UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Paper Economics

Cambridge Working Paper Economics: 1731

# CO-INTEGRATION AND CONTROL: ASSESSING THE IMPACT OF EVENTS USING TIME SERIES DATA

Andrew Harvey and Stephen Thiele

2 July 2017

Control groups can sometimes provide counterfactual evidence for assessing the impact of an event or policy change on a target variable. Fitting a time series model to target and control series offers potential gains over a direct comparison between the target and a weighted average of the controls. More importantly it highlights the assumptions underlying methods such as difference in differences and synthetic control and in doing so suggests ways in which these assumptions may be tested. Our focus is on time series models that are both simple and transparent. Potential gains from fitting such models are analysed and their relative performance is investigated using examples taken from the literature, including the effect of the California smoking law of 1988 and German re-unification. At the same time, some of the draw-backs to current methodology become apparent. It is argued that a time series strategy for the selection of a valid set of controls is to be preferred to one based on data-driven regression methods.

# Co-integration and control: assessing the impact of events using time series data

Andrew Harvey and Stephen Thiele
Faculty of Economics, Cambridge University and
School of Economics and Finance, Queensland University of Technology.
ACH34@CAM.AC.UK

July 2, 2017

**Abstract**

Control groups can sometimes provide counterfactual evidence for assessing the impact of an event or policy change on a target variable. Fitting a time series model to target and control series offers potential gains over a direct comparison between the target and a weighted average of the controls. More importantly it highlights the assumptions underlying methods such as difference in differences and synthetic control and in doing so suggests ways in which these assumptions may be tested. Our focus is on time series models that are both simple and transparent. Potential gains from fitting such models are analysed and their relative performance is investigated using examples taken from the literature, including the effect of the California smoking law of 1988 and German re-unification. At the same time, some of the drawbacks to current methodology become apparent. It is argued that a time series strategy for the selection of a valid set of controls is to be preferred to one based on data-driven regression methods.

KEYWORDS: common trends; difference in differences; intervention analysis; stationarity tests; synthetic control; unobserved components.

# 1 Introduction

Assessing the impact of events or policy changes from time series data is a challenging problem in the social sciences. The simplest way to proceed is by a before and after comparison, but the difficulty is that the behaviour of the target variable might have changed in the absence of any intervention. Hence the need for a time series on a comparative variable, unaffected by the event in question, to act as a control. The counterfactual evidence so obtained provides the basis for assessing the effect of the event.

The most widely used method of analysis is difference-in-differences (DD). The difference in sample means before the event is compared with the difference after. Such studies have increasingly had a time series dimension, with a moderate number of observations both before and after the event; see the summary table in Bertrand et al (2004, Table 1). The key assumptions in these studies are, firstly, that there is a once and for all shift in the target series and secondly that, once allowance has been made for the effect of the event, the underlying difference between the target series and the control series remains the same throughout the sample. The second assumption can be tested if the sample is not too small. As regards the first, there are many applications where a dynamic response, in which there is a gradual adjustment to the intervention, may be more plausible that a 'hard' break. Unless there is prior knowledge about the form of the response - an unlikely scenario - the best approach would seem to be to use a control variable to track the evolution of the changes.

When there is more than one potential control variable, the question arises as to how they should be employed. Abadie, Diamond and Hainmueller – hereafter ADH – have addressed this issue in a series of influential papers. In particular, ADH (2010) assess the effect of the California smoking law of 1989, while ADH (2015) seeks to determine whether per capita income in West Germany fell as a result of German re-unification in 1990. The methodology they propose is, as expressed in ADH (2010, p 493), '... the use of data-driven procedures to construct suitable comparison groups.' A weighted average of the selected series is called a synthetic control (SC). The weighting is determined in the pre-event period by not only choosing the synthetic variable to closely match the movements in the characteristic of interest in the target series, but by also matching other measurable characteristics in the target and control groups.

In this article we approach the issues surrounding the assessment of the effect of an intervention from a time series perspective. Time series issues are rarely addressed in the literature, perhaps because most of the early work on DD focused on the cross-sectional aspects of a two-period study. Bertrand et al (2004) caution against the dangers of ignoring serial correlation in conducting inference in DD studies. The point is well made, but the consequences of failing to appreciate the time series properties of the data could be far more serious than computing a misleading standard error. More often than not, economic and social time series are non-stationary and this presents new problems - and opportunities - for analysis. Of particular importance is the question of whether the series in a group of controls are co-integrated with the target. An assessment of the statistical properties of DD and SC methods can be made by formulating a multivariate time series model in which one of the variables is the target and the others are potential controls. The model provides a yardstick against which the validity and efficiency of other methods can be judged. In contrast to ADH, no additional variables are used. This is also the case in the panel data approach of Hsiao et al (2012). However, Hsiao et al (2012) are concerned primarily with panels of stationary series. They also assume that the dynamics of the intervention follow a stationary process which in our framework is neither necessary nor desirable. Bai et al (2014) extend some of the results in Hsiao et al (2012) to nonstationary time series, but in many other respects their approach is very different from ours. Both papers contain some interesting asymptotic results but while these are reassuring we feel that asymptotics should not be the prime methodology driver in small samples.

The modest length of the available time series means that models must be relatively simple if they are to be useful. They also need to be transparent and to yield results with a clear interpretation. Structural time series models formulated in terms of stochastic trends fulfill this requirement. Such models have been used successfully for many years; see, for example Harvey (1989) and, more recently, Brodersen et al (2015) and Varian (2014). Their potential value in connection with control groups was sketched out in Harvey (1996), following on from the work done by Harvey and Durbin (1986) to assess the impact of the 1983 seat belt law in Great Britain. An example from economics is the study by Angeriz and Arestis (2008) on the effect of the inflation targeting strategy adopted by Sweden, using the European Monetary Union as a control. Vujic, Commandeur and Koopman (2016) show how the approach can be used to investigate the effects of policy changes on crime rates.

In section 2 we analyse the role of dynamics when the number of potential controls is small. Although the SC term is used by ADH in connection with their methodology to select controls and assign them weights, it is used more generally here to denote any weighted average of selected controls. We set out a multivariate model for target and control series. This enables us to determine the conditions under which a SC is valid and whether it is effective in the way that it uses the available information. The potential gain from fitting the full model is then examined with the analysis being supported and illustrated by simulation evidence in Section 3. The implications for the DD estimator are essentially the same. Section 4 discusses how to select a valid set of controls from a large donor pool using time series methods. Our model selection methodology is then applied to the California smoking law and German reunification examples of ADH. We also comment briefly on the Hong Kong example used by Hsiao et al (2012) and analysed further by Gardeazabal and Vega-Bayo (2016). Section 6 explains how to deal with seasonality within the structural time series framework and Section 7 concludes.

# 2 Synthetic control and time series models

A synthetic control is constructed as a combination of contemporaneous values of $N$ series on the characteristic of interest, that is

$$y_t^c = \sum_{i=1}^{N} w_{ij} y_{it} = \mathbf{w}'\mathbf{y}_t, \qquad t = 1, ...., T.$$

where the weights, $w_i, i = 1, ..., N$, in the $N \times 1$ vector $\mathbf{w}$ are chosen in the pre-intervention period. The effect of the intervention is then tracked by $y_{0t} - y_t^c$, $t = \tau, ..., T$, having made allowance for a difference in level. We will regard an SC as *valid* if $y_{0t} - y_t^c$ is stationary.

Questions[1] arise about whether the weights should sum to one and/or be positive. When the problem is approached by formulating a multivariate time series model, all that matters is the dynamic properties of the target variable and the potential controls. The answers to the questions about weights are then clear.

[1]Further work on the criteria for constructing a synthetic control can be found in a recent paper by Doudchenko and Imbens (2016).

Simply using $y_{0t} - y_t^c$ to track the intervention effect ignores the dynamics of the series and any information on dynamics contained in the pre-intervention values of $y_{0t}$. When a model is formulated, it raises the issue as to what might be lost by using a SC as opposed to working with the full model. Broadly speaking there is a trade off between efficiency and robustness, with small sample issues being particularly important. Note that a control $y_t^c$ which does not have the property that $y_{0t} - y_t^c$ is stationary may still be useful provided an appropriate model is constructed.

## 2.1 Constructing a valid synthetic control

Consider a multivariate model

$$\begin{aligned} y_{0t} &= \mu_t + \mu_0 + \varepsilon_{0t}, \quad t = 1, \ldots, T, \\ \mathbf{y}_t &= \boldsymbol{\theta}\mu_t + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{1}$$

in which $\mu_t$ is a common stochastic trend,

$$\mu_t = \mu_{t-1} + \delta_{t-1} + \eta_t, \quad \delta_t = \delta_{t-1} + \zeta_t, \tag{2}$$

taking the value zero at $t = 0$ and $\delta_t$ is the slope. The vectors in the second equation of (1) are all $N \times 1$. The disturbances $\eta_t, \zeta_t$ and $\varepsilon_{0t}$ are normally distributed and serially independent with zero means and variances $\sigma_\eta^2, \sigma_\zeta^2$ and $\sigma_0^2$, while $\boldsymbol{\varepsilon}_t$ is multivariate normal and serially independent with covariance matrix $\boldsymbol{\Sigma}_\varepsilon$. All the disturbances are mutually independent. Setting $\sigma_\eta^2 = 0$ makes the trend an integrated random walk (IRW), whereas setting $\sigma_\zeta^2$ to zero yields a random walk plus drift. The assumptions could be relaxed, for example $(\varepsilon_t, \boldsymbol{\varepsilon}_t')'$ might be a stationary multivariate ARMA process.

The above model provides a vehicle for analyzing the construction of a SC that can track the potential movements in $y_{0t}$ after an intervention. Subtracting $y_t^c$ from the first equation in (1) gives

$$y_{0t} - y_t^c = \mu + (1 - \mathbf{w}'\boldsymbol{\theta})\mu_t + \varepsilon_{0t} - \mathbf{w}'\boldsymbol{\varepsilon}_t, \qquad t = 1, ..., T,$$

where $\mu = \mu_0 - \mathbf{w}'\boldsymbol{\mu}$, but when $(1 \ \mathbf{w}')'$ is a co-integrating vector[2], $\mathbf{w}'\boldsymbol{\theta} = 1$ and $y_{0t} - y_t^c$ is stationary. When $\boldsymbol{\theta} = \mathbf{i}$, where $\mathbf{i}$ is a vector of ones, there is

[2] At the simplest level, a co-integrating vector is one that yields a stationary time series when applied to series that need to be differenced to make them stationary. The concept is a fundamental one in time series econometrics and is covered in most standard texts.

balanced growth. In this case, $\mathbf{w}'\boldsymbol{\theta} = \mathbf{w}'\mathbf{i} = 1$ and so the weights sum to one. On the other hand, if the weights are constrained to sum to one, but $y_{0t} - y_t^c$ contains a stochastic trend component, the SC will not generally be valid, because $y_{0t} - y_t^c$ will not be stationary.

In the panel control approach proposed by Hsiao et al (2012) the weights are given by OLS. Such weights will not, in general, sum to one, but when $\mu_t$ in (1) is replaced by a stationary common factor, there is no longer a clear case for $\boldsymbol{\theta} = \mathbf{i}$. For nonstationary series the weights must sum to one under balanced growth if $y_{0t} - y_t^c$ is to be stationary. The restricted least squares (RLS) solution, which chooses $\mathbf{w}$ to minimize $\sum_{t=1}^{\tau-1}(y_{0t} - \mu - \mathbf{w}'\mathbf{y}_t)^2$ subject to $\mathbf{w}'\mathbf{i} = 1$, is

$$\widehat{\mathbf{w}} = \mathbf{S}_{\mathbf{yy}}^{-1}\mathbf{s}_y + s\mathbf{S}_{\mathbf{yy}}^{-1}\mathbf{i} = \mathbf{w}_{OLS} + s\mathbf{S}_{\mathbf{yy}}^{-1}\mathbf{i}, \tag{3}$$

where $\mathbf{S}_{\mathbf{yy}} = \sum_{t=1}^{\tau-1}\mathbf{y}_t\mathbf{y}_t'$ and $\mathbf{s}_y = \sum_{t=1}^{\tau-1}\mathbf{y}_t y_{0t}$ and $s = (1 - \mathbf{i}'\mathbf{S}_{\mathbf{yy}}^{-1}\mathbf{s}_y)/\mathbf{i}'\mathbf{S}_{\mathbf{yy}}^{-1}\mathbf{i}$. Controls with a large variance tend to be downweighted by the second term. Note that there is nothing to prevent some of the weights being negative. An SC constructed as $\widehat{\mathbf{w}}'\mathbf{y}_t$ will be denoted $\widehat{y}_t^c$.

The easiest way to compute $\widehat{\mathbf{w}}$ is to subtract one of the controls from all the other controls and the target and then do an OLS regression. The coefficient of the subtracted control is equal to unity minus the sum of the coefficients on the other controls. In other words $y_{0t} - y_{it}$ is regressed on a constant and the $N-1$ contrasts, $y_{jt} - y_{it}$, $j \neq i$ to give $\widehat{w}_j$ and $\widehat{w}_i = 1 - \Sigma_{j\neq i}\widehat{w}_j$.

## 2.2 Modeling the intervention

Suppose there is an intervention at time $t = \tau$, where $\tau$ is large enough to enable simple time series models to be constructed. The pattern of the response is rarely known, but if it is assumed that the full effect works its way through after $m \geq 1$ time periods, a set of pulse dummies may be added to complement the step dummy.

Assuming balanced growth, that is $\boldsymbol{\theta} = \mathbf{i}$ in (1),

$$\begin{aligned} y_{0t} &= \mu_t + \mu_0 + \lambda d_t + \textstyle\sum_{j=1}^{m}\lambda_j d_t^* + \varepsilon_{0t}, \quad t = 1, \ldots, T, \\ \mathbf{y}_t &= \mathbf{i}\mu_t + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \end{aligned} \tag{4}$$

where the potential permanent shift in the level of $y_{0t}$ is captured by a step dummy

$$d_t = \begin{cases} 0 & \text{for} \quad t < \tau + m, \\ 1 & \text{for} \quad t \geq \tau + m \end{cases}, \qquad 1 < \tau + m \leq T,$$

whereas the intermediate effects are modeled by $m$ pulse dummies[3]

$$d_t^* = \begin{cases} 0 & \text{for} \quad t \neq \tau + j - 1, \\ 1 & \text{for} \quad t = \tau + j - 1 \end{cases}, \qquad j = 1, ..., m.$$

The effect of the intervention is transitory if $m > 0$ and $\lambda = 0$. If no assumption is made about reaching a new level, then $m = T - \tau + 1$ and only pulse dummies are present.

**Remark 1** *Simple restrictions may sometimes be put on the pattern of pulse dummy coefficients. For example if $\lambda_j = (j/m)\lambda, \;\; j = 1, ..., m,$ the $d_t^{*\prime}s$ are replaced by $d_t \times (j/m)$ from $t = \tau$ to $\tau + m - 1$.*

Subtracting the SC from $y_{0t}$ as given by the first equation in (4) gives the contrast

$$y_{0t} - y_t^c = \mu + \lambda d_t + \sum_{j=1}^{m} \lambda_j d_t^* + \varepsilon_t, \qquad t = 1, ..., T, \tag{5}$$

where $\varepsilon_t = \varepsilon_{0t} - \mathbf{w}'\boldsymbol{\varepsilon}_t$. The dummy variables can then be estimated by regression. When there are only pulse dummies, they are estimated as $\widehat{\lambda}_j = y_{0,\tau+j-1} - y_{\tau+j-1}^c - \widehat{\mu}$, $j = 1, .., T - \tau + 1$, where $\widehat{\mu}$ is the mean of the pre-intervention contrast. If $\mathbf{w}$ is treated as fixed and $\varepsilon_{0t}$ and $\boldsymbol{\varepsilon}_t$ are white noise,

$$Var(\widehat{\lambda}_j) = Var(\widehat{\mu}) + \sigma_0^2 - 2\mathbf{w}'\boldsymbol{\sigma}_\varepsilon + \mathbf{w}'\boldsymbol{\Sigma}_\varepsilon \mathbf{w}, \qquad j = 1, ..., T - \tau + 1, \tag{6}$$

with[4]

$$Var(\widehat{\mu}) = \frac{\sigma_0^2 - 2\mathbf{w}'\boldsymbol{\sigma}_\varepsilon + \mathbf{w}'\boldsymbol{\Sigma}_\varepsilon \mathbf{w}}{\tau - 1}.$$

If $\varepsilon_t$ is serially correlated, it may be modeled by a stationary ARMA process or $Var(\widehat{\lambda}_j)$ replaced by a nonparametric estimator based on the residual correlogram.

**Remark 2** *When there is a hard break, that is no dynamic adjustment, the estimator of $\lambda$ obtained from (5) is just a generalized DD based on $y_t^c$. With a valid control, $\lambda$ can be estimated consistently. Setting the weights to be equal, that is $w_i = 1/N$ results in a simple sum of the means of the $N$ control groups.*

[3] Note that $d_t^* = d_{t+j} - d_{t+j-1}$. Other parameterizations are possible, for example $\sum_{j=0}^{m} \lambda_j d_{t-j}$ can replace $\lambda d_t + \sum_{j=1}^{m} \lambda_j d_t^*$.

[4] In some circumstances it may be possible to set $\mu = 0$, as in ADH(2010), but this is the exception rather than the rule.

The structure of (4) provides insight into the weighting schemes used for the construction of a SC. Suppose the parameters, $\sigma^2_\eta$, $\boldsymbol{\sigma}_\varepsilon$=$Cov(\varepsilon_t, \boldsymbol{\varepsilon}_t)$ and $\boldsymbol{\Sigma}_\varepsilon$=$Var(\boldsymbol{\varepsilon}_t)$, together with the initial conditions, $\mu_0$ and $\boldsymbol{\mu}$, are known and that all the post-intervention dummies are pulses. Smoothed estimates of the common trend, $\mu_{t|T}$, $t = 1, .., T$, can be computed by the Kalman filter and smoother based on all $T$ observations from the control groups and the first $T_0$ from the target; see Durbin and Koopman (2012). The optimal estimator of the target in the post-intervention period is then

$$y_{0t|T} = \mu_0 + \mu_{t|T} + \varepsilon_{0t|T}, \qquad t = \tau, ..., T, \tag{7}$$

where $\varepsilon_{0t|T} = \boldsymbol{\beta}'\boldsymbol{\varepsilon}_{t|T}$, with $\boldsymbol{\beta} = \boldsymbol{\Sigma}_\varepsilon^{-1}\boldsymbol{\sigma}_\varepsilon$ and $\boldsymbol{\varepsilon}_{t|T} = \mathbf{y}_t - \boldsymbol{\mu} - \mathbf{i}\mu_{t|T}$. Thus

$$y_{0t|T} = \mu_0 + \mu_{t|T} + \boldsymbol{\beta}'(\mathbf{y}_t - \boldsymbol{\mu} - \mathbf{i}\mu_{t|T}), \qquad t = \tau, ..., T. \tag{8}$$

The corresponding estimators of the pulses are $\widetilde{\lambda}_j = y_{0t} - y_{0t|T}$, $j = 1, ..., T - \tau + 1$. Since $y_{0t} - y_{0t|T} = (1 - \boldsymbol{\beta}'\mathbf{i})(\mu_t - \mu_{t|T}) + \varepsilon_{0t} - \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t$,

$$Var(\widetilde{\lambda}_j) \simeq (1 - \boldsymbol{\beta}'\mathbf{i})^2 Var(\mu_t - \mu_{t|T}) + \sigma_0^2 - \boldsymbol{\beta}'\boldsymbol{\sigma}_\varepsilon, \qquad j = 1, ..., T - \tau + 1. \tag{9}$$

The approximation comes from the (small) cross-product involving $\mu_t - \mu_{t|T}$ and $\varepsilon_{0t} + \boldsymbol{\beta}'\boldsymbol{\varepsilon}_t$. When $\mu_0$ and $\boldsymbol{\mu}$ are estimated there are additional terms, but these are dominated by the terms already present.

**Remark 3** *When there is no stochastic trend, so* $\mu_t = \mu_{t|T} = 0$, *and* $\mu_0$ *and* $\boldsymbol{\mu}$ *are estimated, the first term in (9) is replaced by* $(\sigma_0^2 - \boldsymbol{\beta}'\boldsymbol{\sigma}_\varepsilon)/(\tau - 1)$. *This formula is exact and may be compared directly with (6).*

**Remark 4** *If there are no controls, the trend can be estimated from the observations before the intervention. Then* $Var(\widetilde{\lambda}_j) \simeq Var(\mu_t - \mu_{t|\tau-1}) + \sigma_0^2$, $j = 1, ..., T - \tau + 1$. *However,* $Var(\mu_t - \mu_{t|\tau-1})$ *is unbounded as* $t \to \infty$. *When there is an eventual permanent shift, at* $\tau + m$, *the observations after stabilization can be used to give a bounded estimate of the trend in the intermediate period. However, the permanent change cannot be estimated consistently (as* $T \to \infty$*).*

Now consider the SC based on the RLS weights, $\widehat{\mathbf{w}}$. In the population these weights are given by

$$\mathbf{w}_\varepsilon' = \boldsymbol{\sigma}_\varepsilon'\boldsymbol{\Sigma}_\varepsilon^{-1} + s_\varepsilon\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1} = \boldsymbol{\beta}' + s_\varepsilon\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1}, \qquad \text{where } s_\varepsilon = (1 - \boldsymbol{\beta}'\mathbf{i})/\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{i}. \tag{10}$$

This result is obtained by noting that when $\mathbf{w}'\mathbf{i} = 1$, $y_{0t} - \mathbf{w}'\mathbf{y}_t = \mu + \varepsilon_t$, where $\varepsilon_t = \varepsilon_{0t} - \mathbf{w}'\boldsymbol{\varepsilon}_t$, and so minimizing $\sum_{t=1}^{\tau-1}(y_{0t} - \mu - \mathbf{w}'\mathbf{y}_t)^2$ subject to $\mathbf{w}'\mathbf{i} = 1$ is the same as minimizing $\sum_{t=1}^{\tau-1} \varepsilon_t{}^2$ subject to $\mathbf{w}'\mathbf{i} = 1$. In the population, this is equivalent[5] to minimizing $Var(\varepsilon_t) = \sigma_0^2 - 2\mathbf{w}'\boldsymbol{\sigma}_\varepsilon + \mathbf{w}'\boldsymbol{\Sigma}_\varepsilon\mathbf{w}$, subject to $\mathbf{w}'\mathbf{i} = 1$.

The weights in (10) may also be written as

$$\mathbf{w}'_\varepsilon = \frac{1}{\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{i}}\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1} + \boldsymbol{\beta}'\left[\mathbf{I} - \frac{1}{\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{i}}\mathbf{i}\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1}\right]. \tag{11}$$

Here the elements in the first term on the right hand side sum to one, whereas those in the second term sum to zero. The first term estimates the common trend, so a comparison with (8) can be made by writting

$$y_{0t}^c = \mu_0 + \widehat{\mu}_t + \boldsymbol{\beta}'(\mathbf{y}_t - \mathbf{i}\widehat{\mu}_t - \boldsymbol{\mu}), \qquad t = \tau, ..., T, \tag{12}$$

where $\widehat{\mu}_t = (\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{i})^{-1}\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1}(\mathbf{y}_t - \boldsymbol{\mu})$. The pulses are estimated as $\lambda_j^c = y_{0,\tau+j-1} - y_{0,\tau+j-1}^c$, $j = 1, .., T - \tau + 1$, and a little manipulation gives

$$Var\left(\lambda_j^c\right) = \frac{(1 - \boldsymbol{\beta}'\mathbf{i})^2}{\mathbf{i}'\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{i}} + \sigma_0^2 - \boldsymbol{\beta}'\boldsymbol{\sigma}_\varepsilon, \qquad j = 1, ..., T - \tau + 1. \tag{13}$$

The first term in (13) comes from estimating the level as $\widehat{\mu}_t$ as opposed to $\mu_{t|T}$ in (9). When $\boldsymbol{\beta}'\mathbf{i} = 1$, $y_{0t}^c$ and $y_{0t|T}$ are the same. On the other hand, when $\boldsymbol{\beta} = \mathbf{0}$, only $\mu_t$ is estimated.

**Remark 5** *When there is only one control,* $Var\left(\lambda_j^c\right) = \sigma_0^2 + \sigma_1^2(1 - 2\beta)$. *With* $\sigma_0 = \sigma_1$ *and a correlation of* $\rho$, $Var(\widetilde{\lambda}_j) \simeq (1-\rho)^2 Var(\mu_t - \widetilde{\mu}_{t|T}) + \sigma_0^2(1-\rho^2)$ *whereas* $Var(\lambda_j^c) = 2\sigma_0^2(1-\rho)$. *The ratio of the last two terms is* $(1+\rho)/2$ *and this is the efficiency of* $\lambda_j^c$, *relative to* $\widetilde{\lambda}_j$, *when there is no stochastic trend (and intercepts are estimated.)*

The fact that $\widehat{\mu}_t$, unlike $\mu_{t|T}$, uses only contemporaneous observations on the controls suggests that the trend component in a SC might benefit from some smoothing, that is $\widehat{\mu}_t$ in (12) is replaced by

$$\widehat{\mu}_{t|T} = \sum_j w_j^* \widehat{\mu}_{t+j}, \tag{14}$$

[5] Bai et al (2014) state that $\mathbf{w}_\varepsilon$ is unique and consistently estimated by OLS, $\mathbf{w}_{OLS}$. This result might be taken to suggest that there is no need to use RLS. However, such a strategy may not be a good one in small samples; see the illustration in the application to German re-unification.

where the weights $w_j^*$ sum to one and the summation is over observations close to $t$. For example, $w_1^* = w_{-1}^* = 0.25$ and $w_0^* = 0.5$. The advantage of fitting a model is that the weights adapt to the propertes of the data rather than being arbitrary.

## 2.3 Estimation by maximum likelihood and regression

Full ML estimation can be carried out on (4). Alternatively a valid SC, $y_t^c$, can be constructed and the system transformed to

$$\begin{aligned} & y_{0t} - y_t^c = \mu + \lambda d_t + \textstyle\sum_{j=1}^m \lambda_j d_t^* + \varepsilon_t, \quad t = 1, \ldots, T, \\ & y_t^c = \mu_t + \mathbf{w}'\boldsymbol{\mu} + \mathbf{w}'\boldsymbol{\varepsilon}_t \\ & \mathbf{y}_t^* - \mathbf{i}y_t^c = \boldsymbol{\mu}^* + \boldsymbol{\varepsilon}_t^* - \mathbf{i}\mathbf{w}'\boldsymbol{\varepsilon}_t, \end{aligned} \tag{15}$$

where the first equation is as in (5), $\mathbf{y}_t^*$ contains $N-1$ of the controls - it doesn't matter which one is dropped - and $\boldsymbol{\mu}^*$ and $\boldsymbol{\varepsilon}_t^*$ are the corresponding elements of $\boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}_t$. Full ML on this system is the same as on (4).

As noted earlier, the SC estimators of the intervention effects are obtained by carrying out a regression on the first equation in (15). An intermediate solution between estimating the joint model for all $N+1$ variables and the single equation for $y_{0t} - y_t^c$ is to estimate the $y_{0t} - y_t^c$ equation jointly with the one for $y_t^c$. In other words, estimate a bivariate model made up of the first two equations in (15).

A system equivalent to (15) is obtained by replacing $y_t^c$ by one of the control variables, $y_{it}$, $i = 1, ..., N$. The full ML estimators of the intervention coefficients are then obtained without constructing $y_t^c$. Furthermore, as suggested earlier by the discussion of RLS, a single equation estimator of the intervention coefficients based on RLS SC can be obtained simply by regressing $y_{0t} - y_{it}$ on the intervention variables, a constant and the $N-1$ differences, $y_{jt} - y_{it}$, $j \neq i$. The estimates are the same as those obtained by full ML estimation of the multivariate model consisting of the equation for $y_{0t} - y_{it}$ together with the $N-1$ for $y_{jt} - y_{it}$, $j \neq i$. However, the computed SEs are different. The single equation constructs the RLS SC indirectly and in doing so allows for the estimation error in the weights.

# 3 Potential gains from model-based estimation

Estimating the effect of an intervention by subtracting the RLS SC from the target series is straightforward and yields valid inferences provided the underlying assumption of balanced growth holds. A correctly specified model can yield more efficient estimators and the purpose of this section is to explore the magnitude of potential gains. We note in passing that although these gains will not be realised by a misspecified model, such a model will nevertheless offer some gains to the extent that it successfully estimates a trend and in doing so provides a more advantageous decomposition than the one offered[6] by $y_{0t}^c$ in (12); see Harvey and Delle Monache (2012, pp 88-99).

Consider a simple model with two controls, that is $N = 2$, and $\mathbf{\Sigma}_\varepsilon = \mathbf{I}_2$. Let $\boldsymbol{\sigma}'_\varepsilon = \left( \begin{array}{cc} \rho & 0 \end{array} \right)$. Then $E(\varepsilon_{0t} \mid \boldsymbol{\varepsilon}_t) = \boldsymbol{\sigma}'_\varepsilon \mathbf{\Sigma}_\varepsilon^{-1} \boldsymbol{\varepsilon}_t = \rho\varepsilon_{1t}$ and

$$\mathbf{w}_\varepsilon = \left[ \begin{array}{c} 1/2 + \rho/2 \\ 1/2 - \rho/2 \end{array} \right] = \left[ \begin{array}{c} (1+\rho)/2 \\ (1-\rho)/2 \end{array} \right].$$

Now consider the following special cases when the trend and constants are known. (i) With $\rho = 1$, $y_t^c = \mathbf{w}'_\varepsilon \mathbf{y}_t = y_{1t}$. The intervention effect is obtained exactly obtained from $y_{0t} - y_{1t}$. Thus $y_{1t}$ is a perfect control and there is nothing to be gained from trying to estimate a multivariate model. (ii) When $\rho = 0$, $y_t^c = \mathbf{w}'_\varepsilon \mathbf{y}_t = (y_{1t} + y_{2t})/2$. In this case $Var(y_{0t} - y_t^c) = 1.5$, but $Var(y_{0t} - \mu_t) = 1$, so if it were possible to estimate the trend exactly, there would be a gain in efficiency. This gain can be realized with a bivariate model for $y_{0t}$ and $y_t^{SC}$. (iii) With $\rho = -1$, $y_t^c = \mathbf{w}'_\varepsilon \mathbf{y}_t = y_{2t}$ and so the first series is discarded. Thus $Var(y_{0t} - y_t^c) = Var(y_{0t} - y_{2t}) = 2$. There is no gain from the bivariate model for $y_{0t}$ and $y_t^c$, but in the trivariate model $y_{1t}$ enables the target to be estimated exactly because it is perfectly (negatively) correlated with the target irregular.

The analysis indicates potential gains for models when the trend is known. We now report a Monte Carlo experiment to determine how much is lost by having to estimate the trend. The model consists of a (common) random walk plus drift with the irregular specified as above. The signal-noise ratio is $q$. Constants and a slope are estimated, but the actual values do not

[6] As a simple illustration, suppose there is only one control. Let $\mu_{t|T} = 0.25y_{t-1} + 0.5y_t + 0.25y_{t+1}$. Let $\sigma_0^2 = \sigma_1^2$ and let $q$ be the signal-noise ratio in a RW. Then this weighted estimator has a smaller variance than the SC if $q < 5 - 8\rho$.

affect the results of the simulations and the same is true of the intervention parameter. Table 1 is for $T_0 = 20$ and $T_1 = 10$, with the MSEs denoting the average over all post-intervention periods, that is $(T-\tau)^{-1}\sum_{j=\tau+1}^{T}[(\lambda_j - \widehat{\lambda}_{0j})^2]$. The estimates for SC and the model (TSM) were obtained using 20,000 replications. The SC model is estimated on the pretreatment sample alone, whereas the TSM is estimated over the full sample, omitting only the post treatment sample for the treated unit. With known parameters, there were 100,000 replications for TSM. In the case of known SC, the MSEs can be computed exactly: it is not difficult to show that the variance is $(3 - 2\rho - \rho^2)/2$ and so for $\rho = -0.5$, 0 and 0.5 the variances are 1.88, 1.50 and 0.88 respectively. The pattern of MSEs for different values of $\rho$ is as suggested by the analysis with the MSEs increasing as $\rho$ moves from 1 to $-1$ but the gain from estimating the full model becoming increasingly apparent, even when the parameters are estimated. When the parameters in the model are given, the SC is never superior, but when parameters are estimated this need not be the case. However, only for $\rho = 0.5$ and $q = 1.5$ is SC better and then only marginally. Finally it can be seen that the model performs relatively better, the smaller is $q$. Indeed with a very large value of $q$, a single control would effectively measure the intervention exactly.

| **Model** | $\rho$ | $q = 0.1$ | $q = 0.25$ | $q = 0.5$ | $q = 1$ | $q = 1.5$ |
|---|---|---|---|---|---|---|
| SC | -0.5 | 1.96 | 1.97 | 1.95 | 1.95 | 1.95 |
| TSM | | 1.15 | 1.29 | 1.41 | 1.57 | 1.66 |
| Known SC | | 1.87 | 1.88 | 1.88 | 1.87 | 1.88 |
| Known TSM | | 1.02 | 1.14 | 1.27 | 1.41 | 1.50 |
| SC | 0 | 1.59 | 1.58 | 1.58 | 1.58 | 1.59 |
| TSM | | 1.25 | 1.31 | 1.37 | 1.44 | 1.49 |
| Known SC | | 1.50 | 1.50 | 1.50 | 1.50 | 1.50 |
| Known TSM | | 1.12 | 1.18 | 1.23 | 1.30 | 1.34 |
| SC | 0.5 | 0.92 | 0.92 | 0.92 | 0.92 | 0.92 |
| TSM | | 0.86 | 0.89 | 0.90 | 0.92 | 0.93 |
| Known SC | | 0.87 | 0.88 | 0.88 | 0.87 | 0.88 |
| Known TSM | | 0.78 | 0.79 | 0.81 | 0.82 | 0.83 |

Table 1. Forecast MSE for treatment - $N = 2$, $T - \tau = 10$, $\gamma = 0$

The model is now extended by letting the correlation between the two controls be $\gamma$ rather than zero. With fixed parameters known, the SC weights

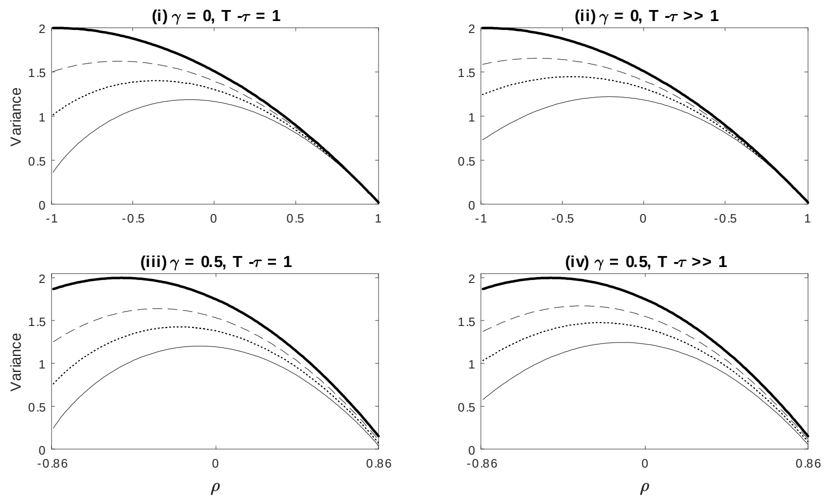



Figure 1: Average MSEs of estimates of interventions

are given by

$$\mathbf{w}_\varepsilon = \begin{bmatrix} \frac{1-\gamma+\rho}{2(1-\gamma)} & \frac{1-\gamma-\rho}{2(1-\gamma)} \end{bmatrix}'$$

Now only the first series is used in the SC when $\gamma + \rho = 1$, whereas with $\gamma - \rho = 1$, only the second is used. Figure 1 shows the results for $\gamma = 0.5$ together[7] with those for $\gamma = 0$. The values of $q$ are as in Table 1 with the lowest curve corresponding to $q = 0.1$. As can be seen, estimation from the full model provides even greater efficiency gains over the full support of $\rho$ (noting that $|\Sigma_\varepsilon| > 0$ only for $|\rho| < \sqrt{3}/2$).

# 4 Selecting the controls

The strategy adopted by ADH and HCW for selecting a set of controls from a large donor pool is to use a data-driven approach in which the pre-intervention observations in the target variable are matched as closely as possible by a small subset of potential donors. The main difference is that whereas ADH use a number of related variables both to select the best controls and to weight them, HCW use only the variables themselves. In addition

[7]Note that when $\gamma + \rho = 1$ only the first series is used in the SC.

ADH constrain the weights of the selected controls to be positive and to sum to one

In selecting a set of controls for nonstationary time series our concern is to ensure that they are on the same growth path as the target. The validity of a potential control may therefore be assessed by a stationarity test on the difference or *contrast* between it and the target. As will be seen from the applications, this strategy can lead to very different results from those obtained in the earlier studies.

### 4.1 Stationarity tests in small samples

Because the number of observations in the pre-intervention period is typically small, we conducted a small Monte Carlo experiment to assess size and power of the tests due to Nyblom and Mäkeläinen (1983) and Kwiatkowski *et al* (1992) - hereafter NM and KPSS - when $T = 20$. The KPSS test depends on a lag, $\ell$, with $\ell = 0$ giving the NM statistic. The model is as in Kwiatkowski *et al.* (1992), namely

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \eta_t \qquad \text{and} \qquad \varepsilon_t = \phi\varepsilon_{t-1} + \xi_t, \end{aligned}$$

where the disturbances $\eta_t$ and $\xi_t$ are mutually independent and generated as $\eta_t \sim N(0, q)$ and $\xi_t \sim N(0, 1 - \phi^2)$. The results in Table 2 are for $\phi$ set to zero or 0.5 and $q = \{0, 0.1, 0.5\}$. A 10% level of significance, for which the critical value is 0.347, is adopted. For each set of parameters, 10,000 simulations were used to calculate the rejection probabilities.

Table 2. KPSS($l$) Rejection Rates for Tests of $I(0)$ versus $I(1)$ with $T = 20$.

| $\phi$ | $q$ | $l = 0$ (NM) | $l = 1$ | $l = 2$ | $l = 3$ | $l = 5$ | $l = 7$ |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.101 | 0.099 | 0.103 | 0.101 | 0.102 | 0.125 |
| | 0.1 | 0.445 | 0.416 | 0.383 | 0.343 | 0.306 | 0.274 |
| | 0.5 | 0.695 | 0.621 | 0.562 | 0.512 | 0.439 | 0.367 |
| 0.5 | 0 | 0.422 | 0.277 | 0.227 | 0.183 | 0.154 | 0.155 |
| | 0.1 | 0.606 | 0.457 | 0.388 | 0.338 | 0.286 | 0.251 |
| | 0.5 | 0.761 | 0.629 | 0.554 | 0.498 | 0.403 | 0.340 |

When there is no autocorrelation in the transitory component, $\varepsilon_t$, the tests have excellent size. As dependence is introduced, the probability of rejection starts to exceed the nominal size of 0.10. Increasing the number of lags is beneficial up to $l = 6$, but beyond that size is adversely affected. For lower lags there is a trade-off between size and power. Kwiatkowski et al (1992) suggest a rule of $l = \lfloor 4(T/100)^{1/4} \rfloor$, which yields $l = 2$ for $T = 20$. Table 2 confirms that $l = 2$ strikes a good balance.

The suggested control selection strategy is one in which KPSS tests are carried out on the contrasts of all units in the donor pool. The units are then ordered according to the magnitude of the KPSS statistic and only selected when the null hypothesis of stationarity is not rejected at a particular level of significance[8]. The weights for a SC are then obtained by RLS. As is apparent from (3), this may result in small weights being assigned to variables where the contrast has a large variance. Such variables may then be dropped from the SC group, though if their contrasts have small KPSS statistics they may still be of value in a multivariate model.

In order to avoid the data-mining pitfalls associated with machine learning, ADH(2014) have a training sample followed by an estimation sample. However, this approach is neither necessary nor desirable for short nonstationary time series where the aim is to determine whether the target and potential controls are on the same path for the whole pre-intervention period.

## 4.2 Statistical significance, diagnostic checking and robustness

For a given group of controls, tests can be carried out on the significance of the intervention variables and the question as to whether the changes have stabilized after a period of time can be addressed[9]. A good starting point is the estimates obtained by subtracting the SC from the target. For nonstationary data, it is clear from (5) that tests based on an assumption of stationarity are valid provided the controls are co-integrated with the target. It is worth making this point because according to Gardeazabal and Vega-Bayo (2016, Section 3) - '...the panel data approach allows for traditional

[8] Note that we have shown that the tests tend to be conservative.

[9] A test for the significance of an intervention could be carried out directly from a contrast using the cointegration breakdown test of Andrews and Kim (2006).

If it is assumed a new level has been reached, the target series residuals can be tested for stationarity. When a hard break is estimated, the KPSS test can include the post intervention observations by adopting the modification in Busetti and Harvey (2001).

inference as long as the outcome variable is stationary, whereas the synthetic control[10] relies on randomization for inference.'

When time series models are fitted, diagnostics for residual serial correlation and other indications of misspecification can be carried out in the pre-intervention period and then when the full sample is used. Of course, such tests are not possible for the target in the post-intervention period when only pulses are estimated. When there are several controls, it may be useful to check whether they remain co-integrated after the intervention.

# 5 Applications

In this section we look at three prominent applications in the literature and examine the various strategies for the selection of a set of controls. Estimates based on the RLS SC and models involving the specification of a stochastic trend are then compared. For the first two applications the underlying framwork is one of balanced growth based on an IRW trend.

## 5.1 Smoking in California

Proposition 99 came into effect in January 1989. It contained a wide range of laws and measures to cut down on smoking. Full details can be found in ADH(2010, pp 497-8). The data used by ADH consists of annual state level observations on per capita cigarette consumption[11].

A plot of the data for California up to 1988 shows a slow increase in the early 1970s followed by a downward trend beginning in the late 1970s. The change in direction is also apparent from a plot of first differences. A univariate time series model therefore needs to include a stochastic slope to allow for the changes in direction. In fitting a univariate local linear trend model, as in (2), the level variance is set to zero so as to minimize the number of parameters to be estimated and to ensure that there is enough flexibility to capture the changes in slope. The estimate of the signal-noise ratio, $\sigma_\zeta^2/\sigma_\varepsilon^2$,

[10] The SC inference referred to in this quote is based on a placebo method introduced by ADH. Paraphrasing ADH (2010, p501), 'Could our results have been obtained entirely by chance?'. They try to answer the question by seeing what happens when each of the controls is taken to be the target. As such it takes account of the control variable selection strategy as well as the assignment of weights.

[11] We follow ADH in not taking logarithms.

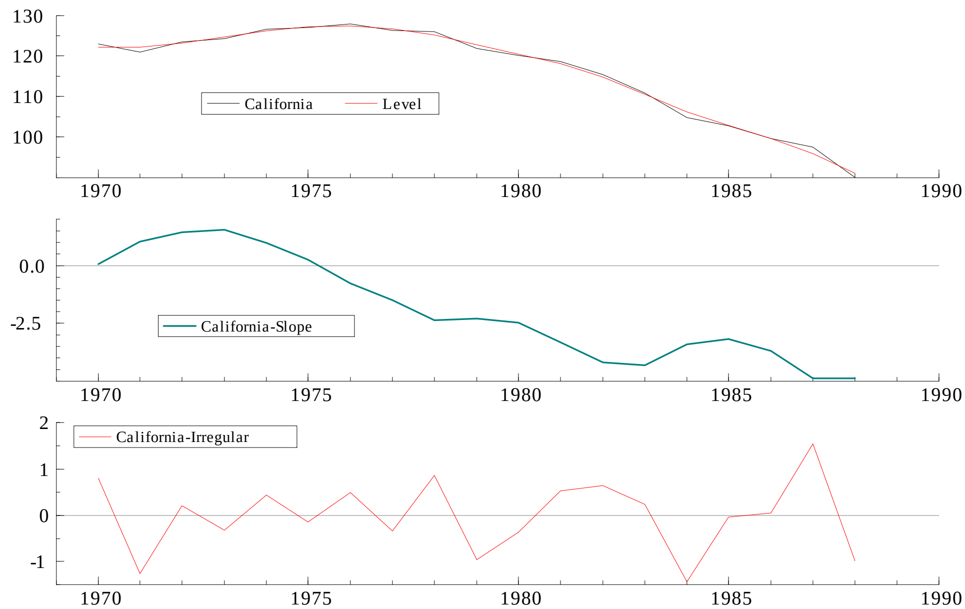



Figure 2: Fit of a local linear trend model to California smoking data from 1970 to 1988. Trend, slope and irregular are obtained by Kalman filter and smoother using the STAMP package of Koopman et al (2008).

is 3.334. There is little evidence of residual serial correlation as $r(1) = 0.067$ while $Q(5) = 8.899$. Figure 2 shows the fitted trend, together with the slope and irregular.

Figure 3 shows the forecasts from 1989 onwards. Although the first few observations are below the predicted values, the difference is not great. Only two are beyond one root mean square error (RMSE) and then only just. Thus it appears that the law had no effect. Bringing controls into the picture tells a different story.

Having discarded some states, mainly because of action they took to combat smoking in the post-intervention period, ADH (2010) obtained a donor pool of 38 states. On the basis of a set of predictor variables, as given in their Table 1, they chose five states which were assigned weights so as to best reproduce the pre-intervention California observations. The states and their weights are: Colorado (0.164), Connecticut (0.069), Montana (0.199), Nevada (0.234) and Utah (0.334).

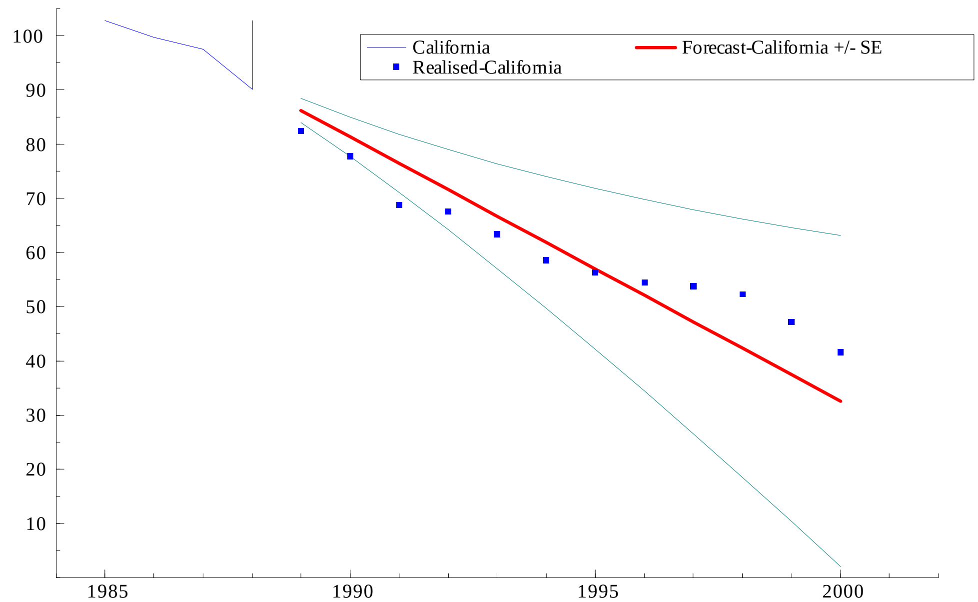



Figure 3: Forecasts from a univariate model for California

If the assumptions underlying the balanced growth model are to be satisfied, the contrast, or difference, between the target and each of the controls should be a stationary time series. Figure 4 shows the contrasts between the five states in the ADH control group and California, together with their correlograms. Colorado and Montana seem to be on the same path, in that they are moving in parallel with California. (It is not necessary for the difference to be zero.) The correlograms support this conclusion in that the autocorrelations after the first soon become small. This is not the case with the other three states. The Utah and Nevada contrasts are both moving towards zero but from opposite directions, so it seems that they are offseting each other.

On the basis of the preceding analysis, we decided to fit a trivariate model to California, Colorado and Montana using the STAMP package of Koopman et al (2008). For the data before 1989, a good fit was obtained. We then estimated the model for all 31 observations, but with 1989 onwards treated as missing for California ( This is essentially the same as estimating a full set of pulse dummies and subtracting them from California.) The smoothed estimates for post 1988 California are shown in Figure 5 together with the

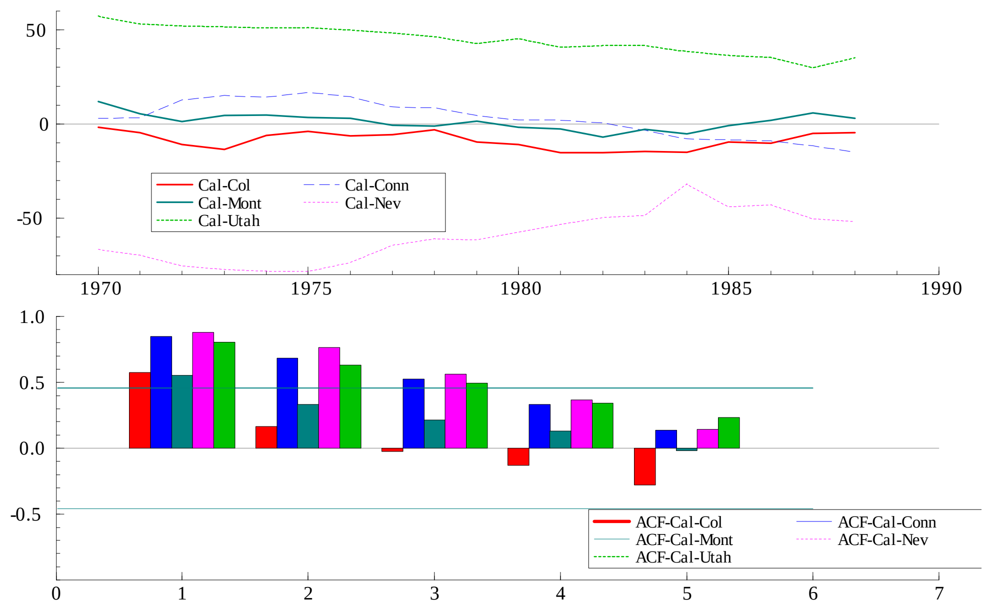



Figure 4: Contrasts between California and Colorado (Col), Connecticut (Conn), Montana (Mont), Nevada (Nev) and Utah, together with their respective correlograms.

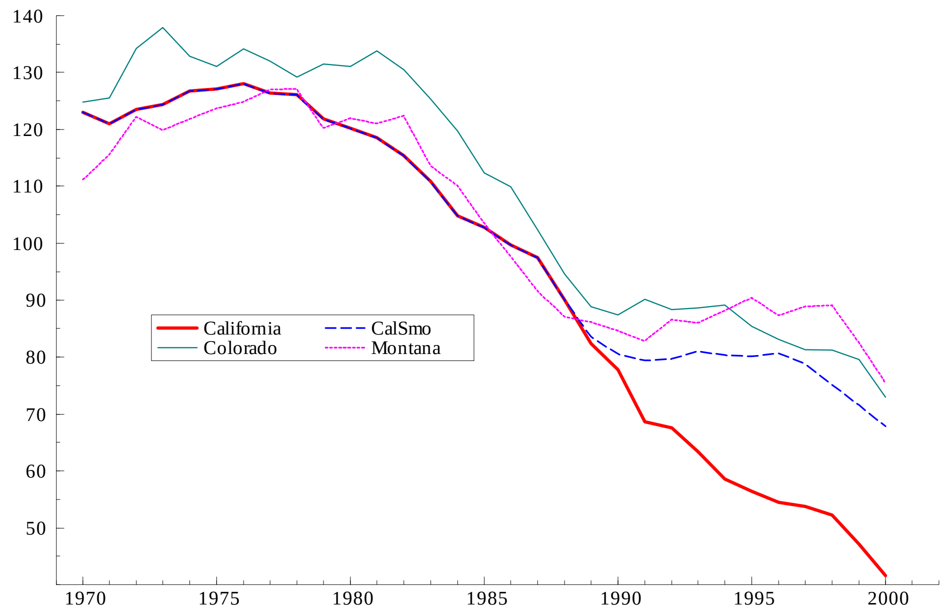
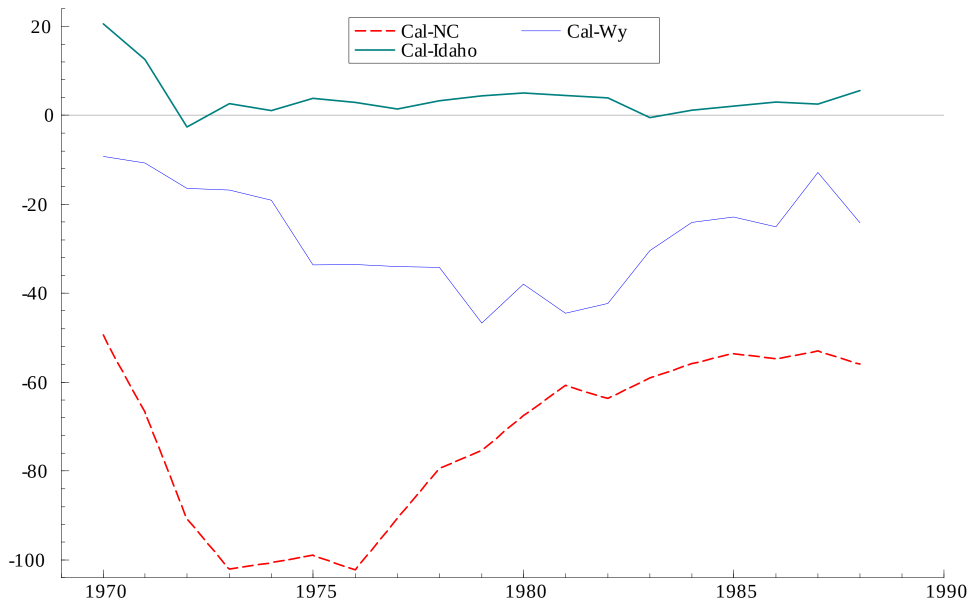



Figure 5: Smoothed estimates with post 1988 California observations treated as missing

series for Colorado and Montana.

The apparent unsuitability of two of the ADH controls prompts a return to the donor pool. The results for KPSS(2) tests applied to the contrasts with California prior to the intervention, are shown in Table 3, together with the contrast for the ADH synthetic control. It is reassuring that Colorado and Montana are in the top four. On the other hand, both Connecticut and Utah - included in the ADH control- have high values. This is consistent with the correlograms of Figure 4 and supports our decision not to include them. Nevada is also rejected at the 10% level of significance but the statistic is smaller and close to that of the ADH SC. Clearly Idaho is a strong contender for a control, as are North Carolina and Wyoming. However, the graph of the contrasts in Figure 4 shows North Carolina and Wyoming to be much more variable than Idaho and therefore perhaps less suitable. The variances of the contrasts presented in the last column of Table 3 are highly informative in this respect.




| | KPSS (2) | Rank | ADH Weight | Variance |
|---|---|---|---|---|
| **Idaho** | **0.218** | **1** | **-** | 24.9 |
| **North Carolina** | **0.249** | **2** | **-** | 364.5 |
| **Colorado** | **0.286** | **3** | 0.164 | 19.7 |
| **Montana** | **0.309** | **4** | 0.199 | 19.4 |
| **Wyoming** | **0.334** | **5** | **-** | 129.9 |
| ADH Synth Cont | 0.421 | - | - | |
| Nevada | 0.422 | 6 | 0.234 | 183.7 |
| Kentucky | 0.506 | 7 | - | |
| North Dakota | 0.522 | 8 | - | |
| Delaware | 0.531 | 9 | - | |
| Indiana | 0.547 | 10 | - | |
| Connecticut | 0.593 | 11 | 0.069 | 95.3 |
| Utah | 0.625 | 15 | 0.334 | 56.7 |

Table 3 Ordered stationarity tests statistics for contrasts with California from 1970 to 1988 ($T = 19$). Bold typeface indicates a failure to reject the null of co-integration at the 10% level of significance; the critical value is 0.347.

On the basis of the KPSS results, it was decided to include Idaho in the

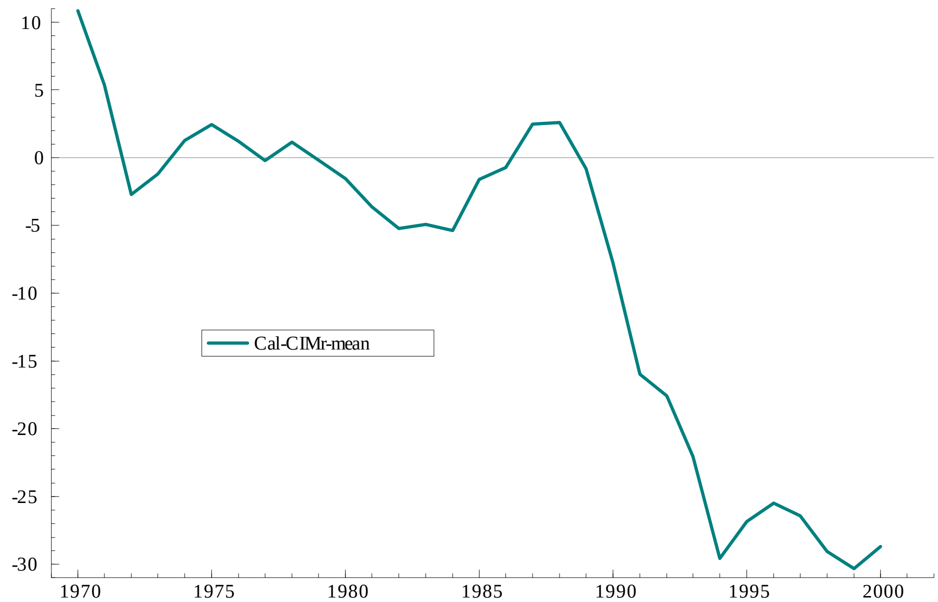



Figure 6: California-SC [Colorado, Idaho, Montana], mean -adjusted

control group. It is re-assuring that after 1988 Colorado, Idaho and Montana continue along a similar path. The RLS weights[12] for the period 1970-88 were 0.385 for Colorado, 0.288 for Idaho and 0.327 for Montana. The OLS estimates summed to 0.939 and were 0.356, 0.275 and 0.308 respectively. It is interesting to note that when Montana was replaced by Wyoming, which has a higher variance, RLS gave estimates of 0.609, 0.410 and $-0.019$. The Wyoming weight is small and with the wrong sign. North Carolina fared similarly when it replaced Wyoming: its coefficient was 0.04. The reason for the small weights is apparent from the second term in (3).

The SC constructed from our RLS weights is contrasted with California in Figure 6. The contrast has been adjusted for the mean prior to the intervention. The SC hovers around zero prior to 1989 and indicates that the effect of the intervention may have stabilized by the mid-90s; see also Figure 5. In what follows a level break is assumed for 1995.

Fitting the multivariate model with contrasts with respect to Colorado

[12] The RLS weights were obtained by regressing the California-Colorado contrast on the Colorado contrasts with Idaho and Montana.

gave a slope disturbance variance of 3.736. The estimate of the irregular covariance matrix can be transformed to yield a covariance matrix for the states themselves ( as opposed to contrasts) to give, in the order Colorado, Montana and Idaho,

$$\widetilde{\mathbf{\Sigma}}_{\varepsilon} = \begin{bmatrix} 14.78 & -5.32 & 3.11 \\ -5.32 & 19.35 & 9.42 \\ 3.11 & 9.42 & 30.98 \end{bmatrix}, \quad \widetilde{\sigma}_{\varepsilon} = \begin{bmatrix} -3.25 \\ -0.72 \\ 4.24 \end{bmatrix} \quad \text{and} \quad \widetilde{\sigma}_0^2 = 2.45,$$

from which $\widetilde{\boldsymbol{\beta}}' = (-0.367, -0.261, 0.253)$ and $\mathbf{i}'\widetilde{\boldsymbol{\beta}} = -0.376$. Inserting the above figures into the population RLS formula, (10), gives $\widetilde{\mathbf{w}}_{\varepsilon} = (0.395, 0.363, 0.242)'$; these weights are close to those obtained from the RLS regression.

The bivariate model was fitted to the contrast between California and the RLS SC. This gave a slope disturbance variance for SC of 2.298 and a (transformed) irregular covariance matrix for California and SC equal to

$$Var\begin{pmatrix} \varepsilon_{0t} \\ \varepsilon_t \end{pmatrix} = \begin{bmatrix} 0.979 & -0.701 \\ -0.701 & 9.379 \end{bmatrix}$$

The single equation for the California constrast with SC had a residual variance of 9.54 which is similar to the figure of 11.76 that the bivariate model produced for the irregular variance of this contrast.

Table 4 shows the intervention effects and their estimated standard errors as obtained from the full multivariate model, the bivariate model and the single equation for the target minus the SC. The estimated SEs indicate a gain from using the models, but remember that they are computed conditional on the parameters[13]. The bivariate model appears to perform as well as the multivariate model. The reason the model estimates have smaller SEs than those obtained by simply subtracting the SC is that the sum of regression coefficents, $\boldsymbol{\beta}$, implied by the estimated covariance matrix of the irregular components is negative. In the multivariate model it is $-0.376$, as in the previous paragraph, and in the bivariate model it is $-0.075$.

[13] Allowance can be made for serial correlation in the SC equation by noting that $r(1) \simeq 0.5$ whereas higher order correlations are close to zero.

| | Multivariate | | | Bivariate | | SC | |
|---|---|---|---|---|---|---|---|
| Year | Estimate | SE | Estimate/SE | Estimate | SE | Estimate | SE |
| 1989 | −2.55 | 1.42 | −1.80 | −4.87 | 1.60 | −0.81 | 3.68 |
| 1990 | −8.20 | 1.90 | −4.31 | −7.41 | 1.89 | −7.75 | 3.68 |
| 1991 | −15.91 | 2.05 | −7.41 | −16.03 | 2.02 | −15.97 | 3.68 |
| 1992 | −16.87 | 2.07 | −8.16 | −17.39 | 2.05 | −17.58 | 3.68 |
| 1993 | −22.52 | 2.00 | −11.31 | −21.78 | 2.00 | −22.06 | 3.68 |
| 1994 | −29.00 | 1.74 | −16.68 | −26.28 | 1.89 | −29.58 | 3.68 |
| **1995** | **−27.85** | **1.47** | **−18.91** | **−28.12** | **1.53** | **−27.82** | **1.68** |

Table 4a Estimated intervention effects for California; level change in 1995

Table 4b shows the regression estimates obtained for the ADH SC. Whereas the differences between the three estimation procedures in Table 4a are never very big, this is not true of the differences between them and the estimates given by the ADH SC. Note that the SEs of the ADH estimates are smaller than the SC SEs in Table 4b. This is because the ADH SC is made up of five states rather than three. Finally it is worth adding that the estimates obtained for 1989 to 1994 simply by subtracting the ADH SC from the target only differ from those shown in Table 4b in the second decimal place.

| | ADH SC | |
|---|---|---|
| Year | Estimate | SE |
| 1989 | −7.57 | 1.86 |
| 1990 | −9.67 | 1.86 |
| 1991 | −13.47 | 1.86 |
| 1992 | −14.07 | 1.86 |
| 1993 | −17.77 | 1.86 |
| 1994 | −22.11 | 1.86 |
| **1995** | **−23.83** | 0.85 |

Table 4b Estimated intervention effects for California with ADH SC

## 5.2 German Re-unification

The Berlin wall came down on November 9th 1989 and German re-unification took place on October 3rd 1990. The question addressed in ADH(2015) is whether re-unification led to a fall in real per capita GDP for what had

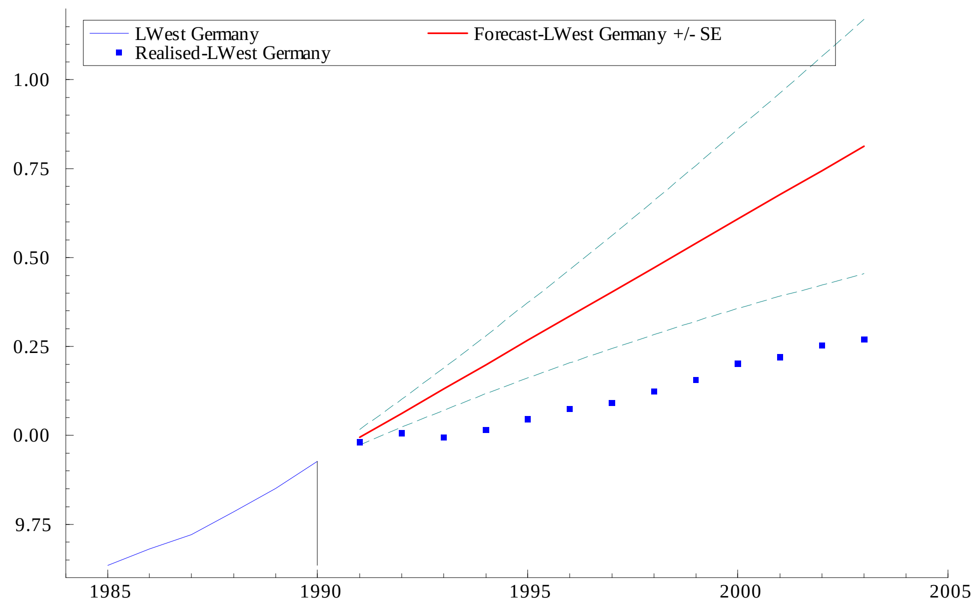



Figure 7: Univariate predictions for annual real per capita GDP in West Germany, in 2002 US dollars.

been West Germany. For the pre-unification characteristics they employ various predictors of economic growth, specifically per capita GDP, inflation, industry share of value added, investment rate, schooling, and a measure of trade openess. Drawing on a donor pool of 16 OECD counties, they construct weights for the SC by cross-validation (CV) over the period 1971 to 1990, using the first ten years for training and the second ten for validation. The final weights, based on the data from 1981 to 1990, are Austria (0.42), Japan (0.16), Netherlands (0.09), Switzerland (0.11) and USA (0.22).

Unlike ADH we follow the usual practice of transforming to logarithms. The predictions for West Germany from a univariate model with IRW trend, shown in Figure 7, indicate a gap that is big and increasing. The question is whether a similar gap will be found when post-unification Germany is compared with countries which were on a similar growth path prior to 1990.

An inspection of Figure 8 immediately reveals a problem with one of the countries selected by ADH, namely Japan. Japan has a bigger growth

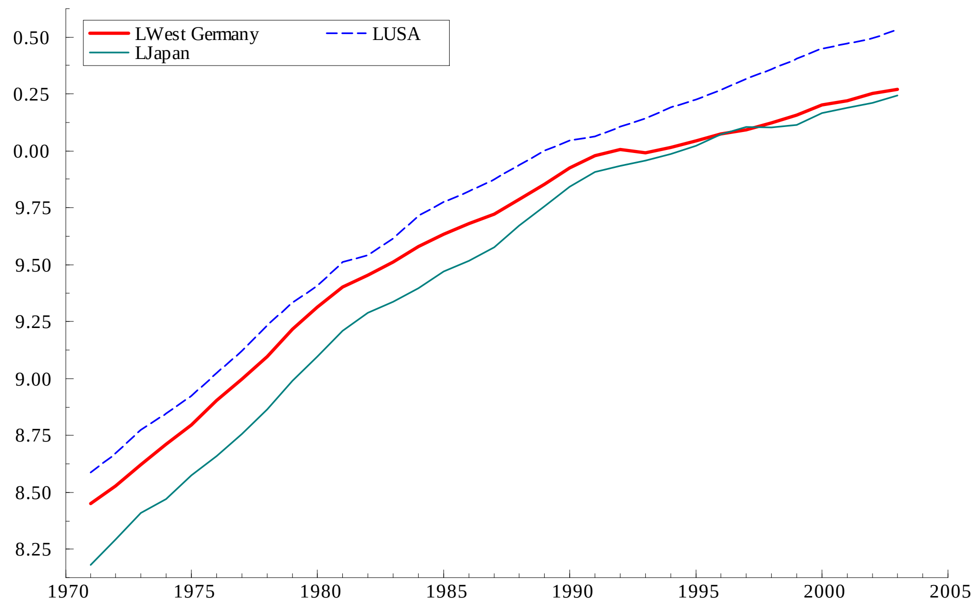



Figure 8: Annual per capita GDP in West Germany, USA and Japan.

rate than Germany up to 1990 and then slows down for reasons primarily connected with an ageing population. Unlike Austria and USA, the Japan contrast shown in Figure 9 is clearly on a downward path before 1990 and this continues after 1989. In other words, Japan is closing the gap with Germany. However, after 1997 the gap starts to increase.

Table 5 shows the KPSS(2) test results for the individual series selected by ADH and the contrasts with West Germany. The first column is for a deterministic versus a stochastic trend, whereas the second is for a fixed slope in an I(1) series versus an I(2) series, that is a stochastic slope. The results are consistent with our argument that Germany, USA and Austria are all best modeled as I(2) series. The KPSS balanced growth co-integration test in the third column supports the notion that before unification, Germany, USA and Austria were on the same growth path[14]. The unsuitability of Japan as a

[14]The KPSS(1) results are similar. We also tested for co-integration with a trend included. This test rejected co-integration in all countries (apart from Netherlands) because the trend is very weak and so the test statistic did not change enough to compensate for the smaller critical value.

control is confirmed and it seems that the Netherlands and Switzerland are also ruled out. Despite this last finding, the aggregate in the ADH SC does appear to be co-integrated with Germany.

| Null | I(0) | I(1) | Coint (Level) | Var$\times 10^{-5}$ | ADH weight |
|---|---|---|---|---|---|
| Trend | Yes | No | No | - | |
| Crit Value (10%) | 0.119 | 0.347 | 0.347 | - | |
| West Germany | 0.184* | 0.445* | - | - | |
| ADH Synth Cont | 0.181* | 0.528** | 0.206 | – | |
| USA | 0.185** | 0.455* | 0.166 | 0.357 | 0.22 |
| Austria | 0.192* | 0.513** | 0.212 | 0.487 | 0.42 |
| Netherlands | 0.189** | 0.434* | 0.665** | | 0.09 |
| Switzerland | 0.125* | 0.152 | 0.587** | | 0.11 |
| Japan | 0.170** | 0.331 | 0.701** | | 0.16 |
| France | 0.181** | 0.429* | 0.056 | 0.164 | |
| Italy | 0.184** | 0.416* | 0.648** | | |

Table 5. KPSS(2) tests for log of annual real per capita GDP from 1971 to1989.
Significance at a 5% and 10% level is indicated by ** and * respectively.

ADH included neither France nor Italy in their control set. However, Table 5 suggests that France should be included, and this is confirmed by Figure 10. After 1989 France continues on the same path as Austria and USA.

The RLS weights for the period 1971-89 were 0.153 for Austria, 0.373 for USA and 0.474 for France. The corresponding OLS estimates were 0.382, 0.396 and 0.212. Thus RLS is not close to OLS, even though the OLS coefficients sum to 0.990. The main reason is that the Franco-German contrast has a much smaller variance than that of Austria and Germany.

Figure 11 shows the contrast of our RLS-SC with Germany. The figure also shows the contrast of Germany with the average of Japan and Switzerland, countries which together make up 27% of the SC of ADH. Our RLS-SC appears stationary before 1990. The same cannot be said of the the average of Japan and Switzerland.

Multivariate, bivariate and single equation models were fitted on the assumption that the unification effect had stabilized by 1999. All three models fit well, with very little residual serial correlation. The estimates of the intervention effects for the full model and the SC regression are shown in Table

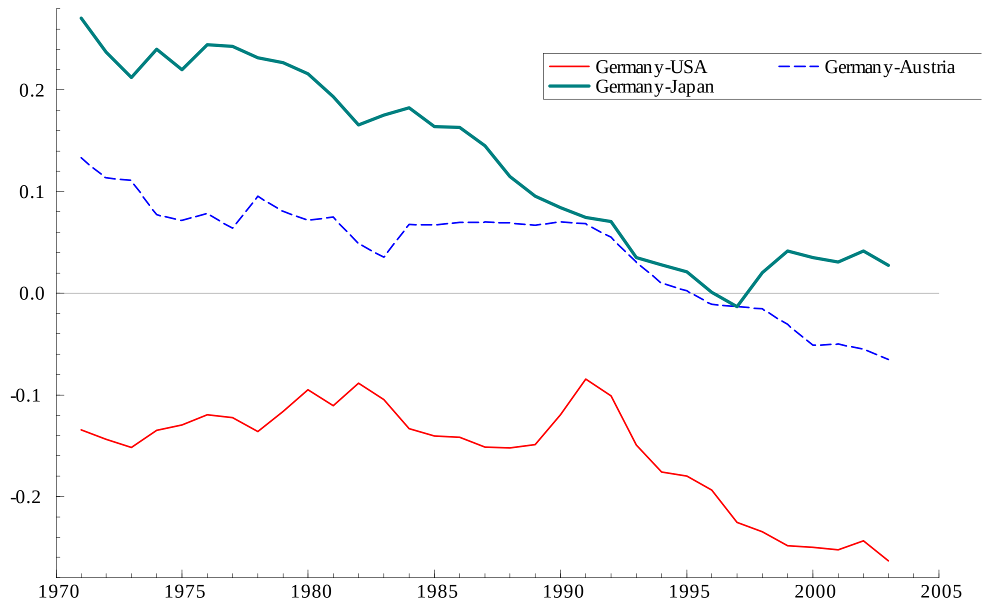



Figure 9: Contrasts between Germany and Austria, Japan and the USA.

6. The bivariate model is omitted because the estimates are almost identical to those of the single equation and the SEs are identical. The full model and SC results are very close as are their SEs. The reason the SC performs so well is that when the bivariate model is estimated it is found that the implied value of $\beta$ is 0.957. Thus this combination of controls is almost perfect. Note that the univariate forecasts in Figure 7 indicate a much bigger gap than is obtained with our preferred control series.

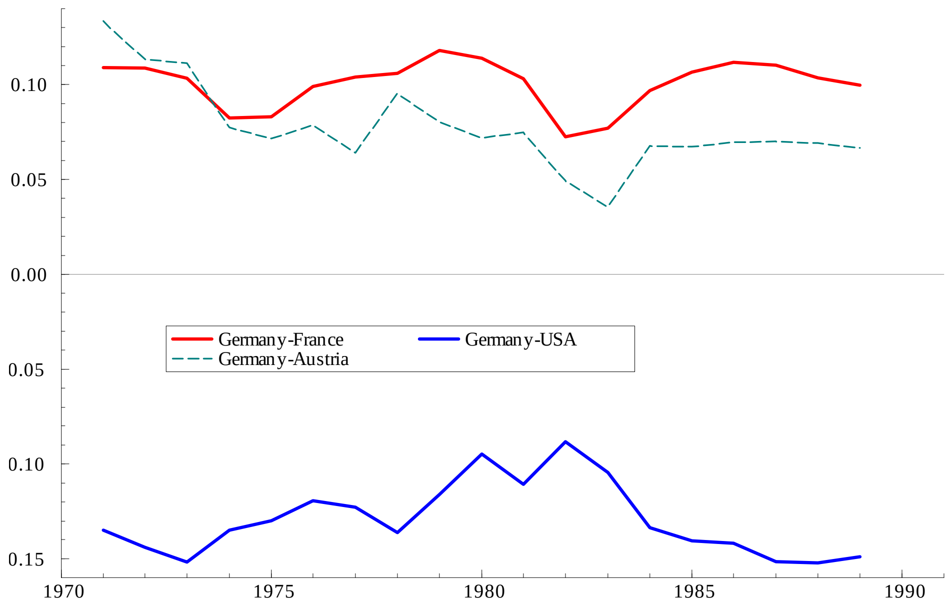



Figure 10: Contrasts for Germany with Austria, France and USA

| | Multivariate | | | SC | |
|---|---|---|---|---|---|
| Year | Estimate | SE | Estimate/SE | Estimate | SE |
| 1991 | 0.0271 | 0.0091 | 2.97 | 0.0276 | 0.0093 |
| 1992 | 0.0123 | 0.0091 | 1.35 | 0.0137 | 0.0093 |
| 1993 | −0.0205 | 0.0091 | −2.45 | −0.0183 | 0.0093 |
| 1994 | −0.0422 | 0.0091 | −4.63 | −0.0387 | 0.0093 |
| 1995 | −0.0495 | 0.0091 | −5.43 | −0.0466 | 0.0093 |
| 1996 | −0.0595 | 0.0091 | −6.52 | −0.0556 | 0.0093 |
| 1997 | −0.0846 | 0.0091 | −9.29 | −0.0820 | 0.0093 |
| 1998 | −0.0934 | 0.0091 | −10.25 | −0.0915 | 0.0093 |
| **1999** | **−0.1168** | **0.0045** | **−26.26** | **−0.1147** | **0.0045** |

Table 6 Estimated intervention effects for Germany. Level break in 1999

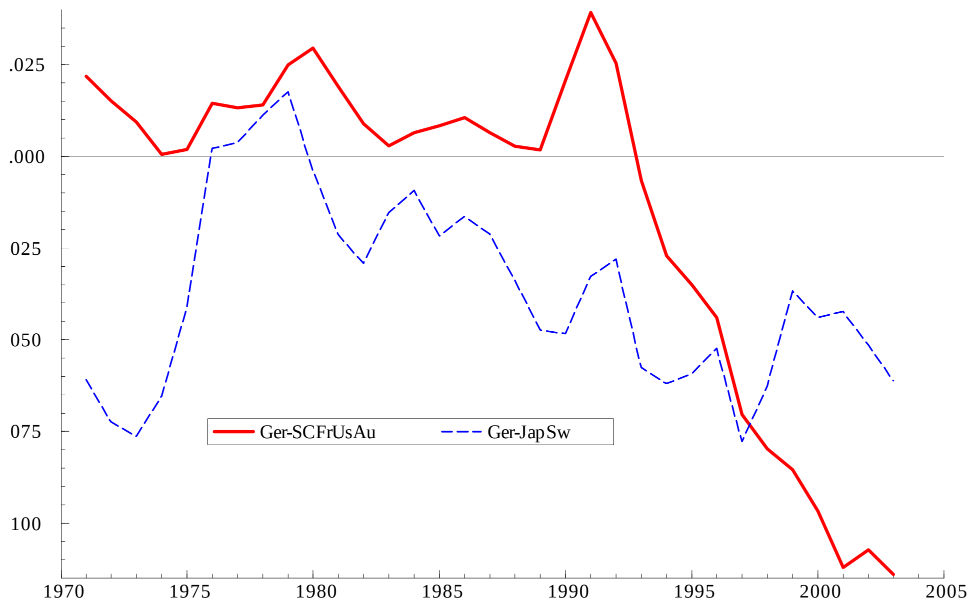



Figure 11: Contrast of Germany with RLS-SC made up of Austria, France and USA and contrast of Germany with the average of Japan and Switzerland.

## 5.3 Hong Kong

Hsiao et al (2012) - HCW - propose a panel control variable approach for assessing the impact on the change in Hong Kong sovereignty in July 1997. They use quarterly growth rates on real GDP which they then treat as stationary and deseasonalized. As noted by Gardeazabal and Vega-Bayo (2016), differencing results in a loss in information and so is not ideal for assessing the effect of an intervention. For example, an immediate break appears as a single outlier and so could easily be missed. Such problems do not arise if seasonality is handled as described in the next section.

The above considerations notwithstanding, it is instructive to examine the control selection methodology used by HCW. On the basis of the 18 observations from 93(1) to 97(2) they choose four countries out of a pool of ten, which was determined according to geographical considerations. The choice was based on regressing Hong Kong on groups of countries and selecting the group that gave the best fit. The estimated regression coefficients, with 't-statistics' in parentheses were as follows: Japan: $-0.675(-6.052)$, Korea: $-0.432$ $(-6.821)$, USA: 0.486 (2.214) and Taiwan: 0.793 (2.558). As can be seen, Japan and Korea have negative weights whereas USA and Taiwan are weighted positively. The sum of the weights is 0.172. The net result, as shown in Figure 12, is a large negative estimate of a temporary intervention effect. The reason is simple. The Asian crisis, which began in July 1997, affected many countries including Korea and the large negative weight for Korea in the panel SC causes it to go up. HCW do not mention the fact that their method has resulted in the construction of such an inappropriate control because they conclude that the intervention effect is temporary and statistically insignificant[15].

# 6 Seasonality[16]

A seasonal component in a structural time series model is usually captured by a multivariate random walk for a vector of seasonal effects, which are constrained to sum to zero; see Koopman et al (2008) or Proietti (2000). Full ML estimation can be carried out on (4) with a seasonal component in

[15] We note in passing that a temporary change in growth rates can give a permanent change in level, but this is not relevant to the point we are making.

[16] This section can be omitted without a loss in continuity.

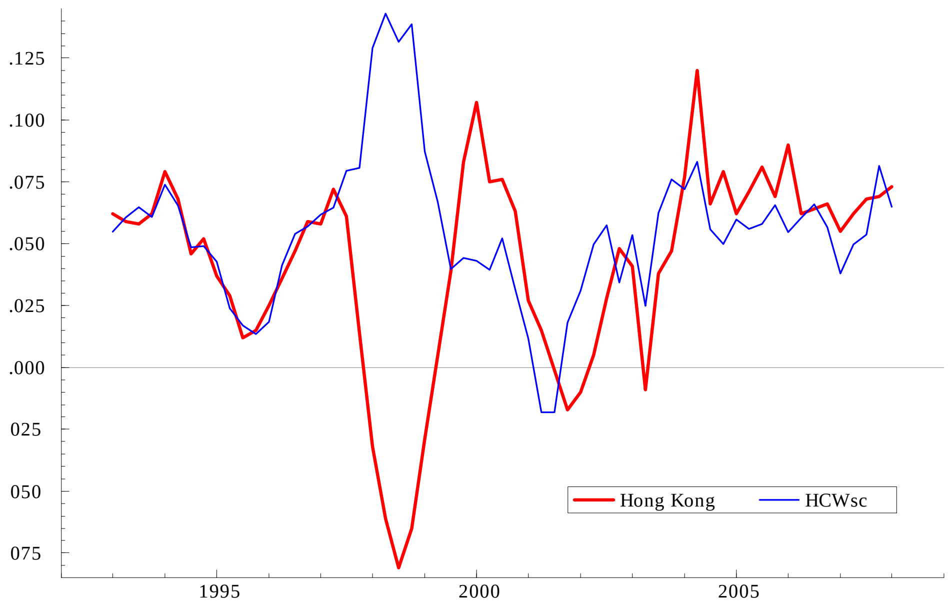



Figure 12: Hong Kong and panel control group

each equation. If there is a pool of potential donors, valid controls may again be selected by stationarity tests. With deterministic seasonality, Phillips and Jin (2002) have shown that the critical values of the KPSS test are unaffected. When seasonality is stochastic, a test can be carried out on seasonal moving averages.

**Remark 6** *When there is a dynamic effect for the intervention, as opposed to an immediate level shift, prior seasonal adjustment of the target series is problematic: it can only be done using the pre-intervention observations with the estimated seasonal pattern just before the intervention being used to seasonally adjust the post-intervention observations. In a model, the estimated seasonal in the target can take account of any observations after the effect of the intervention has stabilized as well as benefitting from correlations with the seasonals in other series.*

A seasonally adjusted SC may be obtained from an equation in which $y_{0t} - y_{it}$ depends on the other contrasts, as in RLS, plus a stochastic seasonal[17].

[17] The form of the seasonal model remains the same in a linear combination of variables.

The intervention variables may then be estimated by extending (5) to include a seasonal component. Alternatively they may be added to the contrast equation used to construct the seasonally adjusted SC, in the same way as was suggested at the end of sub-section 2.3.

An example is provided by the 1983 car seat belt law in Great Britain. Quarterly series on (logarithms of) numbers killed and seriously injured, from 1969(1) to 1984(4) were analysed in Harvey (1986) using rear seat passengers, who were not required to wear belts, as a control for front seat passengers; see also Koopman et al (2008, p 103-4). Figure 13 shows the effect of the seat belt law as estimated by pulse dummies from 83(1) to 84(4) in a bivariate model with balanced growth and by the difference between driver and rear seat passengers, having made allowance for the seasonals[18] as well as the difference in level. The difference estimates are much more variable, reflecting the contrast between $\varepsilon_{0t}$ and $\varepsilon_{1t}$. Furthermore, the effect of the law as measured by the differences seems to increase with time, with the (deseasonalized) average in 1984 being $-0.309$ as opposed to $-0.231$ in 1983. On the other hand, the bivariate model shows no evidence for an increasing response, because in 1983 the change is $-0.253$ and in 1984 it is $-0.267$. Re-estimating the model with an additional level shift in 1984 gave no change in the 83(1) shift (to three significant figures), while for 84(1), the change is $-0.014$ (0.044), which is clearly statistically insignificant.

Figure 14 shows the smoothed series after estimating a hard break in 83(1) from a balanced growth model. The estimate of $\lambda$ was $-0.259$ with an estimated SE of 0.028. There was very little residual serial correlation. The DD estimate of $\lambda$ is $-0.270$ with an estimated SE of 0.031, indicating a slight loss in efficiency.

# 7 Conclusion

In this article we have shown how simple and transparent dynamic models can be used to estimate the effect of an event or policy change on a nonstationary time series using control series. The time series framework also enables us to analyse synthetic control and difference in differences estimators. If these methods are to be valid, the target and a control series must have a common

[18]The seasonals are fixed and estimated by assuming that there is a hard level shift.

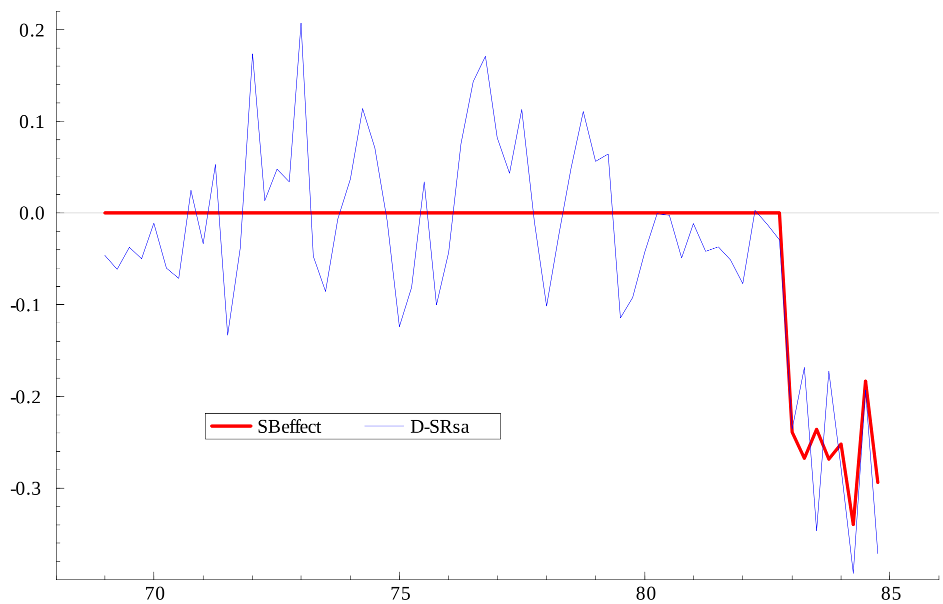



Figure 13: Effect of seat belt law as captured by bivariate model (bold) and difference between drivers and rear seat passengers.

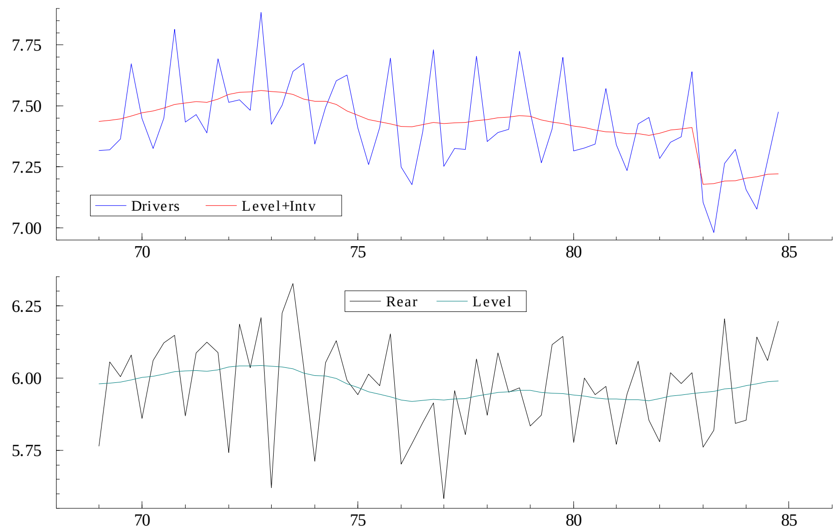



Figure 14: Smoothed levels of (logarithms of ) Drivers and Rear Seat Passengers killed and seriously injured in Great Britain, with allowance made for the seat belt law.

trend with the property that subtracting one series from the other gives a stationary series prior to the intervention. This balanced growth assumption is implicit in studies involving synthetic control and difference in differences estimation but it is rarely tested, or even acknowledged.

In order to determine whether series from a pool of potential donors are valid as controls, we propose selecting series that are found to be co-integrated with the target series on the basis of stationarity tests applied to the contrast with the target. We show the tests to be reliable for small samples. Graphs of the series provide supporting evidence and transparency. Once a set of valid controls has been selected, the weights for a synthetic control are found by a restricted least squares regression. This approach is in stark contrast to data-driven methods that select control series and weight them by regressions involving what are thought to be explanatory variables[19]. The applications to the California smoking law of 1989 and to German re-unification call into

[19]In order to be viable, such methods require that the sample prior to the intervention be large enough to allow it to be broken into training and estimation sub-samples. However, even if this can be done, there is still no guarantee that the series that make up the resulting SC will be valid.

question the validity of some series chosen as controls in the original studies and, at the same time, suggest that a few excluded series may be strong candidates for inclusion.

When a valid synthetic control has been chosen according to time series considerations, the analysis can be extended by setting up a regression equation in which permanent and transitory effects of an intervention can be estimated, together with their standard errors. Extensions to include seasonal effects and serially correlated components are easily implemented. The simplicity of such a synthetic control approach is attractive. However, because a SC only estimates the underlying trend from contemporaneous observations, a model can still offer significant gains in the accuracy with which intervention effects can be estimated. Furthermore the very act of formulating a model highlights the assumptions being made and suggests ways of testing them. Indeed even if the target and control series do not exhibit balanced growth, a suitably specified model still allows valid inference to be made on intervention effects.

**Acknowledgements**

Earlier versions of this paper were given at the UQ/QUT joint workshop in Brisbane, the 2017 York Econometric Symposium, the NIESR in London, the universites of Essex and Oxford and at the VU in Amsterdam. We would like to thank Simon Price, Sam Astill, Robert Taylor, Bent Nielsen, Siem Jan Koopman, Chico Blasques and other participants in these seminars for helpful comments, though, of course, they bear no responsibility for any errors or opinions contained in the paper. Finally we would like to thank Guido Imbens for introducing us to the synthetic control literature and some of the key data sets.

**References**

Abadie, A., Diamond, A., and J. Hainmueller. (2010) Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105, 493-505.

Abadie, A., Diamond, A., and J. Hainmueller. (2014) Comparative politics and the synthetic control method. *American Journal of Political Science*, 59, 2011-25

Angeriz, A. and P. Arestis. (2008). Assessing Inflation Targeting Though Intervention Analysis, *Oxford Economic Papers*, 60, 293-317.

Bai, C., Li, Q., Ouyang, M., 2014. Property Taxes and Home Prices: A Tale of Two Cities. *Journal of Econometrics*, 180, 1–15.

Bertrand, M., Duflo, E. and S. Mullainathan (2004). How much should we trust difference-in-differences estimates? *The Quarterly Journal of Economics*, 119, 249-275.

Brodersen, K. H., Gallusser, F. Koelher, J., Remy, N. and S. L. Scott (2015). Inferring causal impact using Bayesian structural time series models, T*he Annals of Applied Statisti*cs, 9, 247–274.

Busetti, F. and A. C. Harvey (2001) Testing for the Presence of a Random Walk in Series with Structural Breaks. *Journal of Time Series Analysis*, 22, 127-50.

Card, D. and A. Krueger (1994). Minimum wages and employment: Case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review,* 84, 772-93.

Doudchenko, N. and G. W. Imbens. (2016). Balancing, Regression, Difference-in Differences and Synthetic Control Methods: A synthesis. mimeo.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods*. 2nd edition. Oxford: Oxford Statistical Science Series.

Gardeazabal, J. and A. Vega-Bayo (2016). An Empirical Comparison between the Synthetic Control Method and Hsiao's et al's Panel Data Approach to Program Evaluation. *Journal of Applied Econometrics* (forthcoming).

Harvey A.C. (1996). Intervention analysis with control groups, *International Statistical Review,* 64, 313-28.

Harvey A.C. and D. Delle Monache. (2012). Specification and Misspecification of Unobserved Components Models. In Economic Time Series: Modeling and Seasonality, Edited by W. R. Bell, S. H. Holan and T. S. McElroy, 83-108. Chapman and Hall/CRC.

Hsiao, C., Ching H. S., and S. K. Wan (2012). A panel data approach for program evaluation: measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics,* 27, 705–740.

Imbens, G.W. and D.B. Rubin (2015). *Causal Inference in Statistics, Social and Biomedical Sciences.* CUP.

Koopman, S.J., A.C. Harvey, J.A. Doornik and N. Shephard (2008). *STAMP 8.2 Structural Time Series Analyser, Modeller and Predictor*, London: Timberlake Consultants Ltd.

Kwiatkowski, D., Phillips, P.C.B., Schmidt P. and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 44, 159-78.

Nyblom, J. and T. Mäkeläinen (1983). Comparison of tests for the presence of random walk coefficients in a simple linear model, *Journal of the American Statistical Association,* 78*,* 856-64.

Phillips, P.C.B. and S. Jin (2002).The KPSS test with seasonal dummies. *Economics Letters,* 77, 239–243.

Proietti, T. (2000). Comparing seasonal components for structural time series models, *International Journal of Forecasting,* 16, 247-60.

Varian, H. R. (2014). Big data: new tricks for econometrics. *Journal of Economic Perspectives,* 28, 3-28.

Vujic, S., Commandeur, J.J.F. and S.J. Koopman (2016). Intervention time series analysis of crime rates: The case of sentence reform in Virginia, *Economic modelling*, 57, 311-323.