Distributional Counterfactual Analysis with (common) Deterministic trend units *

Ricardo Masini
FGV-EESP

January 2018

(Preliminary version; please do not circulate without permission)

Abstract

The goal of this paper is to extend the counterfactual methodologies in particular Carvalho et al. (2016), by considering the estimation of quantile counterfactuals in the presence of trend units. We derive an asymptotically normal test statistics for the quantile intervention effect and for the distribution effect as a whole. As a by-product we show the consistency of the quantile regression without compactness. Our procedure is illustrated in a detailed simulation experiment as well as in an empirical application in Corporate Finance.

JEL Codes: C22, C23, C32, C33.

Keywords: counterfactual analysis, quantile panel data, quantile counterfactual, ArCo.

*Email: ricardo.masini@fgv.br

1 Introduction

In this paper we propose a new method to carry out counterfactual analysis to evaluate the impact of interventions on the distribution of the variable of interest. Our approach is specially useful in situations where there is a single "treated" unit and no available "controls". Also, we depart from the common setup of microeconometrics studies where together with the variable of interest you also observe some determined (not affected by the treatment) covariates for each unit. Instead, we consider a situation where we follow together with the unit of interest a set of untreated units (peers) during the same time span (time-series panel data).

Motivated by an underlying common factor structure, we aim for situation where the only contemporaneous source of information to construct the counterfactual for the unit of interest is the observed untreated units. In other words, the goal of the proposed method is to put forward a framework to construct an artificial counterfactual based on observed data from a pool of "untreated" peers. This task was carried out from the conditional mean respectively in Hsiao et al. (2012) and Carvalho et al. (2016). The present work can be seem as a generalization to account for distributional affects.

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 and Imbens (1994), Heckman and 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. See, for example, Abadie et al. (2002), Chernozhukov and Hansen (2005), Chernozhukov and Hansen (2006), Chernozhukov and Hansen (2008), Chernozhukov et al. (2014), and Firpo (2007).

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 and 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 and 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 paper 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 paper is organized as follows. Section 2 contains the theoretical framework adopted. Subsection 2.1 presents the overall definitions followed by Subsection 2.2 which discuss the trade-offs between Marginal vs. Conditional distribution modeling. The main results are presented in Subsection 2.3 as part of the asymptotic theory of the proposed estimator. Inference procedures including a distributional test are considered in Section 3. The effects of misspecification in the conditional quantile model (CQM) are discussed in Section 4 where we derive the exact CQM under Gaussianity in Subsection 4.1 and investigate the effects on misspecification in a Monte Carlo study presented in Subsection 4.2. The empirical illustration is described in Section 5. Finally, Section 6 concludes the paper. All proofs are relegated to a technical appendix.

2 Theoretical Framework

A word on notation: we denote random variables by upper case A and its realization by lower case $A = a$. We use bold face for vectors or matrices. All random quantities are defined in a common probability space $(\Omega, \mathcal{F}, P)$. We use $F_A(a) = \mathbb{P}(A \leq a)$ to denote the cdf of A and f_A its density. Also $Q_A(\tau) = \inf\{a : F_A(a) \geq \tau\}$ is the quantile function defined for $\tau[0, 1]$. All the asymptotics are taken as $T \to \infty$.

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 realization a random variable Z_{it}.

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 ($i = 1$). Let D_{it} be a binary variable flagging the periods when the intervention was in place, then we can express the observable variables as interest as

$$Z_{it} = D_{it} Z_{it}^{(1)} + (1 - D_{it}) Z_{it}^{(0)}; \qquad D_{it} = \begin{cases} 1 & \text{if } t \geq T_0 \text{ and } i = 1 \\ 0 & \text{otherwise ,} \end{cases}$$

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

Furthermore, we denote the unit of interest by $Y_t \equiv Z_{1t}$ and the remaining $n - 1$ untreated unit (potential controls) plus a constant regressor by $\boldsymbol{X}_t \equiv (1, Z_{2t}, \ldots, Z_{nt})'$. We treat the peers as untreated, i.e., the intervention had no effect on them formally we require that

Assumption 1. *D_t is independent of $\boldsymbol{X}_t$ for all t*

It 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 as a random variable given by

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

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 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 2. *For each unit $i = 1, \dots, n$ and $t \geq 1$:*

$$Z_{it}^{(0)} = \alpha_i + \delta_i \zeta(t) + \boldsymbol{\lambda}_i' \boldsymbol{F}_t + \eta_{it}$$

where $\alpha_i, \delta_i \in \mathbb{R}$, $\boldsymbol{\lambda}_i \in \mathbb{R}^r$ $d(t)$ are loading of the $(r \times 1)$ common factor $\boldsymbol{F}_t$, $\zeta(t)$ is a deterministic time trend and η is the is idiosyncratic error term considered independent of the factors.

$\boldsymbol{f}_t(f \times 1)$ 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, $\mathbb{E}(\eta_{it} \boldsymbol{f}_j) = \mathbf{0}$, $\forall i, t, j$. 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 2 is quite flexible. It translate into each unit being modeled as a deterministic 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.

2.2 Marginal vs. Conditional Modelling Approach

First let's consider a supposedly more direct approach to test for potential difference in the distribution of Y_t before and after the intervention using its respective empirical distribution function (EDF) to perform a distributional test for instance. Expect for very particular cases, in the presence of a common trend marginal distributions before and after the intervention, say F_0 and F_1 are to constant. Not even during the pre-intervention time and very likely not to be the case for the post intervention times.

As consequence of the deterministic (but unknown) time trend, If we use as a test statistic some metric defined over the difference in EDF, $\widehat{F}_0 - \widehat{F}_1$, 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 (moment or 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 information contained in the peers to conduct a conditional analysis since a combination of non stationary series results in a stationarity result which is in fact the basis for our inference procedure.

Even when the data can be judge stationary. A conditional approach yields a more efficient estimation by exploring the common shock structure to give us tighter estimates of the counterfactual of interest. In most applications, however, we expect a common trend to be driven by a common factor structure (an underlying random walk process for instance) and the intervention affecting the idiosyncratic shock of the treated unit.

Furthermore, we do require a parallel trend assumption as for instance the Difference-in-Difference Approach nor an addictive deterministic treatment such as the Synthetic Control Method. In fact, we can test for any kind of distributional effect on the unit of interest.

In this paper we focus on conditional quantiles (for a conditional expectation approach refer to Carvalho et al. (2016)). We could also choose to model the conditional distributional but both approachs are equivalent to account for the distributional effects of the intervention. 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.

F or 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] $Q_\tau(\boldsymbol{x}) \equiv F^{-1}_{Y|X}(\tau|\boldsymbol{x}) \equiv \inf\{y : F_{Y|X}(y|\boldsymbol{x}) \geq \tau\}$.

Assumption 3. *For each* $\tau \in (0,1)$, $Q_\tau(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{\theta}_0(\tau)$ *for some* $\boldsymbol{\theta}_0(\tau) \in \Theta \equiv \mathbb{R}^n$

The assumption above postulate a correctly specified parametric model for $Q_\tau(\boldsymbol{x})$, which is true for instance when the idiosyncratic terms in Assumption 2 are jointly Gaussian for instance. Failure to this hypothesis (mispecification) are treated in a section below. We could also consider a more flexible setup where we allow both the the functional form and the true parameters to vary with τ, such that $Q_\tau(\boldsymbol{x}) = g_\tau(\boldsymbol{x}'\boldsymbol{\theta}_0(\tau))$ for some know function g_τ. However I decide to show the methodology using a more parsimonious linear model by setting $g_\tau(\boldsymbol{x}, \boldsymbol{\theta}_0(\tau)) = \boldsymbol{x}'\boldsymbol{\theta}_0(\tau)$.

[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 $Q_\tau(\boldsymbol{x}) \equiv F^{-1}_{Y|X}(\tau)$

It can be shown that the parameter $\boldsymbol{\theta}_0 \equiv \boldsymbol{\theta}_0(\tau)$ is a solution (not necessarily unique) to the following population optimization problem $\min_{\theta\in\Theta} \mathbb{E}\rho_\tau(Y_t^{(0)} - \boldsymbol{X}_t'\boldsymbol{\theta})$ where $\rho_\tau(z) = z(\tau - 1\{z < 0\})$ is know as the check function. To ensure that $\boldsymbol{\theta}_0$ is unique we impose the following

Assumption 4. *The distribution functions $F(y) = \mathbb{P}(Y_t \leq y | X_t = x_t)$ are*

(a) F is absolutely continuous with continuous density f

(b) f is uniformly bounded away from 0 and ∞ at the points $F^{-1}(\tau)$

(c) $\mathbb{E}(\boldsymbol{X}_t \boldsymbol{X}_t)$ is positive definite uniformly over $t \geq 1$.

We can let $\tau-$quantile error be $V_t(\tau) \equiv Y_t^{(0)} - \boldsymbol{X}_t'\boldsymbol{\theta}_0(\tau)$ and rewrite the model in the regression format as $Y_t^{(0)} = \boldsymbol{X}_t'\boldsymbol{\theta}_0(\tau) + V_t(\tau)$ with $\mathbb{P}\left(V_t(\tau) \leq 0\right) = \tau$. We can know define an unfeasible estimator given by

$$\widetilde{U}_t(\tau) = Y_t^{(1)} - \boldsymbol{X}_t'\boldsymbol{\theta}_0(\tau), \quad \forall \tau \in (0, 1). \tag{2}$$

The important property of this estimator is that it is by construction quantile unbiased for U_t is the sense that $Q_\tau(\widetilde{U}_t(\tau) - U_t) = 0$ for all $\tau \in (0, 1)$ or, equivalently, that $\widetilde{U}_t = U_t(\tau) + V_t(\tau)$. The estimator is unfeasible to be used in the post intervention sample because $\boldsymbol{\theta}_0$ is unknown can be the consistently estimated using the pre-intervention sample $\{y_t, \boldsymbol{x}_t\}_{t=1}^{T_0-1}$ by the minimizer of the sample counterpart of the minimization above, i.e., $\widehat{\boldsymbol{\theta}}(\tau) = \mathsf{arg\,min}_{\theta\in\Theta} \frac{1}{T_0}\sum_{t=1}^{T_0} \rho_\tau(y_t - \boldsymbol{X}_t'\boldsymbol{\theta}(\tau))$. Therefore we can define the feasible counterpart of (2) that will be used to investigate the intervention effects by

$$\widehat{U}_t(\tau) = Y_t^{(1)} - \boldsymbol{X}_t'\widehat{\boldsymbol{\theta}}(\tau); \qquad \forall \tau \in (0, 1) \tag{3}$$

for $t = T_0, \ldots, T$.

2.3 Asymptotics

In this section we deal with two kind of asymptotic results. In the first one which we call *Partial Asymptotics* we consider an asymptotic approach only for the pre-intervention period as per SC where the number of post intervention period $T_2 \equiv T - T_0$ are kept fixed while $T_0 \to \infty$. This approach is tailored to accommodate situations where the number of pre-intervention period T_0 is much larger then T_2, which justifies the sampling error from the estimation of θ_0 by $\widehat{\theta}$ to be of smaller order then $V_t(\tau)$.

In the second approach named *Full Asymptotics* we establish the asymptotic properties by considering the whole sample increasing, while the proportion between the pre-intervention to the post-intervention sample size is constant. Here $T_0 = \lfloor \lambda_0 T \rfloor$ for fixed

$\lambda_0 \in (0,1)$ such that $T_0 \equiv T_0(T)$ and 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.

2.3.1 Partial Asympotics

The first result we can obtain from letting only the pre-intervention period increase is the consistency of $\widehat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_0$. The consistency result for non-trending regressors in quantile regression appears in several papers and a comprehensive treatment including references can be found in Koencker (2005). In particular we need the following addition assumption

Assumption 5. *There exist positive definite matrices A and $B(\tau)$ such that*

(a) $\boldsymbol{\eta}_t \equiv (\eta_{1,t}, \ldots, \eta_{n,t})' \sim (0, \boldsymbol{\Pi})$ *with* $\boldsymbol{\Pi}$ *diagonal with all elements strictly positive.*

(b) $\boldsymbol{F}_t \sim (0, \boldsymbol{A})$

(c) There exist a positive function $N(T)$ such that $\frac{1}{N(T)} \sum_{t=1}^{T} \zeta(t)$ is bounded away from 0 and ∞.

Condition (a) and (b) of Assumption 5 are necessary to ensure that $\mathbb{E}(\boldsymbol{X}_t \boldsymbol{X}_t)$ is well defined for all $t \geq 1$ and is positive definite. Even though, the limit of $\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(\boldsymbol{X}_t \boldsymbol{X}_t)$ as $T \to \infty$ is rank deficient in the presence of deterministic term. It requires then a different approach to show consistency of the quantile regression parameters to deal with the regressors asymptotic multicolinearity. To the best of my knowledge the result below is novel and it relies on transformation of regressors similar to the one consider in Stock (2010) and Hamilton (1994). Finally, The last condition ensure that we can bound the sum of the deterministic trend. For the polynomial case $\zeta(t) = t^{\alpha}$, for $\alpha \geq 0$ we can set $N(T) = T^{1+\alpha}$ and we have the limit to be $\frac{1}{1+\alpha}$.

Theorem 1. *Under Assumptions 1 to 5, $\widehat{\boldsymbol{\theta}}_T(\tau) \xrightarrow{p} \boldsymbol{\theta}_0(\tau)$ as $T_0 \to \infty$ for all $\tau \in (0,1)$.*

The important consequence of Theorem 6 is that we are able to construct confidence bands for the unobservable $Y_t^{(0)}$ after the intervention. In particular we have as a direct corollary that $\mathbb{P}(Y_t^{(0)} \leq \widehat{Y}_t(\tau)) \longrightarrow \mathbb{P}(Y_t^{(0)} \leq \widetilde{Y}_t(\tau)) = \tau$.

Instead of using directly $\{\widehat{U}_t(\tau)\}_{t \geq T_0}$ as the basis of our inference procedure to test potential difference in the quantiles after the intervention, it will be proven more convenient to use $\{\widehat{M}_t(\tau_1, \tau_2)\}_{t \geq T_0}$ where $\widehat{M}_t(\tau_1, \tau_2) = I(\widehat{Y}_t(\tau_1) \leq Y_t^{(0)} \leq \widehat{Y}_t(\tau_2))$ for $\tau_1 < \tau_2$ and $\widehat{Y}_t(\tau) = \boldsymbol{X}_t' \widehat{\boldsymbol{\theta}}(\tau)$. Clearly its unobservable counterpart $M_t(\tau_1, \tau_2) \sim Bernoulli(\tau_2 - \tau_1)$ under the null hypothesis. Suppose that there is a single post-intervention observation ($T_2 = 1$), in that a case a simple hypothesis test would be to reject H_0 if $\widehat{M}(\tau_1, \tau_2) = 0$

with significance level of $1-(\tau_2-\tau_1)$. Alternatively we could find τ^* such that $\widehat{Y}(\tau^*) = Y^{(1)}$, then report the double-sided p-value as τ^* if $\tau^* \leq 1/2$; or $(1-\tau^*)$ otherwise.

For more than one post-intervention period $T_2 > 1$, provided that $\{V_t\}_{t\geq T_0}$ is iid, we have $S(\tau_1, \tau_2) \equiv \sum_{t>T_0} M_t(\tau_1, \tau_2) \sim Binomial(T_2, \tau)$. Once again a simple test would reject the null whenever $\widehat{S}(\tau_1, \tau_2) < T_2$, i.e. whenever at least one observation in post intervention period falls outside the $[\widehat{Y}_t(\tau_1), \widehat{Y}_t(\tau_2)]$ bracket, which occurs with probability $1-(\tau_2-\tau_1)^{T_2}$. If we consider a symmetric interval such that $\tau_2 = 1-\tau_1$ and set $\tau_1 = \frac{1-(1-\alpha)^{1/T_2}}{2}$ we obtain any arbitrary significance level $\alpha \in (0,1)$. Also, let p_i denote the p-value of the i-th observation calculated as described in the previous paragraph, then the overall test p-value is given by $\max(p_{T_0+1}, \ldots, p_T)$.

Theorem 2. *For a given significance level* $\alpha \in (0, 0.5)$, *let* $\tau = \frac{1-(1-\alpha)^{1/T_2}}{2}$ *and* $\widehat{S}(\tau, 1-\tau)$. *Then a test that reject* H_0 *when* $\widehat{S}(\tau, 1-\tau) < T_2$ *has size size approaching* α *as* $T_0 \to \infty$.

2.3.2 Full Asymptotic

Theorem 3. *Under Assumptions 1 to 5,* $\sqrt{T_0}(\widehat{\boldsymbol{\theta}}_T(\tau) - \boldsymbol{\theta}_0(\tau)) \xrightarrow{d} N(0,1)$ *as* $T_0 \to \infty$ *uniformly in* $\tau \in (0,1)$.

Proof. We prove the result using an argument parallel to Andrews (1994). We start by taking a mean value expansion of $\boldsymbol{S}_T^0(\boldsymbol{\eta}_0,)$ around $\boldsymbol{\eta}_0$, which yields

$$\mathbf{0} = \sqrt{T_0}\boldsymbol{S}_T^0(\theta_0) = \sqrt{T_0}\boldsymbol{S}_T^0(\widehat{\theta}) - \boldsymbol{H}_T^0(\widetilde{\theta})\sqrt{T_0}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$$

Due to the potential trend regressors we need to transform the regressors as in the consistency proof let $\boldsymbol{\eta}(\cdot) = \boldsymbol{C}^{-1}\boldsymbol{\theta}(\cdot)$ is Gaussian. Now if we multiply the last display by $\boldsymbol{\Lambda}_T^{-1}$ and define $\boldsymbol{S}_T(\boldsymbol{\eta}, \tau) := \frac{1}{T_0}\sum_{t=1}^{T_0} \boldsymbol{W}_{t,T}\psi_\tau(Y_t - \widetilde{\boldsymbol{X}}_t'\boldsymbol{\eta})$, $\boldsymbol{S}_T^0(\boldsymbol{\eta}, \tau) := \mathbb{E}\boldsymbol{S}_T(\boldsymbol{\theta}, \tau)$, $\boldsymbol{H}_T(\boldsymbol{\eta}, \tau) := \frac{1}{T_0}\sum_{t=1}^{T_0} \boldsymbol{W}_{t,T}\boldsymbol{W}_{t,T}' f(\widetilde{\boldsymbol{X}}_t'\boldsymbol{\eta})$ and $\boldsymbol{H}_T^0(\boldsymbol{\eta}, \tau) := \mathbb{E}\boldsymbol{H}_T(\boldsymbol{\eta}, \tau)$.

$$\mathbf{0} = \sqrt{T_0}\boldsymbol{S}_T^0(\boldsymbol{\eta}_0) = \sqrt{T_0}\boldsymbol{S}_T^0(\widehat{\boldsymbol{\eta}}) - \boldsymbol{H}_T^0(\widetilde{\boldsymbol{\eta}})\sqrt{T_0}\boldsymbol{\Lambda}_T(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$$

where $\widetilde{\boldsymbol{\eta}}$ lies in the segment from $\widehat{\boldsymbol{\eta}}$ and $\boldsymbol{\eta}_0$.

First, due to the uniform consistency of $\widehat{\boldsymbol{\eta}}$ to $\boldsymbol{\eta}_0$ and the fact that $\boldsymbol{H}_T^0(\cdot)$ is continuous we have $\boldsymbol{H}_T^0(\widetilde{\boldsymbol{\eta}}) = \boldsymbol{H}_T^0(\boldsymbol{\eta}_0) + o_p(1)$ uniformly in $\tau \in \mathcal{T}$ so we can write

$$(\boldsymbol{H}_T^0(\boldsymbol{\eta}_0) + o_p(1))\sqrt{T_0}\boldsymbol{\Lambda}_T(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) = \sqrt{T_0}\boldsymbol{S}_T^0(\widehat{\boldsymbol{\eta}}).$$

Now the left hand side term can be decompose as

$$\begin{aligned}\sqrt{T_0}\boldsymbol{S}_T^0(\widehat{\boldsymbol{\eta}}) &= \left(\sqrt{T_0}\boldsymbol{S}_T(\widehat{\boldsymbol{\eta}}) - \sqrt{T_0}\boldsymbol{S}_T^0(\widehat{\boldsymbol{\eta}})\right) - \sqrt{T_0}\boldsymbol{S}_T(\widehat{\boldsymbol{\eta}}) \\ &= \left(\sqrt{T_0}\boldsymbol{M}_T(\widehat{\boldsymbol{\eta}}) - \sqrt{T_0}\boldsymbol{M}_T(\boldsymbol{\eta}_0)\right) + \sqrt{T_0}\boldsymbol{M}_T(\boldsymbol{\eta}_0) - \sqrt{T_0}\boldsymbol{S}_T(\widehat{\boldsymbol{\eta}})\end{aligned}$$

where $\boldsymbol{M}_T(\boldsymbol{\eta}) := \boldsymbol{S}_T(\boldsymbol{\eta}) - \boldsymbol{S}_T^0(\boldsymbol{\eta})$. Since the process $\sqrt{T_0}\boldsymbol{M}_T(\cdot)$ is stochastic equicontinuos in $(\boldsymbol{\eta}, \tau)$ (proof?) we have that first term in brackets is $o_P(1)$ uniformly in $\tau \in \mathcal{T}$. Also, from Theorem 3.3 of Koenker and Bassett (1978) we have that under the assumptions considered the last term is also $o_p(1)$ uniformly in $\tau \in \mathcal{T}$. Thus we have that $\sqrt{T_0}\boldsymbol{S}_T^0(\widehat{\boldsymbol{\eta}}) = \sqrt{T_0}\boldsymbol{M}_T(\boldsymbol{\eta}_0) + o_p(1)$ uniformly in $\tau \in \mathcal{T}$. Putting thins together we have the asymptotic linear representation

$$(\boldsymbol{H}_T^0(\boldsymbol{\eta}_0(\tau), \tau) + o_p(1))\sqrt{T_0}\boldsymbol{\Lambda}_T(\widehat{\boldsymbol{\eta}}(\tau) - \boldsymbol{\eta}_0(\tau)) = \sqrt{T_0}\boldsymbol{M}_T(\boldsymbol{\eta}_0(\tau), \tau) + o_p(1)$$

We are now left to prove a uniform central limit theorem for the process $\boldsymbol{G}_T(\cdot) := \sqrt{T_0}\boldsymbol{M}_T(\boldsymbol{\eta}_0(\cdot), \cdot) = \frac{1}{\sqrt{T_0}}\sum_{t=1}^{T_0}\boldsymbol{Z}_{t,T}(\cdot)$, where $\boldsymbol{Z}_{t,T}(\cdot) := \boldsymbol{W}_{t,T}V_{0,t}(\cdot) - \mathbb{E}(\boldsymbol{W}_{t,T}V_{0,t}(\cdot))$. For any finite subset $\{\tau_1, \ldots, \tau_k\} \subset \mathcal{T}$ we have form a multivariate central limit theorem that $(\boldsymbol{G}_T(\tau_1), \ldots, \boldsymbol{G}_T(\tau_k))' \xrightarrow{d} (\boldsymbol{G}(\tau_1), \ldots, \boldsymbol{G}(\tau_k))'$ where $\boldsymbol{G}(\cdot)$ is a zero mean Gaussian process defined by its covariance matrix given by $\boldsymbol{\Omega}(\tau_1, \tau_2) := \lim_{T_0\to\infty}\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{E}\boldsymbol{Z}_{t,T}(\tau_1)\boldsymbol{Z}_{t,T}(\tau_2)'$. Let $\boldsymbol{\Delta}_T(\tau_1, \tau_2) := \boldsymbol{G}_T(\tau_1) - \boldsymbol{G}_T(\tau_2)$ Let $d(\tau_1, \tau_2)$ be a metric defined by

$$d^2(\tau_1, \tau_2) := \sup_{T\geq 1}\max_{1\leq i,j\leq n}\mathbb{E}(\boldsymbol{\Delta}_T(\tau_1, \tau_2)\boldsymbol{\Delta}_T(\tau_1, \tau_2)'), \quad \forall(\tau_1, \tau_2) \in \mathcal{T}^2,$$

then the $\boldsymbol{G}_T(\cdot)$ is stochastic equicontinuos with respect to th metric $d(\cdot, \cdot)$ because take any pair (τ_1, τ_2) such that $d(\tau_1, \tau_2) < \delta$, then by Chebischev inequality we have for any $\gamma > 0$

$$\mathbb{P}(\|\boldsymbol{\Delta}_T(\tau_1, \tau_2)\|_2 > \gamma) \leq \frac{\mathbb{E}(\boldsymbol{\Delta}_T'\boldsymbol{\Delta}_T)}{\gamma^2} \leq \frac{nd^2(\tau_1, \tau_2)}{\gamma^2} < \frac{n\delta^2}{\gamma^2}$$

Hence for any given $\epsilon > 0$ and $\gamma > 0$ we can take $\delta = \gamma\sqrt{\epsilon/n}$ such that for all $T \geq 1$ we have the stochastic equicontinuity, namely $\mathbb{P}(\sup_{d(\tau_1,\tau_2)<\delta}\|\boldsymbol{\Delta}_T(\tau_1, \tau_2)\|_2 > \gamma)) < \epsilon$. Therefore $\boldsymbol{G}(\cdot) \Rightarrow \boldsymbol{G}(\cdot)$ as per Dudley (1995) and Van der Vaart and Weller (1996).

Let $\boldsymbol{H}(\cdot) := \lim_{T_0\to\infty}\boldsymbol{H}_T^0(\theta_0, \cdot)$ which exists uniformly since $f(F^-(\tau))$ is uniformly bounded across $\tau \in (0, 1)$ under assumption 4. Finally, putting everything together per Slutsky Theorem we have

$$\boldsymbol{H}(\cdot)\sqrt{T_0}\boldsymbol{\Lambda}^T(\widehat{\boldsymbol{\eta}}(\cdot) - \boldsymbol{\eta}_0(\cdot)) \Rightarrow \boldsymbol{G}(\cdot)$$

□

Lemma 1. *Let* $M_T(\boldsymbol{\theta},\tau) := \frac{1}{T_2}\sum_{t\geq T_0} I(Y_t \leq \boldsymbol{X}_t'\boldsymbol{\theta}) - \tau$, $M_T^0(\boldsymbol{\theta},\tau) := \mathbb{E}M_T(\boldsymbol{\theta},\tau)$ *and* $K_T(\boldsymbol{\theta}) := \frac{1}{T_2}\sum_{t\geq T_0} \boldsymbol{X}_t f(\boldsymbol{X}_t'\boldsymbol{\theta})$ *and* $\boldsymbol{K}_T^0(\boldsymbol{\theta}) := \mathbb{E}\boldsymbol{K}_T(\boldsymbol{\theta})$, *then uniformly in* $\tau \in \mathcal{T}$

$$M_T(\widehat{\boldsymbol{\theta}}(\tau),\tau) = M_T(\boldsymbol{\theta}_0(\tau),\tau) - \boldsymbol{K}_T^0(\boldsymbol{\theta}_0(\tau))(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)) + o_P(1/\sqrt{T})$$

Proof. For fixed $\tau \in \mathcal{T}$, a mean value expansion of $L_T^0(\boldsymbol{\theta}_0(\tau),\tau)$ around $\widehat{\boldsymbol{\theta}}(\tau)$ yields

$$0 = M_T^0(\boldsymbol{\theta}_0(\tau),\tau) = M_T^0(\widehat{\boldsymbol{\theta}}(\tau),\tau) - \boldsymbol{K}_T^0(\widetilde{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)),$$

where $\widetilde{\boldsymbol{\theta}}$ lies in between $\widehat{\boldsymbol{\theta}}(\tau)$ and $\boldsymbol{\theta}_0(\tau)$. By adding and subtracting both $M_T(\widehat{\boldsymbol{\theta}}(\tau),\tau)$ and $M_T(\boldsymbol{\theta}_0(\tau),\tau)$ to last expression and rearranging we are left with

$$M_T(\widehat{\boldsymbol{\theta}}(\tau),\tau) = M_T(\boldsymbol{\theta}_0(\tau),\tau) - \boldsymbol{K}_T^0(\boldsymbol{\theta}_0(\tau))(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)) + R_T,$$

where $R_T := L_T(\widehat{\boldsymbol{\theta}}(\tau),\tau) - L_T(\boldsymbol{\theta}_0(\tau),\tau) + (\boldsymbol{K}_T^0(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{K}_T^0(\boldsymbol{\theta}_0(\tau)))(\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau))$ with $L_T(\boldsymbol{\theta},\tau) := M_T(\boldsymbol{\theta},\tau) - M_T^0(\boldsymbol{\theta},\tau)$. Now by Theorem (asymptotic normality) we have that $\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau) = O_P(1/\sqrt{T})$ uniformly in $\tau \in \mathcal{T}$, also $\boldsymbol{K}_T^0(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{K}_T^0(\boldsymbol{\theta}_0(\tau)) = o_P(1)$ uniformly in $\tau \in \mathcal{T}$ by Theorem (consistency). Then the third term of R_T is $o_P(1/\sqrt{T})$ uniformly.

□

The full asymptotic analysis will be based on the one-sided version of $M(\tau_1,\tau_2)$ used in previous section, i.e., $M_t(\tau) \equiv Y_t^{(1)} - g_\tau(\boldsymbol{X},\boldsymbol{\theta}_0(\tau))$ with its estimator $\widehat{m}_t(\tau)$ defined in (3) and use the empirical distribution function (EDF) of $\widehat{m}_t(\tau) - \Delta_t$ evaluated at zero as our estimator.

$$G_T(\tau) = \frac{1}{T_2}\sum_{t>T_0} I(Y_t^{(1)} \leq \widehat{Y}_t(\tau)) \tag{4}$$

The next result is a consequence of $\widehat{Y}_t$ being consistent for $\widetilde{Y}_t$ in the asymptotics for the pre-intervention period and noticing that (4) is in effect the empirical distribution function (EDF) of V_t under the null

Theorem 4. *Under* H_0 *and assumptions 1 to 5,* $G_T(\tau) \xrightarrow{p} \tau$ *as* $T \to \infty$ *for all* $\tau \in (0,1)$.

Working with the EDF 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 $W_t(\tau) = I(v_t(\tau) \leq 0) - \tau$, under the null and Assumption 2, $\{w_t(\tau)\}_t$ 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 2. Hence, taking into account the uncertainty on the estimation of $\boldsymbol{\theta}_0$ during the pre intervention period, we have

Theorem 5. *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 1-5:*

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

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 5

Corollary 1. *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 1-5:*

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

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

Corollary 2. *Let* $W_T = \{W_T(\tau), \tau \in (0,1)\}$. *Under Assumptions 1-5:*

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

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

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

, *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 Inference

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

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

As a consequence, under the null and Assumption 2, the conditional distribution $F_{Y|\boldsymbol{X}}$ is unaltered. Hence (6) implies the equality of the conditional quantiles of $Y_t|\boldsymbol{X}_t$.

However, (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 $(Y_1|\boldsymbol{X}_1, \ldots, Y_T|\boldsymbol{X}_T)$. For that reason we postulate 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.

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

Once the asymptotic normality of the $\widehat{\tau}_T$ is ensured (Theorem 5) 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 2. *Let $\widehat{\boldsymbol{\Sigma}}_T$ be a consistent estimator for $\boldsymbol{\Sigma}_T$ and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)' \in (0,1)^k$. Under Assumptions 1-5 and $\mathcal{H}_0^\tau$:*

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

, 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 5

Lemma 3. *Under Assumptions 1-5 and $\mathcal{H}_0$:*

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

, 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 continuous transformation of W_T, the next lemma follows from the continuous mapping theorem.

Lemma 4. *For $p \in [1,\infty]$, under Assumptions 1-5 and $\mathcal{H}_0$:*

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

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

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 $W_\infty \equiv \sup_{u\in[0,1]} \mathcal{B}(u)$, 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

4 Misspecification of the Conditonal Quantile Model

Now we examine the consequences of misspecification in the conditional quantile model (CMQ). In the first subsection we derive the exact CQM for the Gaussian factor model which motives (justify) the adoption of the linear CQM extensively used in quantile regression (QR). However, due to the lack of a law of iterated expectation for quantiles[2], QR no longer gives us the minimum mean square error linear approximation of the CQM in case of misspecification. In fact, as shown in Angrist et al. (2006), the QR coefficients minimize the expected weighted mean-squared approximation error, i.e., the square of the difference between the true CQM and a linear approximation with weighting function $w_\tau(X,\theta)$ defined as

$$w_\tau(X,\theta) \equiv \int_0^1 (1-u)f_{Y|X}(uX'\theta + (1-u)Q_{Y|X}(\tau,X))\mathrm{d}u$$

To assess the impact of misspecification we conduct a Monte Carlos study in Subsection 4.2 using different distribution for the innovations of the DGP in Assumption 2 all in a linear specification of the CQM.

[2]Since in general $\mathbb{E}_X(Q(Y|X) \neq Q(Y)$

4.1 Exact Guassian Factor Model

Here we derive the exact conditional quantile model for the DGP of Assumption 2 where we consider that 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 polynomial $\mathbf{\Psi}(L)$, 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 known 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

$$\begin{aligned} \tfrac{1}{\sigma}\boldsymbol{\nu}_T &\sim \mathcal{N}(-\Phi^{-1}(\tau), \Lambda) \\ \lambda_j = \frac{\mathbb{C}(\nu_i, \nu_j)}{\sigma^2} &= \frac{(1, -\boldsymbol{\beta})\mathbf{\Gamma}_j(1, -\boldsymbol{\beta})'}{(1, -\boldsymbol{\beta})\mathbf{\Gamma}_0(1, -\boldsymbol{\beta})'} \end{aligned}$$

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

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

4.2 Monte Carlo

We conducted a Monte Carlo study by simulating the DGP described in Assumption 2 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 1 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.

5 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:[3] 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

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

Table 1: 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
$\Vert\cdot\Vert_\infty$	0.0879	0.0631	0.0432	0.0201	0.0077
$\Vert\cdot\Vert_2$	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
$\Vert\cdot\Vert_\infty$	0.0936	0.0692	0.0469	0.0237	0.0077
$\Vert\cdot\Vert_2$	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
$\Vert\cdot\Vert_\infty$	0.0916	0.0673	0.0438	0.0188	0.0071
$\Vert\cdot\Vert_2$	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 normalized to mean zero and unit variance for each of the distributions considered 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 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 2

Table 2: Analized Cases of Change in Corporate Governance Regime

Treated	Segment	Migration Date /Level	Peers	T	$\lambda_0 = T_0/T$
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 3: 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.

6 Conclusion

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

Technical Appendix

We use the empirical process notation. Let $\mathbb{P}_T(A) := \delta_{Z_t}(A)$ be the empirical measure associated with $\mathbb{P}$ where δ_Z is the Dirac measure. For measurable f let $\mathbb{E}_T f := \frac{1}{T}\sum_{t=1}^T f(Z_t)$, $\mathbb{E} f := \int \frac{1}{T}\sum_{t=1}^T f(Z_t) d\mathbb{P}$ and $\mathbb{Q}_T f := (\mathbb{E}_T - \mathbb{E})f$.

To ease the notation Recall that $Z_t = \delta\zeta(t) + u_t$ and let $\eta := (1, -\theta')'$ such that we write $Y_t - X_t'\theta = Z_t'\eta$.

Lemma 5. *$A_t(\theta) := \mathbb{E}(\eta' Z_t)(1\{\eta' Z_t < 0\}) - \mathbb{E}(\eta' Z_t 1\{\eta' Z_t < 0\}) = o_P(1)$ uniformly in $\theta \in \Theta$, provided that $\eta'\delta \neq 0$ and $\zeta(t) \to \infty$ as $t \to \infty$.*

Proof. For the pointwise convergence, $A_t(\theta) = \eta'\delta\zeta_t[1\{\eta' Z_t < 0\} - \mathbb{P}(\eta' Z_t < 0)] + \eta'\mathbb{E}(U_t)1\{\eta' Z_t < 0\} - \mathbb{E}(U_t 1\{\eta' Z_t < 0\})]$. Both terms converges to zero in probability. To see that, we first show that $B_t(\theta) := \zeta_t[1\{\eta' Z_t < 0\} - \mathbb{P}(\eta' Z_t < 0)] = o_P(1)$. Notice $\mathbb{E}|1\{Z_t'\eta < 0\} - \mathbb{P}(Z_t'\eta < 0)| = 2\mathbb{P}(Z_t'\eta < 0)(1 - \mathbb{P}(Z_t'\eta < 0))$ and $\mathbb{P}(Z_t'\eta < 0) = \mathbb{P}(U_t'\eta < -\zeta(t)\delta'\eta)$. If $\delta'\eta < 0$, then the right hand size diverges to ∞ and the probability converges to zero at a rate $O(\zeta(t)^{-\alpha})$ for some $\alpha > 2$ due to the existence of finite second moment (assumption). On the other hand, if $\delta'\eta > 0$, then the right hand size diverges to $-\infty$ and the probability converges to one and $1 - \mathbb{P}(Z_t\eta < 0)$ converges to zero also at rate $O(\zeta(t)^{-\alpha})$. Therefore $\mathbb{E}|1\{Z_t'\eta < 0\} - \mathbb{P}(Z_t'\eta < 0)| = O(\zeta(t)^{-\alpha})$ and the Markov inequality gives us a upper bound on the probability of $|B_t(\theta)| > \epsilon$ of order $O(\zeta(t)^{1-\alpha})$ which converges to zero.

To show the uniformity of the convergence notice that □

Lemma 6 (Pointwise Convergence)**.** *For all $(\theta, \tau) \in \Theta \times \mathcal{T}$, $\mathbb{Q}_T\rho(\theta, \tau) = o_P(1)$.*

Proof. Recall $Z_t = \delta\zeta(t) + u_t$. Let $\eta := (1, -\theta')'$, then $\mathbb{Q}_T\rho(\theta, \tau) = \mathbb{Q}_T[-\delta'\eta\zeta(t)1\{Z_t'\eta < 0\}] + \mathbb{Q}_T[u_t'\eta(\tau - 1\{Z_t'\eta < 0\})]$. If $\delta'\eta = 0$, the first term is zero and the second converge in probability to zero since $L_t := u_t'\eta(\tau - 1\{Z_t'\eta < 0\})$ is an alpha mixing process with $\mathbb{E}|L_t|^{2+\epsilon} \leq 2\|\eta\|^{2+\epsilon}\mathbb{E}\|Z_t\|^{2+\epsilon} < \infty$. When $\delta'\eta \neq 0$ and $\zeta(t) \to \infty$ as $t \to \infty$, by Markov Inequality we have for an arbitrary $\epsilon > 0$, $\mathbb{P}(|\mathbb{Q}_T[-\delta'\eta\zeta(t)1\{Z_t'\eta < 0\}]| > \epsilon) \leq \epsilon^{-1}\mathbb{E}|\mathbb{Q}_T[-\delta'\eta\zeta(t)1\{Z_t'\eta < 0\}|$, which can be bounded using the triangle inequality by $\epsilon^{-1}\delta'\eta\mathbb{E}_T\zeta(t)\mathbb{E}|1\{Z_t'\eta < 0\} - \mathbb{P}(Z_t'\eta < 0)|$, which converges to zero. , Therefore $\mathbb{E}|1\{Z_t'\eta < 0\} - \mathbb{P}(Z_t'\eta < 0)| = O(\zeta(t)^{-\alpha})$ and the bound becomes $\epsilon^{-1}\delta'\eta\mathbb{E}_T O(\zeta(t)^{1-\alpha}) = o(1)$ as $t \to \infty$. □

Lemma 7. *For fixed $\tau \in \mathcal{T}$, $\mathbb{Q}_T\rho(\theta, \tau) = o_P(1)$ uniformly in $\theta \in K$ where $K \subseteq \Theta$ is a compact.*

Proof. Let $W_t(\theta, \tau) := \rho_t(\theta, \tau) - \mathbb{E}\rho_t(\theta, \tau)$, where $\rho_t(\theta, \tau) := (Y_t - X_t'\theta)(\tau - 1\{Y_t - X'\theta < 0\})$, then we can write $W_t(\theta, \tau) = \eta'(Z_t - \mathbb{E}Z_t)(\tau - 1\{\eta' Z_t < 0\}) + A_t(\theta)$ where $A_t(\theta) :=$

$\mathbb{E}(\eta' Z_t)(1\{\eta' Z_t < 0\}) - \mathbb{E}(\eta' Z_t 1\{\eta' Z_t < 0\})$. We first show that $A_t(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$ as $t \to \infty$, provided that $\delta'\eta \neq 0$ and $\zeta(t) \to \infty$. $A_t(\theta) =$

To upgrade the pointwise convergence (Lemma 5). It is enough to show that $J_T(\theta) := \mathbb{Q}_T \rho(\theta, \tau)$ is stochastic equicountinuos. $|J_T(\theta) - J_T(\theta^*)| \leq \|Z_t - \mathbb{E}Z_t\|_2 \|\theta - \theta^*\|_2$ □

Lemma 8. *Let $g : \Theta \times \mathcal{X} \to \mathbb{R}$ be a random function where (i) Θ is compact; $(X_t)_t$ is an α-mixing sequence of random variables taking value on $\mathcal{X} \subseteq \mathbb{R}^p$ for $p \geq 1$, (ii) $\theta \mapsto g(.,\theta)$ is almost surely continuous there exist a measurable $G(X_t)$ such that (iii) $|g(X_t,\theta)| \leq G(X_t)$ for all $\theta \in \Theta$ and (iv) $\mathbb{E}|G(X_t)|^{2+\gamma} < \infty$ for some $\gamma > 0$, then*

$$\sup_{\theta\in\Theta} \left| \tfrac{1}{T} \sum_{t=1}^{T} g(X_t,\theta) - \mathbb{E}g(X_t,\theta) \right| \xrightarrow{p} 0$$

Proof. For any $\delta > 0$, let $\Delta_t(\theta,\delta) := \sup_{\theta' \in B(\theta,\delta)} g(X_t,\theta') - \inf_{\theta' \in B(\theta,\delta)} g(X_t,\theta')$ where $B(\theta,\delta)$ denotes the open ball centered in θ with radius δ. We want show that $\mathbb{E}\Delta_t(\theta,\delta) \downarrow 0$ as $\delta \downarrow 0$, for all $\theta \in \Theta$. We first show that $\Delta_t(\theta,\delta) \downarrow 0$ almost surely as $\delta \downarrow 0$. Fix $\epsilon > 0$, set $\delta = 1/n$ and let $A_n := \{\omega \in \Omega : \Delta(\theta, 1/n) > \epsilon\}$. Then the almost surely convergence is equivalent to show that the probability of $\limsup_{n\to\infty} A_n$ is zero. Since $(A_n)_n$ is non-increasing (with respect to set inclusion), it is also equivalent to show that $\inf_{n\in\mathbb{N}} A_n$ is a null set, which follows from the almost surely continuity (condition (ii)) of $g(.,\theta)$. Now we can bound $\Delta(\theta,\delta)$ using condition (i)-(iii) since $\Delta(\theta,\delta) \leq 2\sup_{\theta\in\Theta}|g(X,\theta)| \leq 2G(X)$ since a continuous function on compact attains its supremum. Finally by (iv) we have the result applying the Dominated Convergence Theorem to conclude $\lim_{n\to\infty} \mathbb{E}\Delta_t(\theta,\delta) = \mathbb{E}\lim_{n\to\infty} \Delta_t(\theta,\delta) = 0$.

Hence we know that for all $\epsilon > 0$ and for all $\theta \in \Theta$, there is a $\delta := \delta(\theta,\epsilon)$ such that $\mathbb{E}\Delta_t(\theta,\delta) < \epsilon$. Moreover since Θ is compact it can be covered by finite open balls such that $\Theta \subseteq \cup_{k=1}^{K} B(\theta_k,\delta_k)$ where $\delta_k := \delta(\theta_k,\epsilon)$. Then

$$\begin{aligned}
\sup_{\theta\in\Theta} \left[\tfrac{1}{T} \sum_{t=1}^{T} g(X_t,\theta) - \mathbb{E}g(X_t,\theta) \right] &\leq \max_k \sup_{\theta\in B(\theta_k,\delta_k)} \left[\tfrac{1}{T} \sum_{t=1}^{T} g(X_t,\theta) - \mathbb{E}g(X_t,\theta) \right] \\
&\leq \max_k \left[\tfrac{1}{T} \sum_{t=1}^{T} \sup_{\theta\in B(\theta_k,\delta_k)} g(X_t,\theta) - \mathbb{E} \inf_{\theta\in B(\theta_k,\delta_k)} g(X_t,\theta) \right] \\
&\leq \max_k \left[\tfrac{1}{T} \sum_{t=1}^{T} \mathbb{E} \sup_{\theta\in B(\theta_k,\delta_k)} g(X_t,\theta) - \mathbb{E} \inf_{\theta\in B(\theta_k,\delta_k)} g(X_t,\theta) \right] + o_P(1) \\
&\leq \max_k \tfrac{1}{T} \sum_{t=1}^{T} \mathbb{E}\Delta(\theta_k,\delta_k) + o_P(1) \leq \epsilon + o_P(1)
\end{aligned}$$

where the third inequality follows from the Law of Large numbers, since $Y_t := \sup_{\theta\in B(\theta_k,\delta_k)} g(X_t,\theta) - \mathbb{E}\sup_{\theta\in B(\theta_k,\delta_k)} g(X_t,\theta)$ is a zero mean alpha mixing process with $\mathbb{E}|Y_t|^{2+\epsilon} \leq$ □

Lemma 9. *There is a compact $K \subseteq \Theta$ such that a minimizer of $J_T(\boldsymbol{\theta}) := \frac{1}{T}\sum_{t=1}^{T}\rho_\tau(Y_t - X_t'\boldsymbol{\theta})$ with respect to $\theta \in \Theta \subseteq \mathbb{R}^k$ exists with probability approaching 1 uniformly in $\tau \in \mathcal{T} := [\gamma, 1-\gamma]$ for some $\gamma \in (0, 0.5)$*

Proof. Let $H(\boldsymbol{\delta}, \tau) := \frac{1}{T}\sum_{t=1}^{T}\rho_\tau(\nu_t - X_t'\boldsymbol{\delta})$ where $\boldsymbol{\delta} := \theta - \theta_0$. we want to show that $H(\cdot,\cdot)$ is coercive to $+\infty$ with high probability, for that

$$\begin{aligned} H_T(\boldsymbol{\delta}, \tau) &= \tfrac{1}{T}\sum_{t=1}^{T}(V_t - \boldsymbol{X}_t'\boldsymbol{\delta})(\tau - I(V_t - \boldsymbol{X}_t'\boldsymbol{\delta} < 0)) \\ &\geq \tfrac{1}{T}\sum_{t=1}^{T}|V_t - \boldsymbol{X}_t'\boldsymbol{\delta}|\gamma \\ &\geq \gamma\left(\tfrac{1}{T}\sum_{t=1}^{T}|\boldsymbol{X}_t'\boldsymbol{\delta}| - \tfrac{1}{T}\sum_{t=1}^{T}|V_t|\right) \\ &= \gamma\left(\tfrac{1}{T}\sum_{t=1}^{T}\sqrt{\boldsymbol{\delta}'\boldsymbol{X}_t\boldsymbol{X}_t'\boldsymbol{\delta}} - \widehat{\nu}_T\right) \\ &\geq \gamma\left(\sqrt{\boldsymbol{\delta}'\boldsymbol{\Omega}_T\boldsymbol{\delta}} - \widehat{\nu}_T\right) \\ &\geq \gamma\left(\sqrt{\lambda_{\min}(\boldsymbol{\Omega}_T)}\|\boldsymbol{\delta}\|_2 - \widehat{\nu}_T\right), \end{aligned}$$

where $\widehat{\boldsymbol{\Omega}}_T := \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\boldsymbol{X}_t'$ and $\widehat{\nu}_T := \frac{1}{T}\sum_{t=1}^{T}|V_t|$. Now by the Law o Large numbers we have that $\widehat{\boldsymbol{\Omega}}_T - \boldsymbol{\Omega}_T = o_P(1)$ and $\widehat{\nu}_T - \nu = o_P(1)$, where $\boldsymbol{\Omega}_T := \mathbb{E}\widehat{\boldsymbol{\Omega}}_T$ and $\nu := \mathbb{E}\widehat{\nu}_T$. Also, under Assumption 4(c), $\lambda_{\min}(\boldsymbol{\Omega}_T)$ is bound below uniformly by $\underline{\lambda} > 0$ say, then we can write the for $\|\boldsymbol{\delta}\|_2 > r$ for an arbitrary radius $r > 0$ we have

$$\inf_{\|\delta\|_2 > r}\inf_{\tau\in\mathcal{T}} H_T(\boldsymbol{\delta}, \tau) \geq \gamma\left(\sqrt{\underline{\lambda}}r - \nu\right) + o_P(1)$$

From the last expression we conclude that $\lim_{r\to\infty}\inf_{\|\delta\|_2>r}\inf_{\tau\in\mathcal{T}} H_T(\boldsymbol{\delta}, \tau) = \infty$. Hence for any $M > 0$, there exist an $\bar{r}$ such that whenever $\|\boldsymbol{\delta}\|_2 > \bar{r}$ we have that $H_T(\boldsymbol{\delta}, \tau) > M$ with probability approaching one uniformly in $\tau \in \mathcal{T}$ which implies that the lower level set $\{\theta \in \Theta : H_T(\theta, \tau) \leq M\}$ is compact for any $M \in \mathbb{R}$.

The existence of a minimizer is then a question of showing that for at least one M, the lower level set is not empty. Which can be shown by noticing that $H_T(\mathbf{0}, \tau)$ is bounded above with high probability uniformly in $\tau \in \mathcal{T}$ since

$$H_T(\mathbf{0}, \tau) = \tfrac{1}{T}\sum_{t=1}^{T}V_t(\tau - I(V_t < 0)) \leq \tfrac{1}{T}\sum_{t=1}^{T}|V_t| := \widehat{\nu}_t = \nu + o_P(1)$$

Therefore we can conclude that there is a compact $K = \{\delta \in \Theta : \|\delta\| < r\}$ such that a minimizer of $H_T(\boldsymbol{\delta}, \tau)$ denoted by $\widehat{\boldsymbol{\delta}}_T(\tau) = \widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}_0(\tau)$ exists with probability approaching one as $T \to \infty$ $\square$

Lemma 10. $\widehat{\boldsymbol{\theta}}_T(\tau) \xrightarrow{p} \boldsymbol{\theta}_0(\tau)$ *uniformly in* $\tau \in \mathcal{T}$ *where the parameter space is compact* $K \subset \mathbb{R}^k$

Proof. Consider the following objective functions $J_T(\boldsymbol{\theta}, \tau) := \frac{1}{T}\sum_{t=1}^T \rho_\tau(Y_t - X_t'\boldsymbol{\theta})$, $J_T^0(\boldsymbol{\theta}, \tau) := \mathbb{E}J_T(\boldsymbol{\theta}, \tau)$ and $J_0(\boldsymbol{\theta}, \tau) := \lim_{T\to\infty} J_T^0(\boldsymbol{\theta}, \tau)$ whenever the limit exist, otherwise ∞. Also for fixed τ let $\widehat{\boldsymbol{\theta}}_T(\tau)$, $\boldsymbol{\theta}_T^0(\tau)$ and $\boldsymbol{\theta}_0(\tau)$ denote its respective minimizers. We prove the result by showing that $\widehat{\boldsymbol{\theta}}_T(\tau) - \boldsymbol{\theta}_T^0(\tau) = o_P(1)$ and $\boldsymbol{\theta}_T^0 - \boldsymbol{\theta}_0 = o(1)$ uniformly in $\tau \in \mathcal{T}$. The result then follows by the triangle inequality.

The convergence in probability results from a standard consistency argument of M-estimator (Newey and McFaden (1994). First the uniform convergence is ensured by the fact that (i) $\Theta = \mathbb{R}^{n-1} \times (0,1)$ is an open convex set, (ii) J_T converges pointwise to J_T^0 by the Law of Large Number and (iii) J_T is a convex function (which combined with (i) results continuity). The we a uniform convergence over any compact $K \subset \Xi$, namely $\sup_{\xi\in K} |J_T(\xi) - J_T^0(\xi)| = o_P(1)$.

Also, as consequence of the uniform convergence J_T^0 is a continuous convex function with a unique minimize under Assumption 4 we can write for every $\epsilon > 0$, there is a $\delta > 0$ such that if $\|\widehat{\boldsymbol{\theta}}_T(\tau) - \boldsymbol{\theta}_T^0(\tau)\| > \delta$, then $|J_T^0(\widehat{\boldsymbol{\theta}}_T(\tau), \tau) - J_T^0(\boldsymbol{\theta}_T^0(\tau), \tau)| > \epsilon$, which can be majorized by

$$|J_T^0(\widehat{\boldsymbol{\theta}}_T(\tau), \tau) - J_T^0(\boldsymbol{\theta}_T^0(\tau), \tau)| \leq |J_T^0(\boldsymbol{\theta}_T^0(\tau), \tau) - J_T(\boldsymbol{\theta}_T^0(\tau), \tau)| + |J_T(\widehat{\boldsymbol{\theta}}_T(\tau), \tau) - J_T^0(\widehat{\boldsymbol{\theta}}_T(\tau), \tau)|$$

For the second part, a mean value expansion of $J_T^0(\theta_T^0)$ around θ_0 yields

$$\theta_0 - \theta_T^0 = \left[\frac{1}{T}\sum_{t=1}^T \mathbb{E}(X_t X_t' f(X_t'\widetilde{\theta}))\right]^{-1} \frac{1}{T}\sum_{t=1}^T \mathbb{E}\left[X_t \psi_\tau(V_{0,t})\right],$$

where $\widetilde{\theta} \in [\theta_0, \theta_T^0]$ and $V_{0,t} \equiv Y_t - X_t'\theta_0$. Now consider the following linear transformation $\boldsymbol{C}$, if there are no deterministic trend then $\boldsymbol{C} = \boldsymbol{I}_n$ otherwise make the last unit the one with trend and set

$$\boldsymbol{C} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & -\frac{\mu_2}{\mu_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\frac{\mu_{n-1}}{\mu_n} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

so we can then rewrite the mean value expansion above as

$$\eta_0 - \eta_T^0 \equiv C^{-1}(\theta_0 - \theta_T^0) = \left[\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\widetilde{\boldsymbol{X}}_t \widetilde{\boldsymbol{X}}_t' f(Z_t' \widetilde{\eta})) \right]^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\widetilde{\boldsymbol{X}}_t \psi_\tau(V_{0,t}) \right],$$

where $\widetilde{\boldsymbol{X}}_t = \boldsymbol{C}\boldsymbol{X}_t$ and $\widetilde{\eta} \in [\eta_0, \eta_T^0]$.

This transform is useful since now all transformed regressors are trend-free expect for the last one hence they are no longer asymptotic multicolinear. However the converge rate is not the same for the last one so let $\boldsymbol{\Lambda}_T = diag(1, \ldots, 1, \sqrt{T})$ be a diagonal matrix in the presence of trend regressors or identity otherwise,then

$$\boldsymbol{\Lambda}_T(\eta_0 - \eta_T^0) = \left[\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\boldsymbol{W}_{t,T} \boldsymbol{W}_{t,T}' f(Z_t' \widetilde{\eta})) \right]^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\boldsymbol{W}_{t,T} \psi_\tau(V_{0,t}) \right],$$

where $\boldsymbol{W}_{t,T} = \Lambda_T^{-1} \widetilde{\boldsymbol{X}}_t$.

Now the term in brackets converges to a positive definite matrix since $f(\cdot)$ is bounded away from 0 and infinity and the second term converges to zero, hence $\boldsymbol{\Lambda}_T(\eta_0 - \eta_T^0) = o(1)$ which trivially implies that $(\eta_0 - \eta_T^0) = o(1)$. Finally, the continuous Mapping Theorem we have that $\theta_0 - \theta_T^0 = o(1)$, which concludes the proof. □

In other word, we can consistently estimate all the conditional quantiles of $Y^{(0)}|\boldsymbol{X}_t$ since $\boldsymbol{X}_t'\widehat{\boldsymbol{\theta}}(\tau)$ becomes a quantile unbiased estimator of $\tau-$conditional quantile of $Y_t^{(0)}$ in sense of theorem below

Theorem 6. *Under assumptions 1 to 5 as $T_0 \to \infty$*

$$Q_\tau(Y_t^{(0)} - \boldsymbol{X}_t'\widehat{\boldsymbol{\theta}}(\tau)) \longrightarrow Q_\tau(Y_t^{(0)} - \boldsymbol{X}_t'\boldsymbol{\theta}_0(\tau)) = 0, \quad t > T_0$$

to all continuity points of Q_τ

Proof. From Theorem 1 we have $\boldsymbol{X}_t'\widehat{\boldsymbol{\theta}}(\tau)) \xrightarrow{p} \boldsymbol{X}_t'\boldsymbol{\theta}_0(\tau))$, which implies that $Y_t^{(0)} - \boldsymbol{X}_t'\widehat{\boldsymbol{\theta}}(\tau) \xrightarrow{d} Y_t^{(0)} - \boldsymbol{X}_t'\boldsymbol{\theta}_0(\tau)$, which, in turn, implies the result to all continuity points of the cdf of $Y^0|\boldsymbol{X}_t$. □

Lemma 11. *The check function $\rho(z, \tau) :- z(\tau - I(z < 0))$ is convex in $(z, \tau) \in \mathbb{R} \times (0, 1)$*

Proof. Take $\boldsymbol{w}_1 = (z_1, \tau_1)$, $\boldsymbol{w}_2 = (z_2, \tau_2)$ and $\alpha \in [0,1]$ then

$$\begin{aligned}
\rho(\alpha \boldsymbol{w}_1 + (1-\alpha)\boldsymbol{w}_2) &= (\alpha z_1 + (1-\alpha) z_2)(\alpha \tau_1 + (1-\alpha)\tau_2 - I(\alpha z_1 + (1-\alpha) z_2 < 0)) \\
&= \alpha z_1 (\tau - I(\alpha z_1 + (1-\alpha) z_2 < 0)) + (1-\alpha) z_2 (\tau - I(\alpha z_1 + (1-\alpha) z_2 < 0)) \\
&\leq \alpha z_1 (\tau - I(z_1 < 0)) + (1-\alpha) z_2 (\tau - I(z_2 < 0)) \\
&\equiv \alpha \rho_\tau(z_1) + (1-\alpha)\rho_\tau(z_2)
\end{aligned}$$

□

Lemma 12. *The function* $q(\boldsymbol{\theta}, \tau) := \rho(y - \boldsymbol{x}'\boldsymbol{\theta}, \tau)$ *is convex in* $(\boldsymbol{\theta}, \tau) \in \mathbb{R}^n \times (0,1)$ *where* $\boldsymbol{x} \in \mathbb{R}^n$ *and* $y \in \mathbb{R}$.

Proof. Take $\boldsymbol{\gamma}_1 = (\boldsymbol{\theta}_1, \tau_1), \boldsymbol{\gamma}_2 = (\boldsymbol{\theta}_2, \tau_2)$ and $\alpha \in [0,1]$ then

$$\begin{aligned}
q(\alpha \boldsymbol{\gamma}_1 + (1-\alpha)\boldsymbol{\gamma}_2) &= \rho(y - \boldsymbol{x}'(\alpha \boldsymbol{\theta}_1 + (1-\alpha)\boldsymbol{\theta}_2), \alpha \tau_1 + (1-\alpha)\tau_2) \\
&= \rho(\alpha(y - \boldsymbol{x}'\boldsymbol{\theta}_1) + (1-\alpha)(y - \boldsymbol{x}'\boldsymbol{\theta}_2), \alpha \tau_1 + (1-\alpha)\tau_2) \\
&\leq \alpha \rho(y - \boldsymbol{x}'\boldsymbol{\theta}_1, \tau_1) + (1-\alpha)\rho(y - \boldsymbol{x}'\boldsymbol{\theta}_2, \tau_2) \\
&= \alpha q(\boldsymbol{\gamma}_1) + (1-\alpha) q(\boldsymbol{\gamma}_2),
\end{aligned}$$

where the inequality follows from Lemma 5 by setting $z_1 = y - \boldsymbol{x}'\boldsymbol{\theta}_1$ and $z_2 = y - \boldsymbol{x}'\boldsymbol{\theta}_2$. □

Proof of Theorem 5

By Assumption **??** 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

$$\widehat{\tau}_T - \tau = \tfrac{1}{T_2} \sum_{t=T0}^{T} (1\{v_t(\tau) \leq 0\} - \tau) - \tfrac{1}{T_2} \sum_{t=T0}^{T} J_t(\tau) Q_t(\tau) + R(\tau, \widehat{\theta}) \tag{8}$$

, 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 (8) by $\sqrt{T}$ and , then we are left with

$$\sqrt{\frac{T}{T_2}}\left(\frac{1}{\sqrt{T_2}}\sum_{t=T_0}^{T} 1\{v_t(\tau)\leq 0\}-\tau\right)-\left(\frac{1}{T_2}\sum_{t=T0}^{T} J_t(\tau)\nabla g(x,\tilde{\theta})\right)\sqrt{T}\left(\widehat{\theta}(\tau)-\theta_0(\tau)\right)+\sqrt{T}R(\tau,\widehat{\theta}) \tag{9}$$

Note that the term in between is $o_p(1)$ for all non constant regressors 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)

$$\sqrt{\frac{T}{T_1}}\left(\frac{1}{T_2}\sum_{t=T_0}^{T} J_t(\tau)\right)\sqrt{T_1}\left(\widehat{\theta}_c(\tau)-\theta_{c,0}(\tau)\right)=\sqrt{\frac{T}{T_1}}\left(\frac{1}{T_2}\sum_{t=T_0}^{T} J_t(\tau)\right)D(\tau)^{-1}\frac{1}{T_1}\sum_{t=1}^{T_1}\tau-1\{v_t\leq 0\}+o_p(1)$$

, where $D(\tau) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} J_t(\tau)$.

Hence we are left with

$$\sqrt{T}(\widehat{\tau}_T-\tau)=\sqrt{\frac{T}{T_2}}\left(\frac{1}{\sqrt{T_2}}\sum_{t=T_0}^{T} 1\{v_t(\tau)\leq 0\}-\tau\right)-\sqrt{\frac{T}{T_1}}\left(\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1} 1\{v_t(\tau)\leq 0\}-\tau\right)+o_p(1) \tag{10}$$

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 1

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

$$(\Gamma_0)_{ij}=\mathbb{E}(1\{\Delta_t(\tau_i)\leq 0\}1\{\Delta_t(\tau_j)\leq 0\})-\tau_i\tau_j=\mathbb{P}(\Delta_t(\tau_i)\leq 0\cap\Delta_t(\tau_j)\leq 0)-\tau_i\tau_j=\min(\tau_i,\tau_j)-\tau_i\tau_j$$

We can now take stack k equations (10), one for each $\tau = \tau_1,\ldots,\tau_k$ and premultiply by any $\boldsymbol{a}_k \neq \boldsymbol{0} \in \mathbb{R}^k$:

$$\sqrt{T}\boldsymbol{a}'_k(\widehat{\boldsymbol{\tau}}_T-\boldsymbol{\tau})=\sqrt{\frac{T}{T_2}}\left(\frac{1}{\sqrt{T_2}}\sum_{t=T_0}^{T}\boldsymbol{a}'_k\boldsymbol{w}_t\right)-\sqrt{\frac{T}{T_1}}\left(\frac{1}{\sqrt{T_1}}\sum_{t=1}^{T_1}\boldsymbol{a}'_k\boldsymbol{w}_t\right)+o_p(1)$$

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.

References

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

A. Abadie, J.D. Angrist, and G. Imbens. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70:91–117, 2002.

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

A. Abadie, A. Diamond, and J. Hainmueller. Politics and the synthetic control method. *American Journal of Political Science*, 2014. In press.

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

J. Angrist, G. Imbens, and D. Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444–472, 1996.

J.D. Angrist, Ó. Jordá, and G.M. Kuersteiner. Semiparametric estimates of monetary policy effects: String theory revisited. Working Paper 2013-24, Federal Reserve Bank of San Francisco, 2013.

Joshua Angrist, Victor Chernozhukov, and IvÃ¡n FernÃ¡ndez-Val. Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica*, 74(2):539–563, 2006. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/3598810.

A. Belloni, V. Chernozhukov, and C. Hansen. Inference on treatment effects after selection amongst high-dimensional controls. *Review of Economic Studies*, 81:608–650, 2014.

A. Belloni, V. Chernozhukov, I. FernÃ¡ndez-Val, and C. Hansen. Program evaluation with high-dimensional data. *Econometrica*, 2016. In press.

C.V. Carvalho, R. Masini, and M.C. Medeiros. Arco: An artificial counterfactual approach for high-dimensional data. Working paper, Pontifical Catholic University of Rio de Janeiro, 2016.

V. Chernozhukov and C. Hansen. An IV model of quantile treatment effects. *Econometrica*, 73:245–261, 2005.

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

V. Chernozhukov and C. Hansen. Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 141:379–398, 2008.

V. Chernozhukov, I. Fernandez-Val, and B. Melly. Inference on counterfactual distributions. *Econometrica*, 2014. Forthcoming.

S. Dees, F. Di Mauro, M.H. Pesaran, and L.V. Smith. Exploring the international linkages of the Euro area: A Gobal VAR analysis. *Journal of Applied Econometrics*, 22:1–38, 2007.

E. Dubois, J. Héricourt, and V. Mignon. What if the euro had never been launched? a counterfactual analysis of the macroeconomic impact of euro membership. *Economics Bulletin*, 29:2252–2266, 2009.

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

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

C. Hsiao, H. Steve Ching, and S. Ki Wan. 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.

M.H. Pesaran and R.P. Smith. Counterfactual analysis in macroeconometrics: An empirical investigation into the effects of quantitative easing. Discussion Paper 6618, IZA, 2012.

M.H. Pesaran, T.Schuermann, and S.M. Weiner. Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business and Economic Statistics*, 22:129–162, 2004.

M.H. Pesaran, L.V. Smith, and R.P. Smith. 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.

L. Zhang, Z. Du, C. Hsiao, and H. Yin. The macroeconomic effects of the Canada-US free trade agreement on Canada: A counterfactual analysis. *World Economy*, 2014. In Press.