Taylor & Francis
Taylor & Francis Group

Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Counterfactual Analysis and Inference With Nonstationary Data

Ricardo Masini & Marcelo C. Medeiros

To cite this article: Ricardo Masini & Marcelo C. Medeiros (2022) Counterfactual Analysis and Inference With Nonstationary Data, Journal of Business & Economic Statistics, 40:1, 227-239, DOI: 10.1080/07350015.2020.1799814

To link to this article: https://doi.org/10.1080/07350015.2020.1799814

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2022, VOL. 40, NO. 1, 227–239
https://doi.org/10.1080/07350015.2020.1799814

Counterfactual Analysis and Inference With Nonstationary Data

Ricardo Masini[a] and Marcelo C. Medeiros[b]

[a]Sao Paulo School of Economics, Getulio Vargas Foundation, São Paulo, Brazil; [b]Department of Economics, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil

ABSTRACT

Recently, there has been growing interest in developing econometric tools to conduct counterfactual analysis with aggregate data when a single "treated" unit suffers an intervention, such as a policy change, and there is no obvious control group. Usually, the proposed methods are based on the construction of an artificial/synthetic counterfactual from a pool of "untreated" peers, organized in a panel data structure. In this article, we investigate the consequences of applying such methodologies when the data comprise integrated processes of order 1, $I(1)$, or are trend-stationary. We find that for $I(1)$ processes *without* a cointegrating relationship (spurious case) the estimator of the effects of the intervention diverges, regardless of its existence. Although spurious regression is a well-known concept in time-series econometrics, they have been ignored in most of the literature on counterfactual estimation based on artificial/synthetic controls. For the case when at least one cointegration relationship exists, we have consistent estimators for the intervention effect *albeit* with a nonstandard distribution. Finally, we discuss a test based on resampling which can be applied when there is at least one cointegration relationship or when the data are trend-stationary.

ARTICLE HISTORY
Received September 2019
Accepted July 2020

KEYWORDS
ArCo; Cointegration; Counterfactual analysis; Nonstationarity; Policy evaluation; Synthetic control

1. Introduction

The goal of this article is to investigate the consequences of applying popular econometric methods for counterfactual analysis when the data generating mechanism possess stochastic and/or deterministic trends. The framework considered here nests the panel factor (PF) method by Hsiao, Ching, and Wan (2012) and further generalized by Ouyang and Peng (2015), Li and Bell (2017), and Chernozhukov, Wuthrich, and Zhu (2018); the artificial counterfactual (ArCo) of Carvalho, Masini, and Medeiros (2018); and extensions of the synthetic control (SC) proposed by Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010), as discussed in Doudchenko and Imbens (2016) and Ferman and Pinto (2016). Most of the articles on counterfactual analysis for panel data do not take into account the possibility of nonstationarity. In the SC setting, the estimation is traditionally viewed as a cross-section problem and the time-series nature of the data is often ignored.[1]

On the other hand, the PF and the ArCo methods explicitly explore the time dimension. However, Li (2017) and Carvalho, Masini, and Medeiros (2018) assumed that the data are (trend-)stationary while Hsiao, Ching, and Wan (2012) conjectured that if the data are cointegrated, their results will still hold. As we show in this article, this conjecture turns out to be imprecise. Furthermore, inference based on nonstationary data can be terribly misleading and although spurious regression is a well-known concept in time-series econometrics, they have been ignored in most of the literature on counterfactual estimation based on artificial/synthetic controls. For example, Hsiao, Ching, and Wan (2012), Carvalho, Masini, and Medeiros (2018), and Li (2017) considered testing the null hypothesis of no average effect as the time dimension grows.[2] Our results demonstrate that nonstationarity induces strong over-rejection of this null when the critical values are computed using a standard Gaussian distribution.

Recently, there has been growing interest in developing tools to conduct counterfactual analysis with aggregate data when a single unit suffers an intervention, such as a policy change, and there is no clear control group available. An unit can be, for example, a country, a region or state, a municipality, or a firm. In these situations, the solution is to construct an ArCo from a pool of "untreated" peers ("donors pool"). In Hsiao, Ching, and Wan (2012), the counterfactual is constructed from a linear combination of observed variables from peers given by the conditional expectation model. In the SC framework, the counterfactual is built as a convex combination of peers where the weights of the combination are estimated from time-series averages of several variables from the donor pool. Although the methods may seem similar, they differ in the way the linear combination is constructed. In the SC method, the weights are positive and sum to one. On the other hand, Hsiao, Ching, and Wan (2012) did not

[1]More recently, some authors have advocated the use of SC methods without taking time averages of the variables (see, e.g., Doudchenko and Imbens 2016; Ferman and Pinto 2016; Li 2017).

[2]The average is taken over time and not over the cross-section.

CONTACT Marcelo C. Medeiros mcm@econ.puc-rio.br Department of Economics, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro 22451-900, Brazil.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

impose any restrictions. The SC method is now a key ingredient in the toolbox of applied researchers; see Athey and Imbens (2017) for a recent review or Gardeazabal and Vega-Bayo (2017) for an empirical comparison between the PF and SC methods.

1.1. Main Takeaways

This article is a major extension of Carvalho, Masini, and Medeiros (2017). We consider two very distinct scenarios when the data are nonstationary. The first one is the cointegrated/trend-stationary case, in which there is at least one cointegrating relationship among the treated unit and the untreated peers or the data are trend-stationary. This is the most common case in empirical applications and nests models with nonstationary common factors. The second one is the spurious case, when no cointegrating relationship exists and unit-roots are present. The later may sound empirically irrelevant, as in the majority of applications the researcher assumes an underlying factor model which implies cointegration/common trend. However, this is an assumption that must be tested and neglecting potential spurious regressions can yield severely erroneous results. On the other hand, although the cointegration case is the most common setup, we show that, contrary to the common sense, that the estimator for the average intervention effect is not asymptotically normal even in the canonical case where the errors are uncorrelated and normally distributed. Our results are derived under the case where the number of observations grows ($T \longrightarrow \infty$) and the cross-section dimension, n, is kept fixed.

In the cointegration case, when both the pre- and post-intervention samples grow at a proportional rate, we show that our estimator for the average treatment effect is consistent, but not asymptotically normal. The distribution of the test for the null of no average effect is nonstandard even when the distribution of the estimator for the cointegrating vector is mixed normal. This is a new result, which leads to strong over-rejection of the null hypothesis of no average intervention effect when the nonstationary nature of the data is ignored and critical values of test are obtained from a standard Gaussian distribution (Hsiao, Ching, and Wan 2012; Carvalho, Masini, and Medeiros 2018). On the other hand, if the pre-intervention sample diverges to infinity faster than the post-intervention one, than normality can be achieved. The spurious case is more troublesome. We show that the treatment effect estimator diverges. The lack of a cointegrating relationship makes the construction of the artificial control using the pre-intervention period invalid, due to harmful effects of spurious regressions, as extensively discussed in Phillips (1986a). Therefore, we recommend that cointegration should be tested before applying such methods with nonstationary data.

Finally, we discuss two resampling procedures to conduct inference. The first one is an application of the results in Masini and Medeiros (2019) and is developed under the framework where the number of pretreatment observations diverges but the dimension post-treatment sample is fixed. This test can be applied under either cointegration or pure deterministic trends. Furthermore, the testing procedure can be applied even when there is a single observation after the intervention as well as when the data are stationary. A second procedure is also proposed for the case when the full sample diverges to infinity. However, the test will be valid only under stationarity or under the case of trend-stationary variables. It is important to highlight that the deterministic trends considered in this article do not need to be necessarily linear.

Our results highlight that, contrary to what is commonly advocated in the SC literature, it is important to test for cointegration in case of nonstationary data.[3] In general, the counterfactual estimation based on SC methods requires an almost perfect in-sample fit. However, the good pre-intervention fit is an illusion in the case of spurious regressions. It is well known in the time-series literature that the R-squared of spurious regressions tends to one and the t-statistics associated to the regression coefficients diverge as the sample size increases; see Phillips (1986a). The t-statistic for the null of no effect also diverges. In the case when no cointegration relation exists, one possible solution is to consider the data in first-differences.

A detailed simulation study corroborates our theoretical findings and evaluates the asymptotic approximation in finite samples. We also study the effects of imposing restrictions on the linear combination of peers as in the original SC method as well as the use of LASSO estimators as in Carvalho, Masini, and Medeiros (2018), Doudchenko and Imbens (2016), and Li and Bell (2017). As expected, none of these approaches mitigate the harmful effects of nonstationarity. We also conduct a simulation study to evaluate the effects of pretesting for cointegration when applying our resampling inferential procedure. We consider the case when there is only one observation after the intervention and the data can be cointegrated or not (spurious case). As before we compare the unrestricted estimator advocated in this article with restricted versions either imposing the original SC restrictions or the LASSO ones.

1.2. Comparison With the Literature

As far as we know this is the first article to give a full treatment of counterfactual methods when the data are nonstationary for the case where the number of cross-sections is kept fixed. Under some assumptions, Bai, Li, and Ouyang (2014) showed consistency of the panel approach when the data are integrated of first order but the asymptotic distribution was not derived. Ferman and Pinto (2016) studied the bias of the SC method in the case of common nonstationary factors. Li (2018) derived the asymptotic distribution of the average treatment effects under a nonstationary factor model with quite restrictive assumptions. Carvalho, Masini, and Medeiros (2017) considered a more restrictive setting than ours and no inferential procedure is available. Differently from these articles, we provide the asymptotic distribution of the estimator in a much more general setup and develop the results required for inference. Masini and Medeiros (2019) considered the case of nonstationarity in high dimensions and provided complementary results.

[3] See, for instance, Remark 4.4 in Li (2018).

1.3. Organization of the Article

The rest of the article is organized as follows. Section 2 presents the setup considered while Section 3 delivers the theoretical results except for the asymptotic inference procedure which is presented in Section 4. The simulation experiment is shown in Section 5. An empirical example in shown in Section 6. Section 7 concludes the article. Finally, proofs are presented in the online Appendix. Additional technical results are relegated to the supplementary materials.

2. Setup and Estimators

2.1. Basic Setup

Suppose we have n units (countries, states, municipalities, firms, etc.) indexed by $i = 1, \dots, n$. For each unit and for every time period $t = 1, \dots, T$, we observe a realization of a variable y_{it}. We consider a scalar variable just for the sake of simplicity and the results in the article can be easily extended to the multivariate case. Furthermore, we assume that an intervention took place in unit $i = 1$, and only in unit 1, at time T_0+1, where $T_0 = \lfloor \lambda_0 T \rfloor$ and $\lambda_0 \in (0, 1)$.

Let $\mathcal{D}_t$ be a binary variable flagging the periods after the intervention, such that:

$$y_{1t} = \mathcal{D}_t y_{1t}^{(1)} + (1 - \mathcal{D}_t) y_{1t}^{(0)},$$

where $y_{1t}^{(1)}$ denotes the outcome when the unit 1 is exposed to the intervention and $y_{1t}^{(0)}$ is the potential outcome of unit 1 when it is not exposed to the intervention.

We are concerned in testing hypothesis on the potential effects of the intervention in the unit of interest (unit 1) for the post-intervention period. We consider interventions of the form

$$y_{1t}^{(1)} = \delta_t + y_{1t}^{(0)}, \quad t = T_0+1, \dots, T, \tag{2.1}$$

where $\{\delta_t\}_{t=T_0+1}^{T}$ is a deterministic sequence.[4]

The null hypothesis becomes

$$\mathcal{H}_0 : \delta_t = 0, \quad \forall t. \tag{2.2}$$

The null hypothesis of zero average (across post intervention period) treatment effect is

$$\bar{\mathcal{H}}_0 : \Delta_T = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \delta_t = 0. \tag{2.3}$$

Clearly $\mathcal{H}_0$ implies $\bar{\mathcal{H}}_0$. The weaker null is of interest since we have a consistent estimator for Δ_T in contrast to only an asymptotic unbiased estimator for δ_t. It is clear that $y_{1t}^{(0)}$ is not observed from $t = T_0+1$ onward. For this reason, we henceforth call it the *counterfactual*—that is, what would y_{1t} have been like had there been no intervention (potential outcome).

Let $\boldsymbol{y}_{0t} = (y_{2t}, \dots, y_{nt})' := \left(y_{2t}^{(0)}, \dots, y_{nt}^{(0)}\right)'$ be the collection of all untreated variables.[5] Panel based methods, such as the PF and ArCo methodologies, as well as SC extensions, construct the counterfactual by the following model in the absence of an intervention:

$$\begin{aligned} y_{1t}^{(0)} &= \mathcal{M}(\boldsymbol{y}_{0t}; \boldsymbol{\beta}_0) + \nu_t, \quad t = 1, \dots, T, \\ &= \boldsymbol{\beta}_{01}' \boldsymbol{y}_{0t} + \beta_{00} + \nu_t = \boldsymbol{\beta}_0' \boldsymbol{x}_t + \nu_t, \end{aligned} \tag{2.4}$$

where ν_t is an error term , $\boldsymbol{x}_t := \left(\boldsymbol{y}_{0t}', 1\right)'$ and $\boldsymbol{\beta}_0 \in \mathbb{R}^n$. A linear specification (including a constant) for the model $\mathcal{M}(\boldsymbol{y}_{0t})$ is by the most common choice.

The main idea is to estimate (2.4) using just the pre-intervention sample ($t = 1, \dots, T_0$), since in this case $y_{1t}^{(0)} = y_{1t}$. Consequently, the estimated counterfactual is given by

$$\widehat{y}_{1t}^{(0)} = \begin{cases} y_{1t} & \text{if } t = 1, \dots, T_0, \\ \widehat{\boldsymbol{\beta}}_{01}' \boldsymbol{y}_{0t} + \widehat{\beta}_{00} & \text{if } t = T_0+1, \dots, T. \end{cases} \tag{2.5}$$

Under stationarity in the absence of intervention and extra mild assumptions, Hsiao, Ching, and Wan (2012) and Carvalho, Masini, and Medeiros (2018) showed that $\widehat{\delta}_t := y_{1t} - \widehat{y}_{1t}^{(0)}$ is an unbiased estimator for $\delta_t, t = T_0+1, \dots, T$ $T_0 \longrightarrow \infty$ and

$$\widehat{\Delta} = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \widehat{\delta}_t \tag{2.6}$$

is $\sqrt{T}$-consistent for Δ_T and asymptotically normal. However, with stochastic trends, the population parameter $\boldsymbol{\beta}_0$ can no longer be identified as the linear projection parameters of y_{1t} onto $\boldsymbol{y}_{0t}$ and a constant due to the nonstationary of the regressors. Moreover, in the spurious regression case there is no $\boldsymbol{\beta}_0$ such that ν_t will be stationary.

We consider the estimation of (2.4) by ordinary least squares (OLS) using only the pre-intervention sample ($t \leq T_0$) which we denote by $\widehat{\boldsymbol{\beta}}$. The fitted values in the post-intervention period $\widehat{y}_{1t}^{(0)} = \widehat{\boldsymbol{\beta}}' \boldsymbol{x}_t, t > T_0$ will be our estimated counterfactual and the intervention effect will be estimated by $\widehat{\delta}_t := y_{1t} - \widehat{y}_{1t}^{(0)}$ for each period after the intervention.

2.2. Nonstationarity

We model the units in the absence of the intervention as a nonstationary (vector) process $\{\boldsymbol{y}_t^{(0)}\}_{t \geq 1}$ defined on some probability space $(\Omega, \mathcal{F}, \P)$.

Assumption 1. The data generating process (DGP) is indexed by $r \in \{0, 1, \dots, n\}$, where n is also the number of units such that:

(a) For $r = 0$:

$$\boldsymbol{y}_t^{(0)} = \boldsymbol{y}_{t-1}^{(0)} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \quad t \geq 1, \tag{2.7}$$

where $\boldsymbol{y}_0^{(0)} = O_P(1)$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and no cointegration relation exists with unit 1 included.

(b) For $0 < r < n$:

$$\boldsymbol{y}_t^{(0)} = \boldsymbol{y}_{t-1}^{(0)} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \quad t \geq 1, \tag{2.8}$$

where $\boldsymbol{y}_0^{(0)} = O_P(1)$, $\boldsymbol{\mu} \in \mathbb{R}^n$ and there exist $r \geq 1$ cointegration relations with unit 1 included. For $1 \leq r \leq n-1$, there is a $(n \times r)$ matrix $\widetilde{\boldsymbol{\Gamma}}$ with rank r such that

[4] The usual setup in the literature is to consider determinist effects (see, e.g., Chernozhukov, Wuthrich, and Zhu 2018).

[5] We could also have included lags of the variables and/or exogenous regressors into $\boldsymbol{y}_{0t}$ but again to keep the argument simple, we have considered just contemporaneous variables; see Carvalho, Masini, and Medeiros (2018) for more general specifications.

$\widetilde{\boldsymbol{\Gamma}}'\boldsymbol{y}_t^{(0)}$ is integrated of order zero and at least one element of the first row of $\widetilde{\boldsymbol{\Gamma}}$ is nonzero.[6]

(c) For $r = n$:

$$\boldsymbol{y}_t^{(0)} = \boldsymbol{c} + \boldsymbol{\mu} f_t + \boldsymbol{\varepsilon}_t, \quad t \geq 1, \tag{2.9}$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $\boldsymbol{\mu} \in \mathbb{R}^n$, f_t is positive increasing deterministic sequence and $\mu_1 \neq 0$ and $\mu_j \neq 0$ for some $j \neq 1$.

For all cases, $\{\boldsymbol{\varepsilon}_t\}_{t\geq 1}$ is a zero mean covariance-stationary process.

The DGPs defined in Assumption 1 are of the form $\boldsymbol{y}_t^{(0)} = O_P(1) + \boldsymbol{\mu} d_t + \boldsymbol{\eta}_t$, for $t \geq 1$. If $r < n$, then $d_t = t$ and $\boldsymbol{\eta}_t = \sum_{s=1}^t \boldsymbol{\varepsilon}_s$. If $r = n$, then $d_t = f_t$ and $\boldsymbol{\eta}_t = \boldsymbol{\varepsilon}_t$. If we set $f_t = t$ and $\boldsymbol{\mu} \neq \boldsymbol{0}$, we have that DGPs(a) and (b) have a stochastic trend and the DGP(c) have a deterministic linear trend. By setting $\boldsymbol{\mu} = \boldsymbol{0}$ in DGP(a), we have a "pure" random walk. We exclude the case when $\boldsymbol{\mu} = \boldsymbol{0}$ in DGP(c). In that case, the variables are (weakly) stationary.[7]

2.3. Example: Nonstationary Factor Model

We illustrate a possible DGP based on a simple factor model. Suppose that the units in the absence of intervention are modeled via a single factor f_t such that for each unit $i \in \{1, \ldots, n\}$ and every $t \in \{1, \ldots, T\}$ we have

$$y_{it}^{(0)} = c_i + \mu_i \mathrm{f}_t + u_{it}^y, \tag{2.10}$$

where $c_i \in \mathbb{R}$, u_{it}^y is an idiosyncratic error and $\mu_i \in \mathbb{R}$ is the factor loadings for unit i. We impose that the factor follows either a unit root process with a (possibly nonlinear) drift

$$\mathrm{f}_t = \mu^{\mathrm{f}} + \mathrm{f}_{t-1} + u_t^{\mathrm{f}}, \quad t \geq 1 \tag{2.11}$$

for some initial condition $\mathrm{f}_0 = O_P(1)$; or a trend-stationary process

$$\mathrm{f}_t = \mu_t^{\mathrm{f}} + u_t^{\mathrm{f}}, \tag{2.12}$$

where in both cases $\{\mu_t^{\mathrm{f}}\}_{t=1}^\infty$ is a deterministic sequence, not necessarily linear.

Consider that $(u_{1t}^y, \ldots, u_{nt}^y, u_t^{\mathrm{f}})$ is a zero-mean, independent and identically distributed Gaussian random vector. The factor model yields a common trend for those units with nonzero loadings, $\mu_i \neq 0$, and a correlation among the stochastic components of $\boldsymbol{y}_t^{(0)}$ due to u_t^{f}.

Let L be the lag operator and write model (2.11) as

$$\mathrm{f}_t = (1-L)^{-1}\mu^{\mathrm{f}} + (1-L)^{-1}u_t^{\mathrm{f}}. \tag{2.13}$$

Replacing f_t by (2.13) in (2.10), $y_{it}^{(0)}$ is a random walk as in (2.8). On the other hand, model (2.9) can be easily derived by setting $f_t = \mu_t^{\mathrm{f}}$ and $\varepsilon_{it} = \mu_i u_t^{\mathrm{f}} + u_{it}^y$ in (2.9).

We define a pseudo-true model as $y_t = \boldsymbol{\beta}_0'\boldsymbol{x}_t + \nu_t$, where $y_t := y_{1t}^{(0)}$ and $\boldsymbol{x}_t := \left[1, \boldsymbol{y}_{0t}^{(0)\prime}\right]'$. Suppose there are $1 < r+1 \leq n$ units with nonzero loadings ($\mu_i \neq 0$), including unit 1. Without loss of generality, make those the first $r+1$ units. In that case, we have r independent linear relations among those units resulting in a stationary process since we can cancel the trends by setting $\widetilde{\boldsymbol{\Gamma}}'\boldsymbol{y}_t^{(0)}$, where

$$\widetilde{\boldsymbol{\Gamma}}' = \begin{pmatrix} 1 & -\frac{\mu_1}{\mu_2} & 0 & 0 & \\ \vdots & 0 & \ddots & 0 & \boldsymbol{0}_{r\times(n-r-1)} \\ 1 & 0 & 0 & -\frac{\mu_1}{\mu_{r+1}} & \end{pmatrix},$$

and $\boldsymbol{0}_{r\times(n-r-1)}$ is a $r \times (n-r-1)$ matrix of zero elements. After normalizing to obtain the representation $\widetilde{\boldsymbol{\Gamma}}' = \left(I_r : -\boldsymbol{\Gamma}'\right)$, we are left with

$$\boldsymbol{\Gamma}' = \begin{pmatrix} \widetilde{\mu}_1 & \\ \vdots & \boldsymbol{0}_{r\times(n-r-1)} \\ \widetilde{\mu}_r & \end{pmatrix},$$

where $\widetilde{\mu}_i := \frac{\mu_i}{\mu_{r+1}}$ for $i \in \{1, \ldots, r\}$. Then, $\boldsymbol{J}_t = \widetilde{\boldsymbol{\Gamma}}'\boldsymbol{y}_t^{(0)}$ is stationary with a typical element $J_{i,t} = c_i - \widetilde{\mu}_i c_{r+1} + u_{it}^y - \widetilde{\mu}_i u_{r+1,t}^y = \widetilde{c}_i + \widetilde{u}_{it}$, where $\widetilde{c}_i := c_i - \widetilde{\mu}_i c_{r+1}$ and $\widetilde{u}_{it} := u_{it}^y - \widetilde{\mu}_i u_{r+1,t}^y$.

When $r = 1$, the pseudo-true vector of parameters becomes $\boldsymbol{\beta}_0 = \left(c_1 - \frac{\mu_1}{\mu_2}c_2, \frac{\mu_1}{\mu_2}, 0, \ldots, 0\right)'$, and the covariance structure of the vector $\left(u_t^{\mathrm{f}}, u_{1t}^y, \ldots, u_{nt}^y\right)'$ plays no role in determining the coefficients of the pseudo-true model, since there is only one possible linear combination that results in a $I(0)$ process. On the other hand, when $r \geq 2$, we have $\boldsymbol{\beta}_0 = (\widetilde{c}_1 - \boldsymbol{\zeta}'\widetilde{\boldsymbol{c}}_0, \boldsymbol{\zeta}', \widetilde{\mu}_1 - \boldsymbol{\zeta}'\widetilde{\boldsymbol{\mu}}_0, 0, \ldots, 0)'$, where $\widetilde{\boldsymbol{c}}_0 := (\widetilde{c}_2, \ldots, \widetilde{c}_r)'$, $\widetilde{\boldsymbol{\mu}}_0 := (\widetilde{\mu}_2, \ldots, \widetilde{\mu}_r)'$, and $\boldsymbol{\zeta}$ denote the linear projection of $\widetilde{u}_{1t}$ onto $(\widetilde{u}_{2t}, \ldots, \widetilde{u}_{rt})'$. Now it becomes evident that the covariance structure of $\left(u_{1t}^y, \ldots, u_{r+1,t}^y\right)'$ affects the coefficients of the pseudo-true model through $\boldsymbol{\zeta}$. Finally, the error term for the linear regression model is given by $\nu_t = u_{1t}^y - \sum_{i=2}^{r+1} \beta_{0,i} u_{it}^y$.

If $\mu_i = 0$ and $u_{it}^y = u_{it-1}^y + \nu_{it}$, $i = 1, \ldots, n$, in model (2.10), we have the spurious case.

3. Theoretical Results

3.1. Notation

To facilitate the exposure of the theoretical results we establish the following partition scheme. For a (random) matrix $\boldsymbol{M} := (m_{ij})_{1\leq i\leq n, 1\leq j\leq m}$ when we want to make the dimension explicit we write $\underset{(n\times m)}{\boldsymbol{M}}$. We also define the submatrix $\underset{a:b\times c:d}{[\boldsymbol{M}]}$ by

$$\underset{a:b\times c:d}{[\boldsymbol{M}]} := \begin{pmatrix} m_{a,c} & m_{a,c+1} & \cdots & m_{a,d} \\ m_{a+1,c} & m_{a+1,c+1} & \cdots & m_{a+1,d} \\ \vdots & \vdots & \ddots & \vdots \\ m_{b,c} & m_{a+1,c+1} & \cdots & m_{b,d} \end{pmatrix}$$

for integers $1 \leq a \leq b \leq n$ and $1 \leq c \leq d \leq m$. Similarly for a n-dimensional (random) vector $\boldsymbol{v} := (v_1, \ldots, v_n)'$, we write $\underset{(n)}{\boldsymbol{v}}$

[6] Notice that for the case when there is at least one cointegration relation, if the unit one belongs to one of them it belongs to all of them.

[7] A full treatment of the stationary case in a high-dimensional setup can be found in Carvalho, Masini, and Medeiros (2018).

to make the dimension explicit and define the subvector $\underset{a:b}{[\boldsymbol{v}]}$ by

$$\underset{a:b}{[\boldsymbol{v}]} := \begin{pmatrix} v_a \\ v_{a+1} \\ \vdots \\ v_b \end{pmatrix}$$

for integers $1 \leq a \leq b \leq n$. Using this notation we have, for instance, $\boldsymbol{M} = \underset{1:n\times 1:m}{[\boldsymbol{M}]}$ and $\boldsymbol{v} = \underset{1:n}{[\boldsymbol{v}]}$. Also $\underset{1:1\times 1:1}{[\boldsymbol{M}]} = m_{1,1}$ and $\underset{2:n}{[\boldsymbol{v}]} = (v_2, \ldots, v_n)'$.

All the summations are from period 1 to T whenever the limits are left unspecified. For convenience, set $T_2 := T - T_0$ as the number post intervention periods. Recall that T_0 is number of period pre-intervention. We denote convergence in probability and weak converge by "$\xrightarrow{p}$" and "$\Rightarrow$," respectively.

3.2. Main Assumptions

To recover the effects of the intervention, we need the following key assumption.

Assumption 2. $\mathbb{E}(\boldsymbol{y}_t^{(0)}|\mathcal{D}_s) = \mathbb{E}(\boldsymbol{y}_t^{(0)})$, for all t, s.

Roughly speaking, the assumption above is sufficient for the peers to be unaffected by intervention on the unit of interest, that is, the peers are actually untreated.[8]

For a generic process $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$, consider the following assumption.

Assumption 3. Let $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ be a sequence of $(n \times 1)$ random vectors such that

(a) $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ is zero mean weakly (covariance) stationary;
(b) $\mathbb{E}|z_{i1}|^{\xi} < \infty$ for $i = 1, \ldots, n$ and some $2 \leq \xi < \infty$;
(c) $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ is mixing with coefficients such that $\sum_{m=1}^{\infty} \alpha_m^{1-1/\xi} < \infty$ or $\sum_{m=1}^{\infty} \phi_m^{1-2/\xi} < \infty$.

Assumption 3 states conditions under which the multivariate invariance principle is valid for $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ (refer to Proposition 1 in online Appendix A). Assumption 3(a) limits the heterogeneity in the process (at least up to the second moment). Assumption 3(b) is a standard higher moment existence condition which guarantees, along with Assumption 3(c), bounded covariances. Assumption 3(c) restricts the temporal dependence requiring the sequence to be either strong mixing with size $-\frac{\xi}{\xi-2}$ or uniform mixing with size $-\frac{\xi}{2\xi-2}$.

3.3. Variable Transformations

We follow Sims, Stock, and Watson (1990) and consider a variable transformation to derive the correct converge rate of the OLS estimator. Note that these transformations are not necessary in practical applications. For the DGPs considered in Assumption 1, the index r also represents the number of cointegration relations among the n units when $0 < r < n$. As shown in Engle and Granger (1987), the cointegration space characterized by $\widetilde{\boldsymbol{\Gamma}}(n \times r)$ can be normalized to $(\boldsymbol{I}_r : \boldsymbol{\Gamma}')'$, where $\boldsymbol{\Gamma}$ is of dimension $(n-r) \times r$. Since $(\boldsymbol{I}_r : \boldsymbol{\Gamma}')\boldsymbol{y}_t$ is covariance-stationary by definition, we define $\boldsymbol{\alpha} := \mathbb{E}\left[\left(\boldsymbol{I}_r : \boldsymbol{\Gamma}'\right)\boldsymbol{y}_t\right]$ and set $\boldsymbol{z}_t^* := \boldsymbol{C}'\boldsymbol{y}_t$ where

$$\boldsymbol{C} := \boldsymbol{C}(r) := \begin{cases} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} & \boldsymbol{0} \\ \underset{(n-r\times r)}{\boldsymbol{\Gamma}} & \boldsymbol{I}_{n-r} & 0 \\ -\underset{(1\times r)}{\boldsymbol{\alpha}'} & 0 & 1 \end{pmatrix}, & \text{if } 0 < r < n, \text{ and} \\ \boldsymbol{I}_{n+1}, & \text{otherwise.} \end{cases} \tag{3.1}$$

In the presence of a deterministic trend, the last $n-r$ elements of $\boldsymbol{z}_t^*$ might be asymptotic multicollinear. Hence, we propose the following transformation to cancel the drift for all but one, the last variable. If none of the variables has a deterministic trend there is no need to rotate the regressors. Otherwise, without loss of generality, make the last unit the one with trend and define the "rotation" $\boldsymbol{R}(r, \boldsymbol{\mu})$ as

$$\boldsymbol{R} := \boldsymbol{R}(r, \boldsymbol{\mu}) \tag{3.2}$$
$$:= \begin{cases} \begin{pmatrix} \boldsymbol{I}_{n-1} & 0 & 0 \\ -\boldsymbol{h}(0)' & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, & \text{if } (r = 0 \text{ or } r = n) \text{ and } \boldsymbol{\mu} \neq \boldsymbol{0}, \text{ and} \\ \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{0} & 0 & 0 \\ 0 & \boldsymbol{I}_{n-r-1} & 0 & 0 \\ 0 & -\boldsymbol{h}(r)' & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, & \text{for } 1 \leq r \leq n-2 \text{ and } \boldsymbol{\mu} \neq \boldsymbol{0} \\ \boldsymbol{I}_{n+1}, & \text{otherwise,} \end{cases}$$

where $\boldsymbol{h}(r) := \frac{1}{\mu_n}(\mu_{r+1}, \mu_{r+2}, \ldots, \mu_{n-1})'$. Now, we have that $\widetilde{\boldsymbol{z}}_t := \boldsymbol{R}'\boldsymbol{z}_t^*$ is no longer asymptotically multicollinear.

Also for the DGPS with $2 \leq r \leq n$, we have that the first component of $\widetilde{\boldsymbol{z}}_t$ is not necessarily orthogonal to the rest of the stationary components which will be necessary for the asymptotic results. Let g denote the number of stationary components of $\widetilde{\boldsymbol{z}}_t$, excluding the constant , that is, $g := \min(r, n-1)$. Hence, consider the linear projection of $\widetilde{z}_{1t}$ onto $\widetilde{z}_{2t}, \ldots, \widetilde{z}_{gt}$ and a constant: $\widetilde{z}_{1t} = \pi_2\widetilde{z}_{2t} + \cdots + \pi_g\widetilde{z}_{gt} + \delta + z_{1t}$. Define $\boldsymbol{\pi} := (\pi_2, \ldots, \pi_g)'$ and $\boldsymbol{z}_t := \boldsymbol{P}\widetilde{\boldsymbol{z}}_t$, where

$$\boldsymbol{P} := \boldsymbol{P}(r) \tag{3.3}$$
$$:= \begin{cases} \begin{pmatrix} \boldsymbol{1} & \boldsymbol{0} & \boldsymbol{0} & 0 \\ \underset{(g-1\times 1)}{-\boldsymbol{\pi}} & \boldsymbol{I}_{g-1} & \boldsymbol{0} & 0 \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{n-g} & 0 \\ -\delta & \boldsymbol{0} & \boldsymbol{0} & 1 \end{pmatrix}, & \text{if } 2 \leq r \leq n, \text{ and} \\ \boldsymbol{I}_{n+1}, & \text{otherwise.} \end{cases}$$

Finally, all the asymptotic results are carried on the transformed variable $\boldsymbol{z}_t := \boldsymbol{z}_t(r, \boldsymbol{\mu}) := \boldsymbol{H}(r, \boldsymbol{\mu})'\left[\boldsymbol{y}_t^{(0)\prime}, 1\right]'$, which combines all the three transformations described above as

$$\boldsymbol{H}(r, \boldsymbol{\mu}) := \boldsymbol{C}(r)\boldsymbol{R}(r, \boldsymbol{\mu})\boldsymbol{P}(r). \tag{3.4}$$

The OLS estimator, say $\widehat{\boldsymbol{\beta}}$, of y_{1t} on $\boldsymbol{y}_{0t}$ and a constant is related to the OLS estimator $\widehat{\boldsymbol{\gamma}}$ of $\boldsymbol{z}_{1t}$ on $\boldsymbol{z}_{0t}$ and a constant by

[8]For a thorough discussion on Assumption 2, including the potential bias resulting from its failure in the stationary setup refer to Carvalho, Masini, and Medeiros (2018).

Proposition 2 in online Appendix A. Its asymptotic distribution can be derived from Proposition 3 in online Appendix A. Propositions 2 and 3 fully characterize the asymptotic behavior of $\widehat{\boldsymbol{\beta}}$.

In our main result, we consider two kinds of asymptotics. In the first one (part (a) of Theorem 1) which we call *partial asymptotics*, we consider an asymptotic approach only for the pre-intervention period where the number of post-intervention observations $T_2 := T - T_0$ are kept fixed while $T_0 \to \infty$. This approach is tailored to accommodate situations where the number of pre-intervention period T_0 is much larger than T_2, which justifies the sampling error from the estimation of $\boldsymbol{\beta}_0$ by $\widehat{\boldsymbol{\beta}}$ to be of smaller order then ν_t.

Whereas in the second approach (part (b) of Theorem 1), named *full asymptotics*, we establish the asymptotic properties by considering the whole sample increasing in two scenarios. Either the proportion between the pre-intervention to the post-intervention sample size, denote by $\lambda_0 \in (0,1)$, if fixed such that $T_0 = \lfloor \lambda_0 T \rfloor$ and $T_2 := T_2(T)$ or the pre-intervention sample grows faster than the post-intervention one.[9] In this case, the asymptotics are taken as $T \to \infty$.

Theorem 1. Under the conditions of Proposition 3 in online Appendix A and Assumption 2:

(a) As $T_0 \to \infty$, for every $t \in \{T_0+1, \dots, T_2\}$

$$\xi_{T_0}(\widehat{\delta}_t - \delta_t - \nu_t) \Rightarrow -(\boldsymbol{G}\boldsymbol{q})'\boldsymbol{x}_t.$$

(b) As $T \to \infty$ with $T_0/T \to \lambda_0 \in (0,1)$:

$$\xi_T(\widehat{\Delta}_T - \Delta_T) \Rightarrow \frac{1}{1-\lambda_0}(1, \boldsymbol{q}')\boldsymbol{p}.$$

(c) As $T \to \infty$ and $T_0/T \to 1$ (or equivalently $T_2/T_0 \to 0$):

$$\xi_{T_2}(\widehat{\Delta}_T - \Delta_T) \Rightarrow \left[\int_0^1 \boldsymbol{B} \mathrm{d}s\right]_{1:1}, \quad \text{if } r = 0;$$

$$\xi_{T_2}(\widehat{\Delta}_T - \Delta_T) \Rightarrow \left[\int_0^1 \mathrm{d}\boldsymbol{B}\right]_{1:1}, \quad \text{if } 1 \le r \le n.$$

(d) $\zeta_T(\widehat{V}_T - V) \Rightarrow a$ as $T \to \infty$

where $\xi_T := \xi_T(r) := T^{1/2 - I(r=0)}$, $\zeta_T := \zeta_T(r) := T^{1/2-3/2I(r=0)}$, and $\widehat{V}_T := \frac{1}{T_2}\sum_{t=T_0+1}^{T}(\widehat{\delta}_t - \delta_t)^2$. The scalar V, the matrix $\boldsymbol{G}$, the random variable a and the random vectors $\boldsymbol{p}$, $\boldsymbol{q}$, and $\boldsymbol{B}$ are all defined in online Appendix B.

Remark 1. From the partial asymptotic analysis (when only the pre-intervention period is taken to infinity), we have an asymptotic unbiased estimator for each of the post-intervention period under cointegration. Note, however, that the estimator is not consistent. Not surprisingly, for the case where no cointegration relation exist ($r = 0$) the estimator diverges.

[9] We do not consider the opposite scenario when the post-intervention period grows faster than the pre-intervention since it would be unrealistic to most empirical applications.

Remark 2. From the full asymptotic analysis (when the whole sample size is taken to infinity keeping the ratio of the pre to post-intervention fixed) we have a $\sqrt{T}$ consistent estimator for the average intervention effect as a consequence of part (b) for the DGP with at least one cointegration relation or a common deterministic trend ($r > 0$). Again whenever no cointegration relation exist among the unit ($r = 0$) the estimator diverges.

Remark 3. Except for the cases when $T \to \infty$ and $T_0/T \to 1$ where the asymptotic distributions are Gaussian, all other cases have nonstandard limiting distribution as functionals of a Winer process. Furthermore, the presence of the drift nuisance parameters makes those limiting distribution nonpivotal, which requires an alternative inference procedure. Nevertheless, the results in this section give us the rate of convergence of those estimator which is key to the proposed sampling procedure discussed in the next section.

Remark 4. The SC counterfactual estimator and many of its variants, as the augmented synthetic control (ASC) estimator of Ben-Michael, Feller, and Rothstein (2019), are constructed as linear combinations of peers, such as $y_{1t}^{(0)} = \boldsymbol{\omega}'\boldsymbol{y}_{0t}$, where the estimation and restrictions on $\boldsymbol{\omega}$ differ across models/methods. Therefore, define $\Delta(\boldsymbol{\omega})$ as the average intervention effect for a given linear combination $\boldsymbol{\omega}$: $\Delta_T(\boldsymbol{\omega}) = \frac{1}{T-T_0}\sum_{t=T_0+1}^{T}\left(y_{1t} - \boldsymbol{\omega}'\boldsymbol{y}_{0t}\right)$. Not that, under spurious regressions, there is no value of $\boldsymbol{\omega}$ that makes the term inside the parentheses integrated of order zero. Therefore, under the null of no average intervention effect and the case of spurious regression ($r = 0$) it can be easily shown that $\Delta_T(\boldsymbol{\omega})$ diverges for any value of $\boldsymbol{\omega}$ as the sample size grows. As a consequence, in this case no method based on linear combinations of peers will give a reliable counterfactual.

4. Inference

As opposed to the stationary case, the limiting distributions of the estimators considered in Theorem 1 are all nonstandard unless in some specific cases. Critical values could still be computed by simulation as long as the nuisance parameters are identified or a plug-in estimator is used. None of those cases are ideal so we present some alternative asymptotic valid inference procedures.

They are based on the sequence of estimators $\{\widehat{\delta}_t\}_{t=T_0+1}^{T}$. More specifically, we consider any continuous mappings $\boldsymbol{\psi} : \mathbb{R}^{T_2} \to \mathbb{R}^q$ whose argument is the T_2-dimensional vector $(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \dots, \widehat{\delta}_T - \delta_T)'$. Thus, to conduct inference where are interested in distribution of $\widehat{\boldsymbol{\psi}} := \boldsymbol{\psi}(\widehat{\delta}_{T_0+1} - \delta_{T_0+1}, \dots, \widehat{\delta}_T - \delta_T)$. We consider an estimator for the cdf:

$$G_T(\boldsymbol{w}) := ¶(\widehat{\boldsymbol{\psi}} \le w), \quad \boldsymbol{w} \in \mathbb{R}^q, \tag{4.1}$$

where, for a pair of vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^q$, we say that $\boldsymbol{a} \le \boldsymbol{b} \iff a_i \le b_i, \forall i$. Also we only consider the case when $r > 0$, since otherwise the estimators in Theorem 1 diverge.

4.1. *Partial Asymptotic Inference (Resampling)*

The results in this section are based on the procedure proposed in Masini and Medeiros (2019). We consider the case when the pre-intervention period is much larger than the post intervention period, $T_0 \gg T_2$. The results are based on part (a) of Theorem 1. When $T_0 \longrightarrow \infty$, we have that $\widehat{\boldsymbol{\psi}} \xrightarrow{p} \boldsymbol{\psi}_0$, where $\boldsymbol{\psi}_0 := \boldsymbol{\psi}(\nu_{T_0+1}, \ldots, \nu_T)$. Therefore, $G_T(\boldsymbol{w}) \to G(\boldsymbol{w})$ for every continuity point of G defined by

$$G(\boldsymbol{w}) := \P(\boldsymbol{\psi}_0 \leq w), \quad \boldsymbol{w} \in \mathbb{R}^q. \tag{4.2}$$

Consider the construction of $\widehat{\boldsymbol{\psi}}$ using blocks of size T_2 of consecutive observations from the pre-intervention sample. There are $T_0 - T_2 - 1$ of such estimators denoted by

$$\widehat{\boldsymbol{\psi}}_j := \boldsymbol{\psi}(\widehat{\nu}_j, \ldots, \widehat{\nu}_{j+T_2-1}), \quad j = 1, \ldots, T_0 - T_2 + 1,$$

where $\widehat{\nu}_t := y_{1t} - \widehat{\boldsymbol{\beta}}'_{T_0}\boldsymbol{x}_t$ with the subscript T_0 indicates that the estimator is calculated using the only the pre-intervention sample.

For any fixed j, we have that $\widehat{\boldsymbol{\psi}}_j \xrightarrow{p} \boldsymbol{\psi}_j := \boldsymbol{\psi}(\nu_j, \ldots, \nu_{j+T_2-1})$ as $T_0 \to \infty$. Under strictly stationarity of $\{\nu_t\}_{t\geq 1}$, we have that $\boldsymbol{\psi}_j$ is equal in distribution to $\boldsymbol{\psi}_0$ for all j. Therefore, we propose to estimate (4.1) by

$$\widehat{G}_T(\boldsymbol{w}) := \frac{1}{T_0 - T_2 + 1} \sum_{j=1}^{T_0-T_2+1} I(\widehat{\boldsymbol{\psi}}_j \leq \boldsymbol{w}), \quad \boldsymbol{w} \in \mathbb{R}^q.$$

Theorem 2. Under the same conditions of Theorem 1, if further $\{\nu_t\}$ is a strictly stationary process, then as $T_0 \to \infty$:

(a) $\widehat{\boldsymbol{\psi}} \xrightarrow{p} \boldsymbol{\psi}_0$

(b) $\left|\widehat{G}_T(\boldsymbol{w}) - G_T(\boldsymbol{w})\right| \xrightarrow{p} 0$ for every continuity point of $\boldsymbol{w} \mapsto G(\boldsymbol{w})$

(c) If G is continuous, the result (b) holds uniformly in $\boldsymbol{w} \in \mathbb{R}^q$.

By the appropriate choice of $\boldsymbol{\psi}(\cdot)$, Theorem 2 provides a straightforward way to conduct inference even in the case when there is a single observation after the intervention ($T_2 = 1$). We could be interested in testing the intervention effects on all post-intervention period individually by setting $\boldsymbol{\psi}(\boldsymbol{w}) = \boldsymbol{w}$. Or on the average intervention effect across the post-intervention period $\boldsymbol{\psi}(\boldsymbol{w}) = \frac{1}{T_2}\sum_{j=1}^{T_2} w_j$. A reasonable choice to test $\mathcal{H}_0$ using a univariate statistic is to set $\boldsymbol{\psi}(\boldsymbol{w}) = \frac{1}{T_2}\boldsymbol{w}'\boldsymbol{w}$, which is a particular choice among statistics of the form $\boldsymbol{\psi}(\boldsymbol{w}) = \frac{1}{T_2}\sum_{j=1}^{T_2} g(w_j)$ for some nonnegative real valued function $g(\cdot)$ such as $|\cdot|$.

Regardless of the choice, Theorem 2 ensures an asymptotic correct size. For instance, for an univariate $\widehat{\psi}$ we have $\P\left[\widehat{\psi} \leq c_T(1-\alpha)\right] \to 1 - \alpha$, $T_0 \to \infty$, where $c_T(\alpha) := \inf\{w : \widehat{G}_T(w) \geq \alpha\}$. The proof of Theorem 2 follows closely the results in Masini and Medeiros (2019).

4.2. Full Asymptotic Inference (Subsampling)

Now we consider the situation where the size of the pre- and post-intervention period are comparable, $T_0 \approx T_2$, which makes the partial asymptotic argument from the previous section less compelling. Instead we consider full asymptotic approach based on part (b) of Theorem 1. For a given positive integer b split the sample into subsamples of b consecutive observations indexed by $\mathcal{S}_j := \{j, \ldots, j+b-1\}; j \in \mathcal{J} := \{1, \ldots, T-b+1\}$. For each of those subsamples define the index set of post-intervention period keeping the same ratio of original sample $\lambda_0 = T_0/T$, that is, $\mathcal{T}_j := \{\lfloor \lambda_0 b \rfloor + j, \ldots, j+b-1\} \subset \mathcal{S}_j$.

Then, in analogy to $\widehat{\Delta}_T - \Delta_T$ define, for each subsample: $L_j := \frac{1}{\#\mathcal{T}_j}\sum_{t\in\mathcal{T}_j}\left(y_{1t} - \widehat{\boldsymbol{\beta}}'_j\boldsymbol{x}_t - \delta_t\right)$, for $j \in \mathcal{J}$, where $\widehat{\boldsymbol{\beta}}_j$ is the OLS estimator based on a sample indexed by $\mathcal{S}_j \setminus \mathcal{T}_j$ and $\#\mathcal{A}$ denotes the cardinality of $\mathcal{A}$.

According to part (b) of Theorem 1 $\sqrt{b}L_j \Rightarrow p$ for each j as $b \to \infty$. We use $\{L_j\}$ as an estimator for the asymptotic distribution of $\widehat{\Delta}_T$: $F_T(x) := \P(\sqrt{T}(\widehat{\Delta}_T - \Delta_T) \leq x)$, $x \in \mathbb{R}$. The estimator becomes, for each choice of b: $\widehat{F}_{T,b}(x) := \frac{1}{\#\mathcal{J}}\sum_{j\in\mathcal{J}} I(\sqrt{b}L_j \leq x), x \in \mathbb{R}$.

Theorem 3. Under the same conditions of Theorem 1 and $r = n$, provided that $b \to \infty$ as $T \to \infty$ and $b/T \to 0$

$$\sup_{x\in\mathbb{R}} |\widehat{F}_{T,b}(x) - F_T(x)| \xrightarrow{p} 0.$$

Remark 5. Note that the results in the above theorem are only valid in the case of DGP (c) of Assumption 1.

5. Simulations

5.1. Asymptotic Distributions

To evaluate the asymptotic approximation in finite samples, we simulate two different scenarios. In the first one, the treated unit and the peers are cointegrated while in the second case the data are formed by a set of independent random walks. In this later case, the counterfactual is spurious. In both cases, we evaluate the distribution of the estimator for the average intervention effect under the null hypothesis of no intervention at $T_0 = T/2$. We consider $T = 100$ and 1000, and $n = 5$. It is clear that $T = 1000$ is not empirically relevant sample size. Nevertheless, we keep this case to provide evidence in favor of the asymptotic theory developed here. The number of Monte Carlo simulations is set to 10,000. For each scenario and different sample sizes, we report the finite sample distributions of $\widehat{\Delta} = \frac{1}{T-T_0}\sum_{t=T_0+1}^{T}\widehat{\delta}_t$, in comparison to the asymptotic distributions as well as the rejection frequencies, at different significance levels, of the null hypothesis of no intervention effects when nonstationarity is neglected and the test is carried out under standard normal approximation for the t-statistic.

As a complement we also report the empirical rejection rates for the t-test of no intervention effect when the parameters are estimated either by restricted least squares or by the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996). In the first approach, the parameters of the linear combination are restricted to be positive and sum one as in the original SC method, while the LASSO approach was advocated by Carvalho, Masini, and Medeiros (2018) and Doudchenko and Imbens (2016). We report only the case where the linear trend is not included in the regression function. The figures illustrating the results are relegated to the supplementary materials. In the following subsections, we describe the mains findings.

5.1.1. Cointegration

The DGP is defined as

$$y_{1t} = \sum_{i=2}^{n} y_{it} + u_{1t}, \tag{5.1}$$

where $y_{it} = y_{it-1} + u_{it}$, $y_{i0} = 0$, $i = 2, \dots, n$, and $\{u_{1t}, \dots, u_{nt}\}_{t=1}^{T}$ is a sequences of independent and normally distributed zero-mean random variables with unit variance. Furthermore, $\mathbb{E}(u_{jt}u_{is}) = 0$ for all $t = 1, \dots, T$, $s = 1, \dots, T$, $i = 1, \dots, n$, $j = 1, \dots, n$, and $t \neq s$.

Simulation results are shown in Figures S.1–S.3. Figure S.1 shows the empirical versus the theoretical distributions of the scaled coefficient estimates. The distribution of $\widehat{\Delta}_j$, $j = 1, 2$, is presented in Figure S.2 and is compared to the asymptotic results in the article. Figure S.3 compares the size distortions when the normal approximation is used, neglecting nonstationarity, with the case when the correct asymptotic critical values are used.

Two conclusions emerge from the results. First, the simulation corroborates the asymptotic approximation even in small samples. Second, it is clear that neglecting cointegration introduces strong over-rejection of the null hypothesis, leading the researcher to find spurious intervention effects. Finally, it is clear from Figure S.4 that restricting the coefficients does not mitigate the over-rejections is nonstationarity is not taken carefully into account.

5.1.2. Spurious Counterfactual

In this case, the DGP is a vector of independent random walks as follows:

$$y_{it} = y_{it-1} + u_{it}, \tag{5.2}$$

where $y_{i0} = 0$ and $\{u_{it}\}_{t=1}^{T}$ is a sequence of independent and normally distributed zero-mean random variables with unit variance and $\mathbb{E}(u_{it}u_{js}) = 0$ for all $t = 1, \dots, T$, $s = 1, \dots, T$, $i = 1, \dots, n$, $j = 1, \dots, n$, and $t \neq s$.

The simulation results for the spurious case are depicted in Figures S.5–S.7. Figure S.5 presents the empirical versus the theoretical distributions of the coefficients estimates. The distribution of the average intervention effects, $\widehat{\Delta}_j$, $j = 1, 2$, is presented in Figure S.6 and is confronted with the asymptotic results. Finally, Figure S.7 compares the size distortions of the scaled t-test when the normal approximation is used, neglecting nonstationarity, with the case when the correct asymptotic critical values are used. Note that this is not a valid test as the t-stat, without normalization, diverges. The size distortions are presented just for illustrative purposes.

It is clear from the figures that the finite sample distribution can be well approximated by the asymptotic counterpart. Furthermore, the distribution of the scaled t-stat is highly bimodal. Finally, conducting inference in the spurious case is extremely misleading even when restricted estimators are considered as displayed in Figure S.8.

5.2. Inference

We also conduct a simulation to evaluate the finite sample properties of the proposed partial resampling inferential procedure and to evaluate as well the effects of pretesting for cointegration. We simulate $T = \{30, 50, 150, 200, 500\}$ observations of three different models: two involving spurious regression and a third one where all the variables are cointegrated. We evaluate the empirical size of the test following the strategy described below when there is only one observation after the intervention: (1) We run the Engle–Granger methodology to test the null for no-cointegration among the variables. The dependent variable is y_{1t}. We use the Phillips–Ouliaris critical values. (2) If the null of no-cointegration is rejected at level α, we estimate the models in levels and proceed with the inferential procedure advocated in the article. Otherwise, we estimate the model in first-differences.

We consider the following models:

$$y_{it} = \mu_i + y_{it-1} + u_{it},\ i = 1, \dots, n;\ t = 1, \dots, T,$$

where $\{u_t\}_{t=1}^{T}$ is a sequence of independent and normally distributed random variables with zero mean and variance one. Furthermore, $\mathbb{E}(u_{it}u_{jt}) = 0, \forall\, i \neq j$. We set $n = 5$. In this case, the regression of y_{1t} on $y_{2t}, \dots, y_{nt}$ is spurious and there is no cointegration among the variables.

The second model is given by

$$\begin{aligned} y_{1t} &= \mu_1 + y_{1t-1} + u_{1t} \\ y_{2t} &= \sum_{j=3}^{n} y_{jt} + u_{2t} \\ y_{it} &= \mu_i + y_{it-1} + u_{it}, \quad i =, 3, \dots, n;\ t = 1, \dots, T. \end{aligned}$$

The error terms are defined as before. Again, the regression of y_{1t} on $y_{2t}, \dots, y_{nt}$ is spurious but there is a cointegration relation among the donors.

Finally, the third model is

$$\begin{aligned} y_{1t} &= \boldsymbol{\beta}'\boldsymbol{y}_{0t} + u_{1t} \\ y_{it} &= \mu_i + y_{it-1} + u_{it}, \quad i =, 2, \dots, n;\ t = 1, \dots, T, \end{aligned}$$

where $\boldsymbol{y}_{0t} = (y_{2t}, \dots, y_{nt})$ and each element of $\boldsymbol{\beta}$ is sampled, at every Monte Carlo iteration, from a Uniform distribution between -1 and $+1$.

In all the above three models, we consider three cases for the drift term μ_i: (1) $\mu_i = 0$; (2) $\mu = 0.5$; and (3) $\mu \sim \mathsf{N}(0, 0.25)$. The results are displayed in Tables 1–2. The tables report the rejection rates of the partial resampling test (Theorem 2) with one observation after the intervention. We consider different significance levels for the pretest for the null of no-cointegration ($\alpha = \{\text{NA}, 0.001, 0.01, 0.05, 0.10\}$). Note that $\alpha = \text{NA}$ means that pretesting is not conducted. Panel (a) refers to estimation without imposing any restriction on the estimated regression. Panel (b) refers to the restriction that the coefficients of the estimated model must be positive and sum one. Panel (c) reports the results when the model is estimated by LASSO with the penalty term selected by the BIC as in Carvalho, Masini, and Medeiros (2018). Finally, Panel (d) considers the ASC estimator proposed by Ben-Michael, Feller, and Rothstein (2019). The ASC model nests the synthetic differences-in-differences estimator of Arkhangelsky et al. (2019) and has the gold of reducing bias in the original SC estimator.

Several conclusions emerge from the tables. We start analyzing the spurious case (Tables 1 and 3). It is clear that without first pretesting for cointegration, the size distortions are quite

Table 1. Rejection rates under the null (empirical size): spurious regression.

	Panel (a): Unrestricted estimation														
	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim$ N[0, 0.25])				
	$\alpha =$ NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.26	0.13	0.13	0.14	0.14	0.27	0.12	0.13	0.13	0.14	0.27	0.12	0.13	0.13	0.14
50	0.23	0.09	0.09	0.10	0.10	0.24	0.10	0.10	0.10	0.11	0.23	0.09	0.09	0.09	0.10
150	0.20	0.06	0.06	0.06	0.06	0.20	0.06	0.06	0.07	0.07	0.19	0.06	0.06	0.06	0.07
200	0.18	0.06	0.06	0.06	0.06	0.18	0.05	0.05	0.06	0.06	0.19	0.05	0.05	0.06	0.06
500	0.17	0.05	0.05	0.05	0.05	0.18	0.05	0.05	0.05	0.06	0.18	0.05	0.05	0.05	0.06
	Panel (b): Synthetic control restriction														
	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim$ N[0, 0.25])				
	$\alpha =$ NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.24	0.11	0.11	0.12	0.13	0.24	0.10	0.10	0.11	0.12	0.29	0.10	0.10	0.11	0.12
50	0.22	0.09	0.09	0.09	0.10	0.22	0.09	0.09	0.10	0.10	0.27	0.07	0.08	0.08	0.09
150	0.20	0.06	0.06	0.06	0.07	0.20	0.06	0.06	0.06	0.07	0.29	0.06	0.06	0.07	0.08
200	0.19	0.05	0.05	0.06	0.06	0.19	0.06	0.06	0.06	0.07	0.30	0.06	0.06	0.07	0.08
500	0.19	0.05	0.05	0.05	0.06	0.19	0.05	0.05	0.06	0.06	0.34	0.06	0.06	0.07	0.08
	Panel (c): LASSO restriction														
	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim$ N[0, 0.25])				
	$\alpha =$ NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.26	0.09	0.09	0.10	0.11	0.27	0.08	0.09	0.09	0.10	0.26	0.08	0.08	0.09	0.10
50	0.23	0.07	0.07	0.08	0.09	0.23	0.07	0.07	0.08	0.09	0.22	0.07	0.07	0.07	0.08
150	0.20	0.05	0.05	0.05	0.06	0.20	0.06	0.06	0.06	0.07	0.19	0.05	0.06	0.06	0.06
200	0.18	0.05	0.05	0.05	0.06	0.18	0.05	0.05	0.05	0.06	0.19	0.05	0.05	0.05	0.06
500	0.17	0.05	0.05	0.05	0.05	0.18	0.05	0.05	0.05	0.06	0.18	0.05	0.05	0.05	0.06
	Panel (d): Augmented synthetic control														
	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim$ N[0, 0.25])				
	$\alpha =$ NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.24	0.11	0.11	0.12	0.12	0.24	0.10	0.10	0.11	0.12	0.29	0.10	0.10	0.11	0.12
50	0.22	0.09	0.09	0.09	0.10	0.22	0.09	0.09	0.09	0.10	0.27	0.07	0.08	0.08	0.09
150	0.20	0.06	0.06	0.06	0.07	0.20	0.06	0.06	0.06	0.07	0.28	0.06	0.06	0.07	0.08
200	0.19	0.05	0.05	0.06	0.06	0.19	0.05	0.06	0.06	0.07	0.29	0.06	0.06	0.07	0.08
500	0.19	0.05	0.05	0.05	0.06	0.18	0.05	0.05	0.06	0.06	0.32	0.06	0.06	0.07	0.08

NOTE: The table reports the rejection rates of the partial resampling test with one observation after the intervention. We consider different significance levels for the pretest for the null of no-cointegration (α). Note that $\alpha = 0$ means that pretesting is not conducted. Panel (a) refers to estimation without imposing any restriction. Panel (b) refers to the restriction that the coefficients of the estimated model must be positive and sum one. Panel (c) reports the results when the model is estimated by LASSO with the penalty term selected by the BIC. Finally, Panel (d) displays the case of the augmented synthetic control (ASC) estimator.

large and the imposition of restrictions does not attenuate the problem. When pretesting is conducted, the distortions drop as the sample size increases, as expected. Furthermore, imposing restrictions marginally help when the sample is very small ($T = 30$ or $T = 50$).

Now we turn to the cointegration case. First, size distortions can be large when $T = 30$ or $T = 50$ but converges to the nominal size as the sample grows. LASSO restrictions do not seem to help. On the other hand, the test based on the SC and ASC estimators is extremely oversized. This is possibly due to the over restrictive conditions on the parameters of the model.

6. Empirical Example

6.1. Overview

The goal of this empirical illustration is to test whether the share prices of Petrobras as traded at the New York Stock Exchange were affected the policy of the Brazilian government to freeze fuel prices in Brazil and the major corruption scandals involving the main executives of the company during the recent years.

Petroleo Brasileiro S.A. (Petrobras) is a semipublic Brazilian multinational corporation in the petroleum industry headquartered in Rio de Janeiro, Brazil. The company was recently ranked in the 58th position in the most recent Fortune Global 500 list.[10] Petrobras operates in a number of segments in the oil and gas industry worldwide and it was founded by the Brazilian government in 1953 as a national oil company with a legal monopoly on oil produced in Brazil. The pattern of a government founded company controlling all the oil resources of a given nation is not uncommon; it has been replicated many times with Venezuela, Mexico, Norway, Saudi Arabia, and others, with different degrees of success. As a national oil company, Petrobras was initially fully state-owned. However, in November of 1995, the Brazilian Congress amended the constitution to end Petrobras' 43-year-old monopoly. As part of this legislation, the government also authorized the sale of up to 50% minus one of its voting shares, the rest of which was to be held by the government.

[10] *http://fortune.com/global500/*

Table 2. Rejection rates under the null (empirical size): cointegration.

Panel (a): Unrestricted estimation

	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim$ N[0, 0.25])				
	$\alpha =$ NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.14	0.14	0.15	0.15	0.15	0.14	0.14	0.15	0.15	0.15	0.15	0.14	0.16	0.16	0.16
50	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.10	0.10	0.10	0.10	0.11	0.10	0.10	0.10
150	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.07
200	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.05
500	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Panel (b): Synthetic control restriction

	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim$ N[0, 0.25])				
	$\alpha =$ NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.17	0.12	0.15	0.17	0.17	0.38	0.18	0.29	0.36	0.38	0.23	0.14	0.19	0.22	0.23
50	0.17	0.16	0.17	0.17	0.17	0.43	0.40	0.42	0.43	0.43	0.25	0.23	0.25	0.25	0.25
150	0.17	0.17	0.17	0.17	0.17	0.57	0.57	0.57	0.57	0.57	0.28	0.28	0.28	0.28	0.28
200	0.17	0.17	0.17	0.17	0.17	0.60	0.60	0.60	0.60	0.60	0.30	0.30	0.30	0.30	0.30
500	0.18	0.18	0.18	0.18	0.18	0.71	0.71	0.71	0.71	0.71	0.33	0.33	0.33	0.33	0.33

Panel (c): LASSO restriction

	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim$ N[0, 0.25])				
	$\alpha =$ NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.15	0.14	0.15	0.15	0.15	0.14	0.13	0.15	0.15	0.15	0.15	0.14	0.15	0.16	0.16
50	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.10	0.10	0.10	0.10	0.11	0.10	0.10	0.10
150	0.06	0.06	0.06	0.06	0.06	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
200	0.06	0.06	0.06	0.06	0.06	0.09	0.09	0.09	0.09	0.09	0.08	0.08	0.08	0.08	0.08
500	0.05	0.05	0.05	0.05	0.05	0.18	0.18	0.18	0.18	0.18	0.16	0.16	0.16	0.16	0.16

Panel (d): Augmented synthetic control

	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim$ N[0, 0.25])				
	$\alpha =$ NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.15	0.13	0.15	0.16	0.16	0.20	0.14	0.18	0.20	0.20	0.17	0.14	0.16	0.18	0.18
50	0.11	0.12	0.12	0.11	0.11	0.18	0.18	0.18	0.18	0.18	0.14	0.14	0.14	0.14	0.14
150	0.09	0.09	0.09	0.09	0.09	0.20	0.20	0.20	0.20	0.20	0.13	0.13	0.13	0.13	0.13
200	0.08	0.08	0.08	0.08	0.08	0.21	0.21	0.21	0.21	0.21	0.14	0.14	0.14	0.14	0.14
500	0.08	0.08	0.08	0.08	0.08	0.31	0.31	0.31	0.31	0.31	0.20	0.20	0.20	0.20	0.20

NOTE: The table reports the rejection rates of the partial resampling test with one observation after the intervention. We consider different significance levels for the pretest for the null of no-cointegration (α). Note that $\alpha = 0$ means that pretesting is not conducted. Panel (a) refers to estimation without imposing any restriction. Panel (b) refers to the restriction that the coefficients of the estimated model must be positive and sum one. Panel (c) reports the results when the model is estimated by LASSO with the penalty term selected by the BIC. Finally, Panel (d) displays the case of the augmented synthetic control (ASC) estimator.

Consequently, Petrobras is not government owned, but a semipublic company. The Brazilian government owns 54% of the shares and two government investment funds hold a further 10%.

6.2. *Petrobras Under Dilma Rouseff and the Car Wash Operation*

Dilma Rousseff served as the 36th President of Brazil, holding the position from 2011 until her impeachment and removal from office on 31 August 2016. She was the first woman to hold the Brazilian presidency and had previously served as Chief of Staff to former president Luiz Inácio Lula da Silva from 2005 to 2010. Dilma Rousseff's administration was constantly accused of using Petrobras as an instrument to control inflation rates. Regardless of what occurred internationally, the Brazilian government hold domestic fuel prices frozen, especially during the 2013–2014 electoral season, when Dilma Roussef was running for re-election. In a country as dependent on road transportation as Brazil, if fuel prices increase, several other prices also rise by following the upward trend due to the increase in distribution costs. By holding down inflation rates, Dilma Rouseff was able to cash in politically and create an impression that Brazil was not severely affected by the 2008–2009 global crisis. While that might have curbed inflation, it also accumulated billion-dollar losses to the company. Furthermore, Petrobras took a central role in one of the major corruption scandals in history.

Operation Car Wash is an ongoing criminal investigation being carried out by the Federal Police of Brazil. Initially a money laundering investigation, it has expanded to cover allegations of corruption at Petrobras, where executives allegedly accepted bribes in return for awarding contracts to construction firms at inflated prices. The investigation is called *Operation Car Wash* because it was first uncovered at a car wash in Brasília, and has included more than a thousand warrants for search and seizure, temporary and preventive detention, and plea bargain coercive measures, with the aim of ascertaining the extent of a money laundering scheme suspected of moving more than five billion of dollars.

Table 3. Rejection rates under the null (empirical size): y_{1t} not in the cointegration relation.

Panel (a): Unrestricted estimation

	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim \mathrm{N}[0, 0.25]$)				
	$\alpha = \mathrm{NA}$	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.27	0.13	0.13	0.14	0.14	0.28	0.12	0.13	0.13	0.14	0.27	0.12	0.13	0.13	0.14
50	0.23	0.09	0.09	0.09	0.10	0.23	0.09	0.09	0.10	0.10	0.23	0.09	0.09	0.09	0.10
150	0.19	0.06	0.06	0.06	0.07	0.19	0.06	0.06	0.06	0.07	0.19	0.06	0.06	0.06	0.07
200	0.18	0.05	0.05	0.06	0.06	0.18	0.05	0.05	0.05	0.06	0.18	0.05	0.06	0.06	0.06
500	0.17	0.05	0.05	0.05	0.06	0.19	0.05	0.05	0.05	0.06	0.19	0.05	0.05	0.05	0.05

Panel (b): Synthetic control restriction

	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim \mathrm{N}[0, 0.25]$)				
	$\alpha = \mathrm{NA}$	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.23	0.11	0.11	0.11	0.12	0.24	0.11	0.11	0.12	0.13	0.29	0.11	0.11	0.12	0.13
50	0.22	0.08	0.08	0.08	0.09	0.21	0.09	0.09	0.09	0.10	0.28	0.08	0.08	0.09	0.10
150	0.20	0.06	0.06	0.07	0.07	0.19	0.06	0.06	0.07	0.07	0.28	0.06	0.06	0.07	0.08
200	0.19	0.05	0.06	0.06	0.07	0.18	0.05	0.05	0.06	0.07	0.30	0.06	0.06	0.07	0.08
500	0.18	0.05	0.05	0.06	0.06	0.19	0.05	0.05	0.06	0.06	0.34	0.05	0.05	0.06	0.08

Panel (c): LASSO restriction

	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim \mathrm{N}[0, 0.25]$)				
	$\alpha = \mathrm{NA}$	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.26	0.09	0.09	0.10	0.11	0.27	0.08	0.09	0.10	0.11	0.26	0.09	0.09	0.10	0.10
50	0.22	0.07	0.07	0.07	0.08	0.23	0.07	0.07	0.08	0.09	0.23	0.06	0.07	0.07	0.08
150	0.19	0.05	0.05	0.06	0.06	0.18	0.05	0.06	0.06	0.06	0.19	0.05	0.05	0.05	0.06
200	0.18	0.05	0.05	0.05	0.06	0.18	0.05	0.05	0.05	0.05	0.18	0.05	0.05	0.05	0.06
500	0.17	0.05	0.05	0.05	0.05	0.19	0.05	0.05	0.05	0.06	0.19	0.05	0.05	0.05	0.06

Panel (d): Augmented synthetic control

	No drift ($\mu_i = 0$)					Homogeneous ($\mu_i = 0.5$)					Heterogeneous ($\mu_i \sim \mathrm{N}[0, 0.25]$)				
	$\alpha = \mathrm{NA}$	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10	NA	0.001	0.01	0.05	0.10
$T = 30$	0.24	0.10	0.10	0.11	0.12	0.25	0.10	0.10	0.11	0.12	0.28	0.10	0.10	0.11	0.12
50	0.22	0.07	0.08	0.08	0.09	0.22	0.08	0.08	0.09	0.10	0.26	0.07	0.08	0.08	0.09
150	0.19	0.06	0.06	0.06	0.07	0.19	0.06	0.06	0.07	0.07	0.25	0.05	0.06	0.06	0.07
200	0.18	0.05	0.05	0.06	0.06	0.18	0.05	0.05	0.06	0.06	0.26	0.05	0.06	0.06	0.07
500	0.18	0.05	0.05	0.06	0.06	0.19	0.05	0.05	0.06	0.06	0.29	0.05	0.05	0.06	0.07

NOTE: The table reports the rejection rates of the partial resampling test with one observation after the intervention. We consider different significance levels for the pretest for the null of no-cointegration (α). Note that $\alpha = 0$ means that pretesting is not conducted. Panel (a) refers to estimation without imposing any restriction. Panel (b) refers to the restriction that the coefficients of the estimated model must be positive and sum one. Panel (c) reports the results when the model is estimated by LASSO with the penalty term selected by the BIC. Finally, Panel (d) displays the case of the augmented synthetic control (ASC) estimator.

6.3. Results

To estimate the potential losses after the election of Dilma Rousseff and the company's policy to freeze domestic fuel prices, we consider January 2011 as the intervention date. To construct the counterfactual, we consider the prices of other oil companies, stock indexes as well as international oil prices. The oil companies considered are: British Petroleum (BP), CNOOC (CEO), ConocoPhillips (COP), Chevron (CVX), Lukoil (LUKOY), Shell (RDSB), Total (TOT), Exxon Mobil (XOM). The indexes are: Crude oil prices (MCOILBRENTEU), S&P500 (GSPC), and Nasdaq (IXIC). Our sample starts in August 2000 and ends in August 2016.

Table 4 shows the estimated coefficients and the standard errors of the pre-intervention model. Both the dependent and independent variables are in natural logarithms and the model is estimated by OLS. The standard errors are heteroscedastic and autocorrelation robust and are computed with the quadratic spectral kernel with the bandwidth selected by Andrew's method. The table also reports the in-sample R-squared as well as the results for the Phillips–Ouliaris cointegration test. The number between parentheses is the p-value of the test.

The estimated counterfactual is presented in Figure 1. The left panel reports the actual and counterfactual log prices, whereas the right panel of Figure 1 displays the 95% confidence interval after the intervention date computed from the results in Theorem 2. The results suggest that the price policy and the corruption scandals involving Petrobras influenced the company's prices very negatively despite the drop of international oil prices.

7. Conclusions

In this article, we considered the asymptotic properties of popular counterfactual estimators when the data are nonstationary. Our econometric framework encapsulates the panel based methods of Hsiao, Ching, and Wan (2012), the ArCo approach of Carvalho, Masini, and Medeiros (2018), and the SC and its extensions Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010), and Doudchenko and Imbens (2016).

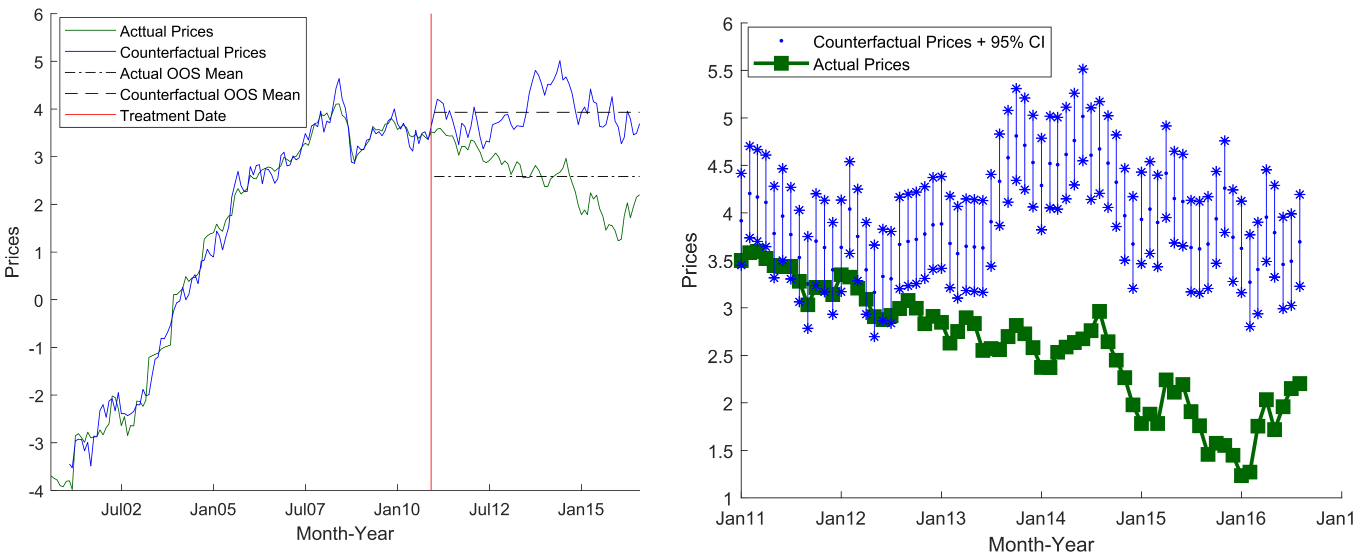

Figure 1. (Left panel) Price evolution: actual and counterfactual. (Right panel) Estimation results: actual and counterfactual prices plus a 95% confidence interval.

Table 4. Pre-intervention model estimation.

Pre-intervention model		
Variable	Coefficient	Std. error
Constant	3.53	2.20
British Petroleum (BP)	0.39	0.29
CNOOC (CEO)	0.07	0.21
ConocoPhillips (COP)	1.80	0.43
Chevron (CVX)	0.33	0.57
Lukoil (LUKOY)	0.66	0.26
Shell (RDSB)	−2.26	0.55
Total (TOT)	3.12	0.73
Exxon Mobil (XOM)	0.30	0.65
Crude oil prices (MCOILBRENTEU)	−0.17	0.28
S&P500 (GSPC)	−6.69	0.74
Nasdaq (IXIC)	4.17	0.70
R-squared	0.99	
P-O (τ-statistic)	−6.85(0.02)	
P-O (z-statistic)	−66.18(0.02)	

NOTE: The table shows the estimated coefficients as well as the standard errors of the pre-intervention regression model. Both the dependent and independent variables are in natural logarithms and the model is estimated by ordinary least squares. The standard errors are heteroscedastic and autocorrelation robust and are computed with the quadratic spectral kernel with the bandwidth selected by Andrew's automatic method. The table also reports the in-sample R-squared as well as the results for the Phillips–Ouliaris cointegration test. The number between parentheses is the p-value of the test.

Two cases are considered. In the first case, there is at least one cointegrating relationship in the data while in the second one the data are formed by a set of independent random walks. The results in the article show that the estimators either diverge or have nonstandard asymptotic distributions. We show strong over-rejection of the null hypothesis of no intervention effect when the nonstationary nature of the data is ignored. Our theoretical results are corroborated by a simulation experiment.

Supplementary Materials

In the online supplements, we report additional material to support the results in the article, which includes additional lemmas and simulation results. In addition to the references already cited in the article, the proofs in the supplement make use of the results in Davidson (1994), Hansen (1992), and Kurtz and Protter (1991). The online appendices contain auxiliary results, details on the notation of Theorem 1, and proof of the main results.

Acknowledgments

This article is a major extension of a previous article titled "The Perils of Counterfactual Analysis with Integrated Processes," which is co-authored with Carlos V. Carvalho. We wish to deeply thank our former co-author Carlos Carvalho for pointing us to this research agenda and for inspiring discussions. We also wish to thank the participants at the 2017 NBER-NSF Time Series Conference, the 2017 Meeting of the Brazilian Econometric Society (SBE), the Workshop Rio-São Paulo of Econometrics, and the 2017 Conference of the International Association for Applied Econometrics for helpful comments. Finally, we wish to thank the co-editor, Jianqing Fan, the associate editor, and three anonymous referees for great comments and suggestions. Nevertheless, any mistakes in the article are our own.

Funding

The research of the second author is partially supported by CNPq and CAPES.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505. [227,237]

Abadie, A., and Gardeazabal, J. (2003), "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93, 113–132. [227,237]

Arkhangelsky, D., Athey, S., Hirshberg, D., Imbens, G., and Wager, S. (2019), "Synthetic Difference in Differences," Working Paper, arXiv no. 1812.09970. [234]

Athey, S., and Imbens, G. (2017), "The State of Applied Econometrics—Causality and Policy Evaluation," *Journal of Economic Perspectives*, 31, 3–32. [228]

Bai, C.-E., Li, Q., and Ouyang, M. (2014), "Property Taxes and Home Prices: A Tale of Two Cities," *Journal of Econometrics*, 180, 1–15. [228]

Ben-Michael, E., Feller, A., and Rothstein, J. (2019), "The Augmented Synthetic Control Method," Working Paper, arXiv no. 1811.04170. [232,234]

Carvalho, C., Masini, R., and Medeiros, M. (2017), "The Perils of Counterfactual Analysis With Integrated Processes," Working Paper, SSRB no. 2894065. [228]

——— (2018), "ArCo: An Artificial Counterfactual Approach for High-Dimensional Panel Time-Series Data," *Journal of Econometrics*, 207, 352–380. [227,228,229,230,231,233,234,237]

Chernozhukov, V., Wuthrich, K., and Zhu, Y. (2018), "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," arXiv no. 1712.09089. [227,229]

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford: Oxford University Press. [238]

Doudchenko, N., and Imbens, G. (2016), "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis," NBER 22791, arXiv no. 1610.07748. [227,228,233,237]

Durlauf, S., and Phillips, P. (1985), "Multiple Time Series Regression With Integrated Processes," *Review of Economic Studies*, 53, 473–495.

Engle, R., and Granger, C. (1987), "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276. [231]

Ferman, B., and Pinto, C. (2016), "Synthetic Controls With Imperfect Pre-Treatment Fit," Working Paper, S ao Paulo School of Economics, FGV. [227,228]

Gardeazabal, J., and Vega-Bayo, A. (2017), "An Empirical Comparison Between the Synthetic Control Method and Hsiao et al.'s Panel Data Approach to Program Evaluation," *Journal of Applied Econometrics*, 32, 983–1002. [228]

Hansen, B. (1992), "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500. [238]

Hsiao, C., Ching, H. S., and Wan, S. K. (2012), "A Panel Data Approach for Program Evaluation: Measuring the Benefits of Political and Economic Integration of Hong Kong With Mainland China," *Journal of Applied Econometrics*, 27, 705–740. [227,228,229,237]

Ibragimov, I. A. (1962), "Some Limit Theorems for Stationary Processes," *Theory of Probability & Its Applications*, 7, 349–382.

Ibragimov, I. A., and Linnik, V. (1971), *Independent and Stationary Sequences of Random Variables*, Wolters–Noordhoff Series of Monographs and Textbooks on Pure and Applied Mathematics.

Kurtz, T., and Protter, P. (1991), "Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations," *The Annals of Probability*, 19, 1035–1070. [238]

Li, K. (2017), "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods," Working Paper, The Wharton School, The University of Pennsylvania. [227]

——— (2018), "Inference for Factor Models Based Average Treatment Effects," Working Paper, The Wharton School, The University of Pennsylvania. [228]

Li, K., and Bell, D. (2017), "Estimation of Average Treatment Effects With Panel Data: Asymptotic Theory and Implementation," *Journal of Econometrics*, 197, 65–75. [227,228]

Masini, R., and Medeiros, M. (2019), "Counterfactual Analysis With Artificial Controls: Inference, High Dimensions and Nonstationarity," Working Paper, SSRN no. 3303308. [228,232,233]

Ouyang, M., and Peng, Y. (2015), "The Treatment-Effect Estimation: A Case Study of the 2008 Economic Stimulus Package of China," *Journal of Econometrics*, 188, 545–557. [227]

Phillips, P. (1986a), "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–340. [228]

——— (1986b), "Weak Convergence to the Matrix Stochastic Integral BdB," Cowles Foundation Discussion Papers 796, Cowles Foundation for Research in Economics, Yale University.

——— (1991), "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

Resnick, S. (1999), *A Probability Path*, Boston, MA: Birkhäuser Boston.

Sims, C., Stock, J., and Watson, M. (1990), "Inference in Linear Time Series Models With Some Unit Roots," *Econometrica*, 58, 113–144. [231]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the LASSO," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [233]