Estimating Causal Effects for Cointegrated Non-stationary Series: An Augmented Synthetic Control Approach

Pablo Crespo* and Ta-Cheng Huang†

This Version: October 2020

Abstract

Synthetic control is an important tool in the set of methodologies for estimating treatment effects. It is, however, dependent on the assumption of trend stationarity. In order to relax the assumption, this paper proposes an alternative approach based on modern techniques for automatic forecasting which learn the trend and seasonality components from the treated unit and correct them using candidate controls in the same spirit as synthetic control when candidate controls are cointegrated. Monte Carlo simulations show that this method is more robust than synthetic control in the presence of non-stationary cointegrated series, and able to identify treatment effects in a variety of forms. An empirical application reexamines the work of Abadie and Gardeazabal (2003) demonstrating the method's ability to replicate their results.

Keywords: Synthetic control; Counterfactuals; Forecasting; General Additive Models; Machine Learning

JEL codes: $C14$, $C58$, $G13$

*PhD Program in Economics, The Graduate Center, City University of New York. Email: pabloalejandrocrespo@gmail.com

†Global Asia Institute, National University of Singapore. Email: tchuang@nus.edu.sg

An earlier version of this paper has been circulated under the title "Trend-corrected Artificial Counterfactual for Cointegrated Non-stationary Series." We are grateful for comments from Dennis Jansen, Qi Li, Markus Ludwig, Liuren Wu, Ke-Li Xu, and participants of 29th Annual Meeting of the Midwest Econometrics Group.

1 Introduction and background

In the case of comparative studies seeking causal effects in the social sciences, treatment/control design quasiexperimentation has been the cause of much discussion in recent years. However, situations in which real feasible controls are not readily available abound. As a response to this Abadie and Gardeazabal (2003) suggest a method for creating artificial counterfactuals called synthetic control. Synthetic control operates on time series data, with a specific time t_0 set to be the time in which a treatment is applied.

The synthetic control method picks a set of candidate controls and creates a linear combination of their respective outcome variables within the convex hull of the outcome variable of the treated unit in periods prior to t_0 to generate a new control unit that mimics the treated unit closely in the pre-treatment stage. In order to determine whether a causal effect exists, the synthetic control is extrapolated to post-treatment periods and then graphic inspection allows the researcher to tell if it diverges from the observations in the treated unit. This extrapolation is possible because the outcome variables for the candidate controls are predicted via a constrained weighted regression on the outcome variable and determining characteristics through quadratic programming. Complementary methodologies can include a penalized form of this problem as Doudchenko and Imbens (2016) show with an elastic net. Thus, the linear combination found to generate the synthetic control is fit on the predicted values of the outcome variables. The constrained parameters in the convex hull of the treatment found in the solution of the regression are used for predicting the outcome values in the post-treatment periods using the information of the candidate controls.

As in most comparative case methodologies, it is assumed that the control units are not affected by the treatment, and that there is no interference between units (Abadie *et al.*, 2010). While most comparative studies often are careful to check whether this assumption is met, the idea of examining the time-series structure of the units is more often ignored. A driving assumption for the use of synthetic control is that the units are trend stationary. Violations of this assumption are studied in Carvalho *et al.* (2016), and

Ferman and Pinto (2016). The most problematic issue with violating this assumption is the over-rejection of a null hypothesis of finding no treatment effect. $I(1)$ series distort forecasts and put in doubt whether the generated synthetic control is indeed reflective of what a counterfactual should be for a treated unit. This occurs even under the circumstances in which a cointegration relationship between treated unit and candidate controls occurs.

The proposed correction in this case would imply using a type of model that can handle mutiple temporal error specifications and can generate good forecasts under non-stationary cointegrated relations. Hence, the model in this paper is a variation on Taylor and Letham (2018) general forecasting at a scale principle. That being that any time series can be given a forecast via a *generalized additive model* (GAM). The use of the GAM includes a trend, seasonality, and a weighted combination of the candidate controls to generate a new artificial counterfactual.

In order to test the efficacy of the GAM, a Monte Carlo experiment is necessary, in which time series that are non-stationary. Then a treatment unit would be generated by pre-defining weights on the control units, pre-defining a cointegrated relation. If no treatment effect is added, the artificial counterfactual should match the treated unit closely. Running the typical synthetic control and the GAM (called trend-corrected control henceforth), the accuracy of both methods can be measure by comparing their respective goodness of fit in relation to the treated unit.

Trend-corrected control uses the full series of the candidate control units as "covariates" in its specification. Since a logical question would be that of why these are used in full, rather than using detrended control series another simulation is run to justify the specification. In this specification, no treatment effect is added and three possible specifications are used. In the first, only shape elements (trend and seasonality) are used for generating the forecast, the second utilizes the pseudo residuals after detrending and deseasoning each control series as covariates, and the third uses the full series. The results suggest that the best forecasts are given across the board with using the full series as covariates.

Another Monte Carlo experiment is used to answer the question of whether this methodology can identify correct magnitudes for treatment effects when they are indeed present. This is done by measuring the distances between the treated unit under generated deterministic effects and the counterfactual. This distance is analyzed and compared to the way the effect is generated. The results indicate that the model does not greatly overstate or understate the treatment effects.

Beyond simulation experiments, an empirical replication of a published study is used as an effective procedure to test how well the proposal works. For that, the Abadie and Gardeazabal (2003) study on the effect of German reunification on the per capita GDP of Germany is used. The results obtained from the proposed model closely resemble those of the original study and are robust to changes in treatment periods.

The remainder of this paper is organized as follows.

2 Base model for synthetic control

Synthetic control proposed in Abadie and Gardeazabal (2003) and Abadie *et al.* (2010) consists of a balanced panel with T periods and $i = 0, 1, \cdots, J$ units, a "treatment" period $t_0 < T$, in which the only unit affected by a treatment is a series $y_{1,t}$ where t is the index for period. The series is defined by

$$y_{1t}^{(1)} = \begin{cases} y_{1t}^{(0)} & \text{for } t = 1, ..., t_0 - 1 \\ y_{1t}^{(0)} + \delta_t & \text{for } t = t_0, ..., T \end{cases}$$

with δ_t representing the treatment effect.

Synthetic control selects possible controls from a donor pool of units that could work as controls. The driving assumption for a unit being part of the donor pool hinges on it being unaffected by the treatment imposed on y_{1t}. These units come from series notated by $y_{0t,i}$ where $i = 1, \ldots, K$ indicates the specific donor. The selection is performed by looking for the donor's weights within the convex hull of the treated unit during the periods before t_0 . This means

$$y_{1t}^{(0)} = \sum_i w_i y_{0t,i} \text{ for } t = 1, \ldots, t_0 - 1$$

where

$$0 < w_i < 1 \text{ and } \sum_i w_i = 1.$$

Hence, in addition of needing for donor pool units to be unaffected by the treatment there is an extra assumption which requires that the "weights" w_i exist. The procedure searches for these weights by solving the following quadratic programming problem

$$\underset{W,V}{\operatorname{argmin}} \sqrt{(X_1 - X_0 W)' V (X_1 - X_0 W)}$$

where W is just a $K \times 1$ vector with w_i as entries, X_1 is the $t_0 - 1 \times 1$ vector of pre-treatment values of the treated unit augmented by any other variables which determine the value of y_{1t}. X_0 is a matrix with K columns, where each column is the vector of the first $t_0 - 1$ observations of each $y_{0t,i}$ augmented by the corresponding variables which augment X_1, and V is a positive definite weighting matrix.

The solution for the weighting vector $\hat{W}$ can then be used for generating a counterfactual. Consider Y_0 to be the matrix composed of the K vectors of the observations of each $y_{0t,i}$. Then the counterfactual is given by

$$\hat{Y}_1^{(0)} = Y_0 \hat{W}.$$

Hence, using the vectors for only the post-treatment periods, the treatment effect can be approximated by the vector

$$\hat{\delta} = Y_1^{(1)} - \hat{Y}_1^{(0)}.$$

3 Trend-corrected artificial counterfactual

Artificial counterfactual construction can be viewed as a forecasting problem. Hence, in the spirit of synthetic control, one can use donor pool units to construct forecasts of the treated unit. If the forecast is appropriate, it should resemble the true counterfactual. Thus to generate the counterfactual it is necessary to use a procedure which can automate the difficult time series forecasting procedures. While Hyndman and Khandakar

(2008) propose automatic procedures for fitting ARIMA models as well as software, the forecasting procedure tends to have large trend errors.

Hence, in the spirit of structural forecasting of Harvey and Peters (1990) and Taylor and Letham (2018), we propose a GAM . A GAM defined by Hastie and Tibshirani (1987) is a sum of components of basis functions estimated at once to approximate complex functional forms in estimation. The GAM used for this problem is defined as

$$Y_1^{(0)}(t) = \gamma_1(t) + \gamma_2(t) + \gamma_3(t) + \varepsilon_t$$

where $\gamma_1(t)$ represents the trend component, $\gamma_2(t)$ the seasonality component and $\gamma_3(t)$ the donor component, and ε_t is an unobservable noise.

Each of the components of the GAM is a model within itself. But their principle is clear in terms of the counterfactual. In the same spirit as synthetic control $\gamma_3(t)$ is used to inform the counterfactual via the donor pool, while the other two components, simpy learn basic shapes of the series from the pre-treatment period to impose in the post-treatment period .

The trend component $\gamma_1(t)$ is estimated as a piecewise linear regression of period t on y_{1t} in the pre-treatment stage and then just extrapolated in the post-treatment stage. Since the changepoint selection problem for piecewise linear regression is difficult to solve, this work uses ℓ_1 trend filtering with order 2 as suggested by Tibshirani *et al.* (2014) for setting them. The model is defined as

$$\hat{\gamma}_1(t) = \operatorname*{argmin}_{g \in \mathbb{R}^n} \frac{\left\| y_{1t} - g(t) \right\|_2^2}{2} + \frac{1}{m!} \lambda \left\| D^{(n,m+1)} g(t) \right\|_1$$

where $\lambda \geq 0$ is the regularization hyperparameter, and $m + 1$ is the order of the model. The penalized matrix is simply the matrix of second differences. Thus, the predictions from the model give a piecewise linear regression which penalizes on the second discrete derivative of changes of y_{1t} in terms of t.

The seasonality component $\gamma_2(t)$ follows a Fourier series for flexibility as proposed by

Harvey and Shephard (1995), which defines it as

$$\gamma_2(t) = \sum_{n=1}^{N} \left[\alpha_{0n} \cos\left(\frac{2\pi n t}{P}\right) + \alpha_{1n} \sin\left(\frac{2\pi n t}{P}\right) \right]$$

for flexibility purposes, N is truncated at 10.

The donor component $\gamma_3(t)$ is defined as

$$\gamma_3(t) = \sum_{i=1}^{K} \beta_i . y_{0t,i}.$$

The last component is simply a linear combination of the donors. Hence, in order to allow for predictability and proper forecasts, if there exists a cointegrating relationship between y_{1t} and all $y_{0t,i}$, the donors are informative and correct the forecasts beyond simple the shapes of the series.

4 Monte Carlo experiments

In order to compare different ways of computing artificial counterfactuals, the treatment and candidate control series need to be generated in a specific way. Series need to satisfy both a cointegrating relation and guarantee the existence of candidate control units within the convex hull of the treated unit. The simulation generation strategy requires a definition for number of members in the donor pool, a set of simulated weights meeting the synthetic control bounds, and a way to correlate donor pool units. The treatment unit is initially generated without effect, such that if the counterfactual fitted on the pre-treatment stage provides good forecasts of the treated unit in the post-treatment stage,one can conclude no effect was identified when there was in fact none. In order to test if the trend-corrected control can identify treatment effects, specific shapes of effects are given to the treated unit only in the post-treatment stage. By computing the distances of the counterfactual to the treated unit and comparing them to the specification of the simulated effects, the efficacy for the identification of the effect can be computed. Additionally, one might ask why the specification of trend corrected control uses the full series of candidate controls as covariates and not detrended series, another experiment in which forecasts only with trend and seasonal components are computed, one with trend, season, and residuals from detrended candidate controls, and finally the specification chosen for the other Monte Carlo experiments.

The donors are generated through unit roots

$$y_{0t,i} = y_{0t-1,i} + \varepsilon_{t,i}$$

where

$$\varepsilon_{t,i} = \lambda_t + \nu_{t,i}.$$

In this setting, λ_t is a vector of stochastic time effects drawn from $\mathcal{N}(0, 0.5)$, and the values for each $\nu_{t,i}$ are drawn from the columns of a multivariate normal distribution matrix ν where the mean vector is a column of zeros and a covariance matrix Σ such that it is generated from a correlation matrix R with off diagonals generated from a uniform (-1,1) random variable. The matrix Σ has diagonal elements set to 0.5 and is arranged to the nearest symmetric positive semidefinite matrix by utilizing the algorithm proposed in Higham (2002).

This design so far guarantees that the donors are $I(1)$ series, and that they exhibit correlation among them. The correlation ensures a realistic scenario in which the researcher would pick candidate controls that exhibit behavior which they expect to be relatively similar to the treated unit in the pre-treatment stage. In order to guarantee cointegration and the existence of weights for synthetic control, a set of $m < K$ uniform random variables is generated and a vector of length $K - m$ with zero entries is generated such that we have the vector U of length K specified as

$$U_i = \begin{cases} Uniform(0,1) & \text{if } 1 \leq i \leq m \\ 0 & \text{otherwise} \end{cases}$$

from which a weighting vector ω with the following components is created

$$\omega_i = \frac{U_i}{\sum_{i=1}^{K} U_i}.$$

In this work, $m = 4$. The reason for picking a small set of non-zero weighting units is because in most applied cases, the number of candidate units picked by synthetic control tends to be small. In addition, it is of interest to examine how well each of the algorithms deals with the increase of units that have real contribution to the treated unit. The sum of all entries in ω is exactly one and all entries are restricted in the convex hull boundaries. The treated unit is then generated by

$$y_{1t} = \sum_{i=1}^{k} \omega_i y_{0t.i} + \eta_t$$

where $\eta_t \sim N(0, 0.5)$. Furthermore, the Monte Carlo runs are generated with the following combinations of treatment times and series lengths

$$(t_0, T) \in \{(60, 80), (60, 100), (90, 120), (90, 150), (120, 160), (120, 200)\}$$

and values of $K \in \{9, 14, 19\}$.

The evaluation metrics for the Monte Carlos are contingent on which experiment we run. In order to evaluate the results of the Monte Carlo, in which no treatment effect is present, the different specifications for countefactuals act as forecasts for the treated unit. The best counterfactual then is one that shows better goodness of fit with the treated unit in the post-treatment period. Thus, for the experiments which test which specification for the donor effect is best, and that one which shows whether trend controlled overcomes the overrejection of the null of no effect, this work uses the measures

$$RMSE = \left[\frac{1}{T - t_0} \sum_{t=t_0}^{T} \left(\hat{y}_{1t}^{(0)} - y_{1t}^{(0)}\right)^2\right]^{\frac{1}{2}} \quad \text{and} \quad MAPE = \frac{1}{T - t_0} \sum_{t=t_0}^{T} \left|\frac{\hat{y}_{1t}^{(0)} - y_{1t}^{(0)}}{y_{1t}^{(0)}}\right|$$

to compare the relative sizes of an advantage between trend-corrected counterfactual and synthetic control, the following comparative measures are used

$$PError = \left(\frac{Error_{SC} - Error_{TC}}{Error_{SC}}\right) \times 100$$

as PRMSE and PMAPE respectively. In addition, for the experiment comparing synthetic control and trend corrected counterfactuals,a variation ratio in the spirit of R^2 can be used

$$R^2 = \frac{\sum_{t=t_0}^{T} \left(\hat{y}_{1t}^{(0)} - \bar{y}_{1t}^{(0)}\right)^2}{\sum_{t=t_0}^{T} \left(y_{1t}^{(0)} - \bar{y}_{1t}^{(0)}\right)^2}$$

and the p-values of an Anderson-Darling test with the null hypothesis of having the true counterfactual and the trend-corrected forecast coming from the same distribution can be used.

When generating effects, this work accounts for 3 types of trends, one in which the treatment effect is constant, one in which it's linear and one which follows a polynomial trend. The effects are simply deterministic functions for $t > t_0$ that are added to the generation scheme in this section for the post treatment series, and is denoted by δ_t. The constant effect is defined as

$$\delta_t = d_0 \text{ where } d_0 = 10$$

the linear effect as

$$\delta_t = d_1(t - t_0) \text{ where } d_1 = 0.2$$

and the polynomial effect as

$$\delta_t = d_0 + d_1(t - t_0) + d_2(t - t_0)^2 \text{ where } d_0 = 1,\ \ d_1 = 0.4,\ \ d_2 = -0.002.$$

To test whether the trend-corrected counterfactual can identify effects without overstating their importance, $\hat{\delta}_t$ can be defined as a vector of distances between the treated unit and the counterfactual

$$\hat{\delta}_t = y_{1t}^{(1)} - \hat{y}_{1t}^{(0)}.$$

Since the constant effect simply creates a parallel series to what the series would be if there were no effect, the effect can be tested by running the regression

$$\hat{\delta}_t = \beta_0 + \beta_1 (t - t_0)$$

for each simulation and seeing if the distribution of the fitted $\hat{\beta}_0$ is centered around the value of d_0 and if $\hat{\beta}_1$ is close to zero. For the linear and polynomial effects, a pseudo R^2 can indicate how well $\hat{\delta}_t$ represents δ_t

$$R^2 = \frac{\sum_{t=t_0}^{T} \left(\hat{\delta}_t - \bar{\delta}_t\right)^2}{\sum_{t=t_0}^{T} \left(\delta_t - \bar{\delta}_t\right)^2}.$$

5 Monte Carlo results

5.1 Specification search

A natural question regarding the specification of the trend-corrected artificial counterfactual is why the donor component is imposed as a combination of the entirety of the donor series rather than detrended versions. Moreover, is there really an improvement from using the donor effect versus using only the trend and seasonality components alone for the forecasts? To this end, the Monte Carlo compares the specification from the last section on the generated series with no effect and compares it to the following two GAM specifications:

$$Y_1^{(0)}(t) = \gamma_1(t) + \gamma_2(t) + \varepsilon_t$$

and

$$Y_1^{(0)}(t) = \gamma_1(t) + \gamma_2(t) + \gamma_4(t) + \varepsilon_t$$

where $\gamma_4(t)$ is defined as the pseudoresiduals from fitting the donor series only using trend and seasonal components.

The results are summarized in Tables 1 to 6, and graphical summaries are shown in Figures 1 and 2. The RMSE and MAPE measures are collected in terms of the post-treatment period meaning that a good forecast with low deviation errors indicates better counterfactuals.

A first view of the figures and tables indicate that in terms of RMSE, the specification of the trend-corrected control shows remarkably better counterfactuals than the alternatives. Furthermore, one can observe that for every level of t_0 a longer forecast horizon yields consistently to worse fits for all specifications. However, the RMSE for the trend-controlled specification is the less altered by longer forecast horizons even at short training times. An extreme case is the lower performing setups for treatment time and forecast horizon with $t_0 = 60$ and the number of donors set at 9. With the RMSE of the shape only counterfactuals increasing about 40.8 percent (from 1.96 to 2.76), the one fitting with pseudoresiduals increasing 26 percent (from 0.84 to 1.06), and the one with the full series only increasing 0.8 percent (from 0.847 to 0.854). Results are comparable across all different set-ups. MAPE shows a similar trend. Overall, one also observes that while all specifications lose forecasting power with longer forecast horizons shorter training times and a higher number of zero weighted donors, the specification chosen and defined in Section 3 outperforms the alternatives.

The inclusion of a donor component is significant according to this experiment. While the pseudoresidual component being one which filters the shape components of the donor pool seems like an intuitive choice for modelling, the results seem to indicate that there is information in the shapes of the donors that can work as a stronger forecast corrector than the pseudoresiduals. This justifies the choice of specification for the GAM as one that uses basic shapes and works in the same spirit as synthetic control.

5.2 Synthetic control comparison with no effect

The ultimate objective of this work is seeking to relax the assumption of trend stationarity which can lead to synthetic control to over reject the null of no effect. In order to illustrate this issue, we see in the simulations that under $I(1)$ series, synthetic control can generate results such as those shown in Figure 3, in which the vertical dotted line denotes t_0. Even if the pre-treatment fits are good, synthetic control can still produce results which might mislead the researcher into interpreting having found an effect when there is in fact none. Trend corrected control handles the same data by generating the results presented in Figure 4. The figures provide an encouraging preliminary result. However, the Monte Carlo experiment with a changing number of unweighted donors, through different forecasting horizons, and training times is necessary to have a complete assessment of the results.

The results of the simulations are summarized in Tables 7 to 10, and graphical summaries are in Figures 5 to 7. Goodness of fit for synthetic control and the trend-controlled counterfactual indicate better counterfactuals. Since the asymptotic behavior in the post-treatment controls leads to hard to define distributions, a combination of the Anderson-Darling test and a pseudo R^2 are presented in tandem. A rejection of the Anderson-Darling test means that the distributions between the countefactual and the observed series of the treated unit do not belong to the same distribution. The pseudo R^2 measure is simply a global ratio of variation about the sample mean of the post-treatment treated unit values.

A first observation from looking at the distribution of RMSE indicate both from the tables and the plots that trend-controlled distributions are to the left of those of synthetic control across the distribution for all cases. The indication being that, at least on this metric, trend-corrected control outperforms synthetic control. The immediate implications are that so far it seems that trend-corrected control is less likely to deviate of a true counterfactual in the post-treatment period. Further inspection shows that longer forecasting horizons for fixed treatment periods lead to increases in the error measure for both models. Meaning that the likelihood of wrongly identifying an effect increases for both. This is expected behavior for any forecasting model, and is further amplified by the non-stationary issue.

Moreover, the decrease in forecast power is more severe for the trend-corrected counterfactual than for the synthetic control. The median RMSE for the trend-corrected counterfactual increase between 10 to 22 percent in their value from short forecasting to long forecasting horizons. Synthetic control shows slightly more robust results exhibiting increases from 3 to 21 percent. It is important to note that such does not imply that the trend-corrected counterfactual ever underperforms synthetic control, showing to be from 22 to 57 percent smaller in the medians.

The rate of the advantage of trend-corrected counterfactual over synthetic control increases the longer the pre-treatment series is used for fitting the counterfactual. This is easily seen from Table 11 , the phenomenon is the same for both PRMSE and PMAPE. The table also shows that the median percent advantage on both measures increases under long forecasting horizons when $t_0 = 60$ and the number of donors is nine but it decreases for all other instances.

When talking about specifically the number of donors, for the same pre-treatment times and forecasting horizons, the behavior is not monotonic. Note that for PRMSE the percent advantage increases for most cases from nine to fourteen donors and then it drops when the number changes from fourteen to nineteen. Exceptions to this behavior are when $t_0 = 60$ with $T = 100$ and $t_0 = 90$ with $T = 150$ in which the number of donors reduces the advantage as it increases. It is of note that these special cases both have long forecasting horizons. PMAPE exhibits a similar behavior that PRMSE does across the same setups. This effect seems to indicate that trend-corrected counterfactual can identify series which have no actual contribution or that are not part of the cointegrating relation up to a certain extent. Its ability to do so is more reduced under longer forecasting horizons. It is still worthy of note that under all setups the proposed method maintains large positive percent advantages suggesting that is still significantly less likely than synthetic control to deviate from the true counterfactual and hence less likely to identify treatment effects when there are none.

The values for the pseudo R^2 are clipped at zero for whenever the measure indicated behavior worse than a horizontal line leading to values larger than one. These are shown in conjunction with the Anderson-Darling p-values. The reason for this joint presentation is because of the lack of knowledge of the asymptotic behavior of the forecast, thus beyond presenting simply a goodness of fit measure, assessing whether the distribution of the counterfactual and the forecast come from the same distribution is a good proxy measure to assess whether the null hypothesis of no effect fails to be rejected or not.

The results encapsulated in Table 10 and Figure 7 suggest that overwhelmingly, synthetic control simulations have low R^2 values and reject the null hypothesis of the synthetic control and actual counterfactual being from the same distribution which is consistent with the violation of stationarity. The same figure suggests that most simulation setups for the trend-corrected counterfactual lie on the top right quadrant. Thus, for a large number of cases the proposed method fails to reject the Anderson-Darling test at a 0.05 level of significance and produces high goodness of fit with $R^2 \geq 0.5$. The results that show that the trend-corrected counterfactual in the rejection area or with low R^2 values are all the cases in which $t_0 = 60$ and a case for $t_0 = 90$ with $T = 150$. This suggests that the trend-corrected counterfactual is not invulnerable to over-rejection of the null of no effect when in the presence of situations with long forecasting horizons and short training periods.

The results of this experiment suggest that under sufficiently long pre-treatment periods and short enough forecasting horizons, and cointegration, the trend-corrected counterfactual is indeed more robust to the issues that synthetic control has in the presence of $I(1)$ series. While not completely invulnerable to the number of irrelevant donors, the method has a capacity to suppress their effect if they are few.

5.3 Effect identification

While the Monte Carlo experiment with no treatment effect implies that the trend-corrected counterfactual is close to a true counterfactual in the experiment's conditions and hence is is implicitly able to identify true treatment effects, it is important to determine whether it can identify these effects without overstating or understating them significantly. The design for this experiment deals with three types of effects: a constant, linear and polynomial trend effects. The results are summarized in Tables 11 to 15, Figures 8 and 9.

The constant effect is tested via regressions for each simulation set up in which $\delta_t = 10$, meaning that the intercept parameter should be centered at around 10 and the slope about zero. From the results of the tables and figures for this test, it is of note that the distribution in all setups, be it for long or short forecasting horizons are centered indeed closely about 10 and are mostly symmetric for the intercept parameter. Looking at the density plots and five number summaries for the slope parameter also suggest the distributions being strongly centered about zero. However, it is important to note that the instances in which $t_0 = 60$ show higher dispersion than the other cases, flattening out the distribution and sometimes creating multi-modality. Yet, these distributions still show to have relatively low bias from the "true" parameter values.

Linear effects are identified through a pseudo R^2 measure defined in Section 4 in which higher values can be interpreted as accuracy on the determination of the magnitude of the effect. The five-number summaries and density plots suggest that the pseudo R^2 distributions are left-skewed about 0.5 with the exception of the cases in which $t_0 = 60$. In these cases, dispersion is also higher than in the other cases with heavier left tails. Thus implying that shorter training times are more unstable than those which are not and thus sometimes having identified effects that are either overstated or understated more often than in any other instance in the experiment.

In a similar fashion polynomial effects are also identified through a pseudo R^2 measure. The results are consistent with those of the linear effect showing similar distribution shapes. However, in this case the distributions are all left-skewed about 0.5. While the cases in which $t_0 = 60$ are centered at lower values than the other instances, one sees that the distributions are centered further to the right than those of the linear effect analysis. Overall, the method provides reliable effect sizes when allowed to train for long enough.

6 Empirical Application

The Monte Carlo experiments provided insight on the effectiveness of the method under astringent conditions from its design. In order to judge the trend-controlled counterfactual, work with real data is necessary. It is additionally convenient to utilize examples that have well known results and have been corroborated and replicated many times. This work picks the well cited analysis from Abadie and Gardeazabal (2003) as a benchmark for replication of results. The study assesses whether German reunification after the fall of the Berlin Wall in 1990 had economic cost for West Germany which had GDP approximately 3 times higher than East Germany. The study confirms indeed that there is economic cost to reunification. A reason for choosing this particular study beyond its well known results are because the outcome variable analyzed is GDP which is known to be non-stationary. However, the donor pool is constricted from OECD countries, and they all have increasing GDP per capita from 1960 to 2003 without largely apparent fluctuations as shown in Figure 10. This implies that the results from Abadie and Gardeazabal (2003) are less likely to have been susceptible to the issues caused by the violation of stationarity. Moreover, the pre-treatment period available for "learning" the series has only 30 periods. The results in Section 5 suggest that short pre-treatment periods are unfavorable for the trend-corrected control. Hence, if the method is able to perform well in terms of placebo tests and confirm the results of the original work, then it is possible to further confirm its usefulness. The data for this replication were obtained from Hainmueller (2014) which has the exact data for replication from the study.

Since the trend-corrected counterfactual depends on there existing a cointegrating relationship between the outcome variable of the treated unit and those of the candidate controls, a good evaluation of the method requires to test if such a relationship exists in practice. Fortunately, the choice of OECD countries picked with non-zero weights by the original paper (Austria, Japan, Netherlands, Switzerland and USA) can be shown to be stationary by running the regression of the series

$$
\begin{aligned}
GDP_{West\,Germany} = \beta_1 GDP_{Austria} + \beta_2 GDP_{Japan} + \beta_3 GDP_{Netherlands} \\
+ \beta_4 GDP_{Switzerland} + \beta_5 GDP_{USA}
\end{aligned}
$$

and performing an augmented Dickey-Fuller test on the residuals. If these turn out to be stationary, then the cointegrating relationship exists as specified. The results of the test are presented in Table 16. Thus, trend-corrected counterfactual is fitted with only the donor pool in the relationship. Since cointegrating relationships can be tested and the Monte Carlo section showed inestability in the presence of many non-cointegrating donors, following these steps is the best suggested practice.

Figure 11 shows a comparison of the results from synthetic control and trend-corrected side to side. The similarities between the two graphs confirm both the effect found in Abadie and Gardeazabal (2003), and the effectiveness of the proposed method. However, this similarities need to be confirmed as pertaining to the ability of the trend-corrected counterfactual of synthesizing a plausible control unit and not simply to a decay in forecast ability across time. This is a particularly important point to examine, as the number of pre-treatment periods is small ($t_0 = 30$).

Since the trend-corrected counterfactual is particularly vulnerable to training under short series, the placebos were placed in order to allow at least 22 periods for fitting. Figure 12 shows the results of changing the treatment periods to 1982, 1984 and 1986 respectively. The figure shows that before the actual treatment period (1990), the placebos show only small divergences from the treated unit series, affirming that the counterfactuals generated seem adequate to use.

The results from this short replication indicate that even under adverse conditions, the proposed model can generate plausible enough results to at the very least serve as a further robustness check for other methods when the series analyzed are $I(1)$ and the treatment unit has a cointegrating relation with members of the donor pool.

7 Concluding Remarks

Trend-corrected artificial counterfactual is a proposed method for approximating counterfactuals which exists in the same spirit as synthetic control but uses an automatic forecasting specification via a GAM to achieve its purpose. This proposal is born to relax the assumption of trend stationarity in synthetic control. Synthetic control is vulnerable to overrejection of the null of no treatment effect under $I(1)$ series, meaning that it can indicate treatment effects when there are none. Using three different Monte Carlo experiments, the method is put to test to confirm that the specification is adequate, that it overcomes the difficulties found in synthetic control, and that in the presence of actual treatment effects, the counterfactual does not understate or overstate their magnitude significantly. For different sets of training lengths, forecasts horizons and number of non cointegrating peers, the proposal shows an overall likelihood of being a correct specification when compared to two alternatives: one using only trend and seasonal components, and one using the same components but adding a linear combination of pseudoresiduals. The second experiment shows that the methodology is far less vulnerable to $I(1)$ series than synthetic control by showing not only better measures of fit against a "true" counterfactual, but also by proving itself belonging to be often from the same distribution. The third experiment uses linear regression and goodness of fit against distances from the artificial counterfactual and generated deterministic effects. The results of this experiment show relatively low bias across all cases.

It is worthy of note that the weaknesses of the model also become evident through the Monte Carlo experiments. The first vulnerability is that the cointegration assumption is strong. This is shown by noting that even if the method maintained relatively good properties in the presence of non-cointegrating donors, it performed unstably and eventually worsened as the number of these grew, making it unadvisable to use donors that do not show cointegrating relationships with the treatment unit. A second important vulnerability is that deviation from the model worsens as the forecasting horizon increases, making it only applicable if the forecasting horizon is significantly shorter than the number of periods in which the model is fit. The third and most inconvenient vulnerability comes in the need for long periods for fitting the series. All these vulnerabilities are consistent with those found in most forecasting methodologies, noting that this is not a technique that is recommendable under every possible scenario when meeting with the issue of $I(1)$ series, but a restricted viable alternative that can often be more reliable than synthetic control results when the trend stationarity assumption is violated.

The seminal work of Abadie and Gardeazabal (2003) is chosen as empirical benchmark for the model. The data in the German reunification analysis is comprised of non-stationary series, albeit increasing ones making the problems with the lack of stationarity be diminished as the validation of the method in the original paper show the conclusion to be robust. Fortunately, synthetic control gives weights to a set of donors that happen to have a cointegrating relationship with the treatment unit and hence become the donor components for the trend-corrected control and allow for easy comparison. However, the model proves to be a particularly tough benchmark for the trend-corrected control due to allowing only 30 periods for fitting. Nevertheless, the results obtained from the method closely follow those obtained in the original paper despite using a different methodology. In a placebo test with differing treatment times it is shown that the divergence from the treated unit after the treatment effect comes from the actual effect rather than simply a loss of forecasting accuracy over time.

The methodology proposed is one that seeks to use the predictive power of modern estimation techniques borrowed from machine learning literature to help further reduce assumptions in causal inference models. While not fully robust to data density, the method can be a viable alternative or at the very least a useful robustness check to complement other tools in the artificial counterfactual toolkit.

References

Abadie, A., Diamond, A. and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, **105** (490), 493–505.

— and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, **93** (1), 113–132.

Carvalho, C., Masini, R. and Cunha Medeiros, M. (2016). The perils of counterfactual analysis with integrated processes. *SSRN working paper*.

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

Ferman, B. and Pinto, C. (2016). *Revisiting the Synthetic Control Estimator*. Mpra paper, University Library of Munich, Germany.

Hainmueller, J. (2014). Replication data for: Comparative Politics and the Synthetic Control Method.

Harvey, A. C. and Peters, S. (1990). Estimation procedures for structural time series models. *Journal of Forecasting*, **9** (2), 89–108.

— and Shephard, N. (1995). Handbook of statistics, volume 11, econometrics: G.s. maddala, c.r. rao and h.d. vinod, eds., 1993, (north holland, amsterdam), 800 pp., us$180.00 df340.00, isbn 0-444-89577-9. *International Journal of Forecasting*, **11** (1), 189–191.

Hastie, T. and Tibshirani, R. (1987). Generalized additive models: Some applications. *Journal of the American Statistical Association*, **82** (398), 371–386.

Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, **22** (3), 329–343.

Hyndman, R. and Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software, Articles*, **27** (3), 1–22.

Taylor, S. J. and Letham, B. (2018). Forecasting at scale. *The American Statistician*, **72** (1), 37–45.

Tibshirani, R. J. *et al.* (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, **42** (1), 285–323.

Tables and Figures

Table 1: RMSE distributions for shape only simulations

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	max	mean
9	60	80	0.78914	1.43590	1.96262	3.38544	8.31973	2.50433
9	60	100	0.98468	1.94527	2.76634	4.51160	11.47539	3.45532
9	90	120	0.82632	1.69201	2.73403	4.60305	11.33523	3.28580
9	90	150	0.91901	2.58588	4.21789	6.73165	15.11855	5.01310
9	120	160	1.14805	2.10810	3.51918	4.75125	12.12152	3.76283
9	120	200	1.42872	2.69864	4.94550	7.33599	13.77090	5.32982
14	60	80	0.65226	1.70133	2.71551	4.02340	7.71511	2.96907
14	60	100	0.85620	2.08563	3.14956	4.77494	10.38072	3.56895
14	90	120	0.96852	1.95131	2.72142	3.87916	12.92792	3.17583
14	90	150	0.86817	2.31842	3.01161	5.28999	12.92096	4.20666
14	120	160	1.00479	1.99635	3.02117	4.59923	9.14653	3.43836
14	120	200	1.13008	2.90600	4.22905	6.30937	16.19830	5.14290
19	60	80	0.76582	1.60886	2.20201	3.58573	9.30600	2.79832
19	60	100	1.06194	2.20436	3.46988	5.06372	9.95823	3.90862
19	90	120	0.81011	1.83065	2.81805	3.96383	9.55248	3.21719
19	90	150	1.15783	2.57742	3.89073	5.74285	12.39690	4.45853
19	120	160	1.00820	1.88676	2.88044	4.83899	9.10227	3.53482
19	120	200	1.34903	2.76846	4.04378	6.14141	14.32905	4.77675

Table 2: RMSE distributions for residual fitted simulations

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	max	mean
9	60	80	0.51744	0.61869	0.83881	1.09349	3.06736	0.95734
9	60	100	0.51670	0.75479	1.05896	1.75788	8.77251	1.58285
9	90	120	0.44434	0.62928	0.80906	1.07740	5.18082	1.04541
9	90	150	0.53669	0.99585	2.06360	3.54813	8.52472	2.53270
9	120	160	0.43357	0.71174	1.03080	1.72036	17.54615	1.62255
9	120	200	0.54667	1.32539	2.60669	5.16300	12.59416	3.46027
14	60	80	0.44231	0.73710	1.00674	1.57174	4.48751	1.30269
14	60	100	0.56569	0.90542	1.61018	2.66030	10.33288	2.03851
14	90	120	0.49472	0.69018	0.92084	1.48073	5.08330	1.22635
14	90	150	0.62114	1.43168	2.04786	3.67197	9.43334	2.68645
14	120	160	0.45123	0.76667	1.04186	1.48344	8.36502	1.38446
14	120	200	0.53554	1.18860	2.33995	4.73978	11.61786	3.01271
19	60	80	0.47762	0.83952	1.24873	1.96029	4.23353	1.50896
19	60	100	0.55315	0.99140	1.64627	3.06868	8.05009	2.23840
19	90	120	0.46295	0.73838	1.05008	1.57136	4.56595	1.24277
19	90	150	0.53242	1.12874	2.54066	4.03470	10.86345	3.11803
19	120	160	0.49033	0.71685	1.30876	2.46783	5.85826	1.75313
19	120	200	0.61163	1.57142	2.71787	4.44330	8.28873	3.12458

Table 3: RMSE distributions for full fitted donor series simulations

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	max	mean
9	60	80	0.43097	0.64645	0.84692	1.11581	2.69815	0.94997
9	60	100	0.49784	0.68143	0.85427	1.23782	3.52636	1.02680
9	90	120	0.40725	0.55585	0.64028	0.75401	1.94999	0.71339
9	90	150	0.45516	0.63649	0.77330	1.18026	3.13339	0.96474
9	120	160	0.41719	0.57988	0.65175	0.79541	1.55984	0.71253
9	120	200	0.49016	0.60823	0.71527	0.91001	2.71487	0.81954
14	60	80	0.41787	0.70582	0.88052	1.18853	2.76755	1.00167
14	60	100	0.53175	0.78760	1.06696	1.67223	3.76562	1.32794
14	90	120	0.45919	0.60832	0.70861	0.92093	1.95844	0.82088
14	90	150	0.52878	0.66631	0.83413	1.09433	2.39658	0.93961
14	120	160	0.46240	0.60066	0.65513	0.80360	1.65332	0.73690
14	120	200	0.47632	0.62528	0.72649	0.93163	5.37172	0.89049
19	60	80	0.47701	0.79408	0.94063	1.22621	3.32827	1.14732
19	60	100	0.54614	0.86567	1.18032	1.88809	5.57098	1.49588
19	90	120	0.46076	0.66824	0.79598	1.00745	2.53512	0.90929
19	90	150	0.52127	0.73588	0.97873	1.49751	2.83381	1.19296
19	120	160	0.43258	0.59925	0.71265	0.95054	1.69668	0.81245
19	120	200	0.49816	0.68565	0.81120	1.11343	2.90566	0.95914

Table 4: MAPE distributions for shape only simulations

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	max	mean
9	60	80	0.07624	0.30047	0.89184	3.22377	95.64540	4.43325
9	60	100	0.15379	0.41855	1.06744	4.13807	186.15164	4.70772
9	90	120	0.06776	0.39418	0.72581	2.05728	31.18369	2.44224
9	90	150	0.09380	0.53553	1.28487	4.43114	46.14471	4.30532
9	120	160	0.06728	0.32737	0.67849	1.91004	98.33533	3.00725
9	120	200	0.08076	0.37998	1.09418	3.44992	79.71574	3.58115
14	60	80	0.04221	0.43180	1.11490	4.78805	109.11012	4.45477
14	60	100	0.14476	0.54443	1.06297	4.01058	2328.28143	27.91972
14	90	120	0.08114	0.36889	0.76431	2.70078	69.90193	3.24436
14	90	150	0.07558	0.36277	0.81208	2.26303	137.79806	4.29325
14	120	160	0.06301	0.26928	0.79411	2.79749	25.83002	2.29077
14	120	200	0.04533	0.69624	2.17715	6.87606	108.71918	6.79175
19	60	80	0.10101	0.32568	0.74681	2.04756	78.57383	3.32530
19	60	100	0.11050	0.36623	1.10881	3.61479	56.25099	3.84678
19	90	120	0.07686	0.34079	0.81913	3.74523	94.63530	4.02190
19	90	150	0.08011	0.42397	0.89420	2.97842	65.12796	4.88408
19	120	160	0.07170	0.40335	0.83783	2.18734	102.74260	4.77775
19	120	200	0.06668	0.43172	0.89063	2.17328	53.78426	2.95248

Table 5: MAPE distributions for residual fitted simulations

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	max	mean
9	60	80	0.05305	0.13817	0.23760	1.01940	106.30539	2.40368
9	60	100	0.06558	0.16595	0.44226	1.38302	46.85956	1.72370
9	90	120	0.03623	0.09896	0.22817	0.61286	8.50888	0.68885
9	90	150	0.02936	0.23206	0.65699	1.91399	17.86707	1.78501
9	120	160	0.03275	0.10584	0.26293	0.81520	15.28490	0.85918
9	120	200	0.04140	0.19477	0.43505	2.23964	36.68060	2.23527
14	60	80	0.04783	0.19458	0.42957	1.16387	60.27193	1.78325
14	60	100	0.07412	0.22406	0.56586	2.81284	1246.37756	16.45917
14	90	120	0.03964	0.11597	0.26890	1.21943	29.53519	1.27532
14	90	150	0.05035	0.20118	0.41314	1.77318	80.64181	2.96381
14	120	160	0.02952	0.11381	0.27760	0.82432	5.06808	0.62797
14	120	200	0.03808	0.28096	1.02821	3.61891	37.82111	3.02370
19	60	80	0.03089	0.19725	0.40284	1.29408	33.13388	1.64235
19	60	100	0.05319	0.17956	0.47737	2.04891	39.39272	2.49399
19	90	120	0.02629	0.16171	0.36220	1.03790	13.64123	1.30946
19	90	150	0.06517	0.21680	0.54811	2.36614	28.01491	2.70363
19	120	160	0.02295	0.14925	0.38106	0.99564	44.34956	2.31556
19	120	200	0.03839	0.21041	0.55324	1.55177	20.95111	1.53157

Table 6: MAPE distributions for full fitted donor series simulations

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	max	mean
9	60	80	0.03119	0.14089	0.27687	1.05335	83.48026	2.28828
9	60	100	0.03947	0.11962	0.36816	1.19979	34.67964	1.21195
9	90	120	0.02791	0.06668	0.16085	0.47360	8.09182	0.52996
9	90	150	0.02803	0.11123	0.25401	0.79449	24.47235	0.91575
9	120	160	0.02935	0.06299	0.14280	0.51291	11.64723	0.54619
9	120	200	0.02881	0.06239	0.15097	0.50235	4.13942	0.47340
14	60	80	0.03557	0.16653	0.37334	1.25558	31.87853	1.31440
14	60	100	0.05796	0.18984	0.35807	2.10094	542.25153	8.63477
14	90	120	0.03335	0.09485	0.22977	0.96691	9.20977	0.80520
14	90	150	0.02804	0.09507	0.25494	0.68329	11.35568	0.72951
14	120	160	0.02336	0.06873	0.12333	0.52625	2.73452	0.40907
14	120	200	0.02870	0.09091	0.43894	0.99290	9.09739	1.04276
19	60	80	0.04319	0.15673	0.29958	0.71270	34.13010	1.38192
19	60	100	0.06398	0.13756	0.38838	1.20620	31.08355	1.77954
19	90	120	0.03189	0.11089	0.26418	0.90615	11.62897	0.85528
19	90	150	0.03125	0.13021	0.26571	0.87274	19.50913	1.25496
19	120	160	0.02352	0.08588	0.16177	0.72996	43.10497	1.40800
19	120	200	0.02686	0.07697	0.17371	0.51519	9.96159	0.58728

Table 7: RMSE for synthetic control on simulations with no treatment effect

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	max	mean
9	60	80	0.46348	0.71910	1.17886	1.88232	6.94243	1.45235
9	60	100	0.53427	0.88645	1.24986	1.82435	5.56186	1.51701
9	90	120	0.43880	0.92192	1.43499	2.18711	6.58607	1.67140
9	90	150	0.55239	1.05028	1.53412	2.41833	6.31954	1.90445
9	120	160	0.49191	0.95729	1.48178	2.24413	5.15087	1.69767
9	120	200	0.50105	1.18015	1.72978	2.46392	7.39993	1.99226
14	60	80	0.52262	0.89099	1.29443	1.99653	4.38489	1.55448
14	60	100	0.53500	0.99364	1.43127	2.26775	8.71648	1.84981
14	90	120	0.53141	1.07756	1.42588	2.27168	6.54282	1.82404
14	90	150	0.58871	1.11188	1.62856	2.36259	6.25511	1.99013
14	120	160	0.57704	1.27737	1.89006	2.82026	6.98552	2.15713
14	120	200	0.72269	1.36060	1.95036	3.44844	9.31937	2.61791
19	60	80	0.49798	0.96843	1.47911	2.08409	5.14541	1.69582
19	60	100	0.58380	1.06720	1.45891	2.17231	6.11633	1.80002
19	90	120	0.53486	1.07903	1.58022	2.41438	5.76935	1.88496
19	90	150	0.65431	1.31025	2.03936	2.92973	7.50374	2.33417
19	120	160	0.51529	1.20337	1.73086	3.01125	6.77017	2.28597
19	120	200	0.54475	1.46979	2.04500	3.02830	7.60490	2.45657

Table 8: RMSE for trend corrected control on simulations with no treatment effect

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	max	mean
9	60	80	0.47770	0.64026	0.87614	1.13569	2.33305	0.95400
9	60	100	0.43154	0.69634	0.86735	1.16975	3.54721	1.02971
9	90	120	0.32012	0.58922	0.65824	0.77007	1.67846	0.71469
9	90	150	0.47225	0.64213	0.75371	1.09754	2.29471	0.88898
9	120	160	0.44132	0.56629	0.63486	0.77869	1.98918	0.70833
9	120	200	0.47215	0.61519	0.70585	0.89135	1.63897	0.81075
14	60	80	0.45326	0.67498	0.88502	1.24782	2.49339	1.00218
14	60	100	0.49193	0.77767	1.01856	1.34076	3.75415	1.18960
14	90	120	0.46582	0.57114	0.70801	0.81815	2.34553	0.76131
14	90	150	0.53083	0.65696	0.82853	1.15283	2.88322	1.04007
14	120	160	0.41447	0.57271	0.69449	0.82811	2.65901	0.75412
14	120	200	0.51040	0.62558	0.76723	1.03941	3.05605	0.90347
19	60	80	0.41122	0.71468	0.95651	1.34749	3.40535	1.12204
19	60	100	0.48727	0.84520	1.16931	1.67917	5.63032	1.43395
19	90	120	0.48293	0.66421	0.81406	1.04524	1.84578	0.89256
19	90	150	0.48062	0.75441	1.02790	1.33303	4.49894	1.17560
19	120	160	0.45052	0.59764	0.72769	0.90356	2.61610	0.82753
19	120	200	0.47765	0.70694	0.83608	1.23373	2.67788	1.02753

Table 9: MAPE for synthetic control on simulations with no treatment effect

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	max	mean
9	60	80	0.04745	0.15997	0.39594	0.89311	287.23884	3.81838
9	60	100	0.04064	0.17927	0.41885	1.56711	28.02064	1.65726
9	90	120	0.03449	0.15606	0.39174	1.39394	365.76863	5.43099
9	90	150	0.04625	0.17712	0.51355	1.49649	29.90947	1.92431
9	120	160	0.02905	0.16453	0.32392	1.01404	137.63296	4.22142
9	120	200	0.04603	0.17232	0.31480	0.93909	17.28009	1.11570
14	60	80	0.03088	0.24432	0.45463	1.98630	66.49292	2.91488
14	60	100	0.06632	0.21418	0.44502	1.32763	30.01348	1.95976
14	90	120	0.05786	0.17877	0.47005	2.04000	10.87292	1.36612
14	90	150	0.03866	0.17439	0.44174	1.31671	36.44003	1.50993
14	120	160	0.02991	0.16988	0.38099	0.99283	73.94772	1.63588
14	120	200	0.05683	0.25011	0.58919	1.86357	101.37268	3.10667
19	60	80	0.03807	0.22130	0.47704	1.87812	26.56683	1.76905
19	60	100	0.07136	0.23230	0.47126	1.31641	19.65374	1.64932
19	90	120	0.05331	0.19799	0.44367	2.29687	22.73646	1.77194
19	90	150	0.03395	0.21114	0.49411	1.93445	42.83328	2.59136
19	120	160	0.05465	0.20173	0.52600	1.28884	106.53358	2.99188
19	120	200	0.02407	0.20694	0.49174	1.77859	678.40354	8.94053

Table 10: Median R^2 and median Anderson-Darling p-values

			Synthetic Control		Trend-corrected	
Number of peers	t_0	T	Anderson-Darling p-values	R^2	Anderson-Darling p-values	R^2
9	60	80	0.00870	0.00000	0.10356	0.40869
9	60	100	0.00584	0.38631	0.09331	0.64934
9	90	120	0.00121	0.12875	0.45532	0.73318
9	90	150	0.00013	0.20765	0.15766	0.79218
9	120	160	0.00185	0.25433	0.50814	0.82657
9	120	200	0.00011	0.36243	0.18405	0.87424
14	60	80	0.00622	0.00000	0.09318	0.28054
14	60	100	0.00085	0.00000	0.04006	0.49277
14	90	120	0.00103	0.00000	0.32685	0.72945
14	90	150	0.00015	0.13858	0.07534	0.71521
14	120	160	0.00003	0.00000	0.41330	0.77922
14	120	200	0.00000	0.01225	0.21470	0.84356
19	60	80	0.00255	0.00000	0.13698	0.19802
19	60	100	0.00069	0.06370	0.01233	0.35236
19	90	120	0.00030	0.00000	0.18857	0.64292
19	90	150	0.00000	0.00000	0.03897	0.74296
19	120	160	0.00005	0.00000	0.29222	0.77332
19	120	200	0.00000	0.00000	0.05676	0.78567

Table 11: Median PRMSE and median PMAPE for simulations without a treatment effect

Number of peers	t_0	T	PRMSE	PMAPE
9	60	80	21.17792	24.58095
9	60	100	28.07255	30.21325
9	90	120	52.20061	51.04480
9	90	150	49.73589	47.87039
9	120	160	53.07136	50.28517
9	120	200	51.82751	52.54214
14	60	80	36.11235	30.65823
14	60	100	22.15180	26.63135
14	90	120	53.87217	57.71388
14	90	150	43.79651	48.11551
14	120	160	61.88194	61.52935
14	120	200	58.03965	61.27640
19	60	80	32.80247	34.93726
19	60	100	20.23104	22.11016
19	90	120	45.27290	50.19871
19	90	150	48.68420	47.76010
19	120	160	56.54778	59.64454
19	120	200	55.51456	56.43091

Table 12: Constant effect Monte Carlo parameter distributions for intercept parameter

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	maximum	mean
9	60	80	8.99942	9.68566	10.01725	10.37308	12.07162	10.04585
9	60	100	9.18966	9.64689	9.99475	10.27452	11.72224	10.02389
9	90	120	9.19764	9.80563	10.01965	10.23495	11.34852	10.02330
9	90	150	8.35178	9.75229	10.10104	10.33289	10.82644	10.05584
9	120	160	9.32970	9.88578	10.11323	10.26437	11.05558	10.08067
9	120	200	9.05791	9.77822	10.03403	10.20787	11.05705	10.01391
14	60	80	8.41367	9.69258	10.04977	10.47736	11.37430	10.06782
14	60	100	7.33638	9.44179	9.92597	10.29861	12.01238	9.89409
14	90	120	8.52311	9.71005	9.99944	10.18920	11.05618	9.96083
14	90	150	8.79134	9.68309	9.94947	10.26960	10.83827	9.96024
14	120	160	9.20167	9.71619	9.99227	10.18024	10.99871	9.97413
14	120	200	8.94543	9.67581	9.92165	10.19719	11.20971	9.95675
19	60	80	8.40945	9.64507	9.99670	10.48546	12.10741	10.04373
19	60	100	8.13932	9.62622	10.12720	10.61515	11.35551	10.09992
19	90	120	8.76901	9.75883	10.03545	10.32268	11.31373	10.04410
19	90	150	8.58043	9.53747	9.99985	10.44683	11.87490	10.01847
19	120	160	8.79590	9.80203	10.03616	10.32836	10.96852	10.04922
19	120	200	8.79558	9.66411	10.04809	10.32038	11.00716	10.00428

Table 13: Constant effect Monte Carlo parameter distributions for slope parameter

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	maximum	mean
9	60	80	-0.22490	-0.04013	-0.00885	0.02720	0.16055	-0.00315
9	60	100	-0.12568	-0.02405	0.00234	0.01981	0.09434	-0.00152
9	90	120	-0.10361	-0.02382	-0.00278	0.01437	0.08471	-0.00509
9	90	150	-0.05491	-0.00953	-0.00158	0.01086	0.04398	-0.00021
9	120	160	-0.03963	-0.00723	-0.00125	0.01232	0.05442	0.00081
9	120	200	-0.03657	-0.00871	-0.00029	0.00637	0.04778	-0.00127
14	60	80	-0.17664	-0.03360	-0.00406	0.03938	0.18684	-0.00102
14	60	100	-0.12216	-0.03051	0.00808	0.03546	0.12839	0.00389
14	90	120	-0.06712	-0.02119	0.00693	0.02290	0.10373	0.00365
14	90	150	-0.06615	-0.01379	0.00005	0.01312	0.04672	-0.00087
14	120	160	-0.04628	-0.00793	-0.00022	0.00992	0.05176	-0.00008
14	120	200	-0.03175	-0.00896	-0.00185	0.00790	0.03018	-0.00119
19	60	80	-0.17075	-0.04709	0.00662	0.04648	0.17498	0.00742
19	60	100	-0.22970	-0.02977	-0.00318	0.02796	0.10677	-0.00477
19	90	120	-0.07067	-0.01929	0.00117	0.02791	0.10803	0.00554
19	90	150	-0.05709	-0.02282	-0.00207	0.01367	0.08928	-0.00103
19	120	160	-0.05503	-0.01248	-0.00081	0.01647	0.07335	0.00242
19	120	200	-0.04986	-0.01292	-0.00227	0.00787	0.05106	-0.00332

Table 14: Constant effect Monte Carlo R^2 distributions for linear effect magnitude

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	maximum	mean
9	60	80	0.00000	0.06238	0.49836	0.67613	0.89206	0.42303
9	60	100	0.00000	0.74867	0.85116	0.91334	0.95635	0.77561
9	90	120	0.00000	0.74172	0.85551	0.88824	0.93286	0.79168
9	90	150	0.65099	0.91849	0.95100	0.96663	0.98563	0.93305
9	120	160	0.64034	0.89062	0.92338	0.94290	0.95950	0.90032
9	120	200	0.55557	0.96009	0.97749	0.98305	0.98890	0.96139
14	60	80	0.00000	0.00000	0.45803	0.63057	0.87558	0.37380
14	60	100	0.00000	0.46571	0.76716	0.88364	0.93318	0.65224
14	90	120	0.00000	0.72106	0.83547	0.88766	0.95699	0.75631
14	90	150	0.36771	0.90507	0.94888	0.96763	0.98051	0.91011
14	120	160	0.62479	0.89556	0.92349	0.94319	0.96596	0.90352
14	120	200	0.78512	0.95734	0.97469	0.98136	0.98883	0.96027
19	60	80	0.00000	0.00000	0.41600	0.61704	0.86845	0.36321
19	60	100	0.00000	0.43350	0.77085	0.86879	0.93928	0.63396
19	90	120	0.00000	0.62509	0.79514	0.87585	0.94638	0.70471
19	90	150	0.35722	0.81201	0.91853	0.95879	0.98273	0.86691
19	120	160	0.43066	0.85501	0.91158	0.93644	0.96320	0.87811
19	120	200	0.52241	0.92198	0.96895	0.97891	0.98988	0.93205

Table 15: Constant effect Monte Carlo R^2 distributions for polynomial effect magnitude

number of peers	t_0	T	minimum	25th percentile	median	75th percentile	maximum	mean
9	60	80	0.00000	0.69678	0.82094	0.88633	0.96634	0.75475
9	60	100	0.07926	0.90074	0.93963	0.96424	0.98187	0.89716
9	90	120	0.54027	0.90875	0.94501	0.96074	0.97654	0.92380
9	90	150	0.82071	0.95755	0.97394	0.98217	0.99265	0.96527
9	120	160	0.85777	0.95579	0.96867	0.97624	0.98357	0.96001
9	120	200	0.69529	0.97249	0.98450	0.98801	0.99223	0.97334
14	60	80	0.00000	0.63720	0.80605	0.86649	0.96120	0.71399
14	60	100	0.00000	0.78166	0.90721	0.95170	0.97368	0.84166
14	90	120	0.54884	0.89558	0.94142	0.95768	0.98347	0.91060
14	90	150	0.67353	0.95111	0.97318	0.98332	0.99003	0.95350
14	120	160	0.85221	0.95707	0.96909	0.97635	0.98659	0.96084
14	120	200	0.85071	0.97042	0.98247	0.98720	0.99234	0.97243
19	60	80	0.00000	0.53743	0.79505	0.87476	0.95465	0.67824
19	60	100	0.00000	0.77508	0.90974	0.94706	0.97518	0.82397
19	90	120	0.48128	0.86901	0.92706	0.95207	0.98082	0.89139
19	90	150	0.67139	0.90382	0.95798	0.97770	0.99112	0.93139
19	120	160	0.77574	0.94194	0.96408	0.97413	0.98550	0.95073
19	120	200	0.67255	0.94642	0.97871	0.98531	0.99285	0.95310

Table 16: ADF test results for confirming cointegrating relationship between the per capita GDP of West Germany and the selected controls from Abadie and Gardeazabal (2003)

Augmented Dickey-Fuller Results	
Dickey-Fuller Statistic	-4.0575
Lag order	3
p-value	0.01632
Alternative hypothesis	Stationary

Figure 1: RMSE comparisons for different specifications

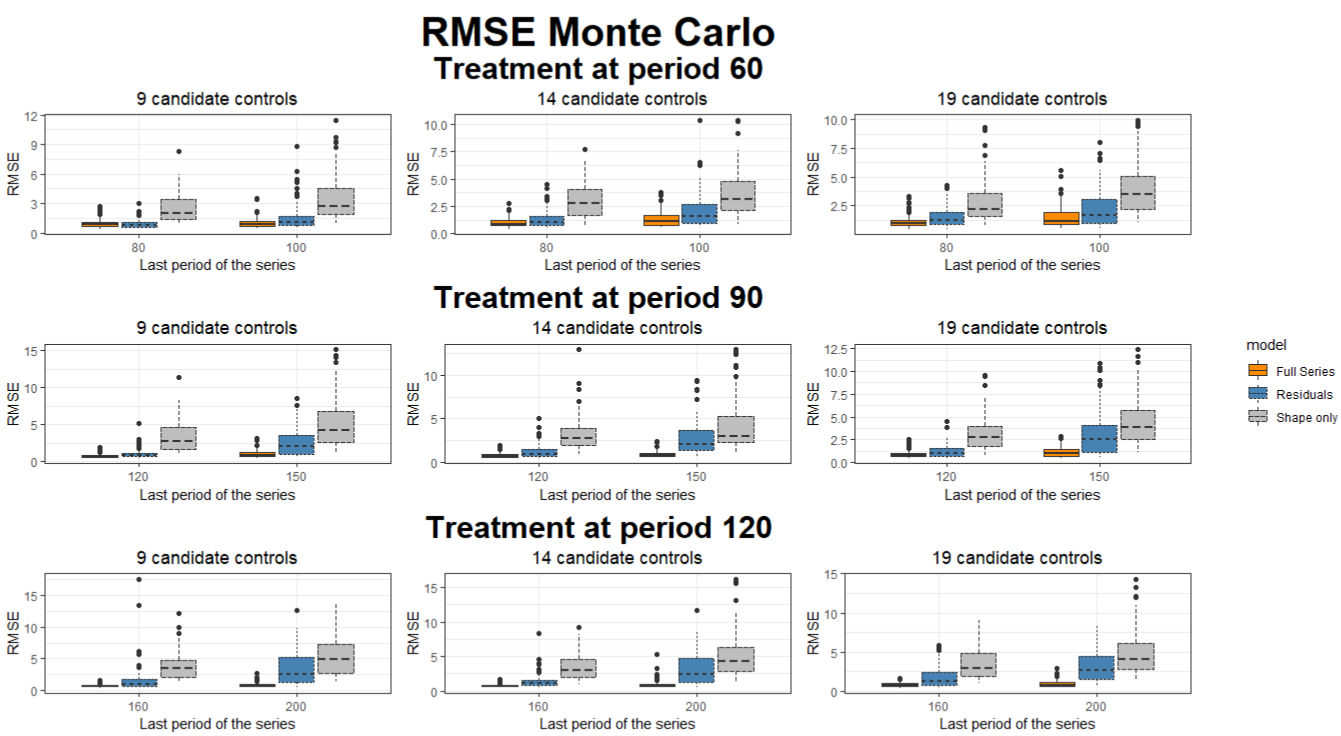

Figure 2: MAPE comparisons for different specifications

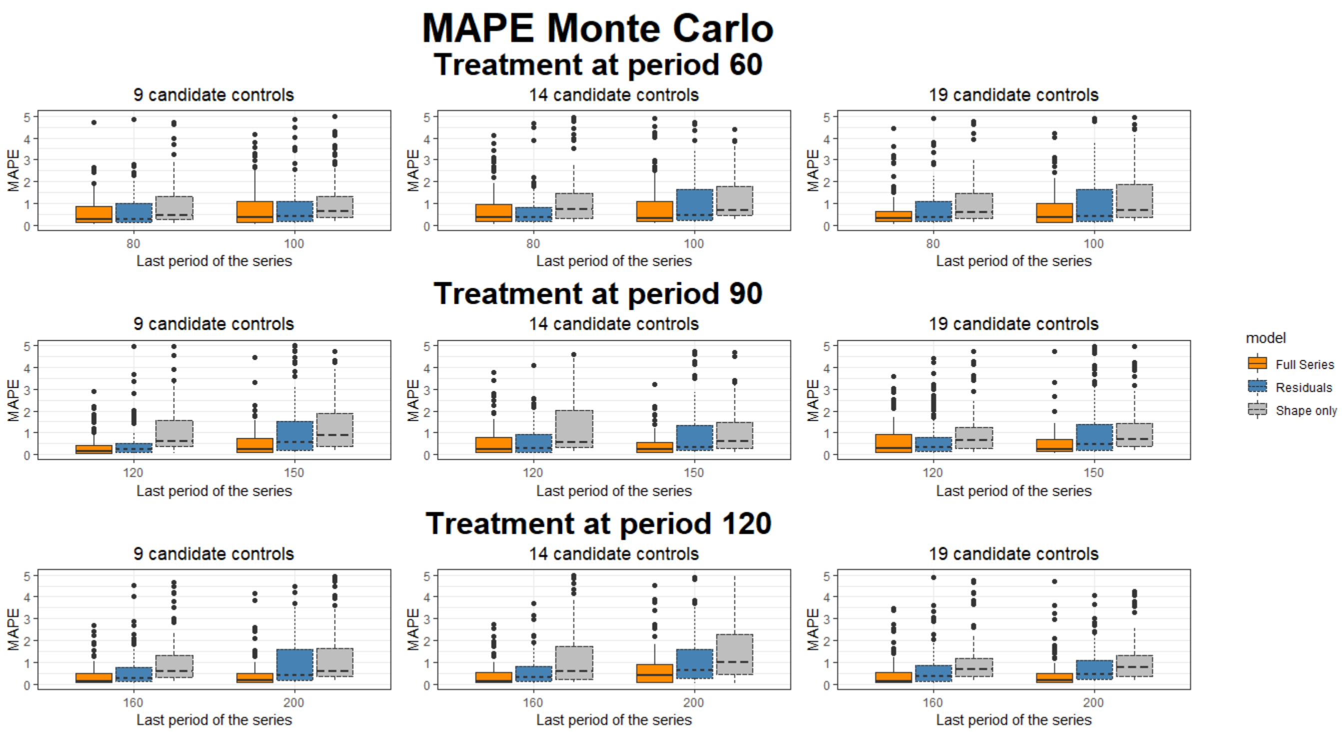

Figure 3: Synthetic Control under generated series indicating a false treatment effect

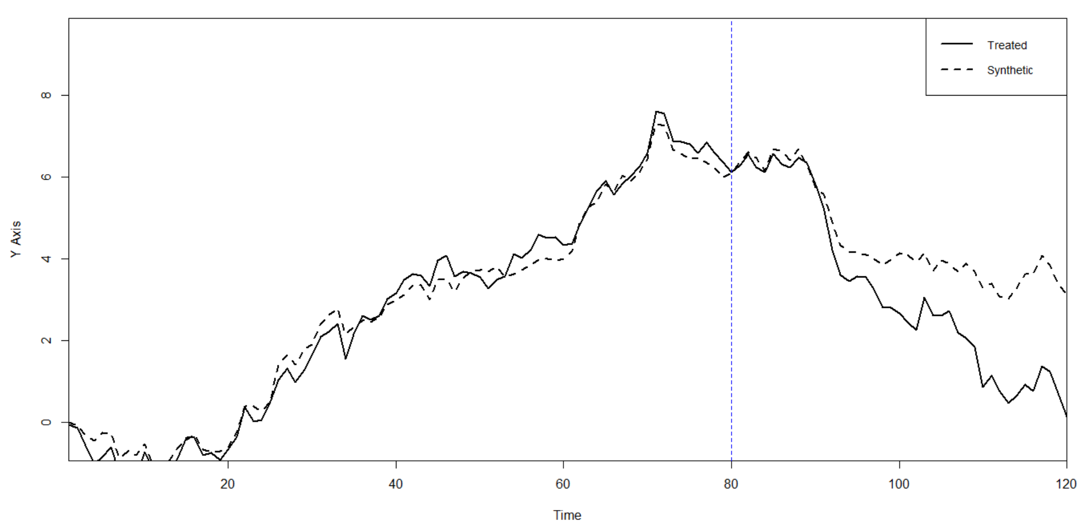

Figure 4: Trend-corrected control under generated series

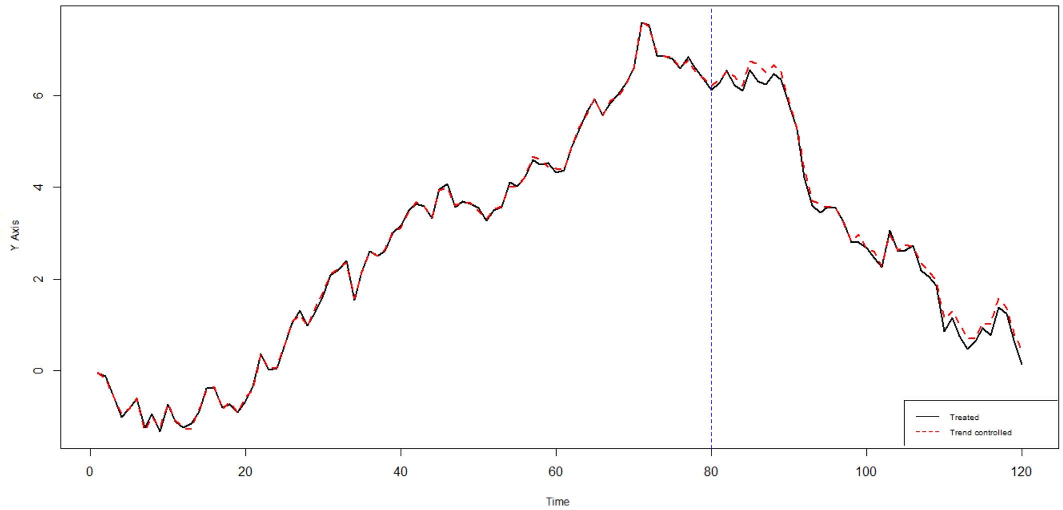

Figure 5: Synthetic control post-treatment RMSE distributions under no effect

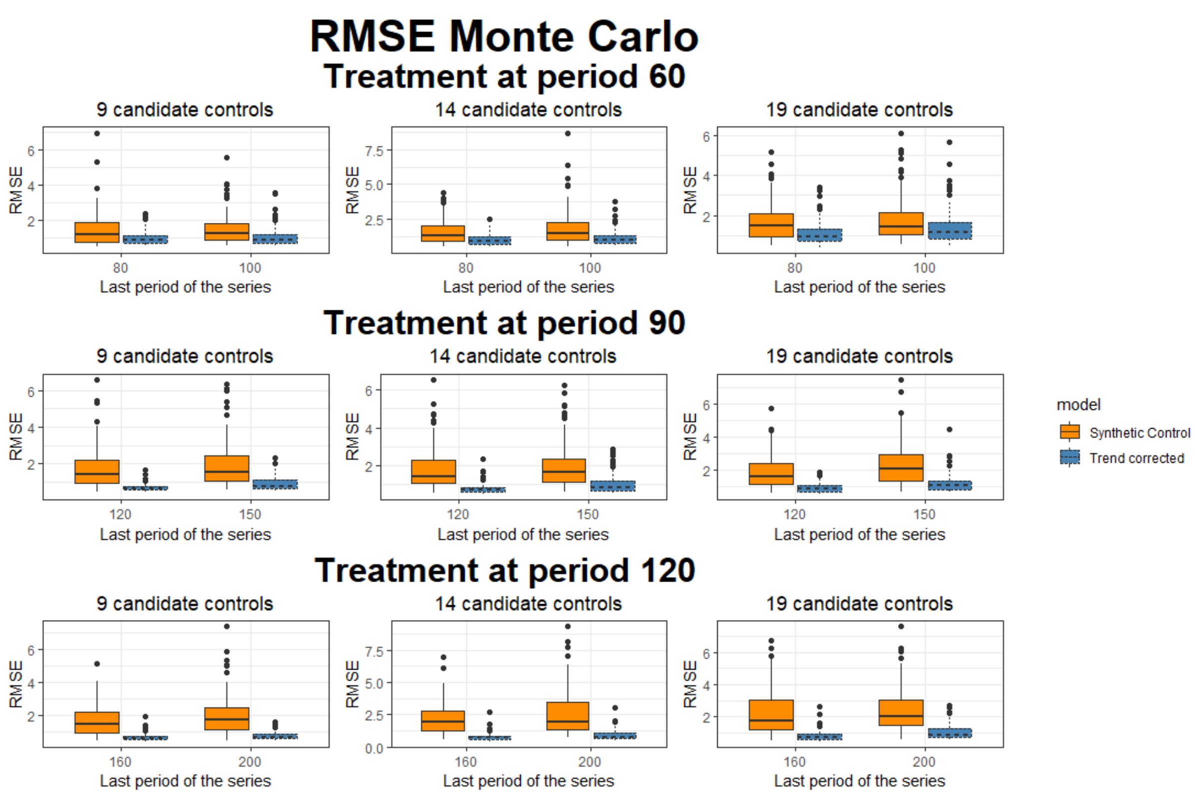

Figure 6: Trend-corrected post-treatment RMSE distributions under no effect

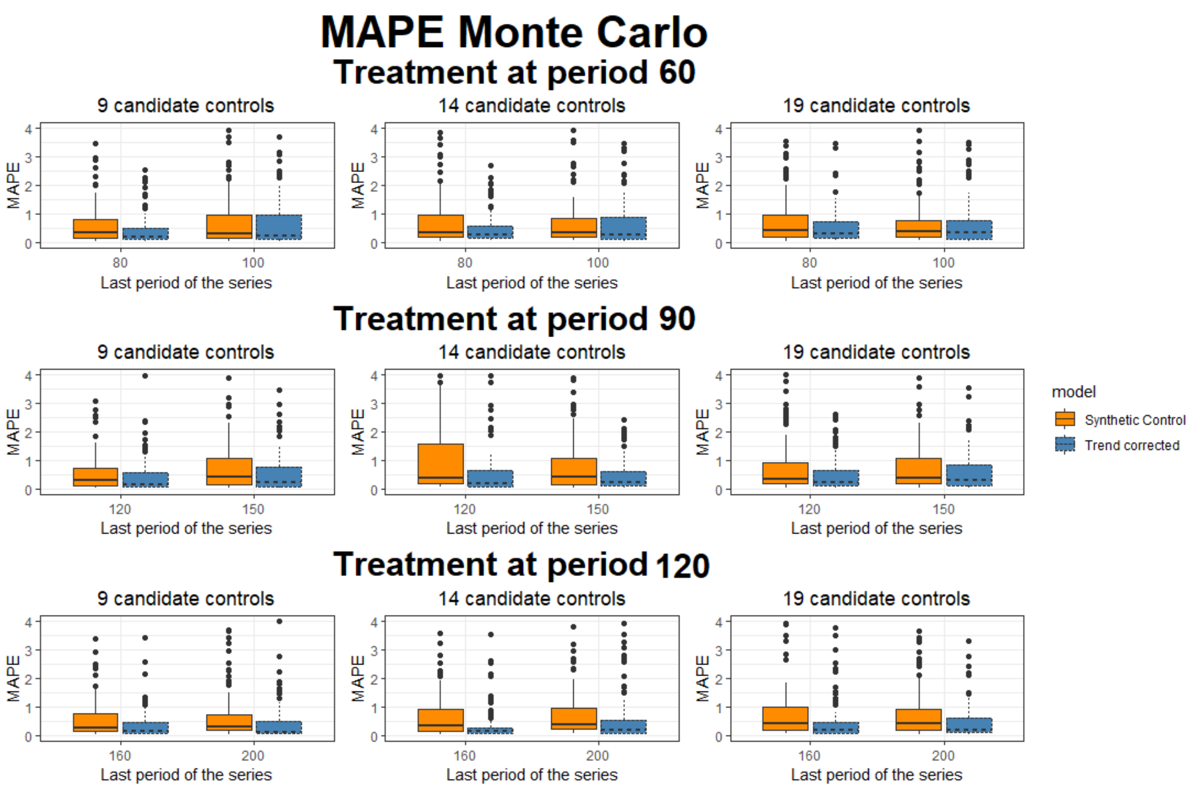

Figure 7: Median Anderson-Darling p-values vs. R^2, horizontal and vertical lines set at 0.5 and 0.05 respectively

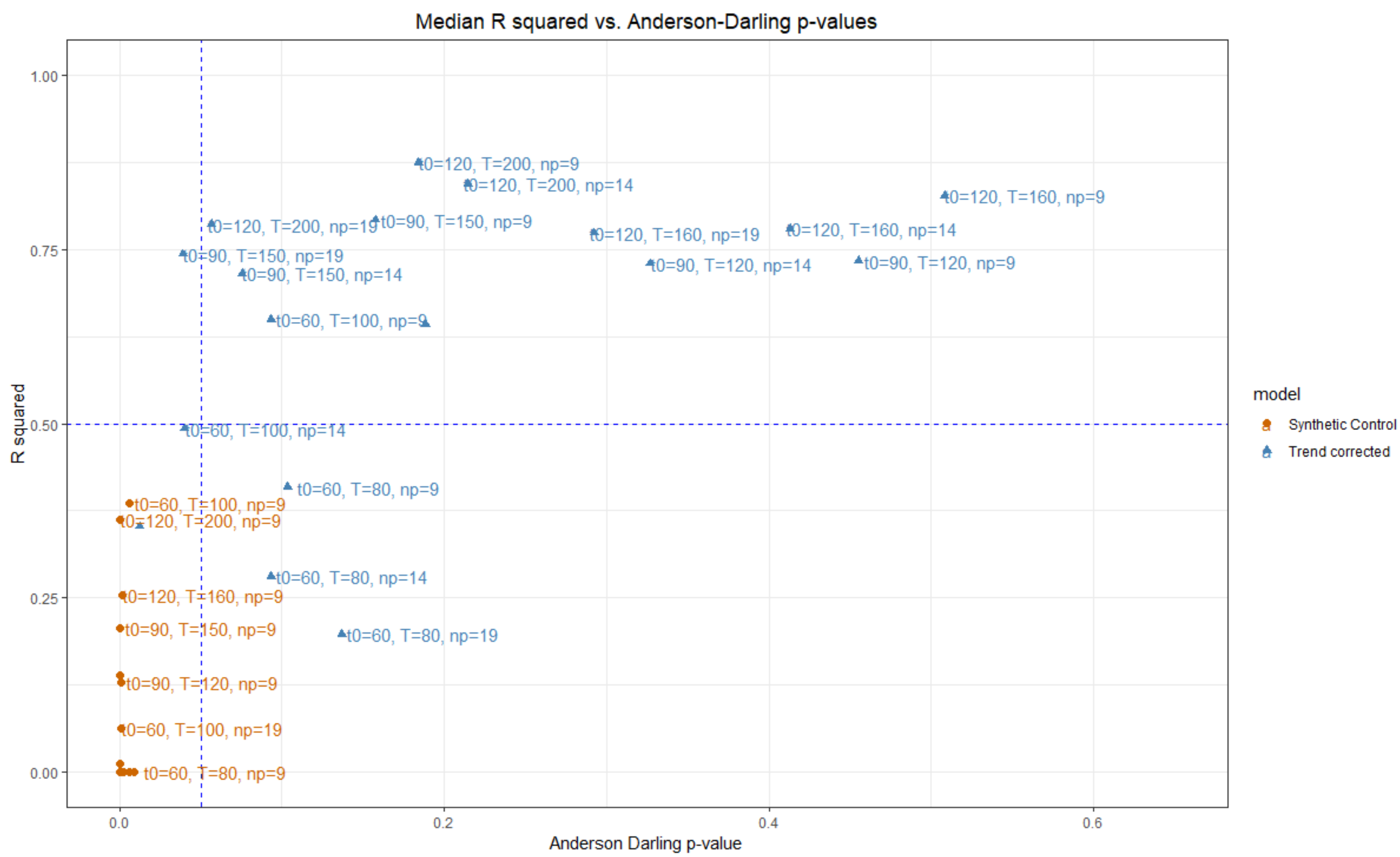

Figure 8: Constant effect testing parameter distributions with centers expected to be at $\beta_0 = 10$ and $\beta_1 = 0$

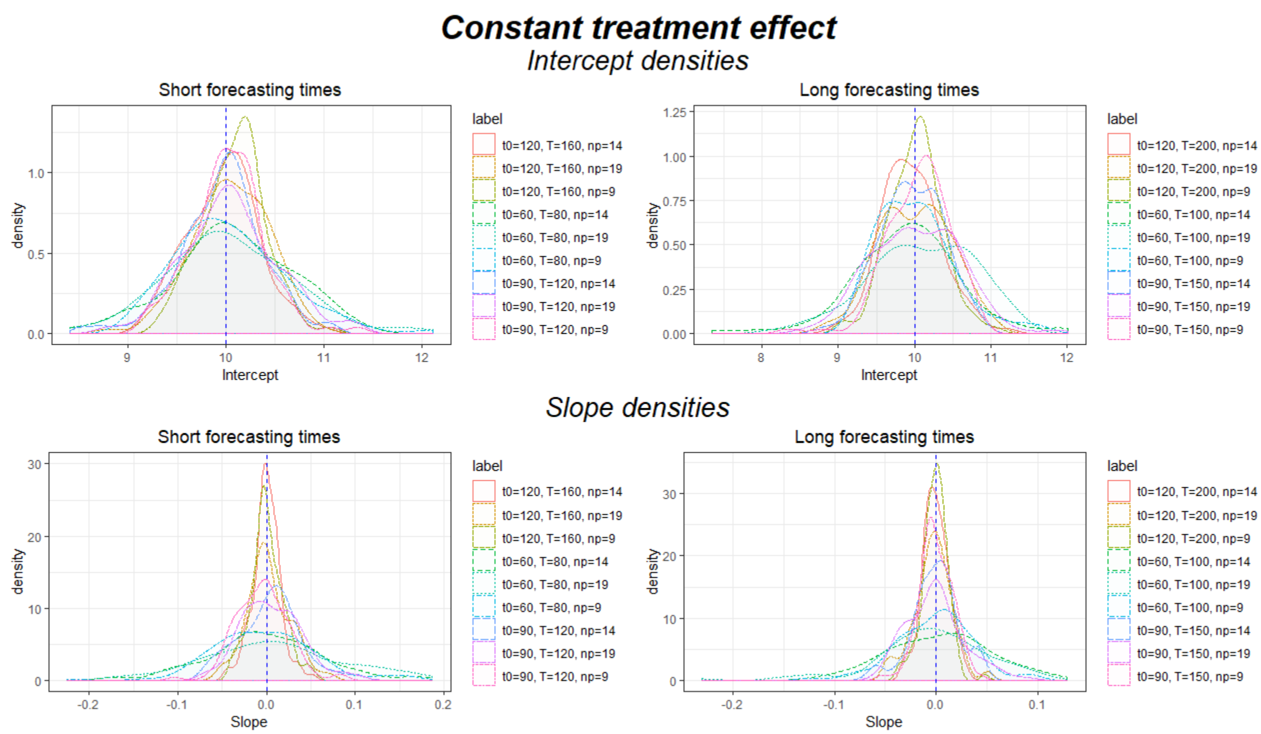

Figure 9: Linear and polynomial effect testing, R^2 distributions

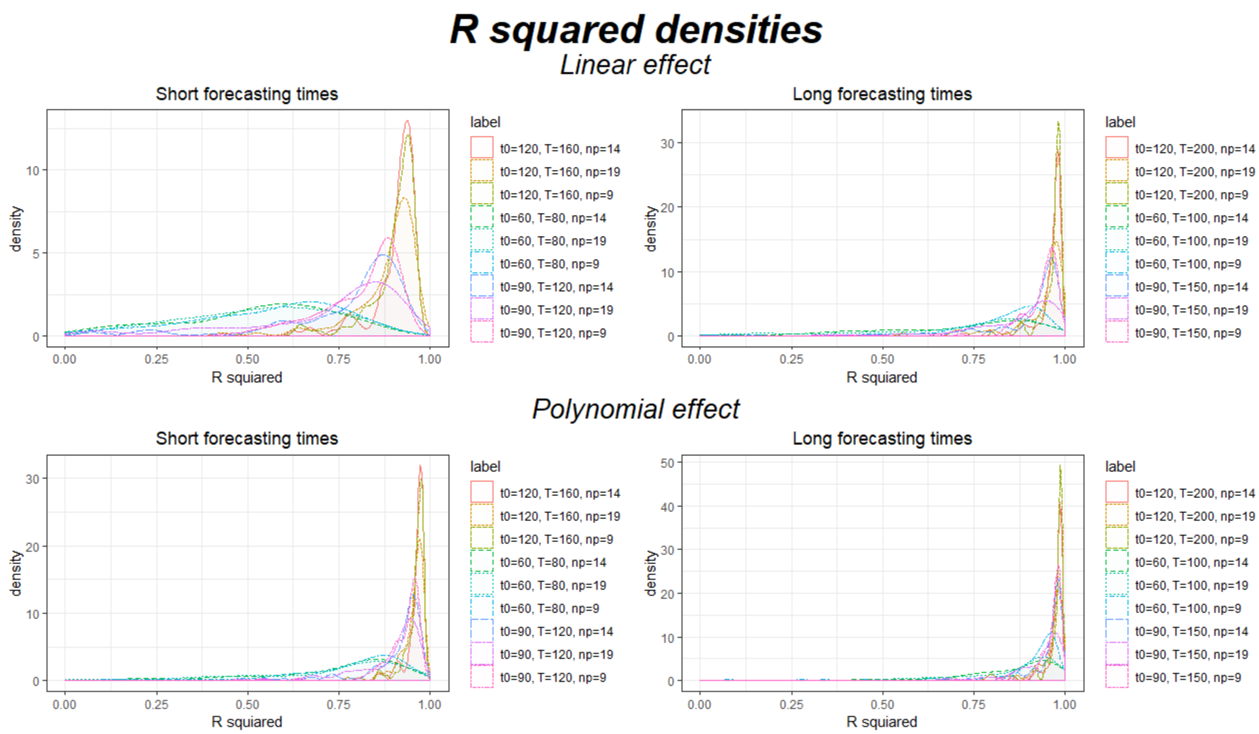

Figure 10: Per Capita GDP OECD countries (PPP, 2002 USD)

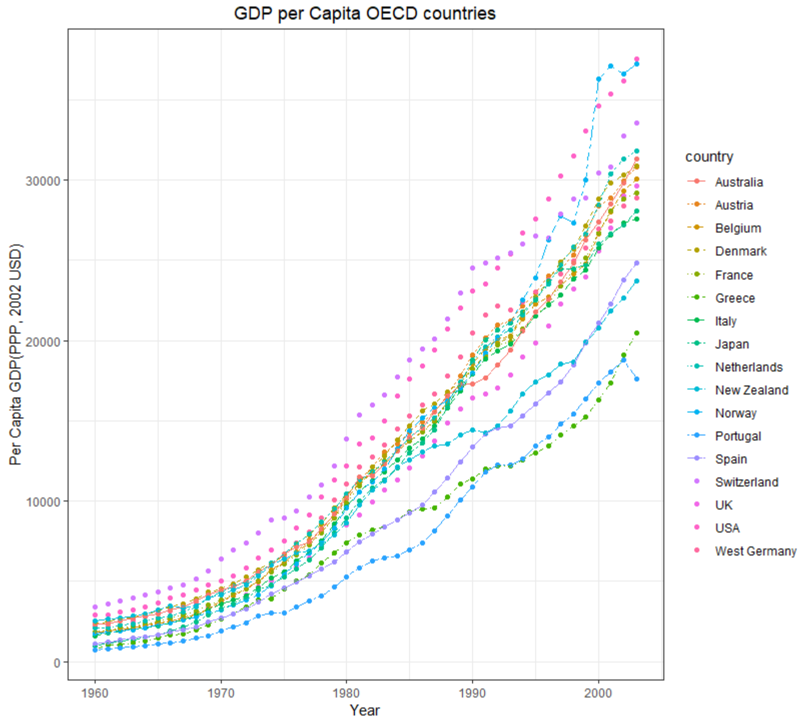

Figure 11: Comparison of results from Abadie and Gardeazabal (2003) and the trend-corrected counterfactual

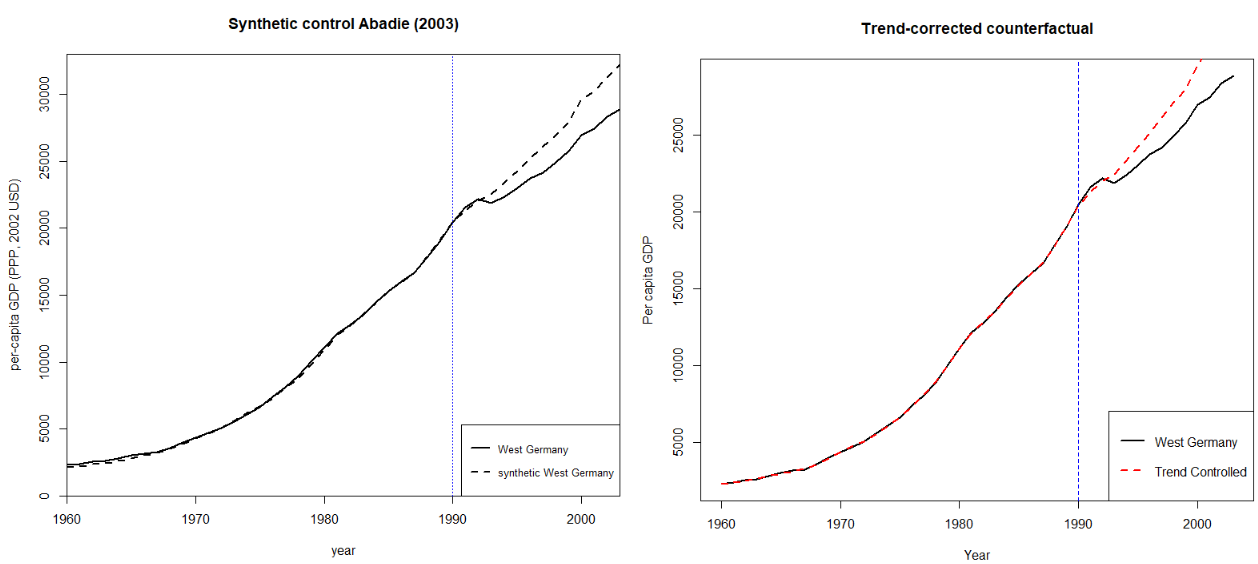

Figure 12: Placebo tests for different treatment times