Using School Lotteries to Evaluate the Value-Added Model

Jonah Deutsch

University of Chicago
Harris School of Public Policy

Slides for 2013 SREE Conference
A Value-Added Model

• Value-Added (VA) model attempts to answer the following question:
  – How much more does a student learn if he/she attends school A vs. school B (or the average school), holding fixed prior achievement and demographics.

• An example:
  \[ Y_{ist} = Y_{i(t-1)} + X_i \beta + \pi_s + \varepsilon_{ist} \]

• Parameters of interest are the school effects themselves.
  – \( \pi_A - \pi_B \) is interpreted as the causal effect of attending school A as opposed to school B

• Estimates will be unbiased if students are not sorted to schools based on unobserved components of their test score.
Empirical Test of VA

• Use source of random variation in the assignment of students to schools generated by a charter school lottery to directly test the school VA model.
  – In other words, I test whether or not the estimated school effects from VA are unbiased
  – Lotteries strictly mandated by state law.

• Compare an experimental estimate (lottery) to a non-experimental one (VA).
  – When doing so, must assure that both methods attempt to estimate same parameter

• However, the non-experimental method (VA) generates an estimate of the effect of attending school A vs. the average school, or school A vs. school B.

• Clearly not the same treatment effects as the one generated by the lottery

• Requires a novel technique
Preview of Findings

• I fail reject the null that the VA model generates unbiased estimates of school effects for both reading and math.

• In reading, the VA estimates are very close to the experimental results, both statistically and practically.

• In math, while the difference is not statistically significant, the results are also consistent with upward bias in the VA model that would be predicted by a typical selection bias scenario.
Outline

• Previous literature
• Analytical framework
• Empirical strategy
• VA models
• Results
• Implications
VA models in the policy world

• Federal Race to the Top Fund rewards states that use student test scores to evaluate teachers and schools, and VA models are increasingly-popular tools with which to do so.
  – Viewed as improvement over No Child Left Behind accountability regime, which used simple test score averages or percent of students scoring above some threshold.
  – In contrast, VA attempts to control for the fact that students enter school, or classrooms, with varying levels of academic preparedness.

• According to 2010 report, over 30 states use VA in accountability policies.
VA in research

• There has been an intense, ongoing debate about the validity of VA

• Number of papers have examined school VA model
  – Do not empirically test it

• Many studies have found persistent, large impacts of teacher quality, as measured by value added, on students’ academic outcomes (e.g. Kane and Staiger, 2008; Rivkin et al., 2005; Aaronson et al., 2007; Chetty et al., 2011)

• Rothstein (2010) finds that teacher VA estimates in year $t$ predicts student growth in year $t-1$, which he concludes falsifies the identifying assumption of VA.

• Only one other paper (Deming et al., 2011) tests a school VA model.
Validity of VA Models
Why do we care?

• Already in use in many states/districts

• If the models are unbiased, the literature has shown that vast improvements can be made, in test score terms and dollar amounts, by replacing worst schools/teachers with average ones.
  – Chetty et al., 2011; Staiger and Rockoff, 2010

• Could allow policy-makers to greatly improve academic and life outcomes.

• Accountability regimes depend on accurate measures of performance as a necessary condition to either:
  – Incentivize productive behaviors
  – Identify quality teachers/schools
Analytical framework

• Under commonly-used exclusion restriction that the lottery outcome impacts test scores only through its effect on the school a student attends
  – Express the effect of winning the lottery as a function of school effects from the VA model

• Estimate VA model using all schools and students in the district
  – These estimates can be used to calculate the effect of winning the lottery.

• If the VA model is unbiased:
  – This VA-based estimate of the effect of winning the lottery will be the same as the actual lottery effect.
Analytical Plan

• Estimator I develop is exactly equivalent to using the VA estimate of the school attended by each lottery participant as their outcome
  – Take mean difference in that outcome between lottery winners and losers
  – Refer to this as VA-based estimate of lottery effect.

• What follows is the justification of this estimator
Analytical Plan

\textbf{VA model:}

\[ Y_{ist} = Y_{i(t-1)} + X_i\beta + \sum_{s=1}^{K} I_{it}^s \pi_s + \varepsilon_{ist} \]

\( I_{it}^s \): indicator equal to 1 if student \( i \) in year \( t \) attended school \( s \), and 0 otherwise.

\textbf{Model for the effect of winning the lottery (ITT effect):}

\[ Y_{ist} = \mu + A_{it}\delta + u_{ist} \]

\( A_{it} \): indicator equal to 1 if student \( i \) in year \( t \) won the lottery, and 0 if the student lost.

\textbf{Effect of winning the lottery (\( \delta \)) :}

\[ \delta = E[Y_{ist}|A_{it} = 1] - E[Y_{ist}|A_{it} = 0] \]

Substitute in expression for \( Y_{ist} \) from the VA model above. Since the first two terms of the VA model are independent of \( A_{it} \), we can write:

\[ \delta = \sum_{s=1}^{K} \{\pi_s(E[I_{it}^s|A_{it} = 1] - E[I_{it}^s|A_{it} = 0])\} + E[\varepsilon_{ist}|A_{it} = 1] - E[\varepsilon_{ist}|A_{it} = 0] \]
Analytical Plan

\[
\delta = \sum_{s=1}^{K} \{ \pi_s (E[I_{it}^s | A_{it} = 1] - E[I_{it}^s | A_{it} = 0]) \} + E[\varepsilon_{ist}|A_{it} = 1] - E[\varepsilon_{ist}|A_{it} = 0]
\]

We can invoke the commonly used exclusion restriction that the lottery impacts test scores only through its effect on determining which school a student attends (i.e. \(\varepsilon_{ist}\) is independent of \(A_{it}\)), to drop the final two terms:

\[
\delta = \sum_{s=1}^{K} \pi_s (E[I_{it}^s | A_{it} = 1] - E[I_{it}^s | A_{it} = 0])
\]

Re-write \(E[I_{it}^s | A_{it} = j] = p_{js}\) for \(j = 0,1\), as the probability of attending school \(s\) conditional on lottery status \(j\):

\[
\delta = \sum_{s=1}^{K} \pi_s (p_{1s} - p_{0s})
\]
Analytical Plan

\[ \delta = \sum_{s=1}^{K} \pi_s (p_{1s} - p_{0s}) \]

We can generate unbiased estimates of \( p_{1s} - p_{0s} \) with the following model:

\[ I_{it}^s = \alpha_0^s + A_{it} \alpha_1^s + e_{it} \quad \text{(there are } s = 1, \ldots, K \text{ of these models)} \]

since \( \alpha_1^s = p_{1s} - p_{0s} \)

Now, we can use the estimates from these models, \( \hat{\alpha}_1^s \), along with the estimates of the VA model (estimated from all students in the district), \( \hat{\pi}_s \), to write:

\[ \hat{\delta}_{VA} = \sum_{s=1}^{K} \hat{\pi}_s \hat{\alpha}_1^s \]

The idea here is that since we know \( \hat{\alpha}_1^s \) is unbiased, then \( \hat{\delta}_{VA} \) will be unbiased if \( \hat{\pi}_s \) is unbiased.
Table 1: Number of winners and losers in the charter school lotteries

<table>
<thead>
<tr>
<th>Year of Lottery</th>
<th>Grade</th>
<th>Lost</th>
<th>Won</th>
<th>Total in year-grade (VA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>45</td>
<td>32</td>
<td>24,660</td>
</tr>
<tr>
<td>2007</td>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>11</td>
<td>39</td>
<td>24,523</td>
</tr>
<tr>
<td>2009</td>
<td>6&lt;sup&gt;th&lt;/sup&gt;</td>
<td>17</td>
<td>88</td>
<td>23,797</td>
</tr>
<tr>
<td>2009</td>
<td>8&lt;sup&gt;th&lt;/sup&gt;</td>
<td>32</td>
<td>21</td>
<td>24,256</td>
</tr>
<tr>
<td>2010</td>
<td>4&lt;sup&gt;th&lt;/sup&gt;</td>
<td>29</td>
<td>6</td>
<td>21,586</td>
</tr>
<tr>
<td>2010</td>
<td>7&lt;sup&gt;th&lt;/sup&gt;</td>
<td>23</td>
<td>38</td>
<td>24,225</td>
</tr>
</tbody>
</table>
Table 2: Randomization Check
No pairwise differences are statistically significant

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Won lottery</td>
<td>-0.142</td>
<td>-0.049</td>
<td>0.083</td>
<td>0.007</td>
<td>-0.020</td>
<td>0.009</td>
<td>0.041</td>
<td>0.059</td>
<td>0.094</td>
<td>-0.048</td>
</tr>
<tr>
<td>SE</td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.05)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.04)</td>
<td>(0.03)</td>
<td>(0.08)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>N</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
</tr>
</tbody>
</table>
Explaining Potential Divergence

• What could potentially drive divergence between the lottery-based estimate and the VA-based estimate?

1) The VA model suffers from selection bias

  – We might reasonably suspect that characteristics such as motivation, parental education, and other factors are positively correlated both with test scores and with the decision to enter a charter school

  – This should lead to upward bias: the VA-based result would be greater than the experimental result
Explaining Potential Divergence

2) There is an individual effect of attending school A vs. school B associated with every student. And:

- The VA model is unbiased, represents average effect, over all students, of attending school A vs. school B.
- Experimental estimate captures average effect only for those students who entered the lottery.

We might think that students who entered the lottery did so because they had more to gain from being in the charter school, and thus:

- Average of lottery entrants’ individual effects of attending the charter school is greater than the overall average effect.
- In this case, the VA-based estimate should be lower than the experimental one.
VA Models

• Another important dimension of VA models is the degree of aggregation

• From a statistical perspective, aggregating over many years reduces sampling variation, leading to more precise estimates

• From a policy perspective, may be desirable to know the performance of each school in each year
  – Research (e.g. McCaffrey et al., 2009) has shown that even after accounting for sampling variation, there is considerable variation within teachers over time.
  – This is likely variation that policy makers want to capture, not average out
  – For example, principals, district policy, and school policy change frequently
    • We might rather be able to use this metric to understand the impacts of these changes than to simply average over them.
  – However, if sampling error is too large, these distinctions may not be possible, and we may be better off aggregating
VA Models

• In this context, if we have a student who entered a lottery for 6th grade in 2009, should we use the VA estimate of the school she attended estimated from:
  – 6th graders in 2009 (current CPS way)
  – 6th graders in 2007 through 2011
  – 4th through 8th graders in 2009
  – Or 4th through 8th graders in 2007 through 2011

• Does it matter?
VA Models

• I explore different types of VA Models:
  – Fixed effects
  – Random effects
  – “Average residual”

• I use empirical Bayes estimates for all three types
Empirical Bayes

• Theory and evidence suggest that schools in the tails of the distribution of estimated school effects will be more likely to be those with larger sampling error (due to smaller sample size).

• To adjust for this, empirical Bayes shrinks estimates to the mean (which is 0 by construction), in proportion to their standard error. Schools with larger sampling error will be shrunk by a greater degree.
Fixed effects model

• Estimate fixed effects coefficients, subtract their mean from each.
  – Apply empirical Bayes shrinkage, following Aaronson et. al. (2007)
Random effects model

• Using OLS, regress test scores on student covariates (i.e. prior test scores, demographics) centered at the school level

• Use residuals as outcome in a random effects (or random intercepts) model with no covariates

• Random intercepts are the empirical Bayes school effect estimates, where variance components and parameters are estimated with maximum likelihood.
“Average Residual” (AR) model

• Using OLS, regress test scores on student covariates.

• Take the mean of the student residuals for each school as the school effect.

• I use the procedure from Kane and Staiger (2008) and Chetty et al. (2011) for teacher VA but adapt it to schools.

• Use empirical Bayes shrinkage, based on covariance of school effects across years.
“Average Residual” (AR) model

• I also use a procedure for school VA from Deming et al. (2011), which estimates a separate model for each year-grade, does not use empirical Bayes (because variance parameters aren’t generated)

• Important distinction between AR model and fixed/random effects models:
  – If school assignment is correlated with student covariates, AR model will face 2 problems:
    • Coefficients on student covariates, and thus the residuals, will be biased
    • Variation that should be attributed to schools will instead be attributed to student covariates
Model Specifics

• All test scores standardized at year-by-grade level

• All lottery effects, both experimental and VA-based, contain dummies for the lottery (i.e. year-by-grade) the student entered

• Since VA model requires baseline test scores, lottery sample is restricted to students in lotteries for 4th through 8th grades.

• Only first post-lottery test score for students in lotteries from 4th through 8th grades in 2006 through 2010.

• Restrictions would not be employed if we were simply interested in the overall effect of this charter school
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Truth” (Lottery)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FE pooled grades</td>
<td>-0.019</td>
<td>-0.013</td>
<td>-0.011</td>
<td>-0.028</td>
<td>-0.018</td>
<td>-0.012</td>
<td>-0.019</td>
<td>-0.025</td>
<td>-0.039</td>
</tr>
<tr>
<td>(S.E.)</td>
<td>(0.065)</td>
<td>(0.009)</td>
<td>(0.015)</td>
<td>(0.012)</td>
<td>(0.019)</td>
<td>(0.005)</td>
<td>(0.007)</td>
<td>(0.022)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>N</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
</tr>
<tr>
<td>Difference</td>
<td>0.006</td>
<td>0.008</td>
<td>-0.008</td>
<td>0.001</td>
<td>0.008</td>
<td>0.000</td>
<td>-0.005</td>
<td>-0.019</td>
<td></td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.064)</td>
<td>(0.062)</td>
<td>(0.063)</td>
<td>(0.060)</td>
<td>(0.065)</td>
<td>(0.064)</td>
<td>(0.063)</td>
<td>(0.067)</td>
<td></td>
</tr>
<tr>
<td>School S.D.</td>
<td>0.120</td>
<td>0.124</td>
<td>0.142</td>
<td>0.142</td>
<td>0.050</td>
<td>0.068</td>
<td>0.162</td>
<td>0.337</td>
<td></td>
</tr>
<tr>
<td>VA schools</td>
<td>560</td>
<td>560</td>
<td>560</td>
<td>540</td>
<td>560</td>
<td>560</td>
<td>560</td>
<td>540</td>
<td>540</td>
</tr>
</tbody>
</table>

Table 3: Reading: lottery- and VA-based results
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“Truth”</td>
<td>FE</td>
<td>FE</td>
<td>FE</td>
<td>AR</td>
<td>AR</td>
<td>AR</td>
<td>AR</td>
<td>RE</td>
</tr>
<tr>
<td></td>
<td>(lottery)</td>
<td>pooled</td>
<td>pool</td>
<td>pool</td>
<td>pooled</td>
<td>pool</td>
<td>years</td>
<td>years</td>
<td></td>
</tr>
<tr>
<td>Won lottery</td>
<td>-0.088</td>
<td>-0.048</td>
<td>-0.065</td>
<td>-0.059</td>
<td>-0.052</td>
<td>-0.032</td>
<td>-0.045</td>
<td>-0.060</td>
<td>-0.098</td>
</tr>
<tr>
<td>(S.E.)</td>
<td>(0.057)</td>
<td>(0.01)</td>
<td>(0.016)</td>
<td>(0.015)</td>
<td>(0.020)</td>
<td>(0.006)</td>
<td>(0.010)</td>
<td>(0.022)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>N</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
<td>381</td>
</tr>
<tr>
<td>Difference</td>
<td>0.040</td>
<td>0.024</td>
<td>0.029</td>
<td>0.036</td>
<td>0.056</td>
<td>0.043</td>
<td>0.028</td>
<td>-0.010</td>
<td></td>
</tr>
<tr>
<td>S.E.</td>
<td>(0.055)</td>
<td>(0.054)</td>
<td>(0.054)</td>
<td>(0.051)</td>
<td>(0.056)</td>
<td>(0.054)</td>
<td>(0.052)</td>
<td>(0.060)</td>
<td></td>
</tr>
<tr>
<td>School S.D.</td>
<td>0.121</td>
<td>0.135</td>
<td>0.143</td>
<td>0.190</td>
<td>0.073</td>
<td>0.105</td>
<td>0.204</td>
<td>0.374</td>
<td></td>
</tr>
<tr>
<td>VA schools</td>
<td>560</td>
<td>560</td>
<td>560</td>
<td>540</td>
<td>560</td>
<td>560</td>
<td>540</td>
<td>540</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Math: lottery- and VA-based results
Conclusions

Reading
• In reading, the divergence between the experimental and VA-based estimates are very small
  – Both statistically and practically insignificant

Math
• The divergence, while not statistically significant, may still be important. Increased power that will come when more lotteries are added will provide more precise answer.
  – If bias is practically significant, consistent with selection bias scenario.
Another Test

• Run VA model, excluding the charter school and all lottery participants

• Generate predicted test scores $\hat{Y}_{ist}$ for all lottery participants who did not attend charter school

• Test $H_0: E[Y_{ist}] - E[\hat{Y}_{ist}] = 0$

• Idea is to see if VA model generates accurate predictions for students who have demonstrated selection into the lottery
## Another Test - Results

<table>
<thead>
<tr>
<th></th>
<th>Reading</th>
<th></th>
<th></th>
<th>Math</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FE</td>
<td>AR</td>
<td>RE</td>
<td>FE</td>
<td>AR</td>
<td>RE</td>
</tr>
<tr>
<td>$E[Y_{ist}] - E[\hat{Y}_{ist}]$</td>
<td>0.040</td>
<td>0.043</td>
<td>-0.005</td>
<td>0.098**</td>
<td>0.101**</td>
<td>0.041</td>
</tr>
<tr>
<td>(S.E.)</td>
<td>(0.043)</td>
<td>(0.044)</td>
<td>(0.044)</td>
<td>(0.035)</td>
<td>(0.035)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>N</td>
<td>167</td>
<td>167</td>
<td>167</td>
<td>167</td>
<td>167</td>
<td>167</td>
</tr>
</tbody>
</table>

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Teacher and School VA Models

- My test has implications for teacher VA models, because school and teacher VA models are inextricably linked:

- Consider a teacher VA model

\[ Y_{ijt} = Y_{i(t-1)} \gamma + X_i \beta + \pi_j + \varepsilon_{ijt} \]

- Unbiased estimation relies on the assumption that students are not sorted to teachers based on unobserved components of their test score (i.e. \( \varepsilon_{ijt} \))

- An equivalent way to express this assumption is:
  
  (1) students are not sorted to teachers within schools

  AND

  (2) students are not sorted to schools

- It is clear that the assumption of the school VA model, (2), is embedded in the assumption of the teacher VA model

- A number of the important studies in the teacher VA literature (e.g. Kane and Staiger, 2008 and Chetty et al., 2011) employ models with these assumptions
Teacher and School VA Models

• Researchers often use only within-school variation to identify teacher effects
  – This eliminates (2) as an assumption
  – More plausible, fewer assumptions always better

• However, for policy practitioners, this alters the interpretation of the metric in important and often-overlooked ways.
  – This empirical strategy only compares teachers within the same school.

• For any policy or resource allocation decision located at the district level or higher, such a metric is probably less useful
  – Rewards or sanctions for teacher performance
  – Retention and promotion decisions
  – Directing extra resources to help the lowest performing teachers (and/or their students)

• There is clearly a trade-off: having a more plausible identification assumption vs. generating ideal parameter for policy-makers.
Future Research

• Add more lotteries for this school to increase power/precision.
• Incorporate lotteries from other schools to increase generalizability.
• Test VA models used in other fields (health).