JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **32**: 983–1002 (2017)
Published online 4 December 2016 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2557

AN EMPIRICAL COMPARISON BETWEEN THE SYNTHETIC CONTROL METHOD AND HSIAO *ET AL.*'S PANEL DATA APPROACH TO PROGRAM EVALUATION

JAVIER GARDEAZABAL* AND AINHOA VEGA-BAYO
Departamento de Fundamentos del Análisis Económico II, University of the Basque Country, Bilbao, Spain

SUMMARY

We compare two program evaluation methodologies: the synthetic control method and the panel data approach. We apply both methods to estimate the effect of the political and economic integration of Hong Kong. The results obtained differ depending on the methodology used. We then conduct a simulation that shows that the synthetic control method results in a post-treatment mean squared error, mean absolute percentage error, and mean error with a smaller interquartile range, whenever there is a good enough match. Copyright © 2016 John Wiley & Sons, Ltd.

Received 11 December 2014; Revised 27 September 2016

Supporting information may be found in the online version of this article.

1. INTRODUCTION

Program evaluation methodologies have long been used by social scientists to measure the effect of different economic or political interventions (treatments). The main problem is that one cannot observe both the outcome under the intervention and in the absence of intervention simultaneously, so we do not know with certainty which methodology gives the best estimate for the intervention effect.

In this paper we focus on comparing two methodologies for program evaluation: the synthetic control method (Abadie and Gardeazabal, 2003) and the panel data approach (Hsiao *et al.*, 2012). They are the only program evaluation methodologies that can handle estimating the effect of an intervention on a single unit, whereas others require more than one treated unit.[1] Moreover, with these two approaches the treated individual is usually an aggregate entity such as a region or country and the treatment is normally a political program or a similar macroeconomic intervention, while most program evaluation methodologies concentrate on microeconomic data. With microeconomic data, the possibility exists of running a field experiment with random assignment as a baseline for comparison of the two methodologies (e.g., LaLonde, 1986; Dehejia and Wahba, 1999). Since this is impossible in the case of macroeconomic treatments, we conduct a simulation experiment that compares both methodologies.

To carry out a comparison of both methodologies and see whether they really differ in their estimations, we initially apply the two procedures to the same dataset: the one used by Hsiao *et al.* (2012) to calculate the effect of the political and economic integration of Hong Kong.[2] The panel data approach

* Correspondence to: Javier Gardeazabal, Departamento de Fundamentos del Análisis Económico II, University of the Basque Country, UPV/EHU,Lehendakari Aguirre 83, 48015 Bilbao, Spain. E-mail: javier.gardeazabal@ehu.es

[1] In recent work, however, Fujiki and Hsiao (2015) extend the method and apply the panel data approach to the case of multiple treated units and disentangle the effects; there are also examples of this for synthetic control such as Barone and Mocetti (2014) and Gardeazabal and Vega-Bayo (2016).

[2] The application of the panel data approach is no more than a replication of the results obtained by Hsiao *et al.* (2012) to ensure that the methodology is correctly applied.

estimates a negative and insignificant treatment effect for the case of political integration, whereas the synthetic control results in a negative and statistically significant political integration effect. In the case of economic integration, qualitatively the results are qualitatively the same: both yield positive and significant results in some periods and non-significant in others; however, they are quantitatively different, since synthetic control estimates a smaller treatment effect than the panel data approach. We also rerun the applications, but making different changes to the donor pool; and we observe that the synthetic control's estimation of the treatment effect is more robust to changes in the set of control units. Both methods are also used to calculate the effect of the 1965 Arab–Israeli conflict and the Sierra Leone civil war. This second empirical application results in an estimated treatment effect that is negative and significant according to both methodologies, but different in terms of the estimated size, depending on the method. The detailed results of this second empirical application are available in an online Appendix (supporting information).

The second part of the paper is devoted to comparing the two methods using a simulation experiment. This allows us to observe the 'real' treated and untreated data, which is otherwise impossible, and to better study the differences between the two methods. The simulations show that the synthetic control method results in a mean squared error (MSE), mean absolute percentage error (MAPE), and mean error with a smaller (interquartile) range whenever there is a good enough match. However, when the treatment periods and the number of controls are small, both methods do equally well. Even in cases when there is not a good enough match in the synthetic counterfactual, the panel data approach does not seem to result in a systematic mean error. We also obtain evidence that the MSE, MAPE, and mean error improve when either a covariate or an individual effect is introduced, and even more so when they are both included. This is true not only for the synthetic control but also for the panel data approach.

In the sections that follow, we first define both methodologies using a common notation. We continue with an empirical application of both methods to the same dataset, i.e., that used by Hsiao *et al.* (2012), in order to measure the effect of the political and economic integration of Hong Kong with different donor pools, and discuss the results obtained. The second half of the paper focuses on running a simulation that enables us to compare both methods and infer which one results in a more accurate estimation of the treatment effect. Section 5 ends with the main conclusions. The online Appendix serves as a complement to explain in further detail the results obtained.

2. DESCRIPTION OF THE TWO METHODOLOGIES

Consider $J + 1$ units over $t = 1, \ldots, T_0, T_0 + 1, \ldots, T$ periods. The first unit is affected uninterruptedly by an intervention in period $T_0 + 1$ until period T, after an initial pre-intervention period $1, \ldots, T_0$. The leftover J units are the controls that form the so-called 'donor pool', and they are not affected by the intervention. Let Y_{jt} denote the outcome variable—the variable in which the intervention effect is being measured—of unit j at period t. Y_{jt}^1 and Y_{jt}^0 denote the potential outcome of unit j at time t under treatment and in the absence of treatment, respectively. We usually do not simultaneously observe both Y_{jt}^1 and Y_{jt}^0, but instead we observe Y_{jt}, which can be written as

$$Y_{jt} = d_{jt} Y_{jt}^1 + (1 - d_{jt}) Y_{jt}^0$$

where d_{jt} is a dummy variable that takes value 1 if unit j is under treatment at time t and value 0 otherwise (Rubin, 1974). In this case and without loss of generality, only the first unit is under intervention, so we have that

$$d_{jt} = \begin{cases} 1 \text{ if } j = 1 \text{ and } t \geq T_0 + 1 \\ 0 \text{ otherwise} \end{cases}$$

DOI: 10.1002/jae

The treatment or intervention effect for the treated unit can therefore be expressed as

$$\alpha_{1t} = Y_{1t}^{1} - Y_{1t}^{0}$$

for $t = T_0 + 1, T_0 + 2, \ldots, T$. For $j = 1$, we observe Y_{1t}^{1} but not Y_{1t}^{0}. The goal of these two program evaluation methodologies is to obtain an estimation for the effect of the intervention, $\hat{\alpha}_{1t}$, during the post-treatment period $T_0 + 1, \ldots, T$ by attempting to replicate the economy of the treated unit in the pre-intervention period $1, \ldots, T_0$; and then obtain an estimate of the outcome variable under no treatment Y_{1t}^{0} during the post-treatment period. As standard with these methods, it is assumed that there is no treatment interference between units; i.e., the outcome of the untreated units is not affected by the treatment of the treated unit. In the paragraphs that follow we explain the approach that each of the two methodologies uses to obtain Y_{1t}^{0}, and by extension, the treatment effect α_{1t}.

2.1. Synthetic Control

The synthetic control is calculated as a convex combination of the units in the donor pool and is, out of all possible synthetic controls, the one that best replicates the outcome variable of the treated unit in the pre-intervention period according to a measure of the distance between the predictors of the treated unit and those of the synthetic control.

The objective of the procedure is to find a vector of weights $\mathbf{W} = (w_2, \ldots, w_{J+1})'$ such that it minimizes the distance in the pre-treatment period:

$$\sqrt{(\mathbf{X_1} - \mathbf{X_0}\mathbf{W})'\mathbf{V}(\mathbf{X_1} - \mathbf{X_0}\mathbf{W})} \tag{1}$$

subject to $w_j \geq 0$ for all $j = 2, \ldots, J+1$ and $w_2 + w_3 + \ldots + w_{J+1} = 1$. $\mathbf{X_1}$ represents a $K \times 1$ vector of pre-intervention values of K predictors for the intervened unit, $\mathbf{X_0}$ is a $K \times J$ matrix that contains the values of the same predictors for all the possible controls, and $\mathbf{V}$ denotes a diagonal matrix with non-negative elements that reflects the relative importance of the predictors. This matrix is estimated so that the outcome variable of the treated unit is best reproduced by the estimated synthetic control $\mathbf{W}^*(\mathbf{V})$.[3] The values of the diagonal elements of $\mathbf{V}$ reflect the relative importance of the different covariates, while each value $w_2, \ldots, w_{J+1}$ is the relative weight of the corresponding control unit in the synthetic control.

One might also think of the objective function over which we minimize the distance in the pre-treatment period, Equation (1), as a decomposition of both covariates and pre-treatment values of the outcome variable. If we partition the predictor data matrices $\mathbf{X}_0$ and $\mathbf{X}_1$ into covariate values $\mathbf{Z}_0, \mathbf{Z}_1$ and pre-treatment values of the outcome variable $\mathbf{Y}_0, \mathbf{Y}_1$:

$$\mathbf{X}_0 = \begin{pmatrix} \mathbf{Z}_0 \\ \mathbf{Y}_0 \end{pmatrix}, \quad \mathbf{X}_1 = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Y}_1 \end{pmatrix}$$

and the predictor weights' matrix $\mathbf{V}$:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_Z & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_Y \end{pmatrix}$$

then the objective function to be minimized, equation (1), becomes

$$\sqrt{(\mathbf{Z}_1 - \mathbf{Z}_0\mathbf{W})'\mathbf{V}_Z(\mathbf{Z}_1 - \mathbf{Z}_0\mathbf{W}) + (\mathbf{Y}_1 - \mathbf{Y}_0\mathbf{W})'\mathbf{V}_Y(\mathbf{Y}_1 - \mathbf{Y}_0\mathbf{W})} \tag{2}$$

[3] See Abadie and Gardeazabal (2003) and Abadie *et al.* (2015) for details on how $\mathbf{V}$ is calculated.

DOI: 10.1002/jae

The synthetic control traditionally minimizes equation (2); we will see how the panel data approach minimizes the second, $(\mathbf{Y}_1 - \mathbf{Y}_0\mathbf{W})'\mathbf{V}_Y(\mathbf{Y}_1 - \mathbf{Y}_0\mathbf{W})$, with $\mathbf{V}_Y$ being the identity matrix.

Let us assume that Y_{jt}^0 is generated by the following factor model:

$$Y_{jt}^0 = \delta_t + \theta_t z_j + \lambda_t \mu_j + \varepsilon_{jt} \tag{3}$$

where δ_t is a time-varying common factor constant across units, θ_t is a vector of unknown constants, λ_t denotes time-varying unobserved common factors (such as cycles or trends), z_j is a vector of observed covariates, μ_j are unknown unit-specific factor loadings, and ε_{jt} is a random error term. If there exist weights $w_2^*, \ldots, w_{J+1}^*$ such that

$$\sum_{j=2}^{J+1} w_j^* Y_{j1} = Y_{11}, \quad \sum_{j=2}^{J+1} w_j^* Y_{j2} = Y_{12}, \quad \ldots, \quad \sum_{j=2}^{J+1} w_j^* Y_{jT_0} = Y_{1T_0}, \quad \text{and} \sum_{j=2}^{J+1} w_j^* z_j = z_1 \tag{4}$$

then one can estimate the effect of the intervention $\hat{\alpha}_{1t}$, per Abadie *et al.* (2010), as the difference between the observed outcome variable Y_{1t} and the estimated synthetic control:

$$\hat{\alpha}_{1t} = Y_{1t} - \sum_{j=2}^{J+1} w_j Y_{jt} \tag{5}$$

In practice, equation (4) rarely holds exactly because, given the data, no set of weights exists such that both sides are equal; the synthetic control is selected so that the equation holds approximately. The size of this discrepancy can be calculated in each particular case, and it is up to the researcher to decide if the synthetic control is close enough to the treated unit.

2.2. Panel Data Approach

The panel data approach developed by Hsiao *et al.* (2012) also attempts to predict Y_{1t}^0 and therefore estimates the treatment effect α_{1t} by exploiting the dependency among cross-sectional units in the donor pool and the treated unit. It is based on the following underlying model. We assume that Y_{it}^0 is generated by a factor model of the form

$$Y_{jt}^0 = \gamma_j + \lambda_t \mu_j + \varepsilon_{jt} \tag{6}$$

where γ_j denotes an individual-specific effect, λ_t is a $(1 \times K)$ vector that denotes time-varying unobserved common factors, μ_j denotes a $(K \times 1)$ vector of constants that can vary across units, and ε_{jt} is a random error term.

Y_{1t}^0 could be predicted using the underlying model Hsiao *et al.* (2012) specify. Instead, they suggest a more practical approach, i.e., using the remaining non-intervened units in the donor pool $Y_{-1t} = (Y_{2t}, \ldots, Y_{J+1t})'$ to predict Y_{1t}^0 using the regression

$$Y_{1t}^0 = \psi + \beta Y_{-1t} + \xi_{1t} \tag{7}$$

This regression is estimated during the pre-treatment only.

The modeling strategy they propose is to use R^2 (or likelihood values) in order to select the best ordinary least squares (OLS) estimate for Y_{1t}^0 using j out of the J units in the donor pool, denoted

DOI: 10.1002/jae

by $M(j)^*$ for $j = 1, \ldots, J$; then choose $M(m)^*$ from $M(1)^*$, …, $M(J)^*$ in terms of a model selection criterion.[4] As in the synthetic control method, the panel data approach only uses data from the pre-treatment period for the estimation of Y_{1t}^0 over the pretreatment period.

For a sufficiently large J or a small T_0, it is possible to have more controls (explanatory variables in the panel data approach) than time periods. In the original paper, Hsiao *et al.* (2012) restricted the donor pool to a subset of possible controls using an argument of geographical and economic proximity. This might not always be possible, or the researcher might want to allow for a more automatic process. Hence we propose the following modification to the modeling strategy previously outlined: use R^2 in order to select the best OLS estimator for Y_{1t}^0 using j out of the J units in the donor pool, denoted by $M(j)^*$ for $j = 1, \ldots, T_0 - g$; then choose $M(m)^*$ from $M(1)^*$, …, $M(T_0 - g)^*$ in terms of a model selection criterion (in our case AICc). Observe that the key difference is that while Hsiao *et al.* (2012) allowed models up to $M(J)^*$, this is now modified to allow models up to $M(T_0 - g)^*$, with $T_0 - g < J$ (see Vega-Bayo, 2016, for an implementation of the method with this modification in R).[5]

2.3. Similarities and Differences Between the Two Methodologies

Given the underlying factor models of the two methods, and if we incorporate the time-varying effects of the synthetic control δ_t from equation (3) into $\lambda_t \mu_j$, one might think of a general model that encompasses the two, with the outcome under no treatment having the following form:

$$Y_{jt}^0 = \gamma_j + \theta_t z_j + \lambda_t \mu_j + \varepsilon_t \tag{8}$$

where γ_j are the unobserved individual effects not explicitly included in the synthetic control model, and $\theta_t z_j$ are the covariate effects not considered by the panel data approach in its original version of the model.[6]

However, there are still important differences between the two methods: First, the synthetic control imposes the restriction that the weights of the countries in the donor pool must add up to one and be non-negative so as not to extrapolate outside the convex hull of the covariates for the untreated units.[7] Second, the panel data approach uses solely data from the outcome variable, whereas the synthetic control is primarily designed to use any covariates that help explain the outcome variable as predictors, and not only pre-treatment values of the outcome variable.

These two differences restrict the way we can conduct the simulation comparison: the second one because the synthetic control method cannot be applied as originally intended, i.e., using covariates other than the pre-treatment values of the outcome variable; the first because, if we force an application where a great deal of extrapolation is needed for a good matching, the synthetic control will not work. This is usually manifested to the researcher through a very large mismatch in the pre-treatment period. That is, the synthetic control does not always have a good match, whereas this does not happen in the case of the panel data approach. We will take both differences into account when carrying out the simulation comparison in Section 4.

---

[4] In our case we use the Akaike Information Criterion with correction for finite sample sizes (AICc). Hsiao *et al.* (2012) conduct the analysis using both AIC and AICc criteria, but in the interest of reducing the number of applications we chose only the latter.

[5] $T_0 - g$ is to allow for at least g degrees of freedom.

[6] Note that in the synthetic control model specified in equation (3) λ_t is usually unrestricted, but if we restrict an element of λ_t to be equal to one, then the corresponding element of μ_j is an individual specific fixed effect.

[7] It is worth mentioning that because the panel data approach is allowed to extrapolate, it mechanically allows for a better fit than the synthetic control.

DOI: 10.1002/jae

3. EMPIRICAL APPLICATION OF BOTH METHODOLOGIES

The initial step in our empirical comparison is to apply both the synthetic control method and the panel data approach first to the political integration of Hong Kong and then to the economic integration, as was done in Hsiao *et al.* (2012), using the same dataset. Thus the application of the panel data approach is a replication of the results obtained by them to ensure the methodology is correctly applied, whereas the synthetic control is a new application. The goal is to compare both methodologies using the same dataset and see whether the results obtained are similar or not. If the estimated treatment effect varies between the two methods, the issue should be explored further.

It is worth noting that the panel data approach allows for traditional inference as long as the outcome variable is stationary, whereas the synthetic control relies on randomization methods for inference. Therefore, for the sake of comparison, we conduct the inference for both methodologies by carrying out the so-called 'placebo studies'.

We use the quarterly real growth rate of 24 countries in the donor pool, computed as the change with respect to the same quarter in the previous year, to predict the quarterly real growth rate of Hong Kong. The dataset, obtained from the supplemental materials provided by Hsiao *et al.* (2012), ranges from 1993:Q1 to 2008:Q1. The pre-treatment period is 1993:Q1–1997:Q2 in the case of the political integration and varies depending on which of the two methods is used in the case of the economic integration. A problem with using this dataset is that, by using real gross domestic product (GDP) growth rate instead of levels as an outcome variable, we might miss part of the effect (see, for example, Gardeazabal and Vega-Bayo, 2016).[8]

The synthetic control for the political integration of Hong Kong is constructed as a convex combination of the countries in the donor pool and the pre-treatment values of GDP (plus their average) as predictors. When applying the synthetic control method, since it can handle all 24 control units, we allow the method to choose the best ones instead of pre-reducing the donor pool as Hsiao *et al.* (2012) does, based on geographical reasons. Table I reports the actual values of the predictors, the sample mean of the countries in the donor pool, the synthetic control and the panel data approach values of the predictors. In general, both the values of the covariates of the synthetic control and the panel data approach are closer to the actual covariates than the sample mean, especially those of the panel data approach. Table II shows the weight that each country in the donor pool has in the synthetic control of the political integration of Hong Kong, as well as the estimated coefficients using the panel data approach, which are also shown in Table A.1 of the online Appendix. Besides showing the coefficients, Table A.1 replicates the results obtained by Hsiao *et al.* (2012) for the political integration. We use their same methodology as well as the same countries in the donor pool, and obtain identical results.

If we compare the columns from Table II that correspond to the political integration, we see that the countries used for the counterfactual are not the same in both cases.[9] This could be because, unlike the synthetic control, Hsiao *et al.* (2012) uses an argument of geographic and economic proximity to reduce the donor pool prior to the estimation; there would otherwise be more explanatory variables than time periods in the estimation.

Both counterfactuals are plotted in Figure 1 and they both replicate properly the economy of Hong Kong. In particular, in Figure 1(a) we can see how the synthetic control does a good job of replicating the economy of Hong Kong prior to the political integration, but not as good as the panel data approach in Figure 1(b). Though they both do a good job in the pre-intervention period, the estimation of the effect appears to be inconsistent between the two methodologies: it seems that it is a lot smaller when

[8] We attempted to reconstruct the levels of GDP using the growth rate and initial values but we could not find the necessary data to do so; we were unable to replicate the growth rates in the dataset using GDP levels from the OECD, probably due to data updating.

[9] Note that the coefficients reported in Table A.1 of the online Appendix correspond to those under the panel coefficients column in Table II.

DOI: 10.1002/jae

Table I. Political and economic integration pre-treatment values

			Political integ.		Economic integ.	
GDP	Actual	Sample mean	Synth.	Panel	Synth.	Panel
Avg. 1993:Q1–1997:Q3	0.049	0.041	0.049	0.056		
Avg. 1993:Q1–2003:Q4	0.031	0.035				0.031
Avg. 2000:Q2–2003:Q4	0.028	0.030			0.028	
1993:Q1	0.062	0.026	0.055	0.055		0.070
1993:Q2	0.059	0.031	0.051	0.061		0.068
1993:Q3	0.058	0.034	0.058	0.065		0.065
1993:Q4	0.062	0.037	0.059	0.061		0.073
1994:Q1	0.079	0.044	0.060	0.074		0.047
1994:Q2	0.068	0.050	0.067	0.066		0.060
1994:Q3	0.046	0.051	0.066	0.049		0.054
1994:Q4	0.052	0.057	0.061	0.049		0.051
1995:Q1	0.037	0.052	0.047	0.043		0.051
1995:Q2	0.029	0.044	0.022	0.024		0.019
1995:Q3	0.012	0.039	0.018	0.017		0.022
1995:Q4	0.015	0.034	0.022	0.014		0.013
1996:Q1	0.025	0.036	0.030	0.019		0.028
1996:Q2	0.036	0.040	0.047	0.042		0.043
1996:Q3	0.047	0.042	0.051	0.054		0.052
1996:Q4	0.059	0.042	0.059	0.057		0.071
1997:Q1	0.058	0.036	0.045	0.062		0.045
1997:Q2	0.072	0.043	0.064	0.065		0.067
1997:Q3	0.061	0.042	0.065	0.080		0.038
1997:Q4	0.014	0.043				0.029
1998:Q1	−0.032	0.035				−0.021
1998:Q2	−0.061	0.010				−0.062
1998:Q3	−0.081	0.004				−0.066
1998:Q4	−0.065	−0.004				−0.075
1999:Q1	−0.029	0.007				−0.020
1999:Q2	0.005	0.029				0.017
1999:Q3	0.039	0.041				0.046
1999:Q4	0.083	0.053				0.069
2000:Q1	0.107	0.072				0.098
2000:Q2	0.075	0.067			0.080	0.082
2000:Q3	0.076	0.060			0.076	0.080
2000:Q4	0.063	0.050			0.063	0.061
2001:Q1	0.027	0.030			0.029	0.026
2001:Q2	0.015	0.019			0.013	0.019
2001:Q3	−0.001	0.008			−0.009	−0.012
2001:Q4	−0.017	0.006			−0.017	−0.027
2002:Q1	−0.010	0.010			−0.007	−0.010
2002:Q2	0.005	0.023			0.018	0.021
2002:Q3	0.028	0.031			0.027	0.020
2002:Q4	0.048	0.034			0.035	0.041
2003:Q1	0.041	0.029			0.034	0.035
2003:Q2	−0.009	0.018			0.010	−0.005
2003:Q3	0.038	0.028			0.032	0.024
2003 Q4	0.047	0.036			0.044	0.037

Note: The table reflects the corresponding pre-treatment period for each of the applications.

using the synthetic control method. The only similarity between the two methodologies is that the effect seems only to last until 1999:Q2 in both cases.

In order to assess the statistical significance of our estimates, we carry out a placebo study by applying the same procedure that was used to estimate the GDP gap of Hong Kong in the political integration to all the countries in the 'donor pool' as done by Abadie *et al.* (2010). That is, we apply the synthetic control procedure to all the countries specified in the part of Table II that corresponds to the political integration, and the panel data method to all the countries in the donor pool, i.e., those specified as a note in Table A.1. Despite using an inference procedure different from the one Hsiao *et al.* (2012) use, the conclusions are the same, as we can see from the results of these placebo studies that are plotted in Figure 2. Note that even though the placebo is carried out for all countries in

DOI: 10.1002/jae

Table II. Political and economic integration weights / coefficients

Country	Political integration		Economic integration	
Country	Synth. weights	Panel coeffs	Synth. weights	Panel coeffs
(Intercept)		0.026		−0.002
Australia	0.155		0.000	0.000
Austria	0.000		0.000	−1.012
Canada	0.000		0.000	0.000
Denmark	0.000		0.029	0.000
Finland	0.000		0.000	0.000
France	0.000		0.000	0.000
Germany	0.000		0.000	0.000
Italy	0.000		0.000	−0.318
Japan	0.000	−0.676	0.225	0.000
Korea	0.000	−0.432	0.000	0.345
Mexico	0.229		0.133	0.313
Netherlands	0.000		0.000	0.000
New Zealand	0.265		0.000	0.000
Norway	0.000		0.240	0.322
Switzerland	0.000		0.000	0.000
UK	0.000		0.000	0.000
USA	0.000	0.486	0.000	0.000
Singapore	0.169	0.000	0.047	0.185
Philippines	0.000	0.000	0.017	0.000
Indonesia	0.182	0.000	0.000	0.000
Malaysia	0.000	0.000	0.011	0.000
Thailand	0.000	0.000	0.043	0.000
Taiwan	0.000	0.793	0.170	0.000
China	0.000	0.000	0.084	0.000

Note: Countries that have a weight or coefficient equal to zero are included in the subset of possible controls, whereas those that do not have any weight or coefficient have not been considered as possible controls.

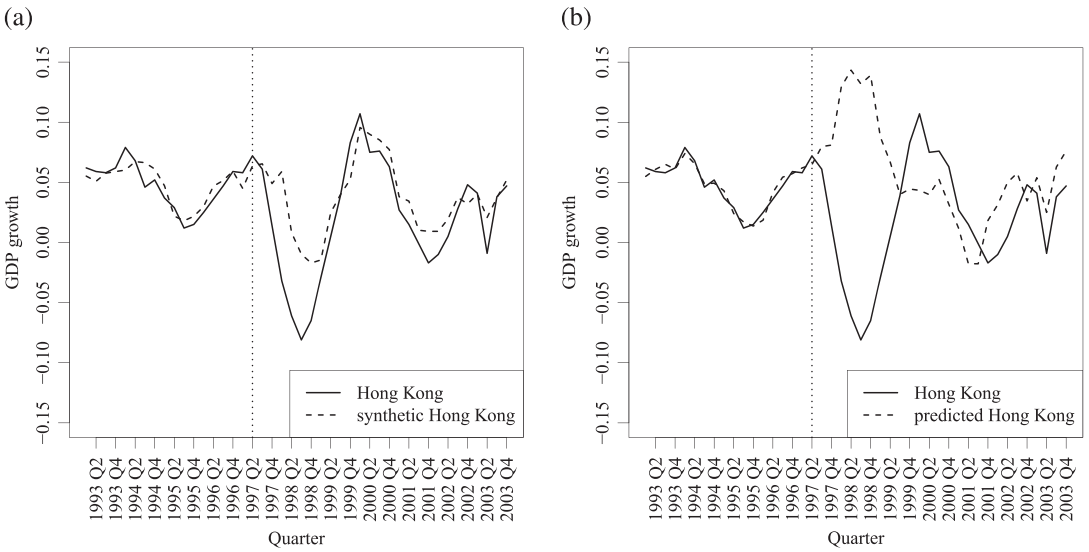

Figure 1. Actual real GDP versus synthetic and predicted real GDP for the political integration of Hong Kong: (a) synthetic control; (b) panel data approach

DOI: 10.1002/jae

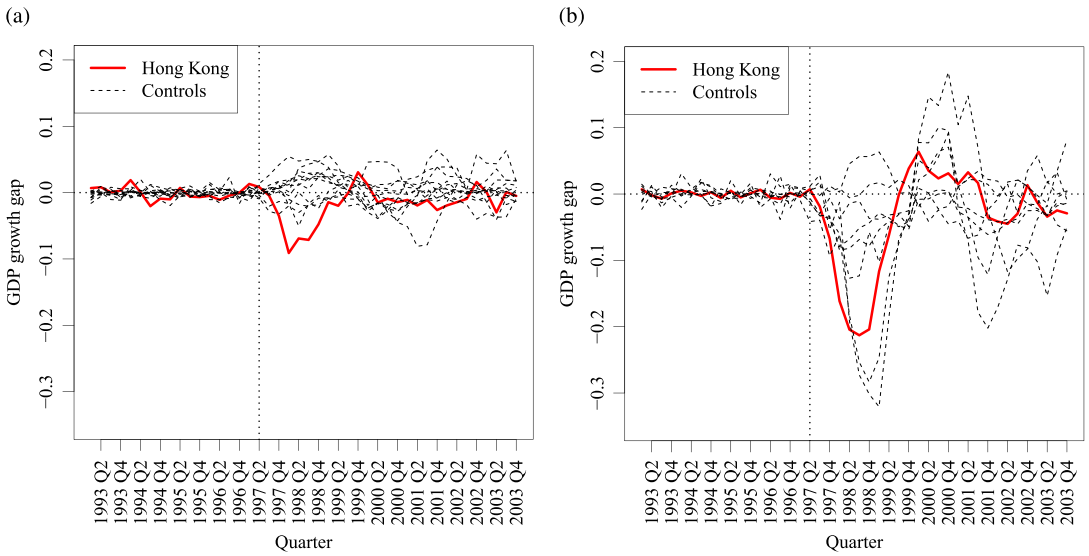

Figure 2. Placebos for the political integration of Hong Kong: (a) synthetic control; (b) panel data approach. [Colour figure can be viewed at wileyonlinelibrary.com]

the donor pool (24 in the case of the synthetic control and 10 in the panel data approach) we only keep those that have the pretreatment MSE lower than twice the pretreatment MSE of Hong Kong, which reduces the placebos to 13 in the case of the synthetic control and eight in the panel data approach.[10] According to Figure 2, the effect of political integration is not significant when using the panel data approach, because the predicted Hong Kong only has the third largest effect out of eight controls. However, the effect is significant under the synthetic control at the 7.69% level: the synthetic Hong Kong has the largest gap out of 13 other units. This discrepancy poses an issue when attempting to estimate the effect of economic integration a few years later. Hsiao *et al.* (2012) argue that, because the effect of the political integration is non-significant, they can pool all the available data in order to estimate the effect of the economic integration, prolonging the pre-treatment period of this second estimation. This argument is not valid for the synthetic control, given our results.

We therefore shorten the pre-treatment period when estimating the effect of the economic integration using the synthetic control (2000:Q2–2003:Q4). For the panel data approach, we keep the pretreatment period to 1993:Q1–2003:Q4, as in the original paper by Hsiao *et al.* (2012). Once again, we construct the synthetic control for the economic integration of Hong Kong as a convex combination of all countries in the donor pool and the pre-treatment values of GDP (plus their average) as predictors.

Table I shows the actual value of the predictors, together with the sample average, the estimated synthetic control and the estimated predictors by the panel data approach. As in the case of political integration, both the values of the synthetic control and those of the panel data approach are, in general, closer to the actual values of the covariates than the sample average. The right half of Table II reports the weights of the donor pool countries in the synthetic control, as well as the estimated coefficients for the panel data approach methodology, for the economic integration of Hong Kong. We can see that the countries used to construct the counterfactuals are different in both methods. This could be for several reasons: the fact that we are estimating over two different pre-treatment period, or because the panel

[10] The pretreatment MSE is measured as the average over the pre-treatment periods of the squared difference between the actual and the counterfactual/predicted outcome.

DOI: 10.1002/jae

data approach is extrapolating outside of the common support. When plotting the counterfactuals in Figure 3, however, we observe that both replicate the economy of Hong Kong in the pre-economic integration period nicely, albeit the panel data approach replicates it slightly better, as in the case of the political intervention. The size of the effect also seems to be smaller according to the synthetic control methodology. Figure 4 shows the results from the placebo study of the economic integration using both methods.

The estimated effect of the economic integration seems to be an outlier in Figure 4, in both parts (a) and (b). Specifically, Figure 4(a) shows that the synthetic Hong Kong is the outlier of 15 placebos but

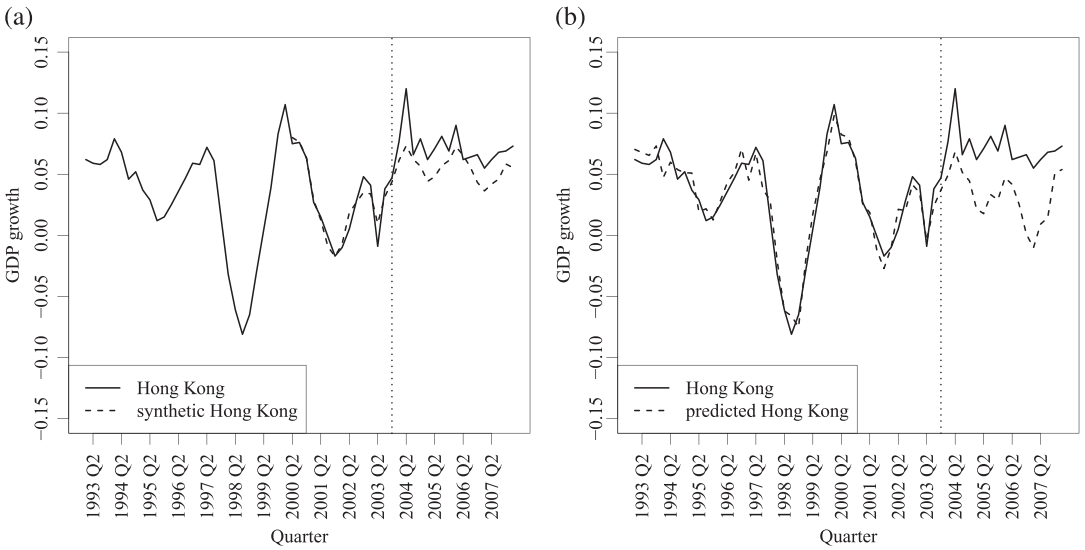

Figure 3. Actual real GDP versus synthetic and predicted real GDP for the economic integration of Hong Kong: (a) synthetic control; (b) panel data approach

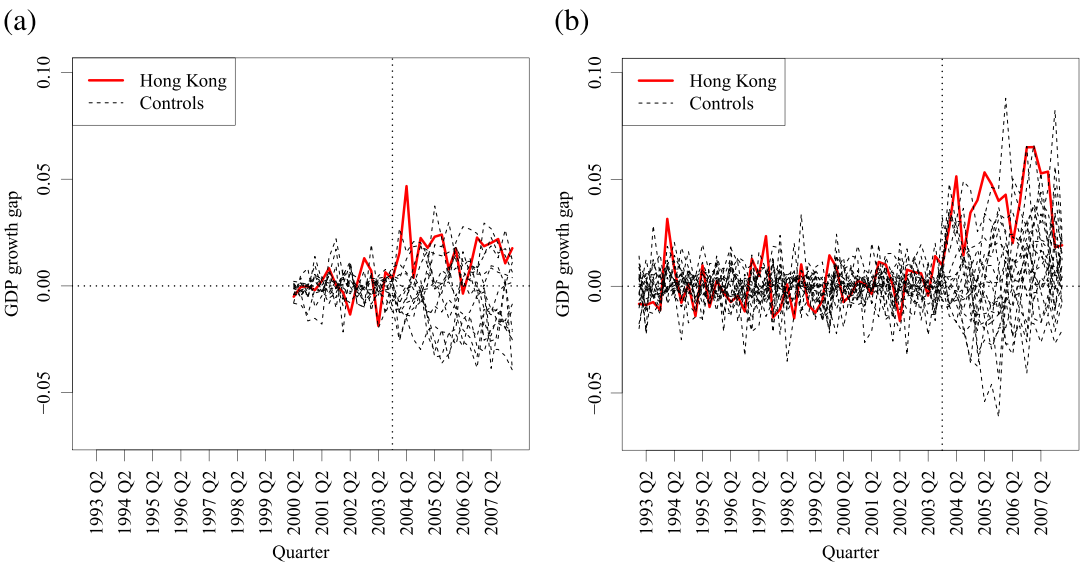

Figure 4. Placebos for the economic integration of Hong Kong: (a) synthetic control; (b) panel data approach. [Colour figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/jae

only for some of the time periods, so its significance is unclear: it seems that the effect is significant using both methods, but not for a few of the time periods (2004:Q3, 2006:Q2, and 2007:Q4). The size of the effect appears to be smaller when estimating it using the synthetic control, and the significance of the effect of the economic integration varies depending on which method is used.

After conducting this baseline application, we also consider the effect that changes in the donor pool has on the estimated treatment effect. We repeat the Hong Kong application but choosing the subset of possible controls in four different ways.

- **Case A** The set of possible controls in the panel data approach includes only those that have a positive weight in the synthetic control.

The set of possible controls is reduced to the subset that has a positive weight in the synthetic control. The results of this estimation are shown in Table A.3 of the online Appendix, while the predicted GDP and placebo studies are depicted in Figure 5. It can be observed that although the predicted path follows a similar one to the synthetic one shown in Figure 1(a), the set of controls in Table A.3 is very different from that in Table II ('Panel coefficients' column). Furthermore, the predicted path is more similar to the synthetic one than that obtained in the baseline application, as depicted in Figure 1(b); and the placebo study now shows that the treatment effect could be significant according to the panel data approach at a 10% level, like the synthetic control methodology as shown in Figure 2(a). Similar results are observed in Figure 6 when repeating the application for the economic integration, for which the controls with positive weights in Table II are used.

- **Case B** The set of possible controls in the synthetic control includes only those that are included in the model in the panel data approach.

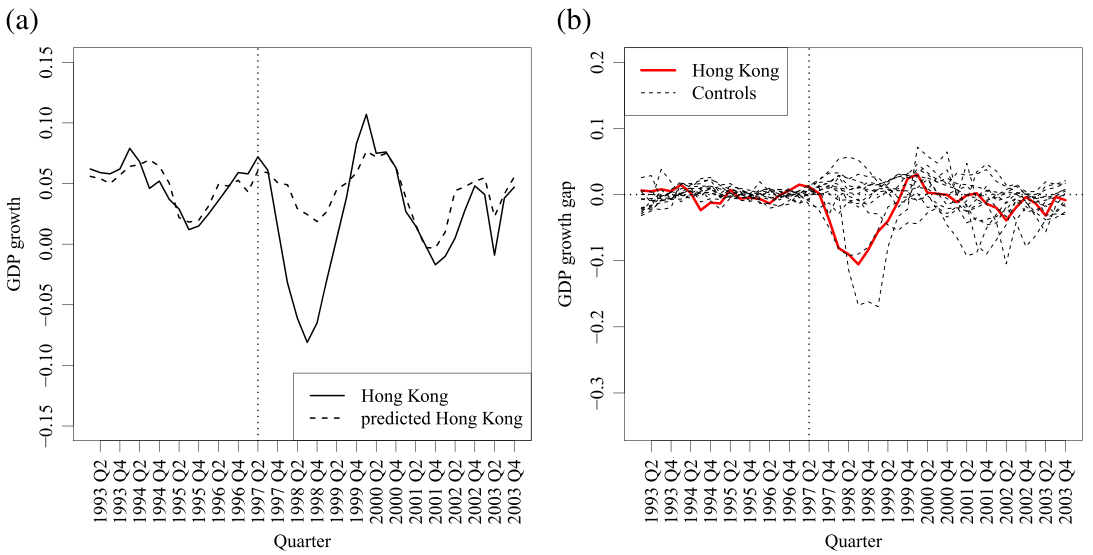

Figure 5. Panel data approach for the political integration of Hong Kong when the subset of possible controls is obtained from $w > 0$ in Synth. Placebos are calculated for all 24 countries in the original donor pool: (a) actual and predicted real GDP; (b) placebos. [Colour figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/jae

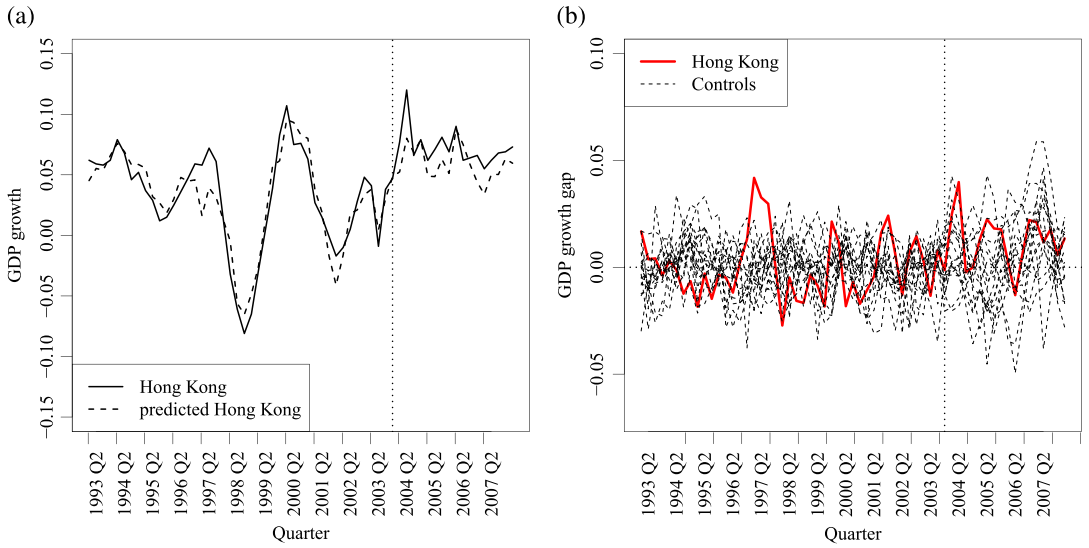

Figure 6. Panel data approach for the economic integration of Hong Kong when the subset of possible controls is obtained from $w > 0$ in Synth. Placebos are calculated for all 24 countries in the original donor pool: (a) actual and predicted real GDP; (b) placebos. [Colour figure can be viewed at wileyonlinelibrary.com]

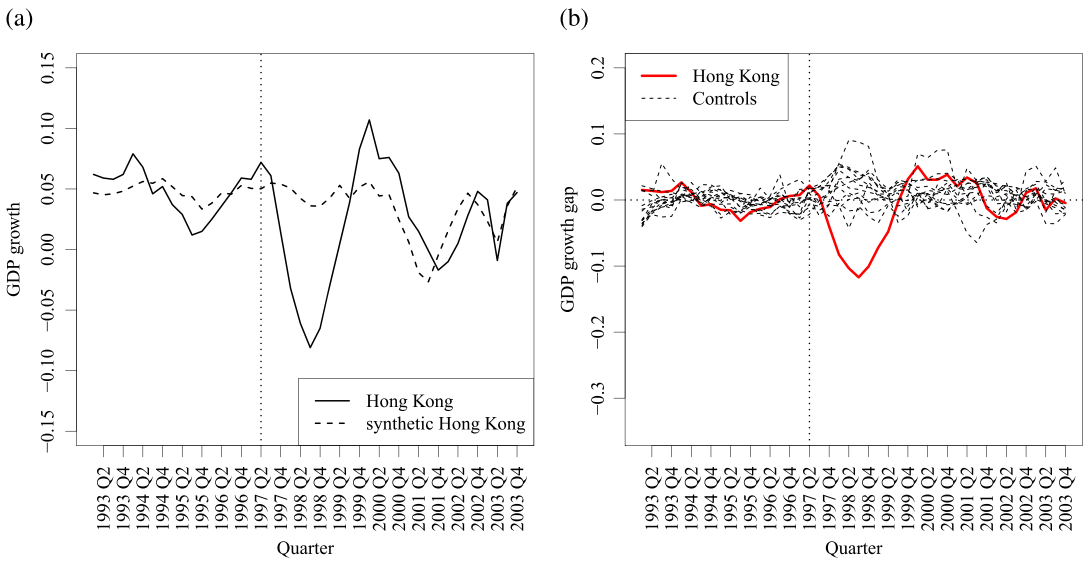

Figure 7. Synthetic control for the political integration of Hong Kong when the subset of possible controls is obtained from the statistically significant controls in the panel data approach. Placebos are calculated for all 24 countries in the original donor pool: (a) actual and synthetic real GDP; (b) placebos. [Colour figure can be viewed at wileyonlinelibrary.com]

We now reverse this application to see how the synthetic control responds to changes in the donor pool; i.e., we apply the synthetic control method using as controls those that are statistically significant when estimating the treatment effect with the panel data approach. For the political integration, we use

DOI: 10.1002/jae

the countries shown in Table A.1 as the donor pool and apply the synthetic control methodology to estimate the treatment effect. We obtain the synthetic path and placebos depicted in Figure 7. The low number of controls in the donor pool poses a problem: the error in the pre-treatment period is somewhat larger than that obtained in the baseline application, although the estimated path appears to be quite similar. However, when estimating the economic integration effect—using the countries from the coefficient column in Table II as the donor pool and applying the synthetic control methodology—we see that this is not the case. Despite the differences in country weights, the synthetic path and placebo study depicted in Figure 8 is strikingly similar to that obtained in the baseline application, shown in Figures 3(a) and 4(a). Thus the fit in the pre-treatment period appears to be worse, but the estimated paths are similar to those obtained in the initial application, so it appears the synthetic control method is more robust to changes in the set of controls.

- **Case C** The set of possible controls in the panel data approach is not preselected by the researcher's criteria.

The results of applying the modified modeling strategy from Section 2.2 to the political integration are shown in Table A.5. The predicted path, together with the placebo study, is shown in Figure 9. We can see that, when comparing these results to those obtained in the baseline application, Table A.1 and Figures 1(b) and 2(b), they are qualitatively similar although quantitatively different.

- **Case D** The set of possible controls in the synthetic control is preselected by the researcher's criteria.

We check how the synthetic control fares compared to the panel data approach when we pre-subset the number of countries in the donor pool on a geographic and economic proximity basis, as Hsiao *et al.* (2012) do for the political integration of Hong Kong. Pre-restricting the donor pool to units

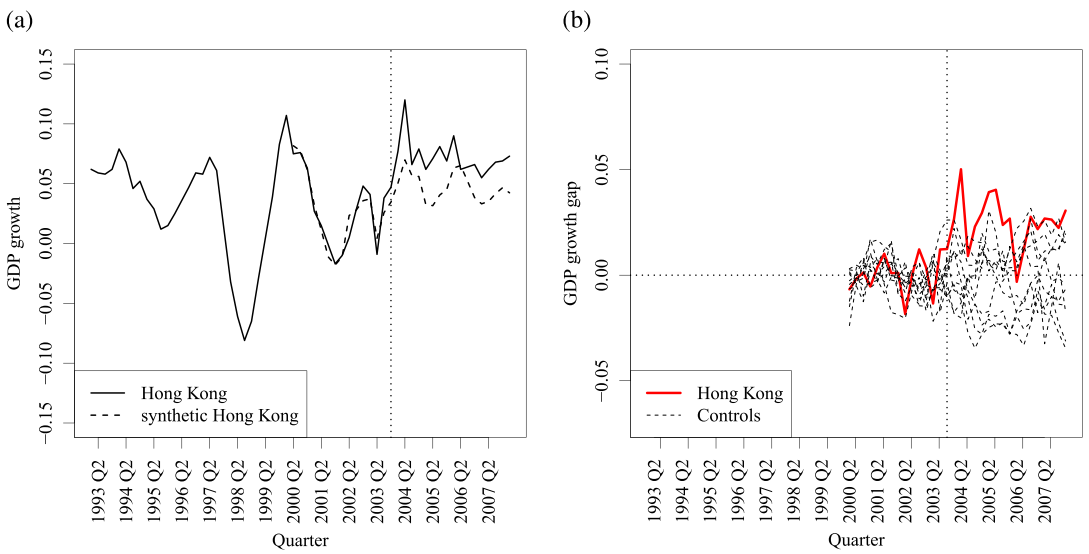

Figure 8. Synthetic control for the economic integration of Hong Kong when the subset of possible controls is obtained from the statistically significant controls in the panel data approach. Placebos are calculated for all 24 countries in the original donor pool: (a) actual and synthetic real GDP; (b) placebos. [Colour figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1002/jae

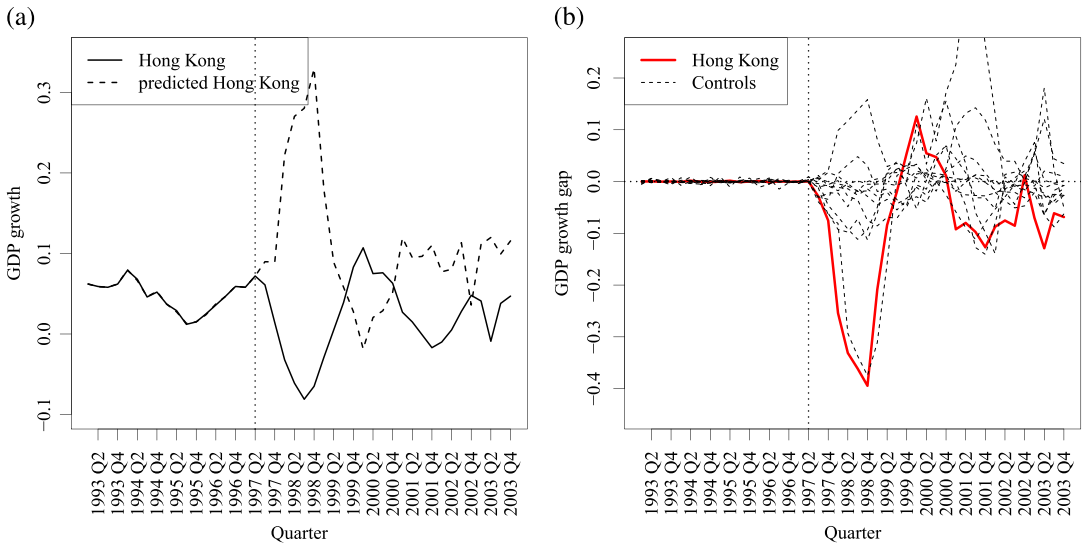

Figure 9. Panel data approach for the political integration of Hong Kong when the subset of possible controls is not pre-selected on a geographical basis: (a) actual and predicted real GDP; (b) placebos. [Colour figure can be viewed at wileyonlinelibrary.com]

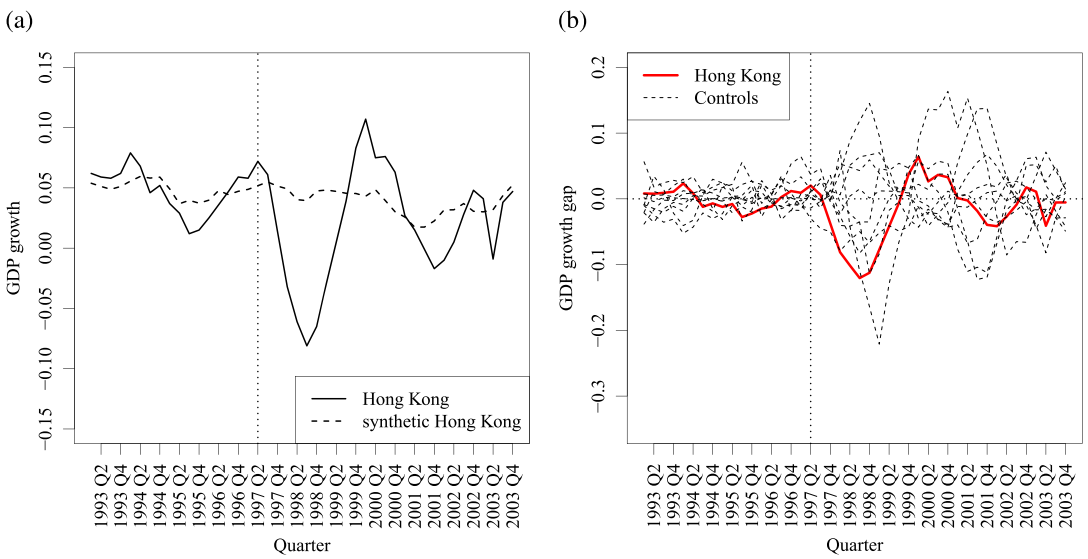

Figure 10. Synthetic control for the political integration of Hong Kong when the subset of possible controls is selected on a geographical basis: (a) actual and predicted real GDP; (b) placebos. [Colour figure can be viewed at wileyonlinelibrary.com]

'similar' to the treatment unit is a widespread practice in the synthetic control literature as well. Both Abadie and Gardeazabal (2003) in their Basque Country application and Abadie *et al.* (2015) in the West Germany reunification case study do this. The results are plotted in Figure 10; we can see how the synthetic control's results are similar to those shown in Figure 7, when the set of possible controls in the synthetic control includes only those that are included in the model in the panel data approach.

DOI: 10.1002/jae

Table III. Average annual treatment effect of the different applications

	Application	Case	Avg. annual tr. effect (Post-treatment period)	Avg. annual tr. effect (First 10Q post-treatment)
Political integration				
Panel	Baseline		-0.0396	-0.0861
	Positive weights	A	-0.0234	-0.0407
	Not restricted	C	-0.0933	-0.1433
Synth.	Baseline		-0.0183	-0.0317
	Significant	B	-0.0143	-0.0435
	Restricted	D	-0.0201	-0.0532
Economic integration				
Panel	Baseline		0.0403	0.0374
	Positive weights	A	0.0117	
Synth.	Baseline		0.0185	0.0189
	Significant	B	0.0255	0.0252

From these four reapplications, we can conclude that the panel data approach is more sensitive to changes in the donor pool than the synthetic control: the estimation of the effect varies more in the case of the panel data approach than with the synthetic control methodology when we change the set of controls in the donor pool. This is shown numerically in Table III: the average annual estimated treatment effect is more volatile in the case of the panel data approach than the synthetic control, suggesting a greater sensitivity to changes in the donor pool.

A second empirical application, detailed in the online Appendix, estimates the treatment effect of the 1965 Arab–Israeli conflict as well as the Sierra Leone civil war on GDP per capita using both methods. In this case, the two methods obtain significantly negative effects for the two armed conflict episodes; however, the estimated size of the effect varies depending on the method used. The evidence from these applications together with the applications to Hong Kong suggests that there are indisputable differences in the estimated treatment effect of the two methods. This is especially obvious when looking at the actual versus predicted paths.

The section that follows attempts to study further the discrepancies observed in the empirical applications using a simulation experiment.

4. SIMULATION EXPERIMENT

Because of the impossibility of observing data under the intervention and in the absence of intervention at the same time, we carry out a simulation that enables us to know the actual value of the (simulated) outcome under no treatment and, therefore, which one of the two methods is more accurate in its estimation. In order to conduct the simulation we use a data generating process (DGP) following the general model specified in equation (8). Starting from this general DGP we will conduct two different simulations.

We assume that the outcome under no treatment is generated as the sum of an individual effect, a covariate, four factor-loaded common trends and cycles, plus an error term:

$$Y_{jt}^{0} = \gamma_j + \theta_t z_j + \sum_{k=1}^{4} \mu_{kj} \lambda_{kt} + \xi_{jt} \tag{9}$$

for $j = 1, 2, \ldots, J+1$ and where $\xi_{jt} \sim \mathcal{N}(0, 0.25)$.[11]

[11] More precise details on the DGP of the simulation are included in the online Appendix.

DOI: 10.1002/jae

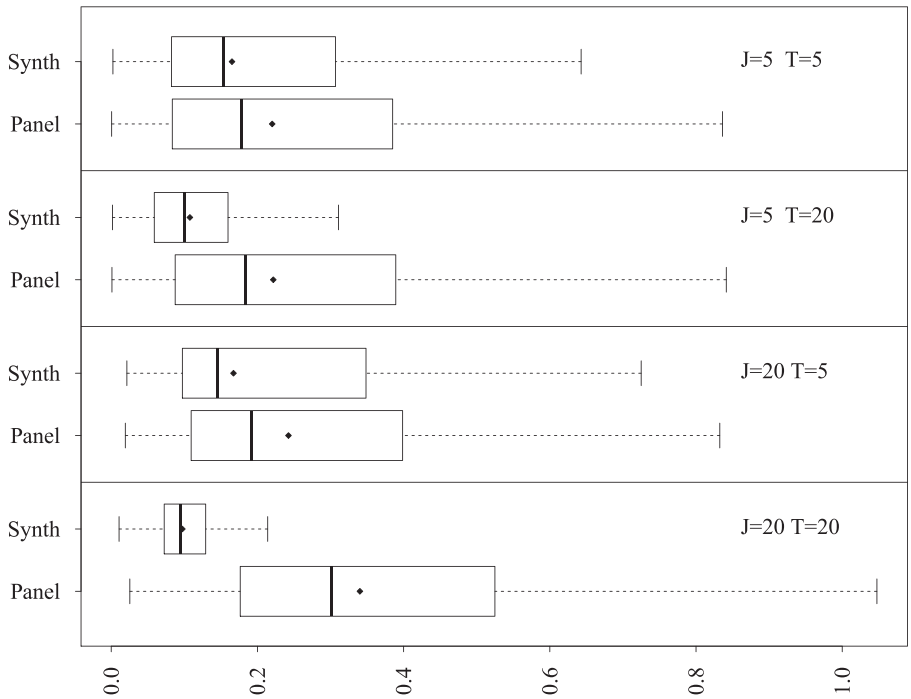

Figure 11. Box plots of the MSE of the simulation with common factors only. The dot represents the mean

For the first simulation, we limit the comparison to the cases that satisfy both methods' restrictions: we only use pre-treatment values as covariates for the estimation, and we only keep the simulation replications that have a good match in the synthetic control. In addition, we restrict $\theta_t = \gamma_j = 0$ so that the DGP will include only the common factors.

The data are generated 10,000 times (replications) according to equation (9) with $\theta_t = \gamma_j = 0$ for two cases of number of controls $J = 5, 20$ and two cases of pre-/post-treatment values $T_0 = 5, 20$.[12] For each replication and for each of the four (J, T_0) combinations, the outcome under no treatment Y_{1t}^0 is estimated using both the synthetic control method and the panel data approach for the four cases. These estimations are done using solely pre-treatment values of the outcome variable and with the modified modeling strategy specified in Section 2.2, taking $g = 2$ in order to allow for at least three degrees of freedom. Although more degrees might be optimal, we are constrained by the number of periods. Once the estimations are carried out, given the synthetic control's restriction, we only keep the replications that have a pre-treatment MAE smaller than 20% of the mean values in the outcome variable; imposing this threshold reduces the replications that we actually keep for the reported results between 5% and 10%.

The main results from this first simulation are summarized in Figures 11, 12, and 13. The figures show, respectively, box plots of the replication for the post-treatment MSE, MAPE, and mean error of the estimated outcome under no treatment with both methods.

Figure 11 shows that the MSE incurred is quite comparable for both methods when $T_0 = 5$, albeit the average post-treatment MSE is smaller for the synthetic control in all four cases. Furthermore,

[12] We use an equal number of pre- and post-treatment periods. The number of time periods and controls are chosen arbitrarily, but taking into account that it is unlikely to have less than five or more than 20 pre- or post-treatment periods nor possible controls, at least in a macroeconomic setting; and that we wanted to consider both a small and a large case.

DOI: 10.1002/jae

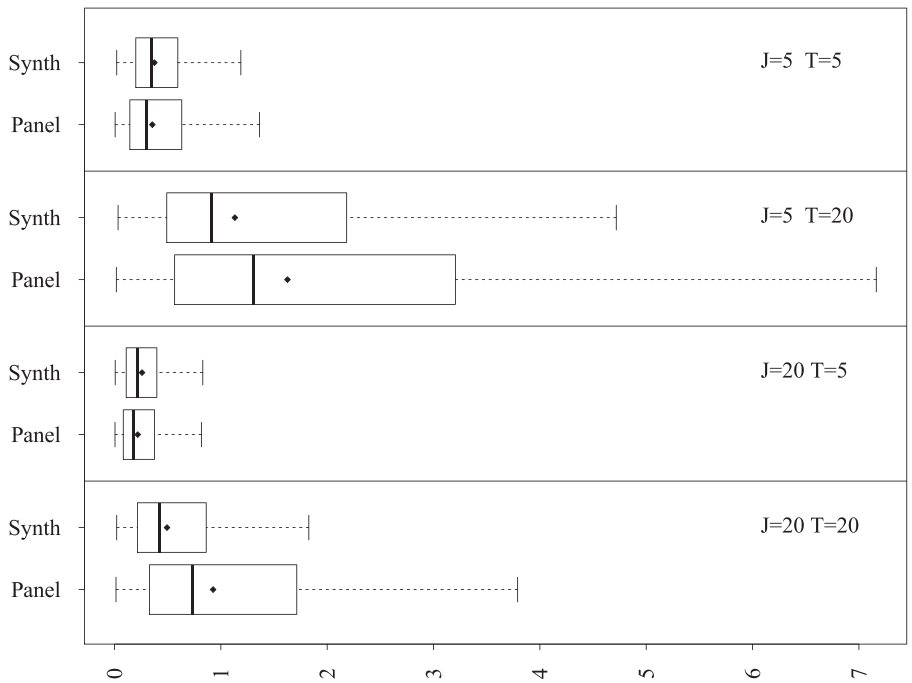

Figure 12. Box plots of the MAPE of the simulation with common factors only. The dot represents the mean

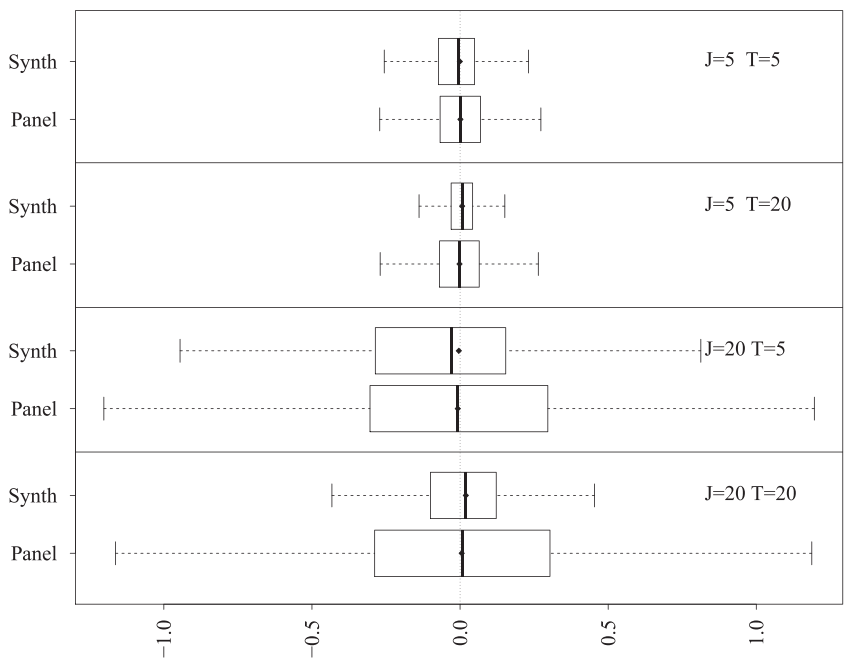

Figure 13. Box plot of the mean error of the simulation with common factors only. The dot represents the mean

DOI: 10.1002/jae

the range of the synthetic control's MSE is much smaller when $T_0 = 20$. Figure 12 paints a similar picture: when $T_0 = 5$, it does not appear to matter, in terms of MSE or MAPE, which one of the two methods the researcher uses. The synthetic control method does better than the panel data approach in terms of MSE and MAPE for the other cases, and even more so for cases with larger time series.

If we take a look at Figure 13, one can observe how, on average, neither of the two methods appears to have a positive or negative mean error. An important point of consideration is that, although in the cases when $T_0 = 20$ the synthetic control appears to have a slightly larger mean error, the (interquartile) range of the mean error is much smaller for the synthetic control. If we focus on the cases where $T_0 = 5$, in terms of the mean error, the synthetic control has a smaller range when $T_0 = 5$ and $J = 20$; however, the median mean error is larger in absolute value than that of the panel data approach. On the other hand, when $T_0 = 5$ and $J = 5$, it appears, as it does for the post-treatment MSE and MAPE, that both methods offer very similar mean errors. This suggests that the researcher could use one or the other, depending, for example, on whether there are data on additional covariates or not, or whether the synthetic control's match is a good one. Both could also be used to check the robustness of the results obtained. Increasing the number of time periods helps the synthetic control method reduce the range of the mean error, but less so for the panel data approach.

Figure 14 shows the mean error of both methods for the rejected cases (i.e., the cases when the synthetic control has a bad match). In this case, the synthetic control seems to have, on average, a positive mean error; that is, it is clearly overestimating the outcome under no treatment. The panel data approach, however, does not have a systematic mean error, although the range is still bigger. Note that these results should be taken with caution, given the low number of rejected replications. Taking all four figures into account, it appears that the synthetic control offers more consistent results than the panel data approach, *whenever there is a good enough match*.

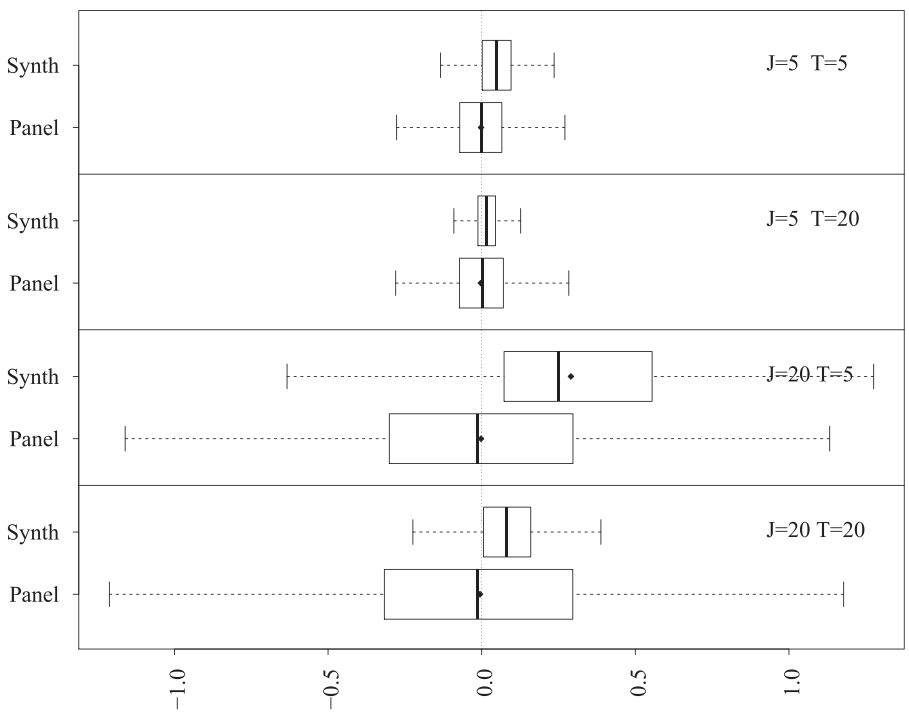

Figure 14. Box plot of the mean error of the simulation for the rejected replications

DOI: 10.1002/jae

For the second simulation, we lift the restriction on $\theta_t = \gamma_j = 0$ and allow for one covariate and an individual effect. However, we restrict the number of common factors to a single common cycle factor. The supplementary results from this second simulation are summarized in Figures A.5, A.6, and A.7 of the online Appendix. All three measures (MSE, MAPE, and mean error) improve when either the covariate or the individual effect is introduced into the DGP, and even more so when they are both included. This is true not only for the synthetic control, as one would expect, but also for the panel data approach. We believe this to be for two main reasons: one, the common cycle factor generated as an AR(1) with $\rho = 0.5$ instead of a parameter closer to unity; and two, simply adding more information or structure to the model allows for a better estimation in both cases.

5. CONCLUSIONS

This paper presents an overview of the synthetic control and the panel data approach, and compares them via 'real data' cases and a simulation. It introduces a minor modification to the method described by Hsiao *et al.* (2012), allowing the potential donor pool to be narrowed automatically instead of by the researcher. We also suggest the possibility of applying this method in levels by introducing inference through placebo studies.

The empirical applications suggest that there is a variation in the estimated treatment effect depending on which method is used. This variation can be either quantitative (as in the Hong Kong applications) or qualitative (as in the armed conflict applications from the online Appendix). The synthetic control's estimations seem to be more robust to changes in the donor pool than the panel data approach, which is more sensitive.

According to the conducted simulations using a general encompassing model, neither of the two methods appears to have a systematically positive or negative mean error; and on average, they both show acceptable results. However, the large (interquartile) range of the MSE, MAPE and mean error obtained for some of the cases in the simulations appears to indicate that the researcher must apply the methods with caution. The synthetic control is especially useful when we have data on additional time periods, as well as on covariate(s).

The simulations also show that it is important to pay close attention to whether there is a non-match in the synthetic control.[13] If this is the case, using the synthetic control is not feasible. The panel data approach can still do the job; however, if the panel data approach is in fact extrapolating we are of course estimating outside of the common support and incurring in extrapolation bias, which might not be desirable. When the condition of having a good match is met, the synthetic control has, in general, more consistent results, in terms of a smaller range for post-treatment MSE, MAPE, and mean error.

ACKNOWLEDGEMENTS

We would like to thank thank Alberto Abadie, two anonymous referees, and the Editor Herman van Dijk for all their thoughtful suggestions; also the Institute for Quantitative Social Sciences at Harvard University, seminar participants at the University of the Basque Country and at the 2014 ZEW Summer Workshop for Young Economists in Mannheim for their helpful comments. Javier Gardeazabal acknowledges financial support from the Spanish Ministerio de Economía y Competitividad ECO2012-35820 and ECO2015-64467-R and the Basque Government IT783-13. Ainhoa Vega-Bayo acknowledges financial support from the Basque Government RBFI-2012-87.

[13] This is usually related to extrapolation; see Abadie *et al.* (2015) for a method to formally detect it.

DOI: 10.1002/jae

REFERENCES

Abadie A, Gardeazabal J. 2003. The economic costs of conflict: a case study of the Basque Country. *American Economic Review* **93**: 113–132.

Abadie A, Diamond A, Hainmueller J. 2010. Synthetic control methods for comparative case studies: estimating the effect of California?s tobacco control program. *American Statistical Association* **105**: 493–505.

Abadie A, Diamond A, Hainmueller J. 2015. Comparative politics and the synthetic control method. *American Journal of Political Science* **59**(2): 495–510.

Barone G, Mocetti S. 2014. Natural disasters, growth and institutions: a tale of two earthquakes. *Journal of Urban Economics* **84**: 52–66.

Dehejia RH, Wahba S. 1999. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association* **94**(448): 1053–1062.

Fujiki H, Hsiao C. 2015. Disentangling the effects of multiple treatments: measuring the net economic impact of the 1995 great Hanshin-Awaji earthquake. *Journal of Econometrics* **186**(1): 66–73.

Gardeazabal J, Vega-Bayo A. 2016. Measuring the economic costs of armed conflict. *University of the Basque Country, Working Papers Series*.

Hsiao C, Steve Ching H, Ki Wan S. 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**(5): 705–740.

LaLonde RJ. 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* **76**(4): 604–620.

Rubin DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* **66**(5): 688–701.

Vega-Bayo A. 2016. An R package for the panel approach method for program evaluation: pampe. *R Journal* **7**(2): 105–121.

DOI: 10.1002/jae