Multivariate quantile impulse response functions

Journal:	*Journal of Time Series Analysis*
Manuscript ID	JTSA-4600.R1
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Montes-Rojas, Gabriel; Universidad De Buenos Aires-Facultad de Ciencias Economicas,
Keywords:	multivariate quantiles, multivariate time-series, vector autoregression, impulse response functions
1. MSC codes :	
2. JEL codes :	C13, C14, C42
Abstract:	A reduced form multivariate quantile autoregressive model is developed to study heterogeneity in the effects of macroeconomic shocks. This framework is used for forecasting and for constructing quantile impulse response functions that explore dynamic heterogeneity in the response of endogenous variables to different shocks. The methodology allows evaluating different quantile paths, defined as the dynamic effects for a fix collection of quantile indexes. The model is applied to study monetary shocks in a three-variable macroeconomic model (output gap, inflation, Fed Funds rate) for the U.S. for the period 1980q1-2010q1.

Journal of Time Series Analysis JTSA-4600 Multivariate quantile impulse response functions*

Gabriel Montes-Rojas

CONICET - IIEP - Universidad de Buenos Aires

and

Universidad Nacional de La Plata

August 21, 2018

Abstract

A reduced form multivariate quantile autoregressive model is developed to study heterogeneity in the effects of macroeconomic shocks. This framework is used for forecasting and for constructing quantile impulse response functions that explore dynamic heterogeneity in the response of endogenous variables to different shocks. The methodology allows evaluating different quantile paths, defined as the dynamic effects for a fix collection of quantile indexes. The model is applied to study monetary shocks in a three-variable macroeconomic model (output gap, inflation, Fed Funds rate) for the U.S. for the period 1980q1-2010q1.

JEL classifications: C13, C14, C42

Keywords: multivariate quantiles, multivariate time-series, vector autoregression, impulse response functions.

*Corresponding author: CONICET and Instituto Interdisciplinario de Economía Política de Buenos Aires, Facultad de Ciencias Económicas, Universidad de Buenos Aires, Av. Córdoba 2122 2do piso, C1120AAQ, Ciudad Autónoma de Buenos Aires, Argentina, email: gabriel.montes@fce.edu.ar. The author is grateful to the co-editor, Dr. David Stoffer, and to two anonymous referees for their constructive comments and suggestions. STATA codes to apply these procedures are available at http://gabrielmontes.com.ar/codes

1 Introduction

An important way to summarize the dynamics of macroeconomic data is to make use of a vector autoregressive (VAR) model. The VAR approach provides statistical tools for data description, forecasting, and structural inference to study rich dynamics in multivariate time-series models.

Nevertheless, the use of a constant-coefficient model as representative of time-series models may not be adequate, as these models ignore the effects that a succession of small and varied shocks may have on the structure of dynamic economic models, particularly for highly aggregated data series. Moreover, these models cannot appropriately account for the presence of asymmetric and heterogeneous dynamic responses. Of particular interest is the asymmetric business cycle dynamics of economic variables, as the occurrence of asymmetries may call into question the usefulness of models with time invariant structures as means of modeling such series.

Quantile regression (QR) is a statistical method for estimating models of conditional quantile functions. This method offers a systematic strategy for examining how covariates influence the location, scale, and shape of the entire response distribution, thereby exposing a variety of heterogeneity in response dynamics. Koenker and Xiao (2006) examine the implications of QR models in time-series. Galvao, Montes-Rojas, and Park (2013) interpret the QR time-series framework as modeling the business cycle, where high (low) conditional realizations of a distributed lag model correspond to high (low) quantiles.

Montes-Rojas (2017) develops a reduced form vector directional quantile estimator based on the multivariate directional quantiles framework of Hallin, Paindaveine, and Šiman (2010). We apply this estimator to a VAR time-series multivariate model, the vector autoregressive quantile (VARQ) model, which generalizes the quantile autoregressive framework proposed by Koenker and Xiao (2006) and Galvao, Montes-Rojas, and Park (2013) to the multivariate case. The proposed model is a solution to a collection of directional quantile models for a fixed orthonormal basis, in which each component represents a directional quantile that corresponds to a particular endogenous variable. The directional quantile theory shows that once a direction is fixed, quantile models are well defined. The model thus delivers a map from the space of the σ-field generated by the information available at a particular time and a unit ball whose dimension is given by the number of endogenous variables to the space of endogenous variables. The VARQ estimator explores heterogeneity by estimating conditional models of each endogenous variable conditional on all other contemporaneous endogenous variables and the set of information available at the time. These conditional models are used to construct a simultaneous system of directional QR models, whose solution is a reduced-form multivariate quantile model.

The VARQ estimator allows us to forecast the future performance of the multivariate time-series, which depends on multivariate quantile indexes. This corresponds to the forecasting of the multivariate density of endogenous variables conditional on the available information. From the forecasting procedure we define an impulse response function (IRF) model that computes the effect of a given perturbation in one or more variables on the entire system, the quantile IRF (QIRF). This procedure explores potential dynamic heterogeneity not covered by the mean-based IRF analysis using mean-based VAR. In particular, we can study the realization of particular sequences of events by analyzing the multivariate quantile indexes, defined as quantile paths.

The proposed framework is different from recent attempts to construct heterogeneous IRFs. Chavleishvili and Manganelli (2016) accommodates a bivariate system of dynamic conditional quantiles of random variables where one of the random variables evolves exogenously to the system. They provide the statistical framework to define structural quantile shocks and the associated quantile impulse response functions using a multivariate generalized autoregressive conditional heteroskedasticity framework.

We estimate the VARQ and perform a QIRF analysis using a three-variable macroeconomic model for the U.S., output gap, inflation and Fed Funds rate, for the period 1980q1-2010q1. We then evaluate the effect of a standard deviation shock in the interest rate, i.e., a monetary shock using the Cholesky decomposition of Christiano, Eichenbaum, and Evans (1996), and explore dynamic heterogeneity applying the QIRFs. This new analysis reveals important asymmetries and heterogeneity in the response to monetary shocks in terms of different quantile paths of high or low conditional output and inflation.

The paper is organized as follows. Section 2 presents the theory of directional quantiles and the definition of the VARQ model. Section 3 develops forecasting procedures using VARQ, and Section 4 develops QIRF. Section 5 shows the application of the VARQ and QIRF models to empirical series. Section 6 concludes.

2 VARQ model

For a given cumulative distribution function F_Y of a univariate random variable Y, the univariate quantile function is well defined. In particular, the τ-quantile for $\tau \in (0,1)$ is defined as $Q_Y(\tau) = \inf\{y \in \mathcal{Y} : \tau \leq F_Y(y)\}$, and if F_Y is continuous, then $Q_Y(\tau) = F_Y^{-1}(\tau)$. In the multivariate case, however, there is no unique definition of a multivariate quantile function.

There is a growing literature on the estimation of QR models for multivariate random variables. Hallin, Paindaveine, and Šiman (2010) and Paindaveine and Šiman (2011, 2012) build on the definition of directional quantiles, whereby quantiles are equipped with a directional vector. Distributional features are thus explored by considering different directional models; see also Fraiman and Pateiro-López (2012) for related work. Wei (2008) develops a bivariate quantile model, following the marginal-conditional structure of Ma and Koenker (2006). White, Kim, and Manganelli (2015) develop an autoregressive model of the quantiles themselves, extending the CAViaR model of Engle and Manganelli (2004) to the multivariate case. In related work, Han, Linton, Oka, and Whang (2016) study quantile dependence among time-series models. Carlier, Chernozhukov, and Galichon (2016) propose a vector QR (linear) model that produces a monotone map, the gradient of a convex function. In a more general setup, Chernozhukov, Galichon, Hallin, and Henry (2015) develop a concept of multivariate quantile based on transportation maps between a distribution of interest with a domain in multivariate real numbers and a unit ball of the same dimension. Finally, another approach is to use copula-based quantile models, as any multivariate distribution can be decomposed into its marginals and a dependence function or copula; see, e.g., Bernard and Czado (2015); however, such an approach requires imposing distributional assumptions.

Consider an m-dimensional process $\boldsymbol{Y}_t = (Y_{1t}, \ldots, Y_{mt})^\top$ and assume that for all $t \in \{0, 1, \ldots\}$, $\boldsymbol{Y}_t \in \mathcal{Y} \subseteq \mathbb{R}^m$. Further, consider a $k \times 1$ vector of covariates $\boldsymbol{X}_t \in \mathcal{X} \subseteq \mathbb{R}^k$. Of particular interest is the case of the covariates generated by the σ-field given by $\{\boldsymbol{Y}_s : s < t\}$ and all other information available at time t. One then deals with a vector autoregressive quantile (VARQ) model. For an autoregressive model of order p, $\boldsymbol{X}_{t-1} = (\boldsymbol{Y}_{t-1}^\top, \ldots, \boldsymbol{Y}_{t-p}^\top)^\top$ and $k = mp$. We can then index VARQ models according to the lag order, i.e., VARQ(p).

Let the vector $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)$ be an index of the $\mathbb{R}^m$ space, which is an element of the open unit ball in $\mathbb{R}^m$ (deprived of the origin) $\mathcal{T}^m = \{\boldsymbol{z} \in \mathbb{R}^m : 0 < \|\boldsymbol{z}\| < 1\}$, where $\|\cdot\|$ denotes the Euclidean norm. Montes-Rojas (2017) considers the reduced form vector directional quantile (VDQ) model given by

$$Q_{\boldsymbol{Y}_t}(\boldsymbol{\tau}|\boldsymbol{x}_{t-1} = \boldsymbol{X}_{t-1}) = \boldsymbol{B}(\boldsymbol{\tau})\boldsymbol{x}_{t-1} + \boldsymbol{A}(\boldsymbol{\tau}), \tag{1}$$

where Q is an $m \times 1$ vector, which corresponds to the multivariate quantiles of the m random variables, $\boldsymbol{B}(\boldsymbol{\tau}) = (\boldsymbol{B}_1(\boldsymbol{\tau}), \ldots, \boldsymbol{B}_m(\boldsymbol{\tau}))^\top$ is an $m \times k$ matrix of coefficients with $\boldsymbol{B}_{j\cdot}(\boldsymbol{\tau})$ for each $j \in \{1, \dots, m\}$, $k \times 1$ vector of coefficients of the jth element in $\boldsymbol{Y}$, and $\boldsymbol{A}(\boldsymbol{\tau})$ is an $m \times 1$ vector of coefficients. Moreover, let $\boldsymbol{B}_{\cdot h}(\boldsymbol{\tau}) = (\boldsymbol{B}_{1h}(\boldsymbol{\tau}), \dots, \boldsymbol{B}_{mh}(\boldsymbol{\tau}))$ the h-lag coefficients for all endogenous variables models, for $h = 1, \dots, p$. Thus, Q is a map $\mathcal{X} \times \mathcal{T}^m \mapsto \mathcal{Y}$. The VDQ applied to an autoregressive context is thus our proposed model for VARQ.

For the time-series context we can define the lag polynomials $(\boldsymbol{B}(\boldsymbol{\tau}, L))$ where L is the lag operator, such that

$$\boldsymbol{B}(\boldsymbol{\tau})\boldsymbol{X}_t = \boldsymbol{B}(\boldsymbol{\tau}, L)\boldsymbol{Y}_t = \sum_{k=1}^{p} \boldsymbol{B}_{\cdot k}(\boldsymbol{\tau}) L^k \boldsymbol{Y}_t$$

and

$$Q_{\boldsymbol{Y}_t}(\boldsymbol{\tau}|\boldsymbol{x}_{t-1}) = \boldsymbol{B}(\boldsymbol{\tau}, L)\boldsymbol{y}_t + \boldsymbol{A}(\boldsymbol{\tau}), \tag{2}$$

where $\boldsymbol{y}_t$ denotes the values of $\boldsymbol{Y}_t$ to be used in the equation.

In order to construct the VARQ model, define $Q_{\boldsymbol{Y}_t}(\boldsymbol{\tau}|\boldsymbol{x}_{t-1}) := \{q_1(\boldsymbol{\tau}|\boldsymbol{x}_{t-1}), \dots, q_m(\boldsymbol{\tau}|\boldsymbol{x}_{t-1})\}^\top$ from the following system of equations,

$$\begin{cases} q_1(\boldsymbol{\tau}|\boldsymbol{x}_{t-1}) & := \boldsymbol{c}_1(\tau_1)^\top q_{-1}(\boldsymbol{\tau}|\boldsymbol{x}_{t-1}) + \boldsymbol{b}_1(\tau_1)^\top \boldsymbol{x}_{t-1} + a_1(\tau_1) \\ \quad \vdots & := \vdots \\ q_m(\boldsymbol{\tau}|\boldsymbol{x}_{t-1}) & := \boldsymbol{c}_m(\tau_m)^\top q_{-m}(\boldsymbol{\tau}|\boldsymbol{x}_{t-1}) + \boldsymbol{b}_m(\tau_m)^\top \boldsymbol{x}_{t-1} + a_m(\tau_m), \end{cases} \tag{3}$$

where $\{\boldsymbol{c}_j(\tau_j)\}_{j=1}^m$ and $\{\boldsymbol{b}_j(\tau_j)\}_{j=1}^m$ are vectors of dimensions $(m-1) \times 1$ and $k \times 1$, respectively, and $\{a_j(\tau_j)\}_{j=1}^m$ are scalars. $\{q_j(\boldsymbol{\tau}|\boldsymbol{x}_{t-1})\}_{j=1}^m$ corresponds to individual time-series QR models of each j component, i.e., Y_{jt}, on all others $-j$ components, i.e., $\boldsymbol{Y}_{-jt}$, and the lags, i.e., $\boldsymbol{X}_t$, where all the components are simultaneously evaluated at $Q(\boldsymbol{\tau}|\boldsymbol{x}_{t-1})$ (for further reference, we use the $-j$ notation to denote a vector or matrix that excludes the jth column or row). In turn these are obtained using Hallin, Paindaveine, and Šiman (2010) directional quantiles for a fix orthonormal basis and the VARQ estimator is a fixed point solution to a system of equations.

Consider the following matrices based on the coefficients of eq. (3): $\boldsymbol{C}(\boldsymbol{\tau}) := \{\boldsymbol{C}_1(\tau_1), \ldots, \boldsymbol{C}_m(\tau_m)\}^\top$ is an $m \times m$ matrix in which the $\{\boldsymbol{C}_j(\tau_j)\}_{j=1}^m$ $m \times 1$-dimensional vectors contain all the elements of the $m-1$ vector of coefficients $\{\boldsymbol{c}_j(\tau_j)\}_{j=1}^m$ augmented with a 0 in the corresponding jth component, $\boldsymbol{b}(\boldsymbol{\tau}) = \{\boldsymbol{b}_1(\tau_1), \ldots, \boldsymbol{b}_m(\tau_m)\}^\top$ is an $m \times k$ matrix, and $\boldsymbol{a}(\boldsymbol{\tau}) = \{a_1(\tau_1), \ldots, a_m(\tau_m)\}^\top$ is an $m \times 1$ vector. Then, the VARQ model is defined as

$$Q_{\boldsymbol{Y}_t}(\boldsymbol{\tau}|\boldsymbol{x}_{t-1}) = \{\boldsymbol{I}_m - \boldsymbol{C}(\boldsymbol{\tau})\}^{-1} \{\boldsymbol{b}(\boldsymbol{\tau})\boldsymbol{x}_{t-1} + \boldsymbol{a}(\boldsymbol{\tau})\} := \boldsymbol{B}(\boldsymbol{\tau})\boldsymbol{x}_{t-1} + \boldsymbol{A}(\boldsymbol{\tau}), \quad (4)$$

where $\boldsymbol{I}_m$ is the m-dimensional identity matrix, $\boldsymbol{B}(\boldsymbol{\tau}) := \{\boldsymbol{I}_m - \boldsymbol{C}(\boldsymbol{\tau})\}^{-1} \boldsymbol{b}(\boldsymbol{\tau})$ and $\boldsymbol{A}(\boldsymbol{\tau}) := \{\boldsymbol{I}_m - \boldsymbol{C}(\boldsymbol{\tau})\}^{-1} \boldsymbol{a}(\boldsymbol{\tau})$. Note that for a fixed $\boldsymbol{\tau}$ the number of parameters to be estimated is that of a structural mean-based VAR model.

The VARQ model describes the multivariate random variables conditional on the past by way of modeling the simultaneous responses. For each of the j equations, the τ_j-quantile model describes the conditional performance of the jth endogenous variable conditional on the values of the others and the available past information. $\boldsymbol{\tau}$ thus represents the individual contribution of every endogenous variable in the system after considering the effects of all the others. Each individual equation corresponds to a quantile autoregressive distributed lag model, e.g., Galvao, Montes-Rojas, and Park (2013). The VARQ model is the simultaneous solution of all equations for the fixed collection of individual univariate quantiles indexed by $\boldsymbol{\tau}$. As noted by Montes-Rojas (2017) this corresponds to a reduced form VAR model that is a functional in $\boldsymbol{\tau}$.

Although the VARQ model is constructed for stationary processes, Koenker and Xiao (2004,2006) consider the case of unit roots in some quantiles and stationarity in others. While we do not pursue the properties of the non-stationary behavior in those cases, the unit root processes can be easily detected by looking at their dynamic behavior.

3 Forecasting using VARQ and quantile paths

The VARQ model implicitly defines a one-period ahead forecasting method for the entire distribution of $\boldsymbol{Y}_{t+1}$ given all the information available at t.

$$Q_{\boldsymbol{Y}_{t+1}}(\boldsymbol{\tau}|\boldsymbol{x}_t) = Q_{\boldsymbol{Y}_{t+1}}(\boldsymbol{\tau}|\{\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_{t-p}\}) = \boldsymbol{B}(\boldsymbol{\tau}, L)\boldsymbol{y}_{t+1} + \boldsymbol{A}(\boldsymbol{\tau}).$$

Define thus $Q_1(\boldsymbol{\tau}|\boldsymbol{x}_t) = Q_{\boldsymbol{Y}_{t+1}}(\boldsymbol{\tau}|\boldsymbol{x}_t)$ as the one-period ahead forecast given all the information available at time t.

Consider now the two-periods ahead forecast, i.e. $t+2$, at quantiles $\boldsymbol{\tau}_2$. Note that this would depend on the response at $t+1$ and the implicit quantile $\boldsymbol{\tau}_1$. In turn then this would depend on both quantiles, $(\boldsymbol{\tau}_2, \boldsymbol{\tau}_1)$. This is defined as a two-periods quantile path, where the collection of indexes correspond to a potential path of the system of endogenous variables. Then

$$Q_2\{(\boldsymbol{\tau}_2, \boldsymbol{\tau}_1)|\boldsymbol{x}_t\} := Q[\boldsymbol{\tau}_2|\{Q_1(\boldsymbol{\tau}_1|\boldsymbol{x}_t), \boldsymbol{y}_t, \ldots, \boldsymbol{y}_{t-p+1}\}].$$

Replacing for the linear quantile models we have

$$= \boldsymbol{B}(\boldsymbol{\tau}_2)\{Q_1(\boldsymbol{\tau}_1|\boldsymbol{x}_t)^\top, \boldsymbol{y}_t^\top, \ldots, \boldsymbol{y}_{t-p+1}^\top\} = \boldsymbol{B}(\boldsymbol{\tau}_2)\{[\boldsymbol{B}(\boldsymbol{\tau}_1)\boldsymbol{x}_t]^\top, \boldsymbol{y}_t^\top, \ldots, \boldsymbol{y}_{t-p+1}^\top\}^\top + \boldsymbol{A}(\boldsymbol{\tau}_2).$$

In particular for the VARQ(1) model

$$Q_2\{(\boldsymbol{\tau}_2, \boldsymbol{\tau}_1)|\boldsymbol{x}_t\} = \boldsymbol{B}(\boldsymbol{\tau}_2)\boldsymbol{B}(\boldsymbol{\tau}_1)\boldsymbol{x}_t + \boldsymbol{B}(\boldsymbol{\tau}_2)\boldsymbol{A}(\boldsymbol{\tau}_1) + \boldsymbol{A}(\boldsymbol{\tau}_2).$$

The h-periods ahead forecast can be written as a function of the forecast of the previous quantiles,

$$Q_h\{(\boldsymbol{\tau}_h, \ldots, \boldsymbol{\tau}_1)|\boldsymbol{x}_t\} = \boldsymbol{B}(\boldsymbol{\tau}_h, L)Q_k\{(\boldsymbol{\tau}_k, \ldots, \boldsymbol{\tau}_1)|\boldsymbol{x}_t\} + \boldsymbol{A}(\boldsymbol{\tau}_h),$$

where $Q_k(.|.) = y_{t-k}$ if $L^k(t+h) \leq t$ and $(\boldsymbol{\tau}_k, \ldots, \boldsymbol{\tau}_1)$, $k = 1, \ldots, h-1$ is the k-periods quantile path. Then we can write

$$Q_h\{(\boldsymbol{\tau}_h, \ldots, \boldsymbol{\tau}_1)|\boldsymbol{x}_t\} = \{\Pi_{k=1}^h \boldsymbol{B}(\boldsymbol{\tau}_k)\}\boldsymbol{x}_t + \sum_{k=1}^{h-1}\{\Pi_{j=k+1}^h \boldsymbol{B}(\boldsymbol{\tau}_j)\}\boldsymbol{A}(\boldsymbol{\tau}_k) + \boldsymbol{A}(\boldsymbol{\tau}_h).$$

This framework allows for forecasting different quantile paths.

A canonical case is fixing $\boldsymbol{\tau}_i = (0.5, \ldots, 0.5)$ for all $i = 1, \ldots, h$, which corresponds to evaluating future values on the conditional median values of the endogenous variables. In general this procedure delivers similar estimates to the mean-based VAR forecasts. In this case each realization is evaluated at the conditional median and the h-periods ahead forecast is also constructed using $h-1, \ldots, 1$ values at the median.

This procedure can be generalized for any $\boldsymbol{\tau}_i = (\tau, \ldots, \tau)$ for all $i = 1, \ldots, h$. In this case high values of τ correspond to the persistent occurrence of the τ conditional quantile in all endogenous variables. Moreover, we do not necessarily need the same τ for all endogenous variables equations. As an example in the empirical application we consider the 0.1 and 0.9 quantiles of output, while we keep the median for inflation and interest rate. As such, we are constructing a potential quantile path where output is either at the low or high end of the business cycle. See Galvao, Montes-Rojas, and Park (2013) for an interpretation of QR time-series models in terms of the business cycle.

Note, however, that if we are interested in the h-periods ahead forecast, this may not depend on the implicit quantile used for the k-step forecast, $k < h$. As such we could integrate out $\boldsymbol{\tau}_k$ by using $\boldsymbol{\tau}_k \sim IID\ U(0,1)^m$ for $k = 1, 2, ..., h-1$. Define $\bar{\boldsymbol{B}} := \mathrm{E}_{\boldsymbol{\tau}}\mathrm{B}(\boldsymbol{\tau})$ and $\bar{\boldsymbol{A}} := \mathrm{E}_{\boldsymbol{\tau}}\mathrm{A}(\boldsymbol{\tau})$. Then,

$$Q_h(\boldsymbol{\tau}|\boldsymbol{x}_t) = \boldsymbol{B}(\boldsymbol{\tau})\bar{\boldsymbol{B}}^{h-1}\boldsymbol{x}_t + \boldsymbol{B}(\boldsymbol{\tau})\left\{\sum_{k=1}^{h-1}\bar{\boldsymbol{B}}^k\bar{\boldsymbol{A}}\right\} + \boldsymbol{A}(\boldsymbol{\tau}).$$

As $h \to \infty$, the long run prediction converges to

$$\lim_{h\to\infty} Q_h(\boldsymbol{\tau}|\boldsymbol{x}_t) = \boldsymbol{B}(\boldsymbol{\tau})(\boldsymbol{I} - \bar{\boldsymbol{B}})^{-1}\bar{\boldsymbol{A}} + \boldsymbol{A}(\boldsymbol{\tau}).$$

Note that $\bar{\boldsymbol{B}}$ and $\bar{\boldsymbol{A}}$ are not necessarily equal to the mean-based reduced form VAR coefficients. Following Koenker and Xiao (2006) we could write a random-coefficients representation of model (3) for which integrating out the quantiles produces the mean-based model. However, the VARQ model is a non-linear transformation of (3) and quantile operators cannot be applied to non-linear functions.

4 Shocks and impulse response functions in the VARQ model

4.1 Impulse responses from VARQ

Ramey (2016) defines shocks as *primitive exogenous* forces that are uncorrelated with each other and they should be *economically meaningful* (pp.52-55). Shocks have the following characteristics: (1) they should be exogenous with respect to the other current and lagged endogenous variables in the model; (2) they should be uncorrelated with other exogenous shocks (...); (3) they should represent either unanticipated movements in exogenous variables or *news* about future movements in exogenous variables. The literature on measuring shocks on time-series models is based on VAR models, where a shock refers to a change in the residual of a conditional model, and in identifying exogenous changes in a structural model. See Ramey (2016); Stock and Watson (2016) for a recent literature review that resumes the state of the art.

By modeling the multivariate quantiles we do not have a structural model nor we have a system of residuals in a reduced form additive model, but rather we replicate the simultaneous movements in the endogenous variables by way of indexing them by $\boldsymbol{\tau}$. The VARQ is a reduced-form model, and then it is adequate for forecasting and impulse response analysis, once shocks have been constructed from some other mechanism, probably based on mean-based structural VAR. Then we consider a different model type shocks where we compute a counterfactual change $\boldsymbol{\delta} \in \mathcal{Y} \subseteq \mathbb{R}^m$ in $\boldsymbol{y}_t$. Our interest lies in evaluating the propagation of those shocks in terms of the multivariate distribution of the m-variate process. We then compute the impulse response function by comparing the multivariate quantiles at $\boldsymbol{x}_t^{\boldsymbol{\delta}} := (\boldsymbol{y}_t + \boldsymbol{\delta}, \boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_{t-p})$ with those at $\boldsymbol{x}_t = (\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_{t-p})$.

Define the $\boldsymbol{\tau}$-quantile impulse response function (QIRF) at $t+1$ for a shock at time t, $\boldsymbol{\delta} \in \mathcal{Y} \subseteq \mathbb{R}^m$, as

$$\mathrm{Qirf}_1(\boldsymbol{\tau}, \boldsymbol{\delta}|\mathrm{x_t}) = \mathrm{Q}_1(\boldsymbol{\tau}|\mathrm{x}_t^{\boldsymbol{\delta}}) - \mathrm{Q}_1(\boldsymbol{\tau}|\mathrm{x_t}) = \mathrm{B}_{\cdot 1}(\boldsymbol{\tau})\boldsymbol{\delta},$$

where Q_1 is the one-period ahead forecast.

Consider now the IRF two-periods ahead, i.e. $t+2$, at quantiles $\boldsymbol{\tau}_2$. Note that this would depend on the response at $t+1$ and the implicit quantile $\boldsymbol{\tau}_1$. In turn then this would depend on both quantiles, $(\boldsymbol{\tau}_2, \boldsymbol{\tau}_1)$, defined as a quantile path.

$$\begin{aligned}
\mathrm{Qirf}_{2(1)}\left\{(\boldsymbol{\tau}_2, \boldsymbol{\tau}_1), \boldsymbol{\delta}|\boldsymbol{x}_t\right\} &= Q_2\left\{(\boldsymbol{\tau}_2, \boldsymbol{\tau}_1)|\boldsymbol{x}_t^{\boldsymbol{\delta}}\right\} - Q_2\left\{(\boldsymbol{\tau}_2, \boldsymbol{\tau}_1)|\boldsymbol{x}_t\right\} \\
&= \begin{cases} (\boldsymbol{B}_{\cdot 2}(\boldsymbol{\tau}_2) + \boldsymbol{B}_{\cdot 1}(\boldsymbol{\tau}_2)\boldsymbol{B}_{\cdot 1}(\boldsymbol{\tau}_1))\boldsymbol{\delta} & p > 1 \\ \boldsymbol{B}_{\cdot 1}(\boldsymbol{\tau}_2)\boldsymbol{B}_{\cdot 1}(\boldsymbol{\tau}_1))\boldsymbol{\delta} & p = 1 \end{cases}.
\end{aligned}$$

The QIRF above is constructed for different quantile paths, where each forecast is evaluated at a given multivariate quantile index, and for a fixed quantile index used for the previous endogenous variables forecasts.

Note however that if we are interested in the two-periods ahead forecast, this may not depend on the implicit quantile used for the one-step forecast. As such we could integrate out $\boldsymbol{\tau}_1$ by using $\boldsymbol{\tau}_1 \sim U(0,1)^m$. Then define

$$\begin{aligned} \text{Qirf}_2(\boldsymbol{\tau}, \boldsymbol{\delta}|\boldsymbol{x}_t) &= Q_2(\boldsymbol{\tau}|\boldsymbol{x}_t^{\boldsymbol{\delta}}) - Q_2(\boldsymbol{\tau}|\boldsymbol{x}_t) \\ &= \begin{cases} (\boldsymbol{B}_{\cdot 2}(\boldsymbol{\tau}) + \boldsymbol{B}_{\cdot 1}(\boldsymbol{\tau})\bar{\boldsymbol{B}}_{\cdot 1})\boldsymbol{\delta} & p > 1 \\ \boldsymbol{B}_{\cdot 1}(\boldsymbol{\tau})\bar{\boldsymbol{B}}_{\cdot 1}\boldsymbol{\delta} & p = 1 \end{cases}. \end{aligned}$$

The difference between $\text{Qirf}_{2(1)}$ and Qirf_2 is that the former corresponds to a particular path of assumed realizations of the multivariate process, while the latter focuses on the distribution two-periods ahead for a forecasted value of one-period ahead.

This procedure above can be generalized for h-periods ahead IRFs, by defining

$$\text{Qirf}_{h(h-1,\dots,1)}\left\{(\boldsymbol{\tau}_h, \boldsymbol{\tau}_{h-1}, \dots, \boldsymbol{\tau}_1), \boldsymbol{\delta}|\boldsymbol{x}_t\right\} = Q_h\left\{(\boldsymbol{\tau}_h, \boldsymbol{\tau}_{h-1}, \dots, \boldsymbol{\tau}_1)|\boldsymbol{x}_t^{\boldsymbol{\delta}}\right\} - Q_h\left\{(\boldsymbol{\tau}_h, \boldsymbol{\tau}_{h-1}, \dots, \boldsymbol{\tau}_1)|\boldsymbol{x}_t\right\}, \tag{5}$$

for a given *path* of multivariate quantiles $(\boldsymbol{\tau}_h, \boldsymbol{\tau}_{h-1}, \dots, \boldsymbol{\tau}_1)$ and shock $\boldsymbol{\delta}$ at time t, and

$$\text{Qirf}_h(\boldsymbol{\tau}, \boldsymbol{\delta}|\boldsymbol{x}_t) = Q_h(\boldsymbol{\tau}|\boldsymbol{x}_t^{\boldsymbol{\delta}}) - Q_h(\boldsymbol{\tau}|\boldsymbol{x}_t), \tag{6}$$

when we integrate out the previous periods that were constructed by iterations. This is different from the mean-based VAR analysis. In this case, by using the iterated expectations property, the effect on h periods ahead is the result of the conditional expectations in the previous periods.

In the long run the QIRF for $h \to \infty$ becomes 0 for stationary models.[1]

A simple simulation exercise appears in the Appendix to illustrate the QIRF model.

[1] The Enders' book (2015, ch.5) has an intuitive explanation for the fact that IRF needs to become 0 as $h \to \infty$. For stationary models the IRF is the collection of the effects of a variable j in a variable k (for all j, k) that occurred at some point $h = 0$ on a different time path $h = 0, 1, \dots$. These effects have to converge to 0 as h gets larger because "shocks cannot have a permanent effect on a stationary series" (Enders, 2015, p.295). The reason is that "[j]ust as an autoregression has a moving average representation, a vector autoregression can be written as a vector moving average (...) [where the model is] expressed in terms of the current and past values of the (...) shocks" (Enders, 2015, p.295). This invertibility condition can only be achieved if the effects of past shocks is smaller as we consider longer horizons.

4.2 Local projections

A robust model for constructing IRFs is based on Jordà (2005) local projections method. The central idea consists in estimating local projections at each period of interest (i.e., $t+h$) rather than extrapolating into increasingly distant horizons from a given model, as it is done with VAR. The advantages of local projections are numerous: (1) they can be estimated by simple regression techniques; (2) they are more robust to misspecification; (3) joint or point-wise analytic inference is simple; and (4) they easily accommodate experimentation with highly nonlinear and exible specications that may be impractical in a multivariate context.

This framework can be easily implemented in a VARQ context by a modeling the VDQ model of $\boldsymbol{Y}_{t+h}$ at each horizon $h = 1, 2, \ldots$ given all the information available at t, that is, all the lags of the endogenous variables up to t (plus exogenous variables if any)

$$Q_h^{lp}(\boldsymbol{\tau}|\boldsymbol{x}_t) := Q_{\boldsymbol{Y}_{t+h}}(\boldsymbol{\tau}|\boldsymbol{x}_t) = \boldsymbol{B}_h(\boldsymbol{\tau})\boldsymbol{x}_t + \boldsymbol{A}_h(\boldsymbol{\tau}). \tag{7}$$

Note that in this case we require to solve a different set of coefficients for each horizon h, which in fact involves directional QR models involving regressing $\boldsymbol{Y}_{j,t+h}$ on $\boldsymbol{Y}_{-j,t+h}$ and $\boldsymbol{x}_t$, for $j = 1, \ldots, m$. Then we could construct the QIRFs as

$$\text{Qirf}_h^{lp}(\boldsymbol{\tau}, \boldsymbol{\delta}|\boldsymbol{x}_t) = Q_h^{lp}(\boldsymbol{\tau}|\boldsymbol{x}_t^{\boldsymbol{\delta}}) - Q_h^{lp}(\boldsymbol{\tau}|\boldsymbol{x}_t). \tag{8}$$

While this is an important alternative for prediction, it does not allow us to study quantile paths. That is, intermediate realizations of the random variables, i.e., for $h-1, h-2, \ldots, 1$, are implicitly evaluated at the mean-based values.

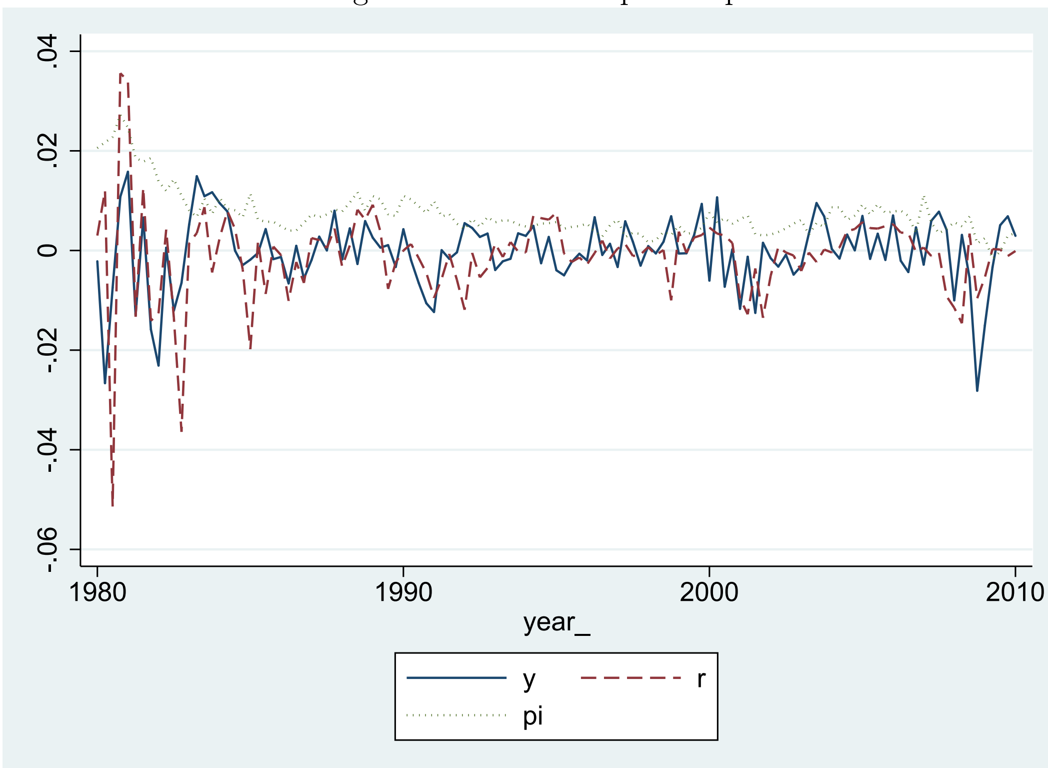
5 Heterogeneity of monetary shocks

We estimate a three-variable (output gap, inflation, Fed Funds rate) VAR(1) model using U.S. quarterly data from 1980q1 to 2010q1 (121 quarters). This simple framework corresponds to the three-variable framework of New Keynesian model rational expectations model of Cho and Moreno (2004, 2006) and Jordà (2005), among others. We do not impose a structural model but rather we consider a simple unrestricted VAR model of order 1 of the three variables. The period is chosen to cover the post-Volcker Fed rules and previous to the zero lower bound, post-financial crisis regime.

The output gap is generated by the first-difference of the Hodrick-Prescott linear filter with linear trend, using the logarithm of the Gross National Product, 1996 constant prices (source: Federal Reserve Bank of St. Louis), denoted y_t. The inflation rate is the log first-difference of the GDP deflator, seasonally adjusted (source: Federal Reserve Bank of St. Louis), denoted π_t. The Fed Funds rate is the monetary policy instrument (source: Board of Governors of the Federal Reserve System), denoted r_t, and corresponds to the first-difference of the 3-months Treasury Bill rate (end of the quarter). The reason we use the first-difference of the interest rate is that over the period of analysis it shows a negative trend and we cannot reject it has a unit root. For this case then $\boldsymbol{Y}_t = (y_t, \pi_t, r_t)$. Figure 1 plots the series considered here, and Table 1 reports summary statistics.

Figure 2 reports the effect of a unit change in r keeping (y, π) unchanged, on the coefficients $\boldsymbol{B}_{(\pi r)1}(\tau_y, \tau_\pi, \tau_r)$, denoted as *bplr* in the figure, and $\boldsymbol{B}_{(yr)1}(\tau_y, \tau_\pi, \tau_r)$denoted as *bylr* in the figure. We also include the OLS estimate given by a regression model of y_t and π_t on $(y_{t-1}, \pi_{t-1}, r_{t-1}$. We consider different specifications with $\tau_y \in \{0.05, 0.10, \ldots, 0.95\}$, $\tau_\pi \in \{0.05, 0.10, \ldots, 0.95\}$ and $\tau_r = 0.50$ for which we evaluate the effect of a unit change in r.

Figure 1: Series 1980q1-2010q1

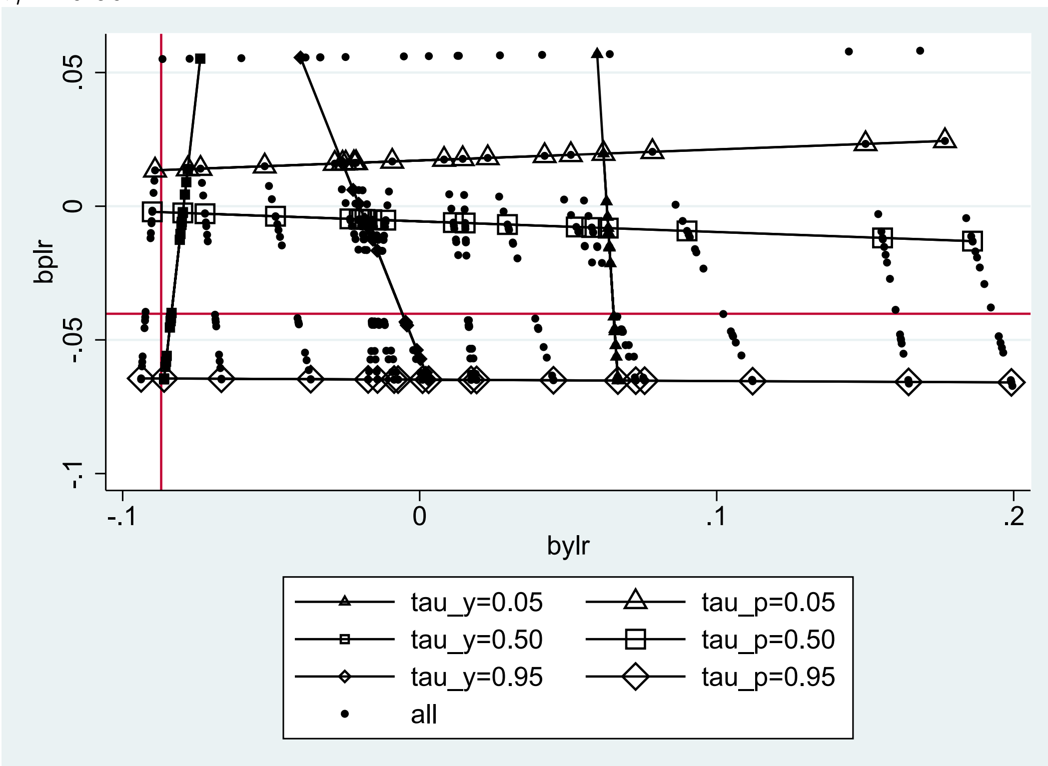


Notes: Output gap, inflation and interest rate series.

Table 1: Summary statistics for the series 1980q1-2010q1

Variable	Obs	Mean	Std. Dev.	Min	Max
y	121	-0.00034663	0.0072249	-0.0281541	0.0158011
π	121	0.0072152	0.0047343	-.0016704	0.0272542
r	121	-0.00962	0.0093464	-0.0514	0.0356
Correlations (y_t, π_t, r_t)					
Variable	y_t	π_t	r_t		
y_t	1.0000				
π_t	-0.1068	1.0000			
r_t	0.3831	0.1275	1.0000		
Correlations (y_t, π_t, r_t) mean-based VAR residuals					
Variable	y_t	π_t	r_t		
y_t	1.0000				
π_t	-0.0023	1.0000			
r_t	0.3329	0.0593	1.0000		

Figure 2: VARQ coefficients for $\tau_y \in \{0.05, 0.10, \ldots, 0.95\}$, $\tau_\pi \in \{0.05, 0.10, \ldots, 0.95\}$ and $\tau_r = 0.50$



Notes: The figure reports the heterogeneity in the effect of the quantile regression coefficient of a lagged change in the interest rate on output (bylr, horizontal axis) and inflation (bplr, vertical axis). Vertical and horizontal lines correspond to the mean-based VAR effects. Lines with small triangles symbols correspond to VARQ coefficients with $\tau_y = 0.05, \tau_\pi \in \{0.05, 0.10, \ldots, 0.95\}, \tau_r = 0.50$; small squares $\tau_y = 0.50, \tau_\pi \in \{0.05, 0.10, \ldots, 0.95\}, \tau_r = 0.50$; small diamonds $\tau_y = 0.95, \tau_\pi \in \{0.05, 0.10, \ldots, 0.95\}, \tau_r = 0.50$; large triangles $\tau_y \in \{0.05, 0.10, \ldots, 0.95\}, \tau_\pi = 0.05, \tau_r = 0.50$; large squares $\tau_y \in \{0.05, 0.10, \ldots, 0.95\}, \tau_\pi = 0.50, \tau_r = 0.50$; and large diamonds $\tau_y \in \{0.05, 0.10, \ldots, 0.95\}, \tau_\pi = 0.9, \tau_r = 0.50$.

Figure 2 unmasks interesting heterogeneity in the responses of output and inflation to changes in the interest rate. In particular, note that while OLS and median effects are small, the highest effects correspond to low τ_y and τ_π quantiles.

Next we compute impulse response functions. We follow the Cholesky identification procedure in Christiano, Eichenbaum, and Evans (1996), using the residuals from a VAR model where we assume the standard ordering: r has no contemporaneous effect on y and π; π has an effect on r but not on y; and y affects both π and r. This implies that shocks to the Fed Funds rate has no contemporaneous effect on the other economic variables. Then

Table 2: VAR system stability

Model	eigen 1	eigen 2	eigen 3
$VAR - OLS$	0.853	0.152	0.067
$VARQ(\tau_y = 0.5, \tau_\pi = 0.1, \tau_r = 0.5)$	0.669	0.131	0.131
$VARQ(\tau_y = 0.1, \tau_\pi = 0.5, \tau_r = 0.5)$	0.813	0.535	0.054
$VARQ(\tau_y = 0.5, \tau_\pi = 0.5, \tau_r = 0.5)$	0.818	0.145	0.145
$VARQ(\tau_y = 0.5, \tau_\pi = 0.9, \tau_r = 0.5)$	0.984	0.153	0.153
$VARQ(\tau_y = 0.9, \tau_\pi = 0.5, \tau_r = 0.5)$	0.820	0.285	0.058

we evaluate the effect of a shock in r, calculated as the standard deviation of this structural shock, on output gap and inflation (also standardized by the standard deviation of their corresponding structural shocks).

Figures 3 and 4 plot the QIRF of this r shock on output gap and inflation dynamics for the VAR-OLS model and for indexes $\boldsymbol{\tau} = (\tau_y, \tau_\pi = 0.50, \tau_r = 0.50)$ with $\tau_y = 0.10, 0.50, 0, 90$ for the former and $\boldsymbol{\tau} = (\tau_y = 0.50, \tau_\pi, \tau_r = 0.50)$ with $\tau_\pi = 0.10, 0.50, 0, 90$. The quantile curves thus represent the potential response of y and π if the VARQ model is evaluated at fixed $\boldsymbol{\tau}$ for all $h = 1, 2, \ldots, 100$. Note, for example for Figure 3, that the case with $\tau_y = 0.10$ corresponds to the simulation of what would be the response of output and inflation to a change in the interest rate (only) if output response were to remain at the bottom 10% conditional quantile. This indeed corresponds to an extreme event as persistent low quantiles would be associated with an unusual depression (as per the 1980-2010 sample). Moreover, the case with $\tau_y = 0.90$ correspond to a case of output response always in the upper 10% conditional quantile, an extraordinary growth as compared to the estimation sample. Moreover, for Figure 4, the analysis correspond to the case of persistent conditional high ($\tau_\pi = 0.90$) or low ($\tau_\pi = 0.10$) inflation.

Table 2 evaluates the dynamic stability of all the different specifications. All eigenvalues modulus are inside the unit circle. The system is close to unit root for $\tau_\pi = 0.90$, which will produce a non-stationary behavior in the QRIF for this case.

The QRIF analysis shows that mean-based OLS and median responses show similar dynamics for both output and inflation. For output gap both responses are very similar and negative, while they are both negative for inflation but mean effects are larger than median ones.

The $\tau_y = 0.90$ output quantile path effects for output are close to zero. This determines that a dynamic path of extraordinary growth would not be affected by changes in the interest rate, although inflationary consequences would be bounded by those of the average effect.

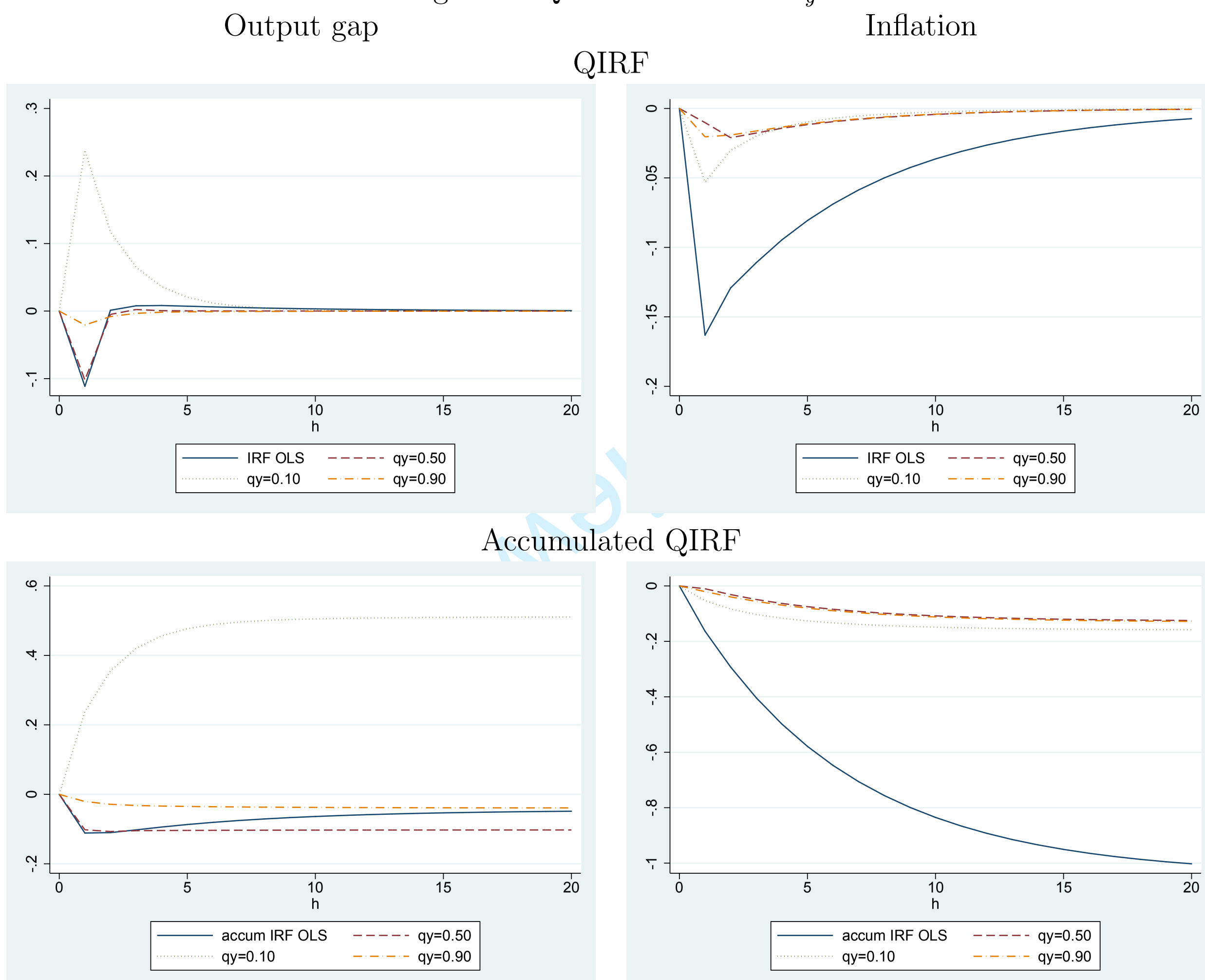
The $\tau_y = 0.10$ output quantile path, however, shows a positive and persistent effect on output. As a result its cumulative effect is the largest. For inflation, however, it shows small effects similar to other quantiles. Thus, if the economy were to remain in a state of permanent recession, as given by persistent realizations in the lower 10th conditional quantile in output gap, increasing the interest rate by 1 std.dev. would increase the output by about 0.5 std.dev. in the long-run as compared to the value if the interest rate would not be changed.

When computing the same graphs for different fixed values of τ_π we observe considerable differences between the case for high and low inflation quantile paths. In particular the case for $\tau_\pi = 0.90$ shows a non-stationary behavior with persistent increasing negative effects on inflation (the maximum modulus eigenvalue in Table 2 is 0.984). This corresponds to a unit root behavior. On the other hand, when using $\tau_\pi = 0.10$ the curves show a positive accumulated effect on inflation.

The model thus highlights potential asymmetry in the dynamic propagation of shocks. The previous analysis determined that high and low quantiles are associated with more persistence and, in general, larger effects of a given shock. This analysis should be combined with the analysis of non-linearity and structural changes in the mean and quantile dynamics. Nevertheless these simple linear models using different quantile paths can be used to evaluate extreme events even when we do not have enough observations to properly evaluate a structural change.

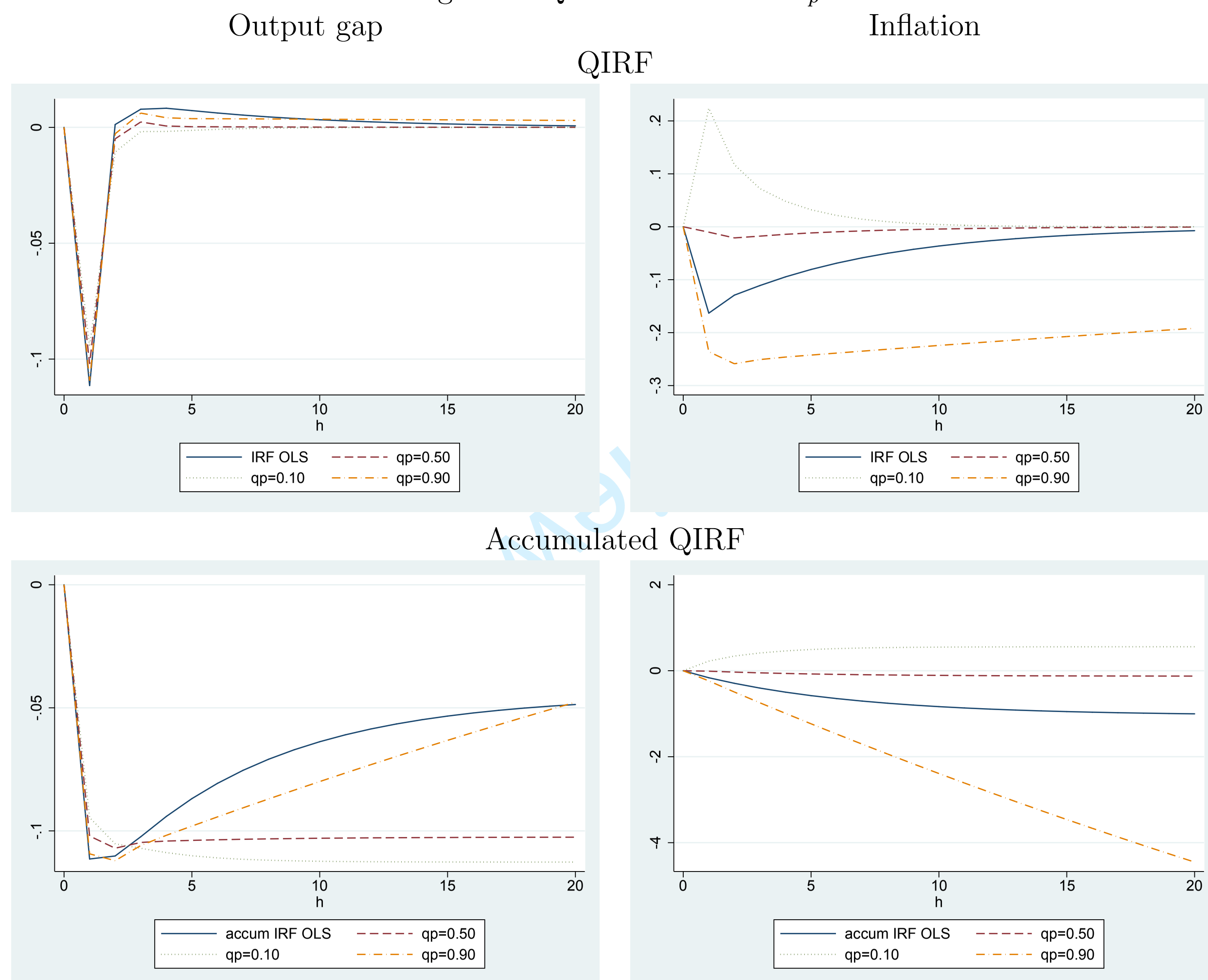
Another issue that is of interest is that while the VARQ model is monotonic on τ, the QIRF effects are not. For instance, while the effects of increasing the interest rate are negative on inflation for the mean and the median, the calculated effects for the $\tau_y = 0.10$ are positive. Note that in general QR models are used to evaluate heterogeneous effects which are themselves the result of unobserved factors. As such, different quantile paths can be the result of omitted variables. In this case, low output quantiles together with a positive shock on the interest rate could be result of attracting foreign capital, which in turn has a positive impact on output with idle productive capacity.

Figure 3: QIRF for different τ_y



Notes: QIRF on output gap and inflation of a std.dev. shock in r_t for $\tau_y \in \{0.10, 0.50, 0.90\}$, $\tau_\pi = 0.50$ and $\tau_r = 0.50$.

Figure 4: QIRF for different τ_p



Notes: QIRF on output gap and inflation of a std.dev. shock in r_t for $\tau_p \in \{0.10, 0.50, 0.90\}$, $\tau_y = 0.50$ and $\tau_r = 0.50$.

6 Conclusion

To our knowledge this is the first attempt to construct impulse response functions from multivariate semiparametric directional quantiles. Within this framework we can explore dynamic heterogeneity in the potential effect of a given shock into the future performance of series.

The present paper can be extended in several directions. First, we have addressed only linear QR models for each separate direction, and, as such, the VARQ model requires a simple matrix inversion. This model could be applied to nonlinear and nonparametric models for each direction, and the VARQ model would thus be a fixed-point solution to a nonlinear system of equations.

Second, the model should be further evaluated in terms of in-sample and out-sample dynamic forecasting. In particular, given fixed covariates for which we would like to forecast, random draws on the m-dimensional unit ball should be able to forecast the m-dimensional density. As QR provides a flexible model to construct univariate density estimations, the VARQ model could be applied to multivariate density frameworks.

Third, we can use the proposed model to evaluate potential structural breaks, as given by dynamic paths that correspond to extreme or unusual events. Quantile configurations based on a given sample can be used to forecast future events for which we do not have enough observations to analyze. That is, extreme events in the past, as given by high or low conditional quantiles provide a useful extrapolation method for future events.

Appendix

Consider a two-variable model, i.e., $m = 2$,

$$Y_{1t} = \beta_{11}Y_{1t-1} + \beta_{12}Y_{2t-1} + \alpha_1 + (1 + \delta_{11}Y_{1t-1} + \delta_{12}Y_{2t-1})\varepsilon_{1t},$$

$$Y_{2t} = \beta_{21}Y_{1t-1} + \beta_{22}Y_{2t-1} + \alpha_2 + (1 + \delta_{21}Y_{1t-1} + \delta_{22}Y_{2t-1})\varepsilon_{2t}.$$

Furthermore assume that $\varepsilon_{jt} \sim iid(0, \sigma_j), j = 1, 2$. This corresponds to a reduced form VAR model. In quantile regression models there is a standard procedure to generate heterogeneity across quantiles that is to generate a 'location-scale shift' model where the covariates (i.e., Y_{1t-1} and Y_{2t-1} in our case) affect the variance (i.e., scale) of the model. As such, heteroskedasticity would imply that the mean-based model differ from that of quantiles. A model with $\delta_{ij} = 0, \forall i, j = 1, 2$, produces no heterogeneity in the quantile effects (a model known as a 'location shift' model). If on the other hand there is a pair i, j such that $\delta_{ij} \neq 0$, there will be heterogeneity in the quantile effects. In particular we have

$$E(Y_{jt}|Y_{1t-1}, Y_{1t-1}) = \beta_{j1}Y_{1t-1} + \beta_{j2}Y_{2t-1} + \alpha_1,$$

and

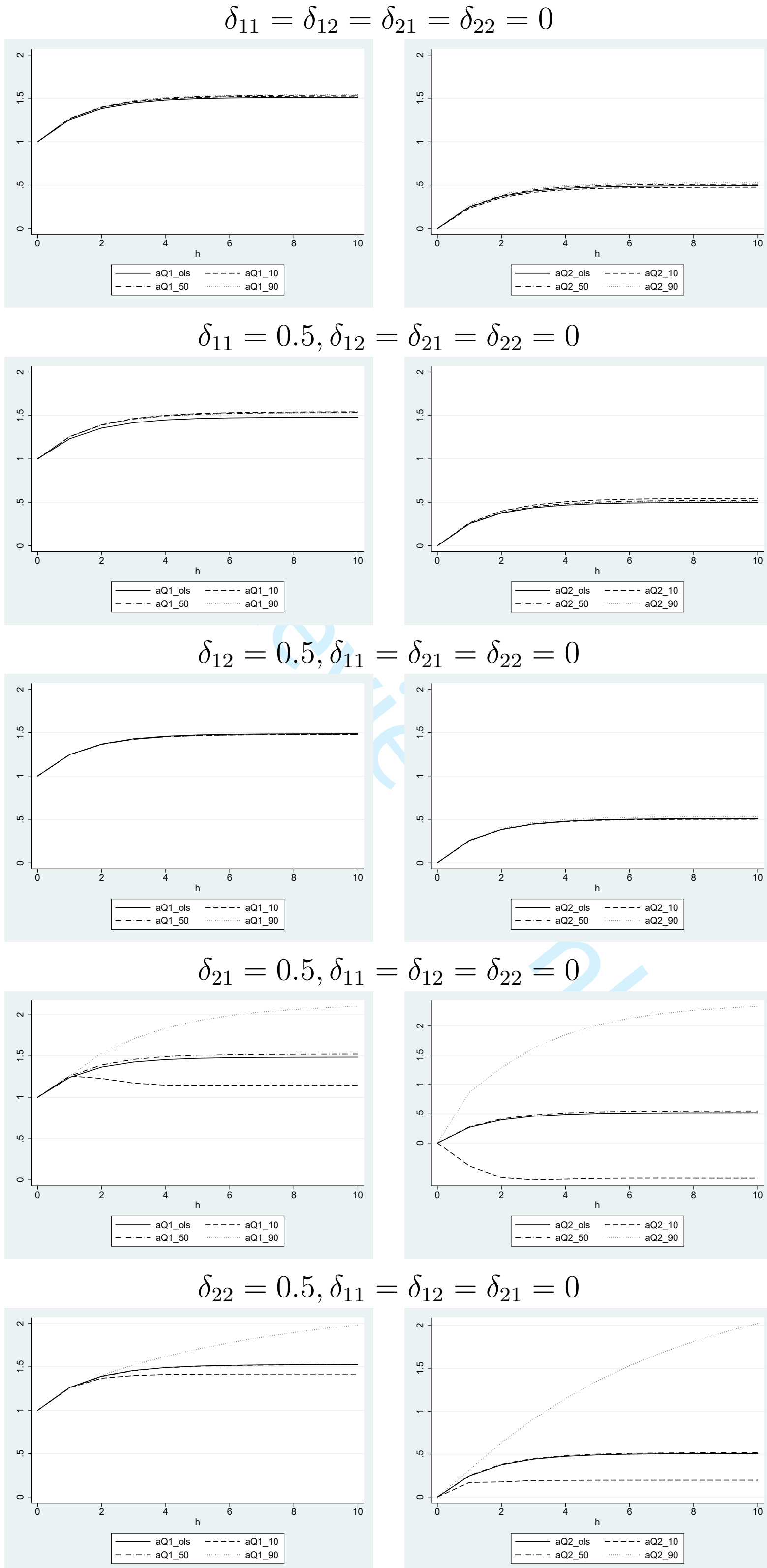
$$Q_{Y_{jt}|Y_{1t-1},Y_{1t-1}}(\tau_j|Y_{1t-1}, Y_{1t-1}) = (\beta_{j1}+\delta_{j1}F_j^{-1}(\tau_j))Y_{1t-1}+(\beta_{j2}+\delta_{j2}F_j^{-1}(\tau_j))Y_{2t-1}+(\alpha_1+F_j^{-1}(\tau_j)),$$

for $j = 1, 2$, where $F_j^{-1}(\tau_j)$ is the τ_j quantile of ε_{jt}. Note that, as noted above, the quantile marginal effects will vary if there is a pair i, j such that $\delta_{ij} \neq 0$.

Consider simulations of $T = 10,000$ observations with $\alpha_1 = \alpha_2 = 1$, $\beta_{ij} = 0.25, \forall i, j = 1, 2$, $\varepsilon_{jt} \sim iid\ Normal(0, 1), j = 1, 2$, with different specifications of the parameters δ_{ij}. We simulate a unit change (shock) in variable 1 and compute IRF (i.e., OLS based VAR model) and QIRF simulations for $\tau_1 = 0.5$ and $\tau_2 \in \{0.1, 0.5, 0.9\}$, using the proposed model in the paper. The QIRF corresponds to different fixed 'quantile paths' for a given pair (τ_1, τ_2).

Figure 5: IRF and QIRF simulations for $\tau_1 = 0.5$ and $\tau_2 \in \{0.1, 0.5, 0.9\}$

Effect in y_1 Effect in y_2



References

Bernard, C., and C. Czado (2015): "Conditional quantiles and tail dependence," *Journal of Multivariate Analysis*, 138, 104–126.

Carlier, G., V. Chernozhukov, and A. Galichon (2016): "Vector quantile regression," *Annals of Statistics*, 44, 1165–1192.

Chavleishvili, S., and S. Manganelli (2016): "Quantile impulse response functions," Working Paper http://www.simonemanganelli.org/Simone/Research_files/QIRF_June2017.pdf.

Chernozhukov, V., A. Galichon, M. Hallin, and M. Henry (2015): "Monge-Kantorovich depth, ranks, quantiles, and signs," CEMMAP Working Paper CWP04/15.

Cho, S., and A. Moreno (2004): "A structural estimation and interpretation of the New Keynesian macro model," Facultad de Ciencias Económicas y Empresariales, Universidad de Navarra, Working Paper 14/03.

——— (2006): "Small-sample study of the New-Keynesian macro model," *Journal of Money, Credit and Banking*, 38, 1461–1481.

Christiano, L., M. Eichenbaum, and C. Evans (1996): "The effects of monetary policy shocks: Evidence from the flow of funds," *Review of Economics and Statistics*, 78, 16–34.

Enders, W. (2005): *Applied Econometric Time Series*. Wiley, Boston.

Engle, R. F., and S. Manganelli (2004): "CAViaR: Conditional autoregressive value at risk by regression quantiles," *Journal of Business and Economic Statistics*, 22, 367–381.

Fraiman, R., and B. Pateiro-López (2012): "Quantiles for finite and infinite dimensional data," *Journal of Multivariate Analysis*, 108, 1–14.

Galvao, A., G. Montes-Rojas, and S. Park (2013): "Quantile autoregressive distributed lag model with an application to house price returns," *Oxford Bulletin of Economics and Statistics*, 75(2), 307–321.

Hallin, M., D. Paindaveine, and M. Šiman (2010): "Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth," *Annals of Statistics*, 38(2), 635–669.

Han, H., O. Linton, T. Oka, and Y.-J. Whang (2016): "The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series," *Journal of Econometrics*, 193(1), 251–270.

Jordà, Ò. (2005): "Estimation and inference of impulse responses by local projections," *American Economic Review*, 95, 161–182.

Koenker, R., and Z. Xiao (2004): "Unit Root Quantile Autoregression Inference," *Journal of the American Statistical Association*, 99, 775–787.

——— (2006): "Quantile autoregression," *Journal of the American Statistical Association*, 101, 980–990.

Ma, L., and R. Koenker (2006): "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134(2), 471–506.

Montes-Rojas, G. (2017): "Reduced form vector directional quantiles," *Journal of Multivariate Analysis*, 158, 20–30.

Paindaveine, D., and M. Šiman (2011): "On directional multiple-output quantile regression," *Journal of Multivariate Analysis*, 102, 193–212.

——— (2012): "Computing multiple-output regression quantile regions," *Computational Statistics and Data Analysis*, 56, 840–853.

Ramey, V. (2016): "Macroeconomic shocks and their propagation," in *John B. Taylor and Harald Uhlig (eds.), Handbook of Macroeconomics, Volume 2*, pp. 71–162. North-Holland, Amsterdam.

Stock, J., and M. Watson (2016): "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregression in macroeconomics," in *John B. Taylor and Harald Uhlig (eds.), Handbook of Macroeconomics, Volume 2*, pp. 415–525. North-Holland, Amsterdam.

Wei, Y. (2008): "An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts," *Journal of the American Statistical Association*, 103(481), 397–409.

White, H., T.-H. Kim, and S. Manganelli (2015): "VAR for VaR: Measuring tail dependence using multivariate regression quantiles," *Journal of Econometrics*, 187, 169–188.

Journal of Time Series Analysis JTSA-4600
Reply to Referee 1

August 15, 2018

I would like to thank you for the very insightful and helpful comments. Given your comments, I have extensively revised the paper and worked on every suggestion you made.

Comments to the Author:

1. **Identification. The Multivariate quantile autoregression relationship that the author consider (and the model characterizes) is a rich and complicate functional relationship. My impression is that such a complicate relationship would correspond to a highly nonlinear quantile regression model. Such a complicate relationship would be difficult to be captured by a "linear-in-parameters" VARQ model suggested in the paper.**
   Can you give one non-trivial example that can be capture by such a "linear-in-parameters" VARQ model?

   Answer: Many thanks for the comment.

   First, I would argue that although the model is a nonlinear one, it is not a complicate or complex one.

   The proposed VARQ model is actually estimated using linear-in-parameters models. The model requires to estimate one linear quantile regression model for each of the m endogenous variables, where for each variable we run a quantile regression on the other contemporaneous endogenous variables and all the lags. Each of these models can be easily estimated as they correspond to linear quantile regression models (`qreg` in STATA or `rq` in R).

The nonlinearity corresponds to a solution of a system of linear equations, see pp.6-7 in the paper. While this is a nonlinear model, I would emphasize that this is not a complicate one. The VARQ model is based on the OLS mean-VAR model, and the nonlinearity is actually found on the mean-VAR model. The reason is the following. A less explored result is that the reduced form VAR model can be constructed by running an OLS model for each endogenous variable on the other contemporaneous endogenous variables and all the lags (i.e., as in eq. (3) replacing the conditional quantile models with conditional expectation models), and then solving a system of equations (i.e., as in eq. (4) replacing the conditional quantile models with conditional expectation models). The reduced-form VAR model is identical to this solution. And this is in fact how the VARQ model is constructed.

A non-trivial example

The Referee asks for an example to study the way the nonlinear relationship appears in the model. Following this comment I have included a new Appendix in the paper with the following simulation exercise. The simulation exercise correspond to a bivariate quantile regression location-scale model.

Consider a two-variable model, i.e., $m = 2$,

$$Y_{1t} = \beta_{11}Y_{1t-1} + \beta_{12}Y_{2t-1} + \alpha_1 + (1 + \delta_{11}Y_{1t-1} + \delta_{12}Y_{2t-1})\varepsilon_{1t},$$

$$Y_{2t} = \beta_{21}Y_{1t-1} + \beta_{22}Y_{2t-1} + \alpha_2 + (1 + \delta_{21}Y_{1t-1} + \delta_{22}Y_{2t-1})\varepsilon_{2t}.$$

Furthermore assume that $\varepsilon_{jt} \sim iid(0, \sigma_j), j = 1, 2$. This corresponds to a reduced form VAR model. In quantile regression models there is a standard procedure to generate heterogeneity across quantiles that is to generate a 'location-scale shift' model where the covariates (i.e., Y_{1t-1} and Y_{2t-1} in our case) affect the variance (i.e., scale) of the model. As such, heteroskedasticity would imply that the mean-based model differ from that of quantiles. A model with $\delta_{ij} = 0, \forall i, j = 1, 2$, produces no heterogeneity in the quantile effects (a model known as a 'location shift' model). If on the other hand there is a pair i, j such that $\delta_{ij} \neq 0$, there will be heterogeneity in the quantile effects. In particular we have

$$E(Y_{jt}|Y_{1t-1}, Y_{1t-1}) = \beta_{j1}Y_{1t-1} + \beta_{j2}Y_{2t-1} + \alpha_1,$$

and

$$Q_{Y_{jt}|Y_{1t-1},Y_{1t-1}}(\tau_j|Y_{1t-1}, Y_{1t-1}) = (\beta_{j1}+\delta_{j1}F_j^{-1}(\tau_j))Y_{1t-1}+(\beta_{j2}+\delta_{j2}F_j^{-1}(\tau_j))Y_{2t-1}+(\alpha_1+F_j^{-1}(\tau_j)),$$

for $j = 1, 2$, where $F_j^{-1}(\tau_j)$ is the τ_j quantile of ε_{jt}. Note that, as noted above, the quantile marginal effects will vary if there is a pair i, j such that $\delta_{ij} \neq 0$.

Consider simulations of $T = 10,000$ observations with $\alpha_1 = \alpha_2 = 1$, $\beta_{ij} = 0.25, \forall i, j = 1, 2$, $\varepsilon_{jt} \sim iid\ Normal(0, 1), j = 1, 2$, with different specifications of the parameters δ_{ij}. We simulate a unit change (shock) in variable 1 and compute IRF (i.e., OLS based VAR model) and QIRF simulations for $\tau_1 = 0.5$ and $\tau_2 \in \{0.1, 0.5, 0.9\}$, using the proposed model in the paper. The QIRF corresponds to different fixed 'quantile paths' for a given pair (τ_1, τ_2).

2. **Computation. A well-known problem in the traditional VAR problem is the large number of unknown parameters to be estimated. Usually additional assumptions and structures are needed to reduce the dimension of parameters in practice.**
   In the case of quantile VAR, in the presence of multiple quantiles $(\tau_1, \ldots \tau_m)$, the number of parameters in the system is even (substantially) larger than the conventional VAR models. How to effectively estimate the proposed models?

   Answer: Many thanks for the comment.

   Let m be the number of endogenous variables and let p be the number of lags to be used in the VAR estimation. Then assuming we use all the p lags for all endogenous variables in all m models, the number of parameters to be estimated in a reduced form VAR model is $m^2p + m$, given that we need mp parameters plus a constant for each of the m equations.

   For a structural VAR form we would need to estimate $m(m-1) + m^2p + m = m^2(p+1)$ if we were to identify all the contemporaneous effects.

   The VARQ model used in the paper and the corresponding quantile impulse-response function (QIRF) can be defined for a fix vector $\boldsymbol{\tau} = (\tau_1, \ldots \tau_m)$. In this case the number of parameters to be estimated for a VARQ is the same as the structural form, $m^2(p+1)$ (see eq. (3), p.6, in the paper), which will be used to define the VARQ model, that is actually a reduced form, and then will deliver $m^2p + m$ parameters (see eq. (4), p.7, in the paper). I include the following clarification in the new version of the paper (p.7):

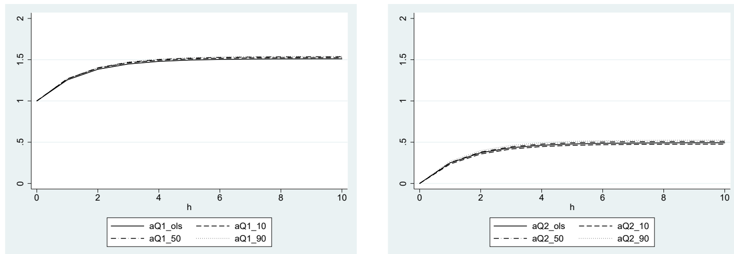
   > "Note that for a fixed $\boldsymbol{\tau}$ the number of parameters to be estimated is that of a structural mean-based VAR model."

   The case considered by the Referee is one in which we evaluate the heterogeneity that could appear in all endogenous variables. While this is the most interesting

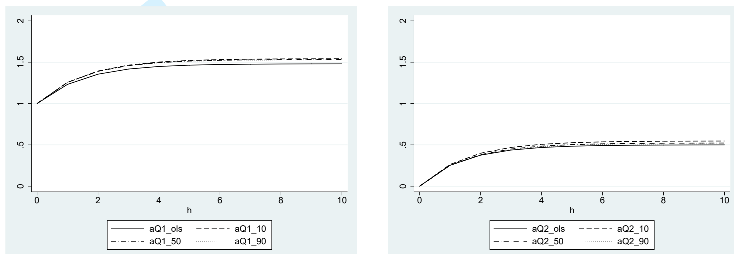
Figure 1: IRF and QIRF simulations for $\tau_1 = 0.5$ and $\tau_2 \in \{0.1, 0.5, 0.9\}$

Effect in y_1 Effect in y_2

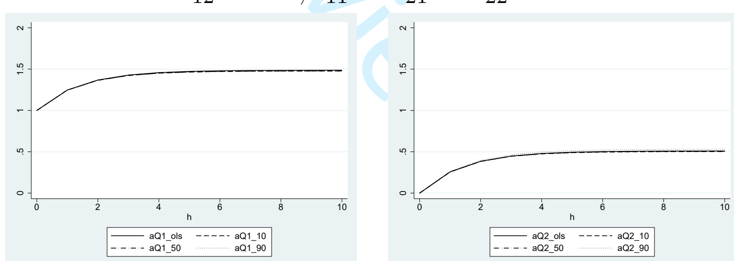
$\delta_{11} = \delta_{12} = \delta_{21} = \delta_{22} = 0$



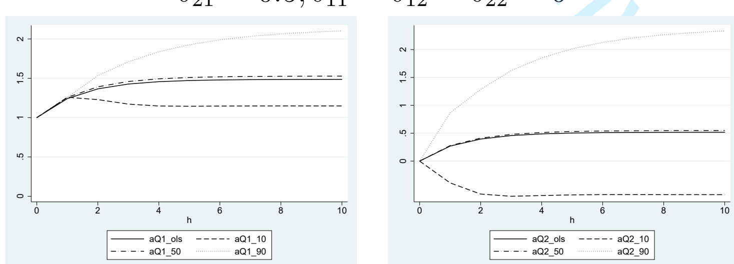
$\delta_{11} = 0.5, \delta_{12} = \delta_{21} = \delta_{22} = 0$



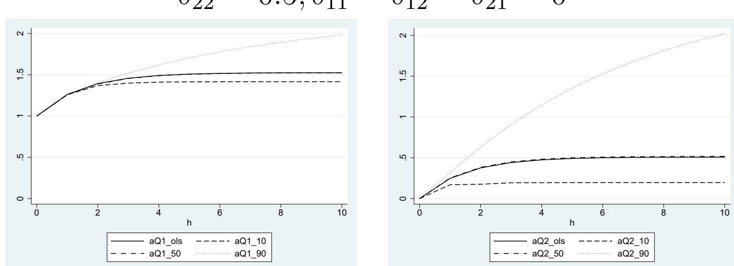
$\delta_{12} = 0.5, \delta_{11} = \delta_{21} = \delta_{22} = 0$



$\delta_{21} = 0.5, \delta_{11} = \delta_{12} = \delta_{22} = 0$



$\delta_{22} = 0.5, \delta_{11} = \delta_{12} = \delta_{21} = 0$



case, I would like to note that the QIRF can be constructed for a single vector a quantiles (which is defined as a quantile path in the paper) as in the previous paragraph. Let q be the number of quantiles considered for each endogenous variable $j = 1, \ldots, m$, which define the grid $\mathcal{T}^{(q)} = (\tau_1, \ldots \tau_q)$. This correspond to each line in equation (3). Then the number of parameters to be estimated for this multiple quantiles model is $q^m(m^2(p+1)$ which will be used to construct $q^m(m^2p + m)$ parameters in the VARQ reduced form.

If q is large then we are constructing a dense grid. As the Referee suggests this model would have too many parameters to be estimated depending on how large m is. However, I would like to emphasize that it is not necessary to consider a dense grid for all of the m variables, as some of them may be fixed at a given quantile. For instance, in the empirical application I consider quantiles in $\{0.05, 0.10, \ldots, 0.95\}$ for output and inflation, and I fix a quantile of 0.50 for the interest rate (see p.14). This corresponds to studying different quantile paths. As such, the number of quantile regression models to be considered is $19 \times 19 \times 1 = 361$, which runs very fast in a standard PCs. When I add heterogeneity in the interest rate I would have 6859. This would require more minutes to run, but programming a loop does not involve a complex model.

In the revised version I included a footnote in the paper where I posted the codes in STATA used for the empirical results in the paper and the simulation exercise.

In closing, we thank you once again for your valuable comments and suggestions. I have made a sincere effort to move the paper in the directions that you propose, and as a result we believe that the paper is greatly improved. I hope that the revised manuscript will meet the requirements of publication in *Journal of Time Series Analysis*.

Journal of Time Series Analysis JTSA-4600
Reply to Referee 2

August 15, 2018

I would like to thank you for the very insightful and helpful comments. Given your comments, I have extensively revised the paper and worked on every suggestion you made.

Specific comments:

1. **Pg6, line -10, "In order to construct the VARQ model, define ..." add "," before after "model".**

   Answer: Thanks. I have modified the text.

2. **Pg 12, line -1, add a brief explanation why QIRF becomes 0 as $h \to \infty$ infinity for stationary model for readers who are not very familiar with IRF. Namely, the impact of shock, which is measured by IRF or QIRF, tends to vanish as the periods become longer and longer.**

   Answer: Many thanks for the comment. I have modified the text to include a simple explanation of why the QRIF should be 0 as $h \to \infty$.

   I include the following footnote 1 in page 12:

   > "The Enders' book (2015, ch.5) has an intuitive explanation for the fact that IRF needs to become 0 as $h \to \infty$. For stationary models the IRF is the collection of the effects of a variable j in a variable k (for all j, k) that occurred at some point $h = 0$ on a different time path $h = 0, 1, \ldots$. These effects have to converge to 0 as h gets larger because "shocks cannot have a permanent effect on a stationary series" (Enders, 1998, p.295). The reason is that "[j]ust as an autoregression has a moving average representation, a vector autoregression can be

written as a vector moving average (...) [where the model is] expressed in terms of the current and past values of the (...) shocks" (Enders, 1998, p.295). This invertibility condition can only be achieved if the effects of past shocks is smaller as we consider longer horizons."

3. **Pg 16, Figure 2, the three lines of τ_y =0.50, 0.05 and 0.95 do not match in type with those listed in the legend. Also, in the legend, group τ_y and τ_p together into two columns.**

Answer: Many thanks for the comment. The previous version had a graph where it cannot be clearly distinguished all the different lines. The new version adds symbols to the graph lines so that they can be clearly identified.

The figure reports the heterogeneity in the effect of the quantile regression coefficient of a lagged change in the interest rate on output (bylr, horizontal axis) and inflation (bplr, vertical axis). I have now used lines with symbols (triangles, squares and diamonds) to better identify the different lines. I have expanded the note to the figure to explain this to the Reader:

Note to Figure 2:
"Notes: The figure reports the heterogeneity in the effect of the quantile regression coefficient of a lagged change in the interest rate on output (bylr, horizontal axis) and inflation (bplr, vertical axis). Vertical and horizontal lines correspond to the mean-based VAR effects. Lines with small triangles symbols correspond to VARQ coefficients with $\tau_y = 0.05, \tau_\pi \in \{0.05, 0.10, \ldots, 0.95\}, \tau_r = 0.50$; small squares $\tau_y = 0.50, \tau_\pi \in \{0.05, 0.10, \ldots, 0.95\}, \tau_r = 0.50$; small diamonds $\tau_y = 0.95, \tau_\pi \in \{0.05, 0.10, \ldots, 0.95\}, \tau_r = 0.50$; large triangles $\tau_y \in \{0.05, 0.10, \ldots, 0.95\}, \tau_\pi = 0.05, \tau_r = 0.50$; large squares $\tau_y \in \{0.05, 0.10, \ldots, 0.95\}, \tau_\pi = 0.50, \tau_r = 0.50$; and large diamonds $\tau_y \in \{0.05, 0.10, \ldots, 0.95\}, \tau_\pi = 0.9, \tau_r = 0.50$."

4. **Pg 16, line 1, Figure ??. Explicitly write down ??.**

Answer: Many thanks for the comment. I apologize for this typo. There was a wrong referencing in the LaTeX file that I compiled. This corresponds to Figure 2. This typo has been corrected.

5. **The data and the codes to obtain the estimation results and produce the figures should be posted on a public website and the website should be provided in the paper in a footnote or in references.**

Answer: Many thanks for the comment. I agree with the Referee. I have included a footnote in the title page stating that the files codes and the data in the empirical example can be downloaded at my webpage: http://gabrielmontes.com.ar/codes

I will make this link permanent. All the results presented in the paper (simulations and empirical application) can be replicated from these codes in STATA.

In closing, we thank you once again for your valuable comments and suggestions. I have made a sincere effort to move the paper in the directions that you propose, and as a result we believe that the paper is greatly improved. I hope that the revised manuscript will meet the requirements of publication in *Journal of Time Series Analysis*.