Supplementary Material for Counterfactual Analysis with Artificial Controls: Inference, High Dimensions and Nonstationarity*

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`

August 2, 2021

Abstract

This supplementary material contains additional simulation results and all the proofs of the results in the main paper.

Keywords: comparative studies, panel data, synthetic control, policy evaluation, intervention, cointegration, spurious regression, resampling, weak sparsity.

***Acknowledgments**: The work of Marcelo C. Medeiros is partly funded by CNPq and CAPES. The authors gratefully acknowledge the invaluable comments and guidance of the guest coeditors, Alberto Abadie and Matias Cattaneo as well as three anonymous referees. The authors are thankful for the comments from Frank Diebold, Jianqing Fan, Marcelo Fernandes, Guido Imbens, Anders B. Kock, Sophocles Mavroeidis, Eduardo F. Mendes, Pedro Souza, Normam Swanson, Michael Wolf, and participants during seminars at Princeton University, Rutgers University, University of Pennsylvania, Warwick University, Oxford University, São Paulo School of Economics, Pontifical Catholic University of Rio de Janeiro, and University of Brasilia. We are also thankful for the comments received during the 2018 Latin American Meeting of the Econometric Society, Guayaquil, Ecuador and the Barcelona GSE summer forum. A special acknowledgment goes to Etiënne Wijler for insightful and technical discussions.

1 Additional Simulation Results

Table S.1 reports the size distortions for the mixed-trend case. The table shows rejection rates under the null hypothesis of no intervention effect under three different nominal size values: 0.01, 0.05 and 0.1. The rejection rates are computed for three estimation frameworks: **LASSO** means that the counterfactual is estimated by LASSO with all the n units included in the model. The penalization parameter λ is chosen as described in Section 4. **Oracle** means that the counterfactual is estimated by ordinary least squares (OLS) using only the s_0 relevant units. Finally, **True** means no estimation, that is, the counterfactual is estimated with the true values of the parameters ($\boldsymbol{\theta}_0$). All distributions are standardized (zero mean and unit variance). Mixed normal means two Normal distributions with probability $(0.3, 0.7)$, mean $(-10, 10)$ and variance $(2, 1)$. The autoregressive of order one, AR(1), structure with coefficient ρ is applied to the common factor innovation U_{1t}^F and the first unit idiosyncratic innovation U_{1t}^Z.

Table S.2 reports several statistics averaged over 10,000 replications for each one of four data generating processes. More specifically, the mean ℓ_1-norm is the average $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_1$, the mean bias is the average bias $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ over the simulations, the mean MSE is the average mean squared error, and the mean Δ is the average intervention effect over the 10 out-of-sample periods. Note that the true value of Δ is zero. MSE Δ is the average squared error over the simulation and, finally, median Δ is the median of the estimates of Δ over the simulations. Each column in the table represents a variation of the baseline scenario, in which we set $T = 100$,$s_0 = 5$, $n = 100$ and $\rho = 0$. Model (1) is given by equations (5.1) and (5.2) where $f_t^F = 0$. Model (2) is given by equations (5.1) and (5.2) where $f_t^F = 1$. Model (3) is given by equations (5.1) and (5.3) where $f_t^F = t$. Model (4) is given by equations (5.1) and (5.3) where $f_t^F = t^2$.

As expected, the ℓ_1-norm, the bias, and the MSE of the estimators decrease with the sample size but increase as the degree of sparsity decreases (s_0 grows), as the number of covariates grows or as the autocorrelation in the errors increases. Nevertheless, the biases are negligible. Concerning the estimator of the average intervention effect (Δ), the estimators are rather precise when the trends are deterministic. On the other hand, with stochastic trends, the biases are small only with no error autocorrelation.

2 Proof of the Main Results

2.1 Proof of Proposition 1

In light of representation (3.2), it is enough to prove result (a) to show that η_{it}/d_{it} vanishes in the appropriate sense as $t \to \infty$. Under DGP (2.4), we have

$$\frac{\eta_{it}}{d_{it}} = \frac{Z_{i0}^{(0)}}{d_{it}} + \frac{\sum_{s=1}^{t} U_{is}}{\sqrt{t}} \frac{\sqrt{t}}{d_{it}} = o_P(1) + O_P(1)o(1) = o_P(1),$$

where the $O_P(1)$ term is a consequence of Assumption 3. Under DGP (2.5), we have that $\eta_{it}/d_{it} = U_{it}/(c_i + f_{it}) \to 0$, almost surely as $f_{it} \to \infty$.

For result (b), we have for DGP (2.4), $Z_{it}^{(0)} = d_{it} + Z_{it}^{(0)} + \sum_{s=1}^{t} U_{it} = O(\sqrt{t}) + O_P(1) + O_P(\sqrt{t}) = O_P(\sqrt{t})$ and for DGP (2.5), $Z_{it}^{(0)} = c_i + f_{it} + U_{it} = O(1) + O(1) + O_P(1) = O_P(1)$.

Finally, under DGP (2.4), if $d_{it} = o(\sqrt{t})$, we have the result by the Central Limit Theorem (ensured by Assumption 3) combined with Slutsky's theorem since $t^{-1/2} Z_{it}^{(0)} = o(1) + t^{-1/2} \sum_{s=1}^{t} U_{it}$.

2.2 Proof of Proposition 2

We start from the reparametrized objective function H defined in (3.4). By definition, $H(\widehat{\boldsymbol{\gamma}}) \le H(\boldsymbol{\gamma})$ for all $\boldsymbol{\gamma}$. Using the fact that $Y_t = \boldsymbol{\gamma}_0' \boldsymbol{W}_t + V_t$ for the transformed variables and letting $\boldsymbol{\Sigma} := \frac{1}{T_0} \sum_{t=1}^{T_0} \boldsymbol{W}_t \boldsymbol{W}_t'$, we have for any $\boldsymbol{\gamma}$:

$$(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \boldsymbol{\Sigma} (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \lambda \|\widehat{\boldsymbol{\gamma}}\|_\nu \le 2(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{W}_t V_t + \lambda \|\boldsymbol{\gamma}\|_\nu, \tag{S.1}$$

where we use the shorthand $\|\boldsymbol{\gamma}\|_\nu := \sum_{i=1}^{p} \nu_i |\gamma_i|$. We can bound from above the first term after the inequality in (S.1) using Hölder's inequality by $\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_1 \|\frac{2}{T} \sum_{t=1}^{T} \boldsymbol{W}_t V_t\|_\infty$, and provided that $\lambda_0 \geq \|\frac{2}{T}\sum_{t=1}^{T} \boldsymbol{W}_t V_t\|_\infty$, we are left with

$$(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \lambda\|\widehat{\boldsymbol{\gamma}}\|_\nu \leq \lambda_0\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_1 + \lambda\|\boldsymbol{\gamma}\|_\nu.$$

Now, let $\mathcal{S} \subseteq \{1\ldots,p\}$ denote an index set such that for any p-dimensional vector $\boldsymbol{v}$, $\boldsymbol{v}_\mathcal{S}$ is the vector containing only the elements of the vector $\boldsymbol{v}$ indexed by $\mathcal{S}$ and $\mathcal{S}^c := \mathcal{S}\setminus\{1,\ldots,p\}$ its complement. For an arbitrary index set $\mathcal{S}$, we use $\|\boldsymbol{\gamma}\|_1 = \|\boldsymbol{\gamma}_\mathcal{S}\|_1 + \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1$ and $\|\boldsymbol{\gamma}\|_\nu = \|\boldsymbol{\gamma}_\mathcal{S}\|_\nu + \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_v$ and the triangle inequality to write

$$\begin{aligned}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \lambda\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}^c}\|_\nu - \lambda_0\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}^c}\|_1 \leq& \\ \lambda_0\|\widehat{\boldsymbol{\gamma}}_\mathcal{S} - \boldsymbol{\gamma}_\mathcal{S}\|_1 + \lambda\|\widehat{\boldsymbol{\gamma}}_\mathcal{S} - \boldsymbol{\gamma}_\mathcal{S}\|_\nu &+ \lambda\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu + \lambda_0\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1.\end{aligned}$$

In addition, consider events defined in (A.3)–(A.5) to conclude that on Ω_2, we have for every $\boldsymbol{\gamma}$ that $\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \geq (1-\lambda_2)\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1$ and $\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \leq (1+\lambda_2)\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1$, which yields

$$\begin{aligned}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + [\lambda(1-\lambda_2) - \lambda_0]\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}^c}\|_1 \leq& \\ [\lambda_0 + \lambda(1+\lambda_2)]\|\widehat{\boldsymbol{\gamma}}_\mathcal{S} - \boldsymbol{\gamma}_\mathcal{S}\|_1 + \lambda\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu &+ \lambda_0\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1.\end{aligned}$$

Set $\underline{\lambda} := \lambda(1-\lambda_2) - \lambda_0$ and sum $\underline{\lambda}\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1$ to both sides of the last inequality and use the triangle inequality to obtain

$$\begin{aligned}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \underline{\lambda}\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}^c} - \boldsymbol{\gamma}_{\mathcal{S}^c}\|_1 \leq& \\ [\lambda_0 + \lambda(1+\lambda_2)]\|\widehat{\boldsymbol{\gamma}}_\mathcal{S} - \boldsymbol{\gamma}_\mathcal{S}\|_1 &+ 2\lambda(\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \vee \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1).\end{aligned}$$

Finally, for $\delta \in [0,1)$, set $\overline{\lambda} := \lambda(1+\lambda_2) + \lambda_0 + \delta\underline{\lambda}$ and sum $\delta\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}_\mathcal{S} - \boldsymbol{\gamma}_{0,\mathcal{S}}\|_1$ to both sides

$$\begin{aligned}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \underline{\lambda}\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}^c} - \boldsymbol{\gamma}_{\mathcal{S}^c}\|_1 + \delta\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}_\mathcal{S} - \boldsymbol{\gamma}_\mathcal{S}\|_1 \leq& \\ \overline{\lambda}\|\widehat{\boldsymbol{\gamma}}_\mathcal{S} - \boldsymbol{\gamma}_\mathcal{S}\|_1 &+ 2\lambda(\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \vee \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1). \qquad \text{(S.2)}\end{aligned}$$

We now consider two cases: (i) if $(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \geq -\delta\underline{\lambda}\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_1 + 2\lambda(\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \vee \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1)$, then the inequality (S.2) implies that $(1-\delta)\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}^c} - \boldsymbol{\gamma}_{\mathcal{S}^c}\|_1 \leq \overline{\lambda}\|\widehat{\boldsymbol{\gamma}}_\mathcal{S} - \boldsymbol{\gamma}_\mathcal{S}\|_1$, which, by the definition of ξ and the compatibility condition on the matrix $\boldsymbol{\Sigma}$, we have

$$\|\widehat{\boldsymbol{\gamma}}_\mathcal{S} - \boldsymbol{\gamma}_\mathcal{S}\|_1 \leq \frac{\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_{\boldsymbol{\Sigma}}\sqrt{|\mathcal{S}|}}{\chi(\Sigma, \mathcal{S}, \xi)}.$$

Using the compatibility condition, the first term on the right-hand side of (S.2) can be upper bounded by

$$\overline{\lambda}\frac{\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_{\boldsymbol{\Sigma}}\sqrt{|\mathcal{S}|}}{\chi(\Sigma,\mathcal{S},\xi)} \leq \frac{\overline{\lambda}^2|\mathcal{S}|}{2\chi^2(\Sigma,\mathcal{S},\xi)} + \frac{1}{2}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_{\boldsymbol{\Sigma}}.$$

Apply the last bound on (S.2)and multiply it by 2 such that

$$\begin{aligned}2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) + 2\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}^c}-\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1 + 2\delta\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}}-\boldsymbol{\gamma}_{\mathcal{S}}\|_1 \leq \\ \frac{\overline{\lambda}^2|\mathcal{S}|}{\chi^2(\Sigma,\mathcal{S},\xi)} + \|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_{\boldsymbol{\Sigma}} + 4\lambda(\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \vee \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1).\end{aligned} \tag{S.3}$$

Notice that for any pair $\boldsymbol{\gamma},\widetilde{\boldsymbol{\gamma}} \in \mathbb{R}^p$, we have the identity

$$2(\widetilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widetilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) = \|\widetilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|_{\boldsymbol{\Sigma}} + \|\widetilde{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_{\boldsymbol{\Sigma}} - \|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|_{\boldsymbol{\Sigma}}. \tag{S.4}$$

Apply (S.4) with $\widetilde{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\gamma}}$ to the first term on the left-hand size of (S.3) such that

$$\begin{aligned}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|_{\boldsymbol{\Sigma}} + 2\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}^c}-\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1 + 2\delta\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}_{\mathcal{S}}-\boldsymbol{\gamma}_{\mathcal{S}}\|_1 \leq \\ \|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|_{\boldsymbol{\Sigma}} + \frac{\overline{\lambda}^2|\mathcal{S}|}{\chi^2(\Sigma,\mathcal{S},\xi)} + 4\lambda(\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \vee \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1).\end{aligned}$$

The result is then obtained by noticing that the sum of the second and third term on the left-hand side of the inequality can be lower bounded by $2\delta\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_1$ because $\delta \in [0,1)$.

Now, if (ii) $(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) \leq -\delta\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_1 + 2\lambda(\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \vee \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1)$, then the identity (S.4) give us directly the result since

$$\begin{aligned}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0\|_{\boldsymbol{\Sigma}} + 2\delta\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_1 &= 2\delta\underline{\lambda}\|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_1 + 2(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma})'\boldsymbol{\Sigma}(\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}_0) - \|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_{\boldsymbol{\Sigma}} + \|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|_{\boldsymbol{\Sigma}} \\ &\leq \|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|_{\boldsymbol{\Sigma}} - \|\widehat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\|_{\boldsymbol{\Sigma}} + 4\lambda(\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \vee \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1) \\ &\leq \|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|_{\boldsymbol{\Sigma}} + 4\lambda(\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \vee \|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1).\end{aligned}$$

2.3 Proof of Proposition 3

The proof follows from Proposition 2. We consider only the case when $b > 0$ since the case for $b = 0$ was done in the main text. First, we use the fact that $\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_\nu \leq (1+\lambda_2)\|\boldsymbol{\gamma}_{\mathcal{S}^c}\|_1$ on the event Ω_2. Additionally, we have that $R_b := \sum_{j=1}^p |\gamma_{0,j}|^b \geq |\mathcal{S}_b|(\frac{\overline{\lambda}^2}{\lambda})^b$ from which we conclude that

$$|\mathcal{S}_b| \leq \left(\frac{\lambda}{\overline{\lambda}^2}\right)^b R_b(\boldsymbol{\gamma}_0) \quad \text{and} \quad \|\boldsymbol{\gamma}^0_{\mathcal{S}_b^c}\| \leq \left(\frac{\overline{\lambda}^2}{\lambda}\right)^{1-b} R_b.$$

Set $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, $\mathcal{S} = \mathcal{S}_q$ in (A.6) and use the previous inequalities to upper bound the right-hand side of (A.6) to obtain (A.7).

For the second result, use the condition $\lambda = k\lambda_0$ to conclude that

$$\underline{\lambda} = (1 - \lambda_2 - 1/k)\lambda := \underline{c}\lambda \quad \text{and} \quad \overline{\lambda} = [1 + \delta + (1 - \delta)(\lambda_2 + 1/k)]\lambda := \overline{c}\lambda.$$

Therefore, $C_2 = \mathbb{1}\{q > 0\}\overline{c}^2$ and $C_1 := \frac{\overline{c}^{2(q-1)}}{2\delta\underline{c}}$.

2.4 Proof of Lemma 1

We divide the proof into three steps. First, we show that under the hypotheses of Theorem 1, the process $\{\boldsymbol{W}_t V_t\}_{t\geq 1}$ can be properly bounded. Then, we show that the event $\Omega_0 \cap \Omega_1$ occurs with high probability. Finally, we derive the results of the Theorem.

2.4.1 Bound Control

We have $\boldsymbol{W}_t = \boldsymbol{L}^{-1}\boldsymbol{X}_t = \boldsymbol{L}^{-1}(\boldsymbol{d}_t + \boldsymbol{\eta}_t)$ where $\boldsymbol{d}_t := (d_{1t}, \ldots, d_{pt})'$ and $\boldsymbol{\eta}_t := (\eta_{1t}, \ldots, \eta_{pt})'$ for $t \geq 1$. Then, for the DGP (2.5) in Assumption 2, recall that $\boldsymbol{\eta}_t = \boldsymbol{U}_t$, $\boldsymbol{d}_t = c + \boldsymbol{\mu} f_t$ and $\boldsymbol{L}$ is just a deterministic diagonal matrix. Hence, the process $\{\boldsymbol{W}_t\}$ is strong mixing with the same coefficient as the process $\{\boldsymbol{U}_t\}$. Moreover the process $\{V_t\}$, as a linear combination of $\boldsymbol{U}_t$, is also strong mixing with the same mixing coefficient as the process $\{\boldsymbol{U}_t\}$. Therefore, the process $\{\boldsymbol{W}_t V_t\}$ is also strong mixing with the same mixing coefficient as the process $\{\boldsymbol{U}_t\}$ under Assumption 3. Additionally, by definition of the scaling matrix $\boldsymbol{L}$, all the components of the vector $\boldsymbol{L}^{-1}\boldsymbol{d}_t$ are bounded between 0 and 1. If the process $\{\boldsymbol{U}_t\}$ fulfills condition (a) of Assumption 3, $\{V_t\}$ also does because $V_t = U_{1t} - \sum_{i=2}^n \theta_{0,i} U_{it}$ and

$$\|V_t\|_{\mathcal{L}^q} \leq |\|U_{1t}\|_{\mathcal{L}^q} + \sum_{i=2}^n |\theta_{0,i}|\|U_{it}\|_{\mathcal{L}^q} = O(\|\boldsymbol{\theta}_0\|_1) = O(1).$$

Then, by the Cauchy-Schwartz inequality, we have that $\{\boldsymbol{W}_t V_t\}$ fulfills the same condition with constant $q/2$ since for some $\epsilon > 0$, we have

$$\sup_{t\in\mathbb{N}}\sup_{i\le p}\mathbb{E}|U_{it}V_t|^{q/2+\epsilon/2} \le \left(\sup_{t\in\mathbb{N}}\sup_{i\le p}\mathbb{E}|U_{it}|^{q+\epsilon}\sup_{t\in\mathbb{N}}\sup_{i\le p}\mathbb{E}|V_t|^{q+\epsilon}\right)^{1/2} < \infty.$$

Furthermore, if $\{(V_t, \boldsymbol{U}_t')'\}$ also fulfills condition (b) of Assumption 3 with the triple (a_1, a_2, a_3) in the exponential bound, then the process $\{\boldsymbol{W}_t V_t\}$ complies with Assumption 3(b) with the triple $(2a_1, a_2, a_3/2)$ since for each component of the vector, $\boldsymbol{U}_t V_t$ is bounded by

$$\mathbb{P}(|U_{it}V_t| > u) \le \mathbb{P}(|U_{it}| > \sqrt{u}) + \mathbb{P}(|V_t| > \sqrt{u}) \le 2a_1 \exp(-a_2 u^{a_3/2}).$$

Now, consider DGP (2.4). Notice that we cannot follow the same proof strategy taken for the DGP (2.5) since in this case, $\{\boldsymbol{W}_t\}$ *cannot* be a mixing process. Therefore, we use Lemma 1 to construct bounds for $\|\sum_{t=1}^{T_0} W_{it}V_t\|_{\mathcal{L}^q}$ and $\|\sum_{t=1}^{T_0} W_{it}W_{jt}\|_{\mathcal{L}^q}$ uniformly in $t \le T_0$ and $1 \le i, j \le p$. For the latter, we have

$$\left\|\sum_{t=1}^{T_0} W_{it}W_{jt}\right\|_{\mathcal{L}^q} \le \sum_{t=1}^{T_0}\frac{d_{it}d_{jt}}{\ell_i\ell_j} + \frac{1}{\ell_j}\left\|\sum_{t=1}^{T_0}\frac{d_{it}}{\ell_i}\eta_{jt}\right\|_{\mathcal{L}^q} + \frac{1}{\ell_i}\left\|\sum_{t=1}^{T_0}\frac{d_{jt}}{\ell_j}\eta_{it}\right\|_{\mathcal{L}^q} + \frac{1}{\ell_i\ell_j}\left\|\sum_{t=1}^{T_0}\eta_{it}\eta_{jt}\right\|_{\mathcal{L}^q}.$$

Since $d_{it}/\ell_i \in [0, 1]$ for all i by definition, the first term is $O(T_0)$. The second and third terms are $O(T_0^{3/2}/l_j)$ and $O(T_0^{3/2}/l_i)$, respectively, by result (b) of Lemma 1 and the last one if $O(T_0^2/(\ell_i\ell_j))$ from result (c) of Lemma 1. Consequently, we conclude that

$$\left\|\sum_{t=1}^{T_0} W_{it}W_{jt}\right\|_{\mathcal{L}^q} = O\left(T_0 \vee \frac{T_0^{3/2}}{\ell_i \wedge \ell_j} \vee \frac{T_0^2}{\ell_i\ell_j}\right) = O(T_0).$$

For the former, we start by the triangle inequality

$$\left\|\sum_{t=1}^{T_0} W_{it}V_t\right\|_{\mathcal{L}^q} \le \left\|\sum_{t=1}^{T_0}\frac{d_{it}}{\ell_i}V_t\right\|_{\mathcal{L}^q} + \frac{1}{\ell_i}\left\|\sum_{t=1}^{T_0}\eta_{it}V_t\right\|_{\mathcal{L}^q}.$$

The first term is $O(\sqrt{T_0})$ by result (a) of Lemma 1. For the second term, we may use result (c) and Hölder's inequality to obtain

$$\left\|\sum_{t=1}^{T_0}\eta_{it}V_t\right\|_{\mathcal{L}^q} \le \left\|\sum_{t=1}^{T_0}\eta_{it}U_{1t}\right\|_{\mathcal{L}^q} + \sum_{j=2}^{n}|\theta_{0,j}|\left\|\sum_{t=1}^{T_0}\eta_{it}U_{jt}\right\|_{\mathcal{L}^q} = O(T_0 \vee T_0\|\boldsymbol{\theta}_0\|_1) = O(T_0).$$

Hence, the second term is $O(T_0/\ell_i)$ by result (a), and therefore

$$\left\|\sum_{t=1}^{T_0} W_{it}V_t\right\|_{\mathcal{L}^q} = O(\sqrt{T_0} \vee T_0/\ell_i) = O(\sqrt{T_0}).$$

2.4.2 Probability Bounds on Ω_0 and Ω_1

In light of the results in the previous subsection, we can set $\lambda_0 = \lambda/2$ with λ as stated in the theorem. For DGP (2.5), results (b) and (c) of Lemma 2 allow us to conclude that for all $c > 0$:

$$\mathbb{P}(\Omega_0^c) = \mathbb{P}\left(\left\|\frac{1}{T_0}\sum_{t=1}^{T_0}\boldsymbol{W}_t V_t\right\|_\infty > \frac{\lambda_0}{2}\right) = \begin{cases} O(c^{-q/2}) & \text{under Assumption 3(a)} \\ O[\exp(-c/2)] & \text{under Assumption 3(b).}\end{cases}$$

We start by showing that $\mathbb{P}(\Omega_1) \to 1$. Recall that $\mathbb{P}(\Omega_1^c) = \mathbb{P}\left(\|\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_0\|_\infty > \lambda_1\right)$. Set $\lambda_1 = \chi_1(\xi, \mathcal{S}, \boldsymbol{\Sigma}_0)/[2(1+\xi)^2 s]$ and $x = \lambda_1\sqrt{T_0}$ in Lemma 2. Results (d) and (e) in Lemma 2 imply that

$$\mathbb{P}(\Omega_1^c) = \begin{cases} O\left[\left(\frac{p^{2/q}s}{\sqrt{T_0}}\right)^q\right] = o(1) & \text{under Assumption 3(a),} \\ O\left\{\exp\left[2\log p - \frac{\chi_1\sqrt{T_0}}{4(1+\xi)^2 s}\right]\right\} = o(1) & \text{under Assumption 3(b),}\end{cases}$$

where the $o(1)$ terms follow by assumption of the theorem since $p^{4/q}s/\sqrt{T_0} = o(1)$ and $s\log p/\sqrt{T_0} = o(1)$.

Additionally, from the relation $\lambda = 2\lambda_0$, we may choose $\lambda_2 > 0$ arbitrarily close to 0 such that ξ in Proposition 2 can be arbitrarily close to 3. For instance, setting $\lambda_2 = 1/10$ yields

$$\frac{\lambda_0 + \lambda(1+\lambda_2)}{\lambda(1-\lambda_2) - \lambda_0} = \frac{1 + 2(1+\lambda_2)}{2(1-\lambda_2) - 1} = \frac{3 + 2\lambda_2}{1 - 2\lambda_2} = 4 =: \xi.$$

Provided that the GIF condition holds, i.e., $\chi_1(4, \mathcal{S}, \boldsymbol{\Sigma}_0) > 0$, we have for λ as stated in the theorem and for all $c > 0$:

$$\mathbb{P}(\Omega_0 \cap \Omega_1) \geq 1 - \begin{cases} O(c^{-q/2}) & \text{under Assumption 3(a),} \\ O[\exp(-c/2)] & \text{under Assumption 3(b).}\end{cases}$$

Similarly, for the DGP (2.4) under Assumption 3(a), by setting λ as stated in the theorem yields

$$\mathbb{P}(\Omega_0^c) = \mathbb{P}\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{W}_t V_t\right\|_\infty > \frac{\lambda_0}{2}\right) = O(c^{-q/2}) \quad \text{and}$$

$$\mathbb{P}(\Omega_1^{*c}) \le \varepsilon.$$

2.5 Proof of Theorem 1

By setting λ according to Assumption 5, we have that $\lambda = O[\psi(p)/\sqrt{T_0}]$ with $\psi(x) = x^{2/q}$ under Assumption 3(a) and $\psi(x) = \log x$ under Assumption 3(b). The second part of Proposition 3 combined with Lemmas 1 and 2 yields

$$\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|_1 = O_P\left[\left(\frac{\psi(p)}{\sqrt{T_0}}\right)^{1-b}\frac{R_b}{\lambda_1}\right].$$

The result (a) then follows from Assumption 6(c).

For the remaining results, we use the fact that

$$\widehat{\delta}_t - \delta_t = V_t + (\widehat{\boldsymbol{\gamma}}_{T_0} - \boldsymbol{\gamma}_0)'\boldsymbol{W}_t, \quad T_0 < t \le T.$$

For (b), we have by Hölder's inequality that $|\widehat{\delta}_t - \delta_t - V_t| = |(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)'\boldsymbol{W}_t| \le \|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|_1 \|\boldsymbol{W}_t\|_\infty$. The order in probability of the first term is given by the result (a), and the second term is $O_P[\psi(p)]$ by Lemma 2(a). Hence, $\widehat{\delta}_t - \delta_t - V_t = O_P[\frac{\psi(p)^{2-b}R_b}{T_0^{(1-b)/2}\lambda_1}] = o_P(1)$ also by Assumption 6(c). For result (c), we have

$$\widehat{\Delta}_T - \Delta_T := \frac{1}{T_1}\sum_{t>T_0}\widehat{\delta}_t - \delta_t = \frac{1}{T_1}\sum_{t>T_0}V_t - (\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)'\frac{1}{T_1}\sum_{t>T_0}\boldsymbol{W}_t.$$

The first term is $O_P(1/\sqrt{T_1})$ under Assumption 3, and the absolute value of the second term is upper bounded by Hölder's inequality since

$$\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|_1 \left\|\frac{1}{T_1}\sum_{t>T_0}\boldsymbol{W}_t\right\|_\infty \le \|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\|_1 \left(\left\|\frac{1}{T_1}\sum_{t>T_0}\boldsymbol{W}_t - \mathbb{E}(\boldsymbol{W}_t)\right\|_\infty + \left\|\frac{1}{T_1}\sum_{t>T_0}\mathbb{E}(\boldsymbol{W}_t)\right\|_\infty\right).$$

The first term in parentheses is $O_P[\psi(p)/\sqrt{T_1}]$ by Lemma 2(b), whereas the second is $O(1)$. Therefore, under the assumptions of the theorem, the term in parentheses is $O_P(1)$. The order in probability of the term outside the parentheses is given by result (a). Hence,

$(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)' \frac{1}{T_1} \sum_{t>T_0} \boldsymbol{W}_t = O_P\left[\left(\frac{\psi(p)}{\sqrt{T_0}}\right)^{1-b} \frac{R_b}{\lambda_1}\right]$ and, therefore

$$\widehat{\Delta}_T - \Delta_T = O_P\left[\left(\frac{\psi(p)}{\sqrt{T_0}}\right)^{1-b} \frac{R_b}{\lambda_1} \vee \frac{1}{\sqrt{T_1}}\right].$$

2.6 Proof of Lemma 2

According to the proposition, let $\mathcal{R}$ be the index set of the stochastic (nondeterministic) w_i. From the definition of Ω_2, we conclude that

$$\Omega_2 = \left\{\sup_{i\in S} \nu_S \leq 1 + \lambda_2\right\} \cap \left\{\inf_{i\in S^c} \nu_{S^c} \geq 1 - \lambda_2\right\} \supseteq \left\{\sup_{i\in \mathcal{H}} |\nu_i - 1| \leq \lambda_2\right\}.$$

To see that it is indeed the case, recall that the intercept is always included in the model (belongs to S). Hence, $\nu_1 = 0 \leq 1 + \lambda_2$ for any $\lambda_2 \in (0, 1)$. For $i > 1$, ν_i is either 1, in that case trivially $1 - \lambda_2 \leq \nu_i \leq 1 + \lambda_2$, or $\nu_i = 1 + \eta_{iT_0}/d_{iT_0}$.

We now show that $\sup_{i\in\mathcal{R}} |\eta_{it}/d_{it}| = o_P(1)$ as $t \to \infty$. For DGP (2.4), we have $\eta_{iT_0}/d_{iT_0} = \left(\frac{1}{\sqrt{T_0}} \sum_{t=1}^{T_0} U_{it}\right) \frac{\sqrt{T_0}}{d_{iT_0}}$ for $i \in \mathcal{H}$ in Assumption 2. Thus,

$$\sup_{i\in\mathcal{H}} |\eta_{iT_0}/d_{iT_0}| \leq \sup_{i\in\mathcal{H}} \left|\frac{1}{\sqrt{T_0}} \sum_{t=1}^{T_0} U_{it}\right| \frac{\sqrt{T_0}}{\inf_{i\in\mathcal{H}} |d_{iT_0}|}.$$

Let $d_{\mathcal{R}}(T_0) := \inf_{i\in\mathcal{R}} |d_{iT_0}|$. Since $\{U_t\}$ is a zero-mean strong mixing process by assumption, we can apply Lemma 2(b) to conclude that

$$\sup_{i\in\mathcal{R}} |\nu_i - 1| = \begin{cases} O_P\left[\frac{(|\mathcal{R}|)^{1/q}\sqrt{T_0}}{d_{\mathcal{R}}(T_0)}\right] = o_P(1) & \text{under Assumption 3(a),} \\ O_P\left[\frac{\sqrt{T_0}\log(|\mathcal{R}|)}{d_{\mathcal{R}}(T_0)}\right] = o_P(1) & \text{under Assumption 3(b).} \end{cases}$$

For DGP (2.5) in Assumption 2, we have that $\eta_{iT_0}/d_{iT_0} = U_{iT_0}/d_{iT_0}$. Then, $\sup_{i\in\mathcal{H}} |U_{iT_0}/d_{iT_0}| \leq \sup_{i\in\mathcal{H}} |U_{iT_0}| / \inf_{i\in\mathcal{H}} |d_{iT_0}|$. Applying Lemma 2(a), we have that

$$\sup_{i\in\mathcal{H}} |\nu_i - 1| = \begin{cases} O_P\left[\frac{(|\mathcal{R}|)^{1/q}}{d_{\mathcal{R}}(T_0)}\right] = o_P(1) & \text{under Assumption 3(a),} \\ O_P\left[\frac{\log(\#\mathcal{R})}{d_{\mathcal{R}}(T_0)}\right] = o_P(1) & \text{under Assumption 3(b),} \end{cases}$$

where all the $o_P(1)$ terms follow from Assumption 4.

2.7 Proof of Theorem 2

Part (a) follows directly from Theorem 1 (b) combined with the continuous mapping theorem. We prove (b) by showing that both $\widehat{\mathsf{Q}}_T(x) - \mathsf{Q}_0(x) = o_P(1)$ and $\mathsf{Q}_T(x) - \mathsf{Q}_0(x) = o(1)$, as $T_0 \to \infty$ for all $x \in \mathcal{C}_0$, the continuity points of $\mathcal{Q}_0(x) := \mathbb{P}(\phi_0 \leq x)$. The result then follows by the triangle inequality. For the latter, as a consequence of result (a), we have $\widehat{\boldsymbol{\phi}} \Rightarrow \boldsymbol{\phi}_0$. For the former, let $\widetilde{\mathsf{Q}}_T(x) := \frac{1}{\tau}\sum_{j=1}^{\tau} \mathbb{1}(\phi_j \leq x\}$ be the unfeasible counterpart of $\widehat{\mathcal{Q}}(x)$, where $\tau := T_0 - T_1 + 1$. We first show that $\widetilde{\mathcal{Q}}_T(x) - \mathsf{Q}_0(x)$ vanishes in probability as $T_0 \to \infty$. Due to the strict stationarity assumption, $\mathbb{E}[\widetilde{\mathsf{Q}}_T(x)] = \frac{1}{\tau}\sum_{j=1}^{\tau} \mathbb{P}(\psi_j \leq x) = \mathbb{P}(\psi_0 \leq x) =: \mathsf{Q}_0(x)$. Hence, $\widetilde{\mathsf{Q}}_T(x)$ is unbiased for $\mathsf{Q}_0(x)$. Therefore, it is enough to show that $\mathbb{E}\left[\widetilde{\mathsf{Q}}_T^2(x)\right]$ converges to zero. Notice that the sequence $\{A_j := \mathbb{1}(\phi_j \leq x)\}_j$ is stationary. For this reason,

$$\mathbb{E}\left[\widetilde{\mathsf{Q}}_T^2(x)\right] = \frac{1}{\tau}\sum_{|k|<\tau}\left(1 - \frac{|k|}{\tau}\right)\boldsymbol{\gamma}_k, \quad \boldsymbol{\gamma}_k := \mathbb{E}(A_1 A_{1+k}).$$

In addition, $0 \leq A_j \leq 1$, so we can bound the first $T_1 - 1$ covariances by 1 and the remaining covariances using a mixing inequality due to Ibragimov (1962); regarding $|k| \geq T_1$, we have $\boldsymbol{\gamma}_k \leq 4\alpha(k - T_1 + 1)$, where $\alpha(m)$ is the mixing coefficient of the process $\{V_t\}_t$. In fact, the sequence $\{A_j(\nu_j, \ldots, \nu_{j+T_1-1}\}_j$ is also strong mixing. Then,

$$\mathbb{E}\left[\widetilde{\mathsf{Q}}_T^2(x)\right] \leq \frac{2T_1 + 1}{\tau} + \frac{8}{\tau}\sum_{k=T_1}^{\tau} \alpha(k - T_1 + 1).$$

Finally, since $T_0 \to \infty$ implies $\tau \to \infty$, we have that the first term converges to zero, and the second term converges to zero due to Assumption 3, which establishes that $\widetilde{\mathsf{Q}}_T(x) - \mathsf{Q}_0(x) = o_P(1)$ for all x.

Now, we write $\widehat{\mathsf{Q}}(x) = \frac{1}{\tau}\sum_{j=1}^{\tau} I[\phi_j + (\widehat{\phi}_j - \phi_j) \leq x]$ and, for any $\epsilon > 0$, we define the event $\mathscr{A}_T(\epsilon) := \{\mathsf{sup}_j \|\widehat{\phi}_j - \phi_j\|_\infty \leq \epsilon\}$. On $\mathscr{A}_T$, we have that

$$\widetilde{\mathsf{Q}}(x - \epsilon\iota) \leq \widehat{\mathsf{Q}}(x) \leq \widetilde{\mathsf{Q}}(x + \epsilon\iota),$$

where $\iota \in \mathbb{R}^b$ is a vector of 1s. If we add a further condition that $\mathscr{B}_T(\epsilon, x) := \{|\widetilde{\mathsf{Q}}(x - \epsilon\iota) -$

$\mathsf{Q}_0(x - \epsilon\iota)| \vee |\widetilde{\mathsf{Q}}(x + \epsilon\iota) - \mathsf{Q}_0(x + \epsilon\iota)| \leq \epsilon\}$, we have

$$\mathsf{Q}_0(x - \epsilon\iota) - \epsilon \leq \widehat{\mathsf{Q}}(x) \leq \mathsf{Q}_0(x + \epsilon\iota) + \epsilon.$$

Now, take $\epsilon \to 0$ to conclude that, conditional on $\mathscr{A}_T \cap \mathscr{B}_T$, we have $|\widehat{\mathsf{Q}}(x) - \mathsf{Q}_0(x)| \leq \epsilon$ for all $x \in \mathcal{C}_0$.

Therefore, it is enough to show that $\mathbb{P}(\mathscr{A}_T \cap \mathscr{B}_T) = 1$ establishes the result (b). $\mathscr{B}_T$ is a sure event as $\widetilde{\mathcal{Q}}(x) \to \mathcal{Q}_0(x)$ for all $x \in C_0$. Regarding $\mathscr{A}_T$, notice that for $1 \leq t \leq T_0$, we have $\widehat{V}_t - V_t = (\widehat{\boldsymbol{\gamma}}_{T_0} - \boldsymbol{\gamma}_0)'\boldsymbol{W}_t$. As a consequence, by Hölder's inequality,

$$\sup_{t \leq T_0} |\widehat{V}_t - V_t| \leq \|\widehat{\boldsymbol{\gamma}}_{T_0} - \boldsymbol{\gamma}_0\|_1 \sup_{t \leq T_0} \|\boldsymbol{W}_t\|_\infty = \|\widehat{\boldsymbol{\gamma}}_{T_0} - \boldsymbol{\gamma}_0\|_1 \sup_{t,i} |W_{it}|.$$

The first term is $O_P[s_0\psi(p)/\sqrt{T_0}]$ by Theorem 1(a), and the second term is $O_P[\psi(pT_0)]$ by Lemma 2(a). Then, under the assumptions of the theorem, we conclude that $\sup_{t \leq T_0} |\widehat{V}_t - V_t| = O_P[s_0\psi(p)\psi(pT_0)/\sqrt{T_0}] = o_P(1)$. Since $\phi(\cdot)$ is continuous, the last result implies $\sup_j \|\widehat{\phi}_j - \phi_j\|_\infty = o_P(1)$.

For (c) and (d), we use the fact that (b) is equivalent (refer to Theorem 6.3.1 of Resnick (1999)) to say that for any subsequence $\{T_j\}$, we can extract a further subsequence $\{T_{j_k}\}$ such that $\widehat{\mathsf{Q}}_{T_{j_k}}(\omega, x) \to \mathsf{Q}_0(x)$ for all $\omega \in \Omega_3$ and $x \in \mathcal{C}_0$ with $\mathbb{P}(\Omega_3) = 1$. For (c), since $\mathsf{Q}_0(x)$ is assumed continuous and for each fixed ω, $\widehat{\mathsf{Q}}_{T_{j_k}}(\omega, x)$ is a cumulative distribution function (cdf), the last convergence can be made uniform by Polya's theorem, i.e., $\sup_{x \in \mathbb{R}^b} |\widehat{\mathsf{Q}}_{T_{j_k}}(\omega, x) - \mathsf{Q}_0(x)| \to 0$ for all $\omega \in \Omega_3$, where $\mathbb{P}(\Omega_3) = 1$. The result then follows by using the equivalence (in the other direction) of Theorem 6.3.1 of Resnick (1999).

For (d), we know that for each $\omega \in \Omega_3$ and $x \in \mathcal{C}_0$, $\widehat{\mathsf{Q}}_{T_{j_k}}(\omega, x) \to \mathsf{Q}_0(x)$ is equivalent to $\widehat{\mathsf{Q}}^{-1}_{T_{j_k}}(\omega, x) \to \mathsf{Q}_0^{-1}(x)$. We refer to Lemma 21.2 of van der Vaart (2000), which implies once again by Theorem 6.3.1 of Resnick (1999) that $\widehat{\mathsf{Q}}_T^{-1}(x) \xrightarrow{p} \mathsf{Q}_0^{-1}(x)$. By the same reasoning $\mathsf{Q}_T^{-1}(x) \to \mathsf{Q}_0^{-1}(x)$ is equivalent to $\mathsf{Q}_T(x) \to \mathsf{Q}_0(x)$ for all $x \in \mathcal{C}_0$. By the triangle inequality, we have $\widehat{\mathsf{Q}}_T^{-1}(x) - \mathsf{Q}_T^{-1}(x) = o_P(1)$ for $x \in \mathcal{C}_0$; then, we write

$$\mathcal{Q}_T\left[\widehat{\mathsf{Q}}_T^{-1}(\tau)\right] = \mathcal{Q}_T\left[\mathsf{Q}_0^{-1}(\tau) + \widehat{\mathcal{Q}}_T^{-1}(\tau) - \mathcal{Q}_0^{-1}(\tau)\right].$$

Then, conditional on the event $\mathscr{D}(\epsilon) := \left\{ \left| \widehat{\mathsf{Q}}_T^{-1}(x) - \mathsf{Q}_0^{-1}(x) \right| \le \epsilon \right\}$, defined for an arbitrary $\epsilon > 0$, and by the monotonicity of $\mathsf{Q}_T(\cdot)$, we have

$$\mathcal{Q}_T \left[\mathcal{Q}_0^{-1}(\tau) - \epsilon \right] \le \mathcal{Q}_T \left[\widehat{\mathsf{Q}}_T(\tau) \right] \le \mathcal{Q}_T \left[\mathcal{Q}_0^{-1}(\tau) + \epsilon \right].$$

Additionally, consider the event

$$\mathscr{E}(\epsilon) := \left\{ \left| \mathcal{Q}_T[\mathcal{Q}_0^{-1}(\tau) - \epsilon] - \mathcal{Q}_0[\mathcal{Q}_0^{-1}(\tau) - \epsilon] \right| \vee \left| \mathcal{Q}_T[\mathcal{Q}_0^{-1}(\tau) + \epsilon] - \mathcal{Q}_0[\mathcal{Q}_0^{-1}(\tau) + \epsilon] \right| \le \epsilon \right\}$$

to write that, conditioned on $\mathscr{D}(\epsilon) \cap \mathscr{E}(\epsilon)$, we have

$$\mathcal{Q}_0 \left[\mathcal{Q}_0^{-1}(\tau) - \epsilon \right] - \epsilon \le \mathcal{Q}_T \left[\widehat{\mathsf{Q}}_T(\tau) \right] \le \mathcal{Q}_T \left[\mathcal{Q}_0^{-1}(\tau) + \epsilon \right] + \epsilon.$$

Taking the limit as $\epsilon \to 0$ to conclude that, for fixed $\tau \in (0,1)$, if $\mathcal{Q}_0^{-1}(\tau) \in \mathcal{C}_0$ and on $\mathscr{D}(\epsilon) \cap \mathscr{E}(\epsilon)$, we have that $\left| \mathcal{Q}_T \left[\widehat{\mathsf{Q}}_T(\tau) \right] - \tau \right| \le \epsilon$, as $\mathcal{Q}_0 \left[\mathcal{Q}_0^{-1}(\tau) \right] = \tau$ for $x \in \mathcal{C}_0$. Finally, the conditioning event happens with probability approaching 1.

3 Auxiliary Lemmas

Due to the lack of different characters, the variable denominations in this appendix are not necessarily consistent with the remainder of the article.

Lemma 1. *Let $\{X_t, t \in \mathbb{N}\}$ be a real-valued zero-mean strong mixing process with mixing coefficient given by $\alpha(m) = \exp(-2cm)$ for some $c > 0$, such that for some $q > 2$, $\sup_{t \in \mathbb{N}} \mathbb{E}|X_t|^{q+\varepsilon} < C_q < \infty$ for some $\varepsilon > 0$. Additionally, define the partial sum $S_t := \sum_{s=1}^{t} X_t$, then*

(a) $\|S_T\|_{\mathcal{L}^q} = O(\sqrt{T})$

(b) $\|\sum_{t=1}^{T} S_t\|_{\mathcal{L}^q} = O(T^{3/2})$

(c) $\|\sum_{t=1}^{T} S_t X_t\|_{\mathcal{L}^{q/2}} = O(T)$ *if* $q > 4$

(d) $\|\sum_{t=1}^{T} S_t^2\|_{\mathcal{L}^q} = O(T^2)$

Proof. Result (a) can be found in Rio (1994); (b) follows from (a) and the triangle inequality since

$$\left\|\sum_{t=1}^{T} S_t\right\|_{\mathcal{L}^q} \leq \sum_{t=1}^{T} \|S_t\|_{\mathcal{L}^q} = \sum_{t=1}^{T}(O(\sqrt{t}) = O(T^{3/2}).$$

For (c), we have that $S_t^2 = (S_{t-1}+X_t)^2 = S_{t-1}^2 + 2S_{t-1}X_t + X_t^2$. After taking summations across t and rearranging, we are left with

$$\sum_{t=1}^{T} S_{t-1}X_t = \frac{1}{2}\left(S_T^2 - \sum_{t=1}^{T} X_t^2\right).$$

Then, by the triangle inequality we have for $q > 4$:

$$\begin{aligned}
2\left\|\sum_{t=1}^{T} S_{t-1}X_t\right\|_{\mathcal{L}^{q/2}} &= \left\|S_T^2 - \sum_{t=1}^{T} X_t^2\right\|_{\mathcal{L}^{q/2}} \\
&= \left\|S_T^2 - \sum_{t=1}^{T}(X_t^2 - \mathbb{E}X_t^2) - \sum_{t=1}^{T} \mathbb{E}X_t^2\right\|_{\mathcal{L}^{q/2}} \\
&\leq \left\|S_T^2\right\|_{\mathcal{L}^{q/2}} + \left\|\sum_{t=1}^{T}(X_t^2 - \mathbb{E}X_t^2)\right\|_{\mathcal{L}^{q/2}} + \sum_{t=1}^{T} \mathbb{E}X_t^2.
\end{aligned}$$

Since the $\mathcal{L}^q$ norm is submultiplicative, the first term is upper bounded by $\|S_T\|_{\mathcal{L}_{q/2}}^2$, which is $O(T)$ by (a). The second term is also $O(T)$ by (a) since $X_t^2 - \mathbb{E}X_t^2$ is a zero-mean strong mixing process with finite moments of order $q/2 + \delta/2$. Finally, the last is $O(T)$, and we conclude that $\|\sum_{t=1}^{T} S_{t-1}X_t\|_{\mathcal{L}^{q/2}} = O(T)$. The result (c) then follows from the triangle inequality because

$$\left\|\sum_{t=1}^{T} S_t X_t\right\|_{\mathcal{L}^{q/2}} \leq \left\|\sum_{t=1}^{T} S_{t-1}X_t\right\|_{\mathcal{L}^{q/2}} + \left\|\sum_{t=1}^{T} X_t^2\right\|_{\mathcal{L}^{q/2}} = O(T).$$

Finally, for (d), we have by the triangle inequality followed by (a):

$$\left\|\sum_{t=1}^{T} S_t^2\right\|_{\mathcal{L}^q} \leq \sum_{t=1}^{T} \left\|S_t^2\right\|_{\mathcal{L}^q} = \sum_{t=1}^{T} O\left(t\right) = O(T^2).$$

□

Lemma 2. *Let $\{\boldsymbol{X}_t := (X_{1t} \ldots X_{pt})', t \in \mathbb{N}\}$ be a $\mathbb{R}^p$-valued zero-mean strong mixing random vector process with mixing coefficient given by $\alpha(m) = \mathsf{exp}(-2cm)$ for some $c > 0$.*

Additionally, consider that the following class of functions

$$\Psi := \{\psi : \mathbb{R} \to \mathbb{R} : \psi(x) = |x|^q, \psi(x) = \exp x^r, q > 2, r > 0\}.$$

Suppose that:

(i) *There exists* $q > 2$ *such that* $\sup_t \sup_{i \le p} \mathbb{E}|X_{it}|^{q+\delta} < C_q < \infty$ *for some* $\delta > 0$ *and*

(ii) *there exist positive constants* a_1, a_2 *and* a_3*, such that* $\sup_t \sup_{i\le p} \mathbb{P}(|X_{it}| > u) \le a_1 \exp(-a_2 x^{a_3})$ *for all* $x > 0$.

Then, for every $x > 0$*, we have*

(a) $\mathbb{P}\left(\|\boldsymbol{X}_t\|_\infty \ge x\right) \le C_1 p/\psi(x)$.

(b) $\mathbb{P}\left(\frac{1}{\sqrt{T}}\left\|\sum_{t=1}^T \boldsymbol{X}_t\right\|_\infty \ge x\right) \le C_2 p/x^q$

(c) $\mathbb{P}\left(\frac{1}{\sqrt{T}}\left\|\sum_{t=1}^T \boldsymbol{X}_t\right\|_\infty \ge x\right) \le R_{1,T}$.

(d) $\mathbb{P}\left[\frac{1}{\sqrt{T}}\left\|\sum_{t=1}^T \boldsymbol{X}_t\boldsymbol{X}_t' - \mathbb{E}(\boldsymbol{X}_t\boldsymbol{X}_t')\right\|_\infty \ge x\right] \le C_3 p^2/x^q$

(e) $\mathbb{P}\left[\frac{1}{\sqrt{T}}\left\|\sum_{t=1}^T \boldsymbol{X}_t\boldsymbol{X}_t' - \mathbb{E}(\boldsymbol{X}_t\boldsymbol{X}_t')\right\|_\infty \ge x\right] \le R_{2,T}$

where $C_j, j = 1, 2, 3$ *are constants depending on* q *and* c*. Additionally,*

$$\begin{aligned}
R_{1,T} &= p\exp\left\{2c_2\left[\sigma + \frac{1}{4c_1^2(\log T)^4}\right] - \frac{x}{2}\right\} \\
&\quad + \sqrt{T}p\left\{\mathbb{1}\left[\frac{x}{2} \le \mu_1\left(\frac{M}{2}\right)\right] + \mathbb{1}\left[\frac{x}{2} > \mu_1\left(\frac{M}{2}\right)\right] a_1 \exp\left[-a_2(M/2)^{a_3}\right]\right\} \\
R_{2,T} &= p^2\exp\left\{2c_2\left[\kappa + \frac{1}{4c_1^2(\log T)^4}\right] - \frac{x}{2}\right\} \\
&\quad + \sqrt{T}p^2\left\{\mathbb{1}\left[\frac{x}{2} \le \omega\sqrt{\mu_2\left(\sqrt{\frac{M}{2}}\right)}\right] + \mathbb{1}\left[\frac{x}{2} > \omega\sqrt{\mu_2\left(\sqrt{\frac{M}{2}}\right)}\right] 2a_1\exp\left[-a_2\left(\frac{M}{2}\right)^{a_3/2}\right]\right\},
\end{aligned}$$

where $M := \frac{\sqrt{T}}{2c_1(\log T)^2}$ *and, for* $k > 0$*,*

$$\mu_k(x) := |\mathbb{E}X_{it}^k \mathbb{1}(|X_{it}| > x)| \le 2\frac{a_1}{a_2^{k/a_3}}\boldsymbol{\gamma}\left(\frac{k}{a_3} + 1, a_2 x^{a_3}\right), \tag{S.1}$$

where $\boldsymbol{\gamma}(s,a) := \int_a^\infty x^{s-1}\,\mathsf{exp}(-x)\mathrm{d}x$ *is the incomplete upper Gamma function. For instance, when* $k = a_3 = 1$, *(S.1) turns out to be* $2\frac{a_1}{a_2^2}(1+a_2x)\,\mathsf{exp}(-a_2x)$.

If we further impose that $\mathsf{log}\, p = o(M^{a_3/2})$, *then, as* $T \to \infty$,

$$R_{1,T} \to p\,\mathsf{exp}\left(2c_2\sigma - \frac{x}{2}\right)$$
$$R_{2,T} \to p^2\,\mathsf{exp}\left(2c_2\kappa - \frac{x}{2}\right).$$

Proof. First, for any $(p_1 \times p_2)$ real-valued random matrix $\boldsymbol{Y}$ and $\psi \in \Psi$, we have by Markov's inequality that, for any $x > 0$,

$$\mathbb{P}(\|\boldsymbol{Y}\|_\infty \geq x) \leq \frac{\mathbb{E}[\psi(\|\boldsymbol{Y}\|_\infty)]}{\psi(x)} \leq \frac{p_1p_2\,\mathsf{sup}_{i\leq p_1; j\leq p_2}\,\mathbb{E}[\psi(|Y_{i,j}|)]}{\psi(x)}. \tag{S.2}$$

Part (a) then follows by setting $\boldsymbol{Y} = \boldsymbol{X}_t$ in (S.2) and applying the definition C_ψ. In the case $\psi(x) = |x|^q$, for part (b), set $\boldsymbol{Y} = \frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{X}_t$ or for part (d), set $\boldsymbol{Y} = \frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{X}_t\boldsymbol{X}_t - \mathbb{E}(\boldsymbol{X}_t\boldsymbol{X}_t')$ in (S.2), and we have Lemma 6 of Carvalho et al. (2018).

For part (c), if $\psi(x) = \mathsf{exp}(x)$, we use a truncation argument. For now, fix $M > 0$ and let $X_{it}^{\leq} := X_{it}\mathbb{1}(|X_{it}| \leq M/2) - \mathbb{E}[X_{it}\mathbb{1}(|X_{it}| \leq M/2)]$ and $X_{it}^{>} := X_{it}\mathbb{1}(|X_{it}| > M/2) - \mathbb{E}[X_{it}\mathbb{1}(|X_{it}| > M/2)]$ for $1 \leq i \leq p$ and $t \geq 1$. Since $\boldsymbol{X}_t$ is zero mean by assumption, we have that $X_{it} = X_{it}^{\leq} + X_{it}^{>}$. Furthermore, by construction, $X_{it}^{\leq}$ is a bounded (by M) zero-mean random variable. Therefore, from Theorem 2 in Merlevède et al. (2009), there exist positive constants c_1 and c_2, depending only on c, such that for all $T \geq 2$ and $0 < q < \frac{1}{c_1M(\mathsf{log}\,T)^2}$, the following inequality holds:

$$\mathsf{log}\,\mathbb{E}\left[\mathsf{exp}\left(q\sum_{t=1}^T X_{i,t}^{\leq}\right)\right] \leq \frac{c_2q^2(T\sigma_i^2 + M^2)}{1 - c_1Mq(\mathsf{log}\,T)^2}, \quad i = 1,\dots,p,$$

where $\sigma_i^2 := \mathsf{sup}_t \sum_{k\in\mathbb{Z}} |\mathbb{E}\left(X_{it}^{\leq}X_{it+k}^{\leq}|\right) < \infty$. If we set $q = \frac{1}{\sqrt{T}}$, take $M = \frac{\sqrt{T}}{2c_1(\mathsf{log}\,T)^2}$ and $\sigma^2 := \mathsf{sup}_{i\leq p}\,\sigma_i^2$, we have

$$\mathsf{log}\,\mathbb{E}\left[\mathsf{exp}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T X_{i,t}^{\leq}\right)\right] \leq 2c_2\left[\sigma^2 + \frac{1}{4c_1^2(\mathsf{log}\,T)^4}\right].$$

Let $\boldsymbol{X}_t^{\leq} := (X_{1t}^{\leq},\dots,X_{pt}^{\leq})'$. Then, applying (S.2) with $\boldsymbol{Y} = \frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{X}_t^{\leq}$ and $\psi(x) = \mathsf{exp}(x)$, we have

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{X}_t^{\leq}\right\|_\infty \geq x\right) \leq p\,\mathsf{exp}\left[2c_2\left(\sigma + \frac{1}{4c_1^2(\log T)^4}\right) - x\right].$$

We now bound $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{X}_t^{>}$, where $\boldsymbol{X}_t^{>} := (X_{1t}^{>}, \ldots, X_{pt}^{>})'$. First, notice that

$$\mathbb{P}\left[|X_{it}\mathbb{1}(|X_{it}| > M/2)| \geq x\right] \leq \mathbb{P}(|X_{it}| > M/2) \leq a_1\,\mathsf{exp}(-a_2(M/2)^{a_3}).$$

Also,

$$|\mathbb{E}[X_{it}\mathbb{1}(|X_{it}| > M/2)]| \leq \int_{\mathcal{X}_i} |x|\mathbb{1}(|x| > M/2)\mathrm{d}F_{it}(x) \leq 2\int_{M/2}^{\infty} xf(x)\mathrm{d}x,$$

where $F_{it}(x) := \mathbb{P}(X_{it} \leq x)$ and $f(x) = a_1a_2a_3x^{a_3-1}\,\mathsf{exp}(-a_2x^{a_3})$, i.e., $f := \frac{\mathrm{d}F}{\mathrm{d}x}$ with $F(x) := 1 - a_1\,\mathsf{exp}(-a_2x^{a_3})$. The last integral cannot be solved analytically when a_3 is not a positive integer. Apart from a change in variable, it is related to the incomplete upper Gamma function as defined above.

Then, by the triangle inequality, we have

$$\begin{aligned}
\mathbb{P}(|X_{it}^{>}| \geq x) &= \mathbb{P}\left\{|X_{it}\mathbb{1}(|X_{it}| > M/2) - \mathbb{E}[X_{it}\mathbb{1}(|X_{it}| > M/2)]| \geq x\right\} \\
&\leq \mathbb{P}\left[|X_{it}\mathbb{1}(|X_{it}| > M/2)| \geq x - \mu_1\left(\frac{M}{2}\right)\right] \\
&\leq \mathbb{1}\left[x \leq \mu_1\left(\frac{M}{2}\right)\right] + \mathbb{1}\left[x > \mu_1\left(\frac{M}{2}\right)\right]\mathbb{P}(|X_{it}| > M/2) \\
&\leq \mathbb{1}\left[x \leq \mu_1\left(\frac{M}{2}\right)\right] + \mathbb{1}\left[x > \mu_1(\frac{M}{2})\right] a_1\,\mathsf{exp}\left[-a_2(M/2)^{a_3}\right].
\end{aligned}$$

Apply the union bound to conclude that

$$\begin{aligned}
\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{X}_t^{>}\right\|_\infty \geq x\right) &\leq \sqrt{T}p\,\sup_t\sup_{i\leq p}\mathbb{P}(|X_{it}^{>}| \geq x) \\
&\leq \sqrt{T}p\left\{\mathbb{1}\left[x \leq \mu_1\left(\frac{M}{2}\right)\right] + \mathbb{1}\left[x > \mu_1\left(\frac{M}{2}\right)\right] a_1\,\mathsf{exp}\left[-a_2(M/2)^{a_3}\right]\right\}.
\end{aligned}$$

Combining both bounds and using the fact that $\{|A+B| \geq x\} \subseteq \{|A| \geq x/2\} \cup \{|B| \geq x/2\}$, we have

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{X}_t\right\|_{\infty}\geq x\right)\leq p\,\mathsf{exp}\left\{2c_2\left[\sigma+\frac{1}{4c_1^2(\mathsf{log}\,T)^4}\right]-\frac{x}{2}\right\}$$
$$+\sqrt{T}p\left\{\mathbb{1}\left[\frac{x}{2}\leq\mu_1\left(\frac{M}{2}\right)\right]+\mathbb{1}\left[\frac{x}{2}>\mu_1\left(\frac{M}{2}\right)\right]a_1\,\mathsf{exp}(-a_2(M/2)^{a_3})\right\}.$$

For (e), set $\psi(x)=\mathsf{exp}(x)$ and $\boldsymbol{Y}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t$ where $\boldsymbol{W}_t:=\boldsymbol{X}_t\boldsymbol{X}_t'-\mathbb{E}(\boldsymbol{X}_t\boldsymbol{X}_t')$ in (S.2) to obtain

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t\right\|_{\infty}\geq x\right)\leq\frac{p^2\,\mathsf{sup}_{1\leq i,j\leq p}\,\mathbb{E}\left[\mathsf{exp}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}W_{i,j,t}\right)\right]}{\mathsf{exp}(x)}.$$

We can conduct a similar truncation argument to the proof of part (c). Let $W_{i,j,t}=W_{i,j,t}^{\leq}+W_{i,j,t}^{>}$ where $W_{i,j,t}^{\leq}:=X_{it}X_{jt}\mathbb{1}\left[(|X_{it}|\vee|X_{jt}|)\leq\sqrt{M/2}\right]-\mathbb{E}\left\{X_{it}X_{jt}\mathbb{1}\left[(|X_{it}|\vee|X_{jt}|)\leq\sqrt{M/2}\right]\right\}$ and $W_{i,j,t}^{>}=X_{it}X_{jt}\mathbb{1}\left[(|X_{it}|\vee|X_{jt}|)>\sqrt{M/2}\right]-\mathbb{E}\left\{X_{it}X_{jt}\mathbb{1}\left[(|X_{it}|\vee|X_{jt}|)>\sqrt{M/2}\right]\right\}$; then by construction, for each $1\leq i,j\leq p$, we have that $\{W_{i,j,t}^{\leq}\}_{t\geq1}$ is a zero mean, bounded by M, strong mixing sequence with the same exponential decay of $\{\boldsymbol{X}_t\}_{t\geq1}$. For that reason,

$$\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t^{\leq}\right\|_{\infty}\geq x\right)\leq p^2\,\mathsf{exp}\left\{2c_2\left[\kappa+\frac{1}{4c_1^2(\mathsf{log}\,T)^4}\right]-x\right\},$$

where $\kappa^2:=\mathsf{sup}_{1\leq i,j\leq p}\,\mathsf{sup}_t\sum_{k\in\mathbb{Z}}|\mathbb{E}\left(W_{i,j,t}W_{i,j,t+k}\right)|<\infty$. For the second term, we have, by Hölder's inequality,

$$\begin{aligned}\left|\mathbb{E}\left(X_{it}X_{jt}\right)\mathbb{1}\left(|X_{it}|\vee|X_{jt}|>\sqrt{M/2}\right)\right|&\leq\mathbb{E}\left[|X_{it}X_{jt}|\mathbb{1}\left(|X_{it}|\vee|X_{jt}|>\sqrt{M/2}\right)\right]\\&\leq\left\{\mathbb{E}\left(X_{it}^2\right)\mathbb{E}\left[X_{jt}^2\mathbb{1}\left(|X_{it}|\vee|X_{jt}|>\sqrt{M/2}\right)\right]\right\}^{1/2}\\&\leq\left\{\mathbb{E}X_{it}^2\mathbb{E}\left[X_{jt}^2\mathbb{1}\left(|X_{jt}|>\sqrt{M/2}\right)\right]\right\}^{1/2}\\&\leq\omega\left[\mu_2\left(\sqrt{M/2}\right)\right]^{1/2},\end{aligned}$$

where $\mathsf{sup}_t\,\mathsf{sup}_i\,\mathbb{E}\left(X_{it}^2\right)\leq\omega^2<\infty$ and $\mu_2(\cdot)$ is defined in (S.1).

Then, by the triangle inequality,

$$
\begin{aligned}
\mathbb{P}(|W^{>}_{i,j,t}| \geq x) &= \mathbb{P}\left\{\left|X_{it}X_{jt}\mathbb{1}\left(|X_{it}| \vee |X_{jt}| > \sqrt{M/2}\right) - \mathbb{E}\left[X_{it}X_{jt}\mathbb{1}\left(|X_{it}| \vee |X_{jt}| > \sqrt{M/2}\right)\right]\right| \geq x\right\} \\
&\leq \mathbb{P}\left\{\left|X_{it}X_{jt}\mathbb{1}\left(|X_{it}| \vee |X_{jt}| > \sqrt{M/2}\right)\right| \geq x - \omega\left[\mu_2\left(\sqrt{M/2}\right)\right]^{1/2}\right\} \\
&\leq \mathbb{1}\left\{x \leq \omega\left[\mu_2\left(\sqrt{M/2}\right)\right]^{1/2}\right\} \\
&\quad + \mathbb{1}\left\{x > \omega\left[\mu_2\left(\sqrt{M/2}\right)\right]^{1/2}\right\}\mathbb{P}\left(|X_{it}| \vee |X_{jt}| > \sqrt{M/2}\right) \\
&\leq \mathbb{1}\{x \leq \omega\left[\mu_2\left(\sqrt{M/2}\right)\right]^{1/2}\} \\
&\quad + \mathbb{1}\left\{x > \omega\left[\mu_2\left(\sqrt{M/2}\right)\right]^{1/2}\right\} 2a_1 \exp\left[-a_2(M/2)^{a_3/2}\right].
\end{aligned}
$$

Once again, apply the union bound to conclude

$$
\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}^{>}_{t}\right\|_{\infty} \geq x\right) \leq \sqrt{T}p^2 \sup_{t\leq T}\sup_{1\leq i,j\leq p}\mathbb{P}(|W^{>}_{i,j,t}| \geq x).
$$

Combining both bounds using the fact that $\{|A+B| \geq x\} \subseteq \{|A| \geq x/2\} \cup \{|B| \geq x/2\}$, we have

$$
\begin{aligned}
\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_{t}\right\|_{\infty} \geq x\right) &\leq p^2 \exp\left\{2c_2\left[\kappa + \frac{1}{4c_1^2(\log T)^4}\right] - \frac{x}{2}\right\} \\
&\quad + \sqrt{T}p^2\mathbb{1}\left\{\frac{x}{2} \leq \omega\left[\mu_2\left(\sqrt{M/2}\right)\right]^{1/2}\right\} \\
&\quad + p^2\mathbb{1}\left\{\frac{x}{2} > \omega\left[\mu_2\left(\sqrt{M/2}\right)\right]^{1/2}\right\} 2a_1 \exp\left[-a_2(M/2)^{a_3/2}\right].
\end{aligned}
$$

For the second part of the Lemma, we use the upper bound for the incomplete upper Gamma function given by Natalini and Palumbo (2000), which states that for $s > 1$, $b > 1$ and $a > \frac{b}{b-1}(s-1)$, we have $\boldsymbol{\gamma}(s,a) < ba^{s-1}\exp(-a)$. Applying this bound in (S.1) with $b = 2$, we have that for all $k > 0$ and $y > 2k/a_3$:

$$
\mu_k(y) := 2\frac{a_1}{a_2^{k/a_3}}\boldsymbol{\gamma}(k/a_3 + 1, a_2 y^{a_3}) < 4a_1 y^k \exp(-a_2 y^{a_3}),
$$

from which we conclude that $\mu_k(y) \to 0$ as $y \to \infty$.

Since $M \to \infty$ is $T \to \infty$, we have for each $x > 0$, there is a $T_x \in \mathbb{N}$ such that

$x > 2\left\{\mu_1(M/2) \vee \omega\left[\mu_2\left(\sqrt{M/2}\right)\right]^{1/2}\right\}$, whenever $T > T_x$. Thus, for $T > T_x$, we have

$$R_{1,T} = p \exp\left\{2c_2\left[\sigma + \frac{1}{4c_1^2(\log T)^4}\right] - \frac{x}{2}\right\} + \sqrt{T}pa_1 \exp\left[-a_2(M/2)^{a_3}\right]$$

$$R_{2,T} = p^2 \exp\left\{2c_2\left[\kappa + \frac{1}{4c_1^2(\log T)^4}\right] - \frac{x}{2}\right\} + \sqrt{T}p^2 2a_1 \exp\left[-a_2\left(\frac{M}{2}\right)^{a_3/2}\right].$$

Hence, as long as $\log p = o(M^{a_3/2})$, we have the second result of the Lemma. □

4 List of Symbols

4.1 The Romans

4.1.1 Lower case

a	Exponent for R_b to expose the condition of the compatibility condition
b	Exponent radius for the weak sparsity definition
$c, c_1, c_2, \ldots$	Generic positive constants
d	Generic Deterministic Trend
e	Exponential
f	Deterministic Trends
g	Generic continuous function for the infetence procedure
h	Cardinality of set $\mathcal{H}$
i	Unit index
j	Regressor index
k	Regressor index 2
ℓ, L	Scaling matrix and its entries
m	Lag of alpha mixing
n	number of units
o, o_p	Landou notation
p	Number of regressors
q	Number of moments
r	Number of I(0) relations
s, s_0	Cardinality of index set
t	Time index
u	
v	
w	Individual weights of the LASSO
x	
y	
z	

4.1.2 Upper case

A	Random element of proof of Theorem 3
B	Standard Brownian motion
F	Factor of the common factor model
G	Generic random vector of Assumption 3 and Definition 1
H	Transformed objective function
M	Generic matrix used in GIF
$I(\cdot)$	Integrated process
J	Linear combination of I(0) processes
O, o, O_P, o_P	Landou notation
Q	LASSO Objective function
R	Remainder of Lemma 2
T, T_0, T_1, T_1	Sample size and Treatment, Pre and Post
U, U^Z, U^F	Innovation
V	Regression error
$X, Y, W, Z^{(0)}, Z^{(1)}$	Units and its transformation

4.2 The Greeks

α	Mixing coefficient
$\boldsymbol{\theta}, \boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}$	Parameter, True and Estimated
$\boldsymbol{\gamma}, \boldsymbol{\gamma}_0, \widehat{\boldsymbol{\gamma}}$	Transformed parameter, True and Estimated
$\delta, \widehat{\delta}, \Delta, \widehat{\Delta}$	Treatment effect, ATE and Estimates
ϵ	Arbitrary small positive constant
ζ	Linear Projecion in the Factor Model
η	The stochastic component of the DGP
θ, Θ	Parameters of the generic model
ι	Vector of 1s
κ	Auxiliary Lemma 1 Appendix
λ, λ_0	Penalty parameter
μ	Constant of the deterministic trend
ν	Combined weight trend
ξ	Cone constant
$\boldsymbol{\pi}$	Projection of I(0) process
ρ	Simulation autocorrelation coefficient
σ	Variance of the innovation
τ	Quantiles
υ	Variance of the defining I(0) process
$\phi, \widehat{\phi}, \phi_j$	The Inference function
χ	GIF Constant
ψ, Ψ	Deterministic Trends
$\Omega, \Omega_0, \Omega_1, \ldots, \omega$	Sample space, events
$\boldsymbol{\gamma}, \widetilde{\boldsymbol{\gamma}}$	Cointegration matrix
$\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_0$	Covariance matrix of WW'

4.3 Miscellaneous

$\mathbb{N}, \mathbb{Z}, \mathbb{R}$	Naturals, integers and real
$\mathscr{C}$	Cone
$\mathscr{H}$	Test hypothesis
$\mathscr{F}$	Sigma algebra
$\mathbb{P}, \mathbb{E}$	Probability and expectation operator
$\mathcal{D}$	Intervention indicator
U	Innovation
$\mathcal{M}$	Generic model
$\mathcal{G}$	Process to define I(0)
$\mathcal{H}$	Set index of growth condition
$\mathcal{S}, \mathcal{S}_0$	Set index
R	index set in the proof of Proposition 3

References

C.V. Carvalho, R. Masini, and M.C. Medeiros. Arco: An artificial counterfactual approach for high-dimensional panel time-series data. *Journal of Econometrics*, 207:352–380, 2018.

A. Ibragimov. Some limit theorems for stationary processes. *Theory of Probability and its Applications*, 7:349–382, 1962.

F. Merlevède, M. Peligrad, and E. Rio. Bernstein inequality and moderate deviations under strong mixing conditions. In C. Houdré, V. Koltchinskii, D.M. Mason, and M. Peligrad, editors, *High Dimensional Probability V: The Luminy Volume*, volume Volume 5, pages 273–292. Institute of Mathematical Statistics, 2009.

P. Natalini and B. Palumbo. Inequalities for the incomplete Gamma function. *Mathematical Inequalities and Applications*, 3:69–77, 2000.

S. Resnick. *A Probability Path*. Birkhäuser Boston, 1999.

E. Rio. Inégalités de moments pour les suites stationnaires et fortement mélangeantes. *Comptes rendus Acad. Sci. Paris, Série I*, 318:355–360, 1994.

A.W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.

Table S.1: **Rejection Rates under the Null (empirical size): Mixed Trends**

Baseline DGP: (5.1) and (5.2) with $T = 100$, independent and identically normally distributed innovations, $n = 200$, $s_0 = 5$, $T_1 = 3$ and $10,000$ Monte Carlo simulations. The test statistic considered is $\phi(x) = \|x\|_2$. All distributions are standardized (zero mean and unit variance). Mixed normal is equal to 2 Normal distributions with probability $(0.3, 0.7)$, mean $(-10, 10)$ and variance $(2, 1)$. The AR(1) structure with coefficient ρ is applied to the common factor innovation U_{1t}^F and the first unit idiosyncratic innovation U_{1t}^Z. The penalization parameter λ is chosen via the Bayesian Information Criterion (BIC). We set the maximum penalty level to be $\|\frac{1}{T_0}\sum_{t=1}^{T_0} Y_t \boldsymbol{X}_t\|_\infty$ with an exponential path down to $\lambda_{\mathsf{min}} = 0.001$ along 100 equally spaced intervals in the `glmnet` package. **Oracle** means OLS estimation in the pre-intervention period with known active regressors S_0 (perfect model selection). **True** means no estimation in the pre-intervention period. True parameter θ_0 was used.

	LASSO			**Oracle**			**True**		
	0.01	**0.5**	**0.1**	**0.01**	**0.05**	**0.1**	**0.01**	**0.05**	**0.1**
				Innovation Distribution					
Normal	0.0406	0.0888	0.1480	0.0415	0.0813	0.1414	0.0288	0.0677	0.1123
$\chi^2(1)$	0.0351	0.0776	0.1482	0.0293	0.0819	0.1313	0.0285	0.0616	0.1066
t-stud(3)	0.0296	0.0860	0.1461	0.0341	0.0858	0.1440	0.0269	0.0652	0.1129
Mixed Normal	0.0448	0.0967	0.1541	0.0345	0.0958	0.1519	0.0248	0.0619	0.1186
				Sample Size					
$T = 50$	0.0475	0.0932	0.1499	0.0486	0.0878	0.1510	0.0363	0.0766	0.1138
100	0.0406	0.0888	0.1480	0.0415	0.0813	0.1414	0.0288	0.0677	0.1123
150	0.0382	0.0814	0.1531	0.0369	0.0835	0.1347	0.0311	0.0712	0.1091
200	0.0391	0.0936	0.1505	0.0369	0.0870	0.1499	0.0319	0.0707	0.1213
500	0.0452	0.1047	0.1606	0.0413	0.1008	0.1542	0.0318	0.0633	0.1211
				Number of Total Units					
$n = 200$	0.0406	0.0888	0.1480	0.0415	0.0813	0.1414	0.0288	0.0677	0.1123
300	0.0277	0.0857	0.1483	0.0285	0.0798	0.1340	0.0235	0.0671	0.1106
500	0.0305	0.0874	0.1488	0.0320	0.0801	0.1397	0.0274	0.0630	0.1214
1000	0.0401	0.0930	0.1455	0.0356	0.0874	0.1477	0.0211	0.0673	0.1158
				Number of Relevant (nonzero) Covariates					
$s_0 = 2$	0.0261	0.0705	0.1272	0.0226	0.0668	0.1218	0.0197	0.0558	0.1063
5	0.0406	0.0888	0.1480	0.0415	0.0813	0.1414	0.0288	0.0677	0.1123
50	0.0502	0.1121	0.1806	0.2544	0.3637	0.4448	0.0181	0.0577	0.1064
97	0.0580	0.1261	0.1958	1.0007	1.0007	1.0009	0.0205	0.0584	0.1069
				Deterministic Component					
$f_t^F = \sqrt{t}$	0.0406	0.0888	0.1480	0.0415	0.0813	0.1414	0.0288	0.0677	0.1123
t	0.0320	0.0815	0.1380	0.0323	0.0816	0.1394	0.0211	0.0623	0.1126
$t^{3/2}$	0.0266	0.0698	0.1196	0.0294	0.0822	0.1387	0.0223	0.0606	0.1091
t^2	0.0267	0.0713	0.1230	0.0293	0.0776	0.1339	0.0189	0.0561	0.1058
				Serial Correlation					
$\rho = 0$	0.0406	0.0888	0.1480	0.0415	0.0813	0.1414	0.0288	0.0677	0.1123
0.5	0.0301	0.0791	0.1323	0.0282	0.0770	0.1324	0.0188	0.0577	0.1020
0.7	0.0280	0.0776	0.1337	0.0269	0.0782	0.1347	0.0214	0.0582	0.1074
0.9	0.0303	0.0756	0.1279	0.0326	0.0830	0.1368	0.0229	0.0638	0.1109
				Postintervention Periods					
$T_1 = 1$	0.0325	0.0754	0.1279	0.0311	0.0719	0.1205	0.0299	0.0699	0.1154
2	0.0292	0.0783	0.1316	0.0275	0.0765	0.1312	0.0224	0.0765	0.1231
3	0.0406	0.0888	0.1480	0.0415	0.0813	0.1414	0.0288	0.0677	0.1123
4	0.0398	0.0937	0.1525	0.0352	0.0885	0.1432	0.0221	0.0610	0.1088
5	0.0520	0.1095	0.1700	0.0474	0.1029	0.1644	0.0298	0.0668	0.1184

Table S.2: **Monte Carlo Results: Estimation**

The table reports several statistics averaged over 10,000 replications for each one of four data generating processes. More specifically, the mean ℓ_1-norm is the average $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_1$, the mean bias is the average bias $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ over the simulations, the mean MSE is the average mean squared error, and the mean Δ is the average intervention effect over the 10 out-of-sample periods. Note that the true value of Δ is zero. MSE Δ is the average squared error over the simulation, and, finally, median Δ is the median of the estimates of Δ over the simulations. Each column in the table represents a variation of the baseline scenario, in which we set $T = 100$,$s_0 = 5$, $n = 100$ and $\rho = 0$. Model (1) is given by equations (5.1) and (5.2) where $f_t^F = 0$. Model (2) is given by equations (5.1) and (5.2) where $f_t^F = 1$. Model (3) is given by equations (5.1) and (5.3) where $f_t^F = t$. Model (4) is given by equations (5.1) and (5.3) where $f_t^F = t^2$.

Model	**Statistic**	**Baseline**	**Sample Size**		**Sparsity**		**Regressors**		**Autocorrelation**	
			$T = 500$	$T = 1000$	$s_0 = 1$	$s_0 = 10$	$n = 50$	$n = 200$	$\rho = 0.2$	$\rho = 0.5$
(1)	mean ℓ_1-norm	1.36	0.26	0.13	0.19	3.04	0.99	1.72	1.46	1.87
	mean bias	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	mean MSE	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
	mean Δ	-0.03	-0.03	0.02	0.01	-0.04	0.01	0.01	0.03	-0.19
	MSE Δ	1.57	0.25	0.17	0.33	3.48	1.00	2.27	2.13	4.99
	median Δ	-0.03	-0.03	0.02	0.01	-0.04	0.01	0.01	0.03	-0.19
(2)	mean ℓ_1-norm	2.46	0.34	0.15	0.63	4.38	1.52	3.55	2.91	3.83
	mean bias	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	mean MSE	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01
	mean Δ	0.10	-0.02	-0.01	-0.28	-0.08	-0.17	-0.30	0.08	-0.17
	MSE Δ	3.20	0.29	0.15	0.93	6.24	1.56	5.72	4.53	13.21
	median Δ	0.10	-0.02	-0.01	-0.28	-0.08	-0.17	-0.30	0.08	-0.17
(3)	mean ℓ_1-norm	3.45	0.66	0.32	1.02	5.82	1.96	4.61	3.68	3.95
	mean bias	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	mean MSE	0.01	0.00	0.00	0.00	0.02	0.01	0.01	0.01	0.01
	mean Δ	0.01	-0.02	0.00	-0.08	0.00	0.13	0.00	-0.11	-0.08
	MSE Δ	4.81	0.39	0.23	1.73	7.41	2.25	7.74	5.87	15.51
	median Δ	0.01	-0.02	0.00	-0.08	0.00	0.13	0.00	-0.11	-0.08
(4)	mean ℓ_1-norm	1.46	0.64	0.58	0.33	2.93	1.24	1.66	1.52	1.93
	mean bias	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	mean MSE	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00
	mean Δ	-0.06	0.01	-0.01	-0.29	-0.03	-0.06	-0.07	-0.06	-0.08
	MSE Δ	0.22	0.12	0.12	0.25	0.30	0.18	0.26	0.32	0.73
	median Δ	-0.06	0.01	-0.01	-0.29	-0.03	-0.06	-0.07	-0.06	-0.08