Combining Matching and Synthetic Control to Trade off Biases from Extrapolation and Interpolation

Maxwell Kellogg[*] Magne Mogstad[†]

Guillaume A. Pouliot[‡] Alexander Torgovitsky[§]

September 23, 2020

Abstract

The synthetic control method is widely used in comparative case studies to adjust for differences in pre-treatment characteristics. A major attraction of the method is that it limits extrapolation bias that can occur when untreated units with different pre-treatment characteristics are combined using a traditional adjustment, such as a linear regression. Instead, the SC estimator is susceptible to interpolation bias because it uses a convex weighted average of the untreated units to create a synthetic untreated unit with pre-treatment characteristics similar to those of the treated unit. More traditional matching estimators exhibit the opposite behavior: they limit interpolation bias at the potential expense of extrapolation bias. We propose combining the matching and synthetic control estimators through model averaging to create an estimator called MASC. We show how to use a rolling-origin cross-validation procedure to train the MASC to resolve trade-offs between interpolation and extrapolation

[*] Kenneth C. Griffin Department of Economics, University of Chicago. Research supported by funding from the National Institute on Aging (T32AG000243).

[†] Kenneth C. Griffin Department of Economics, University of Chicago Statistics Norway, NBER.

[‡] Harris School of Public Policy, University of Chicago.

[§] Kenneth C. Griffin Department of Economics, University of Chicago. Research supported by National Science Foundation grant SES-1846832.

bias. We use a series of empirically-based placebo and Monte Carlo simulations to shed light on when the SC, matching, MASC and penalized SC estimators do (and do not) perform well. Then, we use the MASC re-examine the economic costs of conflicts and find evidence of larger effects than with SC.

Keywords: synthetic control, causal inference, program evaluation, comparative case studies, cross-validation, forecasting, model averaging

1 Introduction

Estimating the causal effect of an intervention (treatment) is a common task across the social sciences. Longitudinal approaches based on difference-in-differences have long been used for this task. However, the credibility of these methods can be strained when the pre-treatment trends or characteristics of the untreated units differ significantly from those of the treated units. This occurs frequently, especially when the units are large aggregates, such as countries or states. For these types of comparative case studies, the synthetic control (SC) method of Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015) provides an alluring alternative.

The motivation of the SC method is to limit the extrapolation bias that can occur when units with different pre-treatment characteristics are combined using a traditional adjustment, such as a linear regression. Instead, the SC estimator *interpolates* by using a convex weighted average of the untreated units to create a synthetic untreated unit with pre-treatment characteristics similar to those of the treated unit. As observed by Abadie et al. (2010, pp. 495–496), this makes the SC estimator susceptible to interpolation bias. In Section 2.2, we formalize this observation by showing that SC will only avoid such bias if the conditional mean of the outcome is linear in pre-treatment characteristics.

As Abadie and L'Hour (2019) observe, the SC estimator belongs to a large class of weighting estimators with weights based on pre-treatment characteristics. Within this class, its vulnerability to interpolation bias is unusual. Most other commonly used estimators, such as nearest neighbor matching, suffer from the opposite drawback of potentially extrapolating too much when suitable untreated units are unavailable. That is, SC controls extrapolation bias while being susceptible to interpolation bias, whereas matching has the opposite properties. This complementarity suggests that an estimator that adaptively combines the SC and matching estimators may be particularly attractive.

In Section 2.5, we propose matching and synthetic control (or MASC) as a model averaging estimator that combines the standard SC and matching estimators. We show how averaging these two purposefully-chosen estimators defends against the weaknesses of both while preserving their strengths. In Section 3, we show how to choose the weight assigned to each estimator in the MASC through cross-validation, as in Wolpert (1992), Breiman (1996) and Hansen and Racine (2012). Our cross-validation criterion uses an evaluation concept referred to as rolling-origin recalibration in the forecasting literature (e.g. Tashman, 2000). One attractive feature of the MASC estimator is that its cross-validated weight can be solved for in closed-form, making it only marginally more difficult to implement than the usual SC estimator. An R package for implementing the MASC is available at https://github.com/maxkllgg/masc.

In Sections 4 and 5, we provide evidence that the MASC estimator performs well in practice. In Section 4, we conduct a placebo study using the data on Spanish terrorism analyzed by Abadie and Gardeazabal (2003). This allows us to evaluate the performance of the matching, SC, penalized SC (Abadie and L'Hour, 2019), and MASC estimators by how well they predict a zero treatment effect for untreated units. We find evidence that MASC performs better than the other three alternatives in this application. This is because MASC is able to adapt to cases where either SC or matching would do well. Then, in Section 5, we use the same data to re-estimate the effect of terrorism on the GDP of the Basque Country. We find larger effect estimates with MASC than with the other three estimators.

Our paper is related to a growing literature on SC (see Abadie, 2019, for a recent survey). The closest work to ours is the paper by Abadie and L'Hour (2019), who propose the penalized SC estimator. The penalized SC and MASC estimators are different, but related in that both assign weights to untreated units while taking into consideration their distance from the treated unit in terms of pre-treatment characteristics. In Section 2.6, we show that the penalized SC estimator is the solution to a constrained version of the problem implicitly solved by the MASC. Thus, the MASC represents a more flexible model than the penalized SC. While this does not necessarily mean it will perform better in practice, our empirical results in Sections 4 and 5 show that, at least for the Spanish data, the penalized SC estimator usually coincides with the standard SC estimator, suggesting that the extra flexibility of MASC can be useful.

Also closely related to our work are the papers by Athey et al. (2019) and Viviano and Bradic (2019), who also consider the benefits of model averaging in the context of comparative case studies. The former authors combine several of the regularized SC and matrix completion estimators developed in Doudchenko and Imbens (2016) and Athey et al. (2018), while the latter authors combine a large number of estimators from the machine learning literature. MASC differs from the estimators in these papers both in details and intent. The purpose of MASC is to directly guard against the types of interpolation biases that can occur with SC, and the extrapolation bias that can occur with matching, by adaptively blending them together. Like Athey et al. (2019), we also find that model averaging tends to work quite well, in concordance with a recurring finding of the economic forecasting literature (see e.g. Stock and Watson, 2004, 2006). A contrast with Athey et al. (2019), Viviano and Bradic (2019), and much of the forecasting literature, is that the estimators we average are purposefully chosen to be complementary. This is exactly the case when data-driven model averaging should be especially beneficial, see, for example, Breiman (1996) or Elliot (2011).

2 Synthetic Control and Matching

2.1 Setup

Suppose that we observe a scalar outcome, Y_{it}, for cross-sectional units denoted by i at times $t = 1, \ldots, T$, as well as a time-invariant binary treatment group indicator, $D_i \in \{0, 1\}$. Units in the treated group become treated at an event date, $t^\star$, so that treatment status in time t is given by $D_{it} \equiv D_i \mathbb{1}[t \geq t^\star]$. Associated with the outcome and treatment are potential outcomes $Y_{it}(0)$ and $Y_{it}(1)$, which are related to the observed outcome via $Y_{it} = D_{it} Y_{it}(1) + (1 - D_{it}) Y_{it}(0)$. Our goal is to estimate the average treatment on the treated (ATT),

$$\mathrm{ATT}_t \equiv \mathbb{E}[Y_{it}(1) - Y_{it}(0) | D_i = 1] = \mathbb{E}[Y_{it} | D_i = 1] - \mathbb{E}[Y_{it}(0) | D_i = 1] \tag{1}$$

where $t \geq t^\star$ is some period after the event date.

Identifying the ATT in (1) is a matter of identifying the mean untreated outcomes for the treated group in the post-period, i.e. $\beta_t \equiv \mathbb{E}[Y_{it}(0) | D_i = 1]$. A common approach for this is to assume that all differences between the treated and untreated units can be eliminated by conditioning on a k–dimensional vector of pre-treatment covariates, $\boldsymbol{X}_i$. This vector will typically include some or all of the pre-treatment outcomes (Y_{it} for $t < t^\star$), as well as potentially other pre-determined characteristics. The formal assumption consists of the following two parts.

Assumption 1. (Selection on observables) If $\boldsymbol{x}$ is in the supports of both $\boldsymbol{X}_i | D_i = 0$ and $\boldsymbol{X}_i | D_i = 1$, then $\mathbb{E}[Y_{it}(0) | D_i = 1, \boldsymbol{X}_i = \boldsymbol{x}] = \mathbb{E}[Y_{it}(0) | D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]$ for all $t \geq t^\star$.

Assumption 2. (Overlap) The support of $\boldsymbol{X}_i | D_i = 1$ is contained in the support of $\boldsymbol{X}_i | D_i = 0$.

Assumption 1 is a mean–independence version of what is variously described in the literature as ignorable treatment assignment (Rosenbaum and Rubin, 1983), unconfoundedness (Imbens and Rubin, 2015), or selection on observables (Barnow et al., 1980; Heckman and Robb, 1985). Together with Assumption 2, it implies that for post-treatment periods $t \geq t^\star$

$$\begin{aligned}\beta_t = \mathbb{E}\left[\,\mathbb{E}[Y_{it}|D_i = 0, \boldsymbol{X}_i]\middle| D_i = 1\right] &\equiv \mathbb{E}\left[\gamma_t(\boldsymbol{X}_i)|D_i = 1\right], \\ \text{where} \quad \gamma_t(\boldsymbol{x}) \equiv \mathbb{E}[Y_{it}&|D_i = 0, \boldsymbol{X}_i = \boldsymbol{x}]. \end{aligned} \tag{2}$$

That is, β_t is point identified by the outcomes for the untreated group, conditional on covariates, after re-weighting by the distribution of these covariates in the treated group. For further discussion, see e.g. Heckman et al. (1997, 1998), Imbens (2004, 2015), or Imbens and Rubin (2015).

Suppose now that we observe a sample of $n + 1$ realizations $\{(y_{i1}, \ldots, y_{iT}, d_i, \boldsymbol{x}_i)\}_{i=1}^{n+1}$ from the distribution of $(Y_{i1}, \ldots, Y_{iT}, D_i, \boldsymbol{X}_i)$. Our focus in this paper is the comparative case study setting considered by Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015), in which there is only a single treated unit. We label this treated unit as $i = 1$, so that $d_1 = 1$, while $d_i = 0$ for all n remaining units $i \geq 2$.

Since we only have a single treated unit, we estimate $\mathbb{E}[Y_{it}|D_i = 1]$ by the realization of Y_{1t} in the post-period. Similarly, since the empirical distribution of $\boldsymbol{X}_i$ given $D_i = 1$ is simply a point mass at $\boldsymbol{x}_1$, we estimate β_t with an estimator of $\gamma_t(\boldsymbol{x}_1)$. Thus, we focus on a class of estimators for the ATT of the form

$$\widehat{\text{ATT}}_t \equiv y_{1t} - \hat{\gamma}_t(\boldsymbol{x}_1), \tag{3}$$

where $\hat{\gamma}_t(\boldsymbol{x}_1)$ is an estimator of $\gamma_t(\boldsymbol{x}_1)$. Note that while Assumptions 1 and 2 justify an interest in estimating the population quantity, $\gamma_t(\boldsymbol{x}_1)$, there is no presumption that any of the untreated units in the sample have pre-treatment covariates $\boldsymbol{x}_1$.

The problem we focus on is how to construct $\hat{\gamma}_t(\boldsymbol{x}_1)$. The estimators we consider are all of the form

$$\hat{\gamma}_t = \sum_{i\geq 2} w_i y_{it} \equiv \boldsymbol{y}_{0t}'\boldsymbol{w} \tag{4}$$

where $\boldsymbol{w} \in \mathbb{R}^n$ are weights applied to the observed outcomes $\boldsymbol{y}_{0t} \in \mathbb{R}^n$ for the untreated units at time t. The weights will always be assumed to live in the $(n-1)$–dimensional simplex

$$\mathcal{S} \equiv \left\{ \boldsymbol{w} \in \mathbb{R}^n : \sum_j w_j = 1 \quad \text{and} \quad w_j \geq 0 \text{ for all } j \right\}, \tag{5}$$

so that $\hat{\gamma}_t$ is a convex weighted average of the outcomes for the untreated units at time t. The question is how to choose the weights, $\boldsymbol{w}$.

2.2 Extrapolation Bias and Interpolation Bias

Consider an estimator of form (4), and write it as

$$\hat{\gamma}_t = \sum_{i\geq 2} w_i \left(\gamma_t(\boldsymbol{x}_i) + u_{it}\right) = \overbrace{\sum_{i\geq 2} w_i \gamma_t(\boldsymbol{x}_i)}^{\text{signal}} + \overbrace{\sum_{i\geq 2} w_i u_{it}}^{\text{noise}}, \tag{6}$$

where $u_{it} \equiv y_{it} - \gamma_t(\boldsymbol{x}_i)$ denotes the deviation between y_{it} and its conditional mean. Our focus in this paper is on the signal term in (6) and under what conditions it can replicate $\gamma_t(\boldsymbol{x}_1)$. That is, we are concerned with the bias of estimators of form (4), as captured by the behavior of (6) when $u_{it} = 0$ for all i. We can decompose this bias into two components:

$$\underbrace{\gamma_t(\boldsymbol{x}_1) - \sum_{i\geq 2} w_i \gamma_t(\boldsymbol{x}_i)}_{\equiv \text{Bias}(\boldsymbol{w})} = \underbrace{\left[\gamma_t(\boldsymbol{x}_1) - \gamma_t\left(\sum_{i\geq 2} w_i \boldsymbol{x}_i\right)\right]}_{\equiv \text{ExtBias}(\boldsymbol{w})} + \underbrace{\left[\gamma_t\left(\sum_{i\geq 2} w_i \boldsymbol{x}_i\right) - \sum_{i\geq 2} w_i \gamma_t(\boldsymbol{x}_i)\right]}_{\equiv \text{IntBias}(\boldsymbol{w})},$$

where $\text{ExtBias}(\boldsymbol{w})$ is the **extrapolation bias** and $\text{IntBias}(\boldsymbol{w})$ is the **interpolation bias**.

To see the justification of these terms, consider a simple case in which there are two untreated units ($n = 3$), and $\boldsymbol{x}_i \equiv x_i$ is scalar ($k = 1$). Figure 1 plots $(x_i, \gamma_t(x_i))$ for $i = 1, 2, 3$, as well as $\gamma_t(x)$ as a function of x. Notice that x_1 lies between x_2 and x_3, so that it is an element of their convex hull.

One way to use the conditional means of the untreated units ($\gamma_t(x_2)$ and $\gamma_t(x_3)$) to approximate that of the treated unit ($\gamma_t(x_1)$) is to linearly interpolate between x_2 and x_3 to obtain

$$\gamma_t^{\text{li}} \equiv \gamma_t(x_2) + (\gamma_t(x_3) - \gamma_t(x_2)) \left(\frac{x_1 - x_2}{x_3 - x_2}\right).$$

This is equivalent to setting w_2 and w_3 to be

$$w_2^{\text{li}} \equiv 1 - \left(\frac{x_1 - x_2}{x_3 - x_2}\right) \quad \text{and} \quad w_3^{\text{li}} \equiv \left(\frac{x_1 - x_2}{x_3 - x_2}\right).$$

Since $x_1 = w_2^{\text{li}} x_2 + w_3^{\text{li}} x_3$, the extrapolation bias associated with the weights $\boldsymbol{w}^{\text{li}} \equiv (w_2^{\text{li}}, w_3^{\text{li}})$ is zero. However, as shown in Figure 1, there is still bias due to interpolation, because $\gamma_t(x)$ is not a linear function of x, and thus

$$\text{IntBias}(\boldsymbol{w}^{\text{li}}) = \gamma_t\left(w_2^{\text{li}} x_2 + w_3^{\text{li}} x_3\right) - \left[w_2^{\text{li}} \gamma_t(x_2) + w_3^{\text{li}} \gamma_t(x_3)\right] \neq 0.$$

Another way to use the untreated units is to simply use the conditional mean for the unit whose value of x_i is closest to x_1. In Figure 1, this is the second untreated unit, $i = 2$. The weights for this approximation strategy are the nearest neighbor weights of $w_2^{\text{nn}} = 1$ and $w_3^{\text{nn}} = 0$, which produce $\gamma_t(x_2)$ as an approximation to $\gamma_t(x_1)$. This approach does not interpolate, so

$$\text{IntBias}(\boldsymbol{w}^{\text{nn}}) = \gamma_t(1 \cdot x_2 + 0 \cdot x_3) - (1 \cdot \gamma_2(x_2) + 0 \cdot \gamma_3(x_3)) = 0.$$

However, it does extrapolate, creating bias to the extent that $\gamma_t(x_1) \neq \gamma_t(x_2)$.

The estimators we consider in this paper aim to minimize interpolation bias, extrapolation bias, or a combination of both, by minimizing bounds on these quantities. Assuming that γ_t is Lipschitz, the magnitude of extrapolation bias can be bounded by

$$|\text{ExtBias}(\boldsymbol{w})| \leq c \left\| \boldsymbol{x}_1 - \sum_{i \geq 2} w_i \boldsymbol{x}_i \right\| \equiv c \times \text{Ext}(\boldsymbol{w}),$$

where $c > 0$ is the Lipschitz constant. Under the same Lipschitz assumption, and assuming that interpolation is actually possible, so that the weights can be chosen to satisfy $\boldsymbol{x}_1 = \sum_{i \geq 2} w_i \boldsymbol{x}_i$ (that is, $\boldsymbol{x}_1$ is in the convex hull of $\{\boldsymbol{x}_i\}_{i \geq 2}$), the magnitude of interpolation bias can be bounded by

$$\begin{aligned} |\text{IntBias}(\boldsymbol{w})| &= \left| \sum_{i \geq 2} w_i \left(\gamma_t(\boldsymbol{x}_1) - \gamma_t(\boldsymbol{x}_i) \right) \right| \\ &\leq \sum_{i \geq 2} w_i |\gamma_t(\boldsymbol{x}_1) - \gamma_t(\boldsymbol{x}_i)| \leq c \sum_{i \geq 2} w_i \|\boldsymbol{x}_1 - \boldsymbol{x}_i\| \equiv c \times \text{Int}(\boldsymbol{w}). \end{aligned}$$

Thus, by choosing $\boldsymbol{w}$ to minimize Ext($\boldsymbol{w}$) and/or Int($\boldsymbol{w}$), one can control extrapolation and/or interpolation bias.

2.3 The Synthetic Control Estimator

The synthetic control (SC) estimator of $\gamma_t(\boldsymbol{x}_1)$ proposed by Abadie and Gardeazabal (2003) and later elaborated by Abadie et al. (2010, 2015) is defined as

$$\hat{\gamma}_t^{\text{sc}} \equiv \boldsymbol{y}_{0t}' \hat{\boldsymbol{w}}^{\text{sc}} \quad \text{where} \quad \hat{\boldsymbol{w}}^{\text{sc}} \equiv \underset{\boldsymbol{w} \in \mathcal{S}}{\arg\min} \left\| \boldsymbol{x}_1 - \boldsymbol{x}_0' \boldsymbol{w} \right\|^2 \equiv \underset{\boldsymbol{w} \in \mathcal{S}}{\arg\min} \, \text{Ext}(\boldsymbol{w})^2, \tag{7}$$

and where we have organized the untreated unit covariates into an $n \times k$ matrix $\boldsymbol{x}_0$. The Euclidean norm in the definition of Ext($\boldsymbol{w}$) might be weighted by some symmetric, positive semidefinite matrix, but we omit this from the notation for simplicity. The SC weights, $\hat{\boldsymbol{w}}^{\text{sc}}$, are chosen so that the weighted average of covariates among the untreated units comes as close as possible to matching the covariate vector of the treated unit, subject to the convexity constraint that they are non-negative and sum to unity.

The SC estimator has a number of attractive properties. By construction, it minimizes the quantity $\text{Ext}(\boldsymbol{w})$ that bounds extrapolation bias. If this quantity can be made zero, then the SC estimator will have no extrapolation bias, by construction. This stands in contrast to linear regression, which is known to be subject to potentially large extrapolation biases depending on how it is specified (see Imbens, 2004, pg. 13, or Abadie, 2019). Another benefit of the SC estimator is that the weights $\hat{\boldsymbol{w}}^{\text{sc}}$ are generally sparse, in the sense that they are only non-zero for a few untreated units (Abadie and L'Hour, 2019). This aids in transparency by providing a way for experts to use contextual knowledge to evaluate the plausibility of the resulting estimates. Also, solving for $\hat{\boldsymbol{w}}^{\text{sc}}$ only requires solving the quadratic program in (7), which is a straightforward convex problem.

One concern with the SC estimator is that it is susceptible to interpolation bias. This was noted by Abadie et al. (2010, pp. 495–496), and has been discussed more recently by Abadie and L'Hour (2019), although those authors emphasize non-uniqueness issues that occur with many treated or untreated units. In Figure 2, we illustrate how interpolation biases can arise with the SC estimator. This figure plots several untreated units in a draw from the empirical Monte Carlo simulation introduced in Section 4.6, which uses the Spanish terrorism data of Abadie and Gardeazabal (2003). The simulation considers a placebo exercise in which the "treated unit" is not in fact treated, so that the ground-truth treatment effect is known to be zero. The left-hand side of Figure 2 plots the outcome paths of the untreated units relative to that of the treated unit, while the right-hand side depicts different estimators, which should be zero in the post-period if they are performing well.

Suppose that we follow the recent tradition in the synthetic control literature of taking $\boldsymbol{X}_i$ to include all pre-treatment outcomes and no other covariates (Doudchenko and Imbens, 2016; Ferman, 2020). We focus on this case throughout the paper both because it allows us to examine the methods graphically, and also because it limits specification searching, but our discussion also applies to the case where $\boldsymbol{X}_i$ includes covariates other than pre-treatment outcomes. In Figure 2, this produces an SC estimator that is comprised of three regions with GDP per capita very different from the placebo region, Asturias. The SC estimator puts zero weight on the two regions whose pre-period paths oscillate closely around Asturias. Instead, it obtains the best pre-period fit by weighting three distant regions. This choice of weights minimizes extrapolation but creates interpolation.

Whether such interpolation leads $\hat{\gamma}_t^{\text{sc}}$ to be biased depends on the structure of the function $\gamma_t(\boldsymbol{x})$. Assuming that $\boldsymbol{x}_1$ lies in the convex hull of $\boldsymbol{x}_0$, the SC estimator will have no interpolation bias ($\text{IntBias}(\hat{\boldsymbol{w}}^{\text{sc}}) = 0$) if and only if

$$\sum_{i\geq 2} \hat{w}_i^{\text{sc}} \gamma_t(\boldsymbol{x}_i) = \gamma_t(\boldsymbol{x}_1) = \gamma_t\left(\sum_{i\geq 2} \hat{w}_i^{\text{sc}} \boldsymbol{x}_i\right). \tag{8}$$

In order for (8) to hold, the function γ_t needs to be linear in $\boldsymbol{x}$ on the empirical support of $\boldsymbol{X}_i$. This is a restrictive functional form assumption about an unknown object.

Suppose, for example, that $Y_{it}(0)$ follows a factor structure, as suggested by Abadie et al. (2010, 2015) and further elaborated by Gobillon and Magnac (2016), Xu (2017), Ferman and Pinto (2019), and Ferman (2020), so that

$$Y_{it}(0) = \boldsymbol{\varphi}_t' \boldsymbol{L}_i + U_{it} \quad \text{with} \quad \mathbb{E}[U_{it}|\boldsymbol{L}_i, Y_{i(t-1)}, \ldots, Y_{i1}] = 0, \tag{9}$$

where $\boldsymbol{\varphi}_t$ is a vector of unknown common factors and $\boldsymbol{L}_i$ is a vector of latent factor loadings. Then

$$\gamma_t(\boldsymbol{x}) = \boldsymbol{\varphi}_t' \boldsymbol{\lambda}(\boldsymbol{x}),$$

where $\boldsymbol{\lambda}(\boldsymbol{x}) \equiv \mathbb{E}[\boldsymbol{L}_i | \boldsymbol{X}_i = \boldsymbol{x}]$. With $\boldsymbol{x}_0' \hat{\boldsymbol{w}}^{\text{sc}} = \boldsymbol{x}_1$,

$$\sum_{i \geq 2} \hat{w}_i^{\text{sc}} \gamma_t(\boldsymbol{x}_i) = \boldsymbol{\varphi}_t' \left(\sum_{i \geq 2} \hat{w}_i^{\text{sc}} \boldsymbol{\lambda}(\boldsymbol{x}_i) \right) = \gamma_t(\boldsymbol{x}_1) + \boldsymbol{\varphi}_t' \left(\sum_{i \geq 2} \hat{w}_i^{\text{sc}} \boldsymbol{\lambda}(\boldsymbol{x}_i) - \boldsymbol{\lambda} \left(\sum_{i \geq 2} \hat{w}_i^{\text{sc}} \boldsymbol{x}_i \right) \right).$$

Thus, in order for (8) to be satisfied, the conditional means of the factor loadings, $\boldsymbol{\lambda}(\boldsymbol{x})$, must be linear in $\boldsymbol{x}$, at least barring knife-edge cases where multiple non-linearities cancel out. Abadie et al. (2010) show that if (9) holds throughout the pre-period, then this linearity will be approximately satisfied.

When (8) is not satisfied, interpolation bias will arise. This is illustrated in Figure 2, where the SC estimator fits the pre-period path of Asturias by interpolating between three regions with very different pre-period paths. The paths are highly nonlinear, suggesting that (8) fails, and that the resulting interpolation bias leads $\hat{\gamma}_t^{\text{sc}}$ to be a poor estimate of the post-treatment outcomes of Asturias. We emphasize that this bias stems from the failure of (8) even assuming perfect fit ($\boldsymbol{x}_1 = \boldsymbol{x}_0' \hat{\boldsymbol{w}}^{\text{sc}}$); it is distinct from the bias due to imperfect fit considered by Ferman and Pinto (2019) and Ben-Michael et al. (2019).

Of course, Figure 2 is just one simulation draw, and one which we have selected to show how interpolation bias can arise. The frequency with which it actually arises in applications is an empirical question. We address this empirical question for the Spanish data in Section 4, where we report the results from our full placebo analysis and Monte Carlo simulations.

2.4 The Matching Estimator

Local nonparametric smoothing estimators are a classical way to estimate $\gamma_t(\boldsymbol{x}_1)$. In general, these estimators can be written as

$$\hat{\gamma}_t^{\text{lo}} \equiv \sum_{i \geq 2} \kappa\left(\|\boldsymbol{x}_i - \boldsymbol{x}_1\|\right) y_{it} \equiv \boldsymbol{y}_{0t}' \hat{\boldsymbol{w}}^{\text{lo}}, \tag{10}$$

where κ is a kernel function that determines the weight applied to each untreated observation. For such an estimator to be local, the function κ should be decreasing, so that untreated units with predetermined characteristics more distant from the treated unit are given less weight. Local smoothing estimators do not require the linearity condition (8) that was required for the SC estimator. Instead, they rely only on γ_t being sufficiently smooth in its continuous components (e.g. Fan and Gijbels, 1992).

Unlike the SC estimator, local smoothing estimators do not necessarily have sparse, convex weights. However, the specific class of k–nearest neighbors estimators (e.g. Cover, 1968) does have weights with these properties. Estimators based on the nearest neighbors idea are widely used for causal inference problems under Assumptions 1 and 2, in which case they are commonly described as matching estimators (e.g. Dehejia and Wahba, 1999; Abadie and Imbens, 2006).

The matching estimator we consider is defined by choosing a positive integer and equally weighting the m untreated units with pre-period characteristics closest to those of the treated unit. That is,

$$\hat{\gamma}_t^{\text{ma}}(m) \equiv \boldsymbol{y}_{0t}' \hat{\boldsymbol{w}}^{\text{ma}}(m), \tag{11}$$

where the weights $\hat{w}_i^{\text{ma}}(m)$ are $1/m$ for the m units with smallest $\|x_i - x_1\|$ and 0 for all other units. For simplicity, we assume there are no ties. Note that this vector of weights can be written as the solution to the optimization problem

$$\hat{\boldsymbol{w}}^{\mathrm{ma}}(m) = \arg\min_{\boldsymbol{w}\in\mathcal{S}} \underbrace{\sum_{i\geq 2} w_i \|\boldsymbol{x}_1 - \boldsymbol{x}_i\|}_{\equiv \mathrm{Int}(\boldsymbol{w})} \quad \text{s.t.} \quad w_i \leq \frac{1}{m} \quad \text{for all } i \geq 2, \tag{12}$$

since the w_i corresponding to the smallest $\|\boldsymbol{x}_1 - \boldsymbol{x}_i\|$ will be pushed up against $1/m$ until the m smallest observations have reached this bound, while all other w_i will be set to 0.

Like the SC estimator, the matching estimator is a sparse, convex weighted average of the post-period outcomes of the untreated units. However, in contrast to the SC estimator, the matching estimator aims to minimize $\mathrm{Int}(\boldsymbol{w})$, and thus minimize interpolation bias. For example, with $m = 2$, the matching estimator equally weights the two regions in Figure 2a that oscillate around the placebo region, since their pre-period outcome paths are close to that of the placebo region. As a consequence, the matching estimator is less susceptible to interpolation bias than the SC estimator, and in this example provides a better estimate of the post-treatment outcomes for the placebo region.

However, the matching estimator is more vulnerable to extrapolation bias than the SC estimator. To see this, consider Figure 3, which reports a draw from our simulation with Navarre as the placebo region. In this draw, the matching estimator uses the single untreated unit that has pre-period path closest to Navarre, even though their pre-period paths are not actually that close, resulting in considerable bias. In contrast, the SC estimator weights two distant regions in a way that provides an excellent fit to the outcome path of Navarre throughout the pre- and post-period. This is consistent with a case in which γ_t is close to linear, so that (8) is close to satisfied, and the SC estimator has little interpolation bias.

2.5 Model Averaging with the MASC Estimator

Both the SC and matching estimators share a number of appealing properties in common. As illustrated in Figures 2 and 3, however, their drawbacks are different and diametrically opposed: the SC estimator controls extrapolation bias but not interpolation bias, while the matching estimator does the opposite. This complementarity suggests that a model averaging estimator will be able to harness the best properties of both the matching and SC estimators. See, for example, the discussion surrounding Theorem 1 of Breiman (1996).

With this motivation, we define the matching and synthetic control (MASC) estimator as

$$\hat{\gamma}_t^{\text{masc}} \equiv \phi\hat{\gamma}_t^{\text{ma}}(m) + (1-\phi)\hat{\gamma}_t^{\text{sc}} \equiv \boldsymbol{y}_{0t}'\hat{\boldsymbol{w}}^{\text{masc}}$$

where $\phi \in [0,1]$ is a tuning parameter, and $\hat{\boldsymbol{w}}^{\text{masc}} \equiv \phi\hat{\boldsymbol{w}}^{\text{ma}}(m)+(1-\phi)\hat{\boldsymbol{w}}^{\text{sc}}$. In Section 3, we provide a cross-validation procedure for choosing ϕ and m. This allows the MASC to control both interpolation and extrapolation biases in a data-driven way. When interpolation bias is the chief concern, the procedure makes the MASC estimator assign more weight to the matching estimator. In Figure 2, it sets $\phi = 1$ and $m = 2$, so that the MASC exactly coincides with the matching estimator with two matches. On the other hand, when extrapolation bias is the concern, the procedure assigns more weight to the SC estimator. For example, in Figure 3, it sets $\phi = 0$, so that the MASC exactly coincides with the SC estimator.

Intermediate cases can also arise, as the simulation draw depicted in Figure 4. In this case, the outcome paths are moderately non-linear, so that the SC estimator suffers from interpolation bias. As the same time, there are no untreated units that closely match the pre-period path of the placebo unit (Rioja), so the matching estimator suffers from extrapolation bias. In contrast, the cross-validation procedure chooses $\phi \approx .5$, which allows the MASC estimator to mix the SC estimator with the matching estimator, mitigating both sources of bias.

2.6 The Penalized Synthetic Control Estimator

A related, but different estimator has recently been proposed by Abadie and L'Hour (2019). Those authors start with the SC estimator and add a penalty that discourages choosing units far from the treated unit. Their penalized SC estimator is defined as

$$\begin{aligned} \hat{\gamma}_t^{\text{pen}} &\equiv \boldsymbol{y}_{0t}'\hat{\boldsymbol{w}}^{\text{pen}} \\ \text{with} \quad \hat{\boldsymbol{w}}^{\text{pen}} &\equiv \operatorname*{arg\,min}_{\boldsymbol{w}\in\mathcal{S}} \, (1-\pi)\|\boldsymbol{x}_1 - \boldsymbol{x}_0'\boldsymbol{w}\|^2 + \pi\left(\sum_{i\geq 2} w_i\|\boldsymbol{x}_i - \boldsymbol{x}_1\|^2\right), \end{aligned} \tag{13}$$

where $\pi \in [0,1]$ is a tuning parameter chosen through cross-validation that controls the penalty incurred by weighting untreated units with pre-treatment characteristics different from the treated unit. When $\pi = 0$, the penalized SC estimator reduces to the usual SC estimator, $\hat{\gamma}_t^{\text{sc}}$, while for $\pi = 1$, it is equal to $\hat{\gamma}_t^{\text{ma}}(m)$ with $m = 1$. (Note that Abadie and L'Hour (2019) parameterize their criterion function slightly differently by normalizing $(1-\pi)$ to 1 and allowing $\pi \geq 0$. The two formulations are equivalent.)

The optimization problem solved by the penalized SC estimator is a constrained version of the one implicitly solved by the MASC estimator. This is because (13) can also be written as

$$\hat{\boldsymbol{w}}^{\text{pen}} = \operatorname*{arg\,min}_{\boldsymbol{w}^a,\boldsymbol{w}^b\in\mathcal{S}} \, (1-\pi)\|\boldsymbol{x}_1 - \boldsymbol{x}_0'\boldsymbol{w}^a\|^2 + \pi\left(\sum_{i\geq 2} w_i^b\|\boldsymbol{x}_i - \boldsymbol{x}_1\|^2\right) \quad \text{s.t.} \quad \boldsymbol{w}^a = \boldsymbol{w}^b,$$

whereas $\hat{\boldsymbol{w}}^{\text{masc}}$ is the solution to this program (with π replaced by ϕ) when m is fixed at 1 and the constraint $\boldsymbol{w}^a = \boldsymbol{w}^b$ is dropped. Note that when this constraint is dropped the problem becomes separable in $\boldsymbol{w}^a$ and $\boldsymbol{w}^b$, and the squares on the norms can be removed without changing the optimal solutions. While the MASC estimator takes a convex combination of the SC and matching estimators—which respectively minimize extrapolation and interpolation bias—the penalized SC estimator minimizes a convex combination of the (squared) SC and matching objective functions. This can lead it to choose an entirely different set of weights.

It is important to mention that we are ignoring a primary motivation provided by Abadie and L'Hour (2019) for the penalized SC estimator, which is its ability to solve the non-uniqueness problem that can arise when solving the SC problem (7). As Abadie and L'Hour (2019) discuss, this problem is usually not an issue when there is a single treated unit, which is the case we consider here. It becomes much more likely to be problematic with multiple treated units. In such settings, one could modify the MASC so that it averages between the matching and *penalized* SC estimators. We expect that the resulting estimator would behave similar to the way the MASC behaves when there is a single treated unit.

2.7 Sparsity

A key motivation for the synthetic control method is sparsity in the weights. In comparative case studies, sparsity facilitates the interpretation of the counterfactual estimate and the recognition and assessment of potential biases (see, e.g. Abadie, 2019).

The sparsity properties of the MASC estimator are inherited by those of the SC and MA estimators. Suppose that n^{sc} of the components of $\hat{\boldsymbol{w}}^{\text{sc}}$ are non-zero. Then at most $n^{\text{sc}} + m$ of the components of $\hat{\boldsymbol{w}}^{\text{masc}}$ are non-zero. In practice, we often find that the untreated units given non-zero weights by the SC and MA weighting schemes are partially overlapping, so that the actual number of non-zero elements is smaller than this. The smallest number of non-zero components that $\hat{\boldsymbol{w}}^{\text{masc}}$ can have is the minimum of m and n^{sc}, which happens if $\phi = 1$ or $\phi = 0$, respectively. Since m can be as small as 1, the MASC weights could have only one non-zero weight. Thus, the MASC is generically neither more nor less sparse than the SC, MA or penalized SC estimators. In the application in Section 5, we find that MASC has four positive weights, while SC, matching and penalized SC all have three.

3 Cross-Validation

3.1 Definitions

In this section, we propose a cross-validation procedure for choosing the tuning parameters for the estimators discussed in the previous section. As in Abadie et al. (2015), our procedure is based on optimizing the fit of the treated unit's outcome series in the pre-treatment period. Whereas those authors used a single training-validation split, our procedure uses a series of one-step ahead forecasts, each of which is estimated using data only from periods prior to the forecast date. This is called rolling-origin recalibration in the forecasting literature (e.g. Tashman, 2000; Bergmeir and Benítez, 2012), which is related to the rolling-window considered by Swanson and White (1997).

We define our folds, $f = 1, \ldots, F$, as consisting of data running between two dates $\underline{t}_f$ and $\bar{t}_f$ in the pre-treatment period. Let $\hat{\gamma}_f(\boldsymbol{\tau})$ denote a generic estimator of the outcome in period $\bar{t}_f + 1$ based on data in fold f, where $\boldsymbol{\tau}$ is a vector of tuning parameters. Our cross-validation procedure chooses $\boldsymbol{\tau}$ to minimize the average one-step ahead forecast error, which we denote as

$$\text{cv}(\boldsymbol{\tau}) \equiv \frac{1}{F} \sum_{f=1}^{F} \left(y_{1(\bar{t}_f+1)} - \hat{\gamma}_f(\boldsymbol{\tau}) \right)^2 . \tag{14}$$

We consider the one-step ahead forecast primarily for concreteness; one could modify the criterion (14) to combine multiple forecast periods under different weights chosen by the researcher. Our R package (https://github.com/maxkllgg/masc) implements such a modification to allow for more general criteria.

Figure 5 illustrates the structure of the rolling-origin cross-validation procedure. In this example, the treatment date is $t^\star = 21$, so that there are 20 pre-treatment periods. A fold with $\bar{t}_f = 19$ uses data from $\underline{t}_f$ up to $\bar{t}_f$ to construct a forecast of the treated unit's outcome in period $\bar{t}_f + 1 = 20$. Fold $f = 18$ uses data from $\underline{t}_f, \ldots, 18$ to forecast at $\bar{t}_f + 1 = 19$, and so on. The criterion is constructed by averaging together the squared prediction errors from all folds $f = 1, \ldots, F$. The attractive part of the rolling-origin cross-validation procedure is that it preserves the temporal structure of the forecasting problem.

The largest that F can be is $t^\star - 2$ if we set $\underline{t}_f = 1$ and $\bar{t}_f = f$ for all $f = 1, \ldots, (t^\star - 2)$. In practice, we use fewer folds than this, and prefer folds that are longer. The bias-variance trade-offs that drive this choice are natural. Folds that end closer to the treatment date ($\bar{t}_f$ closer to $t^\star$) are likely to be more relevant to the post-treatment period. They can also be made longer ($\bar{t}_f - \underline{t}_f$), so that the estimators use more data. On the other hand, we expect that having more folds will decrease the variance of $\mathrm{cv}(\boldsymbol{\tau})$ in repeated samples. These trade-offs are also present in cross-validation with independent and identically distributed data (e.g. Hastie et al., 2009, pg. 242–243). The added complication here is that not all folds are equally valuable, so we prefer ones that use data closer to the actual treatment date.

The parameters $\boldsymbol{\tau}$ differ by estimator. The synthetic control estimator has no tuning parameters. (As mentioned earlier, the Euclidean norm defining the synthetic control or matching estimators could be weighted. Abadie et al. (2010, 2015) view the weights as tuning parameters and choose them using cross-validation. We could do this as well with our criterion (14), but we have elected not to in the current paper because optimizing over the weights introduces computational issues that, while solvable, are not the main focus of our paper (Becker and Klößner, 2017, 2018).) For the matching estimator, $\boldsymbol{\tau}$ is the number of matches, m. The MASC estimator has both m and the model average parameter, ϕ. The penalized synthetic control estimator has the penalty parameter, π.

3.2 Computation

For the MASC estimator, it is straightforward to find the unconstrained minimum of $\mathrm{cv}(\boldsymbol{\tau}) \equiv \mathrm{cv}(\phi, m)$ in ϕ for any fixed m. Using least squares algebra, the solution can be shown to be

$$\hat{\phi}(m) \equiv \frac{\sum_{f=1}^{F}(\hat{\gamma}_f^{\mathrm{ma}}(m) - \hat{\gamma}_f^{\mathrm{sc}})(y_{1,\bar{t}_f+1} - \hat{\gamma}_f^{\mathrm{sc}})}{\sum_{f=1}^{F}(\hat{\gamma}_f^{\mathrm{ma}}(m) - \hat{\gamma}_f^{\mathrm{sc}})^2}. \tag{15}$$

This means that cross-validating the MASC is extremely easy computationally. First, compute $\hat{\phi}(m)$ for a set of potential matches, $m \in \mathcal{M}$. Then for each $m \in \mathcal{M}$, set

$$\hat{\phi}^{\star}(m) \equiv \begin{cases} 0, & \text{if } \hat{\phi}(m) \leq 0 \\ 1, & \text{if } \hat{\phi}(m) \geq 1 \\ \hat{\phi}(m) & \text{otherwise} \end{cases}$$

Finally set $\hat{m}^{\star} \equiv \arg\min_{m \in \mathcal{M}} \mathrm{cv}(\hat{\phi}^{\star}(m), m)$, and set $\hat{\phi}^{\star} \equiv \hat{\phi}^{\star}(\hat{m}^{\star})$. The cross-validated MASC estimator is a weighted average of $\hat{\gamma}^{\mathrm{sc}}$ and $\hat{\gamma}^{\mathrm{ma}}(\hat{m}^{\star})$ with weights $(1 - \hat{\phi}^{\star})$ and $\hat{\phi}^{\star}$, respectively.

For the penalized SC estimator, $\mathrm{cv}(\boldsymbol{\tau}) \equiv \mathrm{cv}(\pi)$ is not necessarily convex in π, which makes it harder to find the global minimum. In the results ahead, we use a grid search to cross-validate the penalized SC estimator.

3.3 Some Statistical Properties of the MASC Estimator

The statistical properties of matching estimators are well understood. In contrast, much less is known about the SC estimator, although this is rapidly changing, with recent advances by Abadie and L'Hour (2019) and Ferman (2020), among others. Since the MASC is an average of the SC and matching estimators, its properties depend on both, but our understanding of these properties is bottlenecked by our understanding of SC. Nevertheless, since the MASC is determined through cross-validation, we can make some limited statements by using the observation that both SC and matching are weakly suboptimal solutions of the cross-validation problem.

The first conclusion we can draw is a bound on the MSPE of the MASC.

Proposition 1. Suppose Assumptions A.1–A.6 in Appendix A are satisfied. Then

$$\mathbb{E}\left[\left(Y_{1t^\star}(0) - \hat{\Gamma}^{\text{masc}}\right)^2\right] \leq O(n^{-2/k}) + \left(\mathbb{E}[U_{1t^\star}^2] + \max_{i\geq 2} \mathbb{E}[U_{it^\star}^2]\right), \tag{16}$$

where $U_{it^\star} \equiv Y_{it^\star}(0) - \gamma_{t^\star}(\boldsymbol{X}_i)$, and $\hat{\Gamma}^{\text{masc}}$ is the stochastic counterpart of $\hat{\gamma}^{\text{masc}}$ in repeated samples.

The proof is presented in the appendix. The bound in Proposition 1 arises by bounding the cross-validation error of MASC with that of the nearest neighbor estimator, which is weakly larger by definition of the MASC. Under a stationarity assumption, the cross-validation criterion faithfully reproduces the MSPE on average, so that this bound implies that the MSPE of MASC is also bounded by that of the nearest neighbor estimator. The MSPE of the nearest neighbor estimator has a bias term and a noise component. Abadie and Imbens (2006, Lemma 2) derived a bound on the former, the square of which is the first term in (16); see also Lemma 2 of Abadie and L'Hour (2019). The latter is bounded by the variances of the residual for the treated unit and its nearest neighbor match, which is the second term in (16).

In a similar vein, we can use the fact that the MASC chooses optimally between the SC and matching estimators to show that it is consistent if either one is consistent.

Proposition 2. Suppose that Assumptions A.3–A.4, and A.7–A.8 in Appendix A are satisfied, and that either $\hat{\gamma}^{\text{sc}}$ or $\hat{\gamma}^{\text{ma}}$ is consistent for $\gamma_{t^\star}(\boldsymbol{X}_1)$. Then $\hat{\gamma}^{\text{masc}}$ is also consistent for $\gamma_{t^\star}(\boldsymbol{X}_1)$.

The proof is presented in the appendix. Proposition 2 shows that the cross-validation approach embedded in the MASC ensures that it will be consistent whenever one of matching or SC is consistent. This is like a double robustness property (e.g. Bang and Robins, 2005). Consistency here requires that at least the pre-treatment period is large, so that the number of folds in the cross-validation procedure is also large. It also potentially requires a large number of untreated units for the consistency of either matching or SC, see e.g. Biau and Devroye (2015, Corollary 11.1) or Ferman (2020). Li (2019) provides a consistency result for SC with a large pre-treatment period, but a small number of untreated units.

4 Placebo Analyses

4.1 Design

We use a series of empirical placebo analyses to examine the behavior of the estimators described in Section 2. These exercises use the same data as in Abadie and Gardeazabal's

(2003) application of the SC method to study the effect of terrorism on per capita GDP in Spain. The data consists of time series on per-capita GDP running from 1955 to 1997 for 17 regions in Spain. The treated unit is the Basque Country, and the treatment is the onset of separatist terrorism, which begins in 1970.

Abadie and Gardeazabal (2003) performed a placebo analysis using Catalonia as the placebo region. Their stated rationale was that Catalonia is similar to the Basque Country, but with lower exposure to terrorism, and particularly salient for their results, since it received the most weight in their application of the SC method. They found that the SC estimator reproduced the actual per capita GDP for Catalonia quite well, at least until the 1980s. They interpreted this as evidence in support of their estimates for the Basque Country.

Using the same logic, we extend this placebo exercise to all of the untreated regions of Spain, with the exception of the Balearic Islands, Extremadura, and Madrid. The reason for excluding these three regions is that the SC estimator provides a particularly poor fit to their pre-period paths. Given the poor fit, one might argue that it is inappropriate to apply SC to these regions. (We thank the associate editors for pointing this out and suggesting that we exclude these three regions. Empirically, we find that including or excluding these regions does not materially change our findings about the relative performance of the various estimators. Results including these regions are provided in Supplemental Appendix D.)

The placebo analyses are conducted separately for each of the remaining 13 untreated regions. We use the same methodology described in Sections 2 and 3, except that now the "treated unit" is a placebo region in which no intervention took place at $t^{\star} = 1970$. For each estimator, we use data from 1955-1969 and cross-validate with $F = 7$ folds, each starting at $\underline{t}_f = 1955$ and ending at $\bar{t}_f \in \{1962, 1963, \ldots, 1968\}$. The number of matches for the matching and MASC estimators is chosen from $\mathcal{M} = \{1, 2, \ldots, 10\}$.

We calculate the mean squared prediction error (MSPE) for each region by taking the differences between its actual and forecasted outcome paths in each of the first four post-treatment years (1970–1973), squaring these differences, and then averaging the four years. (Our findings about the relative performance of the various estimators do not materially change if we use a longer post-period, a point we demonstrate in Figure B.1 of the Supplemental Appendix.) This procedure produces a MSPE for each placebo region and every estimator. As in Abadie and Gardeazabal (2003), we interpret a low MSPE as evidence that an estimator is performing well. Throughout our discussion, we focus on the square root of the MSPE (the RMSPE) so that errors are interpretable in units of GDP per capita.

4.2 Performance Across Estimators

Figure 6 compares the performance of four alternative estimators across each placebo regions in terms of RMSPE. In panel (a), we present the results for the regions with relatively low RMSPE, while we in (b) display the results for the regions with higher RMSPE. Note that, as a consequence, the scales on the vertical axes are different in panels (a) and (b). To help interpret the magnitudes of the estimates, each graph reports the RMSPE both in units of GDP per capita (left vertical axis), and as percent of GDP per capita in the Basque Country in 1969, the year prior to treatment (right vertical axis).

Qualitatively, the MASC tends to outperform the other estimators, including the penalized SC estimator, which frequently coincides with the standard SC estimator. The MASC adapts to regions such as Asturias and Andalusia, where matching performs well but SC does poorly. It also adapts to regions such as Navarre and the Canary Islands, where matching does poorly, but SC does well. In regions like Rioja, Murcia, and Castile-La Mancha, where neither matching nor SC perform well, the MASC outperforms both.

The four estimators exhibit noticeable quantitative differences in performance. On average across the placebo regions, the yearly RMSPE of MASC is \$96.6 per person, equivalent to 1.6 percent of GDP per capita in the Basque Country in 1969. By comparison, the SC estimator has an average yearly RMSPE of \$125.4 per person, which amounts to 2.1 percent of the Basque Country's GDP per capita. The average RMSPE of the penalized SC estimator is very similar to that of the SC estimator. Matching, on the other hand, tends to have larger prediction errors, with an average yearly RMSPE of \$157.2 per person, or 2.6 percent of the Basque Country's GDP per capita. To put these estimates into perspective, the yearly GDP growth of the Basque Country averaged \$159.2 per person across the years 1955-1969. This means that the average yearly prediction error of the estimators ranges from 61 percent (MASC) to 79 percent (SC and penalized SC) and 99 percent (matching) of the annual GDP growth in the treated region prior to the onset of terrorism.

4.3 Pre-Period Fit and Prediction Error of Synthetic Control

The performance of the SC estimator varies across placebo regions. In some regions, its performance is similar to MASC, while in others it performs considerably worse. There are two possible explanations for why SC produces relatively large prediction errors in certain placebo regions. One possibility is that the synthetic unit does not fit the pre-period paths in these regions. The other possible explanation is that the pre-period fit is good, and the prediction error results from the susceptibility of the SC estimator to interpolation bias. To evaluate these explanations, we compare the pre-period fit of the SC estimator to its prediction error in the post-period data. These results are reported in Figure 7.

One finding is that the SC estimator is able to fit the pre-period paths of the placebo regions quite well. Indeed, for every one of the placebo regions, we obtain a pre-period fit (RMSE) that is *lower* than that obtained for the Basque Country, which is the actual treated region. (We discuss these results in Section 5, where we find the RMSE of SC when applied to the Basque Country to be $75.6 per person.) Another finding is that placebo regions with good (or poor) pre-period fits do not necessarily have small (or large) prediction errors. For example, the SC estimator fits the pre-period data the worst in the regions of Murcia and Castile-La Mancha. Yet the prediction error is low in one of these regions, Murcia, and high in the other region, Castile-La Mancha.

4.4 Cross-Validation and Prediction Error of MASC

The results in Figure 6 suggest that MASC performs relatively well compared to the other estimators, at least in the current setting. However, MASC also has non-negligible prediction errors in some placebo regions. One potential reason for poor performance is that a suitable control group simply does not exist; that is, no combination of ϕ and m would lead to low prediction error. Another possibility is that a suitable control group exists, but the cross-validation procedure does a poor job locating it. To distinguish between these two scenarios, we report a region-by-region comparison of the actual RMSPE against the best possible (infeasible) RMSPE that MASC could obtain if ϕ and m were cherry-picked to directly minimize it in the post-treatment period. These results are reported in Figure 8.

Averaging across regions, the minimum infeasible RMSPE is 45 percent lower than the actual RMSPE. However, there is a great deal of heterogeneity across regions. For example, the MASC has the highest RMSPE in the Canary Islands, but Figure 8 shows that this is because there is no suitable control group (choice of ϕ and m) for that region, not because the cross-validation procedure is failing. On the other hand, Figure 8 suggests that MASC has relatively high prediction errors for Asturias and Castile-La Mancha because the cross-validation procedure selects a control group that is far from infeasible optimal one.

In Appendix C, we explore alternative cross-validation procedures, including changing the criterion to consider forecasts that are multiple steps ahead, either individually or averaged, as well as a leave-one-out criterion that measures prediction error in the middle of a fold. Our results there suggest that none of these alternatives achieves a MSPE that is systematically lower across placebo regions than the one-step ahead rolling-origin cross-validation procedure used in our main results.

4.5 Biases due to Interpolation versus Extrapolation

The results in Figure 6 show that MASC tends to outperform the other estimators in the placebo analyses. In Figure 9, we investigate the reason for this by plotting the pre-period fit (RMSE) and post-period prediction error (RMSPE) as functions of ϕ and π for the MASC and penalized SC estimators. We report the average across all placebo regions in dotted lines. We also report Andalusia separately as a solid line, since this is the only placebo region in which SC and penalized SC do not coincide.

When $\phi = \pi = 0$, both the MASC and penalized SC estimators correspond to the standard SC estimator, which minimizes extrapolation bias by maximizing pre-period fit. Panels (a) and (c) of Figure 9 shows how pre-period fit deteriorates monotonically (and mechanically) as ϕ and π increase. At $\phi = \pi = 1$, both the MASC and penalized SC estimators correspond to the matching estimator, which minimizes interpolation bias. Intermediate values of ϕ and π represent a trade-off between extrapolation and interpolation bias. MASC captures this trade-off by assigning weight to both the SC and matching estimators. Penalized SC captures it by changing the relative penalty for giving weight to distant units.

Panels (b) and (d) of Figure 9 show the prediction errors of MASC and penalized SC. These do not need to be monotonic functions of ϕ or π, because these parameters regulate the trade-off between controlling extrapolation and interpolation biases. Indeed, average prediction error for the MASC is minimized at around $\phi = .4$, reflecting the observation in Section 4.2 that MASC is able to adapt to regions where either SC or matching performs well, while blending the two successfully in regions where both perform poorly. In contrast, average prediction error for penalized SC is monotonic in π, a consequence of the fact that the cross-validation procedure almost always chose penalized SC to coincide exactly with SC. The one exception is Andalusia, plotted in the solid line, where penalized SC obtains a lower MSPE than SC by setting $\pi = 1$, rendering it identical to the matching estimator.

4.6 Monte Carlo Simulations

The placebo analyses in this section have been based on a given data set, that is, on one particular realization of the underlying data generating process. This raises two questions. The first is whether the relative performance of the alternative estimators would change when looking across multiple realizations of the same data generating process. The second is how the estimators would perform under alternative data generating processes. To answer both questions, we conducted a Monte Carlo study, which we discuss in detail in Appendix E. Here, we briefly describe the design and results.

The data generating process for the Monte Carlo is a linear factor model with four factors. We take the errors to be normally distributed errors, as in the simulations in

Ferman and Pinto (2019) and Chernozhukov et al. (2020), and model them as an AR(1) process with heteroskedasticity across regions. We first remove year effects from the Spanish data, then fit the residuals to the factor model. Appendix Figure E.6 shows that the model does a good job at reproducing patterns in the original data with a bit of added sampling error.

We use the fitted model to generate simulated data by taking random draws of the stochastic component. We then use the simulated data to compare estimators through the same type of placebo analyses as in Section 4. The only difference is that now we consider the *distribution* of MSPEs across simulation draws for each placebo region. We expect this distribution to be clustered close to zero if the estimator is working well.

The results from this simulation support the four key insights from the placebo analyses based on real data: (i) MASC tends to outperform the other estimators; (ii) the pre-period fit of the SC estimator is not necessarily a strong indicator of its prediction error; (iii) the cross-validation procedure tends to select suitable control groups for MASC when they exist; and (iv) MASC is much more likely than penalized SC to assign positive weight to matching, which reduces interpolation bias, and, ultimately, reduces prediction error.

We then change the data generating process by reducing the number of factors from four to two. This provides a case in which the SC estimator is likely to be a preferred estimator, since a close-fitting synthetic control is easier to find. The fit of the model to the data is, however, much worse, so this simulation is meant as expository, not necessarily realistic. With the two-factor model, the SC estimator does much better, since it more closely satisfies the form (8) needed to not have interpolation bias. The penalized SC continues to behave similar to SC, and MASC is competitive, but slightly worse, since the stochastic errors occasionally make it assigns weight to the matching estimator. Taken together, the simulation results indicate the importance of considering the potential for interpolation bias in SC and related methods, and in particular of assessing the plausibility of the necessary linearity condition, (8).

5 Re-Examining the Economic Costs of Conflict

In this section, we re-analyze the Spanish terrorism application of Abadie and Gardeazabal (2003). The goal is to assess if the alternative estimators yield substantively different estimates of the economic costs of conflict for the Basque Country. We continue to follow the same cross-validation procedure as in Section 4, except that now the Basque Country is the treated unit. The economic costs of conflict are calculated by taking the difference between the Basque Country's actual and forecasted outcome path over the post-period.

5.1 Effect Estimates using MASC

Figure 10 presents the pre-period fit and the post-period estimated costs of terrorism per year for MASC. On average, MASC gives a RMSE of \$97.5 per person in the pre-period. By comparison, the estimated average effect of terrorism in the post-period is an order of magnitude larger, at around \$982 per person, per year over the course of 1970-1997. The estimated effect peaks in 1987, reaching a yearly cost of \$1,558 per person, which is equivalent to 25.6 percent of GDP per capita in the Basque Country in 1969. Cumulating the yearly effects from 1970 to 1997 gives a total loss of \$27,516 per capita. This total effect is nearly five times as large as the GDP per capita in the Basque Country in 1969.

5.2 Comparison of Effect Estimates across Estimators

Figure 11 reports the yearly differences between the effect estimates of MASC and those produced by each alternative estimator.

The SC estimator produces systematically smaller effect estimates than the MASC estimator. On average across the years 1970-1997, the SC estimates of the annual costs per person is $88.2 lower than the MASC estimates of these same costs. To put this difference into perspective, it is about 1.4 percent of the Basque Country's GDP per capita in 1969, or more than half of this region's yearly GDP growth over the years 1955-1969. The differences between the MASC and the SC estimates accumulate from 1970 to 1997 to $2,470 per capita, which is equal to 40 percent of the GDP per capita in the Basque Country in 1969.

The penalized SC estimator gives even smaller effect estimates than SC. On average, the penalized SC estimates of the annual costs per person is $249 lower than the MASC estimates of these costs. The MASC and penalized SC estimates differ the most in 1987, reaching a difference in the estimated costs of terrorism of $433.6 per person, or, equivalently, 7.1 percent of GDP per capita in the Basque Country in 1969. Cumulating the yearly differences from 1970 to 1997 gives a total difference in the economic costs of terrorism of $6,972 per capita. This total difference is larger than the entire GDP per capita of the Basque Country in 1969

The matching estimator gives the largest effect estimates of the four estimators we consider. On average, the matching estimates of the annual costs per person are $106 larger than the MASC estimates. This additional cost is equal to 1.7 percent of the Basque Country's per capita GDP in 1969. The differences between the MASC and the matching estimates accumulate from 1970 to 1997 to $2,965 per capita, which is equal to 49 percent of the GDP per capita in the Basque Country in 1969.

5.3 Choice of Control Group, Sparsity and Pre-Period Fit

Figure 12 shows the weights of the untreated regions that make up the control groups for each estimator. The cross-validation procedure selects three untreated regions for the matching estimator: Madrid, Catalonia, and the Balearic Islands. Each of these regions receive the same weight. The SC estimator also chooses three untreated regions in the construction of the synthetic control. Two of these regions, Madrid and the Balearic Islands, are also selected by the matching estimator, although the weight they receive differs. SC chooses Rioja as the third region, whereas matching chooses Catalonia.

As discussed in Section 2.7, the weights for both the MASC and penalized SC estimators can be sparser than the weights of either the SC or matching estimators, but cannot be sparser than both simultaneously. In our context, it turns out that matching and penalized SC form control groups based on exactly the same three regions. MASC, on the other hand, assigns roughly the same weight to the matching estimator as to the SC estimator. As a result, the control group of the MASC estimator consists of four untreated regions: Madrid, Catalonia, Rioja, and the Balearic Islands.

The denser set of weights for the MASC does not mean that it has superior pre-period fit. On the contrary, the SC estimator has sparser weights yet it maximizes, by construction, the fit in the pre-period. This can be seen in Figure 13, where we plot the year-to-year error of each estimator in the pre-period. The SC estimator has a RMSE of \$75.6 per person in the pre-period. The penalized SC estimator fits the data similarly, whereas the MASC estimator has a worse fit with a RMSE of \$97.5 per person. By comparison, the matching estimator, which has the worst fit, has a RMSE of \$154.3 per person. As the placebo study in Section 4 illustrated, pre-period fit does not necessarily correspond to better performance in the post-period.

6 Conclusion

One of the major impacts of the SC method has been to recast longitudinal comparative case studies as prediction problems. In this paper, we made use of two tools from the machine learning and economic forecasting literature: model averaging and rolling-origin forecast evaluation. By examining the weakness of the SC estimator to interpolation bias, and the weakness of the matching estimator to extrapolation bias, we showed how to use these tools to build a third estimator, the MASC, that is able to effectively avoid both sources of bias. Using both simulated and empirical placebo studies, we found evidence that MASC performs better than either the matching, SC, or penalized SC estimators. We used the MASC estimator to re-examine Abadie and Gardeazabal's (2003) application to the economic costs of conflict in the Basque Country and found noticeably larger effect estimates than with SC.

We have not discussed the delicate issue of statistical inference in comparative case studies. A variety of inferential methods have been recently proposed for SC and related methods. Li (2019) and Chernozhukov et al. (2020) develop asymptotic methods that depend on having many pre– and/or post– periods, while Arkhangelsky et al. (2019) propose a jackknife under additional asymptotics in the number of untreated units. Abadie et al. (2010, 2015), Ferman and Pinto (2017), Firpo and Possebom (2018), Chernozhukov et al. (2019) and Shaikh and Toulis (2019) develop different types of non-asymptotic randomization tests, while Cattaneo et al. (2019) show how to construct prediction intervals using non-asymptotic bounds.

References

Abadie, A. (2019): "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, forthcoming. 2, 9, 16

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of Californias Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505. 1, 5, 8, 9, 10, 11, 18, 32, 66

——— (2015): "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59, 495–510. 1, 5, 8, 10, 17, 18, 32

Abadie, A. and J. Gardeazabal (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country," *The American Economic Review*, 93, 113–132. 1, 2, 5, 8, 9, 21, 22, 23, 29, 32, 66, 67

Abadie, A. and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267. 12, 20, 53, 56

Abadie, A. and J. L'Hour (2019): "A Penalized Synthetic Control Estimator for Disaggregated Data," *Working paper.* 1, 2, 9, 15, 16, 20, 21

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2019): "Synthetic Difference in Differences," *arXiv:1812.09970 [stat].* 32

Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2018): "Matrix Completion Methods for Causal Panel Data Models," Working Paper 25132, National Bureau of Economic Research. 3

Athey, S., M. Bayati, G. Imbens, and Z. Qu (2019): "Ensemble Methods for Causal Effects in Panel Data Settings," *AEA Papers and Proceedings*, 109, 65–70. 3

Bang, H. and J. M. Robins (2005): "Doubly Robust Estimation in Missing Data and Causal Inference Models," *Biometrics*, 61, 962–973. 21

Barnow, B. S., G. G. Cain, A. S. Goldberger, et al. (1980): *Issues in the Analysis of Selectivity Bias*, University of Wisconsin, Inst. for Research on Poverty. 5

Becker, M. and S. Klößner (2017): "Estimating the Economic Costs of Organized Crime by Synthetic Control Methods," *Journal of Applied Econometrics*, 32, 1367–1369. 19

——— (2018): "Fast and Reliable Computation of Generalized Synthetic Controls," *Econometrics and Statistics*, 5, 1–19. 19

Ben-Michael, E., A. Feller, and J. Rothstein (2019): "The Augmented Synthetic Control Method," *arXiv:1811.04170 [econ, stat]*. 11

Bergmeir, C. and J. M. Benítez (2012): "On the Use of Cross-Validation for Time Series Predictor Evaluation," *Information Sciences*, 191, 192–213. 17

Biau, G. and L. Devroye (2015): *Lectures on the Nearest Neighbor Method*, vol. 246, Springer. 21

Breiman, L. (1996): "Stacked Regressions," *Machine Learning*, 24, 49–64. 2, 3, 14

Cattaneo, M. D., Y. Feng, and R. Titiunik (2019): "Prediction Intervals for Synthetic Control Methods," *arXiv:1912.07120 [econ, stat]*. 32

Chernozhukov, V., K. Wuthrich, and Y. Zhu (2019): "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," *arXiv:1712.09089 [econ, stat]*. 32

——— (2020): "Practical and Robust t-Test Based Inference for Synthetic Control and Related Methods," *arXiv:1812.10820 [econ]*. 28, 32, 66

Cover, T. (1968): "Estimation by the Nearest Neighbor Rule," *IEEE Transactions on Information Theory*, 14, 50–55. 12

Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, OUP Oxford. 53, 57, 58, 59

Dehejia, R. H. and S. Wahba (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062. 12

Doudchenko, N. and G. W. Imbens (2016): "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis," Working Paper 22791, National Bureau of Economic Research. 3, 10

Elliot, G. (2011): "Averaging and the Optimal Combination of Forecasts," *Working Paper*. 3

FAN, J. AND I. GIJBELS (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20, 2008–2036. 12

FERMAN, B. (2020): "On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls," *arXiv:1906.06665 [econ]*. 10, 20, 21

FERMAN, B. AND C. PINTO (2017): "Revisiting the Synthetic Control Estimator," *Working paper.* 32

——— (2019): "Synthetic Controls with Imperfect Pre-Treatment Fit," *arXiv:1911.08521 [econ]*. 10, 11, 28, 66

FIRPO, S. AND V. POSSEBOM (2018): "Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets," *Journal of Causal Inference*, 6. 32

GOBILLON, L. AND T. MAGNAC (2016): "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 98, 535–551. 10, 67

HANSEN, B. E. AND J. S. RACINE (2012): "Jackknife Model Averaging," *Journal of Econometrics*, 167, 38–46. 2

HASTIE, T., R. TIBSHIRANI, AND J. H. FRIEDMAN (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, New York, NY: Springer, 2nd ed ed. 18

HECKMAN, J. J., H. ICHIMURA, AND P. TODD (1998): "Matching as an Econometric Evaluation Estimator," *The Review of Economic Studies*, 65, 261–294. 5

Heckman, J. J., H. Ichimura, and P. E. Todd (1997): "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64, 605–654. 5

Heckman, J. J. and R. Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and B. Singer, Cambridge University Press. 5

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *The Review of Economics and Statistics*, 86, 4–29. 5, 9

——— (2015): "Matching Methods in Practice: Three Examples," *Journal of Human Resources*, 50, 373–419. 5

Imbens, G. W. and D. B. Rubin (2015): *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press. 5

Li, K. T. (2019): "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods," *Journal of the American Statistical Association*, 1–16. 21, 32

Rosenbaum, P. R. and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55. 5

Shaikh, A. and P. Toulis (2019): "Randomization Tests in Observational Studies with Staggered Adoption of Treatment," *arXiv:1912.10610 [stat]*. 32

Stock, J. H. and M. W. Watson (2004): "Combination Forecasts of Output Growth in a Seven-Country Data Set," *Journal of Forecasting*, 23, 405–430. 3

——— (2006): "Chapter 10 Forecasting with Many Predictors," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann, Elsevier, vol. 1, 515–554. 3

Swanson, N. R. and H. White (1997): "Forecasting Economic Time Series Using Flexible versus Fixed Specification and Linear versus Nonlinear Econometric Models," *International Journal of Forecasting*, 13, 439–461. 17

Tashman, L. J. (2000): "Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review," *International Journal of Forecasting*, 16, 437–450. 2, 17

Viviano, D. and J. Bradic (2019): "Synthetic Learner: Model-Free Inference on Treatments over Time," *arXiv:1904.01490 [cs, econ, stat]*. 3

Wolpert, D. H. (1992): "Stacked Generalization," *Neural Networks*, 5, 241–259. 2

Xu, Y. (2017): "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models," *Political Analysis*, 25, 57–76. 10

Figure 1: Extrapolation vs. interpolation bias

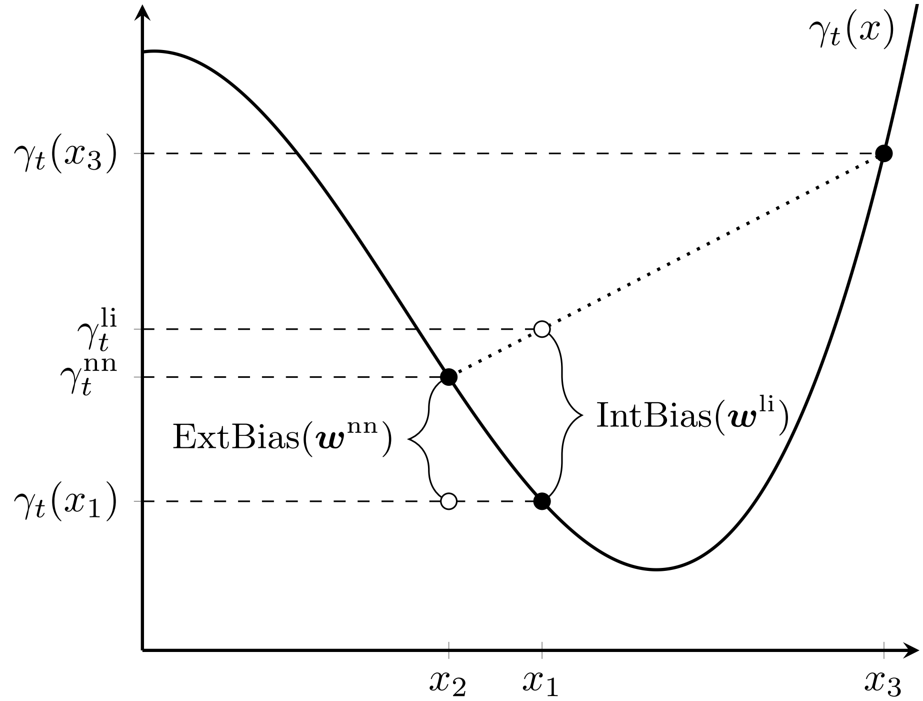
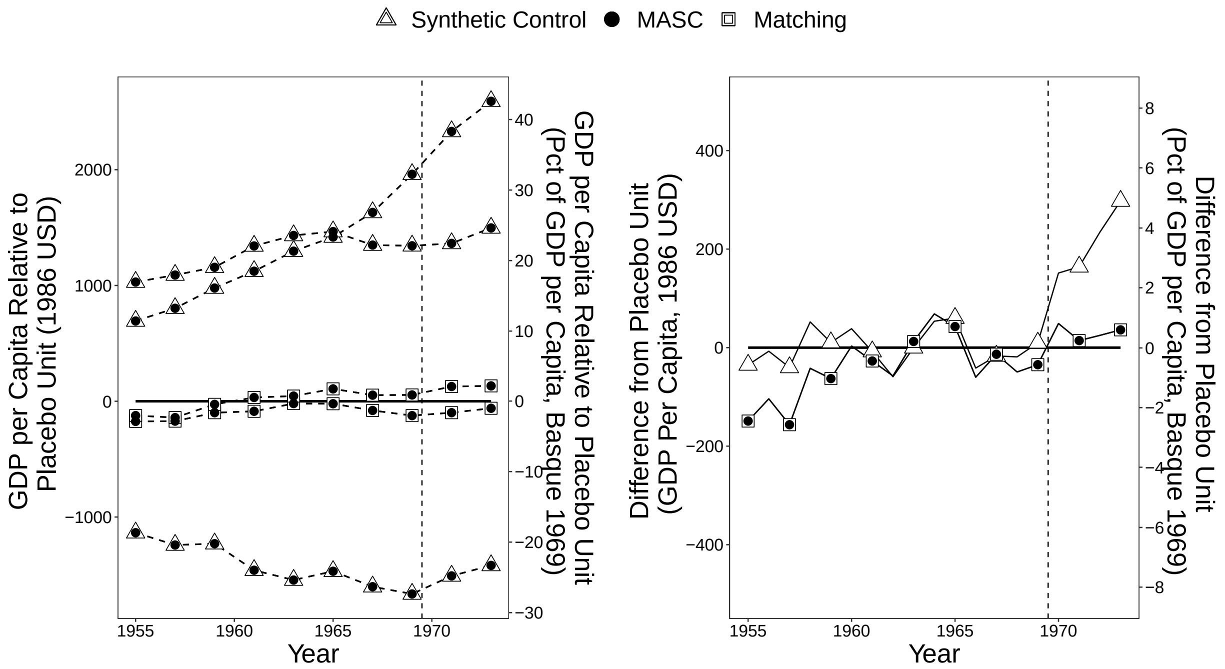


Figure 2: The potential for interpolation bias with the synthetic control estimator

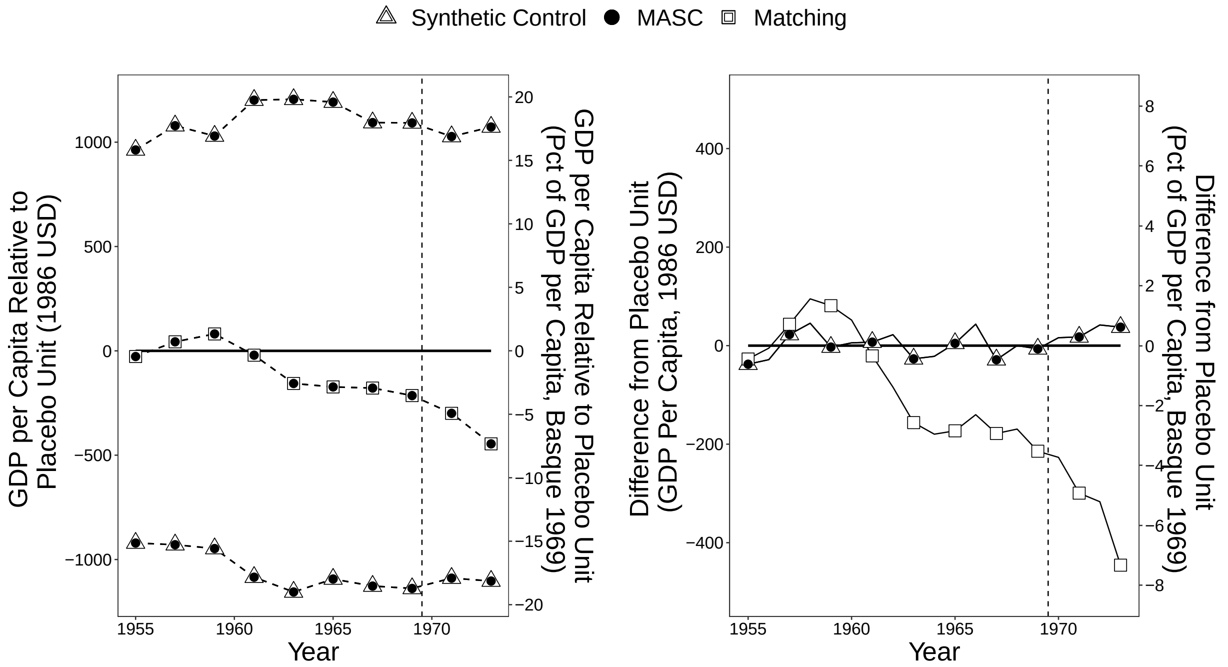


(a) Untreated outcome paths for selected regions.

(b) Estimated outcome paths for the placebo treated region.

Notes: This figure depicts a simulation draw from the empirical Monte Carlo described in Section 4.6, using Asturias as the placebo. The vertical dashed line indicates the beginning of the treatment period. Panel (a) depicts selected control regions which are assigned a weight of at least 0.1 by one of the estimators. Markers indicate which of these controls are assigned weight in the different estimators. Panel (a) plots each estimator using the same markers as in panel (b). In panel (a), paths of GDP per capita over time are plotted relative to the placebo treated region, which is the path lying on the y-intercept. Panel (b) plots the fit and treatment effects (prediction error) for each estimator.

Figure 3: The potential for extrapolation bias with matching

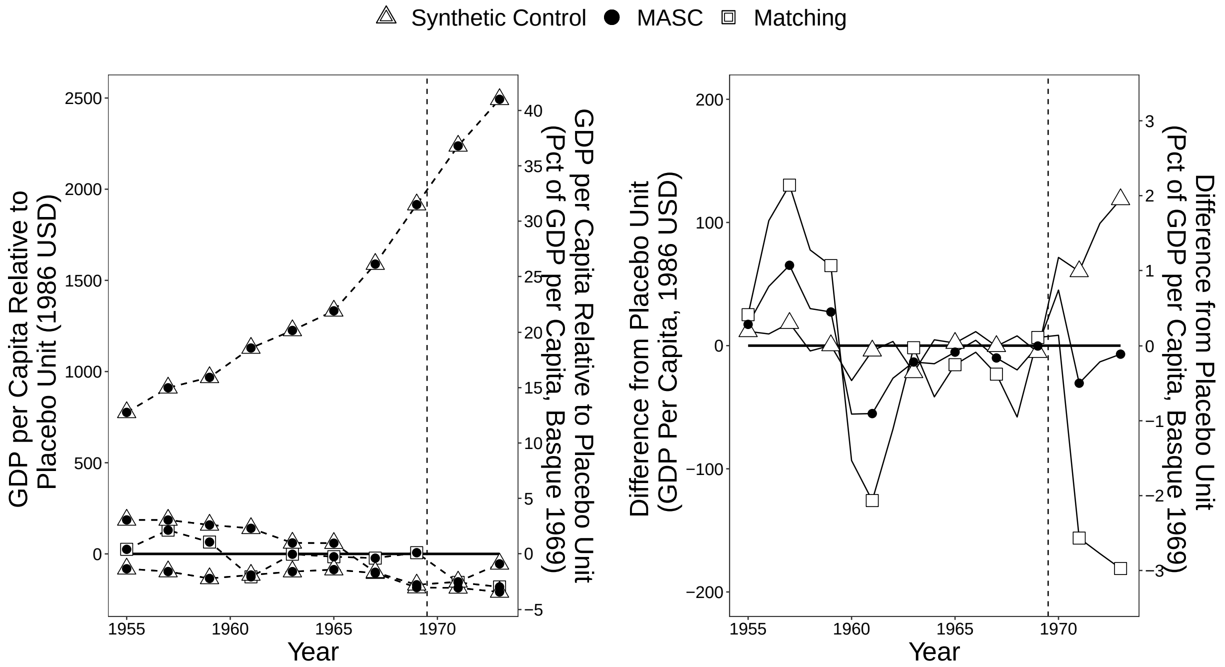


(a) Untreated outcome paths for selected regions.

(b) Estimated outcome paths for the placebo treated region.

Notes: This figure depicts a simulation draw from the empirical Monte Carlo described in Section 4.6, using Navarre as the placebo. The vertical dashed line indicates the beginning of the treatment period. Panel (a) depicts selected control regions which are assigned a weight of at least 0.1 by one of the estimators. Markers indicate which of these controls are assigned weight in the different estimators. Panel (a) plots each estimator using the same markers as in panel (b). In panel (a), paths of GDP per capita over time are plotted relative to the placebo treated region, which is the path lying on the y-intercept. Panel (b) plots the fit and treatment effects (prediction error) for each estimator.

Figure 4: MASC adapts to control both extrapolation and interpolation bias

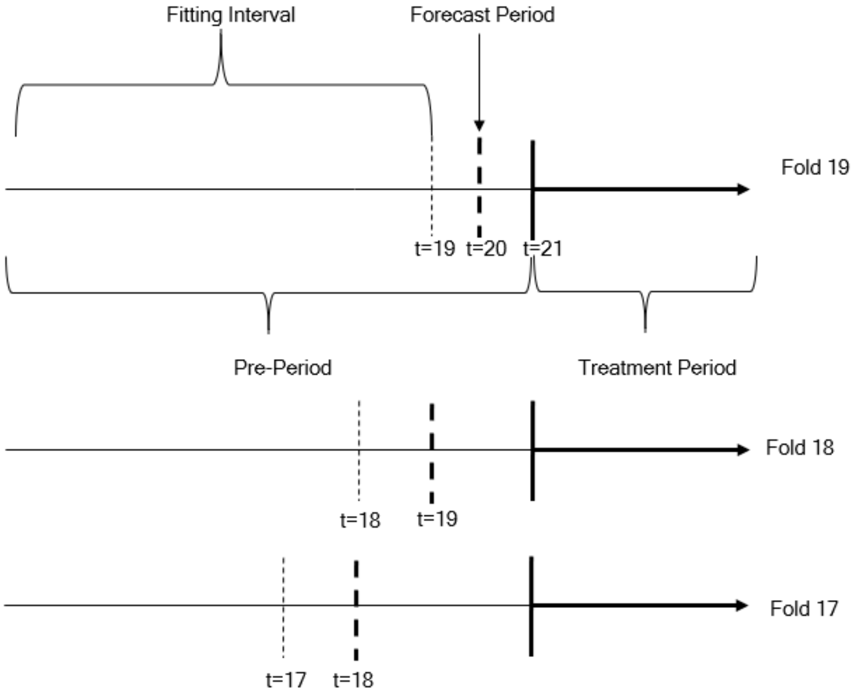


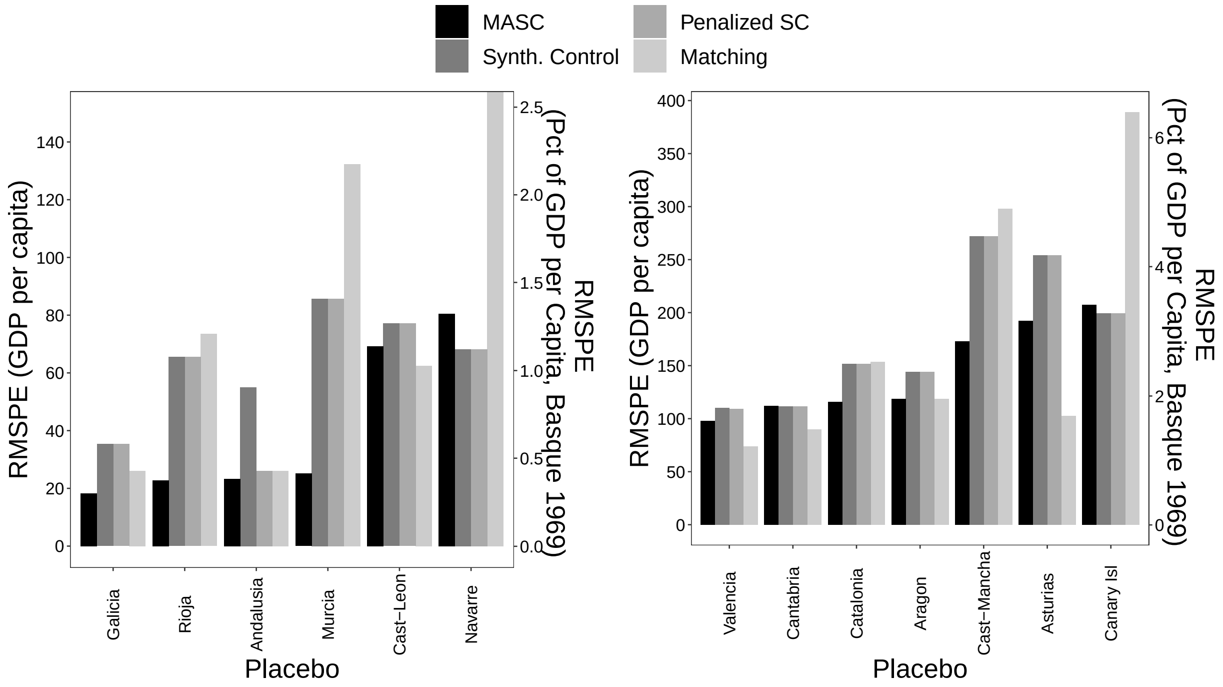
(a) Untreated outcome paths for selected regions.

(b) Estimated outcome paths for the placebo treated region.

Notes: This figure depicts a simulation draw from the empirical Monte Carlo described in Section 4.6, using Rioja as the placebo. The vertical dashed line indicates the beginning of the treatment period. Panel (a) depicts selected control regions which are assigned a weight of at least 0.1 by one of the estimators. Markers indicate which of these controls are assigned weight in the different estimators. Panel (a) plots each estimator using the same markers as in panel (b). In panel (a), paths of GDP per capita over time are plotted relative to the placebo treated region, which is the path lying on the y-intercept. Panel (b) plots the fit and treatment effects (prediction error) for each estimator.

Figure 5: Cross-validation based on rolling-origin recalibration.



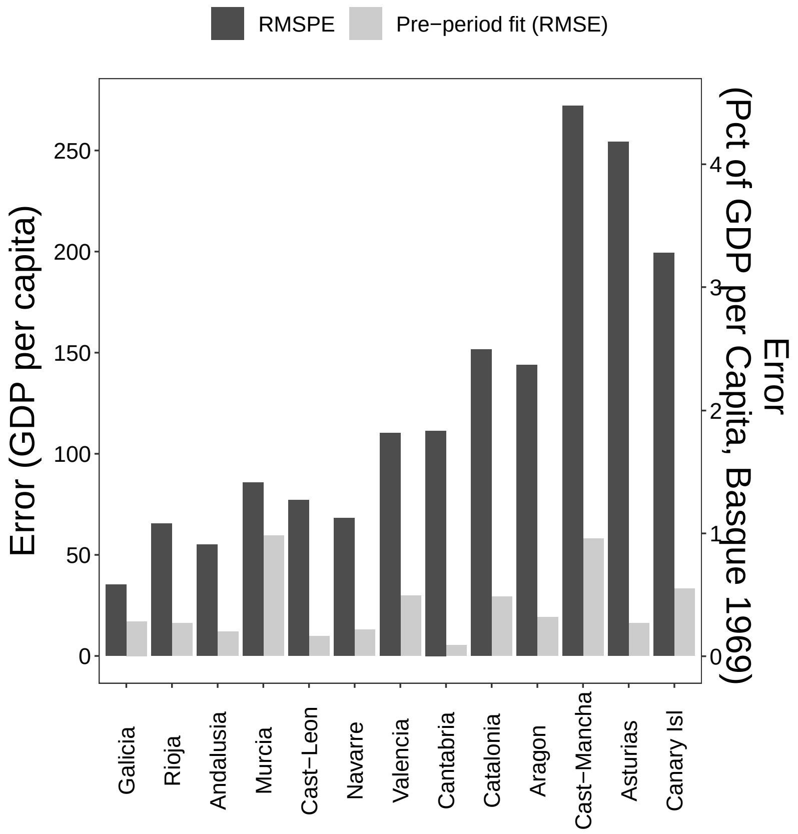


(a) Regions with smaller prediction errors

(b) Regions with larger prediction errors

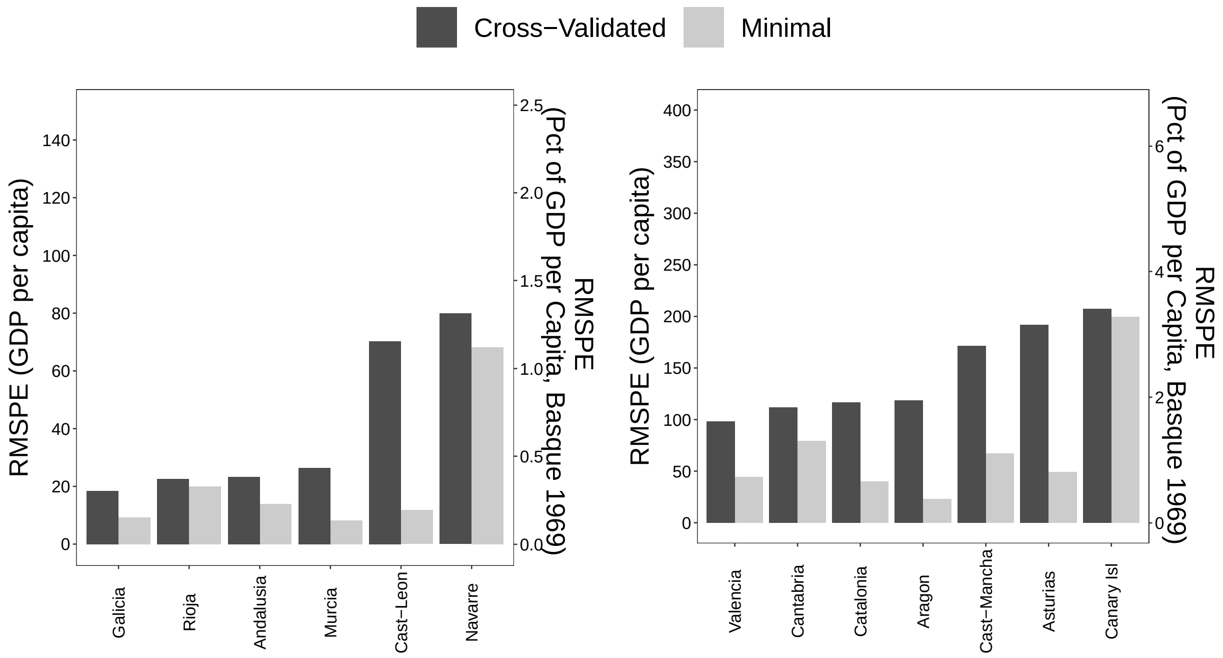
Note: This plot depicts the root means square prediction error from 1970 to 1973 in the 13 main placebo regions. In Navarre, matching returns a MSPE of $498 per person. GDP per capita is measured in 1986 US dollars.

Figure 6: Performance of alternative estimators in the main Spanish placebo analyses



Note: This plot compares the root means square prediction error from 1970 to 1973 to the pre-period fit (root mean square error) in the 13 main placebo regions. GDP per capita is measured in 1986 US dollars.

Figure 7: Pre-period fit and the performance of SC in the main Spanish placebo analyses

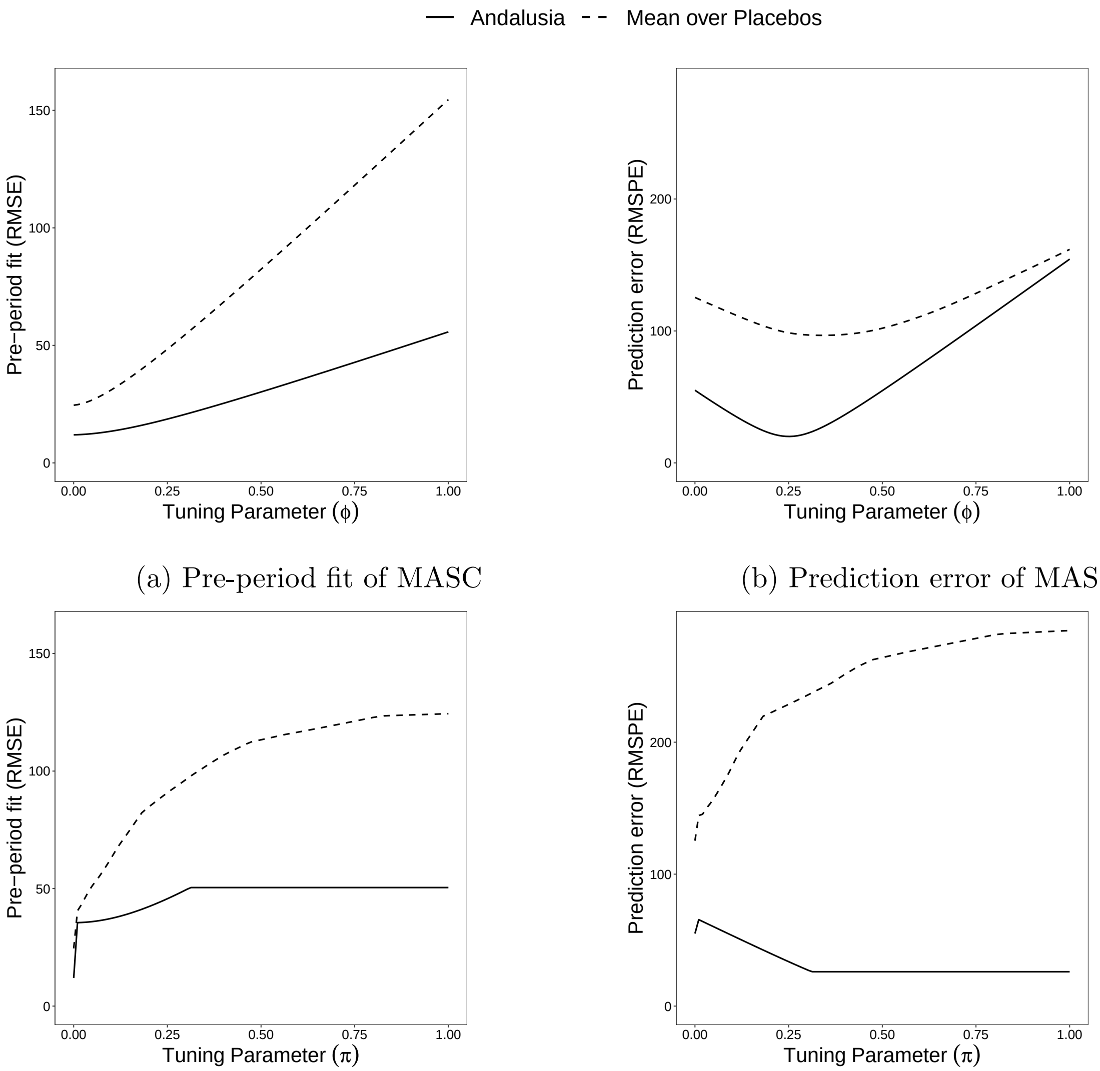


(a) Regions with smaller prediction errors

(b) Regions with larger prediction errors

Note: These plots depict the root means square prediction error from 1970 to 1973 for MASC in the 13 main placebo regions. The darker color plots the value reached by cross-validation. The lighter color plots the minimum value achieved by a control group for MASC (as defined by the tuning parameters ϕ and m). GDP per capita is measured in 1986 US dollars.

Figure 8: Cross-validated and minimum achievable prediction error of MASC in the main Spanish placebo analyses



(a) Pre-period fit of MASC

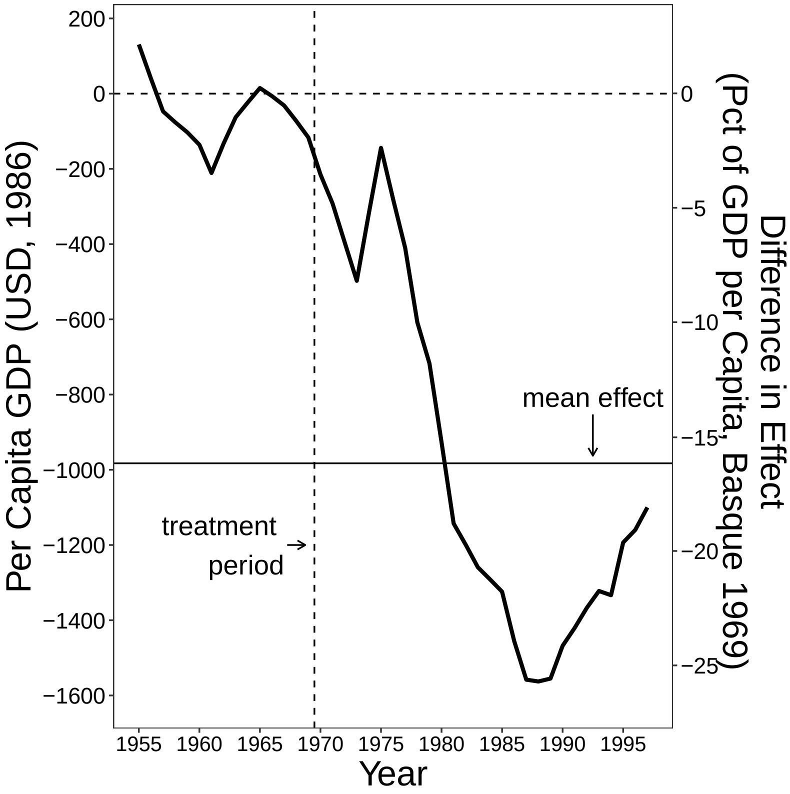
(b) Prediction error of MASC

(c) Pre-period fit of Penalized SC

(d) Prediction error of Penalized SC

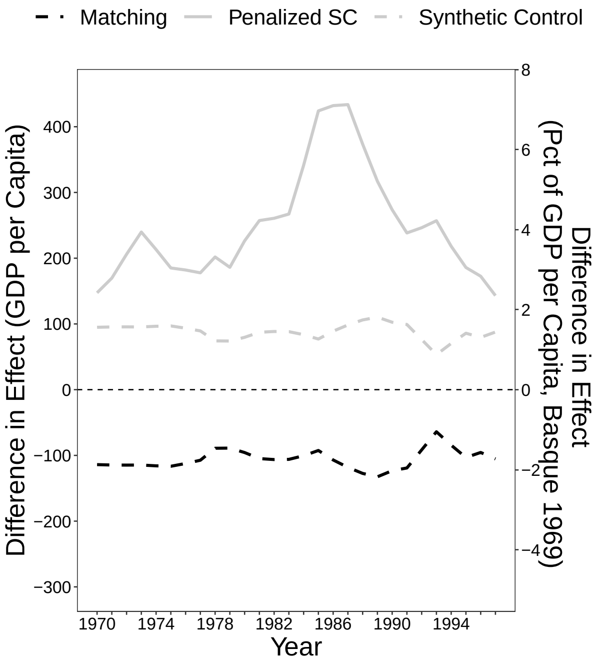
Note: In each graph, a line shows how error (prediction error or pre-period fit) evolves for Andalusia (the solid line) and on average over the 13 placebo regions (the dashed line) as we move from synthetic controls toward matching for the given estimator. Outcomes are measured in 1986 US dollars. For each placebo, the matching estimator (m) is fixed at the value selected by cross-validation. For Andalusia, $m = 6$.

Figure 9: The trade-off between pre-period fit and prediction error for MASC and penalized SC



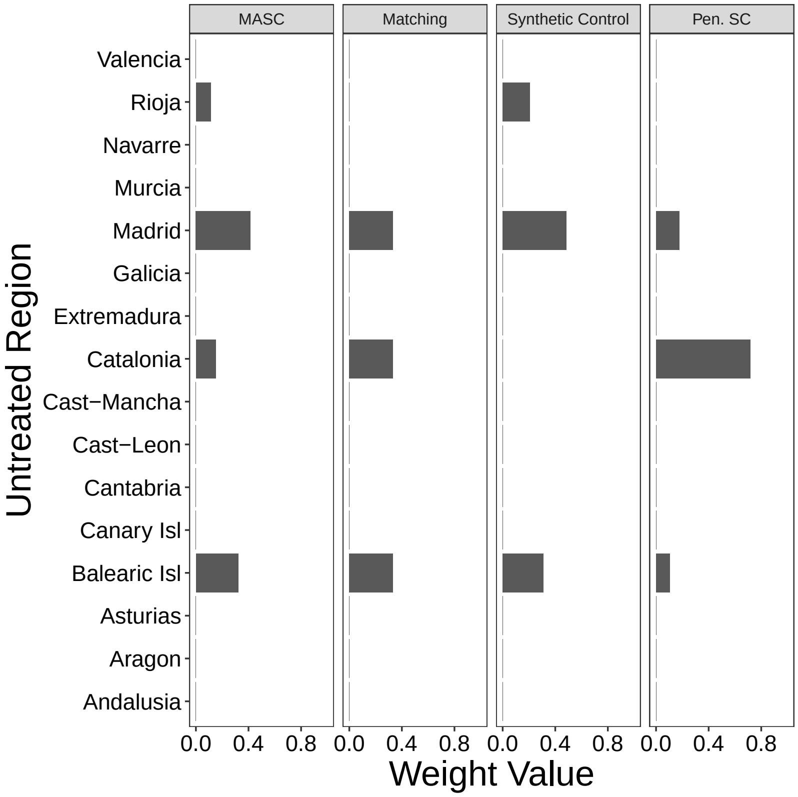
Note: This figure reports the year-to-year pre-period fit of the MASC estimator (to the left of the dashed line) and the post-terrorism estimated treatment effect of the MASC estimator (to the right of the dashed line). The vertical dashed line is drawn halfway between 1969 and 1970 to indicate the beginning of the placebo treatment period in 1970. The solid horizontal line denotes the average yearly per capita cost of terrorism implied by MASC ($982 per person per year).

Figure 10: Pre-terrorism fit and post-period estimated annual costs of terrorism for the MASC estimator



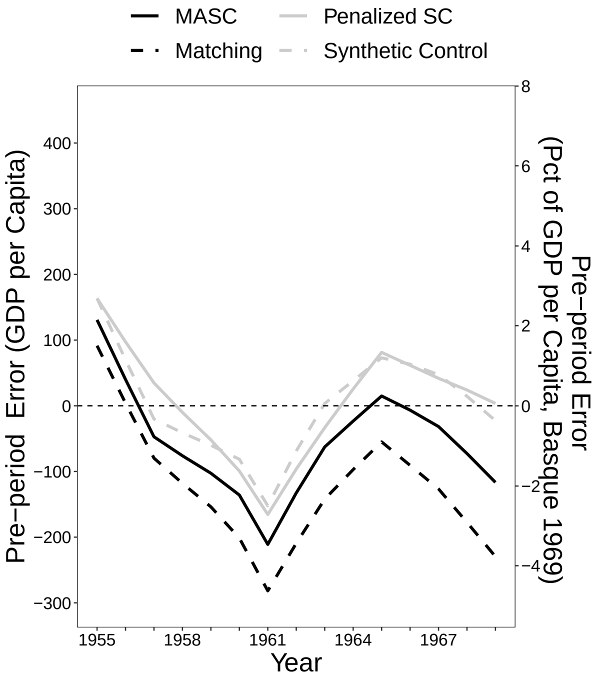
Note: this figure plots the difference between each estimator and MASC in the year-to-year estimates of the cost of terrorism. Differences are reported in 1986 US dollars. A positive value implies a lower estimate for the cost of terrorism than given by the MASC estimator.

Figure 11: Difference from MASC in the cost of terrorism produced by alternative estimators



Note: Each panel if this plot reports the weights of the untreated regions that make up the control group of the given estimator. These weights sum to 1 for each estimator.

Figure 12: Weights on the untreated regions making up the control group for each estimator



Note: GDP per capita is measured in 1986 US dollars. A negative value indicates that the Basque Country's real outcome path lies below the estimator's control group.

Figure 13: Pre-period error of control groups for each estimator

A Technical Appendix

In this section, we discuss the assumptions needed for Propositions 1 and 2 and provide proofs of these results. To help state the assumptions precisely, we let $\boldsymbol{X}_{it}$ denote the predetermined covariates used to construct the forecast at time t or later. In this notation, $\boldsymbol{X}_i$ in the main text is $\boldsymbol{X}_{it^\star}$ and fold f uses $\boldsymbol{X}_{i\bar{t}_f}$ to forecast at $\bar{t}_f + 1$. For clarity, we denote stochastic counterparts of estimators and weights in repeated samples using capital letters to distinguish them notationally from their realizations in the data set at hand. So, for example $\hat{\Gamma}^{\text{masc}}$, $\hat{\Gamma}^{\text{sc}}$, and $\hat{\boldsymbol{W}}^{\text{sc}}$ are random variables, whereas $\hat{\gamma}^{\text{masc}}$, $\hat{\gamma}^{\text{sc}}$, and $\hat{\boldsymbol{w}}^{\text{sc}}$ are their realizations in the sample $\{(y_{i1}, \ldots, y_{iT}, d_i, \boldsymbol{x}_i)\}_{i=1}^{n+1}$ considered in the main text. Also, to simplify some of the subscripts, we let $t_f^\star \equiv \bar{t}_f + 1$ denote the one-step ahead forecasting period for fold f.

A.1 (Equal-length folds) $\bar{t}_f - \underline{t}_f = t_0$ for all $f = 1, \ldots, F$.

A.2 (Stationarity) $(Y_{1t}(0), \boldsymbol{X}_{1t}, \ldots, Y_{nt}(0), \boldsymbol{X}_{nt})$ is second-order stationary.

A.3 (Finite matching, includes nearest neighbor) $\mathcal{M}$ is finite and $1 \in \mathcal{M}$.

A.4 (Independent sampling) $U_{it^\star}$ is independent of $U_{jt^\star}$ and $\boldsymbol{X}_j$ for $i \neq j$.

A.5 (Smoothness) $\gamma_{t^\star}(\boldsymbol{x})$ is Lipschitz in $\boldsymbol{x}$ with Lipschitz coefficient c.

A.6 (Pre-treatment characteristics) $\boldsymbol{X}_i$ is continuously distributed with density that is bounded and bounded away from zero. Its support is compact and convex.

A.7 (Compact parameter space) The support of $Y_{it}(0)$ is contained in the compact set $\mathcal{Y}$ for all i and t.

A.8 (Serial dependence) For each $i = 1, \ldots, n$, $\{Y_{it}, \boldsymbol{X}_{it}\}_{t=-\infty}^{T}$ is α–mixing.

Assumption A.1 simplifies the arguments in Proposition 1 since, together with Assumption A.2, it implies that each cross-validation fold has the same distribution. (Assumption A.1 does not exactly match our empirical implementation, since there we set $\bar{t}_f = 1$ for all f in order to exploit more of the data. However, this assumption greatly helps simplify the accounting without affecting the basic properties of the cross-validation procedure.) In contrast, Assumption A.2 is both a crucial and natural requirement for our cross-validation procedure to succeed. It can be weakened to only require the cross-validation criterion to be stationary; for clarity, we state the stronger sufficient condition that the data is second-order stationary as a whole.

Assumption A.3 requires that the nearest neighbor (matching with $m = 1$) estimator is a feasible outcome of the cross-validation procedure. We use this to bound the MSPE of the MASC by the MSPE of the nearest neighbor estimator in Proposition 1. Assumptions A.4, A.5, and A.6 are then used to bound the MSPE of the nearest neighbor estimator. This part utilizes results by Abadie and Imbens (2006), and in particular their Lemma 2, which provides a bound on the bias of the nearest neighbor estimator.

Assumption A.7 is a standard regularity condition. Assumption A.8 is a condition on serial dependence used in Proposition 2 to ensure that laws of large numbers can be applied. Lower-level sufficient conditions are well-studied, see e.g. Davidson (1994, Section 14).

Proof of Proposition 1. Let $\hat{\Gamma}_f^{\text{masc}}$ denote the MASC estimator on fold f using the same tuning parameters as $\hat{\Gamma}^{\text{masc}}$. That is,

$$\hat{\Gamma}^{\text{masc}}_f = \hat{\Phi}^\star \hat{\Gamma}^{\text{ma}}_f(\hat{M}^\star) + (1 - \hat{\Phi}^\star)\hat{\Gamma}^{\text{sc}}_f,$$

where $\hat{\Gamma}^{\text{ma}}_f(m)$ and $\hat{\Gamma}^{\text{sc}}_f$ are the matching and synthetic control estimators on fold f, and where $(\hat{\Phi}^\star, \hat{M}^\star)$ jointly minimize

$$\text{CV}(\phi, m) \equiv \frac{1}{F}\sum_{f=1}^{F}\left(Y_{1t^\star_f} - \left(\phi\hat{\Gamma}^{\text{ma}}_f(m) + (1-\phi)\hat{\Gamma}^{\text{sc}}_f\right)\right)^2. \tag{17}$$

Since all folds and the final training set have the same length t_0 under Assumption A.1, Assumption A.2 implies that

$$\begin{aligned}\mathbb{E}\left[\text{CV}(\hat{\Phi}^\star, \hat{M}^\star)\right] &= \frac{1}{F}\sum_{f=1}^{F}\mathbb{E}\left[\left(Y_{1t^\star_f}(0) - \hat{\Gamma}^{\text{masc}}_f\right)^2\right] \\ &= \mathbb{E}\left[\left(Y_{1t^\star}(0) - \hat{\Gamma}^{\text{masc}}\right)^2\right] \equiv \text{MSE}(\hat{\Gamma}^{\text{masc}}).\end{aligned}$$

Similarly, the average cross-validation criterion evaluated at the nearest neighbor choice of $\boldsymbol{\tau} = (\phi, m) = (1, 1)$ satisfies $\mathbb{E}[\text{CV}(1, 1)] = \text{MSE}(\hat{\Gamma}^{\text{nn}})$, where $\hat{\Gamma}^{\text{nn}} \equiv \hat{\Gamma}^{\text{ma}}(1)$. Since $(1, 1)$ is a feasible choice for the cross-validation problem under Assumption A.3, we have that

$$\text{MSE}(\hat{\Gamma}^{\text{masc}}) = \mathbb{E}\left[\text{CV}(\hat{\Phi}^\star, \hat{M}^\star)\right] \leq \mathbb{E}\left[\text{CV}(1, 1)\right] = \text{MSE}(\hat{\Gamma}^{\text{nn}}).$$

We establish the claim of the proposition by showing that

$$
\begin{aligned}
\operatorname{MSE}(\hat{\Gamma}^{\mathrm{nn}}) &= \mathbb{E}\left[\left(\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \sum_{i\geq 2} \gamma_{t^\star}(\boldsymbol{X}_i)\hat{W}_i^{\mathrm{nn}}\right) + \left(U_{1t^\star} - \sum_{i\geq 2} U_{it}\hat{W}_i^{\mathrm{nn}}\right)\right)^2\right] \\
&= \mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \sum_{i\geq 2} \gamma_{t^\star}(\boldsymbol{X}_i)\hat{W}_i^{\mathrm{nn}}\right)^2\right] + \mathbb{E}\left[\left(U_{1t^\star} - \sum_{i\geq 2} U_{it^\star}\hat{W}_i^{\mathrm{nn}}\right)^2\right] \\
&\quad + 2\,\mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \sum_{i\geq 2} \gamma_{t^\star}(\boldsymbol{X}_i)\hat{W}_i^{\mathrm{nn}}\right)\left(U_{1t^\star} - \sum_{i\geq 2} U_{it^\star}\hat{W}_i^{\mathrm{nn}}\right)\right] \\
&\leq O(n^{-2/k}) + \left(\mathbb{E}[U_{1t^\star}^2] + \max_{i\geq 2}\mathbb{E}[U_{it^\star}^2]\right) + 0. \qquad (18)
\end{aligned}
$$

To see that the third (cross-product) term in (18) vanishes, write it as

$$
\begin{aligned}
&\mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \sum_{i\geq 2} \gamma_{t^\star}(\boldsymbol{X}_i)\hat{W}_i^{\mathrm{nn}}\right)\left(U_{1t^\star} - \sum_{i\geq 2} U_{it^\star}\hat{W}_i^{\mathrm{nn}}\right)\right] \\
&\quad = \mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \sum_{i\geq 2} \gamma_{t^\star}(\boldsymbol{X}_i)\hat{W}_i^{\mathrm{nn}}\right)\mathbb{E}\left[\left(U_{1t^\star} - \sum_{i\geq 2} U_{it^\star}\hat{W}_i^{\mathrm{nn}}\right)\Big|\{\boldsymbol{X}_i\}_{i=1}^{n+1}\right]\right],
\end{aligned}
$$

where the equality follows because $\hat{\boldsymbol{W}}^{\mathrm{nn}}$ is a function of $\{\boldsymbol{X}_i\}_{i=1}^{n+1}$. Under Assumption A.4,

$$
\begin{aligned}
\mathbb{E}\left[U_{1t^\star}|\{\boldsymbol{X}_i\}_{i=1}^{n+1}\right] &= \mathbb{E}[U_{1t^\star}|\boldsymbol{X}_1] = 0 \\
\text{and}\quad \mathbb{E}\left[\sum_{i\geq 2} U_{it^\star}\hat{W}_i^{\mathrm{nn}}|\{\boldsymbol{X}_i\}_{i=1}^{n+1}\right] &= \sum_{i\geq 2}\hat{W}_i^{\mathrm{nn}}\,\mathbb{E}\left[U_{it^\star}|\boldsymbol{X}_i\right] = 0,
\end{aligned}
$$

since $U_{it^\star} \equiv Y_{it^\star} - \gamma_{t^\star}(\boldsymbol{X}_i)$ satisfies $\mathbb{E}[U_{it^\star}|\boldsymbol{X}_i] = 0$ by construction. The second term of (18) follows because $\hat{W}_i^{\mathrm{nn}}$ is 1 for exactly one $i \geq 2$ and zero for all other i, so that under Assumption A.4,

$$
\mathbb{E}\left[\left(U_{1t^\star} - \sum_{i\geq 2} U_{it^\star}\hat{W}_i^{\mathrm{nn}}\right)^2\right] = \mathbb{E}[U_{1t^\star}^2] + \mathbb{E}\left[\left(\sum_{i\geq 2} U_{it^\star}\hat{W}_i^{\mathrm{nn}}\right)^2\right] \leq \mathbb{E}[U_{1t^\star}^2] + \max_{i\geq 2}\mathbb{E}[U_{it^\star}^2].
$$

As for the first term of (18), it can be bounded as

$$\begin{aligned}
&\mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \sum_{i\geq 2}\gamma_{t^\star}(\boldsymbol{X}_i)\hat{W}_i^{\mathrm{nn}}\right)^2\right] \\
&\quad = \mathbb{E}\left[\left(\sum_{i\geq 2}\hat{W}_i^{\mathrm{nn}}\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \gamma_{t^\star}(\boldsymbol{X}_i)\right)\right)^2\right] \\
&\quad \leq \mathbb{E}\left[\sum_{i\geq 2}\hat{W}_i^{\mathrm{nn}}\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \gamma_{t^\star}(\boldsymbol{X}_i)\right)^2\right] \\
&\quad \leq c^2\,\mathbb{E}\left[\sum_{i\geq 2}\hat{W}_i^{\mathrm{nn}}\left\|\boldsymbol{x}_1 - \boldsymbol{X}_i\right\|^2\right] = c^2\,\mathbb{E}\left[\min_{i\geq 2}\left\|\boldsymbol{X}_1 - \boldsymbol{X}_i\right\|^2\right],
\end{aligned}$$

where the first equality uses the fact that $\hat{\boldsymbol{W}}^{\mathrm{nn}} \in \mathcal{S}$, the first inequality is Jensen's, the second inequality follows from Assumption A.5, and the final equality is an implication of the definition of $\hat{\boldsymbol{W}}^{\mathrm{nn}}$. The claim then follows from Lemma 2 of Abadie and Imbens (2006), which shows that under Assumptions A.6,

$$\mathbb{E}\left[\min_{i\geq 2}\left\|\boldsymbol{X}_1 - \boldsymbol{X}_i\right\|^2\right] = \mathbb{E}\left(\mathbb{E}\left[\min_{i\geq 2}\left\|\boldsymbol{X}_1 - \boldsymbol{X}_i\right\|^2\middle|\boldsymbol{X}_1\right]\right) = O(n^{-2/k}).$$

Q.E.D.

Proof of Proposition 2. Recall from the proof of Proposition 1 that $(\hat{\Phi}^\star, \hat{M}^\star)$ minimizes (17) over (ϕ, m), and that the resulting minimand is

$$\tilde{Q}_F(\hat{\Phi}^\star, \hat{M}^\star) \equiv \mathrm{CV}(\hat{\Phi}^\star, \hat{M}^\star) = \frac{1}{F}\sum_{f=1}^{F}\left(Y_{1t_f^\star} - \hat{\Gamma}_f^{\mathrm{masc}}\right)^2,$$

where the first (re-)definition is to emphasize the dependence of $\mathrm{CV}(\hat{\Phi}^\star, \hat{M}^\star)$ on the number of folds, F. Since both the SC and matching estimators are feasible and thus weakly suboptimal for the same problem, it follows that

$$\tilde{Q}_F(\hat{\Phi}^\star, \hat{M}^\star) \leq \min\left\{\tilde{Q}_F(0, m^\dagger), \tilde{Q}_F(1, \tilde{m})\right\}, \tag{19}$$

where $m^\dagger$ is any arbitrary element of $\mathcal{M}$, and $\tilde{m}$ is any deterministic sequence of matches (potentially depending on n) for which the matching estimator is consistent. In Lemma 1, we show that $\tilde{Q}_F(\phi, m)$ converges to

$$q(\phi, m) \equiv \mathbb{E}\left[\left(Y_{1t^\star}(0) - \hat{\Gamma}^{\text{masc}}(\phi, m)\right)^2\right] \equiv \mathbb{E}\left[\left(Y_{1t^\star}(0) - \left(\phi\hat{\Gamma}^{\text{ma}}(m) + (1-\phi)\hat{\Gamma}^{\text{sc}}\right)\right)^2\right],$$

uniformly over ϕ and m. Applying this result to (19), we have that

$$\tilde{Q}_F(\hat{\Phi}^\star, \hat{M}^\star) \leq \min\left\{q(0, m^\dagger), q(1, \tilde{m})\right\} + o_{\mathbb{P}}(1). \tag{20}$$

If $\hat{\gamma}^{\text{sc}}$ is consistent for $\gamma_{t^\star}(\boldsymbol{X}_1)$, so that $\hat{\Gamma}^{\text{sc}} - \gamma_{t^\star}(\boldsymbol{X}_1) = o_{\mathbb{P}}(1)$, then given Assumption A.7, we also know that $\mathbb{E}[(\hat{\Gamma}^{\text{sc}} - \gamma_{t^\star}(\boldsymbol{X}_1))^2] = o(1)$, see e.g. Davidson (1994, Theorem 12.8). Thus,

$$\begin{aligned} q(0, m^\dagger) &= \mathbb{E}\left[\left(\hat{\Gamma}^{\text{sc}} - \gamma_{t^\star}(\boldsymbol{X}_1) + U_{1t^\star}\right)^2\right] \\ &= \mathbb{E}\left[\left(\hat{\Gamma}^{\text{sc}} - \gamma_{t^\star}(\boldsymbol{X}_1)\right)^2\right] + 2\,\mathbb{E}\left[U_{1t^\star}\left(\hat{\Gamma}^{\text{sc}} - \gamma_{t^\star}(\boldsymbol{X}_1)\right)\right] + \mathbb{E}[U_{1t^\star}^2] = \mathbb{E}[U_{1t^\star}^2] + o(1), \end{aligned}$$

where we used $\mathbb{E}[U_{1t^\star}] = \mathbb{E}[U_{1t^\star}\gamma_{t^\star}(\boldsymbol{X}_1)] = 0$, as well as Assumption A.4, which implies that

$$\mathbb{E}[U_{1t^\star}\hat{\Gamma}^{\text{sc}}] = \mathbb{E}\left[\mathbb{E}\left[U_{1t^\star}|\{\boldsymbol{X}_i\}_{i=1}^{n+1}\right]\hat{\Gamma}^{\text{sc}}\right] = \mathbb{E}\left[\mathbb{E}\left[U_{1t^\star}|\boldsymbol{X}_1\right]\hat{\Gamma}^{\text{sc}}\right] = 0, \tag{21}$$

since $\hat{\Gamma}^{\text{sc}}$ is a function of $\{\boldsymbol{X}_i\}_{i\geq 1}$. By an identical argument, $q(1, \tilde{m}) = \mathbb{E}[U_{1t^\star}^2] + o_{\mathbb{P}}(1)$ if $\hat{\gamma}^{\text{ma}}$ is consistent for $\gamma_{t^\star}(\boldsymbol{X}_1)$. Thus, in either case, (20) reduces to

$$\tilde{Q}_F(\hat{\Phi}^\star, \hat{M}^\star) \leq \mathbb{E}[U_{1t^\star}^2] + o_{\mathbb{P}}(1). \tag{22}$$

On the other hand, from Lemma 1, we also know that

$$\begin{aligned}
\tilde{Q}_F(\hat{\Phi}^\star, \hat{M}^\star) &\xrightarrow{\mathbb{P}} \mathbb{E}\left[\left(Y_{1t^\star}(0) - \hat{\Gamma}^{\text{masc}}\right)^2\right] \\
&= \mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \hat{\Gamma}^{\text{masc}}\right)^2\right] + \mathbb{E}[U_{1t^\star}^2] + \mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \hat{\Gamma}^{\text{masc}}\right) U_{1t^\star}\right] \\
&= \mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \hat{\Gamma}^{\text{masc}}\right)^2\right] + \mathbb{E}[U_{1t^\star}^2], \qquad (23)
\end{aligned}$$

where the second equality used Assumption A.4 to conclude that

$$\mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \hat{\Gamma}^{\text{masc}}\right) U_{1t^\star}\right] = -\,\mathbb{E}\left[\hat{\Gamma}^{\text{masc}}\,\mathbb{E}\left[U_{1t^\star}|\{\boldsymbol{X}_i\}_{i=1}^{n+1}\right]\right] = -\,\mathbb{E}\left[\hat{\Gamma}^{\text{masc}}\,\mathbb{E}\left[U_{1t^\star}|\boldsymbol{X}_1\right]\right] = 0,$$

with similar reasoning as in (21). Combining (22) and (23), we have that

$$\mathbb{E}\left[\left(\gamma_{t^\star}(\boldsymbol{X}_1) - \hat{\Gamma}^{\text{masc}}\right)^2\right] \le o_{\mathbb{P}}(1),$$

which, under Assumption A.7, implies that $\hat{\gamma}^{\text{masc}}$ is consistent for $\gamma_{t^\star}(\boldsymbol{x}_1)$, as claimed. *Q.E.D.*

Lemma 1. Let $R_F(\phi, m) \equiv \tilde{Q}_F(\phi, m) - q(\phi, m)$, where

$$\tilde{Q}_F(\phi, m) \equiv \frac{1}{F}\sum_{f=1}^{F}\left(Y_{1t_f^\star} - \hat{\Gamma}_f^{\text{masc}}(\phi, m)\right)^2 \quad \text{and} \quad q(\phi, m) \equiv \mathbb{E}\left[\left(Y_{1t^\star}(0) - \hat{\Gamma}^{\text{masc}}(\phi, m)\right)^2\right].$$

Under Assumptions A.3, A.7, and A.8, $\sup_{m\in\mathcal{M},\phi\in[0,1]} |R_F(\phi, m)| \to_{\mathbb{P}} 0$ as $F \to \infty$.

Proof of Lemma 1. First, we consider the sequence

$$Z_f(\phi, m) \equiv \left(Y_{1t_f^\star} - \hat{\Gamma}_f^{\text{masc}}(\phi, m)\right)^2 - q(\phi, m).$$

Since $\{Y_{it}, \boldsymbol{X}_{it}\}_{t=-\infty}^{T}$ is α–mixing for each i under Assumption A.8, so too is $F^{-1}Z_f(\phi, m)$ (e.g. Davidson, 1994, Theorem 14.1). Because it has zero-mean and is bounded by Assumption A.7, it is also a mixingale, and in particular

$$\left|\frac{1}{F}\,\mathbb{E}\left[Z_f(\phi, m)|\mathcal{G}_{f-s}\right]\right| \le \frac{1}{F}\psi_s,$$

where $\mathcal{G}_t$ is the σ–field generated from $\{(Y_{is}, \boldsymbol{X}_{is}) : i = 1, \ldots, n, s \leq t\}$ and ψ_s is a finite constant that depends only on s, see Davidson (1994, Theorem 14.2). It then follows immediately from Theorem 19.11 of Davidson (1994) that

$$\mathbb{E}\left[|R_F(\phi, m)|\right] = \mathbb{E}\left[\left|\sum_{f=1}^{F} F^{-1} Z_f(\phi, m)\right|\right] \to 0,$$

as $F \to \infty$ for any fixed ϕ and m, so that $|R_F(\phi, m)| \to_{\mathbb{P}} 0$.

We next establish that $R_F(\phi, m)$ is stochastically equicontinuous as a function of ϕ and m. This is clearly true with respect to $m \in \mathcal{M}$, since $\mathcal{M}$ is a finite set under Assumption A.3. To see that it is stochastically equicontinuous over ϕ, recall that

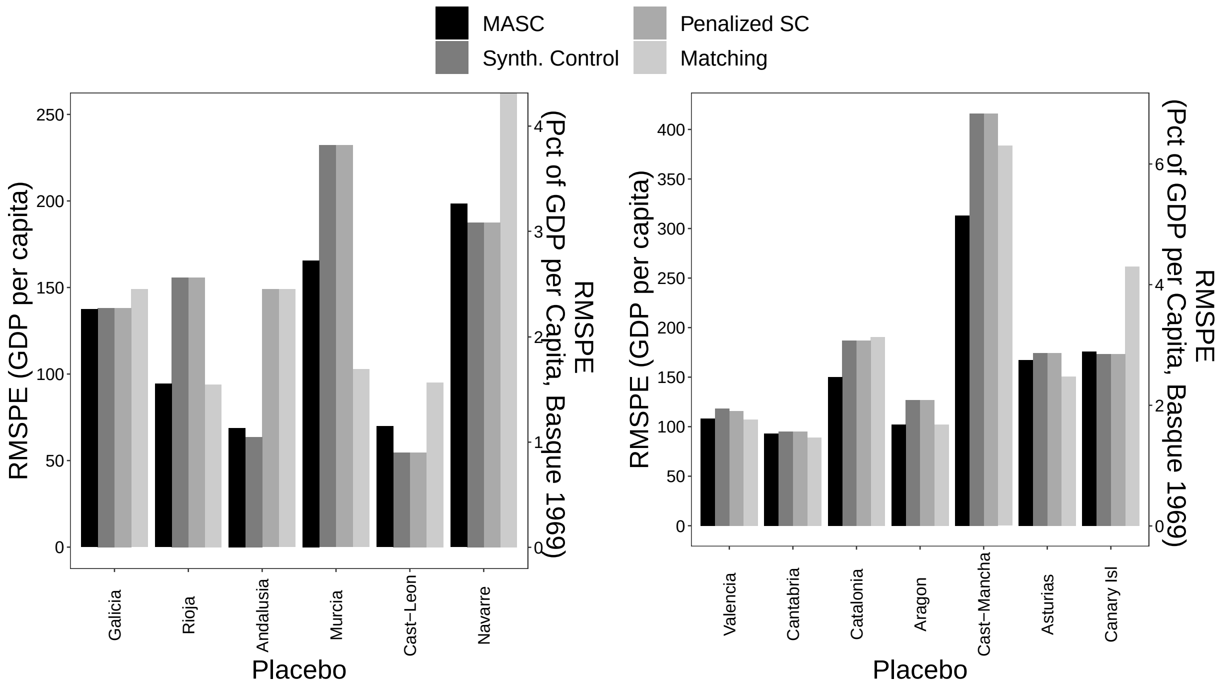
$$\hat{\Gamma}_f^{\text{masc}}(\phi, m) = \phi\hat{\Gamma}_f^{\text{ma}}(m) + (1 - \phi)\hat{\Gamma}^{\text{sc}},$$

so that both $\tilde{Q}(\phi, m)$ and $q(\phi, m)$ are quadratic functions of ϕ. All of the coefficients in these quadratic functions are bounded by virtue of Assumption A.7 and the fact that $\hat{\Gamma}^{\text{ma}}(m)$ and $\hat{\Gamma}^{\text{sc}}$ are composed of convex weights. As a consequence, both $\tilde{Q}(\phi, m)$ and $q(\phi, m)$ are Lipschitz in ϕ over its domain of $[0, 1]$, and thus so too is $R_F(\phi, m)$. Stochastic equicontinuity then follows from e.g. Theorem 21.10 of Davidson (1994). Uniform convergence, that is

$$\sup_{m \in \mathcal{M}, \phi \in [0,1]} |R_F(\phi, m)| \xrightarrow{\mathbb{P}} 0 \quad \text{as } F \to \infty,$$

follows in turn from e.g. Theorem 21.9 of Davidson (1994). *Q.E.D.*

B Additional Results



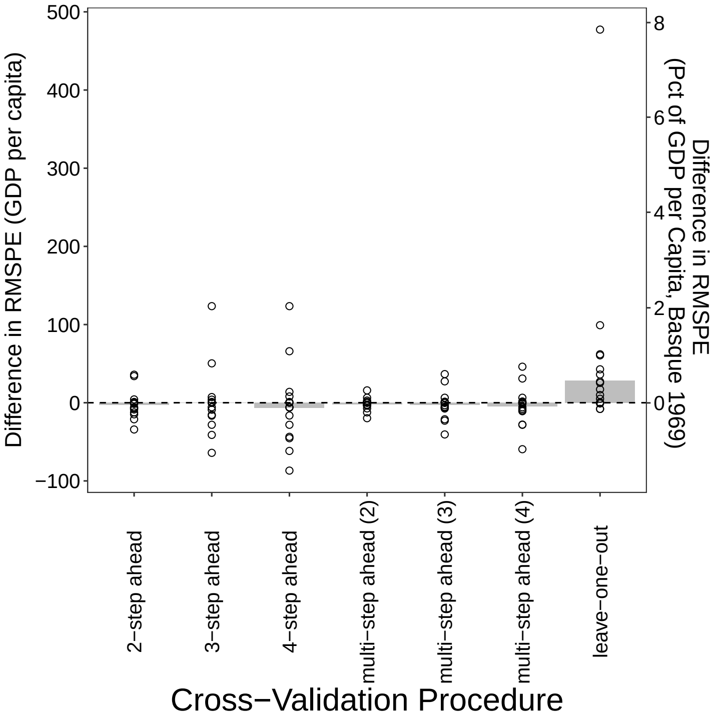
(a) Regions with smaller prediction errors

(b) Regions with larger prediction errors

Note: This plot depicts the root means square prediction error from 1970 to 1979 in the 13 main placebo regions. In Navarre, matching returns a RMSPE of $654.4 per person. GDP per capita is measured in 1986 US dollars.

Figure B.1: Performance of alternative estimators in the main Spanish placebo analyses (10-year prediction error)

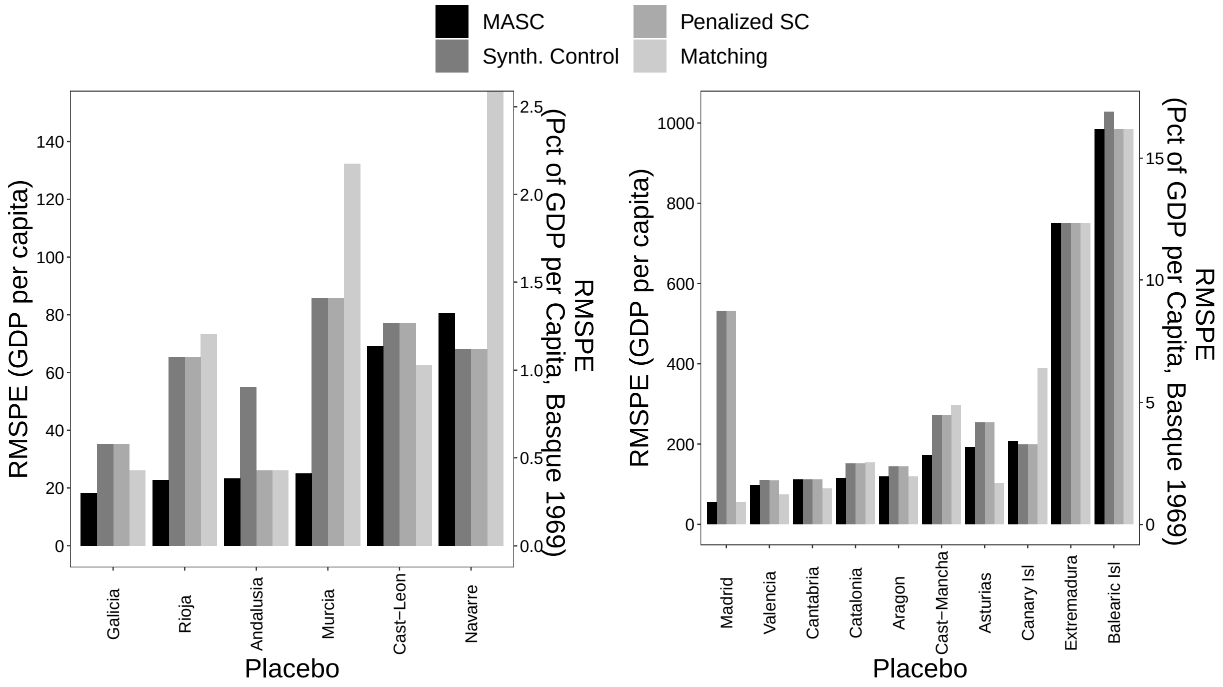
C Alternative Cross-Validation Procedures



Note: Each column of this graph depicts the difference for each of the 13 main placebo regions in the root mean square prediction error (RMSPE) for MASC when using our baseline rolling-origin procedure versus a given alternative cross-validation procedure. Each point represents one of the 13 placebo regions, and the bar represents the average difference across the 13 regions. A positive value indicates that the given cross-validation procedure generates a higher RMSPE than our baseline rolling-origin procedure. K-step ahead procedures forecast, for each fold, only into the kth period ahead. Multi-step ahead procedures forecast, for each fold, from 1 and up to k periods ahead. Leave-one-out predicts the outcome for a single held out point in each fold, and fits the estimator with data before and after the held out point. It uses all pre-period years from 1955 to 1969 as folds, rather than only years from 1962 to 1968. RMSPE is measured from 1970 to 1973, in 1986 US dollars.

Figure C.1: Performance of alternative cross-validation procedures for MASC in the main Spanish placebo analyses, relative to our baseline rolling-origin procedure

D Results Including Regions with Poor SC Fit (Extremadura, Madrid, and the Balearic Islands)

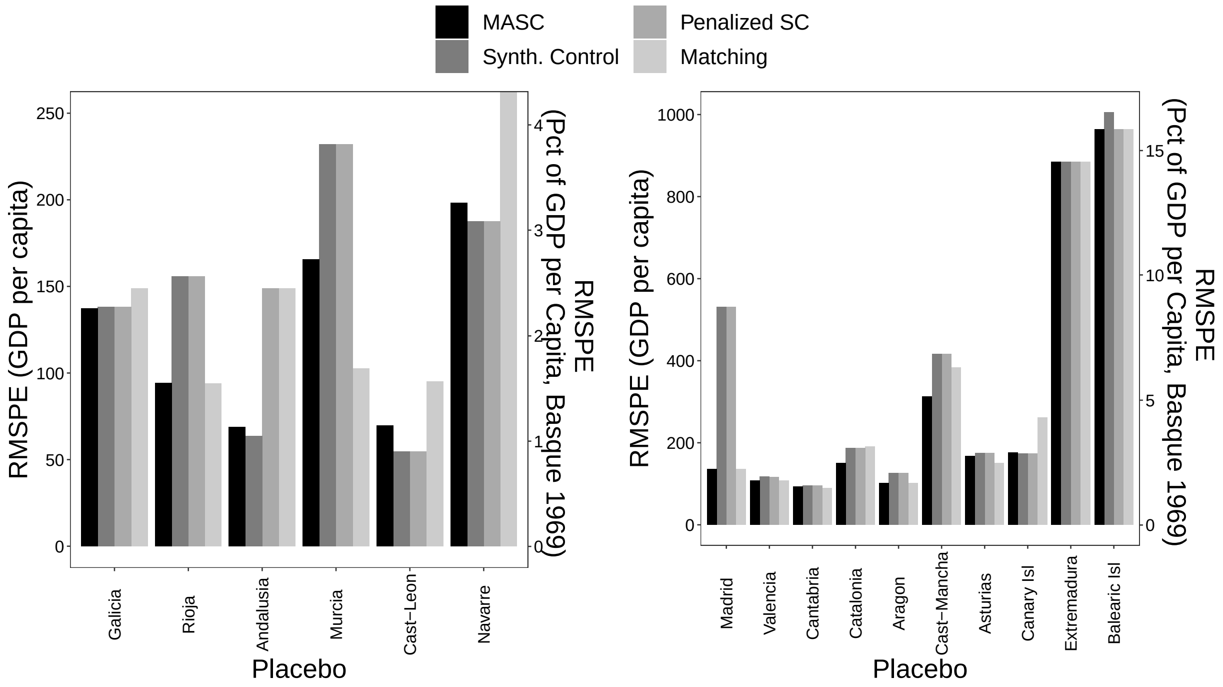


(a) Regions with smaller prediction errors

(b) Regions with larger prediction errors

Note: This plot depicts the root means square prediction error from 1970 to 1973 in all sixteen placebo regions. In Navarre, matching returns a RMSPE of $498 per person. GDP per capita is measured in 1986 US dollars.

Figure D.1: Performance of alternative estimators in all Spanish placebo analyses

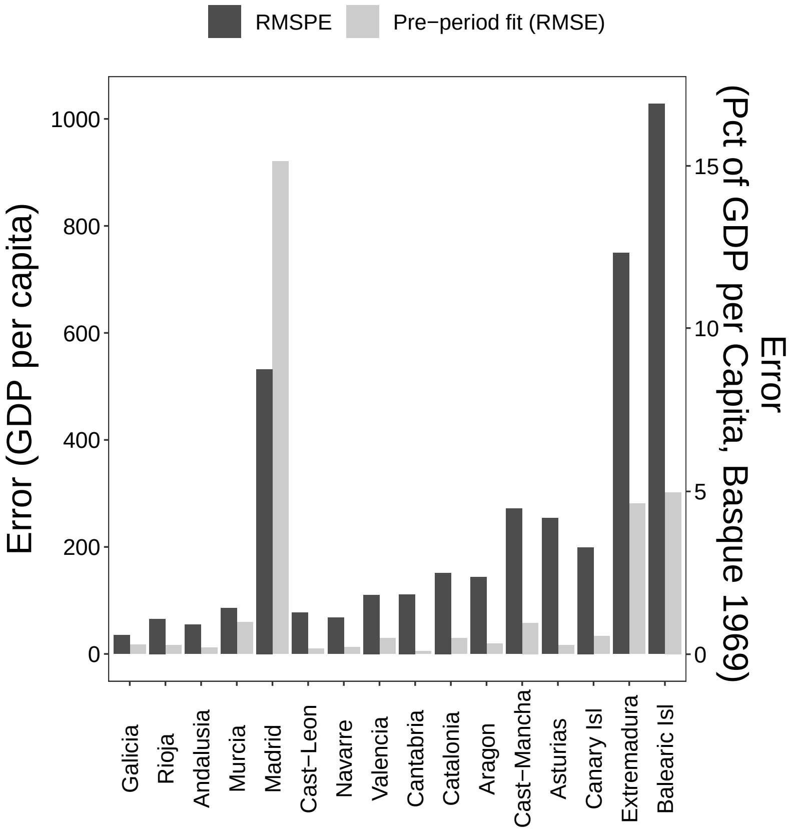


(a) Regions with smaller prediction errors

(b) Regions with larger prediction errors

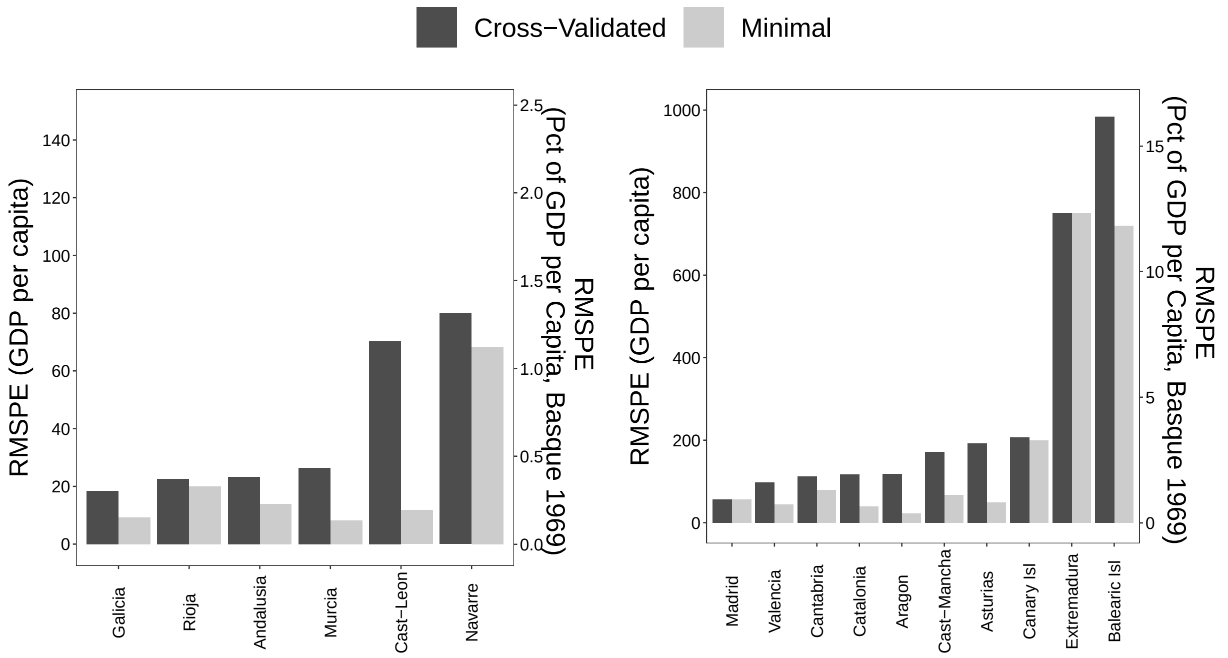
Note: This plot depicts the root means square prediction error from 1970 to 1979 in all sixteen placebo regions. In Navarre, matching returns a RMSPE of $654.4 per person. GDP per capita is measured in 1986 US dollars.

Figure D.2: Performance of alternative estimators in the all Spanish placebo analyses (10-year prediction error)



Note: This plot compares the root means square prediction error from 1970 to 1973 for SC to the pre-period fit (root mean square error) in all sixteen placebo regions. GDP per capita is measured in 1986 US dollars.

Figure D.3: Pre-period fit and the performance of SC in all Spanish placebo analyses



(a) Regions with smaller prediction errors

(b) Regions with larger prediction errors

Note: These plots depict the root means square prediction error from 1970 to 1973 for MASC in all sixteen placebo regions. The darker color plots the value reached by cross-validation. The lighter color plots the minimum value achieved by a control group for MASC (as defined by the tuning parameters ϕ and m). GDP per capita is measured in 1986 US dollars.

Figure D.4: Cross-validated and minimum achievable prediction error of MASC in all Spanish placebo analyses

E Monte Carlo Simulations

E.1 Factor Model of GDP in Regions of Spain

We consider a data generating process consisting of a linear factor model with normally distributed innovations. A previous version of this paper used an autoregressive model with region-specific polynomial time trends; the results do not differ materially. Following Ferman and Pinto (2019) and Chernozhukov et al. (2020), we now consider a factor model of the form

$$\tilde{Y}_{it}(0) = \boldsymbol{\varphi}_t' \boldsymbol{L}_i + Z_{it} \quad \text{where} \quad Z_{it} = \rho_i Z_{i,t-1} + V_{it}, \quad \text{with } V_{it} \sim N(0, \xi_i^2)$$

where $\tilde{Y}_{it}(0)$ is a de-trended outcome measure constructed by removing the period-specific mean from Y_{it}. The innovations V_{it} are uncorrelated across time and regions. Like Chernozhukov et al. (2020), our model includes four latent factors composing the region-specific deterministic trends.

As discussed in Section 2.3, Abadie et al. (2010) argue that (8) holds approximately under the restrictions on potential bias imposed by the linear factor model. Consequently, the SC estimator should suffer little if anything from interpolation bias. Thus, SC should perform at least as well as other estimators provided there exists a suitable synthetic control with a pre-treatment path similar to the path of the placebo region. If no suitable control group exists, alternative estimators may outperform SC.

Using the Abadie and Gardeazabal (2003) Spanish data from 1955 and up to 1973, we fit the factors $\boldsymbol{\varphi}_t$, the factor loadings $\boldsymbol{L}_i$, the region-specific variance in innovations ξ_i^2, and the region-specific autocorrelation coefficient ρ_i. We use the fitted model to redraw the outcome paths for each region of Spain by fixing the factor scores and loadings across simulations and re-drawing region specific innovations, the same as Gobillon and Magnac (2016). The left column of graphs in Figure E.6 show that the fitted model does a good job at faithfully reproducing the patterns in the Spanish data with a bit of sampling error added.

E.2 Multiple Realizations of the Same DGP

The first Monte Carlo exercise addresses whether the relative performance of the four estimators would change when considering multiple realizations of the same data generating process. We use the simulated data to compare estimators through the same type of placebo analyses as in Section 4. The only exception is that we now consider the distribution of RMSPEs across simulation draws for each placebo region. We expect this distribution to be clustered close to zero if the estimator is working well.

The results reinforce four key insights from the placebo analyses in Section 4. First, MASC tends to outperform other estimators. This is apparent in Figure E.1, where we plot the mean, median, 25th percentile, and 75th percentile of the RMSPE for alternative estimators across simulation draws. In panel (a), we report these statistics for the Catalonia placebo, which was the region used by Abadie and Gardeazabal (2003) for their placebo exercise. In Catalonia, the 75th percentile of RMSPE for MASC (\$135) is almost exactly equal to the 25th percentile of RMSPE for SC and penalized SC (\$137), and clearly falls below the 25th percentile of RMSPE for matching (\$148). In panel (b), we report statistics for the average RMSPE across the thirteen primary placebo regions. These still show an edge for MASC, although one that is slightly less pronounced than in Catalonia.

Second, the pre-period fit of the SC estimator is not necessarily a strong indicator of its prediction error. This is apparent in Figure E.2, which shows the expected pre-period fit for

SC is worst in the regions of Murcia and Castile-La Mancha. Yet the expected prediction error is low in Murcia and high in Castile-La Mancha.

Third, MASC more effectively combines matching and synthetic control than penalized SC in this setting. The MASC estimator often assigns positive weight to matching, while the penalized SC estimator tends to behave like the standard SC estimator. This can be seen from Figure E.1, which shows that the distribution of RMSPE for MASC is quite different from those of SC and penalized SC, which are nearly identical.

Fourth and last, the cross-validation procedure tends to do a good job selecting suitable control groups for MASC (as defined by the tuning parameters ϕ and m) when they exist. Figure E.3 shows this by reporting the average RMSPE of the actual parameters chosen by cross-validation against the smallest RMSPE that could be obtained by infeasibly choosing the optimal tuning parameters. Averaging over the thirteen main placebo regions, the minimum expected RMSPE achievable by MASC is 42 percent lower than the actual expected RMSPE achieved by cross-validation. For most regions, it is on average not possible to select a tuning parameter that perform substantially better than the one chosen by cross-validation.

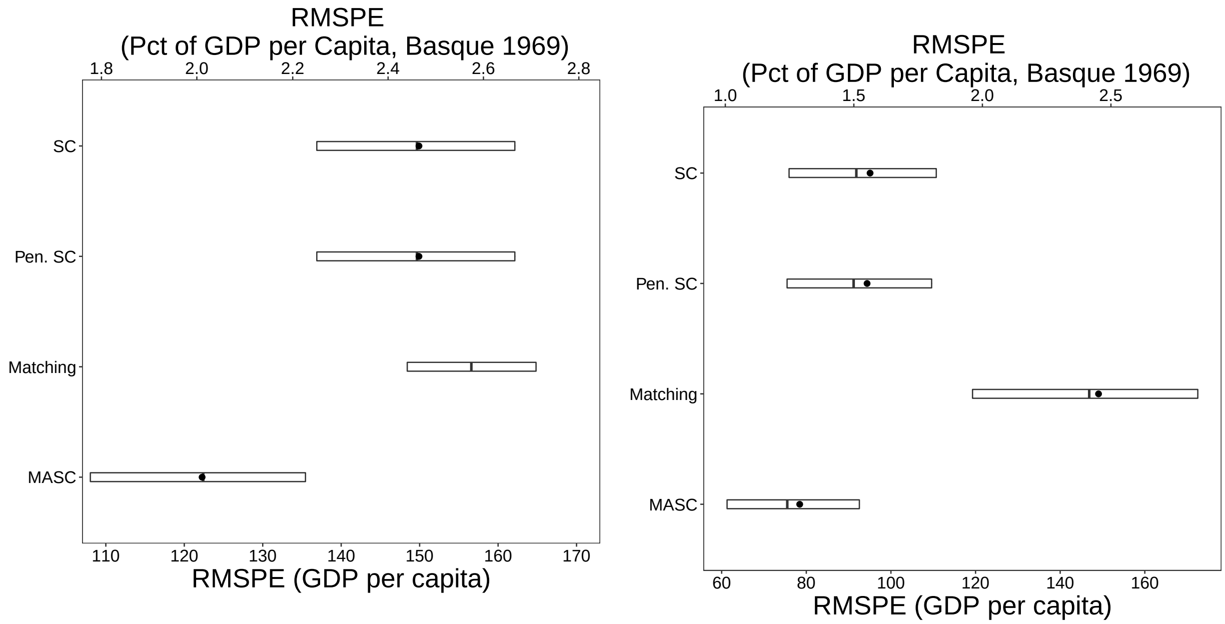
The reason that SC performs poorly in this Monte Carlo turns out to be because four factors is too complex. To illustrate this, we compare the performance of the four estimators when the stochastic component of the factor model is shut down ($Z_{it} = 0$), so that the outcome paths are determined exclusively from $\boldsymbol{\varphi}_t \boldsymbol{L}_i$. In this case, some placebo regions have a synthetic control that fits exactly; these are plotted in panel (a) Figure E.4. As expected, SC picks up the true model in the pre-period for these regions, and thus has zero RMSPE, while matching has errors that range from small to large. The MASC and penalized SC also have zero RMSPE in these regions because they too are able to follow the underlying factor structure into the post-period. In panel (b), we plot placebo regions for which SC does not fit exactly in the pre-period. For these regions, MASC tends to outperform both SC and matching, while penalized SC continues to behave like SC.

E.3 Simplifying the Factor Model

Our second Monte Carlo exercise illustrates conditions under which the SC estimator is likely to be the preferred estimator. We perform the same placebo analyses as in the previous subsection, but now we fit the factor model with two factors instead of four. This exercise is meant to be expository, not necessarily realistic. As shown by the right column of graphs in Figure E.6, the simulated data generated by the two-factor model often looks systematically different from the actual data. With two factors instead of four, there is a perfectly fitting SC for all regions except Castile-La Mancha. In terms of Figure E.4, all regions but this one would be in panel (a).

In this two-factor model, the advantage of SC becomes clear. Figure E.5 shows that SC outperforms all three competitor estimates in all regions which have a close-fitting synthetic control This is because both MASC and penalized SC sometimes erroneously assign weight to the matching estimator when they confuse realizations of the stochastic component of the model with the underlying factor structure. MASC is more susceptible to this than penalized SC, since the latter continues to follow SC closely in this data generating process. Nevertheless, MASC is still competitive with SC in all regions, and performs much better in Castile-La Macha, the one region that lies outside of the convex hull of the others.

Taken together, these simulation results illustrate that one may prefer the SC estimator in settings where one suspects there are a small number of latent factors, so that a suitable synthetic control likely exists.

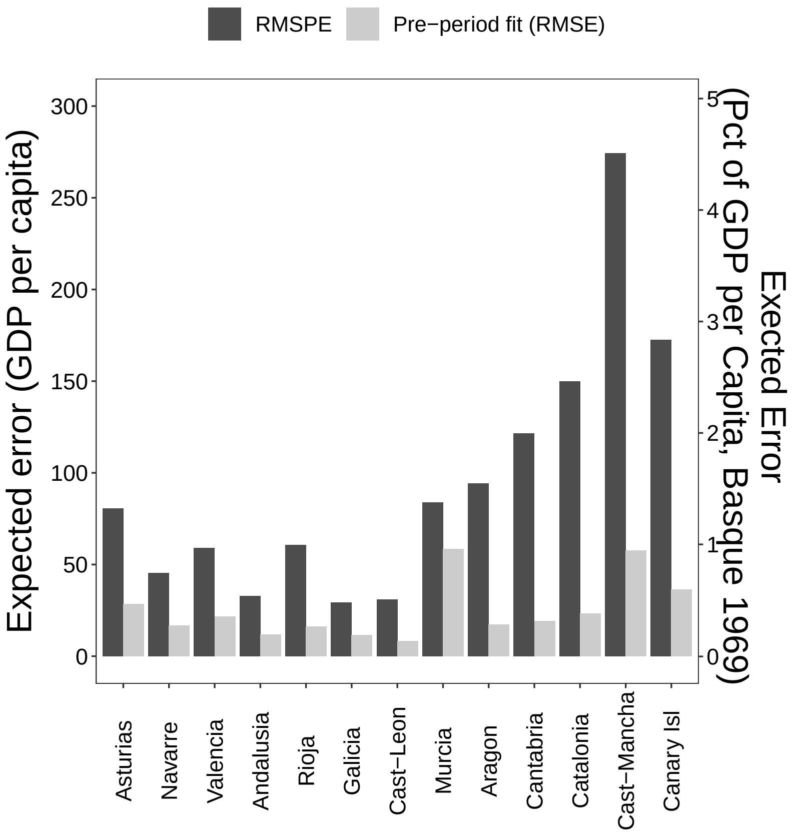


(a) Prediction error for the Catalonia placebo

(b) Average values across placebo exercises

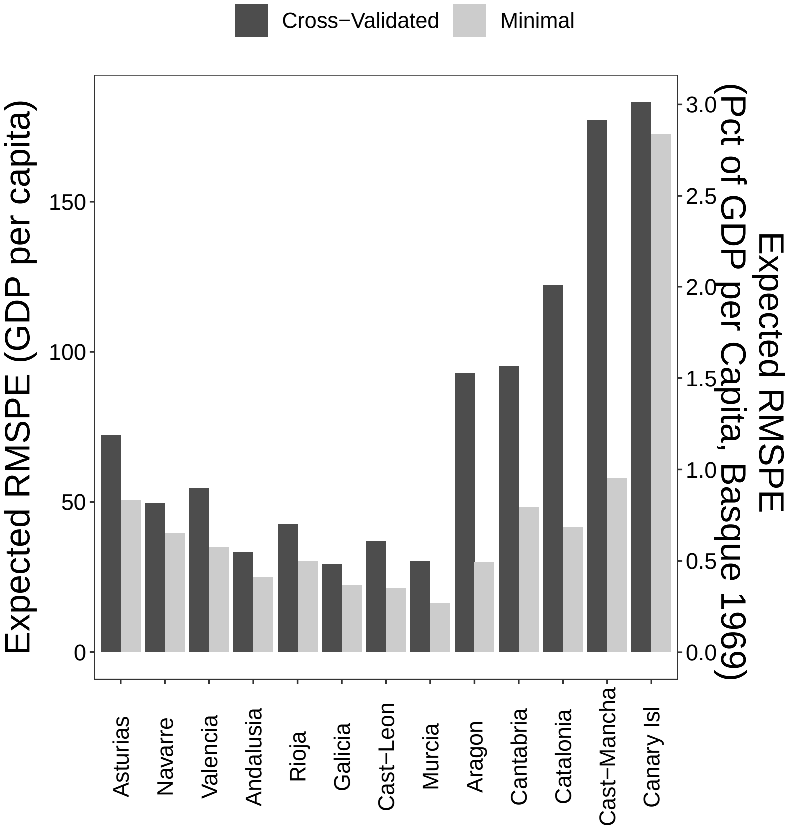
GDP per capita is measured in 1986 US dollars. The outer hinges of the box plot indicate the 25th and 75th percentiles, and the median at the line splitting the box for each estimator. The black dot is the mean. Panel (b) reports the average value for each statistic, averaging over placebo exercises. The mean RMSPE for matching in the Catalonia placebo is omitted ($210 per person per year) Results are based on 1,000 simulated data draws.

Figure E.1: Statistics on prediction error of alternative estimators in placebo analyses of the factor model



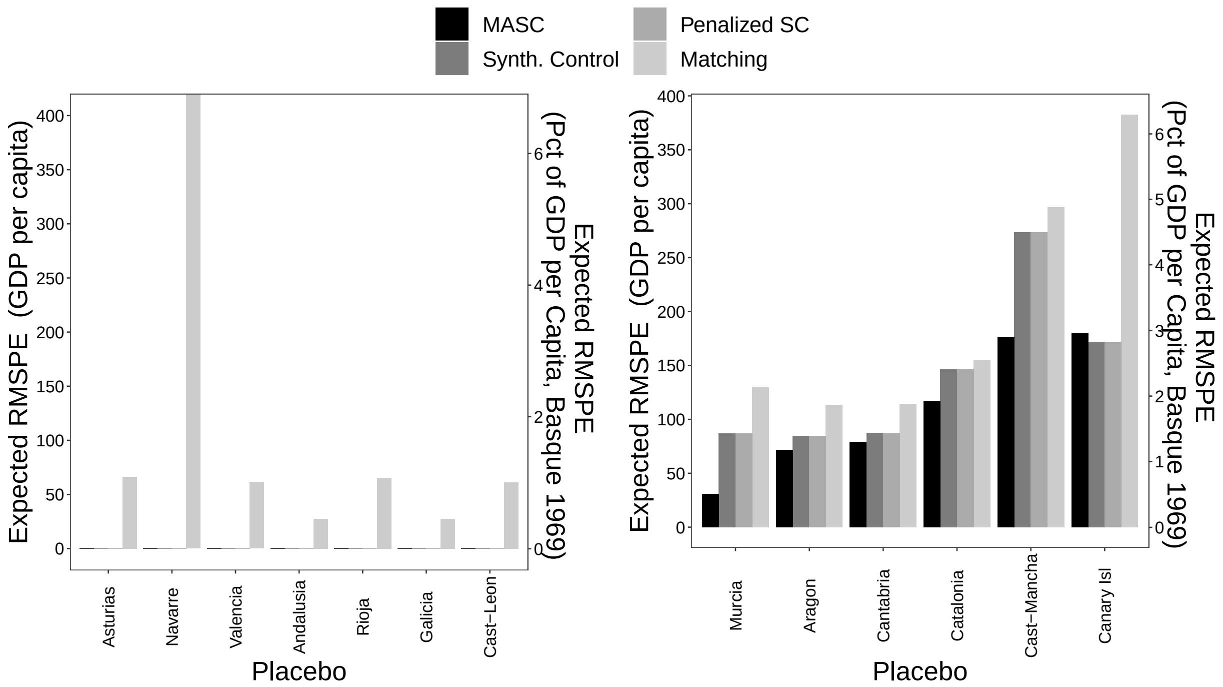
Note: This plot compares the expected root means square prediction error from 1970 to 1973 to the expected pre-period fit (root mean square error) in the 13 main placebo regions. GDP per capita is measured in 1986 US dollars. Results are based on 1,000 simulated data draws.

Figure E.2: Expected pre-period fit and the performance of SC in the placebo analyses of the factor model



Note: This plot depicts the expected root means square prediction error from 1970 to 1973 for MASC in the 13 main placebo regions. The darker color plots the value reached by cross-validation. The lighter color plots the minimum value achieved by a control group for MASC (as defined by the tuning parameters ϕ and m). GDP per capita is measured in 1986 US dollars. Results are based on 1,000 simulated data draws.

Figure E.3: Cross-validated and minimum achievable expected prediction error of MASC in placebo analyses of the factor model

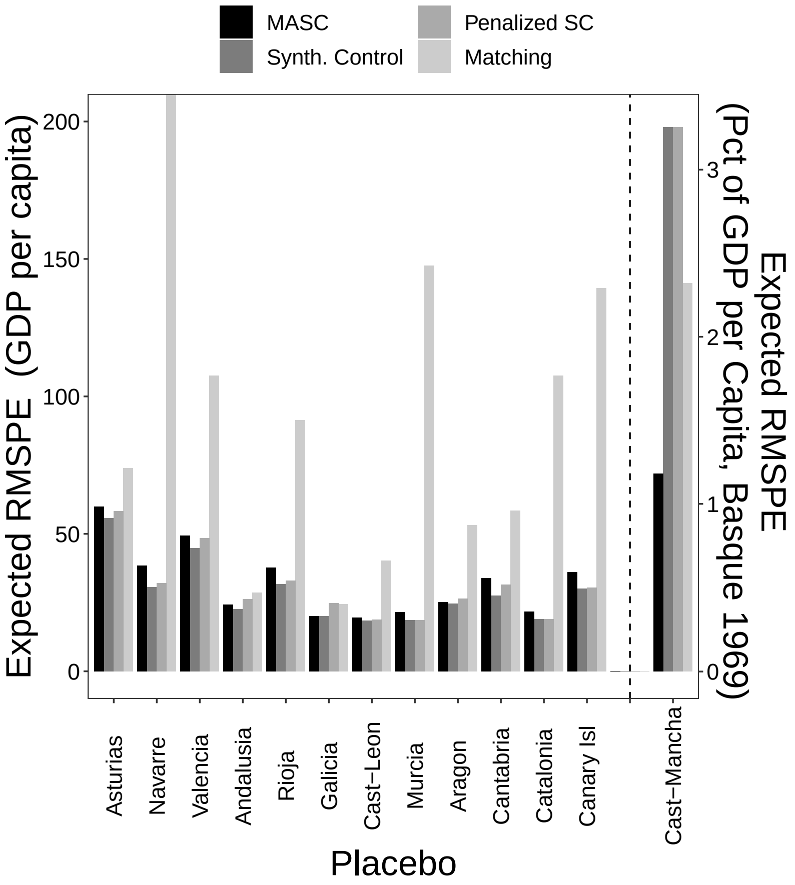


(a) Placebo regions with a suitable synthetic control

(b) Placebo regions without a suitable synthetic control

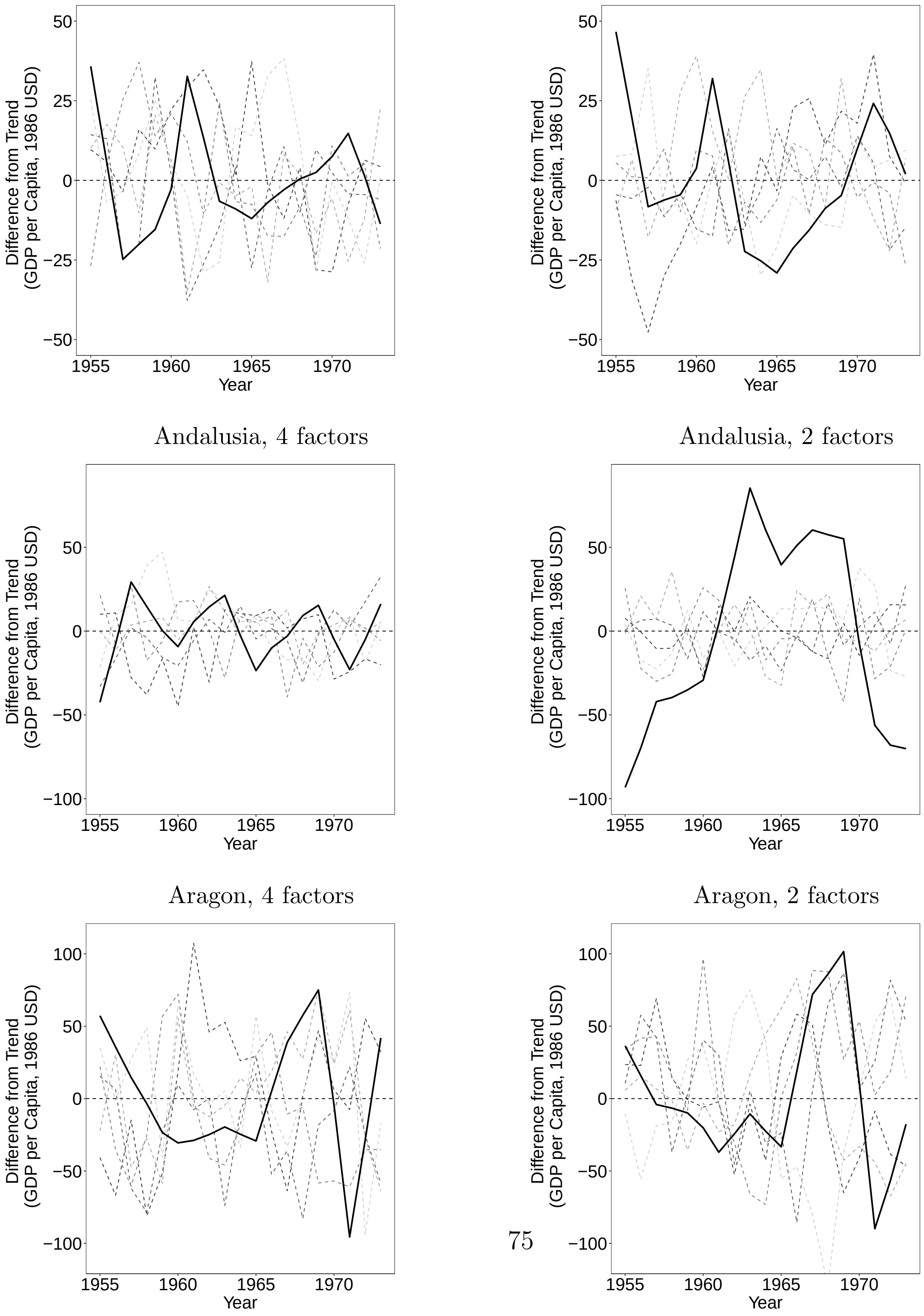
Note: This plot depicts the root means square prediction error over the first four post-period years in the 13 main placebo regions. Data for this exercise consists only of the region-specific trends. Panel (a) depicts regions which have a trend that can be fit perfectly by a synthetic control. Panel (b) depicts regions for which no such synthetic control exists. The RMSPE of matching is $490 for Navarre. Outcomes are measured in 1986 US dollars.

Figure E.4: Performance of alternative estimators in placebo exercises based on trends of the factor model



Note: This plot depicts the root means square prediction error over the first four post-period years in the 13 main placebo regions. Regions to the left of the dashed line have a trend that can be fit perfectly by a synthetic control. No such synthetic control exists for regions to the right of the dashed line. The expected RMSPE of matching is $382 for Navarre. Outcomes are measured in 1986 US dollars. Results are based on 1,000 simulated data draws.

Figure E.5: Performance of alternative estimators in placebo exercises of the two-factor model



Andalusia, 4 factors

Andalusia, 2 factors

Aragon, 4 factors

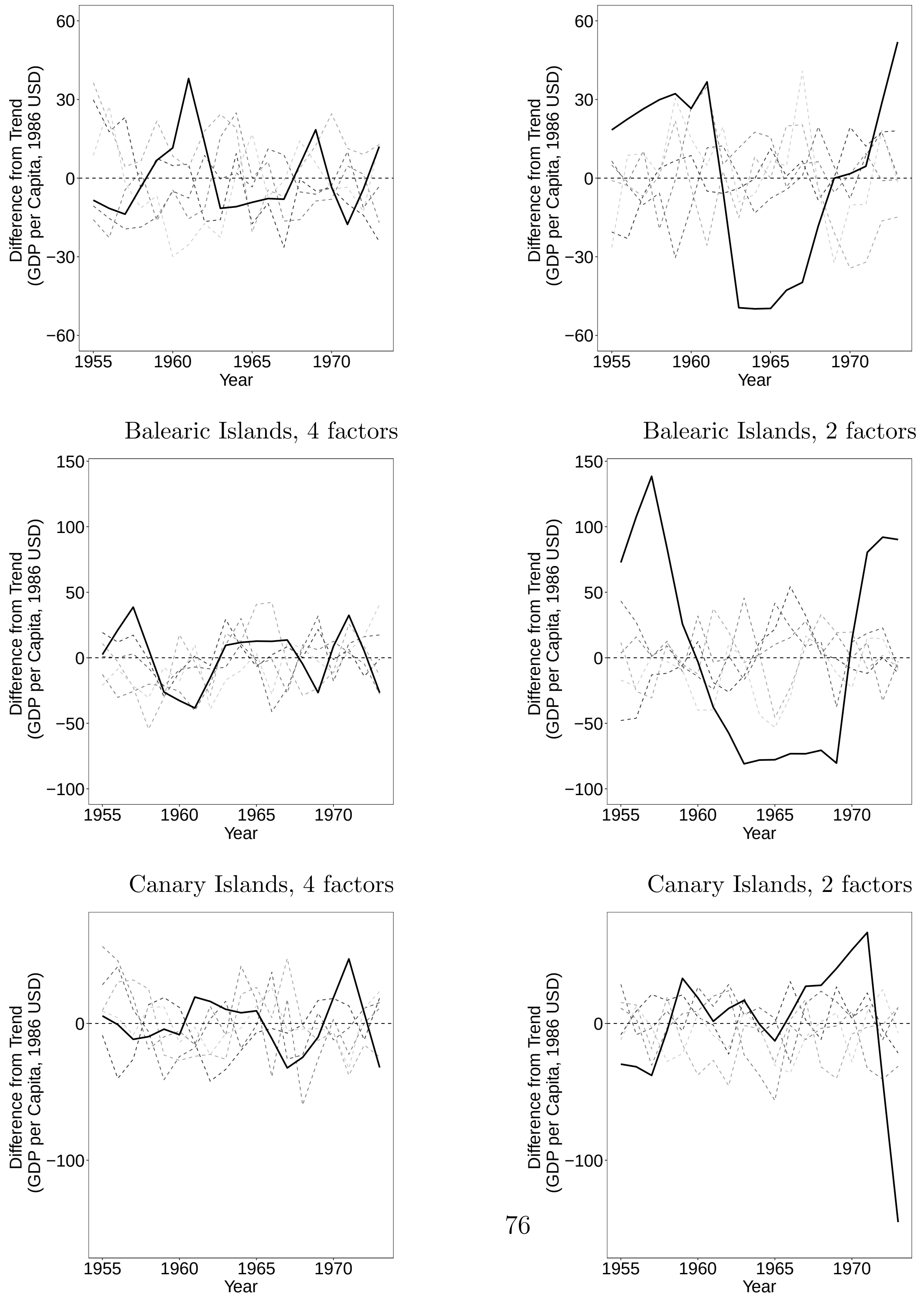
Aragon, 2 factors

Principality of Asturias, 4 factors

Principality of Asturias, 2 factors

Notes: The vertical dashed line is drawn halfway between 1969 and 1970 to indicate the beginning of the placebo treatment period in 1970. The solid black line in each panel indicates the difference of the true time series from the model's fitted trend for the region. The gray dashed lines represent differences from trend for random draws from the factor model.

Figure E.6: Comparing outcome paths to draws from factor models



Balearic Islands, 4 factors

Balearic Islands, 2 factors

Canary Islands, 4 factors

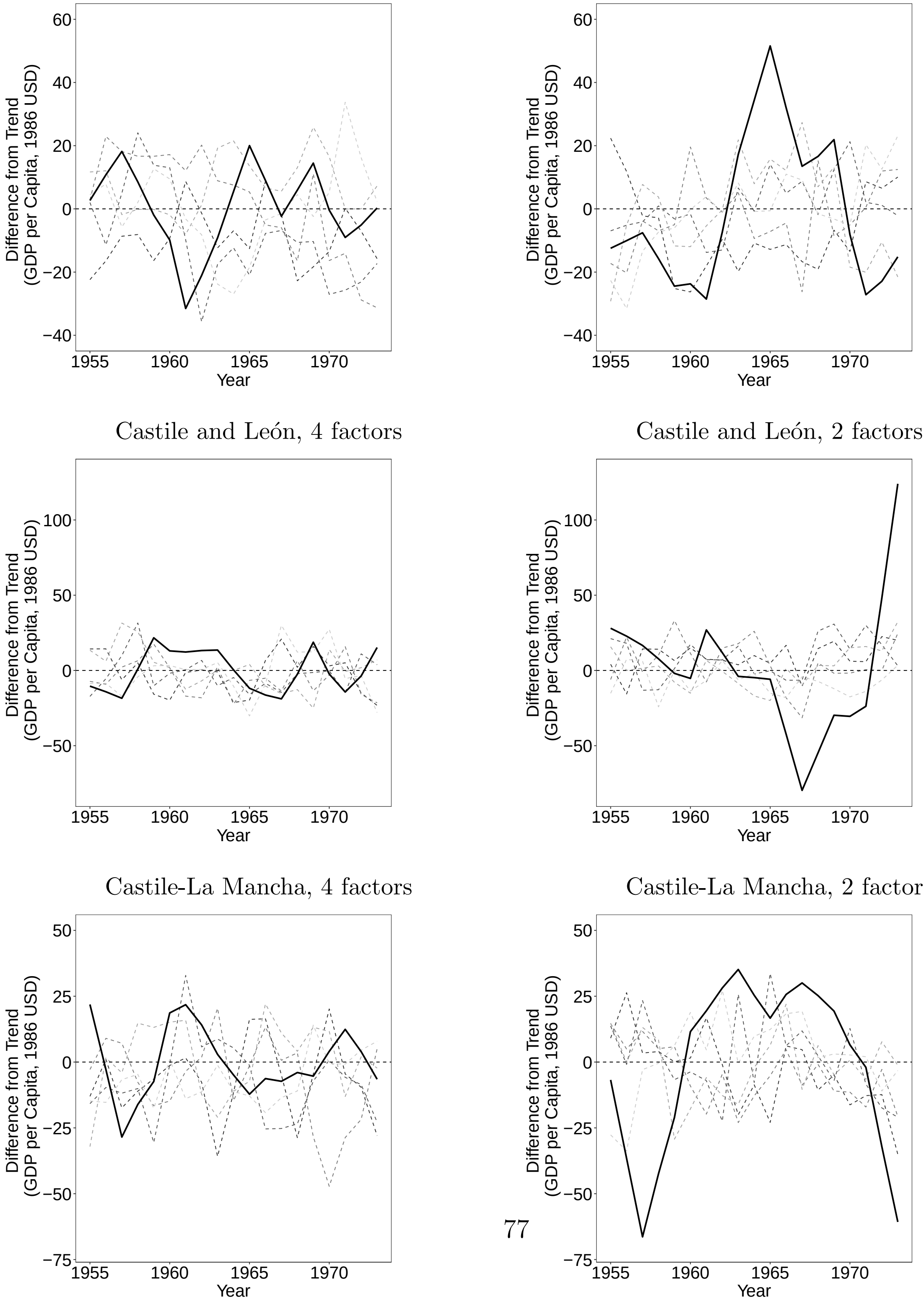
Canary Islands, 2 factors

Cantabria, 4 factors

Cantabria, 2 factors

Notes: The vertical dashed line is drawn halfway between 1969 and 1970 to indicate the beginning of the placebo treatment period in 1970. The solid black line in each panel indicates the difference of the true time series from the model's fitted trend for the region. The gray dashed lines represent differences from trend for random draws from the factor model.

Figure E.6: Comparing outcome paths to draws from factor models



Castile and León, 4 factors

Castile and León, 2 factors

Castile-La Mancha, 4 factors

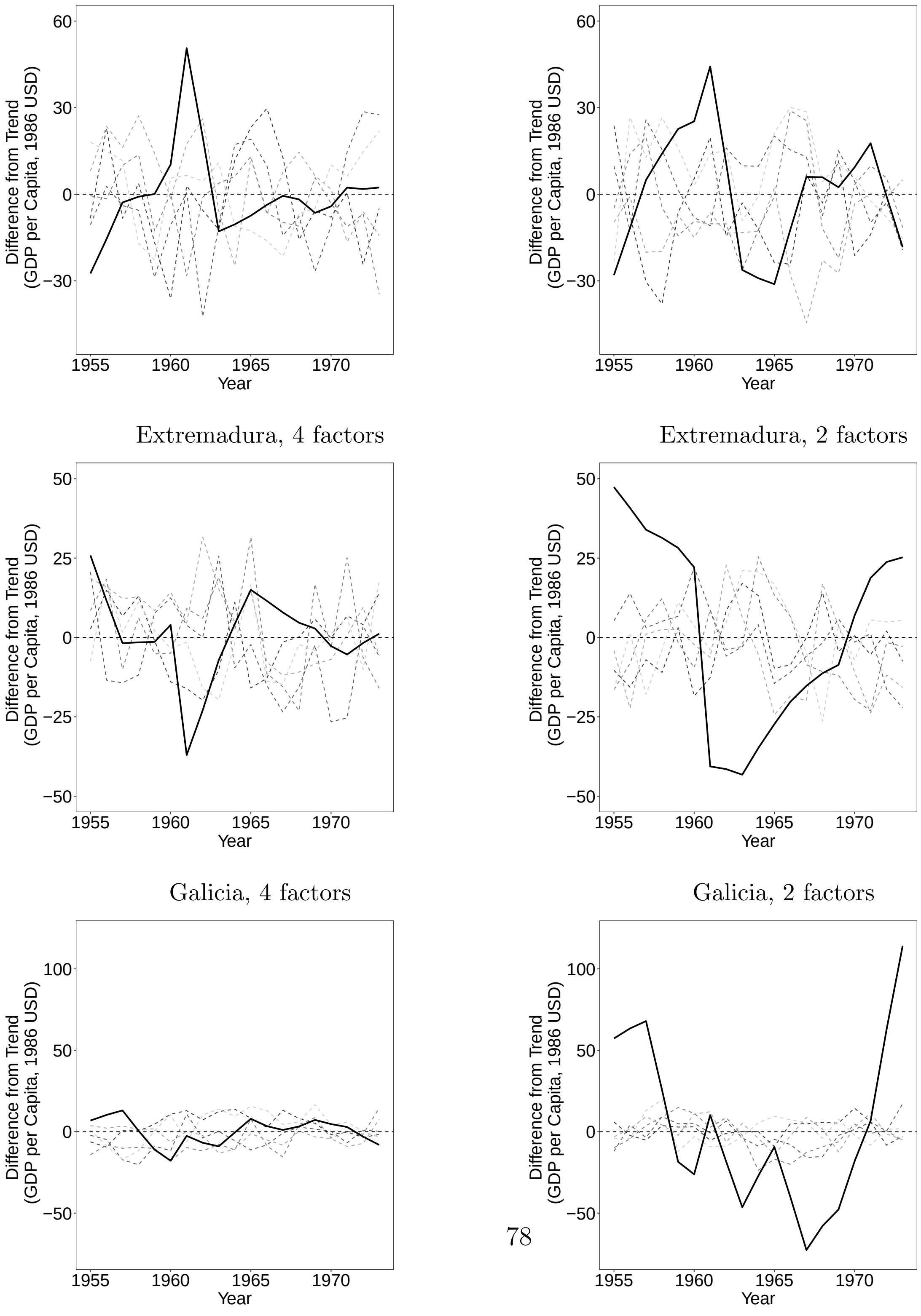
Castile-La Mancha, 2 factors

Catalonia, 4 factors

Catalonia, 2 factors

Notes: The vertical dashed line is drawn halfway between 1969 and 1970 to indicate the beginning of the placebo treatment period in 1970. The solid black line in each panel indicates the difference of the true time series from the model's fitted trend for the region. The gray dashed lines represent differences from trend for random draws from the factor model.

Figure E.6: Comparing outcome paths to draws from factor models



Extremadura, 4 factors

Extremadura, 2 factors

Galicia, 4 factors

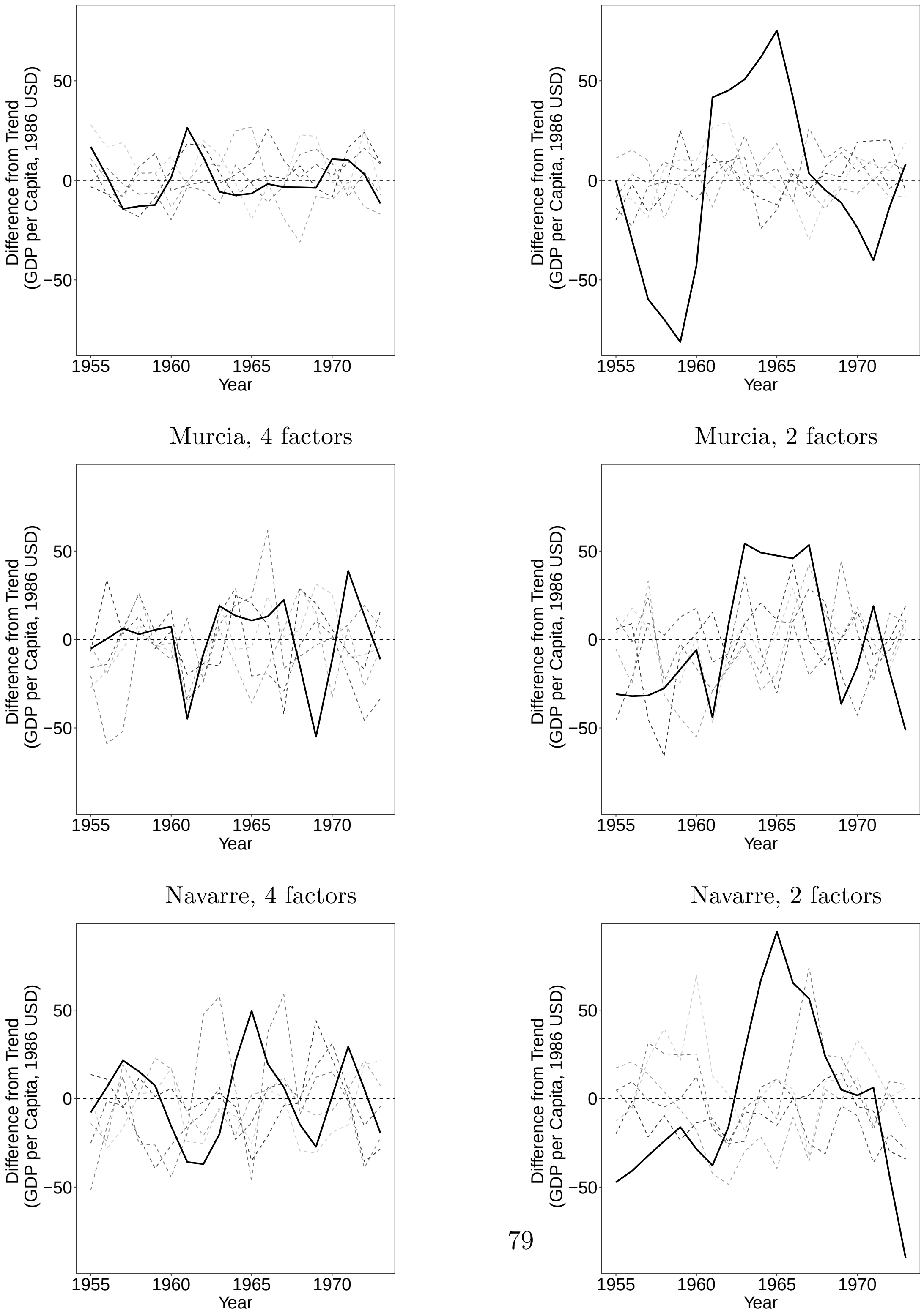
Galicia, 2 factors

Madrid, 4 factors

Madrid, 2 factors

Notes: The vertical dashed line is drawn halfway between 1969 and 1970 to indicate the beginning of the placebo treatment period in 1970. The solid black line in each panel indicates the difference of the true time series from the model's fitted trend for the region. The gray dashed lines represent differences from trend for random draws from the factor model.

Figure E.6: Comparing outcome paths to draws from factor models



Murcia, 4 factors

Murcia, 2 factors

Navarre, 4 factors

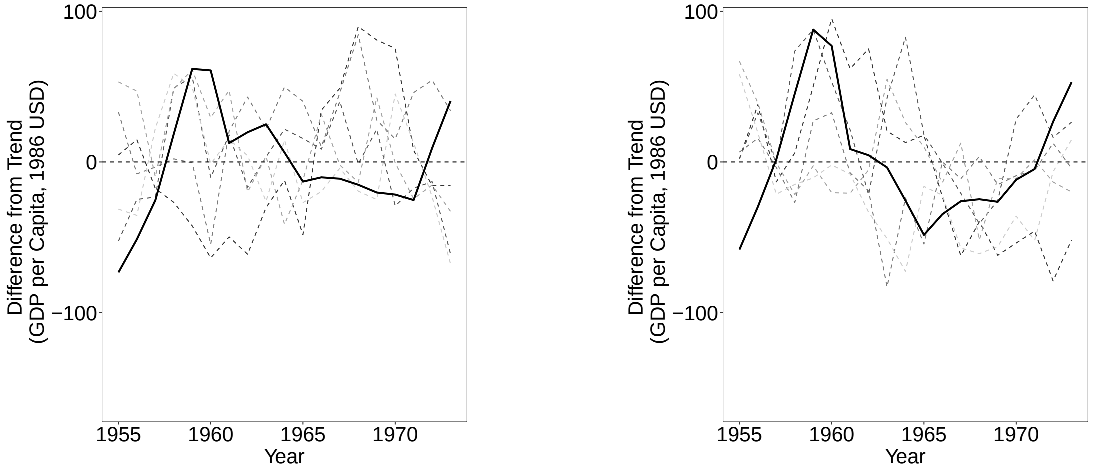
Navarre, 2 factors

Rioja, 4 factors

Rioja, 2 factors

Notes: The vertical dashed line is drawn halfway between 1969 and 1970 to indicate the beginning of the placebo treatment period in 1970. The solid black line in each panel indicates the difference of the true time series from the model's fitted trend for the region. The gray dashed lines represent differences from trend for random draws from the factor model.

Figure E.6: Comparing outcome paths to draws from factor models



Valencia, 4 factors

Valencia, 2 factors

Notes: The vertical dashed line is drawn halfway between 1969 and 1970 to indicate the beginning of the placebo treatment period in 1970. The solid black line in each panel indicates the difference of the true time series from the model's fitted trend for the region. The gray dashed lines represent differences from trend for random draws from the factor model.

Figure E.6: Comparing outcome paths to draws from factor models