International Public Management Journal

ISSN: 1096-7494 (Print) 1559-3169 (Online) Journal homepage: http://www.tandfonline.com/loi/upmj20

The Synthetic Control Method for Comparative Case Studies: An Application Estimating the Effect of Managerial Discretion Under Performance Management

Chris Birdsall

To cite this article: Chris Birdsall (2015): The Synthetic Control Method for Comparative Case Studies: An Application Estimating the Effect of Managerial Discretion Under Performance Management, International Public Management Journal, DOI: 10.1080/10967494.2015.1121178

To link to this article: http://dx.doi.org/10.1080/10967494.2015.1121178

Accepted author version posted online: 25 Nov 2015.
Published online: 25 Nov 2015.

International
Public
Management
Journal

THE SYNTHETIC CONTROL METHOD FOR COMPARATIVE CASE STUDIES: AN APPLICATION ESTIMATING THE EFFECT OF MANAGERIAL DISCRETION UNDER PERFORMANCE MANAGEMENT

CHRIS BIRDSALL
AMERICAN UNIVERSITY

ABSTRACT: *Public management researchers are often interested in estimating the effects of aggregate-level management reforms and policy changes, but they frequently rely on observational data that pose a number of threats to internal validity. The synthetic control method for comparative case studies (SCM) has great potential for advancing public management scholarship by addressing some of the methodological challenges associated with observational data and providing intuitive graphical results that help researchers communicate their findings to audiences unfamiliar with quantitative methods. SCM uses a transparent, data-driven algorithm for selecting a weighted combination of control units that act as a plausible counterfactual for estimating causal effects. This article demonstrates SCM by investigating the effect of enhancing managerial discretion under performance accountability systems in the context of public education. The article also provides a number of possible avenues for future public management research using SCM and practical guidance for applying the method.*

INTRODUCTION

Barry Bozeman once remarked that public management researchers "seem not to stray much beyond the familiar" in their approaches to empirical research (Bozeman 1992, 440). Calls for methodological rigor and creativity appear regularly in public management and administration journals (e.g., Blom-Hansen, Morton, and Serritzlew 2015; Gill and Meier 2000; Gill and Witko 2013; Konisky and Reenock 2013; Margetts 2011; Perry 2012; Pitts and Fernandez 2009), and these calls often carry a sense that the current state of affairs inhibits knowledge building and the development of the field. But what, precisely, is the problem? The problem is that public management researchers often must rely on observational data, which carry

International Public Management Journal, 0(0), pages 1–29
DOI: 10.1080/10967494.2015.1121178

ISSN: 1096-7494 print/1559-3169 online

a number of threats to internal validity that make drawing causal inference difficult (Konisky and Reenock 2013). Researchers often prescribe experimental research designs, and for good reason: The randomized control experiment is the gold standard for estimating causal effects, and there are many potential applications in public management (Blom-Hansen, Morton, and Serritzlew 2015). Random assignment to treatment and control groups provides researchers with reasonable confidence that any effect observed exclusively in a treatment group is evidence of a treatment effect and not systematic differences in characteristics between treatment and control groups. However, randomized control experiments are not always practical or appropriate for answering public management research questions (Blom-Hansen, Morton, and Serritzlew 2015; Konisky and Reenock 2013). Consequently, for many applications, public management scholars must rely on observational data and use estimation techniques that attempt to address issues of selection bias, simultaneity, and reverse causation.

The synthetic control method (SCM), developed in Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010; 2014), is an estimation approach with great potential for helping public management scholars overcome some of the challenges associated with traditional estimation techniques. Take, for example, a situation where a group of researchers is interested in estimating the effect of decentralizing human resource management responsibilities on worker satisfaction in an organization. They begin their research by implementing a difference-in-differences design comparing the organization of interest to a control group of untreated organizations. They soon realize, however, that the organizations in their control group do not provide a suitable counterfactual—that is, a probable estimation of what would have happened to worker satisfaction in the organization if it did not decentralize human resource responsibilities—because they differ widely from the treated organization, both in trends in worker satisfaction as well as important predictors of worker satisfaction. Preexisting differences between the treatment and control groups, then, may bias their results. The researchers are left somewhat stranded because they cannot produce a credible control case with the data they have using traditional techniques. Here is where SCM may be useful.

SCM uses a data-driven procedure for constructing a control case closely resembling the case of interest according to values of the outcome and its predictors in the period before the intervention. Specifically, SCM uses an algorithm that assigns weights to potential control units in the dataset based on their similarity to the treated unit. In the hypothetical case I describe earlier, SCM considers the available control units and derives a weighted combination that is more similar to the organization of interest than any of the control organizations alone or a simple average of the control organizations. Once the researchers implement SCM and have their synthetic control case, estimating the effect of decentralizing human resource responsibilities is simply a matter of comparing post-treatment values of worker satisfaction between the treated organization and the synthetic control. The researchers interpret differences in worker satisfaction in the post-treatment period as evidence of a policy effect. While SCM does not provide methods for conducting classical inference, it does provide a series of falsification tests that allow researchers to assess the validity of their results. An added benefit is SCM displays the results and falsification tests using intuitive figures that do not require statistical knowledge for interpretation. This feature may help researchers communicate their results to practitioners and other stakeholders who may not be familiar with advanced statistical methods.

In this article, I use SCM to investigate a salient issue in public management research: the failure of public organizations to provide managerial discretion in the implementation of performance management systems (Moynihan 2006). A central idea of new public management (NPM) doctrine is that increasing managers' focus on results and providing them greater authority will improve the performance of public organizations (Hood 1991; Moynihan 2006). In the US, however, these NPM tenets have only been partially adopted, with many governments embracing stringent performance accountability systems while neglecting to increase managerial flexibility (Moynihan 2006). This pattern of adoption is well-documented, but its implications for performance have not been thoroughly investigated in the literature.

I investigate this issue in the context of public schools, where school-level performance accountability systems are common, but the management of human and fiscal resources predominately remains centralized at the school district level. In recent years, however, several large school districts have worked to decentralize authority to the school level through a policy known as student-based budgeting (SBB). SBB allocates dollars rather than staff positions to schools, providing principals considerable discretion over human resource and program spending decisions. Examining whether implementing SBB improves student performance may provide evidence about the importance of discretion in performance accountability systems. Toward that end, I use SCM to investigate whether implementing SBB in the Houston Independent School District (HISD) improved student performance.

In this article, I first review the literature on performance management and its partial adoption in the public sector in the United States. Second, I discuss performance management and managerial discretion in the context of public schools, and describe SBB and its implementation in Houston. Third, I describe SCM and the data I use to implement it in this study. Fourth, I discuss the results. Finally, I discuss the implications of the results and suggest potential avenues for further research relating both to performance management and the application of SCM in public management research.

LITERATURE REVIEW

Performance Management

The popularity of performance measurement in the United States is undeniable, as the federal government, nearly every state government, and many local governments implemented some kind of performance information system by 2000 (Brudney, Hebert, and Wright 1999; Joyce 1993; Melkers and Willoughby 1998; Moynihan 2006). Indeed, the diffusion of performance measurement is so complete in the US that scholars have declared that performance joins "motherhood and apple pie as one of the truisms of contemporary American culture'' (Radin 2006, 4), or simply ''everyone is measuring performance'' (Behn 2003, 586). Less popular, however, is the NPM mantra ''let managers manage'' (Kettl 1997), as willingness to loosen control over financial and personnel systems is rare (Brudney, Hebert, and Wright 1999; Moynihan 2005; 2006).

According to the NPM view, enhanced discretionary space for public managers and performance measurement form a mutually supportive relationship, establishing a clear link between responsibility, authority, and reward (Moynihan and Pandey 2006, 121). Indeed, the characteristics associated with centralized, rule-based bureaucracies would seem to inhibit the very kinds of behaviors performance management encourages. Traditional, rule-based bureaucracies, according to the NPM view, are rigid and inflexible, emphasizing compliance and uniformity over flexibility and diversity, and inputs over results (Schick 2003).

While layering a performance accountability system with rewards and sanctions on top of a centralized, rule-based bureaucracy may direct some attention toward meeting policy goals, demands for compliance and following standard operating procedures remain, making it difficult for public managers to use their creativity and knowledge to effect change that will improve performance (Moynihan 2006). More problematic is the possibility that, without new tools and flexibility, managers will turn to undesirable practices to meet performance standards. Indeed, a large body of work on performance management focuses on its unintended consequences, such as gaming (Courty and Marschke 2004), creaming (Grizzle 2002), distraction from important non-mission-based values (Piotrowski and Rosenbloom 2002), and outright cheating (Bohte and Meier 2000). This is not to suggest that neglecting managerial flexibility is the primary source of the problems associated with performance management, but asking managers to do better without new tools and flexibility may frustrate those who already believe they are doing everything they can with the resources they have, encouraging undesirable behavior.

Moynihan (2005) describes the failure to enhance managerial authority as one of the puzzles in the development of performance management systems in the United States. Part of the reason for this failure may be that performance management adopters are more interested in its function as an accountability tool or its symbolic benefits—as they relate to electoral politics or citizen perceptions of government—than in its potential to improve government performance (Heinrich 2012; Hood and Peters 2004; Moynihan 2005). Previous research supports this notion, suggesting that public officials rarely use performance information in decision making (Joyce and Tompkins 2002; Melkers and Willoughby 2005; Moynihan 2006).

The neglect of managerial flexibility in the adoption of performance management in the US is well-documented (Brudney, Hebert, and Wright 1999; Moynihan 2006), but empirical research on its implications is still in its nascent stages. As Nielsen (2013) notes, large bodies of work consider the effects of performance management and the extent of managerial flexibility separately, but his is the first quantitative investigation examining their interaction. In the context of Dutch public schools, Nielsen (2013) finds that the extent of managerial flexibility positively moderates the effect of performance management. To contribute to this stream of research, I turn to the context of performance accountability in public schools in the United States.

Performance and Discretion in Public Schools

Performance in K–12

One of the more controversial applications of performance management practices in the US is in public education (Radin 2006). School-based performance accountability systems specify measurable performance standards for students in core subject areas, such as math and reading, and provide incentives for districts, schools, teachers, and students to meet them (Figlio and Loeb 2011). The incentives often comprise a mix of rewards and sanctions, such as bonuses for teachers and administrators in high-performing schools, or reassignment of teachers and staff or closings in low-performing schools (Figlio and Loeb 2011; Izumi and Evers 2002). Several states adopted performance accountability systems for public schools in the 1990s, and they are nearly universal among public schools in the United States since the implementation of the federal *No Child Left Behind Act* (NCLB) of 2001 (Figlio and Loeb 2011).

School-based performance accountability systems face criticisms similar to those directed at performance management in the public sector in general. Critics claim, for example, that they rely too much on standardized tests, neglecting other important aspects of education, and create perverse incentives for administrators and teachers to cheat (Radin 2006), and previous studies provide evidence supporting these claims (e.g., Jacob and Levitt 2003; 2005). An important shortcoming of many policies is that they specify the school as the unit of accountability for student achievement, but they do not give school-level actors any additional flexibility or decision-making authority for increasing student achievement (Figlio and Loeb 2011). In other words, like most public organizations in the United States, they fail to adopt the second half of the NPM notion of performance management.

Student-Based Budgeting

Most school districts in the US use finance systems that allocate school personnel and instructional materials on the basis of increments of overall student enrollment (Miles and Roza 2006). Schools, for example, may receive one teacher for every additional 25 students or an assistant principal for every 400 students (Ucelli et al. 2002). This model leaves individual schools with little decision-making authority over personnel and program spending, as school districts retain control over school budgets, curriculum, staffing, and scheduling (Miles and Roza 2006; Ouchi 2006; Ucelli et al. 2002). Critics argue that centralizing these decisions at the district level inhibits the ability of principals and other school-level personnel to respond to the educational needs of students. They also argue that allocating staff based on enrollment formulas leads to significant resource disparities across schools, especially in urban districts, as schools on the cusp of a given enrollment threshold may receive fewer resources than a school just over the threshold (Miles and Roza 2006; Stiefel, Rubenstein, and Berne 1998).

Urban school districts in the US, such as Seattle, Cincinnati, San Francisco, and Houston, are responding to concerns about traditional school finance models by implementing student-based budgeting (SBB) (Chambers, Levin, and Shambaugh 2010; Ucelli et al. 2002).[1] The details of SBB policies vary between districts, but they generally feature two primary components (Chambers, Levin, and Shambaugh 2010): First, they allocate dollars, rather than staff positions, based on the number of students enrolled and formulas that weight students according to their educational needs and the costs associated with meeting them. Student-based allocations may, for example, provide additional funding for high-poverty, gifted, disabled, bilingual, or vocational students, recognizing that additional resources may be required to educate these types of students (Ucelli et al. 2002). Second, they provide school-level autonomy for making decisions about program spending, scheduling, and personnel (Ouchi 2006; Snell 2009). The causal theory behind SBB is that providing school-level personnel authority and flexibility over needs-based allocations will improve student performance because they are more attuned to the unique needs of their students than district administrators (Chambers, Levin, and Shambaugh 2010; Curtis, Sinclair, and Malen 2014).

If lack of managerial flexibility is undermining the potential of performance management systems, then investigating the effect of SBB policies where stringent performance accountability systems are in place may provide clues about the difference managerial discretion makes. This article looks at the case of an SBB policy implemented in the context of one of the United States' oldest performance accountability systems for public schools.

Overview of Student-Based Budgeting in Houston Independent School District

Houston Independent School District implemented SBB in academic year 2000–2001 (Snell 2009). The policy applies a weighted student funding formula where schools receive a base allocation for each student depending on grade level, and additional weights may be applied to students with special needs. For example, schools receive additional funding for special education students, gifted and talented students, and those designated poverty or at-risk, among others (Snell 2009). The policy also allows principals discretion in making spending decisions based on these funds. Principals have discretion over class size, program spending, and staff allocations (Houston Independent School District 2011). A principal, for example, may decide to hire an additional counselor rather than an additional assistant principal without any need for approval from a superior (Ouchi 2006). HISD is the only school district in Texas to implement SBB, and is one of only a few in the US to decentralize both budgeting and staffing to the school level (Houston Independent School District 2011).

While Houston is the only school district in Texas with SBB, all school districts operate under the state's performance accountability system adopted by the Texas legislature in 1993 (Izumi and Evers 2002). The state rates schools according to student performance on state-administered standardized tests, dropout rates, and attendance rates (Goertz, Duffy, and Le Floch 2001). A school receives a rating of exemplary, recognized, academically acceptable, or unacceptable/low-performing depending on how well it does according to these measures. The system rates a school acceptable and making adequate yearly progress, for example, if at least 50% of its total students pass each subject of the standardized test, 6% or less drop out, and its attendance rate is at least 94% (Izumi and Evers 2002). Schools receiving a rating of acceptable or higher are eligible to receive rewards in the form of monetary bonuses, while low-performing schools are subject to a series of interventions and sanctions ranging from public hearings and special on-site intervention teams to school closings (Izumi and Evers 2002). In addition to formal rewards and sanctions, a variety of performance measures are publicly available and reported in the media (e.g., Peters 2009), such as high school graduation rates and SAT and ACT participation and results (Izumi and Evers 2002). While these measures do not directly affect schools' ratings, they do communicate aspects of performance that interest parents and the community.

Previous Research

The body of research estimating the effects of SBB is surprisingly small, focusing mostly on whether SBB creates more equitable distribution of resources in school districts. Miles and Roza (2006) find that SBB reduced resource disparities in Cincinnati and Houston school districts, but their analysis does not include control districts. Baker (2009) also examines SBB in Cincinnati and Houston, but uses more sophisticated methods. He finds that implementing SBB in Cincinnati and Houston did not increase spending in schools with high percentages of at-risk and high-poverty students relative to similar schools without SBB, but he does find that it reduced resource disparities overall (Baker 2009). Chambers et al. (2010) find implementing SBB in Oakland and San Francisco coincided with higher levels of per-pupil spending in high poverty schools.

I found only one previous study attempting to identify whether SBB improves student outcomes. Baker and Elmer (2009) find schools in Houston and Seattle school districts do not perform better on state standardized tests than schools in districts without SBB, controlling for student demographics, school spending, and school size. They do find, however, that Houston schools have a more positive year-to-year change in pass rates than other Texas schools in districts without SBB (Baker and Elmer 2009). Baker and Elmer's (2009) findings have important limitations. First, in the case of Houston, their sample does not include pre-implementation data. Their reason for not including pre-implementation data is that Houston schools were subject to allegations of fraudulent boosting of performance measures (Baker and Elmer 2009). They use a binary variable indicating a school is in HISD, which does not provide evidence about whether SBB improves student outcomes. Instead, the coefficient indicates that Houston schools do not outperform other Texas schools, controlling for student demographics, school spending, and school size. The important question is whether SBB allows Houston schools to perform better than they would have without SBB, not necessarily better than schools in other districts. This question cannot be answered without preimplementation data.

Previous scholarly research is limited, but interviews indicate that there is at least a perception among principals that SBB helps them improve student outcomes. For example, one Houston principal explains that SBB gave her the flexibility to fund after-school programs for English-language learners, which she says helped the school achieve a higher status in the state accountability system (Archer 2005). Another Houston principal explains how SBB provides flexibility in hiring personnel: "I have control over whether I want . . . an additional assistant principal or a business manager, whether I want to hire another math teacher or another history teacher. I don't need any approvals from anyone to make these decisions" (Ouchi 2006, 302).

Principal interviews suggest that SBB provides greater school-level autonomy and flexibility, but does that translate into improved student outcomes? Following the logic of NPM reform, I expect that implementing SBB helped Houston increase student achievement relative to how it would have performed without the policy. In the next section, I describe SCM and the data I use for investigating whether SBB improved student outcomes in Houston.

METHODS AND DATA

Public management researchers are often interested in estimating the effects of aggregate-level policy interventions or events affecting aggregate units. These units may be cities, nonprofit organizations, government agencies or other aggregate entities, such as regions or entire countries. The reforms or interventions may be any of a variety of public management implementations, such as performance accountability systems, budget reforms, or collaborative arrangements for public service delivery. A simple approach for estimating the effects of these types of interventions is comparing post-intervention changes in the outcome of interest between the treated unit and one or more control units a researcher believes are similar to the treated unit. The process of selecting control units, however, is often based on a researcher's subjective assessments of affinity with the treated unit and the true extent of their similarity may be less than transparent to both the researcher and the reader. The process of selecting control units has important consequences. If the control units are not sufficiently similar to the case of interest, differences in outcomes between them may be erroneously attributed to the event or intervention of interest when they actually reflect systematic disparities in their characteristics (Abadie, Diamond, and Hainmueller 2014). To illustrate this problem, I describe a simple approach to estimating the effect of SBB in Houston.

A researcher seeking to estimate the effect of SBB in Houston will want to identify a similar school district that did not implement SBB for comparing outcomes before

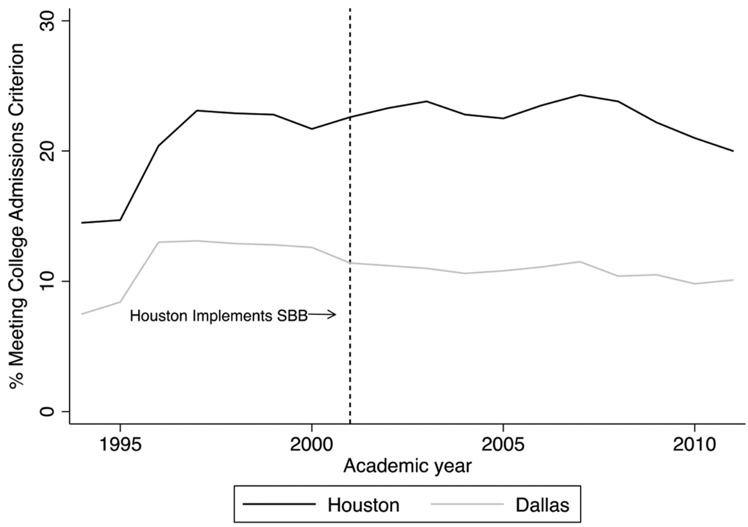

Figure 1. Trends in percent of students meeting college admissions criterion: Houston vs. Dallas.

and after implementation. HISD is the largest school district in Texas and is located in a large metropolitan area. A reasonable choice for comparison, then, might be Dallas Independent School District, the second largest in the state and also located in a large metropolitan area. Similar to Houston, Dallas has a large number of economically disadvantaged and minority students and it is reasonable to expect that its teachers and administrators face similar challenges to their counterparts in Houston. Comparing the trajectories of a student outcome variable before and after Houston implements SBB may provide evidence about whether the policy improved student outcomes in Houston.

Figure 1 shows the percentage of students meeting or exceeding college admissions criterion—a student performance measure described in more detail later in this article—in Houston and Dallas over time. It would be difficult to defend any causal claims based on the evidence provided in Figure 1 because the trends are considerably different in the pretreatment period. If a researcher applied difference-in-differences in this case, they would violate the common trends assumption (Angrist and Pischke 2008). In other words, Dallas, by itself, is not a good counterfactual for Houston. But what if a researcher could borrow only the parts of Dallas that are similar to Houston? This is where SCM becomes a useful research tool. In the next section, I describe the basic intuition of SCM.

Synthetic Control Method

The central idea behind SCM is that a weighted combination of control units selected by data-driven methods provides a better comparison for a treated unit than any single control unit alone (Abadie, Diamond, and Hainmueller 2010). SCM creates a synthetic control by selecting control units from a ‘‘donor pool’’ comprised of units that did not experience the treatment of interest. Specifically, SCM uses a data-driven algorithm to select a weighted combination of control units that most closely approximate values of an outcome and a series of predictor variables of the treated unit in the pretreatment period. SCM promotes transparency in this process by providing output that makes explicit the contribution of each control unit in the synthetic control as well as a comparison of values for predictor variables between the synthetic control and treated unit. The synthetic control will closely match the predictor variable and outcome values of the treated unit in the period leading up to the intervention if SCM is successful. The values of the outcome variable for the synthetic control in the post-treatment period become the counterfactual: what the values of the outcome variable would look like had the policy intervention not occurred. Discrepancies between the treated unit and the synthetic control in the post-treatment period, then, provide evidence and an estimate of a treatment effect.

One of SCM's key advantages is dealing with endogeneity from omitted variable bias due to the presence of unobserved time-invariant and time-varying factors that may affect the outcome of interest (Abadie, Diamond, and Hainmueller 2010; 2014; Billmeier and Nannicini 2013). While estimation techniques such as fixed-effects and difference-in-differences account for unobserved time-invariant—or relatively stable over time—factors, such as those associated with an organization's geography, they cannot account for unobserved time-varying factors, such as changes in the quality of an organization's personnel. SCM helps control for both observed and unobserved time-varying and time-invariant factors affecting the outcome of interest by matching on pre-intervention outcomes (Abadie, Diamond, and Hainmueller 2014). Essentially, only units alike in both observed and unobserved outcome determinants should exhibit similar values of that outcome over an extended period of time (Abadie, Diamond, and Hainmueller 2014).

DATA

I use publicly available school-district-level data from the Texas Education Agency to estimate the effects of SBB in Houston.[2] Texas has over 1,000 school districts with enrollments as low as 20 students and as high as 200,000. Abadie, Diamond, and Hainmueller (2014) recommend restricting datasets to units similar to the treated unit both to improve the performance of SCM and to limit biases associated with applying SCM using large samples.[3] I use school district size as a basis for limiting the sample because previous studies find it has implications for student performance, resource allocations, and management (Andrews, Duncombe, and Yinger 2002). Thus, I restrict the dataset to Texas school districts with at least 25,000 students. The final sample includes 29 school districts, including Houston, for the years 1994 through 2011; the final year data are available for all of the variables.

Outcome Variables

I estimate the effect of SBB in Houston on two measures of student performance. The first measure I use is the percentage of students meeting college admissions criterion. Texas defines meeting college admissions criterion as scoring 1,100 or higher on the Scholastic Achievement Test (SAT) or 24 or higher on the American Achievement Test (ACT). Meeting this criterion does not have any direct implications for individual students, but it does serve as an indicator of whether a student will be competitive if they choose to apply to a college or university. Higher education institutions in the US widely use these tests when making decisions about whether applicants possess the minimum level of knowledge required to succeed in their institutions (Amrein and Berliner 2002). While this measure does not have direct implications for accountability ratings, it is part of the performance reporting requirements for schools and districts, and it serves as an indicator of how well they are preparing students for higher education. Thus, the measure is likely watched closely by parents and policymakers concerned with college-bound student preparedness.

The second measure of student performance is the percentage of students passing state standardized tests in all subjects (reading, writing, and math). Students in grades 3 through 8 and 10 are required to take the tests, and students must pass the grade 10 test to qualify for graduation. Pass rates on the state exams constitute an important component of the state's performance accountability system and are consequential for whether a school faces rewards or sanctions after a school year (Izumi and Evers 2002). This is a commonly used measure for researchers assessing the performance of Texas schools and school districts (e.g., Bohte 2001; Meier and O'Toole 2002).

The performance measures I use, the percentage of students meeting college-admissions criterion and the percentage of students passing state standardized tests, represent upper and lower ends of the student performance distribution. Passing state-administered exams is a minimum competency determined by the state for progressing through the school system, while meeting college admissions criterion on the SAT or ACT is characteristic of a high-performing student. In other words, there are likely important differences between a student barely passing the state exam versus a student barely meeting the college admissions criterion threshold. While education researchers frequently examine only one measure of student performance, such as average test scores, some scholars argue that examining measures that capture basic competencies, such as state standardized tests, as well as measures of more challenging benchmarks, such as the college admissions criterion, provides a richer understanding of student achievement (Andrews, Duncombe, and Yinger 2002; Duncombe, Ruggiero, and Yinger 1996).

Predictor Variables

Selecting predictor variables for matching a synthetic control to a treated unit in an SCM analysis is similar to the process of selecting controls for a traditional regression model. Researchers, then, should use previous research estimating the factors associated with their outcome of interest to guide their choice of predictor variables. I draw on the extensive literature on the determinants of student performance in the US to develop a set of predictor variables for this analysis. Despite a long history dating back to the 1966 Coleman Report (Coleman et al. 1966), there is considerable controversy about the relative importance of factors determining student performance. In general, however, the literature suggests the importance of factors relating to student characteristics, family background, and school inputs (Andrews, Duncombe, and Yinger 2002; Hanushek, Rivkin, and Taylor 1996; Hanushek and Raymond 2005). Common school input controls include class size, measures of teacher quality, district size, and per-pupil expenditures (Andrews, Duncombe, and Yinger 2002; Driscoll, Halcoussis, and Svorny 2003; Duncombe, Miner, and Ruggiero 1995; Hanushek and Raymond 2005). Important student and family background controls include income, English proficiency, and race and ethnicity (Andrews, Duncombe, and Yinger 2002; Hanushek and Raymond 2005).

I use the following school input variables to match Houston with a synthetic control: the percentage of teachers with less than five years experience, average teacher salary, students per teacher, instructional expenditures per student, revenue per student, district size measured as total students, and the percentage of central administrators as a fraction of total full-time district employees, which helps account for a district's administrative capacity. I include the following variables to match factors relating to student and family characteristics: percent low-income students, percent African-American students, percent Hispanic students, and percent bilingual/ESL students. I add the percentage of students taking college admissions exams as a predictor in the college admissions criterion model. Finally, following Abadie, Diamond, and Hainmueller (2010), I include three years (1996, 1998, and 2000) of pretreatment values of the outcome variable for improving the fit between the treated and synthetic units in the pretreatment period.

APPLICATION

Applying SCM to the case of SBB in Houston, I first build a dataset including data for Houston and 28 control districts for several years before and after Houston implements SBB. In this case, I have seven years of pretreatment data and 11 years of post-treatment data, including the year of implementation (academic year 2000–2001). SCM then uses an algorithm to assign weights to potential control districts based on their similarity to Houston according to values of the education production function variables and the outcome variable in the pretreatment period. The algorithm chooses a synthetic control that minimizes the root mean square prediction error (RMSPE), which measures the lack of fit between Houston and the synthetic control, over the pretreatment period.[4] I run separate SCM analyses for both student performance measures.

Constructing Synthetic Houston

Table 1 shows the weights assigned to each school district in the donor pool in each SCM analysis. The resulting synthetic Houston in the state exam pass rates

TABLE 1
District Weights

District	*State Exams*	*College Admissions Criterion*
Killeen	0	0
San Antonio	0.028	0.243
Northeast	0	0
Northside	0	0
Brownsville	0	0
Plano	0	0
Dallas	0.517	0.375
Garland	0	0
Irving	0	0
Mesquite	0	0
Richardson	0	0.038
Ector County	0	0
El Paso	0.071	0
Ysleta	0	0
Fort Bend	0	0
Aldine	0.264	0
Alief	0	0
Cypress-Fairbanks	0	0
Klein	0	0
Pasadena	0.057	0
Spring Branch	0.003	0
Lubbock	0	0
Conroe	0	0
Corpus Christi	0.06	0
Amarillo	0	0
Arlington	0	0
Fortworth	0	0
Austin	0	0.344

analysis is a weighted average of San Antonio, Dallas, El Paso, Aldine, Pasadena, Spring Branch, and Corpus Christi school districts, with all weights summing to 1. All other school districts in the donor pool are weighted zero. The resulting synthetic Houston in college admissions criterion model is a weighted average of San Antonio, Dallas, Richardson, and Austin school districts with all weights summing to 1. All other school districts in the donor pool are weighted zero.

Table 2 compares the pretreatment characteristics of Houston to those of both synthetic controls, Dallas, and the average for all districts in the donor pool. The results in the table indicate that, in the majority of cases, the synthetic controls provide better comparisons for Houston than Dallas or an average of the districts in the donor pool. There are a few notable discrepancies, however, in student demographic variables in the college admissions criterion synthetic control, such as the percent of African-American students and the percent of bilingual students. The pretreatment

TABLE 2
Predictor Means

Variable	*Houston*	*Synthetic (State Exams)*	*Synthetic (College Admissions)*	*Dallas*	*Avg. Control Districts*
% Taking College Admissions Exams	59.757	–	61.058	52.786	64.817
% African-American Students	34.529	33.622	25.130	41.629	14.940
% Hispanic Students	51.586	49.857	52.153	45.529	39.186
% Economically Disadvantaged	66.557	69.360	67.027	73.029	45.635
% Bilingual/ESL	23.100	23.076	18.106	27.286	12.868
% Teachers $\leq$ 5 years exp.	34.357	36.103	30.799	33.214	34.553
Teacher Salary (ln)	10.454	10.438	10.442	10.449	10.412
Students per Teacher	17.729	16.139	16.323	16.743	16.158
Instructional Expenditures per Pupil	2,923.329	2,949.873	2,965.564	2,934.682	2,836.106
Revenue per Pupil	5,133.000	5,144.772	5,481.070	5,233.571	5,319.107
% Central Administrative Staff	0.929	0.076	0.641	0.057	0.477
Total Students (ln)	12.241	11.405	11.399	11.936	10.673
State Exam Pass Rate (1996)	57.300	56.360	–	50.800	67.325
State Exam Pass Rate (1998)	69.500	67.980	–	59.900	77.532
State Exam Pass Rate (2000)	70.300	68.334	–	59.900	79.264
College Admissions Criterion (1996)	20.400	–	21.682	13.000	28.664
College Admissions Criterion (1998)	22.900	–	22.749	12.900	29.382
College Admissions Criterion (2000)	21.700	–	21.961	12.600	29.150
RMPSE		0.636	2.019		

RMSPE in both cases is quite low, 2.019 for state-qualifying exam pass rates and 0.636 for college admissions criterion, indicating that the synthetic controls provide a good fit with Houston in the period leading up to the implementation of SBB (Abadie, Diamond, and Hainmueller 2010). Overall, the results provide confidence that the synthetic controls provide a faithful representation of what the trajectories of the two outcomes would look like if Houston did not implement SBB.

RESULTS

Graphical output is one of SCM's key features. While the method's underlying mechanisms are quite technical, its graphical output is highly intuitive. This feature may be especially helpful for public management researchers wishing to share their research with practitioners and other audiences unfamiliar with interpreting advanced statistical results. Indeed, some scholars claim the increasing use of advanced methods contributes to an increasing divide between scholarship and practice (Meier and Keiser 1996; Raadschelders and Lee 2011). In the remainder of this section, I describe the results of the SCM analyses.

Percentage of Students Meeting College Admissions Criterion

Figure 2 displays trends in the percentage of students meeting college admissions criterion for Houston and its synthetic counterpart during the period 1994–2011. In contrast to Dallas (shown in Figure 1), the synthetic control tracks closely with Houston in the pretreatment period. The close match in the pretreatment period, combined with the overall high degree of balance in values of the predictor variables,

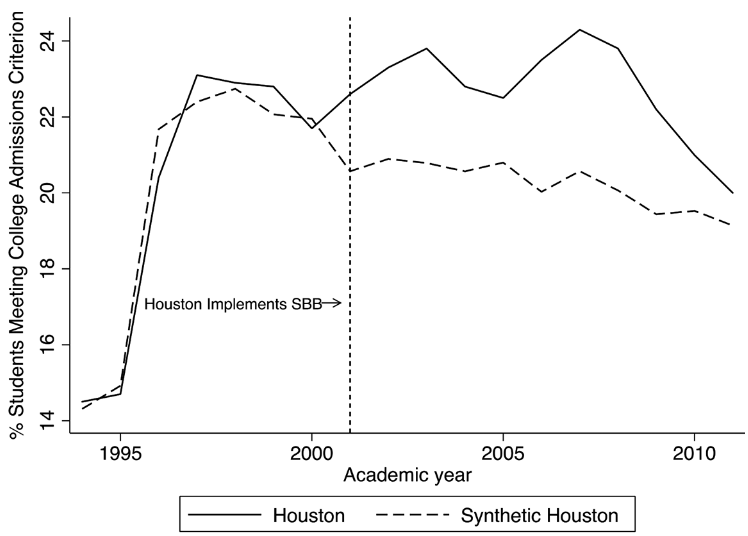

Figure 2. Trends in percent of students meeting college admissions criterion: Houston vs. Synthetic Houston.

suggests that the synthetic control provides a plausible approximation of what trends in the percentage of college admissions criterion would look like if Houston did not implement SBB.

The estimate of the effect of SBB on college admissions criterion is the difference in values between Houston and its synthetic counterpart in the period after Houston implements SBB. Figure 2 shows a divergence in the values that begins immediately in the implementation year (2001). While synthetic Houston experiences a slight decline before leveling off, Houston maintains an upward trajectory, achieving a sizable gap in 2007 before declining until it almost matches synthetic Houston in 2011.

Figure 3 displays trends in state qualifying exam pass rates for Houston and its synthetic counterpart. The pretreatment trends match closely, but so do the post-treatment trends, indicating that SBB did not improve Houston's performance on state exam pass rates. What explains the divergence in results between this measure of student performance and college admissions criterion? First, it is important to note that, in 2003, Texas implemented a new standardized test, replacing the Texas Assessment of Academic Skills (TAAS) with the more challenging Texas Assessment of Knowledge and Skills (TAKS). The steep decline in performance for both Houston and the synthetic control after 2002 likely reflects the increased difficulty of the exam. While all control units experienced this shock, the noise it creates makes it a less than ideal measure for applying SCM. A second interpretation of the divergence is that these measures represent different ends of the performance distribution and one should not necessarily expect them to move together. It may be, for example, that after Houston implemented SBB, principals used their additional discretion to direct resources toward improving students' college readiness.

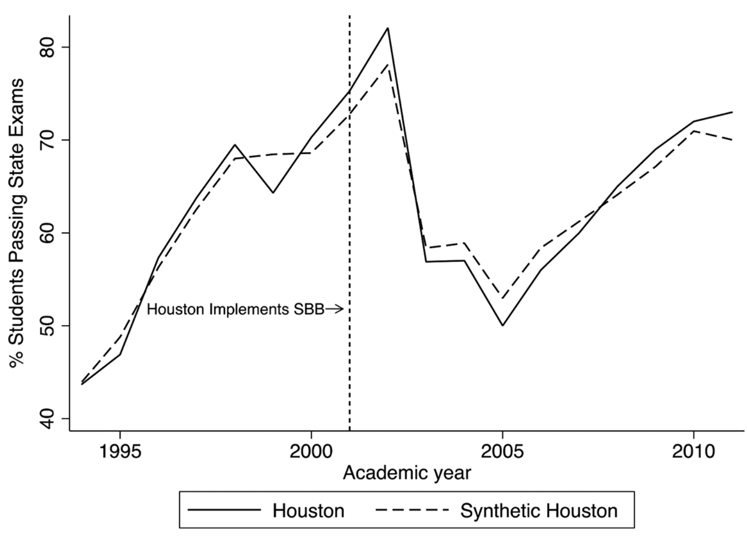

Figure 3. Trends in percent of students passing state exams: Houston vs. Synthetic Houston.

Inference

Traditional inferential techniques for assessing statistical significance are not possible in SCM applications, but Abadie, Diamond, and Hainmueller (2010; 2014) demonstrate placebo tests that researchers may use for accomplishing similar ends. The placebo test works by iteratively assigning the treatment to the units in the donor pool that were not actually treated, creating a distribution of placebo effects that researchers may compare to the effect observed for the unit that actually received the treatment. The intuition behind this test is that if the results are significant, there will be a low probability of observing an effect larger in a non-treated unit than the effect observed in the treated unit. I describe how to implement this test in more detail in the following.

I conduct a series of placebo tests to evaluate the significance of the positive effect observed for SBB on the percentage of students meeting college admissions criterion. The placebo tests use SCM to iteratively assign treatment to districts in the donor pool. The process creates a distribution of estimated gaps in the outcome variable between the districts where no intervention took place and their synthetic counterparts. I then compare these gaps to the gap observed for Houston. If the placebo tests generate gaps similar in magnitude to the one estimated for Houston, my interpretation is that the analysis does not provide significant evidence of an effect of SBB on the percentage of students meeting college admissions criterion.

Figure 4 shows the results for the placebo tests. The light-gray lines show the gaps observed for the control districts that were not treated, while the black line shows the gap observed for Houston. Abadie, Diamond, and Hainmueller (2010) recommend omitting control units with an RMSPE more than two times larger than the treated

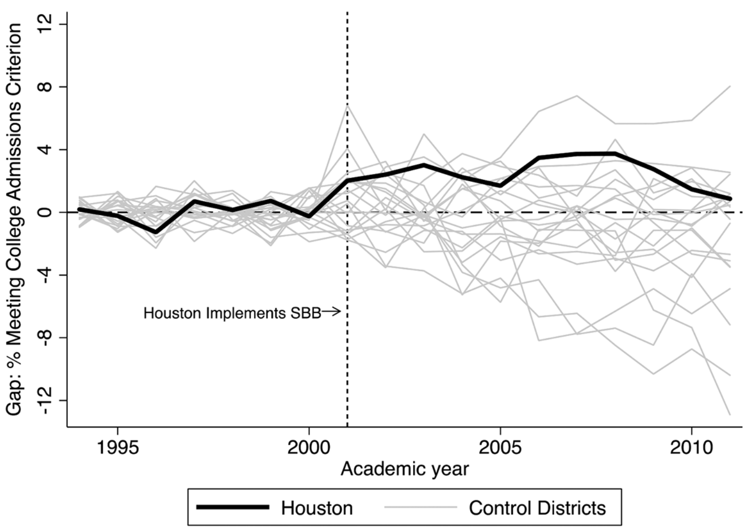

Figure 4. Percent of students meeting college admissions criterion gaps in Houston and placebo gaps in 20 control districts (discards districts with pre-intervention RMSPE two times higher than Houston's).

unit because any gaps observed in the post-treatment period are likely artificially created due to a lack of fit. Thus, I omit eight school districts with RMSPEs more than twice as large as Houston's. Compared to the remaining 20 control districts, the gap for Houston is unusually large with one district systematically outperforming Houston in the post-intervention period. The probability of estimating a gap of the magnitude observed for Houston under random permutation, then, is 1/20 or 5%.

Robustness Test

A potential concern when running SCM is that one control unit may be driving the results and not the policy intervention of interest. In the case of the college admissions criterion result, one may be concerned that, given the heavy weight assigned to Dallas, something going wrong in Dallas is driving the results and not SBB. To test whether Dallas or another control district assigned weight in the synthetic control is driving the results, I run a series of leave-one-out estimates which iteratively reestimate the baseline model, each time omitting one of the districts that received a positive weight in the original estimate (Abadie, Diamond, and Hainmueller 2014). Figure 5 shows the results of the leave-one-out tests. The solid dark line shows Houston's outcome trajectory, the dashed line shows the original synthetic control's outcome trajectory, and the solid light-gray lines show the trajectories of the synthetic leave-one-out estimates. The figure shows that the results are robust to the exclusion of any of the original control districts assigned a weight, as the gap between treated Houston and synthetic Houston remains in each of the leave-one-out estimates. If any of the leave-one-out synthetic controls erased

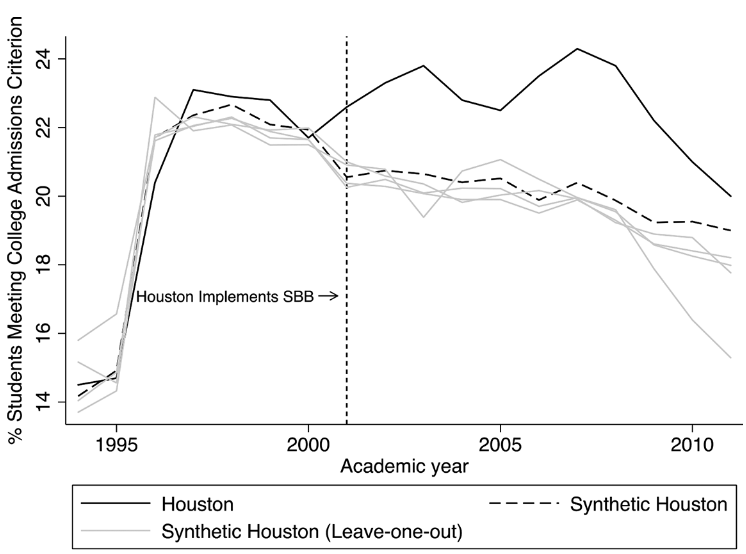

Figure 5. Leave-one-out distribution of the synthetic control for Houston.

the gap with treated Houston, I would conclude that poor performance in one or more of the control districts is responsible for the outcome divergence, not SBB.

DISCUSSION

A key NPM principle is that performance accountability and managerial discretion form a mutually supportive relationship, but governments often neglect to enhance managerial discretion when implementing performance accountability (Moynihan 2006). The goal of this SCM application was to investigate whether providing principals discretion in Houston, a school district operating under a stringent performance accountability system, improved student performance relative to how it would have performed without principal discretion. The SCM results provide mixed evidence on the effect of SBB on student performance in Houston. First, there is no discernible difference in student pass rates on state exams between Houston and its synthetic counterpart in the period after Houston implemented SBB. There is a notable difference, however, in the percentage of students meeting college admissions criterion.

As I discuss in the results section, there are a couple of potential explanations for the divergence in results observed in the two SCM analyses. First, Texas replaced its state exam in 2003, just two years after Houston implemented SBB, with a more rigorous version. The additional noise created by this change may hide any potential effects of SBB on this performance measure. A counterargument, however, is that if additional principal discretion were effective, it should have helped Houston schools deal more effectively with the shock created by the test change relative to control districts where principals have little discretion. A second explanation for the diversion in results is that they represent different ends of the performance distribution. Passing state exams means meeting a basic competency while meeting college admissions criterion means scoring well enough on a college admissions exam to be competitive in the college admissions process. It is possible that principals in Houston use their discretion to help prepare students for the process of applying to colleges and universities. While all principals in Texas likely dedicate as many resources as possible toward helping students pass state exams, as there are direct accountability implications, the additional flexibility granted Houston principals may allow them to direct resources toward other areas that were previously unfeasible due to lack of flexibility.

An important limitation of this analysis is it does not provide evidence about how principals in Houston use their discretion. As Abadie, Diamond, and Hainmueller (2014) explain, however, SCM facilitates qualitative analysis by providing a transparent, quantitative tool to select or validate comparison units. Thus, SCM offers an avenue for future qualitative research investigating how principals in Houston use their discretion in attempts to improve student success—in college admissions exams, for example—compared to principals with less discretion in the control districts selected by SCM.

USING SCM IN PUBLIC MANAGEMENT RESEARCH

The synthetic control method has great potential as a tool for public management research. In this final section, I outline a few research areas where SCM may help advance public management research, summarize the step-by-step procedure for implementing SCM, discuss SCM's limitations, and provide references for further reading and installing software for implementing SCM.

Potential Applications in Public Management Research

In this article, I demonstrated an SCM application in the context of performance management and discretion in public schools, but SCM has potential applications in a variety of public management research areas. Researchers might use SCM to estimate the effects of public management reforms in areas such as financial management and budgeting, performance management, human resources, privatization, and collaboration, among others. In this section, I elaborate on a few potential areas where SCM may be a useful tool for public management scholars.

Collaboration and Networks

Governments increasingly rely on collaborative arrangements—contracts, public-private partnerships, networks, or intergovernmental agreements—for delivering public services, and public management researchers are creating a large body of literature devoted to understanding their dynamics and the factors contributing to their evolution and effectiveness (Isett et al. 2011). The literature on networks in particular is growing exponentially, but methodological challenges are inhibiting progress in some areas (Isett et al. 2011; Moynihan et al. 2011). One of the most important questions in the study of public sector networks (for example, whether networks improve public sector outcomes) remains unanswered (Agranoff and McGuire 2001; Isett et al. 2011; Provan and Kenis 2007; Provan and Milward 2001).

SCM may help researchers overcome several obstacles that limit their ability to assess whether networks and other collaborative arrangements improve public sector outcomes. First, networks often arise under unique, complex circumstances, and it may be difficult to find a plausible control case for comparing gains in performance (Moynihan et al. 2011). A synthetic control closely approximating the relevant characteristics of a jurisdiction before it implements a network may provide researchers with a compelling basis of comparison for the evaluation of a network's performance. A second challenge associated with evaluating the performance of collaborative arrangements is that network development is an iterative process and its effects on outcomes may emerge gradually or change over time (Moynihan et al. 2011). SCM's visual output may help inform researchers' understanding of this developmental process by showing changes in the outcome of interest over time, relative to a synthetic control not using a network for delivering a public service.

Results-Based Reforms

Despite widespread adoption among Western governments and considerable attention in the literature, we still lack definitive evidence about whether results-based reforms improve public sector outcomes (Moynihan et al. 2011). An important obstacle to making progress in this area is that implementing results-based reform is often predicated on a perception of a performance problem (Blom-Hansen, Morton, and Serritzlew 2015). Consequently, there are likely several factors related to both the implementation of a prescription and performance on the outcome of interest (Konisky and Reenock 2013). In other words, there are likely systematic differences in both observable and unobservable characteristics between implementers and non-implementers of performance reforms that may bias results using traditional regression methods (Blom-Hansen, Morton, and Serritzlew 2015; Konisky and Reenock 2013). SCM helps control for observed and unobserved factors that may affect the outcome of interest by matching units on pre-intervention outcomes (Abadie, Diamond, and Hainmueller 2010). Only units that are alike in both observed and observed determinants of an outcome should produce similar trajectories of the outcome over extended periods of time (Abadie, Diamond, and Hainmueller 2014).

International Comparative Public Administration

Public administration and management research is becoming increasingly international, and it is important to develop an understanding of how and why governance differs across countries and how those differences contribute to policy outcomes (Fitzpatrick et al. 2011). A comparative approach has potential to advance public management research by revealing how differences in governance contexts present opportunities and challenges for adopting uniform "best practice" solutions (Fitzpatrick et al. 2011). Variation in political traditions and institutional settings among countries, however, makes comparing different organizational or management approaches difficult (Svensson, Trommel, and Lantink 2008). SCM has a clear application for researchers interested in estimating the effects of national-level reforms, as it can provide a counterfactual case that more closely resembles the country of interest than any single country alone (Abadie, Diamond, and Hainmueller 2014).

Bridging the Quantitative-Qualitative Divide

The synthetic control method also has potential to bridge the quantitative/qualitative divide and help researchers satisfy the call for more mixed method research designs (Groeneveld et al. 2014; Perry 2012; Pitts and Fernandez 2009). Specifically, it brings statistical rigor to comparative case studies, and provides qualitative researchers with a transparent, data-driven method for selecting comparison cases. This point, in particular, should be emphasized because bridging the two traditions can help researchers produce richer research that broadens our understanding of public management.

Connecting Scholarship and Practice

Finally, the synthetic control method may also help the effort toward bridging the divide between scholarship and practice. Scholars cite the increasing use of sophisticated statistical methods that practitioners cannot reasonably be expected to understand as an important factor in the growing divide between scholarship and practice in public administration (Meier and Keiser 1996; Raadschelders and Lee 2011). SCM may help researchers communicate their findings to a broader audience of practitioners by providing intuitive, graphical results that do not require advanced knowledge for interpretation.

Implementing SCM: A Step-By-Step Procedure

Figure 6 shows the step-by-step procedure for implementing SCM. The first step is constructing a balanced panel dataset with one treated unit and several untreated units serving as potential controls in the donor pool, where all units in the dataset have complete data on the outcome variable and predictor variables for the entire period of the study. While SCM may provide an unbiased estimator with only a single pretreatment period, Abadie, Diamond, and Hainmueller (2010; 2014) do not recommend using the method when the number of pretreatment periods is small. The credibility of a synthetic control will depend on how well it tracks with a treated unit over an extended period of time in the pretreatment period. It is advisable, then, for researchers to include as many pretreatment periods as their data allow. I provide further discussion on this issue in the limitations section.

Abadie, Diamond, and Hainmueller (2010; 2014) do not offer specific guidance on the number of control units researchers should include in the donor pool, and the suitable number will likely vary depending on the particular circumstances of the application. They do, however, caution that a large donor pool with many units dissimilar to the treated unit may lead to interpolation bias and overfitting.[3] If researchers are using a particularly large dataset, it is advisable to come up with substantive criteria for reducing the number of control units. A very small donor pool, on the other hand, may raise concerns about whether there are enough control units to construct an adequate synthetic control, and whether there are enough units to implement credible placebo tests. The extent of these concerns, however, will vary between applications and problems may not reveal themselves until researchers attempt to implement SCM.

The second step is selecting a set of predictor variables that SCM will use to assign weights to the synthetic control. Researchers should follow procedures similar to those they would follow for selecting control variables in traditional statistical applications, such as ordinary least squares. The researcher may also select a number of pretreatment periods for matching on the outcome variable. Including pretreatment periods of the outcome variable may improve the fit between the treated and synthetic units. Finally, the researcher must specify which unit in the dataset is the treated unit and the time period the treatment begins.

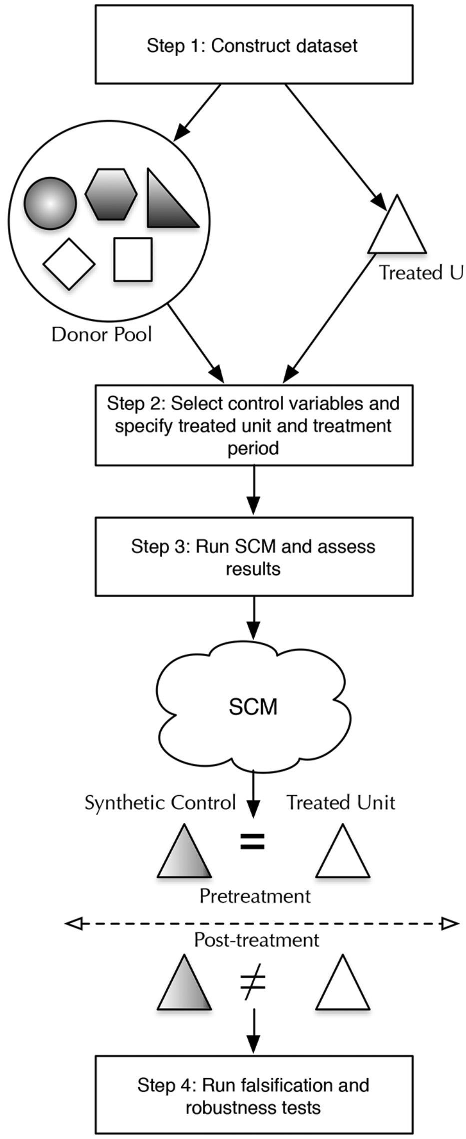

Figure 6. Implementing SCM: A Step-By-Step Procedure.

After selecting a treated unit, treatment period, donor pool, and predictor variables, step 3 is running SCM and assessing the results. SCM presents the researcher with two tables: one showing the weights assigned to the controls and a second table showing mean pretreatment values of the predictor variables for the treated unit and the synthetic control. SCM will also present the researcher with a figure showing the trajectory of the outcome variable for the treated and synthetic units over time with a vertical line representing the year of the intervention or policy. If SCM is successful, the mean values of the predictor variables for the treated and synthetic units should match closely and the figure should depict nearly identical values of the outcome variable over the pretreatment period. If they do not match closely, both in terms of mean predictor values and pretreatment outcome values, it is permissible for researchers to trim the donor pool to units more similar to the treated unit and run synth again. Researchers may also try applying SCM's optimization options, which are explained in the SCM software help files. If researchers are satisfied with their results, the final step is running placebo and falsification tests, as I describe in the inference and robustness test sections. In the next section, I discuss SCM's limitations, which helps inform the process I described in this section.

Limitations

SCM is a promising tool for public management research, but it is not a panacea and it may not always be appropriate. First, SCM's single case focus limits its generalizability. SCM, in other words, does not show what happens if other untreated units experience the intervention of interest. This may be a particularly important shortcoming for researchers hoping to make a strong case for the applicability of a given public management prescription in different contexts. One potential way to provide cross-validation in these cases is conducting multiple SCM studies on units experiencing a given intervention and assessing whether the results show a systematic pattern that can support a general conclusion about an intervention's effectiveness. Researchers using this strategy, however, must be sure to exclude other treated units from donor pools, which becomes more difficult as the number of treated units increases relative to the number of possible untreated controls. Billmeier and Nannicini's (2013) multi-country study of economic liberalization episodes is an excellent example application for researchers interested in this approach.

Governments and organizations may try a litany of different prescriptions for public management or policy problems over time, which may influence the trajectories of outcomes of interest. Researchers using SCM, then, must pay careful attention to other interventions or events, occurring in both the treated and control units, which may also affect the outcome of interest during the period of study. If the treated unit implements an additional intervention aimed at affecting the outcome of interest in the post-implementation period, the researcher must carefully consider whether and how this affects their results. At the same time, it is also important to consider whether units in the donor pool experienced unique interventions or shocks potentially affecting the outcome that were not also experienced by the treated unit (Abadie, Diamond, and Hainmueller 2014). The leave-one-out test helps alleviate concerns about idiosyncratic shocks, but researchers must still be cautious.

Finally, SCM requires a sizable number of pretreatment years for establishing a credible counterfactual. If only a few pretreatment years are used, it limits confidence about the similarity between the synthetic and treated groups and the ability of SCM to control for unobservables. At the same time, a large number of post-intervention years may be needed to demonstrate the extent of a policy effect, especially in cases where the effect of the intervention reveals itself gradually (Abadie, Diamond, and Hainmueller 2014).

Software

Synth, the statistical software for implementing SCM, is available for Stata, MATLAB, and R. Jens Hainmueller maintains a helpful webpage with details on installing Synth in each of these programs, as well as a video tutorial for using Synth in Stata.[5] Another instructional resource is Abadie, Diamond, and Hainmueller's (2011) article on applying SCM in R, which is helpful regardless of whether one is using R.

Further Reading

Researchers interested in using SCM should consult the foundational articles by Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010; 2014). These articles provide thorough descriptions of SCM's technical details and excellent example applications. Other notable SCM applications include Hinrichs' (2012) study on the effects of affirmative action bans in US higher education, Billmeier and Nannicini's (2013) study of economic liberalization episodes, and Cavallo et al.'s (2013) study of the effects of catastrophic natural disasters on economic growth. Each of these articles provides helpful discussions of SCM and its benefits and limitations.

ACKNOWLEDGEMENTS

I am grateful to David Pitts, Seth Gershenson, and Edmund Stazyk for their comments and support. I also thank the anonymous reviewers and symposium editors, Stephan Grimmelikhuijsen, Lars Tummers, and Sanjay Pandey, for their insightful suggestions and support.

NOTES

1. Student-based budgeting is sometimes also referred to as "weighted student funding" or "student-based funding" in the literature.
2. http://ritter.tea.state.tx.us/perfreport/snapshot/index.html.
3. Using SCM with a large sample of units dissimilar to the treated unit increases the risk of interpolation bias and overfitting. Overfitting occurs when the treated unit's characteristics are artificially matched to dissimilar units by combining idiosyncratic variations in a large sample of untreated units (Abadie, Diamond, and Hainmueller 2014).
4. The root mean square prediction error (RMSPE) is the square root of the average of the squared discrepancies in values of the outcome variable between the synthetic control and treated unit.
5. http://web.stanford.edu/~jhain/synthpage.html.

REFERENCES

Abadie, A., A. Diamond, and J. Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105(490): 493–505.

Abadie, A., A. Diamond, and J. Hainmueller. 2011. "Synth: An R Package for Synthetic Control Methods in Comparative Case Studies." *Journal of Statistical Software* 42(13): 1–17.

Abadie, A., A. Diamond, and J. Hainmueller. 2014. "Comparative Politics and the Synthetic Control Method." *American Journal of Political Science* 59(2): 495–510.

Abadie, A. and J. Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93(1): 113–32.

Agranoff, R. and M. McGuire. 2001. "Big Questions in Public Network Management Research." *Journal of Public Administration Research and Theory* 11(3): 295–326.

Amrein, A. L. and D. C. Berliner. 2002. "High-Stakes Testing & Student Learning." *Education Policy Analysis Archives* 10(8): 1–74.

Andrews, M., W. Duncombe, and J. Yinger. 2002. "Revisiting Economies of Size in American Education: Are We any Closer to a Consensus?" *Economics of Education Review* 21(3): 245–62.

Angrist, J. D. and J. Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Archer, J. 2005. "Weighty Decisions." *Education Week* 24(17): 20–21.

Baker, B. D. 2009. "Within-District Resource Allocation and the Marginal Costs of Providing Equal Educational Opportunity: Evidence from Texas and Ohio." *Education Policy Analysis Archives* 17(3): 1–31.

Baker, B. D. and D. R. Elmer. 2009. "The Politics of Off-the-Shelf School Finance Reform." *Educational Policy* 23(1): 66–105.

Behn, R. D. 2003. "Why Measure Performance? Different Purposes Require Different Measures." *Public Administration Review* 63(5): 586–606.

Billmeier, A. and T. Nannicini. 2013. "Assessing Economic Liberalization Episodes: A Synthetic Control Approach." *Review of Economics and Statistics* 95(3): 983–1001.

Blom-Hansen, J., R. Morton, and S. Serritzlew. 2015. "Experiments in Public Management Research." *International Public Management Journal* 18(2): 151–170.

Bohte, J. 2001. "School Bureaucracy and Student Performance at the Local Level." *Public Administration Review* 61(1): 92–99.

Bohte, J. and K. J. Meier. 2000. "Goal Displacement: Assessing the Motivation for Organizational Cheating." *Public Administration Review* 60(2): 173–82.

Bozeman, B. 1992. "Experimental Design in Public Policy and Management Research: Introduction." *Journal of Public Administration Research and Theory* 2(4): 440–42.

Brudney, J. L., F. T. Hebert, and D. S. Wright. 1999. "Reinventing Government in the American States: Measuring and Explaining Administrative Reform." *Public Administration Review* 59(1): 19–30.

Cavallo, E., S. Galiani, I. Noy, and J. Pantano. 2013. "Catastrophic Natural Disasters and Economic Growth." *Review of Economics and Statistics* 95(5): 1549–61.

Chambers, J. G., J. D. Levin, and L. Shambaugh. 2010. "Exploring Weighted Student Formulas as a Policy for Improving Equity for Distributing Resources to Schools: A Case Study of Two California School Districts." *Economics of Education Review* 29(2): 283–300.

Coleman, J. S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. Mood, F. Weinfeld, and R. York. 1966. *Equality of Educational Opportunity*. Washington, DC: US Government Printing Office.

Courty, P. and G. Marschke. 2004. "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." *Journal of Labor Economics* 22(1): 23–56.

Curtis, K., K. Sinclair, and B. Malen. 2014. "Student-Based Budgeting (SBB) as an Education Reform Strategy: Theory and Evidence." 39th Annual AEFP Conference. San Antonio, Texas, March 13–15.

Driscoll, D., D. Halcoussis, and S. Svorny. 2003. "School District Size and Student Performance." *Economics of Education Review* 22(2): 193–201.

Duncombe, W., J. Miner, and J. Ruggiero. 1995. "Potential Cost Savings from School District Consolidation: A Case Study of New York." *Economics of Education Review* 14(3): 265–84.

Duncombe, W., J. Ruggiero, and J. Yinger. 1996. "Alternative Approaches to Measuring the Cost of Education." Pp. 237–356 in *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, DC: The Brookings Institution.

Figlio, D. and S. Loeb. 2011. "Chapter 8: School Accountability." Pp. 383–421 in *Handbook of the Economics of Education*. Amsterdam: North-Holland.

Fitzpatrick, J., M. Goggin, T. Heikkila, D. Klingner, J. Machado, and C. Martell. 2011. "A New Look at Comparative Public Administration: Trends in Research and an Agenda for the Future." *Public Administration Review* 71(6): 821–30.

Gill, J. and K. J. Meier. 2000. "Public Administration Research and Practice: A Methodological Manifesto." *Journal of Public Administration Research and Theory* 10(1): 157–99.

Gill, J. and C. Witko. 2013. "Bayesian Analytical Methods: A Methodological Prescription for Public Administration." *Journal of Public Administration Research and Theory* 23(2): 457–94.

Goertz, M. E., M. C. Duffy, and K. C. Le Floch. 2001. *Assessment and Accountability Systems in the 50 States*. RR-046. Philadelphia, PA: University of Pennsylvania.

Grizzle, G. A. 2002. "Performance Measurement and Dysfunction: The Dark Side of Quantifying Work." *Public Performance & Management Review* 25(4): 363.

Groeneveld, S., L. Tummers, B. Bronkhorst, T. Ashikali, and S. V. Thiel. 2014. "Quantitative Methods in Public Administration: Their Use and Development through Time." *International Public Management Journal* 18(1): 61–86.

Hanushek, E. A. and M. E. Raymond. 2005. "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management* 24(2): 297–327.

Hanushek, E. A., S. G. Rivkin, and L. L. Taylor. 1996. "Aggregation and the Estimated Effects of School Resources." *The Review of Economics and Statistics* 78(4): 611.

Heinrich, C. 2012. "Measuring Public Sector Performance and Effectiveness." Pp. 32–49 in *The SAGE Handbook of Public Administration*. London: Sage Publications.

Hinrichs, P. 2012. "The Effects of Affirmative Action Bans on College Enrollment, Educational Attainment, and the Demographic Composition of Universities." *Review of Economics and Statistics* 94(3): 712–22.

Hood, C. 1991. "A Public Management for All Seasons?" *Public Administration* 69(1): 3–19.

Hood, C. and G. Peters. 2004. "The Middle Aging of New Public Management: Into the Age of Paradox?" *Journal of Public Administration Research and Theory* 14(3): 267–82.

Houston Independent School District. 2011. "HISD Budget Basics." http://www2.houstonisd.org/HISDConnectEnglish/Images/PDF/HISDBudgetBasics_English.pdf

Isett, K. R., I. A. Mergel, K. LeRoux, P. A. Mischen, and R. K. Rethemeyer. 2011. "Networks in Public Administration Scholarship: Understanding Where We Are and Where We Need to Go." *Journal of Public Administration Research and Theory* 21 (Supplement 1): i157–73.

Izumi, L. T. and W. M. Evers. 2002. "State Accountability Systems." Pp. 105–152 in H. J. Walberg and W. M. Evers, eds., *School Accountability: An Assessment by the Koret Task Force on K-12 Education*. Stanford, CA: Hoover Institution Press.

Jacob, B. A. 2005. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89(5–6): 761–96.

Jacob, B. A. and S. D. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *The Quarterly Journal of Economics* 118(3): 843–77.

Joyce, P. G. 1993. "Using Performance Measures for Federal Budgeting: Proposals and Prospects." *Public Budgeting and Finance* 13(4): 3–17.

Joyce, P. G. and S. S. Tompkins. 2002. "Using Performance Information for Budgeting." Pp. 61–96 in K. Newcomer, C. B. Jennings and A. Lomax, eds., *Meeting the Challenges of Performance-Oriented Government*. Washington, DC: American Society for Public Administration.

Kettl, D. F. 1997. "The Global Revolution in Public Management: Driving Themes, Missing Links." *Journal of Policy Analysis and Management* 16(3): 446–62.

Konisky, D. M. and C. Reenock. 2013. "Case Selection in Public Management Research: Problems and Solutions." *Journal of Public Administration Research and Theory* 23(2): 361–93.

Margetts, H. Z. 2011. "Experiments for Public Management Research." *Public Management Review* 13(2): 189–208.

Meier, K. J. and L. R. Keiser. 1996. "Public Administration as a Science of the Artificial: A Methodology for Prescription." *Public Administration Review* 56(5): 459.

Meier, K. J. and L. J. O'Toole. 2002. "Public Management and Organizational Performance: The Effect of Managerial Quality." *Journal of Policy Analysis and Management* 21(4): 629–43.

Melkers, J. and K. Willoughby. 1998. "The State of the States: Performance-Based Budgeting Requirements in 47 Out of 50." *Public Administration Review* 58(1): 66–73.

Melkers, J. and K. Willoughby. 2005. "Models of Performance-Measurement Use in Local Governments: Understanding Budgeting, Communication, and Lasting Effects." *Public Administration Review* 65(2): 180–90.

Miles, K. H. and M. Roza. 2006. "Understanding Student-Weighted Allocation as a Means to Greater School Resource Equity." *Peabody Journal of Education* 81(3): 39–62.

Moynihan, D. P. 2005. "Why and How Do State Governments Adopt and Implement 'Managing for Results' Reforms?" *Journal of Public Administration Research and Theory* 15(2): 219–43.

Moynihan, D. P. 2006. "Managing for Results in State Government: Evaluating a Decade of Reform." *Public Administration Review* 66(1): 77–89.

Moynihan, D. P., S. Fernandez, S. Kim, K. M. LeRoux, S. J. Piotrowski, B. E. Wright, and K. Yang. 2011. "Performance Regimes Amidst Governance Complexity." *Journal of Public Administration Research and Theory* 21 (Supplement 1): i141–55.

Moynihan, D. P. and S. K. Pandey. 2006. "Creating Desirable Organizational Characteristics." *Public Management Review* 8(1): 119–40.

Nielsen, P. A. 2013. "Performance Management, Managerial Authority, and Public Service Performance." *Journal of Public Administration Research and Theory* 24(2): 431–58.

Ouchi, W. G. 2006. "Power to the Principals: Decentralization in Three Large School Districts." *Organization Science* 17(2): 298–307.

Perry, J. L. 2012. "How Can We Improve Our Science to Generate More Usable Knowledge for Public Professionals?" *Public Administration Review* 72(4): 479–82.

Peters, E. 2009. "ACT Scores Drop: *But Marks Still Well Above State Average." *Abilene Reporter-News* (TX), August 20, 4.

Piotrowski, S. J. and D. H. Rosenbloom. 2002. "Nonmission-Based Values in Results-Oriented Public Management: The Case of Freedom of Information." *Public Administration Review* 62(6): 643–57.
Pitts, D. W. and S. Fernandez. 2009. "The State of Public Management Research: An Analysis of Scope and Methodology." *International Public Management Journal* 12(4): 399–420.
Provan, K. G. and P. Kenis. 2007. "Modes of Network Governance: Structure, Management, and Effectiveness." *Journal of Public Administration Research and Theory* 18(2): 229–52.
Provan, K. G. and H. B. Milward. 2001. "Do Networks Really Work? A Framework for Evaluating Public-Sector Organizational Networks." *Public Administration Review* 61(4): 414–23.
Raadschelders, J. C. N. and K. Lee. 2011. "Trends in the Study of Public Administration: Empirical and Qualitative Observations from Public Administration Review, 2000–2009." *Public Administration Review* 71(1): 19–33.
Radin, B. 2006. *Challenging the Performance Movement: Accountability, Complexity, and Democratic Values*. Washington, DC: Georgetown University Press.
Schick, A. 2003. "Opportunity, Strategy and Tactics in Reforming Public Management." *OECD Journal on Budgeting* 2(3): 7–34.
Snell, L. 2009. *Weighted Student Formula Yearbook*. Los Angeles, CA: Reason Foundation.
Stiefel, L., R. Rubenstein, and R. Berne. 1998. "Intra-District Equity in Four Large Cities: Data, Methods and Results." *Journal of Education Finance* 23(4): 447–467.
Svensson, J., W. Trommel, and T. Lantink. 2008. "Reemployment Services in the Netherlands: A Comparative Study of Bureaucratic, Market, and Network Forms of Organization." *Public Administration Review* 68(3): 505–15.
Ucelli, M., E. Foley, T. Emdon, and C. Bond. 2002. *First Steps to a Level Playing Field: An Introduction to Student-Based Budgeting*. Providence, RI: Annenberg Institute for School Reform at Brown University.

ABOUT THE AUTHOR

Chris Birdsall (chris.birdsall@american.edu) is a PhD student in the Department of Public Administration and Policy at American University. His research focuses on public management and higher education.