SUPPLEMENTARY MATERIAL FOR ARCO: AN ARTIFICIAL COUNTERFACTUAL APPROACH FOR HIGH-DIMENSIONAL PANEL TIME-SERIES DATA

Carlos 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
São 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

This supplementary material contains additional material to support the results presented in the paper "ArCo: An Artificial Counterfactual Approach for High-Dimensional Panel Time-Series Data" co-authored by Ricardo P. Carlos Carvalho, Ricardo P. Masini and Marcelo C. Medeiros. The supporting material consists of more detailed simulation results, proofs of the lemmas in the main paper and additional empirical results.
Keywords: counterfactual analysis, comparative studies, multivariate treatment effects, synthetic control, policy evaluation, LASSO, structural break, factor models.
JEL Codes: C22, C23, C32, C33.

1. Introduction

This supplementary material contains additional material to support the results presented in the paper "ArCo: An Artificial Counterfactual Approach for High-Dimensional Panel Time-Series Data" co-authored by Ricardo P. Carlos Carvalho, Ricardo P. Masini and Marcelo C. Medeiros. The supporting material consists of more detailed simulation results, proofs of the lemmas in the main paper and additional empirical results.

The supplementary material is organized as follows. In Section 2 we describe a data generating process to motivate the usefulness of the ArCo method. Section 3 contains additional simulation results. The proofs of the lemmas in the main paper are presented in Section 4. Finally, complementary empirical results are included in Section 5.

[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.

2. A Possible Data Generating Process

Even though we do not impose any specific DGP, the link between the treated unit and its peers can be easily motivated by a very simple, but general, common factor model:

$$\boldsymbol{z}_{it}^{(0)} = \boldsymbol{\mu}_i + \boldsymbol{\Psi}_{\infty,i}(L)\boldsymbol{\varepsilon}_{it}, \quad i = 1, \ldots, n;\ t \geq 1 \tag{S.1}$$

$$\boldsymbol{\varepsilon}_{it} = \boldsymbol{\Lambda}_i \boldsymbol{f}_t + \boldsymbol{\eta}_{it}, \tag{S.2}$$

where $\boldsymbol{f}_t \in \mathbb{R}^f$ is a vector of common unobserved factors such that $\mathsf{sup}_t\, \mathbb{E}(\boldsymbol{f}_t \boldsymbol{f}'_t) < \infty$ and $\boldsymbol{\Lambda}_i$, is a $(q_i \times f)$ matrix of factor loadings. Therefore, we allow for heterogeneous determinist trends of the form $\zeta(t/T)$, where ζ is a integrable function on $[0, 1]$ as in Bai (2009). $\{\boldsymbol{\eta}_{it}\}$,$i = 1, \ldots, n$, $t = 1, \ldots, T$, is a sequence of uncorrelated zero mean random variables. Finally, L is the lag operator and the polynomial matrix $\boldsymbol{\Psi}_{\infty,i}(L) = (\boldsymbol{I}_{q_i} + \boldsymbol{\psi}_{1i}L + \boldsymbol{\psi}_{2i}L^2 + \cdots)$ is such that $\sum_{j=0}^{\infty} \boldsymbol{\psi}_{ji}^2 < \infty$ for all $i = 1, \ldots, n$. $\boldsymbol{I}$ is the identity matrix. Usually, we have $f < n$. Thus, as long as we have a "truly common" factor in the sense of having some rows of $\boldsymbol{\Lambda}_i$ non zero, we expect correlation among the units.

The DGP originated by (S.1) is fairly general and nests several models as by the multivariate Wold decomposition and under mild conditions, any second-order stationary vector process can be written as an infinite order vector moving average process; see Niemi (1979). Furthermore, under a modern macroeconomics perspective, reduced-form for Dynamic Stochastic General Equilibrium (DSGE) models are written as vector autoregressive moving average (VARMA) processes, which, in turn, are nested in the general specification in (S.1) (Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson, 2007; An and Schorfheide, 2007). Gobillon and Magnac (2016) is a special case of the general model described above.

In case of Gaussian errors, the above model will imply that $\mathbb{E}[\boldsymbol{y}_t^{(0)}|\boldsymbol{Z}_{0t}] = \boldsymbol{\Pi}\boldsymbol{Z}_{0t}$. Otherwise, we can choose model $\mathcal{M}$ to be a linear approximation of the conditional expectation. The strategy is to define $\boldsymbol{x}_t$ as a set of transformations of $\boldsymbol{Z}_{0t}$, such as, for instance, polynomials or splines, and write $\boldsymbol{y}_t^{(0)}$ as a linear function of $\boldsymbol{x}_t$.

3. Supplementary Monte Carlo Results

Figures S.1–S.6 present the smoothed histograms for the different estimators considered in the Monte Carlo study.

Figure S.1. Kernel Density - Estimator Comparison with no Trend and no Serial Correlation

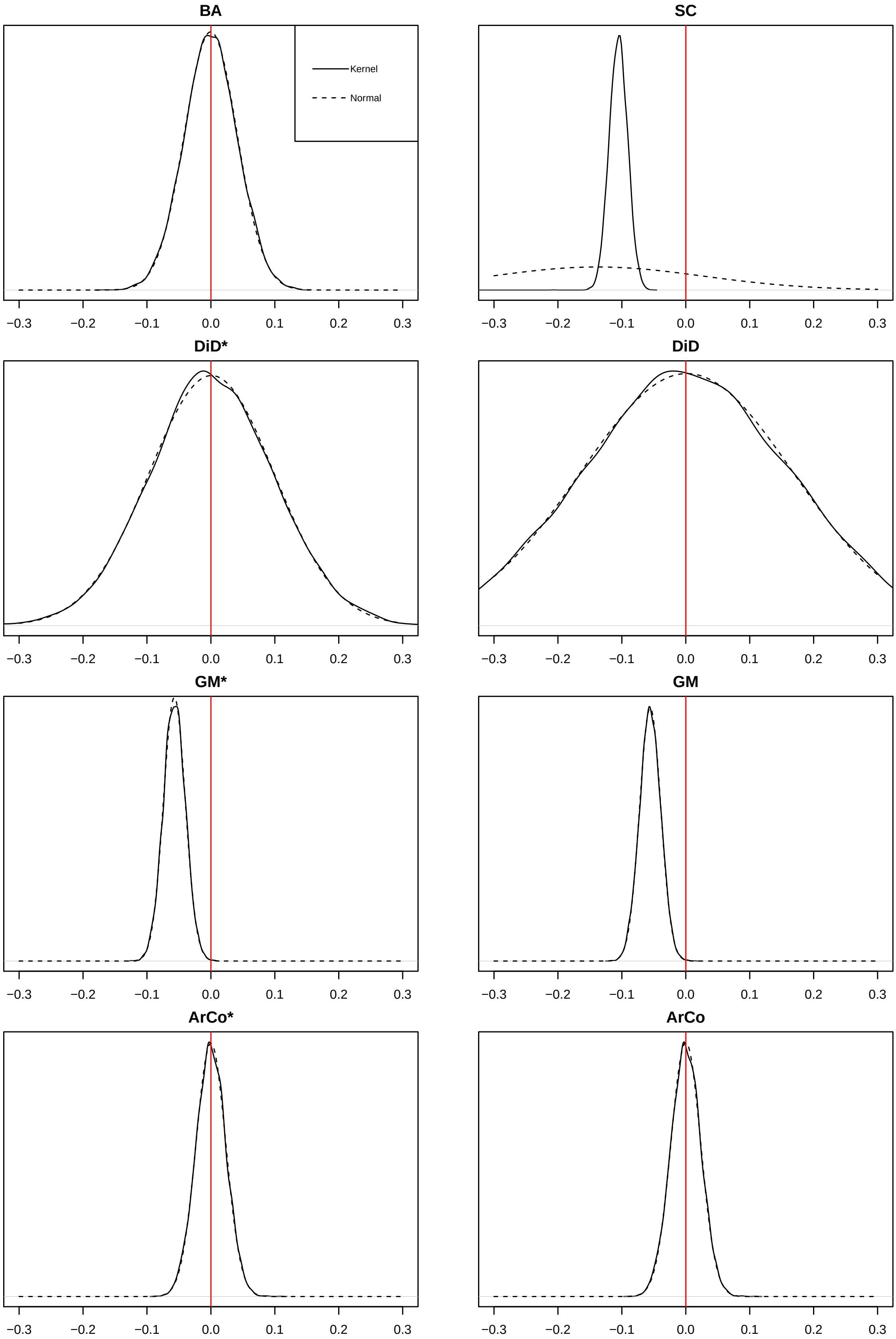


Figure S.2. Kernel Density - Estimator Comparison with no Trend

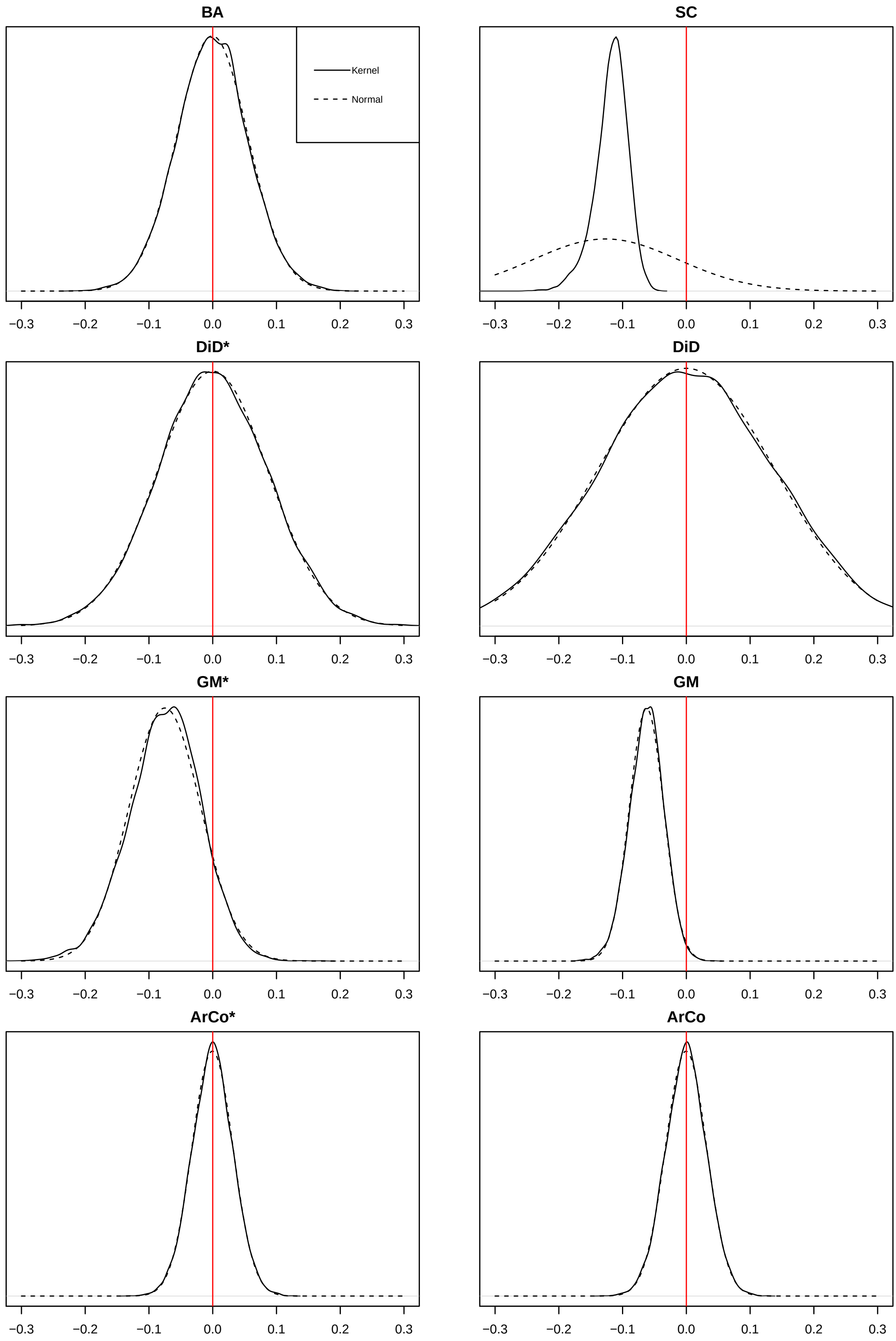


FIGURE S.3. Kernel Density - Estimator Comparison with Common Linear Trend

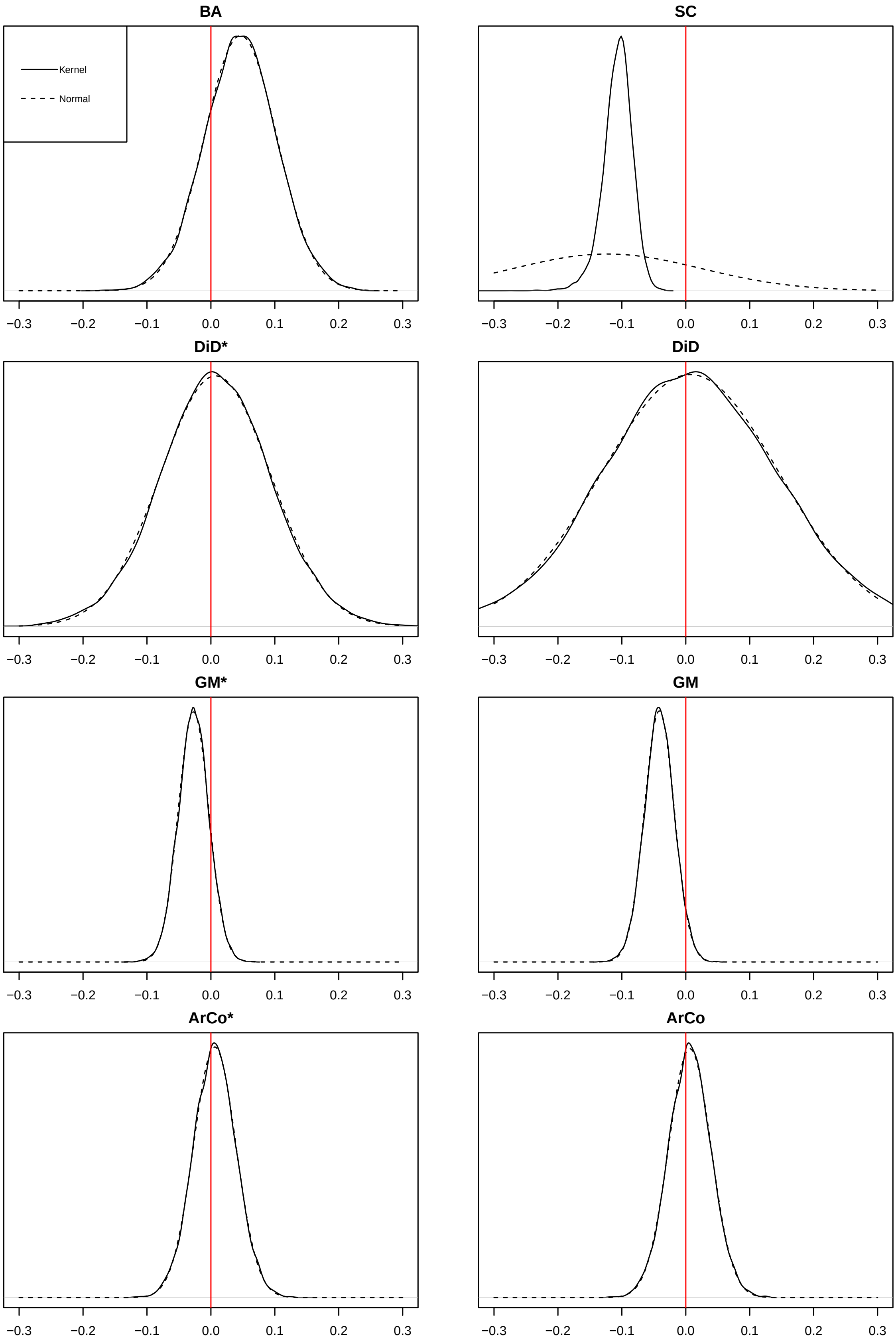


Figure S.4. Kernel Density - Estimator Comparison with Idiosyncratic Linear Trend

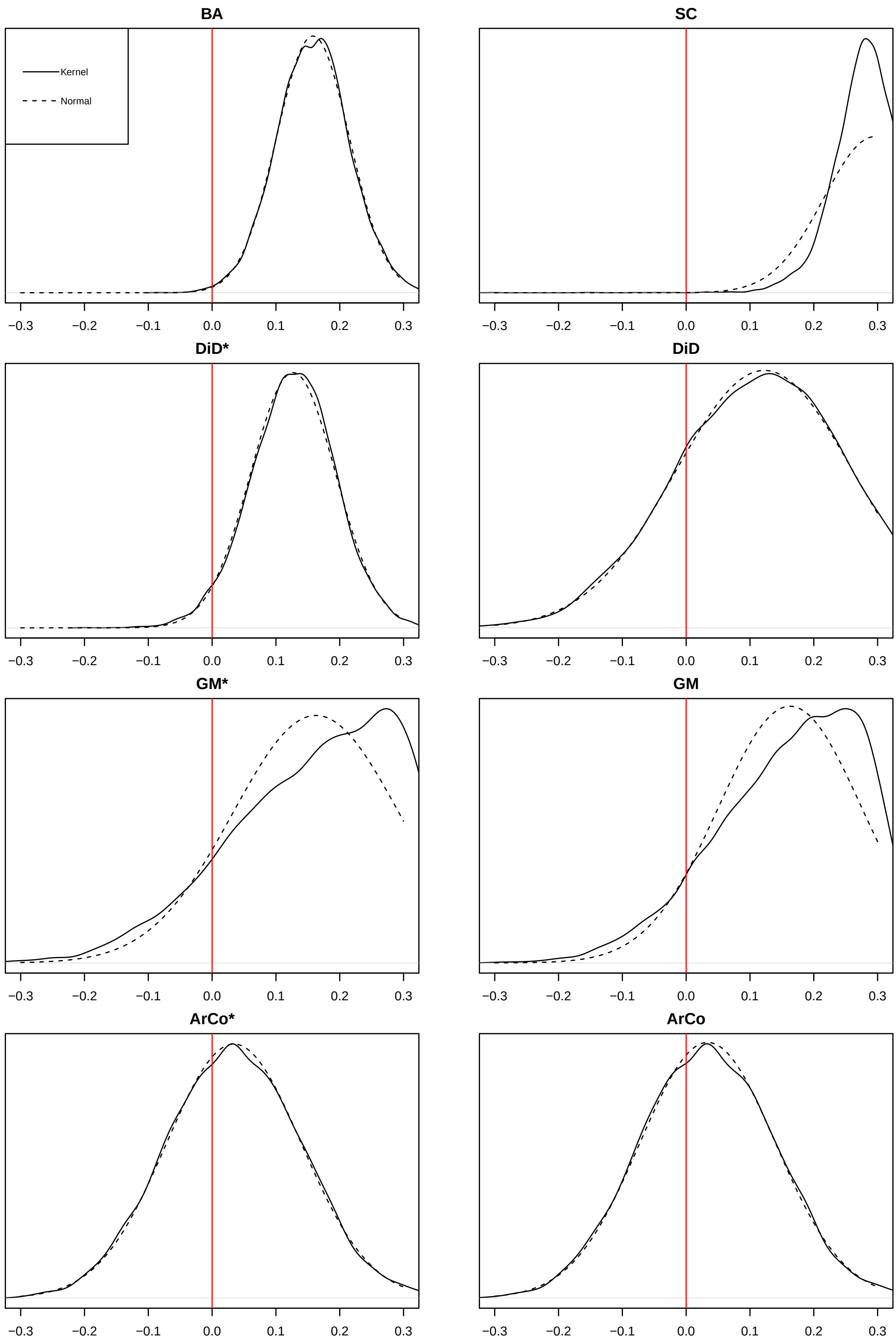


Figure S.5. Kernel Density - Estimator Comparison with Common Quadratic Trend

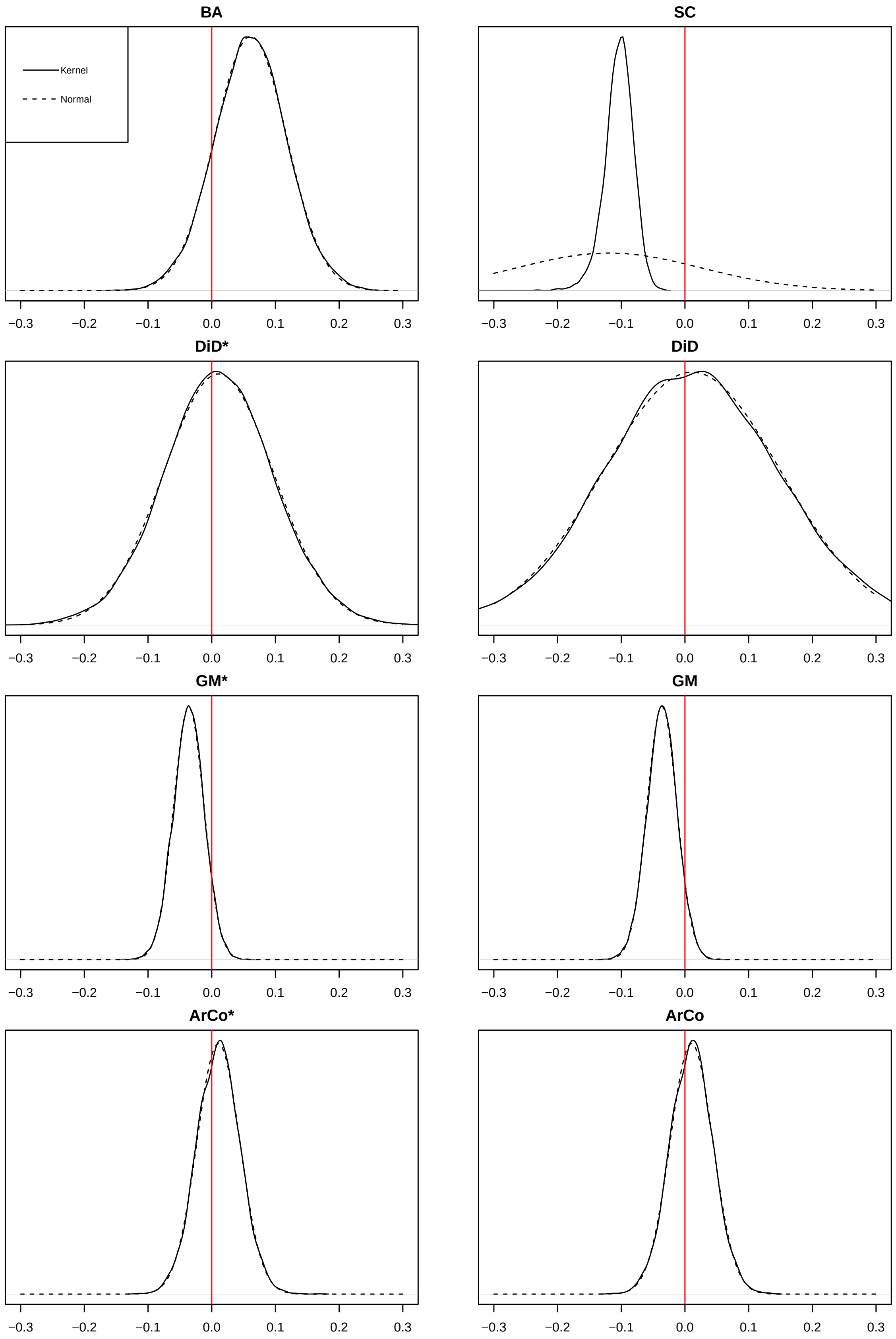
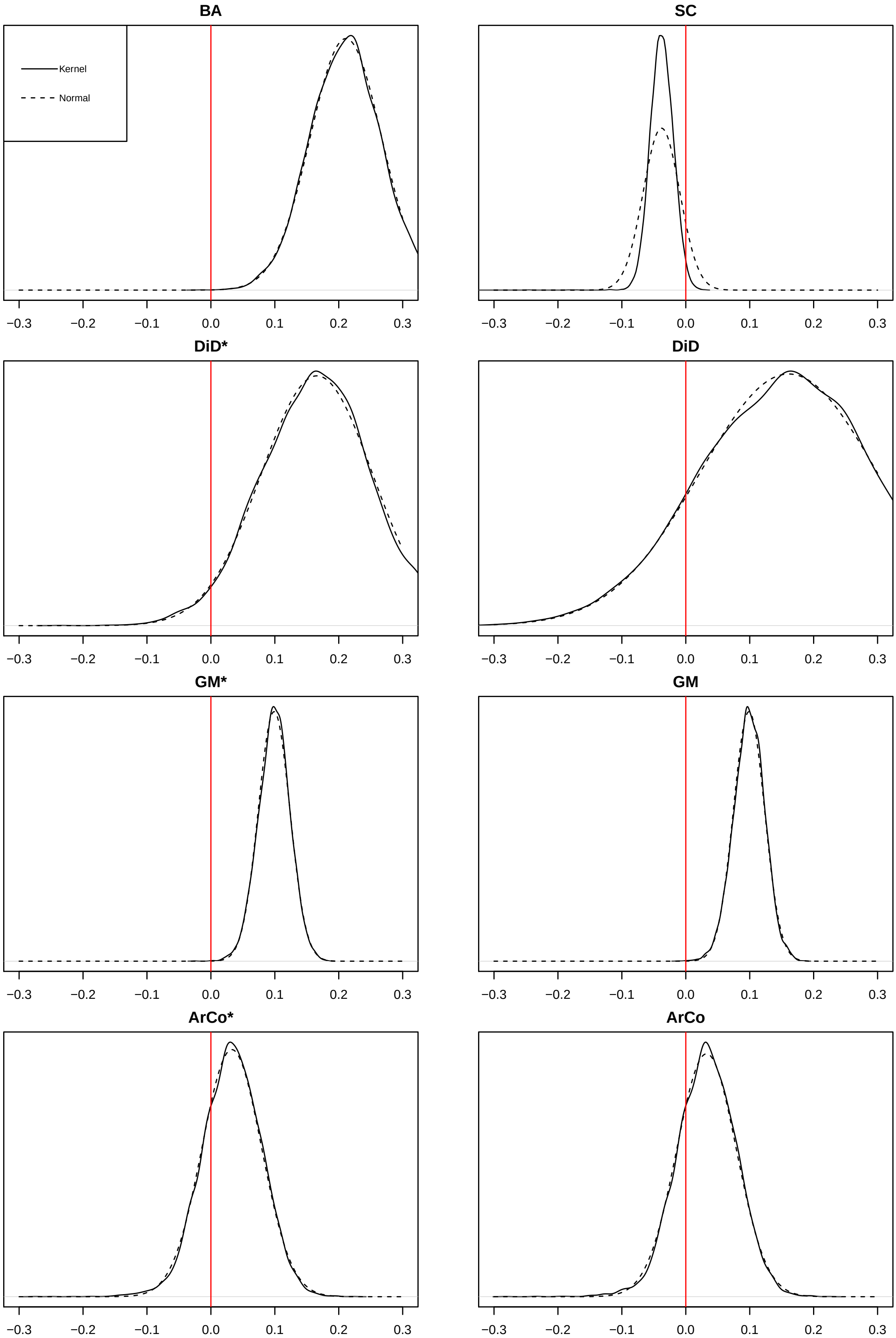


FIGURE S.6. Kernel Density - Estimator Comparison with Idiosyncratic Quadratic Trend



4. Supplementary Theoretical Results

4.1. **Proof of Lemma 1.** From the definition of our estimator in (4) we have:

$$\widehat{\boldsymbol{\Delta}}_T - \boldsymbol{\Delta}_T = \frac{1}{T_2}\sum_{t\geq T_0}\left[\boldsymbol{y}_t - \boldsymbol{\Delta}_T - \widehat{\mathcal{M}}_{t,T_1}\right] = \frac{1}{T_2}\sum_{t\geq T_0}\left[\boldsymbol{y}_t^{(0)} - \widehat{\mathcal{M}}_{t,T_1}\right] = \frac{1}{T_2}\sum_{t\geq T_0}\left[\boldsymbol{\nu}_t - \boldsymbol{\eta}_{t,T_1}\right].$$

After multiplying the last expression by $\sqrt{T}$ we can rewrite it as:

$$\text{(S.3)}\qquad \sqrt{T}\left(\widehat{\boldsymbol{\Delta}}_T - \boldsymbol{\Delta}_T\right) = \underbrace{\frac{\sqrt{T}}{T_2}\sum_{t\geq T_0}\boldsymbol{\nu}_t}_{\equiv \boldsymbol{V}_{2,T}} - \underbrace{\frac{\sqrt{T}}{T_1}\sum_{t\leq T_1}\boldsymbol{\nu}_t}_{\equiv \boldsymbol{V}_{1,T}} - \sqrt{T}\left(\tfrac{1}{T_2}\sum_{t\geq T_0}\boldsymbol{\eta}_{t,T_1} - \tfrac{1}{T_1}\sum_{t\leq T_1}\boldsymbol{\nu}_t\right)$$

By condition (a) in the proposition, the last term in the right hand side converges to zero uniformly in $P \in \mathcal{P}$. Under conditions (b) and (c), each one of the first two terms individually converges in distribution to a Gaussian random variable uniformly in $P \in \mathcal{P}$, which is not enough to ensure that the joint distribution is also Gaussian. However, notice that both $\boldsymbol{V}_{1,T}$ and $\boldsymbol{V}_{2,T}$ are defined with respect to the same random sequence. Hence, not only they are jointly Gaussian but also they are also asymptotically independent since they are summed over non-overlapping intervals:

$$\boldsymbol{V}_T \equiv (\boldsymbol{V}_{1,T}, \boldsymbol{V}_{2,T})' \xrightarrow{d} (\boldsymbol{Z}_1, \boldsymbol{Z}_2)' \equiv \boldsymbol{Z} \sim \mathcal{N}\left\{\boldsymbol{0}, \begin{bmatrix}\lambda_0^{-1}\boldsymbol{\Gamma} & \boldsymbol{0}\\ \boldsymbol{0} & (1-\lambda_0)^{-1}\boldsymbol{\Gamma}\end{bmatrix}\right\},$$

uniformly in $P \in \mathcal{P}$, where $\boldsymbol{\Gamma} \equiv \lim_{T\to\infty}\boldsymbol{\Gamma}_T$.

It follows from Lemma 2(a) that $\boldsymbol{V}_{2,T} - \boldsymbol{V}_{1,T} \xrightarrow{d} \boldsymbol{Z}_2 - \boldsymbol{Z}_1$, uniformly in $P \in \mathcal{P}$. By Lemma 3(a), $\sqrt{T}\left(\widehat{\boldsymbol{\Delta}}_T - \boldsymbol{\Delta}_T\right) \xrightarrow{d} \mathcal{N}\left[\boldsymbol{0}, \frac{\boldsymbol{\Gamma}}{\lambda_0(1-\lambda_0)}\right]$, uniformly in $P \in \mathcal{P}$.

4.2. **Proof of Lemma 2.** The proof is similar to the classical Continuous Mapping Theorem proof but with continuity replaced by uniform continuity. For (a), by the definition of uniform continuity, for any $\epsilon > 0$, there is a $\delta > 0$ such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$ if $d_{\mathcal{D}}(\boldsymbol{x}, \boldsymbol{y}) \leq \delta \Rightarrow d_{\mathcal{E}}\left[\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{g}(\boldsymbol{y})\right] \leq \epsilon$ for some metric $d_{\mathcal{D}}$ and $d_{\mathcal{E}}$, defined on $\mathcal{D}$ and $\mathcal{E}$ respectively. Therefore,

$$\mathbb{P}_P\left\{d_{\mathcal{E}}\left[\boldsymbol{g}(\boldsymbol{X}_T), \boldsymbol{g}(\boldsymbol{X})\right] > \epsilon\right\} \leq \mathbb{P}_P[d_{\mathcal{D}}(\boldsymbol{X}_T, \boldsymbol{X}) > \delta] + \mathbb{P}_P(\boldsymbol{X} \notin \mathcal{C}).$$

The result follows since the first term on the right hand side converges to zero uniformly in $P \in \mathcal{P}$ by assumption and the second is zero for all $P \in \mathcal{P}$ also by assumption.

For (b), given a set $E \in \mathcal{E}$ we have its closure denoted by $\overline{E}$, its complement by E^c and the preimage of $\boldsymbol{g}$ by $\boldsymbol{g}^{-1}(E) \equiv \{\boldsymbol{x} \in \mathcal{D} : \boldsymbol{g}(\boldsymbol{x}) \in E\}$. For closed $F \in \mathcal{E}$ we have that $\boldsymbol{g}^{-1}(F) \subset \overline{\boldsymbol{g}^{-1}(F)} \subset g^{-1}(F) \cup \mathcal{C}^c$ due to the continuity of $\boldsymbol{g}$ on $\mathcal{C}$. Clearly, the event $\{\boldsymbol{g}(\boldsymbol{X}_T) \in F\}$ is the same of $\{\boldsymbol{X}_T \in \boldsymbol{g}^{-1}(F)\}$, then we can write

$$\begin{aligned}\limsup \sup_{P\in\mathcal{P}} \mathbb{P}[\boldsymbol{X}_T \in \boldsymbol{g}^{-1}(F)] &\leq \limsup \sup_{P\in\mathcal{P}} \mathbb{P}[\boldsymbol{X}_T \in \overline{\boldsymbol{g}^{-1}(F)}]\\ &\leq \sup_{P\in\mathcal{P}} \mathbb{P}[\boldsymbol{X} \in \overline{\boldsymbol{g}^{-1}(F)}] \leq \sup_{P\in\mathcal{P}} \mathbb{P}[\boldsymbol{X} \in \boldsymbol{g}^{-1}(F)] + \underbrace{\sup_{P\in\mathcal{P}} \mathbb{P}(\boldsymbol{X} \notin \mathcal{C})}_{=0},\end{aligned}$$

where the second inequality is a consequence of the uniform convergence in distribution of $\boldsymbol{X}_T$ to $\boldsymbol{X}$ and the Portmanteau Lemma (van der Vaart, 1998, Lemma 2.2). The result follows again by the Portmanteau Lemma in the other direction.

4.3. **Proof of Lemma 3.** If $\boldsymbol{X}_T \xrightarrow{p} \boldsymbol{C}$ uniformly in $P \in \mathcal{P}$, then $\boldsymbol{X}_T \xrightarrow{d} \boldsymbol{C}$ uniformly in $P \in \mathcal{P}$ Let $\boldsymbol{Z}_T \equiv (\mathsf{vec}\,\boldsymbol{X}_T, \mathsf{vec}\,\boldsymbol{Y}_T)'$, then $\boldsymbol{Z}_T \xrightarrow{d} \boldsymbol{Z} \equiv (\mathsf{vec}\,\boldsymbol{C}', \mathsf{vec}\,\boldsymbol{Y}')'$ uniformly in $P \in \mathcal{P}$. Now the sum of two real number seen as the mapping $(x, y) \mapsto x + y$ is uniformly continuous. The product mapping $(x, y) \mapsto x.y$ is also uniformly continuous provided that the domain of one of the arguments is bounded. The inverse mapping $x \mapsto 1/x$ can also be made uniformly continuous if the argument is bounded away for zero. Since all the transformations above applied to $\boldsymbol{Z}_T$ are (entrywise) compositions of uniform continuous mapping (hence uniformly continuous), the results follow from Lemma 2(b).

We now state some auxiliary lemmas that will provide bounds in probability used throughout the proof of the main theorem:

4.4. **Proof of Lemma 5.** Since $\widehat{\boldsymbol{\theta}}$ is a minimizer, by definition we have that

$$\frac{1}{T_1}\sum_{t=1}^{T_1}(y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\theta}})^2 + \varsigma\|\widehat{\boldsymbol{\theta}}\|_1 \leq \frac{1}{T_1}\sum_{t=1}^{T_1}(y_t - \boldsymbol{x}_t'\boldsymbol{\theta}_0)^2 + \varsigma\|\boldsymbol{\theta}_0\|_1.$$

Rewriting the expression above using $y_t = \boldsymbol{x}_t'\boldsymbol{\theta}_0 + \nu_t$ yields

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\widehat{\boldsymbol{\Sigma}}}^2 + \varsigma\|\widehat{\boldsymbol{\theta}}\|_1 \leq 2(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\frac{1}{T_1}\sum_{t=1}^{T_1}\boldsymbol{p}_t + \varsigma\|\boldsymbol{\theta}_0\|_1.$$

We can then bound the first term after the inequality using Hölder's inequality as

$$|2(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\frac{1}{T_1}\sum_{t=1}^{T_1}\boldsymbol{p}_t| \leq \|2\frac{1}{T_1}\sum_{t=1}^{T_1}\boldsymbol{p}_t\|_{\max}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1$$

and, on the set $\mathscr{A}(a)$, we have

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\widehat{\boldsymbol{\Sigma}}}^2 + \varsigma\|\widehat{\boldsymbol{\theta}}\|_1 \leq a\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 + \varsigma\|\boldsymbol{\theta}_0\|_1. \tag{S.4}$$

By the triangle inequality we have

$$\|\widehat{\boldsymbol{\theta}}\|_1 = \|\widehat{\boldsymbol{\theta}}[S_0]\|_1 + \|\widehat{\boldsymbol{\theta}}[S_0^c]\|_1 \geq \|\boldsymbol{\theta}_0[S_0]\|_1 - \|\widehat{\boldsymbol{\theta}}[S_0] - \boldsymbol{\theta}_0[S_0]\|_1 + \|\widehat{\boldsymbol{\theta}}[S_0^c]\|_1$$

Using the above expression as a lower bound in the left hand side of (S.4) and

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 = \|\widehat{\boldsymbol{\theta}}[S_0] - \boldsymbol{\theta}_0[S_0]\|_1 + \|\widehat{\boldsymbol{\theta}}[S_0^c]\|_1$$

on the right hand side of (S.4) yields

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\widehat{\boldsymbol{\Sigma}}}^2 + (\varsigma - a)\|\widehat{\boldsymbol{\theta}}[S_0^c]\|_1 \leq (\varsigma + a)\|\widehat{\boldsymbol{\theta}}[S_0] - \boldsymbol{\theta}_0[S_0]\|_1$$

or, equivalently,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\widehat{\boldsymbol{\Sigma}}}^2 + (\varsigma - a)\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 \leq 2\varsigma\|\widehat{\boldsymbol{\theta}}[S_0] - \boldsymbol{\theta}_0[S_0]\|_1. \tag{S.5}$$

Given the compatibility condition for the set S_0 with matrix $\boldsymbol{\Sigma}$, we can bound the left hand side of the above expression as, by definition, there is a $\psi_0 > 0$ such that

$$\|\widehat{\boldsymbol{\theta}}[S_0] - \boldsymbol{\theta}_0[S_0]\|_1 \leq \tfrac{\sqrt{s_0}}{\psi_0}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\boldsymbol{\Sigma}}.$$

Therefore,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}} + (\varsigma - a)\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 \leq 2\varsigma\tfrac{\sqrt{s_0}}{\psi_0}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\boldsymbol{\Sigma}}. \tag{S.6}$$

To relate the norm $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_{\widehat{\boldsymbol{\Sigma}}}$ to the norm $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_{\boldsymbol{\Sigma}}$ we use the set $\mathscr{B}(b) = \{\|\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_{\max} \leq b\}$. Notice that combining (S.5) with the compatibility condition results in

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 \leq \tfrac{2\varsigma}{\varsigma-a}\|\widehat{\boldsymbol{\theta}}[S_0] - \boldsymbol{\theta}_0[S_0]\|_1 \leq \tfrac{2\varsigma}{\varsigma-a}\tfrac{\sqrt{s_0}}{\psi_0}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\boldsymbol{\Sigma}},$$

which, condition on $\mathscr{B}(b)$, yields the following bound

$$\begin{aligned}\left|\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}} - \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\boldsymbol{\Sigma}}\right| &= \left|(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\right| \\ &\leq \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1^2 \\ &\leq b\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1^2 \\ &\leq b\tfrac{4\varsigma^2}{(\varsigma-a)^2}\tfrac{s_0}{\psi_0^2}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\boldsymbol{\Sigma}}\end{aligned}$$

Now we use the previous bound to write

$$\begin{aligned}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\boldsymbol{\Sigma}} &= \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}} + \left(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\boldsymbol{\Sigma}} - \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}}\right) \\ &\leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}} + \left|\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\boldsymbol{\Sigma}} - \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}}\right| \\ &\leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}} + b\tfrac{4\varsigma^2}{(\varsigma-a)^2}\tfrac{s_0}{\psi_0^2}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\boldsymbol{\Sigma}}.\end{aligned}$$

Rearranging terms we get

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\boldsymbol{\Sigma}} \leq \left(1 - \tfrac{4b\varsigma^2}{(\varsigma-a)^2}\tfrac{s_0}{\psi_0^2}\right)^{-1}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}}. \tag{S.7}$$

Using the bound (S.7) in (S.6) yields

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}} + (\varsigma - a)\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 \leq 2\varsigma\tfrac{\sqrt{s_0}}{\psi_0}\left(1 - \tfrac{4b\varsigma^2}{(\varsigma-a)^2}\tfrac{s_0}{\psi_0^2}\right)^{-1/2}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\widehat{\boldsymbol{\Sigma}}}.$$

If we multiply the last expression by 2 and since $\varsigma \geq 2a$ and $32bs_0/\psi_0^2 \leq 1$, the last inequality reduces to

$$2\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}} + \varsigma\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 \leq 4\sqrt{2}\varsigma\tfrac{\sqrt{s_0}}{\psi_0}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_{\widehat{\boldsymbol{\Sigma}}}.$$

Using the fact that $4uv \leq u^2 + 4v^2$ yields

$$2\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}} + \varsigma\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 \leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2_{\widehat{\boldsymbol{\Sigma}}} + 8\varsigma^2\tfrac{s_0}{\psi_0^2}$$

4.5. **Proof of Lemma 6.** From Lemma 5 on $\mathscr{A}(a) \cap \mathscr{B}(b)$, we have that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 \leq \frac{8\varsigma s_0}{\psi_0^2}$, provided that $\varsigma \geq 2a$, $b \leq \frac{\psi_0^2}{32s_0}$ and the compatibility constraint is satisfied for $\boldsymbol{\Sigma} \equiv \mathbb{E}\left(\frac{1}{T_1}\sum_{t=1}^{T_1}\boldsymbol{x}_t\boldsymbol{x}_t'\right)$ with constant $\psi_0 > 0$ (Assumption 2). For convenience set $a = \frac{\varsigma}{2}$ and

$b = \frac{\psi_0^2}{32 s_0}$, then $\mathbb{P}(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 > \frac{8\varsigma s_0}{\psi_0^2}) \leq \mathbb{P}(\mathscr{A}^c \cup \mathscr{B}^c)$ and

$$\begin{aligned}
\mathbb{P}(\mathscr{A}^c \cup \mathscr{B}^c) &\leq \mathbb{P}\left(\left\|\frac{2}{T_1}\sum_{t=1}^{T_1}\boldsymbol{p}_t\right\|_{\max} > \frac{\varsigma}{2}\right) + \mathbb{P}\left(\left\|\frac{1}{T_1}\sum_{t=1}^{T_1}\boldsymbol{M}_t\right\|_{\max} > \frac{\psi_0^2}{32 s_0}\right)\\
&\leq d \max_{1\leq i\leq d}\mathbb{P}\left(\left|\sum_{t=1}^{T_1} p_{i,t}\right| > \frac{\varsigma T_1}{4}\right) + d^2 \max_{1\leq i,j\leq d}\mathbb{P}\left(\left|\sum_{t=1}^{T_1} m_{ij,t}\right| > \frac{\psi_0^2 T_1}{32 s_0}\right)\\
&\leq d\left(\frac{4}{\varsigma T_1}\right)^{\gamma} \max_{1\leq i\leq d}\mathbb{E}\left|\sum_{t=1}^{T_1} p_{i,t}\right|^{\gamma} + d^2\left(\frac{32 s_0}{\psi_0^2 T_1}\right)^{\gamma}\max_{1\leq i,j\leq d}\mathbb{E}\left|\sum_{t=1}^{T_1} m_{ij,t}\right|^{\gamma}\\
&\leq C_1(\gamma)\frac{d}{T_1^{\gamma/2}\varsigma^{\gamma}} + C_2(\gamma,\psi_0)\frac{d^2 s_0^{\gamma}}{T_1^{\gamma/2}}\\
&= C_3(\gamma,\lambda_0)\frac{1}{\kappa^{\gamma}} + o(1),
\end{aligned}$$

where the second inequality follows from the union bound. The third inequality follows from the Markov inequality applied for some $\gamma > 2$. The forth inequality is a consequence of Lemma 3, since (i) by Assumption 3(a) both $\{\boldsymbol{p}_t\}$ and $\{\boldsymbol{M}_t\}$ are strong mixing sequences with exponential decay as measurable functions of $\{\boldsymbol{w}_t\}$; and (ii) by Cauchy-Schwartz inequality combined with Assumption 3(b) we have for some $\delta > 0$:

$$\begin{aligned}
\mathbb{E}|p_{i,t}|^{\gamma+\delta/2} &\leq \left(\mathbb{E}|x_{i,t}|^{2\gamma+\delta}\mathbb{E}|\nu_t|^{2\gamma+\delta}\right)^{\frac{\gamma+\delta/2}{2\gamma+\delta}} \leq c_\gamma, \quad 1\leq i\leq d;\ t\geq 1\\
\mathbb{E}|m_{ij,t} - \mathbb{E}(x_{i,t}x_{j,t})|^{\gamma+\delta/2} &\leq \left(\mathbb{E}|x_{i,t}|^{2\gamma+\delta}\mathbb{E}|x_{j,t}t|^{2\gamma+\delta}\right)^{\frac{\gamma+\delta/2}{2\gamma+\delta}} \leq c_\gamma, \quad 1\leq i,j\leq d;\ t\geq 1.
\end{aligned}$$

Finally, the last equality follows because by Assumption 4(a) we have $\varsigma = \kappa\frac{d^{1/\gamma}}{\sqrt{T}}$ for some $\kappa > 0$ which implies for the first term

$$C_1(\gamma)\frac{d}{T_1^{\gamma/2}\varsigma^{\gamma}} = C_1(\gamma)\left(\frac{T}{T_1}\right)^{\gamma/2}\frac{1}{\kappa^{\gamma}} \leq \frac{C_1(\gamma)\lambda_0^{-\gamma/2}}{\kappa^{\gamma}} + o(1)$$

where we define $C_3(\gamma,\lambda_0) \equiv C_1(\gamma)\lambda_0^{-\gamma/2}$; and the second term $\frac{d^2 s_0^{\gamma}}{T_1^{\gamma/2}} = o(1)$ as a direct consequence of Assumption 4(b) since $s_0\frac{d^{2/\gamma}}{\sqrt{T}} = o(1)$.

4.6. **Proof of Lemma 7.** For a given $\epsilon > 0$, By the union bound, followed by Markov inequality we have:

$$\mathbb{P}\left(\frac{\|\boldsymbol{S}_T\|_{\max}}{d^{1/r}\sqrt{T}} > \epsilon\right) \leq d\max_{1\leq i\leq d}\mathbb{P}\left(\frac{|S_{i,T}|}{d^{1/r}\sqrt{T}} > \epsilon\right) \leq \frac{\max_{1\leq i\leq d}\mathbb{E}|S_{i,T}|^r}{T^{r/2}\epsilon^r},$$

where $S_{i,T} \equiv u_{i,1} + \cdots + u_{i,T}$ is a (scalar) random variable such that by assumption fulfills the condition of Lemma 4 for all $1 \leq i \leq d$, therefore we can write $\max_{1\leq i\leq d}\mathbb{E}|S_{i,T}|^r \leq C_r T^{r/2}$ by Lemma 4, which concludes the proof.

5. Additional Empirical Results

Table S.1. Estimated Effects on food away from home (FAH) Inflation: Placebo Analysis.

	Placebos						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Goiás (GO)	−0.0113	0.1624	0.1606	0.1888	−0.1477	−0.1931	−0.0979
	(0.1811)	(0.1707)	(0.1557)	(0.1642)	(0.2334)	(0.2331)	(0.2032)
Pará (PA)	0.1328	0.2714	0.1933	−0.1419	0.3690	0.3690	0.2789
	(0.2021)	(0.1640)	(0.1708)	(0.2085)	(0.2407)	(0.2407)	(0.2052)
Ceará (CE)	−0.0380	0.2657	0.2223	0.2092	0.1972	0.1972	0.1358
	(0.1484)	(0.1547)	(0.1349)	(0.1368)	(0.1613)	(0.1613)	(0.2506)
Pernambuco (PE)	0.1769	0.1895	0.2698	0.5322	0.1586	0.1586	0.5021
	(0.1949)	(0.1687)	(0.1718)	(0.1741)	(0.2073)	(0.2073)	(0.2174)
Bahia (BA)	0.0125	0.0756	0.1001	0.5707	0.2800	0.2800	0.1737
	(0.2655)	(0.2228)	(0.2433)	(0.3547)	(0.3201)	(0.3201)	(0.2932)
Minas Gerais (MG)	−0.0706	0.1265	0.1417	0.3472	−0.1089	−0.1089	0.0736
	(0.1198)	(0.1007)	(0.1083)	(0.1705)	(0.1560)	(0.1560)	(0.1554)
Rio de Janeiro (RJ)	0.2245	0.2992	0.3126	0.2484	0.1723	0.1723	0.0724
	(0.1165)	(0.1278)	(0.1230)	(0.1245)	(0.1111)	(0.1111)	(0.1300)
Paraná (PR)	0.1409	0.3400	0.2238	0.1441	0.2373	0.2373	0.1732
	(0.2527)	(0.1904)	(0.1582)	(0.2658)	(0.2939)	(0.2939)	(0.2131)
Rio Grande do Sul (RS)	0.4292	0.5422	0.5315	0.4996	0.5325	0.5325	0.4450
	(0.1614)	(0.1653)	(0.1599)	(0.1580)	(0.1627)	(0.1627)	(0.2430)
Inflation	Yes	No	No	No	Yes	Yes	Yes
GDP	No	Yes	No	No	Yes	Yes	Yes
Retail Sales	No	No	Yes	No	No	Yes	Yes
Credit	No	No	No	Yes	No	No	Yes

The table presents the estimated effect of the intervention on the untreated units. Values between parenthesis are the standard error of the estimates.

Table S.2. Estimated Effects on food away from home (FAH) Inflation: The Case without RS.

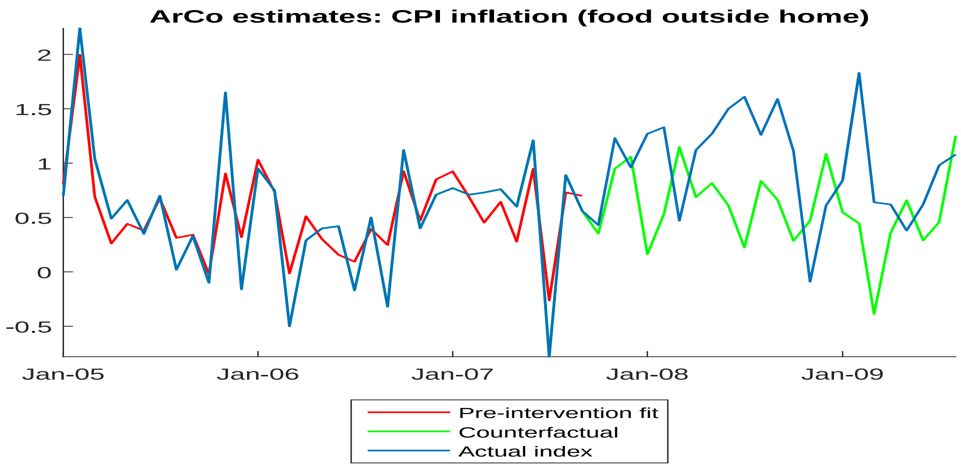
Panel (a): ArCo Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	0.2992 (0.1704)	0.4438 (0.1486)	0.4913 (0.1432)	0.5064 (0.1480)	0.4763 (0.2010)	0.4070 (0.1600)	0.4046 (0.1539)
Inflation	Yes	No	No	No	Yes	Yes	Yes
GDP	No	Yes	No	No	Yes	Yes	Yes
Retail Sales	No	No	Yes	No	No	Yes	Yes
Credit	No	No	No	Yes	No	No	Yes
R-squared	0.6439	0.1213	0.3928	0.1026	0.7960	0.8568	0.8072
Number of regressors	9	8	9	9	17	26	35
Number of relevant regressors	9	3	7	5	14	17	13
Number of observations ($t < T_0$)	33	33	33	33	33	33	33
Number of observations ($t \geq T_0$)	23	23	23	23	23	23	23

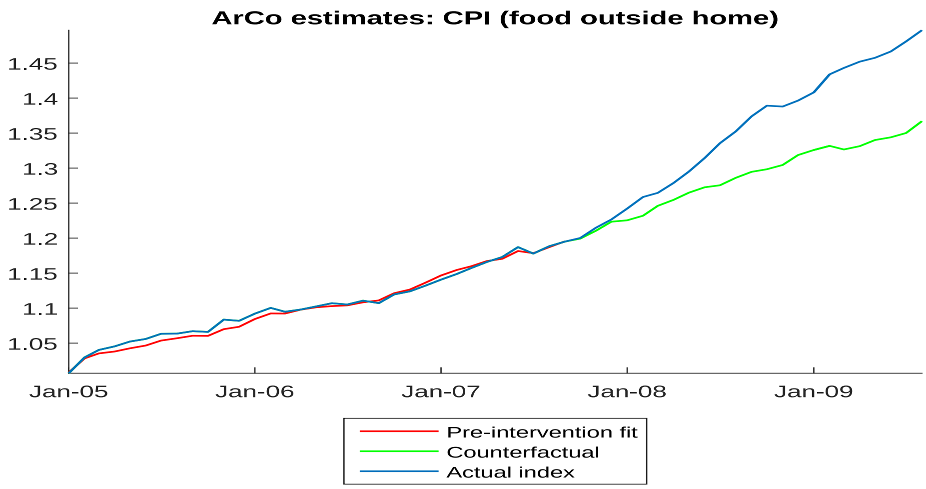
Panel (b): Alternative Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
DiD	0.2524 (0.1466)	0.2407 (0.1456)	0.2494 (0.1467)	0.2412 (0.1556)	0.2387 (0.1457)	0.2520 (0.1466)
GM	0.3694 (0.1234)	0.3788 (0.1243)	0.3595 (0.1246)	0.3775 (0.1227)	0.3660 (0.1228)	–
GDP	Yes	No	No	Yes	Yes	No
Retail Sales	No	Yes	No	Yes	Yes	No
Credit	No	No	Yes	No	Yes	No

The upper panel in the table reports, for different choices of conditioning variables, the estimated average intervention effect after the adoption of the program (*Nota Fiscal Paulista* – NFP). The standard errors are reported between parenthesis. Diagnostic tests do not evidence any residual autocorrelation and the standard errors are computed without any correction. The table also shows the R-squared of the first stage estimation, the number of included regressors in each case as well as the number of selected regressors by the LASSO, and the number of observations before and after the intervention. The lower panel of Table presents some alternative measures of the average intervention effect, namely the Before-and-After (BA), the method proposed by Gobillon and Magnac (2016) (GM) and the difference-in-difference (DiD) estimators.



(a)



(b)

FIGURE S.7. Actual and counterfactual data without RS. The conditioning variables are **inflation**, **DGP growth**, and **retail sales growth**. Panel (a) monthly inflation. Panel (b) accumulated monthly inflation.

References

An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26, 113–172.

Bai, J. (2009): "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

Fernández-Villaverde, J., J. Rubio-Ramírez, T. Sargent, and M. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97, 1021–1026.

Gobillon, L., and T. Magnac (2016): "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 98, 535–551.

Niemi, H. (1979): "On the construction of Wold decomposition for multivariate stationary processes," *Journal of Multivariate Analysis*, 9, 545–559.

van der Vaart, A. (1998): *Asymptotic Statistics*. Cambbridge University Press.