A Report on "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program" by Abadie et al. (2010)

Reviewer 2

February 10, 2026

v1

isitcredible.com

Disclaimer

This report was generated by large language models, overseen by a human editor. It represents the honest opinion of The Catalogue of Errors Ltd, but its accuracy should be verified by a qualified expert. Comments can be made here. Any errors in the report will be corrected in future revisions.

> I am wiser than this person; for it is likely that neither of us knows anything fine and good, but he thinks he knows something when he does not know it, whereas I, just as I do not know, do not think I know, either. I seem, then, to be wiser than him in this small way, at least: that what I do not know, I do not think I know, either.
>
> Plato, *The Apology of Socrates*, 21d

> To err is human. All human knowledge is fallible and therefore uncertain. It follows that we must distinguish sharply between truth and certainty. That to err is human means not only that we must constantly struggle against error, but also that, even when we have taken the greatest care, we cannot be completely certain that we have not made a mistake.
>
> Karl Popper, 'Knowledge and the Shaping of Reality'

Overview

Citation: Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, Vol. 105, No. 490, pp. 493–505.

URL: https://doi.org/10.1198/jasa.2009.ap08746

Abstract Summary: This article investigates the application of synthetic control methods to comparative case studies, discussing their advantages and applying them to estimate the effect of California's Proposition 99 tobacco control program. The study demonstrates that following Proposition 99, tobacco consumption fell markedly in California relative to a comparable synthetic control region.

Key Methodology: Synthetic Control Methods, comparative case studies, placebo tests, permutation inference

Research Question: What was the effect of California's Proposition 99 tobacco control program on tobacco consumption?

Summary

Is It Credible?

This article introduces the synthetic control method (SCM) as a systematic approach to comparative case studies, applying it to estimate the impact of California's Proposition 99 tobacco control program. Abadie et al. argue that SCM offers a transparent, data-driven alternative to the subjective selection of comparison units often found in social science. By constructing a "synthetic California"—a weighted average of control states that best reproduces California's pre-intervention characteristics—they estimate that Proposition 99 reduced annual per-capita cigarette sales by approximately 26 packs by the year 2000 (p. 493). Furthermore, they propose exact inferential techniques, akin to permutation tests, to demonstrate that the probability of obtaining such a result by chance is extremely low, calculated at 0.026 (p. 503).

The methodological contribution of the article is highly credible. The authors provide a formal justification for SCM, demonstrating how it generalizes the traditional difference-in-differences framework to account for time-varying unobserved confounders (p. 495). The transparency of the method is a significant strength; by explicitly reporting the weights assigned to each donor state—such as Utah, Nevada, and Montana—the authors allow readers to scrutinize the composition of the counterfactual (p. 500). The proposed "placebo studies," where the intervention is iteratively assigned to control states to generate a distribution of estimated gaps (prediction errors) for states where no intervention took place, provide a rigorous and intuitive way to assess statistical significance in small-sample settings where standard large-sample inference fails (p. 497).

However, the credibility of the specific empirical estimate—the 26-pack reduction—requires more nuance. The outcome variable is "annual per capita cigarette sales" derived from tax revenues, which the authors use as a proxy for consumption

(p. 499). Proposition 99 increased cigarette taxes, creating a financial incentive for tax avoidance. If Californians shifted purchases to neighboring states or the black market, taxed sales would fall even if consumption remained constant. The authors acknowledge this limitation, noting that tax data "are affected by cigarette smuggling," but argue that smuggling would have to be "massive" to explain the effect size (p. 501). While plausible, this defense is qualitative. Since Nevada receives a significant weight (0.234) in the synthetic control, cross-border purchasing could simultaneously lower California's sales and raise the control group's sales, artificially widening the estimated gap (p. 500).

Furthermore, the assumption of no interference between units is critical. The authors discuss the possibility that anti-smoking sentiment spilled over into control states, which would attenuate their estimate (p. 501). Regarding industry interference, they argue that tobacco companies might have diverted advertising funds *to* California to combat the new regulations, which would also bias the estimate downward. They do not, however, consider the alternative possibility: that the industry, facing a hostile regulatory environment in California, might have strategically retreated and diverted marketing resources *to* unregulated control states to maintain overall revenue. Such a "strategic retreat" would inflate sales in the synthetic control group, again exaggerating the treatment effect. Additionally, the reliance on the entire pre-intervention period (1970–1988) to select weights raises a risk of overfitting, where the model matches noise rather than structural trends. Although the authors mention the possibility of using a training/validation split to mitigate this, they do not implement it in the primary analysis (p. 496). Consequently, while the negative impact of Proposition 99 is almost certainly real given the magnitude of the divergence, the precise estimate of 26 packs likely represents an upper bound, potentially inflated by tax avoidance and unmodeled interference.

The Bottom Line

The methodological framework presented is robust and represents a significant advancement in the rigorous analysis of comparative case studies. The empirical finding that Proposition 99 significantly reduced cigarette sales is credible and supported by strong statistical evidence derived from placebo tests. However, the precise magnitude of the estimated reduction (26 packs) should be interpreted with caution, as it may be inflated by unmeasured tax avoidance behaviors and strategic market shifts that could widen the gap between California and its synthetic counterpart.

Potential Issues

Violation of the no-interference assumption: The validity of the synthetic control method rests on the assumption that the intervention in the treated unit does not affect the outcomes in the control units. The authors acknowledge that this assumption could be violated in their analysis, discussing two potential channels of interference: the spread of anti-tobacco sentiment from California to other states, and the diversion of tobacco industry advertising funds from other states to California. They conclude that both channels would likely cause lower smoking rates in the control states, which would "artificially reduc[e] the magnitude of our estimate" (p. 501). This discussion, however, omits an equally plausible interference channel that would bias the estimate in the opposite direction. Faced with a hostile regulatory environment in California, the tobacco industry could have shifted its promotional resources out of California and into the control states to bolster sales there. Such a strategic response would artificially inflate cigarette sales in the synthetic control, making the gap between it and California appear larger than the true effect of Proposition 99. By framing the potential bias as exclusively attenuating, the article may not fully address the uncertainty surrounding this critical assumption.

Validity of the outcome measure: The study's central outcome variable is "annual per capita cigarette sales" derived from state tax revenues, which serves as a proxy for the construct of interest, tobacco consumption (p. 499). The passage of Proposition 99, which included a 25-cent tax increase, created a direct financial incentive for tax avoidance behaviors such as cross-border purchasing from lower-tax jurisdictions (like Nevada, which receives 23.4% of the weight in the synthetic control) or illicit smuggling (p. 500, Table 2). Any increase in untaxed consumption would be recorded in the data as a decrease in taxed sales, potentially confounding the true effect of the policy on consumption and biasing the estimated effect away from zero. The authors acknowledge this limitation, stating that tax-revenue-based data "are

affected by cigarette smuggling across tax jurisdictions" (p. 499). They later argue that the magnitude of smuggling would have had to be "massive" to explain their large estimated effect and cite literature suggesting cross-border purchases are limited by distance (p. 501). However, this defense is qualitative and does not include a quantitative sensitivity analysis to estimate the plausible magnitude of this bias. Whether this potential measurement error significantly inflates the study's final estimate remains an important question.

Potential confounding from contemporaneous events: The analysis attributes the entire post-1988 divergence between California and its synthetic control to Proposition 99. This interpretation assumes the absence of other major, contemporaneous, California-specific shocks that could have independently affected smoking behavior. The study does not discuss the potential confounding influence of the HIV/AIDS crisis, which was particularly severe in California during this period. The late 1980s and early 1990s saw a massive increase in public health campaigns and health-conscious behavioral changes related to the epidemic. It is plausible that this broader shift in public health awareness could have contributed to the decline in smoking, independent of Proposition 99. If this health shock was more pronounced in California than in the states that comprise the synthetic control (e.g., Utah, Montana, Nevada), the model may risk misattributing some of the effect of this general health shock to the specific tobacco control policy, potentially inflating its estimated impact (p. 500, Table 2).

Potential endogeneity of policy adoption: The study treats the passage of Proposition 99 as an exogenous intervention. However, the policy was a voter-approved initiative, suggesting it was the result of an endogenous political process that likely reflected a pre-existing and growing anti-smoking sentiment within California. While the synthetic control method is designed to account for pre-existing trends by matching on a long history of the outcome variable, it may not adequately capture an unobserved, accelerating dynamic in public attitude that both drove the passage of the law

and would have independently contributed to the subsequent decline in smoking. The article demonstrates an excellent match on the pre-1988 trend, which provides a strong defense against simple trend-based endogeneity (p. 500, Figure 2). However, the more subtle possibility that a "tipping point" in public opinion occurred around 1988, causing both the law's passage and a sharp behavioral shift not predicted by the prior linear trend, remains a potential challenge to the causal interpretation.

Risk of in-sample overfitting: The procedure for constructing the synthetic control involves choosing predictor weights (the V-matrix) to minimize the Mean Squared Prediction Error (MSPE) of the outcome over the entire pre-intervention period of 1970–1988. This approach uses the same data to both train the model (select the weights) and evaluate its pre-treatment fit, which creates a risk of in-sample overfitting. An overfitted model might achieve an artificially close pre-treatment match by fitting to noise rather than to the underlying structural relationship, compromising its ability to generate a reliable counterfactual. The authors acknowledge a method to mitigate this risk, which involves splitting the pre-intervention period into separate training and validation periods, but they do not implement it in their analysis (p. 496). Without such an out-of-sample check, the possibility remains that the excellent pre-treatment fit is partly an artifact of overfitting, which could mechanically amplify the post-intervention divergence.

Scope of interpretation: The study estimates the aggregate effect of Proposition 99, which the authors describe as a multi-component program including a tax increase, funding for media campaigns, and support for local clean-air ordinances (pp. 497–498). The synthetic control method, as applied, estimates the combined effect of this policy bundle and cannot disentangle the relative contributions of its different components. This "black box" approach is a valid research goal for evaluating the program as a whole, but it limits the study's ability to provide specific guidance on which policy levers were most effective. The authors are transparent about their objective to estimate the effect of the overall program, but this feature limits the direct

policy implications of the findings.

Presentation and clerical issues: The article contains minor inconsistencies in its presentation. First, there is a persistent slippage in terminology between the measured outcome, cigarette sales, and the construct of interest, tobacco consumption (e.g., pp. 493, 500). While the authors define their outcome variable as sales-based and acknowledge the measurement issues this entails, the frequent use of "consumption" to describe the findings may overstate the evidence, which is strictly about taxed sales (p. 499). Second, there is a clerical discrepancy in the documentation of the predictor variables. The main text states that predictors were averaged over the 1980–1988 period, but the note to Table 1 on the same page specifies that "beer consumption is averaged 1984–1988" (p. 499). This minor inconsistency in the description of the procedure is not explained.

Future Research

Quantifying measurement error from smuggling: Future research could refine the outcome measure by incorporating estimates of illicit trade and cross-border shopping. By adjusting the tax-revenue data using auxiliary data on tax differentials and distance to lower-tax jurisdictions (particularly for heavily weighted controls like Nevada), researchers could place bounds on the degree to which the sales gap reflects a true reduction in consumption versus a displacement of purchases.

Modeling strategic interference: Future work could test the no-interference assumption by analyzing tobacco industry marketing expenditures during the post-intervention period. If advertising spend significantly increased in the donor states relative to national trends following Proposition 99, this would suggest a strategic displacement of industry effort. Incorporating marketing intensity as a covariate or an auxiliary outcome could help disentangle the policy effect from industry response.

Cross-validation for weight selection: To address the risk of overfitting, future applications of this method should implement the training and validation split suggested but not utilized by the authors. By selecting weights based on an initial pre-intervention period (e.g., 1970–1980) and testing the fit on a subsequent pre-intervention period (e.g., 1981–1988) before assessing the post-treatment effect, researchers can ensure the synthetic control captures a durable structural relationship rather than idiosyncratic noise.

isitcredible.com