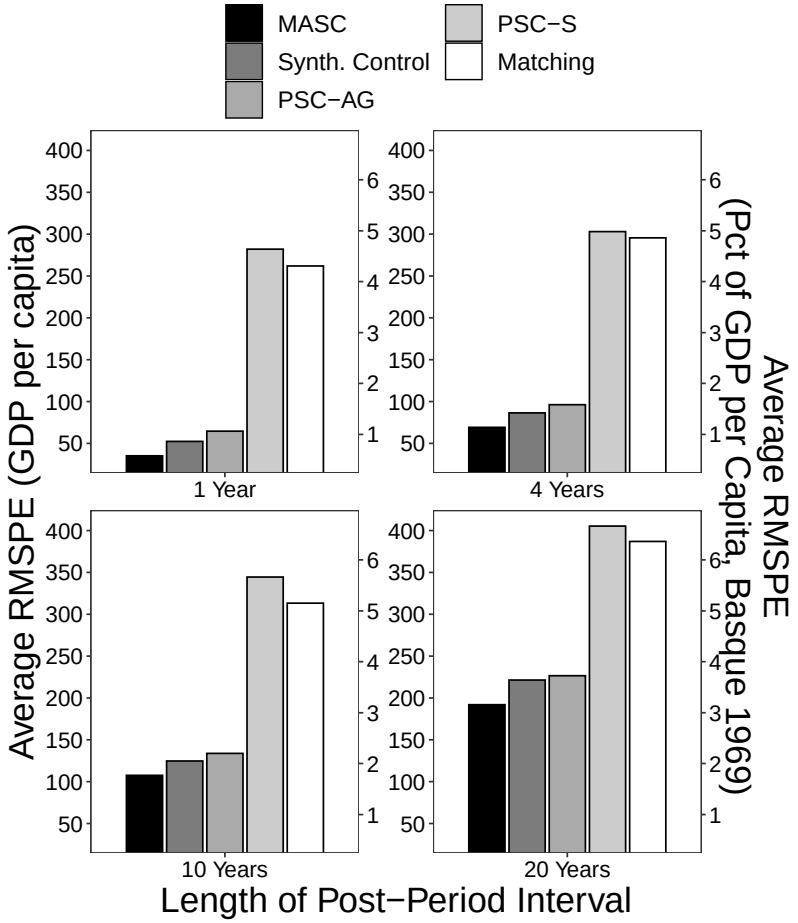
SUPPLEMENTAL APPENDIX TO COMBINING MATCHING AND SYNTHETIC CONTROL TO TRADE OFF BIASES FROM EXTRAPOLATION AND INTERPOLATION

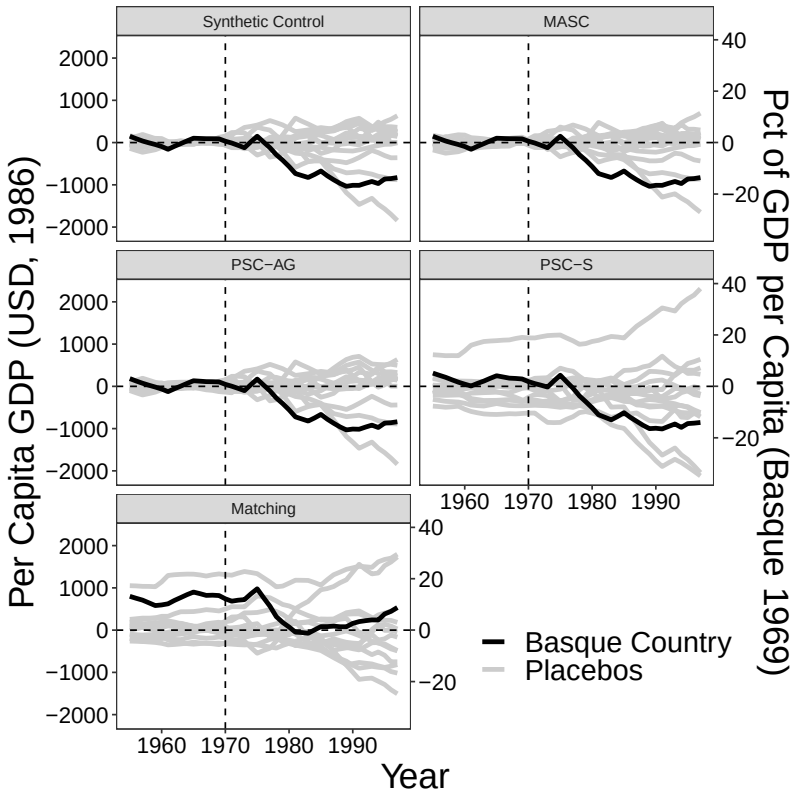
August 30, 2021

A Additional Results



Note: This plot depicts the root means square prediction error over alternative post-period interval lengths that each begin in 1970, averaging over the 13 main placebo regions. GDP per capita is measured in 1986 US dollars.

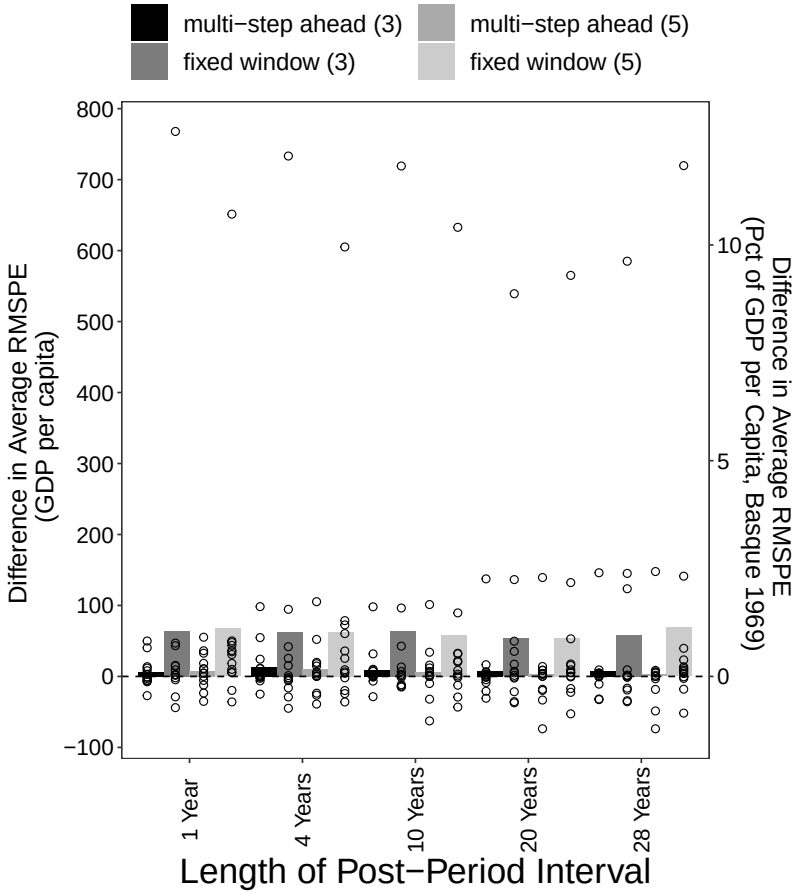
Figure A.1: Shorter-run estimator performance in the main Spanish placebo analyses



Note: The MASC selects $\phi = 0$ for the Basque Country, so MASC and SC imply the same counterfactual for the Basque Country. Each graph reports the year-to-year pre-period fit of the estimator (to the left of the dashed line) and the post-terrorism estimated treatment effect of the estimator (to the right of the dashed line). The placebos plotted in each panel represent the 13 placebo regions studied in Section 4.

Figure A.2: Economic costs of terrorism and placebo effects implied by alternative estimators

B Alternative Cross-Validation Procedures



Note: Plotted values represent differences from our preferred 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 our preferred procedure performs better. Multi-step ahead (f) forecasts from 1 and up to f periods ahead for each fold. Fixed window (f) forecasts over the f-period window for for each fold in which that is possible. For $f = 5$ for instance, this is possible for a single fold (1964).

Figure B.1: Placebo performance of alternative cross-validation procedures for MASC

C Results Including Regions with Poor SC Fit

	Pre-Period Fit					4-Year RMSPE				
Placebo	SC	MASC	PSC-S	PSC-AG	Matching	SC	MASC	PSC-S	PSC-AG	Matching
Andalusia	3	15	95	35	58	12	36	79	8	119
Aragon	24	25	67	24	103	108	106	89	108	70
Asturias	34	75	48	34	120	44	53	39	44	75
Balearic Isl	386	386	1,188	385	1,425	1,108	1,108	1,907	837	2,383
Canary Isl	36	67	336	36	69	218	144	219	218	291
Cantabria	9	18	635	89	115	100	101	722	140	97
Cast-Leon	28	32	442	28	316	21	34	454	21	338
Cast-Mancha	60	100	145	60	72	254	171	79	254	298
Catalonia	18	44	1,057	18	1,295	99	57	1,167	99	1,337
Valencia	37	85	166	66	145	44	40	361	117	84
Extremadura	339	339	554	339	621	750	750	852	750	901
Galicia	21	22	316	21	296	25	20	348	25	303
Madrid	849	899	1,035	849	899	490	455	131	490	455
Murcia	44	65	110	59	92	94	6	134	99	65
Navarre	23	23	76	23	425	79	80	184	79	540
Rioja	26	39	38	32	188	27	52	63	39	225
Average	**121**	**140**	**394**	**131**	**390**	**217**	**201**	**427**	**208**	**474**
	10-Year RMSPE					28-Year RMSPE				
Placebo	SC	MASC	PSC-S	PSC-AG	Matching	SC	MASC	PSC-S	PSC-AG	Matching
Andalusia	51	65	95	53	146	501	512	356	507	551
Aragon	103	102	93	103	98	133	145	227	133	252
Asturias	178	117	173	178	100	903	787	1,001	903	709
Balearic Isl	1,113	1,113	1,977	821	2,323	1,793	1,793	2,543	1,403	2,954
Canary Isl	154	118	354	154	200	271	199	248	271	227
Cantabria	99	98	736	148	76	116	131	1,279	109	556
Cast-Leon	46	40	467	46	298	81	82	624	81	274
Cast-Mancha	382	290	131	382	384	378	271	99	378	328
Catalonia	98	55	1,135	98	1,256	252	268	1,468	252	1,351
Valencia	30	31	338	94	70	259	171	374	373	174
Extremadura	885	885	896	885	950	861	861	705	861	838
Galicia	20	17	385	20	348	105	99	411	105	422
Madrid	491	467	205	491	467	335	1,295	965	335	1,295
Murcia	100	80	104	91	95	329	307	160	362	300
Navarre	210	210	279	210	660	230	230	397	230	1,034
Rioja	151	176	188	163	342	202	187	189	192	267
Average	**257**	**241**	**472**	**246**	**488**	**422**	**459**	**690**	**406**	**721**

Note: Pre-period fit is calculated from 1960 to 1969. For reference, the pre-period fit of SC for the Basque Country is $94. GDP per capita is measured in 1986 US dollars.

Table C.1: Performance of alternative estimators in the all Spanish placebo analyses

D Monte Carlo Simulations

D.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 forms

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

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 for the GDP per capita includes four latent factors composing the region-specific deterministic trends.

The SC estimator should suffer little if anything from interpolation bias in the factor model. 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 1997, 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 coefficients 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). Because the small number of regions in the data limits our ability to fit a more complex latent factor structure that explains the other regional covariates as well, we instead keep those covariates fixed at their original values in the data. Through the lens of this model, it is useful for estimators to incorporate the covariates because information about the factor loadings is obfuscated by the stochastic component of the outcome paths (Z_{it}). The graphs in Figure D.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.

D.2 Multiple Realizations of the Same DGP

The first Monte Carlo exercise addresses whether the relative performance of the five 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 three key insights from the placebo analyses in Section 4. First, MASC tends to have lower RMSPE than other estimators. This is apparent in Figure D.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 Valencia placebo, which is one placebo region in which MASC differs substantially from SC. In Valencia, the MASC estimator tends to has lower RMSPE than SC and PSC-AG; the 75th percentile of RMSPE for MASC ($216) is slightly *below* the 25th percentile for SC or PSC-AG ($221 and $230 respectively) On the other hand, the matching estimator compares more favorably to MASC, with a slightly higher median RMSPE ($187 for matching versus $181 for MASC) and a tighter interquartile range (from $168 to $208 for matching versus $137 to $216 for MASC). However, the matching estimator has a longer right tail that skews the mean RMSPE upward (to $199). The PSC-S estimator has a high RMSPE by contrast, with a 25th percentile of $871. In panel (b), we report statistics on RMSPE pooling across the thirteen primary placebo regions. These still show an edge for MASC, although one that is slightly less pronounced than in Valencia. The MASC exactly corresponds to the original SC estimator in several regions, but (unlike the PSC-AG estimator) MASC tends to have lower RMSPE where they differ. These are regions where prediction errors tend to be smaller, so the 25th percentile of RMSPE is 18 percent smaller for MASC than it is for SC. All estimators tend to have high RMSPE in regions like Asturias and Catalonia (in which SC and MASC both have expected RMSPEs of over $1,000), which causes the relative differences in means and higher quantiles across the estimators to look small when pooling over the placebos.

Second, the pre-period fit of the SC estimator is not necessarily a strong indicator of its prediction error. This is apparent in Figure D.2, which shows that the SC estimator fits similarly well in expectation for Aragon and Asturias. Yet Aragon is one of the regions in which the SC estimator returns its lowest expected prediction error, whereas Asturaias is one of the regions in which it returns its highest expected prediction error.

Third 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 D.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 27 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 has higher MSPE in this Monte Carlo turns out to be that no region of Spain has a suitable synthetic control that reconstructs its outcome path and other characteristics. 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, a suitable synthetic control that perfectly fits the treated unit should have zero prediction error, but as shown in Figure D.4, this occurs for no region. The MASC tends to have lower MSPE than the SC estimator in the regions where the two differ, such as the Canary Islands, Cantabria, Galicia, or Valencia.

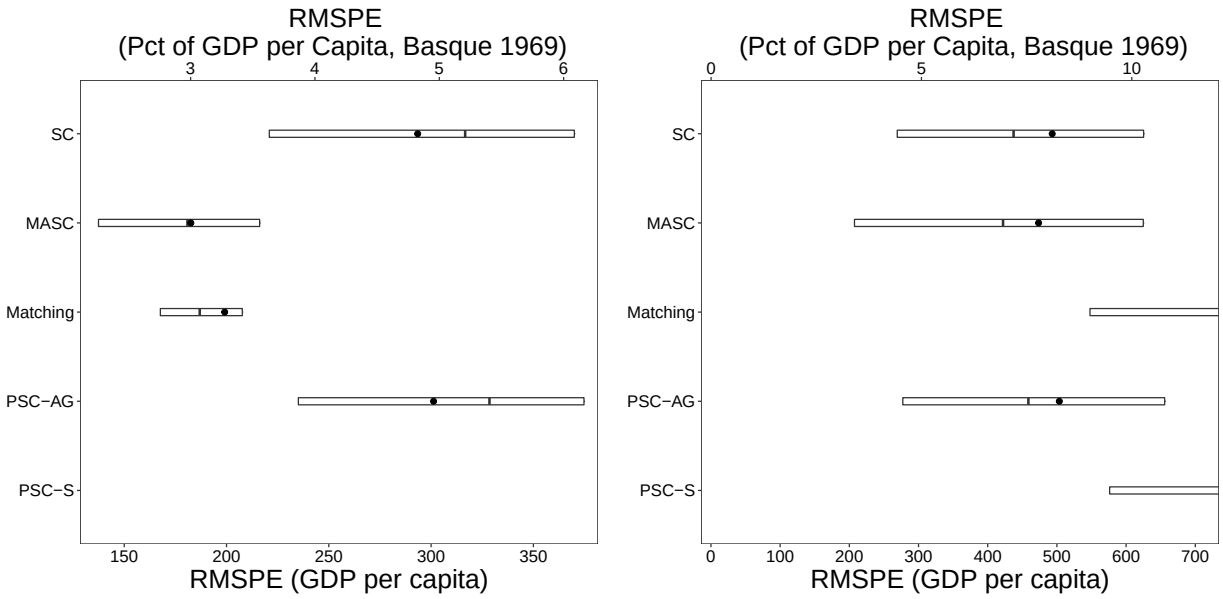
D.3 Simplifying the Factor Model

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 we continue shutting down the stochastic component of the factor model ($Z_{it} = 0$). In the absence of the stochastic component of the outcome path, a suitable synthetic control need only perfectly fit the pre-period *outcome* path of the treated region, not its other characteristics. Therefore in this exercise, we take $\boldsymbol{X}_i$ to include all pre-treatment outcomes and no other covariates, with all estimators equally weighing each year of pre-period outcome data. This greatly reduces the dimensionality of the characteristics that define a region. This exercise is meant to be expository, not necessarily realistic. With only the four latent factors characterizing outcome trends, four regions (Aragon, Castile and León, Valencia, and Galicia) have a suitable synthetic control.

In this model, the advantage of SC becomes clear. Figure D.5 shows that SC perfectly forecasts for the four treated units for which it can find a close-fitting synthetic control. The MASC and penalized SC correctly assign full weight to the SC estimator in these cases, so they perfectly forecast these regions as well (penalized SC, in fact, is nearly identical to SC for all regions). In some regions that lack a suitable synthetic control like Asturias, Cantabria, or Rioja, SC performs comparatively well because the loss from cross-validation for MASC outweighs the benefits from potentially blending SC with matching. Conversely, MASC performs comparatively well in others like Castile-La Mancha, Navarre, or Murcia, where the loss from cross-validation is small relative to the gains from incorporating

matching. Nevertheless, MASC tends to return a higher prediction error across placebos ($180 for MASC versus $164 for SC or penalized SC).

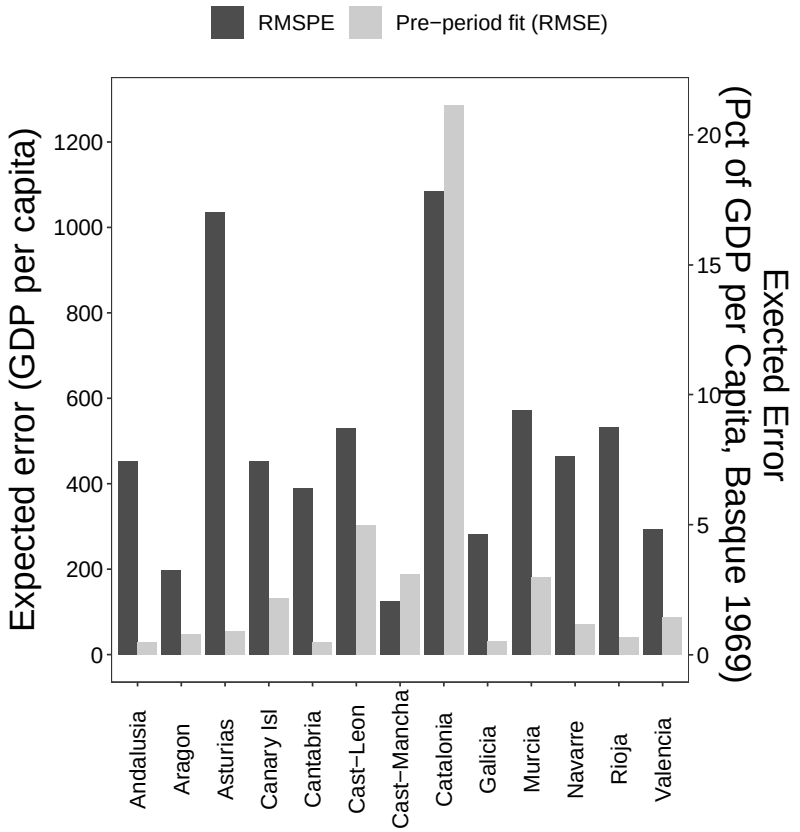
Taken together, these simulation results illustrate that one may prefer the SC estimator in settings where one suspects that a small number of factors mostly drive outcomes, so that a suitable synthetic control (or one close enough to suitable that the loss from cross-validation is large relative to its potential gains for MASC) likely exists.



(a) Prediction error for the Valencia placebo (b) Prediction error for the 13 main placebos

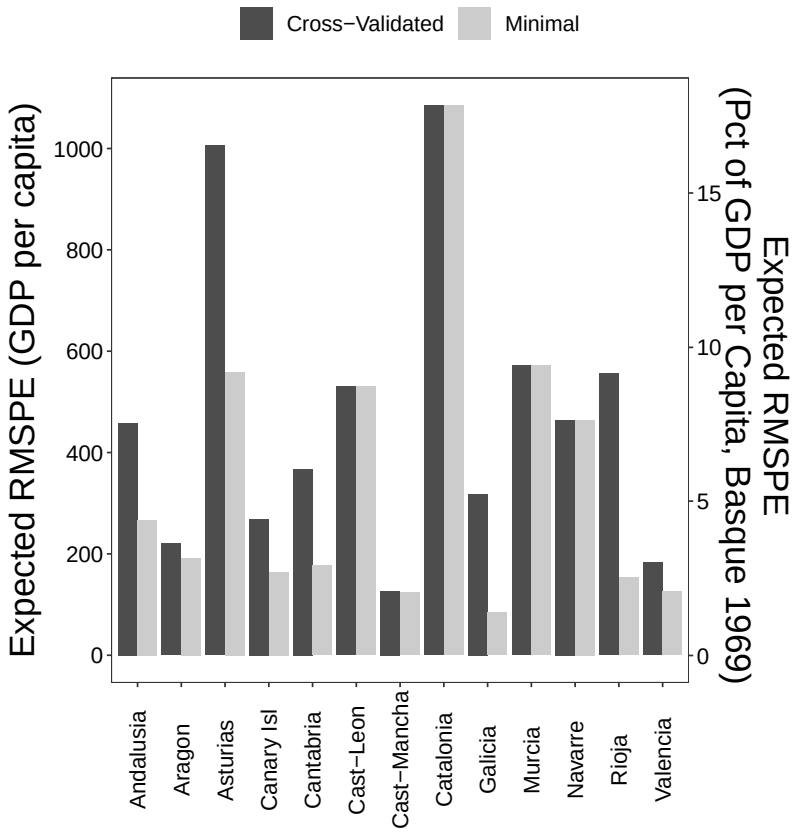
GDP per capita is measured in 1986 US dollars. The outer hinges of the box plot indicate the 25th and 75th percentiles, and the median at the line splitting the box for each estimator. The black dot is the mean. Panel (b) reports statistics after pooling simulation draws across placebos. The statistics for PSC-S in panel (a) omitted for scale; its lower quartile of RMSPE was \$872 per person per year. Likewise, all statistics but the lower quartile are omitted for matching and PSC-S in panel (b) for scale. The mean, median, and upper quartile for matching are \$988, \$878, and \$1374 respectively. The mean, median, and upper quartile for PSC-S are \$1264, \$930, and \$1817 respectively. Results are based on 1,000 simulated data draws.

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



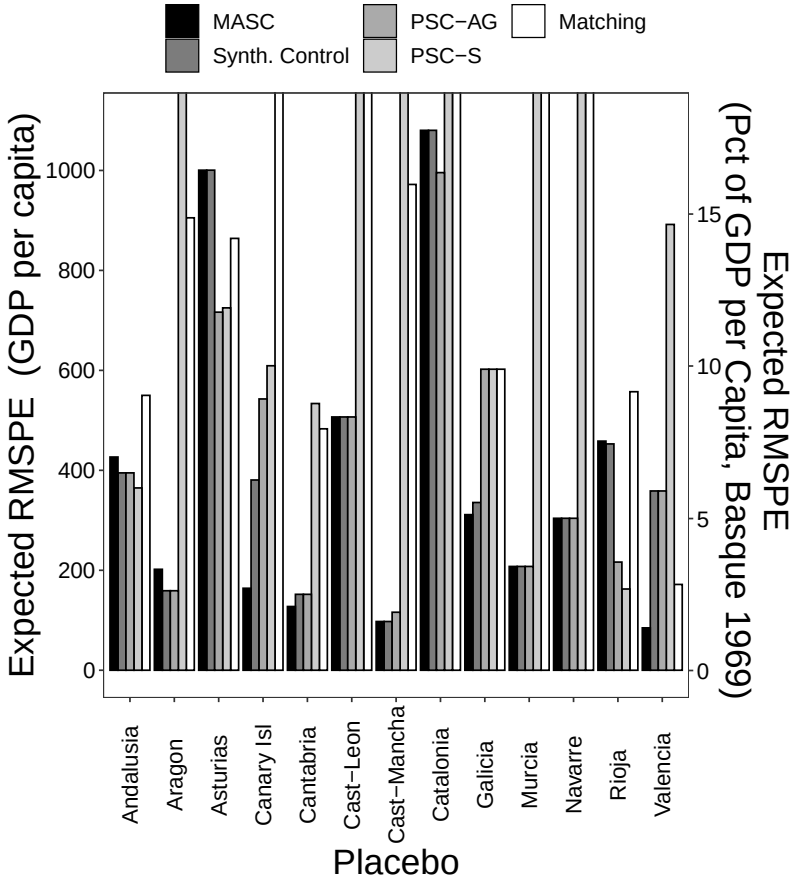
Note: This plot compares the expected root means square prediction error from 1970 to 1997 to the expected pre-period fit (root mean square error) in the 13 main placebo regions. GDP per capita is measured in 1986 US dollars. Results are based on 1,000 simulated data draws.

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



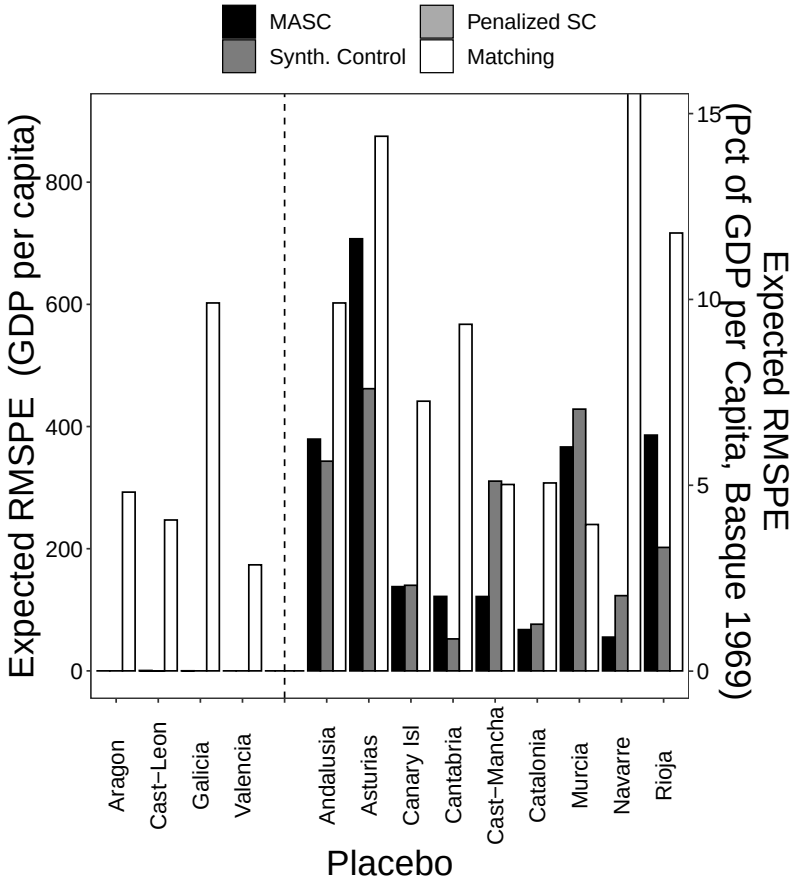
Note: This plot depicts the expected root means square prediction error from 1970 to 1997 for MASC in the 13 main placebo regions. The darker color plots the value reached by cross-validation. The lighter color plots the minimum value achieved by a control group for MASC (as defined by the tuning parameters ϕ and m). GDP per capita is measured in 1986 US dollars. Results are based on 1,000 simulated data draws.

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



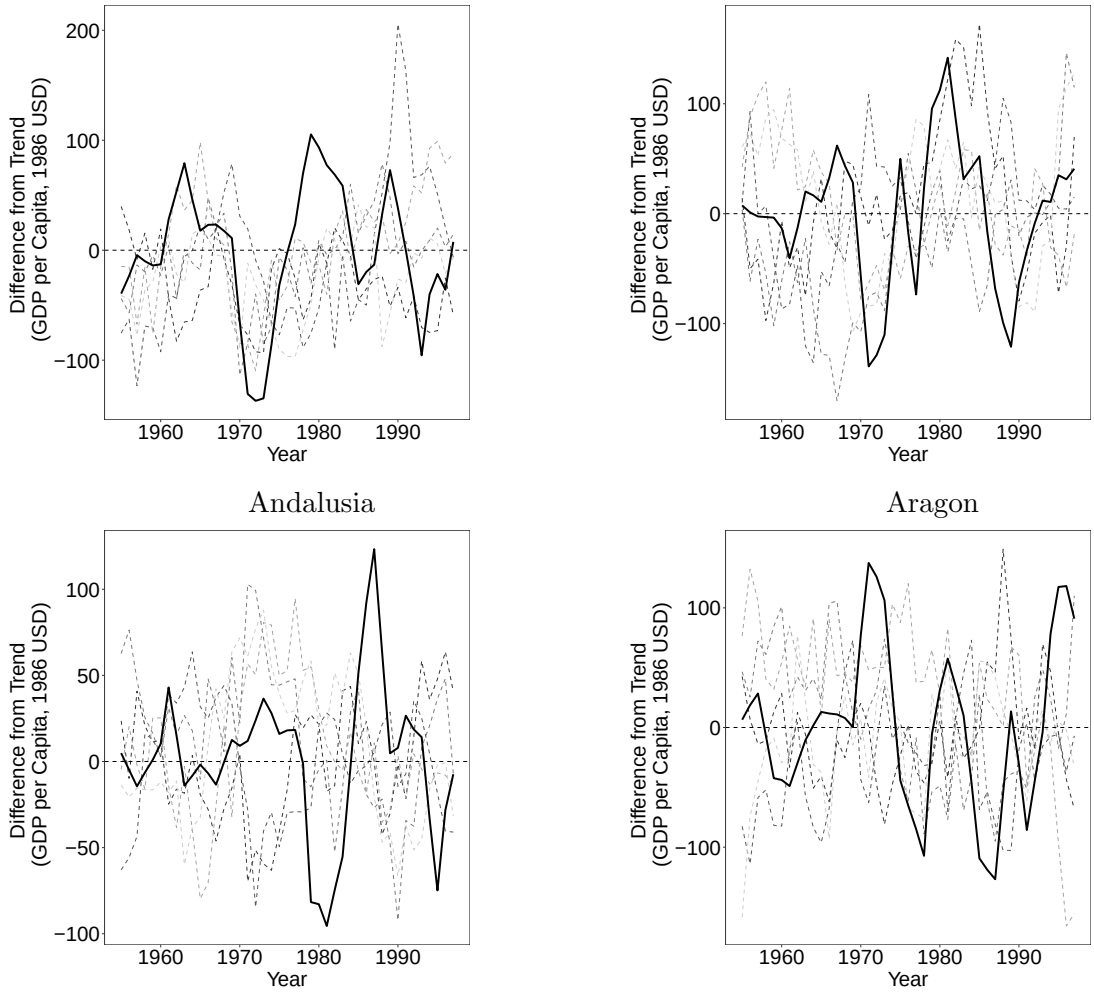
Note: This plot depicts the root means square prediction error from 1970 to 1997 in the 13 main placebo regions. Data for this exercise consists only of the region-specific trends. Outcomes are measured in 1986 US dollars. The censored expected RMSPE values range from $1,618 (for matching in Murcia) to $2,894 (for PSC-S in Catalonia) per person per year.

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



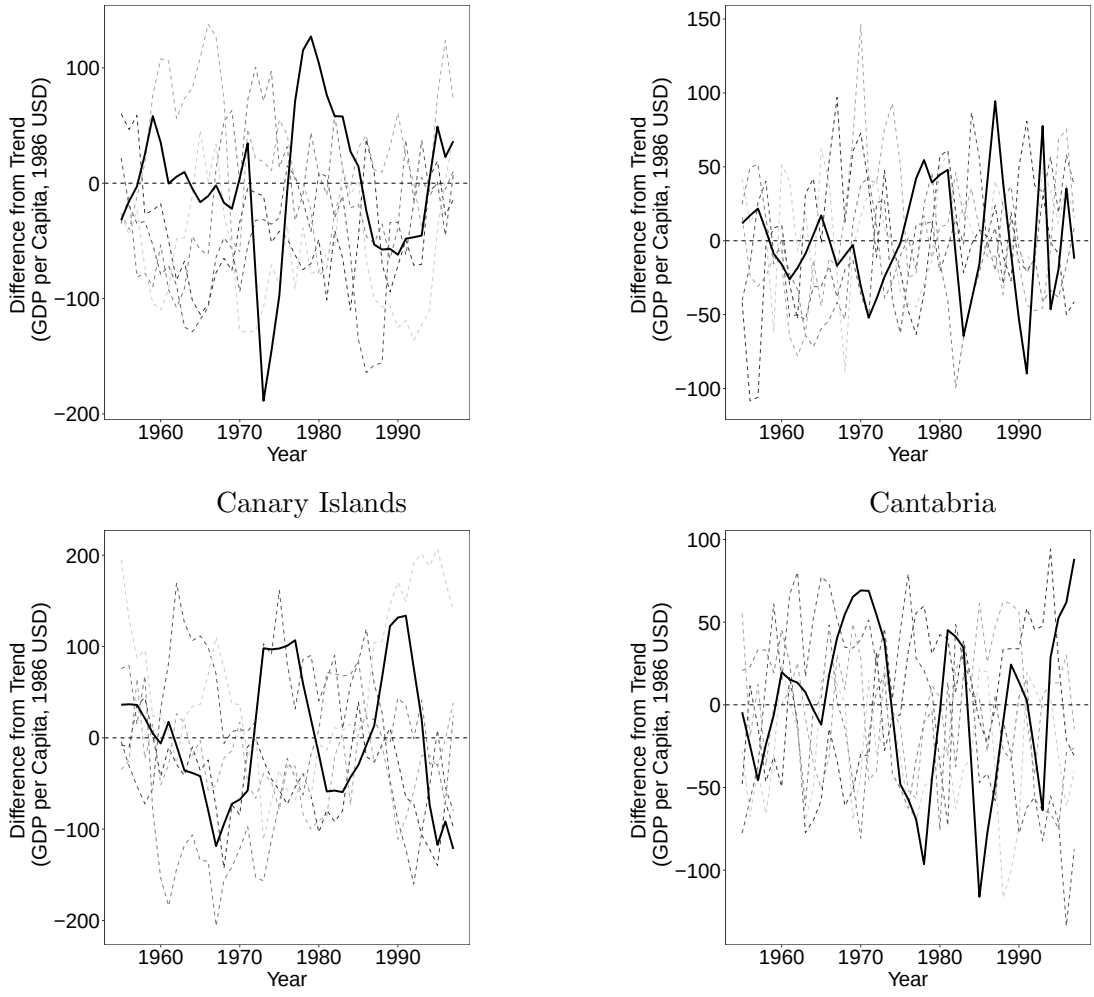
Note: This plot depicts the root means square prediction error from 1970 to 1997 in the 13 main placebo regions. Regions to the left of the dashed line have a trend that can be fit perfectly by a synthetic control. No such synthetic control exists for regions to the right of the dashed line. Estimators in this plot take $\boldsymbol{X}_i$ to include all pre-treatment outcomes and no other covariates, and equally weigh each year of data. The RMSPE of matching is \$1,110 per person per year for Navarre. Outcomes are measured in 1986 US dollars. Results are based on 1,000 simulated data draws.

Figure D.5: Performance of alternative estimators in the placebo exercises based on trends of the factor model, excluding covariates



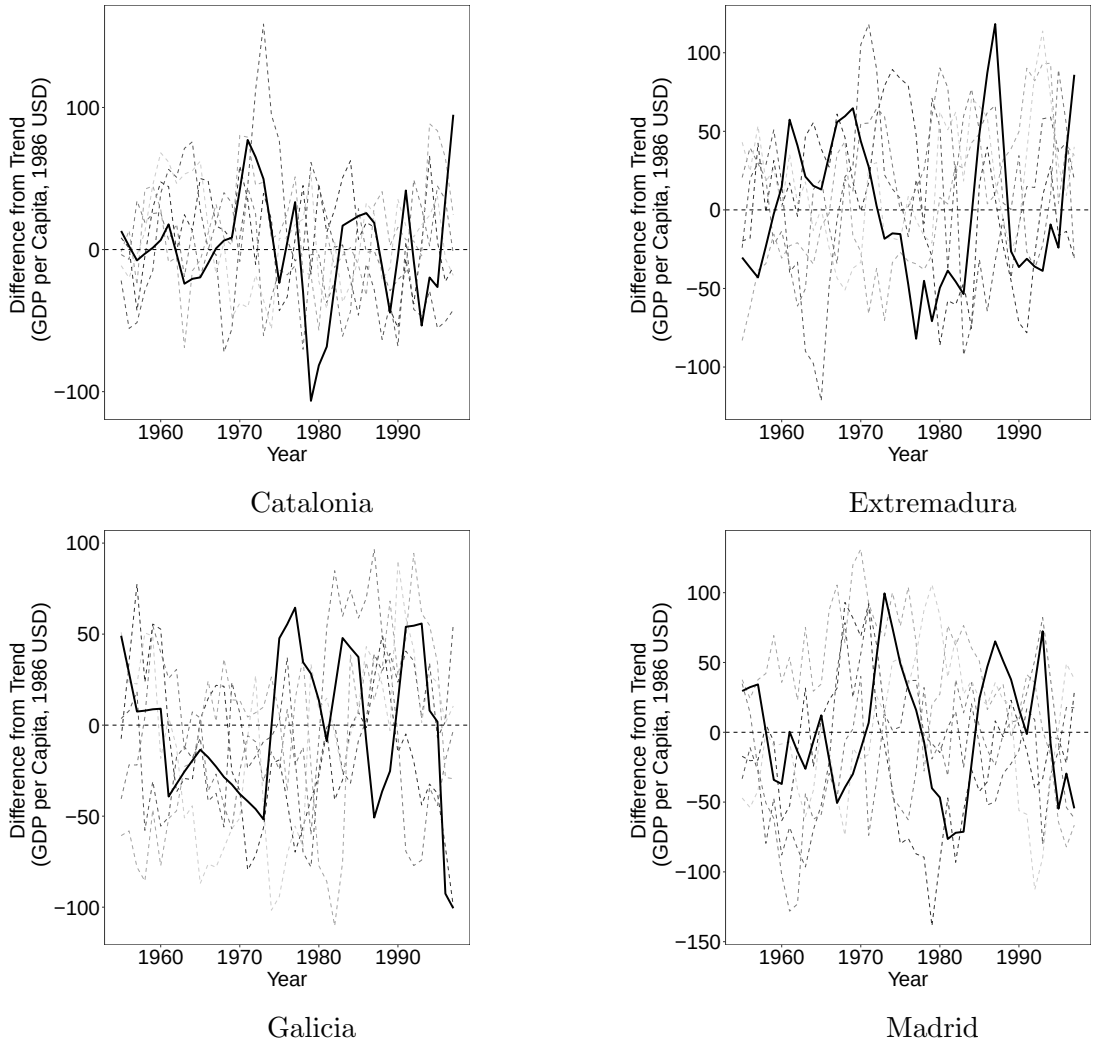
Notes: The solid black line in each panel indicates the difference of the true time series from the model's fitted trend for the region. The gray dashed lines represent differences from trend for random draws from the factor model.

Figure D.6: Comparing outcome paths to draws from the factor model



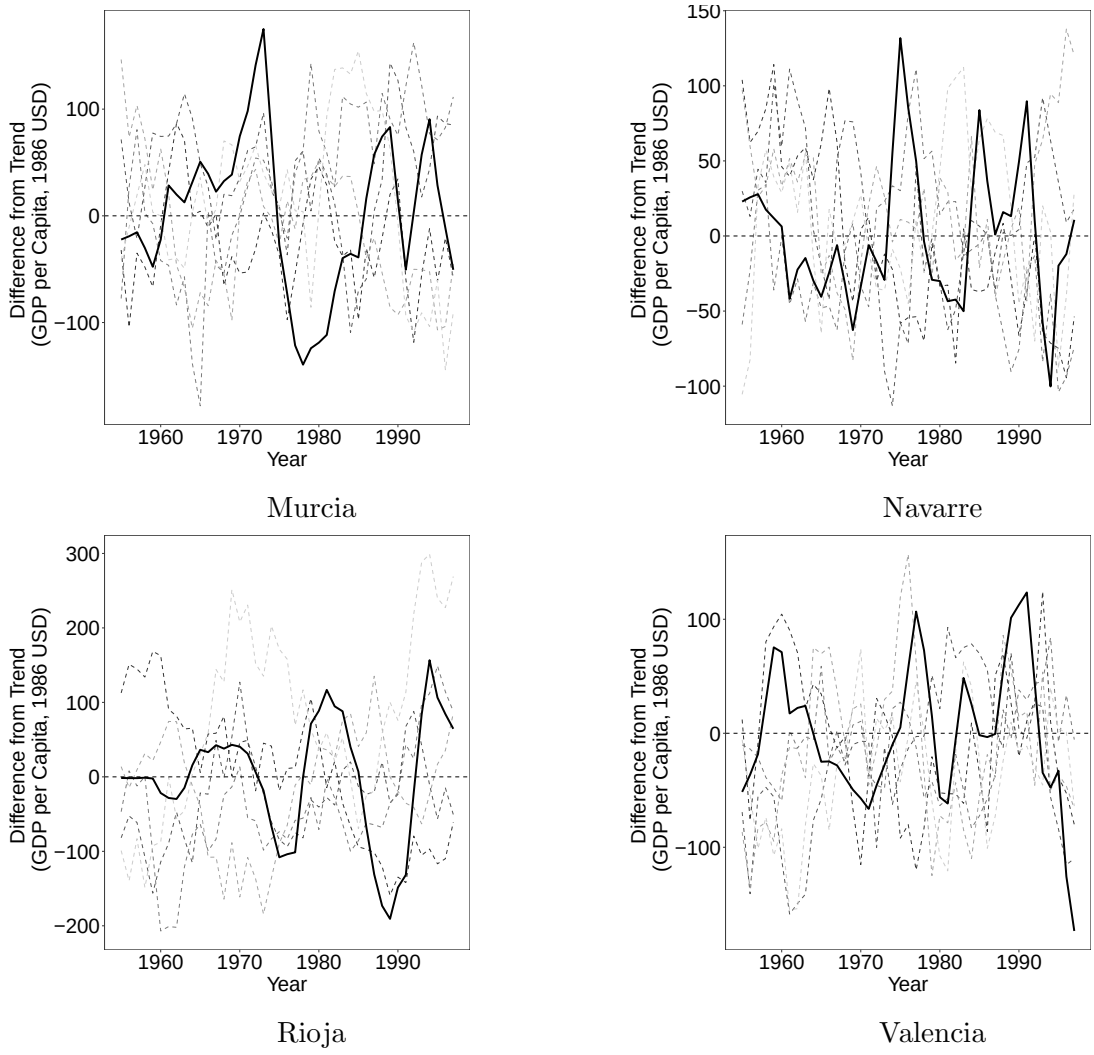
Notes: The solid black line in each panel indicates the difference of the true time series from the model's fitted trend for the region. The gray dashed lines represent differences from trend for random draws from the factor model.

Figure D.6: Comparing outcome paths to draws from the factor model



Notes: 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 D.6: Comparing outcome paths to draws from the factor model



Notes: 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 D.6: Comparing outcome paths to draws from the factor model