

Breitung, Jörg; Bolwin, Lennart; Töns, Justus

**Conference Paper**

## Alternative approaches for estimation and inference in synthetic control designs

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2024: Upcoming Labor Market Challenges

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Breitung, Jörg; Bolwin, Lennart; Töns, Justus (2024) : Alternative approaches for estimation and inference in synthetic control designs, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2024: Upcoming Labor Market Challenges, ZBW - Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/302344>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Alternative approaches for estimation and inference in synthetic control designs

*Jörg Breitung<sup>a</sup>, Lennart Bolwin<sup>b</sup>, Justus Töns<sup>c</sup>*

<sup>a</sup>*University of Cologne*

<sup>b</sup>*German Economic Institute*

<sup>c</sup>*DZ Bank, Frankfurt*

## Abstract

The Synthetic Control (SC) method is widely employed in estimating causal treatment effects in observational studies. Typically, it synthesizes the counterfactual of the treated unit by employing a data-driven weighted average of the remaining units in the post-treatment period. These weights aim to minimize the distance between the treated unit and its counterfactual in the pre-treatment period. To prevent overfitting of the pre-treatment data, the original approach introduced by Abadie, Diamond, and Hainmueller (ADH) imposes a constraint wherein all weights must be weakly positive and sum up to one. Expanding upon [Doudchenko and Imbens \[2016\]](#), we introduce REGSC, an alternative regularized synthetic control approach. This approach involves shrinking individual coefficients toward zero and the sum of coefficients toward one. It amalgamates the benefits of the original SC method and the elastic net, offering a closed-form solution with tunable hyperparameters. The estimator also allows for a straightforward Bayesian representation, providing advantages, particularly in quantifying estimation uncertainty through Bayesian credibility intervals. By appending the donor pool with lagged values of the donors and the counterfactual, the REGSC framework is extended to dynamic contexts. Our dynamic version of the REGSC estimator accommodates nonstationary and cointegrated time series, a case where the original SC method is inconsistent. To assess the relative performance of our estimation approach in static and dynamic data-generating processes we conduct a variety of Monte Carlo experiments and apply it to existing empirical datasets. Our results suggest that the (dynamic) REGSC estimator outperforms other SC methods previously proposed in the literature.

**Keywords:** *Synthetic Control; Causal Inference; Regularization, Autoregressive Distributed Lag Models, Bayesian Estimation*

## 1. Introduction

The synthetic control (henceforth: SC) method was developed by Alberto Abadie, Alexis Diamond, and Jens Hainmueller (ADH) in a series of influential papers ([Abadie and Gardeazabal \[2003\]](#), [Abadie et al. \[2010\]](#), [Abadie et al. \[2015\]](#)). The method is designed to estimate the causal effect of a treatment (resp. intervention) in settings with (at least) a single treatment unit and a number of potential control units. This is achieved by comparing the observed treatment unit to the hypothetical trajectory of the treatment unit in the absence of the treatment. Pre- and post-treatment data are observed for the treatment and control units (also called donors), for the outcome of interest as well as for a set of time-constant covariates. Usually, both the number of potential control units  $J$  and the amount of pre-treatment periods  $T_{pre}$  are small. In many applications  $T_{pre}$  is even smaller than  $J$ , making standard estimation approaches like Ordinary Least Squares (OLS) unreliable or infeasible.

Building on the work of [Doudchenko and Imbens \[2016\]](#) and ADH, we propose a novel regularized SC estimator (henceforth: *REGSC*) that shrinks individual weights towards zero and the sum of the weights towards one. Due to the differentiable penalty term, the estimator has a closed form solution which is especially appealing in the context of low-frequency macroeconomic application of the SC method. Furthermore, our estimator is flexible enough to produce reliable estimates of the counterfactuals and still maintaining the interpretability of the weights. The natural Bayesian representation of our estimator is another mayor strength over existing SC methods. In many applications, the estimated treatment effects are temporally cumulated over the post-treatment periods. In such cases, a simple point estimate of the effect may be misleading and we consider it crucial to assess the estimation uncertainty in terms of forecast intervals. Relying on the Bayesian REGSC representation this can be achieved via Bayesian credibility intervals.

Another contribution of this paper is to accommodate dynamic features of the time series. To achieve this, we expand the donor pool by including lagged values of both the donors and the treatment series. Such expansions often lead to a plethora of explanatory variables, necessitating adequate regularization. In this context we found the elastic net to be particularly promising due to its ability to obtain sparse solutions. Another advantage of dynamic methods is that it is well suited for dealing with possibly nonstationary time series. Note that for nonstationary and not cointegrated time series there does not exist a stable long-run relationship between the treatment and donor series and, therefore, it does not make sense to construct

a synthetic control unit by some linear combination of the donor series. In such cases SC methods need to be applied to the differenced series. On the other hand, if the treatment series are cointegrated with the donor series then taking differences ignores important information on the long-run relationship among the series. In such a scenario, dynamic SC methods offer a significant improvement.

This paper is structured as follows. Section 2 outlines the conceptual framework and motivates the REGSC estimator. The subsequent subsections close the bridge to related estimation techniques like Differences-in-Differences (DiD) and the factor model. Furthermore, we discuss the role of additional explanatory variables and the potential presence of serial dependence in the data. We also demonstrate the estimation of Bayesian credibility intervals to quantify the estimation uncertainty. In section 3, we conduct an extensive Monte Carlo simulation to study the performance of our proposed and various routinely employed estimators in the context of SC. We pay special attention to plausible data-generating processes by considering static factor models as well as stationary vector autoregressions and nonstationary cointegration processes. Section 4 demonstrates the external validity and reveals important differences of the estimators by revisiting two well known applications of ADH. Section 5 concludes and provides an outlook on further work.

## 2. Theoretical framework

The conceptual framework is related to the potential outcome framework as introduced by Neyman [1923] and elaborated by Rubin [1974] and Holland [1986]. Assume that we observe a collection of  $J + 1$  time series  $Y_{j,t}$  with  $j = 0, 1, \dots, J$  and  $t = 1, \dots, T$ . The unit  $j = 0$  is denoted as the *treatment unit* exposed to some intervention (treatment) at period  $T_0 < T$ . The other series with  $j = 1, \dots, J$  belong to the set of *donor units*. Accordingly the temporal ordering is given by:

$$\underbrace{1, 2, \dots, T_0}_{T_{pre} \text{ observations}}, \underbrace{T_0 + 1, T_0 + 2, \dots, T}_{T_{post} \text{ observations}},$$

such that  $T_{pre} = T_0$  periods of pre-treatment data and  $T_{post} = T - T_0$  periods of post-treatment data are observed.

Following the existing literature we rule out any anticipation effects and contamination (i.e., no spillovers to the donor units). As argued by Abadie et al. [2010] in the presence of anticipation effects, the date  $T_0$  should be shifted backward until the no-anticipation assumption seem plausible. If some panel units in the donor pool are affected by the treatment (contamination), these units should be removed

from the donor pool prior to estimation. Our goal is to evaluate the causal effect of the intervention, where the specific form of the effect remains unspecified. This is possible because the main goal of SC estimation lies in the precise estimation of the counterfactual. Formally the treatment effect at time  $\tau > T_0$  is given by

$$\delta_\tau = Y_{0,\tau} - Y_{0,\tau}^N$$

where  $Y_{0,\tau}$  denotes the observed outcome under the treatment and  $Y_{0,\tau}^N$  denotes the (unobserved) potential outcome without treatment, henceforth referred to as the counterfactual outcome. Obviously, the treatment effect is estimated efficiently by replacing  $Y_{0,\tau}^N$  by an efficient estimator. It is well known that the MSE-optimal forecast  $\widehat{Y}_{0,\tau}^N$  of the counterfactual is given by the conditional expectation

$$\widehat{Y}_{0,\tau}^N = \mathbb{E}(Y_{0,\tau}^N | \mathcal{I}_\tau)$$

where  $\mathcal{I}_\tau$  denotes the relevant information set at period  $\tau$ . In what follows the (potential) information set is assumed to be given by

$$\mathcal{I}_\tau = \{\mathbf{z}_\tau, \mathbf{z}_{\tau-1}, \dots, \mathbf{z}_{T_0+1}, \mathbf{y}_{T_0}, \dots, \mathbf{y}_1\} \quad \text{for } \tau > T_0$$

where  $\mathbf{z}_t = (Y_{1,t}, Y_{2,t}, \dots, Y_{J,t})'$  and  $\mathbf{y}_t = (Y_{0,t}, \mathbf{z}_t)'$  such that the information set is comprised by the post-treatment donor series  $\mathbf{z}_\tau$  for  $\tau > T_0$  and the pre-treatment observations of the treated series and the donors  $\mathbf{y}_t$  with  $t \leq T_0$ .

We may extend the information set by including additional predictors comprised in the  $k \times 1$  vector  $\mathbf{x}_t$  that are known not to be affected by intervention but help to predict the counterfactual outcome. We briefly discuss this extension in Section 2.5. However, as such an extension complicates the analysis without adding much to the insights, our primary research focus does not center around scenarios involving additional covariates.

We will further assume that the  $J+1$  time series vector  $\mathbf{y}_t^* = (Y_{0,t}^N, \mathbf{z}_t)'$  is weakly stationary for  $t = 1, \dots, T$ , where  $Y_{0,t}^N = Y_{0,t}$  for  $t \leq T_0$ . In Section 2.7 we discuss extensions to a setting where the autoregressive representation of  $\mathbf{y}_t^*$  possesses roots on the unit circle of the complex plane. It turns out that in this case we require that  $Y_{0,t}^N$  is cointegrated with some linear combination of  $\mathbf{z}_t$ .

### 2.1. The static case

Let us first consider the most simple scenario of one treatment unit  $j = 0$  and two donor units  $j = 1, 2$ . In this subsection we assume that the outcome vector

$\mathbf{y}_t = (Y_{0,t}, Y_{1,t}, Y_{2,t})'$  for  $t = 1, \dots, T_0$  and  $\mathbf{y}_t^* = (Y_{0,t}^N, Y_{1,t}, Y_{2,t})'$  for  $t = T_0 + 1, \dots, T$  are distributed as

$$\mathbf{y}_t \stackrel{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{and} \quad \mathbf{y}_t^* \stackrel{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with  $\boldsymbol{\mu} = (\mu_0, \mu_1, \mu_2)'$  and the positive definite covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_0^2 & \boldsymbol{\sigma}'_{12} \\ \boldsymbol{\sigma}_{12} & \boldsymbol{\Sigma}_2 \end{pmatrix},$$

where  $\sigma_0^2$  denotes the variance of  $\mathbf{y}_0$ ,  $\boldsymbol{\Sigma}_2$  is a  $(2 \times 2)$  covariance matrix of the vector  $(Y_{1,t}, Y_{2,t})'$  and  $\boldsymbol{\sigma}_{12}$  is a  $(2 \times 1)$  vector with elements  $cov(Y_{0,t}, Y_{1,t})$  and  $cov(Y_{0,t}, Y_{2,t})$ .

We are interested in deriving the (mean-square) optimal forecast of the counterfactual  $Y_{0,t}^N$ , which is given by the conditional expectation

$$\begin{aligned} \mathbb{E}(Y_{0,t}^N | Y_{1,t}, Y_{2,t}) &= \mu_0 + w_1(Y_{1,t} - \mu_1) + w_2(Y_{2,t} - \mu_2) \\ &= \mu^* + w_1 Y_{1,t} + w_2 Y_{2,t} \end{aligned}$$

where  $\mu^* = \mu_0 - w_1 \mu_1 - w_2 \mu_2$  and

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \boldsymbol{\Sigma}_2^{-1} \boldsymbol{\sigma}_{12}$$

The original SC estimator imposes the restrictions that  $w_1, w_2 \geq 0$  and  $w_1 + w_2 = 1$  yet our result implies that in the most simple case, there is no inherent reason to do so. Furthermore, we argue that the construction of SC should include a constant term, as otherwise the estimated counterfactual may have a mean outside the convex hull of the donor means. See also [Doudchenko and Imbens \[2016\]](#) for a careful discussion of these restrictions.

To illustrate these findings assume that

$$\mathbf{y}_t \stackrel{iid}{\sim} \mathcal{N} \left( \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.1 & 0.4 \\ 0.1 & 1 & 0.5 \\ 0.4 & 0.5 & 1 \end{pmatrix} \right).$$

For this example the unrestricted optimal weights for the counterfactual result as  $w_1 = -0.1333$ ,  $w_2 = 0.4667$  and  $\mu^* = 0.6667$ .<sup>1</sup> Note that  $w_1$  is negative even though all bivariate correlations between the units are positive. One may argue that this

<sup>1</sup>The computation is postponed to the appendix.

result does not make much sense as the economic interpretation of  $Y_{1,t}$  entering the counterfactual  $\widehat{Y}_{0,t}^N$  with a negative sign is unclear. This demonstrates the trade-off between optimality in a statistical sense and the economic interpretation of the solution. What happens if we impose the restrictions that all weights are positive and sum up to unity? In this case the restricted optimum yields the linear combination  $\widetilde{Y}_{0,t}^N = 0.2Y_{1,t} + 0.8Y_{2,t}$ . The important difference lies in the variance of these estimates. For our example we obtain

$$\begin{aligned} \text{var} \left( Y_{0,t}^N - \widehat{Y}_{0,t}^N \right) &= 0.8267 \\ \text{var} \left( Y_{0,t}^N - \widetilde{Y}_{0,t}^N \right) &= 1.1600. \end{aligned}$$

It is interesting to note that the variance of the restricted estimate is even larger than the unconditional variance of  $Y_0$ . This is possible as  $(w_1, w_2) = (0, 0)$  is not included in the restricted parameter space.

In microeconomic settings it is usually assumed that the units (individuals) in the treatment group and units in the control group are uncorrelated or even independent. In such cases the weights are equal to zero and the optimal forecast of the counterfactual is identical to  $\mu^* = \mu_0$ . This mean is estimated without involving the control units (donors). The only reason for considering the control units in this case may be the additional assumption that  $Y_{0,t}$  and  $Y_{j,t}$  have the same mean for  $j = 1, \dots, J$ .

## 2.2. Regularized SC estimator

In the previous section we considered a framework where all parameters of the distributions are known. In empirical practice this is usually not the case and, therefore, the parameters need to be estimated. In the context of applications of the SC method, usually those samples are fairly small as the variables of interest are often measured at quarterly or even annual intervals. Thus, the number of pre-intervention time periods  $T_0$  is typically small and may even be smaller than the number of units in the donor pool  $J$ . In such scenarios, the unrestricted OLS estimate may face issues of instability, large variances or may not be available if  $T_0 < J$ . To analyse the (asymptotic) properties consider the least-squares estimator for the relevant regression

$$Y_{0,t} = \mu^* + w_1 Y_{1,t} + w_2 Y_{2,t} + \dots + w_J Y_{J,t} + u_t \quad \text{for } t = 1, 2, \dots, T_0.$$

Under standard assumptions for the linear regression it follows that for a *fixed number of regressors* ( $J$ ) the OLS estimator  $\widehat{\mathbf{w}} = (\widehat{w}_1, \dots, \widehat{w}_J)'$  converges in probability

to the optimal weights  $\mathbf{w} = (w_1, \dots, w_J)'$  and has a normal limit distribution for fixed  $J$  and  $T_0 \rightarrow \infty$ . In empirical practice, we typically have a large number of donors candidates such that  $J$  may be of similar magnitude than  $T_0$ . As shown by [Bekker \[1994\]](#), if  $J/T_0$  converges to some constant  $c > 0$  as  $J$  and  $T_0$  tend to infinity, the OLS estimator does not longer converge to the true coefficients. A convenient approach to mitigate this deficiency of the OLS estimator is to impose some regularization. In this context [Doudchenko and Imbens \[2016\]](#) suggest employing an elastic net regression that minimizes the objective function

$$Q(\mathbf{w}, \lambda_1, \lambda_2) = \underbrace{\sum_{t=1}^{T_0} \left( Y_{0,t} - \mu^* - \sum_{j=1}^J w_j Y_{j,t} \right)^2}_{RSS} + \lambda_1 \underbrace{\left( \sum_{j=1}^J w_j^2 \right)}_{Ridge} + \lambda_2 \underbrace{\left( \sum_{j=1}^J |w_j| \right)}_{Lasso}$$

The  $L_2$ -norm (Ridge penalty) as proposed by [Hoerl and Kennard \[1970\]](#) shrinks the coefficients towards zero without performing variable selection in the sense that certain coefficients are set exactly to zero as done by the  $L_1$  (Lasso) penalty. However, the Ridge penalty has the appealing feature that it results in a quadratic minimization problem which admits a computationally fast closed form solution. In contrast, the  $L_1$ -norm (Lasso penalty) as proposed by [Tibshirani \[1996\]](#) implies both, continuous shrinkage and automatic variable selection. As a consequence, the solution of the objective function typically involves entries of  $\mathbf{w}$  that are exactly zero which makes the resulting model sparse and easier to interpret. However, the Lasso has no closed form solution and therefore, numerical optimization techniques are required. The shrinkage parameters  $\lambda_1$  and  $\lambda_2$  can be selected through  $k$ -fold cross validation (CV). This involves storing combinations of  $\lambda_1$  and  $\lambda_2$  that minimize the objective function across  $k$  validation sets and computing the means of those parameters.

We propose a different regularization that we call the regularized synthetic control estimator (REGSC). This estimator augments the OLS objective function by a Ridge penalty and another penalty term that shrinks the sum of coefficient towards one. The penalty term is comparable to a prior on the sum of coefficients as frequently applied in the context of large Bayesian vector-autoregressions (see for instance [Bańbura et al. \[2010\]](#)). Vector autoregressions with a great number of time series can lack stability and imposing the prior belief that the coefficient sum should not exceed a certain threshold has proven to be beneficial. To the best of our knowledge, there is no existing work that advocated the approach of shrinking the sum of coefficients towards one in the context of SC. Our proposed REGSC



estimator has the following the objective function:

$$Q(w, \lambda_1, \lambda_2) = \underbrace{\sum_{t=1}^{T_0} \left( Y_{0,t} - \mu^* - \sum_{j=1}^J w_j Y_{j,t} \right)^2}_{RSS} + \lambda_1 \underbrace{\left( \sum_{j=1}^J w_j^2 \right)}_{\text{Ridge}} + \lambda_2 \underbrace{\left( 1 - \sum_{j=1}^J w_j \right)^2}_{\text{weight penalty}}$$

Due to the individual shrinkage to zero (Ridge) and the joint shrinkage to one (weight penalty), this regularization is closely related to original SC estimator and the elastic net. It combines the strengths of both techniques as it preserves the weight interpretation of the SC approach with weights that (approximately) sum up to one whenever the tuneable hyperparameter  $\lambda_2$  is large.

To ensure unbiasedness in the absence of the constant, we recommend to demean the pre-treatment units as follows: Let  $\tilde{\mathbf{y}}_0 = (Y_{0,1} - \bar{Y}_0, \dots, Y_{0,T_0} - \bar{Y}_0)'$ , where  $\bar{Y}_0 = T_0^{-1} \sum_{t=1}^{T_0} Y_{0,t}$  and  $\tilde{\mathbf{Z}} = (z_1 - \bar{z}, \dots, z_{T_0} - \bar{z})'$ . Following the notation of section 2,  $z_t$  represents the vector of donor values at  $t$ . To demean this vector, we subtract the donor-specific pre-treatment mean from each entry such that  $\bar{\mathbf{z}} = (\bar{Y}_1, \dots, \bar{Y}_J)$ . The minimization of  $Q(w, \lambda_1, \lambda_2)$  has the following closed form solution:

$$\hat{\mathbf{w}}_{\lambda_1, \lambda_2} = \left( \tilde{\mathbf{Z}}' \tilde{\mathbf{Z}} + \lambda_1 \mathbf{I}_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J' \right)^{-1} \left( \tilde{\mathbf{Z}}' \tilde{\mathbf{y}}_0 + \lambda_2 \mathbf{1}_J \right).$$

where  $\mathbf{I}_J$  is the  $J \times J$  identity matrix and  $\mathbf{1}_J$  is a  $J$ -dimensional vector of ones. In the appendix, we show for  $\lambda_1 \rightarrow \infty$  and  $\lambda_1/\lambda_2 \rightarrow \phi$  the weights converge to  $1/(n + \phi) \approx 1/n$ , for large  $n$ , which seems to be a more reasonable target than shrinking towards zero as done by the elastic net. Similar to the case of the elastic net, the shrinkage parameters  $\lambda_1$  and  $\lambda_2$  can be chosen by cross validation, where our experience suggests that  $\lambda_2$  is typically substantially larger than  $\lambda_1$ . Thus, optimizing subject to the restriction  $\lambda_2 \approx 1,000 \cdot \lambda_1$  reduces computation time and already produces reasonable estimates. Note, however, that the choice of  $\lambda_1$  and  $\lambda_2$  will always be case specific. The combination of a closed form solution and tuneable hyperparameters makes the REGSC method highly appropriate for the low-frequency macroeconomic context of SC: It is able to produce weights that are flexible enough, tend to be positive, and can be estimated reliably in small samples. We also tried out to combine the Lasso penalty with the ‘‘weight penalty’’ but we did not find any improvement over the proposed estimator.

Our regularized estimator admits a reasonable Bayesian interpretation. As shown

in the appendix, the REGSC estimator implies a prior distribution of the form

$$\mathbf{w} \sim \mathcal{N}(\mu_0 \mathbf{1}_J, \mathbf{V}_0)$$

where  $\mu_0 = \lambda_2 / (\lambda_1 + J\lambda_2)$

and  $\mathbf{V}_0$  is a matrix with variances  $V_{ii} = \sigma^2(\lambda_1^{-1} - \psi/J)$  for  $i = 1, \dots, J$  and covariances  $V_{ij} = -\sigma^2\psi/J$  for  $i \neq j$ , with  $\psi = \lambda_1^{-1} - (\lambda_1 + J\lambda_2)^{-1}$ .

Note that if  $\frac{\lambda_1}{\lambda_2} \rightarrow 0$ ,  $\mu_0 \rightarrow 1/J$  and  $\psi \rightarrow 1$ . For  $\mu_0$ , this implies an equally weighted prior mean. As  $\psi \rightarrow 1$ , the off-diagonal elements of  $\mathbf{V}_0$  increase in absolute value indicating that for large values of  $\lambda_2$ , the prior has to put more emphasis on the joint distribution of the weights summing up to one. On the contrary, if  $\lambda_2/\lambda_1 \rightarrow 0$ ,  $\mu_0 \rightarrow 0$  and  $\psi \rightarrow 0$ , hence the off-diagonal elements of  $\mathbf{V}_0$  shrink to 0. Further, in case  $\lambda_1$  resp.  $\lambda_2$  approach infinity, the prior becomes dogmatic. In Section 2.8, the Bayesian interpretation is used for the computation of Bayesian credibility intervals.

### 2.3. Difference-in-Difference estimation

Let us briefly compare the REGSC estimator to the difference-in-difference (DiD) estimator, a popular estimation method developed for panel data. The estimator is based on a comparison of the before-after sample means of the treatment and the control group. In our case the treatment group consist on a single unit whereas the control group entails the  $J$  donor series. Another important difference is that the treatment effect is assumed to be different for each time series. Therefore, the “mean” of the post-treatment sample is based on a single observation  $Y_{0,\tau}$  with  $\tau > T_0$ . Accordingly, the DiD estimator boils down to

$$\hat{\delta}_\tau^{DiD} = (Y_{0,\tau} - \bar{Y}_0^{pre}) - (\bar{Y}_\tau^{post} - \bar{Y}_c^{pre})$$

where

$$\bar{Y}_0^{pre} = \frac{1}{T_0} \sum_{t=1}^{T_0} Y_{0,t} = \hat{\mu}_0, \quad \bar{Y}_c^{post} = \frac{1}{J} \sum_{j=1}^J Y_{j,\tau}$$

$$\bar{Y}_c^{pre} = \frac{1}{J \cdot T_0} \sum_{t=1}^{T_0} \sum_{j=1}^J Y_{j,t} = \frac{1}{J} \sum_{j=1}^J \hat{\mu}_j,$$

Comparing the DiD estimator to the REGSC estimator we conclude that the DiD estimator is obtained if  $\hat{w}_j = 1/J$ . Accordingly, our REGSC estimator becomes equivalent to the DiD estimator if  $\lambda_1$  is small and  $\lambda_2$  is large.

### 2.4. Factor models

The SC approach is typically motivated by considering a factor model of the form

$$Y_{it} = \mu_i + \theta_t + \lambda_i f_t + u_t$$

where  $\mu_i$  and  $\theta_t$  are individual and time specific constants (e.g. [Abadie et al. \[2010\]](#) and [Ferman \[2021\]](#)). It is assumed that the common factor  $f_t$  and the idiosyncratic component are uncorrelated and the idiosyncratic errors are mutually uncorrelated. Let us first ignore the time and individual specific constant (which in practice can easily be replaced by sample counterparts). We are interested in the conditional expectation:

$$\begin{aligned} \hat{Y}_{0,\tau}^N &= \mathbb{E}(Y_{0,\tau} | Y_{1,\tau}, \dots, Y_{J,\tau}) \quad \text{for } \tau = T_0 + 1, \dots, T \\ &= \lambda_0 \mathbb{E}(f_\tau | Y_{1,\tau}, \dots, Y_{J,\tau}) \end{aligned}$$

This suggests to estimate the common factor as linear combination of the donors  $\tilde{\mathbf{z}}_\tau = (Y_{1,\tau} - \bar{Y}_1, \dots, Y_{J,\tau} - \bar{Y}_J)'$ . A popular estimator with this property is the principal component (PC) estimator. The approach can easily be generalized to model with  $r > 1$  factors. Let  $V_r$  denote the  $J \times r$  matrix of eigenvectors associated with the  $r$  largest eigenvalues of  $\tilde{\mathbf{Z}}'\tilde{\mathbf{z}}/T_0$ . Accordingly, the vector of  $r$  factors for period  $1 \leq t \leq T$  is estimated as  $\hat{\mathbf{f}}_t = \mathbf{V}'_r \tilde{\mathbf{z}}_t$ . Furthermore, the  $r \times 1$  vector of factor loadings  $\lambda_0$  is obtained from the regression

$$Y_{0,t} = \mu^* + \hat{\mathbf{f}}'_t \lambda_0 + u_t \quad \text{for } t = 1, \dots, T_0$$

Let  $\hat{\mu}^*$  and  $\hat{\lambda}_0$  denote the OLS estimators of  $\mu^*$  and  $\lambda_0$ , respectively. Then the estimated counterfactual results as

$$\tilde{Y}_{0,t}^N = \hat{\mu}^* + \hat{\mathbf{f}}'_t \hat{\lambda}_0$$

Note that  $\hat{\mathbf{f}}'_t \hat{\lambda}_0 = \tilde{\mathbf{w}}'\tilde{\mathbf{z}}_t$  is a particular combination of the donor observations with  $\tilde{\mathbf{w}} = \mathbf{V}'_r \hat{\lambda}_0$ . The factor model can be considered an alternative shrinkage estimator, and we recommend to employ it for the estimation of the counterfactual in SC applications even in absence of a strict factor data-generating process. For instance, in PC regressions, the original set of donors is transformed into a set of linearly uncorrelated variables known as principal components which are linear combinations of the original donors. By using a subset of these principal components that capture the most variability, a shrinkage procedure that focuses on the most important

components and discards the less influential ones is performed. Furthermore, the factor model is particularly appealing when dealing with multicollinearity which is often the case SC applications. In this sense the factor model can be interpreted as another regularized estimator that shrinks the weights towards the implied factor structure.

### 2.5. Additional covariates

The original SC approach (e.g. [Abadie and Gardeazabal \[2003\]](#) and [Abadie et al. \[2010\]](#)) assumes that we have available  $K$  additional time-constant covariates that characterizes important features of the units. Let us represent this additional information by the  $K \times (J + 1)$  matrix  $\mathbf{X}$ . The weight vector is chosen such that

$$\hat{\mathbf{w}}(W) = \arg \min_w (\mathbf{x}_0 - \mathbf{X}_1 w)' \mathbf{V} (\mathbf{x}_0 - \mathbf{X}_1 w)$$

where  $\mathbf{x}_0$  denotes the first columns of  $\mathbf{X}$  (the characteristics of the treatment unit) and  $\mathbf{X}_1$  is the  $K \times J$  matrix of the remaining columns. The matrix  $\mathbf{V}$  is a diagonal matrix of additional weights that is related to the importance of the corresponding characteristics for the minimization problem.

Our REGSC estimator estimator can be seen as a special case of this estimator by treating all of the  $T_0$  observations of  $Y_{j,t}$  as covariates. Accordingly,  $K = T_0$  and the  $(k, j)$  element of  $\mathbf{X}$  is equal to  $Y_{j,t}$ . Furthermore the weight matrix  $\mathbf{V}$  is the identity matrix. An important drawback of the original SC approach is that it does not exploit the information from the past of the  $J + 1$  time series. Let  $\mathbf{Y}_{\text{pre}}$  denote the  $T_0 \times (J + 1)$  matrix of observations coming from the time before the intervention. We can combine the internal and external information by defining the matrix  $\mathbf{X}^+ = (\mathbf{X}, \mathbf{Y}_{\text{pre}})$  and consider the minimization

$$\hat{\mathbf{w}}(V) = \arg \min_w (\mathbf{z}_0^+ - \mathbf{Z}_1^+ w)' \mathbf{V} (\mathbf{z}_0^+ - \mathbf{Z}_1^+ w)$$

where  $\mathbf{z}_0^+$  denotes the first columns of  $\mathbf{Z}^+$  and  $\mathbf{Z}_1^+$  collects the remaining columns. The upper block of the weight matrix  $\mathbf{V}$  that is associated to  $Y_{\text{pre}}$  will typically be an identity matrix whereas the weights of the lower diagonal block are control to the importance of the external information relative to the internal information from the past.

### 2.6. Dynamic models

When modelling macroeconomic time series it is often assumed that the  $(J+1) \times 1$  vector of time series  $\mathbf{y}_t = (Y_{0,t}, \dots, Y_{J,t})'$  can be represented by a vector autoregressive model given by

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\alpha} + \mathbf{A}_1 \mathbf{y}_{t-1} + \dots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t \\ \mathbf{y}_t &= \boldsymbol{\mu} + \mathbf{A}(L)(\mathbf{y}_{t-1} - \boldsymbol{\mu}) + \mathbf{u}_t\end{aligned}$$

where  $\mathbf{A}(L) = \mathbf{A}_1 + \mathbf{A}_2 L + \dots + \mathbf{A}_p L^{p-1}$  denotes the  $(J+1) \times (J+1)$  lag polynomial,  $\boldsymbol{\mu} = (\mu_0, \mu_1, \dots, \mu_J)'$  and  $\mathbb{E}(u_t u_t') = \boldsymbol{\Sigma}$  is a positive definite covariance matrix.

Let us derive the optimal forecast of  $Y_{0,t}$  conditional on  $\mathcal{I}_t = \{\mathbf{y}_\tau^0, \mathbf{y}_{\tau-1}^0, \dots, \mathbf{y}_{T_0+1}^0, \mathbf{y}_{T_0}, \dots, \mathbf{y}_1\}$  for  $\tau \geq T_0$ . Denote by  $\mathbf{Q}$  be the Cholesky factor of the inverse of the covariance matrix such that  $\boldsymbol{\Sigma}^{-1} = \mathbf{Q}'\mathbf{Q}$ , where  $\mathbf{Q}$  is a lower triangular matrix such that

$$\begin{aligned}\widehat{Y}_{0,t}^N &= \mathbb{E}(Y_{0,t}^N | \mathcal{I}_t) \\ &= \mu_0 + \sum_{i=1}^J w_i (Y_{i,t} - \mu_i) + \boldsymbol{\beta}(L)'(\mathbf{y}_{t-1} - \boldsymbol{\mu})\end{aligned}\tag{1}$$

where  $w_i = q_{i+1}/q_1$ ,  $\mathbf{q} = (q_1, q_2, \dots, q_{k+1})'$  is the first row of the matrix  $\mathbf{Q}'$ ,  $\boldsymbol{\beta}(L) = \boldsymbol{\beta}_1 L + \dots + \boldsymbol{\beta}_p L^p$ , and  $\boldsymbol{\beta}_j = w' \mathbf{A}_j / w_1$ . For  $t \geq T_0$  the vector  $\mathbf{y}_t$  results from replacing the treated series by the non-treated counterfactual  $\mathbf{y}_t = (Y_{0,t}^N, Y_{1,t}, \dots, Y_{k,t})$ , where  $\widehat{Y}_{0,t}^N = Y_{0,t}$  for  $t \leq T_0$ . In practice the unknown counterfactual  $Y_{0,t}^N$  needs to be replaced by the estimate  $\widehat{Y}_{0,t}^N$ . Accordingly the sequence  $\widehat{Y}_{0,t}^N$  is obtained from a simple recursion.

An important problem with the optimal solution (1) is that it involves  $(p+1)(k+1)$  parameters which may be difficult to estimate reliably in practice. We therefore adapt some regularization for estimating the parameters. We consider three alternative regularization schemes. First, all coefficients of the dynamic model may be regularized by applying the aforementioned REGSC method as used in the static model and proposed in section 2.2. Second, the REGSC weight regularization is only applied to the vector  $(w_1, \dots, w_J)$ . The additional coefficients attached to the lags of the series are only subject to the ridge penalty. Alternatively, all coefficients of the dynamic model may be regularized by using the elastic net penalty.

### 2.7. Nonstationary time series

In most empirical application of the SC methodology the time series exhibit prominent time trends. Accordingly it is important to consider the case, where the time series are nonstationary. Therefore, we consider the case where some roots of the characteristic equation  $\det[A(z)] = 0$  are equal to one (unit roots).

First assume that the number of roots is equal to  $J + 1$ . In this case the  $J + 1$  time series are assumed to be integrated of order one and there exist no stationary linear combination (no cointegration). In this case the SC methods considered above need to be applied to the *differences* of the time series, as otherwise the estimation suffers from the spurious-regression problem. The differences are used to forecast the differenced counterfactual  $\Delta Y_{0,t}^N$  yielding the difference of the treatment effect  $\delta_\tau^\Delta = \Delta Y_{0,\tau}^I - \Delta Y_{0,\tau}^N$  for  $\tau > T_0$ . The treatment effect of interest is then obtained from cumulating the resulting effects of the differences yielding  $\bar{\delta}_\tau = \delta_{T_0+1}^\Delta + \dots + \delta_\tau^\Delta$ .

Next, assume that the donor set and the treated variable share a common stochastic trend. This is the most favorable situation that was also considered by [Harvey and Thiele \[2020\]](#). In this case there exist  $J$  independent cointegration vectors that render a stationary forecast error of the counterfactual series. We may therefore apply the same estimation procedures in section 2.2 that were originally developed for stationary time series. We also note that if the series are cointegrated with the treated series, then the weights  $w_1, \dots, w_J$  can be estimated super-consistently (that is with the convergence rate  $1/T$ ).

### 2.8. Statistical inference

In this section we briefly discuss statistical tools for assessing the reliability of the estimated treatment effect.

#### *Placebotests*

To assess whether an identified treatment effect can be deemed significant, ADH propose a model-invariant non-parametric inference procedure that is based on Fisher’s permutation test [Fisher \[1935\]](#). The first version of the test permutes the treatment across panel units. Leaving out the treatment unit, the “treatment effect” (resp. placebo effect) is estimated for each of the  $J$  control units although these units are not treated. The observed  $J$  treatment effects are considered to be realisations of the treatment estimator under the null hypothesis of no treatment. The basic assumption for valid inference is that the empirical distribution of the  $J$  placebo treatment effects converge to the true distribution of the forecast error of the counterfactual  $Y_{0,\tau}^N - \tilde{Y}_{0,\tau}^N$ . This is the case if the unconditional distribution

of  $Y_{0,\tau}^N - \tilde{Y}_{0,\tau}^N$  is the same as the unconditional distribution of  $Y_{j,\tau} - \tilde{Y}_{j,\tau}^N$  for all  $j = 1, \dots, J$ . Obviously this is approximately the case in an i.i.d. setting. It is important to notice however, that the treatment effect for unit  $j = 0$  is based on  $J$  donor series, whereas the  $J$  placebo effects are based on  $J - 1$  donors only. Therefore, even in the i.i.d. setting the unconditional distributions of the  $J$  placebo treatments tend to have a (slightly) higher variance for small  $J$ . As it is sufficient to require identical *unconditional* distributions, we can drop the assumption of independence and rely on identical distributions. For normally distributed treatment effects, this implies that the variances are identical and the expectation is equal to zero.

The second version of the test considers time permutations of the treated unit. That is ADH permute the treatment to time periods prior to the true treatment date  $T_0$ . Provided that  $T_0 > J$ , this approach can increase the power of the test, as the empirical distribution is determined by  $T_0$  observations. On the other hand this approach requires that the prediction error variance of the counterfactual is constant in time which may be a more realistic assumption than assuming the forecast error variance to be identical across the control units. Further, time permutations can be used to verify that certain donor units are not affected by the intervention at  $T_0$ , which is a crucial assumption for assigning the donor units. Lastly, as suggested by [Doudchenko and Imbens \[2016\]](#), it is also possible to combine permutations in time and in space. This would imply comparing the estimated treatment effect for  $j = 0$  and  $t = T_0$  to all  $((J + 1) \cdot T_0) - 1$  remaining permutations of time and space.

### *Credibility Intervals*

If a substantial treatment effect is observed, the call for standard errors and prediction intervals emerges naturally. For unbiased estimates, the calculation of corresponding uncertainty metrics is straightforward. However, penalized estimators such as REGSC allow for a trade-off between bias and variance reduction in order to improve the predictive accuracy. Accordingly, standard (frequentist) confidence intervals are not available. In what follows, we show how Bayesian posterior sampling can be used to quantify the estimation uncertainty of our proposed model.

As mentioned in Section 2.2 our proposed regularization is equivalent to imposing a multivariate normal prior distribution of the following form:

$$p(\mathbf{w}) \sim \mathcal{N}(\mu_0 \mathbf{1}_J, V_0)$$

where  $\mu_0 = \lambda_2 / (\lambda_1 + J\lambda_2)$

and  $V_0$  is a matrix with variances  $V_{ii} = \sigma^2(\lambda_1^{-1} - \psi/J)$  for  $i = 1, \dots, J$  and covari-

ances  $V_{ij} = -\sigma^2\psi/J$  for  $i \neq j$ , where  $\psi = \lambda_1^{-1} - (\lambda_1 + J\lambda_2)^{-1}$ . Once the posterior distribution is set up by multiplying the likelihood function with the specific prior, standard MCMC algorithms like Metropolis-Hastings or Hamiltonian Monte Carlo can be performed to obtain valid samples from the marginal posteriors. Subsequently, these samples can be used to construct Bayesian credibility intervals for the weight vector as well as for the predicted quantity.

To illustrate the estimation of Bayesian credibility intervals for the REGSC estimator, we simulated some artificial data. Specifically, we generated data for  $T_{pre} = 50$  periods of pre-treatment and  $T_{post} = 20$  periods of post-treatment data for a single treated unit and  $J = 10$  potential donors. Among these donors, only the first five provide systematic information about the counterfactual and the error is drawn from a standard normal with zero mean and unit variance. For obvious reasons, this knowledge is not accessible when we assess the ability of our estimation procedure to quantify the involved statistical uncertainty by means of Bayesian credibility intervals. Therefore, we assume  $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$  along with a weakly informative inverse gamma distribution with hyperparameters  $a = b = 0.001$  for the error variance. Combining the above REGSC prior with the error distribution and the likelihood gives us the posterior distribution:

$$p(\mathbf{w}, \sigma^2 | \lambda_1, \lambda_2) \propto \underbrace{\prod_{i=1}^{T_0} \mathcal{N}_i(0, \sigma^2)}_{\text{Likelihood}} \cdot \underbrace{\mathcal{IG}(a, b)}_{\sigma^2\text{-Prior}} \cdot \underbrace{\mathcal{N}(\mu, \Sigma)}_{w\text{-Prior}}$$

Note again the multivariate  $J$ -dimensional nature of the REGSC-prior. Due to the weight-penalty, the off-diagonal elements of the covariance matrix are non-zero and the  $J$ -dimensional distribution is not equivalent to the product of  $J$  individual priors as it would be the case for univariate shrinkage procedures like Ridge or Lasso. A more detailed explanation of the Bayesian computation with visualized Markov Chains can be found in the appendix. Note that if the error variance is known, the REGSC estimator is directly accessible via the following analytic form:<sup>2</sup>

$$\begin{aligned} p(w) &\sim \mathcal{N}(\mu, \Sigma) \\ \mu &= \left( \tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J' \right)^{-1} \left( \tilde{Z}'\tilde{y}_0 + \lambda_2 \mathbf{1}_J \right) \\ \Sigma &= \left( \tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J' \right)^{-1} \tilde{Z}' \sigma^2 I_{T_0} \left( \left( \tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J' \right)^{-1} \right)' \tilde{Z}' \end{aligned}$$

Assume we obtained 1,000 valid draws from the marginal posterior densities of

<sup>2</sup>The derivation of the analytical form of the REGSC estimator is postponed to the appendix. Cf. the closed form solution of section 2.2



the weights. Though it is possible to maintain the constant during the Bayesian estimation by appending the design matrix with a column of 1, our experience suggests that the computation without intercept and using demeaned series is more convenient.<sup>3</sup>

By sampling from the posterior distribution, we account for the estimation uncertainty of the weight vector. However, the prediction error for the counterfactual depends also on the error term. Therefore, it is also necessary to include the (normally distributed) error term.<sup>4</sup> To construct the Bayesian credibility interval of the counterfactual treatment series, we draw one observation from the normally distributed constant and the normally distributed error term and add both to the matrix product of the  $70 \times 10$ -dimensional donor matrix and the  $10 \times 1$ -dimensional first draw of the posterior distribution of the weights. This provides us with one potential counterfactual trajectory. By repeating this procedure, say, 1,000 times and disregarding the first and the last 25, we obtain the Bayesian 95% credibility interval. The following figure depicts the REGSC counterfactual with the estimated 95% credibility interval. To quantify the performance of the estimated credibility interval, we also plot the empirical confidence interval that is obtained by adding the (in practice unknown) percentiles of the error distribution (here  $\pm 1.96$ ).

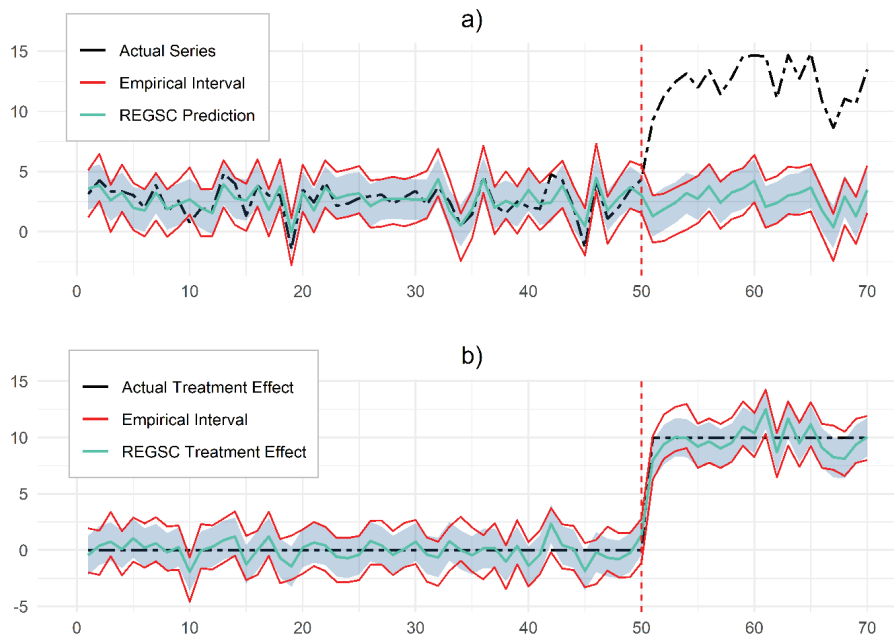
The posterior predictive distribution is an alternative, less conservative way to obtain Bayesian credibility intervals. This involves generating new observations of the dependent variable at each time point given the posterior samples and the likelihood. Again, the 2.5% and the 97.5% percentiles of these new observations form the boundaries of the 95% credibility interval.

### *Bootstrap Intervals*

Bootstrap intervals are another way of quantifying the statistical uncertainty of the estimate. More specifically, bootstrap-based calculations provide an assessment of the variance, reliable estimates of the bias are only available if a reliable unbiased estimate is available. In the REGSC context, we draw with replacement  $T_0$  observations from the  $T_0 \times J$ -dimensional donor pool, compute the REGSC prediction and store the corresponding residuals. This procedure is repeated  $B$  times, which provides us with  $B$  observations of the residual based on which we can compute mean, variance and percentiles. As the temporal ordering of the observations matters in

<sup>3</sup> Let  $\mu_{T_0}$  be the unconditional pre-treatment mean and  $\sigma_{T_0}^2$  be the unconditional pre-treatment variance. Then, the constant is distributed as  $\mu \sim \mathcal{N}(\mu_{T_0}, \sigma_{T_0}^2/T_0)$ .

<sup>4</sup> Let  $\mu_{\epsilon, T_0}$  be the pre-treatment error mean obtained by subtracting the counterfactual from the actual series. Further, let  $\sigma_{\epsilon, T_0}^2$  be the pre-treatment error term variance. Then, the error is distributed as  $\mu_\epsilon \sim \mathcal{N}(\mu_{\epsilon, T_0}, \sigma_{\epsilon, T_0}^2)$ .

**Figure 1.** Bayesian Credibility Intervals

Panel a) plots the actual series (black), the true interval (red) and the REGSC counterfactual (green). At  $T = 51$ , the series experiences a persisting treatment effect of size 10. We see that the Bayesian credibility interval has decent coverage: At only two time points (10, 45), the actual series lies outside the Bayesian interval (shaded area). Panel b) compares the actual treatment effect (black) to the REGSC estimation (green), obtained by subtracting the REGSC counterfactual from the actual series. In the pre-treatment period, the actual treatment effect of zero is within the interval in 48 of 50 cases. In the post-treatment period, the actual treatment effect of 10 falls inside the interval in 18 of 20 cases. Further, we see a high degree of similarity between empirical confidence and estimated credibility interval.

applications of SC, it is recommendable to perform the so-called block-bootstrap which consists of dividing the data into blocks of observations and sampling the blocks randomly with replacement (see e.g. [Hall et al. \[1995\]](#)).

For the parametric bootstrap with normality assumption, we require the bootstrap estimate of the residual standard deviation and the respective percentile of the normal distribution. The bootstrap interval is given by

$$\left[ y_{0t}^N - u_{1-\frac{\alpha}{2}} \hat{\sigma}_{boot}, y_{0t}^N + u_{1-\frac{\alpha}{2}} \hat{\sigma}_{boot} \right],$$

where  $u$  depicts the corresponding percentile of the normal distribution. As the sample sizes in the context of SC are generally small, the normal distribution may be hard to justify and the  $t$ -distribution can provide a suitable and more conservative

basis for the interval. In this case, the interval is given by

$$\left[ y_{0t}^N - t_{1-\frac{\alpha}{2}} \hat{\sigma}_{boot}, y_{0t}^N + t_{1-\frac{\alpha}{2}} \hat{\sigma}_{boot} \right],$$

where  $t$  represents the corresponding percentile of the t-distribution. Both intervals are only asymptotically valid and can be considerably biased in smaller samples. Therefore, the non-parametric percentile interval is another potential approach. As its name suggest, the boundaries of the interval are obtained directly by the respective percentiles of the bootstrap distribution. The percentile interval is defined as follows where  $\hat{G}$  constitutes the percentile function of the bootstrap standard deviation:

$$\left[ \hat{G}_{\frac{\alpha}{2}}^{-1}, \hat{G}_{1-\frac{\alpha}{2}}^{-1} \right]$$

An analogous figure as [1](#) can be found in the appendix. As the bootstrap interval is solely based on the pre-treatment period, the resulting bootstrap interval is somewhat optimistic and exhibits a lower coverage than the Bayesian credibility interval.

### 3. Small sample performance of alternative estimators

In this section, we compare the performance of alternative SC estimators for different data-generating processes (DGP). To this end we generate  $T_{pre} = T_0$  periods of pre-treatment and  $T_{post} = T - T_0$  periods of post-treatment data for a single treated unit and  $J$  donor units. The pre-treatment sample represents the training set for the models, the post-treatment observations define the validation set. To root the simulation framework as close as possible to real-world SC applications, we define  $T_{pre}$  and  $T_{post}$  such that their range is comparable to low-frequency macroeconomic settings, i.e.  $T_{pre} \in \{20, 50, 100\}$  and  $T_{post} \in \{10\}$ .<sup>5</sup> Furthermore we consider three types of DGP, a static factor model, a stationary dynamic vector autoregressive (VAR) process and a nonstationary process incorporating a cointegration structure.

#### 3.1. Static Data Generating Process

##### 3.1.1. Factor model

As the original SC method is based on a particular factor model, see [Abadie et al. \[2010\]](#), we first investigate the performance of alternative SC estimators in a factor model similar to the one considered in [Ferman \[2021\]](#). Rather than analyzing a

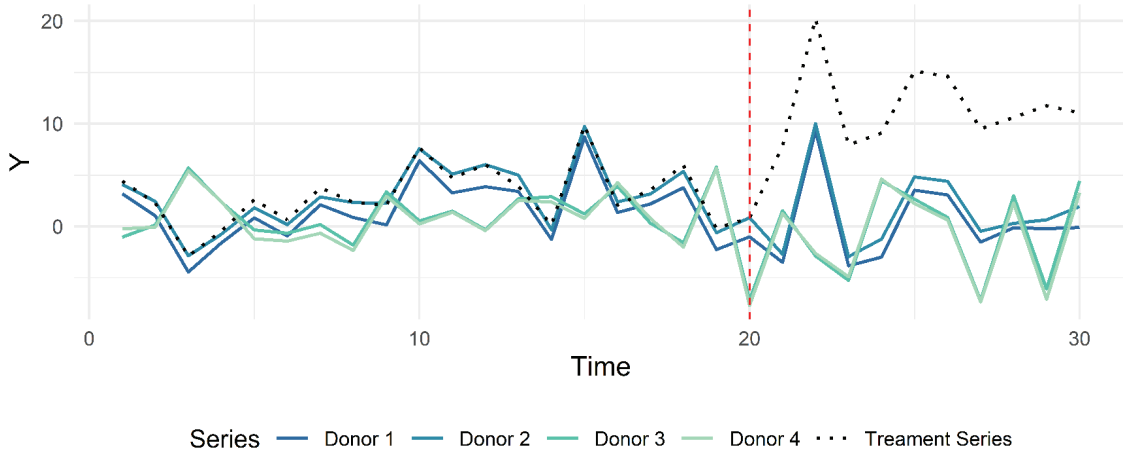
<sup>5</sup>We also considered data sets of varying post-treatment period length but noted that the main determinant of the models accuracy is the pre-treatment sample size.

fully demeaned DGP, we find it more realistic to incorporate a time-invariant and panel-specific intercept into the (potential) outcome. Our representation of the counterfactual is

$$Y_{j,t}^N = \alpha_j + \lambda_j' f_t + \epsilon_{j,t},$$

where  $\lambda_i$  is a  $r \times 1$  vector of factor loadings,  $f_t$  is a unknown  $r \times 1$  vector of common factors  $f_t = (f_{1,t}, \dots, f_{r,t})'$ ,  $\alpha_i$  is the individual specific intercept and  $\epsilon_{j,t}$  is an i.i.d. idiosyncratic shock. Ferman [2021] considers a scenario with two common factors. We proceed analogously and generate data such that the (potential) outcome of the treated unit and the first half of the donor pool load exclusively with unit loading on the first factor, whereas a second factor affects the remaining donors load with unit loading. Accordingly  $\lambda_j = (1, 0)'$  for  $j = 0, 1, \dots, J/2$  and  $\lambda_j = (0, 1)'$  for  $j = J/2 + 1, \dots, J$ . Further, the random variables  $\alpha_j, f_{1,t}, f_{2,t}$  and  $\epsilon_{j,t}$  are realizations of a standard Gaussian white noise process. The following figure exemplifies the functionality of the DGP with  $T_{pre} = 20$ ,  $T_{post} = 10$  and a constant treatment effect of  $\delta_{0,t} = 10$  for  $t \geq T_0$ . To make the factor structure more tangible, we set the

**Figure 2.** Example factor DGP



variance of the factors equal to 10 and the error variance is  $\sigma_\epsilon = 0.1$ . The generated data exhibits a clearly observable factor structure: The treatment unit and the first half of the donors (Donor 1 and 2) as well as the second half of the donors (Donor 3 and 4) share a common factor. Thus, the objective of each employed method is to recover the true factor structure, i.e. irrelevant of the size of donor pool to weight only the first  $\frac{J}{2}$  donors positively. Further, we see that each series possesses an own intercept. Yet in this specific example, the intercept variation is dominated by the factor variation.

For each of the 3 combinations of pre- and post-treatment period length  $T_{pre} \in \{20, 50, 100\}$ ,  $T_{post} \in \{10\}$  and the 6 investigated donor group sizes  $J \in \{5, 10, 15, 20, 25, 30\}$ , we simulated 500 static factor processes as described above. This simulation provides us with a total of 9,000 static factor processes that are analyzed with respect to the following metrics in the post-treatment period:

The RMSE is given by

$$\text{RMSE}_m = \left( \frac{1}{T - T_0} \sum_{t=T_0}^T (y_{0,t} - \delta_{0,t} - \hat{y}_{0,t}(m))^2 \right)^{1/2},$$

and serves as the central loss function, where the index  $m$  refers to the associated estimation approach. Note that the treatment effect  $\delta_{0,t}$  is subtracted from  $y_{0,t}$  to re-calibrate the process towards its origin.

The RMSE can be decomposed into the squared bias and the variance of the forecast. Therefore, we also report the (squared) bias component to assess the size of the systematic estimation error. To distinguish over- and underestimations we look at the distribution of the bias which is especially important when analyzing models without an intercept. On average, such models will exhibit a positive (negative) bias whenever the intercept of the treatment unit falls below (exceeds) the donor intercepts.

The Mincer-Zarnowitz (MZ) regression ([Mincer and Zarnowitz \[1969\]](#)) tests the forecast for a specific form of unbiasedness (also called autocalibration) by regressing the true value on its predicted value in the post-treatment period:

$$y_{0,t} = \beta_0 + \beta_1 \hat{y}_{0,t} \text{ for } t \in T_{post}.$$

If the forecast is rational we expect estimated coefficients close to  $(\beta_0, \beta_1) = (0, 1)$ , an hypothesis that is directly testable by a simple  $F$ -test. We therefore report the relative frequency of Monte Carlo replications for which the  $F$ -test accepted the joint hypothesis  $(\beta_0, \beta_1) = (0, 1)$  at the conventional significance level of 5%.

For the static factor DGP, we employ the following five models:

- 1. Synthetic Control (SC).** The first model is the SC method of ADH with the common restrictions (no intercept, weakly positive weights that sum up to one) yet without additional time-invariant covariates. Therefore, as outlined in section [2.5](#), the matrix of explanatory variables  $\mathbf{X}$  exclusively contains the pre-treatment  $\mathbf{Y}_{j,t}$ -series of all panel units and the weights are chosen such that the pre-treatment distance of  $\mathbf{Y}_{0,t}$  and  $\mathbf{Y}_{0,t}^N$  is minimized. This procedure of appending lagged values

of the dependent variable to the matrix of explanatory variables has already been applied in [Abadie et al. \[2010\]](#), see also section 2.5.

**2. Ordinary Least Squares (OLS).** The second method is a usual (unrestricted) least squares regression that regresses the pre-treatment treatment series on the donor series. The motivation for this approach is that it is efficient if the number of donors ( $J$ ) is small and the number of pre-treatment observations  $T_0$  is large. If  $J/T_0$  is larger than one, the OLS estimator is not available.

**3. Elastic Net Regression (NET).** The third approach we consider is the aforementioned elastic net as proposed by [Doudchenko and Imbens \[2016\]](#). Similar to the simple OLS regression, it regresses the pre-treatment treatment series on the donor series employing the Ridge and Lasso penalties. The constant is unrestricted. For the sake of simplicity and due to the short training periods, we perform a simple 3-fold cross-validation and rely on the *R* package `glmnet` of [Friedman et al. \[2010\]](#). In the conceptual introduction of the elastic net, we stressed the potential drawback of having no closed-form solution. Accordingly, the computation is considerably more demanding.

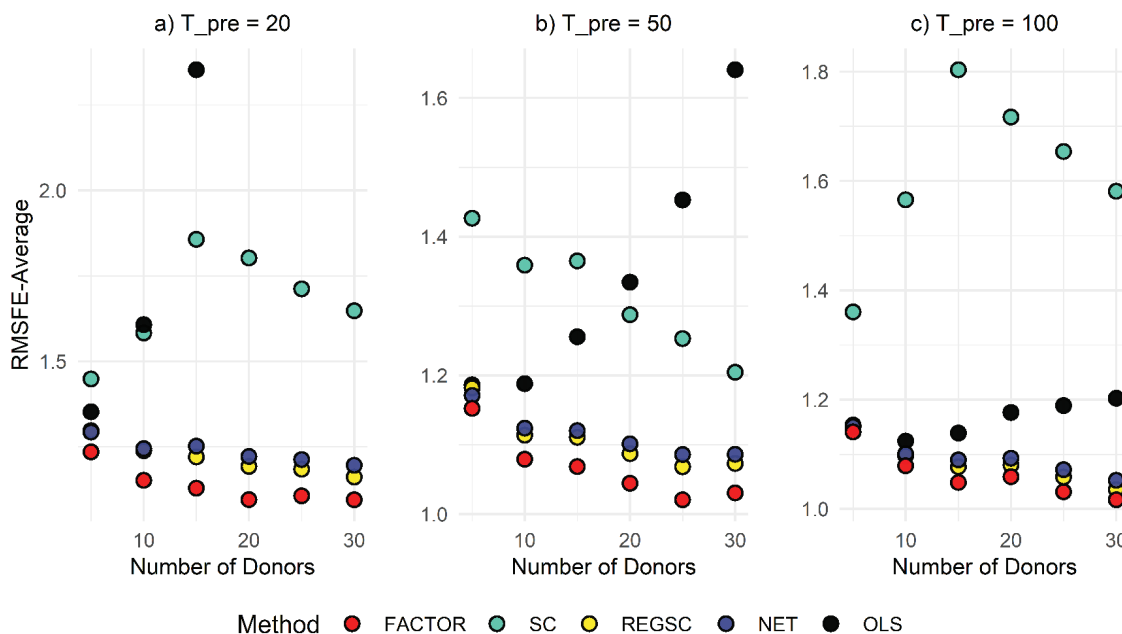
**4. Regularized Synthetic Control (REGSC).** The fourth model under consideration is our proposed regularized synthetic control estimator. It is comparable to the elastic net but substitutes the Lasso-shrinkage by the weight-penalty shrinkage. This substitution is motivated by the natural interpretation of the coefficients as SC weights. Furthermore, using this penalty term tends to produce more positive weights that better correspond to the original idea of a weighting scheme. Its closed-form solution helps to reduce the computational burden. To further reduce computation time, the hyperparameters are selected by applying a 2-fold CV in a two-step random grid-search procedure. Random hyperparameter grid search has proven to be more efficient than manual grid search both theoretically and empirically, see for instance [Bergstra and Bengio \[2012\]](#) for a careful discussion of hyperparameter optimization. In the first step of the procedure, we identify the optimal hyperparameter combination of the initialized grid by applying 2-fold CV. Based on the result of the first step, we enclose the potential optimum in the second step by sequentially holding the first and the second hyperparameter fixed while increasing and decreasing the remaining hyperparameter on a coarser grid.

**5. Factor model (FACTOR)** The data of this process is generated by a factor model and, hence, the principal component estimator suggested in section ?? seems to be most natural for this setup. In the pre-treatment period, we obtain the predictions by regressing the treatment series on the latent factors, obtained from the

first two principal components. The forecasts for the counterfactual in the post-treatment period is obtained by adapting the weighting scheme from the principal component estimator to the set of donors of the post-treatment period, yielding the factors of the post-treatment sample. The factor loading is obtained from a regression of the treatment series on the principal components of the pre-treatment sample. As this model directly builds upon the DGP, it is our benchmark-model and we expect it to perform best among all candidates.

The full simulation results can be found in table format in the appendix 6.7.1 where we group the tables at the level of the six analyzed donor quantities. Here, we give a focus on the the main results of the simulation. To do so, consider the following figure that plots the average RMSE of the models against the size of the donor pool for  $T_{pre} \in \{20, 50, 100\}$  and  $T_{post} \in \{10\}$ .

**Figure 3.** Factor Simulation Performance for  $T_{pre} \in \{20, 50, 100\}$  and  $T_{post} \in \{10\}$



Based on the results in Figure 3 and in the tables of the appendix, the following observations stand out: First, considering unrestricted OLS, we note that the estimator is not applicable in panel a) for  $J \geq 20$ . In panel b) and c), the requirement  $T_{pre} > J$  is satisfied in all cases but the estimator tends to overfit the training data indicated by the positive relation between RMSE and the number of donors. Second, considering the original SC method of ADH, we observe a less clear relation between the models accuracy and relation between  $J$  and  $T_{pre}$ . In panel a), the model provides the best forecast for  $J = 5$ , in panel b) for  $J = 30$  and in c) again for  $J = 5$ .

Furthermore we observe that the method is most precise (compared to  $T_{pre} = 20$  and  $T_{pre} = 100$ ) for  $T_{pre} = 50$  indicating that more pre-treatment observations increase the accuracy but that too many pre-treatment observations deteriorate it. This may be due to the fact that the SC model, as implemented here, fits all pre-treatment values of the treatment series, resulting in an overfitting of the pre-treatment data when  $T_{pre}$  becomes too large. Third, considering the NET, the REGSC and the FACTOR estimators, we see that the factor model dominates all remaining models in all  $J - T_{pre}$ -combinations. This is not surprising as the data are generated by a factor model. All three models successfully balance the trade-off between over- and underfitting as they become increasingly precise for increasing size of the donor pool. Lastly, considering the relative performance of NET and REGSC, we observe that NET is more precise for the three cases of  $J = 5$ . For the remaining 15 cases with  $J > 5$ , the REGSC outperforms the elastic by a small but consistent margin.

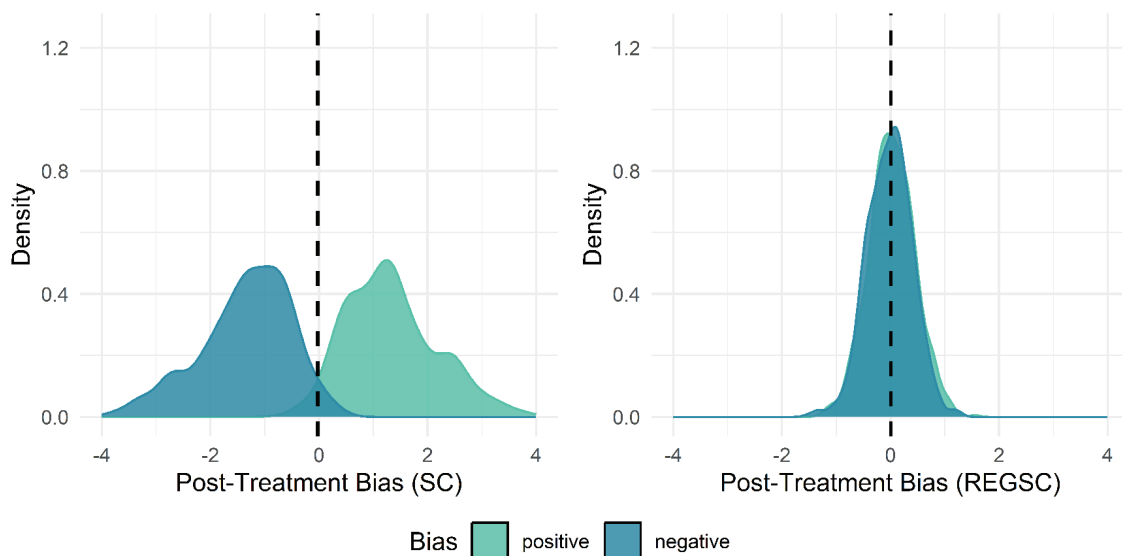
It has been emphasized that RMSE alone should not be the sole precision metric when assessing forecast performance, as it fails to detect over- and underestimation. While the bias can identify iteration-specific over- and underestimations, aggregating it into a single metric like the mean can lead to spurious optimality.<sup>6</sup> To address this issue, we initially examine bias distributions: As previously discussed, models without intercept such as the SC method will show a positive (negative) bias when the mean of the treatment series exceeds (falls below) the means of the donor series. In our simulation, the intercepts of the series are independent and identically distributed (iid) realizations of standard normal distributions. Therefore, the probability that the mean of the treatment series falls below (exceeds) all donor means equals  $\frac{1}{J+1}$ .<sup>7</sup> In order to illustrate the possible bias of the SC method, the following figure presents the distribution of the observed bias of the sample draws where the intercept was most extreme relative to the donor series (for all donor group sizes  $J$ ). We grouped the observed biases according to minimum (positive bias)/ maximum (negative bias) and present the resulting distributions of the biases for the SC and the REGSC model in Figure 4.

The aforementioned bias problem is immediately apparent: the SC method exhibits positive/ negative biases if the treatment intercept falls outside the convex hull of the donors' intercepts. In contrast, the REGSC model (as well as the remaining models that include an intercepts) do not suffer from this problem. In

<sup>6</sup> Consider for instance a model that forecasts  $\{-1, 1\}$ , each in 50% of the cases for a quantity whose optimal forecast is 0 in 100% of the cases. This model is far from optimal but the mean bias is 0.

<sup>7</sup> For each donor quantity  $J$ , there are  $(J + 1)!$  total orderings. In  $\frac{(J+1)!}{J+1}$  of the cases, the intercept of the treatment series is the most extreme. This translates to a probability of  $\frac{1}{J+1}$ .



**Figure 4.** Bias distribution for the SC and the REGSC model

appendix 6.7.1, we plot the bias-distributions for all models and combination of pre- and post-treatment periods lengths. As expected, we see that the bias distributions of all models become more tightly centered around zero the longer the pre-treatment periods. Further the SC model stands out as the only one whose performance does not consistently improve with longer pre-treatment periods. In the appendix, we present the full simulation results for the RMSE and the MZ acceptance rate. In sum, the REGSC model achieves the highest MZ acceptance rate (excluding the factor model) in 9 cases, the elastic net in 8 cases and the SC method in one case. The fact that the MZ-regressions produce somewhat different results than the RMSE is likely due to the small post-treatment period of  $T_{post} = 10$ .

Our findings from the Monte Carlo experiments of the static setup can be summarized as follows: With exception of the unrestricted OLS estimator, all models perform reasonably well in distinguishing systematic pre-treatment patterns from noise and do not indicate any severe risk of overfitting. However, the simplified version of the SC method that matches all pre-treatment values of the dependent variable faces stability issues when  $T_{pre}$  exceeds 50 periods. Further as already stressed by [Abadie et al. \[2010\]](#), we observe biased SC-forecasts if the treatment series intercept does not fall inside the range of the donor intercepts. NET and REGSC model include an intercept and allow more flexible weight coefficients which safeguards against biased forecasts. Furthermore the REGSC tends to outperform the elastic net by a small but consistent margin. Not surprisingly, the FACTOR estimator outperforms all other estimators as the data are generated by a factor model. Therefore in the

next subsection we generate data with a different correlation pattern that explicitly favors the original SC approach.

### 3.1.2. More general correlation patterns

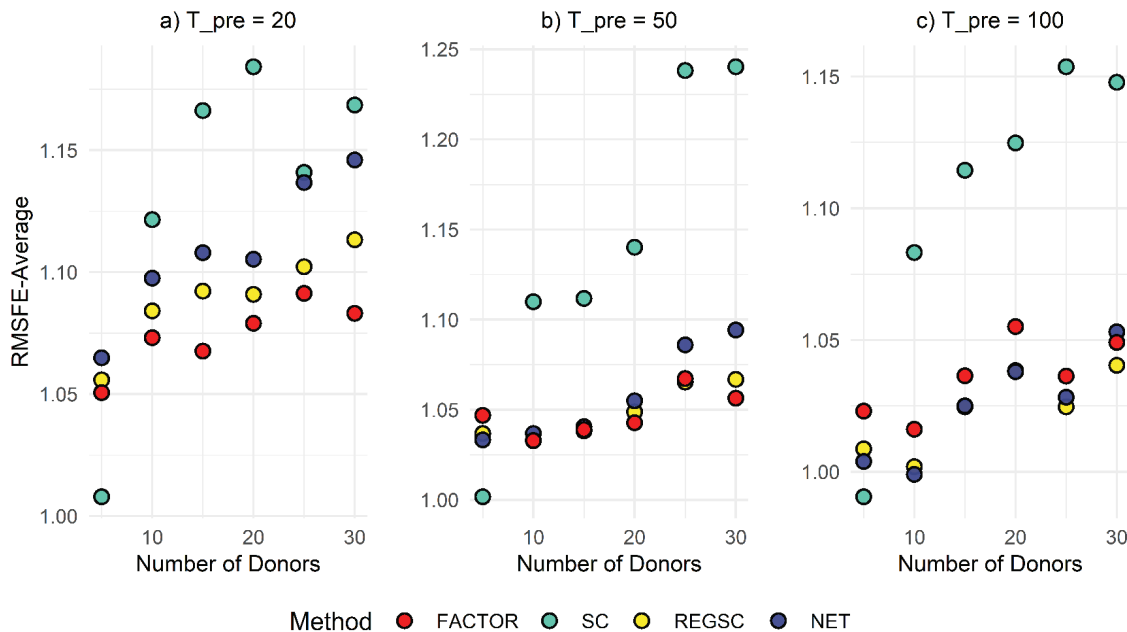
To mitigate the simulation results' reliance on a specific DGP, we assess the model's performance using another static process tailored to favor the SC model. We start by generating donor weights that correspond to the salient features of the original SC approach by randomly drawing  $J/2$  weights from a  $U[0, 1]$  distribution and setting the remaining  $J/2$  donor weights equal to zero. The weights are normalized such that they sum up to unity. The donor series are obtained by generating a  $(T_{pre} + T_{post}) \times (J - 1)$  matrix of normal random variables with covariance matrix  $RR'$ , where  $R$  is a  $(J - 1) \times (J - 1)$  matrix of independent  $U[0, 1]$  distributed elements. The counterfactual is generated as  $Y_{0,t}^N = w'Y_{j,t}^N + \alpha \cdot \epsilon$ , with  $\epsilon$  being a standard normal error and the parameter  $\alpha$  controls the signal-to-noise ratio. We set  $\alpha$  equal to one.<sup>8</sup> Analogously to the previous simulation, we simulate 500 replications, train the model on the first  $T_{pre}$  observations and test their predictive performance on the remaining  $T_{post}$  observations.

Figure 5 reports the sample RMSE across the alternative estimators (excluding OLS as it never was competitive). For a small number of donors ( $J = 5$ ) we find, as anticipated, that the original SC method outperforms all other methods. This is due to the fact that the data generating process adheres to the conditions of the SC method. For a larger number of donors, however, some additional regularization is required, as the variance of the SC method increases dramatically. This is achieved by employing the penalty terms of the REGSC estimator or by imposing the factor structure as for the FACTOR estimator. It is also interesting to note that the NET estimator performs quite poorly for  $T_{pre} = 20$ . In small samples, this estimator heavily shrinks towards zero, leading to a severe bias. Overall, our Monte Carlo experiment confirms the favorable small sample properties of the REGSC and FACTOR estimators.

## 3.2. Stationary Dynamic Data Generating Process

In order to evaluate the performance of the alternative approaches, we now proceed to study these methods in a simulation framework that mimic the real world. Given the focus of previous studies on economic development before and after treatment, it is reasonable to consider changes in GDP, such as GDP growth rates, as the basis of

<sup>8</sup> We also considered DGP with different signal-to-noise ratios and DGP that did not induce sparsity by setting  $J/2$  weights exactly to zero. These adjustments did not change the results.

**Figure 5.** SC Simulation Performance for  $T_{pre} \in \{20, 50, 100\}$  and  $T_{post} \in \{10\}$ 

a realistic SC scenario. To ensure a relatively uniform reference group with sufficient commonalities and correlations, the reference dataset includes all countries from the G20 and the European Union (EU).<sup>9</sup> The dataset is further refined to include only countries with a minimum of 40 years of GDP growth data. The resulting set of 31 selected countries form the basis for generating close-to-reality datasets, simulated using a VAR model.

One country is randomly selected as the treatment unit, and an additional  $J$  similar<sup>10</sup> donors are drawn from the remaining 30 countries. Note that the number of parameters increases dramatically with the dimension of the VAR process, so we are limited to focusing on a small number of countries. Next we estimate a VAR(2) model for these  $J + 1$  selected countries and simulate new data sets without treatment based on the estimated model. For each of the simulated data sets we estimate the counterfactual based on the  $J$  donor series and compare the estimates with the original treatment series. As in the static case, the estimation is based on a variation of pre-treatment periods  $T_{pre} \in \{30, 50, 100\}$  and the post-treatment period is set to  $T_{pre} = 20$ .

<sup>9</sup> The GDP data is extracted from the World Bank's World Development Indicators, which is directly accessible via the WDI-Package [Arel-Bundock \[2022\]](#) in R using the ticker 'NY.GDP.MKTP.KD.ZG' (GDP Growth Rate (annual %)).

<sup>10</sup> The donors are selected in a way such that they share a similar correlation structure with the treatment unit.

For the dynamic setup, we employ all of the previously methods from the static case for reference. Furthermore we introduce estimators that accommodate specific dynamic features, see also section 2.6. In particular we apply the following dynamic approaches:

1. *Unrestricted VAR (VAR)*: The first dynamic estimator uses an unrestricted vector autoregressive (VAR) model. In this approach the lagged donor series and lags of the treatment series are treated as additional donors and the corresponding weights are estimated by OLS. Accordingly, this estimator is the best predictor of the synthetic control as  $T_{pre}$  tends to infinity. In finite samples, however, this estimator suffers from the large number of parameters and typically shows a poor small sample properties. Nevertheless this estimator serves as a benchmark model for the dynamic case as it closely mimics the DGP.

2. *Regularized VAR (REGVAR and NETVAR)*: As the VAR model involves a large number of parameters, it is important in empirical practice to apply some regularization scheme. We therefore extend the regularization of the static setup to the VAR model. The NETVAR method applies the elastic net penalty to all coefficients of the regressors, whereas the REGVAR approach applies the REGSC penalty to all coefficients.<sup>11</sup>

As in the static case, further simulation results can be found in the appendix 6.7.2 while the key aspects are presented here. In Table 1 the simulation results in terms of RMSE are presented for  $T_{pre} = 50$  and  $T_{post} = 20$  periods over 500 iterations for each donor group.<sup>12</sup>

Overall we find that the regularized dynamic approaches outperform the static methods in our dynamic setup. Especially for larger donors sets the dynamic models perform significantly better than their static counterparts and in particular the original SC method. This is highlighted by the boxplots for the RMSE presented in Figure 6. The regularized dynamic estimators not only have the lowest RMSE median but also the smallest interquartile range. It further demonstrates that the performance of the dynamic estimators in comparison with the static estimators increases steadily with an increasing number of donors. Moreover, among the dynamic models, elastic net regularization emerges as the most efficacious regularization method in this simulation setting indicating the strength of the elastic net regularization to

<sup>11</sup> In this dynamic simulation section (other than in the static case) 3-fold cross validation is used since it appeared more effective than 2-fold cross validation. For all dynamic models the lag order is set to  $p = 2$  to align with the simulation settings.

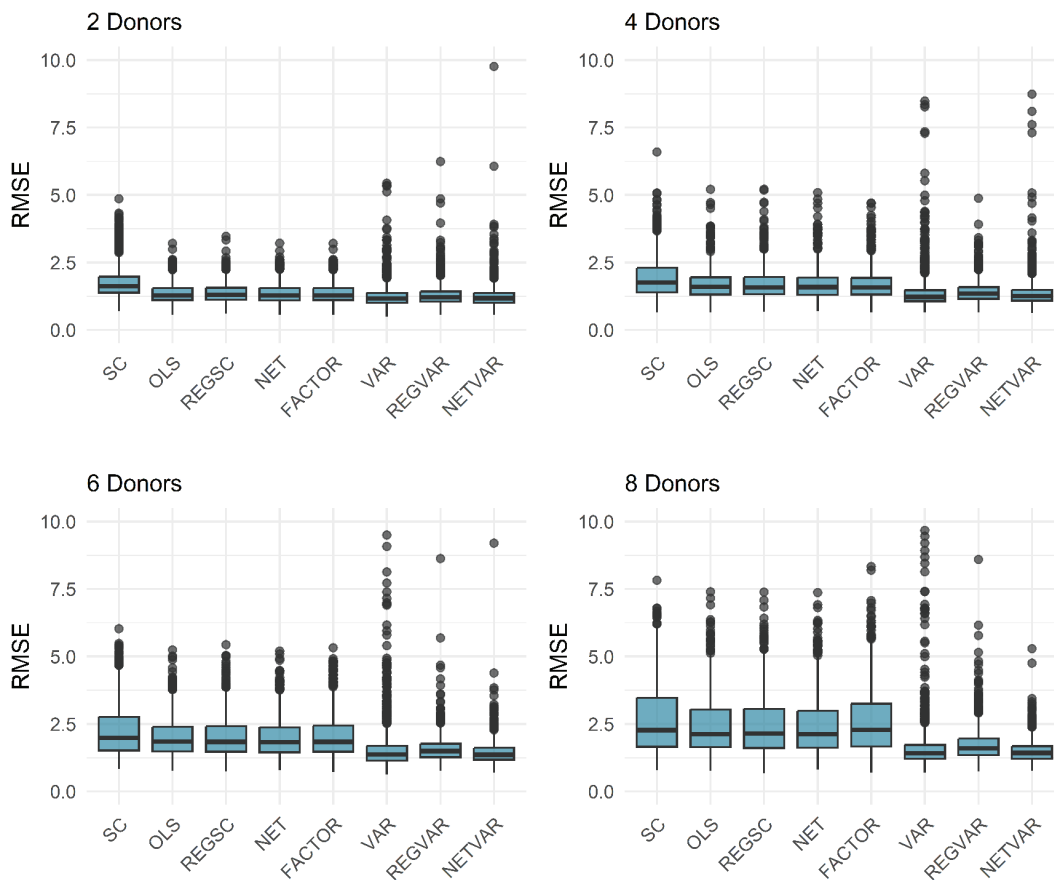
<sup>12</sup> The table contains the average values over all replications for the specific number of donors. For very few iterations (approximately 0.5%), some models produced very large outliers, which are excluded before averaging.

**Table 1.** RMSE for stationary data ( $T_{pre} = 50, T_{post} = 20$ )

| Method | $J = 2$ | $J = 4$ | $J = 6$ | $J = 8$ |
|--------|---------|---------|---------|---------|
| SC     | 1.815   | 2.005   | 2.251   | 2.669   |
| OLS    | 1.349   | 1.688   | 2.007   | 2.444   |
| REGSC  | 1.359   | 1.696   | 2.012   | 2.446   |
| NET    | 1.347   | 1.678   | 1.983   | 2.410   |
| FACTOR | 1.349   | 1.672   | 2.043   | 2.573   |
| VAR    | 1.281   | 1.469   | 1.992   | 1.898   |
| REGVAR | 1.312   | 1.432   | 1.681   | 1.744   |
| NETVAR | 1.277   | 1.490   | 1.632   | 1.554   |

generate sparse solutions.

**Figure 6.** Stationary simulation RMSE comparison for different sizes of the donor Set ( $T_{pre} = 50, T_{post} = 20$ )



### 3.3. Nonstationary dynamics

In many empirical applications the treatment and donor series appear to be nonstationary. It is therefore important to study the ability of the methods to deal with nonstationary data. This section is divided into two parts. The first part examines a nonstationary simulation case which as a comparison is derived from the same GDP data basis as in the static case. The second part examines a new DGP that examines the effects of cointegrated data in a nonstationary time series setting.

The comparison simulation is based on the same data as in the stationary DGP in section 3.2, but now we apply all methods to (the logarithms of) GDP *levels* instead of GDP growth rates. To ensure direct comparability, the GDP growth series are drawn from the same data basis as in the stationary case but are cumulated before estimating the counterfactual. Since (log) GDP possess a VAR(3) representation with time trends, we include a linear trend in the VAR(3) representation when estimating and simulating the dynamic processes in levels. For computing the RMSE

the actual and forecasted counterfactual series is differenced in order to allow for a direct comparison with the results in section 3.2.

**Table 2.** RMSE for nonstationary data ( $T_{pre} = 50$ ,  $T_{post} = 20$ )

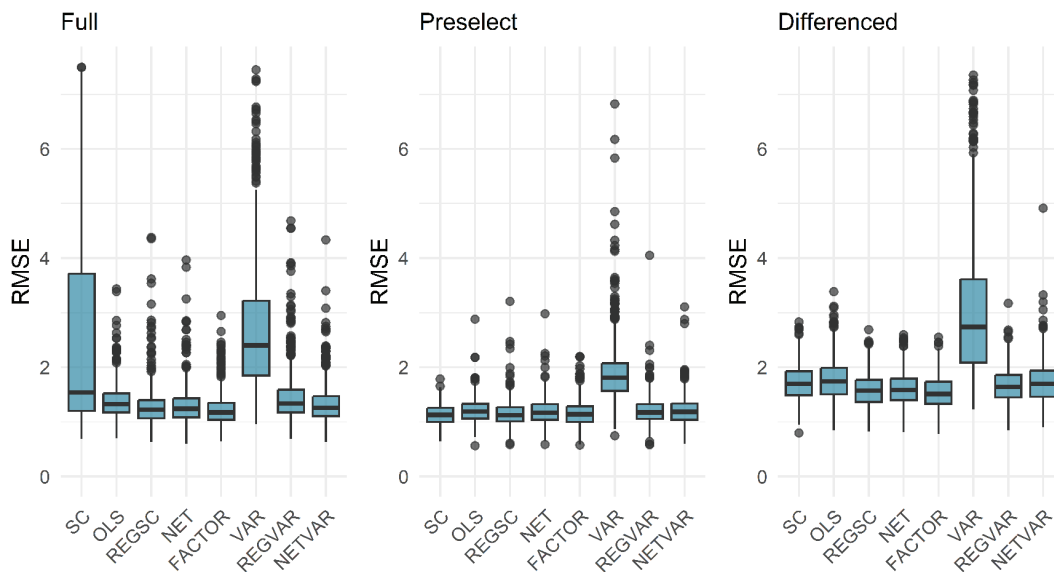
| Method | $J = 2$ | $J = 4$ | $J = 6$ | $J = 8$ |
|--------|---------|---------|---------|---------|
| SC     | 1.6556  | 2.1518  | 2.223   | 2.6359  |
| OLS    | 1.6194  | 2.363   | 2.4976  | 2.7524  |
| REGSC  | 1.5816  | 2.1598  | 2.2576  | 2.6526  |
| NET    | 1.6004  | 2.2647  | 2.3885  | 2.6409  |
| FACTOR | 1.6194  | 2.031   | 2.1988  | 2.7458  |
| VAR    | 1.2948  | 1.4929  | 1.7226  | 2.4175  |
| REGVAR | 1.3848  | 1.7407  | 1.9144  | 2.2169  |
| NETVAR | 1.2425  | 1.4157  | 1.5791  | 1.8216  |

The results demonstrate that the static approaches have severe problems to cope with nonstationary time series. Compared to the (more appropriate) use of differenced time series in section 3.2 the RMSE is much larger, whereas the dynamic methods perform similar whether the series are differenced or not. This is due to the fact that the dynamic approaches include lagged differences and, therefore, the dynamic specification is able to incorporate unit roots in the autoregressive specification if necessary. Hence it is important to check the time series for nonstationarity before applying static SC methods. On the other hand, dynamic methods turn out to be fairly robust against a possible nonstationarity of the data.

The following subsection presents a new DGP is designed to represent a nonstationary environment in which one part of the potential donor data is cointegrated with the treatment unit while the other part is not. The construction of the GDP is similar to the factor model used in the static simulation case in section 3.1.1. To account for the cointegration structure, two independent random walk processes are used as factors, from which the resulting data are derived in such a way that half of the donors are cointegrated with the treatment unit and the other half are not. This setup allows us to analyze three different scenarios. In the first scenario, the data is used as it is, i.e. the full nonstationary data set with both the with the counterfactual cointegrated and non-cointegrated donor sets. In the second scenario, we perform preselection in a way that the non-cointegrated data is removed before the methods are applied. Hence, only the donor sets which are cointegrated with the treatment unit remain. In the third scenario we keep the full data set but take differences in order to account for the nonstationary time structure and obtain stationary data.

Table 3 report the results of the cointegrated DGP for  $J \in \{4, 8, 12, 16, 20\}$  and  $T_{pre} = 50$  and  $T_{post} = 20$ .<sup>13</sup> In general the results are in line with our theoretical reasoning. Compared to the non-cointegrated setup of the previous subsection, the gap between the performance of the static and dynamic methods are considerably smaller. Nevertheless, the dynamic methods outperform the static competitors by a substantial margin. This is also apparent from the boxplots presented in Figure 7, where the RMSE is averaged across all  $J$  and 8, for the individual donor groups.

**Figure 7.** Cointegrated series comparison ( $T_{pre} = 50, T_{post} = 20$ )



Especially the SC and VAR models suffer severely in the full data set scenario indicating problems regarding spurious regression. However, the issues regarding the VAR model may also arise from overfitting due to the comparable large amount of donors<sup>14</sup> and their lagged values which enter the model estimation without regularization. This assumption is also supported by the fact, that the VAR models RMSE is also quite high in the other scenarios.

Overall the results strikingly demonstrate the benefits of variable selection, regularization and cross validation under such a regime. Removing the non-cointegrated data from the donor set, i.e. a priori preselection of variables is beneficial for all models and for all donor amounts under this simulation setting. However, this is also an artifact of the simulation design, where the non-cointegrated data are not informative for estimating the counterfactual. Interestingly, not only does the SC model benefit the most from the preselection due to its very large RMSE beforehand,

<sup>13</sup> Further details are given in appendix 6.7.3.

<sup>14</sup> As before, a lag order of  $p = 2$  is used in estimation



but also in this setup of only cointegrated donors, the SC model performs slightly better than all other models.

Keeping the full donor set intact but establishing stationarity due to differencing the time series also yields significantly better results for the SC model. However, for all other models the cost of loss of information due to the differencing is reflected by slightly higher RMSEs compared to the preselected, but also the full scenario. The simulation results strongly emphasize the importance of addressing the problem of nonstationary in a non-cointegrated time series environment.

**Table 3.** RMSE for cointegrated series ( $T_{pre} = 50, T_{post} = 20$ )

| DGP         | Method | $J = 4$ | $J = 8$ | $J = 12$ | $J = 16$ | $J = 20$ |
|-------------|--------|---------|---------|----------|----------|----------|
| Full        | SC     | 1.2748  | 1.1701  | 5.0211   | 8.5764   | 8.0811   |
|             | OLS    | 1.4048  | 1.2971  | 1.3124   | 1.3975   | 1.4783   |
|             | REGSC  | 1.3978  | 1.2847  | 1.2381   | 1.2457   | 1.2284   |
|             | NET    | 1.4289  | 1.2971  | 1.2415   | 1.2565   | 1.2608   |
|             | FACTOR | 1.3869  | 1.2307  | 1.1683   | 1.1413   | 1.1733   |
|             | VAR    | 2.0577  | 2.1316  | 2.7562   | 3.4028   | 11.1224  |
|             | REGVAR | 1.5814  | 1.4856  | 1.4404   | 1.4129   | 1.4221   |
|             | NETVAR | 1.4041  | 1.3272  | 1.2818   | 1.2803   | 1.3318   |
|             |        | $J = 2$ | $J = 4$ | $J = 6$  | $J = 8$  | $J = 10$ |
| Preselect   | SC     | 1.2035  | 1.1576  | 1.0988   | 1.1060   | 1.0906   |
|             | OLS    | 1.2574  | 1.2097  | 1.1490   | 1.1922   | 1.1972   |
|             | REGSC  | 1.2544  | 1.1908  | 1.0949   | 1.1194   | 1.0911   |
|             | NET    | 1.2663  | 1.2172  | 1.1353   | 1.1814   | 1.1496   |
|             | FACTOR | 1.2465  | 1.1685  | 1.0815   | 1.1046   | 1.0839   |
|             | VAR    | 1.7811  | 1.8383  | 1.7774   | 1.9859   | 2.2580   |
|             | REGVAR | 1.2847  | 1.2287  | 1.1570   | 1.2016   | 1.1511   |
|             | NETVAR | 1.2773  | 1.2242  | 1.1555   | 1.2074   | 1.1642   |
|             |        | $J = 4$ | $J = 8$ | $J = 12$ | $J = 16$ | $J = 20$ |
| Differenced | SC     | 1.7184  | 1.6439  | 1.6593   | 1.9005   | 1.9014   |
|             | OLS    | 1.6785  | 1.6489  | 1.5444   | 1.6819   | 1.7263   |
|             | REGSC  | 1.6847  | 1.6184  | 1.4813   | 1.5551   | 1.5570   |
|             | NET    | 1.6745  | 1.6172  | 1.5014   | 1.5724   | 1.5858   |
|             | FACTOR | 1.6338  | 1.5402  | 1.5232   | 1.5093   | 1.4567   |
|             | VAR    | 1.8922  | 2.0406  | 2.2225   | 2.4952   | 2.8434   |
|             | REGVAR | 1.7536  | 1.6883  | 1.5927   | 1.6059   | 1.6651   |
|             | NETVAR | 1.6907  | 1.6995  | 1.5655   | 1.6484   | 1.6416   |

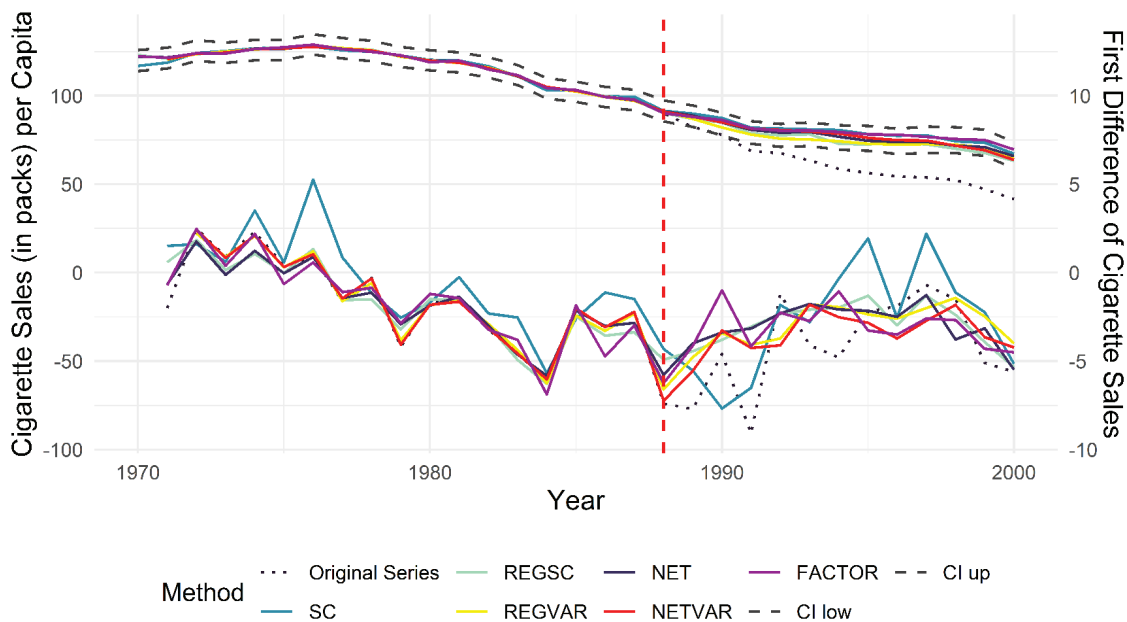
## 4. Applications

In this section, we consider two leading examples of ADH, the effect of California’s tobacco control program [Abadie et al. \[2010\]](#) and the economic cost of the 1990 German Reunification [Abadie et al. \[2015\]](#). We compare the outcomes that result from the different estimation methods and pay special attention to the prevalence of non-stationarity in the time series.

### 4.1. Estimating the Effect of California’s Tobacco Control Program

[Abadie et al. \[2010\]](#) estimate the treatment effect of a large anti-smoking legislation in California called *Proposition 99*. The outcome of interest is per capita smoking in California and 38 U.S. states without legislation serve as Donors. ADH build their estimation on  $T_{pre} = 18$  (1970 – 1987) pre-treatment and  $T_{post} = 13$  (1988 – 2000) post-treatment periods and reckon that *Proposition 99* had a substantial, time-increasing negative effect on per capita cigarette sales of about 26 packs by the year 2000. The following figure plots the re-estimation results of *Proposition 99* in levels and due to the existence of a clear downward time trend, also in first differences. The dashed dark-grey lines in the upper graph depict the 95% credibility interval based on the Bayesian representation of the REGSC model.

**Figure 9.** Cigarette Sales per capita for (synthetic) California



The dashed vertical red line indicates the treatment date. Left axis, upper graph: original series in levels. Right axis, lower graph: transformed series of first differences.

Considering the pre-treatment period, we observe that all models are capable of

reproducing the original series in levels reasonably well. All predictions fall comfortably inside the estimated 95% credibility interval, indicating that the pre-treatment estimation uncertainty is neglectable. A distinct perspective emerges when examining the series of first differences, revealing erratic behavior characterized by significant upward and downward swings. While the SC estimator struggles to accurately capture the actual series, NET, NETVAR, REGSC, REGVAR and FACTOR methods appear to adeptly navigate the complexities of the first difference time series.

Focusing on the post-treatment period and level data, we note that all employed models produce similar estimated treatment effects that range between yearly averages from -14.43 (REGVAR) to -19.74 (NETVAR) packages per capita. Cumulated over the entire time window from 1988 until 2000, this translates into an average reduction of 173.21 (REGVAR) to 236.90 (NETVAR) packages per capita due to *Proposition 99* and our REGSC 95% credibility intervals suggest that the actual cumulated effect of *Proposition 99* could reasonably range from -116 to -282. For the series of first differences, we observe a substantial slowdown in the absolute growth rate of per Capita smoking after *Proposition 99*. NET, REGSC and the factor model generate larger but less similar counterfactuals than in the level-data case. In contrast, the SC model struggles and does neither produce a credible pre- nor a credible post-treatment counterfactual path.

In table 4 we document the full estimation results for all models and both data types. Besides pre-treatment  $R^2$ , end-of-sample ( $\delta_{end}$ ), mean ( $\bar{\delta}$ ) and cumulated treatment effects ( $\sum \delta$ ), we also show the permutation  $p$ -values of ADH obtained by dividing the post-RMSE by the pre-RMSE. By considering the ratio of pre- and post-treatment fit, the approach of ADH implicitly guards against overly conservative test results. Through pre-RMSE-scaling, donor units that are difficult to predict during the pre-treatment period are prevented from yielding extreme test statistics. For the sake of comparability, the results of the models for the differenced data have been cumulated to transform them back to level.

**Table 4.** Estimation Results: Tobacco Control Program

| Type      | Method | $R^2$  | $\delta_{end}$ | $\bar{\delta}$ | $\sum \delta$ | $p$ -value |
|-----------|--------|--------|----------------|----------------|---------------|------------|
| Levels    | SC     | 0.9764 | -25.58         | -18.84         | -226.09       | 0.026      |
|           | REGSC  | 1.0000 | -21.51         | -14.52         | -174.23       | 0.026      |
|           | REGVAR | 0.9983 | -23.68         | -14.51         | -174.08       | 0.179      |
|           | NET    | 0.9984 | -24.21         | -16.36         | -196.32       | 0.026      |
|           | NETVAR | 0.9993 | -22.13         | -16.63         | -199.50       | 0.103      |
|           | FACTOR | 0.9959 | -28.04         | -18.73         | -224.76       | 0.051      |
| 1st Diff. | SC     | 0.8338 | -38.05         | -29.03         | -348.39       | 0.359      |
|           | REGSC  | 0.9914 | -14.43         | -11.45         | -137.41       | 0.205      |
|           | REGVAR | 0.9995 | -15.42         | -10.01         | -120.12       | 0.103      |
|           | NET    | 0.9938 | -13.49         | -11.44         | -137.27       | 0.205      |
|           | NETVAR | 0.9999 | -12.55         | -8.64          | -103.70       | 0.128      |
|           | FACTOR | 0.9943 | -12.28         | -11.36         | -136.35       | 0.128      |

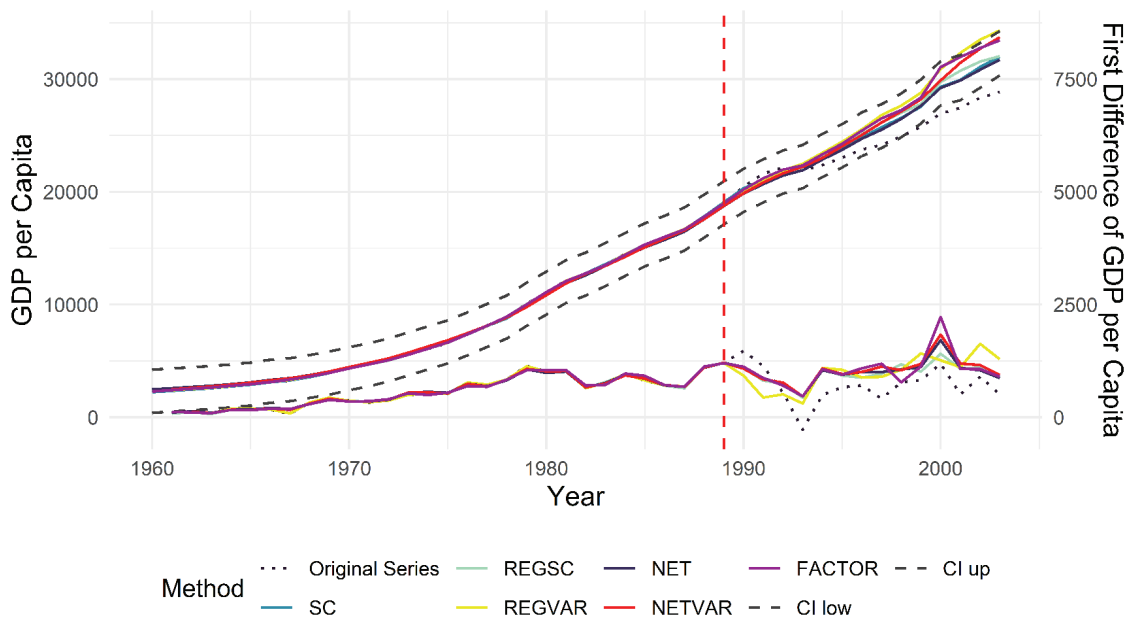
For the level data, we obtain promising  $p$ -values around 5% for the four static models. On the other hand, the  $p$ -values of the two dynamic models seem unrealistically high. Most likely, this observation points towards overfitting issues as if both treatment and donor series are predicted with high accuracy in the pre-treatment period, the permutation test does not consider post-treatment deviation an extreme event. All estimations based on the differenced data suggest smaller and thus also less significant treatment effects. Summarizing this application, we see that independent of the assumptions and estimation methods, all models estimate large negative treatment effects. Further, we observe that model uncertainty is not neglectable as the range of estimated treatment effects varies substantially with the model under consideration. Our credibility intervals for the REGSC model offer orientation in this uncertain environment and it speaks in favor of a significant causal effect of *Proposition 99* that all employed models fall inside the estimated uncertainty bandwidth. Considering the  $p$ -values of the dynamic models and the differenced data, we recommend to follow the parsimony-principle and to employ less parameterized models if the data is noisy and the pre-treatment period is short. However, as the originally SC methods provide a substantially larger effect than most of the remaining models, it is possible that the actual effect of *Proposition 99* was somewhat smaller than initially estimated.

#### 4.2. The Economic Cost of the 1990 German Reunification

The reunification of East and West Germany coincided with a substantial slowdown of GDP per capita growth in West Germany. [Abadie et al. \[2015\]](#) use this

natural experiment as another application of their SC method. In contrast to the *Proposition 99* application, the reunification dataset is somewhat longer as data is observed for  $T_{pre} = 30$  (1960 – 1989),  $T_{post} = 14$  (1990 – 2003) years and  $J = 16$  donors and West Germany. From 1992 onward, ADH identified a negative treatment effect of about \$1,600 per capita and year, translating approximately into an 8% reduction compared to the 1990 baseline level. Similar to the preceding example, the following figure depicts the original and the differenced series for selected SC methods along with 95% credibility intervals for the REGSC approach.

**Figure 10.** GDP per capita for the (synthetic) West Germany



First, it is important to note that the GDP series exhibit a nonstationary pattern, where prior to the German reunification the growth path of per capita GDP was characterized by an exponential trend. This growth pattern is very well captured by the different SC methods. Interesting differences emerge after the reunification, where – as already mentioned by ADH – the primary impact of the reunification appears with a temporal delay of approximately 5 years. The 95% credibility interval suggests that the temporal delay may be even longer as only by the year 1999, the credibility interval indicates a statistically significant treatment effect at the conventional 5% level. This observation indicates an interesting uncertainty trade-off: As the time distance to the treatment date grows, so does the distance between original series and credibility interval. Eventually, the distance is large enough to obtain statistically significant results. However, if this distance has

grown large enough, it becomes questionable whether the model that was trained in the distant past still possesses external validity. Further, it is interesting to note that even though the visual inspection indicates comparable effects, the estimated annual treatment effect across all models and data types is quite different and ranges from -884.98 (REGVAR, differenced data) to -2252.51 (REGVAR, level data). The credibility interval of the average REGSC effect ranges even from -3260 to 623, suggesting a huge uncertainty around the point estimates. This implies an insignificant average treatment effect that is mainly due to the large temporal delay of the effect. Table 5 summarizes the findings and again also presents the  $p$  values of the permutation test.

**Table 5.** Estimation Results: German Reunification

| Type      | Method | $R^2$  | $\delta_{end}$ | $\bar{\delta}$ | $\Sigma \delta$ | $p$ -value |
|-----------|--------|--------|----------------|----------------|-----------------|------------|
| Levels    | SC     | 0.9997 | -3134.1        | -1276.4        | -17870.1        | 0.059      |
|           | REGSC  | 1.0000 | -3187.9        | -1482.6        | -20756.9        | 0.118      |
|           | REGVAR | 1.0000 | -5480.9        | -2252.5        | -31535.2        | 0.235      |
|           | NET    | 0.9999 | -2858.4        | -1058.5        | -14818.7        | 0.353      |
|           | NETVAR | 0.9998 | -4095.6        | -1379.8        | -19316.8        | 0.176      |
|           | FACTOR | 1.0000 | -4593.7        | -2032.5        | -28454.2        | 0.059      |
| 1st Diff. | SC     | 0.9997 | -3537.4        | -1606.1        | -22485.7        | 0.059      |
|           | REGSC  | 0.9999 | -3710.3        | -1474.0        | -20636.4        | 0.118      |
|           | REGVAR | 1.0000 | -4000.9        | -885.0         | -12389.7        | 0.588      |
|           | NET    | 0.9999 | -4186.1        | -1706.1        | -23885.3        | 0.118      |
|           | NETVAR | 0.9999 | -4800.8        | -1870.7        | -26189.8        | 0.118      |
|           | FACTOR | 0.9999 | -4751.7        | -1887.5        | -26424.9        | 0.059      |

Interestingly, while SC-, REGSC- and FACTOR seem to cope with the pre-treatment level data, both versions of the NET struggle to some extent to precisely match the pre-treatment path of per capita GDP. This observation is noteworthy as from a theoretical perspective, the elastic net with static data should nest the solution of the simplified SC model without covariates. The dynamic REGVAR estimator produces cumulated effects that exceed the estimates of the original SC model by almost 100%. Similar to the previous example, the  $p$ -values were computed based on the ratios between pre- and post-treatment fit. Considering the results for the differenced data, we observe an improved predictive performance of the NET estimator and, as expected, a higher precision for the dynamic models compared to their static counterparts. The dynamic models seem to require a more attentive modeling procedure and we again recommend to follow the parsimony principle and opt for the simplest models, i.e. the SC-model and the static REGSC model

for the level data. In summary, we find again that that different models yield different (though in comparable magnitude) treatment effect, stressing once more the crucial aspect of model uncertainty. Our credibility intervals provides again some degree of certainty and demonstrates the above outlined uncertainty-trade-off. Considering the estimation results of the different models, we suggest to regard the estimated effect of the original SC estimator as a conservative estimate, with substantial uncertainty surrounding the point estimates.

## 5. Conclusion

More than 20 years ago, ADH proposed a robust, reliable and easily implemented method for estimating the causal effect in macroeconomic time series that does not require hyperparameter tuning, allows for the inclusion of additional explanatory variables, and is applicable even if the size of the donor pool  $J$  exceeds the number of pre-treatment observations  $T_0$ . It is therefore not surprising that [Athey and Imbens \[2016\]](#) describe the method as “arguably the most important innovation in the policy evaluation literature in the last 15 years”.

In this paper we consider a general model framework for causal inference in a static and dynamic environment. In this framework, the statistical challenge is to devise an efficient estimator for the counterfactual, drawing from the relevant information set. In empirical practice, the crucial issue is to construct a reliable estimator in a situation where the number of donors is of similar magnitude than the number of pre-treatment observations. In this case some regularisation is required. The original SC approach by ADH imposes the restrictions that all weights are positive and sum up to unity, whereas [Doudchenko and Imbens \[2016\]](#) propose an elastic net regularization. Instead of shrinking the weights towards zero our regularization shrinks the sum of weights towards unity, which can be seen a more flexible version of the ADH regularization. As the SC approach is often motivated by using a factor framework, we also propose an SC estimator that is derives from a principal component analysis of the covariance matrix. Our Monte Carlo simulations suggest that both estimators yield a significant improvement over the existing SC methods. Furthermore our regularized estimator admits a natural Bayesian representation that allows us to compute Bayesian credibility intervals to assess the involved estimation uncertainty. Especially in contexts where interest relies in the cumulated treatment effect, it is crucial to assess the uncertainty by providing intervals instead of solely focusing on point estimates.

Recognizing the prevalence of time series dynamics in many applications, we highlight the limitations of static models for estimating the treatment effect. We

therefore propose dynamic versions of the SC estimators that append the donor pool by lagged donor series. This approach significantly increases the possibility of overfitting the pre-treatment data and it is therefore important to adopt a suitable regularization. In the dynamic context, our Monte Carlo experiments show a clear superiority of the elastic net shrinkage which is due to the ability of the elastic net to obtain sparse solutions. Generally, we recommend to prioritize rows over columns as we note that additional donors and covariates that explain the structural characteristics of the panel units (e.g. columns) are less valuable than longer pre-treatment time series (e.g. rows).

Another important issue is that in many empirical applications the time series (for example GDP of different countries) are nonstationary. In this case there may not exist a stable long-run relationship between the treatment and donor series that can be used to construct a reliable SC unit. Accordingly there is a risk that static SC methods suffer from some variant of the spurious regression fallacy. In this case it is important to apply the static methods to the differenced time series. The dynamic approaches circumvent this problem as they are able to accommodate unit roots in the VAR representation of the time series.

Lastly, all SC approaches implicitly assume that the employed model generated the data at hand, so model uncertainty can be neglected. A possible approach for further work would be to investigate how different estimated counterfactuals can be optimally combined, where we again stress the importance of de-meaning intercept-free models to avoid biased predictions. In the context of the SC method, working with demeaned series implies that the donor pool is used to explain the deviations of the treatment series from its pre-treatment mean only. This approach is equivalent to incorporating a constant and recovers the appealing percent interpretation of the coefficients. After the estimation of individual models, a natural candidate for a model stacking approach would be a Bayesian model averaging approach that takes into account both model and estimation uncertainty.



## References

- Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93:113–132, 02 2003. doi: 10.1257/000282803321455188.
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california’s tobacco control program. *Journal of the American Statistical Association*, 105:493–505, 02 2010. doi: 10.1198/jasa.2009.ap08746.
- Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510, 2015. ISSN 00925853, 15405907. URL <http://www.jstor.org/stable/24363579>.
- Vincent Arel-Bundock. *WDI: World Development Indicators and Other World Bank Data*, 2022. URL <https://CRAN.R-project.org/package=WDI>. R package version 2.7.8.
- Susan Athey and Guido Imbens. The state of applied econometrics - causality and policy evaluation. *Journal of Economic Perspectives*, 31, 07 2016. doi: 10.1257/jep.31.2.3.
- Marta Bańbura, Domenico Giannone, and Lucrezia Reichlin. Large bayesian vector auto regressions. *Journal of applied Econometrics*, 25(1):71–92, 2010.
- Paul A Bekker. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62(3):657–81, 1994.
- James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, 13(null):281–305, feb 2012.
- Nikolay Doudchenko and Guido W. Imbens. Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis. NBER Working Papers 22791, National Bureau of Economic Research, Inc, October 2016. URL <https://ideas.repec.org/p/nbr/nberwo/22791.html>.
- Bruno Ferman. On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls. *Journal of the American Statistical Association*, 116(536):1764–1772, October 2021. doi: 10.1080/01621459.2021.196. URL <https://ideas.repec.org/a/taf/jnlasa/v116y2021i536p1764-1772.html>.

- R. A. Fisher. The design of experiments (9th ed.)[1971]. *Macmillan*, 1935.
- Jerome H. Friedman, Trevor Hastie, and Rob Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(i01), 2010. doi: <http://hdl.handle.net/10.1002/jae.2802>.
- Peter Hall, Joel L. Horowitz, and Bing-Yi Jing. On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574, 1995. ISSN 00063444. URL <http://www.jstor.org/stable/2337534>.
- Andrew Harvey and Stephen Thiele. Cointegration and control: Assessing the impact of events using time series data. *Journal of Applied Econometrics*, 36, 12 2020. doi: 10.1002/jae.2802.
- A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67, 1970.
- Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986. doi: 10.1080/01621459.1986.10478354. URL <https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478354>.
- Jacob A Mincer and Victor Zarnowitz. The evaluation of economic forecasts. In *Economic forecasts and expectations: Analysis of forecasting behavior and performance*, pages 3–46. NBER, 1969.
- Jerzy Neyman. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science* 5, 4:465–472, 1923.
- D.B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.
- R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58:267–288, 1996.

## 6. Appendix

### 6.1. The limit of REGSC for $\lambda_1 \rightarrow \infty$ and $\lambda_2 \rightarrow \infty$

For  $\lambda_1 \rightarrow \infty$  and  $\lambda_2 \rightarrow \infty$  the objective function reduce to

$$Q(\lambda_1, \lambda_2) = \lambda_1 \mathbf{w}'\mathbf{w} + \lambda_2 (1 - \mathbf{1}'\mathbf{w})^2$$

The derivative is obtained as

$$\frac{\partial Q(\lambda_1, \lambda_2)}{\partial \mathbf{w}} = 2\lambda_1 \mathbf{w} + 2\lambda_2 (1 - \mathbf{1}'\mathbf{w})$$

By setting the derivative to zero and multiplying with  $\mathbf{1}$  we obtain:

$$\lambda_1 \mathbf{1}'\mathbf{w} + \lambda_2 (n - \mathbf{1}'\mathbf{w}) = 0$$

where  $\mathbf{1}'\mathbf{w} = \sum w_i$ . Solving for  $\mathbf{1}'\mathbf{w}$  we obtain

$$\mathbf{1}'\mathbf{w} = \frac{1}{1 + \lambda_1/\lambda_2}$$

and due to the symmetry of the objective function with respect to the elements of the weight vector we have

$$w_i = 1/(n + n\lambda_1/\lambda_2)$$

### 6.2. Inference

### 6.3. Derivation of the REGSC prior

The objective function of the REGSC estimator is given by

$$\begin{aligned} Q(w, \lambda_1, \lambda_2) &= \sum_{t=1}^{T_0} \underbrace{\left( y_{0,t} - \mu^* - \sum_{j=1}^J w_j y_{j,t} \right)^2}_{RSS} + \lambda_1 \underbrace{\left( \sum_{j=1}^J w_j^2 \right)}_{Ridge} + \lambda_2 \underbrace{\left( 1 - \sum_{j=1}^J w_j \right)^2}_{\text{"inverse"-Ridge}} + C \\ &= RSS + \lambda_1 \mathbf{w}'\mathbf{w} + \lambda_2 (\mathbf{w} - \iota_J)' \mathbf{1}_J \mathbf{1}'_J (\mathbf{w} - \iota_J) + C \end{aligned}$$

where  $\iota_J$  is a  $J \times 1$  vector with identical elements  $1/J$ . The additional constant  $C$  is just introduced for convenience and does not affect the solution. We are looking for a Bayesian interpretation of the likelihood of the form

$$l(w, \lambda_1, \lambda_2) = const - \frac{1}{2\sigma^2} RSS - \frac{1}{2} (\mathbf{w} - \mu_0)' V_0^{-1} (\mathbf{w} - \mu_0)$$

where the second term represents the prior distribution. From the squared  $w_i$ , we find the covariance matrix:

$$\begin{aligned} \frac{1}{\sigma^2} V_0^{-1} &= \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}'_J \\ &= \lambda_1 (I_J - P) + (\lambda_1 + J\lambda_2) P \quad \text{with } P = \frac{2}{J} \mathbf{1}_J \mathbf{1}'_J \\ &= \frac{1}{\lambda_1} (I_J - P) + \frac{1}{\lambda_1 + J\lambda_2} P \\ &= \frac{1}{\lambda_1} I_J + \frac{1}{J(\lambda_1 + J\lambda_2)} \mathbf{1}_J \mathbf{1}'_J \end{aligned}$$

In a similar manner we obtain

$$-\frac{2}{\sigma^2} w' V_0^{-1} \mu_0 = -2\lambda_2 w' \mathbf{1}_J \mathbf{1}'_J \iota_J$$

and, therefore,

$$\begin{aligned} \frac{1}{\sigma^2} V_0^{-1} \mu_0 &= \lambda_2 \mathbf{1}_J \mathbf{1}'_J \iota_J \\ \mu_0 &= \lambda_2 \sigma^2 V_0 \mathbf{1}_J \mathbf{1}'_J \iota_J \\ &= \lambda_2 / (\lambda_1 + J\lambda_2) \end{aligned}$$

This gives the parameter for the prior distribution. Note that the constant  $C$  ensures that also the constant term will match.

#### 6.4. Bayesian Estimation of the REGSC weights

To obtain valid posterior samples for each of the 10 explanatory variables and the error variance, we start by taking the logarithm of the posterior distribution. Afterwards, the posterior is more tangible as it is comprised of the sum of the likelihood and our priors for the coefficients and the error variance. For numerical stability and to expand the support to  $\mathbb{R}$ , we applied a logarithmic transformation to the error variance. For iid-errors, the likelihood simplifies to the sum of  $T_0$  normal densities. The multivariate (logarithmized) normal prior for the coefficients is of dimension  $J$  and cannot be simplified to a  $J$ -dimensional sum of normal distributions as the covariance structure of the prior matters. For the error term, we assumed a weakly informative inverse gamma prior distribution with hyperparameters  $a = b = 0.001$ .

Next, we employed a metropolis-hastings algorithm with random walk updates to obtain samples from the posterior distribution. In each step of the iteration, this involves proposing a new parameter vector based on a normal distribution around

the previously accepted proposal. The variance of this normal is an hyperparameter that can have a large impact on the speed of convergence of the algorithm. Each proposal is assessed according to the posterior value it causes: In case the proposal provides a higher log-posterior value, it is accepted with certainty to ensure the algorithm converges to the true REGSC coefficients. In case the proposal provides a lower log-posterior value, is is accepted with an acceptance probability that is based on the difference between the log-posterior value of the proposal and the previously accepted proposal. This procedure ensures that also areas of lower density are represented accordingly. Bayesian inference is only valid when all Markov Chains have converged simultaneously and when the autocorrelation structure within in obtained Markov Chains is removed. Thus, we simulated a total of 150,000 samples per marginal posterior, burned the first 50,000 and thinned the chains by keeping only every 10th chain element. The following figures shows the converged Markov Chains (black) with the closed form estimates (red) and the 2.5% and the 97.5% percentiles (dashed blue):

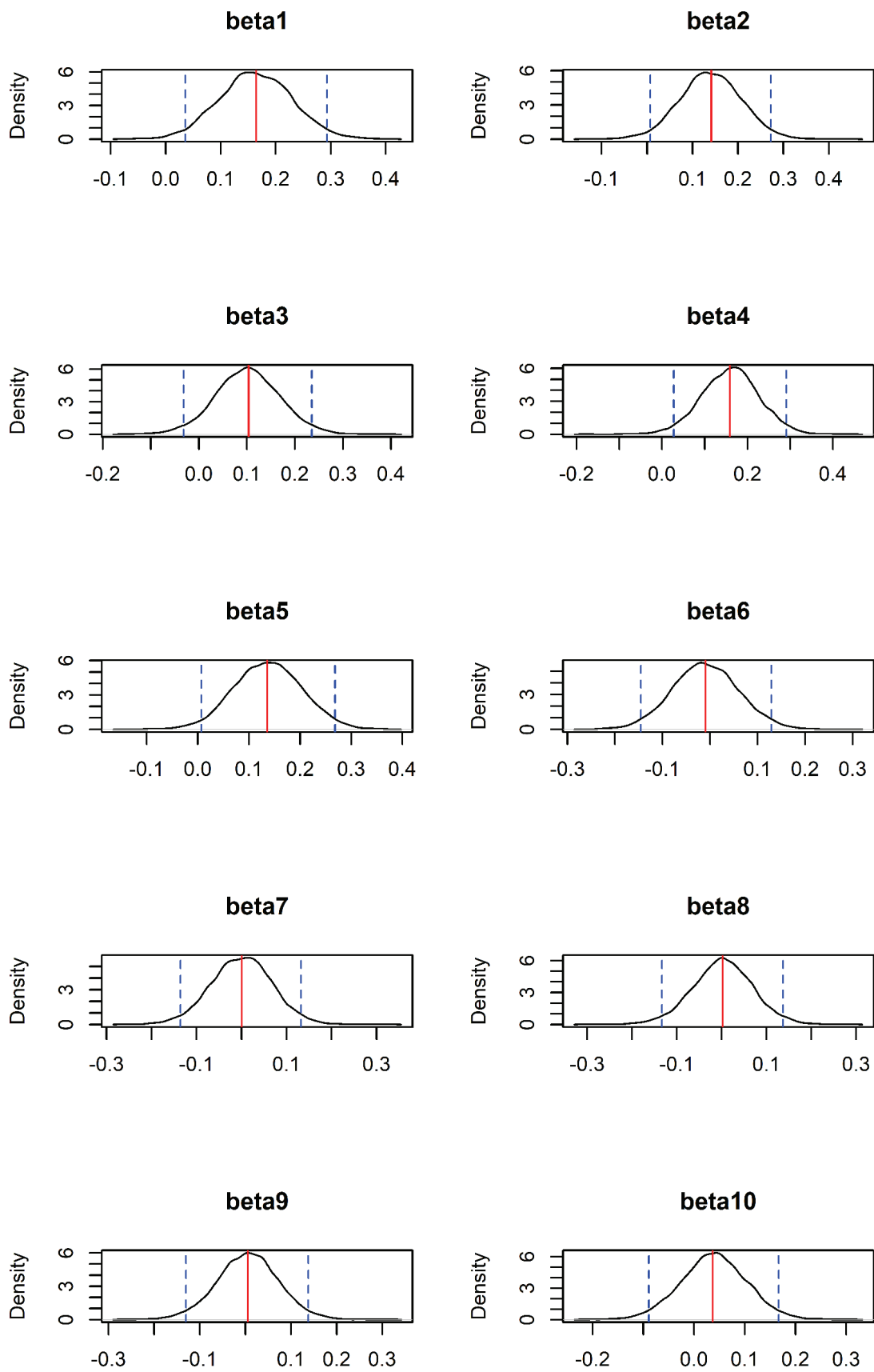


Figure 11. Marginal posterior distributions

### 6.5. Derivation of the analytical REGSC distribution

It has already been shown that the REGSC estimator has the following closed form solution:

$$\hat{w}_{\lambda_1, \lambda_2} = \left( \tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J' \right)^{-1} \left( \tilde{Z}'\tilde{y}_0 + \lambda_2 \mathbf{1}_J \right).$$

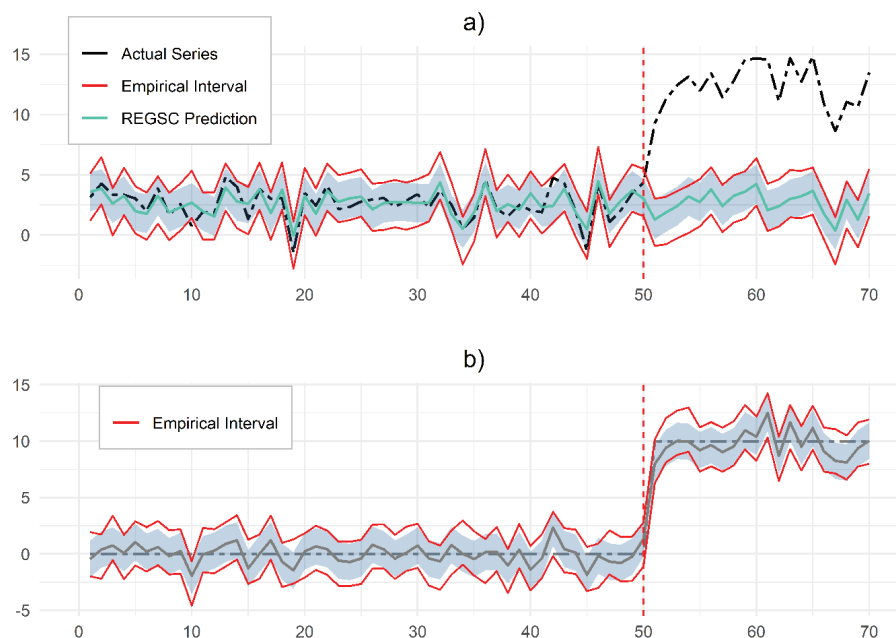
Similar to the variance of the OLS estimator, the variance of the REGSC estimator conditional on the explanatory variables is derived as follows:

$$\begin{aligned} V(\hat{w}_{\lambda_1, \lambda_2}) &= V \left( (\tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1} (\tilde{Z}'\tilde{y}_0 + \lambda_2 \mathbf{1}_J) | \tilde{Z} \right) \\ &= (\tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1} (\tilde{Z}' V(\tilde{y}_0 | \tilde{Z}) (\tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1} \tilde{Z}')' \\ &= (\tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1} \tilde{Z}' V(\epsilon | \tilde{Z}) \left( (\tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1} \tilde{Z}' \right)' \\ &= \sigma^2 (\tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1} \tilde{Z}' \left( (\tilde{Z}'\tilde{Z} + \lambda_1 I_J + \lambda_2 \mathbf{1}_J \mathbf{1}_J')^{-1} \tilde{Z}' \right)' \end{aligned}$$

For a normally distributed error term, it follows that the REGSC estimator is normally distributed with  $\hat{w}_{\lambda_1, \lambda_2} \sim \mathcal{N}(\hat{w}_{\lambda_1, \lambda_2}, V(\hat{w}_{\lambda_1, \lambda_2}))$ .

### 6.6. Bootstrap Interval

Analogous to figure 1, the following figure demonstrates the functionality of the Bootstrap intervals. We see that the Bootstrap interval is overly optimistic, most of the time it is narrower than the true empirical interval.



**Figure 12.** Bootstrap Intervals

Panel a) plots the actual series (black), the true interval (red) and the REGSC counterfactual (green). At  $T = 51$ , the series experiences a persisting treatment effect of size 10. We see that the Bootstrap interval has decent coverage: At only two time points (10, 45), the actual series lies outside the interval. Panel b) compares the actual treatment effect (black) to the REGSC estimation (green), obtained by subtracting the REGSC counterfactual from the actual series. In the pre-treatment period, the actual treatment effect of zero is within the interval in 48 out of 50 cases. In the post-treatment period, the actual treatment effect falls inside the interval in 1 of 20 cases.



## 6.7. Additional results

## 6.7.1. The static case

**Table 6.** Simulation Results of the Static Factor Model with  $\mathbf{J} = \mathbf{5}$  Donors.

| $T_{pre}$ | $T_{post}$ | FACTOR                                      | SC  | REGSC                                       | NET   | OLS  |
|-----------|------------|---|---|---|---|--|
|           |            | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)            |
| 20        | 10         | 1.2348<br>{-0.0103}<br>[0.8940]<br>(0.5889) | 1.4482<br>{-0.0347}<br>[0.7000]<br>(1.0399) | 1.2968<br>{-0.0104}<br>[0.8540]<br>(0.6222) | 1.2926<br>{-0.0037}<br>[0.8766]<br>(0.5745) | 1.3520<br>{0.0143}<br>[0.8240]<br>(1.0234) |
| 50        | 10         | 1.1524<br>{0.0110}<br>[0.9240]<br>(0.6342)  | 1.4264<br>{-0.0285}<br>[0.7160]<br>(1.0640) | 1.1819<br>{0.0113}<br>[0.9060]<br>(0.6435)  | 1.1711<br>{0.0187}<br>[0.9100]<br>(0.5778)  | 1.1862<br>{0.0195}<br>[0.9040]<br>(0.7819) |
| 100       | 10         | 1.1409<br>{0.0200}<br>[0.9240]<br>(0.6091)  | 1.3606<br>{0.0584}<br>[0.7640]<br>(0.9868)  | 1.1529<br>{0.0229}<br>[0.9220]<br>(0.5778)  | 1.1504<br>{0.0216}<br>[0.9220]<br>(0.5481)  | 1.1534<br>{0.0251}<br>[0.9300]<br>(0.6699) |

**Table 7.** Simulation Results of the Static Factor Model with  $\mathbf{J} = \mathbf{10}$  Donors.

| $T_{pre}$ | $T_{post}$ | <b>FACTOR</b>                               | <b>SC</b>                                   | <b>REGSC</b>                                | <b>NET</b>                                  | <b>OLS</b>                                  |
|-----------|------------|---|---|---|---|---|
|           |            | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             |
| 20        | 10         | 1.1519<br>{-0.0157}<br>[0.8880]<br>(0.6850) | 1.5824<br>{-0.0193}<br>[0.5980]<br>(0.8727) | 1.2383<br>{-0.0130}<br>[0.8400]<br>(0.7410) | 1.2445<br>{-0.0190}<br>[0.8402]<br>(0.7231) | 1.6072<br>{0.0124}<br>[0.5800]<br>(2.1206)  |
| 50        | 10         | 1.0791<br>{-0.0159}<br>[0.9500]<br>(0.7413) | 1.3589<br>{0.0260}<br>[0.7720]<br>(1.0677)  | 1.1136<br>{-0.0205}<br>[0.9520]<br>(0.7007) | 1.1239<br>{-0.0179}<br>[0.9380]<br>(0.6334) | 1.1879<br>{-0.0217}<br>[0.9140]<br>(1.0317) |
| 100       | 10         | 1.0790<br>{0.0177}<br>[0.9460]<br>(0.7462)  | 1.5657<br>{-0.0045}<br>[0.6060]<br>(0.8328) | 1.0972<br>{0.0168}<br>[0.9460]<br>(0.7172)  | 1.1007<br>{0.0195}<br>[0.9460]<br>(0.6530)  | 1.1241<br>{0.0179}<br>[0.9440]<br>(0.8755)  |

**Table 8.** Simulation Results of the Static Factor Model with  $\mathbf{J} = 15$  Donors.

| $T_{pre}$ | $T_{post}$ | FACTOR                                      | SC  | REGSC                                       | NET   | OLS   |
|-----------|------------|---|---|---|---|---|
|           |            | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             |
| 20        | 10         | 1.1287<br>{0.0080}<br>[0.8920]<br>(0.7477)  | 1.8565<br>{0.0710}<br>[0.5080]<br>(0.9210)  | 1.2200<br>{0.0058}<br>[0.8580]<br>(0.8326)  | 1.2518<br>{-0.0034}<br>[0.8574]<br>(0.8114) | 2.3528<br>{-0.0210}<br>[0.2620]<br>(5.2601) |
| 50        | 10         | 1.0684<br>{0.0023}<br>[0.9240]<br>(0.7179)  | 1.3649<br>{0.0110}<br>[0.7740]<br>(0.9402)  | 1.1105<br>{0.0077}<br>[0.9260]<br>(0.6938)  | 1.1202<br>{0.0049}<br>[0.9300]<br>(0.6181)  | 1.2556<br>{0.0158}<br>[0.8600]<br>(1.2110)  |
| 100       | 10         | 1.0486<br>{-0.0123}<br>[0.9480]<br>(0.7593) | 1.8036<br>{-0.0236}<br>[0.5440]<br>(0.8418) | 1.0770<br>{-0.0176}<br>[0.9340]<br>(0.7110) | 1.0898<br>{-0.0149}<br>[0.9280]<br>(0.6602) | 1.1387<br>{-0.0220}<br>[0.9300]<br>(0.9729) |

**Table 9.** Simulation Results of the Static Factor Model with  $\mathbf{J} = 20$  Donors.

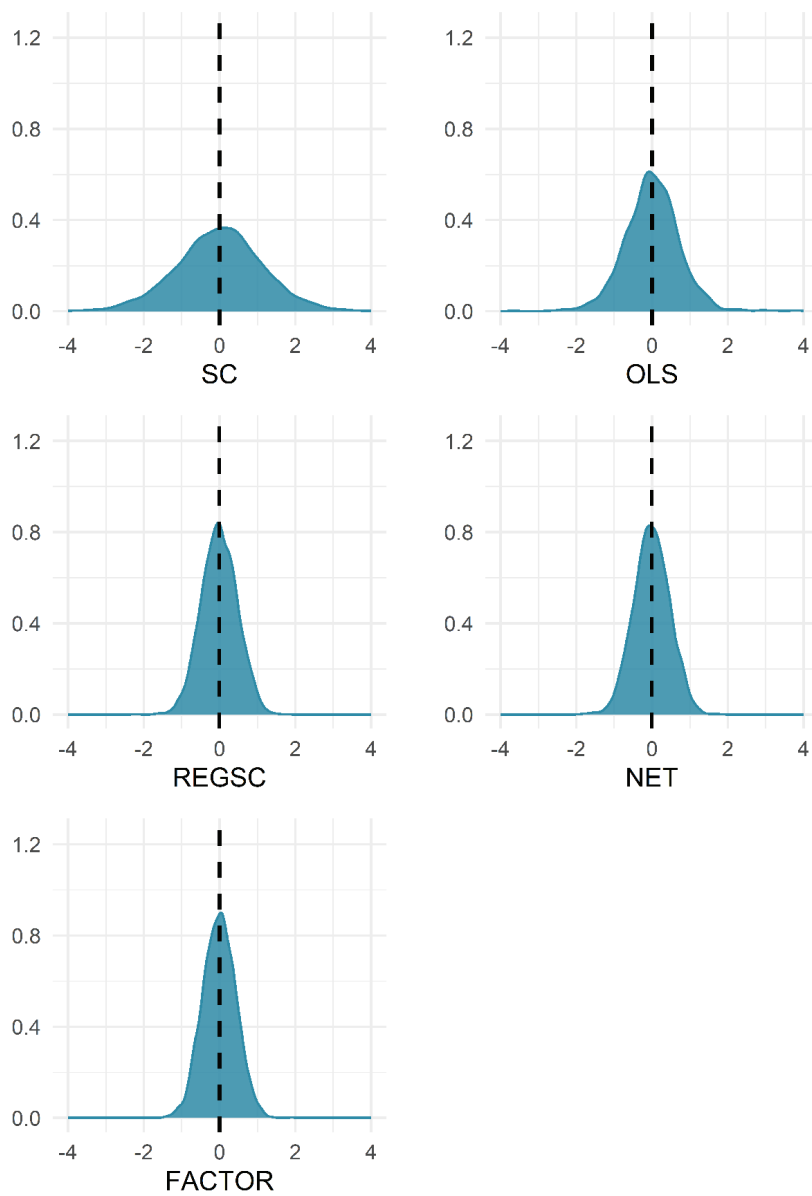
| $T_{pre}$ | $T_{post}$ | FACTOR                                      | SC  | REGSC                                       | NET   | OLS  |
|-----------|------------|---|---|---|---|--|
|           |            | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)            |
| 20        | 10         | 1.0952<br>{-0.0219}<br>[0.9120]<br>(0.7841) | 1.8022<br>{-0.0468}<br>[0.5540]<br>(0.8723) | 1.1924<br>{-0.0226}<br>[0.8760]<br>(0.8030) | 1.2214<br>{-0.0276}<br>[0.8629]<br>(0.8089) | NA<br>NA<br>NA<br>NA                       |
| 50        | 10         | 1.0444<br>{-0.0004}<br>[0.9360]<br>(0.7665) | 1.2875<br>{-0.0687}<br>[0.8100]<br>(0.8944) | 1.0871<br>{0.0034}<br>[0.9180]<br>(0.7122)  | 1.1011<br>{0.0014}<br>[0.9200]<br>(0.6516)  | 1.3345<br>{0.0044}<br>[0.7940]<br>(1.4578) |
| 100       | 10         | 1.0588<br>{0.0124}<br>[0.9600]<br>(0.8199)  | 1.7170<br>{-0.0365}<br>[0.6200]<br>(0.7517) | 1.0803<br>{0.0117}<br>[0.9460]<br>(0.7661)  | 1.0926<br>{0.0170}<br>[0.9340]<br>(0.6983)  | 1.1764<br>{0.0080}<br>[0.9100]<br>(1.1016) |

**Table 10.** Simulation Results of the Static Factor Model with  $\mathbf{J} = 25$  Donors.

| $T_{pre}$ | $T_{post}$ | FACTOR                                      | SC   | REGSC                                      | NET   | OLS   |
|-----------|------------|---|--|--|---|---|
|           |            | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)            | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)            | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             |
| 20        | 10         | 1.1064<br>{0.0018}<br>[0.8900]<br>(0.7039)  | 1.7120<br>{0.0352}<br>[0.5820]<br>(0.7882) | 1.1843<br>{0.0032}<br>[0.8600]<br>(0.7240) | 1.2123<br>{-0.0037}<br>[0.8727]<br>(0.7020) | NA<br>NA<br>NA<br>NA                        |
| 50        | 10         | 1.0206<br>{-0.0059}<br>[0.9180]<br>(0.7826) | 1.2530<br>{0.0267}<br>[0.8120]<br>(0.9777) | 1.0681<br>{0.0006}<br>[0.9100]<br>(0.7508) | 1.0859<br>{-0.0001}<br>[0.9200]<br>(0.6835) | 1.4528<br>{-0.0049}<br>[0.6880]<br>(1.9355) |
| 100       | 10         | 1.0314<br>{0.0028}<br>[0.9440]<br>(0.8144)  | 1.6539<br>{0.0766}<br>[0.6260]<br>(0.7027) | 1.0582<br>{0.0041}<br>[0.9280]<br>(0.7473) | 1.0719<br>{0.0067}<br>[0.9320]<br>(0.6881)  | 1.1890<br>{0.0082}<br>[0.8960]<br>(1.1599)  |

**Table 11.** Simulation Results of the Static Factor Model with  $\mathbf{J} = 30$  Donors.

| $T_{pre}$ | $T_{post}$ | FACTOR                                      | SC  | REGSC                                       | NET   | OLS   |
|-----------|------------|---|---|---|---|---|
|           |            | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             | RMSE<br>{BIAS}<br>[MZ]<br>(VAR)             |
| 20        | 10         | 1.0947<br>{-0.0035}<br>[0.9200]<br>(0.8268) | 1.6478<br>{0.0033}<br>[0.5840]<br>(0.8100)  | 1.1613<br>{-0.0073}<br>[0.8800]<br>(0.7555) | 1.1959<br>{-0.0053}<br>[0.8699]<br>(0.7996) | NA<br>NA<br>NA<br>NA                        |
| 50        | 10         | 1.0306<br>{0.0051}<br>[0.9320]<br>(0.8421)  | 1.2044<br>{-0.0335}<br>[0.8560]<br>(0.9469) | 1.0730<br>{0.0029}<br>[0.9200]<br>(0.7919)  | 1.0862<br>{-0.0018}<br>[0.9160]<br>(0.7065) | 1.6406<br>{-0.0247}<br>[0.5220]<br>(2.4925) |
| 100       | 10         | 1.0167<br>{0.0074}<br>[0.9560]<br>(0.8534)  | 1.5814<br>{-0.0379}<br>[0.5920]<br>(0.6674) | 1.0354<br>{0.0054}<br>[0.9480]<br>(0.7864)  | 1.0524<br>{0.0026}<br>[0.9380]<br>(0.7223)  | 1.2026<br>{0.0069}<br>[0.8740]<br>(1.3139)  |



**Figure 13.** Bias-densities for  $T_{pre} = 20$  and  $T_{post} = 10$

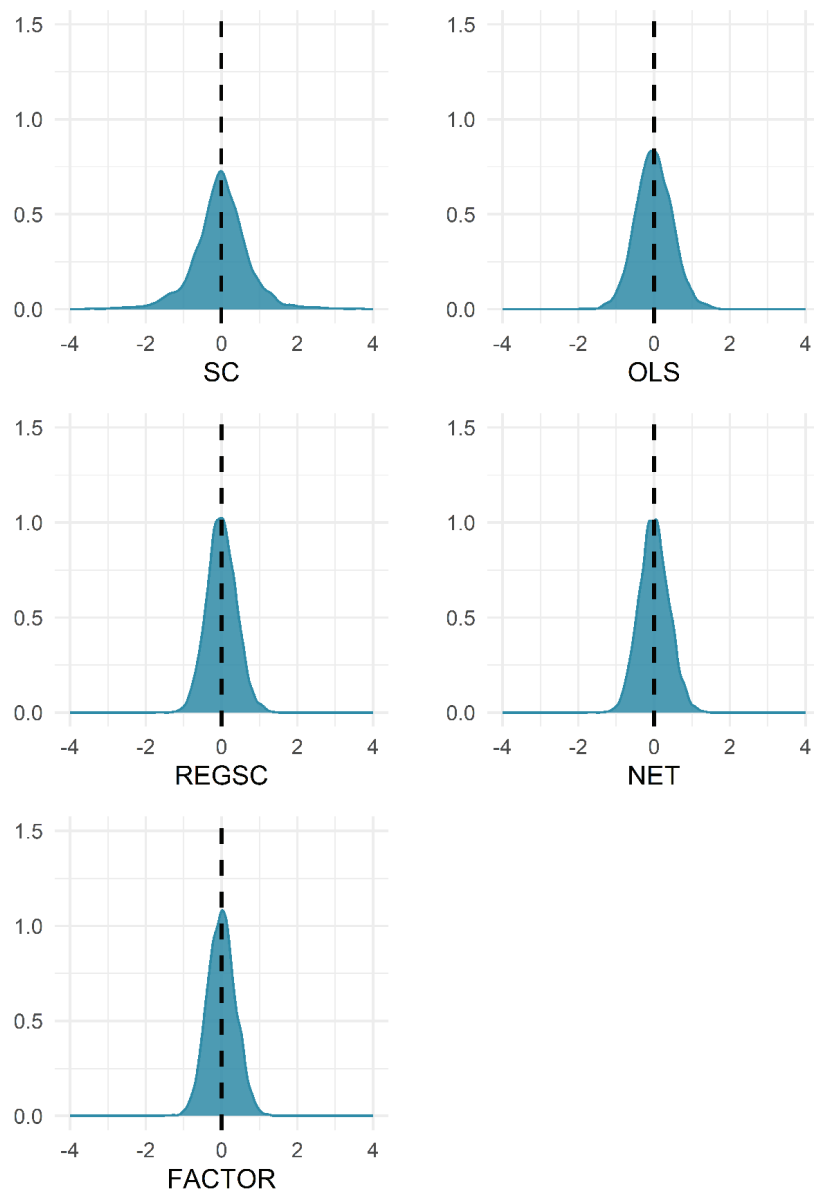
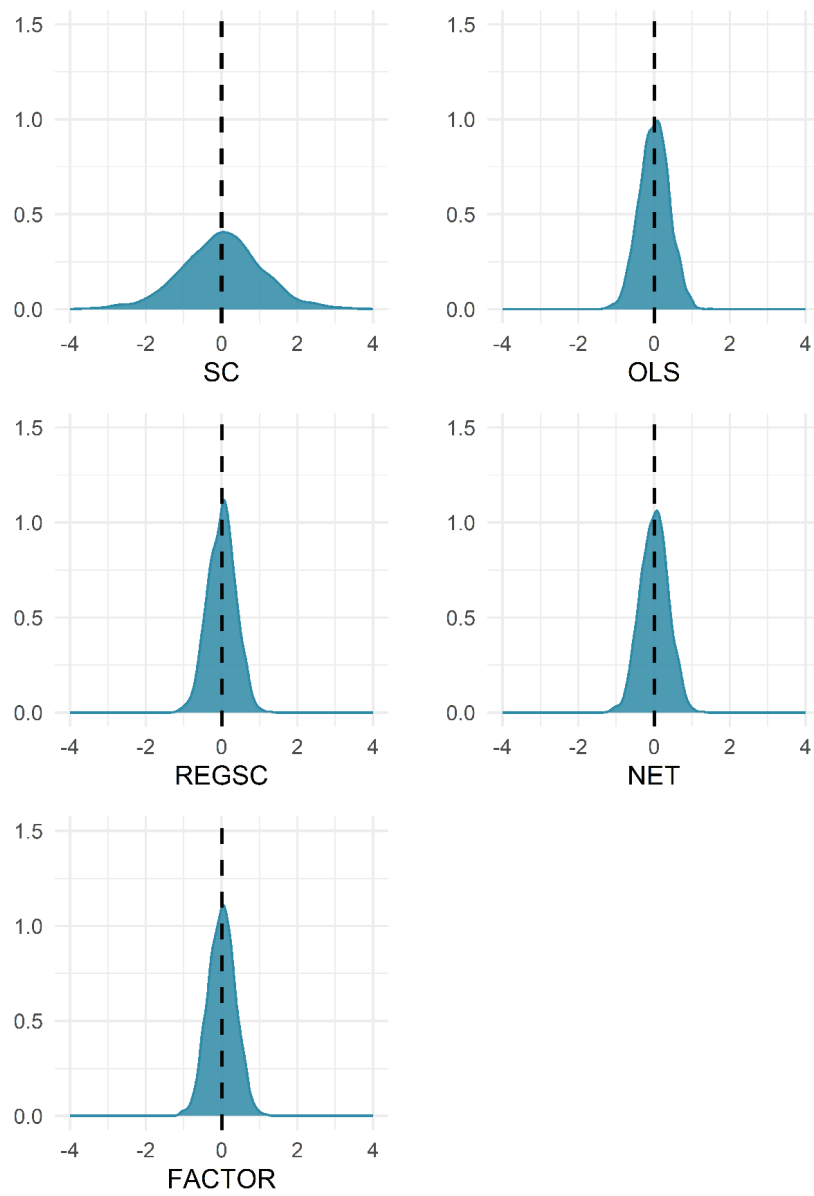


Figure 14. Bias-densities for  $T_{pre} = 50$  and  $T_{post} = 10$



**Figure 15.** Bias-densities for  $T_{pre} = 100$  and  $T_{post} = 10$

**Table 12.** MZ-Acceptance rates for  $T_{pre} \in \{20, 50, 100\}$  and  $J \in \{5, 10, 15, 20, 25, 30\}$ . 2nd best always excludes the benchmark factor model.

| $T_{pre}$ | Donors | SC     | OLS    | REGSC  | NET    | FACTOR | 2nd best* |
|-----------|--------|--------|--------|--------|--------|--------|-----------|
| 20        | 5      | 0.7000 | 0.8240 | 0.8540 | 0.8766 | 0.8940 | NET       |
| 20        | 10     | 0.5980 | 0.5800 | 0.8400 | 0.8402 | 0.8880 | NET       |
| 20        | 15     | 0.5080 | 0.2620 | 0.8580 | 0.8574 | 0.8920 | REGSC     |
| 20        | 20     | 0.5540 | NA     | 0.8760 | 0.8629 | 0.9120 | REGSC     |
| 20        | 25     | 0.5820 | NA     | 0.8600 | 0.8727 | 0.8900 | NET       |
| 20        | 30     | 0.5840 | NA     | 0.8800 | 0.8699 | 0.9200 | REGSC     |
| 50        | 5      | 0.7160 | 0.9040 | 0.9060 | 0.9100 | 0.9240 | NET       |
| 50        | 10     | 0.7720 | 0.9140 | 0.9520 | 0.9380 | 0.9500 | REGSC     |
| 50        | 15     | 0.7740 | 0.8600 | 0.9260 | 0.9300 | 0.9240 | NET       |
| 50        | 20     | 0.8100 | 0.7940 | 0.9180 | 0.9200 | 0.9360 | NET       |
| 50        | 25     | 0.8120 | 0.6880 | 0.9100 | 0.9200 | 0.9180 | NET       |
| 50        | 30     | 0.8560 | 0.5220 | 0.9200 | 0.9160 | 0.9320 | REGSC     |
| 100       | 5      | 0.7640 | 0.9300 | 0.9220 | 0.9220 | 0.9240 | SC        |
| 100       | 10     | 0.6060 | 0.9440 | 0.9460 | 0.9460 | 0.9460 | REGSC     |
| 100       | 15     | 0.5440 | 0.9300 | 0.9340 | 0.9280 | 0.9480 | REGSC     |
| 100       | 20     | 0.6200 | 0.9100 | 0.9460 | 0.9340 | 0.9600 | REGSC     |
| 100       | 25     | 0.6260 | 0.8960 | 0.9280 | 0.9320 | 0.9440 | NET       |
| 100       | 30     | 0.5920 | 0.8740 | 0.9480 | 0.9380 | 0.9560 | REGSC     |



**Table 13.** RMSE average for  $T_{pre} \in \{20, 50, 100\}$  and  $J \in \{5, 10, 15, 20, 25, 30\}$ .  
2nd best always excludes the benchmark factor model.

| $T_{pre}$ | Donors | SC     | OLS    | REGSC  | NET    | FACTOR | 2nd best* |
|-----------|--------|--------|--------|--------|--------|--------|-----------|
| 20        | 5      | 1.4482 | 1.3520 | 1.2968 | 1.2926 | 1.2348 | NET       |
| 20        | 10     | 1.5824 | 1.6072 | 1.2383 | 1.2445 | 1.1519 | REGSC     |
| 20        | 15     | 1.8565 | 2.3528 | 1.2200 | 1.2518 | 1.1287 | REGSC     |
| 20        | 20     | 1.8022 | NA     | 1.1924 | 1.2214 | 1.0952 | REGSC     |
| 20        | 25     | 1.7120 | NA     | 1.1843 | 1.2123 | 1.1064 | REGSC     |
| 20        | 30     | 1.6478 | NA     | 1.1613 | 1.1959 | 1.0947 | REGSC     |
| 50        | 5      | 1.4264 | 1.1862 | 1.1819 | 1.1711 | 1.1524 | NET       |
| 50        | 10     | 1.3589 | 1.1879 | 1.1136 | 1.1239 | 1.0791 | REGSC     |
| 50        | 15     | 1.3649 | 1.2556 | 1.1105 | 1.1202 | 1.0684 | REGSC     |
| 50        | 20     | 1.2875 | 1.3345 | 1.0871 | 1.1011 | 1.0444 | REGSC     |
| 50        | 25     | 1.2530 | 1.4528 | 1.0681 | 1.0859 | 1.0206 | REGSC     |
| 50        | 30     | 1.2044 | 1.6406 | 1.0730 | 1.0862 | 1.0306 | REGSC     |
| 100       | 5      | 1.3606 | 1.1534 | 1.1529 | 1.1504 | 1.1409 | NET       |
| 100       | 10     | 1.5657 | 1.1241 | 1.0972 | 1.1007 | 1.0790 | REGSC     |
| 100       | 15     | 1.8036 | 1.1387 | 1.0770 | 1.0898 | 1.0486 | REGSC     |
| 100       | 20     | 1.7170 | 1.1764 | 1.0803 | 1.0926 | 1.0588 | REGSC     |
| 100       | 25     | 1.6539 | 1.1890 | 1.0582 | 1.0719 | 1.0314 | REGSC     |
| 100       | 30     | 1.5814 | 1.2026 | 1.0354 | 1.0524 | 1.0167 | REGSC     |