Ricardo Pereira Masini

Contributions to the Econometrics of Counterfactual Analysis

Tese de Doutorado

DEPARTAMENTO DE ECONOMIA

Programa de Pós-Graduação em Economia

Rio de Janeiro
April 2016

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO

Ricardo Pereira Masini

Contributions to the Econometrics of Counterfactual Analysis

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Economia of the Departamento de Economia, PUC–Rio as partial fulfillment of the requirements for the degree of Doutor em Economia.

Advisor : Prof. Marcelo Cunha Medeiros
Co–Advisor: Prof. Carlos Viana de Carvalho

Rio de Janeiro
April 2016

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Ricardo Pereira Masini

Contributions to the Econometrics of Counterfactual Analysis

Thesis presented to the Programa de Pós-graduação em Economia of the Departamento de Economia, PUC–Rio as partial fulfillment of the requirements for the degree of Doutor em Economia. Approved by the following commission:

Prof. Marcelo Cunha Medeiros
Advisor
Departamento de Economia — PUC–Rio

Prof. Carlos Viana de Carvalho
Co–advisor
Departamento de Economia — PUC–Rio

Prof. Pedro Carvalho Loureiro de Souza
Departamento de Economia — PUC–Rio

Prof. Leonardo Rezende
Departamento de Economia — PUC–Rio

Prof. Marcelo Jovita Moreira
Departamento de Economia — FGV–EPGE

Prof. Bruno Ferman
Departamento de Economia — FGV–EESP

Prof. Monica Herz
Social Science Center Coordinator — PUC–Rio

Rio de Janeiro, April 1^{st}, 2016

Ricardo Pereira Masini

Graduated in Aeronautical Engineering at Universidade de São Paulo (2002), MBA with finance major at INSEAD - France/Singapore (2008), MSc in Economics at London School of Economics (2011), and now a PhD in Economics at Pontifícia Universidade Católica do Rio de Janeiro (2016).

Ficha Catalográfica

Masini, Ricardo Pereira

Contributions to the Econometrics of Counterfactual Analysis / Ricardo Pereira Masini; advisor: Marcelo Cunha Medeiros; co–advisor: Carlos Viana de Carvalho. — 2016.

131 f.: il. ; 30 cm

1. Tese (doutorado) — Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia.

Inclui bibliografia.

1. Economia – Teses. 2. Counterfactual analysis. 3. Comparative studies. 4. Treatment effects. 5. Synthetic control. 6. LASSO. 7. Factor models. I. Medeiros, Marcelo Cunha. II. Carvalho, Carlos Viana de. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. IV. Título.

CDD: 330

To the girls of my life,Vanessa, Gabriela & Julia.

Acknowledgment

First of all, I will be eternally indebted to my dear wife Vanessa Figaro for all the support throughout my Ph.D. years. I am specially grateful for her understanding during my absence working on the thesis for several weekends and late nights . Also, I would like to thank her for all the long boring hours proof reading all the versions of the manuscripts until the final version (the mistakes are my own).

I would like to express my sincere gratitude to Marcelo Medeiros, who became not only an advisor to me, but a friend. He always believed in the potential of our research and kept motivating me all along. In particular, his guidance and knowledge helped me out in many situations that seemed a dead end. Last but not least, I could not forget to acknowledge your patience despite my stubbornness in many occasions.

Finally, to my beloved parents who have always encouraged me to pursue my dreams in life and from whom I inherited the curiosity which drives my academic aspirations.

Abstract

Masini, Ricardo Pereira; Medeiros, Marcelo Cunha (adviser); Carvalho, Carlos Viana de (co-adviser). **Contributions to the Econometrics of Counterfactual Analysis**. Rio de Janeiro, 2016. 131p. PhD thesis — Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis is composed of three chapters concerning the econometrics of counterfactual analysis. In the first one, we consider a new, flexible and easy-to-implement methodology to estimate causal effects of an intervention on a single treated unit when no control group is readily available, which we called Artificial Counterfactual (ArCo). We propose a two-step approach where in the first stage a counterfactual is estimated from a large-dimensional set of variables from a pool of untreated units using shrinkage methods, such as the *Least Absolute Shrinkage Operator* (LASSO). In the second stage, we estimate the average intervention effect on a vector of variables, which is consistent and asymptotically normal. Moreover, our results are valid uniformly over a wide class of probability laws. As an empirical illustration of the proposed methodology, we evaluate the effects on inflation of an anti tax evasion program. In the second chapter, we investigate the consequences of applying counterfactual analysis when the data are formed by integrated processes of order one. We find that *without* a cointegration relation (spurious case) the intervention estimator diverges, resulting in the rejection of the hypothesis of no intervention effect regardless of its existence. Whereas, for the case when at least one cointegration relation exists, we have a $\sqrt{T}$-consistent estimator for the intervention effect *albeit* with a non-standard distribution. As a final recommendation we suggest to work in first-differences to avoid spurious results. Finally, in the last chapter we extend the ArCo methodology by considering the estimation of conditional quantile counterfactuals. We derive an asymptotically normal test statistics for the quantile intervention effect including a distributional test. The procedure is then applied in an empirical exercise to investigate the effects on stock returns after a change in corporate governance regime.

Keywords

Counterfactual analysis; Comparative studies; Treatment effects; Synthetic control; LASSO; Factor models;

Resumo

Masini, Ricardo Pereira; Medeiros, Marcelo Cunha (orientador) ; Carvalho, Carlos Viana de (co-orientador). **Contribuições para a Econometria de Análise Contrafactual**. Rio de Janeiro, 2016. 131p. Tese de Doutorado — Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese é composta por três capítulos que abordam a econometria de análise contrafactual. No primeiro capítulo, propomos uma nova metodologia para estimar efeitos causais de uma intervenção que ocorre em apenas uma unidade e não há um grupo de controle disponível. Esta metodologia, a qual chamamos de contrafactual artificial (ArCo na sigla em inglês), consiste em dois estágios: no primeiro um contrafactual é estimado através de conjuntos de alta dimensão de variáveis das unidades não tratadas, usando métodos de regularização como LASSO. No segundo estágio, estimamos o efeito médio da intervenção através de um estimador consistente e assintoticamente normal. Além disso, nossos resultados são válidos uniformemente para um grande classe the distribuições. Como uma ilustração empírica da metodologia proposta, avaliamos o efeito de um programa antievasão fiscal. No segundo capítulo, investigamos as consequências de aplicar análises contrafactuais quando a amostra é gerada por processos integrados de ordem um. Concluímos que, na ausência de uma relação de cointegração (caso espúrio), o estimador da intervenção diverge, resultando na rejeição da hipótese de efeito nulo em ambos os casos, ou seja, com ou sem intervenção. Já no caso onde ao menos uma relação de cointegração exista, obtivemos um estimador consistente, embora, com uma distribuição limite não usual. Como recomendação final, sugerimos trabalhar com os dados em primeira diferença para evitar resultados espúrios sempre que haja possibilidade de processos integrados. Finalmente, no último capítulo, estendemos a metodologia ArCo para o caso de estimação de efeitos quantílicos condicionais. Derivamos uma estatística de teste assintoticamente normal para inferência, além de um teste distribucional. O procedimento é, então, adotado em um exercício empírico com o intuito de investigar os efeitos do retorno de ações após uma mudança do regime de governança corporativa.

Palavras–chave

Análise contrafactual; Estudos comparativos; Efeito de tratamento; Controle sintético; LASSO; Modelo de fatores;

Summary

List of Figures

List of Tables

1
ArCo: An Artificial Counterfactual Approach for High-Dimensional Panel Time-Series Data

1.1
Introduction

We propose a method for counterfactual analysis to evaluate the impact of interventions such as regional policy changes, the start of a new government, or outbreaks of wars, just to name a few possible cases. Our approach is specially useful in situations where there is a single treated unity and no available "controls" and is easy to implement in practice[1]. Furthermore, the method is robust to the presence of confounding effects, such as a global shock. The idea is to construct an artificial counterfactual based on a large-dimensional panel of observed time-series data from a pool of untreated peers.

Causality is a research topic of major interest in empirical Economics. Usually, causal statements with respect to the adoption of a given treatment rely on the construction of counterfactuals based on the outcomes from a similar group of individuals not affected by the treatment. Notwithstanding, definitive cause-and-effect statements are usually hard to formulate given the constraints that economists face in finding sources of exogenous variation. However, in micro-econometrics there has been major advances in the literature and the estimation of treatment effects is part of the toolbox of applied economists; see Angrist e Imbens (1994), Angrist et al. (1996), Heckman e Vytlacil (2005), Conley e Taber (2011), Belloni et al. (2014), Ferman e Pinto (2015), and Belloni et al. (2016).

On the other hand, when there is not a natural control group and there is a single treated unit, which is usually the case when handling aggregate (macro) data, the econometric tools have evolved at a slower pace and much of the work has focused on simulating counterfactuals from structural models. However, in recent years, some authors have proposed new techniques inspired partially by the developments in micro-econometrics that are able, under some assumptions, to estimate counterfactuals with aggregate data; see, for instance, Hsiao et al. (2012) and Pesaran e Smith (2012).

[1]Although the results in the chapter are derived under the assumption of single treated unit, they can be easily generalized to the case of multiple units suffering the treatment.

1.1.1 Contributions of the Chapter

The content of this chapter fits into the literature of counterfactual analysis when a control group is not available and usually only one element suffers the treatment. We propose a two-step approach called the **Artificial Counterfactual (ArCo)** method to estimate the average treatment (intervention) effect on the treated unit. Differently from the cross-section literature, the average is taken over the post-intervention period and not over the treated units. In the first step, we estimate a multivariate model based on a high-dimensional panel of time-series data from a pool of untreated peers, measured before the intervention, and without any stringent assumption about the actual Data Generating Process (DGP). Then, we compute the counterfactual by extrapolating the model with data after the intervention. High-dimensionality is relevant when the number of parameters to be estimated is large compared to the sample size. This can happen either when the number of peers and/or the number of variables for each peer is large or when the sample size is small. We use the *Least Absolute Selection and Shrinkage Operator* (LASSO) proposed by Tibshirani (1996) to estimate the parameters. Nonlinearities can be handled by including in the model some transformations of the explanatory variables, such as polynomials or splines. Furthermore, we propose a test of no intervention effects with a standard limiting distribution which is uniformly valid in a wide class of DGPs, either by imposing any stringent restriction on the model parameters, as it is usually the case when the LASSO is the estimation method, or by modifying the estimator as in Belloni et al. (2016). We also show that it is not necessary to consider two-step extensions of the LASSO, such as the adaptive LASSO of Zou (2006), to handle highly collinear regressors. The method is able to simultaneously test for effects in different variables as well as in multiple moments of a set of variables such as the mean and the variance.

In addition, we accommodate situations when the exact time of the intervention is unknown. This is important in the case of anticipation effects. We also propose a $\mathcal{L}_p$ test inspired by the literature on structural breaks Bai (1997), Bai e Perron (1998) and we show that the asymptotic properties of the method remain unchanged. Finally, we derive tests for the case of multiple interventions as well as for contamination effects among units.

The identification of the average intervention effect relies on the common assumption of independence between the intervention and the treated peers but we allow for heterogeneous, possibly nonlinear, deterministic time trends among units. Our results are derived under asymptotic limits on the time dimension (T). However, we allow the number of peers (n) and the number of observed variables for each peer to grow as a function of T.

A thorough Monte Carlo experiment is conducted in order to evaluate the small sample performance of the ArCo methodology in comparison to well-established alternatives, namely: the before-and-after (BA) estimator, the differences-in-differences (DiD) estimator assuming each peer to be an individual in the control group, the panel factor model of Gobillon e Magnac (2016), hereafter PF-GM, and the synthetic Control method, hereafter SC, of Abadie e Gardeazabal (2003) and Abadie et al. (2010). We show that the bias of the ArCo method is, in general, negligible and much smaller than some of the alternatives. Also, the simulations show that the variance and the mean square error of the ArCo estimator is considerably smaller than the ones from its competitors. Moreover, the test for the null of no intervention effect has good size and power properties.

Finally, we illustrate the methodology by evaluating the impacts on inflation of an anti tax-evasion program implemented in October 2007 in Brazil. The mechanism works by giving tax rebates for consumers who ask for sales receipts. Additionally, the registered sales receipts give the consumer the right to participate in monthly lotteries promoted by the government. Similar initiatives relying on consumer auditing schemes were proposed in the European Union and in China. Under the assumptions that (i) a certain degree of tax evasion was occurring before the intervention, (ii) the sellers has some degree of market power and (iii) the penalty for tax-evasion is large enough to alter the seller behaviour, one is expected to see an upward movement in prices due to an increase in marginal cost. Compared to the counterfactual, we show that the program caused an increase of 10.72% in consumer prices over a period of 23 months. This is an important result as most of the studies in the literature focused only of the effects of such policies on reducing tax evasion but neglected the potential harmful effects on inflation.

1.1.2
Connections to the Literature

Hsiao et al. (2012) considered a two-step method where in their first step the counterfactual for a single treated variable of interest is constructed as a linear combination of a low-dimensional set of observed covariates from preselected elements from a pool of peers. The model is estimated by ordinary least squares using data from the pre-intervention period. Their theoretical results have been derived under the hypothesis of correct specification of a linear panel data model with common factors and no covariates. The selection of the included peers in the linear combination is carried out by information criteria. Recently, several extensions of the above methods have been proposed. Ouyang e Peng (2015) relaxed the linear conditional expectation assumption by introducing a semi-parametric estimator. Du e Zhang (2015) made improvements on the selection mechanism for the constituents of the donors pool.

The ArCo method generalize the above papers in important directions. First, by considering LASSO estimation in the first step we allow for a large number of covariates/peers to be included, not requiring any pre-estimation selection which can bias the estimates. Furthermore, shrinkage estimation is quite appealing when the sample size is small compared to the number of parameters to be estimated. It is important to mention that all our convergence results are uniform on a wide class of probability laws under mild conditions as mentioned previously. Second, all our theoretical results are derived under no stringent assumptions about the DGP, which we assume to be unknown. We do not need to estimate the true conditional expectation. This is a nice feature of the ArCo methodology, as usually models are misspecified. Third, we do not restrict the analysis to a single treated variable. We can, for instance, measure the impact of interventions in several variables of the treated unit simultaneously. We also allow for tests on several moments of the variable of interest. Fourth, we also demonstrate that our methodology can still be applied when the intervention time is unknown. Finally, we develop tests for multiple interventions and contamination effects.

When compared to DiD estimators, the advantages of the ArCo methodology are three-fold. First, we do not need the number of treated units to grow. In fact, the workhorse situation is when there is a single treated unit. The second, and most important difference, is that the ArCo methodology has been developed to situations where the $n-1$ untreated units differ substantially from the treated one and can not form a control group even after conditioning on a set of observables. Finally, the ArCo methodology works even without the parallel trends hypothesis[2].

More recently, Gobillon e Magnac (2016) generalize DiD estimators by estimating a correctly specified linear panel model with strictly exogenous regressors and interactive fixed effects represented as a number of common factors with heterogeneous loadings. Their theoretical results rely on double asymptotics when both T and n go to infinity. The number of untreated units must grow in order to guarantee the consistent estimation of the common factors. The authors allow the common confounding factors to have nonlinear

[2]The first difference can be attenuated in light of the recent results of Conley e Taber (2011) and Ferman e Pinto (2015) who put forward inferential procedures when the number of treated groups is small.

deterministic trends, which is an utmost generalization of the linear parallel trend hypothesis assumed when DiD estimation is considered.

The ArCo method differs from Gobillon e Magnac (2016) in many ways. First, as mentioned before, we assume the DGP to be unknown and we do not need to estimate the common factors. Consistent estimation of factors needs that both the time-series and the cross-section dimensions diverge to infinity and can be severely biased in small samples. The ArCo methodology requires only the time-series dimension to diverge. Furthermore, we do not require the regressors to be strictly exogenous which is an unrealistic assumption in most applications with aggregate (time-series) data. We also allow for heterogeneous nonlinear trends but there is no need to estimate them (either explicitly or via common factors). Finally, as in the DiD case, we do not either require the number of treated units to grow or to have a reliable control group (after conditioning on covariates).

Although, both the ArCo and the SC methods construct a counterfactual as a function of observed variables from a pool of peers, the two approaches have important differences. First, the SC method relies on a convex combination of peers to construct the counterfactual which, as pointed out by Ferman e Pinto (2016), biases the estimator. This is clearly evidenced in our simulation experiment. The ArCo solution is a general, possibly nonlinear, function. Even in the case of linearity, the method does not impose any restriction on the parameters. For example, the restriction that weights in the SC methods are all positive seems a bit too strong. Furthermore, the weights in the SC method are usually estimated using time averages of the observed variables for each peer. Therefore, all the time-series dynamics is removed and the weights are determined in a pure cross-sectional setting. In some applications of the SC method, the number of observations to estimate the weights is much lower than the number of parameters to be determined. For example, in Abadie e Gardeazabal (2003) the authors have 13 observations to estimate 16 parameters[3]. A similar issue also appears in Abadie et al. (2010), Abadie et al. (2014). In addition, the SC method was designed to evaluate the effects of the intervention on a single variable. In order to evaluate the effects in a vector of variables, the method has to be applied several times. The ArCo methodology can be directly applied to a vector of variables of interest. In addition, there is no formal inferential procedure for hypothesis testing in the SC method, whereas in the ArCo methodology, a simple, uniformly valid and standard test can be applied. Finally, as discussed

[3]In these cases the estimation is only possible due to the imposed restrictions, which can be seen as a sort of shrinkage.

in Ferman et al. (2016), the SC method does not provide any guidance on how to select the variables which determine the optimal weights.

With respect to the methodology by Pesaran e Smith (2012), the major difference is that the authors construct the counterfactual based on variables that belong to the treated unit and they do not rely on a pool of untreated peers. Their key assumption is that a subset of variables of the treated unit is invariant to the intervention. Although, in some specific cases this could be a reasonable hypothesis, in a general framework this is clearly restrictive.

Recently, Angrist et al. (2013) propose a semiparametric method to evaluate the effects of monetary policy based on the so called policy propensity score. Similar to Pesaran e Smith (2012), the authors only rely on information on the treated unit and no donor pool is available. As before, this is a major difference from our approach. Furthermore, their methodology seems to be particularly appealing to monetary economics but hard to be applied in other settings without major modifications.

It is important to compare the ArCo methodology with the work of Belloni et al. (2014) and Belloni et al. (2016). Both papers consider the estimation of intervention effects in large dimensions. First, Belloni et al. (2014) consider a pure cross-sectional setting where the intervention is correlated to a large set of regressors and the approach is to consider an instrumental variable estimator to recover the intervention effect, as there is no control group available. In the ArCo framework, on the other hand, the intervention is assumed to be exogenous with respect to the peers. Notwithstanding, the intervention may not be (and probably is not) independent of variables belonging to the treated unit. This key assumption enables us to construct honest confidence bands by using the LASSO in the first step to estimate the conditional model. Belloni et al. (2016) proposed a general and flexible extension of the DiD approach for program evaluation in high dimensions. They provide efficient estimators and honest confidence bands for a large number of treatment effects. However, they do not consider the case where there is no control group available. Finally, it is not clear how to apply their methods to aggregate (macro) data.

1.1.3 Potential Applications

There has been a large body of studies that require the estimation of intervention effects with no group of controls.

Measuring the impacts of regional policies is a potential application. For example, Hsiao et al. (2012) measure the impact of economic and political integration of Hong Kong with mainland China on Hong Kong's economy whereas Abadie et al. (2014) estimate spillovers of the 1990 German reunification in West Germany. Pesaran et al. (2007) used the Global Vector Autoregressive (GVAR) framework of Pesaran et al. (2004) and Dees et al. (2007) to study the effects of the launching of the Euro. Gobillon e Magnac (2016) considered the impact on unemployment of a new police implemented in France in the 1990s. The effects of trade agreements and liberalization have been discussed in Billmeier e Nannicini (2013), and Jordan et al. (2014). The rise of a new government or new political regime are, as well, a relevant "intervention" to be studied. For example, Grier e Maynard (2013) considered the economic impacts of the Chavez era.

Other potential applications are new regulation on housing prices as in Bai et al. (2014) and Du e Zhang (2015), new labor laws as considered in Du et al. (2013), and macroeconomic effects of economic stimulus programs Ouyang e Peng (2015). The effects of different monetary policies have been discussed in Pesaran e Smith (2012) and Angrist et al. (2013). Estimating the economic consequences of natural disasters, as in Belasen e Polachek (2008), Cavallo et al. (2013), Fujiki e Hsiao (2015), and Caruso e Miller (2015), is also a promising area of research.

The effects of market regulation or the introduction of new financial instruments on the risk and returns of stock markets has been considered in Chen et al. (2013) and Xie e Mo (2013). Testing the intervention effects in multiple moments of the data can be of special interest in Finance, where the goal could be the effects of different corporate governance policies in the returns and risk of the firms Johnson et al. (2000).

This chapter is organized as follows. In Section 1.2 we present the ArCo method and discuss the conditional model used in the first step of the methodology. In Section 1.3 we derive the asymptotic properties of the ArCo estimator and state our main result. Sub-section 1.3.3 deals with the test for the null hypothesis of no causal effect. Extensions for unknown intervention time, multiple interventions and possible contamination effects are described in Section 1.4. In Section 1.5 we discuss some potential sources of bias in the ArCo method. A detailed Monte Carlo study is conducted in Section 1.6. Section 1.7 deals with the empirical exercise. Finally, Section 1.8 concludes. Tables, figures and all proofs are relegated to the Appendix.

1.2 The Artificial Counterfactual Estimator

1.2.1 Setup

Suppose we have n units (countries, states, municipalities, firms, etc) indexed by $i = 1, \ldots, n$. For each unit and for every time period $t = 1, \ldots, T$, we observe a realization of $\boldsymbol{z}_{it} = (z_{it}^1, \ldots, z_{it}^{q_i})' \in \mathbb{R}^{q_i}$, $q_i \geq 1$. Furthermore, assume that an intervention took place in unit $i = 1$, and only in unit 1, at time $T_0 = \lfloor \lambda_0 T \rfloor$, where $\lambda_0 \in (0, 1)$.

Let $\mathcal{D}_t$ be a binary variable flagging the periods when the intervention was in place. We can express the observable variables of unit 1 as

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

where $\mathcal{D}_t = I(t \geq T_0)$, $I(A)$ is an indicator function that equals 1 if the event A is true, and $\boldsymbol{z}_{1t}^{(1)}$ denotes the outcome when the unit 1 is exposed to the intervention and $\boldsymbol{z}_{1t}^{(0)}$ is the potential outcome of unit 1 when there is no intervention.

We are ultimately concerned with testing hypothesis on the effects of the intervention on unit 1 for $t \geq T_0$. In particular, we consider interventions of the form

$$\boldsymbol{y}_t^{(1)} = \begin{cases} \boldsymbol{y}_t^{(0)}, & t = 1, \ldots, T_0 - 1, \\ \boldsymbol{\delta}_t + \boldsymbol{y}_t^{(0)}, & t = T_0 \ldots, T, \end{cases} \tag{1-1}$$

where $\boldsymbol{y}_t^{(j)} \equiv \boldsymbol{h}(\boldsymbol{z}_{1t}^{(j)})$ for $j \in \{0, 1\}$, $\boldsymbol{h} : \mathbb{R}^{q_1} \mapsto \mathbb{R}^q$ is a measurable function of $\boldsymbol{z}_{1t}$ that will be defined latter, and $\{\boldsymbol{\delta}_t\}_{t=T_0}^T$ is a deterministic sequence. Due to the flexibility of the mapping $\boldsymbol{h}(\cdot)$, interventions modeled as (1-1) are quite general. It includes, for instance, interventions affecting the mean, variance, covariances or any combination of moments of $\boldsymbol{z}_{1t}$. The null hypothesis of interest is

$$\mathcal{H}_0 : \boldsymbol{\Delta}_T = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \boldsymbol{\delta}_t = \boldsymbol{0}. \tag{1-2}$$

The quantity $\boldsymbol{\Delta}_T$ in (3-1) is similar to the traditional *average treatment effect on the treated* (ATET) vastly discussed in the literature[4]. Furthermore, the null hypothesis (3-1) encompasses the case where the intervention is a sequence $\{\boldsymbol{\delta}_t\}_{t=T_0}^T$ under the alternative, which obviously is a special case of uniform treatments by setting $\boldsymbol{\delta}_t = \boldsymbol{\delta}, \forall t \geq T_0$.

The particular choice of the transformation $\boldsymbol{h}(\cdot)$ will depend on which moments of the data the econometrician is interested in testing for effects of the intervention. In other words, the goal will be to test for a break in a set of unconditional moments of the data and check if this break is solely due to the

[4]However, as pointed out in the Introduction, the average is taken over time periods and not over cross-section elements

intervention or has other (global) causes (confounding effects). Typical choices for $\boldsymbol{h}(\cdot)$ are presented as examples below.

Example 1.1 *For the univariate case* ($q_1 = 1$)*, we can use the identity function* $h(a) = a$ *for testing changes in the mean. In fact, provided that the* p*-th moment of the data is finite, we can use* $h(a) = a^p$ *to test any change in the* p*-th unconditional moment.*

Example 1.2 *In the multivariate case* ($q_1 > 1$) *we can consider*

$$\boldsymbol{h}(\boldsymbol{z}_{1t}) = \begin{cases} \boldsymbol{z}_{1t} & \textit{for testing changes in the mean,} \\ \mathsf{vech}\ (\boldsymbol{z}_{1t}, \boldsymbol{z}'_{1t}) & \textit{for testing changes in the second moments.} \end{cases}$$

Example 1.3 *We can also conduct joint tests by combining the different choices for* $\boldsymbol{h}$*. For example, for testing simultaneously for a change in the mean and variance we can set* $\boldsymbol{h}(a) = (a, a^2)'$*. In the multivariate case we can set* $\boldsymbol{y}_t = \mathsf{diag}\ (\boldsymbol{z}_{1t}, \boldsymbol{z}'_{1t})$*.*

Set $\boldsymbol{y}_t = \mathcal{D}_t \boldsymbol{y}_t^{(1)} + (1 - \mathcal{D}_t)\boldsymbol{y}_{yt}^{(0)}$. The exact dimension of $\boldsymbol{y}_t$ depends on the chosen $\boldsymbol{h}(\cdot)$. However, regardless of the choice of $\boldsymbol{h}(\cdot)$, we will consider, without loss of generality, that $\boldsymbol{y}_t \in \mathcal{Y} \subset \mathbb{R}^q$, $q > 0$, and that we have a sample $\{\boldsymbol{y}_t\}_{t=1}^{T}$, being the first $T_0 - 1$ observations before the intervention and the $T - T_0 + 1$ remaining observations after the intervention.

Clearly we do not observe $\boldsymbol{y}_t^{(0)}$ after $T_0 - 1$. We call $\boldsymbol{y}_t^{(0)}$ the *counterfactual*, i.e., what would $\boldsymbol{y}_t$ have been like had there been no intervention (potential outcome). In order to construct the counterfactual, let $\boldsymbol{z}_{0t} = (\boldsymbol{z}'_{2t}, \ldots, \boldsymbol{z}'_{nt})'$ and $\boldsymbol{Z}_{0t} = \left(\boldsymbol{z}'_{0t}, \ldots, \boldsymbol{z}'_{0t-p}\right)'$ be the collection of all the untreated units' observables up to an arbitrary lag $p \geq 0$. The exact dimension of $\boldsymbol{Z}_{0t}$ depends upon the number of peers $(n-1)$, the number of variables per peer, $q_i, i = 2, \ldots, n$, and the choice of p. However, without loss of generality, we assume that $\boldsymbol{Z}_{0t} \in \mathcal{Z}_0 \subseteq \mathbb{R}^d$, $d > 0$.

Consider the following model

$$\boldsymbol{y}_t^{(0)} = \mathcal{M}_t + \boldsymbol{\nu}_t,\ t = 1, \ldots, T, \tag{1-3}$$

where $\mathcal{M}_t \equiv \mathcal{M}(\boldsymbol{Z}_{0t})$, $\mathcal{M} : \mathcal{Z}_0 \to \mathcal{Y}$ is a measurable mapping, and $\mathbb{E}(\boldsymbol{\nu}_t) = \boldsymbol{0}$.[5]

Set $T_1 \equiv T_0 - 1$ and $T_2 \equiv T - T_0 + 1$ as the number of observations before and after the intervention, respectively. One can estimate the model above using the first T_1 observations since, in that case, $\boldsymbol{y}_t^{(0)} = \boldsymbol{y}_t$. Then, the estimate $\widehat{\mathcal{M}}_{t,T_1} \equiv \widehat{\mathcal{M}}_{T_1}(\boldsymbol{Z}_{0t})$ can be used to construct the estimated counterfactual as:

[5]Which can be ensured by either including a constant in the model $\mathcal{M}$ or by centering the variables in a linear specification.

$$\widehat{\boldsymbol{y}}_t^{(0)} = \begin{cases} \boldsymbol{y}_t^{(0)}, & t = 1, \ldots, T_0 - 1, \\ \widehat{\mathcal{M}}_{t,T_1}, & t = T_0, \ldots, T. \end{cases} \tag{1-4}$$

Consequently, we can define:

Definition 1.1 *The Artificial Counterfactual (ArCo) estimator is*

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

where $\widehat{\boldsymbol{\delta}}_t \equiv \boldsymbol{y}_t - \widehat{\boldsymbol{y}}_t^{(0)}$, *for* $t = T_0, \ldots, T$.

Therefore, the ArCo is a two-stage estimator where in the first stage we choose and estimate the model $\mathcal{M}$ using the pre-intervention sample and in the second we compute $\widehat{\boldsymbol{\Delta}}_T$ defined by (1-5). At this point the following remarks are in order.

Remark 1.1 *The ArCo estimator in (1-5) is defined under the assumption that λ_0 (consequently T_0) is known. However, in some cases the exact time of the intervention might be unknown due to, for example, anticipation effects. On the other hand, the effects of a policy change may take some time to be noticed. Although the main results are derived under the assumption of known λ_0, we later show they are still valid when λ_0 is unknown.*

1.2.2
A Key Assumption and Motivations

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

Assumption 1.1 $\boldsymbol{z}_{0t} \perp\!\!\!\perp \mathcal{D}_s$, *for all* t, s.

Roughly speaking the assumption above is sufficient for the peers to be unaffected by intervention on the unit of interest. Independence is actually stronger than necessary. Technically, what is necessary for the results is the mean independence of the chosen model as in $\mathbb{E}(\mathcal{M}_t|\mathcal{D}_t) = \mathbb{E}(\mathcal{M}_t)$. Nevertheless, the latter is implied by Assumption 1.1 regardless of the choice of $\mathcal{M}$. It is worth mentioning that since we allow $\mathbb{E}(\boldsymbol{z}_{1t}|\mathcal{D}_t) \neq \mathbb{E}(\boldsymbol{z}_{1t})$ we might have some sort of selection on observables and/or non-observables belonging to the treated unit. Of course, selection on features of the untreated units is ruled out by Assumption 1.1.

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:

$$\boldsymbol{z}_{it}^{(0)} = \boldsymbol{\mu}_i + \boldsymbol{\Psi}_{\infty,i}(L)\boldsymbol{\varepsilon}_{it}, \quad i = 1,\ldots,n;\ t \geq 1 \tag{1-6}$$

$$\boldsymbol{\varepsilon}_{it} = \boldsymbol{\Lambda}_i \boldsymbol{f}_t + \boldsymbol{\eta}_{it}, \tag{1-7}$$

where $\boldsymbol{f}_t \in \mathbb{R}^f$ is a vector of common unobserved factors such that $\sup_t \mathbb{E}(\boldsymbol{f}_t\boldsymbol{f}_t') < \infty$ 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 (2-6) 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 (2-6) Fernández-Villaverde et al. (2007), An e Schorfheide (2007). Gobillon e 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 $\boldsymbol{x}_t$.

There are at least two major advantages of applying the ArCo estimator instead of just computing a simple difference in the mean of $\boldsymbol{y}_t$ before and after the intervention as a estimator for the intervention effect. The first is an efficiency argument. Note that the "before and after" estimator defined as $\widehat{\boldsymbol{\Delta}}_T^{BA} \equiv \frac{1}{T-T_0+1}\sum_{t=T_0}^{T} \boldsymbol{y}_t - \frac{1}{T_0-1}\sum_{t=1}^{T_0-1} \boldsymbol{y}_t$ is a particular case of our estimator when you have "bad peers", in the sense they are uncorrelated with the unit of interest. In this case, $\mathcal{M}(\cdot) =$ constant and $\widehat{\boldsymbol{\Delta}}_T = \widehat{\boldsymbol{\Delta}}_T^{BA}$. In fact, the additional information provided by the peers helps to reduce the variance of the ArCo estimator.

The second, and more important, argument in favor of the ArCo method is related to its capability of isolate the intervention of interest from aggregate shocks. When attempting to measure the effect of a particular intervention we are usually in a scenario that other aggregate shocks took place at the same time. The ability to disentangle these two effects is vital if one intends to provide a meaningful estimation of the intervention effect. A simple thought experiment illustrates the point: suppose *all* units at time T_0 are hit by a (aggregate) shock that changes all the means by the same amount. If we apply the BA estimator we will eventually encounter this mean break and would erroneously attribute it to the intervention of interest[6]. On the other hand, if we use the ArCo approach, since all the units have changed equally, the estimated effect will be insignificant.

Finally, it is important to stress that the validity of the ArCo procedure does not rely on the traditional parallel trend assumption such as the one usually considered in DiD techniques nor does it assume the trend to be the same for all the units at a given time, as for instance in the SC framework. The necessary assumption for our methodology to work properly is some sort of combination of peers (model $\mathcal{M}$) that can generate an artificial counterfactual whose difference from the real counterfactual is well behaved (in the sense of admitting a Law of Large Numbers and Central Limit theorems). This is usually possible with deterministic trends that do not dominate the stationary stochastic component asymptotically as well as when there is some common structure among units.

1.3 Asymptotic Properties and Inference

1.3.1 Choice of the Pre-intervention Model and a General Result

The first stage of the ArCo method requires the choice of the model $\mathcal{M}$. One should aim for a model that captures most of the information from the available peers. Once the choice is made, the model must be estimated using the pre-intervention sample.

It is important to recognise that we do not assume that the model choice is actually the true model. We can consider that $\boldsymbol{z}_{it}$ is generated by a DGP such as (2-6) irrespective of the choice of $\mathcal{M}$. Ideally, in the mean square error sense, we would like to set $\mathcal{M}$ as the conditional expectation model $\boldsymbol{m}(\boldsymbol{a}) = \mathbb{E}(\boldsymbol{y}_t|\boldsymbol{Z}_{0t} = \boldsymbol{a})$.

[6]Unless the intervention of interest is the aggregate shock but in that case we have invalid peers since they were treated.

Motivated by the fact the dimension of $\boldsymbol{Z}_{0t}$ can grow quite fast in any simple application (by either including more peers, more covariates, or by simply considering more lags) we propose a fully parametric specification in order to approximate $\boldsymbol{m}(\cdot)$ as opposed to try to estimate it non-parametrically. In particular, we approximate it by a linear model (q linear models to be precise) of some transformation of $\boldsymbol{Z}_{0t}$. Consequently, the model is linear in $\boldsymbol{x}_t = \boldsymbol{h}_x(\boldsymbol{Z}_{0t})$, where in $\boldsymbol{x}_t$ we include a constant term. In particular, $\boldsymbol{h}_x$ could be a dictionary of functions such as polynomials, splines, interactions, dummies or any another family of elementary transformations the $\boldsymbol{Z}_{0t}$, in the spirit of sieve estimation Chen (2007). The same approach has been adopted in Belloni et al. (2014) and Belloni et al. (2016).

Hence, $\mathcal{M}_t = \mathsf{diag}\,(\boldsymbol{\theta}'_{0,1}, \ldots, \boldsymbol{\theta}'_{0,q})\boldsymbol{x}_t$, where both $\boldsymbol{x}_t$ and $\boldsymbol{\theta}_{0,j}$, $j = 1, \ldots, q$, are d-dimensional vectors for $j = 1, \ldots, q$. We allow d to be a function of T. Hence, $\boldsymbol{x}_t$ and $\boldsymbol{\theta}_{0,j}$ depend on T but the subscript T will be omitted in what follows. Set $\boldsymbol{r}_t \equiv \boldsymbol{m}_t - \mathcal{M}_t$ as the approximation error and $\boldsymbol{\varepsilon}_t \equiv \boldsymbol{y}_t - \boldsymbol{m}_t$ as the projection error. We can write the model as in (2-3), with $\boldsymbol{\nu}_t = \boldsymbol{r}_t + \boldsymbol{\varepsilon}_t$. The model is then comprised of q linear regressions:

$$y_{jt}^{(0)} = \boldsymbol{x}_t'\boldsymbol{\theta}_{0,j} + \nu_{jt}, \quad j = 1, \ldots, q, \tag{1-8}$$

where $\boldsymbol{\theta}_{0,j}$ are the best (in the MSE sense) linear projection parameters which are properly identified as long as we rule out multicollinearity among $\boldsymbol{x}_t$ (Assumption 1.2).

We consider the sample (in the absence of intervention) as a single realization of the random process $\{\boldsymbol{z}_t^{(0)}\}_{t=1}^T$ defined on a common measurable space $(\Omega, \mathcal{F})$ with a probability law (joint distribution) $P_T \in \mathcal{P}_T$, where $\mathcal{P}_T$ is (for now) an arbitrary class of probability laws. The subscript T makes it explicit the dependence of the joint distribution on the sample size T, but we omit it in what follows. We write $\mathbb{P}_P$ and $\mathbb{E}_P$ to denote the probability and expectation with respect to the probability law $P \in \mathcal{P}$, respectively.

We establish the asymptotic properties of the ArCo estimator by considering the whole sample increasing, while the proportion between the pre-intervention to the post-intervention sample size is constant. The limits of the summations are from 1 to T whenever left unspecified. Recall that $T_1 \equiv T_0 - 1$ and $T_2 \equiv T - T_0 + 1$ are the number of pre and post intervention periods, respectively and $T_0 = \lfloor \lambda_0 T \rfloor$. Hence, for fixed $\lambda_0 \in (0, 1)$ we have $T_0 \equiv T_0(T)$. Consequently, $T_1 \equiv T_1(T)$ and $T_2 \equiv T_2(T)$. All the asymptotics are taken as $T \to \infty$. We denote convergence in probability and in distribution by "$\xrightarrow{p}$" and "$\xrightarrow{d}$", respectively.

First, we state a general result under very high level assumptions which all the other subsequent results rely on. Let $\widehat{\mathcal{M}}_{t,T_1} = (\boldsymbol{x}_t'\widehat{\boldsymbol{\theta}}_{1,T_1}, \dots, \boldsymbol{x}_t'\widehat{\boldsymbol{\theta}}_{q,T_1})'$, for $t \geq T_0$, where $\widehat{\boldsymbol{\theta}}_{j,T_1}$, $j = 1, \dots, q$, is estimated with only the first T_1 pre-intervention observations, and define $\boldsymbol{\eta}_{t,T_1} \equiv \widehat{\mathcal{M}}_{t,T_1} - \mathcal{M}_t$, $t \geq T_0$.

Proposition 1.2 *Under Assumption 1.1, consider further that, uniformly in $P \in \mathcal{P}$ (an arbitrary class of probability laws):*

(a) $\sqrt{T}\left(\frac{1}{T_2}\sum_{t\geq T_0}\boldsymbol{\eta}_{t,T_1} - \frac{1}{T_1}\sum_{t\leq T_1}\boldsymbol{\nu}_t\right) \xrightarrow{p} \mathbf{0}$

(b) $\frac{1}{\sqrt{T_1}}\boldsymbol{\Gamma}_{T_1}^{-1/2}\sum_{t\leq T_1}\boldsymbol{\nu}_t \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q)$, *where* $\boldsymbol{\Gamma}_{T_1} = \mathbb{E}_P\left[\frac{1}{T_1}(\sum_{t\leq T_1}\boldsymbol{\nu}_t)(\sum_{t\leq T_1}\boldsymbol{\nu}_t')\right]$.

(c) $\frac{1}{\sqrt{T_2}}\boldsymbol{\Gamma}_{T_2}^{-1/2}\sum_{t\geq T_0}\boldsymbol{\nu}_t \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q)$, *where* $\boldsymbol{\Gamma}_{T_2} = \mathbb{E}_P\left[\frac{1}{T_2}(\sum_{t\geq T_0}\boldsymbol{\nu}_t)(\sum_{t\geq T_0}\boldsymbol{\nu}_t')\right]$.

Then, uniformly in $P \in \mathcal{P}$, $\sqrt{T}\boldsymbol{\Omega}_T^{-1/2}\left(\widehat{\boldsymbol{\Delta}}_T - \boldsymbol{\Delta}_T\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q)$, where $\mathcal{N}(\cdot,\cdot)$ is the multivariate normal distribution and $\boldsymbol{\Omega}_T \equiv \frac{\boldsymbol{\Gamma}_{T_1}}{T_1/T} + \frac{\boldsymbol{\Gamma}_{T_2}}{T_2/T}$.

Condition (a) above sets a limit for the estimation error to be asymptotic negligible, ensuring the $\sqrt{T}$ rate of convergence of the estimator. Under condition (a) we can write:

$$\widehat{\boldsymbol{\Delta}}_T - \boldsymbol{\Delta}_T = \frac{1}{T_2}\sum_{t\geq T_0}\boldsymbol{\nu}_t - \frac{1}{T_1}\sum_{t\leq T_1}\boldsymbol{\nu}_t + o_p(T^{-1/2}).$$

Finally, conditions (b) and (c) ensure the asymptotic normality of the terms above after appropriate normalization. From the asymptotic variance $\boldsymbol{\Omega}_T$ it becomes evident that an intervention at the middle of the sample, $\lambda_0 = 0.5$, is desirable when $\lim_{T\to\infty}\boldsymbol{\Gamma}_{T_1} = \lim_{T\to\infty}\boldsymbol{\Gamma}_{T_2} \equiv \boldsymbol{\Gamma}$, which happens for instance when $\{\boldsymbol{\nu}_t\}$ is a stationary process. In this case, $\lim_{T\to\infty}\boldsymbol{\Omega}_T = \boldsymbol{\Gamma}/\lambda_0(1-\lambda_0)$.

Recall that if $\mathcal{M} = \boldsymbol{\alpha}_0$, the estimator is equivalent to the BA estimator. Therefore, one advantage of the ArCo is to provide a systematic way to extract as most information as possible from the peers in order to reduce the asymptotic variance of the prediction error. We can make more explicit the peers' contribution in reducing the asymptotic variance of the ArCo estimator by the following matrix inequality (in term of positive definiteness)

$$\mathbf{0} \leq \lim_{T\to\infty}\boldsymbol{\Omega}_T \equiv \boldsymbol{\Omega} \leq \lim_{T\to\infty} T\mathbb{V}\left(\frac{1}{T_2}\sum_{t\geq T_0}\boldsymbol{y}_t^{(0)} - \frac{1}{T_1}\sum_{t\leq T_1}\boldsymbol{y}_t^{(0)}\right) \equiv \widetilde{\boldsymbol{\Omega}},$$

where $\mathbb{V}$ is the variance operator defined for any random vector $\boldsymbol{v}$ as $\mathbb{V}(\boldsymbol{v}) = \mathbb{E}(\boldsymbol{v}\boldsymbol{v}') - \mathbb{E}(\boldsymbol{v})\mathbb{E}(\boldsymbol{v}')$.

The upper bound $\widetilde{\boldsymbol{\Omega}}$ is the long run variance of the variables of the unit of interest (unit 1) weighted by the intervention fraction time λ_0. As a consequence, our estimator variance for any given λ_0, lies in between those two polar cases. One polar case is when there is a perfect artificial counterfactual and the other one is when the peers contribute with no information. Thus, the peer's contribution in reducing the ArCo estimator asymptotic variance could be represented by a R^2-type statistic measuring the "ratio" between the explained long-run variance $\mathbf{\Omega}$ to the total long-run variance $\widetilde{\mathbf{\Omega}}$.

1.3.2 Assumptions and Asymptotic Theory in High-Dimensions

The dimension d of $\boldsymbol{x}_t$ can be potentially very large, even larger than the sample size T, whenever the number of peers and/or the number of variables per peer is large. In these cases it is standard to allow d, and consequently $\boldsymbol{\theta}_j$, $j = 1\dots,q$, to be function of the sample size, such that $d \equiv d_T$ and $\boldsymbol{\theta}_j = \boldsymbol{\theta}_{j,T}$. In order to make estimation feasible, regularization (shrinkage) is usually adopted, which is justified by some sparsity assumption on the vector $\boldsymbol{\theta}_{0,j}$, $j = 1\dots,q$, in the sense that only a small portion of its entries are different from zero.

We propose the estimation of (1-8), equation by equation, by the LASSO approach and we allow that dimension $d > T$ to grow faster than the sample size[7]. Also, since each equation in the model is the same, we drop the subscript j from now on to focus on a generic equation. Therefore, we estimate $\boldsymbol{\theta}_0$ via

$$\widehat{\boldsymbol{\theta}} = \arg\min \left\{ \frac{1}{T_0 - 1} \sum_{t<T_0} (y_t - \boldsymbol{x}_t'\boldsymbol{\theta})^2 + \varsigma \|\boldsymbol{\theta}\|_1 \right\}, \tag{1-9}$$

where $\varsigma > 0$ is a penalty term and $\|\cdot\|_1$ denotes the ℓ_1 norm.

Let $\boldsymbol{\theta}[A]$ denote the vector of parameters indexed by A and S_0 the index set of the non-zero (relevant) parameters $S_0 = \{i : \theta_{0,i} \neq 0\}$ with cardinality s_0. We consider the following set of assumptions.[8]

Assumption 1.2 *(DESIGN) Let* $\mathbf{\Sigma} \equiv \frac{1}{T_1}\sum_{t=1}^{T_1}\mathbb{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$. *There exists a constant* $\psi_0 > 0$ *such that*

$$\|\boldsymbol{\theta}[S_0]\|_1^2 \leq \frac{\boldsymbol{\theta}\mathbf{\Sigma}\boldsymbol{\theta}s_0}{\psi_0^2},$$

for all $\|\boldsymbol{\theta}[S_0^c]\|_1 \leq 3\|\boldsymbol{\theta}[S_0]\|_1$.

Assumption 1.3 *(HETEROGENEITY AND DEPENDENCY) Let* $\boldsymbol{w}_t \equiv (\nu_t, \boldsymbol{x}_t')'$, *then:*

[7]Some efficiency gain could be potentially obtain by a joint estimation, for instance, a SUR (seemly unrelated regression) setting if the regressors of each equation are the not the same. We do not pursue this route here.

[8]Recall that since we drop the equation subscript j, the assumptions below must understood for each equation $j = 1,\dots,q$ separately.

(a) $\{\boldsymbol{w}_t\}$ *is strong mixing with* $\alpha(m) = \exp(-cm)$ *for some* $c \geq \underline{c} > 0$

(b) $\mathbb{E}|w_{it}|^{2\gamma+\delta} \leq c_\gamma$ *for some* $\gamma > 2$ *and* $\delta > 0$ *for all* $1 \leq i \leq d, 1 \leq t \leq T$ *and* $T \geq 1$,

(c) $\mathbb{E}(\nu_t^2) \geq \epsilon > 0$, *for all* $1 \leq t \leq T$ *and* $T \geq 1$.

Assumption 1.4 *(REGULARITY)*

(a) $\varsigma = O\left(\frac{d^{1/\gamma}}{\sqrt{T}}\right)$

(b) $s_0 \frac{d^{2/\gamma}}{\sqrt{T}} = o(1)$

Assumption 1.2 is known as the compatibility condition, which is extensively discussed in Bülhmann e van der Geer (2011). It is quite similar to the restriction of the smallest eigenvalue of $\boldsymbol{\Sigma}$, when one replace $\|\boldsymbol{\theta}[S_0]\|_1^2$ by its upper bound $s_0\|\boldsymbol{\theta}[S_0]\|_2^2$. Notice that we make no compatibility assumption regarding the sample counterpart $\widehat{\boldsymbol{\Sigma}} \equiv \frac{1}{T_1}\sum_{t=1}^{T_1} \boldsymbol{x}_t\boldsymbol{x}_t'$.

Assumption 1.3 controls for the heterogeneity and the dependence structure of the process that generates the sample. In particular Assumption 1.3(a) requires $\{\boldsymbol{w}_t\}$ to be an α-mixing process with exponential decay. It could be replaced by more flexible forms of dependence such as near epoch dependence or $\mathcal{L}_p$-approximability on an α-mixing process as long as we control for the approximation error term. Assumption 1.3(b) bounds uniformly some higher moment which ensures an appropriate Law of Large Numbers, and Assumption 1.3(c) is sufficient for the Central Limit Theorem. The latter bounds the variance of the regression error away from zero, which is plausible if we consider that the fit will never be perfect regardless of how much relevant variables we have in (1-8).

Assumption 3.4(a) and (b) are regularity conditions on the growth rate of the penalty parameter and the number of (relevant/total) parameters, respectively. They are smaller than the analogous results found in the literature for the case of fix design and normality of the error term.[9]

We can now define $\mathcal{P}$ as the class of probability law that satisfies Assumptions 1.2,1.3 and 3.4(b). However, for convenience we explicitly state all those assumptions underlying the results that follows. Here is our main result.

[9]Under those condition, 3.4(a) and (b) become $\varsigma = O\left(\sqrt{\frac{\log d}{T}}\right)$ and $s_0\frac{\log d}{\sqrt{T}} = o(1)$, respectively.

Teorema 1.3 (MAIN) *Let $\mathcal{M}$ be the model defined by* (1-8)*, whose parameters are estimated by* (1-9)*, then under Assumptions 1.1-3.4:*

$$\sup_{P\in\mathcal{P}}\sup_{\boldsymbol{a}\in\mathbb{R}^q}\left|\mathbb{P}_P\left[\sqrt{T}\boldsymbol{\Omega}_T^{-1/2}(\widehat{\boldsymbol{\Delta}}_T-\boldsymbol{\Delta}_T)\leq\boldsymbol{a}\right]-\Phi(\boldsymbol{a})\right|\to 0,\quad \textit{as } T\to\infty,$$

where $\boldsymbol{\Omega}_T$ is defined in Proposition 1.2,the event $\{\boldsymbol{a}\leq\boldsymbol{b}\}\equiv\{a_i\leq b_i,\forall i\}$ and $\Phi(\cdot)$ is the cumulative distribution function of a zero-mean identity covariance normal random vector.

The results above are uniform with respect to the class of probability laws $\mathcal{P}$, which we believe to be large enough to be of some interest. Notice that we do *not* require any strong separation of the parameters away from zero, which is usually accomplished in the literature by imposing a $\theta_{\min}$ which is uniformly bounded away from zero. The uniform convergence above is possible, in our case, as consequence of Assumption 1.1, which translates into the treatment $\mathcal{D}_t$ being uncorrelated with the regressors $\boldsymbol{x}_t$. In other words, the potential non-uniformity issues regarding the estimation of the parameters of $\boldsymbol{\theta}_0$ do not contaminate the estimation of $\boldsymbol{\Delta}_T$, even if the coefficients of the conditional model are of order $O(T^{-1/2})$ as discussed in Leeb and Pötscher (2005,2008,2009).

In a different set-up, Belloni et al. (2014) consider the case where the treatment is correlated with the set of regressors. Consequently, they propose the estimation via a moment condition with the so called *orthogonality property* in order to achieve uniform convergence. Further, Belloni et al. (2016) generalize this idea to conduct uniform inference in a broad class of Z-estimators. In our framework the orthogonality property is a consequence of Assumption 1.1.

1.3.3
Hypothesis Testing under Asymptotic Results

Given the asymptotic normality of $\widehat{\boldsymbol{\Delta}}_T$, it is straightforward to conduct hypothesis testing. It is important, however, to remember the dependence of the results upon knowing the exact point of a possible break and the assurance that the peers are in fact untreated. Fortunately, both conditions can be tested, which is the topic of the next sections. For now will we consider that unit 1 is the only one potentially treated and the moment of the intervention, T_0, is known for certain.

First we need a consistent estimator for the variance $\boldsymbol{\Omega}_T$. More precisely, we need estimators for both $\boldsymbol{\Gamma}_{T_1}$ and $\boldsymbol{\Gamma}_{T_2}$. If we expect to have uncorrelated residuals and given the consistency of $\widehat{\boldsymbol{\theta}}$, we can simply estimate it by the average of the sum of squares of residuals in the pre-intervention model. A popular choice for serially correlated residuals is presented in Andrews (1991) and Newey e West (1987). Both have a similar structure given by the weighted autocovariance estimator as

$$\widehat{\mathbf{\Gamma}}_{T_i} = \widehat{\mathbf{\Gamma}}_{0_i} + \sum_{k=1}^{M} \phi(k) \left(\widehat{\mathbf{\Gamma}}_{k_i} + \widehat{\mathbf{\Gamma}}'_{k_i} \right), \quad i = \{1, 2\}, \tag{1-10}$$

where $\widehat{\mathbf{\Gamma}}_{k_1} \equiv \frac{1}{T_1 - k} \sum_{t=1}^{T_1 - k} \widehat{\boldsymbol{\nu}}_t \widehat{\boldsymbol{\nu}}'_{t+k}$, $\widehat{\mathbf{\Gamma}}_{k_2} \equiv \frac{1}{T_2 - k} \sum_{t=T_0}^{T-k} \widehat{\boldsymbol{\nu}}_t \widehat{\boldsymbol{\nu}}'_{t+k}$, $k = 0, \ldots, M$, and $\widehat{\boldsymbol{\nu}}_t = \boldsymbol{y}_t - \widehat{\mathcal{M}}_{T_0}(\boldsymbol{x}_t) - \widehat{\mathbf{\Delta}}_T I(t \geq T_0)$.

In practice, we still need to specify the maximum number of lags/bandwidth to consider and the weight function. Usually, the later is a kernel function centered at zero. A common choice is a Bartlett kernel where the weights are given simply by $\phi(k) = 1 - \frac{k}{M+1}$. Theorem 2 of Newey e West (1987) and Proposition 1 of Andrews (1991) give general conditions under which the estimator is consistent. Moreover, Andrews (1991) discusses what kind of kernels are allowed and present a sizeable list of options. It also describes a data-driven procedure for bandwidth selection.

Therefore, if we replace $\mathbf{\Omega}_T$ by $\widehat{\mathbf{\Omega}}_T \equiv \frac{\widehat{\mathbf{\Gamma}}_{T_1}}{T_1/T} + \frac{\widehat{\mathbf{\Gamma}}_{T_2}}{T_2/T}$, we can construct honest (uniform) asymptotic confidence intervals and hypothesis testing as follows:

Proposition 1.4 *(Uniform Confidence Interval) Let $\widehat{\mathbf{\Omega}}_T$ be a consistent estimator for $\mathbf{\Omega}_T$ uniformly in $P \in \mathcal{P}$. Under the same conditions of Theorem 1.3, for any given significance level α:*

$$\mathcal{I}_\alpha \equiv \left[\widehat{\Delta}_{j,T} \pm \frac{\widehat{\omega}_j}{\sqrt{T}} \Phi^{-1}(1 - \alpha/2) \right]$$

for each $j = 1, \ldots, q$, where $\widehat{\omega}_j = \sqrt{[\widehat{\mathbf{\Omega}}]_{jj}}$ and $\Phi^{-1}(\cdot)$ is the quantile function of a standard normal distribution. The confidence interval $\mathcal{I}_\alpha$ is uniformly valid (honest) in the sense that for a given $\epsilon > 0$, there exists a T_ϵ such that for all $T > T_\epsilon$:

$$\sup_{P \in \mathcal{P}} \left| \mathbb{P}_P \left(\Delta_{j,T} \in \mathcal{I}_\alpha \right) - (1 - \alpha) \right| < \epsilon.$$

Proposition 1.5 *(Uniform Hypothesis Test) Let $\widehat{\mathbf{\Omega}}_T$ be a consistent estimator for $\mathbf{\Omega}_T$ uniformly in $P \in \mathcal{P}$. Under the same conditions of Theorem 1.3, for a given $\epsilon > 0$, there exists a T_ϵ such that for all $T > T_\epsilon$:*

$$\sup_{P \in \mathcal{P}} \left| \mathbb{P}_P \left(W_T \leq c_\alpha \right) - (1 - \alpha) \right| < \epsilon,$$

where $W_T \equiv T\widehat{\boldsymbol{\Delta}}_T'\widehat{\boldsymbol{\Omega}}_T^{-1}\widehat{\boldsymbol{\Delta}}_T$, $\mathbb{P}(\chi_q^2 \leq c_\alpha) = 1 - \alpha$ *and* χ_q^2 *is a chi-square distributed random variable with* q *degrees of freedom.*

1.4 Extensions

We consider extensions of the framework developed previously. In Section 1.4.1 we deal with the problem of an unknown intervention time and propose a procedure to account for that and develop a consistent estimator for the most likely intervention time. The case of multiple intervention points is treated in Section 1.4.2 and, finally, Section 1.4.3 investigates the presence of treated unit among the controls, which is particularly useful for testing for spillover effects.

1.4.1 Unknown Intervention Timing

There are reasons why the intervention timing might not be known for certainty. It could be due to anticipation effects related to rational expectations regarding an announced change in future policy. Or, on the other hand, a simple delay in the response of the variable of interest. Regardless of the cause of uncertainty about the timing of the intervention, we propose a way to apply the methodology even when T_0 is unknown.

We start by reinterpreting our estimator as a function of λ (or $T_\lambda \equiv \lfloor \lambda T \rfloor$), where $\lambda \in \Lambda$, a compact subset of $(0, 1)$:

$$\widehat{\boldsymbol{\Delta}}_T(\lambda) = \frac{1}{T - T_\lambda + 1} \sum_{t \geq T_\lambda} \widehat{\boldsymbol{\delta}}_{t,T}(\lambda), \quad \forall \lambda \in \Lambda \tag{1-11}$$

PUC-Rio - Certificação Digital Nº 1212340/CA

where $\widehat{\boldsymbol{\delta}}_{t,T}(\lambda) = \boldsymbol{y}_t - \widehat{\mathcal{M}}_T(\lambda)(\boldsymbol{x}_t)$, for $t = T_\lambda, \ldots, T$, and $\widehat{\mathcal{M}}_T(\lambda)$ is the estimate of the model $\mathcal{M}$ based on the first $T_\lambda - 1$ observations. Also, consider a λ-dependent version of our average treatment effect, given by

$$\boldsymbol{\Delta}_T(\lambda) = \frac{1}{T - T_\lambda + 1} \sum_{t=T_\lambda}^{T} \boldsymbol{\delta}_t.$$

For fixed λ, provided that the condition of Proposition 1.2 are satisfied for T_λ (as opposed to just $T_0 \equiv T_{\lambda_0}$), we have the convergence in distribution to a Gaussian. Hence, it is sufficient to consider the following extra assumption.

Assumption 1.5 $\{(\boldsymbol{y}_t', \boldsymbol{x}_t')'\}$ *is a strictly stationary process.*

Assumption 1.5 above is clearly stronger than necessary. For instance, it would be enough to have $\{\boldsymbol{\nu}_t\}$ as a weakly stationary process. However, in order to avoid assumptions that are model dependent (via the choice of $\mathcal{M}$) we state Assumption 1.5 as it is. It follows for instance if the process that generates the observable data in the absence of the intervention $\{\boldsymbol{z}_t^{(0)}\}$ is strictly stationary and both transformations $\boldsymbol{h}(\cdot)$ and $\boldsymbol{h}_x(\cdot)$ are measurable.

In order to analize the properties of the estimator (1-11) it is convenient to define the stochastic process $\{\boldsymbol{S}_T\}$ indexed by $\lambda \in \Lambda$, such that for each $\lambda \in \Lambda$, we have $\boldsymbol{S}_T(\lambda) \equiv \sqrt{T}\boldsymbol{\Gamma}_T^{-1/2}\left[\boldsymbol{\Delta}_T(\lambda) - \boldsymbol{\Delta}_T(\lambda)\right]$. Note that unlike the notation used in Proposition 1.2, we do not include the factors T_1/T and T_2/T inside the asymptotic variance term also since all the results will be under stationarity (Assumption 1.5) we replace $\boldsymbol{\Gamma}_{T_1}$ and $\boldsymbol{\Gamma}_{T_2}$ by its asymptotic equivalent $\boldsymbol{\Gamma}_T$, which is independent of $\lambda \in \Lambda$.

Therefore, the convergence in distribution of $\boldsymbol{S}_T(\boldsymbol{\lambda})$ to a Gaussian for any finite dimension $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)'$ follows directly from Theorem 1.3 combined with Assumption 1.5 and the Cramèr-Wold device. Furthermore the next theorem shows that $\boldsymbol{S}_T$ converges uniformly in $\lambda \in \Lambda$.

Teorema 1.6 *Under the conditions of Proposition 1.2 and Assumption 1.5:*

$$\boldsymbol{S}_T(\lambda) \equiv \sqrt{T}\boldsymbol{\Gamma}_T^{-1/2}\left[\boldsymbol{\Delta}_T(\lambda) - \boldsymbol{\Delta}_T(\lambda)\right] \xrightarrow{d} \boldsymbol{S} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda}}(\lambda, \lambda') = \frac{I_q}{(\lambda \vee \lambda')(1 - \lambda \wedge \lambda')}$, $\forall (\lambda, \lambda') \in \Lambda^2$. *For* $p \in [1, \infty]$, $\|\boldsymbol{S}_T\|_p \xrightarrow{d} \|\boldsymbol{S}\|_p$, *where* $\|f\|_p = \left(\int |f(x)|^p dx\right)^{1/p}$ *if* $1 \leq p \leq \infty$ *and* $\|f\|_\infty = \sup_{x \in \mathcal{X}} |f(x)|$.

The second part of Theorem 1.6 gives us a direct approach to conduct inference in the case of unknown intervention time. We can replace $\boldsymbol{\Gamma}_T$ by a consistent estimator $\widehat{\boldsymbol{\Gamma}}_T$ (as for instance the one discussed in in Section 1.3.3) and conduct inference on $\|\widehat{\boldsymbol{S}}_T\|_p$ under a slightly stronger version of $\mathcal{H}_0$, (which clearly implies $\mathcal{H}_0$):

$$\mathcal{H}_0^\lambda : \boldsymbol{\delta}_t = \boldsymbol{0}, \quad \forall t \geq 1.$$

In practice, as it is the case for the structural breaks tests, we trim the sample to avoid finite sample bias close to the boundaries and select $\Lambda = [\underline{\lambda}, \bar{\lambda}]$. Table C.8 presents the critical values for common choices of $p = \{1, 2, \infty\}$ and trimming values.

The procedure above suggests a natural estimator for the unknown intervention time, which might be useful in situations such as the one discussed in Section 1.4.2 where treatment occurs at multiple unknown intervention times.

We assume a constant intervention such as

Assumption 1.6 $\boldsymbol{\delta}_t = \boldsymbol{\Delta}$, *for* $t = T_0, \ldots, T$, *where* $\boldsymbol{\Delta} \in \mathbb{R}^q$ *is non-random.*

Remark 1.2 *Recall that Assumption 1.6 is not overly restrictive due to the flexibility provided by the transformation* $h(.)$. *The mean of* $\boldsymbol{y}_t$ *might as well represent the variance, covariances or any other moment of interest of the original* $\boldsymbol{z}_{1t}$ *variable.*

Remark 1.3 *Assumption 1.6 implies an instantaneous treatment effect (step function) at* $t = T_0$. *In most cases, however, we might encounter a continuous intervention effect, possibly reaching a distinguishable new steady state value. We could accommodate these cases by trimming this transitory part of the sample, provided we have enough data, and then apply the methodology in the trimmed sample where Assumption 1.6 holds.*

Proposition 1.7 *Under the conditions of Proposition 1.2 and Assumptions 1.5 and 1.6,* $\widehat{\boldsymbol{\Delta}}_T(\lambda) \xrightarrow{p} \phi(\lambda)\boldsymbol{\Delta}$, *where*

$$\phi(\lambda) = \begin{cases} \frac{1-\lambda_0}{1-\lambda} & \text{if } \lambda \leq \lambda_0, \\ \frac{\lambda_0}{\lambda} & \text{if } \lambda > \lambda_0. \end{cases}$$

Since both $\frac{1-\lambda_0}{1-\lambda}$ and $\frac{\lambda_0}{\lambda}$ are bounded between 0 and 1, we have that $\|\mathsf{plim}\, \widehat{\boldsymbol{\Delta}}_T(\lambda)\|_p \leq \|\boldsymbol{\Delta}\|_p$ for all $\lambda \in \Lambda$, where $\|\cdot\|_p$ denotes the ℓ_p norm. Under the maintained hypothesis that $\boldsymbol{\Delta} \neq 0$, we can establish the identification result that $\mathsf{plim}\, \widehat{\boldsymbol{\Delta}}_T(\lambda) = \boldsymbol{\Delta}$ if and only if $\lambda = \lambda_0$. The result above naturally suggests an estimator for λ_0:

$$\widehat{\lambda}_{0,p} = \underset{\lambda \in \Lambda}{\mathsf{arg\,max}}\, J_{T,p}(\lambda) \quad \text{and} \quad J_{T,p}(\lambda) \equiv \|\widehat{\boldsymbol{\Delta}}_T(\lambda)\|_p. \tag{1-12}$$

Teorema 1.8 *Let* $p \in [1, \infty]$. *Under the conditions of Proposition 1.2 and Assumptions 1.5 and 1.6, for* $\boldsymbol{\Delta} \neq \mathbf{0}$, $\widehat{\lambda}_{0,p} = \lambda_0 + o_p(1)$. *If* $\boldsymbol{\Delta} = \mathbf{0}$, $\widehat{\lambda}_{0,p}$ *converges in probability to any* $\lambda \in \Lambda$ *with equal probability.*

1.4.2 Multiple Intervention Points

We can readily extend our analysis to the case of more than one intervention taking place in the unit of interest as long as, in each of them, Assumption 1.6 is valid. Suppose we have S ordered known intervention points corresponding to the fractions of the sample given by $\lambda_0 \equiv 0 < \lambda_1 < \cdots < \lambda_S < 1 \equiv \lambda_{S+1}$.

For each of the intervention points $s = \{1, \ldots, S\}$ we can define the time of each intervention by $T_s \equiv \lfloor \lambda_s T \rfloor$ and construct our estimator in the same way we did for the single intervention case. To simplify notation we define the set of all periods after intervention s but before the intervention $s+1$ by $\tau_s = \{T_s, T_s+1, \ldots, T_{s+1}-1\}$ and set $\#\{A\}$ the number of elements in the set A. Then, we have S estimators given by:

$$\widehat{\boldsymbol{\Delta}}_T^s \equiv \widehat{\boldsymbol{\Delta}}_T(\lambda_s, \widehat{\boldsymbol{\theta}}_s) = \frac{1}{\#\{\tau_s\}} \sum_{t \in \tau_s} \left[\boldsymbol{y}_t - \mathcal{M}_p(\boldsymbol{x}_t, \widehat{\boldsymbol{\theta}}_{s,T}) \right], \qquad s = 1, \ldots, S,$$

where once again $\widehat{\boldsymbol{\theta}}_{s,T}$ is the LASSO estimator using the sample indexed by $t \in \tau_{s-1}$. Note that we could allow the linear model to depend on s, i.e., differ from one intervention point to another. However, a much more parsimonious estimation could be obtained by choosing the same model for all intervention periods.

Under the same set of assumptions for the single intervention case plus Assumption 1.6, we have the sequence of estimators $\{\widehat{\boldsymbol{\Delta}}_T^s\}_{s=1}^S$ consistent for their respective intervention effects $\{\boldsymbol{\Delta}^s\}_{s=1}^S$ and also asymptotically normal. However, we need to make a minor adjustment in the asymptotic covariance matrix to reflect the intervention timing as:

$$\sqrt{T}\boldsymbol{\Gamma}_T^{-1/2}\left(\widehat{\boldsymbol{\Delta}}_T^s - \boldsymbol{\Delta}^s\right) \xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \frac{1}{(\lambda_s - \lambda_{s-1})(\lambda_{s+1} - \lambda_s)}\right], \quad s = 1, \ldots, S.$$

Since under Assumption 1.6 all the interventions are constant, we have that the asymptotic variance $\boldsymbol{\Gamma}$ is the same across all intervention points. Therefore, we can apply the inference for each breaking point as we have described for the single intervention case.

On the other hand, if the intervention points are unknown, we need to first estimate their location as in the single intervention case. Since the intervention points are assumed to be distinct, i.e. $\lambda_i \neq \lambda_j$, $\forall i, j$, it follows from Proposition 1.7 that there exists an interval of size $\epsilon > 0$ around every intervention point such that

$$\widehat{\mathbf{\Delta}}_T^p(\lambda) \xrightarrow{p} \begin{cases} \frac{1-\lambda_p}{1-\lambda}\mathbf{\Delta} & \text{if } \lambda \in [\lambda_p - \epsilon/2, \lambda_p], \\ \frac{\lambda_p}{\lambda}\mathbf{\Delta} & \text{if } \lambda \in (\lambda_p, \lambda_p + \epsilon/2]. \end{cases}$$

Nonetheless, in contrast to the single intervention scenario, in the case of multiple intervention points we need first to estimate how many are they and their respective location to construct $\{\widehat{\mathbf{\Delta}}_T^p\}_{p=1}^P$. One approach is to start with the null hypothesis of no intervention ($s = 0$) against the alternative of a single one. We can then compute $\widehat{\lambda}_1$ as in (1-12) and test the null using $\widehat{\mathbf{\Delta}}_T^0(\widehat{\lambda}_1)$. In case we are able to reject the null, we split the sample at $\widehat{\lambda}_1$ and repeat the procedure in each of the two subsample. Every time we reject the null we split the sample in $\widehat{\lambda}_s$ and proceed sequentially until we no longer reject the null in any subsample.

The sequential procedure described above was advocated by Bai e Perron (1998). It in based on the observation that given a non-zero number of true intervention points, the first loop will encounter the most significant one (in terms of SSR reduction) and proceed sequentially until it finds the last one of them. In case we have multiple intervention points with the same magnitude the method would converge to any of them with equal probability.

Formally, starting from an arbitrary number of $s \geq 0$ intervention points and for a given significance level α we test for each of the $s+1$ subsamples as:

$$\mathcal{H}_0^{(s)} : \mathbf{\Delta} = \mathbf{0} \quad \text{for all } \lambda \in [\lambda_j, \lambda_{j+1})_{j=0}^s\,,$$
$$\mathcal{H}_1^{(s+1)} : \mathbf{\Delta} \neq \mathbf{0} \quad \text{for any } \lambda \in [\lambda_j, \lambda_{j+1})_{j=0}^s\,.$$

Note that the overall significance level of the test is no longer the individual significance level and it has to be adjusted to account for the sequential nature of the procedure.

1.4.3
Testing for the unknown treated unit/Untreated peers

All the analysis carried out so far relies on the knowledge of which unit is the treated one and also, more importantly, on the assumption that the remaining are in fact untreated during the sample period (Assumption 1.1). Yet, there might be cases where we are either unsure or would like to test for those conditions. Given any *finite* subset $\mathcal{I}$ of the available units we would like to test the following hypothesis

$$\mathcal{H}_0^n : \mathbf{\Delta}_T^{(i)} = \mathbf{0} \quad \forall i \in \mathcal{I} \subseteq \{1, \ldots, n\}$$
$$\mathcal{H}_1^n : \mathbf{\Delta}_T^{(i)} \neq \mathbf{0} \quad \text{for some } i \in \mathcal{I}$$

Nothing prevents us from running the same procedure considering each unit $i \in \mathcal{I}$ to be the treated one to obtain $\widehat{\boldsymbol{\Delta}}_T^{(i)}$ as in (1-5) for $i = 1, \ldots, n_{\mathcal{I}}$, where $n_{\mathcal{I}} < \infty$ is the cardinality of the set $\mathcal{I}$. We can then stack all of them in a vector as $\widehat{\boldsymbol{\Pi}}_T(\mathcal{I}) \equiv \left(\widehat{\boldsymbol{\Delta}}_T^{(1)'} \ldots \widehat{\boldsymbol{\Delta}}_T^{(n_{\mathcal{I}})'}\right)'$ as an average estimator for the true average intervention effect vector $\boldsymbol{\Pi}_T(\mathcal{I}) \equiv \left(\boldsymbol{\Delta}_T^{(1)'} \ldots \boldsymbol{\Delta}_T^{((\mathcal{I}))'}\right)'$ where $\boldsymbol{\Delta}_T^{(i)}$ is defined for each unit. Hence,

Proposition 1.9 *Under the conditions of Proposition 1.2, for any* finite *subset* $\mathcal{I} \subseteq \{1, \ldots, n\}$

$$\sqrt{T}\boldsymbol{\Sigma}_{\mathcal{I}}^{-1/2}\left[\widehat{\boldsymbol{\Pi}}_T(\mathcal{I}) - \boldsymbol{\Pi}_T(\mathcal{I})\right] \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}),$$

where $\boldsymbol{\Sigma}_{\mathcal{I}}$ *is a covariance matrix with typical (matrix) element* $(i,j) \in \mathcal{I}^2$ *given by:*

$$\boldsymbol{\Omega}_T^{ij} \equiv T\mathbb{E}\left[\left(\widehat{\boldsymbol{\Delta}}_T^{(i)} - \boldsymbol{\Delta}_T^{(i)}\right)\left(\widehat{\boldsymbol{\Delta}}_T^{(j)} - \boldsymbol{\Delta}_T^{(j)}\right)'\right],$$

with $\boldsymbol{\Omega}_T^{ij} = \frac{\boldsymbol{\Gamma}_{T_1}^{ij}}{T_1/T} + \frac{\boldsymbol{\Gamma}_{T_2}^{ij}}{T_2/T}$, $\boldsymbol{\Gamma}_{T_1}^{ij} = \mathbb{E}\left[\frac{(\sum_{t \leq T_1}\boldsymbol{\nu}_t^i)(\sum_{t \leq T_1}\boldsymbol{\nu}_t^{j'})}{T_1}\right]$, *and* $\boldsymbol{\Gamma}_{T_2}^{ij} = \mathbb{E}\left[\frac{(\sum_{t \geq T_0}\boldsymbol{\nu}_t^i)(\sum_{t \geq T_0}\boldsymbol{\nu}_t^{j'})}{T_2}\right]$.

Therefore, for a given consistent estimator $\widehat{\boldsymbol{\Sigma}}$ *we have under* $\mathcal{H}_0^n$*:*

$$W_T^{\pi} \equiv T\widehat{\boldsymbol{\Pi}}_T'\widehat{\boldsymbol{\Sigma}}_{\mathcal{I}}^{-1}\widehat{\boldsymbol{\Pi}}_T \xrightarrow{d} \chi^2_{nq}.$$

We can obtain a consistent estimator for $\boldsymbol{\Sigma}_{\mathcal{I}}$ repeating the same procedure described in Section 1.3.3 for each pair $(ij) \in \mathcal{I}^2$ to obtain $\widehat{\boldsymbol{\Omega}}^{ij}$ and finally construct the matrix $\widehat{\boldsymbol{\Sigma}}_{\mathcal{I}}$. Hence, for a desired significance level, we can then use W_T^{π} to test $\mathcal{H}_0^n$. Once you remove the (likely) treated unit and re-test it again with the remanning units (peers) the test becomes yet more useful. In case we fail to reject the null, we can interpreted this result as a direct evidence in favour of the hypothesis that the peers are in fact untreated considering the sample at hand. Which ultimately provides support to our key Assumption 1.1.

1.5 Selection Bias, Contamination, Nonstationarity and Other Issues

In this section we discuss some possible sources of bias in the ArCo method. In particular, we consider the potential effects when the intervention does not affect only the outcome of the variable of unit 1. Equivalently, we investigate the consequences whenever Assumption 1.1(b) fails and we expect to have $\mathbb{E}(\boldsymbol{z}_{0t}|\mathcal{D}_t) \neq 0$.

We consider without loss of generality a simpler version of the DGP described in Section 2. Each unit $i = 1, \ldots, n$ under no intervention is represented by $z_{it}^{(0)} = l_i f_t + \eta_{it}$, where η_{it} is a zero mean independent and identically distributed (iid) idiosyncratic shock with variance $\sigma^2_{\eta_i}$. Furthermore, $\mathbb{E}(\eta_{it}\eta_{jt}) = 0$, for all $i \neq j$. Also, the common factor vector f_t is an iid random variables with zero mean and variance σ^2_f.

Set $y_t = z_{1t}$, $\boldsymbol{x}_t = (z_{2t}, \ldots, z_{nt})'$, $\boldsymbol{l}_0 = (l_2, \ldots, l_n)'$ and $\boldsymbol{\sigma}^2_{\eta 0} = (\sigma^2_{\eta_2}, \ldots, \sigma^2_{\eta_n})'$. In this setup we can write

$$\begin{pmatrix} \boldsymbol{y}_t \\ \boldsymbol{x}_t \end{pmatrix} \sim \left[\boldsymbol{0}, \sigma^2_f \begin{pmatrix} l_1^2 + r_1 & l_1 \boldsymbol{l}_0' \\ l_1 \boldsymbol{l}_0 & \boldsymbol{l}_0 \boldsymbol{l}_0' + \mathsf{diag}\,(\boldsymbol{r}_0) \end{pmatrix}\right],$$

where $r_i \equiv \frac{\sigma^2_{\eta_i}}{\sigma^2_f}$ is the noise to signal ratio of unit $i = 1, \ldots, n$ and $\boldsymbol{r}_0 = (r_2, \ldots, r_n)'$.

As a consequence, the best linear projection model is given by $\mathbb{L}(\boldsymbol{y}_t|\boldsymbol{x}_t) = \boldsymbol{x}_t'\boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0 = \left[\boldsymbol{l}_0\boldsymbol{l}_0' + \mathsf{diag}\,(\boldsymbol{r}_0)\right]^{-1}(l_1\boldsymbol{l}_0)$. Furthermore, $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_0 + \nu_t$, where $\mathbb{E}(\boldsymbol{x}_t\nu_t) = \boldsymbol{0}$ by definition, and $\sigma^2_\nu \equiv \mathbb{E}(\nu_t^2) = \sigma^2_f\,(l_1^2 + r_1 - \boldsymbol{\beta}_0' l_1 \boldsymbol{l}_0)$.

Therefore, we have that $\boldsymbol{\beta}_0 \equiv \boldsymbol{\beta}_0(\boldsymbol{l}, \boldsymbol{r})$ and $\sigma^2_\nu \equiv \sigma^2_\nu(\boldsymbol{l}, \boldsymbol{r}, \sigma^2_f)$, where $\boldsymbol{r} = (r_1, \boldsymbol{r}_0')'$ and $\boldsymbol{l} = (l_1, \ldots, l_n)'$.

Suppose now that we have an intervention affecting all units from T_0 onwards, i.e. Assumption 1.1(b) does *not* hold. We consider two situations, one where the intervention is a change in the common factor given by a deterministic sequence $\{c_t^f\}_{t \geq T_0}$ and one where it is completely idiosyncratic $\{c_t^i\}_{t \geq T_0}$ for $i = 1, \ldots, n$, $z_{it}^{(1)} = z_{it}^{(0)} + 1\{t \geq T_0\}\left(c_t^i + l_i c_t^f\right)$.

Consequently, for $t = T_0, \ldots, T$:

$$\begin{aligned}\delta_t = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_0 &= y_t^{(0)} + c_t^1 + l_1 c_t^f - \left(\boldsymbol{x}_t^{(0)} + \boldsymbol{c}_t^0 + \boldsymbol{l}_0 c_t^f\right)'\boldsymbol{\beta}_0 \\ &= c_t^1 + \nu_t - \boldsymbol{c}_t^{0\prime}\boldsymbol{\beta}_0 + (l_1 - \boldsymbol{l}_0'\boldsymbol{\beta}_0)\, c_t^f.\end{aligned}$$

Clearly, under Assumption 1.1(b), we have that $\boldsymbol{c}_t^{(0)} = c_t^f = 0, \forall t$, thus $\mathbb{E}(\delta_t) = c_t^1$ and, ignoring the sampling error of estimating $\boldsymbol{\beta}_0$, the ArCo estimator will be unbiased for the average of c_t^1 for the post intervention period. On the other hand, without those assumptions we have the following bias in normalized statistic

$$b_t \equiv \mathbb{E}\left(\frac{\delta_t - c_t^1}{\sigma_\nu}\right) = \underbrace{\left(\frac{l_1 - \boldsymbol{l}_0'\boldsymbol{\beta}_0}{\sigma_\nu}\right)}_{\equiv \phi_f} c_t^f - \frac{\boldsymbol{c}_t^{0\prime}\boldsymbol{\beta}_0}{\sigma_\nu} \tag{1-13}$$

The factor in the first term of the bias $\phi_f = \phi_f(\boldsymbol{l}, \boldsymbol{r}, \sigma_f^2)$ is a non-linear expression which is hard to express in closed form. However, regardless of the choice of the factor loads $\boldsymbol{l}$ and idiosyncratic shock variances $\boldsymbol{\sigma}_\eta^2 = (\sigma_{\eta_1}^2, \ldots, \sigma_{\eta_n}^2)'$, we have that as $\sigma_f^2 \to \infty$, $r \to 0$ and consequently $R^2 \to 1$. Hence we write $\phi_f = \phi_f(R^2)$. Moreover, $\phi_f(R^2)$ is strictly decreasing in R^2 and approaches zero quite fast as it can be seen in the left scale of Figure B.1. Also $\phi_f = \phi(s_0)$ is also decreasing in the number of relevant variables s_0 for fix R^2.

Hence, if $\boldsymbol{c}_t^0 = \boldsymbol{0}$ but $c_t^f \neq 0$, even with moderate R^2, we have a reasonably small bias which causes the inference to be valid with minor overejection. This is in contrast to the case where we do not include relevant peers in our analysis . In fact, as mentioned previously in the Introduction, that is the main motivation for using the present methodology as opposed to an alternative that does not involve peers (a simple before-and-after estimation of averages for instance). ArCo can effectively isolate the intervention of interest even in the case of partial fulfilment of Assumption 1.1. In the limit of a perfect counterfactual, the bias is zero and the higher is the correlation among the treated unit and the peers, the smaller is the bias.

The second bias term in (1-13) can be seen as a result, for instance, of a global shock that induce breaks in peers in a non-systematic way, which makes this source of bias difficult to handle. To get a better sense, consider for instance the case where the idiosyncratic shock is a fixed proportion of the standard deviation of each unit, i.e. $c_t^i = k\sigma_i, \forall i$ for some $k \in \mathbb{R}$. In that case, $\phi_g = (\boldsymbol{\sigma}'\boldsymbol{\beta}_0/\sigma_\nu)k$, where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)'$. Here the opposite happens, namely $\phi_g(R^2)$ is zero when $R^2 = 0$ and increases in the overall fit of the model. The bias increase is quite sharp as can been seen in the right scale of Figure B.1.

Therefore, whenever one expects $\boldsymbol{c}_t^0 \neq \boldsymbol{0}$, the ArCo methodology does not work properly but the BA estimator does as it can be seen as a particular case of the ArCo estimator with $R^2 = 0$ (for instance by not including any peers) and hence the bias is zero. In general, the ArCo estimator gives the difference between the actual break in the treated unit and what is expected from the peers. A standard solution is to assume that the "treatment assignment" is independent of $\boldsymbol{z}_{0t} = (z_{2t}, \ldots, z_{nt})'$, which is our Assumption 1.1 and the ArCo approach is not subject to selection bias. However, it is important to stress that the "treatment assignment" might be dependent on z_{1t} and our approach is still valid.[10] One way to check if there is no "treatment contamination" is to test the peers for possible breaks after T_0 as discussed in Section 1.4.3.

Other possible source of problems is the use of "non-stationary" processes,

[10]The result is analogous to the average treatment effect on the treated not being biased by selection on (un)observables.

leading to spurious results. In this chapter we focus solely on the case the variables of interest have some sort of “fading memory” behaviour. Thus, if one or more variables are found to be integrated, they must be differenced first in order to achieve stationarity.

1.6 Monte Carlo Simulation

We conducted two sets of Monte Carlo simulations. First, we conduct size and power simulations in order to investigate the finite sample properties of the test. We consider a broad range of cases by combining different innovation distributions, sample sizes, number of peers, number of relevant peers, dependence structure, trends and intervention types. Second, a “horse race” is proposed in order to compare the ArCo estimator with potential alternatives. We consider the SC method of Abadie e Gardeazabal (2003) and Abadie et al. (2010), the PF estimator suggested in Gobillon e Magnac (2016) and the DiD and BA estimators.

1.6.1 Size and Power Simulations

The DGP considered is a version of the common factor model (2-6) with the following baseline scenario: $T = 100$ observations, $n = 100$ units, $q = 1$ one variable per unit, $\lambda_0 = 0.5$ (intervention at the middle of the sample), $s_0 = 5$ relevant (non-zero) parameters with loading factor equal to 1 and $f = 1$ common factor. The common factor and all idiosyncratic shocks are independent and identically normally distributed with zero mean and unit variance. We perform 10,000 simulations.

First, we analyze the influence of the underlying distribution on the test size by holding all the other parameters above fixed and performing the simulation for a chi-square distribution with 1 degree of freedom for asymmetry issues, t-Student distribution with 3 degrees of freedom for fat-tails and a mixed normal distribution for bimodality.[11] As shown in first panel of Table C.2, little influence in the overall size of the test is perceived.

Next we analyze the influence of the number of observations in the test size. We consider $T = \{25, 50, 75, 100\}$. Surprisingly, the size distortions are small even with only 50 observations as shown in the second panel of Table C.2. We stress that since we deal with the intervention at the middle of the sample we have less than $T/2$ observations to fit the high dimensional model.

[11] All innovations are standardized to zero mean and unit variance.

We now investigate the influence of increasing the number of covariates (by increasing either the number of lags or the number of peers)[12]. We set $d = \{100, 200, 500, 1000\}$. The third panel of Table C.2 shows that the test size seems to be unaffected by the increase in model complexity. This should come with no surprise since consistent model selection is not an issue for the methodology to work.

We consider a change of relevant (non-zero) covariates (units) in the pre-intervention model. We consider a case where all the regressors are irrelevant, which reduces (asymptotically) the ArCo to the BA estimator, and we further increase s_0. In the last scenario we consider all regressors non-zero but with decreasing magnitude $1/\sqrt{j}, j = 1, \ldots, 100$. In all cases the LASSO does not overfit the pre-intervention data and the size distortions are small as displayed in Table C.2.

Finally, we consider the case where each unit follows a first-order autoregressive process in order to investigate issues that arise in the presence of serial correlation. In this scenario we include lags of the relevant covariates instead of new peers. The results are shown in the last panel of Table C.2. We note a persistent oversized test, which is more pronounced as the autoregressive coefficient (ρ) becomes closer to 1. The empirical distribution of the estimator (not shown) is, however, very close to normal, and the distortion is a sole consequence of the poor finite sample properties of the variance estimator . In particular it underestimates $\boldsymbol{\Omega}$. We tried several alternatives for $\widehat{\boldsymbol{\Omega}}_T$, including Newey e West (1987), Andrews (1991), Andrews e Monahan (1992), and Haan e Levin (1996). We obtain the best results (last panel of Table C.2) using the procedure proposed in Andrews e Monahan (1992).

It is worth mentioning that the slightly oversized tests are a direct consequence of the persistence of $\{\nu_t\}$ and not necessarily of the persistence of $\{(y_t, \boldsymbol{x}_t')\}$ per se. The problem is attenuated, for instance, when enough lags are included to make $\{\nu_t\}$ closer to a white noise process, or when a linear combination of (potentially highly persistent) $\{(y_t, \boldsymbol{x}_t')\}$ is almost uncorrelated. For pure finite MA processes the usual kernel HAC estimator are known to perform well and the tests are not oversized.

1.6.2 Estimator Comparison

In order to conduct the "horse race" among competitors for counterfactual analysis we consider the following DGP:

[12]The difference is not completely innocuous since we loose one observation to each included lag. Therefore, we include new (uncorrelated) peers and deal with the lag inclusion in the serial correlation scenario.

$$\boldsymbol{z}_{it}^{(0)} = \rho \boldsymbol{A}_i \boldsymbol{z}_{it-1}^{(0)} + \boldsymbol{\varepsilon}_{it}, \quad i = 1, \ldots, n, ; t = 1, \ldots, T, \tag{1-14}$$

where $\boldsymbol{\varepsilon}_{it} = \boldsymbol{\Lambda}_i \boldsymbol{f}_t + \boldsymbol{\eta}_{it}$, $\boldsymbol{f}_t = [1, (t/T)^{\varphi}, v_t]$, $\boldsymbol{z}_{it} \in \mathbb{R}^q$, $\rho \in [0, 1)$, $\varphi > 0$, $\boldsymbol{A}_i(q \times q)$ is a diagonal matrix with diagonal elements strictly between -1 and 1, $\{v_t\}$ is a sequence of iid standardized normal random variables, $\{\boldsymbol{\eta}_{it}\}$ is a sequence of iid normal random vectors with zero mean and covariance matrix $r_f^2 \boldsymbol{I}_{nq}$ where $r_f > 0$ can be interpreted as the noise-to-signal ratio which controls the overall correlation among the units, and $\boldsymbol{\Lambda}_i$ is a $(q \times 3)$ matrix of factor loadings.

Let $\boldsymbol{z}_t$ be the nq dimensional vector obtained by stacking all the $\boldsymbol{z}_{it}^{(0)}$ and $\boldsymbol{\Lambda}$ is the $(nq \times 3)$ matrix after stacking all the $\boldsymbol{\Lambda}_i$. Similarly, define $\boldsymbol{\varepsilon}_t$ by stacking $\boldsymbol{\varepsilon}_{it}$ and $\boldsymbol{A}$ is the $(nq \times nq)$ diagonal matrix composed by the block diagonals $\boldsymbol{A}_i$. We use the notation $\boldsymbol{\Lambda}(j)$ to denote the jth column of $\boldsymbol{\Lambda}$, thus $\boldsymbol{\mu}_{\varepsilon,t} \equiv \mathbb{E}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Lambda}(1) + \boldsymbol{\Lambda}(2)(t/T)^{\varphi}$, $\boldsymbol{\Omega} \equiv \mathbb{V}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Lambda}(3)\boldsymbol{\Lambda}(3)' + r_f^2 \boldsymbol{I}_{nq}$, $\boldsymbol{\mu}_t \equiv \mathbb{E}(\boldsymbol{z}_t) = (\boldsymbol{I}_{nq} - \rho \boldsymbol{A})^{-1} \boldsymbol{\mu}_{\varepsilon,t}$, and $\mathsf{vec}\,(\boldsymbol{\Sigma}) \equiv \mathsf{vec}\,[(\mathbb{V}\boldsymbol{z}_t)] = [\boldsymbol{I}_{(nq)^2} - \rho^2 \boldsymbol{A} \otimes \boldsymbol{A}]^{-1} \mathsf{vec}\,(\boldsymbol{\Omega})$.

We set $y_{it}^{(1)} = y_{it}^{(0)} + \delta_t 1\{t \geq T_0 \text{ and } i = 1\}$, for simplicity we set $\delta_t = \delta$ constant and equal to one standard deviation from the unit of interest (unit 1). We are interested in estimating the average treatment effect

$$\Delta = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \delta_t = \delta.$$

We now briefly state the estimators considered in the Monte Carlo study. Whenever is convenient we use the following partition scheme: $\boldsymbol{z}_{it} = (y_{it}, \boldsymbol{x}_{it}')'$ and $\boldsymbol{z}_{0t} = (\boldsymbol{z}_{2t}', \ldots \boldsymbol{z}_{nt}')$.

Before-and-After (BA)

The difference between the average of the y_{1t} before and after the intervention:

$$\widehat{\Delta}_{BA} = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} y_{1t} - \frac{1}{T_0 - 1} \sum_{t=1}^{T_0 - 1} y_{1t}.$$

Differences-in-Differences (DiD)

The ordinary least squares (OLS) estimator of the dummy coefficient in the following regression models. For the case with covariates,

$$y_{it} = \alpha_0 + \boldsymbol{x}_{it}' \boldsymbol{\beta} + \alpha_1 I(i = 1) + \alpha_2 I(t \geq T_0) + \Delta_{DD^*} I(i = 1, t \geq T_0) + \varepsilon_{it},$$

or, for the case without covariates,

$$y_{it} = \alpha_0 + \alpha_1 I(i=1) + \alpha_2 I(t \geq T_0) + \Delta_{DD} I(i=1, t \geq T_0) + \varepsilon_{it}.$$

Gobillon and Magnac (GM)

The estimator is defined as per Gobillon e Magnac (2016):

$$\widehat{\Delta}_{GM} = \frac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \left(y_{1t} - \widehat{y}_{1t}\right),$$

where $\widehat{y}_{1t}^* = \boldsymbol{x}_{1t}\widehat{\boldsymbol{\beta}} + \widehat{f}_t\widehat{\Lambda}_1$ or without including the covariates $\widehat{y}_{1t} = \widehat{f}_t\widehat{\Lambda}_1$. We choose r the number of factors to be 2 (or 3 if a trend is included).

Synthetic Control (SC)

For simulation purposes we use the algorithm *Synth*[13]. We choose on top of all covariates ($\boldsymbol{x}_{it}$), the average of the dependent variable ($\boldsymbol{y}_{it}$) during the pre-intervention period as a matching variable.

$$\widehat{\Delta}_{SC} = \tfrac{1}{T - T_0 + 1} \sum_{t=T_0}^{T} \left(y_{1t} - \widehat{y}_{1t}\right),$$

where $\widehat{y}_{1t} = \boldsymbol{w}^{*\prime}\boldsymbol{y}_{0t}$. The weight vector $\boldsymbol{w}$ must be non-negative entries that sum to one. It comes from a minimization process involving only values of the selected variables prior to the intervention. In our particular case, we take the pre-intervention average $\bar{\boldsymbol{z}} = \frac{1}{T_0-1}\sum_{t=1}^{T_0-1} \boldsymbol{z}_t$, partition as $\bar{\boldsymbol{z}} = (\bar{\boldsymbol{z}}_1, \bar{\boldsymbol{z}}_0{}')'$ and reshape $\bar{\boldsymbol{z}}_0$ to a matrix $\bar{Z}_0(n-1 \times q)$ where each row are the variables of each of the remaining $n-1$ units

$$\boldsymbol{w}^*(\boldsymbol{V}) = \underset{\boldsymbol{w} \geq 0, \|\boldsymbol{w}\|_1 = 1}{\arg\min} \|\bar{\boldsymbol{z}}_1 - \boldsymbol{w}'\bar{\boldsymbol{z}}_0\|_{\boldsymbol{V}},$$

where $\|\cdot\|_V$ is the norm induced by a positive definite matrix $\boldsymbol{V}$.

Finally, $\boldsymbol{V}$ is chosen as

$$\boldsymbol{V}^* = \arg\min \frac{1}{T_0 - 1} \sum_{t=1}^{T0-1} \left[y_{1t} - \boldsymbol{w}^*(\boldsymbol{V})'\boldsymbol{y}_{0t}\right]^2, \tag{1-15}$$

and we set $\boldsymbol{w}^* \equiv \boldsymbol{w}^*(\boldsymbol{V}^*)$.

The results are presented in Table C.3. The smoothed histograms can be found in Figures B.2–B.7. Overall, the SC and the GM are heavily biased

[13]R package maintained by Jens Hainmueller.

in most cases considered. For the former, this might well be a consequence of the instability of algorithm to find the minimizer of (1-15), since the bias persists even in the absence of time trends, where any fixed linear combination of the peers should give us an unbiased estimator. For the latter it is most likely a consequence of the poor finite sample properties of common factor estimator. It is well understood from Bai (2009) that the consistency depends on the double asymptotics on n and T. On the other hand, BA, DiD and the ArCo seems to have comparable small bias at least in absence of deterministic trends regardless of the presence of serial correlation. The ArCo seems to have better MSE performance. This comes with no surprise since by definition our estimator in the first stage searches for the linear combination that minimizes the MSE.

For the trended cases, first note the BA estimator is severely biased since without using the information of the peers it cannot take into account the time trend effect. For the common trend cases, the DiD estimators have relatively small bias for both the linear and quadratic term. For the former it is expected since a common linear time trend the exactly the kind of DGP that the DiD estimator was designed for. Once again, the ArCo estimators have comparable bias to the DiD estimators for the common trend cases but with significant smaller variance (ranging from 6-16 times smaller). The clear advantage of the ArCo estimation can be seem in the idiosyncratic time trend cases. Even though some small (in finite sample) bias start to show up, it is clear much smaller than all other alternatives.

1.7 The Effects of an Anti Tax Evasion Program on Inflation

In this section we apply the ArCo methodology to estimate the effects of an anti tax evasion program in Brazil on inflation. Although, the causes of business non-compliance and tax evasion have been extensively studied in the literature, as, for example, in Slemrod (2010), little attention has been devoted to measure the indirect effects from enforcing tax compliance.

In Brazil, tax evasion is a major fiscal concern and both the federal and local governments have been proposing new strategies to reduce evasion. Early in 1996, the federal government introduced the SIMPLES[14] system which drastically simplified the tax payments process and helped in reducing the tax burden on small enterprises. Later in 2005, the federal government launched the electronic sales receipt program (*Nota Fiscal Eletrônica*), to further reduce compliance costs to firms.

[14] Integrated System of Tax Payments for Micro and Small Enterprises

In October 2007, the state government of São Paulo in Brazil implemented an anti tax evasion scheme called *Nota Fiscal Paulista* (NFP) program. The NFP program consists of a tax rebate from a state tax named ICMS (tax on circulation of products and services). ICMS is similar to the European VAT and the Canadian GST. However, unlike VAT and GST, ICMS does not apply to services other than those corresponding to interstate and intercity transportation and communication services. The NFP program works as an incentive to the consumer to ask for electronic sales receipts. The registered sales receipts give the consumer the right to participate in monthly lotteries promoted by the government. Furthermore, according to the rules of the program, registered consumers have the right to receive part of the ICMS paid by the seller, as tax rebate, when their tax identifier numbers (CPF) are included in the electronic sales receipts. Similar initiatives relying on consumer auditing schemes were proposed in the European Union and in China; see, for example, Wan (2010). The effectiveness of such programs has been discussed in Fatas et al. (2015) and Brockmann et al. (2016). In the Brazilian state of São Paulo, the NFP program has received extensive support from the population. In January 2008, 413 thousand people were registered in program while in October 2013 there were more than 15 million participants. The amount in Brazilian Reais distributed as rebates also grew rapidly from 44 thousand Reais in January 2008 to an average of 70 million Reais distributed monthly by the end of the same year. Figure B.8 illustrates the NFP participation as well as the value distributed as tax rebates.

Souza (2014) was the first author to discuss whether retailers increased prices in response to the NFP program and consequently whether the program impacted negatively consumers' purchasing power. By using the SC method to construct a counterfactual to the State of São Paulo, Souza (2014) showed that one year after the launching of the NFP program, the accumulated inflation on food away from home (FAH) was 5% higher in the state of São Paulo when compared to the synthetic control. In September 2009, the differences raised to 6.5%. We extend the analysis of Souza (2014) by considering the ArCo methodology as an alternative to the SC method. We also consider the BA, GM, and DiD estimators.

Under the assumptions that (i) a certain degree of tax evasion was occurring before the intervention, (ii) the sellers have some degree of market power and (iii) the penalty for tax evasion is large enough to alter the seller behaviour, one is expected to see an upward movement in prices due to an increase in marginal cost. Therefore, we would like to investigate whether the NFP had an impact on consumer prices in São Paulo. We test this hypothesis below as an empirical illustration of the ArCo methodology. The answer to this kind of question has important implications regarding social welfare effects that are usually neglected in the fiscal debate whenever the aim is to enforce tax compliance.

The NFP was not implemented throughout the sectors in the economy at once. The first sector were restaurants, followed by bakeries, bars and other food service retailers. We do not possess a perfect match for a general consumer price index (IPCA - IBGE) and the sector where the NFP was implemented. However, we can take the IPCA component of food away from home (FAH) as a good indicator for price levels in those sectors. The sample then consists of monthly FAH index for 10 metropolitan areas[15] including São Paulo from January 1995 to September 2009. As a matter of comparison, Souza (2014) estimated a counterfactual by the SC method with assigning the following weights to Belo Horizonte, Recife, Goiânia, and Porto Alegre, respectively: 0.40, 0.27, 0.19, and 0.14. All other donors were assigned zero weights.

In order to compute the counterfactual by the ArCo methodology we consider the following variables from the pool of donors: monthly inflation (FAH), monthly GDP growth, monthly retail sales growth and monthly credit growth. All variables are stationary and no lags or additional transformations are considered. The conditional model is linear and is estimated by LASSO, where the penalty parameter is selected by the Hannan and Quinn (HQ) criterium. The choice of the HQ instead of the BIC, for example, is driven by the fact that the latter delivers conditional models with no variables in most of the cases. The in-sample period (pre-intervention) consists of 33 months while the size of the out-of-sample period is 23.

The factors in the GM methodology are computed from the monthly growth in GDP, retail sales and credit by principal component methods. The number of factors is determined as to explain 80% of the total variance in the data. The BA estimator considers only variables from the treated unit.

The results are depicted in Table C.4. The upper panel in the table reports, for different choices of conditioning variables, the estimated average effect after the adoption of the 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. In all cases, the average effect is significant at the 1% level. The highest R-squared is

[15] Goiânia-GO, Fortaleza-CE, Recife-PE, Salvador-BA, Rio de Janeiro-RJ, São Paulo-SP, Porto Alegre-RS, Curitiba-PR, Belém-PA, Belo Horizonte-MG

achieved when inflation and GDP are used as conditioning variables, followed by a model with inflation, GDP and retail sales. In the first case, column (5) of Table C.4, the monthly average effect is 0.4478%. The aggregate effect during the out-of-sample period is 10.72%. In the second case, column (6) of Table C.4, the monthly average effect is 0.3796% and the aggregate effect is 9.04%. Two facts worth discussing. The first one is the much higher estimated effect when only credit variables are included. This is due to huge outliers (huge increase) observed in credit series in the out-of-sample period for the states of Pernambuco and Rio de Janeiro. If these two states are removed from the donors pool, the monthly average effect drops to 0.5768%. The second point that deserves attention is the much lower effect when only inflation is considered, although the in-sample fit is reasonably good.

Figures B.9 and B.10 show the actual and counterfactual data, both in-sample and out-of-sample. Figure B.9 considers the case where only inflation and GDP growth are considered as conditioning variables while the plots in Figure B.10 consider the case where retail sales growth are also included as a potential regressor in the first stage model.

The lower panel of Table C.4 presents some alternative measures of the average effect, namely the BA, GM and DiD estimators. In all cases the estimated effects are smaller than the ones estimated with the ArCo. The DiD estimators are closer to the SC. The GM falls somehow in between the SC/DiD and the ArCo.

We also run a placebo ArCo estimator to check the robustness of the method. When we do this we find that Porto Alegre seems to have nontrivial breaks after October 2007; see Table C.5. For this reason we re-run the analysis without Porto Alegre in the donor pool. The results are reported in Table C.6. The overall picture seems unchanged.

1.8 Conclusions and Future Research

We proposed a flexible method to conduct counterfactual analysis with aggregate data wish is specially relevant in situations where there is a single treated unit and “controls” are not readily available, such as in regional policy evaluation. The ArCo methodology is very easy to implement and extends and generalize previous proposals in the literature in several aspects: (1) the distribution of test for no-intervention effect is standard and asymptotically honest confidence regions for the average intervention effect can be easily constructed; (2) although the results rely on the number of time-series observations diverging, the LASSO estimator has good finite sample properties,even when the number of estimated parameters are much larger than the sample size; (3) we allow for nonlinear, heterogenous confounding effects; (4) we provide a complete asymptotic theory which can be used to jointly test for intervention effects in a group of variables; (5) The methodology can be applied even if the time of the intervention is not known for certain, which gives us a consistent estimator for the time of the intervention; (6) multiple interventions can be handled; and finally, (6) we also propose a test for the presence of spillover effects among the units.

The current research can be extended in several directions as, for example, the case where the variables are nonstationary (either with cointegration or not). A non-parametric or semiparametric estimation in the pre-intervention model can be also considered.

2
Counterfactual Analysis with Integrated Processes

2.1
Introduction

Over the last few years, there has been a growing interest in the literature in developing econometric tools to conduct counterfactual analysis with aggregate data when a "treated" unit suffers an intervention, such as a policy change, and there is not a clear control group available. In these situations, the proposed solution is to construct an artificial counterfactual from a pool of "untreated" peers ("donors pool"). For example, Hsiao et al. (2012) considered a stationary panel factor model, hereafter PF, where the counterfactual for the treated variable of interest is constructed from a linear combination of observed variables from selected peers given by the conditional expectation model. Another seminal method is the Synthetic Control, hereafter SC, approach of Abadie e Gardeazabal (2003) and Abadie et al. (2010). In the SC framework, the counterfactual variable is build as a convex combination of peers where the weights of the combination are estimated from time-series averages of several variables from the donor pool and is inspired by the matching literature. Although, the above methods seem similar they differ remarkably in the way the linear combination of peers is constructed.

More recently, there has been several extensions of the above methods being proposed in the literature. Ouyang e Peng (2015) extended the PF method by relaxing the linear conditional expectation assumption and introducing a semi-parametric estimator to construct the artificial counterfactual. Du e Zhang (2015) and Gao et al. (2015) made improvements on the selection mechanism for the constituents of the donors pool in the PF method. Fujiki e Hsiao (2015) considered the case of multiple treatments. Carvalho et al. (2016), proposed the Artificial Counterfactual (ArCo), which is a major extension of the PF method and considered, as well, the case of high-dimensional data. Finally, the SC method has been generalized by Xu (2015).

The main purpose of this chapter is to investigate the consequences of applying panel based methods, such as Hsiao et al. (2012) and Carvalho et al. (2016), when the data are non-stationary. The conclusions of the chapter can be also directly extended to SC method, the generalized SC method and the further extensions of the PF method discussed above. Most of the literature on counterfactual analysis for panel data do not take into account the possibility of non-stationarity. One key exception is Bai et al. (2014) where the authors show, under some assumptions, consistency of the panel approach when the data are integrated of first order. However, the paper does not provide the asymptotic distribution of the estimator.

Both the PF and the ArCo (in its simplest form), construct the counterfactual for the treated variable of interest as a linear combination of untreated variables from the peers. The motivation is that there is some common dynamics between the treated unit and the members of the donor pool. We consider two very distinct scenarios: (i) The cointegrated case, where there is at least one cointegrated relation among the units and; (ii) The spurious case, where no integration relation exists. We show that in the first case we have a consistent, but not asymptotically normal, estimator for the different in the drifts before and after the intervention. We also considered under case (i) the possibly of working in first difference of the variable and in fact with a stationary process. It comes with no surprise that the methods can, in that specific case, be applied directly resulting in a consistent asymptotically normal estimator.

The troublesome scenario is case (ii) - the spurious case - where we demonstrate that the treatment effect estimator diverges. The lack of cointegration relation makes the construction of the artificial control using the pre-intervention period invalid, due to harmless effects from spurious regressions as discussed in Phillips (1986). As a consequence, one tends to reject the the hypothesis of no intervention effect too often when the true effect is null.

The chapter is organized as follows. Section 2.2 presents the setup considered in the chapter while Section 2.3 delivers the theoretical results. Section 2.5 concludes the chapter. Finally, all proofs are presented in the Appendix.

2.2 Setup and Estimators

2.2.1 Basic Setup

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

Let $\mathcal{D}_t$ be a binary variable flagging the periods after the intervention.

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

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

where

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

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

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

$$y_{1t}^{(1)} = \delta_t + y_{1t}^{(0)}; \quad t = T_0 \ldots, T, \tag{2-1}$$

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

The null hypothesis becomes

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

The quantity Δ in (3-1) is quite similar to the traditional *average treatment effect on the treated* (ATET) vastly discussed in the literature. It is clear that $y_t^{(0)}$ is not observed from T_0 onwards. For that reason, we call thereafter the *counterfactual*, i.e., what would y have been like had there been no intervention (potential outcome).

In order to construct the counterfactual let $\boldsymbol{y}_{0t} \equiv (y_{2t}, \ldots, y_{nt}')'$ be the collection of all untreated variables.[1] Panel based methods, such as the PF and ArCo methodologies, construct an artificial counterfactual by considering the following model in the absence of an intervention:

$$y_{1t}^{(0)} = \mathcal{M}(\boldsymbol{y}_{0t}) + \nu_t, \quad t = 1, \ldots, T, \tag{2-3}$$

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

The main idea is to estimate (2-3) using just the pre-intervention sample $(t = 1, \ldots, T_0 - 1)$, since in that case $y_{1t}^{(0)} = y_{1t}$. Consequently, the estimated counterfactual is given as:

$$\widehat{y}_{1t}^{(0)} = \widehat{\mathcal{M}}(\boldsymbol{y}_{0t}), \quad t = T_0, \ldots, T, \tag{2-4}$$

[1]We could also have included lags of the variables and/or exogenous regressors into $\boldsymbol{y}_{0t}$ but again to keep the argument simple, we have considered just contemporaneous variables; see Carvalho et al. (2016) for more general specifications.

where $\widehat{\mathcal{M}}(\cdot) \equiv \mathcal{M}(\cdot;\widehat{\boldsymbol{\theta}})$. Under some mild condition is possible to show that $\widehat{\delta}_t \equiv y_t - \widehat{y}_t^{(0)}$, for $t = T_0, \ldots, T$ is an unbiased estimator for $\delta_t, t = T_0, \ldots, T$ as the pre-intervention sample size grows to infinity. Also, under the assumption that the controls are untreated (Assumption 1.1) the average of $\widehat{\delta}_t$ over the post-intervention period:

$$\widehat{\Delta} = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \widehat{\delta}_t, \tag{2-5}$$

is consistent for the average (across time) treatment effect Δ_T and asymptotically normal as $T \to \infty$.

2.2.2 Non-stationarity

Let $\boldsymbol{y}_t^{(0)} \equiv (y_{1t}^{(0)}, \boldsymbol{y}_{0t}^{(0)})'$ denote all the units in the absence of the intervention. Under stationarity of $\boldsymbol{y}_t^{(0)}$ and additional mild assumptions, Hsiao et al. (2012) and Carvalho et al. (2016) show that (2-5) is $\sqrt{T}$-consistent for Δ and asymptotically normal. Suppose now that $\{\boldsymbol{y}_t^{(0)}\}$ is integrated process of order 1, $\mathcal{I}(1)$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and we assume for notational convenience that:[2]

$$\begin{cases} \boldsymbol{y}_t^{(0)} &= \boldsymbol{y}_{t-1}^{(0)} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t, \quad t \geq 1 \\ \boldsymbol{y}_0^{(0)} &= \boldsymbol{0}, \end{cases} \tag{2-6}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ is a drift and $\boldsymbol{\varepsilon_t}$ is a zero mean stationary process with a Wold Representation given by $\boldsymbol{C}(L)\boldsymbol{v}_t$. L denotes the lag operator, $\boldsymbol{C}(L)$ is a $(n \times n)$ matrix polynomial with $\boldsymbol{C}(0) = \boldsymbol{I}_n$ and all eigenvalues of the companion form are inside the unit circle, and $\boldsymbol{v}_t$ is a white noise vector such that

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

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

2.3 Theoretical Results

Before we present our main results let us establish some notation and definitions that we use throughout the rest of the chapter for clarity purposes

[2] Assume $\boldsymbol{y}_0^{(0)} = \boldsymbol{0}$ is without loss of generality. We could either assume $\boldsymbol{y}_0^{(0)}$ to be a any constant or even a random vector with a specific distribution.

2.3.0
Notation and Definitions

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

$$\begin{aligned}\boldsymbol{\Omega}_0(\boldsymbol{v}) &\equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathbb{E}(\boldsymbol{v}_t\boldsymbol{v}_t')\\ \boldsymbol{\Omega}_1(\boldsymbol{v}) &\equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\mathbb{E}(\boldsymbol{\eta}_s\boldsymbol{\eta}_t')\\ \boldsymbol{\Omega}(\boldsymbol{v}) &\equiv \boldsymbol{\Omega}_0(\boldsymbol{v})+\boldsymbol{\Omega}_1(\boldsymbol{v})+\boldsymbol{\Omega}_1(\boldsymbol{v})'\end{aligned}$$

if the limits exist. $\boldsymbol{W}(\cdot)$ denotes a vector Wiener process on $[0,1]^n$. Also for any given (random) matrix $\boldsymbol{M}\in\mathbb{R}^{n\times n}$ and (random) vector $\boldsymbol{m}\in\mathbb{R}^n$ we use the following partition scheme:

$$\boldsymbol{M} = \begin{array}{c c} & \begin{array}{cc} 1 & n-1 \end{array} \\ \begin{array}{c} 1 \\ n-1 \end{array} & \begin{pmatrix} \boldsymbol{M}_{11} & \boldsymbol{M}_{10} \\ \boldsymbol{M}_{01} & \boldsymbol{M}_{00} \end{pmatrix} \end{array} \qquad \boldsymbol{m} = \begin{array}{c} 1 \\ n-1 \end{array} \begin{pmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_0 \end{pmatrix}$$

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

On top of the statistical independence of the intervention with respect the the untreated units (Assumption 1.1), we consider the following key assumption:

Assumption 2.1 *Let $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ be a sequence of $(n\times 1)$ random vectors such that*

(a) $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ is zero mean weakly (covariance) stationary;

(b) $\mathbb{E}|z_{i1}|^{\xi} < \infty$ for $i = 1,\ldots,n$ and some $2 \le \xi < \infty$;

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

Assumption 2.1 state general conditions under which the multivariate invariance principle is valid for the process $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$. Assumption 2.1(a) limits the heterogeneity in the process (at least up to the second moment). Assumption 2.1(b) is just a standard higher moment existence condition for all the n coordinates of the random vector which guarantees, along with Assumption 2.1(c), bounded covariances. Finally, 2.1(c) restrains the temporal dependence requiring the sequence to be either strong mixing with size $-\frac{\xi}{\xi-2}$ or uniform missing with size $-\frac{\xi}{2\xi-2}$.

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

Proposition 2.1 *Let $\boldsymbol{S}_t = \sum_{j=1}^{t} \boldsymbol{z}_j$ be the partial sum of the sequence $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ of $(n \times 1)$ random vectors. Then, under Assumption 2.1,*

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

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

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

The implied convergence in Proposition 2.1(a) is a direct consequence of the stationarity assumption together with the mixing condition as shown by Ibragimov e Linnik (1971). Finally, Proposition 2.1(b) is a multivariate generalization of the univariate invariance principle Durlauf e Phillips (1985).

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

2.3.1 The Cointegrated Case

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

Therefore, the "cointegrated regression" can be written as

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

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

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

$$y_{1t} = \gamma_0 t + \boldsymbol{\beta}_0' \boldsymbol{y}_{0t} + \nu_t \tag{2-7}$$

$$y_{1t} = \alpha_0 + \boldsymbol{\pi}_0' \boldsymbol{y}_{0t} + \zeta_t \tag{2-8}$$

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

We can apply the results of the Lemma A.6 together with the continuous mapping theorem to show the following convergence in distribution:

Lemma 2.1 *Let the process* $\{\boldsymbol{y}_t^{(0)}\}$ *be defined by* (2-6) *have at least one cointegration relation* $(0 < r < n)$. *Also let* $\{\boldsymbol{\eta}_t \equiv (\nu_t, \boldsymbol{\varepsilon}_0')'\}$ *satisfies Assumption 2.1, then for the least squares estimator of the parameters appearing in* (2-7)–(2-8) *using only the pre intervention sample* $(t = 1, ...T_0)$ *as* $T \to \infty$*:*

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

$$T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{d} \boldsymbol{P}_{00}^{-1} \boldsymbol{Q}_{01} \equiv \boldsymbol{h}$$

$$T^{3/2}\left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} \tfrac{3}{\lambda_0^3} \left\{ \left[\boldsymbol{\Omega}^{1/2} \int_0^{\lambda_0} r \mathrm{d}\boldsymbol{W}(\mathrm{r}) \right]_1 - \boldsymbol{h}' \left[\boldsymbol{\Omega}^{1/2} \int_0^{\lambda_0} r \boldsymbol{W}(r) \mathrm{d}r \right]_0 \right\}$$

$$T\left(\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0\right) \xrightarrow{d} \boldsymbol{R}_{00}^{-1} \boldsymbol{V}_{01} \equiv \boldsymbol{p}$$

$$\sqrt{T}\left(\widehat{\alpha} - \alpha_0\right) \xrightarrow{d} \tfrac{1}{\lambda_0} \left\{ \left[\boldsymbol{\Omega}^{1/2} \int_0^{\lambda_0} \mathrm{d}\boldsymbol{W}(r) \right]_1 - \boldsymbol{p}' \left[\boldsymbol{\Omega}^{1/2} \int_0^{\lambda_0} \boldsymbol{W}(r) \mathrm{d}r \right]_0 \right\}.$$

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

$$\widehat{\pi} - \beta_0 \xrightarrow{p} \frac{\gamma_0}{\mu_0}$$

$$T_0^{-1}\left(\widehat{\alpha} - \alpha_0\right) \xrightarrow{p} 0$$

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

where the $(n \times n)$ random matrices are defined as:

$$\begin{aligned}
\boldsymbol{R}(\lambda) &\equiv \boldsymbol{\Omega}^{1/2}\left[\int_0^{\lambda}\boldsymbol{W}(r)\boldsymbol{W}'(r)\mathrm{d}r - \int_0^{\lambda}\boldsymbol{W}(r)\mathrm{d}r\int_0^{\lambda}\boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2}\\
\boldsymbol{P}(\lambda) &\equiv \boldsymbol{\Omega}^{1/2}\left[\int_0^{\lambda}\boldsymbol{W}(r)\boldsymbol{W}'(r)\mathrm{d}r - 3\int_0^{\lambda}r\boldsymbol{W}(r)\mathrm{d}r\int_0^{\lambda}r\boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2}\\
\boldsymbol{V}(\lambda) &\equiv \boldsymbol{\Omega}^{1/2}\left[\int_0^{\lambda}\boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r) - \int_0^{\lambda}\boldsymbol{W}(r)\mathrm{d}r\boldsymbol{W}'(1)\right]\boldsymbol{\Omega}^{1/2} + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_0\\
\boldsymbol{Q}(\lambda) &\equiv \boldsymbol{\Omega}^{1/2}\left[\int_0^{\lambda}\boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r) - \sqrt{3}\int_0^{\lambda}r\boldsymbol{W}(r)\mathrm{d}r\boldsymbol{W}'(1)\right]\boldsymbol{\Omega}^{1/2} + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_0,
\end{aligned}$$

with $\lambda = \lambda_0$ and $\boldsymbol{\Omega} \equiv \boldsymbol{\Omega}(\boldsymbol{\eta})$, $\boldsymbol{\Omega}_1 \equiv \boldsymbol{\Omega}_1(\boldsymbol{\eta})$, $\boldsymbol{\Omega}_0 \equiv \boldsymbol{\Omega}_0(\boldsymbol{\eta})$ as defined in Section 2.3.0.

Remark 2.1 *Whenever there is a drift among the peers and $n > 2$ we have a multicollinearity issue in the least squares estimators, since the drift component dominates the other terms asymptotically. In case of specification* (2-7)*, since we are fitting the trend term $t\gamma$, the multicollinearity appears even for $n = 2$ (only one control). Note that, for the specification* (2-8)*, if we replace γ_0 by its definition $\mu_1 - \beta_0\mu_0$, then as expected $\widehat{\pi} \xrightarrow{p} \frac{\mu_1}{\mu_0}$.*

Remark 2.2 *In fact the estimators* (2-5) *is of little usage whenever we expect to have integrated process with drift. Not only the estimator is not well in large samples, but a simple fitted trend regressor makes a reasonable counterfactual for the unit of interest. Therefore we treat for now on only the the case without drift* ($\boldsymbol{\mu} = \mathbf{0}$).

Similar results to Lemma 2.1(a) appear in Durlauf and Phillips (1985) for instance where the estimator for the non deterministic regressor is super-consistent.

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

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

where $\widehat{\gamma}, \widehat{\boldsymbol{\beta}}, \widehat{\alpha}$ and $\widehat{\boldsymbol{\pi}}$ are the least squares estimators of the parameters appearing in (2-7)–(2-8) using only pre-intervention sample.

Teorema 2.2 *Let the process* $\{\boldsymbol{y}_t^{(0)}\}$ *be defined by* (2-6) *have at least one cointegration relation* $(0 < r < n)$*. Also let* $\{\boldsymbol{\eta}_t \equiv (\nu_t, \boldsymbol{\varepsilon}_0')'\}$ *satisfies the Assumption 2.1, then for the estimators defined in* (2-9) *as* $T \to \infty$*:*

$$\sqrt{T}\left(\widehat{\Delta}_1 - \Delta\right) \xrightarrow{d} \boldsymbol{c}_1 - \boldsymbol{h}'\boldsymbol{d}_0,$$
$$\sqrt{T}\left(\widehat{\Delta}_2 - \Delta\right) \xrightarrow{d} \boldsymbol{a}_1 - \boldsymbol{p}'\boldsymbol{b}_0,$$

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

$$\boldsymbol{a}(\lambda) \equiv \boldsymbol{\Omega}^{1/2}\left[\tfrac{1}{1-\lambda}\int_\lambda^1 \mathrm{d}\boldsymbol{W} - \tfrac{1}{\lambda}\int_0^\lambda \mathrm{d}\boldsymbol{W}\right]$$
$$\boldsymbol{b}(\lambda) \equiv \boldsymbol{\Omega}^{1/2}\left[\tfrac{1}{1-\lambda}\int_\lambda^1 \boldsymbol{W}(r)\mathrm{d}r - \tfrac{1}{\lambda}\int_0^\lambda \boldsymbol{W}(r)\mathrm{d}r\right]$$
$$\boldsymbol{c}(\lambda) \equiv \boldsymbol{\Omega}^{1/2}\left[\tfrac{1}{1-\lambda}\int_\lambda^1 \mathrm{d}\boldsymbol{W} - \tfrac{3(1+\lambda)}{2\lambda^3}\int_0^\lambda r\mathrm{d}\boldsymbol{W}\right]$$
$$\boldsymbol{d}(\lambda) \equiv \boldsymbol{\Omega}^{1/2}\left[\tfrac{1}{1-\lambda}\int_\lambda^1 \boldsymbol{W}(r)\mathrm{d}r - \tfrac{3(1+\lambda)}{2\lambda^3}\int_0^\lambda r\boldsymbol{W}(r)\mathrm{d}r\right],$$

with $\lambda = \lambda_0$ *and* $\boldsymbol{\Omega} \equiv \boldsymbol{\Omega}(\boldsymbol{\eta})$ *as defined in Section 2.3.0.*

Therefore both estimators above are $\sqrt{T}$-consistent for Δ, however with a non-standard limiting distribution. Notice the first term in the limiting distribution of the second specification is in fact the same distribution that appears in Carvalho et al. (2016) for the stationary case. Even though the results above rule out common inference procedures, in Section 2.4 we investigate the results of using a conventional t-stat.

2.3.2 The Spurious Case

We now turn to the case where no cointegration relation exists among $\boldsymbol{y}_t$ prior to the intervention, hence $\boldsymbol{C}(1)$ is full rank. We consider for the pre-intervention period the same specification, (2-7) and (2-8), that were used in the cointegrated case. However, since the "true parameters" no longer exist[3], we cannot express least-squares estimators as diferent form their "true parameters". Hence we have the following result:

Lemma 2.2 *Let the process* $\{\boldsymbol{y}_t^{(0)}\}$ *be defined by* (2-6) *have no cointegration relation* $(r = 0)$*. Also let* $\{\boldsymbol{\varepsilon}_t\}$ *satisfies Assumption 2.1, then for the least squares estimator of the parameters appearing in* (2-7)–(2-8) *as* $T_0 \to \infty$*:*

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

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

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &\xrightarrow{d} \boldsymbol{P}_{00}^{-1}\boldsymbol{P}_{01} \equiv \boldsymbol{f}, \\
\sqrt{T}\widehat{\gamma} &\xrightarrow{d} \tfrac{3}{\lambda_0^3}\left\{\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} r\boldsymbol{W}(r)\mathrm{d}r\right]_1 - \boldsymbol{f}'\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} r\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\}, \\
\widehat{\boldsymbol{\pi}} &\xrightarrow{d} \boldsymbol{R}_{00}^{-1}\boldsymbol{R}_{01} \equiv \boldsymbol{g}, \\
\tfrac{1}{\sqrt{T}}\widehat{\alpha} &\xrightarrow{d} \tfrac{1}{\lambda_0}\left\{\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} \boldsymbol{W}(r)\mathrm{d}r\right]_1 - \boldsymbol{g}'\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} \boldsymbol{W}(r)\mathrm{d}r\right]_0\right\}.
\end{aligned}
$$

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

$$
\begin{aligned}
\widehat{\beta} &\xrightarrow{p} \frac{\mu_1}{\mu_0}, \\
\widehat{\gamma} &\xrightarrow{p} 0, \\
,\widehat{\pi} &\xrightarrow{p} \frac{\mu_1}{\mu_0}, \\
\tfrac{1}{T}\widehat{\alpha} &\xrightarrow{p} 0.
\end{aligned}
$$

In case of $\boldsymbol{\mu} \neq 0$ *and* $n > 2$ *the least squares estimators are not defined asymptotically.*

where the $(n \times n)$ *random matrices* $\boldsymbol{P}(\lambda_0), \boldsymbol{R}(\lambda_0)$ *are defined in Lemma 2.1 but with* $\boldsymbol{\Omega} \equiv \boldsymbol{\Omega}(\boldsymbol{\varepsilon})$.

The limiting distribution of $\widehat{\boldsymbol{\pi}}$ and $\widehat{\boldsymbol{\alpha}}$ are well known from the spurious regression case discussed in Phillips (1986). For $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$, the result is analogous but with a different limiting distribution. In both cases, when $r = 0$ and consequently $\boldsymbol{y}_t$ does *not* cointegrate, we have a spurious regression and both $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\pi}}$ converges, as $T_0 \to \infty$, not to a constant but to a functional of a multivariate Brownian motion. While $\widehat{\boldsymbol{\alpha}}$ diverges, $\widehat{\boldsymbol{\gamma}}$ converges to zero (which is the value of the parameter γ_0 when $\boldsymbol{\mu} = 0$).

Once again we consider the scenario where the researcher conduct the estimation using the estimators defined in (2-9) with $\boldsymbol{y}_t$ in levels.

Teorema 2.3 *Let the process* $\{\boldsymbol{y}_t^{(0)}\}$ *be defined by* (2-6) *have no cointegration relation* $(r = 0)$. *Also let* $\{\boldsymbol{\varepsilon}_t\}$ *satisfies Assumption 2.1, then for the estimators defined in* (2-9) *as* $T \to \infty$:

$$
\begin{aligned}
\tfrac{1}{\sqrt{T}}\left(\widehat{\Delta}_1 - \Delta\right) &\xrightarrow{d} \widetilde{\boldsymbol{f}}'\boldsymbol{d}, \\
\tfrac{1}{\sqrt{T}}\left(\widehat{\Delta}_2 - \Delta\right) &\xrightarrow{d} \widetilde{\boldsymbol{g}}'\boldsymbol{b},
\end{aligned}
$$

where $\widetilde{\boldsymbol{f}} \equiv (1, -\boldsymbol{f}')'$, $\widetilde{\boldsymbol{g}} \equiv (1, -\boldsymbol{g}')'$ *and the* $(n \times 1)$ *random vectors* $\boldsymbol{b}$ *and* $\boldsymbol{d}$ *are defined in Lemma 2.1 but with* $\boldsymbol{\Omega} \equiv \boldsymbol{\Omega}(\boldsymbol{\varepsilon})$.

From the theorem above, it is clear that, unlike in the cointegrated case, $\widehat{\Delta}_j$ diverges as $T \to \infty$ for both specifications. As for the cointegration case we investigate the limiting distribution of a conventional t-statistic in Section 2.4.

2.4 Inference

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

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

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

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

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

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

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

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

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

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

2.4.1 Inference on the Cointegrated Case

Consider now the following stronger version of Assumption 2.1.

Assumption 2.2 *Let $\{\boldsymbol{z}_t\}_{t=1}^{\infty}$ be a sequence of random vectors $(n \times 1)$ such that*

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

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

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

Clearly, Assumption 2.2 implies Assumption 2.1. The fourth order stationarity requirement on $\{\nu_t\}$ translates into weak stationarity of $\{w_t^{(k)} \equiv \nu_t \nu_{t+k}\}$ for any $k \in \mathbb{Z}$. Assumptions 2.2(a)-(c) are sufficient for Assumption A of Andrews (1991) which translate in the summability of the covariances of $w_t^{(k)}$, i.e.

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

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

Lemma 2.3 *If the sequence $\{\nu_t\}$ satisfies Assumption 2.2, then for each $j \in \{1,2\}$,*

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

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

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

Lemma 2.3(b) follows from arguments similar to Newey e West (1987) and Andrews (1991).

Teorema 2.4 *Under the same conditions of Theorem 2.2, but with Assumption 2.1 replaced by 2.2:*

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

$$\tau_1 \xrightarrow{d} \tfrac{\sqrt{1-\lambda_0}}{\omega}\left(\boldsymbol{c}_1 - \boldsymbol{h}'\boldsymbol{d}_0\right)$$
$$\tau_2 \xrightarrow{d} \tfrac{\sqrt{1-\lambda_0}}{\omega}\left(\boldsymbol{a}_1 - \boldsymbol{p}'\boldsymbol{b}_0\right)$$

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

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

where $\omega^2 \equiv \mathbf{\Omega}_{11}$.

Remark 2.3 *Under $\mathcal{H}_0$ we have a $\sqrt{T}$-consistent estimator for the intervention average effect Δ_T albeit with a non-standard asymptotic distribution. In fact by the presence of the second term we can conclude that we systematically over reject asymptotically.*

Remark 2.4 *The "t-test" is also asymptotically consistent as the test statistic diverges under the alternative. Recall that our null hypothesis was defined in* (3-1)*, hence the natural alternative would be $\Delta_T \neq 0$, but since Δ_T could potentially approach zero arbitrally fast as T grows, we restrict the Δ_T to be a non-zero constant. We get similar results by allowing a more flezible intervention profile as long as it does not approach zero faster than $T^{-1/2}$,for instance, by imposing only that $\{\delta_t\}_t$ is such that $\sqrt{T}\Delta_T \to \infty$.*

2.4.2 Inference on the Spurious case

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

Lemma 2.4 *Consider the same conditions of Theorem 2.3, but with Assumption 2.1 replaced by 2.2, then under both $\mathcal{H}_0$ or $\mathcal{H}_1$ as $T \to \infty$:*

(a) $\frac{1}{T}\widehat{\rho}^2_{1k} \xrightarrow{d} \frac{1}{1-\lambda_0}\widetilde{\boldsymbol{f}}'\boldsymbol{L}\widetilde{\boldsymbol{f}}, \quad \forall k$

(b) $\frac{1}{T}\widehat{\rho}^2_{2k} \xrightarrow{d} \frac{1}{1-\lambda_0}\widetilde{\boldsymbol{g}}'\boldsymbol{H}\widetilde{\boldsymbol{g}}, \quad \forall k.$

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

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

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

for $j \in \{1,2\}$*, where*

$$\begin{aligned}
\boldsymbol{H} &\equiv \boldsymbol{\Omega}^{1/2}\left[\int_{\lambda_0}^{1}\boldsymbol{W}(r)\boldsymbol{W}(r)'\mathrm{d}r - \tfrac{1}{1-\lambda_0}\int_{\lambda_0}^{1}\boldsymbol{W}(r)\mathrm{d}r\int_{\lambda_0}^{1}\boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2}\\
\boldsymbol{L} &\equiv \boldsymbol{H} - 2\left[\boldsymbol{k} - \left(\tfrac{1-\lambda_0^3}{3} - \tfrac{(1-\lambda_0)^3}{4}\right)\boldsymbol{j}\right]\boldsymbol{j}'\\
\boldsymbol{j} &\equiv 3\boldsymbol{\Omega}^{1/2}\int_{0}^{\lambda_0} r\boldsymbol{W}(r)\mathrm{d}r\\
\boldsymbol{k} &\equiv \boldsymbol{\Omega}^{1/2}\int_{0}^{\lambda_0} r\boldsymbol{W}(r)\mathrm{d}r\\
c_\phi &\equiv \int_{-\infty}^{\infty}\phi(x)\mathrm{d}x
\end{aligned}$$

Notice that the limiting distribution in (a) and (b) above is independent of k. In fact, it is the same distribution derived in Lemma 1 when we consider $k = 0$. It follows from the fact that the additional term $\sum_{t=1}^{T}\boldsymbol{v}_t\sum_{i=1}^{k}\boldsymbol{\varepsilon}_i'$ is $O_P(T)$. Result (b) for $k = 0$ is similar to the one appering in Phillips (1986). It turns out it is valid for all fixed k and also for specification (2-7) albeit with a different limiting distribution. Using a HAC covariance estimator as proposed by Newey e West (1987) and Andrews (1991), we have an even weaker convergence rate as it goes from T^{-1} to $(J_T T)^{-1}$ as stated in Lemma 6(c)-(d).

Now combining Theorem 2.3 with Lemma 2.4 together with the continuous mapping theorem we have the following result.

Teorema 2.5 *If the process* $\{\boldsymbol{\varepsilon}_t\}$ *satisfies the assumption of Proposition* **??**, *then as* $T \to \infty$*, the estimators defined in* (2-9)*. Under both* $\mathcal{H}_0 : \Delta_T = 0$ *and* $\mathcal{H}_1 = \delta \neq 0$*.*

$$\begin{aligned}
\sqrt{\frac{J_T}{T}}\tau_1 &\xrightarrow{d} \frac{1-\lambda_0}{\sqrt{c_\phi}}\frac{\widetilde{\boldsymbol{f}}'\boldsymbol{d}}{\sqrt{\widetilde{\boldsymbol{f}}'\boldsymbol{L}\widetilde{\boldsymbol{f}}}}\\
\sqrt{\frac{J_T}{T}}\tau_2 &\xrightarrow{d} \frac{1-\lambda_0}{\sqrt{c_\phi}}\frac{\widetilde{\boldsymbol{g}}'\boldsymbol{b}}{\sqrt{\widetilde{\boldsymbol{g}}'\boldsymbol{H}\widetilde{\boldsymbol{g}}}}.
\end{aligned}$$

Remark 2.5 *When conducting a t-test one draws inference on the premisses that* $\tau_j \xrightarrow{d} \mathcal{N}(0,1)$ *under* $\mathcal{H}_0$*. However, as Theorem 2.5 shows,* τ_j *actually diverges under the assumption that* $J_T = o(T^{1/2})$*. Therefore, ignoring the non-stationarity of the data we end up rejecting the null hypothesis too often in finite sample. In fact, as the sample size increases, the probability of rejection the null approaches 1 regarless of the existence of the treatment.*

Remark 2.6 *Notice that the result above is not dependent on the choice of the variance estimator bandwidth. If we use simple variance estimator such as $\widehat{\sigma}_{jT} = \widehat{\rho}_{j0}$ (for the case of iid data), we still have $\tau_j = O_P(\sqrt{T})$. In fact, in this particular case, the t-test diverges in a even faster rate.*

Still under the $\mathcal{H}_0$, but with $\boldsymbol{\mu}_0 \neq 0$, the estimator $\widehat{\Delta}_j$ is not defined asymptotically unless $n = 2$. Even when that is the case, the variance estimator now converges to zero as per (e) and (f) of Lemma 2.4. Consequently the t-stat is not properly defined asymptotically. Thus, as in the cointegrated scenario, the case with drift is of little theoretical insight even for the spurious regression.

Under $\mathcal{H}_1$, but still with $\boldsymbol{\mu}_0 = 0$, the estimator $\widehat{\Delta}_j$ is well defined (even asymptotically) for any n, however, as in the previous case, the variance estimator converges to zero . Nevertheless, in finite sample, we tend to get larger values for τ_j as the sample size increases and truly rejecting the null when its false. For the case where $\boldsymbol{\mu}_0 \neq 0$ once again the t-stat is not properly defined asymptotically.

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

2.4.3 First-Difference

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

$$\boldsymbol{z}_t = \boldsymbol{\mu} + \Delta\mu \boldsymbol{d}_t + \boldsymbol{\varepsilon}_t$$

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

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

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

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

From Theorem 1.3 for the particular case of low dimensional linear specification with $q = 1$ we have:

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

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

Remark 2.7 *The approach above also give us $\sqrt{T}$-consistent estimator for the difference in drifts. However, in contrast to the cointegrated estimator, it is asymptotically normal hence more practical for conducting inference.*

Remark 2.8 *The limiting distribution in first difference is independent of both the prior knowledge of the true values of $\boldsymbol{\mu}$ and the true hypothesis ($\mathcal{H}_0$ or $\mathcal{H}_1$).*

Remark 2.9 *Working in first difference we avoid a true spurious regression since if the integrated process is truly uncorrelated we will end up having $\widehat{\boldsymbol{\theta}} \approx \mathbf{0}$ for the pre-intervention period.*

2.5 Conclusions

In this chapter we consider the asymptotic properties of intervention effects estimators based on the construction of an artificial counterfactual from linear panel data models. The results in the chapter either show that the estimators diverge or have non-standard asymptotic distributions. The main prescription of the chapter is that practitioners should work in first-differences when the data are non-stationary.

3
Conditional Quantile Counterfactual Analysis

3.1
Introduction

In this chapter we propose a new method to carry out counterfactual analysis to evaluate the impact of interventions on the distribution of variables of interest. Our approach is specially useful in situations where there is a single "treated" unit and no available "controls". The goal of the proposed method is the construction of an artificial counterfactual based on observed data from a pool of "untreated" peers. Our approach is a generalization of the work of Hsiao et al. (2012) and Carvalho et al. (2016).

Causality is a major topic of empirical research in Economics. Usually, causal statements with respect of the adoption of a given treatment (intervention) rely on the construction counterfactuals based on the outcomes from a group of individuals not affected by the treatment. Notwithstanding, definitive cause-and-effect statements are usually hard to formulate given the constraints that economists face in finding sources of exogenous variation. However, in micro-econometrics there has been major advances in the literature and the estimation of treatment effects is part of the toolbox of applied economists; see, for example, Angrist et al. (1996), Angrist e Imbens (1994), Heckman e Vytlacil (2005), Belloni et al. (2014), and Belloni et al. (2016). Furthermore, in recent years there has been significant contributions to the estimation of quantile treatment effects when a control group is readily available. See, for example, Abadie et al. (2002) and Firpo (2007) for a low dimensional set up and Chernozhukov e Hansen (2005), Chernozhukov e Hansen (2006), Chernozhukov e Hansen (2008), Chernozhukov et al. (2014) for high dimensional one.

On the other hand, when there is not a natural control group which is usually the case when handling aggregated (macro) data, the econometric tools have evolved in a much slower pace and much of the work has focused on simulating counterfactuals from structural models. However, in recent years, some authors have proposed new techniques inspired partially by the developments in micro-econometrics that are able, under some assumptions, to conduct counterfactual analysis with aggregate (macro) data. Hsiao et al. (2012) put forward a simple panel data method to estimate counterfactuals and studied the impact of economic and political integration of Hong Kong with mainland China on Hong Kong's economy. Zhang et al. (2014) applied the same techniques of Hsiao et al. (2012) to evaluate the impact of Canada-US Free Trade Agreement (FTA) on Canada's GDP, labour productivity and unemployment. Abadie e Gardeazabal (2003) used the SC method to investigate the effects of terrorism on the GDP of the Basque Country while Abadie et al. (2010) and Abadie et al. (2014) applied the the same techniques to measure, respectively, the effects on consumption of a large-scale tobacco control program in California and the economic impact of the 1990 German reunification in West Germany. Pesaran et al. (2007) and Dubois et al. (2009) used the Global Vector Autoregressive (GVAR) framework developed by Pesaran et al. (2004) and Dees et al. (2007) to study the effects of the launching of the Euro. Pesaran e Smith (2012) studied the effects of the quantitative easing (QE) in the United Kingdom with a new methodology partly inspired by the GVAR methods. Finally, Angrist et al. (2013) considered a new semi-parametric method to measure the effects of monetary policy interventions on macroeconomic aggregates. However, none of the above papers considered the case of quantile treatment effects for dynamic data when there is no control group available.

The goal of this chapter is to extend the methodology put forward by Carvalho et al. (2016) by considering the estimation of quantile counterfactuals. We derive an asymptotically normal test statistics for the quantile intervention effect. Our procedure is illustrated in a detailed simulation experiment as well as in an empirical application in Corporate Finance.

The chapter is organized as follows. Section 3.2 presents the estimator and the conditional quantile model. The asymptotic theory is derived in derived in Section 3.3 while inference is considered in Section 3.4. The effects of misspecification is discussed in Section 3.4.1. Section 3.5 shows the Monte Carlo simulations. The empirical illustration is described in Section 3.6. Finally, Section concludes de chapter. All proofs are relegated to the appendix.

3.2 The Estimator

3.2.1 Definitions

Suppose we have n units (countries, states, municipalities, firms, etc) indexed by $i = 1, \ldots, n$. For each unit and for every time period $t = 1, \ldots, T$, we observe a realisation a random variable Z_{it} defined on $(\Omega, \mathcal{F}, P)$

Furthermore we consider that there is *only one* unit that suffers the intervention (treatment) at time $T_0 = \lfloor \lambda_0 T \rfloor$, where $\lambda_0 \in (0,1)$. We assume, without loss of generality, to be the unit one ($i = 1$) and we denote the unit of interest $Y_t \equiv Z_{1t}$. Let D_t be a binary variable flagging the periods when the intervention was in place, then we can express the observable variables of unit of interest as

$$Y_t = D_t Y_t^{(1)} + (1 - D_t) Y_t^{(0)}; \qquad D_t = \begin{cases} 1 & \text{if } t \geq T_0 \\ 0 & \text{otherwise} \end{cases}$$

where, following the literature on treatment effects, $Y_t^{(1)}$ denotes the outcome when the unit i is exposed to the intervention and $Y_t^{(0)}$ when it is not.

The remaning $n - 1$ unit (peers) are potential controls denoted by $\boldsymbol{X}_t \equiv (Z_{2t}, \ldots, Z_{nt})'$. We treat the peers as untreated, i.e., the intervention had no effect on them formally we require that D_t is independent of $\boldsymbol{X}_t$ for all t, which is implied by Assumption 1.1. Once again, Ii is important to not that we do not necessarily require D_t to be independent of Y_t (the unit of interest) only of $\boldsymbol{X}_t$ (the peers) . Since we are only interested in the treatment effect on the treated it is a well known fact, from the treatment effect literature that we can consistently estimate the average effect even when $\mathbb{E}(Y_t | D_t) \neq 0$

We are ultimately interested in the potential effects of this intervention in the unit of interest. Formally defined for the post-intervention period as

$$\Delta_t \equiv Y_t^{(1)} - Y_t^{(0)}; \quad t = T_0, \ldots, T \tag{3-1}$$

Clearly we do not observe $Y_t^{(0)}$ after $T_0 - 1$, for that reason we call thereafter the *counterfactual*, i.e., what would Y_t have been like had there been no intervention (potential outcome). Notice that the intervention effect Δ_t by definition is a random variable possibly with with a time varying distribution (non-stationary). We return to this discussion in subsection 3.2.2.

We construct a proxy variable for $Y_t^{(0)}$ based on the Artificial Counterfactual (ArCo) method by exploiting the relation among the the unit before the intervention. Consider the following data generating process (DGP)

Assumption 3.1 *For each unit $i = 1, \ldots, n$*

$$\begin{aligned} Z_{it}^{(0)} &= \boldsymbol{\Psi}_{\infty,i}(L) \varepsilon_{it} \\ \varepsilon_{it} &= \boldsymbol{\Lambda}_i \boldsymbol{f}_t + \eta_{it} \\ \boldsymbol{f}_t &\sim (\boldsymbol{\mu}_t, \boldsymbol{Q}) \end{aligned}$$

where $\boldsymbol{f}_t (f \times 1)$ is a vector of common unobserved factors such that is serially uncorrelated, with deterministic time trend $\boldsymbol{\mu}_t$ and covariance structure

$\boldsymbol{Q}(f \times f)$. *$\boldsymbol{\Lambda}_i(1 \times f)$ are vectors of factor loadings. The idiosyncratic error term $\eta_{it} \sim (0, \omega_i)$ is also considered serially uncorrelated. Additionally, $\mathbb{E}(\eta_{it}\boldsymbol{f}_j) = \boldsymbol{0}$, $\forall i, t, j$. Finally, L is the lag operator and the polynomial $\Psi_{\infty,i}(L) = (1 + \psi_{1i}L + \psi_{2i}L^2 + \cdots)$ is such that $\sum_{j=0}^{\infty} \psi_{ji}^2 < \infty$ for all i.*

The GDP described by Assumption 3.1 is quite flexible. It translate into each unit being modelled as a determistic idiosyncratic time trend plus a zero-mean weakly dependent stationary (ARMA) process as in

$$Z_{it} = \mu_{it} + \zeta_{it}$$

However, both the time trends and the error terms are linked due to the common factor structure of the DGP. So even though Z_{it} is allowed to be *not* identically distributed (non-stationary) common regression techniques would not result in spurious results.

3.2.2
Conditional Quantile Model

First let's considere a supposedly more direct approach and test for the difference in the distribution of Y_t before and after the intervention. Let F_0 and F_1 the marginal distribution of $Y_t^{(0)}$ and $Y_t^{(1)}$ respectively. Then we could use its empirical distribution function (EDF) to perform a distributional test using some metric defined over $\widehat{F}_0 - \widehat{F}_1$. As consequence of the determinist (but unknown) time trend this simple procedure would mistakenly indicate the presence of a intervention effect whenever a time trend is present. Obviously detrending (as is it common practice in time series analysis) would be naive if we would like to test, for instance, the a intervention effect on the trend itself.

The same problem would occur in the case we would like to test for any unconditional quantile difference before and after the intervention. Any unconditional analysis attempt is bound to suffer from bias specially if the time trend dominates the stochastic term which is usually the case in practice. To circumvent this issue we exploit the the information contained in the peers to conduct a conditional analysis. In particular we focus on conditional quantiles, heuristically we measure the treatment effect by the potential differences that it may cause in the quantiles of the conditional distribution of $Y|\boldsymbol{X}$. In other words, we test for the stability of the distribution function of $Y|\boldsymbol{X}$, which under the hypothesis that the peers are untreated might arguably be caused by the treatment effect on the unit of interest.

Some notation: for the random variable $Y_t^{(0)}|\boldsymbol{X}_t$ let $F_{Y|X}(y|\boldsymbol{x}) = \mathbb{P}(Y_t^{(0)} \leq y|\boldsymbol{X}_t = \boldsymbol{x})$ be the conditional distribution function. Hence we define for a given

$\tau \in [0,1]$ the conditional quantile function (CQF) as[1]

$$Q_\tau(\boldsymbol{x}) = \inf\{y : F_{Y|X}(y|\boldsymbol{x}) \geq \tau\}$$

It can be shown that the CQF (if exists) is the solution to the following minimizaton problem

$$Q_\tau(\boldsymbol{x}) \in \operatorname*{arg\,min}_{Q \in \mathbb{Q}} \mathbb{E}\left[\rho_\tau(Y - Q(\boldsymbol{x})\right]$$

where $\rho_\tau(z) = z(\tau - 1\{z < 0\})$ is know as the check function

Assumption 3.2 *For each $\tau \in [0,1]$*

$$Q_\tau(\boldsymbol{x}) = g_\tau\left(\boldsymbol{x}, \boldsymbol{\theta}_0(\tau)\right)$$

where $g_\tau(\cdot, \boldsymbol{\theta}_0(\tau))) : \mathbb{R}^{n-1} \mapsto \mathbb{R}$ for a unique $\boldsymbol{\theta}_0(\tau) \in \Theta_\tau \subset \mathbb{R}^{p_\tau}$

The assumption above postulate a correctly specified parametric model for $Q_\tau(\boldsymbol{x})$. Failure to this hypothesis (mispecification) are treated in a section below. In the most flexible setup we allow to both the the functional form and the true parameters to vary with τ, however one can get a much more parsimonious model by setting both the same across the quantiles. An even simpler solution is a linear specification such as $g(\boldsymbol{x}, \boldsymbol{\theta}_0) = \boldsymbol{x}'\boldsymbol{\theta}_0$.

We can define the $\tau-$quantile error by $\nu_t(\tau) \equiv Y_t^{(0)} - g(\boldsymbol{X}_t, \boldsymbol{\theta}_0(\tau))$ and rewrite the model in the conventional error format as

$$Y_t^{(0)} = g(\boldsymbol{X}_t, \boldsymbol{\theta}_0(\tau)) + \nu_t(\tau); \qquad \mathbb{P}\left(\nu_t(\tau) \leq 0\right) = \tau$$

It can be shown that the parameter $\boldsymbol{\theta}_0$ is a solution to the following minimizaiton problem

$$\boldsymbol{\theta}_0(\tau) \in \operatorname*{arg\,min}_{\theta \in \Theta} \mathbb{E}\left[\rho_\tau(Y_t^{(0)} - g(\boldsymbol{X}_t, \boldsymbol{\theta}(\tau))\right]$$

hence, using the pre-intervention sample $\{y_t, \boldsymbol{x}_t\}_{t=1}^{T_0-1}$ we can estimate $\boldsymbol{\theta}_0$ solving the sample counterpart of the minimisation above

$$\widehat{\boldsymbol{\theta}}(\tau) = \operatorname*{arg\,min}_{\theta \in \Theta} \frac{1}{T} \sum_{t=T_0}^{T} \left[\rho_\tau(y_t - g(\boldsymbol{x}_t, \boldsymbol{\theta}(\tau))\right]$$

Therefore we define the conditional quantiles ArCo estimator by

[1]this definition is necessary to avoid $Q_\tau(\cdot)$ not to be unique for a given τ, which happen whenever $F_{Y|X}$ has flat regions. If $F_{Y|X}$ is a strictly increasing CDF then $Q_\tau(\boldsymbol{x}) = F_{Y|X}^{-1}(\tau)$

$$\widehat{Y}_t^{(0)}(\tau) = g(\boldsymbol{X}, \widehat{\boldsymbol{\theta}}(\tau)); \qquad t = T_0, \ldots, T \tag{3-2}$$

Also we can define for each $\tau \in [0, 1]$ the intervention effect estimator as

$$\widehat{m}_t(\tau) = Y_t - \widehat{Y}_t^{(0)}(\tau); \qquad t = T_0, \ldots, T \tag{3-3}$$

For completeness we now reproduce from Koenker (2005) some well known condition to ensure the both the consistency and the asymptotic normality of $\widehat{\boldsymbol{\theta}}(\tau)$

Assumption 3.3 *The distribution functions $F_t(y) = \mathbb{P}(Y_t \leq y|X_t = x_t)$ are*

(a) absolutely continuos

(b) with continuos density f_t uniformily bounded away from 0 and ∞ at the points $F_t^{-1}(\tau)$ for $t = 1, 2, \ldots$

Assumption 3.4 *There exist positive definite matrices A and $B(\tau)$ such that*

(a) $\lim_{T\to\infty} \sum_{t=1}^{T} \nabla g_t \nabla g_t' = A$

(b) $\lim_{T\to\infty} \sum_{t=1}^{T} f_t \nabla g_t \nabla g_t' = B(\tau)$

(c) $\max_{t=1,\ldots,T} \|\nabla g_t\| \to 0$

where $\nabla g_t = \frac{\partial g(x_t,\theta)}{\partial \theta}|_{\theta=\theta_0}$ *and* $f_t = f(g(x_t, \theta_0))$

3.3 Asymptotics

Instead of using directly the empirical quantile of $\{\widehat{\Delta}_t(\tau)\}_{t\geq T_0}$ as the basis of our inference procedure to test potential difference in the quantiles after the intervention, it will be proven more convenient to rely on the equivalent result

$$\mathbb{P}(Y_t^{(1)} - g_\tau(\boldsymbol{X}, \boldsymbol{\theta}_0(\tau)) - \Delta_t \leq 0) = \mathbb{E}1\{v_t(\tau) \leq 0\} = \tau$$

Hence we replace $m_t(\tau) \equiv Y_t^{(1)} - g_\tau(\boldsymbol{X}, \boldsymbol{\theta}_0(\tau))$ with its estimator $\widehat{m}_t(\tau)$ defined in (3-3) and use the empirical distribution function (EDF) of $\widehat{m}_t(\tau) - \Delta_t$ evaluated at zero as our estimator

$$\widehat{\tau}_T(\tau) = \tfrac{1}{T-T0+1} \sum_{t=T_0}^{T} 1\{\widehat{m}_t(\tau) - \Delta_t \leq 0\} \tag{3-4}$$

, which allow us to estimate the asymptotic variance without having to estimate the density of $v_t(\tau)$. Ignoring (for now) the sample variance of $\widehat{\boldsymbol{\theta}}(\tau)$, that would be the average of dependent (dependence structure imposed by $\boldsymbol{\Psi}(L)$) Bernoulli trial with probability of success equal to τ under $\mathcal{H}_0$.

Let $w_t(\tau) = 1\{v_t(\tau) \leq 0\} - \tau$, under the null and Assumption 3.1, $\{w_t(\tau)\}_t$ is a stationary process with the j-covariance denoted by $\gamma_j(\tau) \equiv \mathbb{E}(w_t(\tau), w_{t+|j|}(\tau)) = \mathbb{P}(\Delta_t \leq 0, \Delta_{t+|j|} \leq 0)$. From the Bernoulli trial variance we get $\gamma_0(\tau) = \tau(1-\tau)$. The j-correlation is denoted by $\rho_j(\tau) \equiv \gamma_j(\tau)/\gamma_0(\tau)$ and let $\phi(\tau) = \sum_{j=1}^{\infty} 2\rho_j(\tau)$, which is finite by Assumption 3.1. Hence, taking into account the uncertainty on the estimation of $\boldsymbol{\theta}_0$ during the pre intervention period, we have

Teorema 3.1 *For any $\tau \in (0,1)$, let $W_T(\tau) \equiv \sqrt{T\lambda_0(1-\lambda_0)}(\widehat{\tau}_T(\tau) - \tau)$. Under Assumptions 1.1-3.4:*

$$W_T(\tau) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(\tau)\right)$$

where $\mathcal{N}(\mu, \omega^2)$ denotes the normal distribution with mean μ and variance ω^2; and $\sigma^2(\tau) = \tau(1-\tau)(1+\phi(\tau))$.

Since the above theorem is valid for any $\tau \in (0,1)$ and we can any finite set $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)'$ and apply the Cramer-Wold device to derive the multivariate version of Theorem 3.1

Corolário 3.2 *Let $\boldsymbol{W}_T(\boldsymbol{\tau}) = (W_T(\tau_1), \ldots, W_T(\tau_k))'$ for $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)' \in (0,1)^k$ and $k \leq \infty$. Under Assumptions 1.1-3.4:*

$$\boldsymbol{W}_T(\boldsymbol{\tau}) \xrightarrow{d} \mathcal{N}_k\left(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\tau})\right)$$

where $\mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Omega})$ denotes the k-dimensional multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Omega}$; and $\boldsymbol{\Sigma}(\boldsymbol{\tau})$ is a $(k \times k)$ the covariance matrix

$$\boldsymbol{\Sigma}(\boldsymbol{\tau}) = \sum_{j\in\mathbb{Z}} \boldsymbol{\Gamma}_j; \quad \boldsymbol{\Gamma}_j = \mathbb{E}(\boldsymbol{w}_t \boldsymbol{w}'_{t+j}); \quad \boldsymbol{w}_t = (w_{1t}, \ldots, w_{kt})'; \quad w_{it} = 1\{\Delta_t(\tau_i) \leq 0\}$$

with a typical entry of $\boldsymbol{\Gamma}_0$ given by $(\boldsymbol{\Gamma}_0)_{ij} = \min(\tau_i, \tau_j) - \tau_i\tau_j$ for $1 \leq i, j < k$

Further, since the set of indicator functions $\mathcal{I} = \{1\{-x]\}$ is Donsker class we have that the empirical process $W_T = \{W_T(\tau), \tau \in (0,1)\}$ admits a uniform central limit theorem

Corolário 3.3 *Let $W_T = \{W_T(\tau), \tau \in (0,1)\}$. Under Assumptions 1.1-3.4:*

$$W_T \xrightarrow{d} \mathcal{N}_\infty(0, C)$$

where $\mathcal{N}_\infty$ is a infinity dimensional Gaussian distribution with mean 0 and covariance structure given by

$$C(\tau, \tau') = (\min(\tau, \tau') - \tau\tau')\,(1 + \phi(\tau, \tau')), \quad (\tau, \tau') \in [0,1]^2$$

, *where* $\phi(\tau,\tau') = 2\sum_{j=1}^{\infty}\rho_j(\tau,\tau')$, $\rho_j(\tau,\tau') = \frac{\gamma_j(\tau,\tau')}{\gamma_0(\tau,\tau')}$ *and* $\gamma_j(\tau,\tau') = \mathbb{E}(w_t(\tau), w_{t+|j|}(\tau'))$

3.4 Inference

Under the assumption that the intervention had no effect on the unit of interest we postule our the null hypothesis as being

$$\mathcal{H}_0 : \Delta_t = 0 \qquad t = 1, \dots, T \tag{3-5}$$

As a consequence, under the null and Assumption 3.1, the conditional distribution $F_{Y|\boldsymbol{X}}$ is unaltered. Hence (3-6) implies the equality of the conditional quantiles of $Y_t|\boldsymbol{X}_t$.

However, (3-6) is *not* implied by the equality of the conditional quantiles. Since the latter is only with respect to the marginal distribution of $Y_t|\boldsymbol{X}_t$, the intervention might had an effect on the on the jointly distribution of $(Y_1|\boldsymbol{X}_1, \dots, Y_T|\boldsymbol{X}_T)$. For that reason we postule a weaker null hypothesis against which the test is more powerful. We test for the stability of $k < \infty$, $\tau - quantiles$ of the conditional distribution.

$$\mathcal{H}_0^{\tau} : Q_t(\boldsymbol{\tau}) = Q(\boldsymbol{\tau}) \qquad t = 1, \dots, T \tag{3-6}$$

Once the asymptotic normality of the $\widehat{\tau}_T$ is ensured (Theorem 3.1) is straightforward to conduct asymptotic inference. For the a i.i.d sampling we have $\phi_{ij} = 0$ or $\boldsymbol{\Sigma}(\tau) = \boldsymbol{\Gamma}_0$. Note that even uncoreleteness (nor mean independence) are enough for the latter result, since we what is necessary is serial uncorrelation (mean independence) among $\{w_t\}_t$, which is not implied by the by uncorrelatedeness (mean independence) of v_t.

For a general weakly dependent case ϕ_{ij} takes into account the serial correlation structure on w_t which can be consistently estimated using the residuals $\{\boldsymbol{e}_t \equiv \boldsymbol{w}_t - \widehat{\boldsymbol{\tau}}_T\}_t$.The finite sample covariance structure to be estimated given by

$$\boldsymbol{\Sigma}_T \equiv \boldsymbol{\Sigma}_T(\boldsymbol{\tau}) \equiv \sum_{j=-T+T_0}^{T-T0} \frac{T - T_0 + 1 - |j|}{T - T_0 + 1}\boldsymbol{\Gamma}_j$$

Lemma 3.1 *Let* $\widehat{\boldsymbol{\Sigma}}_T$ *be a consistent estimator for* $\boldsymbol{\Sigma}_T$ *and* $\boldsymbol{\tau} = (\tau_1, \dots, \tau_k)' \in (0,1)^k$. *Under Assumptions 1.1-3.4 and* $\mathcal{H}_0^{\tau}$*:*

$$\boldsymbol{W}_T(\boldsymbol{\tau})'\widehat{\boldsymbol{\Sigma}}_T^{-1}\boldsymbol{W}_T(\boldsymbol{\tau})' \xrightarrow{d} \chi_k^2$$

, *where* χ_k^2 *is the chi-square distribution with* k *degrees of freedom*

In a typical application we would like to test for the stability of the interquartile range after the intervention. For instance for a given a pair (τ_1, τ_2) such that $0 \leq \tau_1 < \tau_2 \leq 1$, let $r \equiv \tau_1 - \tau_2$ then we could test the stability of the probability covered r directly using

$$\widehat{r}_T = \tfrac{1}{T - T0 + 1} \sum_{t=T_0}^{T} b_t; \qquad b_t \equiv 1\{\widehat{y}_t^{(0)}(\tau_1) \leq y_t \leq \widehat{y}_t^{(0)}(\tau_2)\}$$

, which as a direct consequence of Theorem 3.1

Lemma 3.2 *Under Assumptions 1.1-3.4 and $\mathcal{H}_0$:*

$$\sqrt{T}\left(\frac{\widehat{r}_T - r}{\sqrt{\frac{r(1-r)(1+\widehat{\phi}_T)}{\lambda_0(1-\lambda_0)}}}\right) \xrightarrow{d} \mathcal{N}(0,1) \tag{3-7}$$

, where $\widehat{\phi}_T = \widehat{\phi}_T(r)$ is a consistent estimator for $\phi_T \equiv \phi_T(r)$, which is the univariate version of (3) with w_t replaced by b_t in the covariance $\{\gamma_j\}_{j\geq 0}$ definition

Any measure of the distance between the test-statistic $W_T \equiv \{W_T(\tau) : \tau \in [0,1]\}$ and the normal distribution $N_\infty(0, C)$ can be used as evidence against the null hypothesis that the the conditional distribution is stable regarding the intervention. Some popular measures of distances are the $\mathcal{L}^p$ norms denoted by $\|\cdot\|_p$ norm for $p \in [1, \infty]$. Since those norms are continuos transformation of W_T, the next lemma follows from the continuos mapping theorem.

Lemma 3.3 *For $p \in [1, \infty]$, under Assumptions 1.1-3.4 and $\mathcal{H}_0$:*

$$\|W_T\|_p \xrightarrow{d} \|\mathcal{N}_\infty(0, C)\|_p$$

, where $\|f\|_p = \left(\int |f(x)|^p dP_X\right)^{1/p}$ if $1 \leq p \leq \infty$ and $\|f\|_\infty = \mathsf{sup}_{x\in\mathcal{X}} |f(x)|$

In particular for $p = 2$ and $p = \infty$ those statistics are the conditional analogous of the square root of Cramer-von-Mises and Kolmogorov-Smirnov (KS) statistic respectively. For a random sample (i.i.d observations) $N_\infty(0, C)$ reduces to a brownian bridge $\mathcal{B}$. Such that the limit distribution is the same of the KS-test, which is given by $W_\infty \equiv \mathsf{sup}_{u\in[0,1]} \mathcal{B}(u)$, which is tabulated or it can be calculated analytically to a arbitrary precision using the Marsaglia Tsang (2003) series

$$P(W_\infty > x) = 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp(-2j^2x^2)$$

Similarly for $p = 2$, we have the limiting distribution of $\|W_T\|_2^2$ given by $W_2 \equiv \int_0^1 \mathcal{B}^2(u)du$, which can also be expanded in a series as

$$P(W_2 > x) = \frac{1}{\pi}\sum_{j=1}^{\infty}(-1)^{j+1}\int_{(2j-1)^2\pi^2}^{4j^2\pi^2}\sqrt{\frac{-\sqrt{y}}{\sin\sqrt{y}}}\frac{\exp\left(-xy/2\right)}{y}dy$$

For the case of weakly dependent data there is no simple analytic solution for the limit distribution of the test statistics. One could conduct distributional inference based on resampling schemes or bootstrap (block bootstrap in that case).

Alternatively under the normality assumption of the innovation we derive in Section 3 a close form for the covariance structure of $\{w\}$ for any particular covariance structure in the raw data. Hence one could fit an simple ARMA model and use those estimated as plug in the λ_j

3.4.1 Misspecification

$$Q_\tau(\boldsymbol{x}) = g\left(\boldsymbol{x}, \boldsymbol{\theta}_0(\tau)\right) + a_\tau(\boldsymbol{x})$$

Consider the assumption where both $\boldsymbol{f}_t$ and $\boldsymbol{\eta}_t$ are normally distributed , in that case

$$\boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \mathbf{\Pi}); \qquad \mathbf{\Pi} = \begin{pmatrix} \mathbf{\Lambda}_1\boldsymbol{Q}\mathbf{\Lambda}_1' + \omega_1 & \mathbf{\Lambda}_1\boldsymbol{Q}\mathbf{\Lambda}_0' \\ \mathbf{\Lambda}_0\boldsymbol{Q}\mathbf{\Lambda}_1' & \mathbf{\Lambda}_0\boldsymbol{Q}\mathbf{\Lambda}_0' + \boldsymbol{\omega}_0 \end{pmatrix}.$$

Giving a, possibly infinity order stable matrix polinomial $\mathbf{\Psi}(L)$, we have the $\boldsymbol{Z}_t = \mathbf{\Psi}(L)\boldsymbol{\varepsilon}_t$ and covariance structure given by

$$\mathbf{\Gamma}_j \equiv \mathbb{C}(\boldsymbol{Z}_t, \boldsymbol{Z}_{t+j}) = \sum_{i=0}^{\infty}\mathbf{\Psi}_i\mathbf{\Pi}\mathbf{\Psi}_{i+j}'$$

Consider the assumption where both $\boldsymbol{f}_t$ and $\boldsymbol{\eta}_t$ are normally distributed , in that case [

It is well know that the conditional distribution of a multivariate normal is also normally distributed as

$$\begin{aligned} Y_t|\boldsymbol{X}_t = \boldsymbol{x} &\sim \mathcal{N}(\alpha + \boldsymbol{x}'\boldsymbol{\beta}, \sigma^2) \\ \boldsymbol{\beta} &= [\mathbf{\Gamma}_0]_{10}[\mathbf{\Gamma}_0]_{00}^{-1} \\ \alpha &= \mu_1 - \boldsymbol{\mu}_0\boldsymbol{\beta} \\ \sigma^2 &= \Omega_{11} - \mathbf{\Omega}_{10}\mathbf{\Omega}_{00}^{-1}\mathbf{\Omega}_{01} \end{aligned}$$

Also for a normal random variable with mean μ and variance σ^2, the quantile function is given by $\mu + \sigma\Phi^{-1}(\tau)$, where $\Phi(\cdot)$ denotes the standard normal distribution function. Hence for our example the conditional quantile functions becomes

$$Q_\tau(\boldsymbol{x}) = \alpha + \boldsymbol{x}'\boldsymbol{\beta} + \sigma\Phi^{-1}(\tau) = \theta_0(\tau) + \boldsymbol{x}'\boldsymbol{\beta}$$

which is linear in the parameters.

Let $\nu_t(\tau) = Y_t - \theta_0(\tau) - \boldsymbol{X}_t'\boldsymbol{\beta} = -\theta_0(\tau) + (1, -\boldsymbol{\beta})\boldsymbol{Z}_t$. Then $\boldsymbol{\nu}_T = (\nu_1, \ldots, \nu_T)'$ is given by

$$\tfrac{1}{\sigma}\boldsymbol{\nu}_T \sim \mathcal{N}(-\Phi^{-1}(\tau), \Lambda)$$

$$\lambda_j = \frac{\mathbb{C}(\nu_i, \nu_j)}{\sigma^2} = \frac{(1, -\boldsymbol{\beta})\boldsymbol{\Gamma}_j(1, -\boldsymbol{\beta})'}{(1, -\boldsymbol{\beta})\boldsymbol{\Gamma}_0(1, -\boldsymbol{\beta})'}$$

In that case we can explicitly express the covariance structure of $\{w_t\}$ by $\gamma_j = \mathbb{P}(\nu_t \leq 0; \nu_{t+j} \leq 0) - \tau^2$. Where the first term can be evaluated for $j \neq 0$ by

$$\mathbb{P}(\nu_t \leq 0; \nu_{t+j} \leq 0) = \Phi\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, -\Phi^{-1}(\tau)\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & \lambda_j \\ \lambda_j & 1 \end{pmatrix}\right]$$

3.5
Monte Carlo

We conducted a Monte Carlo study by simulating the DGP described in Assumption 3.1 applying different configurations around a baseline scenario consisting of 5 units (including the treated one), 100 observations with the treatment at $T_0 = 50$. Table C.7 shows the size for the test for different distributions of the common factor. We include chi-square innovations as well as t-distribution to check the robustness of our asymptotic results to skewness and fat tails respectively. In seems that the the distribution pays little part on determining the test size

Overall the test seems to be rightly sized with greater distortions as we move away from the median. The sup test seems to be consistently slightly undersized, whereas the L_2 slightly oversized. However both distributional test can be considered satisfactory for practical purposes.

3.6
Empirical Illustration

We now apply the methodology described so far to investigate the effects on stock returns after a change in corporate governance regime. The different levels of governance were created by BOVESPA in December, 2000, at the times with three distinct levels:[2] Basic, where no special requirement is made on top of all the rules that already apply to all listed companies in the stock exchange. Level N1, where the participant are required, among other things, to attempt public meeting with analysts and investors at least once an year; keep a minimum of 25% of the company's capital free-floating, Improvement in quarterly reports, including the disclosure of consolidated financial statements and special audit revision. On top of that, to qualify for the level N2, the participant must adopt well established international laws of accounting, create means to mediate partnership disputes,Establishment of a two-year unified mandate for the entire Board of Directors, which must have five members at least, of which at least 20% shall be independent members and, in case of change of ownership, extend the same right of the common shareholders (up to 80% of the value) to the preferential shareholders.

Finally to be listed in the most restrict of corporate governance, level Novo Mercado (NM), the company must have only common stocks.Overall, any movement towards higher levels (from Basic to NM) implies stronger requirements in the listed company, which are mainly design to protect minority shareholders. Since those movements are completely voluntary , it is natural to interpret them as a sign of commitment to better corporate governance practices. The date of the migration would then represent the timing of the intervention (treatment).

We are far from being the pioneers in the attempt to uncover the link between corporate governance and stock returns. To name a few, Mitton (2002) looks at the Southeast Asian 1997 crises to study the relation between the downfall of the stock market and the fact that some of those stock were also listed in the USA via American Depositary Recipients or were audited by well known auditing companies. Lemmon and Lins (2003) compare the stock returns of companies with less concentrated capital structure also considering the Southeast Asian 1997 crises background. In particular for Brazilian market, we have Srour (2005) investigating the relation between stock returns and corporate governance using company data from 1997-2001. Lastly, Almeida (2007) looks at the same scenarios as ours and fit GARCH models to each stock during the transition window.

It seems intuitive that good corporate governance should lead to a decrease in volatility of the returns. While the causes might be different, or at least situation dependent, there are compelling evidence presented in conclusion of all those papers mention above to support such a claim.

We first identify stocks that made the transition. Here we do not

[2]Currently 2 more levels were included: Bovespa Mais and Bovespa Mais Level 2

distinguish between any of the three level (N1,N2 or NM). Any transition from the Basic Level to higher level of corporate governance we treat as a intervention. While this is not entirely satisfactory there is no requirement that each company willing to migrate must be do so level-by-level. Hence we have cases of a company going from Basic to NW at once. Since we do not possess any case of downgrade in the dataset we only investigate upwards movement. Once we identify the unit that made the transition we look from peers (control) in the same sector that did not made any change corporate governance level in the timeframe of interest. We use this criteria to both capture sectorial shock through the peers and isolate the unit of interest from possible spurious correlation among unrelated companies.

The data set consist of daily closing price of hundreds of stocks listed at Bovespa from Jan/00- Dez/09. Of those only 49 made the transition in time spam considered. Restricting to cases, where the unit of interest has at leat one peer in the same business segment that was untreated it reduces to 4 cases to analyze which are described in Table C.9

3.7
Conclusion

In this chapter we have extended the ArCo methodology for the estimation of intervention effects on the quantiles of variables of interest.

Bibliography

ABADIE, A.; DIAMOND, A. ; HAINMUELLER, J.. **Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program**. *Journal of the American Statistical Association*, 105:493–505, 2010.

ABADIE, A.; DIAMOND, A. ; HAINMUELLER, J.. **Politics and the synthetic control method**. *American Journal of Political Science*, 2014. In press.

ABADIE, A.; ANGRIST, J. ; IMBENS, G.. **Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings**. *Econometrica*, 70:91–117, 2002.

ABADIE, A.; GARDEAZABAL, J.. **The economic costs of conflict: A case study of the Basque country**. *American Economic Review*, 93:113–132, 2003.

BELASEN, A.; POLACHEK, S.. **How hurricanes affect wages and employment in local labor markets**. *The American Economic Review: Papers and Proceedings*, 98:49–53, 2008.

BILLMEIER, A.; NANNICINI, T.. **Assessing economic liberalization episodes: A synthetic control approach**. *The Review of Economics and Statistics*, 95:983–1001, 2013.

BELLONI, A.; CHERNOZHUKOV, V. ; HANSEN, C.. **Inference on treatment effects after selection amongst high-dimensional controls**. *Review of Economic Studies*, 81:608–650, 2014.

BELLONI, A.; CHERNOZHUKOV, V.; FERNÁNDEZ-VAL, I. ; HANSEN, C.. **Program evaluation with high-dimensional data**. *Econometrica*, 2016. In press.

BELLONI, A.; CHERNOZHUKOV, V.; CHETVERIKOV, D. ; WEI, Y.. **Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework**. Working Paper 1512.07619, arXiv, 2016.

FERMAN, B.; PINTO, C.. **Inference in differences-in-differences with few treated groups and heteroskedasticity**. Working paper, São Paulo School of Economics - FGV, 2015.

FERMAN, B.; PINTO, C.. **Revisiting the synthetic control estimator**. Working paper, São Paulo School of Economics - FGV, 2016.

FERMAN, B.; PINTO, C. ; POSSEBOM, V.. **Cherry picking with synthetic controls**. Working paper, São Paulo School of Economics - FGV, 2016.

PÖTSCHER, B.; PRUCHA, I.. **Dynamic nonlinear econometric models: Asymptotic theory**. Springer, 1997.

BAI, C.-E.; LI, Q. ; OUYANG, M.. **Property taxes and home prices: A tale of two cities**. *Journal of Econometrics*, 180:1–15, 2014.

CARVALHO, C.; MASINI, R. ; MEDEIROS, M.. **Arco: An artificial counterfactual approach for high-dimensional data**. Working paper, Pontifical Catholic University of Rio de Janeiro, 2016.

HSIAO, C.; CHING, H. S. ; WAN, S. K.. **A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with mainland China**. *Journal of Applied Econometrics*, 27:705–740, 2012.

ANDREWS, D.. **Heteroskedasticity and autocorrelation consistent covariance matrix estimation**. *Econometrica*, 59:817–858, 1991.

ANDREWS, D.; MONAHAN, J.. **An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator**. *Econometrica*, 60:953–966, 1992.

MCLEISH, D.. **Dependent central limit theorems and invariance principles**. *Annals of Probability*, 2:620–628, 1974.

CAVALLO, E.; GALIANI, S.; NOY, I. ; PANTANO, J.. **Catastrophic natural disasters and economic growth**. *The Review of Economics and Statistics*, 95:1549–1561, 2013.

DUBOIS, E.; HÉRICOURT, J. ; MIGNON, V.. **What if the euro had never been launched? a counterfactual analysis of the macroeconomic impact of euro membership**. *Economics Bulletin*, 29:2252–2266, 2009.

FATAS, E.; NOSENZO, D.; SEFTON, M. ; ZIZZO, D.. **A self-funding reward mechanism for tax compliance**. Working Paper 2650265, SSRN, 2015.

RIO, E.. **A new weak dependence condition and applications to moment inequalities**. *Comptes rendus Acad. Sci. Paris, Série I*, 318:355–360, 1994.

SOUZA, F.. **Tax evasion and inflation**. Master's dissertation, *Department of Economics, Pontifical Catholic University of Rio de Janeiro*, http://www.econ.puc-rio.br/biblioteca.php/trabalhos/show/1413, 2014.

CARUSO, G.; MILLER, S.. **Long run effects and intergenerational transmission of natural disasters: A case study on the 1970 ancash earthquake**. *Journal of Development Economics*, 117:134–150, 2015.

BROCKMANN, H.; GENSCHEL, P. ; SEELKOPF, L.. **Happy taxation: increasing tax compliance through positive rewards?** *Journal of Public Policy*, FirstView:1–26, 2016.

CHEN, H.; HAN, Q. ; LI, Y.. **Does index futures trading reduce volatility in the Chinese stock market? a panel data evaluation approach**. *Journal of Futures Markets*, 33:1167–1190, 2013.

FUJIKI, H.; HSIAO, C.. **Disentangling the effects of multiple treatments - measuring the net economic impact of the 1995 great Hanshin-Awaji earthquake**. *Journal of Econometrics*, 186:66–73, 2015.

LEEB, H.; PÖTSCHER, B.. **Model selection and inference: Facts and fiction**. *Econometric Theory*, 21:21–59, 2005.

LEEB, H.; PÖTSCHER, B.. **Sparse estimators and the oracle property, or the return of Hodge's estimator**. *Journal of Econometrics*, 142:201–211, 2008.

LEEB, H.; PÖTSCHER, B.. **On the distribution of penalized maximum likelihood estimators: The LASSO, SCAD, and thresholding**. *Journal of Multivariate Analysis*, 100:1065–2082, 2009.

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

PESARAN, M.; SMITH, R.. **Counterfactual analysis in macroeconometrics: An empirical investigation into the effects of quantitative easing**. Discussion Paper 6618, IZA, 2012.

ZOU, H.. **The adaptive LASSO and its oracle properties**. *Journal of the American Statistical Association*, 101:1418–1429, 2006.

IBRAGIMOV, I.; LINNIK, V.. *Wolters-Noordhoff series of monographs and textbooks on pure and applied mathematics.s*, chapter Independent and stationary sequences of random variables. 1971.

ANGRIST, J.; IMBENS, G.. **Identification and estimation of local average treatment effects**. *Econometrica*, 61:467–476, 1994.

ANGRIST, J.; IMBENS, G. ; RUBIN, D.. **Identification of causal effects using instrumental variables**. *Journal of the American Statistical Association*, 91:444–472, 1996.

ANGRIST, J.; JORDÁ, Ó. ; KUERSTEINER, G.. **Semiparametric estimates of monetary policy effects: String theory revisited**. Working Paper 2013-24, Federal Reserve Bank of San Francisco, 2013.

BAI, J.. **Estimating multiple breaks one at a time**. *Econometric Theory*, 13:315–352, 1997.

BAI, J.. **Panel data models with interactive fixed effects**. *Econometrica*, 77:1229–1279, 2009.

BAI, J.; PERRON, P.. **Estimating and testing linear models with multiple structural changes**. *Econometrica*, 66:47–78, 1998.

FERNÁNDEZ-VILLAVERDE, J.; RUBIO-RAMÍREZ, J.; SARGENT, T. ; WATSON, M.. **ABCs (and Ds) of understanding VARs**. *American Economic Review*, 97:1021–1026, 2007.

HECKMAN, J.; VYTLACIL, E.. **Structural equations, treatment effects and econometric policy evaluation**. *Econometrica*, 73:669–738, 2005.

SLEMROD, J.. **Cheating ourselves: The economics of tax evasion**. *Journal of Economic Perspectives*, 21:25–48, 2010.

WAN, J.. **The incentive to declare taxes and tax revenue: The lottery receipt experiment in china**. *Review of Development Economics*, 14:611–624, 2010.

GRIER, K.; MAYNARD, N.. **The economic consequences of Hugo Chavez: A synthetic control analysis**. *Journal of Economic Behavior and Organization*, 95:1549–1561, 2013.

GOBILLON, L.; MAGNAC, T.. **Regional policy evaluation: Interactive fixed effects and synthetic controls**. *Review of Economics and Statistics*, 2016. forthcoming.

ZHANG, L.; DU, Z.; HSIAO, C. ; YIN, H.. **The macroeconomic effects of the Canada-US free trade agreement on Canada: A counterfactual analysis**. *World Economy*, 2014. In Press.

OUYANG, M.; PENG, Y.. **The treatment-effect estimation: A case study of the 2008 economic stimulus package of China**. *Journal of Econometrics*, 188:545–557, 2015.

PESARAN, M.; SCHUERMANN, T. ; WEINER, S.. **Modeling regional interdependencies using a global error-correcting macroeconometric model**. *Journal of Business and Economic Statistics*, 22:129–162, 2004.

PESARAN, M.; SMITH, L. ; SMITH, R.. **What if the UK or Sweden had joinded the Euro in 1999? an empirical evaluation using a Global VAR**. *International Journal of Finance and Economics*, 12:55–87, 2007.

BÜLHMANN, P.; VAN DER GEER, S.. **Statistics for high dimensional data**. Springer, 2011.

DOUKHAN, P.; LOUHICHI, S.. **A new weak dependence condition and applications to moment inequalities**. *Stochastic Processes and their Applications*, 84:313–342, 1999.

PHILLIPS, P.. **Understanding spurious regressions in econometrics**. *Journal of Econometrics*, 33:311–340, 1986.

ENGLE, R.; GRANGER, C.. **Co-integration and error correction: Representation, estimation, and testing**. *Econometrica*, 55:251–276, 1987.

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

AN, S.; SCHORFHEIDE, F.. **Bayesian analysis of DSGE models**. *Econometric Reviews*, 26:113–172, 2007.

DEES, S.; MAURO, F. D.; PESARAN, M. ; SMITH, L.. **Exploring the international linkages of the Euro area: A Gobal VAR analysis**. *Journal of Applied Econometrics*, 22:1–38, 2007.

DURLAUF, S.; PHILLIPS, P.. **Multiple time series regression with integrated processes**. *Review of Economic Studies*, 53:473–495, 1985.

FIRPO, S.. **Efficient semiparametric estimation of quantile treatment effects**. *Econometrica*, 75:259–276, 2007.

JORDAN, S.; VIVIAN, A. ; WOHAR, M.. **Sticky prices or economically-linked economies: the case of forecasting the Chinese stock market**. *Journal of International Money and Finance*, 41:95–109, 2014.

JOHNSON, S.; BOONE, P.; BREACH, A. ; FRIEDMAND, E.. **Corporate governance in the asian financial crisis**. *Journal of Financial Economics*, 58:141–186, 2000.

XIE, S.; MO, T.. **Index futures trading and stock market volatility in china: A difference-in-difference approach**. *Journal of Futures Markets*, 34:282–297, 2013.

CONLEY, T.; TABER, C.. **Inference with difference in differences with a small number of policy changes**. *Review of Economics and Statistics*, 93:113–125, 2011.

CHERNOZHUKOV, V.; HANSEN, C.. **An IV model of quantile treatment effects**. *Econometrica*, 73:245–261, 2005.

CHERNOZHUKOV, V.; HANSEN, C.. **Instrumental quantile regression inference for structural and treatment effect models**. *Journal of Econometrics*, 132:491–525, 2006.

CHERNOZHUKOV, V.; HANSEN, C.. **Instrumental variable quantile regression: A robust inference approach**. *Journal of Econometrics*, 141:379–398, 2008.

CHERNOZHUKOV, V.; FERNANDEZ-VAL, I. ; MELLY, B.. **Inference on counterfactual distributions**. *Econometrica*, 2014. Forthcoming.

HAAN, W. D.; LEVIN, A.. **Inferences from parametric and non-parametric covariance matrix estimation procedures**, 1996.

NEWEY, W.; WEST, K.. **A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix**. *Econometrica*, 55:703–708, 1987.

CHEN, X.. **Large sample sieve estimation of semi-nonparametric models**. In Heckman, J.; Leamer, E., editors, *Handbook of Econometrics*, volume 6B, pp 5549—-5632. Elsevier Science, 2007.

GAO, Y.; LONG, W. ; WANG, Z.. **Estimating average treatment effect by model averaging**. *Economics Letters*, 135:42–45, 2015.

XU, Y.. **Generalized synthetic control method for causal inference with time-series cross-sectional data**. Working paper, Massachusetts Institute of Technology, 2015.

DU, Z.; YIN, H. ; ZHANG, L.. **The macroeconomic effects of the 35-h workweek regulation in france**. *The B.E. Journal of Macroeconomics*, 13:881–901, 2013.

DU, Z.; ZHANG, L.. **Home-purchase restriction, property tax and housing price in China: A counterfactual analysis**. *Journal of Econometrics*, 188:558–568, 2015.

A
Appendix: Proofs

A.1
Proofs of Chapter 1

We begin by proving an uniform version for the Continuous Mapping Theorem (UCMT) and the Slutsky Theorem (UST). For the next 2 Lemmas, $\boldsymbol{X}_T$, $\boldsymbol{Y}_T$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are random elements taking values on a subset $\mathcal{D}$ of the Euclidean space (real-valued scalar, vector or matrix) defined over the same probabilistic space with distribution P index by $\mathcal{P}$.

Lemma A.1 ***(Uniform Continuous Mapping Theorem)*** *Let* $\boldsymbol{g} : \mathcal{D} \to \mathcal{E}$ *be uniformly continuous at every point of a set* $\mathcal{C} \subseteq \mathcal{D}$ *where* $\mathbb{P}_P(\boldsymbol{X} \in \mathcal{C}) = 1$ *for all* $P \in \mathcal{P}$*.*

(a) If $\boldsymbol{X}_T \xrightarrow{p} \boldsymbol{X}$ *uniformly in* $P \in \mathcal{P}$*, then* $\boldsymbol{g}(\boldsymbol{X}_T) \xrightarrow{p} \boldsymbol{g}(\boldsymbol{X})$ *uniformly in* $P \in \mathcal{P}$*.*

(b) If $\boldsymbol{X}_T \xrightarrow{d} \boldsymbol{X}$ *uniformly in* $P \in \mathcal{P}$*, then* $\boldsymbol{g}(\boldsymbol{X}_T) \xrightarrow{d} \boldsymbol{g}(\boldsymbol{X})$ *uniformly in* $P \in \mathcal{P}$*.*

Proof. 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 the preimage of $\boldsymbol{g}$ denoted by $\boldsymbol{g}^{-1}(E) \equiv \{\boldsymbol{x} \in \mathcal{D} : \boldsymbol{g}(\boldsymbol{x}) \in E\}$. For close $F \in \mathcal{E}$ 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 (Lemma 2.2 Van der Vaart, 2000). The result follows again by the Portmanteau Lemma in the other direction. ∎

Lemma A.2 ***(Uniform Slutsky Theorem)*** *Let* $\boldsymbol{X}_T \xrightarrow{p} \boldsymbol{C}$ *uniformly in* $P \in \mathcal{P}$, *where* $\boldsymbol{C} \equiv \boldsymbol{C}(P)$ *is a non random conformable matrix and* $\boldsymbol{Y}_T \xrightarrow{d} \boldsymbol{Y}$ *uniformly in* $P \in \mathcal{P}$, *then*

(a) $\boldsymbol{X}_T + \boldsymbol{Y}_T \xrightarrow{d} \boldsymbol{C} + \boldsymbol{Y}$ *uniformly in* $P \in \mathcal{P}$

(b) $\boldsymbol{X}_T\boldsymbol{Y}_T \xrightarrow{d} \boldsymbol{C}\boldsymbol{Y}$ *uniformly in* $P \in \mathcal{P}$, *if* $\boldsymbol{C}$ *is bounded uniformly in* $P \in \mathcal{P}$.

(c) $\boldsymbol{X}_T^{-1}\boldsymbol{Y}_T \xrightarrow{d} \boldsymbol{C}^{-1}\boldsymbol{Y}$ *uniformly in* $P \in \mathcal{P}$, *if* $\det(\boldsymbol{C})$ *is bounded away from zero uniformly in* $P \in \mathcal{P}$.

Proof. 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 A.1(b).* ∎

Proof of Proposition 1.2

Proof. Recall that $\mathcal{M}_t \equiv \mathcal{M}(\boldsymbol{x}_t)$, $\boldsymbol{\nu}_t \equiv \boldsymbol{y}_t^{(0)} - \mathcal{M}_t$ for $t \geq 1$ and $\boldsymbol{\eta}_{t,T} \equiv \widehat{\mathcal{M}}_t - \mathcal{M}_t$ for $t \geq T_0$. From the definition of our estimator we have that $\widehat{\boldsymbol{\Delta}}_T - \boldsymbol{\Delta}_T$ is equal to

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

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

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

By condition (a) in the proposition, the last term in the right hand side converges to zero uniformly in $P \in \mathcal{P}$. Under condition (b), 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:

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

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

It follows from Lemma A.1(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 A.2(a), $\sqrt{T}\left(\widehat{\boldsymbol{\Delta}}_T - \boldsymbol{\Delta}_T\right) \xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \frac{\boldsymbol{\Gamma}}{\lambda_0(1-\lambda_0)}\right]$, uniformly in $P \in \mathcal{P}$. ■

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

Lemma A.3 *Let $\{u_t\}_{t\in\mathbb{N}}$ be strong mixing sequence of centered random variables with mixing coefficient with exponential decay. Also for some real $r > 2$, $\sup_t \mathbb{E}|u_t|^{r+\delta} < \infty$ for some $\delta > 0$, then there exist a positive constant C_r (not depending on n) such that*

$$\mathbb{E}|u_1 + \cdots + u_T|^r \leq C_r T^{r/2}.$$

Proof. See Doukhan e Louhichi (1999) and Rio (1994). ■

Lemma A.4 *Under Assumptions 1.2-3.4,* $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 = O_P\left(s_0 \frac{d^{1/\gamma}}{\sqrt{T}}\right)$.

Proof. For real $a, b > 0$ define:

$$\mathscr{A}(a) = \left\{ \left\| \frac{2}{T_1} \sum_{t=1}^{T_1} \boldsymbol{x}_t \nu_t \right\|_{\max} \leq a \right\}, \quad \boldsymbol{p}_t(d \times 1) \equiv \boldsymbol{x}_t \nu_t;$$
$$\mathscr{B}(b) = \left\{ \left\| \frac{1}{T_1} \sum_{t=1}^{T_1} \boldsymbol{M}_t \right\|_{\max} \leq b \right\}, \quad \boldsymbol{M}_t(d \times d) \equiv \boldsymbol{x}_t \boldsymbol{x}_t' - \mathbb{E}(\boldsymbol{x}_t \boldsymbol{x}_t'),$$

where $\|\cdot\|_{\max}$ is the maximum entry-wise norm.

Following Corollary 6.10 of Bühlmann e van der Geer (2011) on $\mathscr{A}(a) \cap \mathscr{B}(b)$, we have that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 \leq \frac{32\varsigma s_0}{\psi_0^2}$, provided that $\varsigma \geq 8a$, $b \leq \frac{\psi_0^2}{32 s_0}$ 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 1.2). For convenience set $a = \frac{\varsigma}{8}$ and $b = \frac{\psi_0^2}{32 s_0}$. Then, we can write $\mathbb{P}\left(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1 > \frac{32\varsigma s_0}{\psi_0^2}\right)$

$$\begin{aligned}
&\leq \mathbb{P}\left(\left\| \frac{2}{T} \sum_{t=1}^{T_1} \boldsymbol{p}_t \right\|_{\max} > \frac{\varsigma}{8} \right) + \mathbb{P}\left(\left\| \frac{1}{T_1} \sum_{t=1}^{T_1} \boldsymbol{M}_t \right\|_{\max} > \frac{\psi_0^2}{32 s_0} \right) \\
&\leq d \max_{1 \leq i \leq d} \mathbb{P}\left(\left| \sum_{t=1}^{T_1} p_{i,t} \right| > \frac{\varsigma T_1}{16} \right) + d^2 \max_{1 \leq i,j \leq d} \mathbb{P}\left(\left| \sum_{t=1}^{T_1} m_{ij,t} \right| > \frac{\psi_0^2 T_1}{32 s_0} \right) \\
&\leq d \left(\frac{16}{\varsigma T_1}\right)^{\gamma} \max_{1 \leq i \leq d} \mathbb{E} \left| \sum_{t=1}^{T_1} p_{i,t} \right|^{\gamma} + d^2 \left(\frac{32 s_0}{\psi_0^2 T_1}\right)^{\gamma} \max_{1 \leq i,j \leq d} \mathbb{E} \left| \sum_{t=1}^{T_1} m_{ij,t} \right|^{\gamma} \\
&\leq C_1(\gamma) \frac{d}{T_1^{\gamma/2} \varsigma^{\gamma}} + C_2(\gamma, \psi_0) \frac{d^2 s_0^{\gamma}}{T_1^{\gamma/2}},
\end{aligned}$$

where the second inequality follows from the union bound. The third inequality follows from the Markov inequality applied for some $\gamma > 2$. The last inequality is a consequence of Lemma 3, since (i) by Assumption 1.3(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 1.3(b) we have for some $\delta > 0$ and $t \geq 1$:

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

The result follows since, by Assumption 3.4(a) $\varsigma = O\left(\frac{d^{1/\gamma}}{\sqrt{T}}\right)$ and by Assumption 3.4(b), $s_0 \frac{d^{2/\gamma}}{\sqrt{T}} = o_P(1)$. ■

Lemma A.5 *Let* $\boldsymbol{S}_T \equiv \sum_{t=1}^{T} \boldsymbol{u}_t$ *where* $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{dt})' \in \mathcal{U} \subset \mathbb{R}^d$ *is a zero mean random vector, such that the process* $(u_{j,t})$ *fulfils the conditions of Lemma A.3 for some real* $r > 2$ *for all* $j \in \{1, \ldots, d\}$. *Then,* $\|\boldsymbol{S}_T\|_{\max} = O_P(d^{1/r}\sqrt{T})$.

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

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

where the last inequality follows from Lemma A.3. ■

Proof of Theorem 1.3

Proof. Recall that $\eta_{t,T} = \boldsymbol{x}_t'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ for $t \geq T_0$, and let $\boldsymbol{\theta}_0 = (\alpha_0, \boldsymbol{\beta}_0')'$, where α is the parameter of the intercept while $\boldsymbol{\beta}$ is the vector of remaining parameters. Similar, let $\boldsymbol{x}_t = (1, \widetilde{\boldsymbol{x}}_t)$. From the definition of the estimator, $\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = \frac{1}{T_1}\sum_{t\leq T_1}\boldsymbol{\nu}_t - \frac{1}{T_1}\sum_{t\leq T_1}\widetilde{\boldsymbol{x}}_t\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$. Combining the last two expressions we can rewrite the estimation error as

$$\begin{aligned}\eta_{t,T} &= \frac{1}{T_1}\sum_{s\leq T_1}\boldsymbol{\nu}_s - \frac{1}{T_1}\sum_{s\leq T_1}\widetilde{\boldsymbol{x}}_s\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \widetilde{\boldsymbol{x}}_t\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\\ &= \frac{1}{T_1}\sum_{s\leq T_1}\boldsymbol{\nu}_s - \left[\frac{1}{T_1}\sum_{s\leq T_1}\widetilde{\boldsymbol{x}}_s - \widetilde{\boldsymbol{x}}_t\right]\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right).\end{aligned}$$

Taking the average over $t = T_0, \ldots, T$, multiplying by $\sqrt{T}$ and rearranging yields:

$$\sqrt{T}\left(\frac{1}{T_2}\sum_{t\geq T_0}\boldsymbol{\eta}_{t,T} - \frac{1}{T_1}\sum_{t\leq T_1}\boldsymbol{\nu}_t\right) = \left(\frac{\sqrt{T}}{T_2}\sum_{t\geq T_0}\widetilde{\boldsymbol{x}}_t - \frac{\sqrt{T}}{T_1}\sum_{t\leq T_1}\widetilde{\boldsymbol{x}}_t\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right).$$

We now show that the last expression is $o_P(1)$ uniformly in $P \in \mathcal{P}$. First, we bound it in absolute term by:

$$\left\|\frac{\sqrt{T}}{T_2}\sum_{t\geq T_0}\widetilde{\boldsymbol{x}}_t - \frac{\sqrt{T}}{T_1}\sum_{t\leq T_1}\widetilde{\boldsymbol{x}}_t\right\|_{\max}\left\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right\|_1.$$

Adding and subtracting the mean, the first term is the sum of two $O_P\left(d^{1/\gamma}\right)$ terms by Lemma A.5 combined with Assumption 1.3(a)-(b). The second term is $O_P\left(s_0\frac{d^{1/\gamma}}{\sqrt{T}}\right)$ by Lemma A.4. Hence, the last term in the above display is $O_P\left(s_0\frac{d^{2/\gamma}}{\sqrt{T}}\right) = o_P(1)$ by Assumption 3.4(b), which verifies condition (a) of Proposition 1.2.

Now $\{\nu_t\}$ is a strong mixing process with mixing coefficient with exponential decay and $\mathsf{sup}_t\,\mathbb{E}|\nu_t|^r < \infty$ for some $r > 4$ by Assumption 1.3(a) and (b). Also, $\mathbb{E}(\nu_t^2)$ is bounded by below uniformly by Assumption

1.3(c). Hence, we have a Central Limit Theorem as per Theorem 10.2 of Pötscher e Prucha (1997). Therefore, conditions (b) and (c) of Proposition 1.2 are verified and the result follows directly from Proposition 1.2.

■

Proof of Propositions 1.4 and 1.5

Proof. Both follows directly from Theorem 1.3 combined with Lemma A.2(c) ■

Proof of Theorem 1.6

Proof. From (A-1) in the Proof of Proposition 1.2, we have for $T_\lambda = \lfloor \lambda T \rfloor$, $\lambda \in \Lambda$ that $\mathbf{\Gamma}^{1/2}\boldsymbol{S}_T(\lambda)$ is equal to:

$$\frac{\sqrt{T}}{T - T_\lambda + 1}\sum_{t \geq T_\lambda} \boldsymbol{\nu}_t - \frac{\sqrt{T}}{T_\lambda - 1}\sum_{t < T_\lambda} \boldsymbol{\nu}_t - \frac{\sqrt{T}}{T - T_\lambda + 1}\sum_{t \geq T_\lambda} \boldsymbol{\eta}_{t,T} + \frac{\sqrt{T}}{T_\lambda - 1}\sum_{t < T_\lambda} \boldsymbol{\eta}_{t,T}.$$

The last two terms are $o_p(1)$ uniformly in $\lambda \in \Lambda$, under the conditions of Proposition 1.2, Assumption 1.5 and the fact that Λ is compact.

For fix $\lambda \in \Lambda$ the pointwise convergence in distribution follows under the conditions of from Proposition 1.2 (for instance under the assumptions of Theorem 1.3). The uniform convergence result then follows from the invariance principle in McLeish (1974) applied to $\boldsymbol{V}_T(\lambda) \equiv \frac{1}{\sqrt{T}}\sum_{t \geq T_\lambda} \boldsymbol{\nu}_t$ and the Continuous Mapping Theorem.

To obtain the covariance structure let $\mathbf{\Gamma}_{s-t} = \mathbb{E}(\boldsymbol{\nu}_t \boldsymbol{\nu}_s')$ for all s, t and note that for any pair $(\lambda, \lambda') \in \Lambda^2$ we have that

$$\begin{aligned}
\frac{1}{T}\sum_{t \geq T_\lambda}\sum_{s \geq T_{\lambda'}} \mathbf{\Gamma}_{s-t} &= \frac{T - T_{\lambda \vee \lambda'} + 1}{T}\left[\frac{1}{T - T_{\lambda \vee \lambda'} + 1}\sum_{t \geq T_\lambda}\sum_{s \geq T_{\lambda'}} \mathbf{\Gamma}_{s-t}\right] \\
&= (1 - \lambda \vee \lambda')\frac{\mathbf{\Gamma}}{\lambda \vee \lambda} + o_p(1),
\end{aligned}$$

where $\lambda \vee \lambda' = \max(\lambda, \lambda')$ and $\lambda \wedge \lambda' = \min(\lambda, \lambda')$. Finally, we have

$$\begin{aligned}\mathbb{E}[\boldsymbol{S}_T(\lambda)\boldsymbol{S}_t'(\lambda')] &= \boldsymbol{\Gamma}^{-1/2}\left[\tfrac{T^2}{(T-T_\lambda+1)(T-T\lambda'+1)}\tfrac{1}{T}\sum_{t\le T_\lambda}\sum_{s\le T_{\lambda'}}\boldsymbol{\Gamma}_{s-t}\right]\boldsymbol{\Gamma}^{-1/2} + o_p(1)\\ &= \left[\frac{1}{(1-\lambda)(1-\lambda')}\right]\frac{(1-\lambda\vee\lambda')}{\lambda\vee\lambda} + o_p(1)\\ &= \frac{1}{(\lambda\vee\lambda)(1-\lambda\wedge\lambda')} + o_p(1) \equiv \boldsymbol{\Sigma}_{\boldsymbol{\lambda}} + o_p(1)\end{aligned}$$

■

Proof of Proposition 1.7

Proof. Below we write T_λ to mean $\lfloor \lambda T \rfloor$. All the convergence in probability are a direct consequence of the Weak Law of Large Numbers ensured by the conditions of Proposition 1 combined with Assumption 1.5: Let $\lambda \le \lambda_0$:

$$\begin{aligned}\widehat{\boldsymbol{\Delta}}_T(\lambda) &\equiv \tfrac{1}{T-T_\lambda+1}\sum_{t=T_\lambda}^{T}\widehat{\boldsymbol{\delta}}_t(\lambda) = \left(\tfrac{T_0-T_\lambda}{T-T_\lambda+1}\right)\frac{\sum_{t=T_\lambda}^{T_0-1}\widehat{\boldsymbol{\Delta}}_t(\lambda)}{T_0-T_\lambda} + \left(\tfrac{T-T_0+1}{T-T_\lambda+1}\right)\frac{\sum_{t=T_0}^{T}\widehat{\boldsymbol{\delta}}_t(\lambda)}{T-T_0+1}\\ &= o_p(1) + \left(\frac{1-\lambda_0}{1-\lambda}\right)\boldsymbol{\Delta}.\end{aligned}$$

Similarly, consider a guess after the true value, $\lambda > \lambda_0$. Then:

$$\begin{aligned}\widehat{\boldsymbol{\Delta}}_T(\lambda) &\equiv \frac{1}{T-T_\lambda+1}\sum_{t=T_\lambda}^{T}\widehat{\boldsymbol{\delta}}_t(\lambda) = \frac{1}{T-T_\lambda+1}\sum_{t=T_\lambda}^{T}\left[\boldsymbol{y}_t - \widehat{\mathcal{M}}(\boldsymbol{x}_t)\right]\\ &= \frac{1}{T-T_\lambda+1}\sum_{t=T_\lambda}^{T}[\boldsymbol{y}_t - \mathcal{M}(\boldsymbol{x}_t)] - \frac{\lambda-\lambda_0}{\lambda}\boldsymbol{\Delta} + o_p(1)\\ &= \frac{1}{T-T_\lambda+1}\sum_{t=T_\lambda}^{T}\left[\boldsymbol{y}_t^{(0)} - \boldsymbol{\alpha}_0 - g(\boldsymbol{\theta}_0)\right] + \frac{\lambda_0}{\lambda}\boldsymbol{\Delta} + o_p(1) = \frac{\lambda_0}{\lambda}\boldsymbol{\Delta} + o_p(1),\end{aligned}$$

where the second equality follows from Assumption 1.6, since a step intervention will only affect (asymptotically) the constant regressor estimation of the model $\mathcal{M}$ by a factor of $\frac{\lambda-\lambda_0}{\lambda_0}$ times the intervention size $\boldsymbol{\Delta}$. To see this let $\boldsymbol{\alpha}_0$ be the constant and $\boldsymbol{\beta}_0$ the remaining parameters. Then,

$$\widehat{\boldsymbol{\alpha}} = \frac{1}{T_\lambda}\sum_{t\le T_\lambda}\boldsymbol{y}_t^{(0)} + \frac{1}{T_\lambda}\sum_{t\le T_\lambda}\boldsymbol{\Delta}I(t\ge T_0) - \frac{1}{T_\lambda}\sum_{t\le T_\lambda}\widetilde{\mathcal{M}}(\widehat{\boldsymbol{\beta}}),$$

where $\mathcal{M}(\boldsymbol{x}_t;\boldsymbol{\theta}_0) \equiv \boldsymbol{\alpha}_0 + \widetilde{\mathcal{M}}(\boldsymbol{x}_t;\boldsymbol{\beta}_0)$. Since the estimation of $\boldsymbol{\beta}_0$ is asymptotically unaffected by a step intervention, under the conditions of Proposition 1.2, $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$. Consequently, $\widehat{\boldsymbol{\alpha}}(\lambda) \xrightarrow{p} \boldsymbol{\alpha} + \frac{\lambda-\lambda_0}{\lambda}\boldsymbol{\Delta}$, $\forall\lambda \in (0,1)$. ■

Proof of Theorem 1.8

Proof. Note that: (i) The limiting function $J_{p,0}(\lambda) \equiv \phi(\lambda)\|\boldsymbol{\Delta}\|_p$ is uniquely maximized at $\lambda = \lambda_0$ under the assumption that $\boldsymbol{\Delta}_T \neq 0$, (ii) The parametric space Λ is compact; (iii) $J_{0,p}(\cdot)$ is a continuous function as consequence of the continuity of $\phi(\cdot)$, (iv) $J_{p,T}(\lambda)$ converges uniformly in probability to $J_{p,0}(\lambda)$ (shown below). Therefore, from Theorem 2.1 of Newey and McFadden (1994) we have that $\widehat{\lambda}_{0,p} \overset{p}{\to} \lambda_0$.

In Theorem 1.6 we show that $\boldsymbol{S}_T$ converges in distribution to $\boldsymbol{S}_T$. Hence, $\boldsymbol{S}_T$ is uniformly tight (in particular with respect to λ). Therefore, $\frac{1}{\sqrt{T}}\boldsymbol{S}_T(\lambda)$ is $o_p(1)$ uniformly in λ. Or equivalently, $\widehat{\boldsymbol{\Delta}}_T(\lambda) \overset{p}{\to} \boldsymbol{\Delta}_T(\lambda)$, uniformly in $\lambda \in \Lambda$.

Now consider any real valued function $f(\cdot)$ that is continuous on a compact set $K \subset \mathbb{R}^k$. In that case $f(\cdot)$ is uniformly continuous on K as every continuous function on a compact domain. By definition then, for a given $\epsilon > 0$, there is a $\delta > 0$ such that for every $(\boldsymbol{x}, \boldsymbol{y}) \in K^2$, $\{|f(\boldsymbol{x}) - f(\boldsymbol{y})| > \epsilon\} \Rightarrow \{\|\boldsymbol{x} - \boldsymbol{y}\| > \delta\}$. Therefore, $\mathbb{P}(|\|\boldsymbol{x}\|_p - \|\boldsymbol{y}\|_p| > \epsilon) \leq \mathbb{P}(\|\boldsymbol{x} - \boldsymbol{y}\| > \delta) + \mathbb{P}(K^c)$.

Finally, note that $\|\cdot\|_p$ is a a continuous function on $\mathbb{R}^q$ so given any $\epsilon > 0$, we can take a arbitrary large compact $K_\epsilon \subset \mathbb{R}^q$ such that $P(K^c) \leq \epsilon$. The result then follows since the first term above converges uniformly to zero in probability. ■

Proof of Proposition 1.9

Proof. Follows directly from Theorem 1.3 applied to each unit of $\mathcal{I}$ individually combined with the Cramèr-Wold device. ■

A.2 Proofs of Chapter 2

Hence, we can derive the following convergence results:

Lemma A.6 *let $\{\boldsymbol{u}_t\}$ is defined as*

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

If the process $\{\boldsymbol{\eta}_t\}$ satisfies Assumption 2.1, then as $T \to \infty$:

(a) $T^{1/2}\bar{\boldsymbol{\eta}} \xrightarrow{d} \boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1)$

(b) $T^{3/2}\check{\boldsymbol{\eta}} \xrightarrow{d} \sqrt{3}\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1)$

(c) $T^{-1/2}\bar{\boldsymbol{u}} \xrightarrow{d} \boldsymbol{\Omega}^{1/2}\int_0^1 \boldsymbol{W}(r)\mathrm{d}r = \frac{1}{3}\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1)$

(d) $T^{1/2}\check{\boldsymbol{u}} \xrightarrow{d} 3\boldsymbol{\Omega}^{1/2}\int_0^1 r\boldsymbol{W}(r)\mathrm{d}r = \frac{2}{5}\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1)$

(e) $T^{-2}\sum_{t=1}^T \boldsymbol{u}_t\dot{\boldsymbol{u}}_t' \xrightarrow{d} \boldsymbol{\Omega}^{1/2}\left[\int_0^1 \boldsymbol{W}(r)\boldsymbol{W}'(r)\mathrm{d}r - \int_0^1 \boldsymbol{W}(r)\mathrm{d}r\int_0^1 \boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2} \equiv \boldsymbol{R}$

(f) $T^{-2}\sum_{t=1}^T \boldsymbol{u}_t\ddot{\boldsymbol{u}}_t' \xrightarrow{d} \boldsymbol{\Omega}^{1/2}\left[\int_0^1 \boldsymbol{W}(r)\boldsymbol{W}'(r)\mathrm{d}r - 3\int_0^1 r\boldsymbol{W}(r)\mathrm{d}r\int_0^1 r\boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2} \equiv \boldsymbol{P}$

(g) $T^{-1}\sum_{t=1}^T \boldsymbol{u}_t\dot{\boldsymbol{\eta}}_t' \xrightarrow{d} \boldsymbol{\Omega}^{1/2}\left[\int_0^1 \boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r) - \int_0^1 \boldsymbol{W}(r)\mathrm{d}r\boldsymbol{W}'(1)\right]\boldsymbol{\Omega}^{1/2} + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_0 \equiv \boldsymbol{V}$

(h) $T^{-1}\sum_{t=1}^T \boldsymbol{u}_t\ddot{\boldsymbol{\eta}}_t' \xrightarrow{d} \boldsymbol{\Omega}^{1/2}\left[\int_0^1 \boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r) - \sqrt{3}\int_0^1 r\boldsymbol{W}(r)\mathrm{d}r\boldsymbol{W}'(1)\right]\boldsymbol{\Omega}^{1/2} + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_0 \equiv \boldsymbol{Q}$

(i) $T^{-1}\bar{\boldsymbol{y}} \xrightarrow{p} \frac{1}{2}\boldsymbol{\mu}$

(j) $\check{\boldsymbol{y}} \xrightarrow{p} \boldsymbol{\mu}$

(k) $T^{-3}\sum_{t=1}^T \dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t' \xrightarrow{p} \frac{1}{12}\boldsymbol{\mu}\boldsymbol{\mu}'$

(l) $T^{-3}\sum_{t=1}^T \ddot{\boldsymbol{y}}_t\ddot{\boldsymbol{y}}_t' \xrightarrow{p} \frac{1}{3}\boldsymbol{\mu}\boldsymbol{\mu}'$

(m) $T^{-1}\bar{\xi} \xrightarrow{p} \frac{1}{2}\gamma$

(n) $T^{-3}\sum_{t=1}^T \dot{\boldsymbol{y}}_t\xi_t \xrightarrow{p} \frac{1}{12}\gamma\boldsymbol{\mu}$

(o) $T^{-3/2}\sum_{t=1}^T \dot{\boldsymbol{y}}_t\boldsymbol{\eta}_t' \xrightarrow{d} \boldsymbol{\mu}\mathcal{N}\left(0, \frac{1}{12}\boldsymbol{\Omega}\right)$

(p) $T^{-3/2}\sum_{t=1}^T \ddot{\boldsymbol{y}}_t\boldsymbol{\eta}_t' \xrightarrow{d} \boldsymbol{\mu}\mathcal{N}\left(\boldsymbol{0}, \frac{1}{3}\boldsymbol{\Omega}\right)$,

where

$$\boldsymbol{\Omega}_0 \equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^T \mathbb{E}(\boldsymbol{\eta}_t\boldsymbol{\eta}_t')$$
$$\boldsymbol{\Omega}_1 \equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^T\sum_{s=1}^{t-1} \mathbb{E}(\boldsymbol{\eta}_s\boldsymbol{\eta}_t')$$
$$\boldsymbol{\Omega} \equiv \lim_{T\to\infty} T^{-1}\mathbb{V}\left(\sum_{t=1}^T \boldsymbol{\eta}_t\right) = \boldsymbol{\Omega}_0 + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_1'$$

and we adopt the following notation

$$\dot{\boldsymbol{u}}_t \equiv \boldsymbol{u}_t - \bar{\boldsymbol{u}}, \qquad \bar{\boldsymbol{u}} \equiv T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t, \tag{A-2}$$

$$\ddot{\boldsymbol{u}}_t \equiv \boldsymbol{u}_t - t\check{\boldsymbol{u}}, \qquad \check{\boldsymbol{u}} \equiv \frac{6}{T(T+1)(2T+1)}\sum_{t=1}^{T} t\boldsymbol{u}_t, \tag{A-3}$$

Proof. Under the assumptions of Proposition 2.1, $\boldsymbol{U}_T(r) \equiv T^{-1/2}\sum_{t=1}^{[rT]}\boldsymbol{\eta}_t \xrightarrow{d} \boldsymbol{\Omega}^{1/2}\boldsymbol{W}(r)$. Hence, for (a)

$$T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\eta}_t = \boldsymbol{U}_T(1) \xrightarrow{d} \boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1) \equiv \mathcal{N}\left(\boldsymbol{0},\boldsymbol{\Omega}\right).$$

For (b), note that

$$T^{-3/2}\sum_{t=1}^{T} t\boldsymbol{\eta}_t \xrightarrow{d} \tfrac{1}{\sqrt{3}}\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1) \equiv \mathcal{N}\left(\boldsymbol{0},\tfrac{1}{3}\boldsymbol{\Omega}\right).$$

Thus,

$$T^{3/2}\check{\boldsymbol{\eta}} = \frac{6T^3}{T(T+1)(2T+1)}T^{-3/2}\sum_{t=1}^{T} t\boldsymbol{\eta}_t \xrightarrow{d} \sqrt{3}\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1) \equiv \mathcal{N}\left(\boldsymbol{0},3\boldsymbol{\Omega}\right).$$

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

$$\begin{aligned}
T^{-3/2}\sum_{t=1}^{T}\boldsymbol{u}_t &= T^{-3/2}\sum_{t=1}^{T}\left(\boldsymbol{u}_{t-1}+\boldsymbol{\eta}_t\right)\\
&= \sum_{t=1}^{T}\int_{\frac{t-1}{T}}^{\frac{t}{T}}\boldsymbol{U}_T(r)\mathrm{d}r + o_p(1)\\
&= \int_0^1 \boldsymbol{U}_T(r)\mathrm{d}r + o_p(1)\\
&\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\int_0^1 \boldsymbol{W}(r)\mathrm{d}r.
\end{aligned}$$

We continue by showing result (c). Write:

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

Summing over $t = 1, \ldots, T$ and rearranging

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

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,

$$\begin{aligned} T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t' &= T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_{t-1}\boldsymbol{u}_{t-1}' + T^{-2}\sum_{t=1}^{T}\left(\boldsymbol{u}_{t-1}\boldsymbol{\eta}_t' + \boldsymbol{\eta}_t\boldsymbol{y}_{t-1}' + \boldsymbol{\eta}_t\boldsymbol{\eta}_t'\right) \\ &= \sum_{t=1}^{T}\int_{\frac{t-1}{T}}^{\frac{t}{T}}\boldsymbol{U}_T(r)\boldsymbol{U}_T'(r)\mathrm{d}r + o_p(1) \\ &= \int_0^1 \boldsymbol{U}_T(r)\boldsymbol{U}_T'(r)\mathrm{d}r + o_p(1) \\ &\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\int_0^1 \boldsymbol{W}(r)\boldsymbol{W}(r)'\mathrm{d}r\boldsymbol{\Omega}^{1/2}. \end{aligned}$$

To prove (d) we write

$$\begin{aligned} T^{-2}\sum_{t=1}^{T}\dot{\boldsymbol{u}}_t\dot{\boldsymbol{u}}_t' &\equiv T^{-2}\sum_{t=1}^{T}\left(\boldsymbol{u}_t - \bar{\boldsymbol{u}}\right)\left(\boldsymbol{u}_t - \bar{\boldsymbol{u}}\right)' \\ &= T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t' - T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\bar{\boldsymbol{y}}' - T^{-2}\bar{\boldsymbol{u}}\sum_{t=1}^{T}\boldsymbol{u}_t' + T^{-1}\bar{\boldsymbol{u}}\bar{\boldsymbol{u}}' \\ &= T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t' - T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\bar{\boldsymbol{u}}' - T^{-1}\bar{\boldsymbol{u}}\bar{\boldsymbol{u}}' + T^{-1}\bar{\boldsymbol{u}}\bar{\boldsymbol{u}}' \\ &= T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t' - \left(T^{-3/2}\sum_{t=1}^{T}\boldsymbol{u}_t\right)\left(T^{-3/2}\sum_{t=1}^{T}\boldsymbol{u}_t\right)' \\ &\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\left[\int_0^1 \boldsymbol{W}(r)\boldsymbol{W}'(r)\mathrm{d}r - \int_0^1 \boldsymbol{W}(r)\mathrm{d}r\int_0^1 \boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2}. \end{aligned}$$

To show (e), we first let $\boldsymbol{h}_t \equiv t\boldsymbol{u}_t = t\sum_{s=1}^{t}\boldsymbol{\eta}_t$ and define

$$\boldsymbol{H}_T(r) \equiv \frac{[rT]}{T}T^{-1/2}\sum_{t=1}^{[rT]}\boldsymbol{\eta}_t \xrightarrow{d} r\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(r).$$

Thus,

$$\boldsymbol{h}_{t-1} = T^{3/2}\boldsymbol{H}_T(\tfrac{t-1}{T}) = T^{5/2}\int_{\frac{t-1}{T}}^{\frac{t}{T}}\boldsymbol{H}_T(r)\mathrm{d}r$$

and $\boldsymbol{h}_t = t\boldsymbol{u}_t = t\left(\boldsymbol{u}_{t-1} + \boldsymbol{\eta}_t\right) = \boldsymbol{h}_{t-1} + \boldsymbol{u}_{t-1} + t\boldsymbol{\eta}_t$. Then,

$$\begin{aligned} T^{-5/2}\sum_{t=1}^{T}\boldsymbol{h}_t &= \sum_{t=1}^{T}\int_{\frac{t-1}{T}}^{\frac{t}{T}}\boldsymbol{H}_T(r)\mathrm{d}r + o_p(1) \\ &= \int_0^1 \boldsymbol{H}_T(r)\mathrm{d}r + o_p(1) \\ &\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\int_0^1 r\boldsymbol{W}(r)\mathrm{d}r. \end{aligned}$$

Therefore, using the previous result:

$$T^{1/2}\breve{\boldsymbol{u}} \equiv \frac{6T^3}{T(T+1)(2T+1)}T^{-5/2}\sum_{t=1}^{T_0}t\boldsymbol{u}_t \xrightarrow{d} 3\boldsymbol{\Omega}^{1/2}\int_0^1 r\boldsymbol{W}(r)\mathrm{d}r.$$

Result (e) is proved by writing

$$\begin{aligned} T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\ddot{\boldsymbol{u}}_t' &\equiv T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\left(\boldsymbol{u}_t - t\breve{\boldsymbol{u}}\right)' \\ &= T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t' - T^{-2}\sum_{t=1}^{T}t\boldsymbol{u}_t\breve{\boldsymbol{u}}' \\ &= T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t' - \frac{T(T+1)(2T+1)}{6T^3}T^{1/2}\breve{\boldsymbol{u}}T^{1/2}\breve{\boldsymbol{u}}' \\ &\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\left[\int_0^1 \boldsymbol{W}(r)\boldsymbol{W}'(r)\mathrm{d}r - 3\int_0^1 r\boldsymbol{W}(r)\mathrm{d}r\int_0^1 r\boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2}. \end{aligned}$$

To prove (f) we need the following result that was demonstrated by **?**

$$T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_{t-1}\boldsymbol{\eta}_t' \xrightarrow{d} \boldsymbol{\Omega}^{1/2}\int_0^1 \boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r)\boldsymbol{\Omega}^{1/2} + \boldsymbol{\Omega}_1.$$

Hence,

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{\eta}_t' &= T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_{t-1}\boldsymbol{\eta}_t' + T^{-1}\sum_{t=1}^{T}\boldsymbol{\eta}_t\boldsymbol{\eta}_t' \\ &\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\int_0^1 \boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r)\boldsymbol{\Omega}^{1/2} + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_0 \end{aligned}$$

Finally, (f) becomes

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}\dot{\boldsymbol{u}}_t\boldsymbol{\eta}_t' &= T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{\eta}_t' + T^{-1}\bar{\boldsymbol{u}}\sum_{t=1}^{T}\boldsymbol{\eta}_t' \\
&= T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{\eta}_t' + \left(T^{-3/2}\sum_{t=1}^{T}\boldsymbol{u}_t\right)\left(T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\eta}_t'\right) \\
&\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\int_0^1 \boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r)\boldsymbol{\Omega}^{1/2} + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_0 + \boldsymbol{\Omega}^{1/2}\int_0^1 \boldsymbol{W}(r)\mathrm{d}r\boldsymbol{W}'(1)\boldsymbol{\Omega}^{1/2} \\
&= \boldsymbol{\Omega}^{1/2}\left[\int_0^1 \boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r) + \int_0^1 \boldsymbol{W}(r)\mathrm{d}r\boldsymbol{W}'(1)\right]\boldsymbol{\Omega}^{1/2} + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_0.
\end{aligned}
$$

For (g):

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\ddot{\boldsymbol{\eta}}_t' &= T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{\eta}_t' - T^{-1}\sum_{t=1}^{T}t\boldsymbol{u}_t\check{\boldsymbol{\eta}}' \\
&= T^{-1}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{\eta}_t' - \frac{T(T+1)(2T+1)}{6T^3}T^{1/2}\check{\boldsymbol{u}}T^{3/2}\check{\boldsymbol{\eta}}' \\
&\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\left[\int_0^1 \boldsymbol{W}(r)\mathrm{d}\boldsymbol{W}'(r) - \sqrt{3}\int_0^1 r\boldsymbol{W}(r)\mathrm{d}r\boldsymbol{W}'(1)\right]\boldsymbol{\Omega}^{1/2} + \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_0.
\end{aligned}
$$

Consider $\boldsymbol{y}_t = \boldsymbol{\mu}t + \boldsymbol{u}_t$. Then, for (h)

$$
\begin{aligned}
T^{-2}\sum_{t=1}^{T}\boldsymbol{y}_t &= \boldsymbol{\mu}T^{-2}\sum_{t=1}^{T}t + T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t \\
&= \boldsymbol{\mu}T^{-1}(T+1)/2 + o_p(1) \\
&= \frac{1}{2}\boldsymbol{\mu} + o_p(1).
\end{aligned}
$$

Remember that $\sum_{t=1}^{T} t\boldsymbol{u}_t = \sum_{t=1}^{T}\boldsymbol{h}_t = O_p(5/2)$. Then,

$$
\begin{aligned}
T^{-3}\sum_{t=1}^{T}\boldsymbol{y}_t\boldsymbol{y}_t' &= T^{-3}\sum_{t=1}^{T}\left(\boldsymbol{\mu}t + \boldsymbol{u}_t\right)\left(\boldsymbol{\mu}t + \boldsymbol{u}_t\right)' \\
&= \boldsymbol{\mu}\boldsymbol{\mu}'T^{-3}\sum_{t=1}^{T}t^2 + \boldsymbol{\mu}\left(T^{-3}\sum_{t=1}^{T}t\boldsymbol{u}_t\right)' + \left(T^{-3}\sum_{t=1}^{T}t\boldsymbol{u}_t\right)\boldsymbol{\mu}' + T^{-3}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}_t' \\
&= \boldsymbol{\mu}\boldsymbol{\mu}'\frac{T(T+1)(2T+1)}{6T^3} + o_p(1) \\
&= \frac{1}{3}\boldsymbol{\mu}\boldsymbol{\mu}' + o_p(1).
\end{aligned}
$$

As a result, for (i) we have

$$\begin{aligned}
T^{-3}\sum_{t=1}^{T}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t' &= T^{-3}\sum_{t=1}^{T}\left(\boldsymbol{y}_t-\bar{\boldsymbol{y}}\right)\left(\boldsymbol{y}_t-\bar{\boldsymbol{y}}\right)' \\
&= T^{-3}\sum_{t=1}^{T}\boldsymbol{y}_t\boldsymbol{y}_t' - \left(T^{-2}\sum_{t=1}^{T}\boldsymbol{y}_t\right)\left(T^{-2}\sum_{t=1}^{T}\boldsymbol{y}_t\right) \\
&= \frac{1}{3}\boldsymbol{\mu}\boldsymbol{\mu}' - \frac{1}{2}\boldsymbol{\mu}\frac{1}{2}\boldsymbol{\mu}' + o_p(1) \\
&= \frac{1}{12}\boldsymbol{\mu}\boldsymbol{\mu}' + o_p(1).
\end{aligned}$$

For (j) we need:

$$\begin{aligned}
\breve{\boldsymbol{y}} &= \frac{6}{T(T+1)(2T+1)}\sum_{t=1}^{T}t\boldsymbol{y}_t \\
&= \frac{6}{T(T+1)(2T+1)}\sum_{t=1}^{T}t^2\boldsymbol{\mu} + \breve{\boldsymbol{u}} \\
&= \boldsymbol{\mu} + o_p(1).
\end{aligned}$$

Consequently,

$$\begin{aligned}
T^{-3}\sum_{t=1}^{T}\ddot{\boldsymbol{y}}_t\ddot{\boldsymbol{y}}_t' &= T^{-3}\sum_{t=1}^{T}\left(\boldsymbol{y}_t-\breve{\boldsymbol{y}}\right)\left(\boldsymbol{y}_t-\breve{\boldsymbol{y}}\right)' \\
&= T^{-3}\sum_{t=1}^{T}\boldsymbol{y}_t\boldsymbol{y}_t' - \left(T^{-2}\sum_{t=1}^{T}\boldsymbol{y}_t\right)T^{-1}\breve{\boldsymbol{y}}' - T^{-1}\breve{\boldsymbol{y}}\left(T^{-2}\sum_{t=1}^{T}\boldsymbol{y}_t'\right) + T^{-1}\breve{\boldsymbol{y}}T^{-1}\breve{\boldsymbol{y}}' \\
&= \frac{1}{3}\boldsymbol{\mu}\boldsymbol{\mu}' + o_p(1).
\end{aligned}$$

From the definitions we have that

$$\begin{aligned}
\dot{\boldsymbol{y}}_t &= \left(t-\tfrac{T+1}{2}\right)\boldsymbol{\mu} + \dot{\boldsymbol{u}}_t \quad \text{and} \\
\ddot{\boldsymbol{y}}_t &= \left(t-1\right)\boldsymbol{\mu} + \ddot{\boldsymbol{u}}_t.
\end{aligned}$$

For that reason,

$$\begin{aligned}
T^{-3/2}\sum_{t=1}^{T}\ddot{\boldsymbol{y}}_t\boldsymbol{\eta}_t' = \boldsymbol{\mu}T^{-3/2}\sum_{t=1}^{T}t\boldsymbol{\eta}_t' - \boldsymbol{\mu}T^{-3/2}\sum_{t=1}^{T}\boldsymbol{\eta}_t' + T^{-3/2}\sum_{t=1}^{T}\ddot{\boldsymbol{u}}\boldsymbol{\eta}_t' \\
\xrightarrow{d} \boldsymbol{\mu}\tfrac{1}{\sqrt{3}}\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1) \equiv \boldsymbol{\mu}\mathcal{N}\left(\boldsymbol{0},\tfrac{1}{3}\boldsymbol{\Omega}\right).
\end{aligned}$$

For (m) we have

$$
\begin{aligned}
T^{-3}\sum_{t=1}^{T}\boldsymbol{y}_t\xi_t &= T^{-3}\sum_{t=1}^{T}\left(\boldsymbol{\mu}t+\boldsymbol{u}_t\right)\left(\nu_t+t\gamma\right)\\
&= \boldsymbol{\mu}T^{-3}\sum_{t=1}^{T}t\nu_t+\boldsymbol{\mu}\gamma T^{-3}\sum_{t=1}^{T}t^2+T^{-3}\sum_{t=1}^{T}\boldsymbol{u}_t\nu_t+\gamma T^{-3}\sum_{t=1}^{T}t\boldsymbol{u}_t\\
&= \tfrac{1}{3}\gamma\boldsymbol{\mu}+o_p(1).
\end{aligned}
$$

Then,

$$
\begin{aligned}
T^{-3}\sum_{t=1}^{T}\dot{\boldsymbol{y}}_t\xi_t &= T^{-3}\sum_{t=1}^{T}\boldsymbol{y}_t\xi_t-\bar{\boldsymbol{y}}T^{-3}\sum_{t=1}^{T}\xi_t\\
&= T^{-3}\sum_{t=1}^{T}\boldsymbol{y}_t\xi_t-T^{-1}\bar{\boldsymbol{y}}T^{-1}\bar{\xi}_t\\
&= \tfrac{1}{3}\gamma\boldsymbol{\mu}-\tfrac{1}{2}\boldsymbol{\mu}\tfrac{1}{2}\gamma+o_p(1)\\
&= \tfrac{1}{12}\gamma\boldsymbol{\mu}+o_p(1).
\end{aligned}
$$

■

Proof of Lemma 2.1

Proof. It is straightforward to express the least-squares estimator as the difference to the true parameter value using notation (A-2)-(A-3) as:

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0=\left(\sum_{t=1}^{T_0}\boldsymbol{y}_{0t}\ddot{\boldsymbol{y}}_{0t}'\right)^{-1}\sum_{t=1}^{T_0}\boldsymbol{y}_{0t}\ddot{\nu}_t, \tag{A-4}$$

$$\widehat{\gamma}-\gamma_0=\check{\nu}-\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)'\check{\boldsymbol{y}}_0, \tag{A-5}$$

$$\widehat{\boldsymbol{\pi}}-\boldsymbol{\beta}_0 =\left(\sum_{t=1}^{T_0}\boldsymbol{y}_{0t}\dot{\boldsymbol{y}}_{0t}'\right)^{-1}\sum_{t=1}^{T_0}\boldsymbol{y}_{0t}\left[\gamma_0\left(t-\tfrac{T+1}{2}\right)+\dot{\nu}_t\right], \quad \text{and} \tag{A-6}$$

$$\widehat{\alpha}-\alpha_0=\tfrac{T+1}{2}\gamma_0+\bar{\nu}-\left(\widehat{\boldsymbol{\pi}}-\boldsymbol{\beta}_0\right)'\bar{\boldsymbol{y}}_0. \tag{A-7}$$

We use the limiting distributions in Lemma A.6 together with the continuous mapping theorem to show all the derivations below. Note that for $\boldsymbol{\mu}=0$, then $\boldsymbol{y}_t=\boldsymbol{u}_t$ and $\gamma_0=0$. As a result,

$$T\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)=\left(\tfrac{1}{T^2}\sum_{t\le T_0}\boldsymbol{y}_{0t}\ddot{\boldsymbol{y}}_{0t}'\right)^{-1}\tfrac{1}{T}\sum_{t\le T_0}\boldsymbol{y}_{0t}\ddot{\nu}_t\xrightarrow{d}\boldsymbol{P}_{00}^{-1}\boldsymbol{Q}_{01},$$

$$
\begin{aligned}
T^{3/2}\left(\widehat{\gamma}-\gamma_0\right) &= \tfrac{6T^3}{T_0(T_0+1)(2T_0+1)}\left[\left(\tfrac{1}{T^{3/2}}\sum_{t\leq T_0} t\nu_t\right) - T\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\left(\tfrac{1}{T^{5/2}}\sum_{t\leq T_0} t\boldsymbol{y}_{0t}\right)\right] \\
&\xrightarrow{d} \tfrac{3}{\lambda_0^3}\left\{\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} r\mathrm{d}\boldsymbol{W}(\mathrm{r})\right]_1 - \boldsymbol{Q}_{10}\boldsymbol{P}_{00}^{-1}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} r\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\},
\end{aligned}
$$

$$
\begin{aligned}
T\left(\widehat{\boldsymbol{\pi}}-\boldsymbol{\beta}_0\right) &= \left(\tfrac{1}{T^2}\sum_{t\leq T_0}\boldsymbol{y}_{0t}\dot{\boldsymbol{y}}_{0t}'\right)^{-1}\tfrac{1}{T}\sum_{t\leq T_0}^{T_0}\boldsymbol{y}_{0t}\dot{\nu}_t \\
&\xrightarrow{d} \boldsymbol{R}_{00}^{-1}\boldsymbol{V}_{01},
\end{aligned}
$$

and

$$
\begin{aligned}
\sqrt{T}\left(\widehat{\alpha}-\alpha_0\right) &= \tfrac{T}{T_0}\left[\left(\tfrac{1}{\sqrt{T}}\sum_{t\leq T_0}\nu_t\right) - T\left(\widehat{\boldsymbol{\pi}}-\boldsymbol{\beta}_0\right)'\left(\tfrac{1}{T^{3/2}}\sum_{t\leq T_0}\boldsymbol{y}_0\right)\right] \\
&\xrightarrow{d} \tfrac{1}{\lambda_0}\left\{\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}\mathrm{d}\boldsymbol{W}(r)\right]_1 - \boldsymbol{V}_{10}\boldsymbol{R}_{00}^{-1}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\}.
\end{aligned}
$$

For $\mu_0 \neq 0$ and $n = 2$,

$$
\begin{aligned}
\widehat{\pi}-\beta_0 &= \left(T_0^{-3}\sum_{t=1}^{T_0} y_{0t}\dot{y}_{0t}\right)^{-1} T_0^{-3}\sum_{t=1}^{T_0} y_{0t}\left[\gamma_0\left(t-\tfrac{T+1}{2}\right)+\dot{\nu}_t\right] \\
&= \left(T_0^{-3}\sum_{t=1}^{T_0} y_{0t}\dot{y}_{0t}\right)^{-1}\left[\tfrac{T(T+1)(2T+1)}{6T^3}\breve{y}_0 - \tfrac{T(T+1)}{2T^2}T^{-1}\bar{y}_0\right]\gamma_0 + o_P(1) \\
&= \left(\tfrac{1}{12}\mu_0^2\right)^{-1}\left[\tfrac{1}{3}\mu_0 - \tfrac{1}{2}\tfrac{1}{2}\mu_0\right]\gamma_0 + o_P(1) \\
&= \frac{\gamma_0}{\mu_0} + o_P(1)
\end{aligned}
$$

$$
\begin{aligned}
T_0^{-1}\left(\widehat{\alpha}-\alpha_0\right) &= \frac{T_0+1}{2T_0}\gamma_0 + T_0^{-1}\bar{\nu} - \left(\widehat{\pi}-\beta_0\right)T_0^{-1}\bar{\boldsymbol{y}}_0 \\
&= \frac{1}{2}\gamma_0 - \frac{\gamma_0}{\mu_0}\frac{\mu_0}{2} + o_P(1) \\
&= o_P(1).
\end{aligned}
$$

■

Proof of Theorem 2.2

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

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

Therefore,

$$\begin{aligned}
\widehat{\Delta}_1 - \Delta &= \tfrac{1}{T_2}\sum_{t>T_0}\left(\widehat{\delta}_{1t} - \delta_t\right)\\
&= \tfrac{1}{T_2}\sum_{t>T_0}\nu_t - \tfrac{T+T_0+1}{2}(\widehat{\gamma} - \gamma_0) - \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)'\tfrac{1}{T_2}\sum_{t>T_0}\boldsymbol{y}_{0t}\\
&= \left[\tfrac{1}{T_2}\sum_{t>T_0}\nu_t - \varphi(T,T_0)\sum_{t\le T_0}t\nu_t\right] - \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)'\left[\tfrac{1}{T_2}\sum_{t>T_0}\boldsymbol{y}_{0t} - \varphi(T,T_0)\sum_{t\le T_0}t\boldsymbol{y}_{0t}\right],
\end{aligned}$$

where $\varphi(T, T_0) \equiv \frac{3(T+T_0+1)}{T_0(T_0+1)(2T_0+2)}$; and

$$\begin{aligned}
\widehat{\Delta}_2 - \Delta &= \tfrac{1}{T_2}\sum_{t=T_0}^{T}\left(\widehat{\delta}_{2t} - \delta_t\right)\\
&= \tfrac{1}{T_2}\sum_{t=T_0}^{T}\nu_t - \widehat{\alpha} - (\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0)'\,\tfrac{1}{T_2}\sum_{t=T_0}^{T}\boldsymbol{y}_{0t}\\
&= \left[\tfrac{1}{T_2}\sum_{t=T_0+1}^{T}\nu_t - \tfrac{1}{T_0}\sum_{t=1}^{T_0}\nu_t\right] - (\widehat{\boldsymbol{\pi}} - \boldsymbol{\beta}_0)'\left[\tfrac{1}{T_2}\sum_{t=T_0+}^{T}\boldsymbol{y}_{0t} - \tfrac{1}{T_0}\sum_{t=T_0}^{T}\boldsymbol{y}_{0t}\right]
\end{aligned}$$

From the expression above is easy to see that for the case $\boldsymbol{\mu} = 0(\gamma_0 = 0)$ both estimators are consistent under the null $\Delta\mu = 0$. In fact,

$$\begin{aligned}
\sqrt{T}\left(\widehat{\Delta}_1 - \Delta\right) = \tfrac{T}{T_2}&\left(\tfrac{1}{\sqrt{T}}\sum_{t>T_0}\nu_t\right) - T^2\varphi(T,T_0)\left(\tfrac{1}{T^{3/2}}\sum_{t\le T_0}t\nu_t\right)\\
&- T\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)'\left[\tfrac{T}{T_2}\left(\tfrac{1}{T^{3/2}}\sum_{t>T_0}\boldsymbol{y}_{0t}\right) - T^2\varphi(T,T_0)\left(\tfrac{1}{T^{5/2}}\sum_{t\le T_0}t\boldsymbol{y}_{0t}\right)\right]\\
&\xrightarrow{d} \tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_{\lambda_0}^{1}\mathrm{d}\boldsymbol{W}\right]_1 - \tfrac{3(1+\lambda_0)}{2\lambda_0^3}\left[\boldsymbol{\Omega}^{1/2}\int_{0}^{\lambda_0}r\,\mathrm{d}\boldsymbol{W}\right]_1\\
&- \boldsymbol{Q}_{10}\boldsymbol{P}_{00}^{-1}\left\{\tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_{\lambda_0}^{1}\boldsymbol{W}(r)\mathrm{d}r\right]_0 - \tfrac{3(1+\lambda_0)}{2\lambda_0^3}\left[\boldsymbol{\Omega}^{1/2}\int_{0}^{\lambda_0}r\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\}\\
&\equiv \boldsymbol{c}_1 - \boldsymbol{Q}_{10}\boldsymbol{P}_{00}^{-1}\boldsymbol{d}_0.
\end{aligned}$$

For the second specification we have:

$$\begin{aligned}
\sqrt{T}\left(\widehat{\Delta}_2-\Delta\right) &= \tfrac{T}{T_2}\left(\tfrac{1}{\sqrt{T}}\sum_{t>T_0}\nu_t\right)-\sqrt{T}\widehat{\alpha}-T\left(\widehat{\boldsymbol{\pi}}-\boldsymbol{\beta}_0\right)'\tfrac{T}{T_2}\left(\tfrac{1}{T^{3/2}}\sum_{t>T_0}\boldsymbol{y}_{0t}\right)\\
&\xrightarrow{d} \tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1)-\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(\lambda_0)\right]_1\\
&\quad-\tfrac{1}{\lambda_0}\left\{\left[\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(\lambda_0)\right]_1-\boldsymbol{V}_{10}\boldsymbol{R}_{00}^{-1}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\}\\
&\quad-\boldsymbol{V}_{10}\boldsymbol{R}_{00}^{-1}\tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_{\lambda_0}^{1}\boldsymbol{W}(r)\mathrm{d}r\right]_0\\
&= \tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(1)\right]_1-\tfrac{1}{(1-\lambda)\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\boldsymbol{W}(\lambda_0)\right]_1\\
&\quad-\boldsymbol{V}_{10}\boldsymbol{R}_{00}^{-1}\left\{\tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_{\lambda_0}^{1}\boldsymbol{W}(r)\mathrm{d}r\right]_0-\tfrac{1}{\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\}\\
&\equiv \boldsymbol{a}_1-\boldsymbol{V}_{10}\boldsymbol{R}_{00}^{-1}\boldsymbol{b}_0.
\end{aligned}$$

■

Proof of Lemma 2.2

Proof. The least square estimator are

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\sum_{t\le T_0}\boldsymbol{y}_{0t}\ddot{\boldsymbol{y}}_{0t}'\right)^{-1}\sum_{t\le T_0}\boldsymbol{y}_{0t}\ddot{y}_{1t}\\
\widehat{\gamma} &= \check{y}_1-\widehat{\boldsymbol{\beta}}'\check{\boldsymbol{y}}_0\\
\widehat{\boldsymbol{\pi}} &= \left(\sum_{t=1}^{T_0}\boldsymbol{y}_{0t}\dot{\boldsymbol{y}}_{0t}'\right)^{-1}\sum_{t=1}^{T_0}\boldsymbol{y}_{0t}\dot{y}_{1t}\\
\widehat{\alpha} &= \bar{y}_1-\widehat{\boldsymbol{\pi}}'\bar{\boldsymbol{y}}_0
\end{aligned}$$

For the case $\boldsymbol{\mu}=0$, we have that $\boldsymbol{y}_t=\boldsymbol{u}_t$. As a consequence, by the continuous mapping theorem combined with the results of Lemma A.6:

$$\widehat{\boldsymbol{\beta}}=\left[\tfrac{1}{T^2}\sum_{t\le T_0}\boldsymbol{u}_t\ddot{\boldsymbol{u}}_t'\right]_{00}^{-1}\left[\tfrac{1}{T^2}\sum_{t\le T_0}\boldsymbol{u}_t\ddot{\boldsymbol{u}}_t'\right]_{01}\xrightarrow{d}\boldsymbol{P}_{00}^{-1}\boldsymbol{P}_{01},$$

$$\begin{aligned}
\sqrt{T}\widehat{\gamma} &= \tfrac{6T^3}{T_0(T_0+1)(2T_0+1)}\left[\left(\tfrac{1}{T^{5/2}}\sum_{t\le T_0}ty_{1t}^{(0)}\right)-\widehat{\boldsymbol{\beta}}'\left(\tfrac{1}{T^{5/2}}\sum_{t\le T_0}t\boldsymbol{y}_{0t}\right)\right]\\
&\xrightarrow{d}\tfrac{3}{\lambda_0^3}\left\{\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}r\boldsymbol{W}(r)\mathrm{d}r\right]_1-\boldsymbol{P}_{10}\boldsymbol{P}_{00}^{-1}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}r\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\},
\end{aligned}$$

$$\widehat{\boldsymbol{\pi}} = \left[\tfrac{1}{T^2}\sum_{t=1}^{T}\boldsymbol{u}_t\dot{\boldsymbol{u}}_t'\right]_{00}^{-1}\left[\tfrac{1}{T^2}\sum_{t=1}^{T}\boldsymbol{u}_t\dot{\boldsymbol{u}}_t'\right]_{01} \xrightarrow{d} \boldsymbol{R}_{00}^{-1}\boldsymbol{R}_{01},$$

and

$$\begin{aligned}\tfrac{1}{\sqrt{T}}\widehat{\alpha} &= \tfrac{T}{T_0}\left[\left(\tfrac{1}{T^{3/2}}\sum_{t\le T_0}\boldsymbol{y}_{1t}^{(0)}\right) - \widehat{\boldsymbol{\pi}}'\left(\tfrac{1}{T^{3/2}}\sum_{t\le T_0}\boldsymbol{y}_0\right)\right]\\ &\xrightarrow{d} \tfrac{1}{\lambda_0}\left\{\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}\boldsymbol{W}(r)\mathrm{d}r\right]_1 - \boldsymbol{R}_{10}\boldsymbol{R}_{00}^{-1}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\}.\end{aligned}$$

■

Proof of Theorem 2.3

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

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

Therefore,

$$\begin{aligned}\widehat{\Delta}_1 - \Delta &= \tfrac{1}{T_2}\sum_{t>T_0}\left(\widehat{\delta}_{1t} - \delta_t\right)\\ &= \tfrac{1}{T_2}\sum_{t>T_0} y_{1t}^{(0)} - \tfrac{T+T_0+1}{2}\widehat{\gamma} - \widehat{\boldsymbol{\beta}}'\tfrac{1}{T_2}\sum_{t>T_0}\boldsymbol{y}_{0t}\\ &= \left[\tfrac{1}{T_2}\sum_{t>T_0} y_{1t}^{(0)} - \varphi(T,T_0)\sum_{t\le T_0} t y_{1t}^{(0)}\right] - \widehat{\boldsymbol{\beta}}'\left[\tfrac{1}{T_2}\sum_{t>T_0}\boldsymbol{y}_{0t} - \varphi(T,T_0)\sum_{t\le T_0} t\boldsymbol{y}_{0t}\right]\end{aligned}$$

and,

$$\begin{aligned}\widehat{\Delta}_2 - \Delta &= \tfrac{1}{T_2}\sum_{t>T_0}\left(\widehat{\delta}_{2t} - \delta_t\right)\\ &= \tfrac{1}{T_2}\sum_{t>T_0} y_{1t}^{(0)} - \widehat{\alpha} - \widehat{\boldsymbol{\pi}}'\tfrac{1}{T_2}\sum_{t>T_0}\boldsymbol{y}_{0t}\\ &= \left[\tfrac{1}{T_2}\sum_{t>T_0} y_{1t}^{(0)} - \tfrac{1}{T_0}\sum_{t\le T_0} y_{1t}^{(0)}\right] - \widehat{\boldsymbol{\pi}}'\left[\tfrac{1}{T_2}\sum_{t>T_0}\boldsymbol{y}_{0t} - \tfrac{1}{T_0}\sum_{t\le T_0}\boldsymbol{y}_{0t}\right]\end{aligned}$$

Combining the results from Lemma 2 with the Continuous Mapping

Theorem we have the following convergence in distribution:

$$
\begin{aligned}
\tfrac{1}{\sqrt{T}}\left(\widehat{\Delta}_1-\Delta\right) &= \tfrac{T}{T_2}\left(\tfrac{1}{T^{3/2}}\sum_{t>T_0} y_{1t}^{(0)}\right) - T^2\varphi(T,T_0)\left(\tfrac{1}{T^{5/2}}\sum_{t\leq T_0} t y_{1t}^{(0)}\right)\\
&\quad - \widehat{\boldsymbol{\beta}}'\left[\tfrac{T}{T_2}\left(\tfrac{1}{T^{3/2}}\sum_{t>T_0}\boldsymbol{y}_{0t}\right) - T^2\varphi(T,T_0)\left(\tfrac{1}{T^{5/2}}\sum_{t\leq T_0} t\boldsymbol{y}_{0t}\right)\right]\\
&\xrightarrow{d} \tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_{\lambda_0}^{1}\boldsymbol{W}(r)\mathrm{d}r\right]_1 - \tfrac{3(1+\lambda_0)}{2\lambda_0^3}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} r\boldsymbol{W}(r)\mathrm{d}r\right]_1\\
&\quad - \boldsymbol{P}_{10}\boldsymbol{P}_{00}^{-1}\left\{\tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_{\lambda_0}^{1}\boldsymbol{W}(r)\mathrm{d}r\right]_0 - \tfrac{3(1+\lambda_0)}{2\lambda_0^3}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} r\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\}\\
&\equiv \boldsymbol{d}_1 - \boldsymbol{P}_{10}\boldsymbol{P}_{00}^{-1}\boldsymbol{d}_0,
\end{aligned}
$$

and

$$
\begin{aligned}
\tfrac{1}{\sqrt{T}}\left(\widehat{\Delta}_2-\Delta\right) &= \tfrac{T}{T_2}\left(\tfrac{1}{T^{3/2}}\sum_{t>T_0} y_{1t}^{(0)}\right) - \tfrac{T}{T_0}\left(\tfrac{1}{T^{3/2}}\sum_{t\leq T_0} y_{1t}^{(0)}\right)\\
&\quad - \widehat{\boldsymbol{\pi}}'\left[\tfrac{T}{T_2}\left(\tfrac{1}{T^{3/2}}\sum_{t>T_0}\boldsymbol{y}_{0t}\right) - \tfrac{T}{T_0}\left(\tfrac{1}{T^{3/2}}\sum_{t\leq T_0}\boldsymbol{y}_{0t}\right)\right]\\
&\xrightarrow{d} \tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_{\lambda_0}^{1}\boldsymbol{W}(r)\mathrm{d}r\right]_1 - \tfrac{1}{\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}\boldsymbol{W}(r)\mathrm{d}r\right]_1\\
&\quad - \boldsymbol{R}_{10}\boldsymbol{R}_{00}^{-1}\left\{\tfrac{1}{1-\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_{\lambda_0}^{1}\boldsymbol{W}(r)\mathrm{d}r\right]_0 - \tfrac{1}{\lambda_0}\left[\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0}\boldsymbol{W}(r)\mathrm{d}r\right]_0\right\}\\
&\equiv \boldsymbol{b}_1 - \boldsymbol{R}_{10}\boldsymbol{R}_{00}^{-1}\boldsymbol{b}_0.
\end{aligned}
$$

■

Proof of Lemma 2.3

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

$$
\begin{aligned}
\widehat{\nu}_{1t} &= \dot{\nu}_t - (\widehat{\gamma}-\gamma_0)(t-\tfrac{T+T_0+1}{2}) - (\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t} + \dot{\delta}_t\\
\widehat{\nu}_{2t} &= \dot{\nu}_t - (\widehat{\boldsymbol{\pi}}-\boldsymbol{\beta}_0)'\dot{\boldsymbol{y}}_{0t} + \dot{\delta}_t.
\end{aligned}
$$

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

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

Both $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\pi}}-\boldsymbol{\beta}_0$ are $O_P(\frac{1}{T})$ by Lemma 2.1. Also, $\sum \dot{\boldsymbol{y}}_{0t}\dot{\boldsymbol{y}}'_{0t+k} = O_P(T^2)$; and $\sum \dot{\nu}_t \dot{\boldsymbol{y}}_{0t+k} = O_P(T)$ all as a consequence of Lemma A.6. Thus for $j \in \{1,2\}$, we have:

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

where the last equality involves no more than some algebraic manipulation using the definition of $\dot{\nu}_t$ and $\ddot{\nu}_t$ and neglecting the $o_P(1)$ terms. Therefore, by the Law Large Numbers, which is ensured under Assumption 2.2,

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

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

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

Under Assumption A of Andrews (1991), which is implied by Assumption 2.2, the second term is $o_P(1)$. Assumption B of Andrews (1991), which ensures the first term to be $o_P(1)$ is not fulfilled directly by specification (2-7) due to the trend regressor. However, what is really necessary for the result is to bound the mean value expansion of the first term, which in our case, is simply given by

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

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

■

Proof of Theorem 2.4

We can decompose the t-statistic as:

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

Under $\mathcal{H}_0$ the second term is zero and the first term converges in distribution by the Slutsky Theorem since the numerator of the term between parentheses converges in distribution according to Theorem 2.2, and the denominator converges in probability according to the Lemma 2.3, hence

$$\begin{aligned}
\tau_1 &\xrightarrow{d} \tfrac{\sqrt{1-\lambda_0}}{\omega}\left[\boldsymbol{c}_1 - \boldsymbol{Q}_{10}\boldsymbol{P}_{00}^{-1}\boldsymbol{d}_0\right] \\
\tau_2 &\xrightarrow{d} \tfrac{\sqrt{1-\lambda_0}}{\omega}\left[\boldsymbol{a}_1 - \boldsymbol{V}_{10}\boldsymbol{R}_{00}^{-1}\boldsymbol{b}_0\right]
\end{aligned}$$

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

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

Lemma A.7 *If the process $\{\boldsymbol{\varepsilon}_t\}$ satisfies the Assumption 2.1, then as $T \to \infty$, for any $k \geq 0$*

$$T^{-2}\sum_{t=1}^{T}\boldsymbol{u}_t\boldsymbol{u}'_{t+k} \xrightarrow{d} \boldsymbol{\Sigma}^{1/2}\int_0^1 \boldsymbol{W}(r)\boldsymbol{W}'(r)\mathrm{d}r\boldsymbol{\Sigma}^{1/2}.$$

Proof. We consider for $k \geq 0$ that $\boldsymbol{v}_{t+k} = \boldsymbol{v}_t + \sum_i^k \boldsymbol{\varepsilon}_{t+i}$. Then,

$$T^{-2}\sum_{t=1}^{T}\boldsymbol{v}_t\boldsymbol{v}'_{t+k} = T^{-2}\sum_{t=1}^{T}\boldsymbol{v}_t\boldsymbol{v}'_t + T^{-2}\sum_{t=1}^{T}\boldsymbol{v}_t\sum_{i=1}^{k}\boldsymbol{\varepsilon}'_{t+i}.$$

We show that $T^{-1}\sum_{t=1}^{T}\boldsymbol{v}_t\boldsymbol{\varepsilon}'_{t+i} = O_P(1)$ for every $i \in \{1,\ldots,k\}$. For that purpose, define for any integer j,

$$\boldsymbol{U}_T^j(r) = \left(\tfrac{1}{T}\right)^{1/2}\sum_{t=1}^{[rT]}\boldsymbol{\varepsilon}_{t+j}.$$

Hence,

$$T^{-1}\sum_{t=1}^{T}\boldsymbol{y}_{t-1}\boldsymbol{\varepsilon}'_{t+j} = \sum_{t=1}^{T}\boldsymbol{U}_T^0\left(\tfrac{t-1}{T}\right)\int_{\frac{t-1}{T}}^{\frac{T}{t}}\mathrm{d}\boldsymbol{U}_T^j(r) = \int_0^1 \boldsymbol{U}_T^0(r)\mathrm{d}\boldsymbol{U}_T^j(r).$$

Let $\boldsymbol{\Omega}_j \equiv \lim_{T\to\infty} T^{-1}\mathbb{E}\left(\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t \sum_{t=1}^{T} \boldsymbol{\varepsilon}_{t+j}'\right)$. Clearly, $\boldsymbol{\Omega}_0 = \boldsymbol{\Omega}$. Then,

$$\begin{bmatrix} \boldsymbol{U}_T^0(r) \\ \boldsymbol{U}_T^j(r) \end{bmatrix} \xrightarrow{d} \tilde{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{W}(r)\tilde{\boldsymbol{\Sigma}}^{1/2} \equiv \begin{bmatrix} \boldsymbol{U}^0(r) \\ \boldsymbol{U}^j(r) \end{bmatrix} \quad \text{where } \tilde{\boldsymbol{\Sigma}} \equiv \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}_j \\ \boldsymbol{\Sigma}_j' & \boldsymbol{\Sigma} \end{bmatrix}.$$

For $j \geq 0$, the process $\boldsymbol{\varepsilon}_{t+j}$ is a martingale with respect to the process $\boldsymbol{y}_{t-1}$. Thus, we have a sufficient condition to apply Theorem 2.1 developed by Kurtz and Protter (1991) and also restated in Hansen (1992) that

$$T^{-1}\sum_{t=1}^{T} \boldsymbol{y}_{t-1}\boldsymbol{\varepsilon}_{t+j}' = \int_0^1 \boldsymbol{U}_T^0(r)\mathrm{d}\boldsymbol{U}_T^j(r) \xrightarrow{d} \int_0^1 \boldsymbol{U}^0(r)\mathrm{d}\boldsymbol{U}^j(r).$$

Note that the stochastic integral above is not easy to evaluate except for when $j = 0$. In that case we have the particular result shown in **?** and used to prove part (e) of Lemma A.6 above. However, for our purposes, is enough to known that the distribution exists and hence the term is $O_P(1)$ for any non-negative j. Therefore, for every $i \in \{1, \ldots, k\}$, we have $T^{-2}\sum_{t=1}^{T} \boldsymbol{v}_{t-1}\boldsymbol{\varepsilon}_{t+i-1}' = o_P(1)$. Thus, we have the desired result as a finite sum of $o_P(1)$ terms.

■

Proof of Lemma 2.4

Proof. First we show the following result: For $\lambda < \lambda'$:

$$\boldsymbol{w}_t(\lambda, \lambda') \equiv \boldsymbol{u}_t - \tfrac{1}{T_{\lambda'}-T_\lambda} \sum_{T_\lambda < s \leq T_{\lambda'}} \boldsymbol{u}_s$$
$$\boldsymbol{x}_t(\lambda, \lambda') \equiv \boldsymbol{u}_t - \tfrac{1}{T_2-T_1} \sum_{T_1 \leq s \leq T_2} \boldsymbol{u}_s$$

, where $T_\lambda = \lfloor \lambda T \rfloor$ and $\sum_{(\lambda,\lambda']} \equiv \sum_{T_\lambda < t \leq T_{\lambda'}}$, then

$$\begin{aligned}
\tfrac{1}{T^2}\sum_{(\lambda,\lambda']} \boldsymbol{w}_t\boldsymbol{w}_t' &= \tfrac{1}{T^2}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\boldsymbol{u}_t' - \tfrac{1}{T^2(T_{\lambda'}-T_\lambda)}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t \sum_{(\lambda,\lambda']} \boldsymbol{u}_s' - \tfrac{1}{T^2(T_{\lambda'}-T_\lambda)}\sum_{(\lambda,\lambda']} \boldsymbol{u}_s \sum_{(\lambda,\lambda']} \boldsymbol{u}_t' \\
&\quad + \tfrac{1}{T^2(T_{\lambda'}-T_\lambda)^2}\sum_{(\lambda,\lambda']}\sum_{(\lambda,\lambda']} \boldsymbol{u}_s \sum_{(\lambda,\lambda']} \boldsymbol{u}_k' \\
&= \tfrac{1}{T^2}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\boldsymbol{u}_t' - \tfrac{1}{T^2(T_{\lambda'}-T_\lambda)}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t \sum_{(\lambda,\lambda']} \boldsymbol{u}_t' \\
&= \tfrac{1}{T^2}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\boldsymbol{u}_t' + \tfrac{T}{T_{\lambda'}-T_\lambda}\left(\tfrac{1}{T^{3/2}}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\right)\left(\tfrac{1}{T^{3/2}}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\right)' \\
&\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\left[\int_\lambda^{\lambda'} \boldsymbol{W}(r)\boldsymbol{W}(r)'\mathrm{d}r + \tfrac{1}{\lambda'-\lambda}\int_\lambda^{\lambda'} \boldsymbol{W}(r)\mathrm{d}r \int_\lambda^{\lambda'} \boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2} \\
&\equiv \boldsymbol{R}(\lambda,\lambda')
\end{aligned}$$

$$\begin{aligned}
\tfrac{1}{T^2}\sum_{(\lambda,\lambda']} \boldsymbol{x}_t\boldsymbol{x}_t' &= \tfrac{1}{T^2}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\boldsymbol{u}_t' - \tfrac{1}{T^2(T_{\lambda'}-T_\lambda)}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t \sum_{(\lambda,\lambda']} \boldsymbol{u}_s' - \tfrac{1}{T^2(T_{\lambda'}-T_\lambda)}\sum_{(\lambda,\lambda']} \boldsymbol{u}_s \sum_{(\lambda,\lambda']} \boldsymbol{u}_t' \\
&\quad + \tfrac{1}{T^2(T_{\lambda'}-T_\lambda)^2}\sum_{(\lambda,\lambda']}\sum_{(\lambda,\lambda']} \boldsymbol{u}_s \sum_{(\lambda,\lambda']} \boldsymbol{u}_k' \\
&= \tfrac{1}{T^2}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\boldsymbol{u}_t' - \tfrac{1}{T^2(T_{\lambda'}-T_\lambda)}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t \sum_{(\lambda,\lambda']} \boldsymbol{u}_t' \\
&= \tfrac{1}{T^2}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\boldsymbol{u}_t' + \tfrac{T}{T_{\lambda'}-T_\lambda}\left(\tfrac{1}{T^{3/2}}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\right)\left(\tfrac{1}{T^{3/2}}\sum_{(\lambda,\lambda']} \boldsymbol{u}_t\right)' \\
&\xrightarrow{d} \boldsymbol{\Omega}^{1/2}\left[\int_\lambda^{\lambda'} \boldsymbol{W}(r)\boldsymbol{W}(r)'\mathrm{d}r + \tfrac{1}{\lambda'-\lambda}\int_\lambda^{\lambda'} \boldsymbol{W}(r)\mathrm{d}r \int_\lambda^{\lambda'} \boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2} \\
&\equiv \boldsymbol{R}(\lambda,\lambda')
\end{aligned}$$

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

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

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

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

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

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

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

where

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

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

$$
\begin{aligned}
\tfrac{1}{T}\widehat{\rho}_{10}^2 &= \tfrac{1}{TT_2}\sum_{t>T_0}\widehat{\nu}_{1t}^2\\
&= \tfrac{1}{TT_2}\left[\widehat{\boldsymbol{\theta}}_1'\left(\sum_{t>T_0}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t'\right)\widehat{\boldsymbol{\theta}}_1 - 2\widehat{\gamma}\widehat{\boldsymbol{\theta}}_1'\sum_{t>T_0}\left(t-\tfrac{T+T_0+1}{2}\right)\dot{\boldsymbol{y}}_t + \widehat{\gamma}^2\sum_{t>T_0}\left(t-\tfrac{T+T_0+1}{2}\right)^2\right]\\
&= \tfrac{T}{T_2}\widehat{\boldsymbol{\theta}}_1'\left[\left(\tfrac{1}{T^2}\sum_{t>T_0}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t'\right) - 2\left(\tfrac{1}{T^{5/2}}\sum_{t>T_0}\left(t-\tfrac{T+T_0+1}{2}\right)\dot{\boldsymbol{y}}_t\right)(\sqrt{T}\check{\boldsymbol{y}})'\right.\\
&\qquad\left. + \tfrac{1}{T^3}\sum_{t>T_0}\left(t-\tfrac{T+T_0+1}{2}\right)^2(\sqrt{T}\check{\boldsymbol{y}})(\sqrt{T}\check{\boldsymbol{y}})'\right]\widehat{\boldsymbol{\theta}}_1\\
&= \tfrac{T}{T_2}\widehat{\boldsymbol{\theta}}_1'\left\{\left(\tfrac{1}{T^2}\sum_{t>T_0}\dot{\boldsymbol{y}}_t\dot{\boldsymbol{y}}_t'\right) - \left[2\left(\tfrac{1}{T^{5/2}}\sum_{t>T_0}t\dot{\boldsymbol{y}}_t\right) - \tfrac{1}{T^3}\sum_{t>T_0}\left(t-\tfrac{T+T_0+1}{2}\right)^2(\sqrt{T}\check{\boldsymbol{y}})\right](\sqrt{T}\check{\boldsymbol{y}})'\right\}\widehat{\boldsymbol{\theta}}_1\\
&\xrightarrow{d} \tfrac{1}{1-\lambda_0}\widetilde{\boldsymbol{f}}'\left\{\boldsymbol{H} - 2\left[\boldsymbol{k} - \left(\tfrac{1-\lambda_0^3}{3} - \tfrac{(1-\lambda_0)^3}{4}\right)\boldsymbol{j}\right]\boldsymbol{j}'\right\}\widetilde{\boldsymbol{f}},
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{H} &\equiv \boldsymbol{\Omega}^{1/2}\left[\int_{\lambda_0}^1 \boldsymbol{W}(r)\boldsymbol{W}(r)'\mathrm{d}r - \tfrac{1}{1-\lambda_0}\int_{\lambda_0}^1 \boldsymbol{W}(r)\mathrm{d}r\int_{\lambda_0}^1 \boldsymbol{W}'(r)\mathrm{d}r\right]\boldsymbol{\Omega}^{1/2}\\
\boldsymbol{j} &\equiv 3\boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} r\boldsymbol{W}(r)\mathrm{d}r\\
\boldsymbol{k} &\equiv \boldsymbol{\Omega}^{1/2}\int_0^{\lambda_0} r\boldsymbol{W}(r)\mathrm{d}r
\end{aligned}
$$

Similarly, for the second specification we have:

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

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

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

and we have shown that first and second terms inside the brackets of the expressions above are $O_P(1)$ by Lemma A.7 and Lemma A.6 respectively.

Finally, the remaninder terms of $\boldsymbol{M}_{1k}$ are all $o_P(1)$ by simply by applying the convergence results presented in Lemma A.6. Therefore, we have proved part (a) and (b).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Hence, we are left with

$$T^{-1}\widehat{\sigma}_{jT}^2 = T^{-1}\widehat{\rho}_{j0}\sum_{|k|<T}\phi\left(\tfrac{k}{J_T}\right) + o_P(1).$$

If we multiply the above expression by J_T^{-1}, we get

$$(J_T T)^{-1}\widehat{\sigma}_{jT}^2 = T^{-1}\widehat{\rho}_{j0}\left(J_T^{-1}\sum_{|k|<T}\phi\left(\tfrac{k}{J_T}\right)\right) + o_P(1).$$

By taking the limit as $T\to\infty$ we get the desired result. ■

Proof of Theorem 2.5

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

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

As long as $\Delta_T = o(\sqrt{T})$, we have that the second term in last expression is $o_P(1)$. The result than follows from Theorem 2.3, Lemma 2.4 and the continuous mapping theorem.

A.3
Proofs of Chapter 3

Proof of Theorem 3.1

Proof. By assumption 3.3 g is differentiable so by the mean value theorem

$$Q_t(\tau) = g(x,\widehat{\theta}(\tau)) - g(x,\theta_0(\tau)) = \nabla g(x,\tilde{\theta})\left(\widehat{\theta}(\tau) - \theta_0(\tau)\right) \quad \text{where } \tilde{\theta} \in \|\theta - \theta_0\|$$

Let $T_2 \equiv T - T_0 + 1$, then for a given $\tau \in (0,1)$

$$\widehat{\tau}_T = \tfrac{1}{T_2}\sum_{t=T0}^{T} 1\{\widehat{\Delta}_t(\tau) \leq 0\} = \tfrac{1}{T_2}\sum_{t=T0}^{T} 1\{v_t(\tau) - Q_t(\tau) \leq 0\}$$

, where the last term can be decompose as

$$\widehat{\tau}_T - \tau = \tfrac{1}{T_2}\sum_{t=T0}^{T}\left(1\{v_t(\tau) \leq 0\} - \tau\right) - \tfrac{1}{T_2}\sum_{t=T0}^{T} J_t(\tau)Q_t(\tau) + R(\tau,\widehat{\theta}) \tag{A-8}$$

, where $J_t(\tau) \equiv f(g(x_t,\theta_0))$ and $f(\xi)$ is the density function of distribution function $F(\xi) = \mathbb{P}(v_t \leq \xi)$

Under the null, the first term is $o_p(1)$ by the LGN, the last term multiplied by $\sqrt{T}$ was shown to be $o_p(1)$ by Koul (1969) and appears also in Chen and Lockhart (2001). The term in between is also $o_p(1)$ as long as $\widehat{\theta}$ is consistent for θ_0, which demonstrate the consistency of $\widehat{\tau}$.

For the asymptotic normality multiply (A-8) by $\sqrt{T}$ and , then we are left with

$$\sqrt{\tfrac{T}{T_2}}\left(\tfrac{1}{\sqrt{T_2}}\sum_{t=T_0}^{T} 1\{v_t(\tau) \leq 0\} - \tau\right) - \left(\tfrac{1}{T_2}\sum_{t=T0}^{T} J_t(\tau)\nabla g(x,\tilde{\theta})\right)\sqrt{T}\left(\widehat{\theta}(\tau) - \theta_0(\tau)\right) + \sqrt{T}R(\tau,\widehat{\theta})$$

Note that the term in between is $o_p(1)$ for all non constant regressores of $g(\cdot)$. Let θ_c be constant regressor parameters and $T_1 \equiv T_0 - 1$, then the term in between can be written using Bahadur representation (1966)

$$\sqrt{\tfrac{T}{T_1}}\left(\tfrac{1}{T_2}\sum_{t=T_0}^{T} J_t(\tau)\right)\sqrt{T_1}\left(\widehat{\theta}_c(\tau) - \theta_{c,0}(\tau)\right) = \sqrt{\tfrac{T}{T_1}}\left(\tfrac{1}{T_2}\sum_{t=T_0}^{T} J_t(\tau)\right)D(\tau)^{-1} \tfrac{1}{T_1}\sum_{t=1}^{T_1} \tau - 1\{v_t \leq 0\} + o_p(1)$$

, where $D(\tau) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} J_t(\tau)$.

Hence we are left with

$$\sqrt{T}(\widehat{\tau}_T - \tau) = \sqrt{\tfrac{T}{T_2}}\left(\tfrac{1}{\sqrt{T_2}}\sum_{t=T_0}^{T} 1\{v_t(\tau) \leq 0\} - \tau\right) - \sqrt{\tfrac{T}{T_1}}\left(\tfrac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1} 1\{v_t(\tau) \leq 0\} - \tau\right) + o_p(1)$$

let $w_t(\tau) \equiv 1\{v_t(\tau) \leq 0\} - \tau$ and $\sigma^2(\tau) = \lim_{T\to\infty} \mathbb{E}(\sum_{t=1}^{T} w_t(\tau))^2 < \infty$ by assumption, then by the CLT we have

$$\sqrt{T}(\widehat{\tau}_T - \tau) \Rightarrow \sqrt{\tfrac{1}{1-\lambda_0}}\mathcal{N}\left(0, \sigma^2(\tau)\right) + \sqrt{\tfrac{1}{\lambda_0}}\mathcal{N}\left(0, \sigma^2(\tau)\right) \equiv \mathcal{N}\left(0, \frac{\sigma^2(\tau)}{\lambda_0(1-\lambda_0)}\right)$$

■

Proof of Corollary 3.2

Proof. First let $w_{it} = 1\{\Delta_t(\tau_i) \leq 0\} - \tau_i$, and $\Gamma_j = \mathbb{E}(\boldsymbol{w}_t \boldsymbol{w}_{t+j}')$ for $j \in \mathbb{Z}$ where $\boldsymbol{w}_t = (w_{1t}, \ldots, w_{kt})'$, hence

$$\begin{aligned}(\Gamma_0)_{ij} &= \mathbb{E}(1\{\Delta_t(\tau_i) \leq 0\}1\{\Delta_t(\tau_j) \leq 0\}) - \tau_i\tau_j \\ &= \mathbb{P}(\Delta_t(\tau_i) \leq 0 \cap \Delta_t(\tau_j) \leq 0) - \tau_i\tau_j \\ &= \min(\tau_i, \tau_j) - \tau_i\tau_j\end{aligned}$$

We can now take stack k equations (**??**), one for each $\tau = \tau_1, \ldots, \tau_k$ and premultiply by any $\boldsymbol{a}_k \neq \boldsymbol{0} \in \mathbb{R}^k$:

$$\sqrt{T}\boldsymbol{a}_k'(\widehat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}) = \sqrt{\tfrac{T}{T_2}}\left(\tfrac{1}{\sqrt{T_2}}\sum_{t=T_0}^{T} \boldsymbol{a}_k'\boldsymbol{w}_t\right) - \sqrt{\tfrac{T}{T_1}}\left(\tfrac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1} \boldsymbol{a}_k'\boldsymbol{w}_t\right) + o_p(1)$$

But $\boldsymbol{a}_k'\boldsymbol{w}_t$ is an ergodic stationary process, hence by the CLT each of the terms in parenthesis converge in distribution to normal random variable with mean 0 and variance $\boldsymbol{a}_k'\boldsymbol{\Sigma}(\boldsymbol{\tau})\boldsymbol{a_k}$, where $\boldsymbol{\Sigma}(\boldsymbol{\tau}) \equiv \sum_{j\in\mathbb{Z}} \boldsymbol{\Gamma}_j$. Hence by the Cramer-Wold device the corollary follows. ■

B
Appendix: Figures

Figure B.1: Bias Factor defined on (1-13) for $l_i = \sigma_{\eta_i} = 1$ for all $i = 1, \ldots, n$.

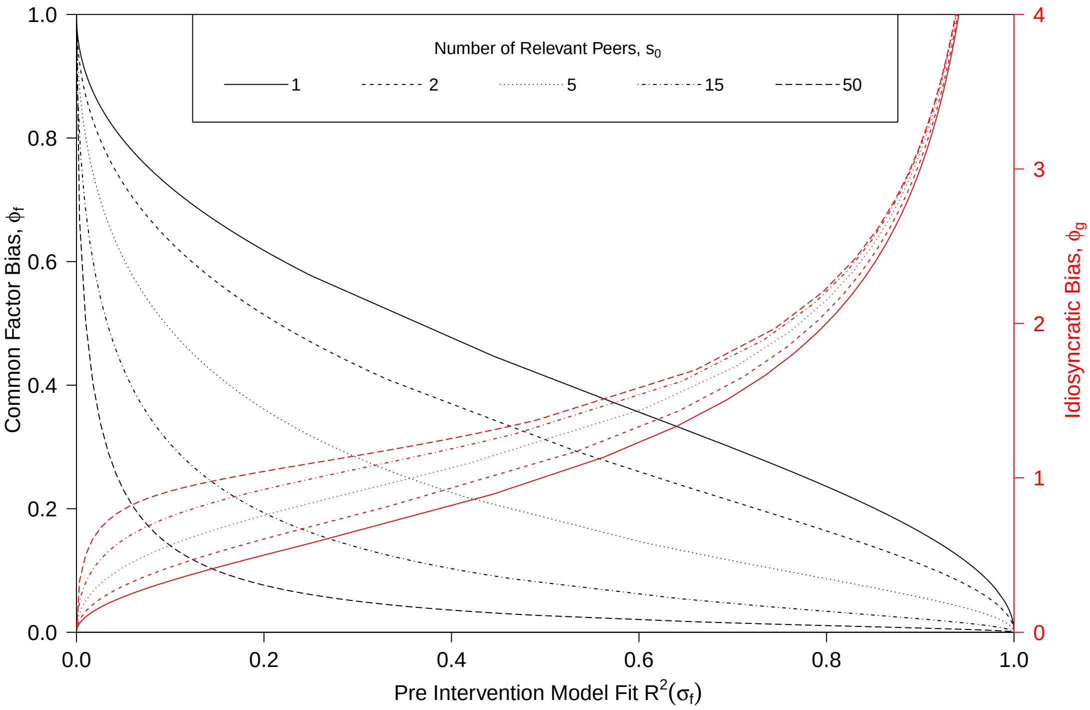


Figure B.2: Kernel Density - Estimator Comparison with no Trend and no Serial Correlation

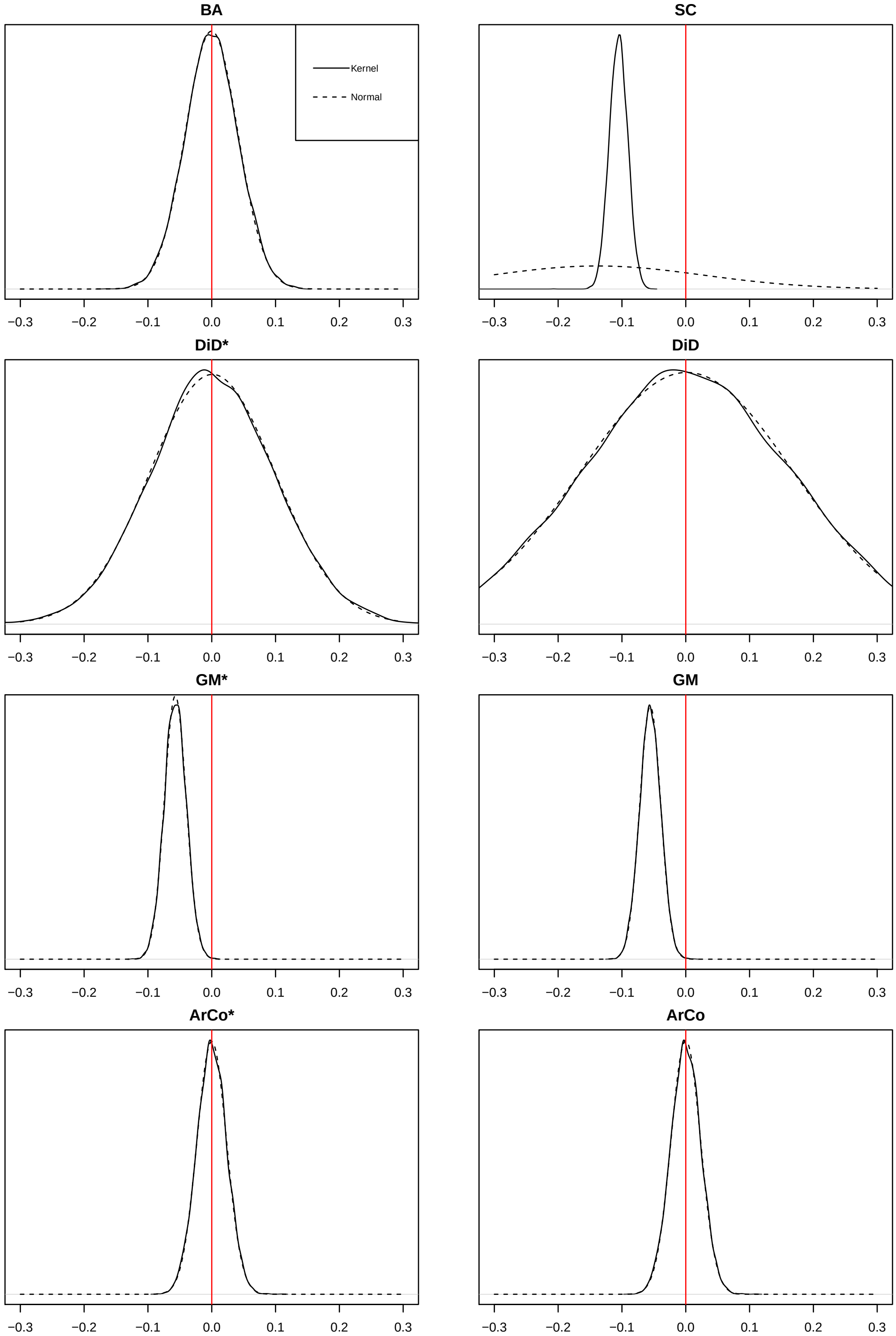


Figure B.3: Kernel Density - Estimator Comparison with no Trend

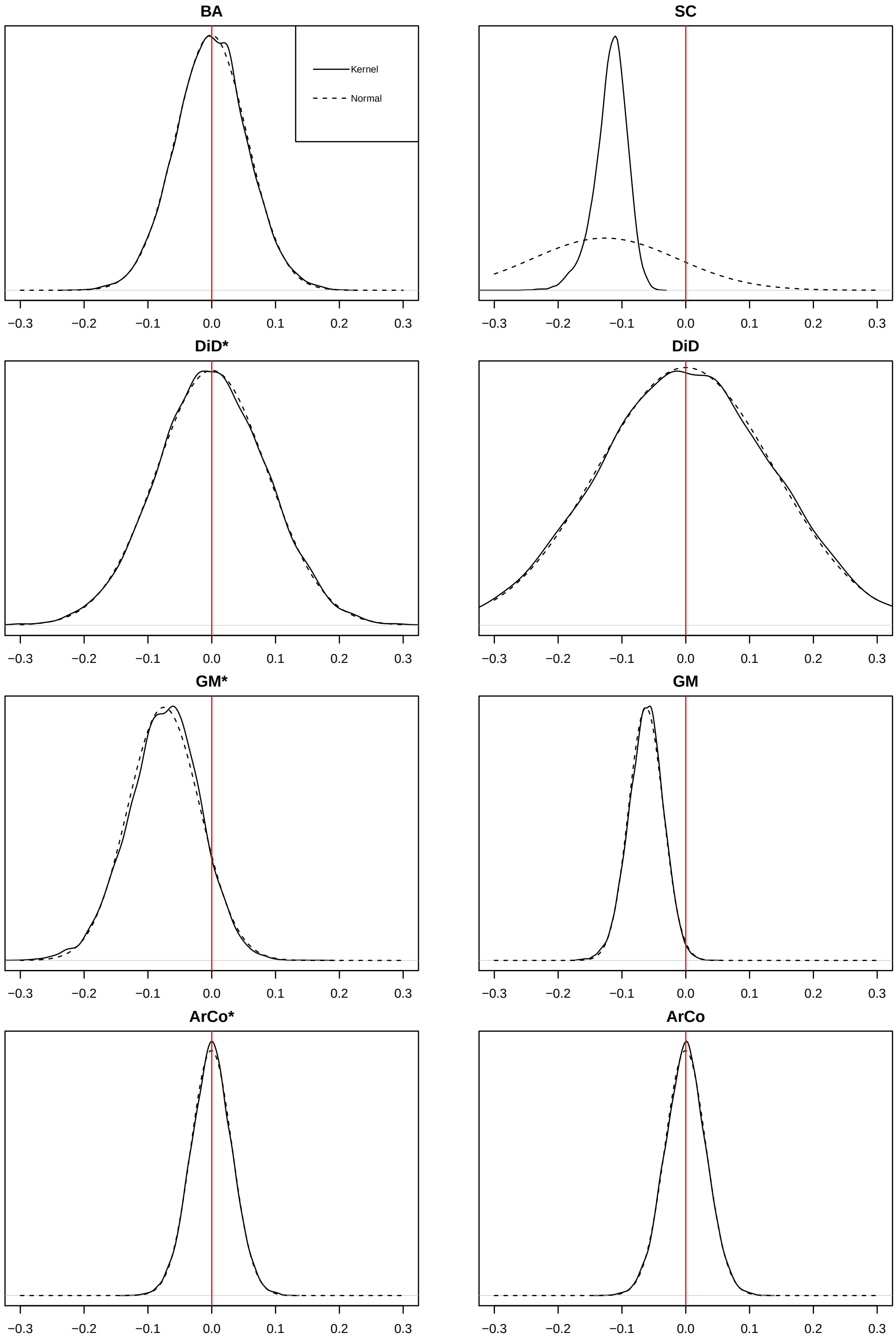


Figure B.4: Kernel Density - Estimator Comparison with Common Linear Trend

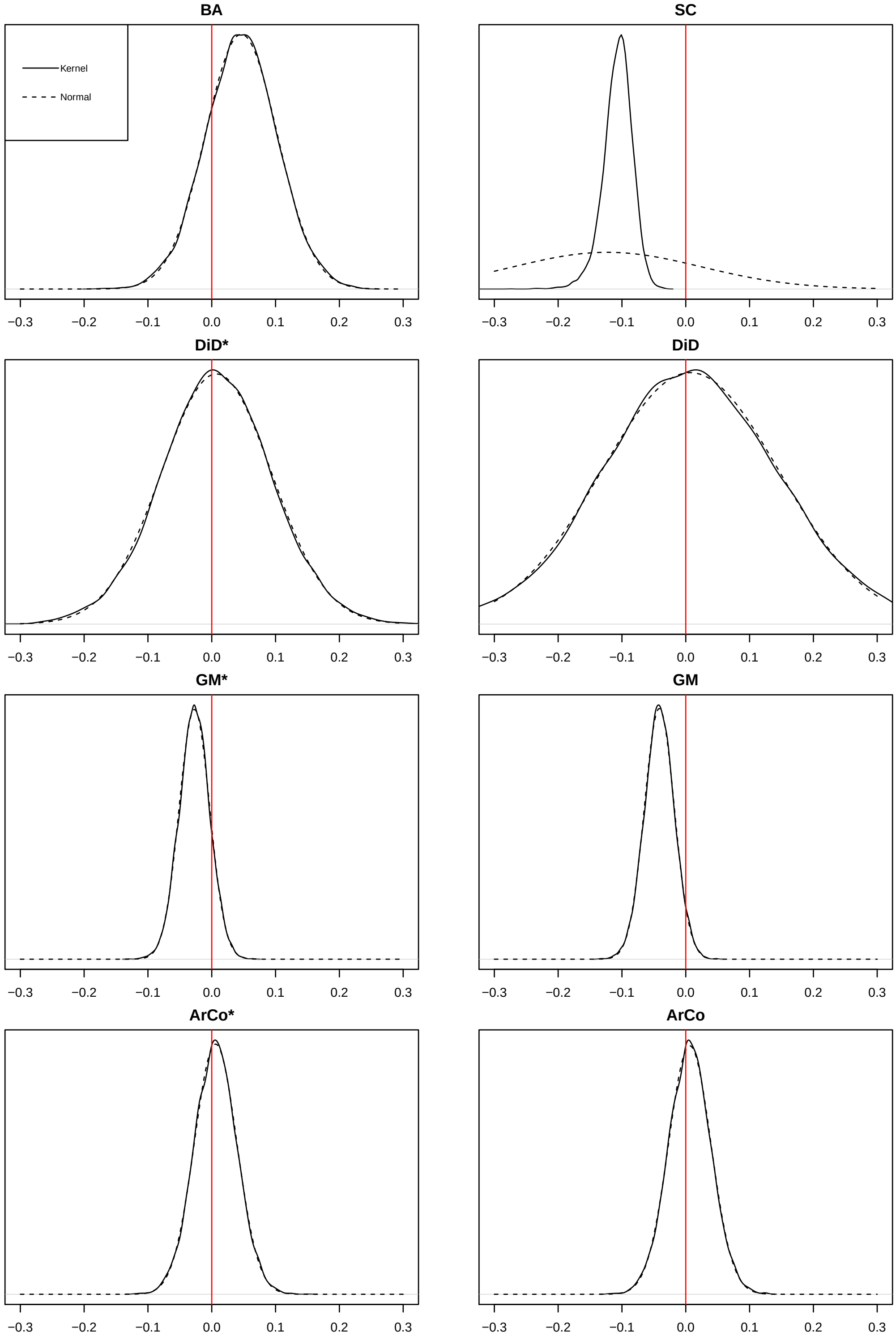


Figure B.5: Kernel Density - Estimator Comparison with Idiosyncratic Linear Trend

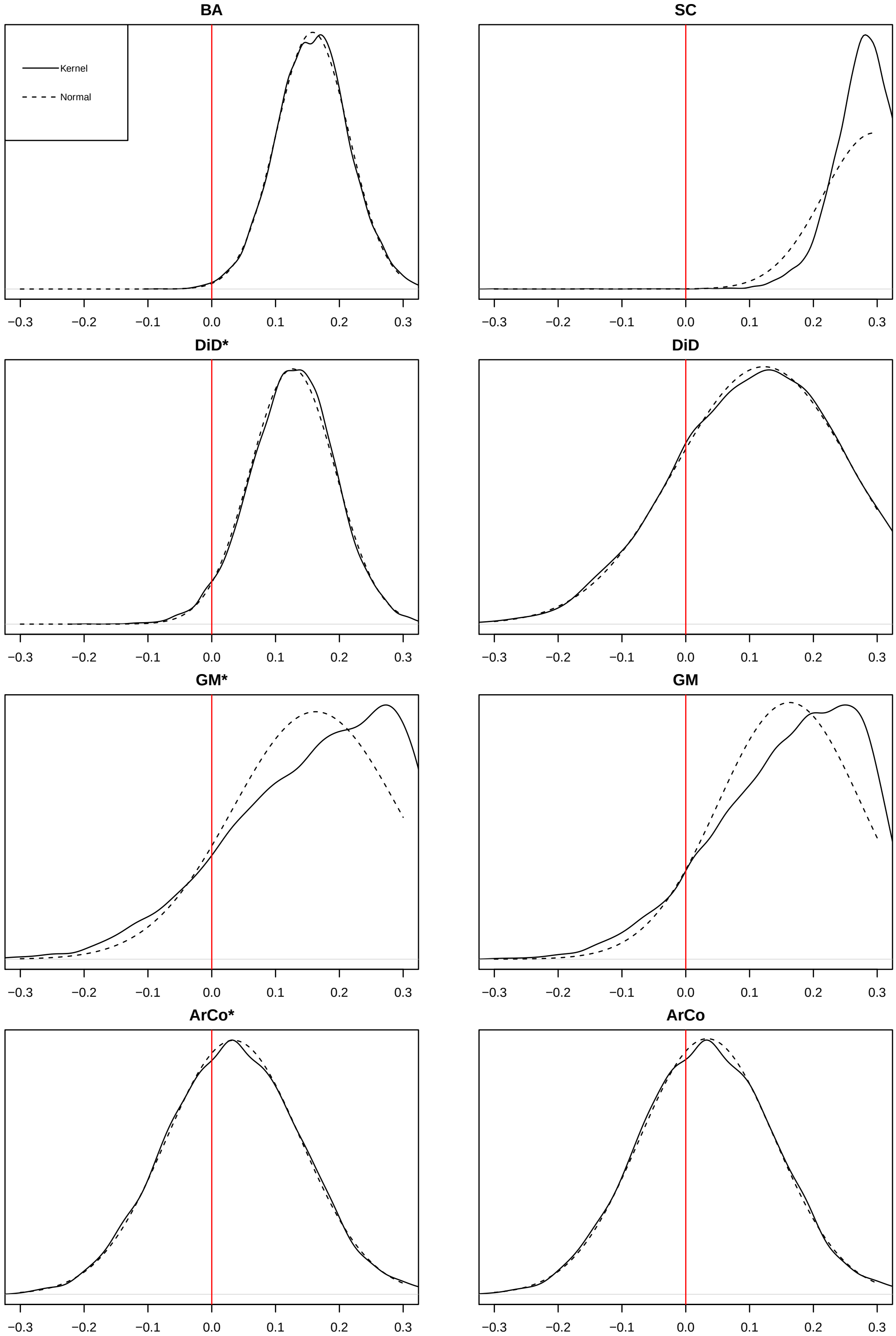


Figure B.6: Kernel Density - Estimator Comparison with Common Quadratic Trend

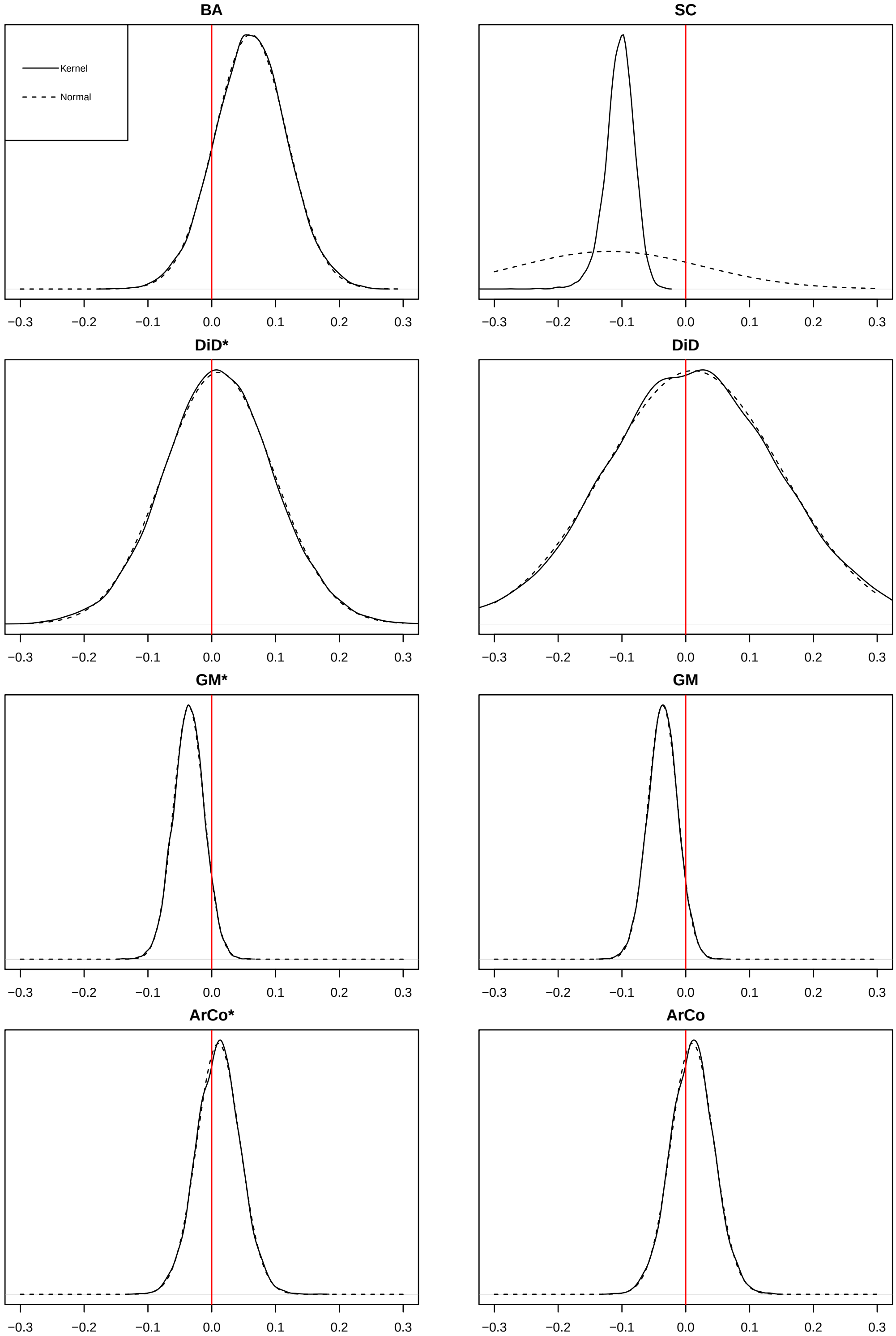


Figure B.7: Kernel Density - Estimator Comparison with Idiosyncratic Quadratic Trend

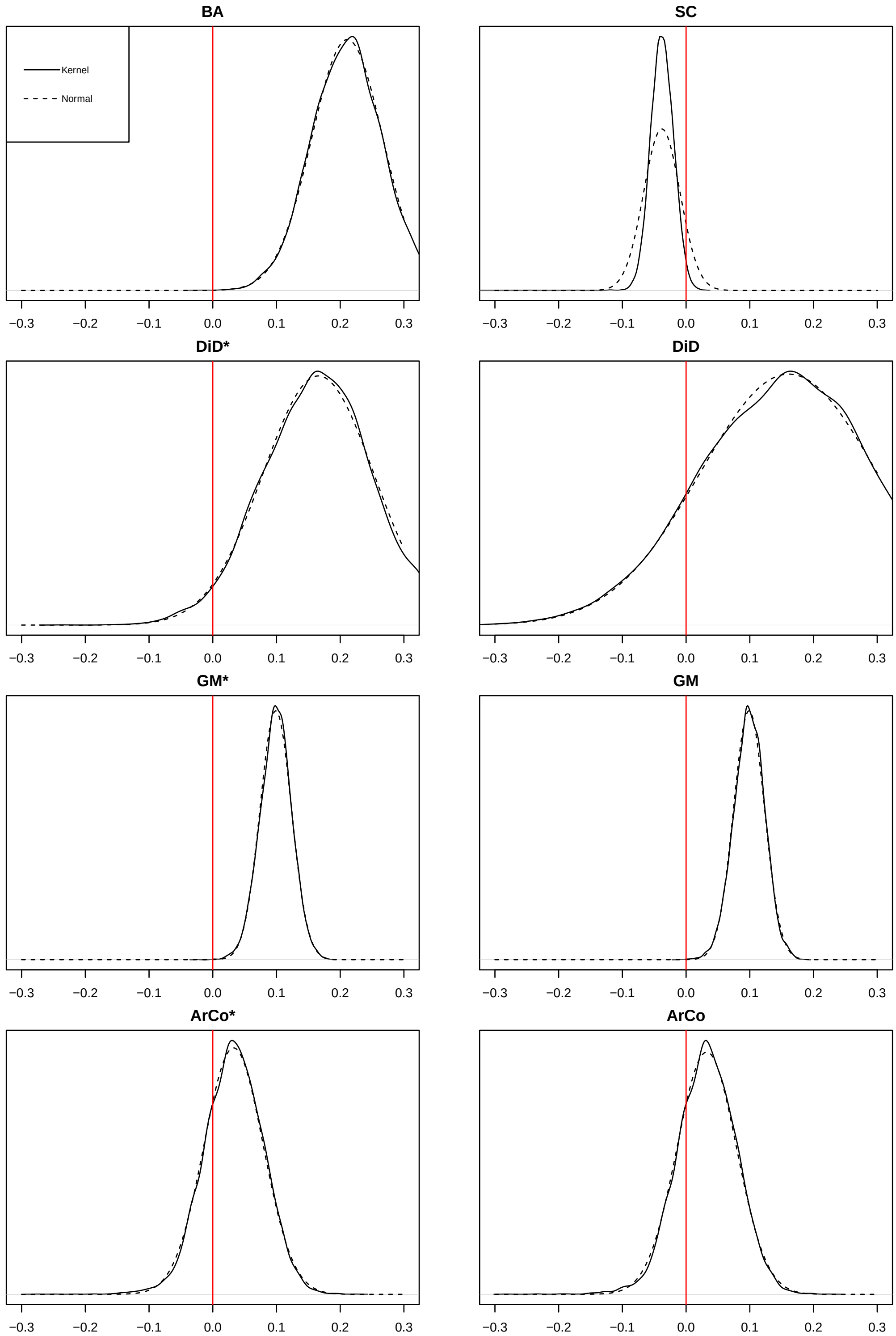
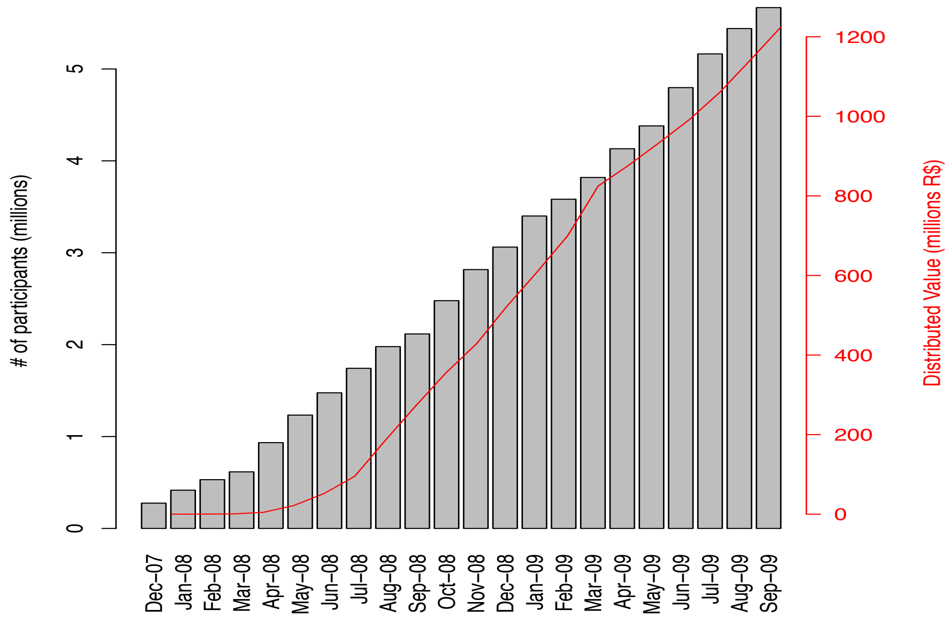
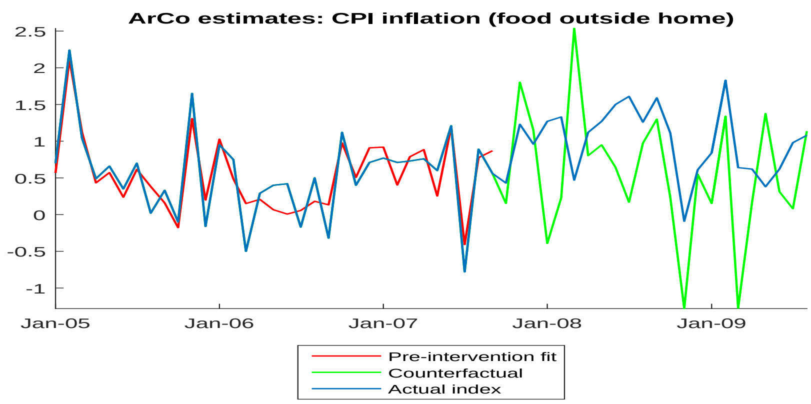
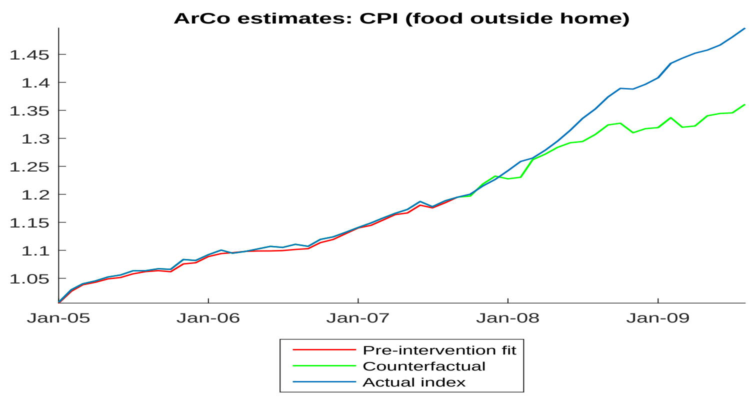


Figure B.8: NFP Participation (left) and Value distributed (right)



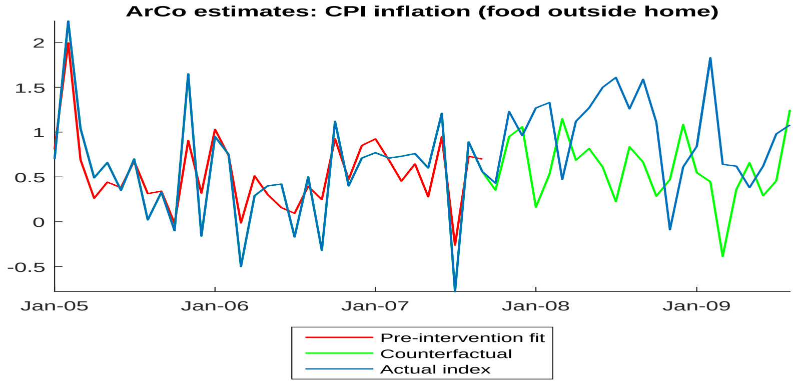


B.9(a):

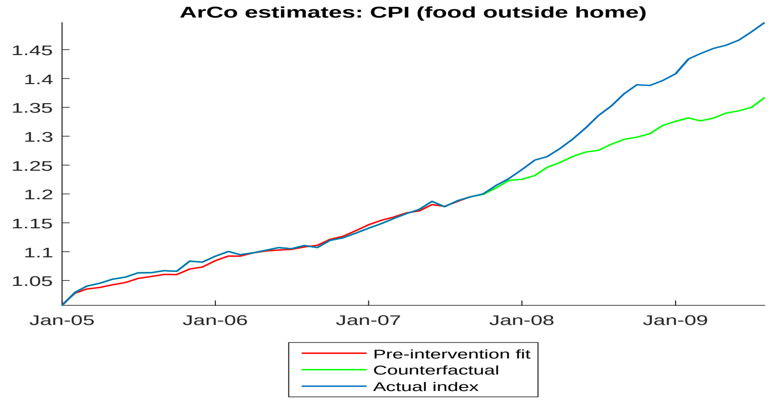


B.9(b):

Figure B.9: Actual and counterfactual data. The conditioning variables are **inflation** and **DGP growth**. Panel (a) monthly inflation. Panel (b) accumulated monthly inflation.



B.10(a):



B.10(b):

Figure B.10: 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.

C
Appendix: Tables

Table C.1: Rejection Rates under the Alternative (Test Power)

	$\boldsymbol{\alpha = 0.1}$	**0.075**	**0.05**	**0.025**	**0.01**
	Step Intervention[1] $\delta_t = c\,\sigma_1 1\{t \geq T_0\}$				
$c = 0.15$	0.2045	0.1695	0.1287	0.0805	0.0436
0.25	0.3783	0.3266	0.2686	0.1890	0.1108
0.35	0.5769	0.5235	0.4545	0.3465	0.2414
0.5	0.8314	0.7945	0.7440	0.6478	0.5227
0.75	0.9876	0.9831	0.9741	0.9520	0.9094
1	0.9998	0.9995	0.9992	0.9983	0.9943
	Linear Increasing $\delta_t = c\,\sigma_1 \frac{t-T_0+1}{T-T_0+1} 1\{t \geq T_0\}$				
$c = 1$	0.8318	0.7938	0.7379	0.6397	0.5121
1.25	0.9877	0.9813	0.9717	0.9459	0.8948
1.5	0.9997	0.9997	0.9990	0.9969	0.9922
	Linear Decreasing $\delta_t = c\,\sigma_1 \frac{T-t+1}{T-T_0+1} 1\{t \geq T_0\}$				
$c = 1$	0.8298	0.7956	0.7434	0.6492	0.5107
1.25	0.9868	0.9818	0.9720	0.9490	0.8985
1.5	0.9995	0.9994	0.9989	0.9968	0.9933

All simulations above as per DGP in (2-6) with the parameters in the baseline scenario as described in the footnote of Table C.2.

[1] All interventions intensity are measured as a factor $c > 0$ of the standard deviation of unit of interest, σ_1.

Table C.2: Rejection Rates under the Null (Test Size)

	Bias	**Var**[a]	$\widehat{s}_0$	$\boldsymbol{\alpha = 0.1}$	**0.05**	**0.01**
	Innovation Distribution [b]					
Normal	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
$\chi^2(1)$	-0.0014	1.1004	5.9287	0.1227	0.0652	0.0154
t-stud(3)	0.0035	1.1026	5.6437	0.1077	0.0543	0.0103
Mixed-Normal	0.0069	1.1267	5.5457	0.1134	0.0607	0.0136
	Sample Size					
$T = 100$	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
75	-0.0030	1.1449	6.3992	0.1075	0.0546	0.0124
50	0.0021	1.1747	6.1219	0.1092	0.0626	0.0155
25	-0.0050	0.8324	3.2463	0.1330	0.0763	0.0226
	Number of Total Covariates					
$d = 100$	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
200	-0.0016	1.1655	5.7314	0.1102	0.0565	0.0135
500	-0.0043	1.2112	5.6625	0.1119	0.0556	0.0114
1000	0.0012	1.2477	5.5275	0.1054	0.0566	0.0115
	Number of Relevant (non-zero) Covariates					
$s_0 = 0$	0.0038	1.0981	0.6105	0.1059	0.0550	0.0136
5	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
10	0.0003	1.0373	9.5813	0.1103	0.0581	0.0120
100	0.0003	-	20.1624	0.1114	0.0574	0.0145
	Determinist Trend $(t/T)^{\varphi}$					
$\varphi = 0$	0.0006	1.1304	5.4076	0.1057	0.0555	0.0128
0.5	0.0142	1.1245	5.6285	0.1101	0.0598	0.0199
1	0.0183	1.1313	5.5030	0.1188	0.0613	0.0168
2	0.0221	1.1398	5.4259	0.1273	0.0675	0.0261
	Serial Correlation[c]					
$\rho = 0.2$	-0.0001	1.4109	5.5246	0.1160	0.0640	0.0158
0.4	0.0002	1.6909	5.9276	0.1223	0.0678	0.0184
0.6	0.0031	1.8895	6.9012	0.1440	0.0871	0.0283
0.8	0.0033	1.9977	7.9464	0.1546	0.0927	0.0329

Baseline DGP: (2-6) with $T = 100$, iid normally distributed innovations; $T_0 = 50$; $n = 100$ units; $d = n = 100$ covariates (including the constant); $s_0 = 5$, $q = 1$; $10,000$ Monte-Carlo simulations per case. The penalization parameter is chosen via Bayesian Information Criteria (BIC). We set the maximum number of included variables to be $T^{0.8}$ in the glmnet package in R.

[a] Relative to the variance of the oracle/OLS estimator in the fist stage knowing the relevant regressors.

[b] All distributions are standardized (zero mean and unit variance); Mixed normal equal to 2 Normal distributions with probability $(0.3, 0.7)$, mean $(-10, 10)$ and variance $(2, 1)$.

[c] All units are simulated as AR(1) processes. The variance estimator is computed as Andrews e Monahan (1992) with an AR(1) pre-whitening followed by a standard HAC estimator with Quadratic Spectral Kernel on the residuals. Optimal bandwidth selection for AR(1) as per Andrews (1991).

Table C.3: Estimators Comparison

	BA	**SC**	**DiD***	**DiD**	**GM***	**GM**	**ArCo***	**ArCo**
	No Time Trend ($\varphi = 0$) and No Serial Correlation ($\rho = 0$)							
Bias[1]	-0.001	-0.678	0.005	0.008	-0.280	-0.273	0.000	0.000
Var	3.151	50.555	17.870	51.444	0.544	0.510	1.001	1.000
MSE	3.152	86.075	17.871	51.449	6.601	6.255	1.001	1.000
	No Time Trend ($\varphi = 0$)							
Bias	-0.003	-0.596	0.000	0.000	-0.353	-0.294	-0.002	-0.002
Var	2.997	12.293	7.215	18.506	3.057	0.705	0.998	1.000
MSE	2.996	27.634	7.214	18.502	8.438	4.427	0.998	1.000
	Common Linear Time Trend ($\varphi = 1$)							
Bias	0.218	-0.579	0.034	0.033	-0.128	-0.195	0.028	0.029
Var	2.900	19.590	6.741	17.720	0.522	0.499	1.007	1.000
MSE	4.677	32.165	6.558	17.159	1.151	1.985	1.004	1.000
	Idiosyncratic Linear Time Trend ($\varphi = 1$)							
Bias	0.744	1.391	0.597	0.577	0.766	0.766	0.161	0.158
Var	0.288	0.564	0.392	1.720	1.499	1.113	0.996	1.000
MSE	2.270	7.544	1.651	2.771	3.493	3.142	0.999	1.000
	Common Quadratic Time Trend ($\varphi = 2$)							
Bias	0.288	-0.562	0.051	0.053	-0.170	-0.170	0.049	0.048
Var	2.809	18.486	6.571	17.199	0.512	0.488	1.007	1.000
MSE	5.583	28.407	6.105	15.837	1.520	1.498	1.010	1.000
	Idiosyncratic Quadratic Time Trend ($\varphi = 2$)							
Bias	0.994	-0.179	0.780	0.758	0.465	0.465	0.154	0.153
Var	1.443	0.377	3.499	8.878	0.282	0.274	0.992	1.000
MSE	14.786	0.701	10.868	14.002	3.216	3.210	0.998	1.000

$S = 10,000$ simulations from DGP (1-14); $T = 100$ observations; Intervention at $T_0 = 50$ only on the first variable of the first unit of intensity one standard deviation; r_f chosen such that $R^2 = 0.5$; $n = 5$ units; $q = 3$ variables per unit; innovations are iid normally distributed; $\rho = 0.5$ and $\mathsf{diag}(\boldsymbol{A})$ are independent draws from uniform $[-1, 1]$; All the loads (for the constant, the time trend and the stochastic factor) are independent draws from uniform distribution $[-5, 5]$, except for the common trend cases where the time trend loads are equal to unit for all variables of all units and for the cases with no time trend where they are all set to zero.

* Estimators using the $q - 1$ covariates of unit 1. Hence, unfeasible if we expect the intervention to affect all the variables in unit 1

[1] Bias measured as a ratio to the intervention intensity, defined by one standard deviation of the first variable of the first unit; Variance and MSE measured as a ratio to the ArCo Variance and MSE, respectively.

Table C.4: Estimated Effects on food away from home (FAH) Inflation.

Panel (a): ArCo Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	0.2500 (0.1726)	0.4441 (0.1487)	0.4870 (0.1414)	0.7973 (0.2431)	0.4478 (0.2017)	0.3796 (0.1613)	0.4046 (0.1539)	0.4422 (0.1467)
Inflation	Yes	No	No	No	Yes	Yes	Yes	No
GDP	No	Yes	No	No	Yes	Yes	Yes	No
Retail Sales	No	No	Yes	No	No	Yes	Yes	No
Credit	No	No	No	Yes	No	No	Yes	No
R-squared	0.6849	0.1240	0.3856	0.3106	0.7993	0.8948	0.8072	0
Number of regressors	10	9	10	10	19	29	39	0
Number of relevant regressors	10	3	6	9	16	15	13	0
Number of observations ($t < T_0$)	33	33	33	33	33	33	33	33
Number of observations ($t \geq T_0$)	23	23	23	23	23	23	23	23

Panel (b): Alternative Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
BA	0.4472 (0.1464)	0.4478 (0.1466)	0.4390 (0.1471)	0.4538 (0.1464)	0.4501 (0.1467)	0.4422 (0.1467)
DiD	0.2195 (0.1467)	0.2111 (0.1460)	0.2171 (0.1467)	0.2112 (0.1460)	0.2088 (0.1461)	0.2194 (0.1467)
GM	0.3699 (0.1237)	0.3785 (0.1246)	0.3759 (0.1234)	0.3759 (0.1234)	0.3607 (0.1226)	−−
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 e Magnac (2016) (GM) and the difference-in-difference (DiD) estimators.

Table C.5: 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 C.6: 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 e Magnac (2016) (GM) and the difference-in-difference (DiD) estimators.

Table C.7: Rejection Rates under the null (size)

	Normal Distribution				
(τ_1, τ_2)	$\alpha = 0.1$	0.075	0.05	0.025	0.01
(0,0.5)	0.1067	0.0687	0.0400	0.0236	0.0066
(0.33,0.66)	0.1093	0.0674	0.0394	0.0189	0.0037
(0.25,0.75)	0.1302	0.0867	0.0548	0.0339	0.0092
(0.2,0.8)	0.1414	0.0982	0.0641	0.0437	0.0154
(0.15,0.85)	0.1858	0.1333	0.0954	0.0621	0.0272
(0.1,0.9)	0.2358	0.1725	0.1278	0.0885	0.0637
$\lVert \cdot \rVert_\infty$	0.0879	0.0631	0.0432	0.0201	0.0077
$\lVert \cdot \rVert_2$	0.1194	0.0899	0.0598	0.0282	0.0107
	t-Student distribution with 3 dof				
(τ_1, τ_2)	$\alpha = 0.1$	0.075	0.05	0.025	0.01
(0,0.5)	0.1077	0.0670	0.0419	0.0249	0.0069
(0.33,0.66)	0.1087	0.0648	0.0366	0.0209	0.0040
(0.25,0.75)	0.1276	0.0864	0.0544	0.0326	0.0109
(0.2,0.8)	0.1449	0.1017	0.0702	0.0449	0.0168
(0.15,0.85)	0.1831	0.1343	0.0942	0.0629	0.0253
(0.1,0.9)	0.2515	0.1842	0.1348	0.0934	0.0627
$\lVert \cdot \rVert_\infty$	0.0936	0.0692	0.0469	0.0237	0.0077
$\lVert \cdot \rVert_2$	0.1215	0.0918	0.0614	0.0292	0.0117
	Chi-square distribution with 1 dof				
(τ_1, τ_2)	$\alpha = 0.1$	0.075	0.05	0.025	0.001
(0,0.5)	0.1049	0.0682	0.0413	0.0224	0.0066
(0.33,0.66)	0.1096	0.0673	0.0396	0.0205	0.0048
(0.25,0.75)	0.1279	0.0822	0.0519	0.0305	0.0108
(0.2,0.8)	0.1344	0.0931	0.0616	0.0404	0.0163
(0.15,0.85)	0.1807	0.1278	0.0932	0.0598	0.0220
(0.1,0.9)	0.2419	0.1777	0.1301	0.0887	0.0603
$\lVert \cdot \rVert_\infty$	0.0916	0.0673	0.0438	0.0188	0.0071
$\lVert \cdot \rVert_2$	0.1231	0.0963	0.0626	0.0282	0.0115
	Uniform distribution				
(τ_1, τ_2)	$\alpha = 0.1$	0.075	0.05	0.025	0.001
(0,0.5)	0.1045	0.0664	0.0403	0.0216	0.0058
(0.33,0.66)	0.1141	0.0691	0.0391	0.0198	0.0045
(0.25,0.75)	0.1342	0.0896	0.0560	0.0342	0.0110
(0.2,0.8)	0.1443	0.0976	0.0664	0.0419	0.0172
(0.15,0.85)	0.1775	0.1273	0.0882	0.0616	0.0249
(0.1,0.9)	0.2376	0.1745	0.1280	0.0900	0.0615

NB: $T = 100$ observations, $T_0 = 50$ ($\lambda_0 = 0.5$). $n = 4$ units. 10000 Monte-Carlo simulations per case. All disturbances are normalised to mean zero and unit variance for each of the distributions considered

Table C.8: Critical Vales for Unknown Intervention Time Inference: $\mathbb{P}(\|\boldsymbol{S}\|_p > c) = 1 - \alpha$

		Confidence Level					
	$\Lambda = [\underline{\lambda}, \bar{\lambda}]$	$\alpha = 0.2$	0.15	0.1	0.05	0.0025	0.001
$p = 1$	$[0.5, 0.95]$	2.5679	2.7824	3.0732	3.5457	3.9844	4.5346
	$[0.1, 0.9]$	2.4332	2.6569	2.9550	3.4530	3.9218	4.4805
	$[0.15, 0.85]$	2.3786	2.6164	2.9375	3.4482	3.9138	4.4728
	$[0.2, 0.8]$	2.3366	2.5833	2.9167	3.4399	3.9115	4.4655
$p = 2$	$[0.5, 0.95]$	3.0633	3.2814	3.5706	4.0228	4.4378	4.9674
	$[0.1, 0.9]$	2.8230	3.0441	3.3340	3.8138	4.2602	4.7792
	$[0.15, 0.85]$	2.7052	2.9400	3.2448	3.7391	4.1859	4.7235
	$[0.2, 0.8]$	2.6169	2.8579	3.1795	3.6787	4.1466	4.7159
$p = \infty$	$[0.5, 0.95]$	8.6192	9.1867	9.9400	11.1562	12.2190	13.5604
	$[0.1, 0.9]$	6.4807	6.8974	7.4353	8.2781	9.0400	10.0020
	$[0.15, 0.85]$	5.6000	5.9506	6.4041	7.1014	7.7328	8.5187
	$[0.2, 0.8]$	5.0630	5.3815	5.7957	6.4303	7.0047	7.7473

NB: All critical values were obtained as the quantile of the empirical distribution using 100,000 draws from a multivariate normal distribution with covariance $\boldsymbol{\Sigma_\Lambda}$ via a grid of 500 points between $\underline{\lambda}$ and $\bar{\lambda}$ inclusive.

Table C.9: Analized Cases of Change in Corporate Governance Regime

Treated	Segment	Migration Date /Level	Peers	T	$\lambda_0 = \frac{T_0}{T}$
BBAS3	Banking	28-Jun-2006 (NM)	ITUB BBDC4 SANB4	280	0.46
ETER3	Construction Material	2-Mar-2005 (N2)	CCHI3 HAGA4	150	0.67
SBSP3	Sewage and Water Dist.	24-Apr-2002 (NM)	SAPR4 HAGA4 CABB3	135	0.54
RSID3	Building and Incorporation	27-Jan-2006 (NM)	GEN4 CYRE3	127	0.43

NB: T is the sample size, whenever possible we try to trim the sample size to have the intervention in the middle (minimum variance as described above); T_0 is the time of the intervention.

Table C.10: Estimation Resutls ($\widehat{r} = \widehat{\tau}_2 - \widehat{\tau}_1$)

	Coverage Probability (τ_1, τ_2)				
	(0,0.5)	(0.15, 0.85)	(0.2, 0.8)	(0.25, 0.75)	(0.33, 0.66)
$r_0 = \tau_2 - \tau_1$	0.5	0.70	0.6	0.5	0.33
BBAS3	0.4636 (0.5426)	0.8477 (0.0071)	0.7152 (0.0493)	0.6556 (0.0093)	0.3907 (0.2804)

NB: p-value in parentheses. Standard error estimation using under iid assumption.