TEXTO PARA DISCUSSÃO

No. 654

The Perils of Counterfactual Analysis with Integrated Processes

Carlos Carvalho
Ricardo Masini
Marcelo Medeiros

DEPARTAMENTO DE ECONOMIA
www.econ.puc-rio.br

The Perils of Counterfactual Analysis with Integrated Processes

December 27, 2016

Carlos V. Carvalho[1]
Central Bank of Brazil and
Department of Economics, Pontifical Catholic University of Rio de Janeiro
E-mail: cvianac@econ.puc-rio.br

Ricardo Masini
Sao Paulo School of Economics, Getulio Vargas Foundation
E-mail: ricardo.masini@fgv.br

Marcelo C. Medeiros
Department of Economics, Pontifical Catholic University of Rio de Janeiro
E-mail: mcm@econ.puc-rio.br

Abstract

Recently, there has been a growing interest in developing econometric tools to conduct counterfactual analysis with aggregate data when a "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 counterfactual from a pool of "untreated" peers, organized in a panel data structure. In this paper, we investigate the consequences of applying such methodologies when the data are formed by integrated process of order 1. We find that *without* a cointegration relation (spurious case) the intervention estimator diverges resulting in the rejection of the hypothesis of no intervention effect regardless of its existence. Whereas, for the case when at least one cointegration relation exists, we have a $\sqrt{T}$-consistent estimator for the intervention effect *albeit* with a non-standard distribution. However, even in this case, the test of no intervention effect is extremely oversized if nonstationarity is ignored. When a drift is present in the data generating processes, the estimator for both cases (cointegrated and spurious) either diverges or is not well defined asymptotically. As a final recommendation we suggest to work in first-differences to avoid spurious results.

JEL Codes: C22, C23, C32, C33.

Keywords: counterfactual analysis, comparative studies, panel data, ArCo, synthetic control, policy evaluation, intervention, cointegration, factor models, spurious regression, nonstationarity.

[1]The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Central Bank of Brazil.

1 Introduction

The goal of this paper is to investigate the consequences of applying popular econometric methods for counterfactual analysis when the data are non-stationary. The econometric framework considered in the paper nests the panel factor (PF) method by Hsiao, Ching, and Wan (2012), the Artificial Counterfactual (ArCo) of Carvalho, Masini, and Medeiros (2016), and extensions of the Synthetic Control (SC) originally proposed by Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010) as discussed in Doudchenko and Imbens (2016). Most of the literature on counterfactual analysis for panel data do not take into account the possibility of non-stationarity. For example, in the SC setting, the econometric estimation is viewed as a cross-section problem and the time-series nature of the data is ignored. On the other hand, the PF and the ArCo methods explicitly explore the time dimension. However, Carvalho, Masini, and Medeiros (2016) assume that the data is stationary while Hsiao, Ching, and Wan (2012) conjecture that if the data are cointegrated, their results will still hold. This conjecture is wrong as we demonstrate in this paper.

Over the last few years, there has been a growing interest in the literature in developing econometric tools to conduct counterfactual analysis with aggregate data when a "treated" unit suffers an intervention, such as a policy change, and there is not a clear control group available. In these situations, the proposed solution is to construct an artificial counterfactual from a pool of "untreated" peers ("donors pool"). For example, in Hsiao, Ching, and Wan (2012) the counterfactual for the treated variable of interest is constructed from a linear combination of observed variables from selected peers given by the conditional expectation model. In the SC framework, the counterfactual variable is build 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 and is inspired by the matching literature. Although, the above methods seem similar they differ remarkably in the way the linear combination of peers is constructed. For instance, in the SC method the weights in the linear combination of peers are positive and sum to one. On the other hand, Hsiao, Ching, and Wan (2012) do not impose any restrictions. The SC method is now a key ingredient in the toolbox of applied researchers interested in policy evaluation; see, for a example, Athey and Imbens (2016) for a recent review.

More recently, there has been several extensions of the above methods being proposed in the literature. Ouyang and Peng (2015) extended the PF method by relaxing the linear conditional expectation assumption and introducing a semi-parametric estimator to construct the artificial counterfactual. Du and Zhang (2015) and Gao, Long, and Wang (2015) made improvements on the selection mechanism for the constituents of the donors pool in the PF method. Fujiki and Hsiao (2015) considered the case of multiple treatments. Carvalho, Masini, and Medeiros (2016), proposed the ArCo, which is a major extension of the PF method and considered, as well, the case of high-dimensional data. Finally, the SC method has been generalized by Xu (2015) and Doudchenko and Imbens (2016). As in Hsiao, Ching, and Wan (2012) and Carvalho, Masini, and Medeiros (2016), Doudchenko and Imbens (2016) relax the restrictions on the weights of the SC method and, similarly to Carvalho, Masini, and Medeiros (2016) advocate the use of shrinkage methods to estimate the linear combination of peers.

As far as we know this is the first paper to give a full treatment of counterfactual methods when the data is nonstationary. One key exception is Bai, Li, and Ouyang (2014) where the authors show, under some assumptions, consistency of the panel approach when the data are integrated of first order. However, the paper does not provide the asymptotic distribution of the estimator nor propose a test of hypothesis.

We consider two very distinct scenarios: (i) the cointegrated case, where there is at least one cointegration relation among the units and; (ii) the spurious case, where no integration relation exists. We show that in the first case we have a consistent, but not asymptotically normal estimator for the average intervention effect. The distribution of the test for the null of no effect is nonstandard even when the distribution of the estimator for the cointegration vector is mixed normal. This leads to strong over-rejection of the null hypothesis when the non-stationary nature of the data is ignored. Case (ii) - the spurious case - is even more troublesome. We demonstrate that the treatment effect estimator diverges. The lack of cointegration relation makes the construction of the artificial control using the pre-intervention period invalid, due to harmless effects from spurious regressions as discussed in Phillips (1986a). As a consequence, the t-statistic for the null of no effect diverges. The solution is consider the data in first-differences.

A detailed simulation study corroborates our theoretical findings and evaluates the asymptotic approximation in finite samples. We consider samples as small as 100 observations. 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 shrinkage estimators as in Carvalho, Masini, and Medeiros (2016) and Doudchenko and Imbens (2016). As expected none of these approaches mitigate the harmful effects of nonstationarity.

The rest of the paper 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 4. The simulation experiment is shown in Section 5. Section 6 concludes the paper. Finally, all proofs and figures are presented in the Appendix.

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 paper 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. As a result, we can express the observed y_{1t} as

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

where

$$\mathcal{D}_t = \begin{cases} 1 & \text{if } t > T_0 \\ 0 & \text{otherwise,} \end{cases}$$

and $y_{1t}^{(1)}$ 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 ultimately concerned in testing hypothesis on the potential effects of the intervention in the unit of interest (unit 1) for the post-intervention period. In particular, we consider interventions of the form

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

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

The null hypothesis becomes

$$\mathcal{H}_0 : \Delta_T = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \delta_t = 0. \tag{2}$$

The quantity Δ_T in (2) is quite similar to the traditional *average treatment effect on the treated* (ATET) vastly discussed in the literature. It is clear that $y_{1t}^{(0)}$ is not observed from $t = T_0 + 1$ onwards. For this reason, we call thereafter the *counterfactual*, i.e., what would y_{1t} have been like had there been no intervention (potential outcome).

In order to construct the counterfactual let $\boldsymbol{y}_{0t} \equiv (y_{2t}, \ldots, y_{nt})'$ be the collection of all untreated variables.[2] Panel based methods, such as the PF and ArCo methodologies, as well as the SC extensions discussed in Doudchenko and Imbens (2016), construct an artificial counterfactual by considering the following model in the absence of an intervention:

$$y_{1t}^{(0)} = \mathcal{M}(\boldsymbol{y}_{0t}) + \nu_t, \quad t = 1, \ldots, T, \tag{3}$$

where $\mathcal{M} : \mathcal{Y}_0 \times \boldsymbol{\Theta} \to \mathbb{R}$ measurable mapping index by the $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The main idea is to estimate (3) using just the pre-intervention sample ($t = 1, \ldots, T_0$), since in this case $y_{1t}^{(0)} = y_{1t}$. Consequently, the estimated counterfactual is given as:

$$\widehat{y}_{1t}^{(0)} = \begin{cases} y_{1t} & \text{if } t = 1, \ldots, T_0, \\ \widehat{\mathcal{M}}(\boldsymbol{y}_{0t}) & \text{if } t = T_0 + 1, \ldots, T, \end{cases} \tag{4}$$

where $\widehat{\mathcal{M}}(\cdot) \equiv \mathcal{M}(\cdot; \widehat{\boldsymbol{\theta}})$.

Let $\boldsymbol{y}_t^{(0)} \equiv \left(y_{1t}^{(0)}, \boldsymbol{y}_{0t}^{(0)\prime}\right)'$ denote all the units in the absence of the intervention. Under stationarity of $\boldsymbol{y}_t^{(0)}$ and additional mild assumptions, Hsiao, Ching, and Wan (2012) and Carvalho, Masini, and Medeiros (2016) show that $\widehat{\delta}_t \equiv y_t - \widehat{y}_t^{(0)}$ is an unbiased estimator for $\delta_t, t = T_0 + 1, \ldots, T$ as the pre-intervention sample size grows to infinity and

$$\widehat{\Delta} = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \widehat{\delta}_t, \tag{5}$$

is $\sqrt{T}$-consistent for Δ_T and asymptotically normal.

2.2 Non-stationarity

Suppose now that $\{\boldsymbol{y}_t^{(0)}\}$ is integrated process of order 1, $\mathcal{I}(1)$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and assume for notational convenience that:[3]

$$\begin{cases} \boldsymbol{y}_t^{(0)} & = \boldsymbol{y}_{t-1}^{(0)} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \quad t \geq 1 \\ \boldsymbol{y}_0^{(0)} & = \boldsymbol{0}, \end{cases} \tag{6}$$

[2] 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 (2016) for more general specifications.

[3] We assume $\boldsymbol{y}_0^{(0)} = \boldsymbol{0}$ without loss of generality. We could either assume $\boldsymbol{y}_0^{(0)}$ to be a any constant or even a random vector with a specific distribution. In that case a constant regressor must appear in both specification but the results will be unaffected.

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is a drift and $\boldsymbol{\varepsilon_t}$ is a zero mean covariance stationary process (Assumption 2) with a Wold Representation given by $\boldsymbol{C}(L)\boldsymbol{v_t}$. L denotes the lag operator, $\boldsymbol{C}(L)$ is a $(n \times n)$ matrix polynomial with $\boldsymbol{C}(0) = \boldsymbol{I}_n$, and $\boldsymbol{v}_t$ is a white noise vector such that

$$\mathbb{E}(\boldsymbol{v}_t\boldsymbol{v}_s') = \begin{cases} \boldsymbol{\Lambda}, & \text{if } t = s, \\ 0, & \text{otherwise,} \end{cases}$$

where $\boldsymbol{\Lambda}$ is a positive definite symmetric covariance matrix.

3 Theoretical Results

3.1 Notation and Definitions

For any zero mean vector process $\{\boldsymbol{v}_t\}$ define on a common probability space, we define the following matrices:

$$\begin{aligned}
\boldsymbol{\Omega}_0(\boldsymbol{v}) &\equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\boldsymbol{v}_t\boldsymbol{v}_t'), \\
\boldsymbol{\Omega}_j(\boldsymbol{v}) &\equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-j} \mathbb{E}(\boldsymbol{v}_s\boldsymbol{v}_t'), \quad j = 1, 2, \ldots \\
\boldsymbol{\Omega}(\boldsymbol{v}) &\equiv \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left(\sum_{t=1}^{T}\boldsymbol{v}_t\sum_{t=1}^{T}\boldsymbol{v}_t'\right),
\end{aligned}$$

if the limits exist.

Also for any pair $(\lambda, \lambda') \in [0,1]^2$ with $\lambda < \lambda'$ and $\varrho = 0, 1, ...$, we define the following $(n \times n)$ random matrices:

$$\begin{aligned}
\boldsymbol{A}(\varrho, \lambda, \lambda', \boldsymbol{v}) &\equiv \boldsymbol{\Omega}(\boldsymbol{v})^{\frac{1}{2}}\left[\int_{\lambda}^{\lambda'} \boldsymbol{W}\boldsymbol{W}'\mathrm{d}r - \vartheta(\varrho, \lambda, \lambda')\int_{\lambda}^{\lambda'} r^{\varrho}\boldsymbol{W}\mathrm{d}r\int_{\lambda}^{\lambda'} r^{\varrho}\boldsymbol{W}'\mathrm{d}r\right]\boldsymbol{\Omega}(\boldsymbol{v})^{\frac{1}{2}} \\
\boldsymbol{B}(\varrho, \lambda, \lambda', \boldsymbol{v}) &\equiv \boldsymbol{\Omega}(\boldsymbol{v})^{\frac{1}{2}}\left[\int_{\lambda}^{\lambda'} \boldsymbol{W}\mathrm{d}\boldsymbol{W}' - \vartheta(\varrho, \lambda, \lambda')\int_{\lambda}^{\lambda'} r^{\varrho}\boldsymbol{W}\mathrm{d}r\int_{\lambda}^{\lambda'} r^{\varrho}\mathrm{d}\boldsymbol{W}\right]\boldsymbol{\Omega}(\boldsymbol{v})^{\frac{1}{2}} \\
&\quad + (\lambda' - \lambda)\left[\boldsymbol{\Omega}_1(\boldsymbol{v}) + \boldsymbol{\Omega}_0(\boldsymbol{v})\right]
\end{aligned}$$

and $(n \times 1)$ random vectors:

$$\begin{aligned}
\boldsymbol{a}(\varrho, \lambda, \lambda', \boldsymbol{v}) &\equiv \vartheta(\varrho, \lambda, \lambda')\boldsymbol{\Omega}(\boldsymbol{v})^{\frac{1}{2}}\int_{\lambda}^{\lambda'} r^{\varrho}\boldsymbol{W}\mathrm{d}r \\
\boldsymbol{b}(\varrho, \lambda, \lambda', \boldsymbol{v}) &\equiv \vartheta(\varrho, \lambda, \lambda')\boldsymbol{\Omega}(\boldsymbol{v})^{\frac{1}{2}}\int_{\lambda}^{\lambda'} r^{\varrho}\mathrm{d}\boldsymbol{W},
\end{aligned}$$

where

$$\vartheta(\varrho, \lambda, \lambda') \equiv \frac{1 + 2\varrho}{\lambda'^{1+2\varrho} - \lambda^{1+2\varrho}},$$

and $\boldsymbol{W} \equiv \boldsymbol{W}(r), r \in [0,1]$, denotes a standard vector Wiener process on $[0,1]^n$.

Finally, for any given (random) matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ and (random) vector $\boldsymbol{m} \in \mathbb{R}^n$ we use the following partition scheme:

$$\boldsymbol{M} = \begin{matrix} & \begin{matrix} 1 & n-1 \end{matrix} \\ \begin{matrix} 1 \\ n-1 \end{matrix} & \begin{pmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{10} \\ \boldsymbol{M}_{01} & \boldsymbol{M}_{00} \end{pmatrix} \end{matrix} \qquad \text{and} \qquad \boldsymbol{m} = \begin{matrix} 1 \\ n-1 \end{matrix}\begin{pmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_0 \end{pmatrix}.$$

Whenever the vector or matrix has additional arguments, we use the alternative, but equivalent, notation $[\boldsymbol{M}]_{10} \equiv \boldsymbol{M}_{10}$, or $[\boldsymbol{m}]_1 \equiv \boldsymbol{m}_1$ to make the same partition scheme above without polluting the notation.

All the summations are from period 1 to T whenever the limits are left unspecified. We establish the asymptotic properties of the estimator by considering the whole sample increasing, while the proportion between the pre-intervention to the post-intervention sample size is constant. For convenience set $T_2 \equiv T - T_0$ as the number post intervention periods. Recall that $T_0 = \lfloor \lambda_0 T \rfloor$. Hence, for fixed $\lambda_0 \in (0,1)$ we have $T_0 \equiv T_0(T)$. Consequently, $T_2 \equiv T_2(T)$. All the asymptotics are taken as $T \to \infty$. We denote convergence in probability and weak converge by "$\xrightarrow{p}$" and "$\Rightarrow$", respectively.

3.2 Main Assumptions

In order to recover the effects of the intervention we need the following key assumption.

Assumption 1. $\mathbb{E}(\mathcal{M}_t|\mathcal{D}_t) = \mathbb{E}(\mathcal{M}_t)$, *where* $\mathcal{M}_t \equiv \mathcal{M}(\boldsymbol{y}_{0t})$.

Roughly speaking, the assumption above is sufficient for the peers to be unaffected by intervention on the unit of interest, i.e., the peers are actually untreated[4]. It is worth mentioning that since we allow $\mathbb{E}(\boldsymbol{y}_{1t}|\mathcal{D}_t) \neq \mathbb{E}(\boldsymbol{y}_{1t})$ we might have some sort of selection on observables and/or non-observables belonging to the treated unit. Of course, selection on features of the untreated units is ruled out by Assumption 1.

Assumption 2. *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* $i = 1, \dots, n$ *and some* $2 \leq \xi < \infty$*;*

(c) $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ *is mixing with mixing coefficients satisfying either* $\sum_{m=1}^{\infty} \alpha_m^{1-1/\xi} < \infty$ *or* $\sum_{m=1}^{\infty} \phi_m^{1-2/\xi} < \infty$.

Assumption 2 state general conditions under which the multivariate invariance principle is valid for the process $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$. Assumption 2(a) limits the heterogeneity in the process (at least up to the second moment). Assumption 2(b) is just a standard higher moment existence condition for all the n coordinates of the random vector which guarantees, along with Assumption 2(c), bounded covariances. Finally, 2(c) restrains the temporal dependence requiring the sequence to be either strong mixing with size $-\frac{\xi}{\xi-2}$ or uniform mixing with size $-\frac{\xi}{2\xi-2}$.

The following result is well-known and it will be stated here just for the sake of clarity of the developments in the forthcoming sections.

Proposition 1. *Let* $\boldsymbol{S}_t = \sum_{j=1}^{t} \boldsymbol{z}_j$ *be the partial sum of the sequence* $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ *of* $(n \times 1)$ *random vectors. Then, under Assumption 2,*

(a) $\boldsymbol{\Sigma} = \lim_{T\to\infty} T^{-1}\mathbb{E}(\boldsymbol{S}_T\boldsymbol{S}_T')$ *exists and is positive definite,*

(b) $\boldsymbol{Z}_T(r) \equiv T^{-1/2}\boldsymbol{S}_{[rT]} \Rightarrow \boldsymbol{\Sigma}^{1/2}\boldsymbol{W}(r)$,

where $[\cdot]$ *denotes the integer part and* $\boldsymbol{W}(\cdot)$ *is a vector Wiener process on* $[0,1]^n$.

[4] For a throughout discussion on Assumption 1, including the potential bias resulting from its failure in the stationary setup refer to Carvalho, Masini, and Medeiros (2016).

The implied convergence in Proposition 1(a) is a direct consequence of the stationarity assumption together with the mixing condition as shown by Ibragimov and Linnik (1971). Finally, Proposition 1(b) is a multivariate generalization of the univariate invariance principle (Durlauf and Phillips, 1985).

Let r denotes the rank of $\boldsymbol{C}(1)$. As shown in Engle and Granger (1987), a necessary condition for $\boldsymbol{y}_t^{(0)}$ to have $r \in \{1,\ldots,n-1\}$ cointegration relations is that the rank of $\boldsymbol{C}(1)$ be $n-r$, i.e., rank deficient. When $r=0$ there is no cointegration and when $r=n$ the vector $\boldsymbol{y}_t^{(0)}$ is stationary in levels. Therefore, we consider datasets that are generated, in the absence of a intervention, either by a cointegrated system of order 1 or that are just a collection of unrelated $\mathcal{I}(1)$ processes.

3.3 The Cointegrated Case

If we have r cointegration relations, then there exists a $(n \times r)$ matrix $\boldsymbol{\Gamma}$ with rank r such that $\boldsymbol{\Gamma}'(\boldsymbol{y}_t^{(0)} - t\boldsymbol{\mu})$ is $\mathcal{I}(0)$, where. Since every linear combination of the columns of $\boldsymbol{\Gamma}$ is also a cointegration vector for $\boldsymbol{y}_t^{(0)}$. We can define $(1, -\boldsymbol{\beta}_0')' = \boldsymbol{\Gamma}\boldsymbol{\chi}$ for some $\boldsymbol{\chi} \neq \boldsymbol{0} \in \mathbb{R}^r$ such that $(1, -\boldsymbol{\beta}_0')(\boldsymbol{y}_t^{(0)} - t\boldsymbol{\mu}) \equiv \nu_t \sim \mathcal{I}(0)$. Note that even after the normalization of the first element the resulting linear combination is not the only possible stationary process (unless $r=1$). However, as we will show below, the least squares procedure will give consistent estimators for the combination that give the stationary process with the smallest variance.

Therefore, the "cointegrated regression" can be written as

$$y_{1t}^{(0)} = \gamma_0 t + \boldsymbol{\beta}_0' \boldsymbol{y}_{0t}^{(0)} + \nu_t, \qquad \text{for } t \geq 1$$

where $\gamma_0 \equiv \mu_1 - \boldsymbol{\beta}_0'\boldsymbol{\mu}_0$.

Since for the pre-intervention period, $t = 1, \ldots, T_0$, we have the observable $\boldsymbol{y}_t = \boldsymbol{y}_t^{(0)}$ we can use the pre-intervention sample to estimate the unknown parameters. We will consider two distinct specifications for the pre-intervention period: (i) the correct specification with a time trend included and (ii) the misspecified case with no time trend.

$$y_{1t} = \gamma_0 t + \boldsymbol{\beta}_0' \boldsymbol{y}_{0t} + \nu_t \tag{7}$$
$$y_{1t} = \alpha_0 + \boldsymbol{\pi}_0' \boldsymbol{y}_{0t} + \zeta_t \tag{8}$$

Clearly, $\alpha_0 = 0$ and $\zeta_t = \nu_t + \gamma_0 t$. Thus, ζ_t is non-stationary unless $\gamma_0 = 0$.

We can apply the results of the Lemma 5 together with the continuous mapping theorem to show the following convergence in distribution:

Lemma 1. *Let the process* $\{\boldsymbol{y}_t^{(0)}\}$ *be defined by* (6) *have at least one cointegration relation* $(0 < r < n)$. *Also assume that* $\{\boldsymbol{\eta}_t \equiv (\nu_t, \boldsymbol{\varepsilon}_0')'\}$ *satisfies Assumption 2. Then, for the least squares estimator of the parameters appearing in* (7)–(8) *using only the pre intervention sample* $(t = 1, \ldots, T_0)$, *as* $T \to \infty$:

(a) For $\boldsymbol{\mu} = 0$,

$$\begin{aligned}
T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) &\Rightarrow \boldsymbol{P}_{00}^{-1}\boldsymbol{Q}_{01} \equiv \widetilde{\boldsymbol{h}} \\
T^{3/2}\left(\widehat{\gamma} - \gamma_0\right) &\Rightarrow \left[\boldsymbol{b}(1,0,\lambda_0,\boldsymbol{\eta})\right]_1 - \widetilde{\boldsymbol{h}}'\left[\boldsymbol{a}(1,0,\lambda_0,\boldsymbol{\eta})\right]_0 \\
T\left(\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0\right) &\Rightarrow \boldsymbol{R}_{00}^{-1}\boldsymbol{V}_{01} \equiv \widetilde{\boldsymbol{p}} \\
\sqrt{T}\left(\widehat{\alpha} - \alpha_0\right) &\Rightarrow \left[\boldsymbol{b}(0,0,\lambda_0,\boldsymbol{\eta})\right]_1 - \widetilde{\boldsymbol{p}}'\left[\boldsymbol{a}(0,0,\lambda_0,\boldsymbol{\eta})\right]_0.
\end{aligned}$$

(b) For $\mu_0 \neq 0$ and $n = 2$,

$$\widehat{\pi} - \beta_0 \xrightarrow{p} \frac{\gamma_0}{\mu_0}; \qquad \frac{1}{T}(\widehat{\alpha} - \alpha_0) \xrightarrow{p} 0.$$

In case of $\boldsymbol{\mu} \neq 0$ for either, specification (7) or $n > 2$, the least squares estimators are not defined asymptotically.

$\boldsymbol{P} \equiv \boldsymbol{A}(1, 0, \lambda_0, \boldsymbol{\eta})$; $\boldsymbol{R} \equiv \boldsymbol{A}(0, 0, \lambda_0, \boldsymbol{\eta})$; $\boldsymbol{Q} \equiv \boldsymbol{B}(1, 0, \lambda_0, \boldsymbol{\eta})$; $\boldsymbol{V} \equiv \boldsymbol{B}(0, 0, \lambda_0, \boldsymbol{\eta})$ as defined in Section 3.1.

Remark 1. *Whenever there is a drift among the peers and $n > 2$ we have a multicollinearity issue in the least squares estimators, since the drift component dominates the other terms asymptotically. In case of specification (7), since we are fitting the trend term $t\gamma$, the multicollinearity appears even for $n = 2$ (only one control). Note that, for specification (8), if we replace γ_0 by its definition $\mu_1 - \beta_0\mu_0$, then, as expected, $\widehat{\pi} \xrightarrow{p} \frac{\mu_1}{\mu_0}$.*

Remark 2. *In fact the estimators (5) is of little usage whenever we expect to have integrated process with drift. Not only the estimator is not well in large samples, but a simple fitted trend regressor makes a reasonable counterfactual for the unit of interest. Therefore we treat for now on only the the case without drift ($\boldsymbol{\mu} = \mathbf{0}$).*

Similar results to Lemma 1(a) appear in Durlauf and Phillips (1985).

We now consider the estimation for the intervention effect in two specifications described above: (i) The true model as in (7); and (ii) a model that would naturally arise if we choose to ignore (or be unaware of) the non-stationarity in the data. As shown above, the distribution of the regression estimators is dependent on the presence of a drift term. As a consequence, the intervention effect estimator is defined, for each specification $j = \{1, 2\}$, as:

$$\widehat{\Delta}_j = \frac{1}{T_2} \sum_{t > T_0} y_{1t} - \widehat{y}_{1t}^{(j)} \quad \text{where } \widehat{y}_{1t}^{(j)} = \begin{cases} \widehat{\gamma} t + \widehat{\boldsymbol{\beta}}' \boldsymbol{y}_{0t} & \text{if } j = 1 \\ \widehat{\alpha} + \widehat{\boldsymbol{\pi}}' \boldsymbol{y}_{0t} & \text{if } j = 2 \end{cases} \tag{9}$$

where $\widehat{\gamma}, \widehat{\boldsymbol{\beta}}, \widehat{\alpha}$ and $\widehat{\boldsymbol{\pi}}$ are the least squares estimators of the parameters appearing in (7)–(8) using only pre-intervention sample.

Theorem 1. *Under Assumption 1, let the process $\{\boldsymbol{y}_t^{(0)}\}$ be defined by (6) have at least one cointegration relation ($0 < r < n$). Also assume that $\{\boldsymbol{\eta}_t \equiv (\nu_t, \boldsymbol{\varepsilon}_0')'\}$ satisfies Assumption 2. Then, for the estimators defined in (9) as $T \to \infty$:*

$$\sqrt{T}\left(\widehat{\Delta}_1 - \Delta_T\right) \Rightarrow \boldsymbol{h}'\boldsymbol{c},$$
$$\sqrt{T}\left(\widehat{\Delta}_2 - \Delta_T\right) \Rightarrow \boldsymbol{p}'\boldsymbol{d},$$

where the $(n \times 1)$ random vectors above are defined as:

$$\boldsymbol{c} \equiv \boldsymbol{c}(\lambda_0) = \begin{pmatrix} \left[\boldsymbol{b}(0, \lambda_0, 1, \boldsymbol{\eta}) - \frac{1+\lambda_0}{2}\boldsymbol{b}(1, 0, \lambda_0, \boldsymbol{\eta})\right]_1 \\ \left[\boldsymbol{a}(0, \lambda_0, 1, \boldsymbol{\eta}) - \frac{1+\lambda_0}{2}\boldsymbol{a}(1, 0, \lambda_0, \boldsymbol{\eta})\right]_0 \end{pmatrix}$$
$$\boldsymbol{d} \equiv \boldsymbol{d}(\lambda_0) = \begin{pmatrix} \left[\boldsymbol{b}(0, \lambda_0, 1, \boldsymbol{\eta}) - \boldsymbol{b}(0, 0, \lambda_0, \boldsymbol{\eta})\right]_1 \\ \left[\boldsymbol{a}(0, \lambda_0, 1, \boldsymbol{\eta}) - \boldsymbol{a}(0, 0, \lambda_0, \boldsymbol{\eta})\right]_0 \end{pmatrix},$$

with $\boldsymbol{a}(\cdot, \cdot, \cdot, \cdot)$, $\boldsymbol{b}(\cdot, \cdot, \cdot, \cdot)$ defined in Section 3.1, $\boldsymbol{h} \equiv (1, -\widetilde{\boldsymbol{h}}')'$, $\boldsymbol{p} \equiv (1, -\widetilde{\boldsymbol{p}}')'$, with $\widetilde{\boldsymbol{h}}$ and $\widetilde{\boldsymbol{p}}$ as defined in Lemma 1.

Therefore, both estimators above are $\sqrt{T}$-consistent for Δ, however with a non-standard limiting distribution. Even though the results above rule out common inference procedures, in Section 4 we investigate the results of using a conventional t-statistic.

3.4 The Spurious Case

We now turn to the case where no cointegration relation exists among $\boldsymbol{y}_t$ prior to the intervention, hence $\boldsymbol{C}(1)$ is full rank. We consider for the pre-intervention period the same specification, (7) and (8), that were used in the cointegrated case. However, since the "true parameters" no longer exist[5], we cannot express least-squares estimators as differences from their "true parameters". Hence we have the following result:

Lemma 2. *Let the process $\{\boldsymbol{y}_t^{(0)}\}$ be defined by (6) have no cointegration relation ($r = 0$). Also assume that $\{\boldsymbol{\varepsilon}_t\}$ satisfies Assumption 2. Then, for the least squares estimator of the parameters appearing in (7)–(8), as $T \to \infty$:*

(a) For $\boldsymbol{\mu} = 0$

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &\Rightarrow \boldsymbol{P}_{00}^{-1}\boldsymbol{P}_{01} \equiv \widetilde{\boldsymbol{f}}, \\ \sqrt{T}\widehat{\gamma} &\Rightarrow \boldsymbol{f}'\boldsymbol{a}(1,0,\lambda_0,\boldsymbol{\varepsilon}), \quad \textit{where } \boldsymbol{f} \equiv (1, -\widetilde{\boldsymbol{f}}')' \\ \widehat{\boldsymbol{\pi}} &\Rightarrow \boldsymbol{R}_{00}^{-1}\boldsymbol{R}_{01} \equiv \widetilde{\boldsymbol{g}}, \\ \frac{1}{\sqrt{T}}\widehat{\alpha} &\Rightarrow \boldsymbol{g}'\boldsymbol{a}(0,0,\lambda_0,\boldsymbol{\varepsilon}), \quad \textit{where } \boldsymbol{g} \equiv (1, -\widetilde{\boldsymbol{g}}')' \end{aligned}$$

(b) For $\mu_0 \neq 0$ and $n = 2$,

$$\widehat{\pi} \xrightarrow{p} \frac{\mu_1}{\mu_0}; \qquad \frac{1}{T}\widehat{\alpha} \xrightarrow{p} 0.$$

In case of $\boldsymbol{\mu} \neq 0$ for either, specification (7) or $n > 2$, the least squares estimators are not defined asymptotically.

The random matrices $\boldsymbol{P}, \boldsymbol{R}$ are defined in Lemma 1 but with $\boldsymbol{\eta}$ replaced by $\boldsymbol{\varepsilon}$ and with $\boldsymbol{a}(\cdot,\cdot,\cdot,\cdot)$, $\boldsymbol{b}(\cdot,\cdot,\cdot,\cdot)$ defined in Section 3.1.

The limiting distribution of $\widehat{\boldsymbol{\pi}}$ and $\widehat{\boldsymbol{\alpha}}$ are well known from the spurious regression case discussed in Phillips (1986a). For $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$, the result is analogous but with a different limiting distribution. In both cases, when $r = 0$ and consequently $\boldsymbol{y}_t$ does *not* cointegrate, we have a spurious regression and both $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\pi}}$ converges, as $T \to \infty$, not to a constant but to a functional of a multivariate Brownian motion. While $\widehat{\boldsymbol{\alpha}}$ diverges, $\widehat{\boldsymbol{\gamma}}$ converges to zero (which is the value of the parameter γ_0 when $\boldsymbol{\mu} = 0$).

Once again we consider the scenario where the researcher conduct the estimation using the estimators defined in (9) with $\boldsymbol{y}_t$ in levels.

Theorem 2. *Under Assumption 1, let the process $\{\boldsymbol{y}_t^{(0)}\}$ be defined by (6) have no cointegration relation ($r = 0$). Also assume $\{\boldsymbol{\varepsilon}_t\}$ satisfies Assumption 2. Then for the estimators defined in (9), as $T \to \infty$:*

$$\begin{aligned} \frac{1}{\sqrt{T}}\left(\widehat{\Delta}_1 - \Delta\right) &\Rightarrow \boldsymbol{f}'\boldsymbol{e}, \\ \frac{1}{\sqrt{T}}\left(\widehat{\Delta}_2 - \Delta\right) &\Rightarrow \boldsymbol{g}'\boldsymbol{l}, \end{aligned}$$

where the $(n \times 1)$ random vectors above are defined as:

$$\begin{aligned} \boldsymbol{e} \equiv \boldsymbol{e}(\lambda_0) &= \boldsymbol{a}(0,\lambda_0,1,\boldsymbol{\varepsilon}) - \frac{1+\lambda_0}{2}\boldsymbol{a}(1,0,\lambda_0,\boldsymbol{\varepsilon}), \\ \boldsymbol{l} \equiv \boldsymbol{l}(\lambda_0) &= \boldsymbol{a}(0,\lambda_0,1,\boldsymbol{\varepsilon}) - \boldsymbol{a}(0,0,\lambda_0,\boldsymbol{\varepsilon}), \end{aligned}$$

with $\boldsymbol{a}(\cdot,\cdot,\cdot,\cdot)$, $\boldsymbol{b}(\cdot,\cdot,\cdot,\cdot)$ defined in Section 3.1; Also $\boldsymbol{f}$ and $\boldsymbol{g}$ as defined in Lemma 2.

[5] In the sense that no (linear) combination of the units result in a stationary process.

From the theorem above, it is clear that, unlike in the cointegrated case, $\widehat{\Delta}_j$ diverges as $T \to \infty$ for both specifications. As for the cointegration case we investigate the limiting distribution of a conventional t-statistic in following section.

4 Inference

Given the asymptotic results from the last section we would like to further investigate the consequences of conducting usual inference. In particular, we investigate the limiting distribution of a conventional t-statistic such as

$$\tau_j \equiv \frac{\widehat{\Delta}_j}{\sqrt{\widehat{\mathbb{V}}(\widehat{\Delta}_j)}}, \quad j = \{1, 2\}, \tag{10}$$

where the denominator is supposed to be a an estimator for the standard deviation of $\widehat{\Delta}_j$. For that, define the centred residuals for the post intervention regression period, $t = T_0 + 1, \ldots, T$, as

$$\widehat{\nu}_{1t} = y_{1t} - \widehat{\gamma} t - \widehat{\boldsymbol{\beta}}' \boldsymbol{y}_{t0} - \widehat{\Delta}_1$$
$$\widehat{\nu}_{2t} = y_{1t} - \widehat{\alpha} - \widehat{\boldsymbol{\pi}}' \boldsymbol{y}_{t0} - \widehat{\Delta}_2.$$

Then, for each $j = 1, 2$, we have the following covariance estimators for $\rho_k^2 \equiv \mathbb{E}(\nu_t \nu_{t+k})$, where $k = \{-T + T_0 + 1, \ldots, T - T_0 - 1\}$:

$$\widehat{\rho}_{jk}^2 = \begin{cases} \frac{1}{T-T_0} \sum_{t=T_0+1}^{T-k} \widehat{\nu}_{jt} \widehat{\nu}_{jt+k} & \text{if } k \geq 0, \\ \frac{1}{T-T_0} \sum_{t=T0+1}^{T+k} \widehat{\nu}_{jt} \widehat{\nu}_{jt-k} & \text{if } k < 0. \end{cases}$$

Therefore, for some choice of a kernel function $\phi(\cdot)$ and bandwidth J_T such that $J_T \to \infty$ as $T \to \infty$, we have

$$\widehat{\sigma}_j^2 \equiv \widehat{\sigma}_j^2(J_T) = \sum_{|k|<T} \phi(k/J_T) \widehat{\rho}_{jk}^2. \tag{11}$$

Finally, our estimator for the variance of $\widehat{\Delta}_j$ becomes

$$\widehat{\mathbb{V}}(\widehat{\Delta}_j) \equiv \frac{\widehat{\sigma}_j^2}{T - T_0}.$$

4.1 Inference on the Cointegrated case

Consider now the following stronger version of Assumption 2.

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

(a) *$\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ is zero-mean fourth-order stationary process;*

(b) *$\mathbb{E}|\boldsymbol{z}_1|^{4\xi} < \infty$ and some $\xi > 1$;*

(c) *$\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ is strong mixing with the mixing coefficients such that $\sum_{m=1}^{\infty} m^2 \alpha_m^{1-2/\xi} < \infty$.*

Clearly, Assumption 3 implies Assumption 2. The fourth-order stationarity requirement on $\{\nu_t\}$ translates into weak stationarity of $\{w_t^{(k)} \equiv \nu_t \nu_{t+k}\}$ for any $k \in \mathbb{Z}$. Assumptions 3(a)-(c) are sufficient for Assumption A of Andrews (1991) which translate in the summability of the covariances of $w_t^{(k)}$, i.e.

$$\lim_{T \to \infty} T^{-1} \mathbb{V} \left(\sum_{|k|<T} \sum_{t=1}^{T-|k|} \nu_t \nu_{t+|k|} \right) < \infty.$$

Thus, we have a weak law of large number by Chebyshev's Inequality applied for each k which is result (a) in the following lemma.

Lemma 3. *If the sequence $\{\nu_t\}$ satisfies Assumption 3, then for each $j \in \{1,2\}$,*

(a) $\widehat{\rho}^2_{jk} \xrightarrow{p} \rho^2_k, \quad \forall k.$

If in addition, $\int_{-\infty}^{\infty} |\phi(x)| \mathrm{d}x < \infty$ and $J_T^2/T \to 0$ as $T \to \infty$, then

(b) $|\widehat{\sigma}^2_{jT} - \sum_{|k|<T} \rho^2_k| \xrightarrow{p} 0.$

Lemma 3(b) follows from arguments similar to Newey and West (1987) and Andrews (1991).

Theorem 3. *Under the same conditions of Theorem 1, but with Assumption 2 replaced by 3:*

(a) Under the null $\mathcal{H}_0 : \Delta_T = 0$,

$$\tau_1 \Rightarrow \frac{\sqrt{1-\lambda_0}}{\omega} \boldsymbol{h}'\boldsymbol{c}$$
$$\tau_2 \Rightarrow \frac{\sqrt{1-\lambda_0}}{\omega} \boldsymbol{p}'\boldsymbol{d}$$

(b) Under the alternative, $\mathcal{H}_1 : \Delta_T = \delta \neq 0$, both estimators ($j = 1,2$) diverge as

$$\frac{1}{\sqrt{T}} \tau_j \xrightarrow{p} \sqrt{1-\lambda_0} \frac{\delta}{\omega},$$

where the $(n \times 1)$ vectors above are defined in Theorem 1 and $\omega^2 \equiv \boldsymbol{\Omega}_{11}$.

Remark 3. *Under $\mathcal{H}_0$ we have a $\sqrt{T}$-consistent estimator for the intervention average effect Δ_T albeit with a non-standard asymptotic distribution. In fact, by the presence of the second term we can conclude that we systematically over reject asymptotically.*

Remark 4. *The t-test is also asymptotically consistent as the test statistic diverges under the alternative. Recall that our null hypothesis was defined in (2), hence the natural alternative would be $\Delta_T \neq 0$, but since Δ_T could potentially approach zero arbitrally fast as T grows, we restrict the Δ_T to be a non-zero constant. We get similar results by allowing a more flexible intervention profile as long as it does not approach zero faster than $T^{-1/2}$, for instance, by imposing that $\{\delta_t\}$ is such that $\sqrt{T}\Delta_T \to \infty$.*

4.2 Inference on the Spurious case

Since hypothesis testing is not carried directly on $\widehat{\Delta}_j$, it is useful to derive an expression for the limiting distribution of a common t-stat such as the one considered in the cointegrated case. First we need the following result

Lemma 4. *Consider the same conditions of Theorem 2, but with Assumption 2 replaced by 3. Then, under both $\mathcal{H}_0$ or $\mathcal{H}_1$, as $T \to \infty$:*

(a) $\frac{1}{T}\widehat{\rho}^2_{1k} \Rightarrow \frac{1}{1-\lambda_0} \boldsymbol{f}'\boldsymbol{L}\boldsymbol{f}, \quad \forall k \in \mathbb{Z},$

(b) $\frac{1}{T}\widehat{\rho}^2_{2k} \Rightarrow \frac{1}{1-\lambda_0} \boldsymbol{g}'\boldsymbol{H}\boldsymbol{g}, \quad \forall k \in \mathbb{Z}.$

If in addition, $\int_{-\infty}^{\infty} |\phi(x)|\mathrm{d}x < \infty$ *and* $J_T^2/T \to 0$ *as* $T \to \infty$*, then*

(c) $\frac{1}{J_T T}\widehat{\sigma}_{1T}^2 \Rightarrow \frac{c_\phi}{1-\lambda_0}\boldsymbol{f}'\boldsymbol{L}\boldsymbol{f}$,

(d) $\frac{1}{J_T T}\widehat{\sigma}_{2T}^2 \Rightarrow \frac{c_\phi}{1-\lambda_0}\boldsymbol{g}'\boldsymbol{H}\boldsymbol{g}$,

for $j \in \{1,2\}$*, where* $\boldsymbol{H} \equiv \boldsymbol{A}(0,\lambda_0,1,\boldsymbol{\varepsilon})$, $\boldsymbol{L} \equiv \boldsymbol{H} - 2\boldsymbol{q}\boldsymbol{s}' + \varsigma(\lambda_0)\boldsymbol{s}\boldsymbol{s}'$, $\boldsymbol{q} \equiv \frac{1-\lambda_0^3}{3}\boldsymbol{a}(1,\lambda_0,1,\boldsymbol{\varepsilon}) - \frac{1-\lambda_0^2}{2}\boldsymbol{a}(0,\lambda_0,1,\boldsymbol{\varepsilon})$, $\boldsymbol{s} \equiv \boldsymbol{a}(1,0,\lambda_0,\boldsymbol{\varepsilon})$*;* $\varsigma(\lambda) \equiv \frac{1-\lambda^3}{12} - \frac{\lambda(1-\lambda)}{4}$, $c_\phi \equiv \int_{-\infty}^{\infty}\phi(x)\mathrm{d}x$*;* $\boldsymbol{f}$ *and* $\boldsymbol{g}$ *are defined in Lemma 2, and* $\boldsymbol{a}(\cdot,\cdot,\cdot,\cdot)$*, and* $\boldsymbol{A}(\cdot,\cdot,\cdot,\cdot)$ *defined in Section 3.1.*

Notice that the limiting distribution in (a) and (b) above is independent of k. In fact, it is the same distribution derived in Lemma 1 when we consider $k = 0$. It follows from the fact that the additional term $\sum_{t=1}^{T}\boldsymbol{v}_t\sum_{i=1}^{k}\boldsymbol{\varepsilon}_i'$ is $O_P(T)$. Result (b) for $k = 0$ is similar to the one appering in Phillips (1986a). It turns out it is valid for all fixed k and also for specification (7) albeit with a different limiting distribution. Using a HAC covariance estimator as proposed by Newey and West (1987) and Andrews (1991), we have an even weaker convergence rate as it goes from T^{-1} to $(J_T T)^{-1}$ as stated in Lemma 6(c)-(d).

Now combining Theorem 2 with Lemma 4 together with the continuous mapping theorem we have the following result.

Theorem 4. *Consider the same conditions of Theorem 2, but with Assumption 2 replaced by 3. Then, under both* $\mathcal{H}_0 : \Delta_T = 0$ *and* $\mathcal{H}_1 = \delta \neq 0$*.*

$$\sqrt{\frac{J_T}{T}}\tau_1 \Rightarrow \frac{1-\lambda_0}{\sqrt{c_\phi}}\frac{\boldsymbol{f}'\boldsymbol{e}}{\sqrt{\boldsymbol{f}'\boldsymbol{L}\boldsymbol{f}}}$$
$$\sqrt{\frac{J_T}{T}}\tau_2 \Rightarrow \frac{1-\lambda_0}{\sqrt{c_\phi}}\frac{\boldsymbol{g}'\boldsymbol{l}}{\sqrt{\boldsymbol{g}'\boldsymbol{H}\boldsymbol{g}}},$$

where the $(n \times 1)$ *random vectors* $\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{e}$*, and* $\boldsymbol{l}$ *are defined in Theorem 2; and the* $(n \times n)$ *random matrices* $\boldsymbol{L}$*,* $\boldsymbol{H}$ *and the constant* c_ϕ *are defined in Lemma 4.*

Remark 5. *When conducting a t-test one draws inference on the premises that* $\tau_j \Rightarrow \mathcal{N}(0,1)$ *under* $\mathcal{H}_0$*. However, as Theorem 4 shows,* τ_j *actually diverges under the assumption that* $J_T = o(T^{1/2})$*. Therefore, ignoring the non-stationarity of the data we end up rejecting the null hypothesis too often. In fact, as the sample size increases, the probability of rejection the null approaches 1 regardless of the existence of the treatment.*

Remark 6. *Notice that the result above is not dependent on the choice of the variance estimator bandwidth. If we use simple variance estimator such as* $\widehat{\sigma}_{jT} = \widehat{\rho}_{j0}$ *(for the case of iid data), we still have* $\tau_j = O_P(\sqrt{T})$*. In fact, in this particular case, the t-test diverges in a even faster rate.*

In summary, for the spurious case, we end up rejecting the $\mathcal{H}_0$ regardless of the existence of an intervention effect when panel based methods for counterfactual analysis are applied in levels. The result is similar in spirit of the one found by Phillips (1986a). However, in the spurious regression case we are usually interested in the t-stat related to the β coefficients of the regression. In the present case, the interest lies in average of the error of the predicted model $\widehat{\Delta}_j$.

4.3 First-Difference

A simple alternative approach would be to work with the first difference $\boldsymbol{z}_t \equiv \boldsymbol{y}_t - \boldsymbol{y}_{t-1}$, and have, by definition, a stationary dataset either in the cointegrated case or in the spurious one. Hence:

$$\boldsymbol{z}_t^{(0)} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t$$

The difference would be that for the cointegrated case the covariance matrix of $\boldsymbol{\Gamma} \equiv \mathbb{V}(\boldsymbol{\varepsilon}_t)$ is rank deficient $(n-r)$ and for the spurious case is full rank since $r = 0$. Nevertheless, we can apply the panel-based methodologies for stationary process unaltered. The pre intervention model becomes

$$z_{1t} = \lambda_0 + \boldsymbol{\theta}_0' \boldsymbol{z}_{0t} + \omega_t \qquad t = 2, \ldots, T_0$$

where $\boldsymbol{\theta}_0 = \boldsymbol{\Gamma}_{00}^{-1}\boldsymbol{\Gamma}_{01}$ and $\lambda_0 = \mu_1 - \boldsymbol{\beta}'\boldsymbol{\mu}_0$. For the post -intervention period $t = T_0+2, \ldots T$, we can take the average of the $\widehat{z}_{1t} = \widehat{\lambda} + \widehat{\boldsymbol{\theta}}' \boldsymbol{z}_{0t}$ as the estimator for $\mathbb{E}(z_1) \equiv \mu_1^*$ and construct the following estimator for the difference in the drifts $\Delta\mu = \mu_1 - \mu_1^*$

$$\begin{aligned}
\widehat{\Delta}_F &= \frac{1}{T - T_0 - 1} \sum_{t=T_0+2}^{T} \left(z_{1t} - \widehat{\lambda} - \widehat{\boldsymbol{\theta}}' \boldsymbol{z}_{0t} \right) \\
\widehat{\boldsymbol{\theta}} &= \left(\sum_{t=2}^{T_0} \dot{\boldsymbol{z}}_{0t} \dot{\boldsymbol{z}}_{0t}' \right)^{-1} \sum_{t=2}^{T_0} \dot{\boldsymbol{z}}_{0t} \dot{z}_{1t} \\
\widehat{\lambda} &= \bar{z}_1 - \widehat{\boldsymbol{\theta}} \bar{\boldsymbol{z}}_0.
\end{aligned}$$

From Theorem 1 of Carvalho, Masini, and Medeiros (2016) for the particular case of a linear specification we have

$$\sqrt{T} \frac{\left(\widehat{\Delta}_F - \Delta \right)}{\widehat{\sigma}_F \left[\lambda_0 (1 - \lambda_0) \right]^{-1/2}} \Rightarrow \mathcal{N}(0, 1),$$

where $\widehat{\sigma}_F^2$ is a consistent estimator for $\sigma_F^2 \equiv \lim_{T\to\infty} T^{-1} \mathbb{V}(\sum_{t=1}^{T} \omega_t)$, defined in (11) for the post intervention residuals.

Remark 7. *The approach above also give us $\sqrt{T}$-consistent estimator for the difference in drifts. However, in contrast to the cointegrated estimator, it is asymptotically normal hence more practical for conducting inference.*

Remark 8. *The limiting distribution in first difference is independent of both the prior knowledge of the true values of $\boldsymbol{\mu}$ and the true hypothesis ($\mathcal{H}_0$ or $\mathcal{H}_1$).*

Remark 9. *Working in first difference we avoid a true spurious regression since if the integrated process is truly uncorrelated we will end up having $\widehat{\boldsymbol{\theta}} \approx \mathbf{0}$ for the pre-intervention period.*

5 Simulations

In order to evaluate the asymptotic approximation in finite samples as well as to evidence the harmful effects of neglecting non-stationarity when conducting counterfactual inference, 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 as well as the t-statistics under the null hypothesis of no intervention at $T_0 = T/2$. We consider $T = 100$ and $1,000$, and $n = 5$. The number of Monte Carlo simulations is set to 2,000. For each scenario and different sample sizes, we report the finite sample distributions of $\widehat{\Delta}_j$ as in (9) and τ_j as in (10) for $j = 1, 2$, 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 (2016) and Doudchenko and Imbens (2016). We report only the case where the linear trend is not included in the regression function.

5.1 Cointegration

The data generating process (DGP) is defined as

$$y_{1t} = \sum_{i=2}^{n} y_{it} + u_t, \tag{12}$$

where

$$y_{it} = y_{it-1} + e_{it}, \tag{13}$$

$y_{i0} = 0$, $i = 2, \ldots, n$, and $\{u_t\}_{t=1}^{T}$ and $\{e_{it}\}_{t=1}^{T}$ are sequences of independent and normally distributed zero-mean random variables with unit variance. Furthermore, $\mathbb{E}(u_t e_{is}) = 0$ and $\mathbb{E}(e_{it} e_{js}) = 0$ for all $t = 1, \ldots, T$, $s = 1, \ldots, T$, $i = 1, \ldots, n$, $j = 1, \ldots, n$, and $t \neq s$.

The simulation results are shown in Figures 1–4. Figure 1 shows the empirical versus the theoretical distributions of the scaled coefficients estimates as in Lemma 1. The distribution of $\widehat{\Delta}_j$, $j = 1, 2$, is presented in Figure 2 and is compared to the asymptotic results of Theorem 1. The empirical distribution of the t-statistic (10) is presented in Figure 3 and is as well compared to the asymptotic approximation as in Theorem 3. Finally, Figure 4 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 de 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 5 that restricting the coefficients does not mitigate the over-rejections is nonstationarity is not taken carefully into account.

5.2 Spurious Counterfactual

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

$$y_{it} = y_{it-1} + e_{it}, \tag{14}$$

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

The simulation results for the spurious case are depicted in Figures 6–9. Figure 6 presents the empirical versus the theoretical distributions of the coefficients estimates as in Lemma 2. The distribution of the average intervention effects, $\widehat{\Delta}_j$, $j = 1, 2$, is presented in Figure 7 and is confronted to the asymptotic results of Theorem 2. The empirical distribution of the t-statistic (10) is presented in Figure 8 and is compared to the asymptotic approximation as in Theorem 4. Finally, Figure 9 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 10.

6 Conclusions

In this paper we considered the asymptotic properties of popular counterfactual estimators when the data are of nonstationary. More specifically, our econometric framework encapsulates the panel based methods of Hsiao, Ching, and Wan (2012), the artificial counterfactual approach of Carvalho, Masini, and Medeiros (2016) and the synthetic control and its extensions Abadie and Gardeazabal (2003); Abadie, Diamond, and Hainmueller (2010); Doudchenko and Imbens (2016).

Two cases are considered. In the first case there is at least one cointegration relation in the data while in the second one the data are formed by a set of independent random walks. The results in the paper either show that the estimators diverge or have non-standard asymptotic distributions. We show a strong over-rejection of the null hypothesis of the null of no intervention effect when the non-stationary nature of the data is ignored. Out theoretical results are corroborated by a simulation experiment. The main prescription of the paper is that practitioners should work in first-differences when the data are nonstationary.

Techincal Appendix

In what follows use the auxiliary process $\{\boldsymbol{u}_t\}$ defined as

$$\boldsymbol{u}_t = \boldsymbol{u}_{t-1} + \boldsymbol{\eta}_t, \quad t \geq 1$$
$$\boldsymbol{u}_0 = 0.$$

For $(\lambda, \lambda') \in [0,1]^2$ with $\lambda < \lambda'$, we denote the summation $\sum_{T_\lambda < t \leq T_{\lambda'}}$ by $\sum_{(\lambda,\lambda']}$ where $T_\lambda = \lfloor \lambda T \rfloor$ and we adopt the following notation for $\varrho = 0, 1, \ldots$:

$$\sum_{(\lambda,\lambda']} \dot{\boldsymbol{u}}_t(\varrho) \equiv \sum_{(\lambda,\lambda']} \left[\boldsymbol{u}_t - t^\varrho \bar{\boldsymbol{u}}(\varrho)\right], \qquad \bar{\boldsymbol{u}}(\varrho) \equiv \left(\sum_{(\lambda,\lambda']} t^{2\varrho} \right)^{-1} \sum_{(\lambda,\lambda']} t^\varrho \boldsymbol{u}_t. \tag{15}$$

Notice that $\bar{\boldsymbol{u}}(0)$ is a simple average. Hence, $\dot{\boldsymbol{u}}_t(0)$ is the deviation from the mean.

First we state and prove in the Lemma below several convergence results that will be applied the the subsequent proofs. Some of these results have been shown elsewhere; see for instance Durlauf and Phillips (1985) and Phillips (1986a,b).

Lemma 5. *Let the sequence $\{\boldsymbol{u}_t\}_{t=1}^T$ be defined as above. If the process $\{\boldsymbol{\eta}_t\}$ satisfies Assumption 2, then as $T \to \infty$:*

(a) $\frac{1}{T^{1/2+\varrho}} \sum_{(\lambda,\lambda']} t^\varrho \boldsymbol{\eta}_t \Rightarrow \boldsymbol{\Omega}^{1/2} \int_\lambda^{\lambda'} r^\varrho \mathrm{d}\boldsymbol{W}$

(b) $\frac{1}{T^{3/2+\varrho}} \sum_{(\lambda,\lambda']} t^\varrho \boldsymbol{u}_t \Rightarrow \boldsymbol{\Omega}^{1/2} \int_\lambda^{\lambda'} r^\varrho \boldsymbol{W} \mathrm{d}r$

(c) $\frac{1}{T} \sum_{(\lambda,\lambda']} \boldsymbol{u}_t \boldsymbol{\eta}'_{t+j} \Rightarrow \boldsymbol{\Omega}^{1/2} \left[\int_\lambda^{\lambda'} \boldsymbol{W} \mathrm{d}\boldsymbol{W}\right] \boldsymbol{\Omega}^{1/2} + (\lambda' - \lambda)\boldsymbol{\Omega}_j, \quad j > 0$

(d) $\frac{1}{T^2} \sum_{(\lambda,\lambda']} \boldsymbol{u}_t \boldsymbol{u}'_{t+k} \Rightarrow \boldsymbol{\Omega}^{1/2} \left[\int_\lambda^{\lambda'} \boldsymbol{W}\boldsymbol{W}' \mathrm{d}r\right] \boldsymbol{\Omega}^{1/2}, \quad k \in \mathbb{Z}$

(e) $T^{\varrho - 1/2} \bar{\boldsymbol{u}}(\varrho, \lambda, \lambda') \Rightarrow \left(\frac{1+2\varrho}{\lambda'^{1+2\varrho} - \lambda^{1+2\varrho}}\right) \boldsymbol{\Omega}^{1/2} \int_\lambda^{\lambda'} r^\varrho \boldsymbol{W} \mathrm{d}r \equiv \boldsymbol{a}(\varrho, \lambda, \lambda', \boldsymbol{\eta})$

(f) $T^{\varrho + 1/2} \bar{\boldsymbol{\eta}}(\varrho, \lambda, \lambda') \Rightarrow \left(\frac{1+2\varrho}{\lambda'^{1+2\varrho} - \lambda^{1+2\varrho}}\right) \boldsymbol{\Omega}^{1/2} \int_\lambda^{\lambda'} r^\varrho \mathrm{d}\boldsymbol{W} \equiv \boldsymbol{b}(\varrho, \lambda, \lambda', \boldsymbol{\eta})$

(g) $\frac{1}{T^2} \sum_{(\lambda,\lambda']} \dot{\boldsymbol{u}}_t(\varrho) \dot{\boldsymbol{u}}_t(\varrho)' \Rightarrow \boldsymbol{\Omega}^{1/2} \left[\int_\lambda^{\lambda'} \boldsymbol{W}\boldsymbol{W}' \mathrm{d}r - \left(\frac{1+2\varrho}{\lambda'^{1+2\varrho} - \lambda^{1+2\varrho}}\right) \int_\lambda^{\lambda'} r^\varrho \boldsymbol{W} \mathrm{d}r \int_\lambda^{\lambda'} r^\varrho \boldsymbol{W}' \mathrm{d}r\right] \boldsymbol{\Omega}^{1/2} \equiv \boldsymbol{A}(\varrho, \lambda, \lambda', \boldsymbol{\eta})$

(h) $\frac{1}{T} \sum_{(\lambda,\lambda']} \dot{\boldsymbol{u}}_t(\varrho) \boldsymbol{\eta}'_t \Rightarrow \boldsymbol{\Omega}^{1/2} \left[\int_\lambda^{\lambda'} \boldsymbol{W} \mathrm{d}\boldsymbol{W}' - \left(\frac{1+2\varrho}{\lambda'^{1+2\varrho} - \lambda^{1+2\varrho}}\right) \int_\lambda^{\lambda'} r^\varrho \boldsymbol{W} \mathrm{d}r \int_\lambda^{\lambda'} r^\varrho \mathrm{d}\boldsymbol{W}\right] \boldsymbol{\Omega}^{1/2} + (\lambda' - \lambda)(\boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_0) \equiv \boldsymbol{B}(\varrho, \lambda, \lambda', \boldsymbol{\eta})$.

Also, if we include a drift such that $\boldsymbol{v}_t \equiv \boldsymbol{\mu} t + \boldsymbol{u}_t, \quad t \geq 1$, then:

(i) $T^{\varrho - 1} \bar{\boldsymbol{v}}(\varrho, 0, 1) \xrightarrow{p} \frac{2\varrho+1}{\varrho+2} \boldsymbol{\mu}$

(j) $\frac{1}{T^3} \sum_{t=1}^T \dot{\boldsymbol{v}}_t(\varrho) \dot{\boldsymbol{v}}_t(\varrho)' \xrightarrow{p} \frac{1}{3} \left(\frac{\varrho-1}{\varrho+2}\right)^2 \boldsymbol{\mu}\boldsymbol{\mu}'$

(k) $\frac{1}{T^{3/2}} \sum_{t=1}^T \dot{\boldsymbol{v}}_t(\varrho) \dot{\boldsymbol{\eta}}_t(\varrho)' \Rightarrow \boldsymbol{\mu} \int_0^1 \left[r - \left(\frac{2\varrho+1}{\varrho+2}\right) r^\varrho\right] \mathrm{d}\boldsymbol{W}' \boldsymbol{\Omega}^{1/2}$

Finally, if we include a linear tendency in process $\{\boldsymbol{\eta}_t\}$ *such that* $\boldsymbol{\xi}_t \equiv \boldsymbol{\eta}_t + \boldsymbol{\gamma} t$, *then:*

(l) $T^{\varrho-1}\bar{\boldsymbol{\xi}}(\varrho) \xrightarrow{p} \frac{2\varrho+1}{\varrho+2}\boldsymbol{\gamma}$

(m) $\frac{1}{T^3}\sum_{t=1}^{T} \dot{\boldsymbol{v}}_t(\varrho)\dot{\boldsymbol{\xi}}_t(\varrho)' \xrightarrow{p} \frac{1}{3}\left(\frac{\varrho-1}{\varrho+2}\right)^2 \boldsymbol{\mu}\boldsymbol{\gamma}'$,

where $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\eta})$ *and* $\boldsymbol{\Omega}_j \equiv \boldsymbol{\Omega}_j(\boldsymbol{\eta})$.

Proof. Let $\boldsymbol{U}_T(r) \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{\eta}_t$ with sample path on the Skorohod space $D[0,1]$. Then, as a consequence of Proposition 1, $\boldsymbol{U}_T \Rightarrow \boldsymbol{\Omega}^{1/2}\boldsymbol{W}$.

For (a) write $\frac{1}{\sqrt{T}}\boldsymbol{\eta}_t = \boldsymbol{U}_T\left(\frac{t}{T}\right) - \boldsymbol{U}_T\left(\frac{t-1}{T}\right) \equiv \int_{\frac{t-1}{T}}^{\frac{t}{T}} \mathrm{d}\boldsymbol{U}_T(r)$, then:

$$\frac{1}{T^{1/2+\varrho}}\sum_{(\lambda,\lambda']} t^{\varrho}\boldsymbol{\eta}_t = \sum_{(\lambda,\lambda']}\left(\frac{t}{T}\right)^{\varrho}\int_{\frac{t-1}{T}}^{\frac{t}{T}} \mathrm{d}\boldsymbol{U}_T(r) = \int_{\frac{T_\lambda}{T}}^{\frac{T_{\lambda'}}{T}} \left(\frac{\lfloor rT \rfloor}{T}\right)^{\varrho} \mathrm{d}\boldsymbol{U}_T \Rightarrow \boldsymbol{\Omega}^{1/2}\int_{\lambda}^{\lambda'} r^{\varrho}\mathrm{d}\boldsymbol{W},$$

where the convergence in distribution follows from Theorem 2.2 of Kurtz and Protter (1991).

For (b), note that $\boldsymbol{u}_{t-1} = \sqrt{T}\boldsymbol{U}_T\left(\frac{t-1}{T} \le r < \frac{t}{T}\right)$. Consequently, $\boldsymbol{u}_{t-1} = T^{3/2}\int_{\frac{t-1}{T}}^{\frac{t}{T}} \boldsymbol{U}_T(r)\mathrm{d}r$. Then,

$$\begin{aligned}\frac{1}{T^{3/2+\varrho}}\sum_{(\lambda,\lambda']} t^{\varrho}\boldsymbol{u}_t = \frac{1}{T^{3/2}}\sum_{(\lambda,\lambda']}\left(\frac{t}{T}\right)^{\varrho}(\boldsymbol{u}_{t-1}+\boldsymbol{\eta}_t) &= \sum_{(\lambda,\lambda']}\int_{\frac{t-1}{T}}^{\frac{t}{T}}\left(\frac{\lfloor rT \rfloor}{T}\right)^{\varrho}\boldsymbol{U}_T(r)\mathrm{d}r + o_P(1)\\ &= \int_{\frac{T_\lambda}{T}}^{\frac{T_{\lambda'}}{T}}\left(\frac{\lfloor rT \rfloor}{T}\right)^{\varrho}\boldsymbol{U}_T(r)\mathrm{d}r + o_P(1)\\ &\Rightarrow \boldsymbol{\Omega}^{1/2}\int_{\lambda}^{\lambda'} r^{\varrho}\boldsymbol{W}(r)\mathrm{d}r.\end{aligned}$$

For (c), define $\boldsymbol{U}_T^j(r) \equiv \left(\frac{1}{T}\right)^{1/2}\sum_{t=1}^{[rT]} \boldsymbol{\eta}_{t+j}$ for any positive integer j. Hence:

$$\frac{1}{T}\sum_{(\lambda,\lambda']} \boldsymbol{y}_{t-1}\boldsymbol{\eta}'_{t-1+j} = \sum_{(\lambda,\lambda']}\boldsymbol{U}_T^0\left(\frac{t-1}{T}\right)\int_{\frac{t-1}{T}}^{\frac{t}{T}} \mathrm{d}\boldsymbol{U}_T^j(r) = \int_{\frac{T_\lambda}{T}}^{\frac{T_{\lambda'}}{T}} \boldsymbol{U}_T^0(r)\mathrm{d}\boldsymbol{U}_T^j(r).$$

Let $\boldsymbol{\Sigma}_j \equiv \lim_{T\to\infty} T^{-1}\mathbb{E}\left(\sum_{t=1}^{T}\boldsymbol{\eta}_t\sum_{t=1}^{T}\boldsymbol{\eta}'_{t+j}\right)$. It is straightforward to show that

$$\begin{bmatrix}\boldsymbol{U}_T^0\\ \boldsymbol{U}_T^j\end{bmatrix} \Rightarrow \widetilde{\boldsymbol{\Sigma}}_j^{1/2}\boldsymbol{W} \equiv \begin{bmatrix}\boldsymbol{U}^0\\ \boldsymbol{U}^j\end{bmatrix} \quad \text{, where } \widetilde{\boldsymbol{\Sigma}}_j(n^2\times n^2) \equiv \begin{bmatrix}\boldsymbol{\Sigma}_0 & \boldsymbol{\Sigma}_j\\ \boldsymbol{\Sigma}'_j & \boldsymbol{\Sigma}_0\end{bmatrix}.$$

Note that $\boldsymbol{\Sigma}_0 = \boldsymbol{\Omega}$.

Therefore, it is possible to apply a generalization of Theorem 2.2 of Kurtz and Protter (1991). See, for instance, Theorem 30.13 in Davidson (1994) or Hansen (1992) for the case of $j = 1$. Consequently,

$$\int_{\frac{T_\lambda}{T}}^{\frac{T_{\lambda'}}{T}} \boldsymbol{U}_T^0(r)\mathrm{d}\boldsymbol{U}_T^j(r) \Rightarrow \boldsymbol{\Omega}^{1/2}\left[\int_{\lambda}^{\lambda'} \boldsymbol{W}\mathrm{d}\boldsymbol{W}\right]\boldsymbol{\Omega}^{1/2} + (\lambda'-\lambda)\boldsymbol{\Omega}_j.$$

Also the stochastic integral above for the case of $j = 1$ is the same one appearing in Phillips (1986b).

For (d), we start by considering $k=0$:

$$\boldsymbol{u}_t\boldsymbol{u}_t' = (\boldsymbol{u}_{t-1}+\boldsymbol{\eta}_t)(\boldsymbol{u}_{t-1}+\boldsymbol{\eta}_t)' = \boldsymbol{u}_{t-1}\boldsymbol{u}_{t-1}' + \boldsymbol{u}_{t-1}\boldsymbol{\eta}_t' + \boldsymbol{\eta}_t\boldsymbol{u}_{t-1}' + \boldsymbol{\eta}_t\boldsymbol{\eta}_t'.$$

Summing over and rearranging we are left with

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{u}_{t-1}\boldsymbol{\eta}_t' + \boldsymbol{\eta}_t\boldsymbol{u}_{t-1}' + \boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right) &= \frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{u}_t\boldsymbol{u}_t' - \boldsymbol{u}_{t-1}\boldsymbol{u}_{t-1}'\right)\\ &= \frac{1}{T}\left(\boldsymbol{u}_T\boldsymbol{u}_T' - \boldsymbol{u}_0\boldsymbol{u}_0'\right)\\ &\Rightarrow \boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1)\boldsymbol{W}(1)'\boldsymbol{\Omega}^{1/2}.\end{aligned}$$

Therefore, $T^{-2}\sum_{t=1}^{T}\left(\boldsymbol{u}_{t-1}\boldsymbol{\eta}_t' + \boldsymbol{\eta}_t\boldsymbol{u}_{t-1}' + \boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right) = o_P(1)$.

Finally,

$$\begin{aligned}\frac{1}{T^2}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{u}_t' &= \frac{1}{T^2}\sum_{(\lambda,\lambda']}\boldsymbol{u}_{t-1}\boldsymbol{u}_{t-1}' + \frac{1}{T^2}\sum_{(\lambda,\lambda']}\left(\boldsymbol{u}_{t-1}\boldsymbol{\eta}_t' + \boldsymbol{\eta}_t\boldsymbol{y}_{t-1}' + \boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right)\\ &= \int_{\frac{T_\lambda}{T}}^{\frac{T_{\lambda'}}{T}}\boldsymbol{U}_T(r)\boldsymbol{U}_T'(r)\mathrm{d}r + o_P(1)\\ &\Rightarrow \boldsymbol{\Omega}^{1/2}\int_{\lambda}^{\lambda'}\boldsymbol{W}(r)\boldsymbol{W}(r)'\mathrm{d}r\boldsymbol{\Omega}^{1/2}.\end{aligned}$$

For $k\in\mathbb{Z}$ we have that $\boldsymbol{u}_{t+k} = \boldsymbol{u}_t + \mathrm{sgn}(k)\sum_{i=1}^{|k|}\boldsymbol{\eta}_{t+i}$. Then,

$$\frac{1}{T^2}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{u}_{t+k}' = \frac{1}{T^2}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{u}_t' + \mathrm{sgn}(k)\frac{1}{T^2}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\sum_{i=1}^{|k|}\boldsymbol{\eta}_{t+i}'.$$

We have show in (c) that $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{\eta}_{t+i}' = O_P(1)$ for every $i\in\{1,\ldots,|k|\}$. Thus, we have the desired result as the second term is a finite sum of $o_P(1)$ terms.

To prove (e) and (f) we use the following result from power series

$$\sum_{t=1}^{T}t^k = \frac{1}{k+1}T^{k+1} + o(T^{k+1}), \qquad k=0,1,2..$$

to show that

$$\frac{T^{1+2\varrho}}{\sum\limits_{(\lambda,\lambda']}t^{2\varrho}} = \frac{1+2\varrho}{\left(\frac{T_{\lambda'}}{T}\right)^{1+2\varrho} - \left(\frac{T_\lambda}{T}\right)^{1+2\varrho} + o(1)} \longrightarrow \frac{1+2\varrho}{\lambda'^{1+2\varrho} - \lambda^{1+2\varrho}}.$$

Then, for (e):

$$\begin{aligned}T^{\varrho-1/2}\bar{\boldsymbol{u}}(\varrho,\lambda,\lambda') &= \left(\frac{T^{1+2\varrho}}{\sum\limits_{(\lambda,\lambda']}t^{2\varrho}}\right)\left(\frac{1}{T^{3/2+\varrho}}\sum_{(\lambda,\lambda']}t^{\varrho}\boldsymbol{u}_t\right)\\ &\Rightarrow \left(\frac{1+2\varrho}{\lambda'^{1+2\varrho} - \lambda^{1+2\varrho}}\right)\boldsymbol{\Omega}^{1/2}\int_{\lambda}^{\lambda'}r^{\varrho}\boldsymbol{W}\mathrm{d}r,\end{aligned}$$

and for (f):

$$T^{\varrho+1/2}\bar{\boldsymbol{\eta}}(\varrho,\lambda,\lambda') = \left(\frac{T^{1+2\varrho}}{\sum_{(\lambda,\lambda']} t^{2\varrho}}\right)\left(\frac{1}{T^{1/2+\varrho}}\sum_{(\lambda,\lambda']} t^{\varrho}\boldsymbol{\eta}_t\right)$$
$$\Rightarrow \left(\frac{1+2\varrho}{\lambda'^{1+2\varrho}-\lambda^{1+2\varrho}}\right)\boldsymbol{\Omega}^{1/2}\int_{\lambda}^{\lambda'} r^{\varrho}\mathrm{d}\boldsymbol{W}$$

To show (g), we omit the $(\varrho,\lambda,\lambda')$ argument in what follows:

$$\begin{aligned}
\frac{1}{T^2}\sum_{(\lambda,\lambda']}\dot{\boldsymbol{u}}_t\dot{\boldsymbol{u}}_t' &\equiv \frac{1}{T^2}\sum_{(\lambda,\lambda']}\left(\boldsymbol{u}_t - t^{\varrho}\bar{\boldsymbol{u}}\right)\left(\boldsymbol{u}_t - t^{\varrho}\bar{\boldsymbol{u}}\right)' \\
&= \frac{1}{T^2}\left[\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{u}_t' - \sum_{(\lambda,\lambda']} t^{\varrho}\boldsymbol{u}_t\bar{\boldsymbol{u}}' - \bar{\boldsymbol{u}}\sum_{(\lambda,\lambda']} t^{\varrho}\boldsymbol{u}_t' + \sum_{(\lambda,\lambda']} t^{2\varrho}\bar{\boldsymbol{u}}\bar{\boldsymbol{u}}'\right] \\
&= \frac{1}{T^2}\left[\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{u}_t' - \left(\sum_{(\lambda,\lambda']} t^{2\varrho}\right)\bar{\boldsymbol{u}}\bar{\boldsymbol{u}}'\right] \\
&= \frac{1}{T^2}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{u}_t - \left(\frac{\sum_{(\lambda,\lambda']} t^{2\varrho}}{T^{2\varrho+1}}\right)\left(T^{\varrho-1/2}\bar{\boldsymbol{u}}\right)\left(T^{\varrho-1/2}\bar{\boldsymbol{u}}\right)' \\
&\Rightarrow \boldsymbol{\Omega}^{1/2}\left[\int_{\lambda}^{\lambda'}\boldsymbol{W}\boldsymbol{W}'\mathrm{d}r - \left(\frac{1+2\varrho}{\lambda'^{1+2\varrho}-\lambda^{1+2\varrho}}\right)\int_{\lambda}^{\lambda'} r^{\varrho}\boldsymbol{W}\mathrm{d}r\int_{\lambda}^{\lambda'} r^{\varrho}\boldsymbol{W}'\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2}
\end{aligned}$$

To prove (h), we first use the result (c) to show that

$$\frac{1}{T}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{\eta}_t' = \frac{1}{T}\sum_{(\lambda,\lambda']}\boldsymbol{u}_{t-1}\boldsymbol{\eta}_t' + \frac{T_{\lambda'}-T_{\lambda}}{T}\left(\frac{1}{T_{\lambda'}-T_{\lambda}}\sum_{(\lambda,\lambda']}\boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right)$$
$$\Rightarrow \boldsymbol{\Omega}^{1/2}\int_{\lambda}^{\lambda'}\boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r)\boldsymbol{\Omega}^{1/2} + (\lambda'-\lambda)(\boldsymbol{\Omega}_0+\boldsymbol{\Omega}_1).$$

Finally, we have:

$$\begin{aligned}
\frac{1}{T}\sum_{(\lambda,\lambda']}\dot{\boldsymbol{u}}_t(\varrho)\boldsymbol{\eta}_t' &= \frac{1}{T}\left[\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{\eta}_t' - \bar{\boldsymbol{u}}(\varrho)\sum_{(\lambda,\lambda']} t^{\varrho}\boldsymbol{\eta}_t'\right] \\
&= \frac{1}{T}\left[\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{\eta}_t' - \left(\sum_{(\lambda,\lambda']} t^{2\varrho}\right)\bar{\boldsymbol{u}}(\varrho)\bar{\boldsymbol{\eta}}(\varrho)'\right] \\
&= \frac{1}{T}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{\eta}_t' - \left(\frac{\sum_{(\lambda,\lambda']} t^{2\varrho}}{T^{2\varrho+1}}\right)\left(T^{\varrho-1/2}\bar{\boldsymbol{u}}(\varrho)\right)\left(T^{\varrho+1/2}\bar{\boldsymbol{\eta}}(\varrho)\right)' \\
&\Rightarrow \boldsymbol{\Omega}^{1/2}\left[\int_{\lambda}^{\lambda'}\boldsymbol{W}\mathrm{d}\boldsymbol{W}' - \left(\frac{1+2\varrho}{\lambda'^{1+2\varrho}-\lambda^{1+2\varrho}}\right)\int_{\lambda}^{\lambda'} r^{\varrho}\boldsymbol{W}\mathrm{d}r\int_{\lambda}^{\lambda'} r^{\varrho}\mathrm{d}\boldsymbol{W}\right]\boldsymbol{\Omega}^{1/2} \\
&\quad + (\lambda'-\lambda)(\boldsymbol{\Omega}_0+\boldsymbol{\Omega}_1).
\end{aligned}$$

Now for (i), take $\bar{\boldsymbol{v}}(\varrho) \equiv \bar{\boldsymbol{v}}(\varrho, 0, 1)$ then:

$$\begin{aligned}
T^{\varrho-1}\bar{\boldsymbol{v}}(\varrho) &= \boldsymbol{\mu} T^{\varrho-1}\left(\sum_{t=1}^{T} t^{2\varrho}\right)^{-1}\sum_{t=1}^{T} t^{\varrho+1} + T^{\varrho-1}\bar{\boldsymbol{u}}(\varrho) \\
&= \boldsymbol{\mu} T^{\varrho-1}\left(\frac{T^{2\varrho+1}}{2\varrho+1}\right)^{-1}\frac{T^{\varrho+2}}{\varrho+2} + \frac{1}{\sqrt{T}}\left(T^{\varrho-1/2}\bar{\boldsymbol{u}}(\varrho)\right) + o(1) \\
&= \frac{2\varrho+1}{\varrho+2}\boldsymbol{\mu} + o_P(1).
\end{aligned}$$

For (j), first note that

$$\begin{aligned}
\frac{1}{T^3}\sum_{t=1}^{T}\boldsymbol{v}_t\boldsymbol{v}_t' &= \frac{1}{T^3}\sum_{t=1}^{T}(\boldsymbol{\mu}t+\boldsymbol{u}_t)(\boldsymbol{\mu}t+\boldsymbol{u}_t)' = \boldsymbol{\mu}\boldsymbol{\mu}'\frac{1}{T^3}\sum_{t=1}^{T}t^2 + o_P(1) \\
&= \boldsymbol{\mu}\boldsymbol{\mu}'\frac{T(T+1)(2T+1)}{6T^3} + o_P(1) = \frac{1}{3}\boldsymbol{\mu}\boldsymbol{\mu}' + o_P(1),
\end{aligned}$$

then taking from (f) we can write:

$$\begin{aligned}
\frac{1}{T^3}\sum_{t=1}^{T}\dot{\boldsymbol{v}}_t(\varrho)\dot{\boldsymbol{v}}_t(\varrho)' &= \frac{1}{T^3}\sum_{t=1}^{T}\boldsymbol{v}_t\boldsymbol{v}_t - \left(\frac{\sum_{t=1}^{T}t^{2\varrho}}{T^{2\varrho+1}}\right)\left(T^{\varrho-1}\bar{\boldsymbol{v}}\right)\left(T^{\varrho-1}\bar{\boldsymbol{v}}\right)' \\
&= \frac{1}{3}\boldsymbol{\mu}\boldsymbol{\mu}' - \frac{1}{2\varrho+1}\left(\frac{2\varrho+1}{\varrho+2}\right)^2\boldsymbol{\mu}\boldsymbol{\mu}' + o_P(1) \\
&= \frac{(\varrho-1)^2}{3(\varrho+2)^2}\boldsymbol{\mu}\boldsymbol{\mu}' + o_P(1).
\end{aligned}$$

To prove (k) first we show that

$$\tfrac{1}{T^{3/2}}\sum_{t=1}^{T}\boldsymbol{v}_t\boldsymbol{\eta}_t' = \tfrac{1}{T^{3/2}}\sum_{t=1}^{T}(\boldsymbol{u}_t+\boldsymbol{\mu}t)\boldsymbol{\eta}_t' = \tfrac{1}{T^{3/2}}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{\eta}_t' + \boldsymbol{\mu}\tfrac{1}{T^{3/2}}\sum_{t=1}^{T}t\boldsymbol{\eta}_t' \Rightarrow \boldsymbol{\mu}\int_0^1 r\mathrm{d}\boldsymbol{W}'\boldsymbol{\Omega}^{1/2},$$

then we can use the latter result combined with result (f) and (i) to conclude that:

$$\begin{aligned}
\tfrac{1}{T^{3/2}}\sum_{t=1}^{T}\dot{\boldsymbol{v}}_t(\varrho)\dot{\boldsymbol{\eta}}_t(\varrho)' = \tfrac{1}{T^{3/2}}\sum_{t=1}^{T}\dot{\boldsymbol{v}}_t(\varrho)\boldsymbol{\eta}_t' &= \tfrac{1}{T^{3/2}}\sum_{t=1}^{T}\boldsymbol{v}_t\boldsymbol{\eta}_t' - \tfrac{\sum_{t=1}^{T}t^{2\varrho}}{T^{2\varrho+1}}(T^{\varrho-1}\bar{\boldsymbol{v}}(\varrho))(T^{\varrho+1/2}\bar{\boldsymbol{\eta}}(\varrho)') \\
&\Rightarrow \boldsymbol{\mu}\int_0^1 r\mathrm{d}\boldsymbol{W}'\boldsymbol{\Omega}^{1/2} - \left(\tfrac{1}{2\varrho+1}\right)\left(\tfrac{2\varrho+1}{\varrho+2}\boldsymbol{\mu}\right)\left(\tfrac{2\varrho+1}{1}\int_0^1 r^{\varrho}\mathrm{d}\boldsymbol{W}'\boldsymbol{\Omega}^{1/2}\right) \\
&= \boldsymbol{\mu}\int_0^1\left[r - \left(\tfrac{2\varrho+1}{\varrho+2}\right)r^{\varrho}\right]\mathrm{d}\boldsymbol{W}'\boldsymbol{\Omega}^{1/2}.
\end{aligned}$$

For (l), we expand as in the proof (i) to obtain:

$$\begin{aligned}
T^{\varrho-1}\bar{\boldsymbol{\xi}}(\varrho) &= \boldsymbol{\gamma} T^{\varrho-1}\left(\sum_{t=1}^{T}t^{2\varrho}\right)^{-1}\sum_{t=1}^{T}t^{\varrho+1} + T^{\varrho-1}\bar{\boldsymbol{\eta}}(\varrho) \\
&= \boldsymbol{\gamma} T^{\varrho-1}\left(\tfrac{T^{2\varrho+1}}{2\varrho+1}\right)^{-1}\tfrac{T^{\varrho+2}}{\varrho+2} + \tfrac{1}{T^{3/2}}\left(T^{\varrho+1/2}\bar{\boldsymbol{\eta}}(\varrho)\right) + o(1) \\
&= \tfrac{2\varrho+1}{\varrho+2}\boldsymbol{\gamma} + o_P(1).
\end{aligned}$$

Finally, for (m) we have using the result (i) and (l):

$$\frac{1}{T^3}\sum_{t=1}^{T}\dot{\boldsymbol{v}}(\varrho)\dot{\boldsymbol{\xi}}(\varrho)' = \left(\frac{\sum_{t=1}^{T}t^2}{T^3}\right)\boldsymbol{\mu}\boldsymbol{\gamma}' - \left(\frac{\sum_{t=1}^{T}t^{2\varrho}}{T^{2\varepsilon+1}}\right)(T^{\varrho-1}\bar{\boldsymbol{v}}(\varrho))(T^{\varrho+1/2}\bar{\boldsymbol{\eta}}(\varrho)') + o_P(1)$$
$$= \tfrac{1}{3}\boldsymbol{\mu}\boldsymbol{\gamma}' - \tfrac{2\varrho+1}{(\varrho+2)^2}\boldsymbol{\mu}\boldsymbol{\gamma}' + o_P(1)$$
$$= \tfrac{1}{3}\left(\tfrac{\varrho-1}{\varrho+2}\right)^2\boldsymbol{\mu}\boldsymbol{\gamma}' + o_P(1).$$

□

Proof of Lemma 1

Proof. Using notation (15) we can express the least-squares estimator as the difference to the true parameter value as:

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \left[\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(1)\dot{\boldsymbol{y}}_{0t}(1)'\right]^{-1}\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(1)\dot{\nu}_t(1),$$
$$\widehat{\gamma} - \gamma_0 = \bar{\nu}(1) - \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)'\bar{\boldsymbol{y}}_0(1),$$
$$\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0 = \left[\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(0)\dot{\boldsymbol{y}}_{0t}(0)'\right]^{-1}\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(0)\dot{\zeta}_t(0),$$
$$\widehat{\alpha} - \alpha_0 = \bar{\zeta}(0) - (\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0)'\,\bar{\boldsymbol{y}}_0(0).$$

For the case of $\boldsymbol{\mu} = \mathbf{0}$ (no drift), consequently $\gamma_0 = 0$, we have $\zeta_t = \nu_t$ and we can apply the limiting distributions in Lemma 5 together with the continuous mapping theorem to conclude:

$$T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = \left[\frac{1}{T^2}\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(1)\dot{\boldsymbol{y}}_{0t}(1)'\right]^{-1}\left[\frac{1}{T}\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(1)\dot{\nu}_t(1)\right] \Rightarrow \boldsymbol{P}_{00}^{-1}\boldsymbol{Q}_{01} \equiv \widetilde{\boldsymbol{h}},$$
$$T^{3/2}\left(\widehat{\gamma} - \gamma_0\right) = T^{3/2}\bar{\nu}(1) - T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)'\sqrt{T}\bar{\boldsymbol{y}}_0(1) \Rightarrow [\boldsymbol{b}(1,0,\lambda_0,\boldsymbol{\eta})]_1 - \widetilde{\boldsymbol{h}}'\,[\boldsymbol{a}(1,0,\lambda_0,\boldsymbol{\eta})]_0\,,$$
$$T\left(\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0\right) = \left[\frac{1}{T^2}\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(0)\dot{\boldsymbol{y}}_{0t}(0)'\right]^{-1}\left[\frac{1}{T}\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(0)\dot{\zeta}_t(0)\right] \Rightarrow \boldsymbol{R}_{00}^{-1}\boldsymbol{V}_{01} \equiv \widetilde{\boldsymbol{p}},$$
$$\sqrt{T}\left(\widehat{\alpha} - \alpha_0\right) = \sqrt{T}\bar{\zeta}(0) - T\left(\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0\right)'\frac{1}{\sqrt{T}}\bar{\boldsymbol{y}}_0(0) \Rightarrow [\boldsymbol{b}(0,0,\lambda_0,\boldsymbol{\eta})]_1 - \widetilde{\boldsymbol{p}}'\,[\boldsymbol{a}(0,0,\lambda_0,\boldsymbol{\eta})]_0\,.$$

For $\boldsymbol{\mu} \neq 0$ note that both $\frac{1}{T_0^3}\sum_{t=1}^{T_0}\dot{\boldsymbol{y}}_{0t}(1)\dot{\boldsymbol{y}}_{0t}(1)'$ and $\frac{1}{T_0^3}\sum_{t=1}^{T_0}\dot{\boldsymbol{y}}_{0t}(0)\dot{\boldsymbol{y}}_{0t}(0)'$ converge to singular matrices according to Lemma 5(m). Hence, unless $\boldsymbol{y}_{0t}$ is a scalar the OLS estimators are not well defined asymptotically. Moreover, for the first specfication even for the scalar case the OLS estimator is not defined since if we set $\varrho = 1$ we get that $\frac{1}{T_0^3}\sum_{t=1}^{T_0}\dot{\boldsymbol{y}}_{0t}(1)\dot{\boldsymbol{y}}_{0t}(1)' \stackrel{p}{\longrightarrow} 0$. For the case of unique regressor in the second specification we are left with:

$$\widehat{\pi} - \beta_0 = \left[\frac{1}{T_0^3}\sum_{(0,\lambda_0]}\dot{y}_{0t}(0)\dot{y}_{0t}(0)'\right]^{-1}\left[\frac{1}{T_0^3}\sum_{(0,\lambda_0]}\dot{y}_{0t}(0)\dot{\zeta}_t(0)\right] \stackrel{p}{\longrightarrow} \left(\frac{\mu_0^2}{12}\right)^{-1}\frac{\mu_0\gamma_0}{12} = \frac{\gamma_0}{\mu_0},$$
$$\frac{1}{T}\left(\widehat{\alpha} - \alpha_0\right) = \frac{T_0}{T}\left[\frac{1}{T_0}\bar{\zeta}(0) - (\widehat{\pi} - \beta_0)'\frac{1}{T_0}\bar{y}_0(0)\right] \stackrel{p}{\longrightarrow} \lambda_0\left[\frac{\gamma_0}{2} - \frac{\gamma_0}{\mu_0}\frac{\mu_0}{2}\right] = 0.$$

□

Proof of Theorem 1

Proof. For the post intervention period $t = T_0 + 1, \ldots, T$ we can write:

$$\widehat{\delta}_{1t} - \delta_t = y_{1t} - \widehat{\gamma}t - \widehat{\boldsymbol{\beta}}'\boldsymbol{y}_{0t} - \delta_t = \nu_t - (\widehat{\gamma} - \gamma_0)\, t - \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' \boldsymbol{y}_{0t}$$
$$\widehat{\delta}_{2t} - \delta_t = y_{1t} - \widehat{\alpha} - \widehat{\boldsymbol{\pi}}'\boldsymbol{y}_{0t} - \delta_t = \nu_t - \widehat{\alpha} - (\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0)' \boldsymbol{y}_{0t}.$$

Therefore for the first specification:

$$\begin{aligned}\sqrt{T}\left(\widehat{\Delta}_1 - \Delta\right) &= \frac{\sqrt{T}}{T_2}\sum_{t>T_0}\left(\widehat{\delta}_{1t} - \delta_t\right)\\ &= \sqrt{T}\bar{\nu}(0,\lambda_0,1) - \frac{T+T_0+1}{2T}T^{3/2}(\widehat{\gamma} - \gamma_0) - T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)'\frac{1}{\sqrt{T}}\bar{\boldsymbol{y}}_0(0,\lambda_0,1)\\ &\Rightarrow [\boldsymbol{b}(0,\lambda_0,1,\boldsymbol{\eta})]_1 - \frac{1+\lambda_0}{2}\left([\boldsymbol{b}(1,0,\lambda_0,\boldsymbol{\eta})]_1 - \widetilde{\boldsymbol{h}}'\left[\boldsymbol{a}(1,0,\lambda_0,\boldsymbol{\eta})\right]_0\right)\\ &\quad - \widetilde{\boldsymbol{h}}'[\boldsymbol{a}(0,\lambda_0,1,\boldsymbol{\eta})]_0 \equiv \boldsymbol{h}'\boldsymbol{c}.\end{aligned}$$

Similarly, for the second specification:

$$\begin{aligned}\sqrt{T}\left(\widehat{\Delta}_2 - \Delta\right) &= \frac{\sqrt{T}}{T_2}\sum_{t=T_0}^{T}\left(\widehat{\delta}_{2t} - \delta_t\right)\\ &= \sqrt{T}\bar{\nu}(0,\lambda_0,1) - \sqrt{T}(\widehat{\alpha} - \alpha_0) - T\left(\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0\right)'\frac{1}{\sqrt{T}}\bar{\boldsymbol{y}}_0(0,\lambda_0,1)\\ &\Rightarrow [\boldsymbol{b}(0,\lambda_0,1,\boldsymbol{\eta})]_1 - \left([\boldsymbol{b}(0,0,\lambda_0,\boldsymbol{\eta})]_1 - \widetilde{\boldsymbol{p}}'\left[\boldsymbol{a}(0,0,\lambda_0,\boldsymbol{\eta})\right]_0\right)\\ &\quad - \widetilde{\boldsymbol{p}}'[\boldsymbol{a}(0,\lambda_0,1,\boldsymbol{\eta})]_0 \equiv \boldsymbol{p}'\boldsymbol{d}.\end{aligned}$$

□

Proof of Lemma 2

Proof. Using notation (15) we can express the least-squares estimator for the spurious case as:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \left[\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(1)\dot{\boldsymbol{y}}_{0t}(1)'\right]^{-1}\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(1)\dot{y}_{1t}(1),\\ \widehat{\gamma} &= \bar{y}_1(1) - \widehat{\boldsymbol{\beta}}'\bar{\boldsymbol{y}}_0(1) = (1, -\widehat{\boldsymbol{\beta}}')\bar{\boldsymbol{y}}(1),\\ \widehat{\boldsymbol{\pi}} &= \left[\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(0)\dot{\boldsymbol{y}}_{0t}(0)'\right]^{-1}\sum_{(0,\lambda_0]}\dot{\boldsymbol{y}}_{0t}(0)\dot{y}_{1t}(0),\\ \widehat{\alpha} &= \bar{y}_1(0) - \widehat{\boldsymbol{\pi}}'\bar{\boldsymbol{y}}_0(0) = (1, -\widehat{\boldsymbol{\pi}}')\bar{\boldsymbol{y}}(0).\end{aligned}$$

For the case of $\boldsymbol{\mu} = \mathbf{0}$ (no drift) we can apply the limiting distributions in Lemma 5 together with the continuous mapping theorem to conclude:

$$\widehat{\boldsymbol{\beta}} = \left[\frac{1}{T^2}\sum_{(0,\lambda_0]} \dot{\boldsymbol{y}}_{0t}(1)\dot{\boldsymbol{y}}_{0t}(1)'\right]^{-1}\left(\frac{1}{T^2}\sum_{(0,\lambda_0]} \dot{\boldsymbol{y}}_{0t}(1)\dot{y}_{1t}(1)\right) \Rightarrow \boldsymbol{P}_{00}^{-1}\boldsymbol{P}_{01} \equiv \widetilde{\boldsymbol{f}},$$

$$\sqrt{T}\widehat{\gamma} = (1, -\widehat{\boldsymbol{\beta}}')\sqrt{T}\bar{\boldsymbol{y}}(1) \Rightarrow (1, -\widetilde{\boldsymbol{f}}')\boldsymbol{a}(1, 0, \lambda_0, \boldsymbol{\varepsilon}) \equiv \boldsymbol{f}'\boldsymbol{a}(1, 0, \lambda_0, \boldsymbol{\varepsilon}),$$

$$\widehat{\boldsymbol{\pi}} = \left[\frac{1}{T^2}\sum_{(0,\lambda_0]} \dot{\boldsymbol{y}}_{0t}(0)\dot{\boldsymbol{y}}_{0t}(0)'\right]^{-1}\left[\frac{1}{T^2}\sum_{(0,\lambda_0]} \dot{\boldsymbol{y}}_{0t}(0)\dot{y}_{1t}(0)\right] \Rightarrow \boldsymbol{R}_{00}^{-1}\boldsymbol{R}_{01} \equiv \widetilde{\boldsymbol{g}},$$

$$\frac{1}{\sqrt{T}}\widehat{\alpha} = (1, -\widehat{\boldsymbol{\pi}}')\frac{1}{\sqrt{T}}\bar{\boldsymbol{y}}(0) \Rightarrow (1, -\widetilde{\boldsymbol{g}}')\boldsymbol{a}(0, 0, \lambda_0, \boldsymbol{\varepsilon}) \equiv \boldsymbol{g}'\boldsymbol{a}(0, 0, \lambda_0, \boldsymbol{\varepsilon}).$$

For the case of $\boldsymbol{\mu} = \mathbf{0}$, as shown in the proof of Lemma 1, the estimators are not well defined except for the no trend specification with only one regressor, in the case:

$$\widehat{\pi} = \left[\frac{1}{T_0^3}\sum_{(0,\lambda_0]} \dot{y}_{0t}(0)\dot{y}_{0t}(0)'\right]^{-1}\left[\frac{1}{T_0^3}\sum_{(0,\lambda_0]} \dot{y}_{0t}(0)\dot{y}_{1t}(0)\right] \xrightarrow{p} \left(\frac{\mu_0^2}{12}\right)^{-1}\frac{\mu_0\mu_1}{12} = \frac{\mu_1}{\mu_0},$$

$$\frac{1}{T}\widehat{\alpha} = \frac{T_0}{T}(1, -\widehat{\pi}')\frac{1}{T_0}\bar{\boldsymbol{y}}(0) \xrightarrow{p} \lambda_0(1, -\mu_1/\mu_0)\frac{\boldsymbol{\mu}}{2} = 0.$$

□

Proof of Theorem 2

Proof. For the post intervention period $t = T_0 + 1, \ldots, T$ we have:

$$\widehat{\delta}_{1t} - \delta_t = y_{1t} - \widehat{\gamma}t - \widehat{\boldsymbol{\beta}}'\boldsymbol{y}_{0t} - \delta_t = y_{1t}^{(0)} - \widehat{\gamma}t - \widehat{\boldsymbol{\beta}}'\boldsymbol{y}_{0t}$$
$$\widehat{\delta}_{2t} - \delta_t = y_{1t} - \widehat{\alpha} - \widehat{\boldsymbol{\pi}}'\boldsymbol{y}_{0t} - \delta_t = y_{1t}^{(0)} - \widehat{\alpha} - \widehat{\boldsymbol{\pi}}'\boldsymbol{y}_{0t}.$$

Therefore, for the first specification:

$$\begin{aligned}\frac{1}{\sqrt{T}}\left(\widehat{\Delta}_1 - \Delta\right) &= \frac{1}{\sqrt{T}T_2}\sum_{t>T_0}\left(\widehat{\delta}_{1t} - \delta_t\right)\\ &= \frac{1}{\sqrt{T}}\bar{y}_1(0, \lambda_0, 1) - \frac{T + T_0 + 1}{2T}\sqrt{T}\widehat{\gamma} - \widehat{\boldsymbol{\beta}}'\frac{1}{\sqrt{T}}\bar{\boldsymbol{y}}_0(0, \lambda_0, 1)\\ &= (1, -\widehat{\boldsymbol{\beta}}')\frac{1}{\sqrt{T}}\bar{\boldsymbol{y}}(0, \lambda_0, 1) - \frac{T + T_0 + 1}{2T}\sqrt{T}\widehat{\gamma}\\ &\Rightarrow \boldsymbol{f}'\boldsymbol{a}(0, \lambda_0, 1, \boldsymbol{\varepsilon}) - \frac{1 + \lambda_0}{2}\boldsymbol{f}'\boldsymbol{a}(1, 0, \lambda_0, \boldsymbol{\varepsilon}) \equiv \boldsymbol{f}'\boldsymbol{e}.\end{aligned}$$

Similarly, for the second specification:

$$\begin{aligned}\frac{1}{\sqrt{T}}\left(\widehat{\Delta}_2 - \Delta\right) &= \frac{1}{\sqrt{T}T_2}\sum_{t>T_0}\left(\widehat{\delta}_{2t} - \delta_t\right)\\ &= \frac{1}{\sqrt{T}}\bar{y}_1(0, \lambda_0, 1) - \frac{1}{\sqrt{T}}\widehat{\alpha} - \widehat{\boldsymbol{\pi}}'\frac{1}{\sqrt{T}}\bar{\boldsymbol{y}}_0(0, \lambda_0, 1)\\ &= (1, -\widehat{\boldsymbol{\pi}}')\frac{1}{\sqrt{T}}\bar{\boldsymbol{y}}(0, \lambda_0, 1) - \frac{1}{\sqrt{T}}\widehat{\alpha}\\ &\Rightarrow \boldsymbol{g}'\boldsymbol{a}(0, \lambda_0, 1, \boldsymbol{\varepsilon}) - \boldsymbol{g}'\boldsymbol{a}(0, 0, \lambda_0, \boldsymbol{\varepsilon}) \equiv \boldsymbol{g}'\boldsymbol{l}.\end{aligned}$$

□

Proof of Lemma 3

Proof. For the post intervention period $t = T_0 + 1, \dots, T$:

$$\widehat{\nu}_{1t} = \dot{\nu}_t - (\widehat{\gamma} - \gamma_0)\left(t - \frac{T + T_0 + 1}{2}\right) - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t} + \dot{\delta}_t$$
$$\widehat{\nu}_{2t} = \dot{\nu}_t - (\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t} + \dot{\delta}_t.$$

Since either under $\mathcal{H}_0$ or $\mathcal{H}_1$, $\dot{\delta} = 0$, we have for $k = \{0, 1, \dots, T-1\}$

$$\widehat{\nu}_{1t}\widehat{\nu}_{1t+k} = \dot{\nu}_t\dot{\nu}_{t+k} - \dot{\nu}_t(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t+k} - (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t}\dot{\nu}_{t+k} + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t}\dot{\boldsymbol{y}}'_{0t+k}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$$
$$\widehat{\nu}_{2t}\widehat{\nu}_{2t+k} = \dot{\nu}_t\dot{\nu}_{t+k} - \dot{\nu}_t(\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t+k} - (\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t}\dot{\nu}_{t+k} + (\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t}\dot{\boldsymbol{y}}'_{0t+k}(\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0).$$

Both $\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ and $(\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0)$ are $O_P(\frac{1}{T})$ by Lemma 1. Also, $\sum \dot{\boldsymbol{y}}_{0t}\dot{\boldsymbol{y}}'_{0t+k} = O_P(T^2)$ and $\sum \dot{\nu}_t\dot{\boldsymbol{y}}_{0t+k} = O_P(T)$ all as a consequence of Lemma 5. Thus for $j \in \{1, 2\}$, we have:

$$\sum_{t=T_0+1}^{T-k} \widehat{\nu}_{jt}\widehat{\nu}_{jt+k} = \sum_{t=T_0+1}^{T-k} \dot{\nu}_t\dot{\nu}_{t+k} + O_P(1) = \sum_{t=T_0+1}^{T-k} \nu_t\nu_{t+k} + O_P(1),$$

where the last equality involves no more than some algebraic manipulation using the definition of $\dot{\nu}_t$ and $\ddot{\nu}_t$ and neglecting the $o_P(1)$ terms. Therefore, by the Law Large Numbers, which is ensured under Assumption 3,

$$\widehat{\rho}^2_{jk} \equiv \frac{1}{T_2}\sum_{t=T_0+1}^{T-k} \widehat{\nu}_{1t}\widehat{\nu}_{1t+k} \xrightarrow{p} \mathbb{E}\left(\nu_t\nu_{t+k}\right) \equiv \rho^2_k, \quad \forall k.$$

For part (b), the result follows from an argument parallel to one presented in Andrews (1991). Let $\widetilde{\sigma}^2$ be the pseudo-estimator analogous to the estimator $\widehat{\sigma}^2_j$ but with sequence $\widehat{\nu}_{jt}$ replaced by the unobservable sequence $\{\nu_t\}$ and let $\sigma^2 = \sum_{|k|<T} \rho^2_k$. Hence by the triangle inequality we have

$$|\widehat{\sigma}^2_j - \sigma^2| \leq |\widehat{\sigma}^2_j - \widetilde{\sigma}^2| + |\widetilde{\sigma}^2 - \sigma^2|.$$

Under Assumption A of Andrews (1991), which is implied by Assumption 3, the second term is $o_P(1)$. Assumption B of Andrews (1991), which ensures the first term to be $o_P(1)$ is not fulfilled directly by specification (7) due to the trend regressor. However, what is really necessary for the result is to bound the mean value expansion of the first term, which in our case, is simply given by

$$\frac{\sqrt{T}}{J_T}(\widehat{\sigma}^2_j - \widetilde{\sigma}^2) = \frac{1}{J_T}\sum_{|k|<T} \kappa(\frac{k}{J_t})\frac{1}{T_2}\sum_{t>T_0+|k|} \frac{\partial s(\tilde{\gamma}, \tilde{\boldsymbol{\beta}})}{\partial \gamma}(\widehat{\gamma} - \gamma_0) + \frac{\partial s(\tilde{\gamma}, \tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}'}(\widehat{\boldsymbol{\beta}} - \beta_0).$$

Since by Lemma $\widehat{\gamma} - \gamma_0 = O_P(T^{-3/2})$, a sufficient condition to bound the first term becomes $\mathsf{sup}_{t\geq 1}\, \mathbb{E}\left\|T^{-1}\frac{\partial \nu}{\partial \gamma}\right\|^2 \leq \infty$, which is clearly satisfied by our specification. The final requirement are the same that appears in Theorem 1 of Andrews (1991) and is fulfilled by most of the kernel functions used in the literature.

□

Proof of Theorem 3

Proof. We can decompose the t-statistic as:

$$\tau_j \equiv \sqrt{T_2}\frac{\widehat{\Delta}_j}{\widehat{\sigma}_j} = \sqrt{T_2}\left[\frac{(\widehat{\Delta}_j - \Delta_T)}{\widehat{\sigma}_j} + \frac{\Delta_T}{\widehat{\sigma}_j}\right] = \sqrt{\frac{T_2}{T}}\left(\frac{\sqrt{T}(\widehat{\Delta}_j - \Delta_T)}{\widehat{\sigma}_j}\right) + \frac{\sqrt{T_2}\Delta_T}{\widehat{\sigma}_j}$$

Under $\mathcal{H}_0$ the second term is zero and the first term converges in distribution by the Slutsky Theorem since the numerator of the term between parentheses converges in distribution according to Theorem 1, and the denominator converges in probability according to the Lemma 3, hence

$$\tau_1 \Rightarrow \frac{\sqrt{1-\lambda_0}}{\omega}\boldsymbol{h}'\boldsymbol{c},$$
$$\tau_2 \Rightarrow \frac{\sqrt{1-\lambda_0}}{\omega}\boldsymbol{p}'\boldsymbol{d},$$

whereas, under $\mathcal{H}_1$ the second term diverges at rate $\sqrt{T}$ since

$$\frac{1}{\sqrt{T}}\tau_j = \sqrt{\frac{T_2}{T}}\frac{\delta}{\widehat{\sigma}_j} \xrightarrow{p} \sqrt{1-\lambda_0}\frac{\delta}{\omega}.$$

□

Proof of Lemma 4

Proof. Let $\widehat{\boldsymbol{\theta}}_1 \equiv (1, \widehat{\boldsymbol{\beta}}')'$ and $\widehat{\boldsymbol{\theta}}_2 \equiv (1, \widehat{\boldsymbol{\pi}}')'$, then we can write the post intervention centered residuals for the first specification as:

$$\begin{aligned}
\widehat{\nu}_{1t} &\equiv y_{1t} - t\widehat{\gamma} - \widehat{\boldsymbol{\beta}}'\boldsymbol{y}_{0t} - \widehat{\Delta}_1 \\
&= \left(y_{1t}^{(0)} - \frac{1}{T_2}\sum_{t>T_0} y_{1t}\right) - \widehat{\boldsymbol{\beta}}'\left(\boldsymbol{y}_{0t} - \frac{1}{T_2}\sum_{t>T_0}\boldsymbol{y}_{0t}\right) - \widehat{\gamma}\left(t - \frac{1}{T_2}\sum_{t>T_0} t\right) + \left(\delta_t - \frac{1}{T_2}\sum_{t>T_0}\delta_t\right) \\
&= \dot{y}_{1t}^{(0)} - \widehat{\boldsymbol{\beta}}'\dot{\boldsymbol{y}}_{0t} - \widehat{\gamma}\left(t - \frac{T+T_0+1}{2}\right) + \dot{\delta}_t \\
&= (1, -\widehat{\boldsymbol{\beta}}')\dot{\boldsymbol{y}}_t^{(0)} - \widehat{\gamma}\left(t - \frac{T+T_0+1}{2}\right) + \dot{\delta}_t \\
&\equiv \widehat{\boldsymbol{\theta}}_1'\dot{\boldsymbol{y}}_t^{(0)} - \widehat{\gamma}\left(t - \frac{T+T_0+1}{2}\right) + \dot{\delta}_t,
\end{aligned}$$

where $\dot{\boldsymbol{u}} \equiv \dot{\boldsymbol{u}}(0, \lambda_0, 1)$. Similarly for the specification 2:

$$\begin{aligned}
\widehat{\nu}_{2t} &\equiv y_{1t} - \widehat{\alpha} - \widehat{\boldsymbol{\pi}}'\boldsymbol{y}_{0t} - \widehat{\Delta}_2 \\
&= \left(y_{1t}^{(0)} - \frac{1}{T_2}\sum_{t>T_0} y_{1t}\right) - \widehat{\boldsymbol{\pi}}'\left(\boldsymbol{y}_{0t} - \frac{1}{T_2}\sum_{t>T_0}\boldsymbol{y}_{0t}\right) + \left(\delta_t - \frac{1}{T_2}\sum_{t>T_0}\delta_t\right) \\
&= \dot{y}_{1t}^{(0)} - \widehat{\boldsymbol{\pi}}'\dot{\boldsymbol{y}}_{0t} + \dot{\delta}_t \\
&= (1, -\widehat{\boldsymbol{\pi}})\dot{\boldsymbol{y}}_t^{(0)} + \dot{\delta}_t \\
&\equiv \widehat{\boldsymbol{\theta}}_2'\dot{\boldsymbol{y}}_t^{(0)} + \dot{\delta}_t
\end{aligned}$$

Note that $\dot{\boldsymbol{y}}_{t+k}^{(0)} = \dot{\boldsymbol{y}}_t^{(0)} + \sum_{i=1}^k \boldsymbol{\varepsilon}_{t+i}$, for $t \geq T_0$ and $k \geq 0$ and under $\mathcal{H}_0$ or $\mathcal{H}_1$, $\dot{\delta}_t = 0$, thus:

$$\widehat{\nu}_{1t+k} = \widehat{\nu}_{1t} + \widehat{\boldsymbol{\theta}}_1' \sum_{i=1}^{k} \boldsymbol{\varepsilon}_{t+i} - \widehat{\gamma}k,$$
$$\widehat{\nu}_{2t+k} = \widehat{\nu}_{2t} + \widehat{\boldsymbol{\theta}}_2' \sum_{i=1}^{k} \boldsymbol{\varepsilon}_{t+i},$$

Therefore, for $j \in \{1, 2\}$:

$$\frac{1}{T}\widehat{\rho}_{jk}^2 = \frac{1}{T}\widehat{\rho}_{j0}^2 + \frac{T}{T_2}\widehat{\boldsymbol{\theta}}_j' \boldsymbol{M}_{jk} \widehat{\boldsymbol{\theta}}_j',$$

where

$$\begin{aligned}\boldsymbol{M}_{1k} \equiv & \left(\frac{1}{T^2}\sum_{t=T_0+1}^{T-k} \dot{\boldsymbol{y}}_t \sum_{i=1}^{k} \boldsymbol{\varepsilon}_{t+i}'\right) - (\sqrt{T}\check{\boldsymbol{y}}) \left[\frac{1}{T^{5/2}}\sum_{t=T_0+1}^{T-k}\left(t - \frac{T+T_0+1}{2}\right)\sum_{i=1}^{k}\boldsymbol{\varepsilon}_{t+i}'\right] \\ & - k\left(\frac{1}{T^{5/2}}\sum_{t=T_0+1}^{T-k}\dot{\boldsymbol{y}}_t\right)(\sqrt{T}\check{\boldsymbol{y}})' + k\left[\frac{1}{T^3}\sum_{t=T_0+1}^{T-k}\left(t - \frac{T+T_0+1}{2}\right)\right](\sqrt{T}\check{\boldsymbol{y}})(\sqrt{T}\check{\boldsymbol{y}})' \\ & - \left(\frac{1}{T^2}\sum_{t=T-k+1}^{T}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t'\right) \\ \boldsymbol{M}_{2k} \equiv & \left(\frac{1}{T^2}\sum_{t=T_0+1}^{T-k} \dot{\boldsymbol{y}}_t \sum_{i=1}^{k} \boldsymbol{\varepsilon}_{t+i}'\right) - \left(\frac{1}{T^2}\sum_{t=T-k+1}^{T}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t'\right)\end{aligned}$$

Hence, to show that $\frac{1}{T}\widehat{\rho}_{j0}^2$ and $\frac{1}{T}\widehat{\rho}_{jk}^2$ for $j = \{1, 2\}$, share the same limiting distribution for any k is sufficient to show that $\frac{1}{T}\widehat{\rho}_{j0}^2$ converges in distribution and that $\boldsymbol{M}_{jk} = o_P(1), \forall k$, since $\widehat{\boldsymbol{\theta}}_j$ are shown to be $O_P(1)$. For the first case:

$$\begin{aligned}\frac{1}{T}\widehat{\rho}_{10}^2 &= \frac{1}{TT_2}\sum_{t>T_0}\widehat{\nu}_{1t}^2 \\ &= \frac{1}{TT_2}\left[\widehat{\boldsymbol{\theta}}_1'\left(\sum_{t>T_0}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t'\right)\widehat{\boldsymbol{\theta}}_1 - 2\widehat{\gamma}\widehat{\boldsymbol{\theta}}_1'\sum_{t>T_0}\left(t - \frac{T+T_0+1}{2}\right)\dot{\boldsymbol{y}}_t + \widehat{\gamma}^2\sum_{t>T_0}\left(t - \frac{T+T_0+1}{2}\right)^2\right] \\ &= \frac{T}{T_2}\left[\widehat{\boldsymbol{\theta}}_1'\left(\frac{1}{T^2}\sum_{t>T_0}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t'\right)\widehat{\boldsymbol{\theta}}_1 - 2\widehat{\boldsymbol{\theta}}_1'\left(\frac{1}{T^{5/2}}\sum_{t>T_0}t\dot{\boldsymbol{y}}_t\right)(\sqrt{T}\widehat{\gamma}) + \frac{1}{T^3}\sum_{t>T_0}\left(t - \frac{T+T_0+1}{2}\right)^2(\sqrt{T}\widehat{\gamma})^2\right]\end{aligned}$$

We have from Lemmas 2 and 5 that:

$$\sqrt{T}\widehat{\gamma} \Rightarrow \boldsymbol{f}'\boldsymbol{a}(1, 0, \lambda_0, \boldsymbol{\varepsilon}) \equiv \boldsymbol{f}'\boldsymbol{s}$$
$$\frac{1}{T^{5/2}}\sum_{t>T_0}t\dot{\boldsymbol{y}}_t \Rightarrow \frac{1-\lambda_0^3}{3}\boldsymbol{a}(1, \lambda_0, 1, \boldsymbol{\varepsilon}) - \frac{1-\lambda_0^2}{2}\boldsymbol{a}(0, \lambda_0, 1, \boldsymbol{\varepsilon}) \equiv \boldsymbol{q}$$
$$\frac{1}{T^3}\sum_{t>T_0}\left(t - \frac{T+T_0+1}{2}\right)^2 \to \frac{1-\lambda_0^3}{12} - \frac{\lambda_0(1-\lambda_0)}{4} \equiv \varsigma(\lambda_0).$$

Then, by the continuous mapping theorem

$$\frac{1}{T}\widehat{\rho}_{10}^2 \Rightarrow \frac{1}{1-\lambda_0}\boldsymbol{f}'\left[\boldsymbol{H} - 2\boldsymbol{q}\boldsymbol{s}' + \varsigma(\lambda_0)\boldsymbol{s}\boldsymbol{s}'\right]\boldsymbol{f}.$$

Similarly, for the second specification we have:

$$\begin{aligned}\frac{1}{T}\widehat{\rho}_{20}^2 &= \frac{1}{TT_2}\sum_{t>T_0}\widehat{\nu}_{2t}^2 \\ &= \frac{T}{T_2}\widehat{\boldsymbol{\theta}}_2'\left(\frac{1}{T^2}\sum_{t>T_0}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t'\right)\widehat{\boldsymbol{\theta}}_2 \\ &\Rightarrow \frac{1}{1-\lambda_0}\boldsymbol{g}'\boldsymbol{H}\boldsymbol{g}.\end{aligned}$$

Now we show that $\boldsymbol{M}_{jk} = o_P(1), \forall k, j \in \{1, 2\}$. Clearly the last term of both expressions vanishes in probability as $T \to \infty$. As for the first term in both expressions, note that for each $i \in \{1, \dots, k\}$:

$$\frac{1}{T^2}\sum_{t=T_0+1}^{T-k}\dot{\boldsymbol{y}}_t\boldsymbol{\varepsilon}_{t+i}' = \frac{1}{T}\left[\frac{1}{T}\sum_{t=T_0+1}^{T-k}\boldsymbol{y}_t\boldsymbol{\varepsilon}_{t+i}' - \frac{T}{T_2}\left(\frac{1}{T^{3/2}}\sum_{t=T_0+1}^{T}\boldsymbol{y}_t\right)\left(\frac{1}{\sqrt{T}}\sum_{t=T_0+1}^{T-k}\boldsymbol{\varepsilon}_{t+i}'\right)\right],$$

and we have shown that the first, second and third terms inside the brackets of the expressions above are $O_P(1)$ by Lemma 5 result (c),(b) and (a), respectively. Finally, the remainder terms of $\boldsymbol{M}_{1k}$ are all $o_P(1)$ simply by applying the convergence results presented in Lemma 5. Therefore, we have proved part (a) and (b).

For parts (c) and (d), since $\widehat{\rho}_{jk} = \widehat{\rho}_{j-k}$ and the covariance kernels are normalized such that $\phi(0) = 1$, we write:

$$\begin{aligned}\frac{1}{J_T T}\widehat{\sigma}_j^2 &\equiv \frac{1}{J_T T}\widehat{\rho}_{j0}^2 + 2\frac{1}{J_T}\sum_{k=1}^{T-1}\phi\left(\frac{k}{J_T}\right)\frac{1}{T}\widehat{\rho}_{jk}^2 \\ &= \frac{1}{J_T T}\widehat{\rho}_{j0}^2 + 2\frac{1}{J_T}\sum_{k=1}^{T-1}\phi\left(\frac{k}{J_T}\right)\left(\frac{1}{T}\widehat{\rho}_{j0}^2 + \frac{T}{T_2}\widehat{\boldsymbol{\theta}}_j'\boldsymbol{M}_{jk}\widehat{\boldsymbol{\theta}}_j'\right) \\ &= \left(\frac{1}{T}\widehat{\rho}_{j0}^2\right)\left[\frac{1}{J_T}\sum_{|k|<T}\phi\left(\frac{k}{J_T}\right)\right] + 2\frac{T}{T_2}\widehat{\boldsymbol{\theta}}_j'\left[\frac{1}{J_T}\sum_{k=1}^{T-1}\phi\left(\frac{k}{J_T}\right)\boldsymbol{M}_{jk}\right]\widehat{\boldsymbol{\theta}}_j,\end{aligned}$$

The first term in parentheses converges in distribution as shown above, the second converges to C_ϕ by Assumption, hence it is left to show that the term in brackets of the expression above are $o_P(1)$ since $\widehat{\boldsymbol{\theta}}_j$ is $O_P(1)$. We show that convergence in probability using the Markov's inequality and the fact that $\mathbb{E}\|\boldsymbol{M}_{j,k}\|$ can be bounded by a positive decreasing sequence. We show for the second specification ($j = 2$), the argument is entirely analogous to the first one. First, we need the following bounds

$$\begin{aligned}\mathbb{E}\|\boldsymbol{P}_{jt,T}\| \le b_p < \infty \quad \forall j, t \le T, T, \quad \boldsymbol{P}_{jt,T} \equiv \frac{1}{T}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t',\\ \mathbb{E}\|\boldsymbol{R}_{jt,T}(i)\| \le \bar{b}_T < \infty \quad \forall j, t \le T, i, \quad \boldsymbol{R}_{jt,T} \equiv \frac{1}{T}\dot{\boldsymbol{y}}_t\boldsymbol{\varepsilon}_t'.\end{aligned}$$

Assuming $\boldsymbol{y}_0 = \boldsymbol{0}$ we can write

$$\dot{\boldsymbol{y}}_t = \sum_{s=1}^{t}\left(\frac{s-1}{T}\right)\boldsymbol{\varepsilon}_s \equiv \sum_{s=1}^{t}g_1(s,T)\boldsymbol{\varepsilon}_s$$

Since the function $g_1(\cdot,\cdot)$ is bounded between 0 and 1 we can write

$$
\begin{aligned}
\mathbb{E}\|\boldsymbol{P}_{jt,T}\| = \mathbb{E}\|T^{-1}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t'\| &= \mathbb{E}\left\|T^{-1}\sum_{s=1}^{T} g_1(s,T)\boldsymbol{\varepsilon}_s \sum_{s=1}^{T} g_1(s,T)\boldsymbol{\varepsilon}_s'\right\| \\
&= \mathbb{E}\left\|T^{-1}\sum_{s=1}^{t}\sum_{l=1}^{t} g_1(s,T)g_1(l,T)\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_l'\right\| \\
&\leq T^{-1}\sum_{s=1}^{t}\sum_{l=1}^{t} g_1(s,T)g_1(l,T)\mathbb{E}\left\|\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_l'\right\| \\
&\leq T^{-1}\sum_{s=1}^{t}\sum_{l=1}^{t}\mathbb{E}\left\|\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_l'\right\| \\
&\leq T^{-1}\sum_{s=1}^{T}\sum_{l=1}^{T}\mathbb{E}\left\|\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_l'\right\| \\
&\leq \lim_{T\to\infty} T^{-1}\sum_{s=1}^{T}\sum_{l=1}^{T}\mathbb{E}\left\|\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_l'\right\| \equiv b_p,
\end{aligned}
$$

where the last limit exists under Assumptions (a)-(c) of Lemma 3. For the second bound we have

$$
\begin{aligned}
\mathbb{E}\|\boldsymbol{R}_{jt,T(i)}\| = \mathbb{E}\|T^{-1}\ddot{\boldsymbol{y}}_t\boldsymbol{\varepsilon}_{t+i}'\| &= \mathbb{E}\left\|T^{-1}\sum_{s=1}^{T} g_1(s,T)\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_{t+i}'\right\| \\
&\leq T^{-1}\sum_{s=1}^{t} g_1(s,T)\mathbb{E}\left\|\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_{t+i}'\right\| \\
&\leq T^{-1}\sum_{s=1}^{t}\mathbb{E}\left\|\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_{t+i}'\right\| \\
&\leq T^{-1}\sum_{s=1}^{T}\mathbb{E}\left\|\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_{T+i}'\right\|.
\end{aligned}
$$

Note that the last term above is $o_P(1)$ because the summation is finite due to Assumptions (a)-(c) of Lemma 3. Thus, for a fixed T and i there exist a bound $b_T(i)$ such that $\mathbb{E}\|\boldsymbol{R}_{jt,T(i)}\| \leq b_T(i) < \infty$ for every $t \leq T$ and $b_T(i) \to \infty$. Moreover, due to the mixing condition (Lemma 3(c)) we know that when $i = 1$ we have the largest bounds over all i for a given T so we define $\bar{b}_T \equiv b_T(1)$.

Now we show $\mathcal{L}_p$ convergence so for any $\epsilon > 0$. Let

$$
\mathscr{A}_T = \left\{\omega \in \Omega : \left\|\frac{1}{T-T_0}\sum_{k=1}^{T-1}\phi\left(\frac{k}{J_T}\right)\sum_{t=T-k+1}^{T}\boldsymbol{P}_{jt,T}(\omega)\right\| > \epsilon\right\} \quad \text{and}
$$
$$
\mathscr{B}_T = \left\{\omega \in \Omega : \left\|\frac{1}{T-T_0}\sum_{k=1}^{T-1}\phi\left(\frac{k}{J_T}\right)\sum_{t=T_0+1}^{T-k}\sum_{i=1}^{k}\boldsymbol{R}_{jt,T}(i)(\omega)\right\| > \epsilon\right\}.
$$

For $\mathscr{A}_T$ by the Markov's inequality

$$\begin{aligned}
\mathbb{P}(\mathscr{A}_T) &\leq \frac{1}{\epsilon}\mathbb{E}\left\|\frac{1}{T-T_0}\sum_{k=1}^{T-1}\phi\left(\frac{k}{J_T}\right)\sum_{t=T-k+1}^{T}\boldsymbol{P}_{jt,T}\right\| \\
&\leq \frac{1}{(T-T_0)\epsilon}\sum_{k=1}^{T-1}\left|\phi\left(\frac{k}{J_T}\right)\right|\sum_{t=T-k+1}^{T}\mathbb{E}\left\|\boldsymbol{P}_{jt,T}\right\| \\
&\leq \frac{1}{(T-T_0)\epsilon}\sum_{k=1}^{T-1}\left|\phi\left(\frac{k}{J_T}\right)\right|\sum_{t=T-k+1}^{T} b_p \\
&\leq \frac{b_p}{(T-T_0)\epsilon}\sum_{k=1}^{T-1}k\left|\phi\left(\frac{k}{J_T}\right)\right|.
\end{aligned}$$

Note that the kernels are uniformly bounded such that for non-negative integer h:

$$\lim_{T\to\infty}\frac{1}{J_T^{h+1}}\sum_{|k|<T}\left|\phi\left(\frac{k}{J_T}\right)\right| = C_h \quad \text{where } C_h \equiv \int_{-\infty}^{\infty} x^h\left|\phi\left(x\right)\right|\mathrm{d}x.$$

As a result, as long as $J_T = o(T^{1/2})$ we have

$$\mathbb{P}(\mathscr{A}_T) \leq \frac{b_p}{\epsilon}\frac{T}{T-T_0}\frac{J_T^2}{T}\left[J_T^{-2}\sum_{k=1}^{T-1}k\left|\phi\left(\frac{k}{J_T}\right)\right|\right] \to 0.$$

For $\mathscr{B}_T$, by the Markov's inequality

$$\mathscr{A}_T = \left\{\omega\in\Omega : \left\|\frac{1}{T-T_0}\sum_{k=1}^{T-1}\phi\left(\frac{k}{J_T}\right)\sum_{t=T-k+1}^{T}\boldsymbol{P}_{jt,T}(\omega)\right\| > \epsilon\right\} \quad \text{and}$$

$$\mathscr{B}_T = \left\{\omega\in\Omega : \left\|\frac{1}{T-T_0}\sum_{k=1}^{T-1}\phi\left(\frac{k}{J_T}\right)\sum_{t=T_0+1}^{T-k}\sum_{i=1}^{k}\boldsymbol{R}_{jt,T}(i)(\omega)\right\| > \epsilon\right\}$$

For $\mathscr{A}_T$, by the Markov's inequality

$$\begin{aligned}
\mathbb{P}(\mathscr{B}_T) &\leq \frac{1}{\epsilon}\mathbb{E}\left\|\frac{1}{T-T_0}\sum_{k=1}^{T-1}\phi\left(\frac{k}{J_T}\right)\sum_{t=T_0+1}^{T-k}\sum_{i=1}^{k}\boldsymbol{R}_{jt,T}(i)\right\| \\
&\leq \frac{1}{(T-T_0)\epsilon}\sum_{k=1}^{T-1}\left|\phi\left(\frac{k}{J_T}\right)\right|\sum_{t=T_0+1}^{T-k}\sum_{i=1}^{k}\mathbb{E}\left\|\boldsymbol{R}_{jt,T}(i)\right\| \\
&\leq \frac{\bar{b}_T}{\epsilon}\sum_{k=1}^{T-1}k\left|\phi\left(\frac{k}{J_T}\right)\right| \\
&\leq \frac{1}{\epsilon}(T\bar{b}_T)\frac{J_T^2}{T}\left(\frac{1}{J_T^2}\sum_{k=1}^{T-1}k\left|\phi\left(\frac{k}{J_T}\right)\right|\right) \to 0.
\end{aligned}$$

The last inequality holds because by definition $\lim_{T\to\infty} T\bar{b}_T = \lim_{T\to\infty}\sum_{t=1}^{T}\mathbb{E}\|\varepsilon_t, \varepsilon_{T+1}\| < \infty$ and under assumption that $J_T = o(T^{1/2})$.

□

Proof of Theorem 4

For both specification $j = \{1, 2\}$, we have:

$$\sqrt{\frac{J_T}{T}}\tau_j \equiv \sqrt{\frac{J_T T_2}{T}}\frac{\widehat{\Delta}_j}{\widehat{\sigma}_j} = \sqrt{\frac{T_2}{T}}\left[\frac{\frac{1}{\sqrt{T}}(\widehat{\Delta}_j - \Delta_T)}{\frac{1}{\sqrt{TJ_T}}\widehat{\sigma}_j}\right] + \frac{\frac{1}{\sqrt{T}}\Delta_T}{\frac{1}{\sqrt{TJ_T}}\widehat{\sigma}_j}.$$

As long as $\Delta_T = o(\sqrt{T})$, we have that the second term in last expression is $o_P(1)$. The result than follows from Theorem 2, Lemma 4 and the continuous mapping theorem.

References

Abadie, A., A. Diamond, and J. Hainmueller (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.

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

Andrews, D. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Athey, S., and G. Imbens (2016): "The State of Applied Econometrics - Causality and Policy Evaluation," Discussion paper, arXiv:1607.00699v1.

Bai, C.-E., Q. Li, and M. Ouyang (2014): "Property taxes and home prices: A tale of two cities," *Journal of Econometrics*, 180, 1–15.

Carvalho, C., R. Masini, and M. Medeiros (2016): "ArCo: An Artificial Counterfactual Approach for High-Dimensional Data," Working paper, Pontifical Catholic University of Rio de Janeiro.

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

Doudchenko, N., and G. Imbens (2016): "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis," 22791, NBER.

Du, Z., and L. Zhang (2015): "Home-purchase restriction, property tax and housing price in China: A counterfactual analysis," *Journal of Econometrics*, 188, 558–568.

Durlauf, S., and P. Phillips (1985): "Multiple time series regression with integrated processes," *Review of Economic Studies*, 53, 473–495.

Engle, R., and C. Granger (1987): "Co-integration and error correction: Representation, estimation, and testing," *Econometrica*, 55, 251–276.

Fujiki, H., and C. Hsiao (2015): "Disentangling the effects of multiple treatments Measuring the net economic impact of the 1995 great Hanshin-Awaji earthquake," *Journal of Econometrics*, 186, 66–73.

Gao, Y., W. Long, and Z. Wang (2015): "Estimating average treatment effect by model averaging," *Economics Letters*, 135, 42–45.

Hansen, B. (1992): "Convergence to Stochastic Integrals for ependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500.

Hsiao, C., H. S. Ching, and S. K. Wan (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.

Ibragimov, I., and V. Linnik (1971): *Wolters-Noordhoff series of monographs and textbooks on pure and applied mathematics.s*chap. Independent and stationary sequences of random variables.

Kurtz, T., and P. Protter (1991): “Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations,” *The Annals of Probability*, 19, 1035–1070.

Newey, W., and K. West (1987): “A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica*, 55, 703–708.

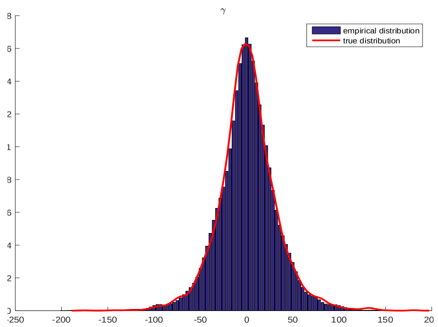
Ouyang, M., and Y. Peng (2015): “The treatment-effect estimation: A case study of the 2008 economic stimulus package of China,” *Journal of Econometrics*, 188, 545–557.

Phillips, P. (1986a): “Understanding spurious regressions in econometrics,” *Journal of Econometrics*, 33, 311–340.

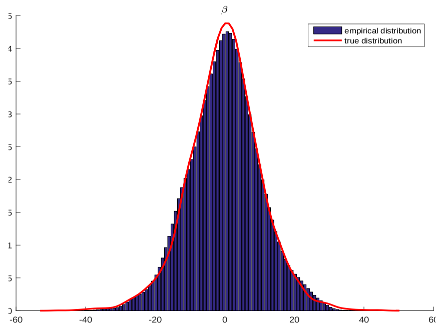
——— (1986b): “Weak Convergence to the Matrix Stochastic Integral BdB,” Cowles Foundation Discussion Papers 796, Cowles Foundation for Research in Economics, Yale University.

Tibshirani, R. (1996): “Regression shrinkage and selection via the LASSO,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 267–288.

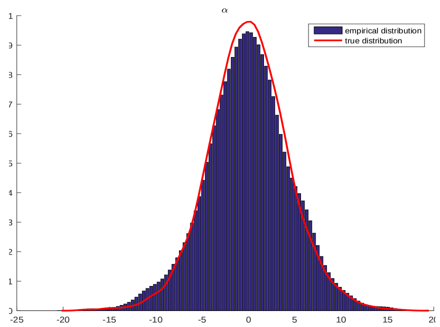
Xu, Y. (2015): “Generalized Synthetic Control Method for Causal Inference with Time-Series Cross-Sectional Data,” Working paper, Massachusetts Institute of Technology.



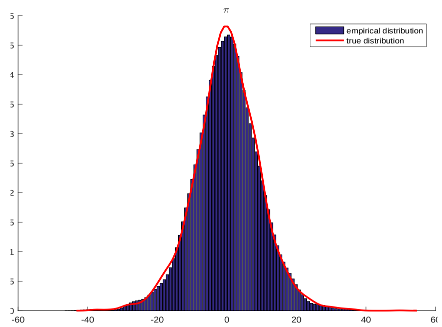
(a) γ, $T = 100$



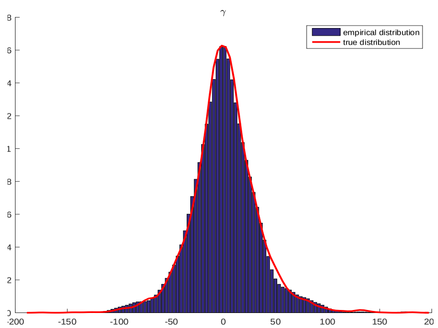
(b) β, $T = 100$



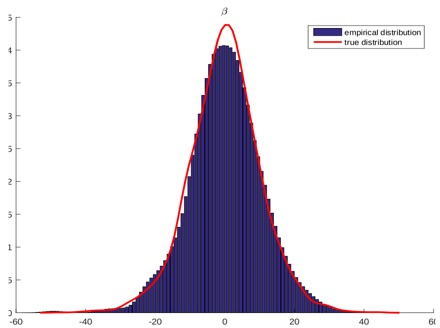
(c) α, $T = 100$



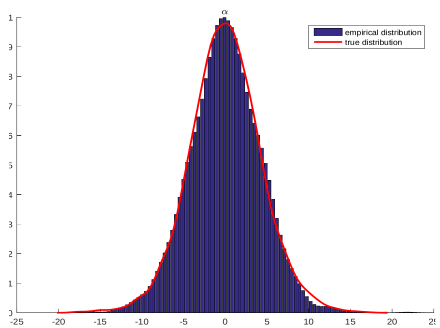
(d) π, $T = 100$



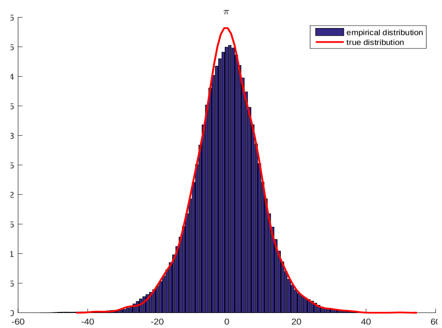
(e) γ, $T = 100$



(f) β, $T = 100$

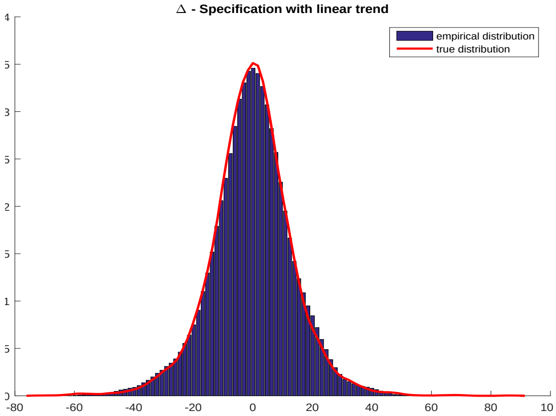


(g) α, $T = 1000$

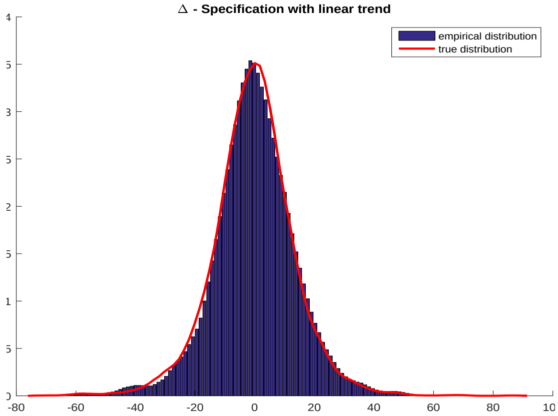


(h) π, $T = 1000$

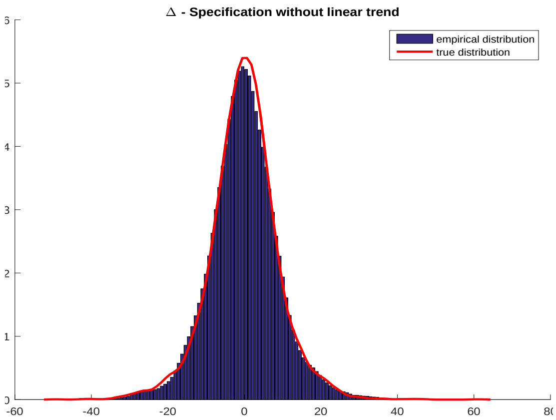
Figure 1: Empirical (bars) and asymptotic (solid line) distributions of the estimates of regression coefficients in the case of cointegration for $T = 100$ and $T = 1000$ as in Lemma 1. The distributions are scaled as in the lemma. Panel (a): $T^{3/2}(\widehat{\gamma} - \gamma_0)$, $T = 100$; Panel (b): first element of $T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, $T = 100$; Panel (c): $\sqrt{T}(\widehat{\alpha} - \alpha_0)$, $T = 100$; Panel (d): first element of $T(\widehat{\pi} - \boldsymbol{\beta}_0)$, $T = 100$; Panel (e): $T^{3/2}(\widehat{\gamma} - \gamma_0)$, $T = 1000$; Panel (f): first element of $T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$, $T = 1000$; Panel (g): $\sqrt{T}(\widehat{\alpha} - \alpha_0)$, $T = 1000$; and Panel (h): first element of $T(\widehat{\pi} - \boldsymbol{\beta}_0)$, $T = 1000$.



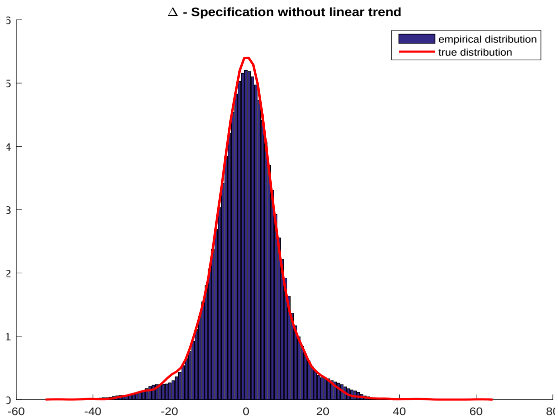
(a) $T = 100$, with trend



(b) $T = 1000$, with trend

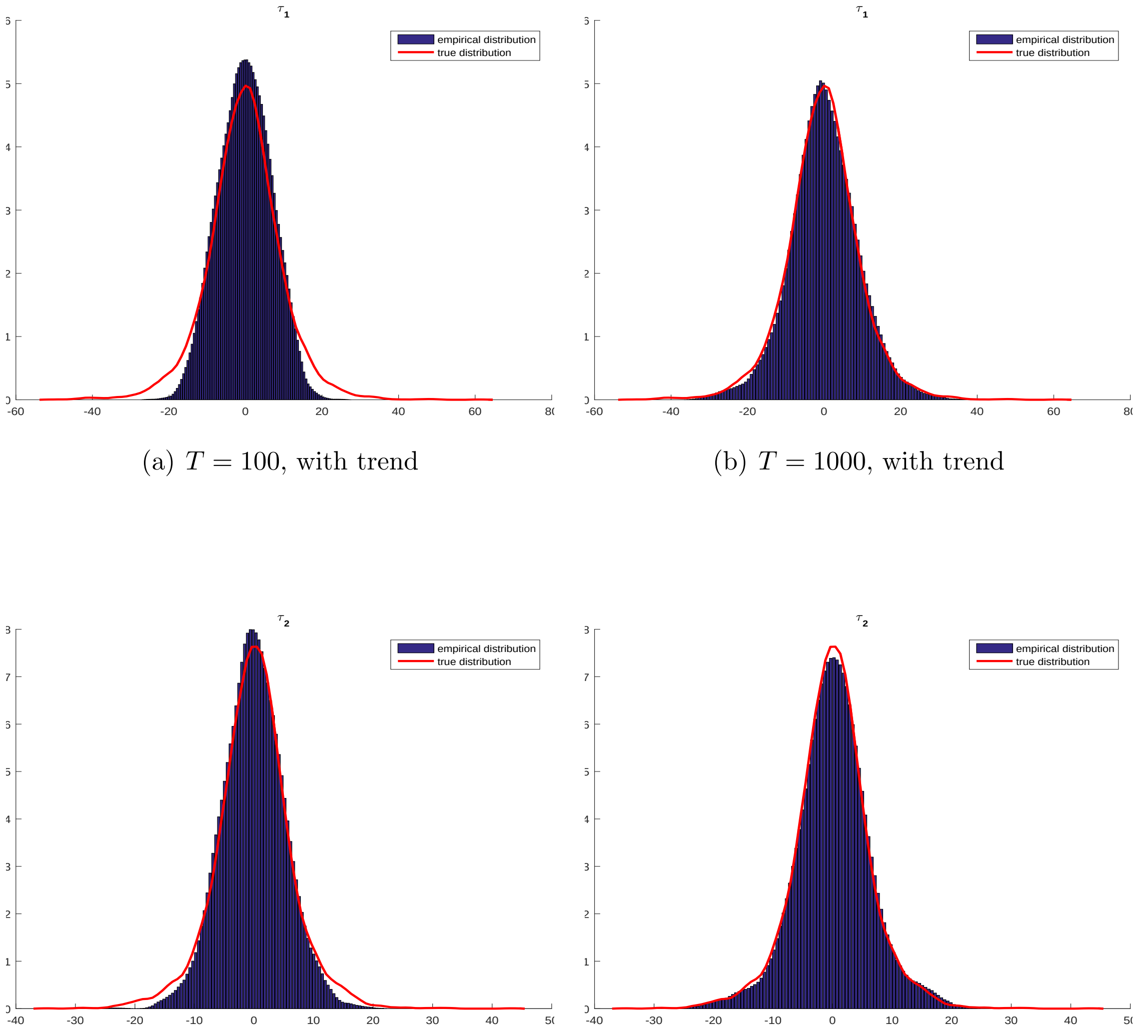


(c) $T = 100$, without trend



(d) $T = 1000$, without trend

Figure 2: Empirical (bars) and asymptotic (solid line) distributions of the counterfactual effects in the cointegration case as stated in Theorem 1 for $T = 100$ and $T = 1000$. The distributions are scaled as in the Theorem. Panel (a): trend included in the estimated equation and $T = 100$; Panel (b): trend included in the estimated equation and $T = 1000$; Panel (c): trend excluded from the estimated equation and $T = 100$; Panel (d): trend excluded from the estimated equation and $T = 1000$.



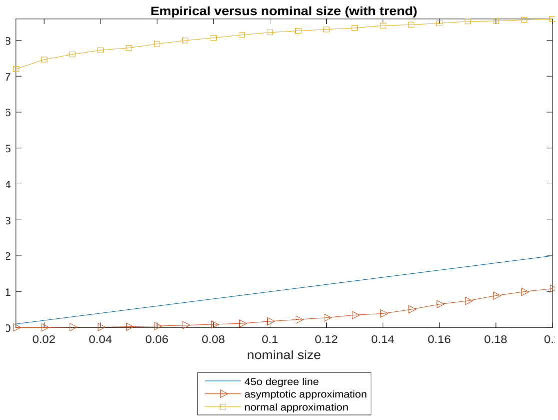
(a) $T = 100$, with trend

(b) $T = 1000$, with trend

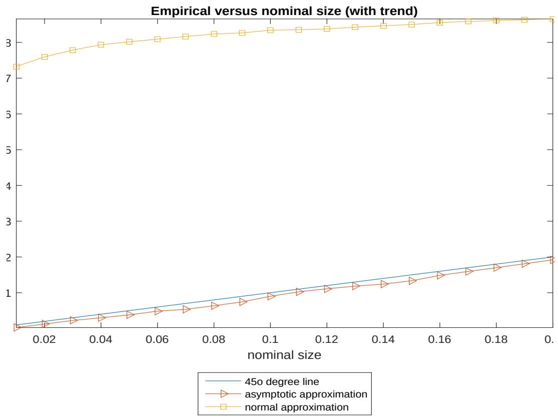
(c) $T = 100$, without trend

(d) $T = 1000$, without trend

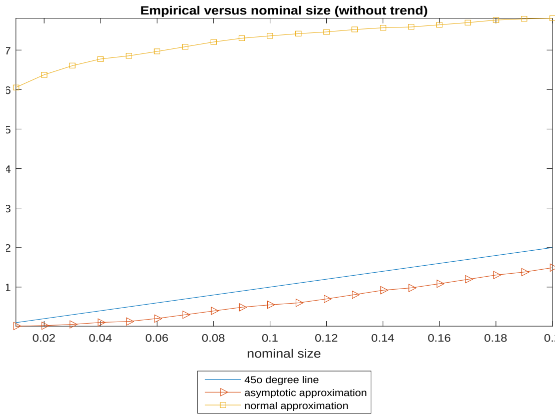
Figure 3: Empirical (bars) and asymptotic (solid line) distributions of the t-statistics in the cointegration case as stated in Theorem 3 for $T = 100$ and $T = 1000$. Panel (a): trend included in the estimated equation and $T = 100$; Panel (b): trend included in the estimated equation and $T = 1000$; Panel (c): trend excluded from the estimated equation and $T = 100$; Panel (d): trend excluded from the estimated equation and $T = 1000$.



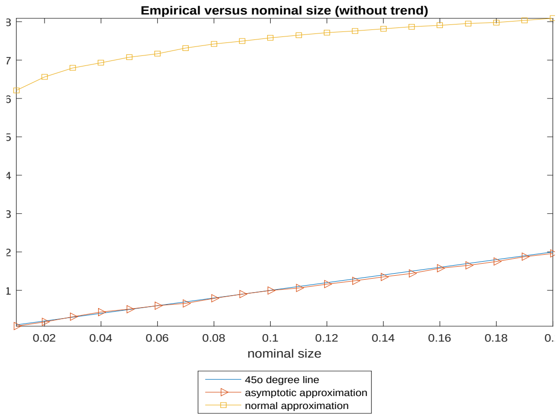
(a) $T = 100$, with trend



(b) $T = 1000$, with trend

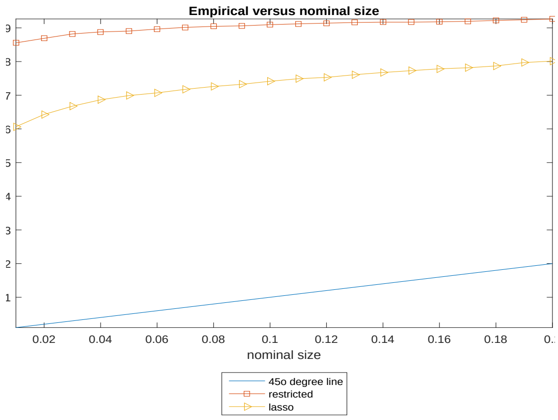


(c) $T = 100$, without trend

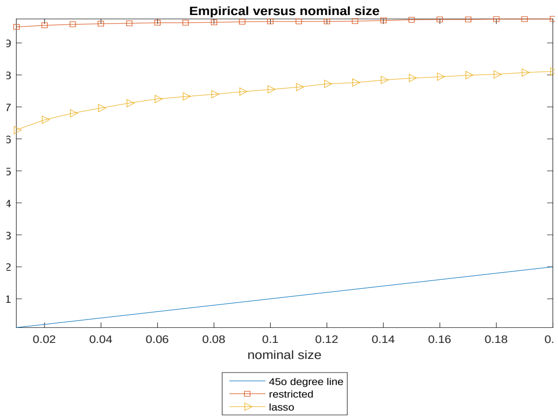


(d) $T = 1000$, without trend

Figure 4: Size distortion plots of the t-test in the cointegration case for $T = 100$ and $T = 1000$ under the asymptotic approximation for the t-statistic distribution (lines with triangles) and the normal approximation (lines with squares). Panel (a): trend included in the estimated equation and $T = 100$; Panel (b): trend included in the estimated equation and $T = 1000$; Panel (c): trend excluded from the estimated equation and $T = 100$; Panel (d): trend excluded from the estimated equation and $T = 1000$. The horizontal axis represents the nominal size and the vertical axis represents the empirical size.

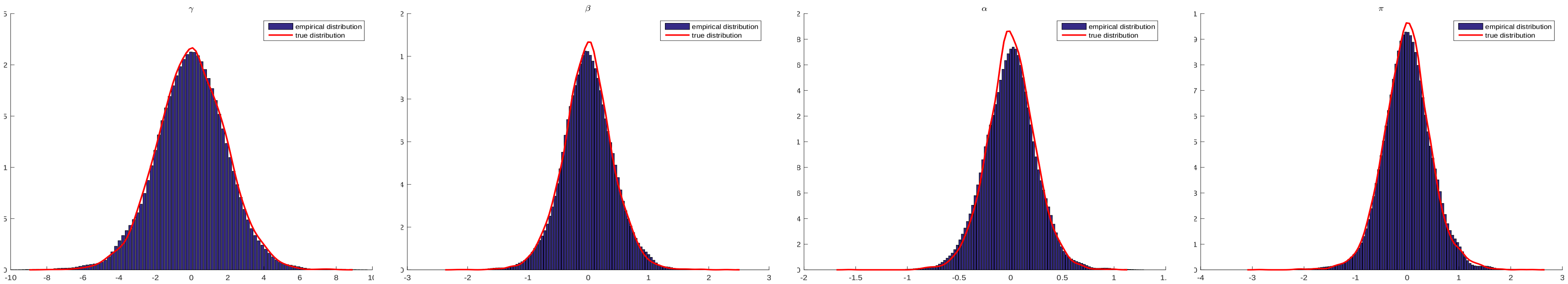


(a) $T = 100$, with trend

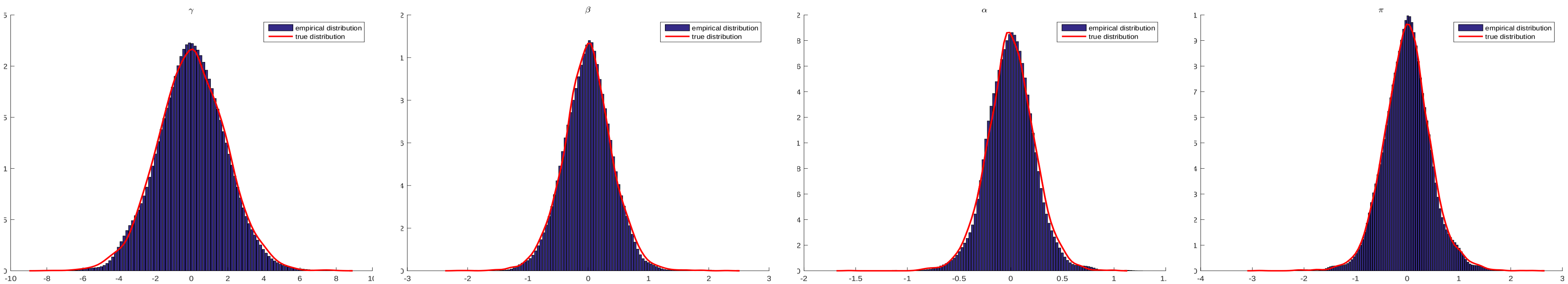


(b) $T = 1000$, with trend

Figure 5: Empirical rejection rates (size) in the cointegration case when the coefficients of the linear combination of peers is restricted. Two different sample sizes are considered: $T = 100$ (panel (a)) and $T = 1000$ (panel (b)).

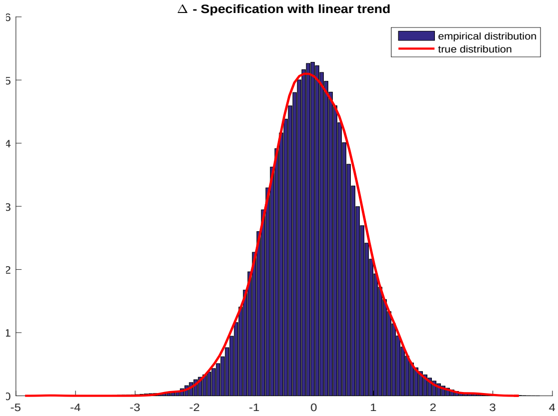


(a) γ, $T = 100$ (b) β, $T = 100$ (c) α, $T = 100$ (d) π, $T = 100$

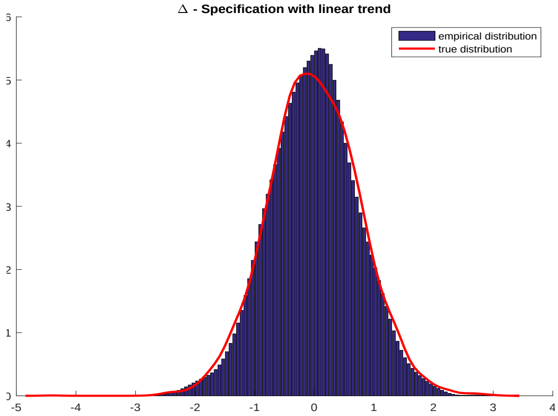


(e) γ, $T = 1000$ (f) β, $T = 1000$ (g) α, $T = 1000$ (h) π, $T = 1000$

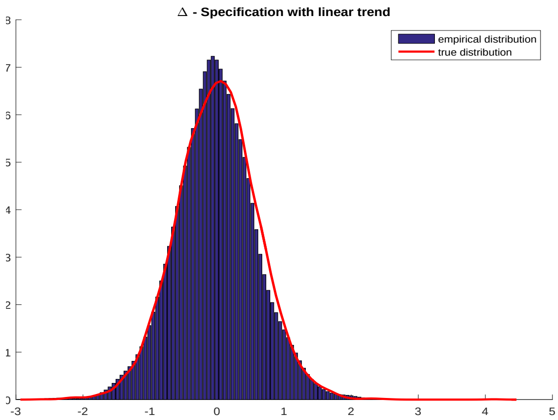
Figure 6: Empirical (bars) and asymptotic (solid line) distributions of the estimates of regression coefficients in the spurious case for $T = 100$ and $T = 1000$ as in Lemma 2. The distributions are scaled as in the lemma. Panel (a): $\sqrt{T}\widehat{\gamma}$, $T = 100$; Panel (b): first element of $\widehat{\boldsymbol{\beta}}$, $T = 100$; Panel (c): $T^{-1/2}\widehat{\alpha}$, $T = 100$; Panel (d): first element of $\widehat{\pi}$, $T = 100$; Panel (e): $\sqrt{T}\widehat{\gamma}$, $T = 1000$; Panel (f): first element of $\widehat{\boldsymbol{\beta}}$), $T = 1000$; Panel (g): $T^{-1/2}\widehat{\alpha}$, $T = 1000$; and Panel (h): first element of $\widehat{\pi}$, $T = 1000$.



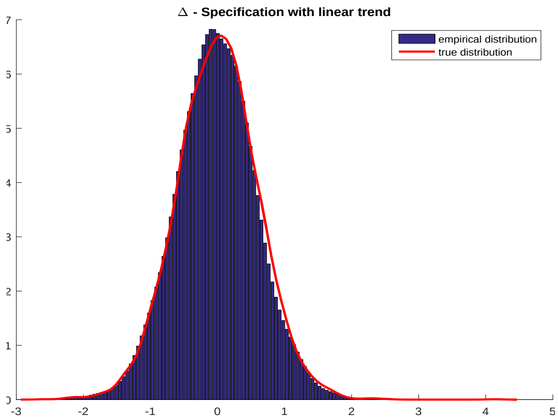
(a) $T = 100$, with trend



(b) $T = 1000$, with trend

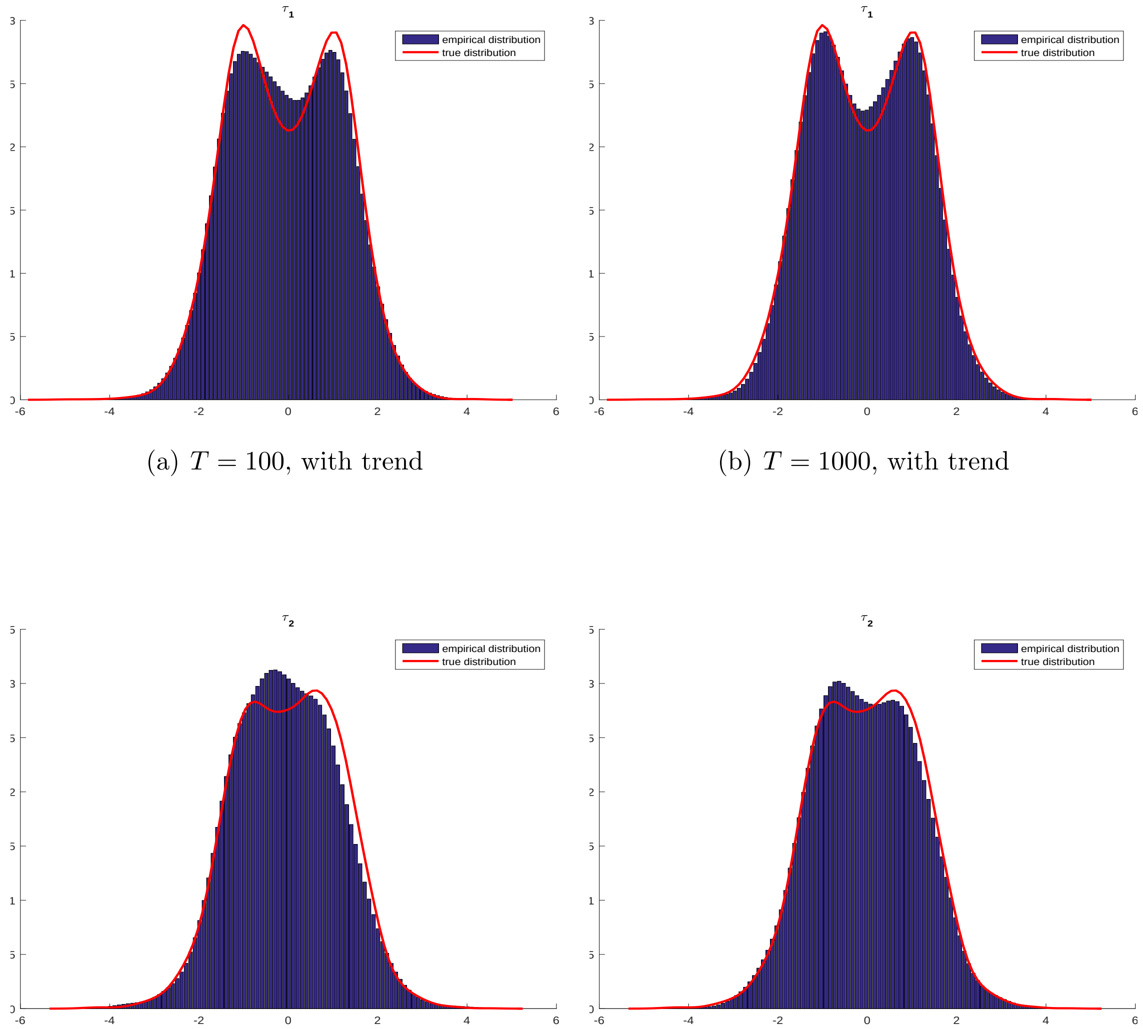


(c) $T = 100$, without trend



(d) $T = 1000$, without trend

Figure 7: Empirical (bars) and asymptotic (solid line) distributions of the counterfactual effects in the spurious case as stated in Theorem 2 for $T = 100$ and $T = 1000$. The distributions are scaled as in the Theorem. Panel (a): trend included in the estimated equation and $T = 100$; Panel (b): trend included in the estimated equation and $T = 1000$; Panel (c): trend excluded from the estimated equation and $T = 100$; Panel (d): trend excluded from the estimated equation and $T = 1000$.



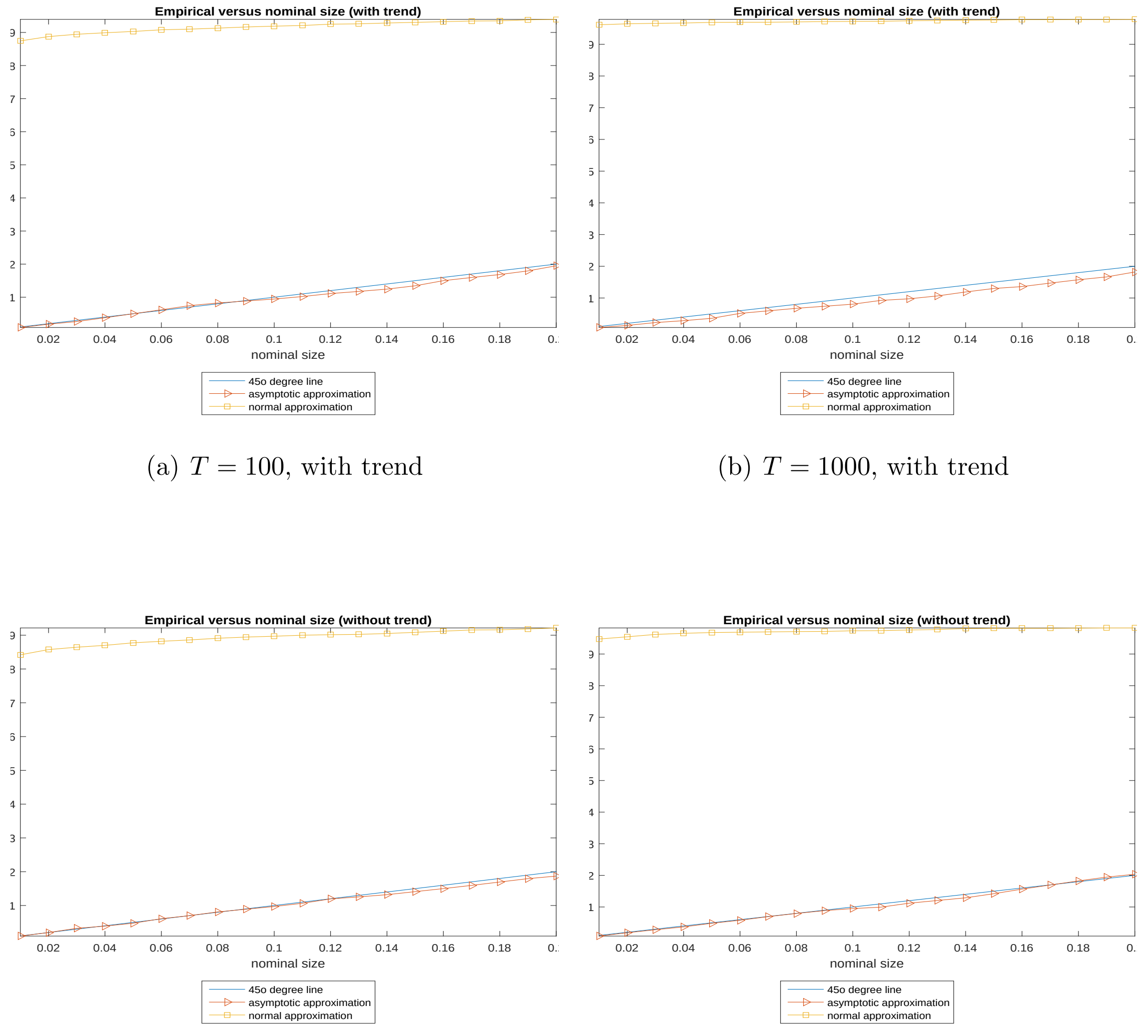
(a) $T = 100$, with trend

(b) $T = 1000$, with trend

(c) $T = 100$, without trend

(d) $T = 1000$, without trend

Figure 8: Empirical (bars) and asymptotic (solid line) distributions of the t-statistics in the spurious case as stated in Theorem 4 for $T = 100$ and $T = 1000$. The distributions are scaled as in the Theorem. Panel (a): trend included in the estimated equation and $T = 100$; Panel (b): trend included in the estimated equation and $T = 1000$; Panel (c): trend excluded from the estimated equation and $T = 100$; Panel (d): trend excluded from the estimated equation and $T = 1000$.



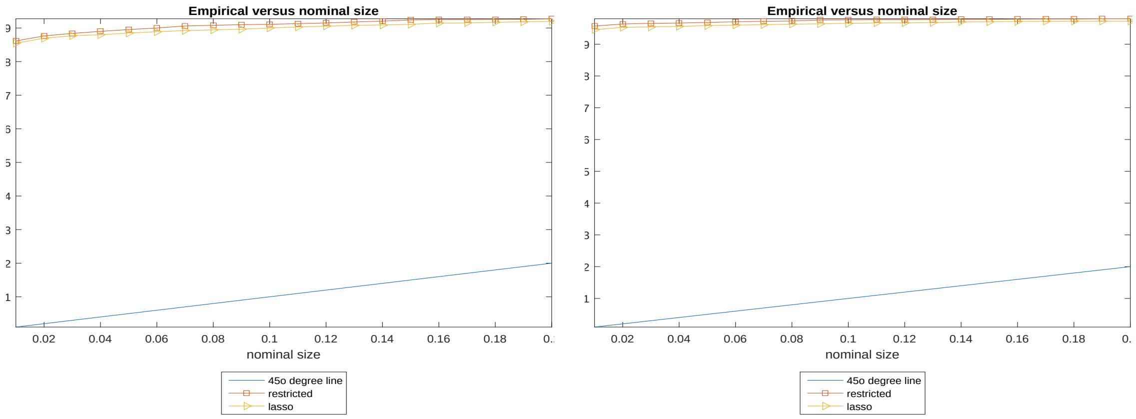
(a) $T = 100$, with trend

(b) $T = 1000$, with trend

(c) $T = 100$, without trend

(d) $T = 1000$, without trend

Figure 9: Size distortion plots of the **scaled** t-test in the cointegration case for $T = 100$ and $T = 1000$ under the asymptotic approximation for the **scaled** t-statistic distribution (lines with triangles) and the normal approximation (lines with squares). Panel (a): trend included in the estimated equation and $T = 100$; Panel (b): trend included in the estimated equation and $T = 1000$; Panel (c): trend excluded from the estimated equation and $T = 100$; Panel (d): trend excluded from the estimated equation and $T = 1000$. The horizontal axis represents the nominal size and the vertical axis represents the empirical size.



(a) $T = 100$, with trend

(b) $T = 1000$, with trend

Figure 10: Empirical rejection rates (size) in the spurious case when the coefficients of the linear combination of peers is restricted. Two different sample sizes are considered: $T = 100$ (panel (a)) and $T = 1000$ (panel (b)).