

NBER WORKING PAPER SERIES

SYNTHETIC CONTROL METHODS FOR COMPARATIVE CASE STUDIES:
ESTIMATING THE EFFECT OF CALIFORNIA'S TOBACCO CONTROL PROGRAM

Alberto Abadie
Alexis Diamond
Jens Hainmueller

Working Paper 12831
<http://www.nber.org/papers/w12831>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2007

All authors are affiliated with Harvard's Institute for Quantitative Social Science (IQSS). We thank Jake Bowers, Dan Hopkins, and seminar participants at the 2006 APSA Meetings in Philadelphia for helpful comments. Funding for this research was generously provided by NSF grant SES-0350645 (Abadie). The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2007 by Alberto Abadie, Alexis Diamond, and Jens Hainmueller. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

Alberto Abadie, Alexis Diamond, and Jens Hainmueller

NBER Working Paper No. 12831

January 2007

JEL No. C21,C23,H75,I18,K32

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. 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. The methods proposed in this article produce informative inference regardless of the number of available comparison units, the number of available time periods, and whether the data are individual (micro) or aggregate (macro). Software to compute the estimators proposed in this article is available at the authors' web-pages.

Alberto Abadie
John F. Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
and NBER
alberto_abadie@harvard.edu

Jens Hainmueller
Department of Government
Harvard University
1737 Cambridge Street
Cambridge, MA 02138
jhainm@fas.harvard.edu

Alexis Diamond
Political Economy and Government
Harvard University
1737 Cambridge Street
Cambridge, MA 02138
adiamond@fas.harvard.edu

I. INTRODUCTION

Economists and other 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.

Because the outcome of interest is an aggregate, this type of study requires only aggregate data, which are widely available (for example, at the school, city, or region level). 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 outcomes of interest (like in Card, 1990, and Card and Krueger, 1994).

Given the widespread availability of aggregate/macro data, 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 economics and other social sciences, perhaps because the 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, if the analysis employs aggregate data, traditional inferential techniques, which measure uncertainty about aggregate parameters, do not apply. Then, it becomes apparent that an additional source of uncertainty should be taken into account to inform inference: the 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. Notice that the use of individual micro data, as opposed to aggregate data, only increases the amount of uncertainty if the outcome of interest is an aggregate.

This article addresses current methodological shortcomings of case study analysis. We advocate the use of data-driven procedures to construct adequate 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, however, 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 regions often provides a better comparison for the region exposed to the intervention than any single region 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.

We describe a simple econometric model that justifies the synthetic control approach and demonstrate that the conditions of the model are more general than the conditions under which traditional linear panel data or difference-in-differences estimators are valid. In addition, we propose new methods that allow researchers to perform inferential exercises about the effects of the event or intervention of interest that are valid regardless of the

number of available comparison units, the number of available time periods, and whether aggregate or individual data are used for the analysis.

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 paper, we demonstrate the statistical significance of our estimates.

Cross-country regressions are often criticized because they put countries side-by-side regardless of whether they have similar or radically different characteristics (see, for example, Temple, 1999). The synthetic control method provides an appealing data-driven procedure to select control groups for the study of the effects of events or interventions that take place at the level of a country. To illustrate the application of the techniques proposed in this article to cross-country data, we include an appendix where we use the synthetic control method to estimate the economic impact of the 1990 German unification in West Germany.

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 the application of the synthetic control method to the study of the economic effects of the German reunification.

II. SYNTHETIC CONTROL METHODS FOR COMPARATIVE CASE STUDIES

A. *Comparative Case Studies*

A case study focuses on a particular occurrence of the event or intervention 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. For example, in his classic study of the economic impact of immigration, Card (1990) analyzes the behavior of the Miami labor market in the wake of the 1980 Mariel Boatlift, when the Mariel immigrants increased the size of the labor force in Miami by 7 percent in a matter of a few months. In comparative case studies, researchers compare units affected by the event or intervention of interest to a group of unaffected 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).¹

To simplify the exposition, we proceed as if only one unit or region is subject to the intervention of interest.² 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. Also without loss of generality and to simplify notation, we 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. We can always decompose Y_{it}^N as:

$$Y_{it}^N = \delta_t + \nu_{it},$$

¹In comparative case studies, the main emphasis is usually on the identification of the impact of the particular event or intervention on hand (internal validity), sometimes at the cost of limited immediate generalizability to other settings (external validity). The case study method is inductive by nature: it attempts to learn about general phenomena by looking closely at particular instances of those phenomena.

²Otherwise, we could first aggregate the data from the regions exposed to the intervention.

for units $i = 1, \dots, J + 1$, and time periods $t = 1, \dots, T$, where δ_t is a common time period effect, and ν_{it} has mean equal to zero for all t . Let T_0 be 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 $t \in \{1, \dots, T_0\}$ and all $i \in \{1, \dots, N\}$, we have that $Y_{it}^I = Y_{it}^N$.³ Let $\alpha_{it} = Y_{it}^I - Y_{it}^N$ be the effect of the intervention for unit i at time t , if unit i is exposed to the intervention in periods $T_0 + 1, T_0 + 2, \dots, T$ (where $1 \leq T_0 < T$). Therefore:

$$Y_{it}^I = \delta_t + \alpha_{it} + \nu_{it}.$$

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

$$Y_{it} = \delta_t + \alpha_{it}D_{it} + \nu_{it}.$$

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 $(\alpha_{1T_0+1}, \dots, \alpha_{1T})$. 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 ν_{it} is given by a factor model:

$$\nu_{it} = Z_i\theta_t + \lambda_t\mu_i + \varepsilon_{it}, \tag{1}$$

where Z_i is a $(1 \times r)$ vector of observed covariates (not affected by the intervention), θ_t is a $(r \times 1)$ vector of unknown parameters, μ_i is an unobserved unit-specific effect, λ_t is an

³Of course, this is done without loss of generality. If the anticipation of the intervention impacts the outcome before the intervention is implemented, we can always redefine T_0 to be the first period in which the outcome may possibly react to the (anticipated) intervention.

unknown common factor, and the error terms ε_{it} are unobserved transitory shocks at the region level with zero mean for all i .⁴ It is important to realize that this model does not rule out the existence of time-varying measured determinants of ν_{it} . The vector Z_i may contain pre- and post-intervention values of time-varying variables, as long as they are not affected by the intervention. For example, if $T = 2$, $T_0 = 1$, $Z_i = (Z_{i1}, Z_{i2})$,

$$\theta_1 = \begin{pmatrix} \beta \\ 0 \end{pmatrix} \quad \text{and} \quad \theta_2 = \begin{pmatrix} 0 \\ \beta \end{pmatrix},$$

then $Z_i\theta_t = Z_{it}\beta$. Notice also that the model in (1) does not restrict Z_i , μ_i , and ε_{it} to be independent.

Consider a $(J \times 1)$ vector of weights $W = (w_2, \dots, w_{J+1})'$ such that $w_j \geq 0$ for $j = 2, \dots, J+1$ and $w_2 + \dots + w_{J+1} = 1$. 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:

$$\sum_{j=2}^{J+1} w_j Y_{jt} = \delta_t + \sum_{j=2}^{J+1} w_j \nu_{jt}.$$

Let $K = (k_1, \dots, k_{T_0})'$ be a $(T_0 \times 1)$ vector. Consider an arbitrary linear combination of pre-intervention outcomes: $\bar{Y}_i^K = \sum_{s=1}^{T_0} k_s Y_{is}$. For example, if $k_1 = k_2 = \dots = k_{T_0-1} = 0$ and $k_{T_0} = 1$, then $\bar{Y}_i^K = Y_{iT_0}$, the value of the outcome variable in the period previous to the intervention. Alternatively, if $k_1 = k_2 = \dots = k_{T_0} = 1/T_0$, then $\bar{Y}_i^K = T_0^{-1} \sum_{s=1}^{T_0} Y_{is}$, the simple average of the outcome variable in the pre-intervention periods. Notice that

$$\begin{aligned} \sum_{j=2}^{J+1} w_j \bar{Y}_j^K - \bar{Y}_1^K &= \sum_{j=2}^{J+1} w_j \sum_{s=1}^{T_0} k_s Y_{js} - \sum_{s=1}^{T_0} k_s Y_{1s} \\ &= \left(\sum_{s=1}^{T_0} k_s \delta_s + \sum_{j=2}^{J+1} w_j \sum_{s=1}^{T_0} k_s \nu_{js} \right) - \left(\sum_{s=1}^{T_0} k_s \delta_s + \sum_{s=1}^{T_0} k_s \nu_{1s} \right) \\ &= \sum_{j=2}^{J+1} w_j \sum_{s=1}^{T_0} k_s (\nu_{js} - \nu_{1s}). \end{aligned}$$

⁴Without loss of generality, assume that Z_i and μ_i have zero mean. Otherwise, we can always include $\bar{Z}_i\theta_t + \lambda_t\bar{\mu}_i$ in δ_t (where \bar{Z}_i and $\bar{\mu}_i$ are the averages of Z_i and μ_i over the $J+1$ regions), so ν_{it} has mean equal to zero for each t .

Therefore,

$$\begin{aligned} \sum_{j=2}^{J+1} w_j \bar{Y}_j^K - \bar{Y}_1^K &= \left(\sum_{j=2}^{J+1} w_j Z_j - Z_1 \right) \left(\sum_{s=1}^{T_0} k_s \theta_s \right) \\ &+ \left(\sum_{j=2}^{J+1} w_j \mu_j - \mu_1 \right) \left(\sum_{s=1}^{T_0} k_s \lambda_s \right) + \sum_{j=2}^{J+1} w_j \sum_{s=1}^{T_0} k_s (\varepsilon_{js} - \varepsilon_{1s}). \end{aligned}$$

In addition,

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

Suppose that we can choose $(w_2^*, \dots, w_{J+1}^*)'$ such that:

$$\sum_{j=2}^{J+1} w_j^* \bar{Y}_j^K = \bar{Y}_1^K, \quad \text{and} \quad \sum_{j=2}^{J+1} w_j^* Z_j = Z_1. \quad (2)$$

Then, if $\sum_{s=1}^{T_0} k_s \lambda_s \neq 0$, we obtain:

$$\sum_{j=2}^{J+1} w_j^* Y_{jt} = Y_{1t}^N - \frac{\lambda_t}{\sum_{s=1}^{T_0} k_s \lambda_s} \sum_{j=2}^{J+1} w_j^* \sum_{s=1}^{T_0} k_s (\varepsilon_{js} - \varepsilon_{1s}) + \sum_{j=2}^{J+1} w_j^* (\varepsilon_{jt} - \varepsilon_{1t}).$$

Therefore, for $t \in \{T_0 + 1, \dots, T\}$,

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

is an unbiased estimator of α_{1t} .

Equation (2) can hold exactly only if (\bar{Y}_1^K, Z_1) belongs to the convex hull of $\{(\bar{Y}_2^K, Z_2), \dots, (\bar{Y}_{J+1}^K, Z_{J+1})\}$. 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.

This simple model can be easily extended in several directions. Notice first that equation (1) assumes the existence of a single unobserved factor. Multiple unobserved factors can be controlled for by using as many (linearly independent) combinations of pre-intervention

outcomes in the first part of equation (2), instead of a single one, to select the synthetic control. If M linear combinations of pre-intervention outcomes, $\bar{Y}_i^{K_1}, \dots, \bar{Y}_i^{K_M}$, are used to select the synthetic control, the first part of equation (2) becomes, $\sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_1} = \bar{Y}_1^{K_1} \dots \sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_M} = \bar{Y}_1^{K_M}$.⁵

Moreover, the simple linear model presented in this section does not need to 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 sample to regions with (\bar{Y}_j^K, Z_j) sufficiently close to (\bar{Y}_1^K, Z_1) under some distance metric. As explained below, in contrast with more traditional regression methods, which typically rely on asymptotic limit theorems for inference, the availability of a small number of regions to construct the synthetic control does not invalidate our inferential procedures.

On the other hand, it is worth noticing that, even taken at face value, equation (1) generalizes the usual difference-in-differences (fixed-effects) model commonly applied in the empirical literature. The difference-in-differences model allows for the presence of unobserved confounders but restricts the effect of those confounders to be constant in time. In contrast, the model presented in this section allows the effects of confounding unobserved characteristics to vary with time. Notice that 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 .

Relative to more 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 regions with weights that sum to one, the synthetic control method makes explicit (1) the relative contribution of each control region to the counterfactual of interest; and (2) the similarity of the synthetic con-

⁵More precisely, suppose that ν_{it} includes F unobserved factors, so for $1 \leq i \leq J+1$ and $1 \leq t \leq T$, $\nu_{it} = Z_i \theta_t + \lambda_t \mu_i + \varepsilon_{it}$, where $\lambda_t = (\lambda_{t1}, \dots, \lambda_{tF})$ is a $(1 \times F)$ vector of unknown common factors and μ_i is an $(F \times 1)$ vector of factor loadings. Consider M vectors of dimension $(T_0 \times 1)$, K_m , with $1 \leq m \leq M$ and s -element equal to k_{ms} . Let $\bar{Y}_i^{K_m} = \sum_{s=1}^{T_0} k_{ms} Y_{is}$, and suppose that we can choose $(w_2^*, \dots, w_{J+1}^*)$ such that $\sum_{j=2}^{J+1} w_j^* Z_j = Z_1$ and $\sum_{j=2}^{J+1} w_j^* \bar{Y}_j^{K_m} = \bar{Y}_1^{K_m}$, for all $m \in \{1, \dots, M\}$. If the $(M \times F)$ matrix with (m, f) -element $\sum_{s=1}^{T_0} k_{ms} \lambda_{sf}$ has full column rank, then $\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$ is an unbiased estimator of α_{1t} , for $t \in \{T_0 + 1, \dots, T\}$.

trol and the treated region in terms of pre-intervention outcomes and other predictors of post-intervention outcomes.

Finally, it is worth noticing that because the choice of a synthetic control does not require that the researchers have access to post-intervention outcomes, the synthetic control method allows researchers to decide on study design issues 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.

C. Implementation

Assume that there are J regions not exposed to the intervention so they can serve as controls. We consider any weighted average of non-exposed regions as a potential (synthetic) control. Let W be a $(J \times 1)$ vector of positive weights that sum to one. That is, $W = (w_2, \dots, w_{J+1})'$ with $w_j \geq 0$ for $j = 2, \dots, J + 1$ and $w_2 + \dots + w_{J+1} = 1$. Each value of W represents a weighted average of the available control regions and, therefore, a synthetic control.

The outcome variable of interest is observed for T periods for the region affected by the intervention Y_{1t} , ($t = 1, \dots, T$) and the unaffected regions Y_{jt} , ($t = 1, \dots, T, j = 2, \dots, J + 1$). Let T_0 be the number of pre-intervention periods and $T_1 = T - T_0$ 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 $X_1 = (Z_1, \bar{Y}_1^{K_1}, \dots, \bar{Y}_1^{K_M})'$ 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 $(Z_j, \bar{Y}_j^{K_1}, \dots, \bar{Y}_j^{K_M})'$. The vector W^* is chosen to minimize some distance (or pseudo-distance), $\|X_1 - X_0W\|$, between X_1 and X_0W , subject to $w_2 \geq 0, \dots, w_{J+1} \geq 0, w_2 + \dots + w_{J+1} = 1$. In particular, we will consider $\|X_1 - X_0W\|_V = \sqrt{(X_1 - X_0W)'V(X_1 - X_0W)}$, where V is some $(k \times k)$

symmetric and positive semidefinite matrix, although other choices are also possible.

Although our inferential procedures are valid for any choice of V , the choice of V influences the mean square error of the estimator (that is, the expectation of $(Y_1 - Y_0W^*)'(Y_1 - Y_0W^*)$). Because V is symmetric and positive semidefinite, there exist two $(k \times k)$ matrices, U and A , such that the rows of U , $\{u_n\}_{n=1}^k$, form an orthonormal basis of \mathbb{R}^k , A is diagonal with all diagonal elements, $\{a_{nn}\}_{n=1}^k$, equal or greater than zero, and $V = U'AU$. As a result, the vector W^* minimizes $(H_1 - H_0W)'A(H_1 - H_0W)$, subject to $w_2 \geq 0, \dots, w_{J+1} \geq 0$, $w_2 + \dots + w_{J+1} = 1$, where $H_1 = UX_1$ and $H_0 = UX_0$. In other words, the matrix V assigns weight a_{nn} to the linear combination of characteristics in X_0 and X_1 with coefficients u_n . To minimize the mean square error of our estimators we want to choose a matrix V that assigns large weights to combinations of the pre-intervention variables that have large predictive power on the outcome variable.⁶ Of course, 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 affected region in the pre-intervention periods. For example, Abadie and Gardeazabal (2003) 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. 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.⁷

⁶For example, it is easy to check that if $\theta_t = \theta$ and $\lambda_t = \lambda$ for all $t > T_0$, and ε_{it} are identically distributed, then the mean square error is minimized with $V = (\theta', \lambda)' \cdot (\theta', \lambda)$, and the only linear combination of covariates that is assigned a positive weight has coefficients $(\theta', \lambda) / \|(\theta', \lambda)\|$.

⁷In cases when little is known about the relative predictive power of the pre-intervention variables, researchers may decide to normalize the variables in X_0 and X_1 using V equal to the inverse of the estimated variance-covariance matrix of the variables in X_0 and X_1 (Mahalanobis distance) or equal to a diagonal matrix with the elements in the main diagonal equal to the inverses of the sample variances of the variables in X_0 and X_1 (normalized Euclidean distance).

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 the value of aggregate data does not reduce to zero our 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. Even when aggregate data are available, there is still an additional source of uncertainty that should be taken into account to inform inference. This additional uncertainty is the uncertainty 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 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 produce informative inference regardless of the number of available comparison units, the number of

available time periods, and whether the data are individual (micro) or aggregate (macro). However, the quality of some of the inferential exercises proposed in this article increases with the number of available comparison units. 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. Starting in the late 1960's, the Basque clandestine organization ETA implemented a terrorist campaign that lasted more than 30 years and resulted in more than 800 deaths. Although terrorist attacks took place in almost every region of Spain, most of the attacks and casualties occurred within the Basque Country. To estimate the economic growth that the Basque Country would have experienced in the absence of terrorism, Abadie and Gardeazabal (2003) construct a synthetic control region as the combination of other regions in Spain that best reproduced the values of economic growth predictors for the Basque Country at the start of the terrorist campaign. Abadie and Gardeazabal (2003) show that, during the terrorism years, per capita income in the synthetic Basque Country without terrorism was up to 12 percent higher than per capita income in the actual Basque Country with 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. In contrast to the Basque Country, per capita income in the synthetic Catalonia closely tracked the observed per capita income in Catalonia. This exercise demonstrates that a combination of other Spanish regions chosen to match the economic characteristics of Catalonia before the outbreak of terrorism provides an adequate control for Catalonia.

In this paper, we extend the idea of a placebo study to produce quantitative inference in comparative case studies. As in classical 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. By construction, this exercise produces exact inference regardless of the number of available comparison regions, time periods, and whether the data are individual or aggregate. However, inference becomes more informative if the number of potential comparison regions is large.

For cases in which the number of available comparison regions is small, one can use the longitudinal dimension of the data to produce additional placebo studies, as in Bertrand, Duflo, and Mullainathan (2004) where the dates of the placebo interventions are set at random.^{8, 9}

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.¹⁰ 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

⁸See Appendix B for an application of an in-time placebo to the study of the economic impact of the 1990 German reunification.

⁹See Donald and Lang (2006) for related work on inference in difference-in-differences models.

¹⁰See Gruber (2001) for a survey on tobacco consumption and regulation in the United States. Ireland imposed a workplace smoking ban in 2004. This was followed by Italy in 2005, and Scotland in 2006. Belgium, Australia, and the United Kingdom have workplace smoking bans scheduled for 2007 (Borio, 2005).

state (Siegel, 2002).¹¹ 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 spawned 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 Health Education Account funds 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

¹¹Proposition 99 assigned tax revenues to six accounts: Physician Services (35 percent), Health Education(20 percent), Hospital Services (10 percent), Research (5 percent), Public Resources (5 percent), and Unallocated (25 percent) (Glantz and Balbach 2000).

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 1995, with a 50-cent tax increase, and Oregon in 1997, 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 October 6, 2006, there were 17 states and the District of Columbia and 519 municipalities across the country with laws in effect requiring 100 percent smoke-free workplaces, bars, or restaurants. Similar laws have been enacted but are not yet effective in other states (ANRF, 2006).

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.¹²

The most recently published study similar to ours is Fichtenberg and Glantz (2000), in the *New England Journal of Medicine*. This article uses 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 smoking data 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

¹²See also Goel and Nelson (2006) for a recent literature review on the effectiveness of anti-smoking legislation.

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. Borrowing from the statistical matching literature, we refer to the set of potential controls for California as the “donor pool”. Because the synthetic California is meant to reproduce the smoking rates that would have been observed for California in the absence of Proposition 99, it is imperative that 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).¹³ 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 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 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. 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

¹³Notice that, even if the remaining tax increases substantially reduced smoking in any of the control states that gets assigned a positive W -weight, this should if anything attenuate the treatment effect estimate that we obtain for California.

years of lagged smoking consumption (1975, 1980, and 1988). Appendix A provides data sources.^{14, 15}

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 an adequate 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 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

¹⁴Average retail prices of cigarettes vary quite a bit across the United States. For example, in 1989, average retail prices ranged from \$1.16 in Kentucky to \$1.74 in Nevada.

¹⁵Results are robust regardless of which and how many predictor variables we include. The list of predictors used for robustness checks include: unemployment, income inequality, poverty, welfare transfers, crime rates, drug related arrest rates, state cigarette taxes, population density, and numerous variables to capture the demographic, racial, and social structure of states. Inclusion of these predictors leaves our results virtually unaffected. The weights associated with additional predictors in the matrix V usually are close to zero because the few predictors used in the streamlined baseline model already account for most of the variation in cigarette consumption over time.

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 pretreatment 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 an adequate control group for California. In particular, prior to the passage of Proposition 99 average beer consumption and cigarette retail prices were lower in the 38 control states than in California. Moreover, prior to the passage of Proposition 99 average cigarette sales per-capita were substantially higher 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 99.¹⁶

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 reproduce extremely

¹⁶Table 1 highlights an attractive 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).

well the trajectory of this variable in California for the entire pre-Proposition 99 period. Combined with the high balance on all smoking predictors (Table 1), this suggests that the synthetic California provides a sensible approximation to the per-capita cigarette packs 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 began to diverge noticeably. While cigarette consumption in the synthetic California continued on its moderate downward trend, the real California experienced a sharp downward kink. 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.

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.¹⁷

¹⁷Part 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.

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

To evaluate the statistical 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 we interpret 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, we interpret that our analysis provides significant evidence of a negative effect of Proposition 99 on cigarette sales in California.

The idea of the placebo test 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 (see, for example, Lehmann, 1997).

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. 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 3,437. 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 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, and 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 higher than five times the pre-Proposition 99 MSPE for 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 6 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 6, 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 implementations of comparative case studies have been plagued by inferential challenges and ambiguity about the choice of valid control groups. Building on an idea in Abadie and Gardeazabal (2003), this paper proposes the use of synthetic control methods to overcome these shortcomings. Our method allows for causal inference in observational settings with a single treated unit. We use a data-driven procedure to construct a weighted combination of potential comparison regions that approximate the most relevant characteristics of the units exposed to the intervention.

We show that the conditions under which the synthetic control estimator is valid are more general than the conditions required for traditional linear panel data or difference-in-differences estimators. In addition, we propose a method to produce informative inference regardless of the number of available comparison units, the number of available time periods, and whether the data are individual (micro) or aggregate (macro). 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 re-label 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. 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: THE ECONOMIC IMPACT OF THE GERMAN REUNIFICATION
IN WEST GERMANY

In this appendix, we illustrate the application of the synthetic control method to cross-country data. For this purpose, we apply the synthetic control method to estimate the economic impact of the 1990 German reunification in the former West Germany (Federal Republic of Germany). Using the synthetic control method, we construct a synthetic West Germany as a convex combination of other advanced industrialized countries chosen to resemble the values of economic growth predictors for West Germany prior to the reunification. Our sample of potential controls includes Australia, Austria, Belgium, Canada, Denmark, Finland, France, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and the United States.

We provide a list of all variables used in the analysis at the end of this appendix, along with data sources. For each variable we have made sure that the German data refers exclusively to the territory of the former West Germany. For that purpose, when necessary, our dataset was supplemented with data from the German Federal Statistical Office (*Statistisches Bundesamt*). The outcome variable is real per-capita GDP (PPP-adjusted, measured in 2002 U.S. Dollars). We rely on a standard set of predictors commonly used in the economic growth literature. We use the investment rate measured as the ratio of real domestic investment (private plus public) to GDP. Our proxy for human capital is the percentage of high school graduates in the total population aged 25 and older. Our predictors also include the inflation rate and the share of value added by the industrial sector. In addition, we include trade openness, measured by the sum of exports and imports as a percentage of GDP, among our predictors of economic growth. The inclusion of additional growth predictors did not change our results substantively.

We chose the synthetic West Germany to approximate the values of the economic growth predictors for the actual West Germany prior to the German reunification. Table A.1 compares the pre-reunification characteristics of the actual West Germany to those of our synthetic West Germany, and also to those of the population-weighted average of the 20

OECD countries in the donor pool. The statistics in this table show that the synthetic West Germany approximates the pre-1990 values of the economic growth predictors for West Germany far more accurately than the average of OECD countries. This suggests that the synthetic West Germany provides a better control for the actual West Germany than the average of other OECD countries.

Table A.2 shows the weights of each country in the synthetic West Germany. The synthetic West Germany is a weighted average of Austria, the United States, Switzerland, The Netherlands, and Japan, with weights decreasing in this order. All other countries in the donor pool are assigned zero weights.

Figure A.1 displays the GDP per-capita trajectory of West Germany and its synthetic counterpart for the 1960-2003 period. The synthetic West Germany almost exactly reproduces the trend of this variable in the actual West Germany for the entire pre-reunification period. This remarkable fit for the pre-treatment period, along with the high balance on GDP predictors (Table A.1), suggests that our synthetic West Germany provides a sensible estimate of the counterfactual GDP per-capita trend that West Germany would have experienced in the absence of the German reunification.

Our estimate of the effect of the German reunification on GDP per-capita in West Germany is given by the difference between the actual West Germany and its synthetic version. We find that the German reunification did not have much of an effect on West German GDP per-capita in the first two years immediately following reunification. In this initial period GDP per-capita in the synthetic West Germany is slightly lower than in the actual West Germany, which is broadly in line with arguments about an initial demand boom (see, for example, Meinhardt et al. (1995)). From 1992 onwards, however, the two lines diverge substantially. While the actual West Germany embarks on a recessionary trend, the trajectory of the synthetic West Germany keeps ascending, albeit at a slightly lower pace than in the pre-unification period. The divergence of the two series continues to grow until the end of the sample period. Our results suggest a pronounced negative effect of the reunification on West German income. They suggest that for the entire 1990-2003

period, GDP per-capita was reduced by about 1,460 USD per year on average, a decline of approximately 6 percent.

To evaluate the statistical significance of our estimate, we conduct a placebo study where the treatment is reassigned not across units but in time. Thus, instead of running a placebo study by reassigning reunification to different countries, we maintain West Germany as the treated unit but reassign reunification to an earlier point in time. The idea is to evaluate how likely it is to obtain results of the magnitude that we obtain in Figure A.1 when we apply our methods in a sample period before the reunification.

Here we show results for the case when reunification is reassigned to the year 1980, ten years earlier than it actually occurred. Thus, we run the same model as before, but the pre-treatment period is constrained to be 1960-1980. We also lag our predictors variables accordingly. The results for reassigning reunification to years other than 1980 are substantially identical. They are omitted to economize on space.

Figure A.2 displays the results of our in-time placebo study. Notice that the synthetic West Germany almost exactly reproduces the evolution of GDP per capita in the actual West Germany for the entire pre-treatment period. Most importantly, the GDP per-capita trajectories of West Germany and its synthetic counterpart do not diverge considerably after 1980. That is, in contrast to the actual 1990 German reunification, our 1980 placebo reunification has no perceivable effect. This placebo study shows that a weighted combination of other OECD countries can be used to predict accurately the evolution of GDP per-capita in West Germany prior to the reunification. This suggests that the gap estimated in Figure A.1 reflects the impact of the German reunification and not a potential lack of predictive power of the synthetic control.

The data sources employed for this application are:

- GDP per capita (PPP 2002 \$US). Source: OECD National Accounts (retrieved via the OECD Health Database in October 2006). Data for West Germany was obtained from Statistisches Bundesamt 2005 (Arbeitskreis “Volkswirtschaftliche Gesamtrechnungen der Länder”) and converted using PPP monetary conversion factors (retrieved from

the OECD Health Database in October 2006).

- Investment Rate: Ratio of real domestic investment (private plus public) to real GDP. Source: Barro and Lee (1994).
- Schooling: Percentage of secondary school completed in the total population aged 25 and older. Source: Barro and Lee (2000).
- Industry: industry share of value added. Source: World Bank WDI Database 2005 and Statistisches Bundesamt 2005.
- Inflation: annual percentage change in consumer prices (base year 1995). Source: World Development Indicators Database 2005 and Statistisches Bundesamt 2005.
- Trade Openness: Export plus Imports as percentage of GDP. Source: World Bank: World Development Indicators CD-ROM 2000.

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.
- ANRF (2006), "Municipalities with 100% Smokefree Laws", Available at: <http://www.no-smoke.org/pdf/100ordlisttabs.pdf>. Accessed on Nov. 15, 2006.
- BARRO, R.J. and LEE, J. (2000), "International Data on Educational Attainment: Updates and Implications," *CID Working Paper No. 42, April 2000 – Human Capital Updated Files*.
- 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.
- BORIO, G. (2005), *Tobacco Timeline: Chapter Eight, The Twenty-First Century, the New Millennium*, http://www.tobacco.org/resources/history/Tobacco_History21.html.
- 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.
- 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.
- DONALD, S.G. and LANG, K. (2006), "Inference with Difference in Differences and Other Panel Data", *Review of Economics and Statistics*, (forthcoming).
- 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/>.
- GOEL, R. K. and NELSON, M. (2006) "The Effectiveness of Anti-Smoking Legislation: A Review", *Journal of Economic Surveys*, vol. 20, no. 3, 325-355.
- GRUBER, J. (2001), "Tobacco at the Crossroads: The Past and Future of Smoking Regulation in the United States", *Journal of Economic Perspectives*, vol. 15, no. 2, 193-212.
- 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.
- KING, G. and ZHENG, L. (2006), "The Dangers of Extreme Counterfactuals", *Political Analysis*, vol. 14, no. 2, 131-159.
- LEHMANN, E.L. (1997), *Testing Statistical Hypotheses* (2nd edition). Berkeley: University of California Press.
- 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.
- MEINHARDT, V., SEIDEL, B., STILE, F., and TEICHMANN, D., (1995), "Transferleistungen in die neuen Bundesländer und deren wirtschaftliche Konsequenzen", *Deutsches Institut für Wirtschaftsforschung, Sonderheft 154*.
- ORZECZOWSKI 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.
- 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.

TEMPLE, J. (1999), "The New Growth Evidence." *Journal of Economic Literature*, vol. 37(1), 112-156.

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.

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

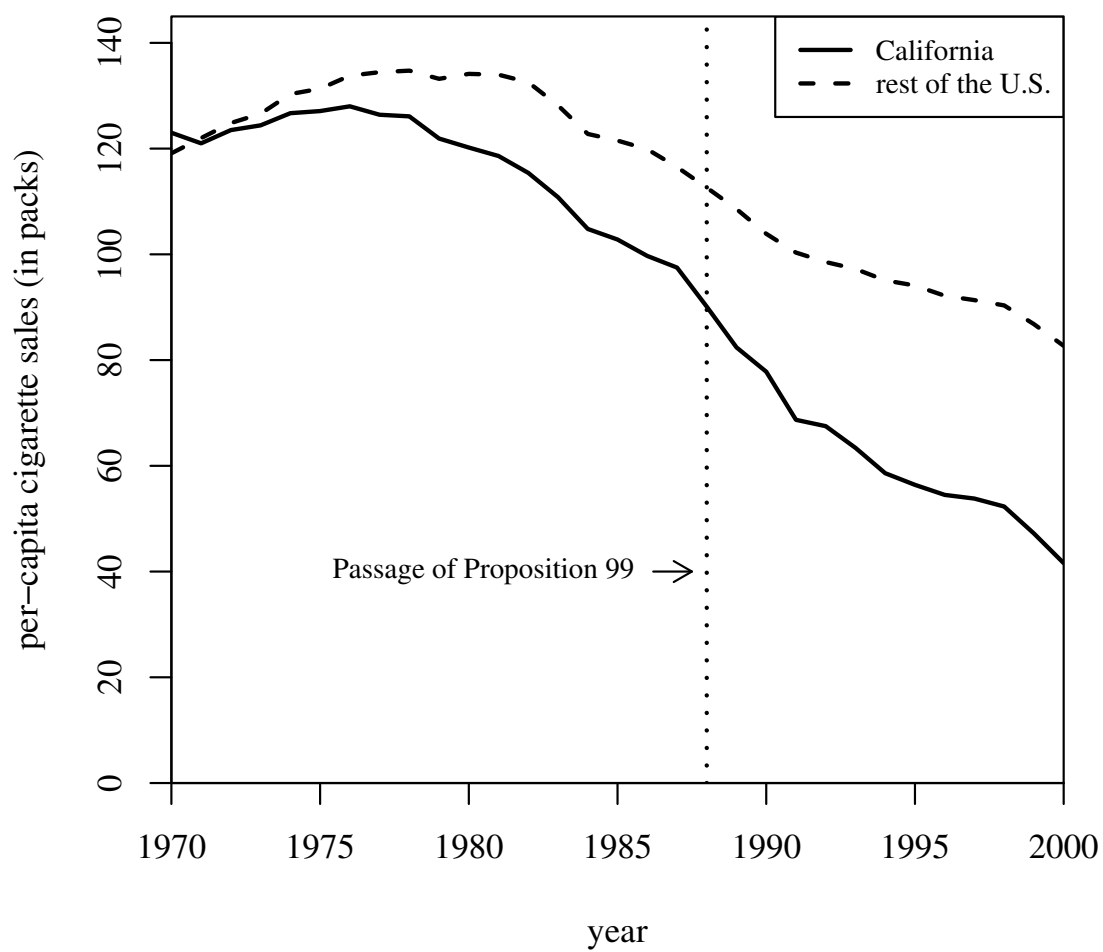


Figure 2: Trends in Per-Capita Cigarette Sales: California vs. synthetic California

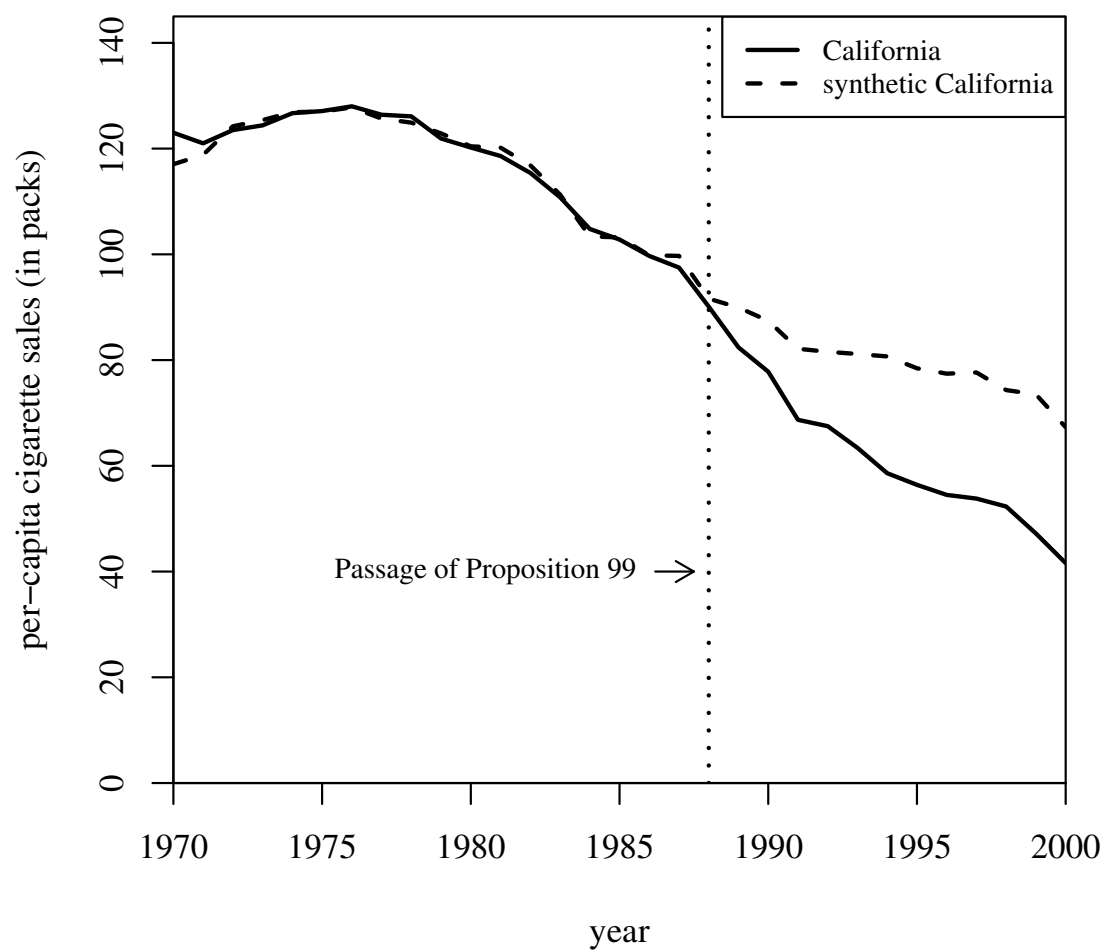


Figure 3: Per-Capita Cigarette Sales Gap Between California and Synthetic California

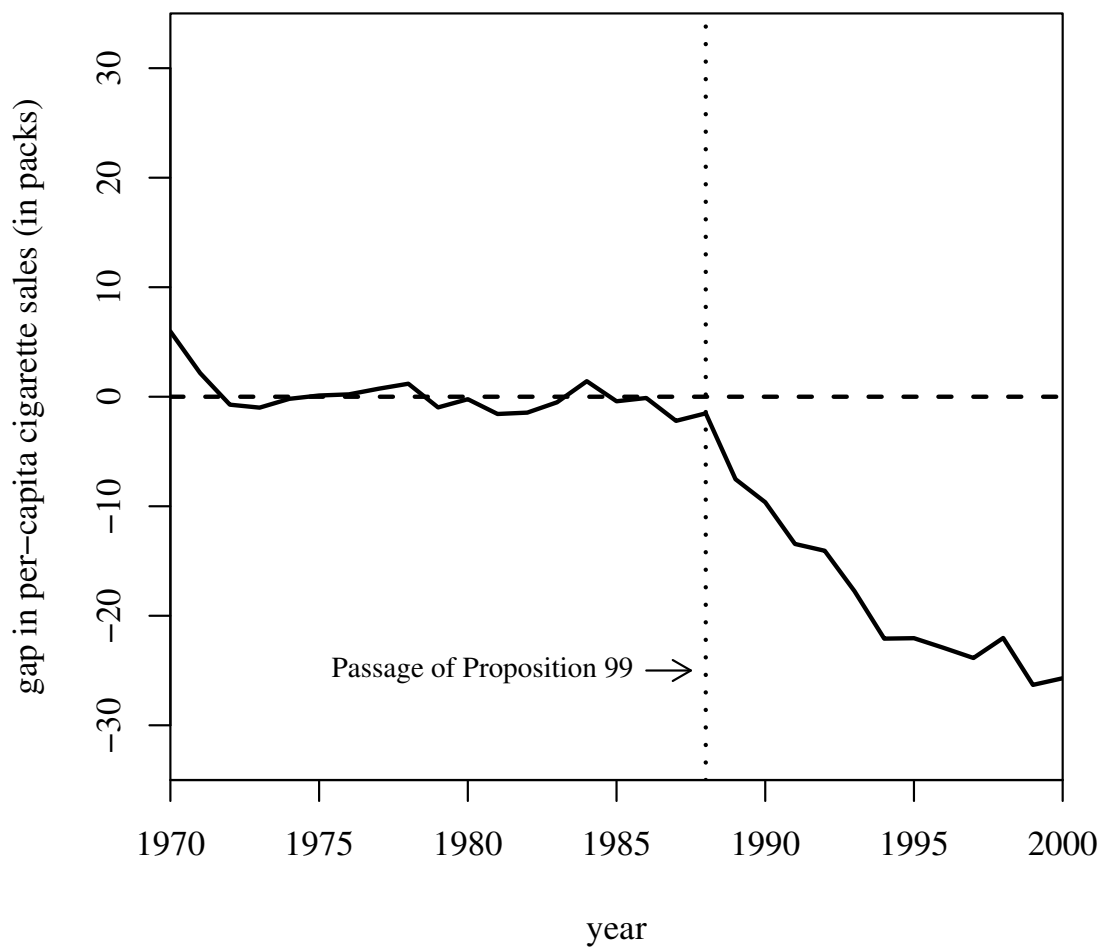


Figure 4: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in all 38 Control States

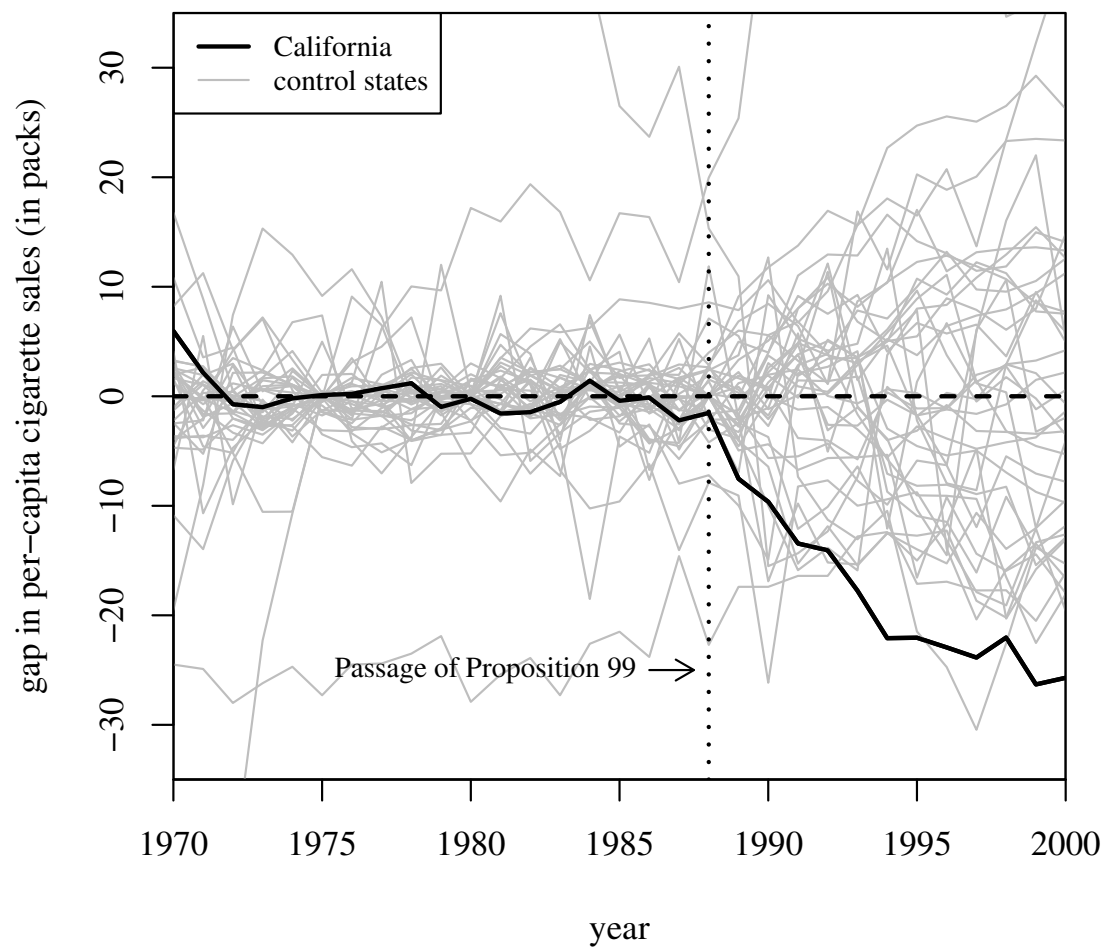


Figure 5: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 34 Control States (Discards States with Pre-Proposition 99 MSPE Twenty Times Higher than California's)

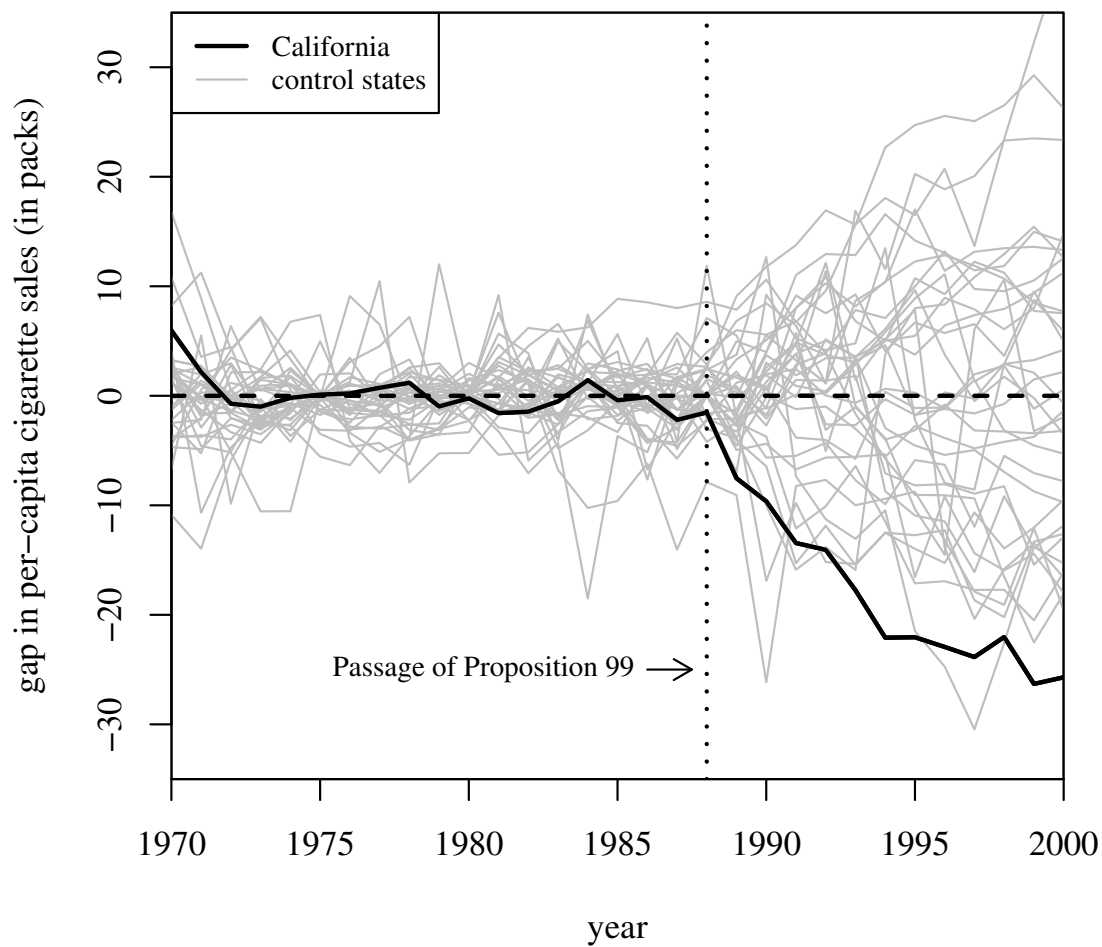


Figure 6: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 29 Control States (Discards States with Pre-Proposition 99 MSPE Five Times Higher than California's)

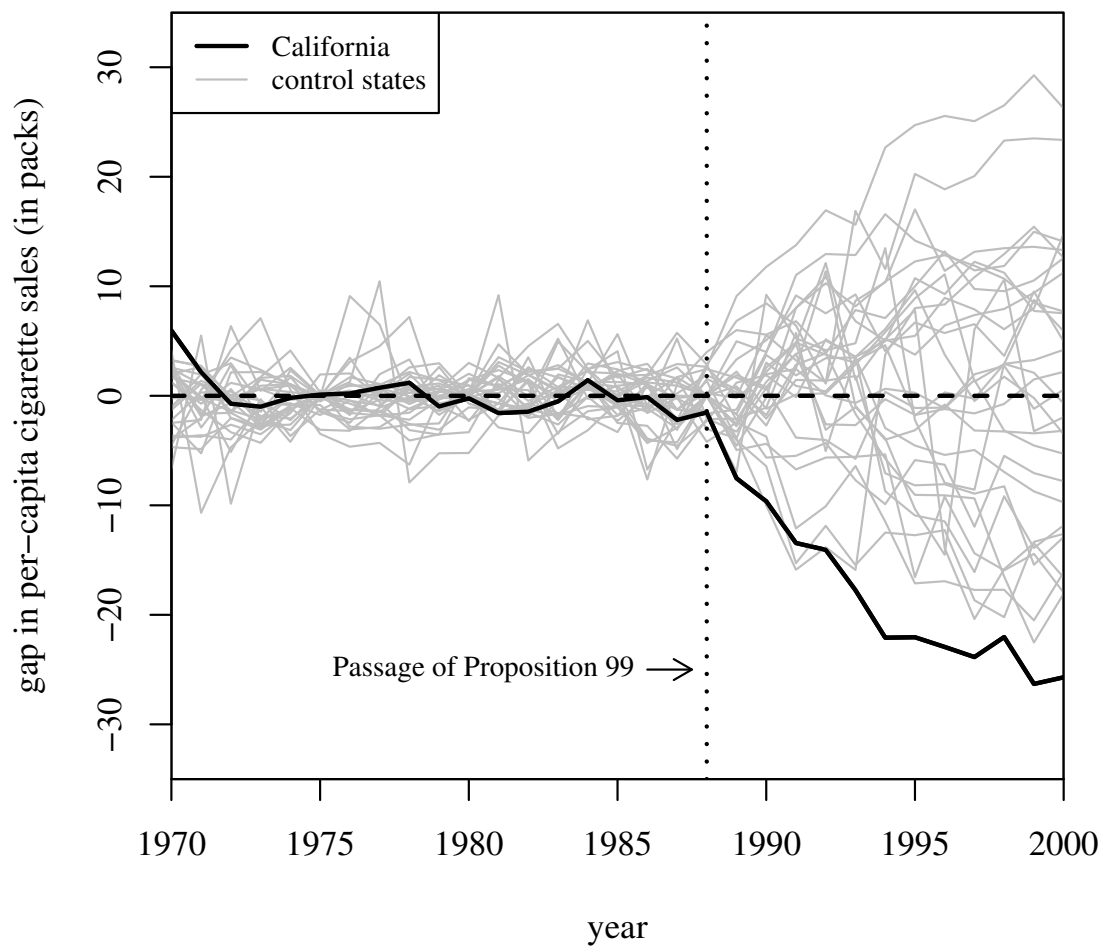


Figure 7: Per-Capita Cigarette Sales Gaps in California and Placebo Gaps in 19 Control States (Discards States with Pre-Proposition 99 MSPE Two Times Higher than California's)

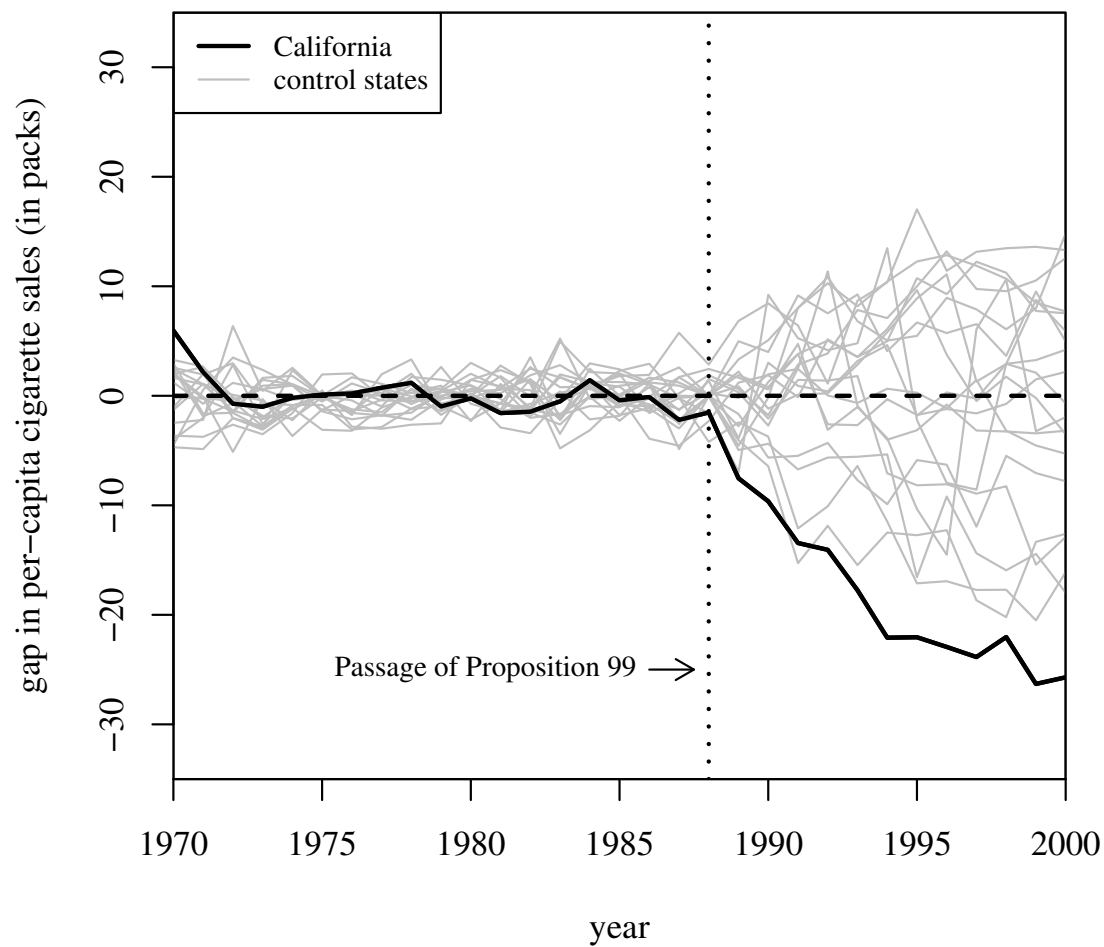


Figure 8: Ratio of Post-Proposition 99 MSPE and Pre-Proposition 99 MSPE: California and 38 Control States

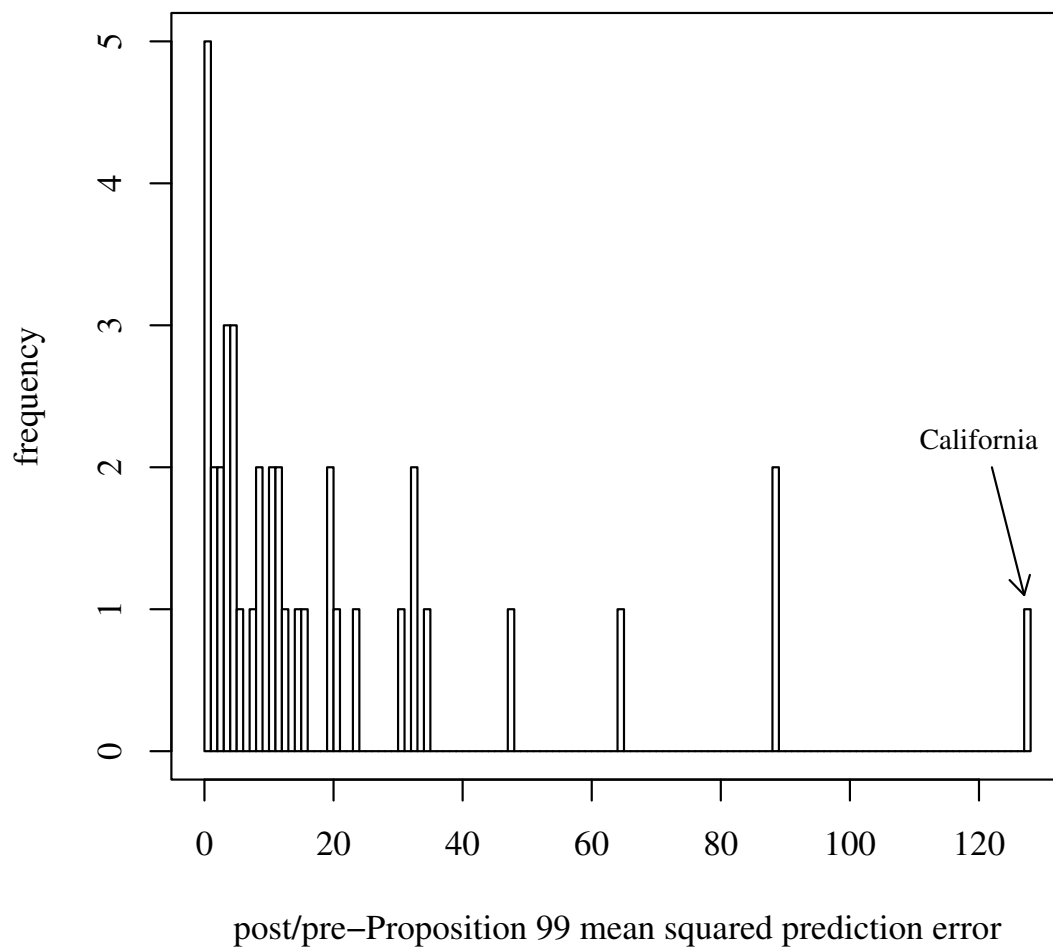


Figure A.1: Trends in Per-Capita GDP: West Germany vs. Synthetic West Germany

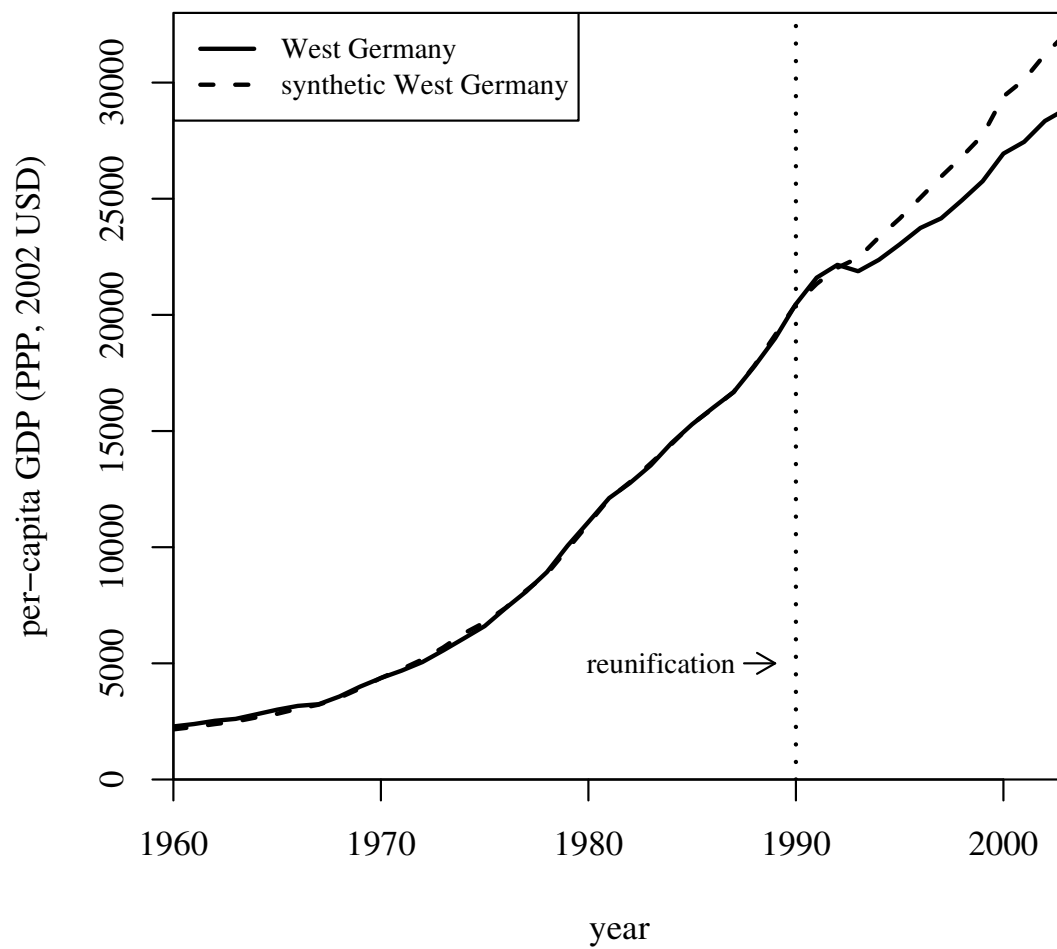


Figure A.2: Placebo in Time. Trends in Per-Capita GDP: West Germany vs. Synthetic West Germany

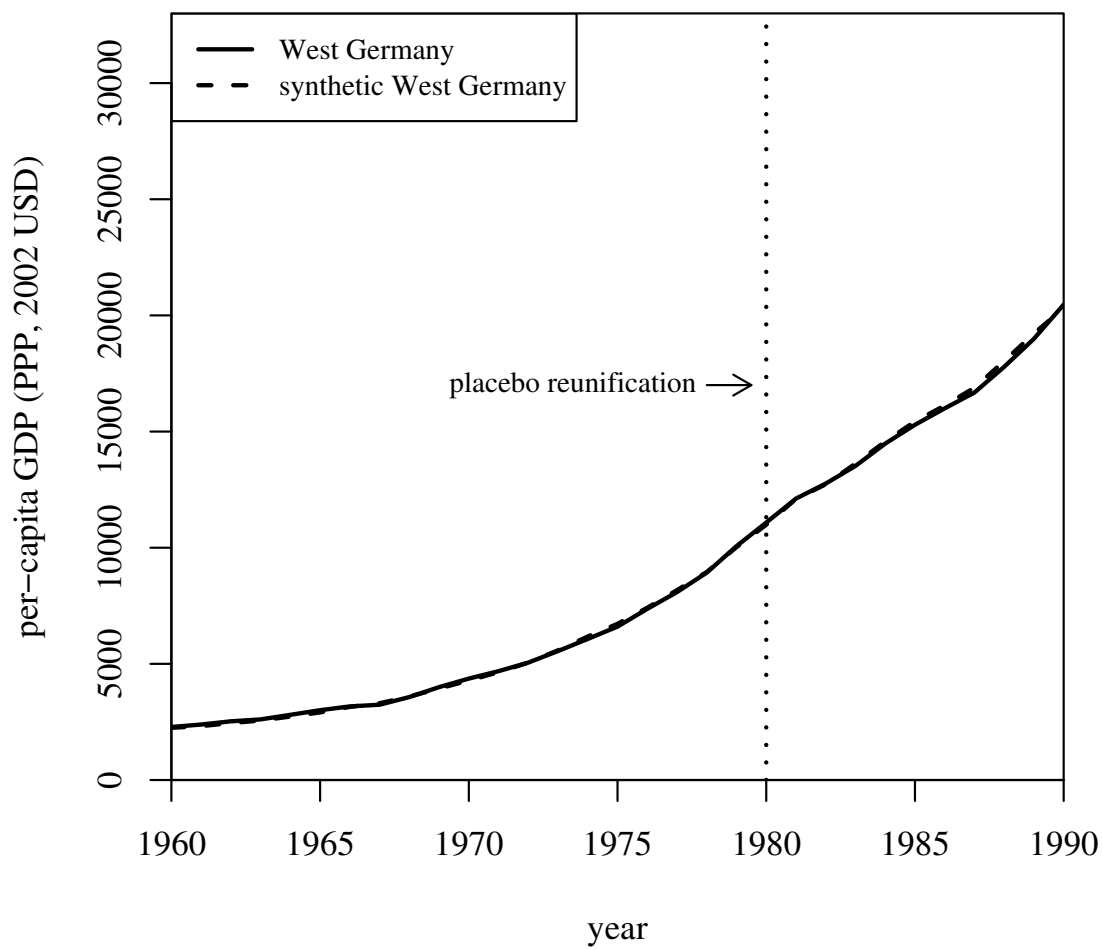


Table 1: Cigarette Sales Predictor Means

Variables	California		Average of
	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).

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

Table A.1: Economic Growth Predictor Means
before the German Reunification

	West Germany	Synthetic West Germany	OECD Sample excl. West Germany
GDP per-capita	10774.85	10808.02	10161.86
Inflation rate	3.80	5.17	9.01
Trade openness	50.59	58.41	33.47
Schooling	29.20	31.21	22.37
Investment rate	27.00	27.00	25.44
Industry share	34.69	34.64	34.28

Note: GDP, inflation rate, and trade openness are averaged for the 1970–1989 period. Industry share is averaged for the 1980–1989 period. Investment rate is averaged for the 1980–1985 period. Schooling is from year 1985.

Table A.2: Country Weights in the Synthetic West Germany

Country	Weight	Country	Weight
Australia	0	Japan	0.127
Austria	0.421	Netherlands	0.137
Belgium	0	New Zealand	0
Canada	0	Norway	0
Denmark	0	Portugal	0
Finland	0	Spain	0
France	0	Sweden	0
Greece	0	Switzerland	0.153
Ireland	0	United Kingdom	0
Italy	0	United States	0.161