Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program

Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program

Alberto Abadie – Harvard University and NBER
Alexis Diamond – IFC, World Bank Group
Jens Hainmueller – MIT

June 2009

Abstract

Building on an idea in Abadie and Gardeazabal (2003), this article investigates the application of synthetic control methods to comparative case studies. We discuss the advantages of these methods and apply them to study the effects of Proposition 99, a large-scale tobacco control program that California implemented in 1988. We demonstrate that following Proposition 99 tobacco consumption fell markedly in California relative to a comparable synthetic control region. We estimate that by the year 2000 annual per capita cigarette sales in California were about 26 packs lower than what they would have been in the absence of Proposition 99. Using new inferential methods proposed in this article, we demonstrate the significance of our estimates. Given that many policy interventions and events of interest in social sciences take place at an aggregate level (countries, regions, cities, etc.) and affect a small number of aggregate units, the potential applicability of synthetic control methods to comparative case studies is very large, especially in situations where traditional regression methods are not appropriate.

Alberto Abadie, John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge MA 02138. E-mail: alberto_abadie@harvard.edu. Alexis Diamond, The International Finance Corporation, 2121 Pennsylvania Avenue, Washington, DC 20433. E-mail: adiamond@ifc.org. Jens Hainmueller, Department of Political Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: jhainmueller@mit.edu. In addition, all authors are affiliated with Harvard's Institute for Quantitative Social Science (IQSS). We thank Joshua Angrist, Jake Bowers, Javier Gardeazabal, Bryan Graham, Dan Hopkins, Rustam Ibragimov, Guido Imbens, Michael Lovenheim, Paul Rosenbaum, Don Rubin, Jas Sekhon, Gary Solon, Chris Winship, and seminar participants at Harvard, Syracuse, Uppsala, the NBER Labor Meetings, and the 2006 APSA Meetings in Philadelphia for helpful comments. Funding for this research was generously provided by NSF grant SES-0350645 (Abadie). This paper was awarded the 2007 Gosnell Prize for the best work in methods presented at any political science conference during the preceding year.

Companion software developed by the authors (*Synth* package for MATLAB, R, and Stata) is available at the authors' webpages.

I. Introduction

Social scientists are often interested in the effects of events or policy interventions that take place at an aggregate level and affect aggregate entities, such as firms, schools, or geographic or administrative areas (countries, regions, cities, etc.). To estimate the effects of these events or interventions, researchers often use comparative case studies. In comparative case studies, researchers estimate the evolution of aggregate outcomes (such as mortality rates, average income, crime rates, etc.) for a unit affected by a particular occurrence of the event or intervention of interest and compare it to the evolution of the same aggregates estimated for some control group of unaffected units. Card (1990) studies the impact of the 1980 Mariel Boatlift, a large and sudden Cuban migratory influx in Miami, using other cities in the southern United States as a comparison group. In a well-known study of the effects of minimum wages on employment, Card and Krueger (1994) compare the evolution of employment in fast-food restaurants in New Jersey and its neighboring state Pennsylvania around the time of an increase in New Jersey's minimum wage. Abadie and Gardeazabal (2003) estimate the effects of the terrorist conflict in the Basque Country on the Basque economy using other Spanish regions as a comparison group.

Comparing the evolution of an aggregate outcome (e.g., state-level crime rate) between a unit affected by the event or intervention of interest and a set of unaffected units requires only aggregate data, which are often available. However, when data are not available at the same level of aggregation as the outcome of interest, information on a sample of disaggregated units can sometimes be used to estimate the aggregate outcomes of interest (like in Card, 1990, and Card and Krueger, 1994).

Given the widespread availability of aggregate/macro data (for example, at the school, city, or region level), and the fact that many policy interventions and events of interest in the social sciences take place at an aggregate level, comparative case study research has broad potential. However, comparative case study research remains limited in the social sciences, perhaps because its empirical implementation is subject to two elusive problems. First, in comparative case studies there is typically some degree of ambiguity about how comparison units are chosen. Researchers often select comparison groups on the basis of subjective measures of affinity between affected and unaffected units. Second, comparative case studies typically employ data on a sample of disaggregated units and inferential techniques that measure *only* uncertainty about the aggregate values of the data in the population. Uncertainty about the values of aggregate variables can be eliminated completely if aggregate data are available. However, the availability of aggregate data does not imply that the effect of the event or intervention of interest can be estimated without error. Even if aggregate data are employed, there remains uncertainty about the ability of the control group to reproduce the counterfactual outcome trajectory that the affected units would have experienced in the absence of the intervention or event of interest. This type of uncertainty is not reflected by the standard errors constructed with traditional inferential techniques for comparative case studies.

This article addresses current methodological shortcomings of case study analysis. We advocate the use of data-driven procedures to construct suitable comparison groups, as in Abadie and Gardeazabal (2003). Data-driven procedures reduce discretion in the choice of the comparison control units, forcing researchers to demonstrate the affinities between the affected and unaffected units using observed quantifiable characteristics. In practice, it is often difficult to find a single unexposed unit that approximates the most relevant characteristics of the unit(s) exposed to the event of interest. The idea behind the synthetic control approach is that a combination of units often provides a better comparison for the unit exposed to the intervention than any single unit alone. For example, in their study of the economic impact of terrorism in the Basque Country, Abadie and Gardeazabal (2003) use a combination of two Spanish regions to approximate the economic growth that the Basque Country would have experienced in the absence of terrorism. Card (1990) implicitly uses a combination of cities in the southern United States to approximate the evolution that the Miami labor market would have experienced in the absence of the Mariel Boatlift.

Relative to traditional regression methods, transparency and safeguard against extrapolation are two attractive features of the synthetic control method. Because a synthetic control is a weighted average of the available control units, the synthetic control method makes explicit (1) the relative contribution of each control unit to the counterfactual of interest; and (2) the similarities (or lack thereof) between the unit affected by the event or intervention of interest and the synthetic control, in terms of pre-intervention outcomes and other predictors of post-intervention outcomes. Because the weights can be restricted to be positive and sum to one, the synthetic control method provides a safeguard against extrapolation.

In addition, because the choice of a synthetic control does not require access to post-intervention outcomes, the synthetic control method allows researchers to decide on study design without knowing how those decisions will affect the conclusions of their studies. Rubin (2001) and others have advocated that the ability to make decisions on research design while remaining blind to how each particular decision affects the conclusions of the study is an important device for promoting research honesty in observational studies.

We describe a simple model that justifies the synthetic control approach. The model extends the traditional linear panel data (difference-in-differences) framework, allowing that the effects of unobserved variables on the outcome vary with time. In addition, we propose a new method to perform inferential exercises about the effects of the event or intervention of interest. The inferential exercises proposed in this article produce potentially informative inference regardless of the number of available comparison units, the number of available time periods, and the level of aggregation of the data.

We apply the synthetic control method to study the effects of California's Proposition 99, a large-scale tobacco control program implemented in California in 1988. We demonstrate that following the passage of Proposition 99 tobacco consumption fell markedly in California relative to a comparable synthetic control region. We estimate that by the year 2000 annual per capita cigarette sales in California were about 26 packs lower than what they would have been in the absence of Proposition 99. Using new inferential methods proposed in this article, we demonstrate the significance of our estimates.

The rest of the article is organized as follows. Section II describes the main ideas behind the synthetic control approach to comparative case studies of aggregate events. In section III we apply synthetic control methods to estimate the effect of California's Proposition 99. Section IV concludes. Appendix A lists the data sources for the application in section III. Appendix B contains technical details.

II. Synthetic Control Methods for Comparative Case Studies

A. Comparative Case Studies

Case studies focus on particular occurrences of events or interventions of interest. Often, the motivation behind case studies is to detect the effects of an event or policy intervention on some outcome of interest by focusing on a particular instance in which the magnitude of the event or intervention is large relative to other determinants of the outcome, or in which identification of the effects of interest is facilitated by some other characteristic of the intervention. In comparative case studies, researchers compare one or more units exposed to the event or intervention of interest to one or more unexposed units. Therefore, comparative case studies are only feasible when some units are exposed and others are not (or when their levels of exposure differ notably).

To simplify the exposition, we proceed as if only one unit or region is subject to the intervention of interest (otherwise, we could first aggregate the data from the regions exposed to the intervention). In addition, we adopt the terms "region" or "unit" and "intervention" or "treatment", which can be substituted for "country", "state", "city", etc. and "event", "shock", "law", etc., respectively for specific applications.

B. A Motivating Model

The following simple model provides a rationale for the use of synthetic control methods in comparative case study research. Suppose that we observe $J + 1$ regions. Without loss of generality, suppose also that only the first region is exposed to the intervention of interest, so that we have J remaining regions as potential controls. Borrowing from the statistical matching literature, we refer to the set of potential controls as the "donor pool". Also without loss of generality and to simplify notation, assume that the first region is uninterruptedly exposed to the intervention of interest after some initial intervention period.

Let Y_{it}^N be the outcome that would be observed for region i at time t in the absence of the intervention, for units $i = 1, \ldots, J+1$, and time periods $t = 1, \ldots, T$. Let T_0 be the number of pre-intervention periods, with $1 \leq T_0 < T$. Let Y_{it}^I be the outcome that would be observed for unit i at time t if unit i is exposed to the intervention in periods $T_0 + 1$ to T. We assume that the intervention has no effect on the outcome before the implementation period, so for $t \in \{1, \ldots, T_0\}$ and all $i \in \{1, \ldots, N\}$, we have that $Y_{it}^I = Y_{it}^N$. In practice, interventions may have an impact prior to their implementation (for example, via anticipation effects). In those cases, T_0 could be redefined to be the first period in which the outcome may possibly react to the intervention. Implicit in our notation is the usual assumption of no interference between units (see Rosenbaum, 2007, for a detailed discussion of the assumption of no interference between units). That is, we assume that outcomes of the untreated units are not affected by the intervention implemented in the treated unit. In section III we discuss this assumption in the context of our empirical investigation.

Let $\alpha_{it} = Y_{it}^I - Y_{it}^N$ be the effect of the intervention for unit i at time t, and let D_{it} be an indicator that takes value one if unit i is exposed to the intervention at time t, and value zero otherwise. The observed outcome for unit i at time t is

$$Y_{it} = Y_{it}^N + \alpha_{it} D_{it}.$$

Because only the first region (region "one") is exposed to the intervention and only after period T_0 (with $1 \leq T_0 < T$), we have that:

$$D_{it} = \begin{cases} 1 & \text{if } i = 1 \text{ and } t > T_0, \\ 0 & \text{otherwise.} \end{cases}$$

We aim to estimate $(\alpha_{1T_0+1}, \ldots, \alpha_{1T})$. For $t > T_0$,

$$\alpha_{1t} = Y_{1t}^I - Y_{1t}^N = Y_{1t} - Y_{1t}^N.$$

Because Y_{1t}^I is observed, to estimate α_{1t} we just need to estimate Y_{1t}^N. Suppose that Y_{it}^N is given by a factor model:

$$Y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it}, \tag{1}$$

where δ_t is an unknown common factor with constant factor loadings across units, Z_i is a $(r \times 1)$ vector of observed covariates (not affected by the intervention), θ_t is a $(1 \times r)$ vector of unknown parameters, λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is an $(F \times 1)$ vector of unknown factor loadings, and the error terms ε_{it} are unobserved transitory shocks at the region level with zero mean.

Consider a $(J \times 1)$ vector of weights $W = (w_2, \ldots, w_{J+1})'$ such that $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Each particular value of the vector W represents a potential synthetic control, that is, a particular weighted average of control regions. The value of the outcome variable for each synthetic control indexed by W is:

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \delta_t + \theta_t \sum_{j=2}^{J+1} w_j Z_j + \lambda_t \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}.$$

Suppose that there are $(w_2^*, \ldots, w_{J+1}^*)$ such that:

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}, \ \ldots \ , \sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0}, \ \text{ and } \ \sum_{j=2}^{J+1} w_j^* Z_j = Z_1. \tag{2}$$

In Appendix B, we prove that if $\sum_{t=1}^{T_0} \lambda_t' \lambda_t$ is non-singular, then,

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt} = \sum_{j=2}^{J+1} w_j \sum_{s=1}^{T_0} \lambda_t \left(\sum_{n=1}^{T_0} \lambda_n' \lambda_n \right)^{-1} \lambda_s' (\varepsilon_{js} - \varepsilon_{1s}) - \sum_{j=2}^{J+1} w_j^* (\varepsilon_{jt} - \varepsilon_{1t}). \tag{3}$$

In Appendix B we prove also that, under standard conditions, the average of the right hand side of equation (3) will be close to zero if the number of pre-intervention periods is large relative to the scale of the transitory shocks. This suggests using

$$\widehat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$$

for $t \in \{T_0 + 1, \ldots, T\}$ as an estimator of α_{1t}.

Equation (2) can hold exactly only if $(Y_{11}, \ldots, Y_{1\,T_0}, Z_1')$ belongs to the convex hull of $\{(Y_{21}, \ldots, Y_{2\,T_0}, Z_2'), \ldots, (Y_{J+1\,1}, \ldots, Y_{J+1\,T_0}, Z_{J+1}')\}$. In practice, it is often the case that no set of weights exists such that equation (2) holds exactly in the data. Then, the synthetic control region is selected so that equation (2) holds approximately. In some cases, it may not even be possible to obtain a weighted combination of untreated units such that equation (3) holds approximately. This would be the case if $(Y_{11}, \ldots, Y_{1T_0}, Z_1')$ falls far from the convex hull of $\{(Y_{21}, \ldots, Y_{2T_0}, Z_2'), \ldots, (Y_{J+1\,1}, \ldots, Y_{J+1\,T_0}, Z_{J+1}')\}$. Notice, however, that the magnitude of such discrepancy can be calculated for each particular application. So for each particular application, the analyst can decide if the characteristics of the treated unit are sufficiently matched by the synthetic control. In some instances, the fit many be poor and then we would not recommend using a synthetic control.

Even if there is a synthetic control that provides a good fit for the treated units, interpolation biases may be large if the simple linear model presented in this section does not hold over the entire set of regions in any particular sample. Researchers trying to minimize biases caused by interpolating across regions with very different characteristics may restrict the donor pool to regions with similar characteristics to the region exposed to the event or intervention of interest.

Notice that, even if taken at face value, equation (1) generalizes the usual difference-in-differences (fixed-effects) model that is often applied in empirical studies in the social sciences. The traditional difference-in-differences (fixed-effects) model can be obtained if we impose that λ_t in equation (1) is constant for all t. That is, the difference-in-differences model allows for the presence of unobserved confounders but restricts the effect of those confounders to be constant in time, so they can be eliminated by taking time differences. In contrast, the factor model presented in this section allows the effects of confounding unobserved characteristics to vary with time. Under this model, taking time differences does not eliminate the unobserved confounders, μ_j. However, a synthetic control such that

$$\sum_{j=2}^{J+1} w_j^* Z_j = Z_1 \quad \text{and} \quad \sum_{j=2}^{J+1} w_j^* \mu_j = \mu_1, \tag{4}$$

would provide an unbiased estimator of Y_{1t}^N. Choosing a synthetic control in this manner is, of course, not feasible because $\mu_1, \ldots, \mu_{J+1}$ are not observed. However, under fairly standard conditions (see Appendix B), the factor model in equation (1) implies that a synthetic control can fit Z_1 and a long set of pre-intervention outcomes, $Y_{11}, \ldots, Y_{1T_0}$, only as long as it fits Z_1 and μ_1, so equation (4) holds approximately.

Synthetic controls can provide useful estimates in more general contexts than the factor model considered thus far. Consider, for example, the following autoregressive model with time-varying coefficients:

$$\begin{aligned} Y^N_{i\,t+1} &= \alpha_t\, Y^N_{i\,t} + \beta_{t+1}\, Z_{i\,t+1} + u_{i\,t+1}, \\ Z_{i\,t+1} &= \gamma_t\, Y^N_{i\,t} + \Pi_t\, Z_{i\,t} + v_{i\,t+1}, \end{aligned} \tag{5}$$

where $u_{i\,t+1}$ and $v_{i\,t+1}$ have mean zero conditional on $\mathcal{F}_t = \{Y_{j\,s}, Z_{j\,s}\}_{1\leq j\leq N,\ s\leq t}$. Suppose that we can choose $\{w^*_j\}_{2\leq j\leq N}$ such that:

$$\sum_{j=2}^{J+1} w^*_j\, Y_{j\,T_0} = Y_{1\,T_0}, \quad \text{and} \quad \sum_{j=2}^{J+1} w^*_j\, Z_{j\,T_0} = Z_{1\,T_0}. \tag{6}$$

Then, the synthetic control estimator is unbiased even if data for only a single pretreatment period are available. See Appendix B for details.

C. Implementation

Let W be a $(J \times 1)$ vector of positive weights that sum to one. That is, $W = (w_2, \ldots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \ldots, J+1$ and $w_2 + \cdots + w_{J+1} = 1$. Each value of W represents a weighted average of the available control regions and, therefore, a synthetic control. Notice that, although we define our synthetic controls as convex combinations of unexposed units, negative weights or weights larger than one can be used at the cost of allowing extrapolation.

The outcome variable of interest is observed for T periods for the region affected by the intervention Y_{1t}, $(t = 1, \ldots, T)$ and the unaffected regions Y_{jt}, $(j = 2, \ldots, J+1, t = 1, \ldots, T)$. Let $T_1 = T - T_0$ be the number of post-intervention periods. Let Y_1 be the $(T_1 \times 1)$ vector of post-intervention outcomes for the exposed region, and Y_0 be the $(T_1 \times J)$ matrix of post-intervention outcomes for the potential control regions.

Let the $(T_0 \times 1)$ vector $K = (k_1, \ldots, k_{T_0})'$ define a linear combination of pre-intervention outcomes: $\bar{Y}^K_i = \sum_{s=1}^{T_0} k_s Y_{is}$. For example, if $k_1 = k_2 = \cdots = k_{T_0 - 1} = 0$ and $k_{T_0} = 1$, then $\bar{Y}^K = Y_{iT_0}$, the value of the outcome variable in the period immediately prior to the intervention. If $k_1 = k_2 = \cdots = k_{T_0} = 1/T_0$, then $\bar{Y}_i^K = T_0^{-1}\sum_{s=1}^{T_0} Y_{is}$, the simple average of the outcome variable for the pre-intervention periods. Consider M of such linear combinations defined by the vectors $K_1, \ldots, K_M$. Let $X_1 = (Z_1', \bar{Y}_1^{K_1}, \ldots, \bar{Y}_1^{K_M})'$ be a $(k \times 1)$ vector of pre-intervention characteristics for the exposed region, with $k = r + M$. Similarly, X_0 is a $(k \times J)$ matrix that contains the same variables for the unaffected regions. That is, the j-th column of X_0 is $(Z_j', \bar{Y}_j^{K_1}, \ldots, \bar{Y}_j^{K_M})'$. The vector W^* is chosen to minimize some distance (or pseudo-distance), $\|X_1 - X_0 W\|$, between X_1 and $X_0 W$, subject to $w_2 \geq 0, \ldots, w_{J+1} \geq 0$, $w_2 + \cdots + w_{J+1} = 1$. One obvious choice for $\bar{Y}_i^{K_1}, \ldots, \bar{Y}_i^{K_M}$ is $\bar{Y}_i^{K_1} = Y_{i1}, \ldots, \bar{Y}_i^{K_{T_0}} = Y_{i\,T_0}$, that is, the values of the outcome variable for all the available pre-intervention periods. In practice, however, the computation of the weights $w_2^*, \ldots, w_{J+1}^*$ can be simplified by considering only a few linear combinations or pre-intervention outcomes and checking whether equation (2) holds approximately for the resulting weights.

To measure the discrepancy between X_1 and $X_0 W$, we will employ $\|X_1 - X_0 W\|_V = \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$, where V is some $(k \times k)$ symmetric and positive semidefinite matrix, although other choices are also possible. If the relationship between the outcome variable and the explanatory variables in X_1 and X_0 is highly nonlinear and the support of the explanatory variables is large, interpolation biases may be severe. (For example, an equally weighted combination of a 65%-White 35%-Nonwhite state and a 95%-White 5%-Nonwhite state will approximate the outcome of a 80%-White 20%-Nonwhite state if that outcome is approximately linear in the racial composition of the states. However, if the outcome is highly nonlinear in racial composition, the quality of the approximation may be poor.) In that case, W^* can be chosen to minimize $\|X_1 - X_0 W\|$ plus a set of penalty terms specified as increasing functions of the distances between X_1 and the corresponding values for the control units with positive weights in W. Alternatively, as mentioned in section II.B, interpolation biases can be reduced by restricting the comparison group to units that are similar to the exposed units in term of the values of X_1.

Although our inferential procedures are valid for any choice of V, the choice of V influences the mean square error of the estimator. The optimal choice of V assigns weights to linear combinations of the variables in X_0 and X_1 to minimize the mean square error of the synthetic control estimator. Sometimes this choice can be based on subjective assessments of the predictive power of the variables in X_1 and X_0. The choice of V can also be data-driven. One possibility is to choose V such that the resulting synthetic control region approximates the trajectory of the outcome variable of the affected region in the pre-intervention periods. Following Abadie and Gardeazabal (2003), in the empirical section of this article we choose V among positive definite and diagonal matrices such that the mean squared prediction error of the outcome variable is minimized for the pre-intervention periods (see Abadie and Gardeazabal, 2003, Appendix B for details). Alternatively, if the number of available pre-intervention periods in the sample is large enough, researchers may divide them into an initial training period and a subsequent validation period. Given a V, $W^*(V)$ can be computed using data from the training period. Then, the matrix V can be chosen to minimize the mean squared prediction error produced by the weights $W^*(V)$ during the validation period.

D. Inference

The standard errors commonly reported in regression-based comparative case studies measure uncertainty about aggregate data. For example, Card (1990) uses data from the U.S. Current Population Survey to estimate native employment rates in Miami and a set of comparison cities around the time of the Mariel Boatlift. Card and Krueger (1994) use data on a sample of fast-food restaurants in New Jersey and Pennsylvania to estimate the average number of employees in fast-food restaurants in these two states around the time when the minimum wage was increased in New Jersey. The standard errors reported in these studies reflect only the unavailability of aggregate data on employment (for native workers in Miami and other cities, and in fast-food restaurants in New Jersey and Pennsylvania, respectively). This mode of inference would logically produce zero standard errors if aggregate data were used for estimation. However, perfect knowledge of aggregate data does not eliminate all uncertainty about the parameters of interest. That is, even if aggregate data are used for estimation, in most cases researchers would not believe that there is no remaining uncertainty about the value of the parameters of interest. The reason is that not all uncertainty about the value of the estimated parameters come from lack of knowledge of aggregate data. In comparative case studies, an additional source of uncertainty derives from ignorance about the ability of the control group to reproduce the counterfactual of how the treated unit would have evolved in the absence of the treatment. This type of uncertainty is present regardless of whether aggregate data are used for estimation or not. The use of individual micro data, as opposed to aggregate data, only increases the total amount of uncertainty if the outcome of interest is an aggregate.

Large sample inferential techniques are not well-suited to comparative case studies when the number of units in the comparison group and the number of periods in the sample are relatively small. In this article, we propose exact inferential techniques, akin to permutation tests, to perform inference in comparative case studies. The methods proposed here can be used whether data are individual (micro) or aggregate (macro), and do not require a large number of comparison units in the donor pool.

The inferential techniques proposed in this article extend Abadie and Gardeazabal (2003) in several directions. In their study of the economic effects of terrorism, Abadie and Gardeazabal (2003) use a synthetic control region to estimate the economic growth that the Basque Country would have experienced in the absence of terrorism. To assess the ability of the synthetic control method to reproduce the evolution of a counterfactual Basque Country without terrorism, Abadie and Gardeazabal (2003) introduce a placebo study, applying the same techniques to Catalonia, a region similar to the Basque Country but with a much lower exposure to terrorism. Similar falsification tests have been used to assess the effects of computers on the distribution of wages (DiNardo and Pischke, 1997), the effect of the Mariel Boatlift on native unemployment in Miami (Angrist and Krueger, 1999), and the validity of the rational addiction model for cigarette consumption (Auld and Grootendorst, 2004). This type of "placebo tests" or "falsification tests" appear under different names in the literature. Angrist and Krueger (1999) discuss empirical tests of this type under the heading "refutability" tests. Rosenbaum (2002a) discusses the use of outcomes “known to be unaffected by the treatment” to evaluate the presence of hidden biases. Placebo studies are also closely related to uniformity trials in agricultural research (Cochran, 1937).

In this paper, we extend the idea of a placebo study to produce quantitative inference in comparative case studies. The idea of the placebo test proposed here is akin to the classic framework for permutation inference, where the distribution of a test statistic is computed under random permutations of the sample units’ assignments to the intervention and non-intervention groups. As in permutation tests, we apply the synthetic control method to every potential control in our sample. This allows us to assess whether the effect estimated by the synthetic control for the region affected by the intervention is large relative to the effect estimated for a region chosen at random. This inferential exercise is exact in the sense that, regardless of the number of available comparison regions, time periods, and whether the data are individual or aggregate, it is always possible to calculate the exact distribution of the estimated effect of the placebo interventions. Notice also that the inferential exercise proposed here produces classical randomization inference for the case where the intervention is indeed randomized across regions, a rather restrictive condition. More generally, our inferential exercise examines whether or not the estimated effect of the actual intervention is large relative to the distribution of the effects estimated for the regions not exposed to the intervention. This is informative inference if under the hypothesis of no intervention effect the estimated effect of the intervention is not expected to be abnormal relative to the distribution of the placebo effects. In this sense, our inferential procedure is related to that of DiNardo and Pischke (1997) and Auld and Grootendorst (2004). DiNardo and Pischke (1997) compare the wage differential associated with computer skills (as reflected in the on-the-job computer use) to the wage differentials associated with the use of other tools (pencils, telephones, calculators) that do not proxy for skills that are scarce in the job market. Similarly, to assess the validity of the rational addiction model, Auld and Grootendorst (2004) compare the result of a test of rational addiction for cigarette consumption to the results of the same test applied to substances that are not considered addictive (milk, eggs, oranges, apples).

For cases in which the number of available comparison regions is very small, one can use the longitudinal dimension of the data to produce placebo studies, as in Bertrand, Duflo, and Mullainathan (2004) where the dates of the placebo interventions are set at random. Heckman and Hotz (1989) provides an earlier application of in-time placebos.

Our work is related to recent developments in inferential methods for difference-in-difference models (see Wooldridge (2003), Athey and Imbens (2006) and Donald and Lang (2007)) Section 6.5 in Wooldridge and Imbens (2008) provides a recent survey of this literature. Also closely related to our work, Conley and Taber (2008) propose an alternative method to do inference in comparative cases studies based on consistent estimation of the distribution of regression residuals for the case where the number of regions in the control group is large. Rosenbaum (2002a,b) provides a detailed discussion of the use of permutation inference in randomized experiments and observational studies.

III. Estimating the Effects of California's Proposition 99

A. Background

Anti-tobacco legislation has a long history in the United States, dating back at least as far as 1893, when Washington became the first state to ban the sale of cigarettes. Over the next 30 years 15 other states followed with similar anti-smoking measures (Dinan and Heckelman, 2005). These early anti-tobacco laws were primarily motivated by moral concerns; health issues were secondary (Tate, 1999). Almost 100 years later, after these early laws had long since been repealed, widespread awareness of smoking's health risks launched a new wave of state and federal anti-tobacco laws across the United States and, ultimately, overseas. Leading this wave, in 1988, was a voter initiative in California known as Proposition 99, the first modern-time large-scale tobacco control program in the United States.

Proposition 99 increased California's cigarette excise tax by 25 cents per pack, earmarked the tax revenues to health and anti-smoking education budgets, funded anti-smoking media campaigns, and spurred local clean indoor-air ordinances throughout the state (Siegel, 2002). Upon initial implementation, Proposition 99 produced more than $100 million per year in anti-tobacco projects for schools, communities, counties, and at the state level. Almost $20 million a year became available for tobacco-related research. As Glantz and Balbach (2000) put it, "[t]hese programs dwarfed anything that any other state or the federal government had ever done on tobacco."

Proposition 99 triggered a wave of local clean-air ordinances in California. Before Proposition 99 no city or town in California required restaurants to be 100 percent smoke-free. From 1989 to 2000 approximately 140 such laws were passed (Siegel, 2002). By 1993 local ordinances prohibiting smoking in the workplace protected nearly two-thirds of the workers in California (Glantz and Balbach, 2000). In 1994 the State of California passed additional legislation that banned smoking in enclosed workplaces. By 1996 more than 90 percent of California workers were covered by a smoke-free workplace policy (Siegel, 2002). Non-smokers' rights advocates view the wave of local ordinances passed under the impetus of Proposition 99 as an important step in the effort to undercut the then existing social support network for tobacco use in California (Glantz and Balbach, 2000).

The tobacco industry responded to Proposition 99 and the spread of clean-air ordinances by increasing its political activity in California at both the state and local levels. Tobacco lobby groups spent 10 times as much money in California in 1991-1992 as they had spent in 1985-1986 (Begay et al., 1993). In addition, after the passage of Proposition 99, tobacco companies increased promotional expenditures in California (Siegel, 2002).

In 1991 California passed Assembly Bill 99, a new piece of legislation implementing Proposition 99. Contrary to the original mandate of Proposition 99, Assembly Bill 99 diverted a significant fraction of Proposition 99 tobacco tax revenues into medical services with little or no connection to tobacco (Glantz and Balbach, 2000). Also in 1991 a new governor began to exert increasing control over California's anti-smoking media campaign. In 1992 Governor Pete Wilson appointed a new Department of Health Services director and halted the media campaign, which provoked a lawsuit by the American Lung Association (ALA). The ALA won the suit and the campaign was back by the end of 1992, although with a reduced budget (Siegel, 2002).

Even so, Proposition 99 was widely perceived to have successfully cut smoking in California. From the passage of Proposition 99 through 1999 adult smoking prevalence fell in California by more than 30 percent, youth smoking levels dropped to the lowest in the country, and per capita cigarette consumption more than halved (California Department of Health Services, 2006). Prior to 1988 per capita cigarette consumption in California trailed the national average by 22.5 packs; ten years later per capita consumption was 40.4 packs lower than the national average (Siegel, 2002).

Following early reports of California's success with Proposition 99, other states adopted similar policies. In 1993 Massachusetts raised taxes on cigarettes from 26 to 51 cents per pack to fund a Health Protection Fund for smoking prevention and cessation programs. Similar laws passed in Arizona in 1994, with a 50-cent tax increase, and Oregon in 1996, where the tax on cigarettes rose from 38 to 68 cents per pack (Siegel, 2002). In November 1998 the tobacco companies signed a $206 billion Master Settlement Agreement that led the industry to impose an immediate 45-cent increase in cigarette prices nationwide (Capehart, 2001). As of April 20, 2009, 30 states, the District of Columbia, and 792 municipalities across the country had laws in effect requiring 100 percent smoke-free workplaces, bars, or restaurants (ANRF, 2009).

Previous studies have investigated the impact of Proposition 99 on smoking prevalence using a variety of methods. Breslow and Johnson (1993), Glantz (1993), and Pierce et al. (1998) show that cigarette consumption in California after the passage of Proposition 99 in 1988 was lower than the average national trend and lower than the linearly extrapolated pre-program trend in California. Hu, Sung and Keeler (1995) use time-series regression to disaggregate the effects of Proposition 99's tax hike and media campaign on per capita cigarette sales.

A related literature has studied the effect of smoking bans on smoking prevalence. Woodruff et al. (1993) show that smoking prevalence in California in 1990 was lower among workers affected by workplace smoking restrictions than among unaffected workers. More generally, Evans, Farrelly, and Montgomery, (1999), Farrelly, Evans, and Sfekas (1999), and Longo et al. (2001) have provided evidence on the effectiveness of workplace smoking bans.

In a study closely related to the analysis in this section, Fichtenberg and Glantz (2000) use least-squares regression to predict smoking rates in California as a function of the smoking rate for the rest of the United States. The regressions in Fichtenberg and Glantz (2000) estimate the effect of Proposition 99 as a time trend in per capita cigarette consumption starting after the implementation of Proposition 99 in 1989. Fichtenberg and Glantz (2000) allow also for a change in this trend after 1992, when the anti-tobacco media campaign was first temporally eliminated and then reestablished but with reduced funds. Using this regression specification, Fichtenberg and Glantz (2000) estimate that during the period 1989-1992 Proposition 99 accelerated the rate of decline of per capita cigarette consumption in California by 2.72 packs per year. Due to program cut-backs after 1992, Fichtenberg and Glantz (2000) estimate that during the period 1993-1997 Proposition 99 accelerated the rate of decline of per capita cigarette consumption in California by only 0.67 packs per year.

B. Data and Sample

We use annual state-level panel data for the period 1970-2000. Proposition 99 was passed in November 1988, giving us 18 years of pre-intervention data. Our sample period begins in 1970 because it is the first year for which data on cigarette sales are available for all our control states. It ends in 2000 because at about this time anti-tobacco measures were implemented across many states, invalidating them as potential control units. Moreover, a decade-long period after the passage of Proposition 99 seems like a reasonable limit on the span of plausible prediction of the effect of this intervention.

Recall that the synthetic California is constructed as a weighted average of potential control states, with weights chosen so that the resulting synthetic California best reproduces the values of a set of predictors of cigarette consumption in California before the passage of Proposition 99. Because the synthetic California is meant to reproduce the smoking rates that would have been observed for California in the absence of Proposition 99, we discard from the donor pool states that adopted some other large-scale tobacco control program during our sample period. Four states (Massachusetts, Arizona, Oregon, and Florida) introduced formal statewide tobacco control programs in the 1989-2000 period and they are excluded from the donor pool. We also discard all states that raised their state cigarette taxes by 50 cents or more over the 1989 to 2000 period (Alaska, Hawaii, Maryland, Michigan, New Jersey, New York, Washington). Notice that, even if smaller tax increases substantially reduced smoking in any of the control states that gets assigned a positive weight in the synthetic control, this should if anything attenuate the treatment effect estimate that we obtain for California. Finally, we also exclude the District of Columbia from our sample. Our donor pool includes the remaining 38 states. Our results are robust, however, to the inclusion of the discarded states.

Our outcome variable of interest is annual per capita cigarette consumption at the state level, measured in our dataset as per capita cigarette sales in packs. We obtained these data from Orzechowski and Walker (2005) where they are constructed using information on state-level tax revenues on cigarettes sales. This is the most widely used indicator in the tobacco research literature, available for a much longer time-period than survey-based measures of smoking prevalence. A disadvantage of tax-revenue-based data relative to survey data on smoking prevalence is that the former is affected by cigarette smuggling across tax jurisdictions. We discuss this issue later in the section. We include in X_1 and X_0 the values of predictors of smoking prevalence for California and the 38 potential controls, respectively. Our predictors of smoking prevalence are: average retail price of cigarettes, per capita state personal income (logged), the percentage of the population age 15-24, and per capita beer consumption. These variables are averaged over the 1980-1988 period and augmented by adding three years of lagged smoking consumption (1975, 1980, and 1988). Appendix A provides data sources.

Using the techniques described in Section II, we construct a synthetic California that mirrors the values of the predictors of cigarette consumption in California before the passage of Proposition 99. We estimate the effect of Proposition 99 on per capita cigarette consumption as the difference in cigarette consumption levels between California and its synthetic versions in the years after Proposition 99 was passed. We then perform a series of placebo studies that confirm that our estimated effects for California are unusually large relative to the distribution of the estimate that we obtain when we apply the same analysis to all states in the donor pool.

C. Results

Figure 1 plots the trends in per capita cigarette consumption in California and the rest of the United States. As this figure suggests, the rest of the United States may not provide a suitable comparison group for California to study the effects of Proposition 99 on per capita smoking. Even before the passage of Proposition 99 the time series of cigarette consumption in California and in the rest of the United States differed notably. Levels of cigarette consumption were similar in California and the rest of the United States in the early 1970's. Trends began to diverge in the late 1970's, when California's cigarette consumption peaked and began to decline while consumption in the rest of the United States was still rising. Cigarette sales declined in the 1980's, but with larger decreases in California than in the rest of the United States. In 1988, the year Proposition 99 passed, cigarette consumption was about 27 percent higher in the rest of the United States relative to California. Following the law's passage, cigarette consumption in California continued to decline. To evaluate the effect of Proposition 99 on cigarette smoking in California, the central question is how cigarette consumption would have evolved in California after 1988 in the absence of Proposition 99. The synthetic control method provides a systematic way to estimate this counterfactual.

As explained above, we construct the synthetic California as the convex combination of states in the donor pool that most closely resembled California in terms of pre-Proposition 99 values of smoking prevalence predictors. The results are displayed in Table 1, which compares the pre-treatment characteristics of the actual California with that of the synthetic California, as well as with the population-weighted average of the 38 states in the donor pool. We see that the average of states that did not implement a large-scale tobacco-control program in 1989-2000 does not seem to provide a suitable control group for California. In particular, prior to the passage of Proposition 99 average beer consumption and cigarette retail prices were lower in the average of the 38 control states than in California. Moreover, prior to the passage of Proposition 99 average cigarette sales per capita were substantially higher on average in the 38 control states than in California. In contrast, the synthetic California accurately reproduces the values that smoking prevalence and smoking prevalence predictor variables had in California prior to the passage of Proposition 99.

Table 1 highlights an important feature of synthetic control estimators. Similar to matching estimators, the synthetic control method forces the researcher to demonstrate the affinity between the region exposed to the intervention of interest and the regions in the donor pool. As a result, the synthetic control method safeguards against estimation of "extreme counterfactuals," that is, those counterfactuals that fall far outside the convex hull of the data (King and Zheng, 2006). As explained in section II.C, we chose V among all positive definite and diagonal matrices to minimize the mean squared prediction error of per capita cigarette sales in California during the pre-Proposition 99 period. The resulting value of the diagonal element of V associated to the log per capita GDP variable is very small, which indicates that, given the other variables in Table 1, log GDP per capita does not have substantial power predicting the per capita cigarette consumption in California before the passage of Proposition 99. This explains the discrepancy between California and its synthetic version in terms of log GDP per capita.

Table 2 displays the weights of each control state in the synthetic California. The weights reported in Table 2 indicate that smoking trends in California prior to the passage of Proposition 99 is best reproduced by a combination of Colorado, Connecticut, Montana, Nevada, and Utah. All other states in the donor pool are assigned zero W-weights.

Figure 2 displays per capita cigarette sales for California and its synthetic counterpart during the period 1970-2000. Notice that, in contrast to per capita sales in other U.S. states (shown in Figure 1), per capita sales in the synthetic California very closely track the trajectory of this variable in California for the entire pre-Proposition 99 period. Combined with the high degree of balance on all smoking predictors (Table 1), this suggests that the synthetic California provides a sensible approximation to the number of cigarette packs per capita that would have been sold in California in 1989-2000 in the absence of Proposition 99.

Our estimate of the effect of Proposition 99 on cigarette consumption in California is the difference between per capita cigarette sales in California and in its synthetic version after the passage of Proposition 99. Immediately after the law's passage, the two lines begin to diverge noticeably. While cigarette consumption in the synthetic California continued on its moderate downward trend, the real California experienced a sharp decline. The discrepancy between the two lines suggests a large negative effect of Proposition 99 on per capita cigarette sales. Figure 2 plots the yearly estimates of the impacts of Proposition 99, that is, the yearly gaps in per capita cigarette consumption between California and its synthetic counterpart. Figure 2 suggests that Proposition 99 had a large effect on per capita cigarette sales, and that this effect increased in time. The magnitude of the estimated impact of Proposition 99 in Figure 2 is substantial. Our results suggest that for the entire 1989-2000 period cigarette consumption was reduced by an average of almost 20 packs per capita, a decline of approximately 25 percent.

In order to assess the robustness of our results, we included additional predictors of smoking prevalence among the variables used to construct the synthetic control. Our results stayed virtually unaffected regardless of which and how many predictor variables we included. The list of predictors used for robustness checks included state-level measures of unemployment, income inequality, poverty, welfare transfers, crime rates, drug related arrest rates, cigarette taxes, population density, and numerous variables to capture the demographic, racial, and social structure of states.

Our analysis produces estimates of the effect of Proposition 99 that are considerably larger than those obtained by Fichtenberg and Glantz (2000) using linear regression methods. In particular, Fichtenberg and Glantz (2000) estimate that by 1997 Proposition 99 had reduced per capita cigarette sales in California by about 14 packs per year. Our estimates increase this figure substantially, to 24 packs per year. Part of this difference is likely to be explained by the fact that Fichtenberg and Glantz (2000) use per capita cigarette sales in the rest of the United States to reproduce how this variable would have evolved in California in the absence of Proposition 99. As explained above, after the enactment of Proposition 99 in California, other states, like Massachusetts and Florida passed similar tobacco control legislation. While we eliminate these states as potential controls, Fichtenberg and Glantz (2000) do not do so, which is likely to attenuate their estimates.

There are several ways in which the assumption of no interference between units of section II could be violated in the context of our analysis of the effects of Proposition 99. In our judgment, these potential violations do not appear to be severe, and in some cases would likely attenuate the estimated effect of Proposition 99. Perhaps the most important concern in this regard is that the increase in anti-tobacco sentiment created in California by Proposition 99 could have spread to other states, contaminating the donor pool. Another concern is that in response to Proposition 99 the tobacco industry could have diverted funds from planned advertising campaigns in other states to California. In both cases, interference would likely cause lower levels of smoking in the control states, attenuating our estimate of Proposition 99. On the other hand, it is possible that the rise in tobacco taxes implemented under Proposition 99 increased cigarette smuggling or cross-border purchases from nearby jurisdictions. However, Lovenheim (2008) and DeCicca, Kenkel, and Liu (2008) provide evidence that large distances to lower tobacco price jurisdictions keep low the level of cross-border cigarette purchases for California. There is much less information about organized smuggling, although it has been argued that the extent of this activity in the US is likely to be small and in decline (e.g., Kleine, 1993). An increase in the number of cigarettes smuggled into California after the passage of Proposition 99 would exacerbate our estimates. However, given the large magnitude of the effects that we estimate in this article, the increase in cigarettes smuggled into California after Proposition 99 would have had to have been massive in order to explain our estimates.

D. Inference about the effect of the California Tobacco Control Program

To evaluate the significance of our estimates, we pose the question of whether our results could be driven entirely by chance. How often would we obtain results of this magnitude if we had chosen a state at random for the study instead of California? To answer this question, we use placebo tests. Similar to Abadie and Gardeazabal (2003) and Bertrand, Duflo, and Mullainathan (2004), we run placebo studies by applying the synthetic control method to states that did not implement a large-scale tobacco control program during the sample period of our study. If the placebo studies create gaps of magnitude similar to the one estimated for California, then our interpretation is that our analysis does not provide significant evidence of a negative effect of Proposition 99 on cigarette sales in California. If, on the other hand, the placebo studies show that the gap estimated for California is unusually large relative to the gaps for the states that did not implement large-scale tobacco control program, then our interpretation is that our analysis provides significant evidence of a negative effect of Proposition 99 on cigarette sales in California.

To assess the significance of our estimates, we conduct a series of placebo studies by iteratively applying the synthetic control method used to estimate the effect of Proposition 99 in California to every other state in the donor pool. In each iteration we reassign in our data the tobacco control intervention to one of the 38 control states, keeping California in the donor pool. That is, we proceed as if one of the states in the donor pool would have passed a large-scale tobacco control program in 1988, instead of California. We then compute the estimated effect associated with each placebo run. This iterative procedure provides us with a distribution of estimated gaps for the states in which no intervention took place.

Figure 4 displays the results for the placebo test. The gray lines represent the gap associated with each of the 38 runs of the test. That is, the gray lines show the difference in per capita cigarette sales between each state in the donor pool and its respective synthetic version. The superimposed black line denotes the gap estimated for California. As the figure makes apparent, the estimated gap for California during the 1989-2000 period is unusually large relative to the distribution of the gaps for the states in the donor pool.

As Figure 4 indicates, the synthetic method provides an excellent fit for per capita cigarette sales in California prior to the passage of Proposition 99. The pre-intervention mean squared prediction error (MSPE) in California (the average of the squared discrepancies between per capita cigarette sales in California and in its synthetic counterpart during the period 1970-1988) is about 3. The pre-Proposition 99 median MSPE among the 38 states in the donor pool is about 6, also quite small, indicating that the synthetic control method is able to provide a good fit for per capita cigarette consumption prior to Proposition 99 for the majority of the states in the donor pool. However, Figure 4 indicates also that per capita cigarette sales during the 1970-1988 period cannot be well-reproduced for some states by a convex combination of per capita cigarette sales in other states. The state with worst fit in the pre-Proposition 99 period is New Hampshire, with a MSPE of 3437. The large MSPE for New Hampshire does not come as a surprise. Among all the states in the donor pool, New Hampshire is the state with the highest per capita cigarette sales for every year prior to the passage of Proposition 99. Therefore, there is no combination of states in our sample that can reproduce the time series of per capita cigarette sales in New Hampshire prior to 1988. Similar problems arise for other states with extreme values of per capita cigarette sales during the pre-Proposition 99 period.

If the synthetic California had failed to fit per capita cigarette sales for the real California in the years before the passage of Proposition 99, we would have interpreted that much of the post-1988 gap between the real and the synthetic California was also artificially created by lack of fit, rather than by the effect of Proposition 99. Similarly, placebo runs with poor fit prior to the passage of Proposition 99 do not provide information to measure the relative rarity of estimating a large post-Proposition 99 gap for a state that was well-fitted prior to Proposition 99. For this reason, we provide several different versions of Figure 4, each version excluding states beyond a certain level of pre-Proposition 99 MSPE.

Figure 5 excludes states that had a pre-Proposition 99 MSPE of more than 20 times the MSPE of California. This is a very lenient cutoff, discarding only four states with extreme values of pre-Proposition 99 MSPE for which the synthetic method would be clearly ill-advised. In this figure there remain a few lines that still deviate substantially from the zero gap line in the pre-Proposition 99 period. Among the 35 states remaining in the figure, the California gap line is now about the most unusual line, especially from the mid 1990's onward.

Figure 6 is based on a lower cutoff, excluding all states that had a pre-Proposition 99 MSPE of more than five times the MSPE of California. Twenty-nine control states plus California remain in the figure. The California gap line is now clearly the most unusual line for almost the entire post-treatment period.

In Figure 7 we lower the cutoff even further and focus exclusively on those states that we can fit almost as well as California in the period 1970-1988, that is, those states with pre-Proposition 99 MSPE not higher than twice the pre-Proposition 99 MSPE for California. Evaluated against the distribution of the gaps for the 19 remaining control states in Figure 7, the gap for California appears highly unusual. The negative effect in California is now by far the lowest of all. Because this figure includes 19 control states, the probability of estimating a gap of the magnitude of the gap for California under a random permutation of the intervention in our data is 5 percent, a test level typically used in conventional tests of statistical significance.

One final way to evaluate the California gap relative to the gaps obtained from the placebo runs is to look at the distribution of the ratios of post/pre-Proposition 99 MSPE. The main advantage of looking at ratios is that it obviates choosing a cutoff for the exclusion of ill-fitting placebo runs. Figure 8 displays the distribution of the post/pre-Proposition 99 ratios of the MSPE for California and all 38 control states. The ratio for California clearly stands out in the figure: post-Proposition 99 MSPE is about 130 times the MSPE for the pre-Proposition 99 period. No control state achieves such a large ratio. If one were to assign the intervention at random in the data, the probability of obtaining a post/pre-Proposition 99 MSPE ratio as large as California's is $1/39 = 0.026$.

IV. Conclusion

Comparative case study research has broad potential in the social sciences. However, the empirical implementation of comparative case studies is plagued by inferential challenges and ambiguity about the choice of valid control groups. In this paper, we propose data-driven procedures to select synthetic comparison units in comparative case studies. We show that the synthetic control estimator is valid under fairly standard conditions. In addition, we propose a method to produce inference in comparative cases studies that incorporates uncertainty about the validity of the control unit. Moreover, we provide software to implement the estimators proposed in this article.

We demonstrate the applicability of the synthetic control method by studying the effects of Proposition 99, a large-scale tobacco control program that California passed in 1988. Our results suggest the effects of the tobacco control program are much larger than prior estimates have reported. We show that if one were to relabel the intervention state in the dataset at random, the probability of obtaining results of the magnitude of those obtained for California would be extremely small, 0.026.

Appendix A: Data Sources

In this appendix, we describe the data used in our analysis and provide sources.

- Per-capita cigarette consumption (in packs). Source: Orzechowski and Walker (2005). These data are based on the total tax paid on sales of packs of cigarettes in a particular state divided by its total population.

- Average retail price per pack of cigarettes (in cents). Source: Orzechowski and Walker (2005). Price figures include state sales taxes, if applicable.

- Per-capita state personal income (logged). Source: Bureau of the Census, United States Statistical Abstract. Converted to 1997 dollars using the Consumer Price Index.

- State population and percent of state population aged 15-24. Source: U.S. Census Bureau.

- Per-capita beer consumption. Source: Beer Institute's Brewer's Almanac. Measured as the per capita consumption of malt beverages (in gallons).

Appendix B: Technical Details

Consider the first model in section II.B,

$$Y_{it}^N = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \varepsilon_{it},$$

where $\lambda_t = (\lambda_{t1}, \ldots, \lambda_{tF})$ is a $(1 \times F)$ vector of common factors, for $t = 1, \ldots, T$, and $\mu_i = (\mu_{i1}, \ldots, \mu_{iF})'$ is an $(F \times 1)$ vector of factor loadings, for $i = 1, \ldots, J+1$. The weighted average of the outcome in the donor pool, using weights $\{w_j\}_{2 \le j \le J+1}$ is:

$$\sum_{j=2}^{J+1} w_j Y_{jt}^N = \delta_t + \theta_t \left(\sum_{j=2}^{J+1} w_j Z_j \right) + \lambda_t \left(\sum_{j=2}^{J+1} w_j \mu_j \right) + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}.$$

As a result,

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j Y_{jt}^N = \theta_t \left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j \right) + \lambda_t \left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j \right) + \sum_{j=2}^{J+1} w_j (\varepsilon_{1t} - \varepsilon_{jt}).$$

We assume that the terms ε_{it} are independent across units and in time. The analysis can, however, be extended to more general settings. Notice that even with ε_{it} independent across units and in time, the unobserved residual $u_{it} = \lambda_t \mu_i + \varepsilon_{it}$ may be correlated across units and in time because the presence of the term $\lambda_t \mu_i$. Assume also that the terms ε_{it} are mean-independent of $\{Z_i, \mu_i\}_{i=1}^{J+1}$. Let Y_i^P be the $T_0 \times 1$ vector with t-th element equal to Y_{it}. Similarly, let ε_i^P be the $(T_0 \times 1)$ vector with t-th element equal to ε_{it}. Finally, let θ^P and λ^P be the $(T_0 \times r)$ matrix and $(T_0 \times F)$ matrix with t-th rows equal to θ_t and λ_t, respectively. We obtain,

$$Y_1^P - \sum_{j=2}^{J+1} w_j Y_j^P = \theta^P \left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j \right) + \lambda^P \left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j \right) + \sum_{j=2}^{J+1} w_j (\varepsilon_1^P - \varepsilon_j^P).$$

Let $\xi(M)$ be the smallest eigenvalue of:

$$\frac{1}{M} \sum_{t=T_0-M+1}^{T_0} \lambda_t' \lambda_t.$$

Assume that $\xi(M)$ is bounded away from zero: $\xi(M) \ge \underline{\xi} > 0$, for each positive integer, M. Assume also that there exists a constant, $\bar{\lambda}$, such that $|\lambda_{tf}| \le \bar{\lambda}$ for all $t = 1, \ldots, T$,

$f = 1, \ldots, F$. Therefore, because $\lambda^{P\prime}\lambda^P$ is not singular:

$$\begin{aligned}
Y_{1t}^N - \sum_{j=2}^{J+1} w_j Y_{jt}^N &= \lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\left(Y_1^P - \sum_{j=2}^{J+1} w_j Y_j^P\right) \\
&+ (\theta_t - \lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\theta^P)\left(Z_1 - \sum_{j=2}^{J+1} w_j Z_j\right) \\
&- \lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\left(\varepsilon_1^P - \sum_{j=2}^{J+1} w_j \varepsilon_j^P\right) + \sum_{j=2}^{J+1} w_j(\varepsilon_{1t} - \varepsilon_{jt}).
\end{aligned}$$

Suppose that there exist $\{w_2^*, \ldots, w_{J+1}^*\}$ such that equation (2) holds. Then

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j^* Y_{jt}^N = R_{1t} + R_{2t} + R_{3t},$$

where

$$R_{1t} = \lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\sum_{j=2}^{J+1} w_j^* \varepsilon_j^P, \quad R_{2t} = -\lambda_t(\lambda^{P\prime}\lambda^P)^{-1}\lambda^{P\prime}\varepsilon_1^P,$$

and $R_{3t} = \sum_{j=2}^{J+1} w_j^*(\varepsilon_{jt} - \varepsilon_{1t})$. Consider the case of $t > T_0$. Then, R_{2t} and R_{3t} have mean zero. Notice that,

$$R_{1t} = \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} \lambda_t \left(\sum_{n=1}^{T_0} \lambda_n' \lambda_n\right)^{-1} \lambda_s' \varepsilon_{js}.$$

Because $\sum_{t=1}^{T_0} \lambda_t'\lambda_t$ is symmetric and positive definite, so is its inverse. Then, applying the Cauchy-Schwarz Inequality, we obtain:

$$\begin{aligned}
\left(\lambda_t \left(\sum_{n=1}^{T_0} \lambda_n'\lambda_n\right)^{-1} \lambda_s'\right)^2 &\leq \left(\lambda_t \left(\sum_{n=1}^{T_0} \lambda_n'\lambda_n\right)^{-1} \lambda_t'\right)\left(\lambda_s \left(\sum_{n=1}^{T_0} \lambda_n'\lambda_n\right)^{-1} \lambda_s'\right) \\
&\leq \left(\frac{\bar{\lambda}^2 F}{T_0\, \underline{\xi}}\right)^2.
\end{aligned}$$

Let

$$\bar{\varepsilon}_j^L = \sum_{s=1}^{T_0} \lambda_t \left(\sum_{n=1}^{T_0} \lambda_n'\lambda_n\right)^{-1} \lambda_s' \varepsilon_{js}$$

for $j = 2, \ldots, J+1$.

Assume that, for some even p, the p-th moments of $|\varepsilon_{jt}|$ exist for $j = 2, \ldots, J+1$ and $t = 1, \ldots, T_0$. Using Hölder's Inequality:

$$\sum_{j=2}^{J+1} w_j^* |\bar{\varepsilon}_j^L| \leq \left(\sum_{j=2}^{J+1} w_j^* |\bar{\varepsilon}_j^L|^p \right)^{1/p} \leq \left(\sum_{j=2}^{J+1} |\bar{\varepsilon}_j^L|^p \right)^{1/p}.$$

Therefore, applying again Hölder's Inequality:

$$E\left[\sum_{j=2}^{J+1} w_j^* |\bar{\varepsilon}_j^L| \right] \leq \left(E\left[\sum_{j=2}^{J+1} |\bar{\varepsilon}_j^L|^p \right] \right)^{1/p}.$$

Now, using Rosenthal's Inequality:

$$E|\bar{\varepsilon}_j^L|^p \leq C(p) \left(\frac{\bar{\lambda}^2 F}{\underline{\xi}} \right)^p \max\left\{ \frac{1}{T_0^p} \sum_{t=1}^{T_0} E|\varepsilon_{jt}|^p, \left(\frac{1}{T_0^2} \sum_{t=1}^{T_0} E|\varepsilon_{jt}|^2 \right)^{p/2} \right\},$$

where $C(p)$ is the p-th moment of minus one plus a Poisson random variable with parameter equal to one (see Ibragimov and Sharakhmetov, 2002). Let $\sigma_{jt}^2 = E|\varepsilon_{jt}|^2$, $\sigma_j^2 = (1/T_0)\sum_{t=1}^{T_0} \sigma_{jt}^2$, $\bar{\sigma}^2 = \max_{j=2,\ldots,J+1} \sigma_j^2$, and $\bar{\sigma} = \sqrt{\bar{\sigma}^2}$. Similarly, let $m_{p,jt} = E|\varepsilon_{jt}|^p$, $m_{p,j} = (1/T_0)\sum_{t=1}^{T_0} m_{p,jt}$, and $\bar{m}_p = \max_{j=2,\ldots,J+1} m_{p,j}$. We obtain that, for $t > T_0$,

$$E|R_{1t}| \leq C(p)^{1/p} \left(\frac{\bar{\lambda}^2 F}{\underline{\xi}} \right) J^{1/p} \max\left\{ \frac{\bar{m}_p^{1/p}}{T_0^{1-1/p}}, \frac{\bar{\sigma}}{T_0^{1/2}} \right\}.$$

Last equation shows that the bias of the estimator can be bounded by a function that goes to zero as the number of pre-treatment periods increases.

Consider now the autoregressive model in equation (5). Notice that

$$Y_{i\,T_0+1}^N = \big(\alpha_{T_0} + \beta_{T_0+1}\,\gamma_{T_0}\big) Y_{i\,T_0} + \beta_{T_0+1}\,\Pi_{T_0}\,Z_{i\,T_0} + \beta_{T_0+1}\,v_{i\,T_0+1} + u_{i\,T_0+1},$$

where $\{u_{it}, v_{it}\}_{T_0+1 \leq t \leq T_0+n}$ have mean zero conditional on $\mathcal{F}_{T_0}$. Working recursively, it can be shown that conditional on $Y_{i\,T_0}$ and $Z_{i\,T_0}$, $Y_{i\,T_0+n}^N$ is a linear function of $\{u_{it}, v_{it}\}_{T_0+1 \leq t \leq T_0+n}$, for $n \geq 1$. Then, because $\{w_j^*\}_{2 \leq j \leq N}$ is a deterministic function of $\mathcal{F}_{T_0}$ and $\{u_{it}, v_{it}\}_{T_0+1 \leq t \leq T_0+n}$ have mean zero conditional on $\mathcal{F}_{T_0}$, the bias of the synthetic control estimator is equal to zero if equation (6) holds.

References

Abadie, A. and Gardeazabal, J. (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country", *American Economic Review* , vol. 93, no. 1, 112-132.

Angrist, J.D. and Krueger, A.B (1999), "Empirical Strategies in Labor Economics," in A. Ashenfelter and D. Card eds. Handbook of Labor Economics, vol. 3. Elsevier Science.

ANRF (2009), "Municipalities with 100% Smokefree Laws", available at: `http://www.no-smoke`.org/pdf/100ordlisttabs.pdf. Accessed on June 23, 2009.

Athey, S. and Imbens, G.W. (2006), "Identification and Inference in Nonlinear Difference-in-Differences Models", *Econometrica* , vol. 74, no. 2, 431-497.

Auld, M.C. and Grootendorst, P. (2004),"An Empirical Analysis of Milk Addiction", *Journal of Health Economics*, vol. 23, 1117-1133.

Begay, M., Traynor, M., and Glantz, S. (1993), "The Tobacco Industry, State Politics, and Tobacco Education in California", *American Journal of Public Health*, vol. 83, no. 9, 1214-1221.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004), "How Much Should we Trust Differences-in-Differences Estimates?", *Quarterly Journal of Economics*, vol. 119, no. 1, 249-275.

Breslow, L. and Johnson, M. (1993) "California's Proposition 99 on Tobacco, and its Impact", *Annual Review of Public Health*, vol. 14, 585-604.

California Department of Health Services (2006), "Fast Facts", PS 11, `http://www.dhs.ca.gov/opa/FactSheets/PDF/ps11.pdf`.

Capehart, T. (2001) "Trends in the Cigarette Industry after the Master Settlement Agreement", *USDA Electronic Outlook Report*, October, TBS-250-01.

Card, D. (1990), "The Impact of the Mariel Boatlift on the Miami Labor Market", *Industrial and Labor Relations Review*, vol. 44, 245-257.

Card, D. and Krueger, A. B. (1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania", *American Economic Review*, vol. 84, 772-793.

Cochran, W.G. (1937), "A Catalogue of Uniformity Trial Data", *Supplement to the Journal of the Royal Statistical Society*, vol. 4(2), 233-253.

Conley, T.G. and Taber, C.R. (2008), "Inference with Difference in Differences with a Small Number of Policy Changes", mimeo, University of Chicago.

DeCicca, P., Kenkel, D., and Liu, F. (2008), "Excise Tax Avoidance: The Case of State Cigarette Taxes", mimeo, McMaster University, Canada.

Dinan, J. and Heckelman, J. (2005), "The Anti-Tobacco movement in the Progressive Era: A Case Study of Direct Democracy in Oregon", *Explorations in Economic History*, vol. 42, no. 4, 529-546.

DiNardo, J.E. and Pischke, J.S. (1997), "The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?", *Quarterly Journal of Economics*, vol. 112, 291-303.

Donald, S.G. and Lang, K. (2007), "Inference with Difference in Differences and Other Panel Data", *Review of Economics and Statistics*, vol. 89, no. 2, 221-233.

Evans, W., Farrelly, M., and Montgomery, E. (1999), "Do Workplace Smoking Bans Reduce Smoking?", *American Economic Review*, vol. 89, no. 4, 728-747.

Farrelly, M., Evans, W., and Sfekas, A. (1999), "The Impact of Workplace Smoking Bans: Results from a National Survey", *Tobacco Control*, vol. 8, 272-227.

Fichtenberg, C. and Glantz, S. (2000), "Association of the California Tobacco Control Program with Declines in Cigarette Consumption and Mortality from Heart Disease", *New England Journal of Medicine*, vol. 343, no. 24, 1772-1777.

Glantz, S. (1993), "Changes in Cigarette Consumption, Prices, and Tobacco Industry Revenues Associated with California's Proposition 99", *Tobacco Control*, vol. 2, 311-314.

Glantz, S. and Balbach, E. (2000), *Tobacco War: Inside the California Battles.* Berkeley: University of California Press. `http://ark.cdlib.org/ark:/13030/ft167nb0vq/`.

Heckman, J.J. and Hotz, V.J. (1989), "Choosing Among Alternatives Nonexperimental Methods For Estimating The Impact of Social Programs", *Journal of The American Statistical Association*, vol. 84, 862-874.

Hu, T., Sung, H., and Keeler, T. (1995), "Reducing Cigarette Consumption in California: Tobacco Taxes vs. an Anti-Smoking Media Campaign", *American Journal of Public Health*, vol. 85, no. 9, 1218-1222.

Ibragimov, R. and Sharakhmetov, S. (2002), "The Exact Constant in the Rosenthal Inequality for Random Variables with Mean Zero", *Theory of Probability and Its Applications*, vol. 46, 127-131.

King, G. and Zheng, L. (2006), "The Dangers of Extreme Counterfactuals", *Political Analysis*, vol. 14, no. 2, 131-159.

Kleine, R. (1993) "The Declining Role of Interstate Cigarette Smuggling in the United States", *Tobacco Control*, vol. 2, 38-40.

Longo, D., Johnson, J., Krusea, R., Brownson, R., and Hewett, J. (2001), "A Prospective Investigation of the Impact of Smoking Bans on Tobacco Cessation and Relapse", *Tobacco Control*, vol. 10, 267-272.

Lovenheim, M.F., (2008), "How Far to the Border?: The Extent and Impact of Cross-Border Casual Cigarette Smuggling", *National Tax Journal*, vol. 61, no. 1, 7-33.

ORZECHOWSKI and WALKER (2005), The Tax Burden on Tobacco. Historical Compilation, Vol 40, 2005. Arlington, VA: Orzechowski & Walker.

PIERCE, J., GILPIN, E., EMERY, S., FARKAS, A., ZHU, S., CHOI, W., BERRY, C., DISTEFAN, J., WHITE, M., SOROKO, S., and NAVARRO, A. (1998), Tobacco Control in California: Who's Winning the War? An Evaluation of the Tobacco Control Program, 1989-1996, La Jolla, CA: University of California at San Diego, Chapter 2.

ROSENBAUM, P.R. (2002a), *Observational Studies*, second edition. Springer, New York.

ROSENBAUM, P.R. (2002b), "Covariance Adjustment in Randomized Experiments and Observational Studies", *Statistical Science*, vol. 17, no. 3, 286-327.

ROSENBAUM, P.R. (2007), 'Inference Between Units in Randomized Experiments", *Journal of the American Statistical Association*, vol. 102, no. 477, 191-200.

RUBIN, D.B. (2001), "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation", *Health Services and Outcomes Research Methodology*, vol. 1, 169-188.

SIEGEL, M. (2002), "The Effectiveness of State-Level Tobacco Control Interventions: A Review of Program Implementation and Behavioral Outcomes", *Annual Review of Public Health*, vol. 23, 45-71.

TATE, C. (1999), *Cigarette Wars: The Triumph of the Little White Slaver*, NY: Oxford University Press.

WOODRUFF, T., ROSBROOK, B., PIERCE, J., and GLANTZ, S. (1993) "Lower Levels of Cigarette Consumption Found in Smoke-Free Workplaces in California", *Archives of Internal Medicine*, vol. 153, 1485-1493.

WOOLDRIDGE, J. M. (2003) "Cluster-Sample Methods in Applied Econometrics", *American Economic Review*, vol. 93(2), 133-138.

WOOLDRIDGE, J. M. and IMBENS, G.W. (2008) "Recent Developments in the Econometrics of Program Evaluation", NBER Working Paper No. W14251.

Figure 1: Trends in Per-Capita Cigarette Sales: California vs. the Rest of the United States

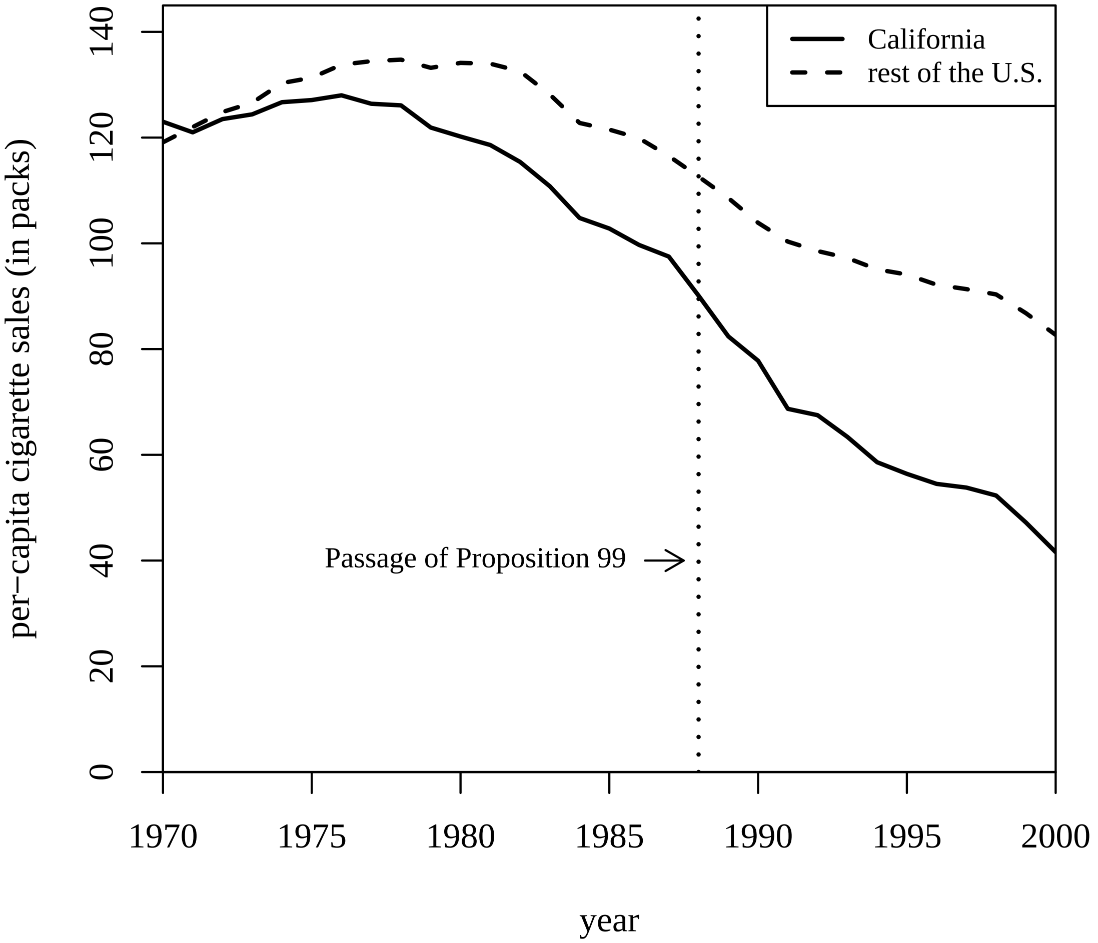

Figure 2: Trends in Per-Capita Cigarette Sales: California vs. Synthetic California

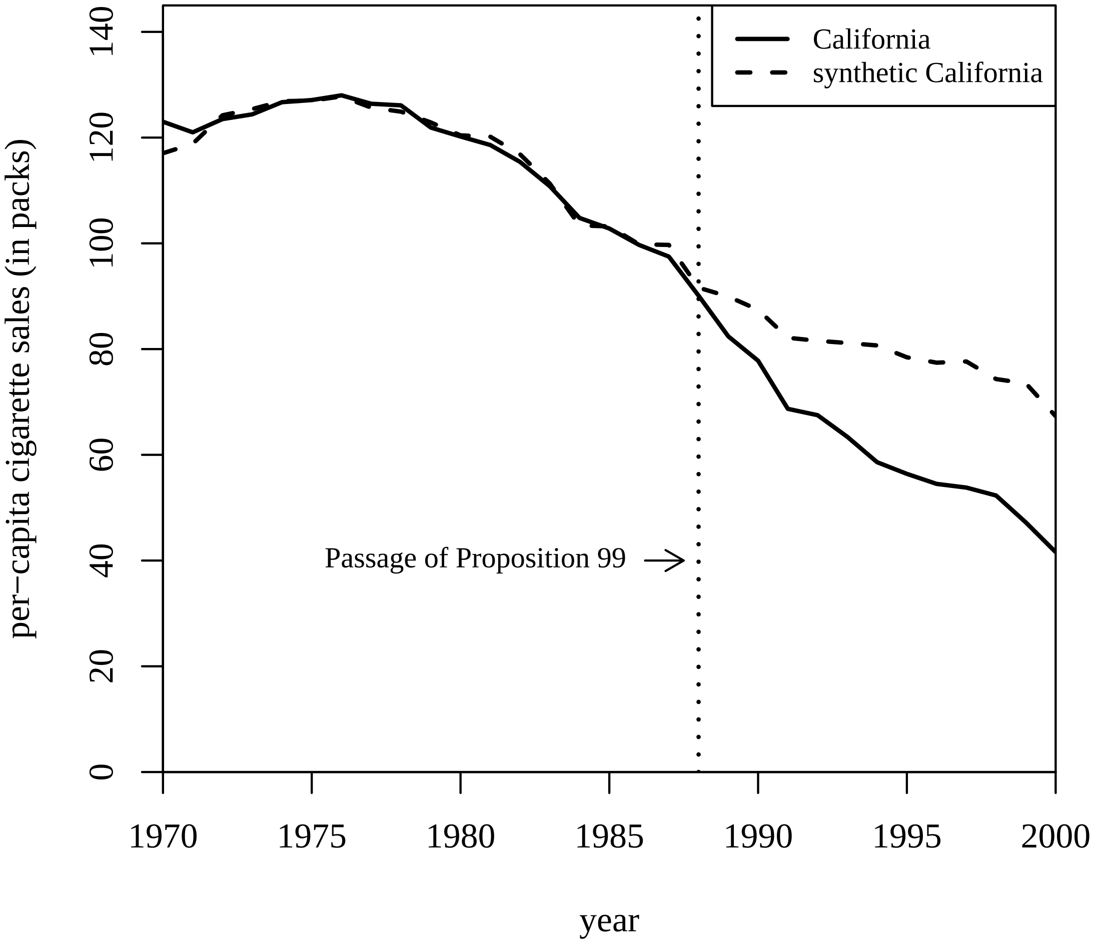

Figure 3: Per-Capita Cigarette Sales Gap Between California and Synthetic California

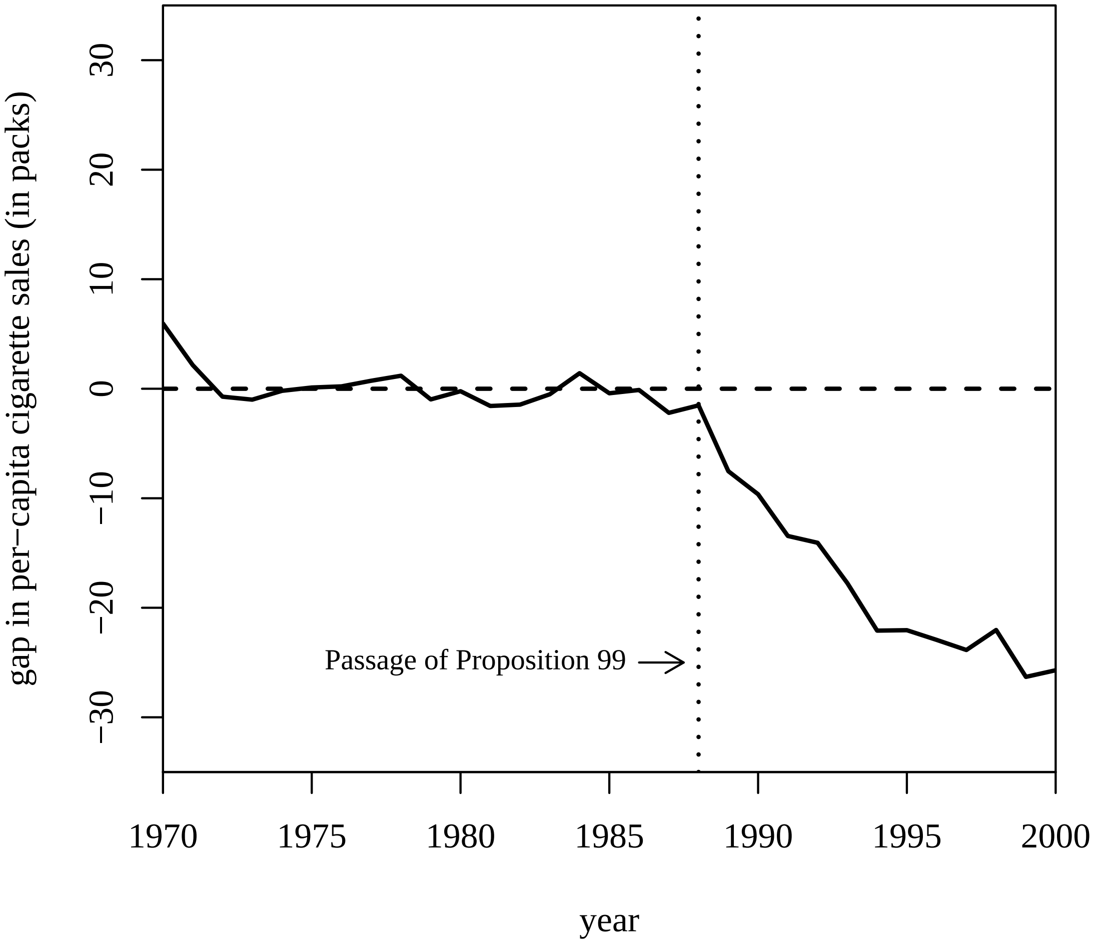

Figure 4: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in all 38 Control States

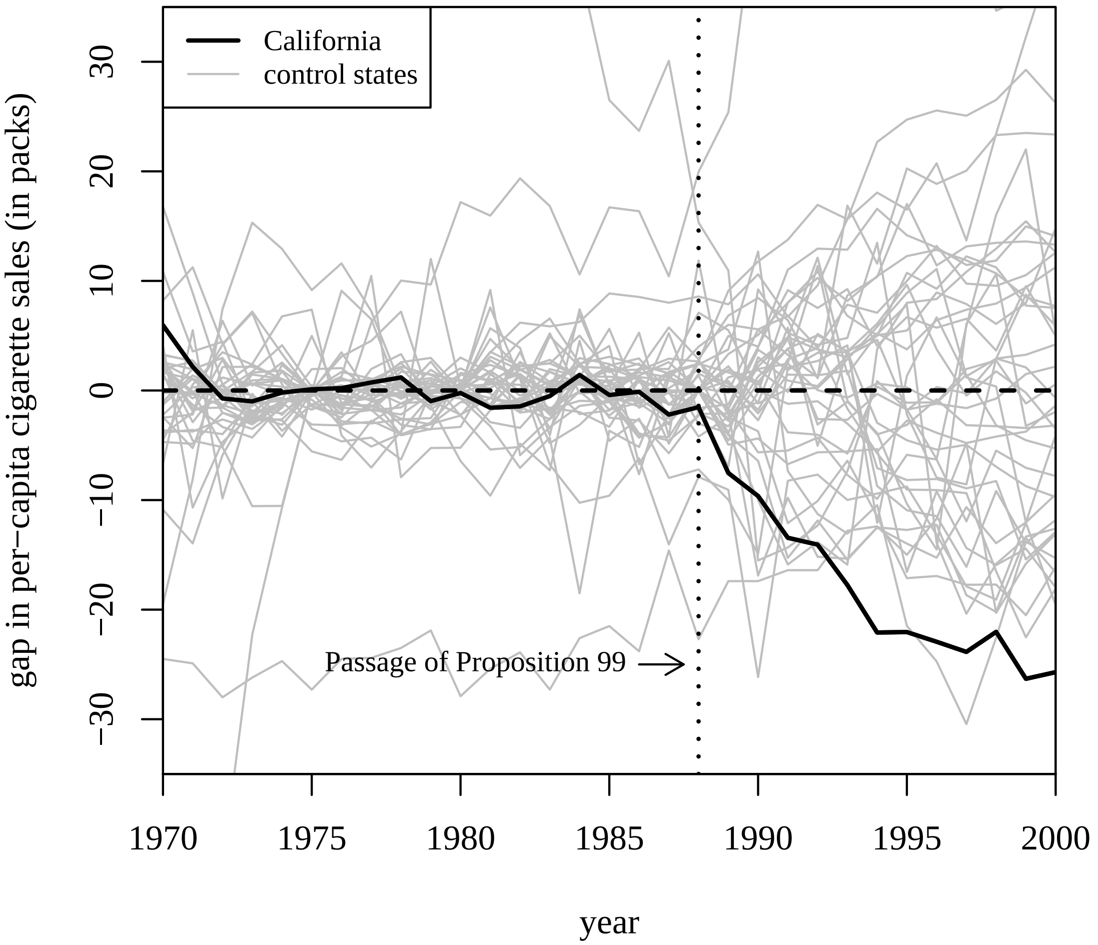

Figure 5: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 34 Control States (Discards States with Pre-Proposition 99 MSPE Twenty Times Higher than California's)

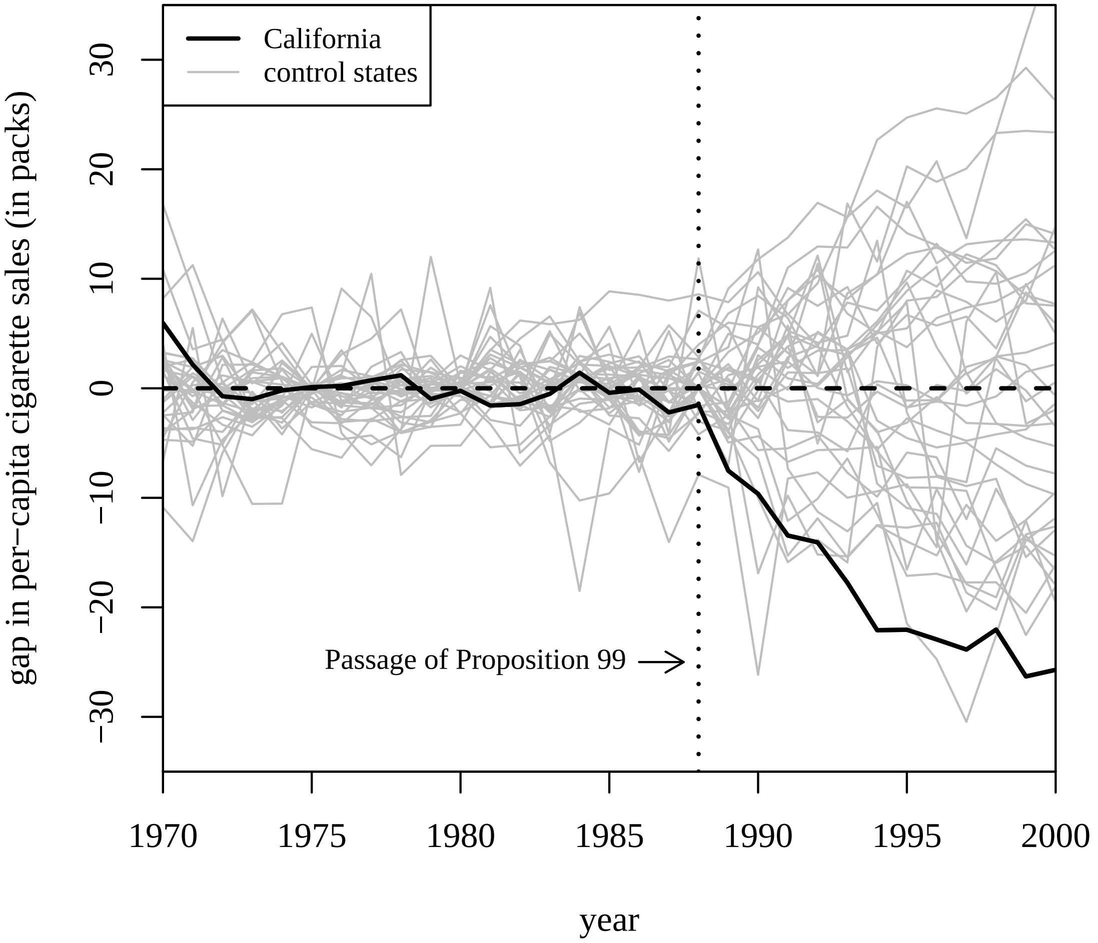

Figure 6: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 29 Control States (Discards States with Pre-Proposition 99 MSPE Five Times Higher than California's)

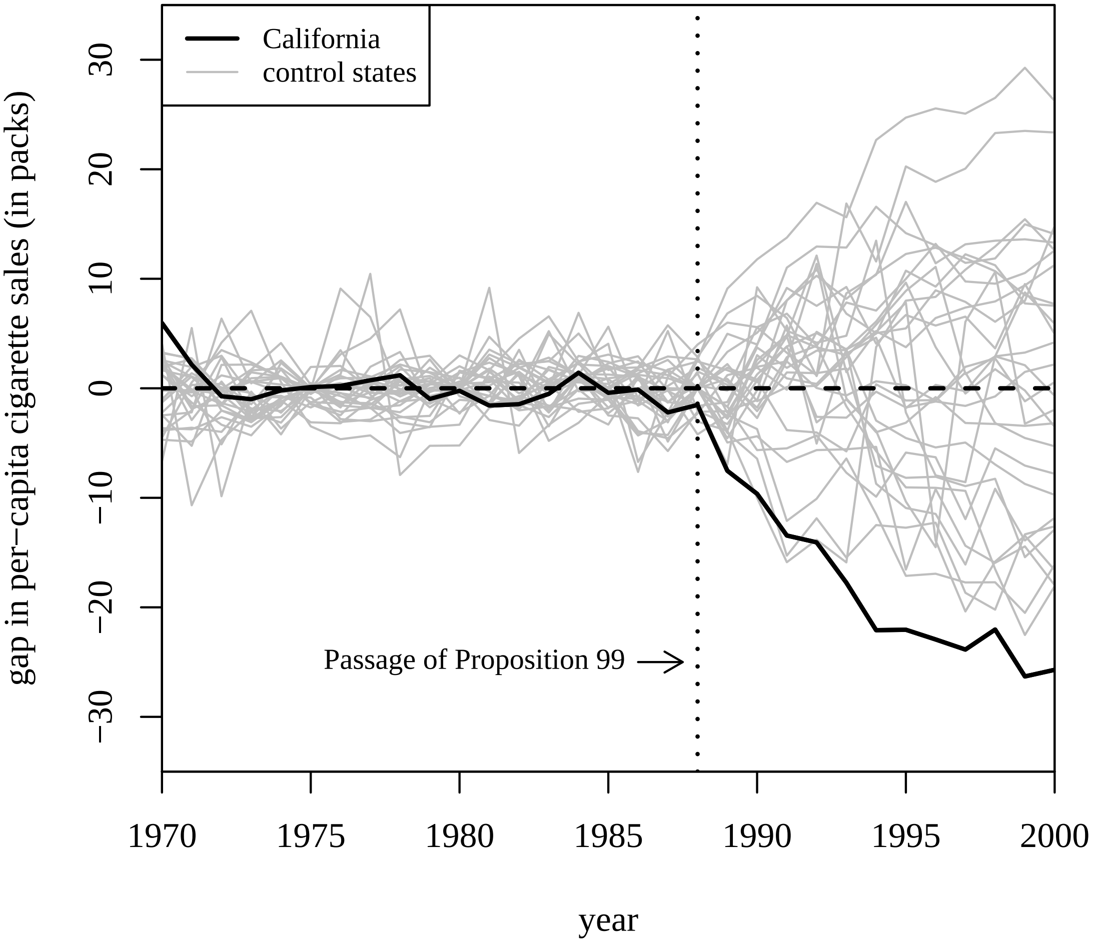

Figure 7: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 19 Control States (Discards States with Pre-Proposition 99 MSPE Two Times Higher than California's)

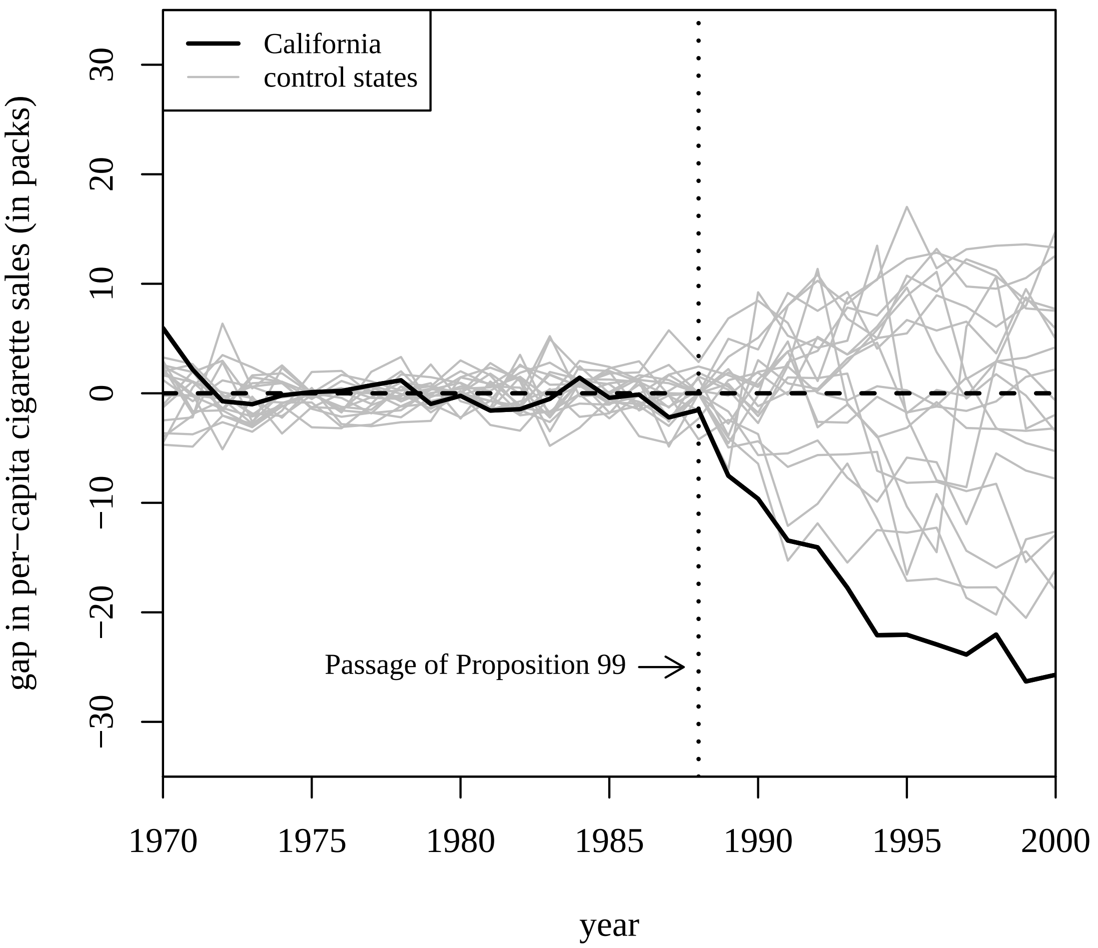

Figure 8: Ratio of Post-Proposition 99 MSPE and Pre-Proposition 99 MSPE: California and 38 Control States

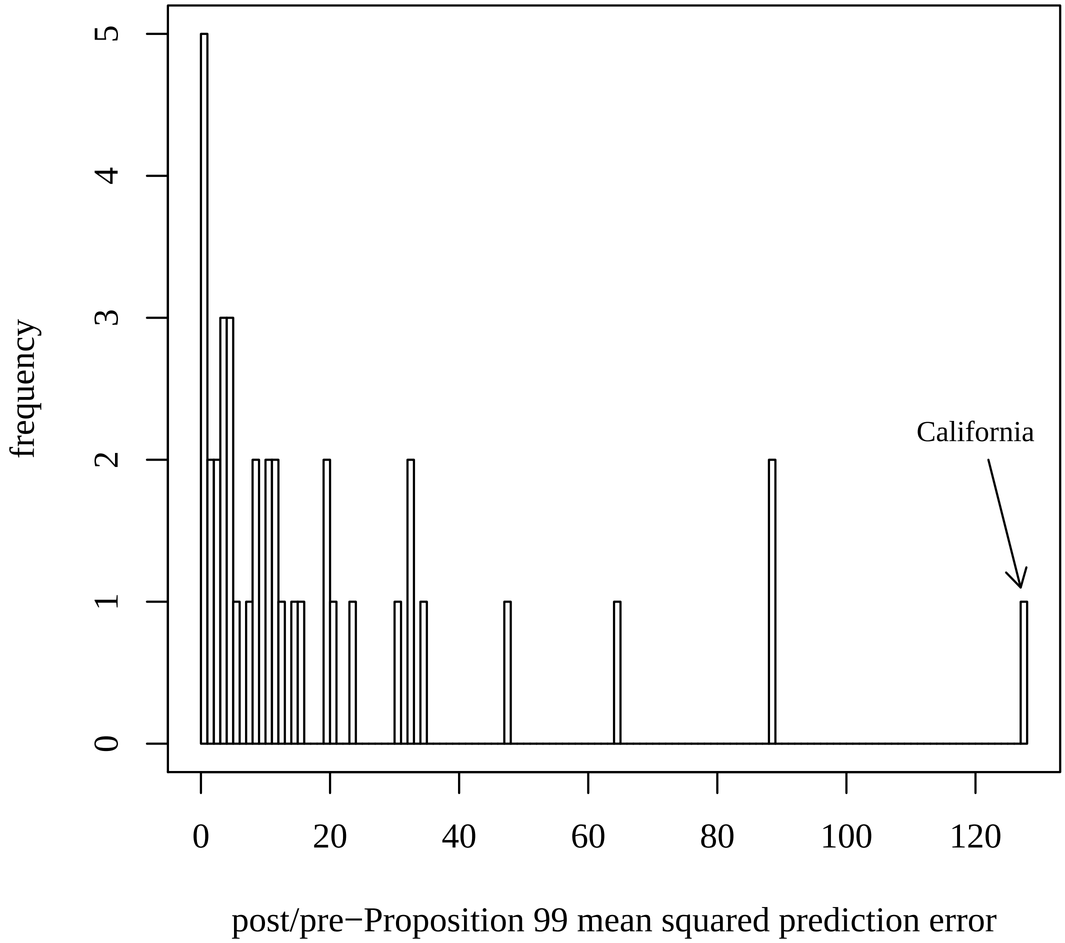

Table 1: Cigarette Sales Predictor Means

	California		Average of
Variables	Real	Synthetic	38 control states
Ln(GDP per capita)	10.08	9.86	9.86
Percent aged 15-24	17.40	17.40	17.29
Retail price	89.42	89.41	87.27
Beer consumption per capita	24.28	24.20	23.75
Cigarette sales per capita 1988	90.10	91.62	114.20
Cigarette sales per capita 1980	120.20	120.43	136.58
Cigarette sales per capita 1975	127.10	126.99	132.81

Note: All variables except lagged cigarette sales are averaged for the 1980-1988 period (beer consumption is averaged 1984-1988). Cigarette sales are measured in packs.

Table 2: State Weights in the Synthetic California

State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	-	Nebraska	0
Arizona	-	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	-
Connecticut	0.069	New Mexico	0
Delaware	0	New York	-
District of Columbia	-	North Carolina	0
Florida	-	North Dakota	0
Georgia	0	Ohio	0
Hawaii	-	Oklahoma	0
Idaho	0	Oregon	-
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	-	Vermont	0
Massachusetts	-	Virginia	0
Michigan	-	Washington	-
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0