Article

Evaluation Review
2017, Vol. 41(6) 593-619

DOI: 10.1177/0193841X17742233
journals.sagepub.com/home/erx
SAGE

Using Synthetic Controls to Evaluate the Effect of Unique Interventions: The Case of Say Yes to Education

Robert Bifulco[1], Ross Rubenstein[2], and Hosung Sohn[3]

Abstract

Background: "Place-based" scholarships seek to improve student outcomes in urban school districts and promote urban revitalization in economically challenged cities. Say Yes to Education is a unique district-wide school reform effort adopted in Syracuse, NY, in 2008. It includes full-tuition scholarships for public and private universities, coupled with extensive wraparound support services in schools. **Objectives:** This study uses synthetic control methods to evaluate the effect of Say Yes on district enrollment and graduation rates. It also introduces the synthetic control method and provides guidance for its use in evaluating single-site

[1] Department of Public Administration and International Affairs, Maxwell School of Citizenship and Public Affairs, Syracuse University, Syracuse, NY, USA
[2] Department of Public Management and Policy, Andrew Young School of Policy Studies, Georgia State University, Atlanta, GA, USA
[3] School of Public Service, College of Social Sciences, Chung-Ang University, Seoul, Republic of Korea

Corresponding Author:
Hosung Sohn, School of Public Service, College of Social Sciences, Chung-Ang University, 84 Heukseok-ro, Heukseok-dong, Seoul 156-756, Republic of Korea.
Email: hsohn@cau.ac.kr

interventions. **Method:** Combining school district-level data from the National Center for Education Statistics' Common Core of Data and New York State School Report Cards, this article uses synthetic control methods to construct a synthetic comparison district to estimate counterfactual enrollment and graduation trends for Syracuse. **Results:** We find that Say Yes to Education was associated with enrollment increases in the Syracuse City School District, a district that had previously experienced decades of sustained enrollment declines. We do not find consistent evidence of changes in graduation rates following adoption of the program. **Conclusions:** Graduation rate analyses demonstrate that estimates of treatment effects can be sensitive to choices that the researcher has to make in applying synthetic control methods, particularly when pretreatment outcome measures appear to have considerable amounts of noise.

Keywords
Say Yes to Education, place-based policy, synthetic control method

Introduction

In this article, we apply synthetic control methods to estimate the impacts of the Say Yes to Education program (Say Yes) in Syracuse, NY.[1] The Say Yes intervention includes the promise of a scholarship covering the full cost of tuition to all students who qualify for higher education and graduate from public schools in the city, as well as a set of school reform and student support initiatives. We focus here on the effects of Say Yes on district enrollments and graduation rates. Our purpose is to evaluate the program's effects while also examining key features of the synthetic control method by assessing the sensitivity of effect estimates to various choices made in applying the method.

Education policy makers and evaluation specialists are often interested in the effects of unique interventions implemented across an entire school district. Estimating the effect of these district-wide interventions poses difficult challenges, however. Any difference in outcomes between the district that adopts the intervention and other districts might be explained by a host of district-specific characteristics other than the intervention. In studies where a treatment is applied to multiple units, it is often possible to select a sample of untreated units that has a distribution of potentially confounding variables that is sufficiently similar to the distribution among the treated units. As a result, average differences in outcomes between the treated units and a sample of the untreated units cannot be explained by differences between the two groups in the distribution of other variables. When there is only one treated unit, however, it is often difficult to identify an untreated district that is sufficiently similar to the treated unit on all the potentially confounding variables. Also, when interventions affect all schools, teachers, or students simultaneously, information on schools, teachers, or students within the district will not help an evaluator develop credible estimates of the counterfactual.

Abadie, Diamond, and Hainmueller (2010) propose a strategy for obtaining effect estimates that can be applied in evaluations of unique interventions, which they refer to as synthetic control methods. They argue that a composite comparison unit can be constructed as a weighted combination of untreated units, providing a more valid counterfactual estimate than comparison to any single untreated comparison unit. The authors also suggest an approach to assessing uncertainty in the estimates provided by synthetic control methods, which we describe more fully below.

Synthetic control methods have clear similarities to methods referred to by some as comparative interrupted times series analysis or, alternatively, as fixed-effects regressions with controls for unit-specific trends. Synthetic control methods differ from these parametric regression approaches, though, in the selection of comparison units, treatment of transitory shocks in pretreatment outcome measures, and the assessment of uncertainty in effect estimates. Abadie et al. (2010) describe several advantages of synthetic controls relative to other methods. The methods they recommend provide a replicable way for researchers to select comparison units without any reference to posttreatment outcomes. Relative to parametric regression methods, synthetic control methods are transparent regarding how closely the comparison units match the treated unit on preintervention outcomes and other covariates, as well as the relative influence of each untreated unit in the estimation of the counterfactual outcome. Also, synthetic control methods prevent researchers from relying on extrapolation of observed relationships between outcomes of interest and potentially confounding covariates to estimate the counterfactual.

Although the synthetic control method provides advantages over regression-based methods in the selection of comparison units when only a single treatment unit exists, we show here that estimates of treatment effects can be sensitive to choices that the researcher makes in applying the method, especially when pretreatment outcome measures appear to have considerable amounts of noise.

The remainder of this article is organized as follows. In the next section, we provide background on the Say Yes program in Syracuse and review previous research examining similar "place-based" scholarship programs. We then describe the data used in our analysis. In Synthetic Control Methods section, we provide an overview of synthetic control methods, followed by results of applying synthetic control methods to estimate the effect of Say Yes on enrollment and graduation rates in Results From Synthetic Control Methods section. The sixth section offers conclusions both about the effects of Say Yes on enrollment and graduation rates in Syracuse and on the usefulness of synthetic control methods for the evaluation of educational programs.

Background and Previous Literature

Say Yes to Education is an ambitious initiative that combines "place-based" college scholarships with intensive student supports during elementary and secondary school. A partnership between the Say Yes to Education Foundation, the Syracuse City School District (SCSD), and Syracuse University, the initiative was announced in spring 2008 and first implemented in Syracuse in the 2008–2009 school year. Say Yes guarantees full-tuition coverage at any New York State public college or university and over 80 private institutions nationwide for any graduate of the SCSD who attends district schools during the last 3 years of high school. In addition, the initiative assists school improvement efforts and provides resources to expand and coordinate wraparound supplementary services for students at all grade levels, including extended day and year programming, school- and community-based health services, family counseling services, and pro bono legal clinics. Since being implemented in Syracuse, the program has expanded to Buffalo, NY (2013), and Guilford County, NC (2015).

The goals of Say Yes are 2-fold. First, by offering generous educational benefits to residents, the scholarship offer creates an important locational advantage that can help to attract new residents and businesses and retain families who might otherwise have left the district. A leading indicator of successful community revitalization, then, would be enrollment increases in the city public schools above those expected in the absence of the program.

Second, the program provides a strong and clear incentive as well as supports to help students interested in higher education to graduate high school and achieve at a level that will prepare them for college. This goal could be achieved through several mechanisms. The wraparound school services provided by Say Yes could help to improve student achievement and keep students on track for graduation. The incentive of free college education could motivate some students at risk of dropping out to stay in school and graduate. Peer effects and changes in school culture could promote a focus on graduation and college going that may affect all students whether or not they are eligible for scholarships by attending Syracuse public schools for 10th through 12th grades.[2] Finally, program benefits could induce college-ready students to relocate to the district.

The past decade has seen growth in "place-based" scholarships similar to the Say Yes scholarship offer. Unlike an earlier wave of "I Have a Dream" programs which focused intensive supports and promises of free college tuition on specific schools or grades within a school (Kahne & Bailey, 1999), place-based scholarships primarily focus on entire school districts, typically in urban areas. The programs typically award up to full tuition at public and sometimes private colleges and universities to students attending public schools in a specific city. A compilation of such programs found over 80 operating around the country, though the details of the programs varied widely (W.E. Upjohn Institute for Employment Research, 2016).

The available research has focused primarily on two important short-term potential effects. First, have the programs increased enrollments in city public schools? In general, the studies have found positive enrollment effects, though their methods have differed. Much of the research has focused on one of the first such programs, the Kalamazoo Promise in Michigan. For example, Bartik, Eberts, and Huang (2010) use pre-Kalamazoo Promise enrollment trends to estimate counterfactual post-Promise enrollments under various assumptions about entrance and exit rates and estimate enrollment increases of up to 25%, had pre-Promise entry and exit rates continued in the years after the start of the program. Hershbein (2013) uses student-level data for Kalamazoo and aggregate data for other Michigan districts and reports a 40% increase in new entrants to Kalamazoo schools in the first year of the Promise, with the majority coming from other Michigan public school districts and a quarter coming from other states. Similarly, Ritter and Ash (2016) compare enrollment trends after the start of the El Dorado (Arkansas) Promise and find a reversal of preexisting downward trends in El Dorado while comparison districts saw continued declines. Sohn, Rubenstein, Murchie, and Bifulco (2017) use difference-in-differences and triple-difference models to examine enrollment changes in Syracuse and Buffalo, NY, after the start of the Say Yes program, comparing both cities to Rochester, NY, and to surrounding suburbs. They find significant enrollment increases in both Say Yes cities. Enrollment increases between 1.2% and 3.2% were estimated for Syracuse.

Two studies to date have examined enrollment trends across multiple districts with place-based scholarships. Bartik and Sotherland (2015) compare migration patterns in eight areas with place-based scholarships to matched sets of “commuting zones” and “Migration Public Use Microdata Areas.” They find significant decreases in out-migration in areas that include the scholarships but no increases in in-migration. LeGower and Walsh (2014) compare districts with place-based scholarship to surrounding districts in the same metropolitan areas and find enrollment increases of approximately 4%.

A second research focus has been whether the programs have led to increases in college preparation and matriculation among recipients of the scholarship offer. Bartik, Herschbein, and Lachowska (2015) construct a difference-in-differences analysis by comparing students eligible for the Kalamazoo Promise by virtue of continuous enrollment in the Kalamazoo Public School for Grades 9–12 to those ineligible for scholarships. They find increases in college matriculation of 14% and in 6-year credential attainment of 25%. Bozick, Gonzalez, and Engberg (2015) examine students before and after the start of the Pittsburgh Promise program and find no overall increase in college matriculation but a shift to 4-year institutions.

The Syracuse Say Yes program differs from many other place-based scholarships in important ways. First, the scholarship offer requires shorter residency in the district than most other similar scholarships and, unlike many place-based scholarships, also includes a large number of private universities. Second, in contrast to some place-based scholarship programs, the Say Yes initiative emphasizes school improvement by providing a wide range of school-based student supports. Thus, as one of the most generous college scholarship programs in the nation, and with growing interest around the country in tuition-free college programs, estimates of the effects of Say Yes are of interest to policy makers, the program operators, and potential funders. Moreover, as the above review suggests, no broad consensus exists regarding the appropriate comparison groups for evaluation of these programs, particularly in the absence of detailed student-level data. Interrupted time series designs rely on extrapolation of preexisting trends in a single district, while difference-in-differences designs suffer potential bias if pretreatment trends are not parallel and the comparison districts are not sufficiently similar on observable measures. While previous research has relied on regression-based approaches, the synthetic control method offers a systematic approach to choosing appropriate comparison units using aggregate data.

Sample and Data

To estimate the impacts of Say Yes on enrollment and graduation rates, we use data from two sources. District enrollments (K–12) for the school years 1998–1999 to 2011–2012 are from the common core of data (CCD) maintained by the National Center for Education Statistics. Data on high school graduation rates before and after the Say Yes program are from the New York State Report Cards (NYSRC) managed by the New York State Education Department. The NYSRC provides district-level information on 4-year cohort high school graduation rates starting from the school year 2001–2002 (i.e., the cohort first entering ninth grade in 1998) to 2010–2011 (i.e., the cohort first entering ninth grade in 2007).

Cohort graduation rates are calculated by longitudinally tracking students who first enter the district in ninth grade. Students who enrolled in the district after October 31 of the cohort year or whose last enrollment record shows that they had transferred to a school outside the district are eliminated from the cohort. The 4-year graduation rate is the proportion of students in the cohort who earned a diploma 4 years after the cohort entered ninth grade. In addition to enrollment and graduation measures, shares of free-lunch eligible, African American, and Hispanic students are drawn from the CCD.

Synthetic Control Methods

Overview of Synthetic Control Methods

Synthetic control methods are described by Abadie et al. (2010), and we begin this section by summarizing that description. Suppose we observe $J + 1$ districts at several time periods $t = 1, \ldots, T$, and that one of these districts receives a treatment at the start of one of those time periods and for each of the succeeding periods. The other districts remain untreated. A synthetic control for the treated district can be created by assigning a set of weights $W = w_2, \ldots, w_{J+1}$ to the J untreated units such that $w_j \geq 0$ for $j = 2, \ldots, J + 1$ and $w_2 + \cdots + w_{J+1} = 1$. The value of the outcome for the synthetic control in time period t is computed as the weighted average $\sum_{j=2}^{J+1} w_j Y_{jt}$.

Now, suppose there is a set of weights $w_2^*, \ldots, w_{J+1}^*$ such that $\sum_{j=2}^{J+1} w_j^* Y_{jt}$ equals the value of the outcome for the treated district in each pretreatment time period and $\sum_{j=2}^{J+1} w_j^* \mathbf{Z}_{jt}$ equals the value of each of a set of baseline covariates, $\mathbf{Z}$, for the treated district. Abadie et al. (2010) demonstrate that, if the number of preintervention time periods is large relative to the scale of transitory shocks in the pretreatment outcomes, then $\sum_{j=2}^{J+1} w_j^* Y_{jt}$ will provide estimates of counterfactual outcomes for the treated district in the posttreatment time periods with bias close to zero. In practice, a set of weights that create a synthetic control that exactly matches the treated districts on all pretreatment outcome and covariate values may not exist. If a set of weights can be identified that matches the treated district closely on pretreatment values, however, then the resulting synthetic control will provide credible estimates of the counterfactual posttreatment outcomes. Abadie et al. (2010) argue that an advantage of the synthetic control method is that it allows the researcher to examine how well the synthetic control matches the treated unit and decide whether the match is sufficiently close to allow conclusions about effects of treatment.

The primary technical challenge in implementing the method is to assign weights to each comparison group district that constitutes the synthetic control. These weights are selected in a way that optimizes the match between the treatment district and the resulting synthetic control on pretreatment outcomes and covariates. Let $\mathbf{X}_1$ be a vector of pretreatment covariate values and some set of linear combinations of the outcome values for the treated district in the pretreatment time periods, and let $\mathbf{X}_0 W$ be a vector of the same pretreatment covariate values and linear combinations of the pretreatment outcome values for a synthetic control. Abadie et al. (2010) suggest selecting a set of weights that minimize some measure of the distance between $\mathbf{X}_1$ and $\mathbf{X}_0 W$ subject to the constraints that the weight for each untreated district is nonnegative and that all of the weights sum to one. These constraints ensure that estimates are not based on extrapolations of the relationship between pretreatment characteristics and posttreatment outcomes beyond the region of common support on the pretreatment characteristics. Alternative methods for selecting weights can be distinguished by the way the distance between $\mathbf{X}_1$ and $\mathbf{X}_0 W$ is measured. Abadie and Gardeazbal (2003) and Abadie et al. (2010) choose weights that minimize the mean square prediction error for outcome measures during the preintervention periods, an approach we use in the analyses below.

Note that we could simply choose *W* to minimize the differences in the pretreatment outcomes without including covariates. Because preintervention outcome variables will contain some error, including the covariates can provide better out-of-sample forecasts than matching on preintervention outcomes alone. That is, a synthetic control that matches the treatment unit closely on covariate values as well as pretreatment outcome measures will provide a less biased estimate of the counterfactual, as compared to a synthetic control that matches only on pretreatment outcomes.

Choices in the Application of Synthetic Control Methods

Abadie et al. (2010) refer to the set of untreated units from which the control districts are drawn as the "donor pool." Even if a synthetic control that closely matches the treated district can be constructed from a given donor pool, the resulting effect estimates may suffer from interpolation bias if some members of the donor pool used to compute the synthetic control are substantially different than the treated district. For instance, if a unit with much lower values of the outcome during the pretreatment period is averaged together with a unit with much higher values on the pretreatment outcome, the synthetic control may match the treated unit quite closely. But these two units, which are substantially different from the treated unit, may not provide a good estimate of the counterfactual outcomes of the treated unit during the posttreatment period. For this reason, researchers may want to restrict the donor pool to units with similar characteristics to the treated unit.[3] In fact, a subsequent study (Abadie, Diamond, & Hainmueller, 2015) that estimates the economic impact of the German reunification using synthetic control methods emphasizes this point.

Restricting the donor pool in this way raises two issues, however. First, the researcher has to establish criteria for determining which units are sufficiently similar to the treated unit. Second, because the methods for constructing a synthetic control do not allow for extrapolation, if there are few units in the donor pool with values of pretreatment outcomes and covariates sufficiently close to the values of a treated unit, then the method might not be able to create a synthetic control that is a close match to the treated unit.

In the analysis below, we examine the sensitivity of effect estimates to the choice of a donor pool. Specifically, we use two donor pools—one of which is restricted to districts we judge to be sufficiently similar to Syracuse in ways most likely to affect enrollment trends and another which is more comprehensive.

The SCSD is an urban school district with high shares of students from low-income and minority families. The New York State Education Department classifies the SCSD as one of the Big Five city school districts in the state, and other than New York City, the others (Buffalo, Rochester, and Yonkers) are the most comparable to Syracuse along several dimensions. The SCSD is the smallest of the Big Five city districts, however, and there are other smaller city districts that also have much in common with it. We, therefore, construct two donor pools—one restricted to districts similar to the SCSD and the other including a more comprehensive set of districts. To construct the restricted donor pool, we first limit the sample to Buffalo, Rochester, Yonkers, and the districts the New York State Association of Small City School Districts defines as "small city" districts. The resulting restricted donor pool includes 22 districts.

For the more comprehensive donor pool, we include all districts in New York State that are classified as city or suburban in the CCD. Due to data availability, the numbers of districts in the comprehensive donor pools are different for enrollment and graduation rates analyses. For enrollment analyses, the comprehensive donor pool contains 275 districts, whereas for graduation analyses, it includes 236 districts. An important assumption is that districts in the donor pool did not contemporaneously implement Say Yes or similar programs. We know of no statewide programs implemented at the same time as Say Yes, and Say Yes itself did not expand to Buffalo until after the analysis period.

Researchers must also choose the linear combinations of pretreatment outcome observations the synthetic control should seek to match. One natural choice is the value of the outcome variable for all the available pretreatment periods. In practice, including all available pretreatment outcomes measures does not necessarily result in a closely matched synthetic control. For instance, some of the pretreatment outcome values might be disproportionately influenced by transitory shocks, and it is not clear that closely matching those values will improve estimates of the counterfactual outcomes for the treated district during the posttreatment period. The existing literature provides little guidance about how to choose the set of linear combinations of pretreatment outcome observation periods.

To explore the sensitivity of effect estimates to the choice of pretreatment outcome observations, we construct synthetic controls using six different sets of pretreatment years, which are detailed in Table 1. To allow for assigning weights to districts that are similar to SCSD in terms of socioeconomic characteristics, we also include in all of our specifications the average shares of free-lunch eligible, African American, and Hispanic

Table 1. Alternative Specifications of Pretreatment Years.

Specification	Description
1	First and last year of pretreatment periods
2	First, middle, and last year of pretreatment periods
3	Middle and last year of pretreatment periods
4	Last pretreatment year and the average of outcomes in all other pretreatment years
5	Each pretreatment year
6	Each year from the middle to the end of the pretreatment periods

students over the entire pretreatment period, which for the enrollment analysis ranges from 1998–1999 to 2007–2008 and for the graduation analysis covers the 1998 through 2004 cohorts. The inclusion of covariates (e.g., the average share of free-lunch eligible) when estimating the weights reduces the likelihood that some districts quite dissimilar demographically to Syracuse could contribute to the counterfactual outcome.

Inference Procedure

Because effect estimates are based on comparison of a single treated unit with a synthetic control, an asymptotic approximation is not well suited for assessing the uncertainty in the estimates. Abadie et al. (2010) propose a permutation test to assess the likelihood that the effect size found in the treated unit is sufficiently unlikely to occur by chance. Specifically, they recommend constructing a synthetic control for each unit in the sample, using the remaining districts as the donor pool for each other district. The true treatment effect in each of these untreated units is presumably zero. By comparing the "effect estimate" in each of the untreated units to that found in the treated unit, one can assess how likely it is to obtain an effect estimate that large by chance.

A stylized example helps to illustrate the procedure. Assume we have 1 treated unit and 99 untreated donor units. The treated unit receives the treatment at time t. We construct a synthetic control for the treated unit and find that outcome value in the treated unit increased by the amount ΔY. We then construct synthetic controls for each of the untreated donor units, assuming that these untreated units also received the treatment at time t, and measure the increase in the outcome value in each, relative to its synthetic control, in the posttreatment period. If we find that none of the 99 untreated district has an increase as large as or larger than ΔY in the posttreatment period, this is analogous to a 99% confidence level the increase in the treated district was not due to chance.

More specifically, for each iteration of the synthetic control procedure, we compute the prediction errors for each year, square them (because the errors can be either positive or negative), and take the square root of the average of the squared prediction errors in the pretreatment period and the square root of the average of the squared prediction errors in the posttreatment period. These are called the pre- and posttreatment root mean squared prediction errors (RMSPEs). The ratio of the posttreatment RMSPE to the pretreatment RMSPE is used as a test statistic. The higher this test statistic, the greater the deviation in the posttreatment period between the district and its synthetic control and, therefore, the stronger the evidence of a policy impact. Calculating the percentage of test statistics from all iterations of this procedure that are as large as or larger than that obtained for Syracuse provides a p-value. Because we assume that the effect of Say Yes in all of the donor pool districts is zero, this p-value can be interpreted as the probability of obtaining effect estimates as large as that obtained for Syracuse if the true treatment effect were zero (Abadie, Dimaond, & Hainmueller, 2010; Cameron & Miller, 2015).

Results From Synthetic Control Methods

Estimated Effects on Enrollment

Figure 1 displays the time series in enrollment counts and high school cohort graduation rates for the SCSD and for the (unweighted) averages of our two donor pools. For each year, the SCSD enrollment count and the graduation rates differ substantially from the average of the donor pools. Syracuse also differs from the both donor pools in the percentage of minority and low-income students (comparisons available upon request). These comparisons suggest that these sets of districts may not provide appropriate control groups. Moreover, one cannot draw any clear conclusions from these figures about whether Say Yes had an impact on enrollment and graduation rates.

Implementation of the synthetic control method begins by assigning a weight to each district in the donor pool that specifies that district's contribution to the construction of the synthetic Syracuse (weighted average of control units). We include three covariates (the average percentage of the district's students who are African American, Hispanic, and eligible for free lunch, as a proxy for poverty) and outcomes (enrollment or graduation rates) in the pretreatment years. Weights are assigned, so that the difference

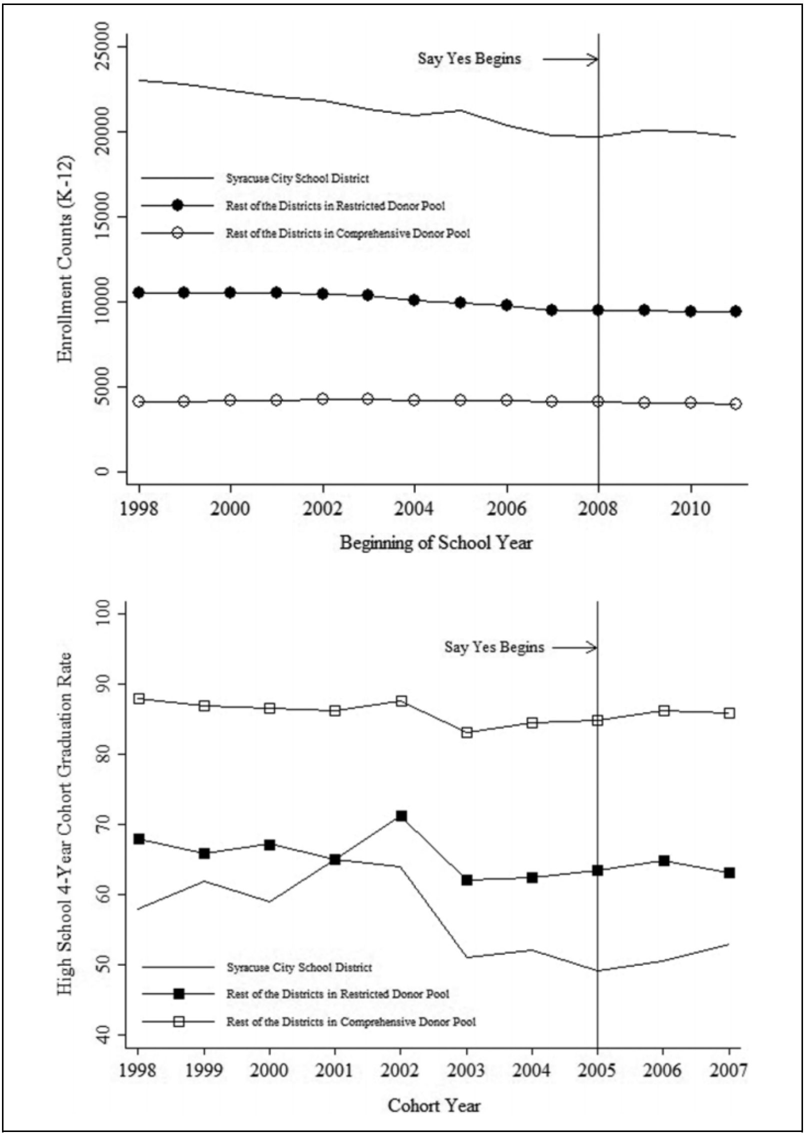

Figure 1. Actual trends in enrollment and graduation rates: Syracuse City School District versus donor districts.

Table 2. Assignment of Weights (Enrollment Analysis).

District Name	Assigned Weights					
	Specif. 1	Specif. 2	Specif. 3	Specif. 4	Specif. 5	Specif. 6
Panel A: Restricted donor pool						
Albany CSD	.000	.021	.000	.000	.005	.000
Brentwood UFSD	.000	.000	.000	.000	.116	.129
Buffalo CSD	.000	.078	.022	.034	.029	.000
Niagara Falls CSD	.484	.288	.498	.499	.411	.404
Rochester CSD	.502	.406	.479	.467	.438	.467
Utica CSD	.014	.207	.001	.000	.000	.000
Panel B: Comprehensive donor pool						
Albany CSD	.000	.000	.000	.000	.005	.000
Brentwood UFSD	.000	.000	.000	.000	.116	.134
Buffalo CSD	.117	.174	.091	.065	.029	.004
Elmira CSD	.000	.000	.307	.148	.000	.000
Hopevale UFSD	.101	.197	.053	.021	.000	.000
Mount Vernon CSD	.000	.061	.000	.000	.000	.000
Niagara Falls CSD	.248	.000	.124	.324	.411	.401
Rochester CSD	.386	.341	.425	.442	.438	.461
Smithtown CSD	.149	.156	.000	.000	.000	.000
Utica CSD	.000	.069	.000	.000	.000	.000

Note. Specif., CSD, and UFSD denote "specification," "city school district," and "union free school district," respectively. Districts that do not appear in the table do not receive positive weights equal to or greater than 0.001 in any of the specifications. Restricted donor pool includes 22 districts, whereas comprehensive donor pool includes 275 districts.

in covariates and preintervention outcomes between Syracuse and synthetic Syracuse is as small as possible. Note that this procedure chooses weights without using any information on outcomes in the posttreatment period.

Estimation of weights for each district in the donor pool depends on which pretreatment years are included in the synthetic control minimization problem. In Table 2, we present the estimated weights that are obtained from different specifications. For example, when we include pretreatment outcomes for only the first and last years, as well as the average shares of free-lunch eligible, African American, and Hispanic students over the entire pretreatment period in the matching vector Z (i.e., Specification 1), Niagara Falls receives a weight of .484 and Rochester receives a weight of .502. If instead each year of pretreatment outcomes is included in the matching vector Z (i.e., Specification 5), then Brentwood receives a weight of 0.116 in addition to Niagara Falls and Rochester.[4]

As can be seen from Table 2, when the restricted donor pool is used, only 6 of the 22 donor pool districts receive positive weights in any of the specifications, and regardless of the specification, Niagara Falls and Rochester receive a combined weight of at least 70%. Of the three districts that have larger enrollments than the SCSD, Rochester is the closest in enrollment with about 12,500 more students than the SCSD in the last year prior to the initiation of Say Yes, and the enrollment of Niagara Falls is about 13,000 fewer students than the SCSD in that year. Both districts, like Syracuse, are majority non-White and have experienced steadily declining enrollments. Niagara Falls and Rochester also receive more than 50% of the weight in all specifications except the second specification when the comprehensive donor pool is used.

Figure 2 displays enrollments for the SCSD and synthetic Syracuse for each year from the fall of 1998 to the fall of 2011. The first year of the Say Yes initiative is 2008. Each graph in the figure displays enrollments for the SCSD and the synthetic controls drawn from the restricted and comprehensive donor pools for one of the six specifications of pretreatment outcome variables. For Specifications 1–4, the enrollments for synthetic controls match the SCSD enrollment well in each pretreatment year except 2005. In that year, enrollment in the SCSD shows an unusual and temporary spike above the pretreatment period trend. Because Specifications 1–4 do not include the 2005 enrollment as a variable in the matching vector, and 2005 deviates from the pretreatment trend, the 2005 enrollment measure is not well matched by the synthetic controls. Specifications 5 and 6, in contrast, do include the 2005 enrollment measure and as a result achieve a closer match on that pretreatment measure. The closer match on the 2005 enrollment measure comes, however, at the cost of a less close match than Specifications 1–4 on enrollment in the year immediately preceding the initiation of Say Yes in Syracuse.

The estimated effects of Say Yes are computed as the difference between enrollments in the SCSD and the synthetic controls during each posttreatment year and are displayed in Table 3.[5] The effect estimates are quite similar across the first four specifications for both donor pools and across Specifications 5 and 6. The estimates from Specifications 5 and 6 are, however, lower than the estimates from the first four specifications. The primary difference between the first four specifications and Specifications 5 and 6 is that the latter specifications include enrollment in 2005 as part of the matching vector, while the first four specifications do not.

The effect estimates obtained from the restricted and comprehensive donor pool are quite similar to each other, which is not surprising given the similarity in weights assigned to districts when the two donor pools are

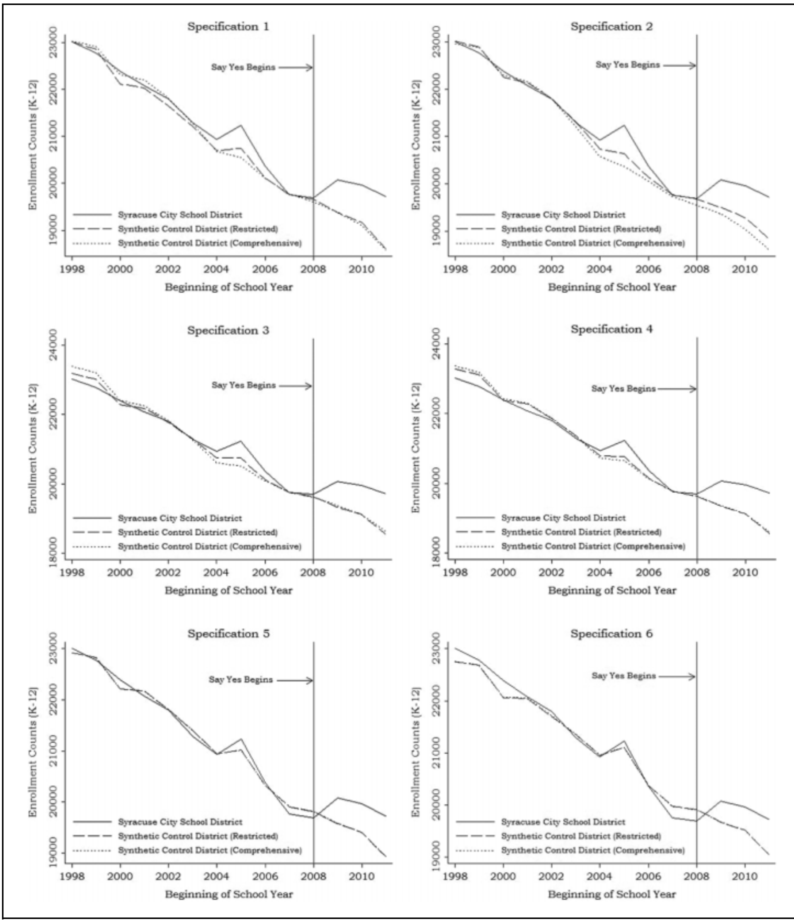

Figure 2. Trends in enrollment by model specifications. See Table 1 for description of pretreatment years included in each specification.

used. The choice of donor pool does not influence effect estimates in this case because the SCSD has one of the largest enrollments in both pools and the SCSD also had a more marked downward trend in enrollment than most other districts. As a result, the weights assigned to particular districts are similar whether the restricted or comprehensive donor pool is used.

Despite the differences in estimates between the first four specifications and Specifications 5 and 6, each tells a qualitatively similar story about the

Table 3. Estimated Effects on K–12 Enrollments, RMSPE, and *p*-value.

Specification	Year 1	Year 2	Year 3	Year 4	RMSPE	*p*-value
Panel A: Restricted donor pool						
Specification 1	30	704	789	1,110	214.16	.087
Specification 2	24	576	676	889	220.09	.217
Specification 3	69	739	840	1,166	209.04	.044
Specification 4	64	730	839	1,165	227.19	.130
Specification 5	−114	500	560	795	114.27	.304
Specification 6	−216	405	445	672	161.23	.130
Panel B: Comprehensive donor pool						
Specification 1	87	693	859	1,132	252.35	.076
Specification 2	147	715	920	1,117	318.44	.243
Specification 3	76	702	845	1,075	323.72	.098
Specification 4	67	713	840	1,121	280.36	.091
Specification 5	−114	500	560	795	114.27	.562
Specification 6	−213	404	445	668	156.37	.120

Note. Restricted donor pool includes 22 districts, whereas comprehensive donor pool includes 275 districts. Years 1–4 correspond to the effect estimates. *p*-value implies a probability of getting a post/pretreatment RMSPE ratio as large as the post/pretreatment RMSPE ratio of Syracuse if one assigns the treatment at random in the data. Specifications are the same as in Table 1. Pretreatment period includes years 1998–2007. All models are run with percent Black, percent Hispanic, and percent free lunch eligible as covariates. RMSPE = root mean squared prediction error.

effect of Say Yes on enrollments. Specifically, the estimates indicate positive effects, which are first realized in the second year after Say Yes begins and which grow somewhat by the fourth year of the Say Yes intervention. By Year 4, the estimated effects on enrollment range from 672 to 1,166 students, which is between 3.2% and 5.6% of the total enrollment in the 2007–08 school year, the last year prior to the adoption of Say Yes, and several have *p*-values less than .10.

The differences in effect estimates between Specifications 1–4 and Specifications 5 and 6 result largely from the fact that when the first four specifications are used, the enrollment of the synthetic control in the last year prior to the initiation of Say Yes matches the SCSD enrollment for that year quite closely while in Specifications 5 and 6, the enrollment of the synthetic control is higher than the enrollment in the SCSD in that year. This pattern suggests that we can reduce the differences between the two sets of estimates by computing effect estimates as difference-in-differences rather than mere posttreatment differences between the treated district and the synthetic control. Specifically, we can compute effect estimates as:

Table 4. Difference-in-Differences Estimates of Effects on K–12 Enrollments.

Specification	Year 1	Year 2	Year 3	Year 4
Panel A: Restricted donor pool				
Specification 1	31	704	789	1,110
Specification 2	36	588	688	901
Specification 3	57	728	828	1,155
Specification 4	69	736	844	1,171
Specification 5	31	645	705	939
Specification 6	−3	618	658	885
Panel B: Comprehensive donor pool				
Specification 1	104	710	876	1,149
Specification 2	113	680	885	1,082
Specification 3	76	702	845	1,075
Specification 4	77	723	850	1,131
Specification 5	31	645	705	939
Specification 6	−1	616	657	880

Note. Restricted donor pool includes 22 districts, whereas comprehensive donor pool includes 275 districts. Years 1–4 correspond to the effect estimates. Specifications are the same as in Table 1. Pretreatment period includes years 1998–2007.

$$(Y_{j=1,t} - Y_{j=1,t=0}) - \left(\sum_{j=2}^{J+1} w_j^* Y_{jt} - \sum_{j=2}^{J+1} w_j^* Y_{j,t=0}\right),$$

where $Y_{j=1,\,t}$ is the outcome (i.e., enrollment) for the treated district, $j = 1$, in a particular posttreatment year t; $Y_{j=1,\,t=0}$ is the outcome for that same district in the last year prior to the treatment, $t = 0$; $\sum_{j=2}^{J+1} w_j^* Y_{jt}$ is the synthetic control outcome in year t; and $\sum_{j=2}^{J+1} w_j^* Y_{j,\,t=0}$ is the synthetic control outcome in year $t = 0$.

Table 4 presents these difference-in-differences estimates. This approach does substantially reduce the disparity in effect estimates between the first four specifications and Specifications 5 and 6. The estimated Year 4 effects in this table vary over a quite narrow range between 885 and 1,171, which is between 4.5% and 5.9% of the enrollment for the SCSD in the last year prior to the initiation of Say Yes.[6] Given the consistency across specifications, we prefer these estimates. The estimated effects are slightly larger than estimates from Sohn et al. (2017) that used difference-in-differences and triple-difference regression analyses to examine enrollment changes in Syracuse.

Estimated Effects on Graduation Rates

Table 5 displays the weights used to construct synthetic controls in the analysis of cohort graduation rates. In contrast to the analysis of enrollments, the weights used to construct synthetic controls for graduation rates differ substantially across specifications. For instance, when the restricted donor pool is used, Schenectady receives weights that vary from .078 to .442. While Buffalo receives weights greater than 0.600 in four of the six specifications, it receives zero weight in Specification 3. Also, the weights assigned to districts differ considerably depending on whether the restricted or comprehensive donor pool is used. For example, Schenectady receives substantial weights in most of the specifications when the restricted donor pool is used but does not receive positive weights in any of the specifications when the comprehensive donor pool is used.

Figure 3 shows that the pretreatment time series for the cohort graduation rate in the SCSD is more volatile than the enrollment time series (Figure 2).[7] Consequently, the distribution of graduation rates changes substantially from year to year, making the assignment of weights sensitive to which pretreatment years are included in the matching vector. Figure 3 also shows that the pretreatment graduation rates for the synthetic control matches the graduation rates for the SCSD less well than in the case of enrollment. The quality of the match is influenced by the amount of volatility in the outcome measure during the pretreatment period.

Table 6 presents the estimated effects on cohort graduation rates. Three things stand out. First, the results are quite sensitive to specification choices. While virtually all the estimates using the restricted donor pool suggest negative impacts, the estimated effects range from −3.4 percentage points to –9.6 percentage points in Year 1, from –0.8 to –8.7 percentage points in Year 2, and from 0.8 to –6.4 percentage points in Year 3. The results are even less consistent for the comprehensive donor pool. For example, estimates of Year 3 impacts range from a substantial reduction in the graduation rate of 3.9 percentage points to a substantial increase of 6.9 percentage points. Second, in many cases, estimates are sensitive to whether the restricted or comprehensive donor pool is used. For some specifications of pretreatment outcomes, the comprehensive donor pool provides substantially more negative effect estimates; for other specifications, substantially more positive effect estimates. Third, the errors in prediction of pretreatment graduation rates are large. As a result, the chance of getting large impact estimates would be low even if the true effects of Say Yes on

Table 5. Assignment of Weights (Cohort Graduation Rate Analysis).

	Assigned Weights					
District Name	Specif. 1	Specif. 2	Specif. 3	Specif. 4	Specif. 5	Specif. 6
Panel A: Restricted donor pool						
Albany CSD	.064	.000	.000	.000	.000	.000
Buffalo CSD	.323	.800	.000	.608	.646	.609
East Ramapo CSD	.000	.000	.000	.000	.000	.027
Hempstead UFSD	.000	.021	.000	.000	.134	.000
Hudson CSD	.000	.000	.002	.000	.000	.001
Niagara Falls CSD	.000	.101	.160	.000	.000	.000
Poughkeepsie CSD	.000	.000	.396	.000	.000	.000
Rochester CSD	.351	.000	.000	.157	.000	.000
Schenectady CSD	.261	.078	.442	.232	.219	.363
Panel B: Comprehensive donor pool						
Albany CSD	.042	.000	.000	0.005	.012	.000
Buffalo CSD	.679	.477	.368	0.683	.000	.428
Binghamton CSD	.000	.000	.000	0.000	.131	.000
East Ramapo Central SD	.000	.000	.000	.000	.000	.066
Elmira CSD	.190	.000	.000	.193	.000	.000
Glen Cove CSD	.000	.000	.000	.000	.052	.000
Greenburgh Central SD	.000	.000	.000	.000	.000	.019
Greenburgh Eleven UFSD	.090	.166	.015	.072	.219	.189
Hawth.-Cedar Kn. UFSD	.000	.000	.000	.000	.111	.000
Hempstead UFSD	.000	.000	.000	.000	.025	.000
Niagara Falls CSD	.000	.168	.000	.000	.000	.000
Poughkeepsie CSD	.000	.000	.392	.047	.058	.000
Rensselaer CSD	.000	.189	.226	.000	.269	.059
Rochester CSD	.000	.000	.000	.000	.002	.000
Watervliet CSD	.000	.000	.000	.000	.000	.222
Westbury UFSD	.000	.000	.000	.000	.000	.018

Note. Specif., CSD, and UFSD denote "specification," "city school district," and "union free school district," respectively. Districts that do not appear in the table do not receive positive weights equal to or greater than 0.001 in any of the specifications. Restricted donor pool includes 22 districts, whereas comprehensive donor pool includes 236 districts.

graduation rates were large. This uncertainty is reflected in the high p-values that are reported for most of the specifications, particularly when the restricted donor pool is used. In sum, the synthetic control methods are

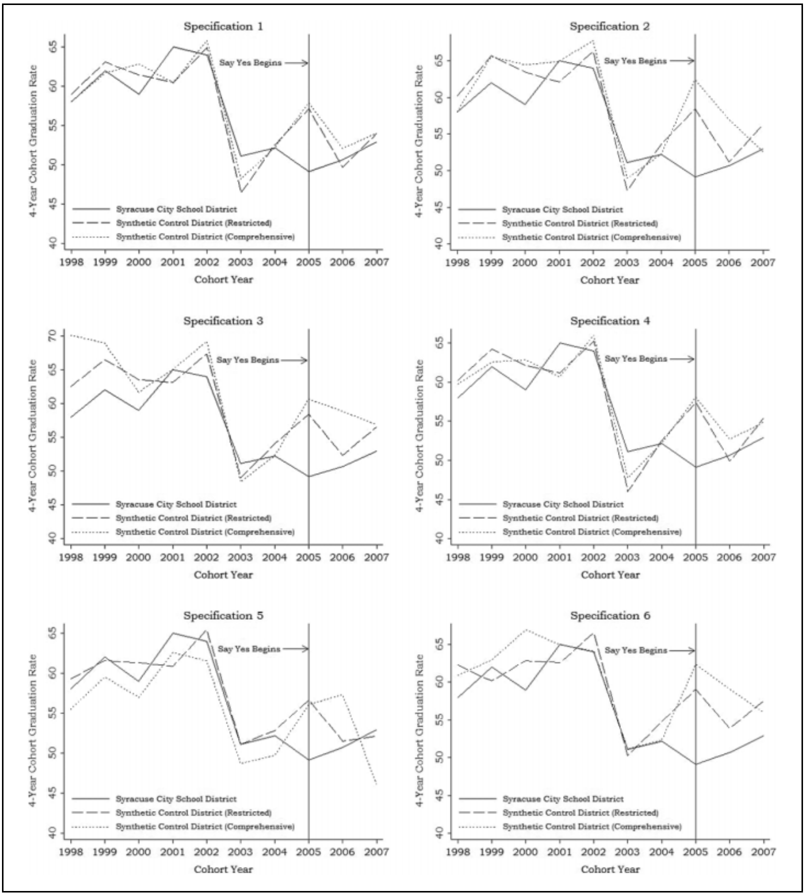

Figure 3. Trends in 4-year cohort graduation rate by model specifications. See Table 1 for description of pretreatment years included in each specification.

unable to provide informative estimates of the effects of Say Yes on this outcome due to the volatility of the pretreatment graduation rates.

Table 7 presents the results of difference-in-differences effect estimates computed using each of the different synthetic controls. Unlike in the analysis of enrollment, using difference-in-differences does not consistently or substantially reduce the variation in impact estimates across different specifications of the pretreatment outcome. Thus, it remains clear that the synthetic control method does not provide reliable estimates of program

Table 6. Estimated Effects on Cohort Graduation Rates, RMSPE, and p-value.

Specification	Year 1	Year 2	Year 3	RMSPE	p-value
Panel A: Restricted donor pool					
Specification 1	−3.373	−1.015	−0.469	2.53	.696
Specification 2	−9.611	−1.763	−3.344	3.14	.435
Specification 3	−9.255	−8.684	−6.442	6.88	.348
Specification 4	−6.076	−0.908	−1.911	2.21	.348
Specification 5	−7.505	−0.837	0.842	1.95	.565
Specification 6	−8.062	−2.268	−3.646	3.26	.174
Panel B: Comprehensive donor pool					
Specification 1	−8.753	−1.421	−1.071	2.52	.101
Specification 2	−13.274	−6.286	0.367	2.84	.055
Specification 3	−11.527	−8.127	−3.870	5.79	.215
Specification 4	−8.904	−2.038	−1.913	2.66	.106
Specification 5	−6.917	−6.630	6.850	2.48	.557
Specification 6	−13.202	−8.398	−2.984	3.06	.076

Note. Restricted donor pool includes 22 districts, whereas comprehensive donor pool includes 236 districts. Years 1–3 correspond to the effect estimates. p-value implies a probability of getting a post/pretreatment RMSPE ratio as large as the post/pretreatment RMSPE ratio of Syracuse if one assigns the treatment at random in the data. Specifications are the same as Table 1. Pretreatment period includes years 2001–2007. RMSPE = the root mean squared prediction error.

Table 7. Difference-in-Differences Estimates of Effects on Cohort Graduation Rates.

Specification	Year 1	Year 2	Year 3
Panel A: Restricted donor pool			
Specification 1	−3.049	−0.691	−0.145
Specification 2	−7.655	0.193	−1.388
Specification 3	−9.007	−8.435	−6.193
Specification 4	−6.084	−0.916	−1.918
Specification 5	−6.864	−0.196	1.482
Specification 6	−7.682	−1.887	−3.265
Panel B: Comprehensive donor pool			
Specification 1	−8.696	−1.364	−1.014
Specification 2	−13.236	−6.248	0.405
Specification 3	−11.446	−8.046	−3.789
Specification 4	−8.893	−2.027	−1.902
Specification 5	−9.386	−9.100	4.380
Specification 6	−13.058	−8.254	−2.841

Note. Restricted donor pool includes 22 districts, whereas comprehensive donor pool includes 236 districts. Years 1–3 correspond to the effect estimates. Specifications are the same as Table 1. Pretreatment period includes years 2001–2007.

impacts on cohort graduation rates because the pretreatment time series is short relative to the time-specific shocks in the outcome variable.

Conclusions

This article uses synthetic control methods to evaluate the effects of a place-based scholarship program and to highlight critical issues evaluators face in using the technique to construct valid counterfactual estimates. In the analysis of enrollments, where the time series of outcome measures does not appear to be strongly influenced by transitory shocks, synthetic control methods provide consistent evidence of meaningful public school enrollment increases in Syracuse relative to control districts after the start of the Say Yes to Education program. Moreover, these effect estimates did not vary substantially when choices of donor pool and pretreatment outcome years were changed, providing an important check on the robustness of the results. Because the synthetic control method disciplines the choice of comparison units used to estimate counterfactual posttreatment outcomes, we are able to derive effect estimates that are largely insensitive to the sample of units used in the estimation and remove a potentially important source of bias in effect estimates.

In contrast to enrollments, the time series of graduation rates is quite volatile. In this case, the effect estimates provided by synthetic control methods were sensitive to both the choice of pretreatment years included in the matching algorithm and the choice of donor pool districts. Given these results, and that synthetic controls can yield minimally biased effect estimates only when the number of preintervention time periods is large relative to the scale of transitory shocks (Abadie et al., 2010), it is difficult to draw conclusions about the effects of the treatment on graduation rates. It is important to note, however, that estimates derived from parametric regressions can also be sensitive to both functional form and sample choices when pretreatment outcomes trends are noisy.

More substantively, it is quite possible that meaningful effects on graduation rates in Syracuse will take time to develop. Students already at risk of dropping out when the program was announced may be unmotivated by the scholarship offer or, more importantly, may already be too far behind to graduate. As students who received additional supports in elementary and middle school begin to reach graduation age, it is possible that graduation rates may start to increase.

Finally, we find consistent evidence that, after decades of population and enrollment declines, public school enrollments in the city of Syracuse began to rebound following the start of the Say Yes to Education program. While the estimated 4-year increases of between 3% and 6% may appear modest, it must be considered in the context of a city such as Syracuse that lost one third of its population between 1950 and 2000 (Office of the New York State Comptroller, 2004). Thus, these enrollment increases suggest that providing large visible amenities such as college scholarships and support services can help to change the trajectory of urban school districts. Whether these increases are leading indicators of a more comprehensive educational and economic turnaround is an important topic for future evaluation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Only recently did synthetic control methods become popular in estimating the treatment effects (e.g., Billmeier & Nannicini, 2013; Bohn, Lofstrom, & Raphael, 2014; Campos, Coricelli, & Moretti, 2014; Trandafir, 2014).
2. In the absence of student-level data, we cannot measure the percentage of students who qualify for the scholarships in Syracuse. In Buffalo, we estimate that approximately 89% of 12th graders in the post-Say Yes years were eligible. Because Buffalo requires attendance in Grades 9–12 rather than 10–12 to qualify for a scholarship, this may represent a lower bound on the percentage of eligible students in Syracuse.
3. Selecting weights for each district in the synthetic control to ensure close matches between the synthetic control and the treated unit on pretreatment covariates as well as pretreatment outcome measures can also help to reduce this type of interpolation bias.
4. It is common for most donor units to receive a weight of zero. For example, Abadie, Diamond, and Hainmeuller (2010) report that 33 of 38 states in their donor pool receive a weight of zero in constructing a synthetic California.
5. Note that the RMSPE and p-values in Table 3 are computed by comparing the average across the 4 post-Say Yes years in Syracuse to the average across the years in the synthetic control district. Therefore, each specification produces one RMSPE and one p-value. The null hypotheses tested in this procedure is that the true average effect across the posttreatment years is zero.

6. To the extent that families that migrate into Syracuse come from districts that receive some weight in the synthetic control analysis, the enrollment of the synthetic control may itself be influenced by Say Yes. Given the location of the districts that receive positive weights in construction of the counterfactual, we suspect that the influence on Say Yes on district enrollments is minimal. Studies of a similar scholarship program suggest that enrollment increases are drawn primarily from neighboring districts and private schools (Hershbein, 2013).
7. New York State made several changes to graduation requirement over the period, such as increasing the number of end-of-course Regents Exams that students must pass. The last major change began with the 9th grade class of 2001. New York State end-of-course testing also predates the federal No Child Left Behind Act of 2001. To the best of our knowledge, no major state policy changes coincided with the start of the Say Yes program, or with the dip in graduation rates in 2003 in Figure 1.

References

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program. *Journal of the American Statistical Association*, *105*, 493–505.

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*, 495–510.

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, *93*, 113–132.

Bartik, T. J., Eberts, R., & Huang, W. (2010). *The Kalamazoo promise, and enrollment and achievement trends in Kalamazoo public schools*. Presented at the PromiseNet 2010 Conference, June 16–18, Kalamazoo, MI.

Bartik, T., Hershbein, B., & Lachowska, M. (2015). The effects of the Kalamazoo Promise scholarship on college enrollment, persistence, and completion. Upjohn Institute Working Paper 15-229. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Bartik, T. J., & Sotherland, N. (2015). *Migration and housing price effects of place-based college scholarships (November 2015) (Upjohn Institute, Working Papers, 15-245)*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Billmeier, A., & Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, *95*, 983–1001.

Bohn, S., Lofstrom, M., & Raphael, S. (2014). Did the 2007 legal Arizona workers act reduce the state's unauthorized immigrant population? *Review of Economics and Statistics*, *96*, 258–269.

Bozick, R., Gonzalez, G., & Engberg, J. (2015). Using a merit-based scholarship program to increase rates of college enrollment in an urban school district: The case of the Pittsburgh promise. *Journal of Student Financial Aid*, 45, *Article 2*.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*, 317–372. IZA Institute of Labor Economics, http://ftp.iza.org/dp8162.pdf

Campos, F., Coricelli, F., & Moretti, L. (2014). *Economic growth and political integration: Estimating the benefits from membership in the European union using the synthetic counterfactuals method* (IZA Discussion Paper No. 8162).

Hershbein, B. (2013). *A second look at enrollment changes after the Kalamazoo promise*. Presentation at Lumina Foundation, Indianapolis, IN.

Kahne, J., & Bailey, K. (1999). The role of social capital in youth development: The case of "I have a dream" programs. *Educational Evaluation and Policy Analysis*, *21*, 321–343.

LeGower, M., & Walsh, R. (2014). Promise scholarship programs as place-making policy: Evidence from school enrollment and housing prices. *Journal of Urban Economics*, *101*, 74–89.

Office of the New York State Comptroller. (2004). *Population trends in New York State's cities*. Albany, NY: Division of Local Government Services and Economic Development.

Ritter, G. W., & Ash, J. W. (2016). How the "promise" of a college scholarship transformed the culture in El Dorado schools. *Phi Delta Kappan*, *97*, 13–19.

Sohn, H., Rubenstein, R., Murchie, J., & Bifulco, R. (2017). Assessing the effects of place-based scholarships on urban revitalization: The case of say yes to education. *Educational Evaluation and Policy Analysis*, *39*, 198–222.

Trandafir, M. (2014). The effect of same-sex marriage laws on different-sex marriage: Evidence from the Netherlands. *Demography*, *51*, 317–340.

W.E. Upjohn Institute for Employment Research. (2016). *Local, place-based scholarship programs*. Retrieved December 2016, from http://www.upjohn.org/sites/default/files/promise/Lumina/Promisescholarshipprograms.pdf

Author Biographies

Robert Bifulco is chair and associate dean at the Department of Public Administration and International Affairs, Syracuse University. His research interests include education finance and policy, program evaluation, school segregation, and state and local government finance.

Ross Rubenstein is professor and Dan E. Sweat Distinguished Chair in Educational and Community Policy at the Andrew Young School of Policy Studies, Georgia

State University. His research interests include school finance, college scholarship programs, and evaluation of education reforms.

Hosung Sohn is an assistant professor at the School of Public Service, Chung-Aung University. His research area is in the field of public finance and program evaluation using experimental and quasi-experimental research designs.