

ESSAYS IN CAUSAL INFERENCE AND SYNTHETIC CONTROL

by

SIMEON A. MINARD

A DISSERTATION

Presented to the Department of Economics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2019

DISSERTATION APPROVAL PAGE

Student: Simeon A. Minard

Title: Essays in Causal Inference and Synthetic Control

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

Glen R. Waddell	Chair
Jeremy Piger	Core Member
Ben Hansen	Core Member
David Wagner	Institutional Representative

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School
-----------------------	--

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2019

© 2019 Simeon A. Minard

DISSERTATION ABSTRACT

Simeon A. Minard

Doctor of Philosophy

Department of Economics

June 2019

Title: Essays in Causal Inference and Synthetic Control

This dissertation includes previously unpublished co-authored material. The first chapter of this dissertation outlines a new method of estimating the synthetic control technique that has a number of desirable properties. The second chapter causally infers a positive property crime impact of supply side drug intervention, and important policy result. The third and final chapter outlines an method of variable selection in linear regression to be used to decrease bias and increase precision.

CURRICULUM VITAE

NAME OF AUTHOR: Simeon A. Minard

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR

Western Washington University, Bellingham, WA

DEGREES AWARDED:

Doctor of Philosophy, Economics, 2019, University of Oregon

Master of Science, Economics, 2015, University of Oregon

Bachelor of Arts, Economics, 2013, Western Washington University

AREAS OF SPECIAL INTEREST:

Applied Econometrics

Applied Microeconomics

Labor

GRANTS, AWARDS AND HONORS:

Department of Economics Graduate Teaching Award, University of Oregon,
2017

Kleinsorge Fellowship Award, University of Oregon, 2014

ACKNOWLEDGEMENTS

I would like to thank my family and friends, and all members of my committee, especially Dave.

To my family, and Dave.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
II. DISPERSION WEIGHTED SYNTHETIC CONTROLS	3
Introduction	3
A new synthetic control procedure	8
Empirics	14
Conclusion	41
III. THE UNINTENDED CONSEQUENCES OF SUPPLY SIDE DRUG INTERVENTION: EVIDENCE FROM DEA CHEMICAL CLASSIFICATION	43
Introduction	43
Background	45
Data	47
Identification	50
Results	57
Further Analysis	61
Conclusion	68

Chapter	Page
IV. BIAS REDUCTION THROUGH VARIABLE SELECTION	70
Introduction	70
Model selection	72
Simulations	73
Conclusion	86
V. CONCLUSION	87
REFERENCES CITED	88

LIST OF FIGURES

Figure	Page
1. Examples of “overall” and “relative” dispersion	13
2. Contamination: all available donors	18
3. Contamination: Treatment and MSE	19
4. Contamination: kernel densities	20
5. Unobserved trends: all available donors	23
6. Unobserved trends: treatment and MSE	24
7. Unobserved trends: kernel densities	25
8. Unobserved error: all available donors	27
9. Unobserved error: precision properties	28
10. DWSC with a “Bad controls” problem	30
11. Per-capita cigarette sales (packs) in California, 1970 to 2000	32
12. California smoking: donor inclusion	34
13. California: DWSC robustness	35
14. Reported rape offences in Rhode Island, 1970 to 2009	38
15. Rhode Island: DWSC robustness	39
16. Rhode Island: donor inclusion	40
17. Frequency histogram: cocaine admission proportion	51
18. Frequency histogram: total admissions and cocaine admissions	52
19. Geographic representation of treatment	54
20. Time series representation of treatment	56
21. Event study	58
22. Pre-post crime against admissions	64

Figure	Page
23. Removal of low addiction states	65
24. Treatment estimates with more pre-treatment years	66
25. Distributional co-variate balance tests: Baseline data-generating process	76
26. Bias entering through an omitted interaction (of x_2 and x_3), which correlates differently among treated units	78
27. Bias entering through an omitted non-linearity (x_1^2) that varies differently among treated units	80
28. Bias entering through multiple omitted interactions (x_1x_2 and x_4^2) that vary differently among treated units	82
29. Variable selection can reduce bias entering through unobservables: Positive bias	83
30. Variable selection can reduce bias entering through unobservables: Negative bais	84
31. Where there is no bias, variable selection increases precision	85

LIST OF TABLES

Table	Page
1. Cocaine Usage in 2006 (Percentage of Respondents NSDUH 2006) . . .	48
2. Pre-Treatment Summary Statistics	49
3. Property Crime Rate (2005 Cocaine Admissions per 100,000 population)	53
4. Property Crime Rate (2005 Cocaine Admission Proportion)	59
5. Property Crime Rate (2005 Cocaine Admission Proportion) (Population Weighted)	60
6. Property Crime Rate (2005 Cocaine Admission Proportion)	62
7. Violent Crime Rates (2005 Cocaine Admission Proportion)	63
8. Property Crime Rate (2005 Alcohol Admission Proportion)	67

CHAPTER I

INTRODUCTION

In Dispersion Weighted Synthetic Controls (co-authored with Glen Waddell) we propose a new approach to synthetic-control methods, through which we regularize the consideration of variation in available control units and the stability properties of the synthetic control. Specifically, we introduce two penalties directly into the objective function, allowing for the endogenous down-weighting of donors to the synthetic control with outcomes that exhibit different patterns of variation before and after treatment, and donors tending to be distant from the synthetic controls average each period. While nesting a typical approach, we offer an intuitively appealing method for applied researchers to evaluate the reasonableness of a variety of synthetic controls and consider the sensitivity of results.

The Unintended Consequences of Supply-Side Drug Intervention: Evidence from DEA Chemical Classification (JMP) focuses on the crime impacts of supply-side drug policy. Supply-side drug intervention plays a central role in both local and federal anti-drug policy, yet disruption of drug markets may lead to unintended changes in the behavior of individuals who participate these markets. Specifically, addicted users facing increased prices and reduced availability may turn to financially motivated crime in order to continue their consumption patterns. I find evidence that states with high levels of cocaine addiction experienced significant increases in property crime relative to states with lower levels of addiction after a nationwide supply shock in the cocaine market that reduced availability. This effect should be accounted for in both formulation and execution of supply side

interventions as well as in the cost-benefit assessment of supply side drug policy as a whole.

In *Bias Reduction through Variable Selection* (co-authored with Glen Waddell) we focus on variable selection techniques to remove bias from regression parameters. In regressions that seek to determine a causal impact, the treatment of potentially confounding variables is very important for assuring unbiased estimates, and is often approached through the so called balance test. We show that while observable covariates may appear balanced in terms of means, there are other dimensions where imbalance can lead to bias specifically higher-order polynomials and interactions of these variables. As the number of potential controls may be large (and may even exceed the number of observations) we advocate the use of model-selection procedures Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC). We motivate this process through a number of simulated environments, showing decreases in bias and improvements in precision in cases of both omitted observable interactions and omitted unobservable variables

CHAPTER II

DISPERSION WEIGHTED SYNTHETIC CONTROLS

I would like to acknowledge Glen Waddell, who contributed to the early stages of developing the initial concept, and provided guidance through the development of the estimator. He also contributed to the writing and presentation of the paper. I formulated the estimator mathematically, developed it computationally, performed all computational simulations and created all figures, and contributed to the writing and presentation of the paper.

Introduction

Athey and Imbens (2017) call synthetic control methods (SCMs), “the most important innovation in the evaluation literature in the last 15 years.” Generally, SCMs refers to the construction of a weighted average of untreated units that projects the path that a treated unit would have followed in the absence of treatment—it is quickly becoming a preferred approach to policy evaluation in the absence of appropriate individual control units. However, as approaches to SCMs have yet to be standardized, the scope for researchers to make specification choices can undermine SCM results.¹

In traditional approaches to estimating synthetic controls, it is assumed that the relationship being identified is stable. That is, one assumes that the pre-treatment fit with the treated unit is sufficient to imply that the post-treatment

¹ Ferman et al. (2018) (i.e., “the cherry picking paper”) will surely stand as one of the formative papers in the SCM literature, and points directly to this in suggesting that “with no clear guidance on the choice of predictor variables used to estimate the [synthetic control] weights, there are opportunities for the researcher to search for specifications with statistically significant results, undermining one of the main advantages of the method.”

levels of the synthetic control approximates the treated entity’s counterfactual levels in the treatment period. While one cannot benchmark a synthetic control to the post-treatment behavior of the treated unit absent treatment, given the fundamental econometric problem, there is interpretable information in the behavior of untreated units in these periods.

While nesting a typical approach to SCMs, we offer an intuitively appealing flexibility for applied researchers to evaluate the reasonableness and stability properties of synthetic controls in their chosen environment, and consider the sensitivity of estimated treatment effects amid the choices inherent to SCMs. Specifically, we propose a systematic regularization across two dimensions, motivated by properties that we anticipate being particularly desirable in a control group—one’s tolerance for untreated units contributing to the synthetic control to be distant from the synthetic control’s average each period (i.e., dispersion), and one’s tolerance for the outcomes of untreated units to exhibit different patterns of variation before and after treatment (i.e., relative dispersion).²

In Figure 1, for example, we plot a 2×2 matrix of four abstract notions of high and low overall dispersion, and high and low relative dispersion. To our eye, a set of controls with low dispersion in both dimensions might be appealing—the bottom-right cell of the figure. The point being, as one approaches a set of untreated units that looks other than that depicted in that preferred cell, one can endogenize the down-weighting of donors to the synthetic control who are generally distant (moving top to bottom in the figure), or the down-weighting of donors who

² While estimating standard errors in the context of SCMs is itself a developing literature, we imagine that increases in the precision of an estimated synthetic control imply increases, ultimately, in the precision of estimated treatment effects.

are systematically more- or less-distant from the synthetic control after treatment than they were before (moving left to right in the figure).

To be clear, we do not think of this as a fix, necessarily, but as a procedure that reveals the sensitivity of SCM-derived estimates of treatment to the consideration of synthetic-control stability. In the end, we recommend that researchers plot estimates of the effect of treatment across the two parameters we introduce (i.e., controlling “dispersion” and “relative dispersion”)—this also implies that one considers a very large number of synthetic controls. Conveniently, over the range of parameters, our procedure is entirely driven by pre-treatment MSE at one end and converges toward identifying the best-available individual control at the other.³ Further, we see a comparison of the treated unit to this “best” or “closest” control as directly informative—we might consider asking of SCM papers, more generally, whether adopting SCMs is responsible for *attenuating* or *amplifying* the estimated treatment effect relative to such a benchmark. For example, in the California experiment of Abadie et al. (2010), comparing California cigarette sales to the individual control with lowest pre-treatment MSE (i.e., Montana) yields an estimated treatment of 25.4 fewer packs per capita, annually, with the 1988 anti-smoking campaign—this is the *largest* from among the point estimates retrieved over the range of our two parameters. In the Rhode Island experiment of Cunningham and Shah (2018), it’s quite the opposite, where comparing Rhode Island rape reports to the individual control with lowest pre-treatment MSE (i.e., New Hampshire) yields a relative decline in the treated state of only 1.6 rapes per 100,000, which is the smallest estimate we retrieve of the effect of the 2010

³ Among the early innovators of SCMs, Abadie et al. (2010) retrieves an estimate of the causal effect of a 1988 anti-smoking campaign in California on per-capita smoking sales. We return to consider this result as part of our empirical applications. (See Abadie and Gardeazabal (2003), and Abadie et al. (2015) for other formative SCM analyses.)

decriminalization of prostitution. Any additions to “Synthetic Rhode Island” *increase* the magnitude of what one would find in this best-available case. This is a distinction worth noting, we believe (as will be the slope of the estimated treatment effect in parameters, which we discuss below with the figures we propose).

With the post-treatment variation of individual donors directly contributing to the estimated treatment effect in proportion to their weight, it seems uncontroversial that we take exceptional care over post-treatment behavior of individual contributors as we determine those weights. However, given the near-absolute importance that typical SCMs place on the *pre*-treatment fit of the synthetic to the treated unit, it is noteworthy that we allow for the *post*-treatment variation in potential donors directly in the determination of synthetic-control weights. Yet, in our procedure, pre-treatment MSE plays no-less important a role, as we propose that the presentation of treatment effects always be accompanied by the presentation of pre-treatment MSE, where we can evaluate their co-movements as we vary the *ex ante* importance of the synthetic control’s stability in the objective function. In this way, we more-fully explore the variation in available control units in the building of appropriate efficacy tests, while maintaining the ability to speak back into more-standard approaches.

Our procedure also moves somewhat toward transparency. For example, SCMs can often lead to multiple synthetic controls having near-equivalent measures of pre-treatment fit while at the same time producing estimated treatment effects that differ wildly. This is not uncommon, in our experience, and a source of sensitivity that has lead to doubt in our own attempts to evaluate SCM results. In our procedure, we will argue that near-equivalence in estimated treatment and in pre-treatment MSE (despite putting increasing wright on stability properties)

should breed confidence, and is something that should be demonstrated in SCM analyses. Likewise, to the extent a particular application produces bounds across a parameter space, we have learned something. Or, to the extent a single point estimate is still desired from a given analysis, we find it an appealing feature of our procedure that we can potentially foreclose on some synthetic controls over others on principled grounds, while adding transparency to the context from which that inference is made.

As we approach our procedure, we acknowledge the broader literature and direct readers to Doudchenko and Imbens (2016) for foundational context regarding the relations between synthetic controls, difference-in-differences, and matching methods. In what follows, we first walk through the specifics of what we refer to as a dispersion-weighted synthetic control (DWSC), which in large part comes from our own attempts to build confidence in our own policy evaluations with synthetic-control methods. We then, in Section 2.3, report the results of our procedure applied in two policy-relevant empirical settings. First, we apply the procedure to simulated environments—in one, we consider a “contaminated-control” problem, in another, we consider scenarios where type is unobservable, and untreated units of different “type” are similar enough in pre-treatment to be given weight in the synthetic control by traditional approaches, and in another, we consider SCM with only “bad controls” available that each violate “common trends.” Second, we apply DWSC to one of the canonical SCM settings: the 1988 anti-smoking campaign in California evaluated in Abadie et al. (2010). Third, we consider a more-recent SCM result: the change in reported-rape offences in Rhode Island around their legalization of prostitution, evaluated Cunningham and Shah (2018). We offer some concluding remarks in Section V.

A new synthetic control procedure

In seeking out both transparency and an informative regularization of a synthetic control procedure, we allow for two penalties in the objective function that yields the synthetic control. First, we allow for a penalty on individual donors who exhibit changes (before and after treatment) in how they fit within a given synthetic control. That is, we build into our methodology the endogenous down-weighting of individual donors to the extent they behave *differently* in the post-treatment period than they had in the pre-treatment period. This has strong intuition, we believe, insofar as one is more confident that one has established a “reasonable control” when the estimator produces a set of donors who at least vary similarly with each other before *and after* treatment. Second, we allow for a penalty on individual donors according to their overall mean-squared distance from the synthetic control. That is, in the determination of donor weights, we allow for the endogenous down-weighting of donors who are relative outliers within the synthetic control, allowing researchers to consider robustness to “tighter” synthetic controls.⁴

The objective function

Consider the typical context in which the employ of SCMs for identification seems advantageous—observations of a treated unit’s outcomes, Y_{it} , and a set of K untreated entities over the same time interval $t \in [1, T]$, with treatment occurring after some $T_0 > 1$ (i.e., T_0 is the last pre-treatment observation). As per usual,

⁴ To one’s concern that we may introduce potentially undesirable consequences as we consider distance to the synthetic control, note that we will still rely on pre-treatment MSE capturing in the objective function the difference to the treated unit itself. As comparison to the treated unit in the post-treatment period is untenable, neither of the two channels through which we will allow for the down-weighting of donors can use that as a benchmark for comparison in the post period. It’s this that drives us to using the synthetic control itself.

the treated unit is not observed in the treatment period absent treatment, and SCMs handle this fundamental econometric problem with a convex combination of control entities generating a single “synthetic control” that by assumption then forecasts the counterfactual series that the treated unit would have followed in the absence of treatment. Any deviations from this counterfactual are then attributed to treatment. As we’ve already noted, weighted post-treatment movements in donor entities are hardwired into the estimated treatment effect, which will motivate that we consider post-treatment movements in the synthetic control more directly.

As a measure of pre-treatment fit, we follow the convention of adopting pre-treatment mean squared error (MSE). Specifically, with vector $w = [w_1, w_2, \dots, w_k]$ collecting the set of weights to be determined across K untreated entities, and Y_{0t} notating the outcome of the treated unit itself, we define pre-treatment MSE as

$$\mathcal{M}(w) = \frac{\sum_{t=1}^{T_0} \left(Y_{0t} - \sum_{i=1}^K w_i Y_{it} \right)^2}{T_0}. \quad (2.1)$$

The solution to minimization $\mathcal{M}(w)$ we can define as some set of weights \bar{w} .⁵ We augment this simple minimization of pre-treatment MSE with the addition of two penalties.

First, in the objective function we allow for the endogenous down-weighting of donors when their post-treatment behavior differs from their pre-treatment

⁵ Implied in our procedure will be that the unobservable components of outcomes are of first-order importance, more so than particular covariate movements. Though important to distinguish, this is second-order to our contribution, noting that there are examples of synthetic-control estimation that do not rely on covariate inclusion (e.g., see Doudchenko and Imbens (2016)), (Ferman et al., 2018) offers a discussion of the role of covariates and, ultimately, arguments supporting the exclusion of covariates in SCMs, and Ferman and Botosaru (2017) demonstrates the unbiasedness of synthetic controls matching on pre-treatment outcomes. (Moreover, note that the inclusion of covariates requires that some pre-treatment periods be excluded (Kaul et al., 2018), which introduces significant scope to the econometrician’s choice of which outcomes to include.)

behavior. The intuition is straightforward, we believe—it is reasonable to question why some donors to the synthetic control may be behaving differently after treatment, and to consider the sensitivity of a single point estimate to the down-weighting of deviants. In this way, we think of this as not a prescription for a fix, per se, but as a procedure that reveals the sensitivity of SC-derived point estimates to the consideration of synthetic-control stability. In this dimension, we enable the down-weighting of donors who are more (or less) distant from the synthetic control after treatment than they were before treatment. Specifically, we can define the relative dispersion of each donor, $\mathcal{R}_j(w)$, as

$$\mathcal{R}_j(w) = \left[\frac{\sum_{t=1}^{T_0} \left(Y_{jt} - \sum_{i=1}^K w_i Y_{it} \right)^2 / T_0 - \sum_{t=T_0+1}^T \left(Y_{jt} - \sum_{i=1}^K w_i Y_{it} \right)^2 / (T - T_0)}{\sum_{t=1}^{T_0} \left(Y_{jt} - \sum_{i=1}^K w_i Y_{it} \right)^2 / T_0} \right]^2, \quad (2.2)$$

and, summing across donors, define a measure of relative dispersion, $\mathcal{R}(w)$, as

$$\mathcal{R}(w) = \sum_{j=1}^K w_j \mathcal{R}_j(w), \quad (2.3)$$

which assures that only weighted donors contribute to the objective function. We parameterize this potential penalty in the objective function with $\rho \in [0, 1)$.

Second, we allow for the endogenous down-weighting of donors contributing excessively to the overall dispersion of the synthetic control. That is, defining the overall variation within the synthetic control as

$$\mathcal{D}(w) = \frac{\sum_{t=1}^T \sum_{j=1}^K w_j \left(Y_{jt} - \sum_{i=1}^K w_i Y_{it} \right)^2}{T}, \quad (2.4)$$

we allow for the down-weighting of donors k who contribute excessively to this variation, generally, which produces a tighter-fitting synthetic control (with no particular notion of pre/post balance). We parameterize the relative importance of this more-general dispersion in the synthetic control with $\delta \in [0, 1]$.

Together, then, (for given ρ and δ) we can cast the estimation procedure with the objective function,

$$\arg \min_w \underbrace{(1 - \rho - \delta) \mathcal{M}(w)}_{\text{Pre-treatment MSE}} + \underbrace{\rho \mathcal{R}(w) \frac{\mathcal{M}(\bar{w})}{\mathcal{R}(\bar{w})}}_{\text{Pre/post relative ("rho") dispersion}} + \underbrace{\delta \mathcal{D}(w) \frac{\mathcal{M}(\bar{w})}{\mathcal{D}(\bar{w})}}_{\text{Overall ("delta") dispersion}}$$

$$s.t. \quad \rho \geq 0, \delta \geq 0, \rho + \delta < 1, \sum_{i=1}^K w_i = 1, w_i \geq 0 \forall i.$$
(2.5)

In (2.5), weights w are chosen by standard numerical optimization techniques to minimize a convex combination of pre-treatment mean squared error and our measures of dispersion, \mathcal{R} and \mathcal{D} . In order to make the size of the penalties and MSE comparable, we scale both penalties by the ratio of pre-treatment MSE to the penalty, evaluated at the solution to minimizing the pre-treatment MSE in (2.1), which we notate as \bar{w} .⁶

Some things of note

We have five quick notes to make, before proceeding to consider applications.

First, we note that the procedure we propose is not unlike that of Doudchenko and Imbens (2016), who introduce penalties on the L1 (sum of absolute value of weights) and L2 (sum-of-squared weights) norms of the weights.

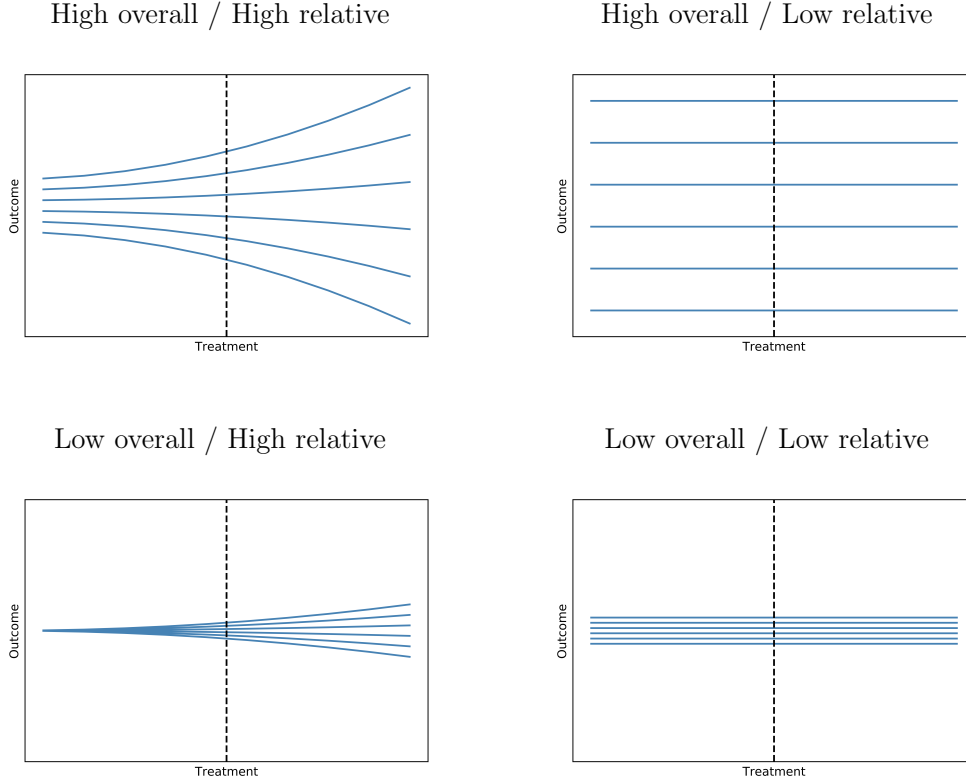
⁶ Recall that the set of weights \bar{w} minimizes $\mathcal{M}(w)$ is the same as that which minimizes 2.5 given $\rho = \delta = 0$.

In our procedure, while also increasing transparency and demonstrating robustness across a range of estimates, we retain the weight restrictions of the original Abadie et al. (2010) but allow the researcher to weigh more heavily the importance of stability among donors to the the synthetic control. In general, we share the prior that relaxing these restrictions can lead to substantial improvements in the estimator. However, penalties such as those we introduce are difficult to conceptualize without those restrictions. Negative weights, for example, create problems given our notions of what properties we value in a synthetic control and we feel comfortable penalizing (similarity in levels, and similarity pre/post similarity). Put in other words, we are willingly sacrificing the benefits to relaxing weight restrictions in exchange for desirable and informative properties in the procedure.

Second, that we have now introduced ρ and δ as parameters, Figure 1 can be recast with the added intuition that, as one approaches data with properties other than that depicted in that preferred cell, it is with increases in δ that one achieves lower overall dispersion of the data contributing to the synthetic (moving down) and with increases in ρ that one achieves lower relative dispersion of the data contributing to the synthetic (moving to the right).

Third, note that benchmarking the penalties to the synthetic control itself will protect against the down-weighting of donors to the synthetic control that truly do represent the counterfactual. Our procedure allows for the down-weighting of untreated units to the extent they are outliers in the set of all potential donors. To the extent their movement is common across potential donors, the penalty does not so quickly bind and they are not down-weighted. We address related concerns in one of our simulated environments below, assigning the treated unit to one of

FIGURE 1.
Examples of “overall” and “relative” dispersion



two unobservable types—the procedure we introduce increasingly weights those of the same type as the treated unit, while the traditional method does not.

Fourth, we simply wish to formalize parameter conditions. Recall that choosing $\rho = 0$ and $\delta = 0$ nests an estimator that matches only on pre-treatment outcomes, and in that way captures the typical synthetic control design.⁷ Note, however, that there are K potential solutions to (2.5) when either $\rho = 1$ or $\delta = 1$. That is, at either limit, any degenerate set of weights (i.e., all weight on a single donor) yields an objective function that evaluates to zero and the estimator cannot

⁷ The `synth` package, for example, will produce estimates that minimize (2.5) subject to $\rho = \delta = 0$, when all pre-treatment periods are included. (`synth` is currently available in Matlab, R, and Stata, through Jens Hainmueller.)

distinguish any of the K donors from one another. However, as one approaches $\rho = 1$ or $\delta = 1$ and the residual weight of $1 - \rho - \delta$ is still on pre-treatment MSE, the estimator will collapse on the single donor that best-matches the *pre-treatment* outcomes of the treated unit itself, and thereby avoid choosing donors with attractive dispersion properties but poor fit with the treated unit. Formally, then, we define the parameter space as $\rho + \delta \in [0, 1)$, recalling that the value in our procedure is, again, in the variation in the estimated treatment effect across the *range* of these parameters.⁸ By extension, across this parameter space, pre-treatment MSE cannot be larger than the pre-treatment MSE of the single best-fitting control.

Finally, synthetic control method is often referred to as a generalization of the difference-in-differences (DD) method. However, this is only true if data are de-meaned before estimation, accounting for the “fixed effect” implied in difference-in-differences estimators. We follow the general convention in synthetic controls of not demeaning our data, with recognition that de-meaning may well become commonplace (Ferman and Pinto, 2016). Importantly, we expect that in a context where data are de-meaned, the importance of ρ would rise, while the importance of δ would fall. Further, the interpretation of “best-available” individual control would move toward aligning with what one might think of as the best available for a DD design—similar in co-movements, but not necessarily in levels.

Empirics

In this section, we produce three sets of results. First, we consider the estimator’s behavior in three simulated environments—an environment where a

⁸ We have no strong prior on the relative importance one might imagine for ρ and δ .

subset of controls is contaminated by treatment, an environment where the donors become increasingly disperse, and an environment where there are donors are unlikely to have common trends with the treated unit. The bias in the traditional synthetic control is revealed in a lack of robustness across the allowable ranges of ρ and δ .

Second, we reconsider the Abadie et al. (2010) (ADH) analysis of California’s 1988 anti-smoking campaign, retrieving estimates of the causal effect of the policy intervention on per-capita smoking sales that substantiate the original result, and with some added confidence (we argue). In short, we find even the point estimate to be insensitive to our procedure—as we increase the importance of both ρ or δ type dispersions in the determination of the synthetic control, there is very little movement in the estimated treatment effect and, if anything, the point estimate *increases* in magnitude on approach to the best-available individual control.

Third, we reconsider one of the results in Cunningham and Shah (2018), a well-known and recent SCM paper in which the 2003 decriminalization of prostitution lead to a decrease in reported rape offences. Considering ρ and δ type dispersions in this example reveals a sensitivity that traditional SCMs and point estimates that seemingly *decrease* in magnitude on approach to the best-available individual control.

In all cases, we propose that in the plotting of the estimated treatment effect and the MSE across permutations of ρ and δ , researchers can evaluate the reasonableness and sensitivity of a synthetic control in their chosen environment.

Simulated environments

It is easy to introduce bias into SCMs that only consider pre-treatment MSE. Here, we provide four cases that strike us as instructive—a contaminated-control problem, two scenarios in which there are unobservable types within the potential-donor pool, and one in which there are just a bunch of “bad controls” that would each fail the “common trends” assumption of difference-in-differences designs. In each, and for each parameter $\{\rho, \delta\}$ in $0(.01).99$, we summarize the relevant properties based on 1,000 simulated draws of the idiosyncratic components of the associated data-generating processes.

Contamination among potential donors

Data used in comparative case studies can often be subject to the criticism that some of the potential control entities may also have experienced treatment, which lead to bias in the estimated treatment effect. For the purpose of our simulation we care less about why this may be—typically it is a geographic neighbour, for example, that we anticipate experiencing some fallout from treatment. For example, Oregon’s legalization of recreational marijuana introduced variation in marijuana-related outcomes in Washington—retailers along the Oregon border experienced a 41-percent decline in sales following Oregon’s market opening (Hansen et al., 2017). We fully anticipate that institutional knowledge will continue to be brought to bear on analysis, and that the elimination of controls thought to be contaminated can proceed as per usual. However, where institutional knowledge is expensive, or we wish to communicate beyond those who have this knowledge, or where transparency will be appreciated, our procedure endogenously down-weights contaminated controls to the extent that there are evident *changes* in how potential

donors contribute to the SC’s dispersion after treatment. In this simulated context, we show how the introduction of a relative-dispersion penalty, in particular, can endogenously down-weight the contaminated control and eliminate bias, if not merely identify the sensitivity.

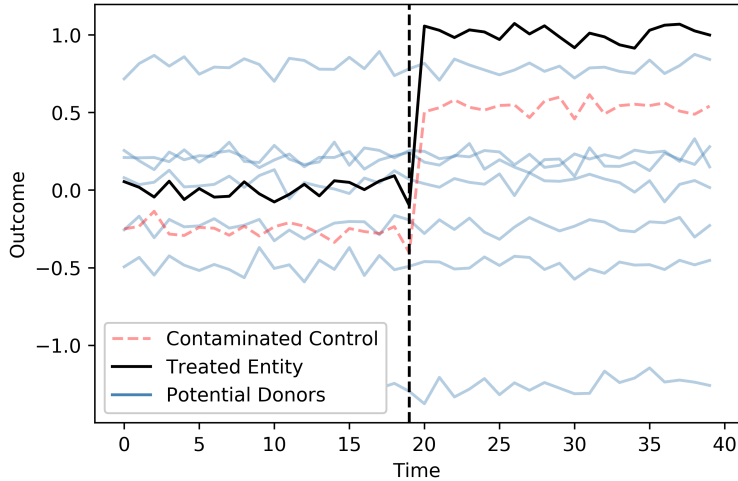
Here, we simulate a simple environment in which there is a true treatment (we set equal to one), but it spills into another untreated unit—a “contaminated-control” problem. The simulated data consists of one treated unit and eight potential controls, one of which experiences some fraction of the treatment experienced by the true treated unit. We otherwise let outcomes be sensitive only to a randomly determined intercept (from $N(0,1)$) and idiosyncratic shocks (from $N(0,0.1)$). The outcomes are observed for 40 periods, with treatment occurring at time 20, which introduces a level increase of 1 only for the treated unit, and 0.8 for the single “contaminated” control.⁹ (A single random draw from this data-generating process is depicted in Figure 2.)

In Panel A of Figure 3, we plot the estimated treatment effect across parameters, identifying the true treatment effect with a solid line, which the DWSC approaches smoothly in our simulated environment. This data-generating process also reveals a potential risk, evident in the declining treatment effect at extreme parameter values. That is, the estimator can tend toward single entities at some level, which we are inclined to think of as finding the single-best control, which would be desirable. However, in a data-generating process such as this, note that the best-available control could well be the contaminated control itself, which emphasizes that caution still be brought to bear. With high ρ , for example, the synthetic control here converges to the contaminated control $1/8^{th}$ of the time,

⁹ Contaminations of larger (smaller) sizes are likewise eliminated, but at lower (higher) values of ρ .)

FIGURE 2.

Contamination among potential controls: A single draw of the treated unit and all available donors



Notes: Data-generating process described in Section 2.3.

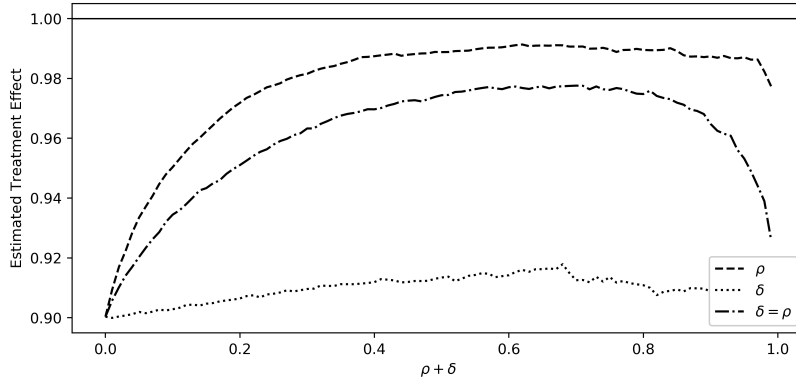
which is why the average estimated treatment drops again. (At $\rho = \delta = 0$, the contaminated control is given weight of .125. This is equivalent to the mean across iterations as $\delta \rightarrow 1$, where $1/8^{th}$ of the time the weight given to the contaminated control approaches 1.) In Panel B of Figure 3, we evidence the tradeoff of higher penalties on dispersion (i.e., higher pre-treatment MSE).

In Figure 4, we produce kernel densities of the treatment-effect estimates, across select parameter values. In both panels—in Panel A we vary ρ , and in Panel B we vary δ —the unpenalized SC draws from the set of potential donors similarly, with the contaminated control receiving weight in the synthetic control with equal probability. (Therein lies the problem, and source of bias.) As was implied in Figure 3, as we increase ρ , the contaminated control is down-weighted—when eliminated fully, precision around the true treatment parameter arrives. However, when the contaminated control is not eliminated, it can end up representing a

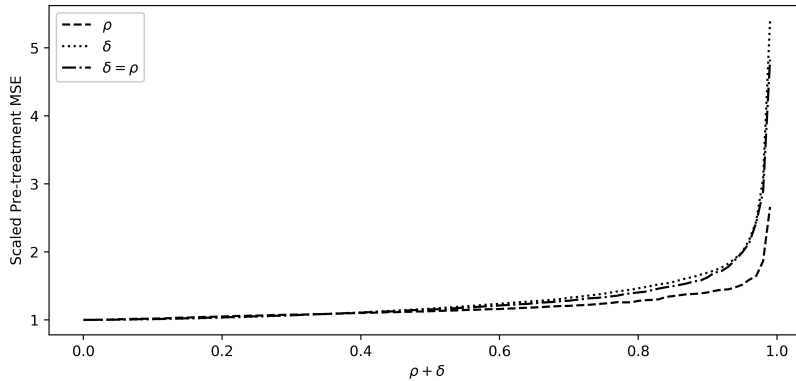
FIGURE 3.

Contamination among potential controls: Estimated treatment and pre-treatment MSE

Panel A: Mean treatment effect (true = 1), across parameters



Panel B: Pre-treatment MSE (normalized to 1 at $\rho = \delta = 0$)

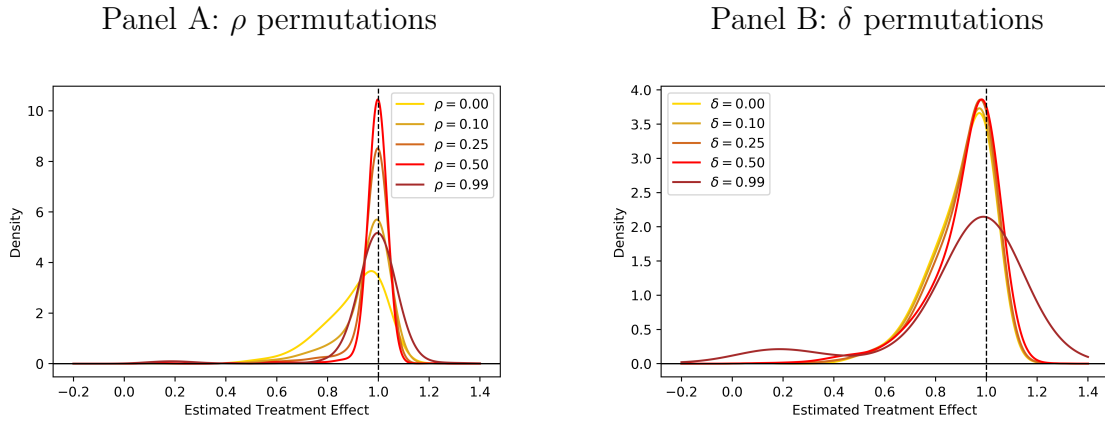


Notes: In all reported results we plot the range of parameter spaces across 0(.1).99. Data-generating process described in Section 2.3.

larger share of the synthetic control, and thereby induce a bi-modality—this second mode is precisely at the difference between the true treatment effect and the outcome of the contaminated control (i.e., $1 - .8 = .2$). This is more evident in Panel B, across increases in δ . As the overall dispersion in this simulated environment is driven largely by the level differences across donors, the exclusion of the contaminated control need not contribute significantly to reducing overall dispersion. On the other hand, the contamination does contribute substantially

FIGURE 4.

Contamination among potential controls: The effect of parameters on kernel densities around mean estimated treatment effect (true = 1)



Notes: In all reported results we plot the range of parameter spaces across 0(.1).99. Data-generating process described in Section 2.3.

to relative dispersion, which leaves ρ as the more-direct and more-effective margin when there are contaminated controls among the potential donors.

Unobservable types, different in their trends

Here, we imagine an unobserved heterogeneity that one might anticipate being challenging to a synthetic control environment—the treated unit being one of two unobservable types, largely overlapping in the pre-treatment period. In this case, we assume that the true treatment effect is zero. While all entities in the data-generating process will follow similar paths—hence, the challenge—we put one-third of potential donors on a path that slowly drifts away from the other two-thirds of potential donors, as though there are unobservable components contributing to their outcomes that do make them less and less comparable over time. Though smooth, we purposefully configure the divergence so as to become increasingly apparent, which highlights the potential for an MSE-driven synthetic

control to give weight to these entities when not considering their post-treatment behavior.

Specifically, we posit 21 entities, each observed over 40 periods, with “treatment” falling in the last half of observations, 14 entities are “A” types, following $Y_{it}^A = 0.02t - 0.002t^2 + \epsilon_{it}$ and 7 entities are “B” types, following $Y_{it}^B = 0.07t - 0.005t^2 + \epsilon_{it}$, where ϵ_{it} are drawn from $N(0,0.1)$. Including the “treated” unit among the “A” types tips toward having plenty of donors similar to the teated unit from which to construct a reasonable synthetic control. Likewise, then, including the “treated” unit among the “B” types tips toward having fewer donors similar to the teated. We will consider both scenarios, and in so doing find that the unpenalized environment assigns weight to donors of the “wrong” type in both scenarios, while penalizing dispersion down-weights “wrong” types—again, this is true regardless of which type we assign to the treated unit. (As anticipated, given the data-generating process in this environment, there is greater sensitivity evident around permutations of δ , controlling overall dispersion, than to ρ .) This scenario is depicted in Figure 5, as both the theoretical data-generating process and single representative draws.

In Figure 6, we plot the estimated treatment effects and pre-treatment MSE across the parameter space, in each case identifying the true treatment effect (zero) with a solid line. In Panel A, the treated unit is assigned to Type A. In Panel B, the treated unit is assigned to Type B. In both, the traditional approach to estimating synthetic controls is biased away from the true treatment effect, as “other” types receive positive weight to the extent that they help minimize pre-treatment MSE. However, in either case, permutations of our dispersion-weighting parameters have the DWSC approach the true treatment effect smoothly.

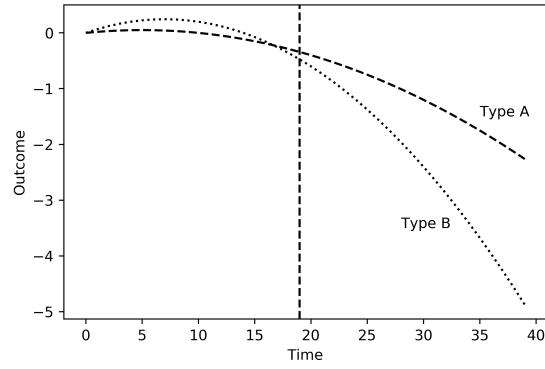
Moreover, in parameters, the fraction of weight given to donors of the same type increases—this is shown in the second row of plots, with convergence to A types on the left and B types on the right. Where a traditional approach to deriving a synthetic control struggles to identify type appropriately, *regardless of the treated unit’s type*, our procedure has the synthetic control collapsing on the same—the set of potential donors that we would anticipate being the appropriate counterfactuals, *a priori*. Likewise, the estimated treatment converges to its true level as either parameter increases. In the third row, we demonstrate the associated tradeoff also evidence the fundamental tradeoff associated with greater synthetic-control stability—higher pre-treatment MSE.

In Figure 7, we produce kernel densities that speak to the distribution around the mean treatment effect. In this setting it is informative, as it reveals a certain tension in the procedure. First, these kernels make evident the bias in traditional approach (i.e., $\rho = 0$ and $\delta = 0$, each in yellow). Second, with increases in either dispersion penalty, the expected parameter clearly moves toward unbiasedness, but there is also the initial suggestion of a bi-modality, as the underlying types are sorted out. Third, as one might anticipate, the distribution of treatment parameters is collapsing faster when we penalize overall dispersion (δ) than when we penalize relative dispersion (ρ). Yet, also evident is the increase in variation around the point estimate for extreme parameter values—this is also anticipated, as it coincides with the synthetic control collapsing on the best-available control as $\rho \rightarrow 1$, or $\delta \rightarrow 1$. However, this eventual loss of precision is notably less pronounced in extreme values of *rho* than in extreme values of δ , consistent with the synthetic control (and the estimated treatment effect itself) being less sensitive to ρ permutations (see Figure 6). While less generalizable, note also that there are

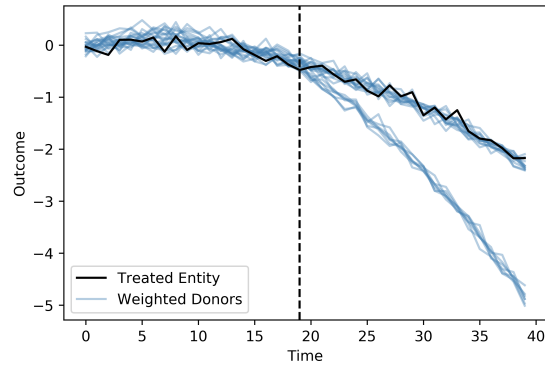
FIGURE 5.

Unobserved types, different in trends: The treated unit and all available donors

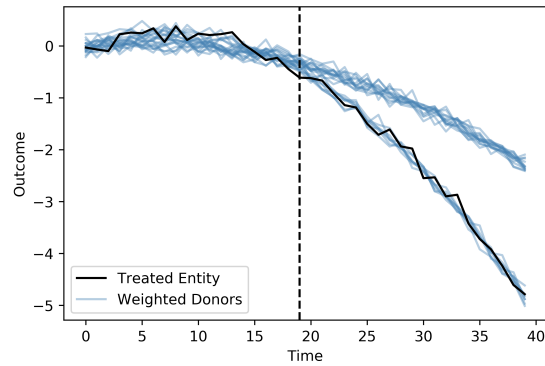
Panel A: Theoretical data-generating processes



Panel B: A representative draw when treated unit is assigned to Type A



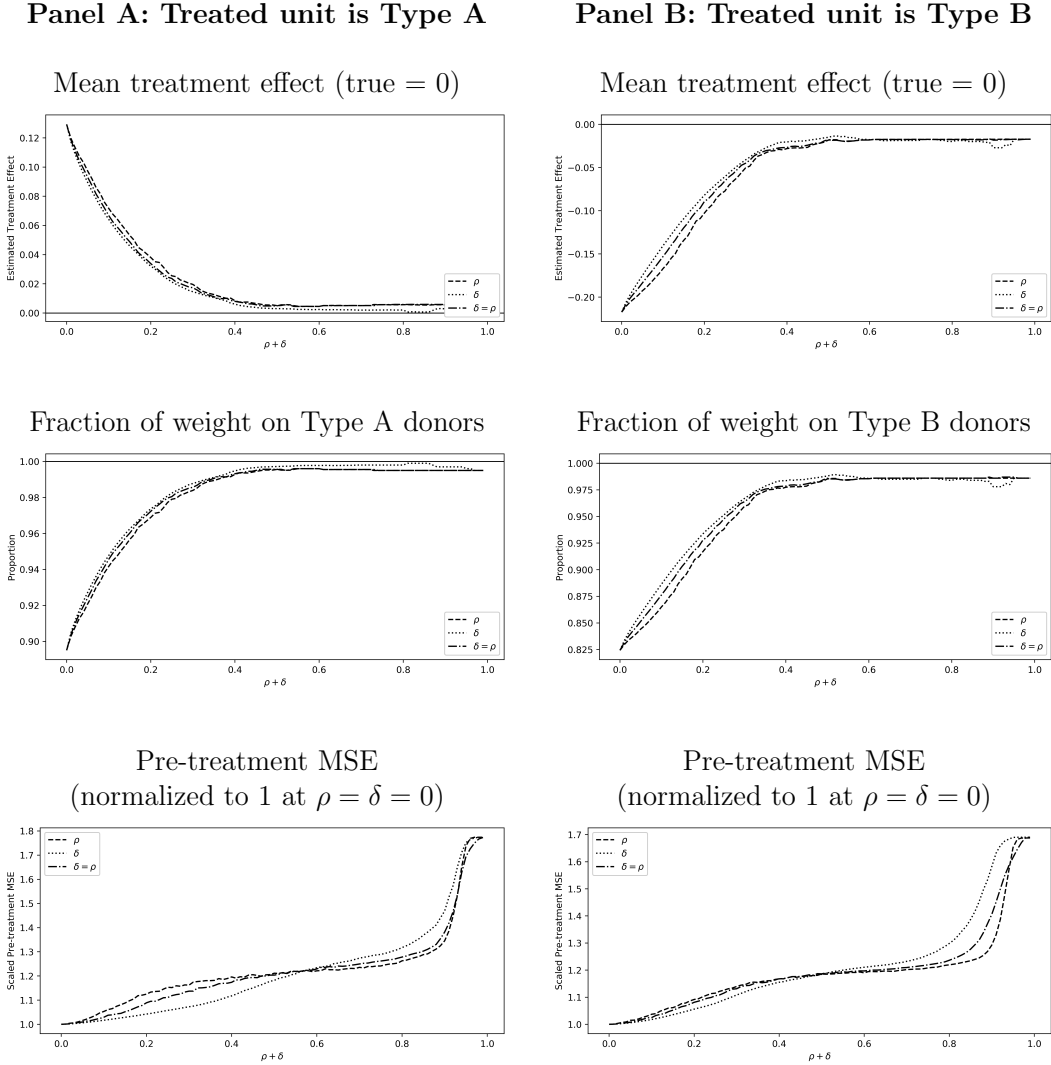
Panel C: A representative draw when treated unit is assigned to Type B



Notes: Data-generating processes described in Section 2.3.

FIGURE 6.

Unobserved types, different in trends: Estimated treatment and pre-treatment MSE



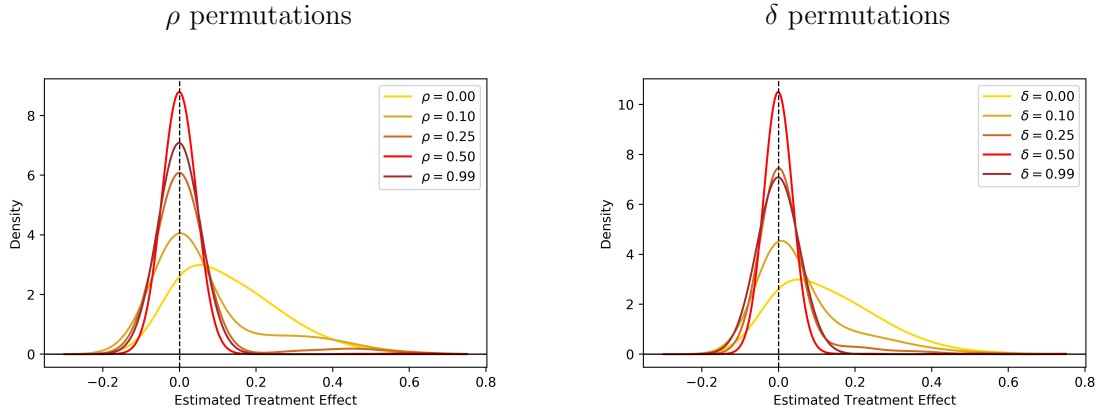
Notes: In all reported results we plot the range of parameter spaces across 0(.1).99. Data-generating process described in Section 2.3.

more “A” types available in the donor pool than “B” types. As such, the synthetic control is initially less sensitive to parameters when the treated unit is an A Type, and collapses to the best-available control at higher parameter values—when the treated unit is a B Type, the best-available is reached sooner.

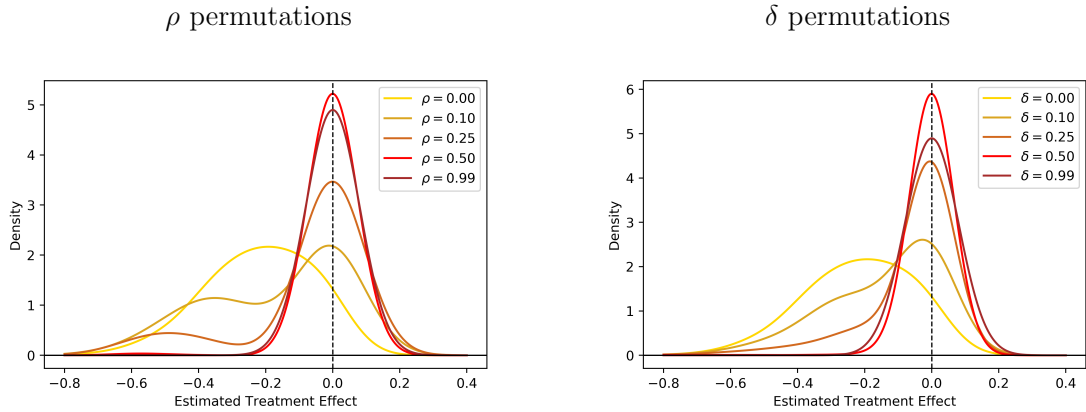
FIGURE 7.

Unobserved type, different in trends: The effect of parameters on kernel densities around mean estimated treatment effect (true = 0)

Panel A: Treated unit is Type A



Panel B: Treated unit is Type B



Notes: In all reported results we plot the range of parameter spaces across $0(.1).99$. Data-generating process described in Section 2.3.

Unobservable types, different in their idiosyncratic errors

We now imagine an unobserved heterogeneity in which δ may play the first-order role. Specifically, we imagine the treated unit being one of two unobservable types, where instead of those types diverging and creating bias proportional to their

weight in the synthetic control, we consider the estimator’s behavior in light of having added noise to the system.

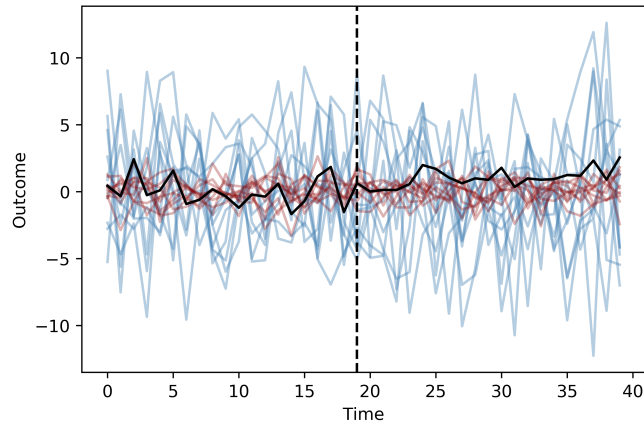
Specifically, we posit 21 entities, each observed over 40 periods, with “treatment” (true=1) falling in the last half of observations, 10 entities are “high-variance” types, following $Y_{it}^H = \epsilon_{it}^H$, where ϵ_{it}^H is drawn from $N(0,4)$ and 10 entities are “low-variance” types, following $Y_{it}^L = \epsilon_{it}^L$, where ϵ_{it}^L are drawn from $N(0,1)$. Unlike the problem of Section 2.3, where our procedure had the synthetic control collapse on the same type as the treated unit, here we will find the procedure having the synthetic control converge to those donors with the smaller variance, whether the treated unit is itself found among the low- or high-variance types. (Representative draws from this data-generating process are depicted in Figure 8.)

In Figure 9, we plot the relevant properties of the estimator across 1,000 draws. In Panel A, we assign the treated unit to be among the low-variance types. In Panel B we assign the treated unit to be among the high-variance types. Unlike the environment above—recall that DWSC converged toward using donors of the same type as the treated unit—in the first row we see that penalizing dispersion has the estimator converge to using low-variance donors, *regardless of the treated unit’s type*. In the second row, we plot the variance in estimated treatment, also across parameters, demonstrating an increase in precision as the synthetic control collapses on low-variance types. However, there is a distinct tradeoff in this environment, that can ultimately show up as *increasing* variation in the estimated treatment effect. Namely—in the third row we produce plots of the *number* of donors receiving positive weight—the number of donors contributing to the synthetic control is clearly decreasing in parameters. As such, the effect

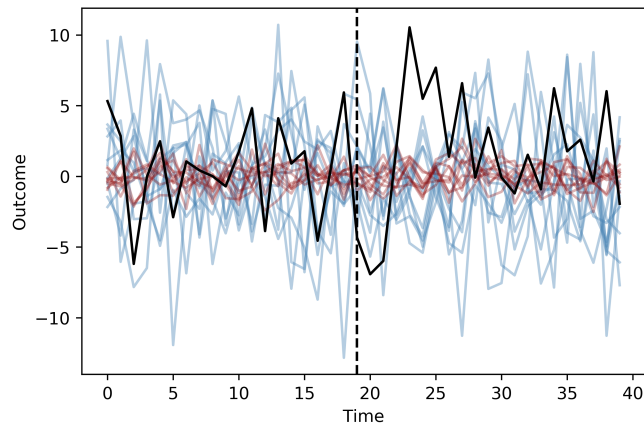
FIGURE 8.

Unobserved types, different in idiosyncratic error: The treated unit and all available donors

Panel A: A representative draw when the treated unit is a low-variance type ($\text{true} = 1$)



Panel B: A representative draw when the treated unit is a high-variance type ($\text{true} = 1$)



Notes: Data-generating processes described in Section 2.3.

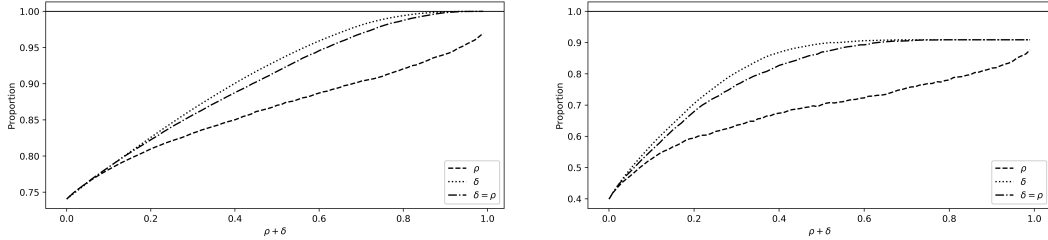
on precision need not be beneficial. In this environment, the “number of donors” ultimately overtakes the precision associated with the DWSC finding low-variance

FIGURE 9.

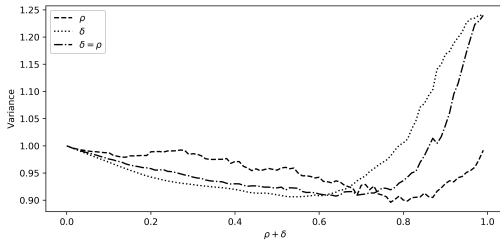
Unobserved types, different in idiosyncratic error: Precision-related properties

Panel A: Treated unit low-variance **Panel B: Treated unit high-variance**

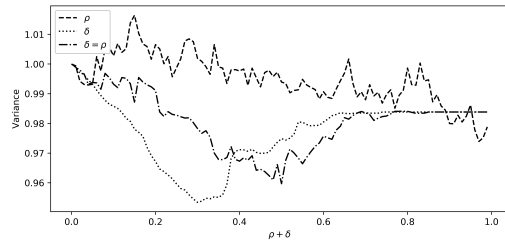
Fraction of weight on low-variance donors Fraction of weight on low-variance donors



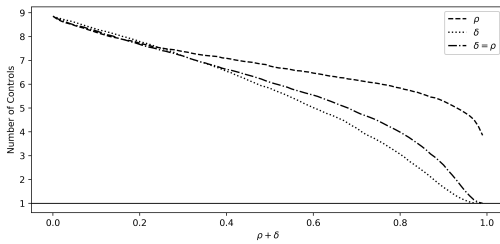
Variance in treatment effect
(normalized to 1 at $\rho = \delta = 0$)



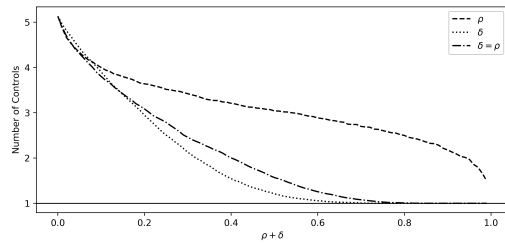
Variance in treatment effect
(normalized to 1 at $\rho = \delta = 0$)



Number of weighted donors



Number of weighted donors



Notes: In all reported results we plot the range of parameter spaces across 0(.1).99. Data-generating process described in Section 2.3.

donors—this tradeoff is more likely to bind the closer are the two types in their variance properties.¹⁰

¹⁰ As the difference in variance across types increases, the benefits associated with “finding low-types” decrease yet the costs associated with “number of donors” falling remain.

“Bad” pre-trends

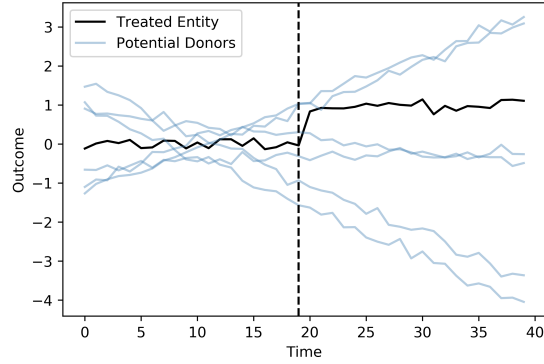
As one last simulation, we imagine the condition in which the potential donors are trending differently and where SCM might be the go-to approach to evaluating the effect of treatment. In some sense, we imagine hardwiring what some have described as the typical picture of “bad” controls that SCM can potentially work within. Within a simulated environment, of course, the mean behavior of an SCM approach will not be biased—if potential donors are merely on different linear trends, this is not a “bad controls” problem at all. As such, documenting the variance properties of the estimator is our interesting here. A representative draw of this environment is depicted in Panel A of Figure 10, where we posit 6 potential donors, each observed over 40 periods, with “treatment” (true=1) falling in the last half of observations, where, otherwise, the treated unit follows $Y_t = \epsilon_t \sim N(0,0.1)$. Each donor follows $Y_{it} = \alpha_i + \beta_i t + \epsilon_{it}$, where $\alpha_i \sim N(0,1.5)$, and $\beta_i \sim N(0,0.2)$, and $\epsilon_{it} \sim N(0,0.1)$.

Given the linear path of each potential donor, there is a set of weights w such that the linear combination of donors well-approximates the path of the treated group for any given set of trending donors—given that they do not deviate from their paths in our simulated environment (other than through an idiosyncratic term), the counterfactual is also well-approximated in the post-treatment period and the estimated treatment is unbiased (Panel B). As anticipated, our procedure identifies treatment effect with increasing variance around the 1,000 point estimates we simulate, which we show in Panel C.

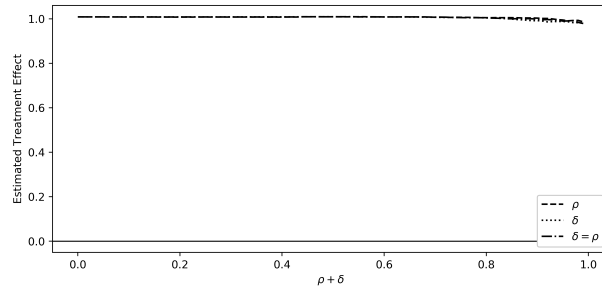
Without different patterns of behavior among donors before and after treatment, the standard SCM is not ill-equipped to retrieving the treatment effect from such a data-generating process. That said, DWSC applied to such a data-

FIGURE 10.
DWSC with a “Bad controls” problem

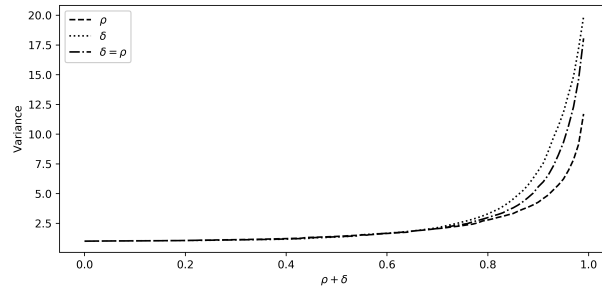
Panel A: A representative draw (true = 10)



Panel B: Mean treatment effect (true = 10)



Panel C: Variance of treatment effect
(normalized to 1 at $\rho = \delta = 0$)



Notes: Data-generating processes described in Section 2.3.

generating process will still afford a confidence, insofar as the estimates effect of treatment is robust to ρ and δ permutations.

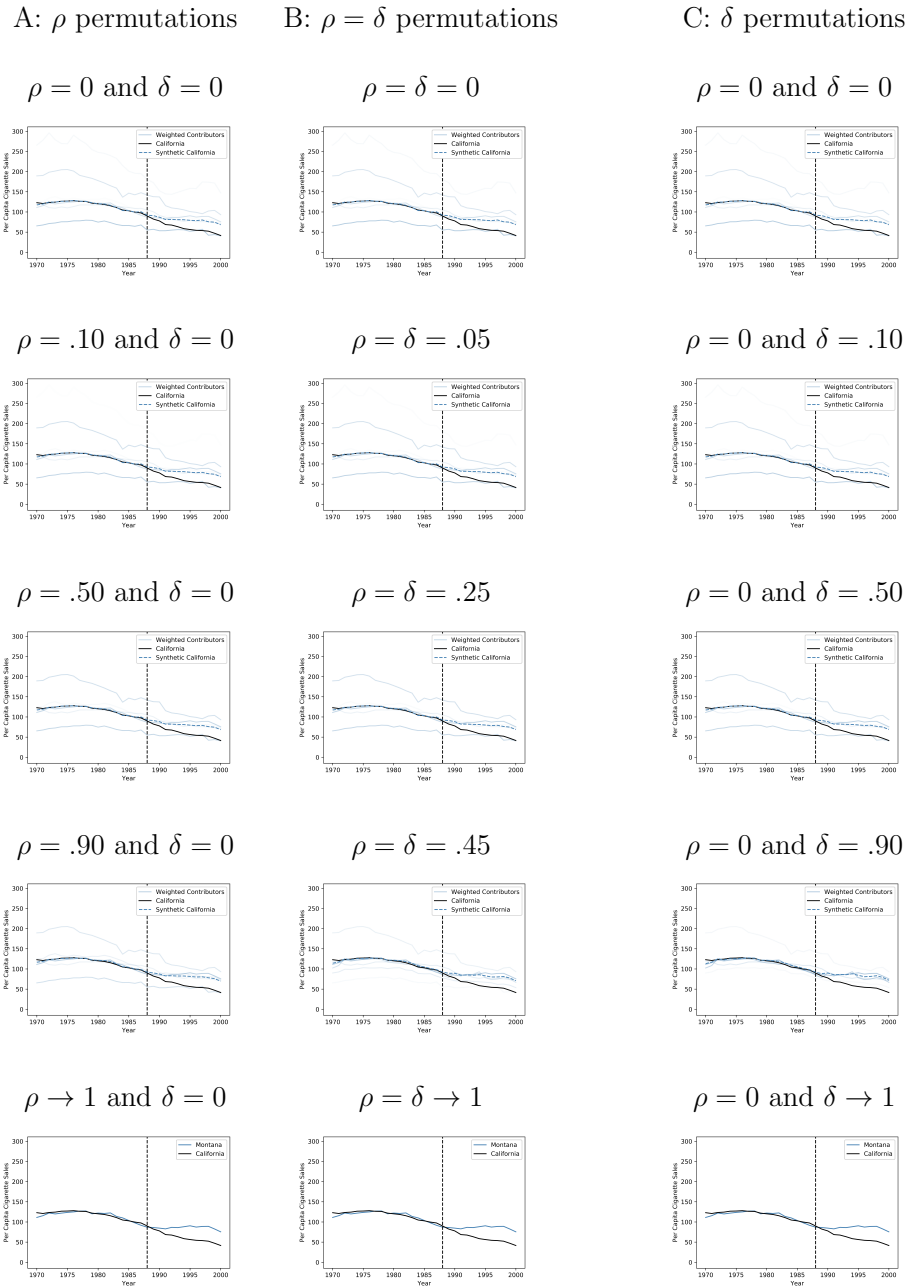
Cigarette sales and anti-smoking initiatives

Here, we apply our estimator to the data of Abadie et al. (2010), re-considering the effect of California’s 1988 anti-smoking initiative known as Proposition 99. In Figure 11, we offer 15 different plots—together they span the $\rho + \delta$ parameter space, while in each we produce the time series of California cigarette sales (solid black), cigarette sales in the Synthetic California (dashed blue), and cigarette sales in each of the donors to that synthetic (solid blue). For these individual donors, we plot them with levels of intensity that are proportional to the weights (w) given to each state in the Synthetic California.

While the top row (i.e., $\rho + \delta = 0$) is qualitatively similar to that of Abadie et al. (2010), our weights do differ somewhat, which highlights the role played by covariates in their analysis.¹¹ In the second through fourth rows, we have allowed for the endogenous down-weighting of donors to the synthetic control subject to ρ and δ dispersions—we plot select examples for permutations of ρ (Panel A), of $\rho = \delta$ (Panel B), and of δ alone (Panel C). Recall that it is donors with *differences* in dispersion that are down-weighted as ρ increases and we move down Panel A, as the estimator increasingly prefers donors that contribute *similarly* to SC-dispersion on either side of treatment. As δ increases (moving down Panel C), the estimator prefers donors who are, on average, proximate to the SC’s average each period. In our experience, the limiting cases of $\rho + \delta \rightarrow 1$ is helpful in identifying something of a single best-available control. In the last row, we see that two of the three columns

¹¹ It is the addition of covariates (and necessarily, then, the dropping of several observations of outcomes) that explains the exclusion of New Hampshire, for example. In all of our analysis we exclude covariates but include all pre-treatment outcomes, eliminating researcher choice and thereby increasing comparability across analyses.

FIGURE 11.
Per-capita cigarette sales (packs) in California, 1970 to 2000



Notes: For given parameters, we plot the surviving donors from the analysis described in Section 2.3, with each line's intensity proportional to the assigned weights.

the synthetic control has collapsed to the single best-available control from among those in the set of potential donors.

While changes in the composition of the synthetic control are evident, the overall picture in this application is one of stability in the combination of donor-states, and even in their relative weights—they just don’t change very much. The robust nature of the synthetic control in terms of its composition of states across parameters is also made evident in Figure 12, where we offer a different view of the donor weights resulting from our permutations.

In Panel A of Figure 13, we produce treatment effects across three separate conditions—changing the importance of relative dispersion (ρ), overall dispersion (δ) and increasing the dispersion penalties subject to an equal-weighted condition (i.e., $\rho = \delta$).¹² In each of their origins (i.e., at $\rho = \delta = 0$), we are approximating the typical synthetic control, as we have highlighted above. (The point estimate at our $\rho = \delta = 0$ origin is -19.5, and very close to the point estimate of Abadie et al. (2010), which is -20.) The noteworthy point, however, is the robustness of the estimated treatment effect to our permutations, so much so that the three plots are largely on top of one another, and not straying from the published result. This panel speaks to a stability in the original inference of Abadie et al. (2010)—if anything, we come away from this exercise imagining that “20-percent reductions in cigarette sales” is a lower bound on the estimated treatment effect were one to penalize within-SC changes in dispersion.

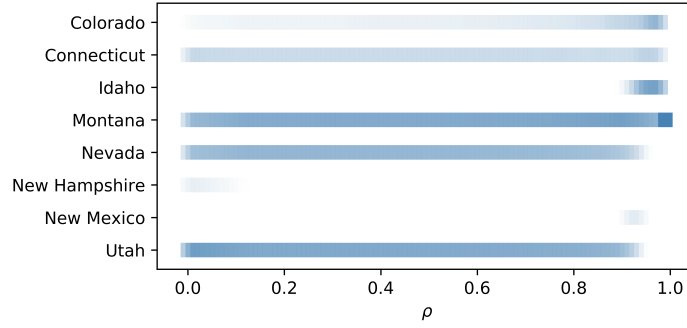
Possibly as important, though implicit in the stability of the estimated treatment, is how little pre-treatment fit is given up in these parameter permutations. The fundamental tradeoff as ρ or δ increase is to give up pre-

¹² In all reported results we plot the range of parameter spaces across 0(.1).99.

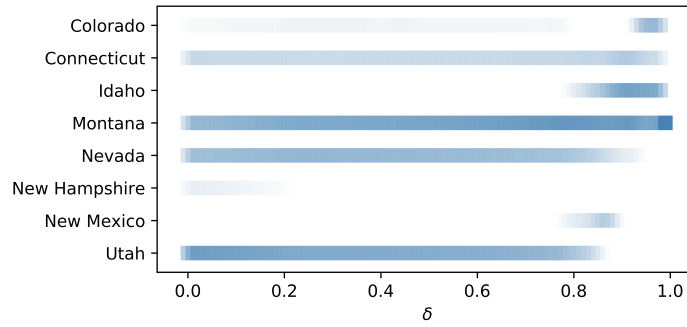
FIGURE 12.

Donor-inclusion plots: Contributions to the dispersion-weighted synthetic controls around California's anti-smoking campaign

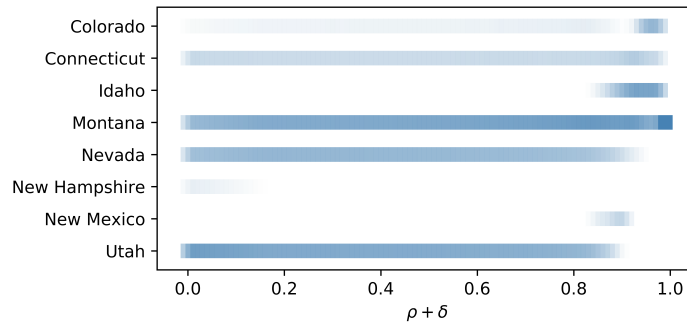
Panel A: ρ permutations



Panel B: δ permutations



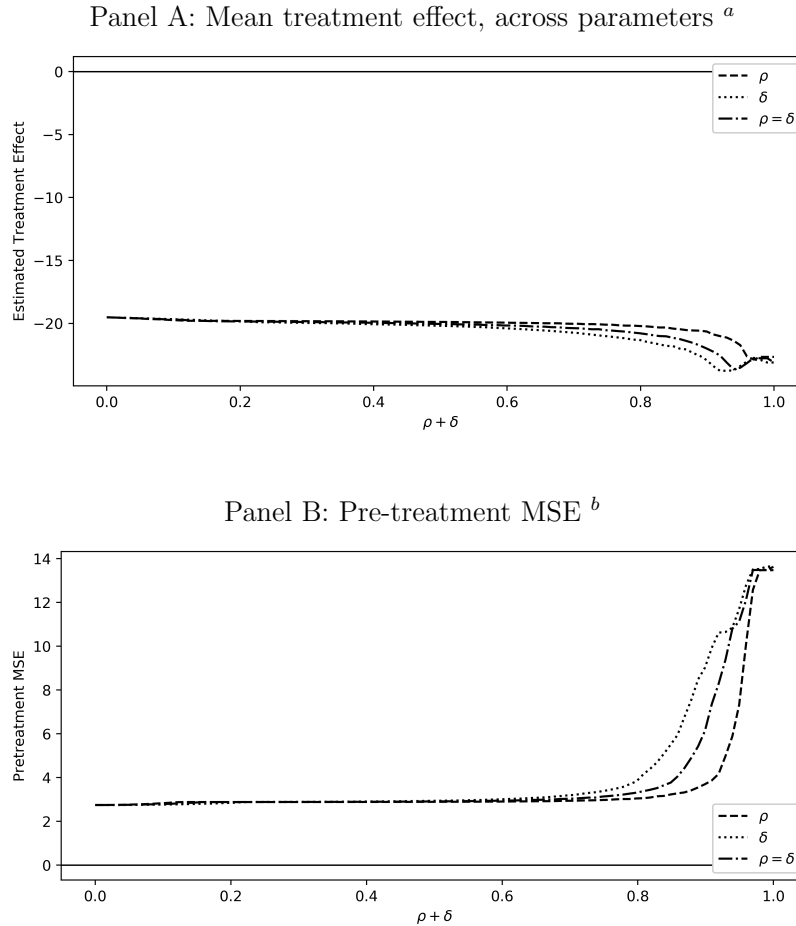
Panel C: $\rho = \delta$ permutations



Notes: In all reported results we plot the range of parameter spaces across 0(.1).99. The synthetic control in Abadie et al. (2010) is the weighted average of Utah (0.334), Nevada (0.234), Montana (0.199), Colorado (0.164), and Connecticut (0.069).

FIGURE 13.

The robustness of the Abadie *et al.* (2010) “California result” to dispersion-weighted synthetic control (DWSC)



Notes: In each panel, we plot the estimated treatment effect across $[0, 1]$, for $\rho \in [0, 1]$ (while setting $\delta = 0$), $\delta \in [0, 1]$ (while setting $\rho = 0$), and $\rho + \delta \in [0, 1]$ (while setting $\rho = \delta$). In all reported results we plot the range of parameter spaces across $0(.1).99$. ^a Abadie *et al.* (2010) reports that cigarette consumption was reduced by an average of almost 25.4 packs (per capita, annually). ^b Abadie *et al.* (2010) reports a pre-treatment MSE of roughly 3.

treatment fit for greater SC-stability. In Panel B of Figure 13, we capture the changes in MSE across the same space, and the flatness in Panel B is encouraging, and consistent with the cost side of increasing the importance of either dispersion penalty not being overly onerous in these data. To our eye, we would like to see this sort of robustness, as it coincides with an appealing stability in the control group to which California is ultimately compared.

While instructive, Figure 11 is exceedingly inefficient for presentation purposes, as might be Figure 12. Thus, unless there are particular points to be made, we propose that researchers communicate something akin to Figure 13.

Reported rape offences and the decriminalization of prostitution

The recent SCM paper of Cunningham and Shah (2018) is very well known, and thus provides an appealing opportunity to consider the DWSC procedure around the unexpected decriminalization of prostitution in 2003 on reported rape offences. It is a nice example of the sort of policy variation that one with an eye for clever sources of variation should jump on, and to which SCMs might naturally be applied.¹³ Below, we follow our procedure in estimating the effect of Rhode Island’s decriminalization on rape reports, with data from the Uniform Crime Reports, Federal Bureau of Investigation.¹⁴

In Figure 15, we produce our collection of plots, across parameters, which ultimately identifies New Hampshire as the single-best control. However, the instability in building a synthetic control out of the available time series of donors around this natural experiment is most evident when we plot the estimated

¹³ In short, the effect of decriminalization on reported rape offences is an empirical question, as decriminalization yields safer work spaces to the extent firms are more willing to invest, but sex workers could also be more willing to report to and cooperate with police upon assault.

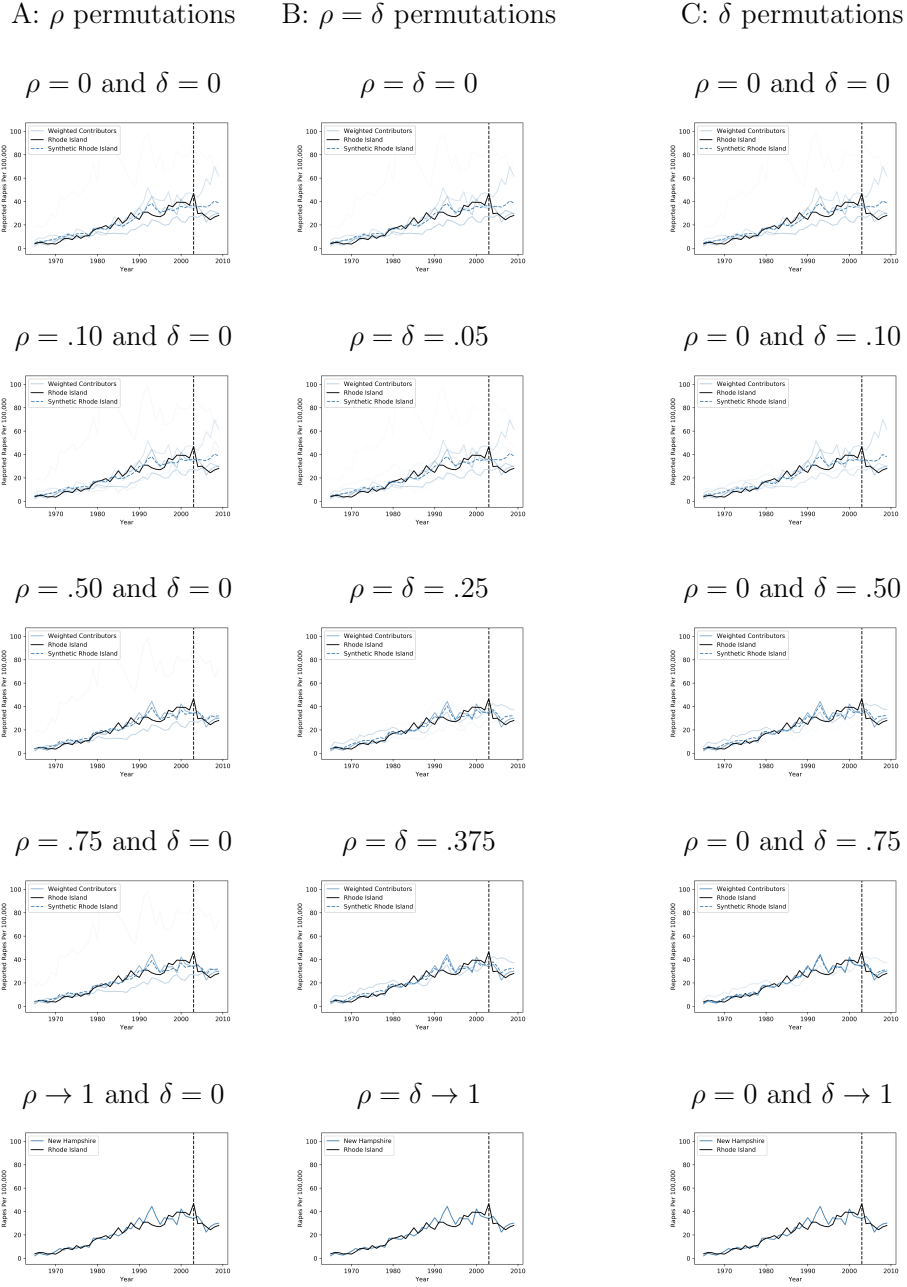
¹⁴ Defining a synthetic control for Rhode Island (the treated state), among the conclusions made in Cunningham and Shah (2018) is that reports of rape fall by roughly 32 percent with the decriminalization of prostitution, from 34.1 per 100,000, to 20. As we proceed, we do so knowingly make several departures from Cunningham and Shah (2018), even when conditioning on $\rho = \delta = 0$. For example, in the determination of weights, we do include all pre-treatment rape reports between 1970 and 2002 (they match only on 1979, 1995, 2001, 2002, and 2003). We do not include averages of any rape reports (they match on mean rape over 1992 through 1995, mean rape over 2001 and 2002, and mean rap over 2002 and 2003). We model outcomes in levels (while in their preferred specification, they model two-year averages “to minimize the volatility in the series”). Our procedure applied to the smoothed data suggest that the deviation coincident with Rhode Island’s decriminalization of prostitution is smaller than that in the level-data.

treatment and MSE across parameters, which we do in Figure 15. The sensitivity of the SC's composition as we allow for more weight on the stability properties of the synthetic control is somewhat striking, as tipping points are clearly reached and individual donor states are discarded from the set of donors. This was not at all a property of the California smoking data, which was quite well behaved through all of our perturbations—this context alone is important to drawing inference from a particular set of ρ and $\delta = 0$.

Again, there are only slight pre-treatment MSE costs suffered over large regions of the parameter space. Yet, over the same space, treatment estimates vary significantly. Unless there are very strong priors made on what contributes to a good control, of the sort that would preclude consideration of the stability properties as we have done, this is indicative of an environment that leaves the estimated treatment itself the discriminating element across a very wide range of potential SCs with similar in their pre-treatment MSE.

While our analysis supports that, relative to a synthetic controls, reported rapes fall with decriminalization in Rhode Island, our range of estimates is somewhat wide—declines of roughly 9 per 100,000 without any consideration for dispersion, falling quickly as we increase the stability of the synthetic control in either ρ or δ dimensions (to roughly -5), and, as dispersion is weighted more heavily, estimates are as low as 1.6 fewer rape offences reported per 100,000. The sensitivity of the composition of the synthetic control across parameters is also made evident in Figure 16. While New Hampshire remains a significant contributor throughout most of the parameter space (and to a lesser extent Iowa), the jumps in estimated treatment in Figure 15 are clearly coinciding with the exit and entry of other states from the set of weighted donors.

FIGURE 14.
Reported rape offences in Rhode Island, 1970 to 2009

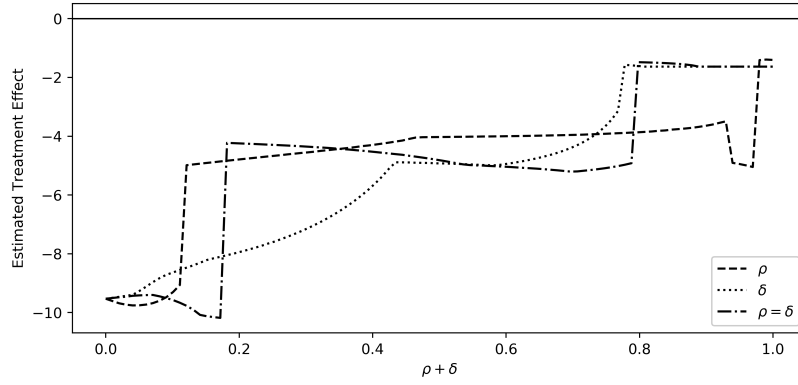


Notes: Reported rape offences per 100,000. For given parameters, we plot the surviving donors from the analysis described in Section 2.3, with each line's intensity proportional to the assigned weights.

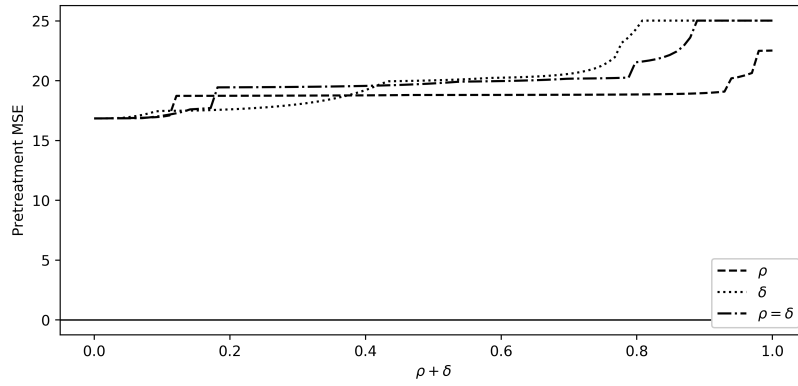
FIGURE 15.

Reported rape offenses and Rhode Island's decriminalization of prostitution, robustness to dispersion-weighted synthetic control (DWSC)

Panel A: Mean treatment effect, across parameters



Panel B: Pre-treatment MSE

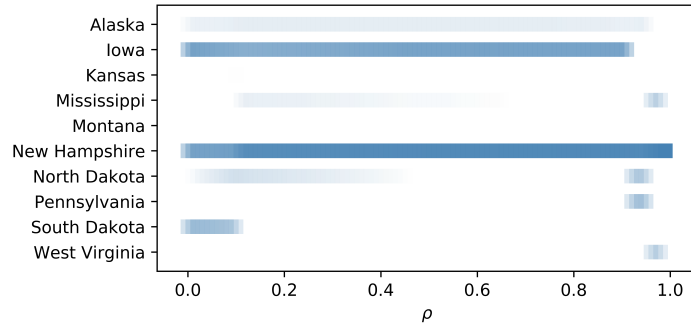


Notes: In each panel, we plot the estimated treatment effect across $[0, 1)$, for $\rho \in [0, 1)$ (while setting $\delta = 0$), $\delta \in [0, 1)$ (while setting $\rho = 0$), and $\rho + \delta \in [0, 1)$ (while setting $\rho = \delta$). In all reported results we plot the range of parameter spaces across $0(.1).99$.

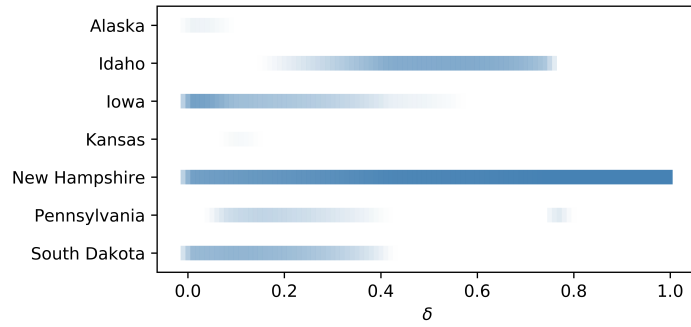
FIGURE 16.

Donor-inclusion plots: Contributions to the dispersion-weighted synthetic controls around Rhode Island's decriminalization of prostitution

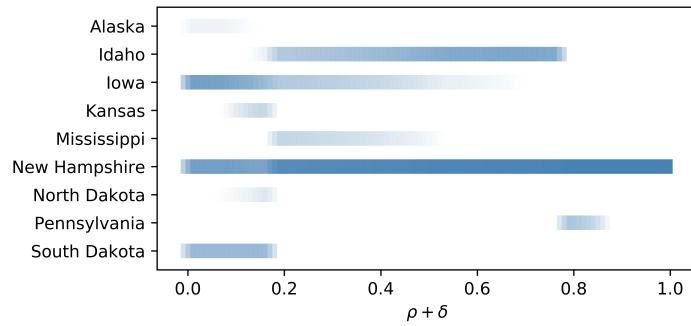
Panel A: ρ permutations



Panel B: δ permutations



Panel C: $\rho = \delta$ permutations



Notes: In all reported results we plot the range of parameter spaces across 0(.1).99. The synthetic control in Cunningham and Shah (2018) is the weighted average of South Dakota (0.356), Idaho (0.342), New Hampshire (0.162), and North Dakota (0.140).

Conclusion

Our intent is not to introduce one new synthetic control to fix all synthetic controls—there is much work still being done in the area, with different emphases and approaches (Athey et al., 2017; Xu, 2017) and with the researcher community still refining our understanding of how individual approaches map across each other (Doudchenko and Imbens, 2016; Rothstein et al., 2018). Rather, we propose a procedure that reasonably nests a common approach to running synthetic controls, while allowing for parameter choices that control the weight given to stability among the weighted donors—a dispersion-weighted synthetic control (DWSC). We ultimately propose that researchers produce figures that plot estimated treatment effects and pre-treatment MSEs across the feasible ranges of these stability influencing parameters (something like Figure 13 or Figure 15), although “donor-inclusion plots” (e.g., figures 12 and 16) are also informative.¹⁵

While we assume, throughout, that institutional knowledge will remain informative in assessing the potential credibility of individual donors, to rely on institutional knowledge (or ocular econometrics) to catch such events is not efficiently repeatable, while leaving a dimension of scope available to researchers that may itself be concerning (Ferman et al., 2018). To be clear, then, the value in our procedure is not in either limit but rather in the behavior of the estimated treatment effect across parameters—its sensitivity to the *ex ante* importance of SC-stability, to the researcher or as is suitable in the environment. If there is a reasonable stability to the synthetic control across the parameter spaces we allow for, then our confidence should increase. (We imagine confidence in the California

¹⁵ We also anticipate that inference will be performed through permutation tests, as in Abadie et al. (2010) and much of the synthetic-control literature to date. We leave the particulars of that area of investigation to future work.

smoking result increasing here, for example.) Overall, if the figures we propose reveal an instability in the synthetic control itself, or that point estimates change even though pre-treatment MSE does not, or that estimated treatment is falling in magnitude as the stability of the synthetic control increases—if any of these—then we imagine additional limits being put on resulting conclusions.

(Python script to implement the above procedure is available from the authors' webpages.)

CHAPTER III

THE UNINTENDED CONSEQUENCES OF SUPPLY SIDE DRUG INTERVENTION: EVIDENCE FROM DEA CHEMICAL CLASSIFICATION

Introduction

The negative social consequences of illicit-drug markets are well documented and wide ranging, stemming from the behavior of both traffickers and users of the substances. Illegal drug markets are often associated with all manner of crime. For example, Evans et al. (2018) finds that crack-cocaine markets are associated with violence long after their inception, while Castillo et al. (2014) finds that cocaine scarcity leads to violence in areas of Mexico most associated with cocaine trafficking. Drug users (and addicts especially) are often thought to commit crimes at a higher rate than non-users, and are especially prone to committing financially motivated crime. Nurco et al. (1991) observes that the use of heroin and cocaine are strongly associated with criminality, with use of cocaine in particular being substantially higher than the general population among prisoners, parolees, probationers, and arrestees.

In light of this, law enforcement agencies worldwide devote substantial resources to disrupting the manufacture, transportation, and distribution of illicit substances. These types of intervention are generally intended to decrease drug use by reducing the availability of drugs, and weakening the structure of their markets. By reducing supply, the price of drugs should increase, and the equilibrium quantity consumed should decrease. This philosophy underlies an enormous portion of state attempts to combat the negative consequences of drug

use. Unfortunately, this approach to policy may not work as intended. Rather than substituting away from drug use, users may select into criminal activity (or increase the intensity of their current criminal activity) to pay the higher prices.

This drug price/crime-effect is examined in Silverman and Spruill (1977) where the authors find a positive elasticity of property crime with respect to heroin prices in the city of Detroit. However, this analysis suffers from two problems. First the analysis is constrained to a single city which may not represent behavior on a larger scale. Second, as the price variation is not associated with supply side movement, we cannot interpret this relationship causally. The positive correlation of prices and crime may occur because increases in property crime yield higher drug prices and not vice versa. ¹

Here, I analyze a national supply side shock to the cocaine market in the United States stemming from the DEA placing regulations on the manufacture and distribution of sodium permanganate, a chemical used in the production of cocaine. Similar shocks on the methamphetamine market are examined in Dobkin and Nicosia (2009) and Dobkin et al. (2014). While the authors find some short term effectiveness of pseudoephedrine regulation in terms of increasing prices and decreasing purity, they find that prices and purity rebound relatively quickly to levels close to their pre-treatment values. In terms of impacts on criminal behavior, they report mild increases in robbery, but do not generally observe across the board increases in financially motivated crime.

In this analysis, I leverage variation in addition to the supply shock, in the form of geographic differences in addiction rates. This is driven by the observation that areas with higher levels of cocaine addiction should be more sensitive –

¹For example, increases in crime unrelated to drug prices could lead to higher income for drug users, and thus higher drug demand.

in expectation – to national drug prices. When accounting for this additional variation, I find that the Drug Enforcement Administration’s (DEA) regulation of sodium permanganate led to higher levels of property crime in areas with more cocaine addiction, compared to areas with low addiction. This impact is on the order of 6.8% - 9.1% of pretreatment property crime rates. No impact appears to be present in violent crime, nor does it manifest when interacting the shock with geographic variation in alcohol addiction, further suggesting that the systematic variation observed in this analysis is particular to the cocaine drug channel.

In Section 3.2, I provide background on the policy variation in question, while in Section 3.3, I describe the origin and structure of the data used for the analysis. The approach to econometric identification is presented in Section 3.4. Empirical results are presented in Section 3.5, and further discussed in Section 3.6. Section V concludes.

Background

In December of 2004, the DEA proposed the classification of sodium permanganate as a list II chemical under the Controlled Substances Act of 1970. Classification of this nature institutes “Know your customer” responsibilities, manufacturing inventory and use reports, 15 day advanced DEA notice for imports and exports, effective security controls, and required reporting for unusual sales or losses.² These measures are intended to avoid diversion of the chemical to the illicit production of cocaine. A public comment period lasted until May 2005,

²Sometimes shortened as KYC, know your customer laws require that producers and distributors of the chemical verify the identity of their customers and assess the likelihood of that the product is used for illegal purposes.

the decision was finalized in September 2006, and firms were expected to be in compliance in December of 2006.

Motivating this proposal was sodium permanganate's "direct substitutability for potassium permanganate in the illicit production of cocaine, as well as recent cocaine related drug busts where the chemical was found."³ Cunningham et al. (2015) evaluates four different cocaine precursor chemical regulations of this nature, and concludes that they did to various extents disrupt the cocaine market. The sodium permanganate regulation in particular was found to be especially effective in both decreasing purity and increasing prices. Thus, I hope to exploit the variation in price resulting from this intervention, along with geographically distributed measures of cocaine addiction, to assess any evidence of the drug price/crime effect.

The magnitude of this effect will depend heavily on the elasticity of drug demand. Casual drug users may be price sensitive, though users with addiction problems are likely much less so. Saffer and Chaloupka (1999) find evidence of extremely price inelastic demand for cocaine, estimating it at -0.28. This stands in contrast to measures of price elasticity of methamphetamine which are found to be substantially higher at -1.766 even among methamphetamine dependent users Chalmers et al. (2009). Given the strong association with cocaine use and criminal behavior, increases in drug price may merely place financial pressure on users of the drug, and in doing so increase their probability of committing financially motivated crime.

Cocaine use is often thought to be associated with persons of higher socioeconomic status, who are much less likely to fund a substance habit with

³Potassium permanganate, a similar chemical, can also used in the production of cocaine, and underwent the same classification much earlier, in 1989.

property crime. This may however be a misconception. Table 1 displays data from the 2006 National Survey of Drug Use and Health (NSDUH) which conducts a national survey with a wide variety of questions on drug use. When surveyed on cocaine use in the last month, unemployed persons above the age of 18 had the highest proportion of cocaine use (3.4%) compared to those with part time employment (1.3%), and and those with full time employment (1.0%). For the same question, those with less than a high school education had the highest proportion of cocaine use in the last month as well (1.4%) compared to those with a high school education (1.0%,) those with some college (1.3%), and those with a college degree (0.7%). Further, use of crack cocaine, often associated with property crime, was still high with 0.3% of persons reporting use in the past month, and 2.1% of unemployed respondents reporting use in the past month. Both the proportion of users and their demographic characteristics lend credence to the idea that disruptions in the cocaine market could plausibly impact crime rates.

Data

The data used in this analysis come from a variety of sources. Crime data are pulled from the FBI's Uniform Crime Reports (UCR), which collects known index crimes from police agencies across the United States. They include measures of overall property and violent crime, as well as the individual crime categories that make up these aggregates. Because each individual policing agency can choose when (and even if) to report, monthly data can be exceptionally noisy and prone to measurement error. Crime measures are thus aggregated to the state year level, necessitating that this become the unit of observation. Crimes are measured in rates, specifically the number of crimes per 100,000 population. The individual

TABLE 1.
Cocaine Usage in 2006 (Percentage of Respondents NSDUH 2006)

	Cocaine	Crack Cocaine
Unemployed	3.4	2.1
Part Time	1.3	0.3
Full Time	1.0	0.2
< High School	1.4	0.5
High School Degree	1.0	0.2
Some College	1.3	0.4
College Degree	0.7	0.2

categories of crime considered are larceny, burglary, motor vehicle theft, robbery, homicide, aggravated assault, and rape.

Addiction measures are constructed using the the Substance Abuse and Mental Health Services Administrations (SAMHSA) Treatment Episode Data Set – Admissions (TEDS - A).⁴ This dataset catalogues every individual admission into a substance abuse rehabilitation center in the United States.⁵ A variety of information is collected on each individual, including geographic identifiers and substances related to their admission.

Economic controls are pulled from the Federal Reserve Economic Data (FRED) website, while population and demographic information come from the

⁴The exact measures are explained in Section 3.4.

⁵Missing data in crucial years from the District of Columbia and Arizona exclude these states from this analysis.

TABLE 2.
Pre-Treatment Summary Statistics

	Mean	Standard Deviation	Minimum	Maximum
Property Crime Rate	3499.738	858.7925	1767	5849.8
Larceny Theft Rate	2426.393	555.9958	1336	3977.2
Burglary Rate	707.0099	231.0313	309.3	1241.6
Motor Vehicle Theft Rate	366.3357	187.2067	94.5	1116
Violent Crime Rate	405.9429	181.2505	78.2	828.1
Murder Rate	4.680272	2.499123	.6	13.2
Aggravated Assault Rate	262.0337	132.4219	42.6	627
Rape Rate	33.08776	9.226726	13.9	55.5
Robbery Rate	106.1459	59.17499	6.8	257.2
Cocaine Addiction Proportion	.2636611	.1245882	.0517621	.5195239
Population	5878864	6373103	493754	3.62e+07
Percent White Population	83.47112	12.56999	24.29768	97.87047
Percent Black Population	10.77166	9.625486	.4487852	36.99424
Unemployment Rate	4.84657	1.092486	2.3	8.141666
Real Median Household Income	58103.99	8686.396	40116	79735
State Minimum Wage	5.302517	.7503486	2.65	7.35
Home Ownership	70.33571	5.098342	53.4	81.3
<i>N</i>				294

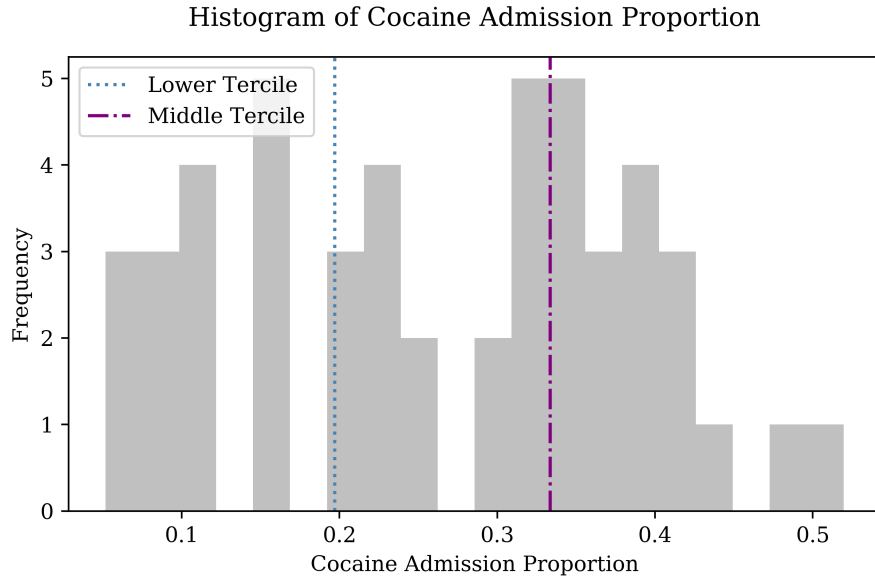
Surveillance, Epidemiology and End Results Program (SEER). These controls include the unemployment rate, minimum wage, real median household income, homeownership rate, and ethnic background for each state. The dataset is limited to the years 2000-2011 in order to minimize conflict with other similar chemical regulations occurring in the mid 1990s.

Identification

In this analysis, I wish to identify the effect of a supply side shock in the cocaine market on financially motivated crime through a measure of the intensity of cocaine addiction. The measure of cocaine addiction used is of central importance to the analysis. While one may be tempted to measure addiction using the rate of cocaine related admissions per 100,000 population, I argue that this measure in isolation is problematic. The overall number of admissions in a state can stem from a variety of factors other than the number of addicted users. Specifically, state funding for substance abuse treatment, cultural attitudes towards treatment, and the propensity for the justice system to refer offenders to treatment can all substantially impact the number of admissions. In order to avoid merely measuring the impact of overall treatment admissions, I construct a cocaine admission proportion measure, dividing the number of cocaine admissions by the number of overall admissions.⁶ This then measures the extent of cocaine use among the “addicted” population and assures that identification is not driven instead by variation in overall admissions. If the number of substance abuse admissions is not truly a measure of the extent of overall addiction in a state (and is instead a

⁶While this is the addiction measure used in a majority of econometric specifications, I could use the number of cocaine related admissions per 100,000 population and control for other, non-cocaine related admissions. I find qualitatively similar results when estimating this model, which I report in Table 3.

FIGURE 17.
 Frequency histogram of cocaine admission proportion, the central measure of sensitivity
 used in Table 4



function of the above-mentioned factors), the cocaine admission proportion is a better measure of the extent of cocaine addiction in a location.

Figures 17 and 18 show the distribution of both this measure and its components, while Figure 19 represents these measures geographically along with a representation of the pre-post treatment difference in property crime.⁷ A noticeable relationship between cocaine admission proportion and property crime differences is present.

⁷Admissions per 100,000 population and Cocaine Admissions per 100,000.

FIGURE 18.

Frequency histogram of total admissions per 100,000 population (left) and total cocaine admissions per 100,000 population (right), the denominator and numerator (respectively) of the sensitivity measure above.

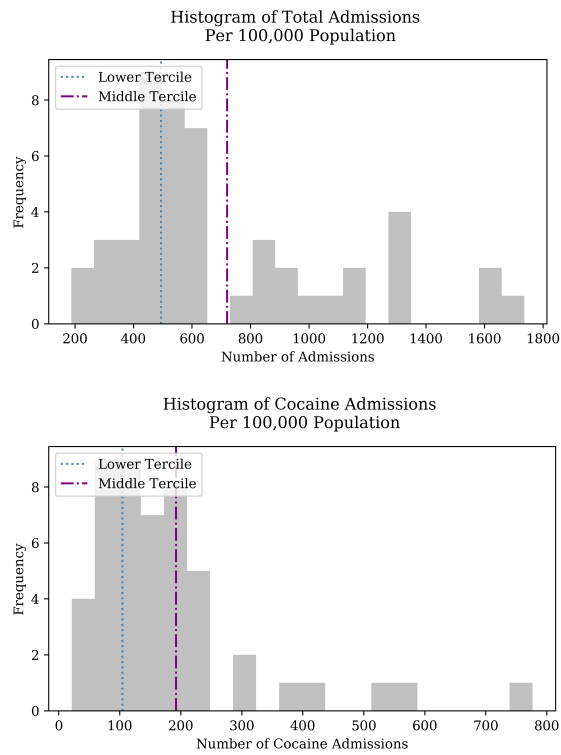


TABLE 3.
Property Crime Rate (2005 Cocaine Admissions per 100,000 population)

	(1)	(2)	(3)	(4)
	Property Crime Rate	Property Crime Rate	Property Crime Rate	Property Crime Rate
Cocaine Admissions × Post Treatment	0.767** (0.349)	0.755** (0.283)	0.821** (0.327)	0.724** (0.282)
Non Cocaine Admissions × Post Treatment	-0.231 (0.197)	-0.310* (0.174)	-0.316 (0.199)	-0.307 (0.187)
Effect Size (Pre-Treatment SD)	0.155	0.153	0.166	0.147
Impact Size (At Pre-Treatment Mean)	3.92%	3.86%	4.20%	3.70%
<i>N</i>	588	588	588	588
State & Year FE	Y	Y	Y	Y
Controls	N	N	Y	Y
State Specific Trends	N	Y	N	Y

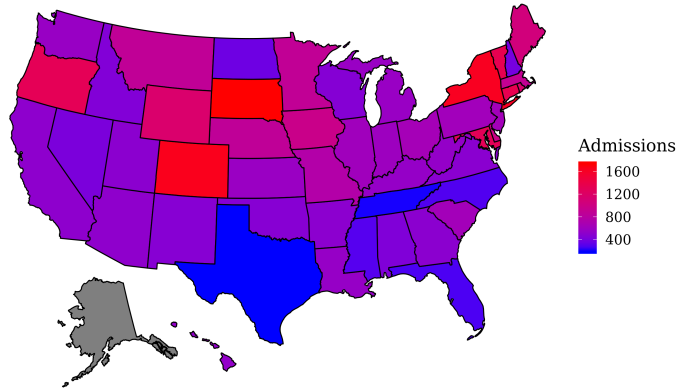
Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

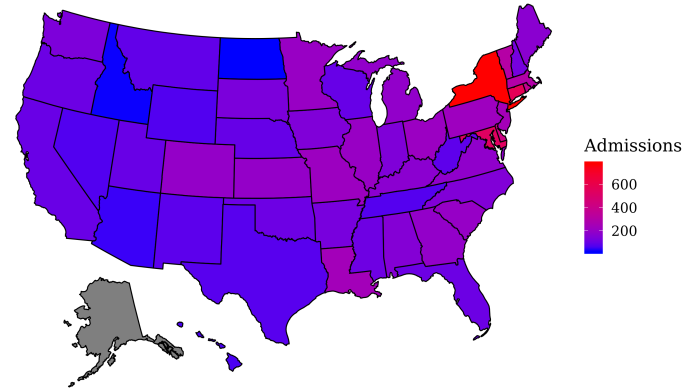
FIGURE 19.

Geographic representations of treatment and outcome, treatment admissions per 100,000 population (upper left), cocaine related admissions per 100,000 (upper right), proportion of admissions cocaine related (lower left), and pre-post difference in property crime rates (lower right).

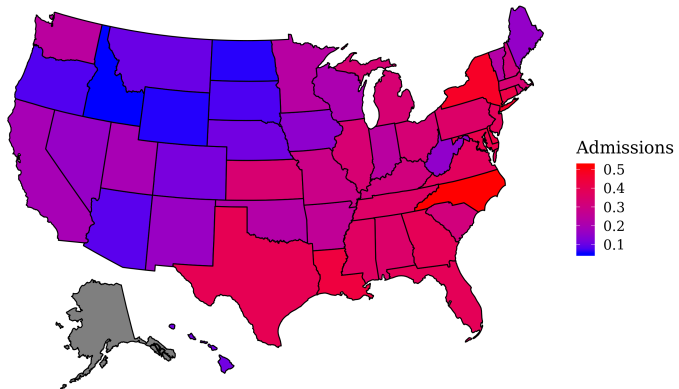
Admissions Per 100,000 Population



Cocaine Admissions Per 100,000 Population



Proportion of Cocaine Admissions



Pre Post Treatment Difference in Property Crime

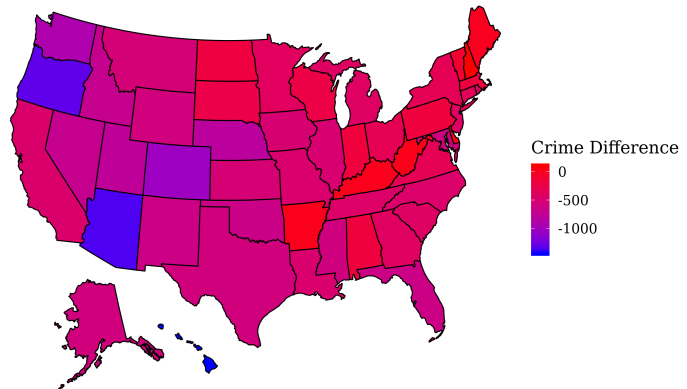


Figure 20 illustrates the benefits of this approach to identification. In the first panel of Figure 20 I show the average property crime rate⁸ across all states in the sample. No discernible change is apparent until observing the second panel, which splits states into three separate terciles based on their cocaine admission proportion (referred to as “Low addiction”, “Mid addiction”, and “High addiction”), and plots the average behavior of their property crime rates over time—these too suggest that the evolution of property crime is different across states with different rates of cocaine admission. Specifically, this suggests that states with higher cocaine addiction have higher property crime rates than those with lower cocaine addiction in the post treatment period. The three groups seem to display relatively similar pre-treatment behavior, not showing evidence of a pre-existing difference in their property crime rates.

To identify this effect econometrically, I use cocaine admission proportions in 2005, and estimate the following difference-in-differences model:

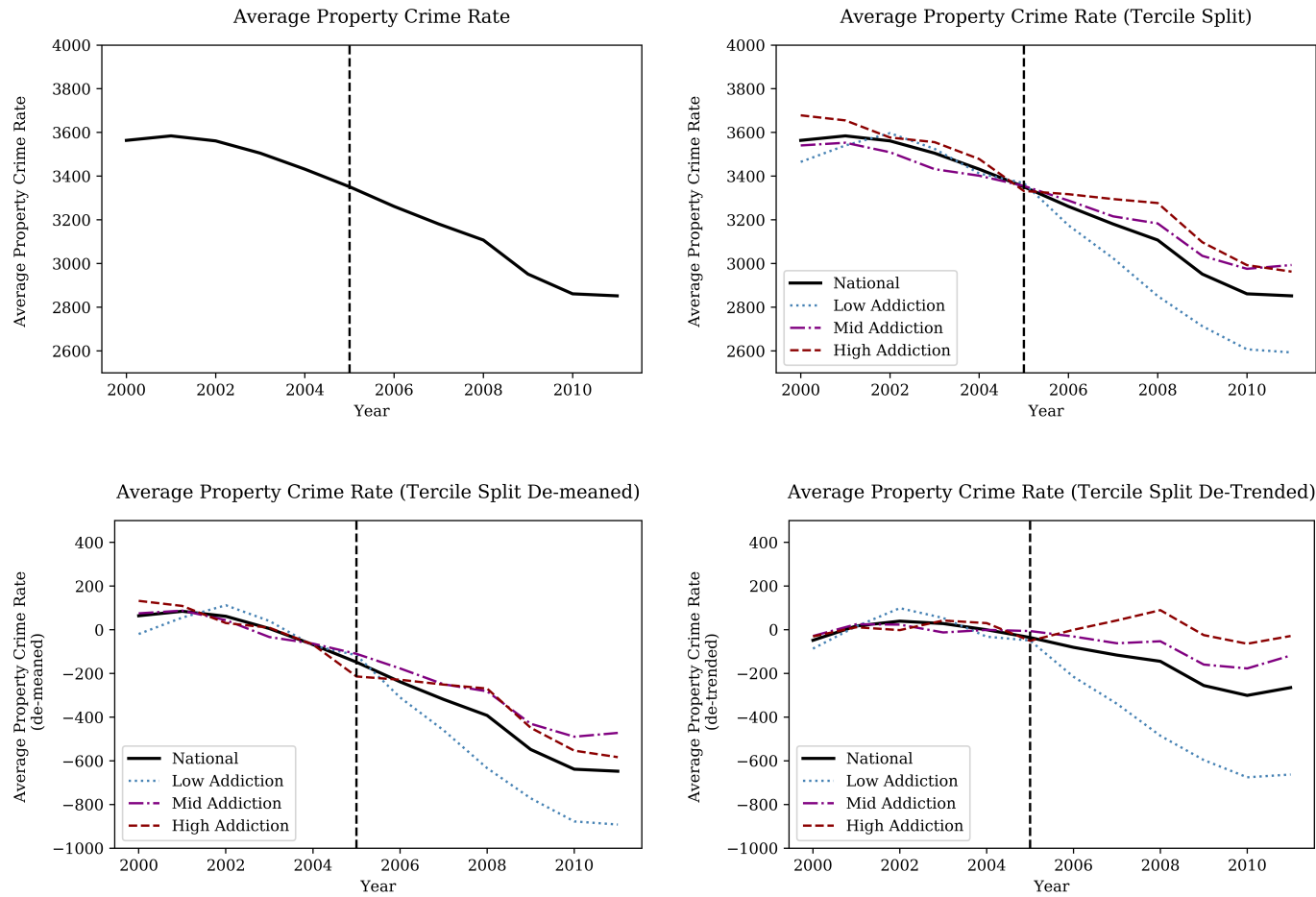
$$\text{Property Crime Rate}_{it} = \beta \cdot \text{Admission Proportion}_{i,2005} \times \text{Post Treatment}_t + \gamma \cdot \text{Admission Proportion}_{i,2005} + \theta \cdot \text{Post Treatment}_t + \alpha \cdot X_{it} + \delta_t + F_i + \epsilon_{it}, \quad (3.1)$$

where $\alpha \cdot X_{it}$ is a vector of controls (as described in section 3.3), δ_t is a year fixed effect, and F_i is a state fixed effect. The coefficient of interest is β , which measures the impact that pre-treatment cocaine addiction proportion had on property crime rates in the post treatment period. Interpretation of this coefficient requires scaling

⁸Property Crimes Per 100,000 population

FIGURE 20.

Time series of average property crime rates split across cocaine addiction terciles. National property crime rate average (upper left), national property crime rate average split across addiction terciles (upper right), national property crime rate average split across addiction terciles with pre-treatment mean subtracted (lower left), and national property crime rate average split across addiction terciles with pre-treatment trend removed (lower right)



by the specific state's cocaine admission proportion. As such, any reported impact or effect sizes will use the sample mean of this proportion.

Results

Estimating this model (as shown in Table 4) shows substantial increases in property crime in higher cocaine admission proportion states (relative to those with low cocaine admission proportion). The effect is robust across multiple specifications, including those with and without controls, and those with and without state specific time trends. The impact size measures from a 6.8% to a 9.1% increase in property crimes per 100,000 per year, a considerable rise in crime.⁹ The estimates remain effectively unchanged when weighting observations by population.

Turning to the dynamics of treatment, Figure 21 splits treatment into an event study.¹⁰ There is little to no impact of cocaine admission proportion in the pre-treatment years with a marked rise immediately following treatment that continues into 2008, and then levels off (yet stays large and persistent) for the remainder of the sample.

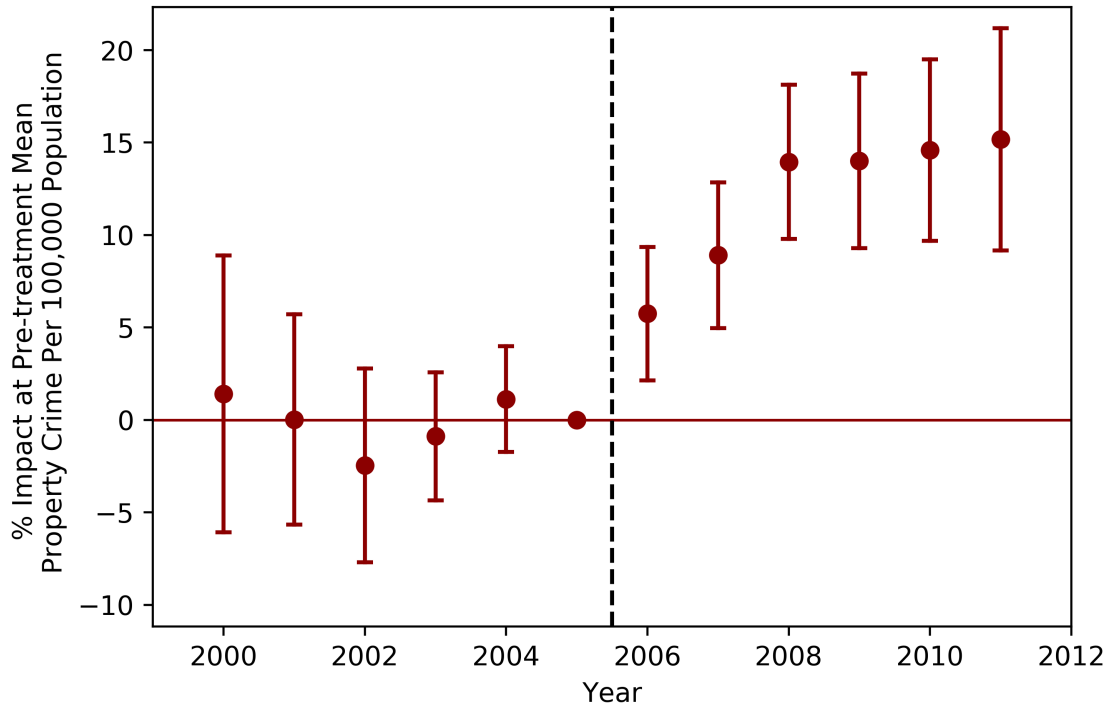
Dividing property crime into its components, I see statistically and economically significant impacts on both larceny and burglary, with the impact on motor vehicle theft being not statistically significant. When analyzing the effects on violent crime, I see no effects on the overall rate, murder rate, or assault rate. However, I do see significant increases in the robbery rate, and significant decreases in rape. The across the board increases in financially motivated crime (barring motor vehicle theft, which is still estimated as a positive impact) are

⁹Effects calculated at the pretreatment mean of both property crime and cocaine proportion.

¹⁰This specification includes state and year fixed effects, the full set of controls, and state specific time trends.

FIGURE 21.

Event study, treatment defined as 2005 cocaine admission proportion, including state and year fixed effects, state specific time trends, and controls. Estimated treatment parameters are scaled into impact size for interpretability.



consistent with a disruption in the cocaine market driving users to financially motivated crime. The disruption does not seem to impact violent crime, save for the decrease in rape. However, this could still be consistent with an addiction story. Cunningham and Shah (2017) finds that increases in prostitution stemming from inadvertent legalization of indoor sex work led to decreases in rape. If rather than larceny, some users turn to prostitution, this could still be a result of the supply-side intervention due to increased prostitution.

TABLE 4.
Property Crime Rate (2005 Cocaine Admission Proportion)

	(1)	(2)	(3)	(4)
	Property Crime Rate	Property Crime Rate	Property Crime Rate	Property Crime Rate
Cocaine Proportion × Post Treatment	991.2** (430.0)	999.3*** (317.4)	1222.2*** (415.3)	921.5*** (338.5)
Effect Size (Pre-Treatment SD)	0.292	0.294	0.360	0.271
Impact Size (At Pre-Treatment Mean)	7.38%	7.44%	9.09%	6.86%
<i>N</i>	588	588	588	588
State & Year FE	Y	Y	Y	Y
Controls	N	N	Y	Y
State Specific Trends	N	Y	N	Y

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 5.
Property Crime Rate (2005 Cocaine Admission Proportion) (Population Weighted)

	(1)	(2)	(3)	(4)
	Property Crime Rate	Property Crime Rate	Property Crime Rate	Property Crime Rate
Cocaine Proportion × Post Treatment	879.9** (429.3)	1056.4*** (336.6)	972.2** (425.0)	967.0** (363.6)
Effect Size (Pre-Treatment SD)	0.259	0.311	0.286	0.285
Impact Size (At Pre-Treatment Mean)	6.55%	7.86%	7.23%	7.1%
<i>N</i>	588	588	588	588
State & Year FE	Y	Y	Y	Y
Controls	N	N	Y	Y
State Specific Trends	N	Y	N	Y

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Further Analysis

The cocaine addiction proportion in the above specifications is calculated using only admissions from 2005, the last year before treatment. This is done in order to gain an addiction measure as close in time to treatment as possible. However, the estimated results could merely be an artifact of noisy or idiosyncratic variation in that particular year of the TEDS-A dataset. In Figure 24 the main regression is run with the addiction proportion calculated by successively adding more and more pre-treatment years to the proportion measure. This should to some extent smooth our measure of addiction. The estimated treatment effect and its statistical significance change very little, if at all, offering reassurance that results do not stem merely from the selection of the addiction rate in 2005 as treatment.

The estimated result could also stem from a factor related to using any addiction proportion in this model. Were I to construct a similar measure for a different substance and see similar statistical results, this would cast serious doubt on the mechanism driving the estimates being truly related to the cocaine market intervention. To test this, a specification identical to equation 3.1 is run, however the admission proportion is now calculated with alcohol related admissions. Small and statistically insignificant effects are found across the same variety of specifications and are presented in Table 8.

In Figure 19, the geographic variation in cocaine admission proportion is shown along side the geographic pre-post treatment change in property crime rates. To better illustrate how this relationship translates into an estimated treatment effect, these two quantities are included in a scatter plot in Figure 22. In order to ensure that the estimated effect is not based entirely on the comparison of

TABLE 6.
Property Crime Rate (2005 Cocaine Admission Proportion)

	(1)	(2)	(3)	(4)
	Property Crime Rate	Larceny Theft Rate	Burglary Rate	Motorvehicle Theft Rate
Cocaine Proportion \times Post Treatment	921.5*** (338.5)	628.7*** (214.7)	161.2** (67.49)	128.8 (78.71)
Effect Size (Pre-Treatment SD)	0.271	0.293	0.1866	0.141
Impact Size (At Pre-Treatment Mean)	6.86%	6.77%	6.01%	8.74%
N	588	588	588	588

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All specifications include controls, state and year fixed effects, and state-specific time trends

TABLE 7.
Violent Crime Rates (2005 Cocaine Admission Proportion)

	(1)	(2)	(3)	(4)	(5)
	Violent Crime Rate	Murder Rate	Assault Rate	Rape Rate	Robbery Rate
Cocaine Proportion \times Post Treatment	31.37 (49.91)	0.649 (0.933)	2.869 (34.55)	-14.63** (6.606)	42.49** (19.65)
Effect Size (Pre-Treatment SD)	0.034	.031	0.005	-0.336	0.119
Impact Size (At Pre-Treatment Mean)	1.89%	3.17%	0.27%	-11.28%	9.65%
N	588	588	588	588	588

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

All specifications include controls, state and year fixed effects, and state-specific time trends

extreme values, I slowly remove the bottom of the sample (states with low cocaine admission proportion), and re-estimate equation 3.1. This obviously removes observations and lowers the precision of the estimator, but the behavior of the estimator across these different subsets should give insight on the estimator's reliance on low values of admission proportion. While the statistical significance of the estimator does not withstand the entire procedure, it stays very similar in sign and magnitude. Estimates and confidence intervals are shown in Figure 23.

FIGURE 22.

Pre-post property crime differences plotted against 2005 proportion of admissions cocaine related.

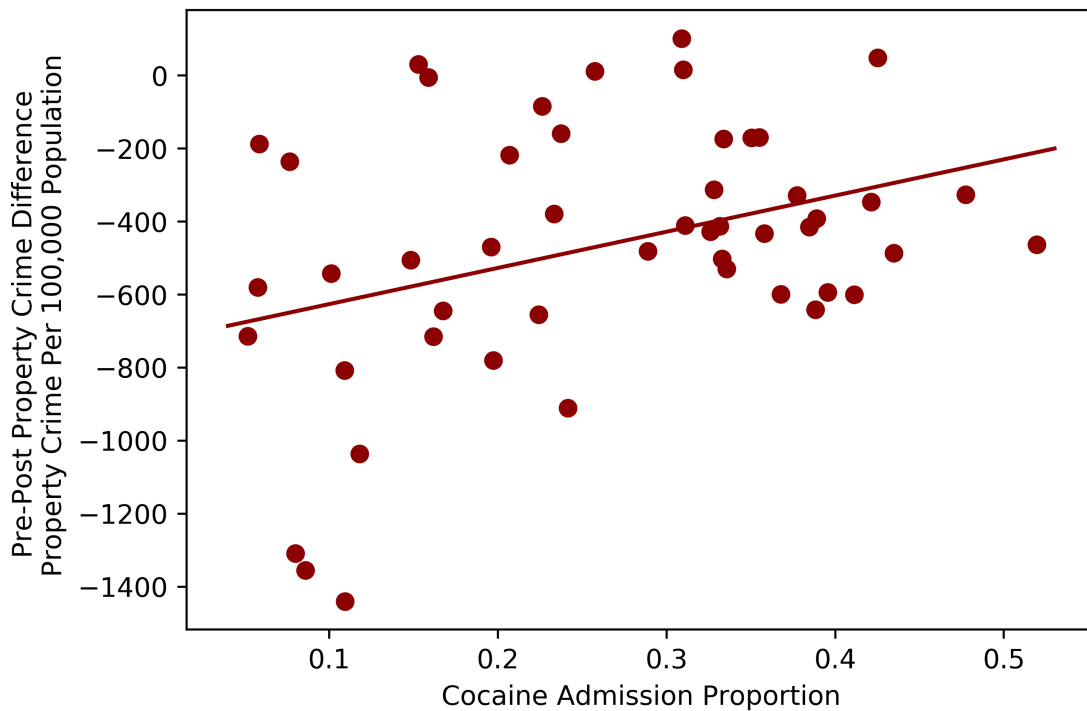


FIGURE 23.

Gradual Removal of Low Addiction States, coefficient estimate vs. minimum cocaine admission proportion, with number of observations represented in the bar graph

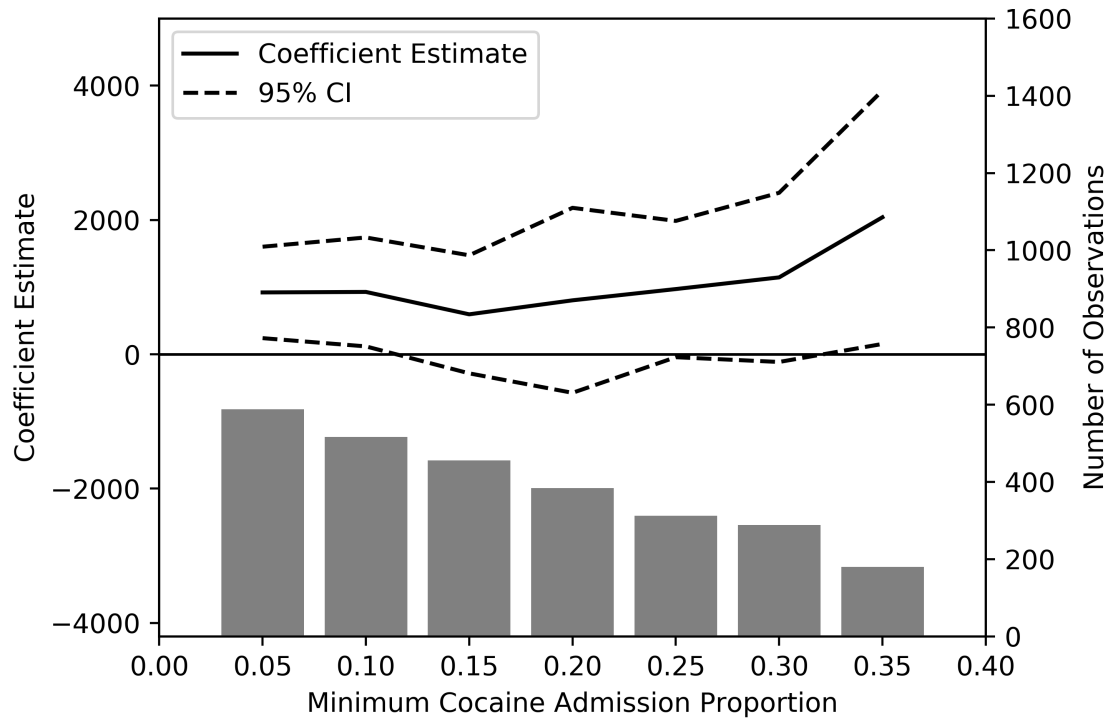


FIGURE 24.

Treatment effect estimates from specification calculating treatment with successively more pretreatment years.

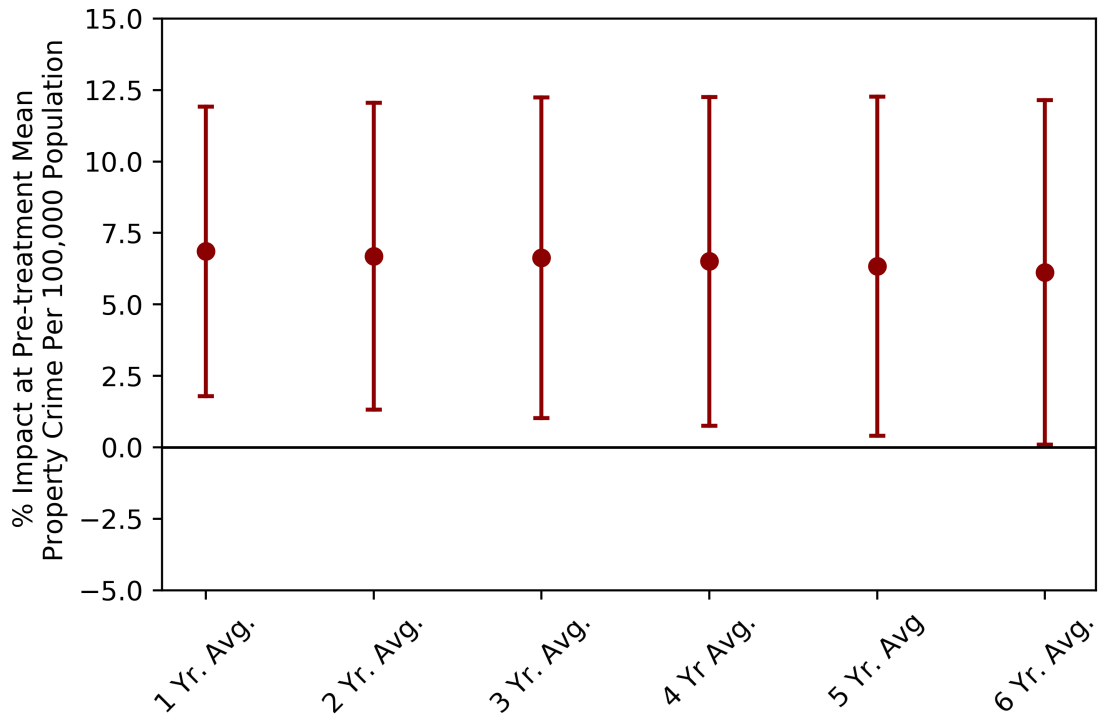


TABLE 8.
Property Crime Rate (2005 Alcohol Admission Proportion)

	(1)	(2)	(3)	(4)
	Property Crime Rate	Property Crime Rate	Property Crime Rate	Property Crime Rate
Alcohol Proportion \times Post Treatment	291.6 (529.9)	-122.5 (343.4)	106.7 (566.9)	-120.4 (385.8)
Effect Size (Pre-Treatment SD)	0.203	-0.085	0.074%	-.084%
Impact Size (At Pre-Treatment Mean)	5.13%	-2.15%	1.87%	-2.21%
<i>N</i>	588	588	588	588
State & Year FE	Y	Y	Y	Y
Controls	N	N	Y	Y
State Specific Trends	N	Y	N	Y

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Conclusion

The evidence presented here raises significant concerns about the continued use of supply-side drug policy. Using the estimated parameters, this intervention led to roughly 699,913 more larceny cases per year in the United States.¹¹ Using the most-conservative estimates of the cost of crime collected in McCollister et al. (2010), I find the yearly cost in larceny alone to be 563 million dollars.

The changes in property crime are large, and indicate that policies of this nature impose serious indirect costs on society. The direct costs of the policy, from every indication seemed relatively minor, as DEA reports indicated that manufacturers and distributors offered little to no resistance to the change. However, the costs associated with the drug price/crime effect are tied directly to the effectiveness of the policy itself, regardless of the manner in which it is implemented. This implies that significant costs of this type of policy are unavoidable.

Mitigation may be possible, through either law enforcement activity and vigilance in areas most likely to be affected, or increased support and treatment for addicts leading up to, or in the wake of a serious supply side intervention. Generally, much of the negative consequences of illegal drug use could be avoided by decreasing the number of addicted users, and would diminish the negative effects of existing supply side policy. As it stands, merely placing pricing pressure alone does not seem to achieve this goal, and merely perpetuates the criminality of these individuals, with major costs spilling over to the rest of society. Moving forward, a

¹¹There were roughly 6.78 million larceny cases reported in the UCR in 2005, the last year before the treatment.

re-emphasis on, and re-allocation of resources to demand side drug policy seems a more effective approach.

CHAPTER IV

BIAS REDUCTION THROUGH VARIABLE SELECTION

I would like to acknowledge Glen Waddell, who contributed to the computational simulations and figure making, as well as the writing and presentation of the paper. I developed to the procedure explained in the paper, contributed to the computational simulation and figures, as well as the writing and presentation of the paper.

Introduction

In the policy evaluation literature, omitted-variable bias is classically described as the omission of some x_i that correlates both with treatment assignment ($D_i = 1$) and with y_i , such that its omission from a model of y_i leads to the effect of x_i loading on to treatment assignment. Moreover, we fail to retrieve an unbiased estimate of treatment to the extent $\text{corr}(x_i, \epsilon_i) \neq 0$ and $\text{corr}(x_i, D_i) \neq 0$. Variables that meet these conditions are generally referred to as confounders.

It would suffice, of course, for $\text{corr}(x_i, \epsilon_i)$ or for $\text{corr}(x_i, D_i)$ to be zero—in theory we well understand that our omitted-variable bias (OVB) suspicions subside with either. Yet, we note the heavy weighting in practice toward establishing only that $\text{corr}(x_i, D_i) = 0$ before proceeding to consider causal interpretations of $\hat{\beta}$. Researchers report “balance tests,” for example, as demonstrations of there being no significant differences in observable covariates across treatment and control groups. To establish that such balance exists (fingers crossed) is typically the entry fee paid by all empirical micro-economists. If race and gender do differ in levels between the treated and control group, for example, and maybe education

a little bit, we include those factors in what follows. If they don't, we often throw them in anyway, with a nod to prior literature, or to theory, or to something. As a notion, though, we are hopeful that not many observable factors differ, as it breeds suspicion that unobservables may also differ systematically between treatment and control, a problem from which there is little comeback. Random assignment of treatment assuages this worry, of course. However, even then we practice similar traditions.¹

We suggest that a second, complementary practice be adopted as part of this tradition. Clearly, we should be convinced of the causal interpretation available whether it is through convincing ourselves that $\text{corr}(x, D) = 0$ or that $\text{corr}(x_i, \epsilon_i) = 0$. With the support of that logic, then, we recommend the use of simple and approachable variable-selection techniques, focusing on the explanatory power of potential covariates on the outcome in question. If certain covariates (or functions covariates) are shown to be predictive of y_i , they should be included in estimation—in expectation, this can increase the precision with which we estimate the effect of treatment, for example. To the contrary, if one focuses solely on the correlation of treatment with covariates—as one tends to in balance-test justifications for covariate selection—including a variable for which $\text{corr}(x_i, D_i) \neq 0$ but $\text{corr}(x_i, \epsilon_i) = 0$ will needlessly *decrease* the precision with which we estimate the effect of treatment. Regardless, the somewhat haphazard approach to covariate selection—including a set of covariates whether or not they are balanced across treatment and control—can easily be improved upon.

In the end, with the assistance of some machinery, we will demonstrate that even biases that originate in the unobservable component of the data-generating

¹ For more on balance tests—largely regarding their inappropriate use, frankly—see Mutz et al. (2019); Begg (1990); Senn (1994).

process are at times partially correctable, making this approach to model selection informative as a sensitivity exercise, if not profoundly powerful as a diagnosis exercise.

In Section 4.2 we consider the methods and the computational apparatus we have in mind. In Section 4.3 we consider several different simulated environments, with known properties that allow us to walk through the technique in ways that applied researchers might be inclined to think of these sorts of problems. We then offer briefly concluding thoughts.

Model selection

There are many approaches to model selection, with procedures and criteria such as likelihood-based selection criteria like the Bayesian Information Criterion (BIC) (Schwarz, 1978) and Akaike Information Criterion (AIC) (Akaike, 1998), to penalized-regression methods like Lasso (Tibshirani, 1996; Shortreed and Ertefaie, 2017) and Ridge regressions. Due to its widespread use, we focus on the Bayesian Information Criterion.²

The BIC is increasing in the number of variables included in the model, k , and is decreasing in fit, and takes the form

$$BIC = n \cdot \ln \left(\frac{RSS}{n} \right) + k \cdot \ln(n) ,$$

² Ertefaie et al. (2018) demonstrates a technique that considers outcome and treatment assignment simultaneously.

where n is the sample size and RSS is the residual sum of squares—lower values of BIC are preferred. In general, such a rule is an actionable procedure by which a variable must justify its inclusion in the model by adding *enough* fit.³

In our simulated environments below, we will be running many iterations, recording how often covariates and their interactions are chosen for inclusion. Of course, running the procedure in practice produces a list of covariates and interactions that are to be included in the predictive model—a binary indication—having evaluated all possible models and arriving at the one producing the lowest BIC.⁴

Simulations

Below, we will consider environments in which we have engineered omitted-variable biases that originate in the observable domain—omitting a relevant interaction of covariates, omitting a higher-order polynomial, and then a combination of these two. We then consider bias originating in the unobservable domain, where we again reflect on the ability to control meaningfully for variation in outcomes that better enables the identification of the treatment parameter, despite originating in the unobservable component.

In each of the simulated environments we consider here, one can imagine the following operating as somewhat of a baseline data-generating process (DGP). In

³ While this is generally perceived as a need for a variable to be sufficiently predictive of y , many criteria of this form exist—AIC and adjusted- R^2 are other variants, but differ in the magnitude of their penalty for variable inclusion. BIC has the largest penalty of the three, and is thereby set up to choose the most parsimonious model.

⁴ While our simulated environments do not suffer from this problem, we imagine that in some data-generating processes the set of all feasible models will be too large to be computationally feasible. In such instances, we recommend forward or backward selection procedures, a stepwise process that considers only a subset of potential models.

each setting, we run 2,500 iterations, drawing samples of 2,000 units each time. In each of those 2,500 iterations, we assign treatment randomly to half the sample, and randomly draw covariates x_k from joint normals with means μ_k , variances σ_k^2 , and known correlations. For example, where true treatment is captured in τ , the DGP may be

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon ,$$

having chosen parameters of β_k , μ_k , σ_k^2 and a variance-covariance matrix. That variance-covariance matrix will be key, and with each scenario we will be explicit about our definition of those relationships. Importantly, though, in every simulated environment we produce, all co-variates pass a traditional balance test—more than x_k being equal in mean across treatment and control, but actually being equivalent distributions. For example, in Figure 25 we reproduce the full distributions of each covariate in a baseline DGP, having set $\beta_k = 1$, $\mu_k = 0$, $\sigma_k^2 = 1$, and $cov(x_j, x_k) = 0 \ \forall \ j \neq k$. As prescribed, treatment is random with respect to covariates.

OVB originating in non-linear transformations of observables

Omitted interactions

In Figure 26, we consider a case of bias entering through an omitted interaction—we'll use x_2 and x_3 for this—that correlates differently among treated units. In each iteration of this setting we assign treatment randomly to half the sample, and randomly draw covariates x_1 through x_4 from joint normals with means $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$, and correlations as illustrated in Panel A—in essence, we keep things clean other than to introduce some correlation among treated units in the interaction of x_2 and x_3 , through which

the bias enters. The DGP is then,

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \beta_{23} x_2 x_3 + \epsilon ,$$

where $\tau = 1$, $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$, and $\beta_{23} = 1$. With all co-variates passing a balance test, the naive model we imagine running is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon .$$

In Panel B of Figure 26, we summarize 2,500 iterations of this DGP—in particular, the frequencies with which variable selection has chosen each level and potential interaction to be included in fitting a model of y . For ease, we’ve highlighted the “offending” interaction that the BIC procedure has (rightly) chosen to include in the model of y , 2,500 of 2,500 times—here, that variable is $x_2 x_3$. Notably, no other interaction has been included more often than 19 times.

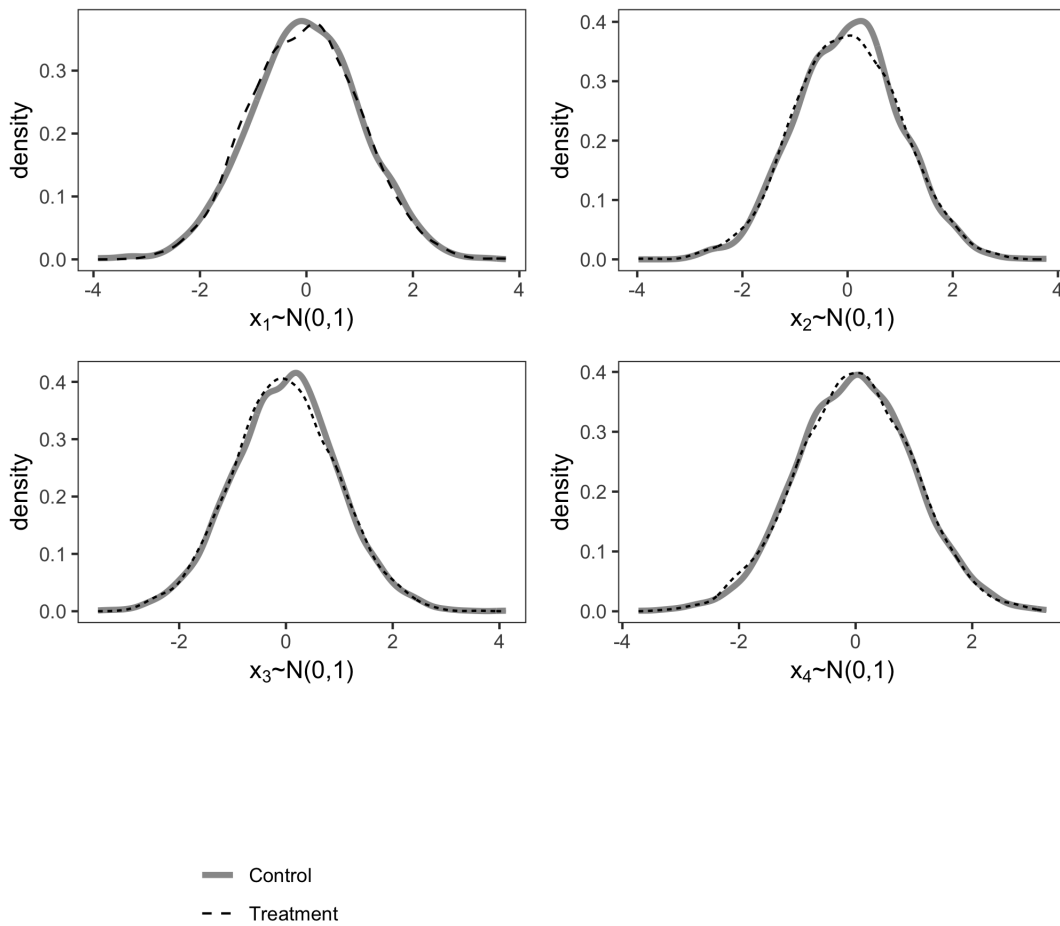
In Panel C, we produce two kernel densities, one capturing the collection of $\hat{\tau}$ identified in iterations of the naive model, and the other capturing the collection of $\hat{\tau}$ identified in iterations of the model that includes the variables selected by BIC. As we’ve suggested, in every case, this selection includes relaxing the implicit restriction that $\beta_{23} = 0$, estimating a parameter on x_{23} . In the small number of iterations (25) where variable selection has included other interactions, we do include them when they are chosen. Given the DGP we’ve engineered, the naive model retrieves, as it should, an estimated treatment parameter that is biased upward. However, following the prescriptions of the variable-selection model leads to $\hat{\tau}$ that are centered around its true value ($\tau = 1$). The a parameter is also estimated more precisely.

FIGURE 25.
 Distributional co-variate balance tests: Baseline data-generating process

In each iteration we assign treatment randomly to half the sample, and randomly draw covariates x_1 through x_4 from joint normals with means $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$, and $cov(x_j, x_k) = 0 \ \forall \ j \neq k$. The DGP is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon,$$

where $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$.



Omitted higher-order polynomials

In Figure 27, we consider a variant of omitted interaction—the bias entering through an omitted non-linearity (x_1^2) that varies differently among treated units. In each iteration of this setting the DGP is given by

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \beta_{11} x_1^2 + \epsilon ,$$

where $\beta_{11} = .1$, and all other parameters are unchanged. Again all co-variates pass a balance test, so the naive model we imagine estimating is $y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon$.

In Panel B of Figure 27, we summarize 2,500 iterations of the DGP and model, and the associated frequencies with which variable selection has chosen each level and potential interaction to be included in fitting a model of y . We’ve again highlighted the “offending” omission, that the BIC procedure has chosen to include in the model 2,500 of 2,500 times. (Here, no other interaction has been included more often than 25 times.) In Panel C we produce kernel densities for the naive and for the “preferred” model identified through variable selection—in most instances, this simply includes adds x_1^2 , but in some instances it also added other interactions. Where the naive model retrieves an inflated notion of the efficacy of treatment, following the variable-selection prescription has produced an unbiased estimate of $\hat{\tau} = 1$. As before, precision around this parameter has also increased.

Multiple omitted variables

In Figure 28, we consider bias entering through multiple omitted variables ($x_1 x_2$ and x_4^2) that vary differently among treated units. While we’re at it, we also muddy up the variance-covariance matrix somewhat, which we report in Panel A—we note that correlations differ for treated and control groups. In each iteration of

FIGURE 26.

Bias entering through an omitted interaction (of x_2 and x_3), which correlates differently among treated units

In each iteration we assign treatment randomly to half the sample, and randomly draw covariates x_1 through x_4 from joint normals with means $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$, and correlations as illustrated in Panel A. The DGP is then

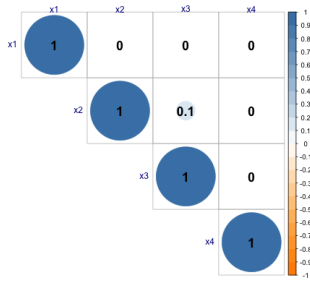
$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \beta_{23} x_2 x_3 + \epsilon,$$

where $\tau = 1$, $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$, and $\beta_{23} = 1$. With all co-variates passing a balance test, the naive model estimated is then

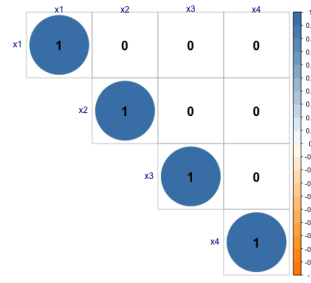
$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon.$$

Panel A: Covariate correlation

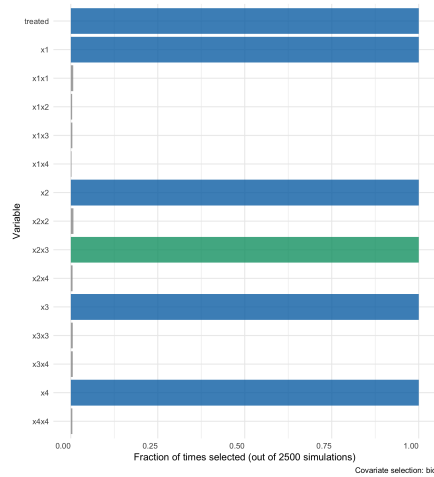
(i) Treated observations



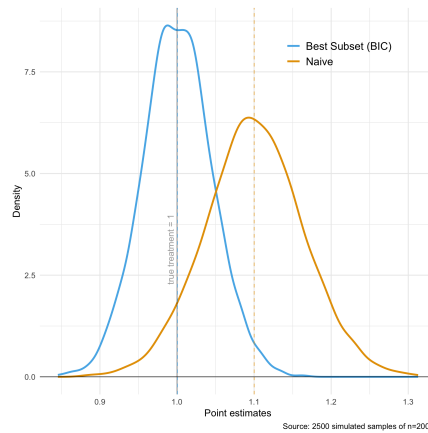
(ii) Control observations



Panel B: Inclusion probabilities



Panel C: Treatment-effect estimates



this setting, the DGP is given by

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \beta_{12} x_1 x_2 + \beta_{44} x_4^2 + \epsilon ,$$

where $\beta_{12} = \beta_{44} = 1$, and all other parameters are unchanged. Despite the additional noise and complexity, variable selection has chosen to include $x_1 x_2$ and x_4^2 , with no other interaction being picked up more often than 25 times. In Panel C, the kernel densities identify that the inclusion of these interactions corrects the OVB present in the naive model.⁵

OVB through unobservables

In figures 29 and 30, we consider biases entering through *unobservable* components. In these settings, we again assign treatment randomly to half the sample, and randomly draw covariates from joint normals. However, here we draw a total of five— x_1 through x_5 .⁶ In the DGP we adopt, we allow for a level effect of x_5 , and its interaction with other covariates in the model,

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^5 \beta_k x_k + \sum_{j=1}^4 \beta_{j5} x_j x_5 + \epsilon .$$

⁵ Vansteelandt et al. (2012) address the case in which there is a large number of confounders that each have small predictive power in y . In this case, the joint exclusion of these many small contributors can lead to bias. To guard against this, one could, for example, force the inclusion of the full set of covariates linearly, and then perform variable selection only on the higher-order terms—this would result in a “no worse than” position compared to the default approach that is absent variable-selection procedures.

⁶ We parameterize as we have before, with means $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$, variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = 1$, and correlations (all off-diagonals are nonzero) as illustrated in Panel A of Figure ???. We also set $\tau = 1$, $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 1$, and set $\beta_{15} = \beta_{25} = \beta_{35} = \beta_{45} = 1$.

FIGURE 27.

Bias entering through an omitted non-linearity (x_1^2) that varies differently among treated units

In each iteration we assign treatment randomly to half the sample, and randomly draw covariates x_1 through x_4 from joint normals with means $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$ among control observations and $\sigma_1^2 = 2$, and $\sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$ among treated observations. Correlations of zero for both treatment and control observations. The DGP is then

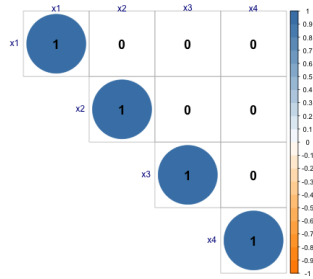
$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \beta_{11} x_1^2 + \epsilon,$$

where $\tau = 1$, $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$, and $\beta_{11} = .1$. With all co-variates passing a balance test, the naive model estimated is then

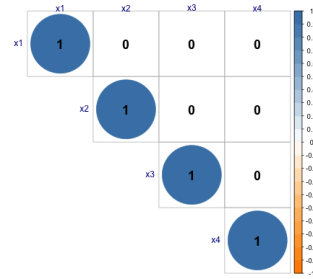
$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon.$$

Panel A: Covariate correlation

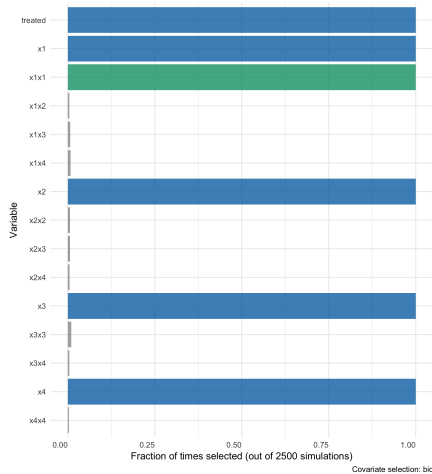
(i) Treated observations



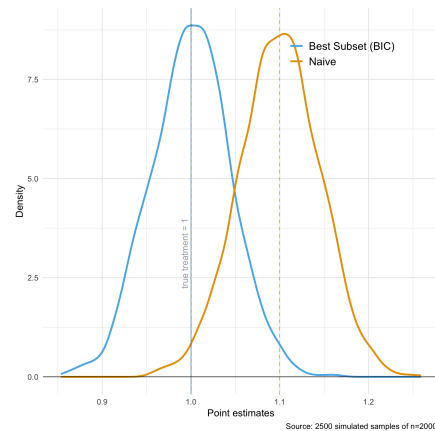
(ii) Control observations



Panel B: Inclusion probabilities



Panel C: Treatment-effect estimates



Supposing that x_5 is unobservable to the econometrician, however, we've created a setting in which the naive model remains $y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon$. (Notably, all co-variates— x_5 among them—still pass a balance test.) Moreover, we're now in a setting in which the variable-selection procedure, which is also restricted to the observables, cannot weight x_5 or its interactions directly.

In Figure 29 we hardwire into the data-generating process unambiguously positive bias. As x_5 and ϵ correlate differently in treated and control units, the estimate of τ is biased up. Yet, even though x_5 is unobservable, it's influence in y is partially estimable—given its correlation with observable covariates, the variable-selection routine better explains variation in y with weight on interactions of the observables. Moreover, the same source of variation that plagues the estimation of $\hat{\tau}$ —that x co-vary differently among treated and control units—then allows for the absorption of that variation so to not load onto $\hat{\tau}$. In Figure 30 we instead hardwire into the DGP an unambiguously negative bias—variable selection partially corrects this bias.

In Figure 31 we consider a mix of offsetting biases, that net out to zero (or come close, on average). In this case, we see evidence of gains in precision. However, inclusion of a variable need not cancel offsetting biases equally, which can result in $\hat{\tau}$ moving away from τ (Steiner and Kim, 2016). While direct diagnosis of this problem is impossible, were different parameter estimates evident across this and baseline specifications, one would imagine a certain lack of confidence in having identified treatment.

FIGURE 28.

Bias entering through multiple omitted interactions (x_1x_2 and x_4^2) that vary differently among treated units

In each iteration we assign treatment randomly to half the sample, and randomly draw covariates x_1 through x_4 from joint normals with means $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$, and correlations as illustrated in Panel A. The DGP is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \beta_{12} x_1 x_2 + \beta_{34} x_3 x_4 + \epsilon,$$

where $\tau = 1$, $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$, and $\beta_{12} = \beta_{44} = 1$. With all co-variates passing a balance test, the naive model estimated is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon.$$

Panel A: Covariate correlation

(i) Treated observations

(ii) Control observations



Panel B: Inclusion probabilities

Panel C: Treatment-effect estimates

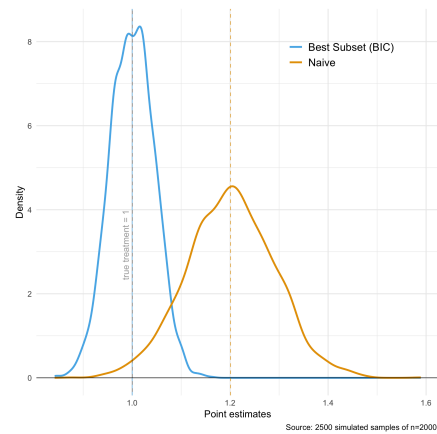
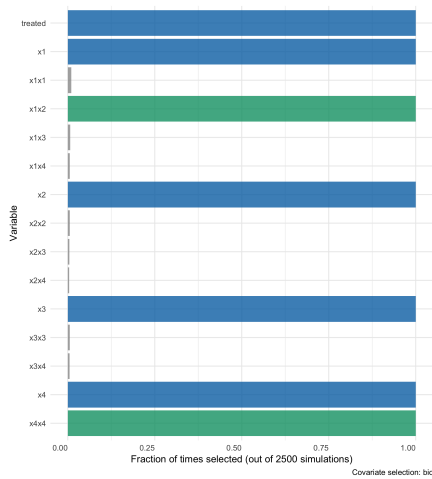


FIGURE 29.

Variable selection can reduce bias entering through unobservables: Positive bias

In each iteration we assign treatment randomly to half the sample, and randomly draw covariates x_1 through x_5 from joint normals with means $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$, variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = 1$, and correlations as illustrated in Panel A. The DGP is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^5 \beta_k x_k + \sum_{k=1}^4 \beta_{k5} x_k x_5 + \epsilon,$$

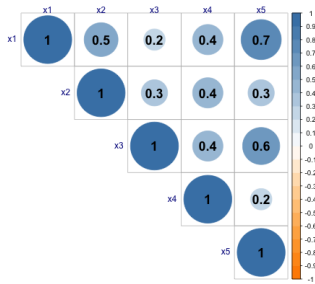
where $\tau = 1$, $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 1$, and $\beta_{15} = \beta_{25} = \beta_{35} = \beta_{45} = 1$. As x_5 is assumed to be unobservable to the econometrician, the naive model estimated is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon.$$

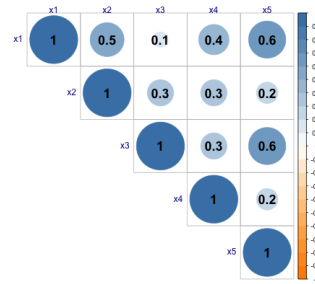
Notably, all co-variables— x_5 among them—pass a balance test.

Panel A: Covariate correlation

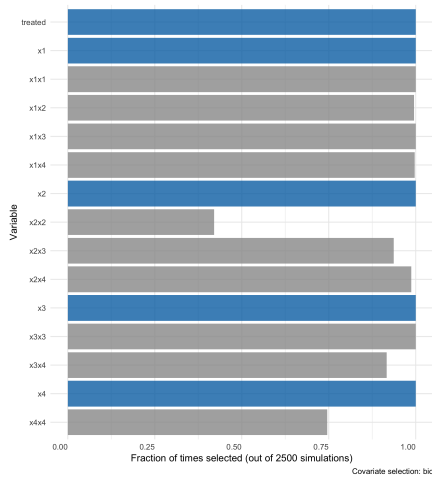
(i) Treated observations



(ii) Control observations



Panel B: Inclusion probabilities



Panel C: Treatment-effect estimates

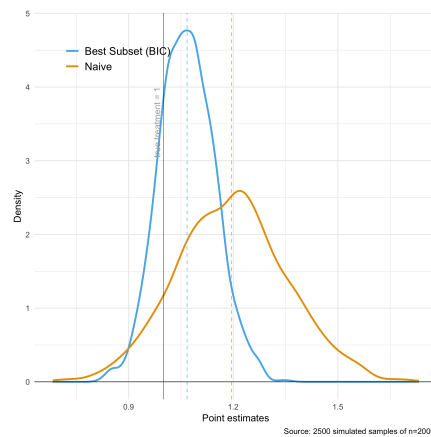


FIGURE 30.

Variable selection can reduce bias entering through unobservables: Negative bias

In each iteration we assign treatment randomly to half the sample, and randomly draw covariates x_1 through x_5 from joint normals with means $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$, variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = 1$, and correlations as illustrated in Panel A. The DGP is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^5 \beta_k x_k + \sum_{k=1}^4 \beta_{k5} x_k x_5 + \epsilon,$$

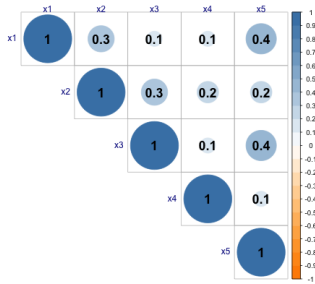
where $\tau = 1$, $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 1$, and $\beta_{15} = \beta_{25} = \beta_{35} = \beta_{45} = 1$. As x_5 is assumed to be unobservable to the econometrician, the naive model estimated is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon.$$

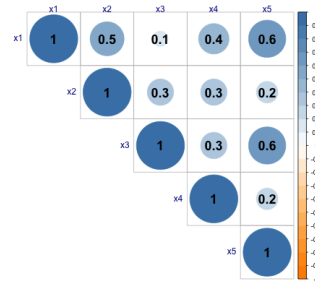
Notably, all co-variables— x_5 among them—pass a balance test.

Panel A: Covariate correlation

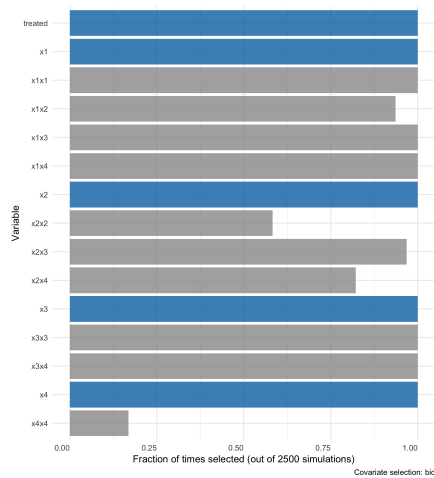
(i) Treated observations



(ii) Control observations



Panel B: Inclusion probabilities



Panel C: Treatment-effect estimates

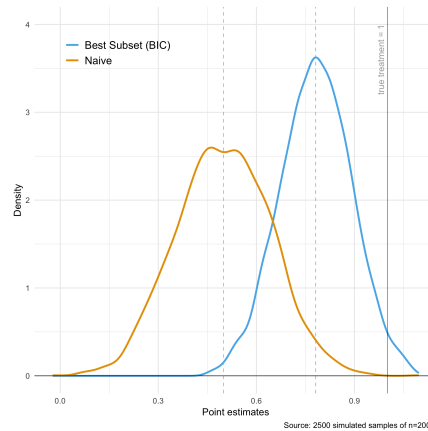


FIGURE 31.

Where there is no bias, variable selection increases precision

In each iteration we assign treatment randomly to half the sample, and randomly draw covariates x_1 through x_5 from joint normals with means $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 0$, variances $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = 1$, and correlations as illustrated in Panel A. The DGP is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^5 \beta_k x_k + \sum_{k=1}^4 \beta_{k5} x_k x_5 + \epsilon,$$

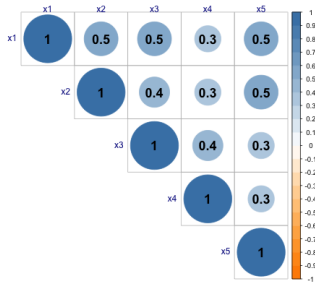
where $\tau = 1$, $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 1$, and $\beta_{15} = \beta_{25} = \beta_{35} = \beta_{45} = 1$. As x_5 is assumed to be unobservable to the econometrician, the naive model estimated is then

$$y = \beta_0 + \tau \mathbb{1}(Treated) + \sum_{k=1}^4 \beta_k x_k + \epsilon.$$

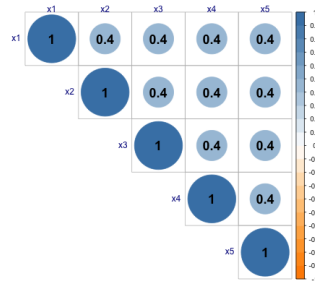
Notably, all co-variables— x_5 among them—pass a balance test.

Panel A: Covariate correlation

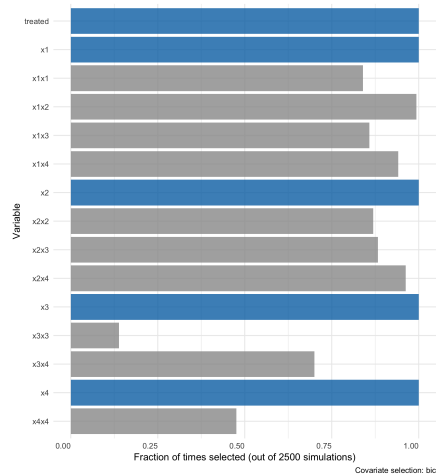
(i) Treated observations



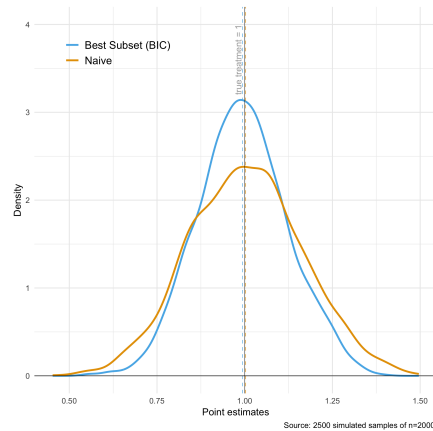
(ii) Control observations



Panel B: Inclusion probabilities



Panel C: Treatment-effect estimates



Conclusion

As applied researchers, we are surprisingly unguided in our approach to variable selection. However, available procedures are low-cost, with minimal risks, and can 1) identify sources of bias due to functions of observables, 2) increase precision in estimated treatment effects by more-efficiently conditioning of outcomes, or 3) to the extent they are correlate systematically with functions of observables, proxy for bias-inducing variation originating in the unobservables.

CHAPTER V

CONCLUSION

In chapter 1, a technique is proposed for improving estimation of synthetic controls, while in chapter 2, I find an important unintended consequence of supply side drug policy. Finally, in chapter 3 a technique is proposed for improving variable selection in regressions seeking to find treatment effects. Together, these papers contribute to the discipline of causal inference in economics.

REFERENCES CITED

- Abadie, A., Diamond, A., and 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(490):493–505.
- Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.
- Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132.
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer.
- Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2017). Matrix completion methods for causal panel data models.
- Athey, S. and Imbens, G. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.
- Begg, C. B. (1990). Significance tests of covariate imbalance in clinical trials. *Controlled clinical trials*, 11(4):223–225.
- Castillo, J., Mejía, D., and Restrepo, P. (2014). Scarcity without leviathan: The violent effects of cocaine supply shortages in the mexican drug war.
- Chalmers, J., Bradford, D., Jones, C., et al. (2009). How do methamphetamine users respond to changes in methamphetamine price? *BOCSAR NSW Crime and Justice Bulletins*, page 16.
- Cunningham, J. K., Callaghan, R. C., and Liu, L.-M. (2015). Us federal cocaine essential (precursor) chemical regulation impacts on us cocaine availability: an intervention time–series analysis with temporal replication. *Addiction*, 110(5):805–820.
- Cunningham, S. and Shah, M. (2017). Decriminalizing indoor prostitution: Implications for sexual violence and public health. *The Review of Economic Studies*, 85(3):1683–1715.
- Cunningham, S. and Shah, M. (2018). Decriminalizing indoor prostitution: Implications for sexual violence and public health. *The Review of Economic Studies*, 85(3):1683–1715.

- Dobkin, C. and Nicosia, N. (2009). The war on drugs: methamphetamine, public health, and crime. *American Economic Review*, 99(1):324–49.
- Dobkin, C., Nicosia, N., and Weinberg, M. (2014). Are supply-side drug control efforts effective? evaluating otc regulations targeting methamphetamine precursors. *Journal of Public Economics*, 120:48–61.
- Doudchenko, N. and Imbens, G. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Working Paper 22791, National Bureau of Economic Research.
- Ertefaie, A., Asgharian, M., and Stephens, D. A. (2018). Variable selection in causal inference using a simultaneous penalization method. *Journal of Causal Inference*, 6(1).
- Evans, W. N., Garthwaite, C., and Moore, T. J. (2018). Guns and violence: The enduring impact of crack cocaine markets on young black males. Technical report, National Bureau of Economic Research.
- Ferman, B. and Botosaru, I. (2017). On the role of covariates in the synthetic control method.
- Ferman, B. and Pinto, C. (2016). Revisiting the synthetic control estimator.
- Ferman, B., Pinto, C., and Possebom, V. (2018). Cherry picking with synthetic controls.
- Hansen, B., Miller, K., and Weber, C. (2017). Drug trafficking under partial prohibition: Evidence from recreational marijuana.
- Kaul, A., Klößner, S., Pfeifer, G., and Schieler, M. (2018). Synthetic control methods: Never use all pre-intervention outcomes together with covariates.
- McCollister, K. E., French, M. T., and Fang, H. (2010). The cost of crime to society: New crime-specific estimates for policy and program evaluation. *Drug and alcohol dependence*, 108(1):98–109.
- Mutz, D. C., Pemantle, R., and Pham, P. (2019). The perils of balance testing in experimental design: Messy analyses of clean data. *The American Statistician*, 73(1):32–42.
- Nurco, D. N., Hanlon, T. E., and Kinlock, T. W. (1991). Recent research on the relationship between illicit drug use and crime. *Behavioral Sciences & the Law*, 9(3):221–242.
- Rothstein, J., Ben-Michael, E., and Feller, A. (2018). The role of the propensity score in the synthetic control method.

- Saffer, H. and Chaloupka, F. (1999). The demand for illicit drugs. *Economic inquiry*, 37(3):401–411.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.
- Senn, S. (1994). Testing for baseline balance in clinical trials. *Statistics in medicine*, 13(17):1715–1726.
- Shortreed, S. M. and Ertefaie, A. (2017). Outcome-adaptive lasso: Variable selection for causal inference. *Biometrics*, 73(4):1111–1122.
- Silverman, L. P. and Spruill, N. L. (1977). Urban crime and the price of heroin. *Journal of Urban Economics*, 4(1):80–103.
- Steiner, P. M. and Kim, Y. (2016). The mechanics of omitted variable bias: Bias amplification and cancellation of offsetting biases. *Journal of causal inference*, 4(2).
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Vansteelandt, S., Bekaert, M., and Claeskens, G. (2012). On model selection and model misspecification in causal inference. *Statistical methods in medical research*, 21(1):7–30.
- Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.