Supplementary Material

Counterfactual Analysis and Inference with Nonstationary Data

Ricardo P. Masini and Marcelo C. Medeiros

S.1 Introduction

In this online supplement we report additional material to support the results in the paper "Counterfactual Analysis and Inference with Nonstationary Data". In Section S.2 we state and prove auxiliary lemmas used to derive the theoretical results in the paper. In Section S.3 we present additional simulation results.

S.2 Auxiliary Lemmas

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

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

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 $\varrho = 0, 1, 2, \ldots$

First we state and prove in the Lemma below several convergence results that will be applied in the subsequent proofs. We also use throughout the following well-known results from power series

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

to show that

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

Lemma 1. *Let the sequence $\{\boldsymbol{u}_t\}_{t=1}^T$ be defined as above. If the process $\{\boldsymbol{\eta}_t\}$ satisfies Assumption 3, then as $T \to \infty$:*

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

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

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

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

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

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

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,

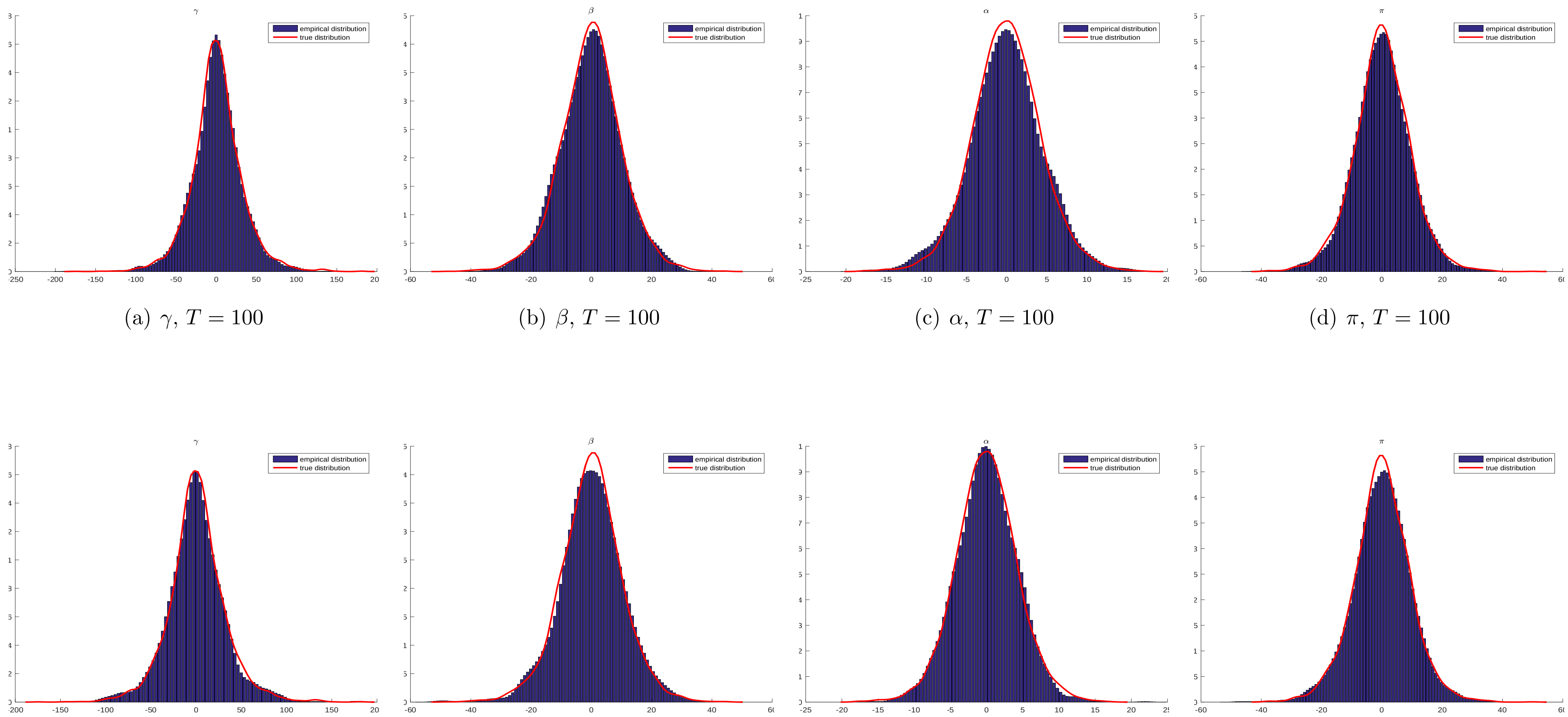
$$\frac{1}{T^2}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{u}_{t+k}' = \frac{1}{T^2}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\boldsymbol{u}_t' + \mathrm{sgn}(k)\frac{1}{T^2}\sum_{(\lambda,\lambda']}\boldsymbol{u}_t\sum_{i=1}^{|k|}\boldsymbol{\eta}_{t+i}'.$$

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. □

S.3 Simulation Results

S.3.1 Asymptotic Distributions

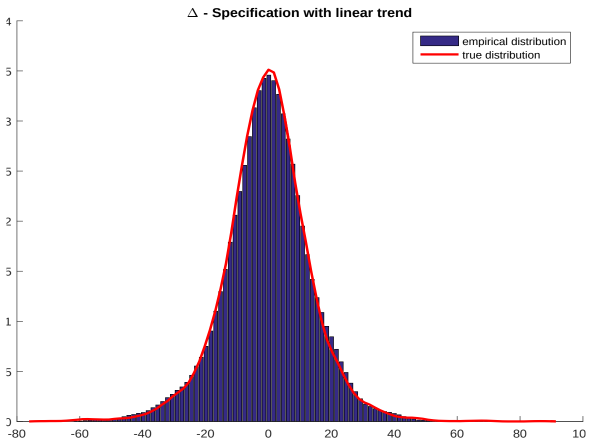
To evaluate the asymptotic approximation in finite samples, we simulate two different scenarios. In the first one, the treated unit and the peers are cointegrated while in the second case the data are formed by a set of independent random walks. In both cases we evaluate the distribution of the estimator for the average intervention effect under the null hypothesis of no intervention at $T_0 = T/2$. We consider $T = 100$ and $1,000$, and $n = 5$. The number of Monte Carlo simulations is set to 10,000. For each scenario and different sample sizes, we report the finite sample distributions of $\widehat{\Delta} = \frac{1}{T-T_0}\sum_{t=T_0+1}^{T}\widehat{\delta}_t$, in comparison to the asymptotic distributions as well as the rejection frequencies, at different significance levels, of the null hypothesis of no intervention effects when nonstationarity is neglected and the test is carried out under standard normal approximation for the t-statistic. As a complement we also report the empirical rejection rates for the t-test of no intervention effect when the parameters are estimated either by restricted least squares or by LASSO.



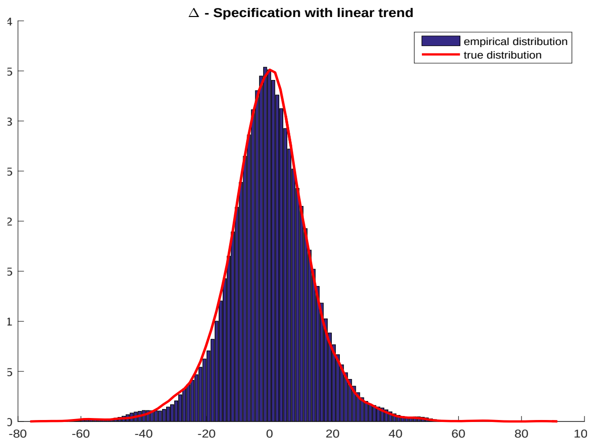
(a) γ, $T = 100$ (b) β, $T = 100$ (c) α, $T = 100$ (d) π, $T = 100$

(e) γ, $T = 100$ (f) β, $T = 100$ (g) α, $T = 1000$ (h) π, $T = 1000$

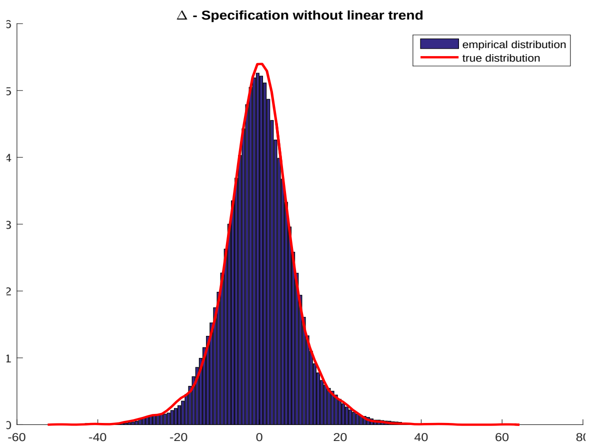
Figure S.1: Empirical (bars) and asymptotic (solid line) distributions of the estimates of regression coefficients in the case of cointegration for $T = 100$ and $T = 1000$. The distributions are scaled as in the lemma. Panel (a): $T^{3/2}(\widehat{\gamma} - \gamma_0)$, $T = 100$; Panel (b): first element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



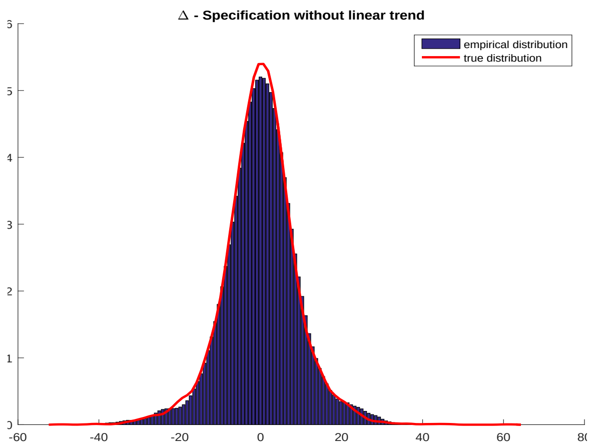
(a) $T = 100$, with trend



(b) $T = 1000$, with trend

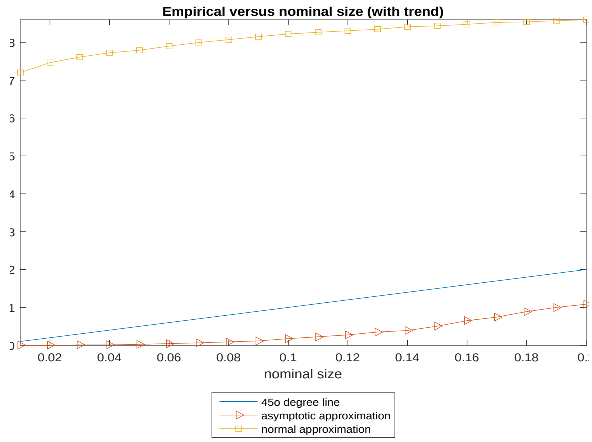


(c) $T = 100$, without trend

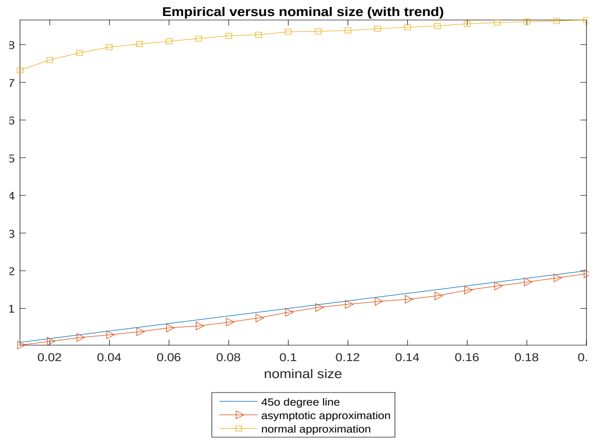


(d) $T = 1000$, without trend

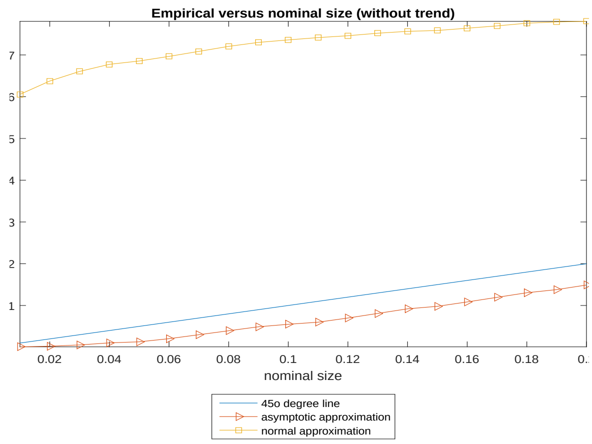
Figure S.2: Empirical (bars) and asymptotic (solid line) distributions of the counterfactual effects in the cointegrating case for $T = 100$ and $T = 1000$. The distributions are scaled as in the Theorem. Panel (a): trend included in the estimated equation and $T = 100$; Panel (b): trend included in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



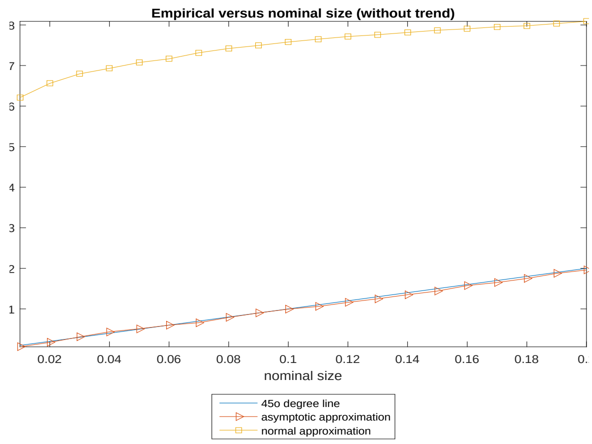
(a) $T = 100$, with trend



(b) $T = 1000$, with trend

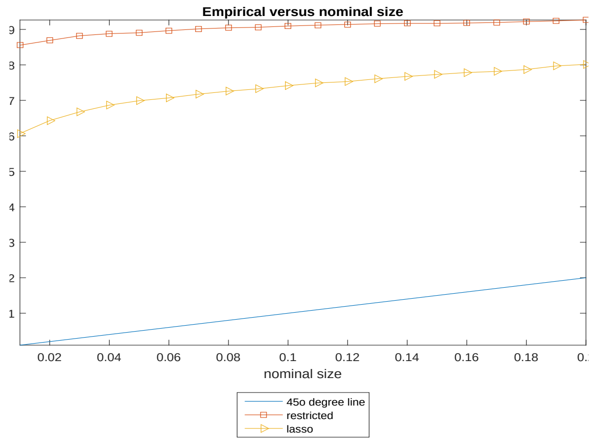


(c) $T = 100$, without trend

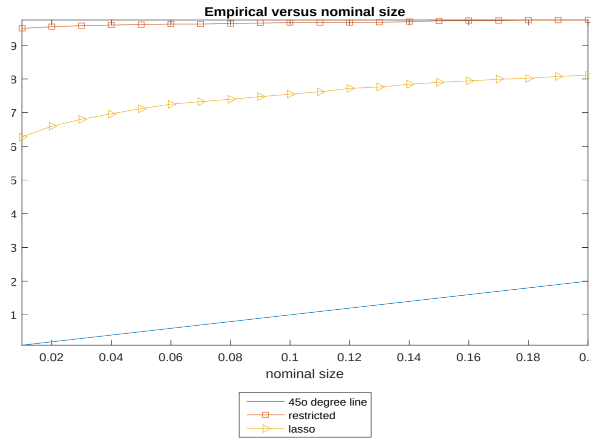


(d) $T = 1000$, without trend

Figure S.3: Size distortion plots of the t-test in the cointegrating 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.

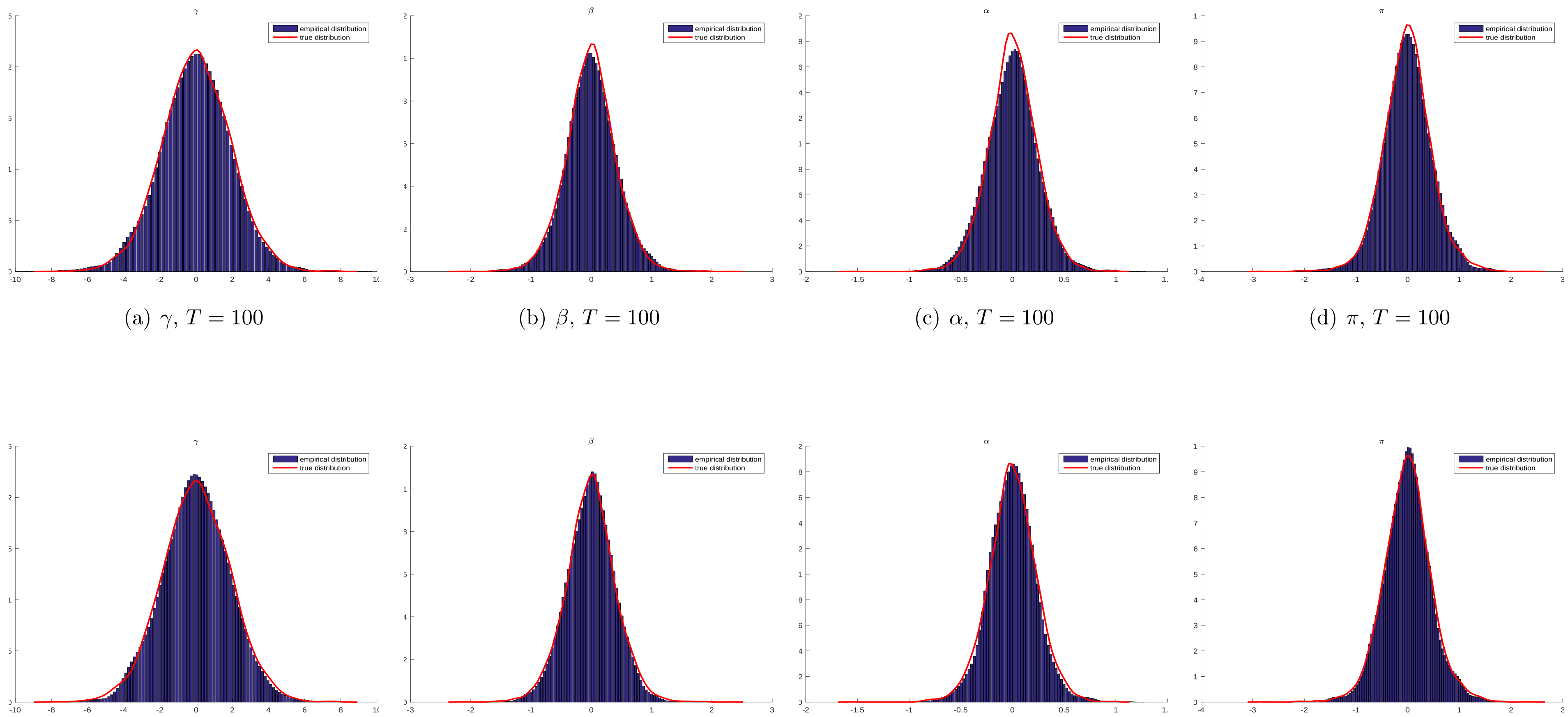


(a) $T = 100$, with trend



(b) $T = 1000$, with trend

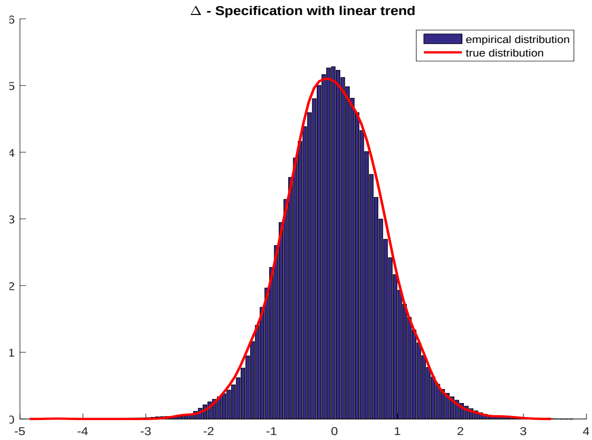
Figure S.4: Empirical rejection rates (size) in the cointegrating case when the coefficients of the linear combination of peers are restricted. Two different sample sizes are considered: $T = 100$ (panel (a)) and $T = 1000$ (panel (b)).



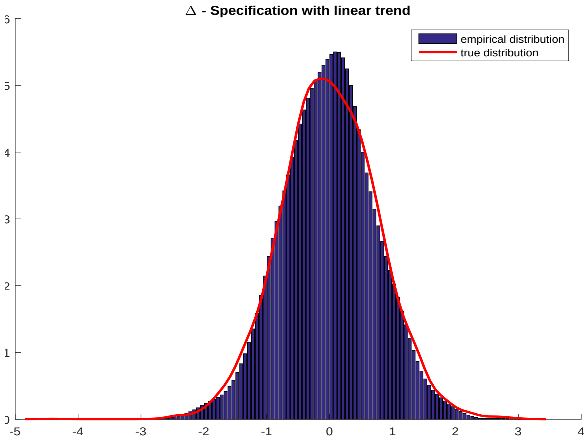
(a) γ, $T = 100$ (b) β, $T = 100$ (c) α, $T = 100$ (d) π, $T = 100$

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

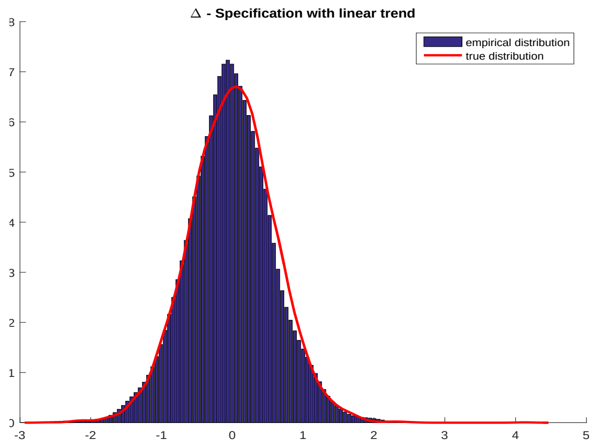
Figure S.5: Empirical (bars) and asymptotic (solid line) distributions of the estimates of regression coefficients in the spurious case for $T = 100$ and $T = 1000$. The distributions are scaled as in the lemma. Panel (a): $\sqrt{T}\widehat{\gamma}$, $T = 100$; Panel (b): first element of $\widehat{\boldsymbol{\beta}}$, $T = 100$; 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 $\widehat{\pi}$, $T = 1000$.



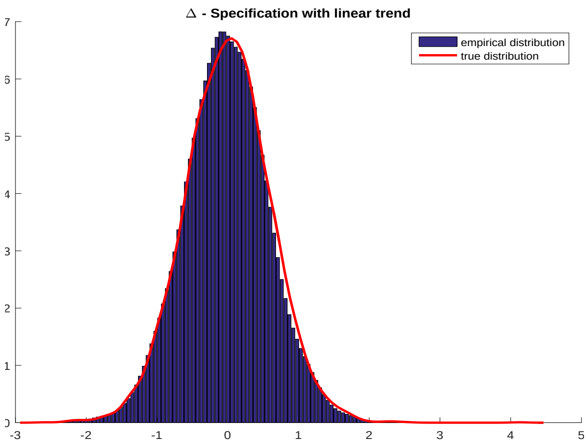
(a) $T = 100$, with trend



(b) $T = 1000$, with trend

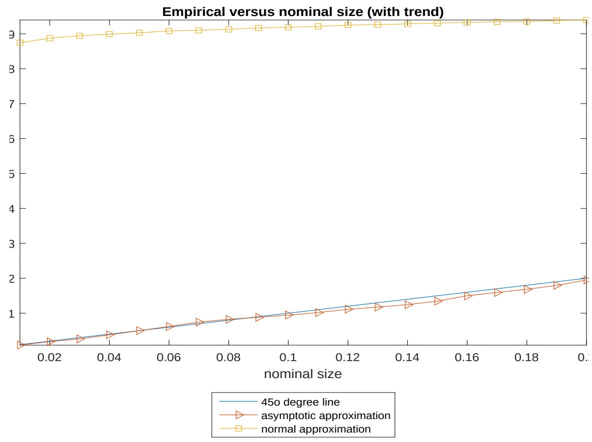


(c) $T = 100$, without trend

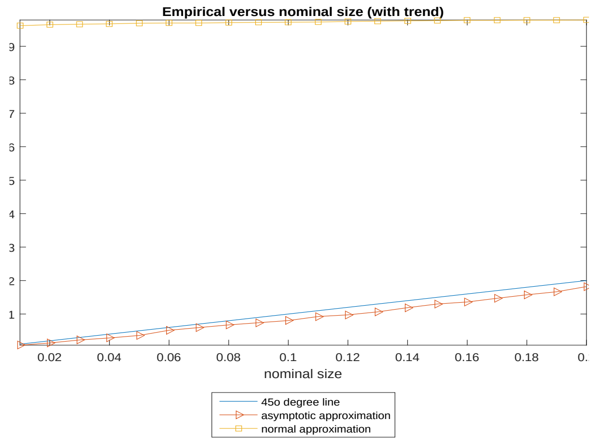


(d) $T = 1000$, without trend

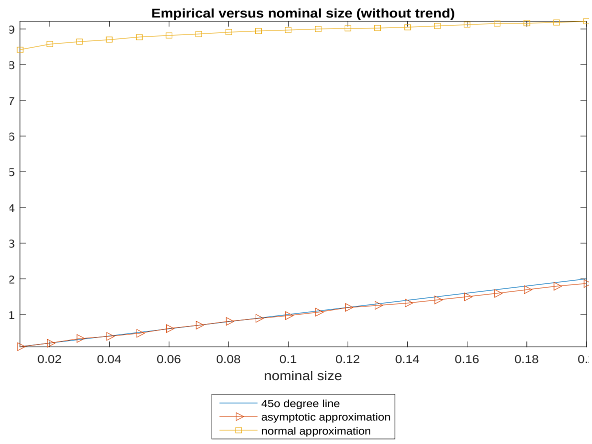
Figure S.6: Empirical (bars) and asymptotic (solid line) distributions of the counterfactual effects in the spurious case for $T = 100$ and $T = 1000$. The distributions are scaled as in the Theorem. Panel (a): trend included in the estimated equation and $T = 100$; Panel (b): trend included in the estimated equation and $T = 1000$; Panel (c): trend excluded from the estimated equation and $T = 100$; Panel (d): trend excluded from the estimated equation and $T = 1000$.



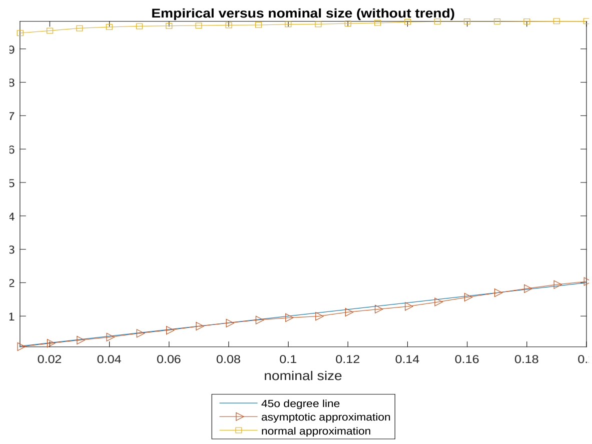
(a) $T = 100$, with trend



(b) $T = 1000$, with trend

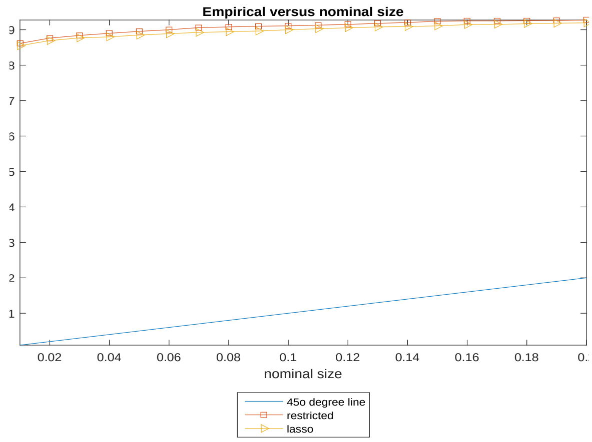


(c) $T = 100$, without trend

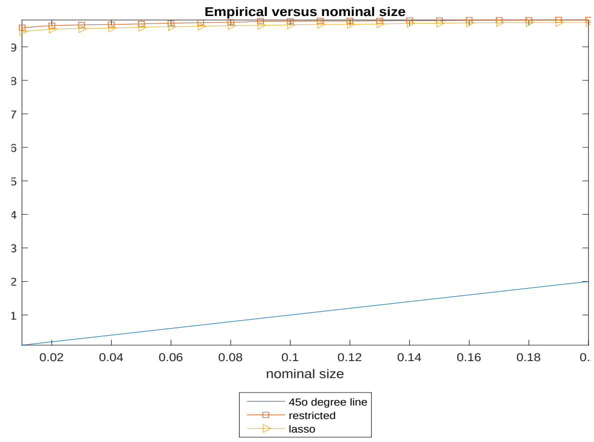


(d) $T = 1000$, without trend

Figure S.7: Size distortion plots of the **scaled** t-test in the cointegrating 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 S.8: 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)).