most DPS and NYC students are disadvantaged, with over 70 percent 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 the full 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 virtually disappear. 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.³

³Balance 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. The risk controls here 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 don’t 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 un-

The bottom rows of Table 1 show that most Denver and New York schools are over-subscribed. Specifically, at least one student has non-degenerate risk at 68 out of 83 Denver middle schools, 448 out of 674 NYC middle schools, and 380 out of 537 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 under-subscribed, while others 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.

5.1 Evaluating RC VAM

The test characterized by equation (14) is used to compare the predictive validity of RC VAM estimates with results from three other OLS VAM estimators relying on fewer controls. The first of these is a benchmark model, labelled *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 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 (e.g., Chetty et al. 2014a). A third model, labelled *risk only*, omits conventional VAM controls but includes assignment propensity scores, p_i . RC VAM estimates come from regressions that combine conventional VAM controls with controls for school assignment risk.

Uncontrolled VAM estimates are clearly contaminated by selection bias. This can be seen in Table 2, which reports the components of the test statistic (14) for sixth grade math scores in each city. For the purpose of VAM testing, schools are classified into 20 bins defined by ventiles of the distribution of estimated conventional value-added. The testing equation, (11), is estimating using bin-level (rather than single-school) offers and propensity scores. This aggregation may increase test power relative to models using many school-specific offers as instruments for $\alpha_{j(i)}$ in (15) (see Roodman (2009) for a discussion of the connection between number of over-identifying restrictions and test power). Aggregation also facilitates a graphical depiction of VAM validity. 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 0.59 in NYC (shown in Panel B). These estimates are statistically different from one, while the overidentification and omnibus tests clearly reject the null hypothesis of zero bias in the uncontrolled model. Supplemental Appendix Table A3 reports results using

related to assigned school value-added conditional on assignment risk. This makes it unlikely that selective attrition biases estimates of reduced form school-offer effects.

alternative bin schemes; the choice of bin size matters little.

VAM research to date suggests that control 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 VAMs estimated with conventional controls boost the forecast coefficient markedly, yielding estimates of 1.12 in DPS and 0.93 in NYC. The former estimate is statistically indistinguishable from one, while the latter is marginally significantly different from one, 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, VAMs estimated with risk controls alone similarly improve greatly on the uncontrolled estimates. The risk-only model generates forecast coefficients of 0.75 and 0.82 in DPS and NYC. In contrast with the test results for conventional VAM, 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 schools, estimated RC VAM forecast coefficients are close to one, 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 0.75 and 0.63. 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 used to construct 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 (and the improvement of RC VAM over risk-only VAM) is therefore especially impressive.⁴

The tests reported in Tables 2 and 3 can be interpreted as asking whether VAM estimates

⁴Lagged score controls used to compute the estimates in Table 2 come from 5th grade, taken the year before the start of middle school; lagged score controls used to compute the estimates in Table 3 come from 8th grade, taken the year before the start of high school.

predict the causal effects of randomized school offers. Figure 1 presents a graphical summary of these predictions. Specifically, the figure plots reduced form offer effects for each value-added bin (the ρ_ℓ coefficients in equation (17)) 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 controls or risk controls pushes the forecast slopes towards one and reduces dispersion around the best-fitting lines. Including both sets of controls results in tightly-estimated relationships that are indistinguishable from the 45-degree line. Figure A1 and Table A4 in the Supplemental Appendix show broadly similar results for reading test scores.

5.2 IV VAM Estimates of Value Added

Our IV VAM application focuses on the NYC middle school sample, which includes the largest numbers of students and schools with assignment risk in the three samples we’ve looked at. The vector of mediators used to predict causal value-added includes a screened school indicator and the alternative 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 Table 2. 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 suitable for an asymptotic sequence in which the number of instruments grows in proportion to sample size.

Using an uncontrolled VAM estimate to predict causal value-added yields an estimated IV VAM forecast coefficient of only 0.25. This result, reported in the first column of Table 4, is qualitatively consistent with the test statistics reported in Table 2, which show strong evidence of forecast bias and a failure of uncontrolled VAM to satisfy the corresponding overidentification test.

The IV VAM procedure also generates an estimated screened-school effect on causal value added. This is negative and significantly different from zero in column 1. 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 from the forecast regression as $\sigma_\beta^2 = \text{Var}(M_j'\varphi) + \sigma_\nu^2$). 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 screened status ($\sigma_\nu = 0.13$) and a total value-added standard deviation of $\sigma_\beta = 0.19$. The latter estimate highlights the substantial variation in quality across NYC middle 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 the IV VAM forecast coefficients, to around 0.9 and 0.6 respectively, while also decreasing the estimated residual standard deviations. Screened school effects in these columns are smaller and not significantly different from zero, suggesting the larger negative estimate in the first column is driven by the diminished predictive power of uncontrolled VAM. IV VAM models generate roughly comparable estimates of σ_β in columns 1, 2, and 4 (around 0.2), though the estimate in column 3 is notably larger at around 0.28, perhaps reflecting failure of Assumption MIV in this case. It’s noteworthy that the forecast coefficient estimates for uncontrolled and risk-only VAM in Tables 2 and 4 differ markedly, though not so for conventional VAM. This reflects the near-unbiasedness of conventional VAM estimates. When the VAM estimates used as mediator are biased, 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 Table 3 in highlighting the essential unbiasedness of RC VAM. Including the RC VAM estimates as an IV VAM mediator yields a precisely estimated forecast coefficient of 0.98, with an estimated residual standard deviation of virtually zero. 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 one, though including these two highly correlated mediators reduces the precision of estimated forecast coefficients. This is impressive since conventional VAM is found to be a reliable predictor of causal value-added on its own; our estimates indicate that conventional VAM does not predict causal value-added conditional on RC VAM. The screened school coefficient estimates in these columns are likewise precisely estimated zeros.

On balance, there would seem to be little scope for IV VAM posterior predictions to improve on RC VAM. Conventional VAMs also generate generally accurate predictions of causal effects. It’s important to note, however, that both of these models rely on controls for lagged test scores measured one year prior to the outcome. In practice, these controls may be unavailable. For example, the 2020 COVID-19 pandemic has disrupted standardized testing in many districts, resulting in a year of missing test scores. 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 measurements that control for

test scores lagged by more than one year. To investigate the consequences of longer lags, column 6 of Table 4 reports IV VAM estimates from a procedure where the mediator is a set of conventional VAM estimates computed using more distant lagged score controls. The estimates in this case replace 5th grade scores with 3rd grade scores in a conventional VAM for sixth-grade outcomes. Using earlier lagged score controls reduces the forecast coefficient to 0.80 with a forecast residual standard deviation of $\sigma_\nu = 0.07$. This seems a noteworthy 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.

5.3 The MSE of Estimated VAMs

MSE provides a natural standard of comparison for IV VAM and OLS VAM estimators computed using alternative sets of controls. Although RC VAM and other OLS value-added estimators generate estimates of value-added for each school, with no further computation required, an empirical Bayes estimator can reduce MSE in this context by shrinking the OLS estimates toward zero. Importantly, MSE here is with reference to OLS VAM parameters, initially setting aside the question of whether these have a causal interpretation.

Given a set of OLS VAM estimates, $\hat{\alpha}_j$, the MSE-minimizing predictions of the corresponding OLS coefficients, α_j , are given by:

$$\alpha_j^* = \lambda_j \hat{\alpha}_j, \quad \lambda_j = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + s_j^2}, \quad (24)$$

where σ_α^2 is the variance of the OLS coefficients and s_j^2 is the sampling variance of $\hat{\alpha}_j$. The shrinkage formula in (24) coincides with the conditional expectation $E[\alpha_j | \hat{\alpha}_j]$ when $(\alpha_j, \hat{\alpha}_j)$ are joint Normal and uncorrelated across schools, and gives a best linear approximation to this conditional expectation outside of Normality. As in Morris (1983), an empirical Bayes estimator based on OLS plugs estimates of σ_α^2 and s_j into equation (24). To that end, s_j is taken to be the standard error of $\hat{\alpha}_j$, while σ_α^2 is estimated by the variance of estimated $\hat{\alpha}_j$'s minus the average squared standard error.

We turn next to MSE calculations with reference to a causal VAM target. The MSE of α_j^* as an estimator of causal value-added, β_j , is given by:

$$E[(\alpha_j^* - \beta_j)^2] = E[\lambda_j^2 s_j^2] + \sigma_\alpha^2 E[(\varphi - \lambda_j)^2] + \sigma_\nu^2. \quad (25)$$

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 somewhat more involved calculation in the Supplemental Appendix derives the MSE of IV VAM posterior predictions, with a

parallel decomposition into bias and sampling variance. As with α_j^* , the MSE of IV VAM is computed by plugging estimates of forecast coefficients, residual variances, and sampling variances of VAM parameters into the relevant formula.

Figure 2 compares the root mean squared error (RMSE) of OLS VAM and IV VAM posterior predictions of causal value added, computed by applying (25) and the appendix formulas to the parameter estimates summarized in Table 4.⁵ RMSE is minimized by the RC VAM posterior, at around 0.06 standard deviations in the distribution of middle school math scores. Because RC VAM is essentially unbiased, the IV VAM posterior for this model coincides with that for OLS. Not surprisingly, sampling variance contributes far more to the MSE of RC VAM than bias. What bias there is arises from shrinkage (i.e., $\lambda_j \neq 1$).

OLS VAM estimates generate posterior predictions with higher RMSE than RC VAM, but this gap is reduced by IV VAM. The gains are most impressive for risk-only VAM, the IV version of which cuts RMSE almost in half. This result can be read as saying that IV VAM is a good substitute for lagged score controls, a point with high empirical relevance. By contrast, IV VAM improves much less on conventional estimates, a result consistent with Table 2. Interestingly, the RMSE of the conventional model exceeds the RMSE 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.

The last two columns in Figure 2 compare RMSE for OLS and IV VAM versions of conventional VAM estimated with older lagged score controls. Consistent with the estimates in the last column of Table 4, use of 3rd instead of 5th grade score controls increases RMSE substantially. IV VAM reduces the resulting bias markedly, cutting RMSE by about 20% relative to the OLS analog predictions and closing roughly half the gap with RC VAM. This highlights the value of IV VAM estimation in applications with missing or degraded lagged achievement controls.

6 Summary and Conclusions

VAM estimates may help families choose schools wisely, perhaps with life-changing consequences. Policy-makers and educators likewise base high-stakes decisions related to school access, expansion, and closure on VAM; the federal government and many states require this. Given these stakes, how should the consequences of attendance at individual schools be estimated? Many school accountability frameworks report rankings which amount to unad-

⁵The estimates in Table 4 imply different total variances of value-added. We therefore divide RMSE for each model by the ratio of the estimate of σ_β for that model to the corresponding estimate for the RC VAM model. This puts RMSE on a common scale across models.

justed mean outcomes for each school. Such poorly controlled VAM estimates confuse school quality with the socioeconomic status and racial make-up of student bodies. In large urban districts like those examined here, schools sporting the highest test scores and graduation rates tend to have an outsize share of non-minority students. These schools are also found in wealthier neighborhoods.

The primary econometric challenge in this context is how to eliminate or at least 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 fully 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 under-subscription. Moreover, the assumptions justifying RC VAM are easily validated by testing whether RC VAM residuals are orthogonal to offers of seats at over-subscribed schools. Application of this test to schools in Denver and NYC suggest 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’ve also introduced an IV VAM estimator that exploits reduced forms for causal VAM 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). The results reported here show that IV VAM can improve on poorly-controlled estimates of school quality and on estimates that rely on older and perhaps less relevant controls.

Farther 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’s 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.

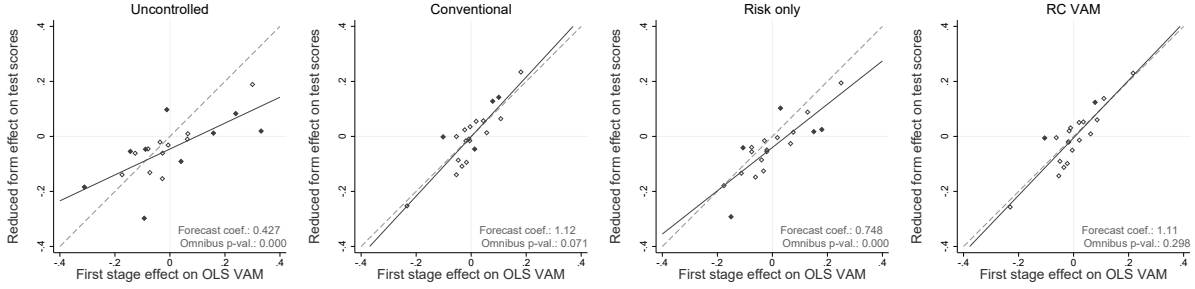
References

- ABALUCK, J., M. M. C. BRAVO, P. HULL, AND A. STARC (2020): “Mortality Effects and Choice Across Private Health Insurance Plans,” NBER Working Paper 27578.
- ABDULKADIROĞLU, A., J. D. ANGRIST, Y. NARITA, AND P. A. PATHAK (2017): “Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation,” *Econometrica*, 85, 1373–1432.
- (2019): “Breaking Ties: Regression Discontinuity Design Meets Market Design,” MIT SEII Working Paper.
- ABDULKADIROĞLU, A., P. A. PATHAK, J. SCHELLENBERG, AND C. R. WALTERS (2020): “Do Parents Value School Effectiveness?” *American Economic Review*, 110, 1502–39.
- ANGRIST, J. D. (1991): “Grouped-Data Estimation and Testing in Simple Labor-Supply Models,” *Journal of Econometrics*, 47, 243–266.
- ANGRIST, J. D., P. D. HULL, P. A. PATHAK, AND C. R. WALTERS (2016): “Interpreting Tests of School VAM Validity,” *American Economic Review: Papers & Proceedings*, 106, 388–392.
- (2017): “Leveraging Lotteries for School Value-Added: Testing and Estimation,” *Quarterly Journal of Economics*, 132, 871–919.
- BACHER-HICKS, A., T. J. KANE, AND D. O. STAIGER (2014): “Validating Teacher Effect Estimates Using Changes in Teacher Assignments in Los Angeles,” NBER Working Paper 20657.
- BEKKER, P. (1994): “Alternative Approximations to the Distributions of Instrumental Variable Estimators,” *Econometrica*, 62, 657–681.
- BORUSYAK, K. AND P. HULL (2020): “Non-Random Exposure to Exogenous Shocks: Theory and Applications,” NBER Working Paper 27845.
- CALONICO, S., M. D. CATTANEO, M. H. FARRELL, AND R. TITIUNIK (2019): “Regression Discontinuity Designs Using Covariates,” *Review of Economics and Statistics*, 101, 442–451.
- CHETTY, R., J. N. FRIEDMAN, AND J. E. ROCKOFF (2014a): “Measuring the Impact of Teachers I: Evaluating Bias in Teacher Value-Added Estimates,” *American Economic Review*, 104, 2593–2563.
- (2014b): “Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood,” *American Economic Review*, 104, 2633–2679.
- CHETTY, R. AND N. HENDREN (2018): “The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates,” *Quarterly Journal of Economics*, 133, 1163–1228.

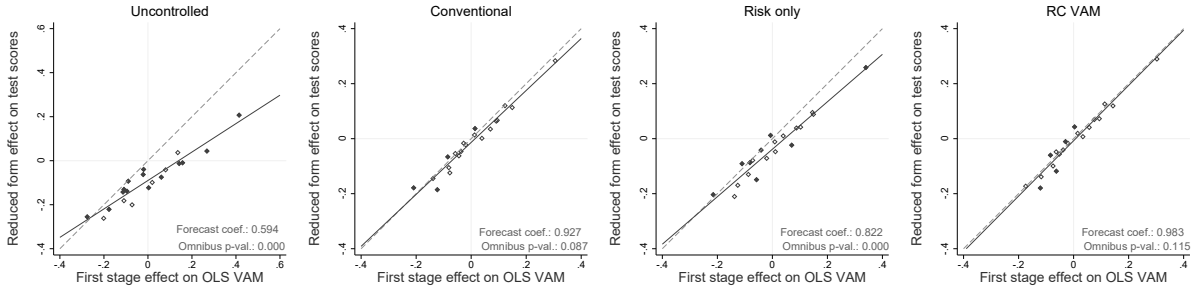
- DALE, S. B. AND A. B. KRUEGER (2002): “Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables,” *Quarterly Journal of Economics*, 117, 1491–1527.
- (2014): “Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data,” *Journal of Human Resources*, 49, 323–358.
- DEMING, D. (2014): “Using School Choice Lotteries to Test Measures of School Effectiveness,” *American Economic Review: Papers & Proceedings*, 104, 406–411.
- DONALD, S. G. AND W. K. NEWKEY (2001): “Choosing the Number of Instruments,” *Econometrica*, 69, 1161–1191.
- HAUSMAN, J. A. (1978): “Specification Tests in Econometrics,” *Econometrica*, 46, 1251–1271.
- (1983): “Specification and Estimation of Simultaneous Equation Models,” in *Handbook of Econometrics*, ed. by Z. Griliches and M. D. Intriligator, North-Holland, vol. 1, 391–448.
- HULL, P. D. (2018): “Estimating Hospital Quality with Quasi-Experimental Data,” Available at SSRN: <https://ssrn.com/abstract=3118358>.
- KANE, T. J., J. E. ROCKOFF, AND D. O. STAIGER (2008): “What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City,” *Economics of Education Review*, 27, 615–631.
- KANE, T. J. AND D. O. STAIGER (2008): “Estimating Teacher Impacts on Student Achievement: an Experimental Evaluation,” NBER Working Paper 14607.
- KOLESÁR, M., R. CHETTY, J. FRIEDMAN, E. GLAESER, AND G. W. IMBENS (2015): “Identification and Inference With Many Invalid Instruments,” *Journal of Business and Economic Statistics*, 33, 474–484.
- MORRIS, C. N. (1983): “Parametric Empirical Bayes Inference: Theory and Applications,” *Journal of the American Statistical Association*, 78, 47–55.
- MOUNTJOY, J. AND B. HICKMAN (2020): “The Returns to College(s): Estimating Value-Added and Match Effects in Higher Education,” Becker Friedman Institute Working Paper.
- NAGAR, A. L. (1959): “The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations,” *Econometrica*, 27, 575–595.
- ROBINSON, P. M. (1988): “Root-N-Consistent Semiparametric Regression,” *Econometrica*, 56, 931–954.
- ROODMAN, D. (2009): “A Note on the Theme of Too Many Instruments,” *Oxford Bulletin of Economics and Statistics*, 71, 135–158.
- ROSENBAUM, P. R. AND D. B. RUBIN (1983): “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70, 41–55.
- SARGAN, J. (1958): “The Estimation of Economic Relationships Using Instrumental Variables,” *Econometrica*, 26, 393–415.

Figure 1. Visual Instrumental Variables Tests for Bias (Math)

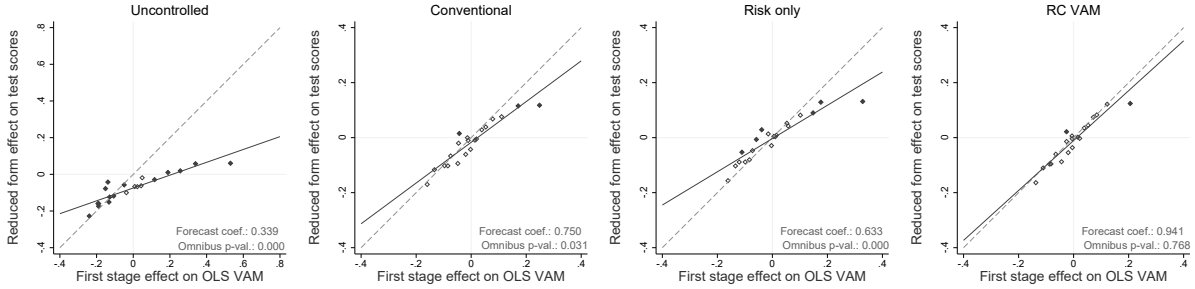
A. Denver middle schools



B. NYC middle schools



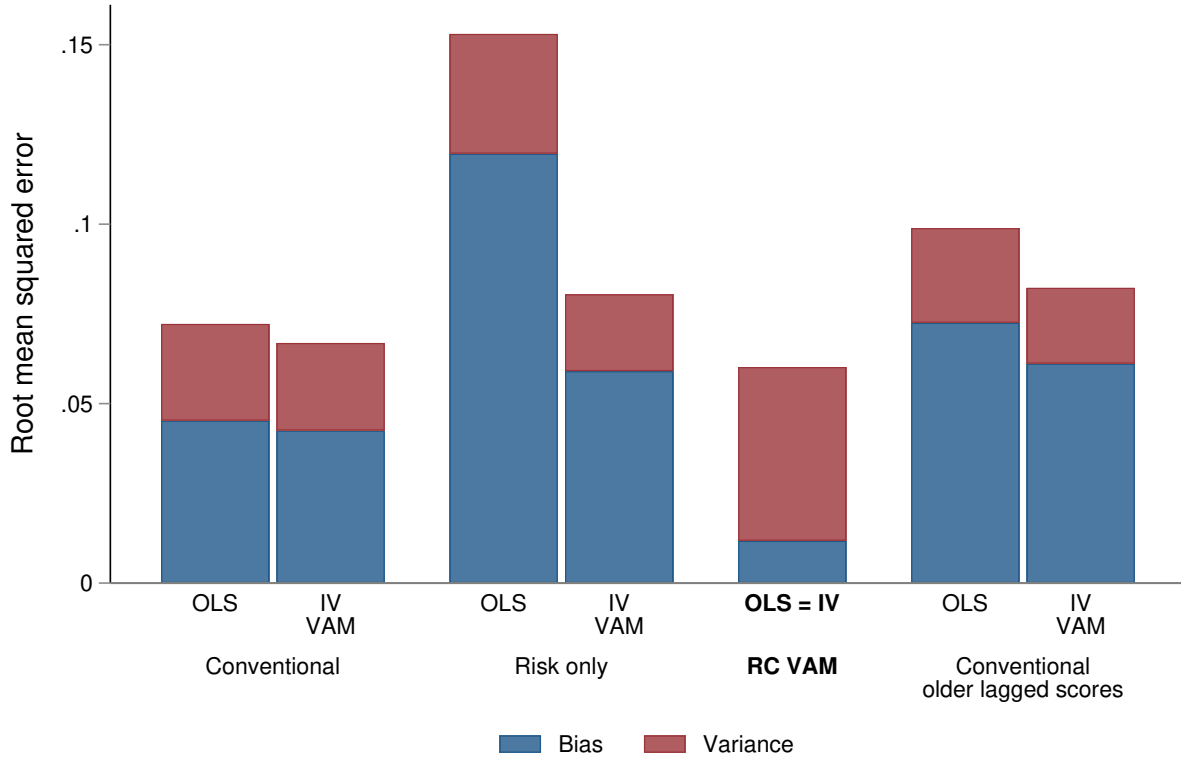
C. NYC high schools



— Forecast coef. line - - - 45 degree line

Notes: This figure plots reduced-form estimates against value-added first stages from each of 20 school assignment bins. Outcomes are 6th grade math CSAP and CMAS scores for Denver, 6th 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 zero and standard deviation one 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, while dashed lines indicate the 45-degree line.

Figure 2. RMSE of Value-Added Estimates for NYC Middle Schools



Notes: This figure plots root mean squared error (RMSE) for posterior predictions of value-added generated by the models and outcomes used for Table 4. 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 indicates RMSE. Blue and red shading mark the shares of MSE due to bias and variance, respectively. Conventional and RC VAM posterior predictions are from models using 5th grade tests as lagged score controls, as in Table 2. The conventional VAM posterior with older lagged scores relies on 3rd grade tests for lagged score controls, as in column 6 of Table 4. Section 5 details the calculations used to produce this figure.

Table 1. Descriptive Statistics

	Denver middle schools		NYC middle schools		NYC high schools	
	All	With risk	All	With risk	All	With risk
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Demographics</i>						
Hispanic	0.592	0.580	0.413	0.445	0.382	0.431
Black	0.126	0.140	0.232	0.254	0.283	0.279
White	0.210	0.202	0.153	0.110	0.136	0.106
Female	0.493	0.494	0.492	0.483	0.525	0.522
Free/reduced price lunch	0.723	0.703	0.753	0.787	0.775	0.793
Special education	0.102	0.087	0.208	0.217	0.127	0.067
English language learner	0.393	0.415	0.098	0.099	0.093	0.087
<i>Baseline scores</i>						
Math (standardized)	0.000	0.079	0.000	-0.052	0.000	-0.038
ELA (standardized)	0.000	0.071	0.000	-0.044	0.000	-0.035
<i>Enrollment</i>						
Screened	0.000	0.000	0.055	0.039	0.110	0.065
Lottery	1.000	1.000	0.945	0.961	0.890	0.935
Students	37,101	8,116	185,988	46,410	121,583	32,432
Schools	83	76	674	623	537	516
Lotteries (schools with risk)		68		448		380

Notes: This table describes the study samples in Denver and NYC. Column 1 reports descriptive statistics for Denver students enrolled in 6th grade in the 2012-13 through 2018-19 school years. Column 3 reports statistics for NYC middle school students enrolled in 6th grade in the 2016-17 through 2018-19 school years. Column 5 reports statistics for NYC high school students enrolled in 9th 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 5th grade for middle school samples and 8th grade for high school samples. Baseline scores are standardized to be mean zero and standard deviation one in the student-level test score distribution, separately by year.

Table 2. VAM Bias Tests for Middle School Math Scores

	Uncontrolled	Conventional	Risk only	RC VAM
	(1)	(2)	(3)	(4)
Panel A. <i>Denver middle schools</i>				
Forecast coefficient	0.427 (0.059)	1.12 (0.106)	0.748 (0.085)	1.11 (0.102)
First stage F statistic	48.5	104	41.7	97.1
<i>Bias tests:</i>				
Forecast bias	95.3 [0.000]	1.21 [0.272]	8.72 [0.003]	1.21 [0.271]
Overidentification (19 d.f.)	79.0 [0.000]	28.7 [0.070]	53.6 [0.000]	21.6 [0.305]
Omnibus (20 d.f.)	174 [0.000]	29.9 [0.071]	62.3 [0.000]	22.8 [0.298]
N (testing)			7,661	
N (VAM estimation)			37,101	
Panel B. <i>NYC middle schools</i>				
Forecast coefficient	0.594 (0.030)	0.927 (0.041)	0.822 (0.039)	0.983 (0.044)
First stage F statistic	153	648	171	530
<i>Bias tests:</i>				
Forecast bias	188 [0.000]	3.13 [0.077]	20.8 [0.000]	0.152 [0.697]
Overidentification (19 d.f.)	84.5 [0.000]	25.9 [0.134]	52.9 [0.000]	27.6 [0.091]
Omnibus (20 d.f.)	273 [0.000]	29.0 [0.087]	73.7 [0.000]	27.8 [0.115]
N (testing)			44,758	
N (VAM estimation)			185,988	

Notes: 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, 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 zero and standard deviation one 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.

Table 3. VAM Bias Tests for SAT Math Scores, NYC High Schools

	Uncontrolled (1)	Conventional (2)	Risk only (3)	RC VAM (4)
Forecast coefficient	0.339 (0.030)	0.750 (0.064)	0.633 (0.055)	0.941 (0.080)
First stage F statistic	148	206	80.9	148
<i>Bias tests:</i>				
Forecast bias	476 [0.000]	15.1 [0.000]	44.3 [0.000]	0.538 [0.463]
Overidentification (19 d.f.)	26.8 [0.110]	18.2 [0.511]	19.0 [0.456]	14.6 [0.748]
Omnibus (20 d.f.)	503 [0.000]	33.3 [0.031]	63.3 [0.000]	15.1 [0.768]
N (testing)			30,158	
N (VAM estimation)			121,583	

Notes: 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.

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

	(1)	(2)	(3)	(4)	(5)	(6)
Mediators						
Uncontrolled VAM	0.253 (0.029)					
Conventional VAM		0.914 (0.043)			-0.134 (0.131)	
Risk only VAM			0.596 (0.040)			
RC VAM				0.982 (0.040)	1.14 (0.071)	
Conventional VAM (older lagged scores)						0.801 (0.043)
Screened sector dummy	-0.132 (0.050)	-0.025 (0.037)	-0.065 (0.043)	0.007 (0.035)	0.020 (0.034)	-0.057 (0.040)
Standard deviations						
Value-added	0.188	0.204	0.277	0.213	0.214	0.207
Forecast residual	0.128	0.052	0.094	0.013	0.012	0.073
First stage F statistic	28.2	39.6	26.4	38.4	39.3	32.4
N			46,410			

Notes: 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 5th grade tests as lagged score controls, as in Table 2. The conventional VAM mediator used to compute the estimates in column 6 relies on 3rd grade tests for lagged score controls. Test scores are standardized to be mean zero and standard deviation one in the student-level test score distribution, separately by year. Standard errors are reported in parentheses.

A Appendix Proofs

A.1 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 (4) 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 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.

A.2 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

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

so the population regression of Z_{ij} on $p_{i1}, \dots, p_{iJ}$ 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.

A.3 Equation (14)

Note that $Y - D\hat{\alpha} = Y - D\hat{\alpha}\hat{\varphi} + D\hat{\alpha}(\hat{\varphi} - 1)$ and that

$$(Y - D\hat{\alpha}\hat{\varphi})' \tilde{P}_Z D\hat{\alpha}(\hat{\varphi} - 1) = 0,$$

since the first-stage fitted values, $\tilde{P}_Z D\hat{\alpha}$, are necessarily orthogonal to the 2SLS residuals $Y - D\hat{\alpha}\hat{\varphi}$. Therefore,

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

A.4 Lemma 3

Using the law of iterated expectations:

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

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

Simple and Credible Value-Added Estimation Using Centralized School Assignment: Supplemental Appendix

Joshua Angrist, MIT & NBER

Peter Hull, U Chicago & NBER

Parag Pathak, MIT & NBER

Christopher Walters, UC Berkeley & NBER

January 2021

Contents

I. IV VAM Consistency	1
II. Empirical Bayes Formulas	3
Accounting for First-Stage Estimation Error	4
Posterior Means Without Under-subscription	5
Posterior Means that Account for OLS Sampling Variance	5
III. Mean Squared Error Calculations	6
IV. Appendix Figures and Tables	8

I. IV VAM Consistency

The IV VAM estimation framework draws on Kolesár et al. (2015). Equations (20) and (21) generalize equation (3.1) in Kolesár et al. (2015) to incorporate multiple endogenous variables $M_{j(i)k}$. The Kolesár et al. (2015) consistency assumptions (1, 2(i), 3, 4, and 5) have parallels here as follows:

Assumption IVR. *The instruments and residuals $\tilde{Z}_i \in \mathbb{R}^{L_N}$, $\tilde{\varepsilon}_i \in \mathbb{R}$, $\tilde{v}_i \in \mathbb{R}^K$, for $i = 1, \dots, N$, $N = 1, \dots$ are triangular arrays of random variables with $(\tilde{Z}_i, \tilde{\varepsilon}_i, \tilde{v}_i)$, $i = 1, \dots, N$ exchangeable. $(\tilde{Z}, 1)$ is full column rank with probability one.*

Assumption REG. *$(\tilde{\varepsilon}_i, \tilde{v}_i)' \mid \tilde{Z}$ are iid with zero mean, a positive definite covariance matrix, and finite fourth moments.*

Assumption NIV. *The number of instruments satisfies $L/N = \alpha + o(N^{-1/2})$ for $\alpha \in [0, 1)$.*

Assumption CPM. *The $(1+K) \times (1+K)$ matrix $\Lambda_N/N = (\delta \ \psi)' \tilde{Z}'_{\perp} \tilde{Z}_{\perp} (\delta \ \psi)/N$ converges in probability to a positive semidefinite concentration parameter matrix Λ , with $\Lambda_{2 \dots K, 2 \dots K}$ positive definite. Furthermore, $E[\Lambda_N/N] \rightarrow \Lambda$.*

Assumption ZC. $\Lambda_{1k} = 0$ for $k = 2, \dots, 1+K$.

Here $\tilde{Z}_i$ denotes the i th observation on the centered offer matrix $\tilde{Z}$, $\tilde{v} = (\tilde{v}_{i1}, \dots, \tilde{v}_{iK})$ collects the first stage residuals, and $\tilde{Z}_{\perp}$ collects demeaned $\tilde{Z}_{i\ell}$. The setup for the instruments and residuals (Assumption IVR), regularity conditions (Assumption REG), assumption on the number of instruments (Assumption NIV), and first assumption on the concentration parameter matrix (Assumption CPM) are standard in Bekker many-instrument asymptotics. As in Kolesár et al. (2015) we strengthen the last condition slightly and adopt their novel zero correlation condition (Assumption ZC) which implies the large-sample orthogonality of exclusion restriction violations and first stage effects. We again extend this condition to allow for multiple endogenous variables, and note that it is satisfied under the mean-independence of ν we use in Lemma 3. Formally, for each k , we have:

$$\begin{aligned} E[\Lambda_{N,1k}/N] &= E[\delta' \tilde{Z}'_{\perp} \tilde{Z}_{\perp} \psi_k / N] \\ &= E[E[\nu' \mid \Pi, M, \tilde{Z}] \Pi' \tilde{Z}'_{\perp} \tilde{Z}_{\perp} \Pi M_k / N] \\ &= 0 \end{aligned}$$

when $E[\nu \mid \Pi, M, \tilde{Z}] = 0$. It follows that $\Lambda_{1k} = 0$ for each k under the convergence condition in Assumption CPM.

Under these five assumptions, Kolesár et al. (2015) show the consistency of a bias-corrected 2SLS estimator of the form

$$\hat{\varphi} = \left(\tilde{M}'_{\perp} (I - kR_{\tilde{Z}_{\perp}}) \tilde{M}_{\perp} \right)^{-1} \left(\tilde{M}'_{\perp} (I - kR_{\tilde{Z}_{\perp}}) Y_{\perp} \right),$$

where $\tilde{M}$ is an $N \times K$ matrix collecting observations of demeaned endogenous variables $M_{j(i)k}$, $Y_{\perp}$ is an $N \times 1$ vector collecting observations of the outcome Y_i , k is a scalar, and $R_{\tilde{Z}_{\perp}}$ is the residual-maker matrix of $\tilde{Z}_{\perp}$; i.e. $R_{\tilde{Z}_{\perp}} = I - P_{\tilde{Z}_{\perp}}$ for the projection matrix $P_{\tilde{Z}_{\perp}} = \tilde{Z}_{\perp}(\tilde{Z}'_{\perp}\tilde{Z}_{\perp})^{-1}\tilde{Z}'_{\perp}$. Kolesár et al. (2015) establish $\hat{\varphi} \xrightarrow{p} \varphi$ for $k = 1/(1 + \alpha)$. This aligns $\hat{\varphi}$ with the estimator that Donald and Newey (2001) propose for many-instrument settings, building on Nagar (1959). Kolesár et al. (2015) also consider a modified version of this estimator which accounts for many control variables which we use in the empirical applications. In practice we find both bias-corrected estimators to yield very similar results to conventional 2SLS, which sets $k = 1$.

To estimate σ_{ν} , we modify an estimator for Λ_{11} which Kolesár et al. (2015) show to be consistent under additional assumptions and which they propose (and we use) for estimating the asymptotic variance of $\hat{\varphi}$. The modification follows from the fact that when $E[\nu \mid \Pi, W, \tilde{Z}] = 0$ and $Var(\nu \mid \Pi, M, \tilde{Z}) = \sigma_{\nu}^2 I$ as in Corollary 2,

$$\begin{aligned} E[\Lambda_{N,11}/N] &= E[\delta' \tilde{Z}'_{\perp} \tilde{Z}_{\perp} \delta / N] \\ &= E[\nu' \Pi' \tilde{Z}'_{\perp} \tilde{Z}_{\perp} \Pi \nu / N] \\ &= tr(E[\Pi' \tilde{Z}'_{\perp} \tilde{Z}_{\perp} \Pi]) \sigma_{\nu}^2 / N. \end{aligned}$$

Thus, if $\hat{\Lambda}_{11}$ is a consistent estimator of Λ_{11} (implying $\hat{\Lambda}_{11} \xrightarrow{p} E[\Lambda_{N,11}/N]$ under Assumption CPM) and $\hat{t}$ is consistent for $tr(E[\Pi' \tilde{Z}'_{\perp} \tilde{Z}_{\perp} \Pi]) / N$ (i.e. $\hat{t} - tr(E[\Pi' \tilde{Z}'_{\perp} \tilde{Z}_{\perp} \Pi]) / N \xrightarrow{p} 0$) then by the continuous mapping theorem we have the consistent estimator $\hat{\sigma}_{\nu} = \sqrt{\hat{\Lambda}_{11} / \hat{t}} \xrightarrow{p} \sigma_{\nu}$. Following Kolesár et al. (2015), we use

$$\hat{\Lambda}_{11} = \max\{(Y - \tilde{M}\hat{\varphi})'(P_{\tilde{Z}_{\perp}} / (N - L) - R_{\tilde{Z}_{\perp}}(L/N^2))(Y - \tilde{M}\hat{\varphi}), 0\}$$

and further use

$$\hat{t} = tr(\hat{\Pi} \tilde{Z}'_{\perp} (P_Z - \tilde{Z}_{\perp} \hat{\psi} (\hat{\psi}' \tilde{Z}'_{\perp} \tilde{Z}_{\perp} \hat{\psi})^{-1} \hat{\psi}' \tilde{Z}'_{\perp}) \tilde{Z}'_{\perp} \hat{\Pi}) / N.$$

These formulas are derived from Lemmas 1 and 2 in Bekker (1994), as detailed in Appendix A of Kolesár et al. (2015).

Finally, it's worth noting that the propensity score formulas derived in Abdulkadiroğlu

et al. (2017) and Abdulkadiroğlu et al. (2019) use a large-market approximation that fixes the number of schools while increasing the number of students participating in the match. This large-market sequence scales applicants and school capacities in proportion, thereby generating admissions cutoffs that are fixed across matches. The Bekker sequence, by contrast, increases the number of instruments and therefore (implicitly) the number of schools. It remains to derive propensity scores for matching markets under this sequence. Importantly, however, the number of students per school in the matching markets studied here appears to be large enough for a large-market sequence to yield estimated propensity scores that balance student characteristics conditional on school offers.

II. Empirical Bayes Formulas

Proposition 1 follows from the observation that under the assumptions $\hat{\rho}$ and β are jointly normally distributed, conditional on $(M, \Pi, \tilde{Z})$. Specifically, since $\rho = \Pi\beta$ and $\beta = M\varphi + \nu$,

$$\begin{bmatrix} \hat{\rho} \\ \beta \end{bmatrix} \mid (M, \Pi, \tilde{Z}) \sim N \left(\begin{bmatrix} \Pi M\varphi \\ M\varphi \end{bmatrix}, \begin{bmatrix} \Pi\Pi'\sigma_\nu^2 + \Sigma & \Pi\sigma_\nu^2 \\ \Pi'\sigma_\nu^2 & \sigma_\nu^2 I \end{bmatrix} \right).$$

Thus, $E[\beta \mid \hat{\rho}, M, \Pi, \tilde{Z}]$ is given by:

$$\begin{aligned} \beta^* &= E[\beta \mid M, \Pi, \tilde{Z}] + Cov(\beta, \hat{\rho} \mid M, \Pi, \tilde{Z}) Var(\hat{\rho} \mid M, \Pi, \tilde{Z})^{-1} (\hat{\rho} - E[\hat{\rho} \mid M, \Pi, \tilde{Z}]) \\ &= M\varphi + \Pi'\sigma_\nu^2 (\Pi\Pi'\sigma_\nu^2 + \Sigma)^{-1} (\hat{\rho} - \Pi M\varphi) \\ &= \Omega\hat{\rho} + (I - \Omega\Pi)M\varphi, \end{aligned}$$

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

Corollary 2 follows from the observation that if $E[\nu \mid M, \Pi, \tilde{Z}] = 0$ and $Var(\nu \mid M, \Pi, \tilde{Z}) = \sigma_\nu^2 I$ then the β^* in equation (23) is the fitted value from a regression of β on $\hat{\rho}$, conditional on $(M, \Pi, \tilde{Z})$. Standard arguments show that this regression gives a minimum-MSE linear approximation to the conditional expectation $E[\beta \mid \hat{\rho}, M, \Pi, \tilde{Z}]$, and so solves (22) in the class of linear predictors. Specifically, β^* solves:

$$\min_{b_0, B_1} E \left[((b_0 + B_1\hat{\rho}) - \beta)'((b_0 + B_1\hat{\rho}) - \beta) \mid M, \Pi, \tilde{Z} \right],$$

where b_0 and B_1 are a constant vector and coefficient matrix, respectively. As noted in the text, empirical Bayes value-added posteriors are obtained by plugging estimates of Π , Σ , φ , and σ_ν^2 into the equation for β^* .

Accounting for First-Stage Estimation Error

We also derive a posterior mean accounting for estimation error in the first stage matrix, Π , assuming joint-Normal reduced form and first-stage estimation errors:

$$\begin{bmatrix} \hat{\rho} \\ \text{Vec}(\hat{\Pi}) \end{bmatrix} | (M, \Pi, \tilde{Z}, \rho) \sim N \left(\begin{bmatrix} \rho \\ \text{Vec}(\Pi) \end{bmatrix}, \begin{bmatrix} \Sigma & \Psi \\ \Psi' & \Xi \end{bmatrix} \right),$$

as would be obtained in a conventional asymptotic approximation. Consider the MSE minimization problem

$$\min_{b_0, B_1, B_2} E \left[\left(b_0 + B_1 \hat{\rho} + B_2 \text{Vec}(\hat{\Pi}) - \beta \right)' \left(b_0 + B_1 \hat{\rho} + B_2 \text{Vec}(\hat{\Pi}) - \beta \right) | M, \Pi, \tilde{Z} \right].$$

This problem is solved by the conditional regression of β on $\hat{\rho}$ and $\text{Vec}(\hat{\Pi})$ given M and Π :

$$\begin{aligned} \beta^* &= E[\beta | M, \Pi, \tilde{Z}] \\ &+ \text{Cov} \left(\beta, \begin{bmatrix} \hat{\rho} \\ \text{Vec}(\hat{\Pi}) \end{bmatrix} | M, \Pi, \tilde{Z} \right) \text{Var} \left(\begin{bmatrix} \hat{\rho} \\ \text{Vec}(\hat{\Pi}) \end{bmatrix} | M, \Pi, \tilde{Z} \right)^{-1} \\ &\times \begin{bmatrix} \hat{\rho} - E[\hat{\rho} | M, \Pi, \tilde{Z}] \\ \text{Vec}(\hat{\Pi}) - E[\text{Vec}(\hat{\Pi} | M, \Pi, \tilde{Z})] \end{bmatrix} \\ &= M\varphi + \begin{bmatrix} \Pi'\sigma_\nu^2 & 0 \end{bmatrix} \begin{bmatrix} \Pi\Pi'\sigma_\nu^2 + \Sigma & \Psi \\ \Psi' & \Xi \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho} - \Pi M\varphi \\ \text{Vec}(\hat{\Pi}) - \text{Vec}(\Pi) \end{bmatrix}. \end{aligned}$$

Plugging the estimates of Π , Σ , Ψ , Ξ , φ , and σ_ν^2 in to this formula yields

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

where

$$\hat{\Omega} = \hat{\Pi}' \left(\hat{\Pi}\hat{\Pi}' + (\hat{\Sigma} - \hat{\Psi}\hat{\Xi}^{-1}\hat{\Psi}')/\hat{\sigma}_\nu^2 \right)^{-1}.$$

The difference between this weighting matrix and the weighting matrix in Proposition 1 comes from the $\hat{\Psi}\hat{\Xi}^{-1}\hat{\Psi}'$ term, which adjusts the reduced form sampling variance matrix $\hat{\Sigma}$ when estimates of $\hat{\rho}$ and $\hat{\Pi}$ are correlated.

Posterior Means Without Under-subscription

When $L = J$, the feasible version of equation (23) can be written:

$$\beta^* = \hat{W}\hat{\beta} + (I - \hat{W})M\hat{\varphi},$$

where $\hat{W} = \hat{\Omega}\hat{\Pi}$ is a positive definite matrix and $\hat{\beta} = \hat{\Pi}^{-1}\hat{\rho}$ is a two-stage least squares estimate of β using risk-adjusted offers as instruments for school enrollment. When M is a vector of OLS coefficients, $\hat{\beta}^*$ is therefore a matrix-weighted average of OLS and IV estimates of β , with the former scaled by the estimated forecast coefficient $\hat{\varphi}$. The weights depend on the estimated first-stage matrix $\hat{\Pi}$, the estimated forecast residual variance $\hat{\sigma}_\nu^2$, and estimation error Σ .

Posterior Means that Account for OLS Sampling Variance

Let $M = \hat{\alpha} = \alpha + e_\alpha$ be a vector of noisy OLS value-added estimates. We assume the vector of estimation error, e_α , is mean-zero and uncorrelated with α and ν conditional on Π and $\tilde{Z}$. We also assume $E[\alpha\alpha' | \Pi, \tilde{Z}] = \sigma_\alpha^2 I$ and write $E[e_\alpha e'_\alpha | \Pi, \tilde{Z}] = V_\alpha$. Then, with $\beta = \varphi\alpha + \nu$ and $E[\alpha | \Pi, \tilde{Z}] = [\nu | \Pi, \tilde{Z}] = 0$, the posterior mean of β given the reduced-form and OLS estimates becomes:

$$\begin{aligned} \beta^* &= Cov \left(\beta, \begin{bmatrix} \hat{\rho} \\ \hat{\alpha} \end{bmatrix}' \mid M, \Pi, \tilde{Z} \right) Var \left(\begin{bmatrix} \hat{\rho} \\ \hat{\alpha} \end{bmatrix} \mid M, \Pi, \tilde{Z} \right)^{-1} \begin{bmatrix} \hat{\rho} \\ \hat{\alpha} \end{bmatrix} \\ &= \begin{bmatrix} \Pi'(\varphi^2\sigma_\alpha^2 + \sigma_\nu^2) & \varphi\sigma_\alpha^2 I \end{bmatrix} \begin{bmatrix} \Pi\Pi'(\varphi^2\sigma_\alpha^2 + \sigma_\nu^2) + \Sigma & \Pi\varphi\sigma_\alpha^2 + V_{\rho\alpha} \\ \Pi'\varphi\sigma_\alpha^2 + V'_{\rho\alpha} & \sigma_\alpha^2 I + V_\alpha \end{bmatrix}^{-1} \begin{bmatrix} \hat{\rho} \\ \hat{\alpha} \end{bmatrix}, \end{aligned}$$

where $V_{\rho\alpha}$ denotes the covariance of reduced-form and OLS estimation error. This can be written

$$\beta^* = \Omega_{\hat{\rho}}\hat{\rho} + \Omega_{\hat{\alpha}}\hat{\alpha},$$

where

$$\begin{aligned} \Omega_{\hat{\rho}} &= \left(\Pi'(\varphi^2\sigma_\alpha^2 + \sigma_\nu^2) - \varphi\sigma_\alpha^2 (\sigma_\alpha^2 I + V_\alpha)^{-1} (\Pi'\varphi\sigma_\alpha^2 + V'_{\rho\alpha}) \right) \\ &\quad \times \left(\Pi\Pi'(\varphi^2\sigma_\alpha^2 + \sigma_\nu^2) + \Sigma - (\Pi\varphi\sigma_\alpha^2 + V_{\rho\alpha}) (\sigma_\alpha^2 I + V_\alpha)^{-1} (\Pi'\varphi\sigma_\alpha^2 + V'_{\rho\alpha}) \right)^{-1} \end{aligned}$$

and

$$\Omega_{\hat{\alpha}} = (I - \Omega_{\hat{\rho}}\Pi) \varphi \sigma_{\alpha}^2 (\sigma_{\alpha}^2 I + V_{\alpha}) - \Omega_{\hat{\rho}} V_{\rho\alpha} (\sigma_{\alpha}^2 I + V_{\alpha}) .$$

This formula is the same as equation (17) in Angrist et al. (2017) when the mediating vector M includes OLS VAM and the VAM parameters are normalized so that $E[\alpha_j] = E[\beta_j] = 0$.

With unbiased OLS VAM, so that $\varphi = 1$ and $\sigma_{\nu}^2 = 0$, and with homoskedastic estimation error, $V_{\rho\alpha} = \Pi V_{\alpha}$ by the usual Gauss-Markov logic. Then,

$$\begin{aligned} \Pi'(\varphi^2 \sigma_{\alpha}^2 + \sigma_{\nu}^2) - \varphi \sigma_{\alpha}^2 (\sigma_{\alpha}^2 I + V_{\alpha})^{-1} (\Pi' \varphi \sigma_{\alpha}^2 + V_{\rho\alpha}') &= \left(\Pi' \sigma_{\alpha}^2 - \sigma_{\alpha}^2 (\sigma_{\alpha}^2 I + V_{\alpha})^{-1} (\Pi' \sigma_{\alpha}^2 + V_{\alpha} \Pi') \right) \\ &= 0 \end{aligned}$$

such that $\Omega_{\hat{\rho}} = 0$ and no weight is put on the reduced-form vector. Furthermore,

$$\Omega_{\hat{\alpha}} = \sigma_{\alpha}^2 (\sigma_{\alpha}^2 I + V_{\alpha}) .$$

In this case $\beta^* = \Omega_{\hat{\alpha}} \hat{\alpha}$ can be seen to be the conventional OLS-VAM posteriors. For example, when V_{α} is diagonal, $\Omega_{\hat{\alpha}} = \text{diag}(\lambda_1, \dots, \lambda_J)$ and we obtain the $\alpha_j^* = \lambda_j \hat{\alpha}_j$ in equation (24).

III. Mean Squared Error Calculations

By the law of iterated expectations, the mean squared error of the conventional shrinkage estimator in equation (25) is given by:

$$E [(\alpha_j^* - \beta_j)^2] = E [Var(\alpha_j^* | \alpha_j, \beta_j, \lambda_j)] + E \left[(E[\alpha_j^* | \alpha_j, \beta_j, \lambda_j] - \beta_j)^2 \right] .$$

The first term is the expected sampling variance of the value-added estimator, while the second equals expected squared bias conditional on the underlying VAM parameters. Suppose the sampling error s_j (and therefore λ_j) is independent of α_j and ν_j across schools. The variance term is given by

$$E [Var(\alpha_j^* | \alpha_j, \beta_j, \lambda_j)] = E [\lambda_j^2 Var(\hat{\alpha}_j | \alpha_j, \beta_j, \lambda_j)] = E [\lambda_j^2 s_j^2] ,$$

and the bias component is given by

$$\begin{aligned}
E \left[\left(E[\alpha_j^* | \alpha_j, \beta_j, \lambda_j] - \beta_j \right)^2 \right] &= E \left[(\lambda_j \alpha_j - \varphi \alpha_j - \nu_j)^2 \right] \\
&= E[(\lambda_j - \varphi)^2 \alpha_j^2] - 2E[(\lambda_j - \varphi) \alpha_j \nu_j] + E[\nu_j^2] \\
&= E[(\varphi - \lambda_j)^2] \sigma_\alpha^2 + \sigma_\nu^2,
\end{aligned}$$

where the third line uses independence of (α_j, ν_j) from λ_j along with $E[\alpha_j \nu_j] = 0$. Adding these two components yields equation (25).

We calculate the mean squared error of the IV VAM posterior predictions in Appendix A.4, accounting for estimation error in both the reduced form $\hat{\rho}$ and OLS-VAM mediators $M = \hat{\alpha}$. Write the reduced form as

$$\hat{\rho} = \Pi \beta + e_\rho = \Pi \alpha \varphi + \Pi \nu + e_\rho,$$

where e_α is again mean-zero given Π , with $E[e_\alpha e'_\alpha | \Pi] = V_\alpha$ and $E[e_\alpha e'_\rho | \Pi] = V_{\alpha\rho}$. The difference between the IV VAM posterior and causal value-added is then

$$\begin{aligned}
\beta^* - \beta &= \Omega_{\hat{\rho}}(\Pi \alpha \varphi + \Pi \nu + e_\rho) + \Omega_{\hat{\alpha}}(\alpha + e_\alpha) - \alpha \varphi - \nu \\
&= ((\Omega_{\hat{\rho}} \Pi - I) \varphi + \Omega_{\hat{\alpha}}) \alpha + (\Omega_{\hat{\rho}} \Pi - I) \nu + \Omega_{\hat{\rho}} e_\rho + \Omega_{\hat{\alpha}} e_\alpha.
\end{aligned}$$

Average mean squared error across the J schools can then be written

$$\begin{aligned}
J^{-1} E [(\beta^* - \beta)' (\beta^* - \beta)] &= J^{-1} tr(((\Omega_{\hat{\rho}} \Pi - I) \varphi + \Omega_{\hat{\alpha}}) ((\Omega_{\hat{\rho}} \Pi - I) \varphi + \Omega_{\hat{\alpha}})' \sigma_\alpha^2 \\
&\quad + (\Omega_{\hat{\rho}} \Pi - I) (\Omega_{\hat{\rho}} \Pi - I)' \sigma_\nu^2 \\
&\quad + \Omega_{\hat{\rho}} \Sigma \Omega_{\hat{\rho}}' + \Omega_{\hat{\alpha}} V_\alpha \Omega_{\hat{\alpha}}' + \Omega_{\hat{\rho}} V_{\alpha\rho}' \Omega_{\hat{\alpha}}' + \Omega_{\hat{\alpha}} V_{\rho\alpha} \Omega_{\hat{\rho}}')
\end{aligned}$$

We calculate RMSE by plugging estimates of $\Omega_{\hat{\rho}}$, $\Omega_{\hat{\alpha}}$, Π , φ , Σ , σ_ν^2 , σ_α^2 , V_α , and $V_{\alpha\rho}$ into this expression and taking its squared root. As with α_j^* , the average mean squared error of β_j^* is the sum a bias term

$$J^{-1} tr \left(((\Omega_{\hat{\rho}} \Pi - I) \varphi + \Omega_{\hat{\alpha}}) ((\Omega_{\hat{\rho}} \Pi - I) \varphi + \Omega_{\hat{\alpha}})' \sigma_\alpha^2 + (\Omega_{\hat{\rho}} \Pi - I) (\Omega_{\hat{\rho}} \Pi - I)' \sigma_\nu^2 \right)$$

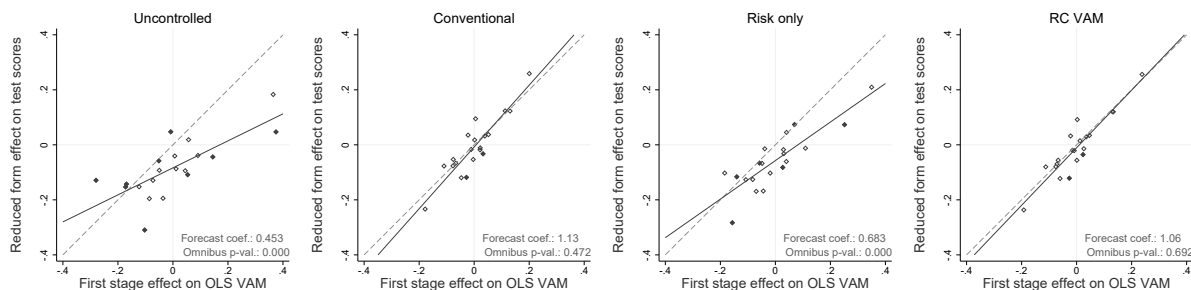
and a variance term,

$$J^{-1} tr \left(\Omega_{\hat{\rho}} \Sigma \Omega_{\hat{\rho}}' + \Omega_{\hat{\alpha}} V_\alpha \Omega_{\hat{\alpha}}' + \Omega_{\hat{\rho}} V_{\alpha\rho}' \Omega_{\hat{\alpha}}' + \Omega_{\hat{\alpha}} V_{\rho\alpha} \Omega_{\hat{\rho}}' \right).$$

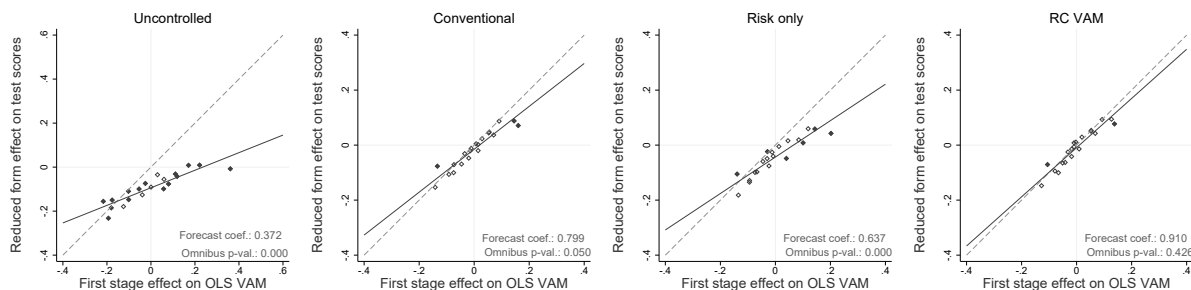
IV. Appendix Figures and Tables

Figure A1. Visual Instrumental Variables Tests for Bias (Reading)

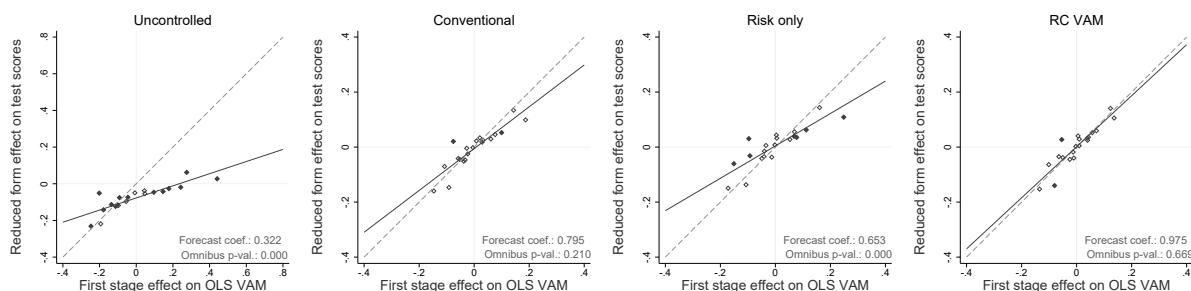
A. Denver middle schools



B. NYC middle schools



C. NYC high schools



— Forecast coef. line - - - - 45 degree line

Notes: This figure plots risk-adjusted offer reduced-form estimates against value-added first stages from each of 20 school assignment offer bins. Outcomes are 6th grade reading CSAP and CMAS scores for Denver, 6th grade ELA New York State Assessment scores for NYC middle schools, and SAT reading/writing scores for NYC high schools. 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, while dashed lines indicate the 45-degree line.

Table A1. Statistical Tests for Balance

	Denver middle schools		NYC middle schools		NYC high schools	
	Uncontrolled (1)	Controlled (2)	Uncontrolled (3)	Controlled (4)	Uncontrolled (5)	Controlled (6)
Demographics						
Hispanic						
Offered VAM	-0.509 (0.020)	-0.042 (0.065)	-0.173 (0.008)	0.030 (0.025)	-1.47 (0.024)	0.002 (0.013)
Any offer	-0.029 (0.007)	0.012 (0.029)	-0.016 (0.003)	-0.013 (0.009)	-0.02 (0.004)	-0.131 (0.104)
Black						
Offered VAM	0.083 (0.013)	0.055 (0.044)	-0.42 (0.007)	-0.003 (0.023)	-1.79 (0.021)	-0.006 (0.011)
Any offer	-0.043 (0.005)	-0.039 (0.022)	-0.136 (0.003)	0.011 (0.008)	-0.014 (0.003)	0.056 (0.097)
White						
Offered VAM	0.411 (0.018)	-0.056 (0.055)	0.339 (0.006)	0.004 (0.013)	1.19 (0.019)	0.003 (0.009)
Any offer	0.085 (0.005)	0.065 (0.019)	0.072 (0.002)	0.004 (0.005)	0.011 (0.002)	-0.002 (0.055)
Female						
Offered VAM	-0.068 (0.020)	0.080 (0.065)	0.008 (0.008)	0.005 (0.026)	0.027 (0.027)	0.145 (0.106)
Any offer	0.011 (0.007)	-0.046 (0.032)	0.032 (0.003)	-0.02 (0.009)	0.083 (0.004)	0.012 (0.013)
Free/reduced price lunch						
Offered VAM	-0.529 (0.020)	0.052 (0.061)	-0.33 (0.007)	0.034 (0.020)	-1.31 (0.023)	-0.003 (0.011)
Any offer	-0.074 (0.006)	-0.031 (0.025)	-0.091 (0.003)	0.000 (0.007)	-0.017 (0.003)	-0.161 (0.088)
Special education						
Offered VAM	-0.050 (0.011)	-0.016 (0.036)	-0.096 (0.006)	-0.014 (0.022)	-0.258 (0.011)	0.004 (0.007)
Any offer	-0.001 (0.004)	0.033 (0.019)	-0.072 (0.003)	0.006 (0.008)	-0.346 (0.003)	-0.043 (0.056)
English language learner						
Offered VAM	-0.246 (0.019)	-0.020 (0.063)	-0.019 (0.004)	0.034 (0.017)	-0.412 (0.014)	-0.006 (0.008)
Any offer	-0.072 (0.007)	-0.069 (0.031)	-0.016 (0.002)	0.002 (0.006)	-0.028 (0.002)	-0.016 (0.067)
Baseline scores						
Math (standardized)						
Offered VAM	0.892 (0.040)	0.187 (0.130)	1.15 (0.015)	0.002 (0.046)	6.95 (0.041)	0.054 (0.158)
Any offer	0.221 (0.012)	-0.057 (0.057)	0.341 (0.006)	-0.032 (0.017)	0.339 (0.007)	0.012 (0.018)
ELA (standardized)						
Offered VAM	0.753 (0.040)	0.109 (0.126)	0.935 (0.016)	-0.015 (0.049)	6.01 (0.043)	0.330 (0.159)
Any offer	0.192 (0.013)	-0.065 (0.056)	0.320 (0.006)	-0.012 (0.017)	0.359 (0.006)	-0.010 (0.018)
N	37,101	8,116	185,988	46,410	121,583	32,432

Notes: This table reports balance statistics, estimated by regressing baseline covariates on the estimated conventional OLS VAM of the offered school and an indicator for any offer. Columns 2, 4, and 6 control for expected OLS VAM, any offer risk, and running variable controls in NYC samples. Robust standard errors are reported in parentheses.

Table A2. Differential Attrition

Controls:	Expected VAM	Expected VAM, any offer risk
	(1)	(2)
Panel A. <i>Denver middle schools</i>		
Offered VAM	0.028 (0.039)	0.017 (0.039)
Any offer		0.025 (0.020)
N	9,250	9,250
Panel B. <i>NYC middle schools</i>		
Offered VAM	0.037 (0.017)	0.044 (0.017)
Any offer		0.038 (0.006)
N	53,313	53,313
Panel C. <i>NYC high schools</i>		
Offered VAM	-0.031 (0.082)	-0.055 (0.085)
Any offer		0.018 (0.010)
N	53,068	53,068

Notes: This table reports differential attrition estimates. The regression reported in column 1 regresses an indicator for follow-up in the sample on the estimated conventional OLS VAM of the offered school, controlling for expected OLS VAM and running variable controls in NYC samples. The regression in column 2 additionally regresses follow-up on an indicator for any offer and an any offer risk control. Robust standard errors are reported in parentheses.

Table A3. Sensitivity of Tests for Bias to Instrument Binning

	Denver middle schools				NYC middle schools				NYC high schools			
	Uncontrolled (1)	Conventional (2)	Risk only (3)	RC VAM (4)	Uncontrolled (5)	Conventional (6)	Risk only (7)	RC VAM (8)	Uncontrolled (9)	Conventional (10)	Risk only (11)	RC VAM (12)
<i>A. 4 Offer Bins</i>												
Forecast coefficient	0.668 (0.071)	1.15 (0.111)	0.936 (0.096)	1.15 (0.111)	0.611 (0.034)	0.936 (0.047)	0.851 (0.046)	1.000 (0.051)	0.359 (0.035)	0.790 (0.074)	0.679 (0.065)	0.967 (0.091)
First stage F statistic	122	364	135	340	547	2042	560	1643	550	771	318	592
Bias tests:												
Forecast bias	22.2 [0.000]	1.71 [0.191]	0.440 [0.507]	1.80 [0.180]	130 [0.000]	1.83 [0.176]	10.6 [0.001]	0.000 [0.998]	331 [0.000]	8.07 [0.004]	24.7 [0.000]	0.130 [0.719]
Overidentification (3 d.f.)	3.86 [0.277]	4.33 [0.228]	2.29 [0.514]	2.88 [0.411]	41.1 [0.000]	6.17 [0.103]	22.7 [0.000]	11.2 [0.010]	10.2 [0.017]	1.72 [0.633]	2.81 [0.422]	1.14 [0.767]
Omnibus (4 d.f.)	26.1 [0.000]	6.04 [0.196]	2.73 [0.604]	4.68 [0.322]	171 [0.000]	8.00 [0.092]	33.3 [0.000]	11.2 [0.024]	341 [0.000]	9.8 [0.044]	27.5 [0.000]	1.27 [0.866]
N (testing)		6,437				37,008				23,216		
N (VAM estimation)		37,101				185,988				121,583		
<i>B. 10 Offer Bins</i>												
Forecast coefficient	0.492 (0.058)	1.07 (0.105)	0.745 (0.084)	1.09 (0.105)	0.621 (0.032)	0.918 (0.043)	0.838 (0.042)	0.976 (0.046)	0.358 (0.031)	0.779 (0.066)	0.677 (0.058)	0.981 (0.083)
First stage F statistic	85.8	196	80.6	171	251	1114	287	911	279	392	148	278
Bias tests:												
Forecast bias	75.4 [0.000]	0.424 [0.515]	9.20 [0.002]	0.705 [0.401]	138 [0.000]	3.63 [0.057]	15.2 [0.000]	0.267 [0.605]	420 [0.000]	11.2 [0.001]	30.7 [0.000]	0.054 [0.817]
Overidentification (9 d.f.)	37.6 [0.000]	14.5 [0.107]	30.8 [0.000]	11.4 [0.250]	52.0 [0.000]	16.1 [0.064]	29.7 [0.001]	19.3 [0.023]	11.9 [0.218]	5.11 [0.824]	5.79 [0.760]	2.42 [0.983]
Omnibus (10 d.f.)	113 [0.000]	14.9 [0.135]	40.0 [0.000]	12.11 [0.278]	190 [0.000]	19.7 [0.032]	44.9 [0.000]	19.6 [0.034]	432 [0.000]	16.3 [0.091]	36.5 [0.000]	2.47 [0.991]
N (testing)		7,505				42,866				27,629		
N (VAM estimation)		37,101				185,988				121,583		
<i>C. 20 Offer Bins</i>												
Forecast coefficient	0.427 (0.059)	1.12 (0.106)	0.748 (0.085)	1.11 (0.102)	0.594 (0.030)	0.927 (0.041)	0.822 (0.039)	0.983 (0.044)	0.339 (0.030)	0.750 (0.064)	0.633 (0.055)	0.941 (0.080)
First stage F statistic	48.5	104	41.7	97.1	153	648	171	530	148	206	80.9	148
Bias tests:												
Forecast bias	95.3 [0.000]	1.21 [0.272]	8.72 [0.003]	1.21 [0.271]	188 [0.000]	3.13 [0.077]	20.8 [0.000]	0.152 [0.697]	476 [0.000]	15.1 [0.000]	44.3 [0.000]	0.538 [0.463]
Overidentification (19 d.f.)	79.0 [0.000]	28.7 [0.070]	53.6 [0.000]	21.6 [0.305]	84.5 [0.000]	25.9 [0.134]	52.9 [0.000]	27.6 [0.091]	26.8 [0.110]	18.2 [0.511]	19.0 [0.456]	14.6 [0.748]
Omnibus (20 d.f.)	174 [0.000]	29.9 [0.071]	62.3 [0.000]	22.8 [0.298]	273 [0.000]	29.0 [0.087]	73.7 [0.000]	27.8 [0.115]	503 [0.000]	33.3 [0.031]	63.3 [0.000]	15.1 [0.768]
N (testing)		7,661				44,758				30,158		
N (VAM estimation)		37,101				185,988				121,583		

Notes: This table reports tests for bias in OLS value-added models (VAMs). See notes to Table 2 for a description of models and test procedures. Panel A reports estimates that use school assignment instruments binned by quartile of estimated conventional VAM; Panel B reports estimates that use instruments binned by decile; and Panel C reports estimates that use instruments binned by ventile (same as in Tables 2 and 3). Standard errors are reported in parentheses; test p-values are reported in brackets.

Table A4. Tests for Bias in OLS Value-Added Models (Reading)

	Uncontrolled (1)	Conventional (2)	Risk only (3)	RC VAM (4)
Panel A. <i>Denver middle schools</i>				
Forecast coefficient	0.453 (0.056)	1.13 (0.108)	0.683 (0.074)	1.06 (0.101)
First stage F statistic	49.6	94.1	51.3	100.0
<i>Bias tests:</i>				
Forecast bias	94.3 [0.000]	1.48 [0.225]	18.5 [0.000]	0.388 [0.533]
Overidentification (19 d.f.)	55.3 [0.000]	18.3 [0.503]	35.3 [0.013]	16.0 [0.655]
Omnibus (20 d.f.)	150 [0.000]	19.8 [0.472]	53.8 [0.000]	16.4 [0.692]
N (testing)		7,932		
N (VAM estimation)		37,101		
Panel B. <i>NYC middle schools</i>				
Forecast coefficient	0.372 (0.033)	0.799 (0.062)	0.637 (0.054)	0.910 (0.071)
First stage F statistic	155	567	113	408
<i>Bias tests:</i>				
Forecast bias	363 [0.000]	10.4 [0.001]	44.9 [0.000]	1.61 [0.205]
Overidentification (19 d.f.)	50.9 [0.000]	21.0 [0.338]	38.0 [0.006]	18.9 [0.466]
Omnibus (20 d.f.)	414 [0.000]	31.4 [0.050]	82.9 [0.000]	20.5 [0.426]
N (testing)		45,110		
N (VAM estimation)		185,988		
Panel C. <i>NYC high schools</i>				
Forecast coefficient	0.322 (0.037)	0.795 (0.083)	0.653 (0.071)	0.975 (0.101)
First stage F statistic	126	179	62.5	131
<i>Bias tests:</i>				
Forecast bias	343 [0.000]	6.09 [0.014]	23.6 [0.000]	0.059 [0.807]
Overidentification (19 d.f.)	31.8 [0.033]	18.7 [0.473]	24.6 [0.173]	16.7 [0.607]
Omnibus (20 d.f.)	375 [0.000]	24.8 [0.210]	48.2 [0.000]	16.8 [0.669]
N (testing)		30,754		
N (VAM estimation)		121,583		

Notes: This table reports tests for bias in OLS value-added models (VAMs). See 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.