Quantile Impulse Response Analysis with Applications in Macroeconomics and Finance

Whayoung Jung*　　Ji Hyung Lee†

October, 2021

Abstract

This paper studies the dynamic responses of the conditional quantiles and their applications in macroeconomics and finance. We build a multi-equation autoregressive conditional quantile model and propose a new construction of quantile impulse response functions (QIRFs). The tool set of QIRFs provides detailed distributional evolution of an outcome variable to economic shocks. We show the left tail of economic activity is the most responsive to monetary policy and financial shocks. The impacts of the shocks on *Growth-at-Risk* (the 5% quantile of economic activity) during the global financial crisis are assessed. We also examine how the economy responds to a hypothetical financial distress scenario.

Keywords: Quantile Impulse Response, Growth-at-Risk, Monetary Policy, Financial Shocks

JEL classification: C22

1 Introduction

The conditional mean of an outcome variable has been the primary object of study in economics, as it summarizes the central response to explanatory variables. The scientific interests of policymakers and researchers, however, go beyond the conditional mean. Extreme events and business cycles have significant effects on the economy, so we also need to study the tail or shoulder of the outcome distribution. It is therefore important to obtain a more complete picture of the dynamic responses of the conditional distribution.

As an alternative to the conventional mean regression, Koenker and Bassett (1978) proposed quantile regression (QR). Since QR estimates heterogeneous regression coefficients for a response variable across its conditional distribution, it provides a richer interpretation in regression analysis (see Koenker (2005) for the textbook treatment). Recent development in time series QR models enables researchers to study dynamics at various parts of an outcome distribution.

*Korea Capital Market Institute, 143 Uisadang-daero, Yeongdeungpo-gu, Seoul 07332, Republic of Korea, Email: junghy0103@kcmi.re.kr

†Department of Economics, University of Illinois, 1407 W. Gregory Dr., 214 David Kinley Hall, Urbana, IL 61801, USA. Email: jihyung@illinois.edu

This paper investigates how quantiles of endogenous variables respond over time to a shock in a vector autoregressive (VAR) model. We build a QR model accommodating important dynamics of macro/financial time series data. Certain cross-sectional and time series characteristics, such as dispersion and persistence, are important in their distributional evolution. While those characteristics are not fully measured by observable variables, they can be effectively captured using quantiles. Thus, we modify an autoregressive conditional quantile specification, and model the conditional quantile of innovations in a VAR model as a function of past observable variables as well as its own past quantiles. We adopt the CAViaR-type model by Engle and Manganelli (2004), but extend it to include the level impact of macro variables and the effect of time-varying volatility, see Remark 2.1 for a detailed discussion.

Based on the proposed multi-equation QR model, we construct quantile impulse response functions (QIRFs) applying the quantile framework to impulse response functions (IRFs). There is no consensus on how to define the quantile response. We suggest an alternative definition of the QIRF which is compatible with the conventional mean IRF. This QIRF describes a heterogeneous shock-response mechanism across the distribution, complementing the conventional IRF analysis in macro/finance. Thus the QIRF provides effective tools for empirical policy analysis.

We also provide estimation and statistical inference for the QIRF. We suggest a three-step estimation procedure based on the conventional VAR and a QR model. Moreover, valid econometric inferential tools based on both asymptotics and the residual-based moving block bootstrap are provided. While some researchers estimate quantile responses using the local projection method proposed by Jordà (2005), theoretical results for the application of local projections to QR models have not been fully explored. This paper contributes to statistical inference of quantile responses.

In our empirical application, we first investigate quantile responses of US macroeconomic variables to monetary policy and financial shocks. We find that economic activity has the most heterogeneous response across quantiles, while the response of financial variables is relatively homogeneous. An expansionary monetary policy shock shifts the distribution of economic activity to the right. The shock significantly reduces downside risk to growth but merely affects upside risk. On the contrary, a financial shock shifts the economic activity distribution to the left. The left tail quantiles are substantially more responsive than the median or upper quantiles. This empirical result is in line with Adrian, Boyarchenko, and Giannone (2019) who show that deteriorating financial conditions strongly increase the downside risks to growth, but not the upside risks. Moreover, a monetary policy shock has much more persistent effects on *Growth-at-Risk*, defined as the conditional 5% quantile of economic activity, than on its mean. The dynamic response of Growth-at-Risk to a financial shock decays in a similar way to the mean IRF of economic activity.

Secondly, we quantitatively assess how much downside and upside risks to growth were affected by the financial and monetary policy shocks during and after the Global Financial Crisis (GFC). Financial shocks during August 2007–June 2009 decreased the 5% quantile of Chicago Fed National Activity Index (CFNAI) by 1.5 on average over 2008–2010.[1] However, the decrease in its 95%

[1] CFNAI is a monthly index for US economic activity. Section 5.1 explains the index in detail.

quantile due to the shocks over the same period was much less: -0.8. Monetary policy shocks during July 2009–December 2015 increased the 5% quantile by 0.4 on average over 2010–2015. The increase suggests the unconventional monetary policy after the GFC effectively reduced downside risks to growth. On the other hand, the upper quantile was hardly affected by the monetary policy.

Thirdly, a measure of financial conditions (National Financial Conditions Index, NFCI) exhibits explosive dynamics at its right tail quantiles (tighter financial conditions).[2] When severe financial conditions continue for several months, it creates substantial downside risk to the economy. This locally explosive behavior of financial conditions illustrates that a sharp deterioration of financial markets may lead to a financial crisis in a short period of time.

This paper relates to several strands of literature. From the perspective of econometrics, we extends times series QR models describing heterogeneous dynamics at different parts of the conditional distribution. Over the past few decades, QR methods have been widely applied to time series models to study asymmetric dynamics. Koenker and Zhao (1996), Xiao and Koenker (2009), Koenker and Xiao (2006), and Xiao (2009) estimate QR models for autoregressive conditional heteroskedasticity (ARCH), generalized ARCH (GARCH), autoregressive (AR), and cointegrated processes, respectively. While the conditional quantile is modeled as a linear function of past observations in those models, Engle and Manganelli (2004) develop autoregressive conditional quantile specifications in which the conditional quantile depends not only on past observations but also on unobservable past conditional quantiles. White, Kim, and Manganelli (2015, WKM henceforth) further extend the model to multivariate and multi-quantile models.[3] Modifying their autoregressive conditional quantile specification, our QR model effectively incorporates latent information such as dispersion and persistence of distribution into the evolution of the conditional distribution.

This paper is also related to Chang, Kim, and Park (2021) and Chang, Miller and Park (2021) who study time series of cross-sectional distributions. Chang, Kim and Park (2021) analyze the effects on income distribution of macroeconomic policy shocks, and Chang, Miller and Park (2021) study the dynamics of the global temperature distributions. Our paper modifies the usual VAR approach to study the conditional quantile dynamics, while Chang, Kim, and Park (2021) and Chang, Miller and Park (2021) develop new methodology of mixed autoregression which combines the conventional VAR and functional autoregression (FAR). Please also see Chang, Kim and Park (2016) and Chang et al. (2020) for the earlier related studies.

In terms of empirical applications, our paper closely relates to recent literature investigating asymmetric impacts of economic state variables on upside and downside risks to growth.[4] Adrian, Boyarchenko, and Giannone (2019) find that the lower quantiles of economic activity are substan-

[2]NFCI is an index for US financial conditions measuring the tightness of financial markets. A high value of NFCI represents tighter financial conditions. See Section 5.1 for more details about the index.

[3]See also Xiao (2012) and Linton and Xiao (2018) for recent advances in time series QR models and their applications.

[4]The QR methods are also used to study heterogeneous dynamics for various macroeconomic and financial variables. Chevapatrakul, Kim, and Mizen (2009) evaluate the quantile response of interest rates to inflation and the output gap. Galvao, Montes-Rojas, and Park (2013) study the effects of income and interest rates on UK house price returns across quantiles. Mumtaz and Surico (2015) investigate the dynamic relationship between interest rates and the conditional quantile of consumption.

tially affected by financial conditions, while the upper quantiles are not. Adrian et al. (2018) study the effects of financial conditions on Growth-at-Risk defined as the conditional 5% quantile of GDP growth. They show that looser financial conditions increase the lower quantiles of GDP growth in the short run, but decrease the lower quantiles in the medium term. Loria, Matthes, and Zhang (2019) estimate how upside and downside risks to growth respond to various shocks. They find that the lower quantiles of GDP growth are affected more than other quantiles by all shocks under study (monetary policy, credit spread, and productivity shocks). Although we use a different QR model and a different data set, our empirical findings are largely consistent with these studies.

This paper also contributes to recent studies constructing QIRFs. There have been a few papers investigating the evolution of the conditional quantile in response to a shock. WKM propose the pseudo-QIRF to investigate the response of Value-at-Risk (VaR) based on a GARCH model. The QIRF of Montes-Rojas (2019) estimates the response of quantile paths applying the vector directional quantile model to vector autoregression. The QIRF proposed by Chavleishvili and Manganelli (2019) describes the impact of a shock on the quantile of future quantiles in a VAR model.[5] Han, Jung, and Lee (2019) estimates the quantile response of financial asset returns under the GARCH framework using the local projection method, and Kim, Lee, and Mizen (2019) investigate the quantile response of macroeconomic variables in a VAR framework. Adopting the generalized impulse response function by Koop, Pesaran, and Potter (1996), this paper suggests the QIRF which is conceptually comparable to the standard mean IRF. Section 3.2 compares our QIRF to the existing QIRFs in recent studies.

The rest of the paper is organized as follows. Section 2 introduces the QR model for the innovations in a VAR model. Section 3 proposes the definition and construction of the QIRF, with a brief comparison to the recent literature. Section 4 discusses estimation of the model and QIRF, then provides inferential methods based on asymptotics and the residual-based moving block bootstrap. In Section 5, we study QIRFs of the US economy. In particular, we study the dynamic quantile responses of economic activity during and after the GFC. We also examine the quantile responses of macroeconomic variables in a distress scenario where a deterioration of financial conditions continues. Section 6 concludes, and the online Appendix includes technical assumptions and proofs.

2 Quantile Regression Model for a Structural VAR Analysis

In this section, we introduce the autoregressive conditional quantile model in a VAR framework. Let $\mathbf{y}_t = [y_{1t}\ y_{2t}\ \ldots\ y_{nt}]^\top$ be an $n \times 1$ vector of variables of interest. In a typical structural VAR analysis, $\mathbf{y}_t$ consists of the conditional mean and the innovations:

$$\mathbf{y}_t = \underbrace{A_0 + A_1\mathbf{y}_{t-1} + \ldots + A_p\mathbf{y}_{t-p}}_{\text{The conditional mean}} + \underbrace{\mathbf{u}_t,}_{\text{The innovations}} \qquad \mathbf{u}_t \sim (0, \Sigma_{\mathbf{u}}), \tag{1}$$

[5] They derive the quantile response employing *the law of iterated quantiles* (p.15).

where $\mathbf{u}_t = [u_{1t}\ u_{2t}\ \dots\ u_{nt}]^\top = \Theta_0 \boldsymbol{\epsilon}_t$ for a vector of structural shocks $\boldsymbol{\epsilon}_t = [\epsilon_{1t}\ \epsilon_{2t}\ \dots\ \epsilon_{nt}]^\top \sim (0, I)$ such that $\Sigma_{\mathbf{u}} = \Theta_0 \Theta_0^\top$. The conditional mean and the innovations determine the location and shape of the outcome distribution, respectively.

While the conditional mean is usually the main interest in the VAR literature, researchers begin to pay more attention to time-varying volatility of the innovations. A large literature in macroeconomics employs stochastic volatility (SV) models to take account of volatility dynamics exhibited in macroeconomic variables. In the SV model, variations in volatility are attributed to *a random process*.[6] However, recent studies suggest that the conditional volatility can be accounted for by observable state variables. For example, Adrian, Boyarchenko, and Giannone (2019) find that the conditional volatility of GDP growth is correlated with its conditional mean and financial conditions. Their finding suggests that variations in GDP growth volatility can be explained by GDP growth and financial conditions.

In this paper, we build a QR model explaining systematic dynamics of the innovations. While most of the VAR studies assume a specific distribution such as the multivariate Gaussian for $\mathbf{u}_t$, we model the conditional quantile of the innovations without such assumptions. Moreover, we allow for the conditional heteroskedasticity of unknown form as in Brüggemanna, Jentsch, and Trenkler (2016). We take an empirically driven modeling approach using the QR model. This modeling framework allows us to investigate asymmetry in the downside and upside risks, unlike the typical model with symmetric second-moment dynamics. When evolution of the distribution is accompanied by skewness dynamics, this QR approach can be more effective.

Define a natural filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. All information available at time t is represented by the information set $\mathcal{F}_t$. For $i = 1, 2, ..., n$ and $\tau \in (0, 1)$, let $Q_{u_{it}}(\tau|\mathcal{F}_{t-1})$ denote the τ-quantile of u_{it} conditional on $\mathcal{F}_{t-1}$ such that $Pr(u_{it} \leq Q_{u_{it}}(\tau|\mathcal{F}_{t-1})\big|\mathcal{F}_{t-1}) = \tau$. Our QR model describes the evolution of the conditional distribution using the following autoregressive conditional quantile specifications. For u_{it}, its conditional τ-quantile is modeled as a function of $\mathbf{y}_{t-1}$ and its own past quantiles:

$$\begin{aligned} Q_{u_{it}}(\tau|\mathcal{F}_{t-1}) &= c_{i,\tau} + \mathbf{a}_{i,\tau}^\top \mathbf{y}_{t-1} \\ &\quad + \sum_{k=1}^{l} \left[b_{k,i,\tau} \left(Q_{u_{i,t-k}}(\tau_U|\mathcal{F}_{t-k-1}) - Q_{u_{i,t-k}}(\tau_L|\mathcal{F}_{t-k-1}) \right) + d_{k,i,\tau} Q_{u_{i,t-k}}(\tau|\mathcal{F}_{t-k-1}) \right], \end{aligned} \tag{2}$$

for some τ_U and τ_L such that $0 < \tau_L < \tau_U < 1$. The vector coefficient $\mathbf{a}_{i,\tau}$ measures how the conditional quantile responds to observable economic state variables, $\mathbf{y}_{t-1}$, typically correlated with the business cycle. Note that $\mathbf{a}_{i,\tau}^\top \mathbf{y}_{t-1}$ represents the impacts of macro variables on the shape of the distribution, not on the location as the conditional mean of u_{it} is zero.

The summation terms in (2) describe autoregressive dynamics of u_{it} along its own quantiles, which are elaborated in the following remark.

[6] In the literature of VAR models with SV, volatility dynamics are usually modeled as geometric random walks such as $\log \sigma_t = \log \sigma_{t-1} + \eta_t$ where η_t is a mean-zero stochastic error term. See, e.g., Stock and Watson (2002) and Primiceri (2005).

Remark 2.1 *In Equation (2), $b_{k,i,\tau}$ and $d_{k,i,\tau}$ represent the effect of dispersion of the conditional distribution and the quantile persistence, respectively. As a measure of dispersion, we use the distance between the conditional τ_H-quantile and τ_L-quantile. To get the idea, consider a scale model of $y_t = \sigma_t \varepsilon_t$ with $\sigma_t \in \mathcal{F}_{t-1}$ and $\varepsilon_t \sim iid\ F_\varepsilon$, then it is easy to show that*

$$Q_{u_{it}}(\tau_H|\mathcal{F}_{t-1}) - Q_{u_{it}}(\tau_L|\mathcal{F}_{t-1}) = \left(F_\varepsilon^{-1}(\tau_H) - F_\varepsilon^{-1}(\tau_L)\right)\sigma_t,$$

where $F_\varepsilon^{-1}(\cdot)$ is the inverse cumulative distribution function (CDF). Therefore, $Q_{u_{it}}(\tau_H|\mathcal{F}_{t-1}) - Q_{u_{it}}(\tau_L|\mathcal{F}_{t-1})$ is the conditional volatility scaled by $F_\varepsilon^{-1}(\tau_H) - F_\varepsilon^{-1}(\tau_L)$.[7] Since the distance between the conditional τ_H and τ_L quantiles serves as a measure of volatility, appropriate values need to be chosen for τ_H and τ_L. In this paper, we let $\tau_H = 84\%$ and $\tau_L = 16\%$, which correspond to left and right shoulders of a distribution.[8] The proposed QR model provides rich flexibility in modeling the evolution of the conditional distribution. The lagged conditional quantiles incorporate information not captured by observable variables. Dispersion and persistence play important roles in distribution dynamics in practice. Thus, we use the past conditional quantiles to include the dispersion and autoregressive terms (persistence).

A stable VAR(p) process $\{\mathbf{y}_t\}$ has the following Wold moving-average (MA) representation: $\mathbf{y}_t = \boldsymbol{\mu} + \sum_{k=0}^{\infty} \Phi_k \mathbf{u}_{t-k}$ where $\boldsymbol{\mu} = \left(I - \sum_{k=1}^{p} A_k\right)^{-1} A_0$, $\Phi_0 = I$ and $\Phi_i = \sum_{k=1}^{\min(i,p)} \Phi_{i-k} A_k$. As the conditional quantile itself is an autoregressive process, $Q_{u_{it}}(\tau|\mathcal{F}_{t-1})$ in (2) has a representation in terms of past innovations replacing $\mathbf{y}_{t-1}$ with $\boldsymbol{\mu} + \sum_{k=0}^{\infty} \Phi_k \mathbf{u}_{t-1-k}$. The evolution of the conditional distribution of $\mathbf{u}_t$ is described based on its past history $\{\mathbf{u}_k\}_{k=-\infty}^{t-1}$.

Let $\mathbf{Q}_{\mathbf{u}_t}(\tau|\mathcal{F}_{t-1}) = [Q_{u_{1t}}(\tau|\mathcal{F}_{t-1})\ Q_{u_{2t}}(\tau|\mathcal{F}_{t-1})\ \ldots\ Q_{u_{nt}}(\tau|\mathcal{F}_{t-1})]^\top$. Define an $n \times 1$ matrix $c_\tau = [c_{1,\tau}\ c_{2,\tau}\ \ldots\ c_{n,\tau}]^\top$ and $n \times n$ matrices

$$A_\tau = \begin{bmatrix} \mathbf{a}_{1,\tau}^\top \\ \mathbf{a}_{2,\tau}^\top \\ \vdots \\ \mathbf{a}_{n,\tau}^\top \end{bmatrix},\ B_{k,\tau} = \begin{bmatrix} b_{k,1,\tau} & 0 & \ldots & 0 \\ 0 & b_{k,2,\tau} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & b_{k,n,\tau} \end{bmatrix},\ D_{k,\tau} = \begin{bmatrix} d_{k,1,\tau} & 0 & \ldots & 0 \\ 0 & d_{k,2,\tau} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{k,n,\tau} \end{bmatrix}.$$

[7] For the intuition behind the QR model, let us consider the following linear autoregressive conditional heteroscedasticity (ARCH) volatility dynamics: $\sigma_t = \gamma + \alpha y_{t-1} + \beta\sigma_{t-1}$. Then, the quantile dynamics belongs to the specification in (2):

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = c_\tau + a_\tau y_{t-1} + b_\tau \left(Q_{y_{t-1}}(\tau_U|\mathcal{F}_{t-2}) - Q_{y_{t-1}}(\tau_L|\mathcal{F}_{t-2})\right)$$

where $c_\tau = F_\epsilon^{-1}(\tau) \cdot \gamma$, $a_\tau = F_\epsilon^{-1}(\tau) \cdot \alpha$, and $b_\tau = F_\epsilon^{-1}(\tau) \cdot \beta / \left(F_\epsilon^{-1}(\tau_H) - F_\epsilon^{-1}(\tau_L)\right)$. In addition, our QR model accounts for quantile persistence arises from asymmetric dynamics (or skewness dynamics). Asymmetric dynamics have not been theoretically examined as much as volatility dynamics, thus it is not easy to provide such an example in explicit form. However, a growing number of literature are suggesting empirical evidence of asymmetric dynamics. Hence, we incorporate quantile persistence effect, $d_{k,i,\tau} Q_{u_{i,t-k}}(\tau|\mathcal{F}_{t-k-1})$, in our QR model.

[8] Those quantiles are less volatile than tail quantiles and far enough apart from each other for a volatility measure. Moreover, the [16%, 84%] interval is commonly provided for posterior probability bands in Bayesian inference, and the interval covers approximately two standard deviations around the mean in the case of a Normal distribution.

With the matrix notation, a multi-equation system of (2) can be concisely expressed as

$$\begin{aligned}\mathbf{Q}_{\mathbf{u}_t}(\tau|\mathcal{F}_{t-1}) &= c_\tau + A_\tau \mathbf{y}_{t-1} \\ &+ \sum_{k=1}^{l} \left[B_{k,\tau} \left(\mathbf{Q}_{\mathbf{u}_{t-k}} (\tau_U|\mathcal{F}_{t-k-1}) - \mathbf{Q}_{\mathbf{u}_{t-k}} (\tau_L|\mathcal{F}_{t-k-1}) \right) + D_{k,\tau} \mathbf{Q}_{\mathbf{u}_{t-k}} (\tau|\mathcal{F}_{t-k-1}) \right].\end{aligned} \tag{3}$$

Remark 2.2 *Some QR models study the quantile dependence across variables using high frequency financial time series data. See, e.g., WKM, Li, Li, and Tsai (2015), and Han et al. (2016). However, our model does not allow interactions between quantiles of different variables. The first main reason is that the frequency of macroeconomic data is relatively low, such as quarterly or monthly observations. If we allow non-zero off-diagonal entries in $B_{k,\tau}$ and $D_{k,\tau}$, the estimation becomes infeasible due to the small number of observations relative to the number of parameters. Compared to related macroeconomic literature, the diagonal assumption on $B_{k,\tau}$ and $D_{k,\tau}$ is not too restrictive. In the literature of VAR models with time-varying volatility, volatility dynamics are usually modeled as geometric random walks of its own such as $\log \sigma_{i,t} = \log \sigma_{i,t-1} + \eta_{i,t}$ where $\sigma_{i,t}$ and $\eta_{i,t}$ represent the volatility of structural shock to variable i and the stochastic error, respectively. In those literature, volatility interactions across variables are not allowed either. Secondly, the macroeconomic variables show relatively long-term fluctuation and co-dependence, so the degree of tail dependence across variables is lower than high frequency financial data. Thus, $B_{k,\tau}$ and $D_{k,\tau}$ in (3) are assumed to be diagonal matrices in our model.*[9]

Under model (1) and (3), there is a one-to-one relation between $\mathbf{Q}_{\mathbf{y}_t}(\tau|\mathcal{F}_{t-1})$ and $\mathbf{Q}_{\mathbf{u}_t}(\tau|\mathcal{F}_{t-1})$ as the distribution of $\mathbf{y}_t$ is governed by $\mathbf{u}_t$. Thus, the conditional quantile of $\mathbf{y}_t$ has the following functional form:

$$\begin{aligned}\mathbf{Q}_{\mathbf{y}_t}(\tau|\mathcal{F}_{t-1}) &= A_0 + \sum_{k=1}^{p} A_k \mathbf{y}_{t-k} + c_\tau + A_\tau \mathbf{y}_{t-1} \\ &+ \sum_{k=1}^{l} \left[B_{k,\tau} \left(\mathbf{Q}_{\mathbf{u}_{t-k}} (\tau_U|\mathcal{F}_{t-k-1}) - \mathbf{Q}_{\mathbf{u}_{t-k}} (\tau_L|\mathcal{F}_{t-k-1}) \right) + D_{k,\tau} \mathbf{Q}_{\mathbf{u}_{t-k}} (\tau|\mathcal{F}_{t-k-1}) \right].\end{aligned}$$

3 Quantile Impulse Response Function

In this section, we construct QIRFs to investigate how the conditional quantile of an outcome variable responds to a shock over time. If the shock affects only the location of an outcome distribution, quantile responses will be homogeneous across quantiles. If the shock changes the shape of its distribution, however, QIRFs will show heterogeneous impacts on each quantile. The complete picture of the QIRF mechanism, therefore, complements the conventional IRF.

[9] Although quantile interactions across variables are not accounted for in the model, the cross-sectional tail dependence can be examined indirectly. Adrian and Brunnermeier (2016) measure such tail co-dependence using CoVaR defined as changes in tail risk conditional on another tail event relative to the median state. Similarly, conditional on one variable in particular tail events, responses of another variable's tail risk are investigated in Section 5.5.

3.1 Definition and Construction of QIRF

Since there has not been an agreement on the definition of the quantile response, we propose an alternative definition using the structure of the QR model in Section 2. We measure the quantile response by comparing the conditional quantiles from the following two dynamic paths:

$$\begin{array}{l}\{\ldots,\ \mathbf{y}_{t-2},\ \mathbf{y}_{t-1},\ \mathbf{y}_t,\ \boldsymbol{\mu}_{t+1}(\mathbf{y}_t),\ \boldsymbol{\mu}_{t+2}(\mathbf{y}_t),\ \boldsymbol{\mu}_{t+3}(\mathbf{y}_t),\ \ldots\},\\ \{\ldots,\ \mathbf{y}_{t-2},\ \mathbf{y}_{t-1},\ \tilde{\mathbf{y}}_t,\ \boldsymbol{\mu}_{t+1}(\tilde{\mathbf{y}}_t),\ \boldsymbol{\mu}_{t+2}(\tilde{\mathbf{y}}_t),\ \boldsymbol{\mu}_{t+3}(\tilde{\mathbf{y}}_t),\ \ldots\},\end{array} \tag{4}$$

where $\tilde{\mathbf{y}}_t = \mathbf{y}_t + \mathbf{u}_t = \mathbf{y}_t + \Theta_0 \boldsymbol{\epsilon}_t$ and

$$\boldsymbol{\mu}_{t+s}(\mathbf{y}_t) = \begin{cases} E\left[\mathbf{y}_{t+1} \mid \mathbf{y}_t; \mathcal{F}_{t-1}\right], & \text{for } s = 1, \\ E\left[\mathbf{y}_{t+s} \mid \boldsymbol{\mu}_{t+s-1}(\mathbf{y}_t),\ \ldots,\ \boldsymbol{\mu}_{t+1}(\mathbf{y}_t), \mathbf{y}_t; \mathcal{F}_{t-1}\right], & \text{for } s \geq 2. \end{cases}$$

The two paths are identical up to time $t-1$. At time t, one path is hit by a shock $\boldsymbol{\epsilon}_t$, but the other is not. After time t, realizations of $\{\mathbf{y}_{t+s}\}_{s\geq 1}$ are assumed to be the conditional mean based on their own history in each time path. Koop, Pesaran, and Potter (1996) illustrate that the conventional mean IRF can be defined as the difference between the conditional means from the two time paths:

$$IRF^{(s)} := \begin{cases} E\left[\mathbf{y}_{t+1} \mid \tilde{\mathbf{y}}_t; \mathcal{F}_{t-1}\right] - E\left[\mathbf{y}_{t+1} \mid \mathbf{y}_t; \mathcal{F}_{t-1}\right], & \text{for } s = 1, \\ E\left[\mathbf{y}_{t+s} \mid \boldsymbol{\mu}_{t+s-1}(\tilde{\mathbf{y}}_t),\ \ldots,\ \boldsymbol{\mu}_{t+1}(\tilde{\mathbf{y}}_t), \tilde{\mathbf{y}}_t; \mathcal{F}_{t-1}\right] & \\ \qquad -E\left[\mathbf{y}_{t+s} \mid \boldsymbol{\mu}_{t+s-1}(\mathbf{y}_t),\ \ldots,\ \boldsymbol{\mu}_{t+1}(\mathbf{y}_t), \mathbf{y}_t; \mathcal{F}_{t-1}\right], & \text{for } s \geq 2. \end{cases}$$

They propose a generalized impulse response function applying a similar impulse response concept to nonlinear models.

Following their intuition, we define the QIRF as the difference between the conditional quantiles from the two time paths.

Definition 3.1

$$QIRF_\tau^{(s)} := \begin{cases} Q_{\mathbf{y}_{t+1}}\left(\tau \mid \tilde{\mathbf{y}}_t; \mathcal{F}_{t-1}\right) - Q_{\mathbf{y}_{t+1}}\left(\tau \mid \mathbf{y}_t; \mathcal{F}_{t-1}\right), & \textit{for } s = 1, \\ Q_{\mathbf{y}_{t+s}}\left(\tau \mid \boldsymbol{\mu}_{t+s-1}(\tilde{\mathbf{y}}_t),\ \ldots,\ \boldsymbol{\mu}_{t+1}(\tilde{\mathbf{y}}_t), \tilde{\mathbf{y}}_t; \mathcal{F}_{t-1}\right) & \\ \qquad -Q_{\mathbf{y}_{t+s}}\left(\tau \mid \boldsymbol{\mu}_{t+s-1}(\mathbf{y}_t),\ \ldots,\ \boldsymbol{\mu}_{t+1}(\mathbf{y}_t), \mathbf{y}_t; \mathcal{F}_{t-1}\right), & \textit{for } s \geq 2. \end{cases}$$

Definition 3.1 can be interpreted as a quantile version of the IRF.[10] Under this definition, QIRFs

[10] As explained in Section 2, $\mathbf{Q}_{\mathbf{u}_t}(\tau|\mathcal{F}_{t-1})$ can be expressed as a linear function of the past innovations. Accordingly, the QIRF could be defined as $\frac{\partial \mathbf{Q}_{\mathbf{u}_t}(\tau|\mathcal{F}_{t-1})}{\partial \boldsymbol{\epsilon}_t'}$ for a structural shock $\boldsymbol{\epsilon}_t$, in the same way as the IRF is defined as $\frac{\partial \mathbf{y}_{t+s}}{\partial \boldsymbol{\epsilon}_t'}$. Under such a definition, the resulting QIRF still has the same representation as (5). We use Definition 3.1 since it provides a more intuitive interpretation of the QIRF.

are recursively expressed as

$$
\begin{aligned}
QIRF_{\tau}^{(s)} =& IRF^{(s)} + A_{\tau} IRF^{(s-1)} \\
&+ \sum_{k=1}^{r} \left[B_{k,\tau} \left(QIRF_{\tau_U}^{(s-k)} - QIRF_{\tau_L}^{(s-k)} \right) + D_{k,\tau} \left(QIRF_{\tau}^{(s-k)} - IRF^{(s-k)} \right) \right],
\end{aligned} \tag{5}
$$

where $r = \min\{s-1, l\}$ and

$$
IRF^{(s)} = \begin{cases} \Theta_0 \boldsymbol{\epsilon}_t, & \text{for} \quad s = 0, \\ \sum_{k=1}^{\min\{s,p\}} A_k IRF^{(s-k)}, & \text{for} \quad s \geq 1. \end{cases}
$$

Using the QIRF, heterogeneous dynamics across quantiles can be closely examined. When tail risks are more responsive to volatility, then these dynamics are described by the QIRF at tail quantiles as $B_{k,\tau}$ captures such heterogeneous responses. When a shock has a more persistent impact on specific parts of the distribution, their QIRFs account for this effect as $D_{k,\tau}$ measures the degree of persistence.

3.2 Comparison to Recent QIRF Studies

Recently, there have been a few attempts to investigate quantile dynamics based on QR models. In this subsection, we compare our QIRF with other approaches in the related literature.

Some studies estimate quantile responses applying local projections by Jordà (2005), though econometric theories have not been thoroughly examined yet. Adrian et al. (2018) apply the local projection method to a standard QR model, whereas Han, Jung, and Lee (2019) apply the method to the autoregressive conditional quantile model of WKM. Loria, Matthes, and Zhang (2019), on the other hand, use local projections indirectly. They first estimate quantiles of a dependent variable using a standard QR model, then construct quantile responses applying the local projection method to an ordinary least squares (OLS) regression in which the response variable is the estimated quantile. Instead of using the local projection, we take a different approach to the construction of quantile responses: the QIRF is constructed based on a multi-equation describing the evolution of the system. As theoretical results for the application of local projections to the quantile framework have not been developed yet, our approach can complement their local projection methods. In particular, our model can be effective when unobservable latent information, which are not easily controlled for in local projections, plays nontrivial role in the evolution of quantiles.

As the attention to downside and upside risks to economic variables increase, a growing number of researchers are constructing QIRFs in a VAR framework. However, there is not yet an agreement about how to define the quantile response, and each study has defined it in different ways. Similar to our paper, WKM and Montes-Rojas (2019) define their QIRFs as the difference between the conditional quantiles from two time paths: one path is affected by a shock, but the other path (as the benchmark) is not. However, their formulation of the time paths is different from ours. For the pseudo-QIRF of WKM, the two time paths are identical except at the time when a shock hits the system. This scenario does not account for the effect of the shock on subsequent conditional distributions. As a result, their pseudo-QIRF underestimates the magnitude of a shock on quantile responses.[11] Montes-Rojas (2019), on the other hand, assumes persistent realizations of the lower (or upper) quantile for the time paths compared in his QIRF construction.[12] Thus, his QIRF describes the cumulative impact of shocks as if the economy is under continuous distress. Moreover, the quantile responses are not directly comparable across quantiles because the response at each quantile is against a series of different shocks.

The QIRF of Kim, Lee, and Mizen (2019) is conceptually the same as the QIRF of this paper. Their QIRF measures the expected change in the conditional quantile due to a shock. But, their underlying QR model is different from ours: their model specifies the conditional quantile of $\mathbf{y}_t$ where as ours specifies that of $\mathbf{u}_t$. As a result, their QIRF is constructed in a different way from ours. Chavleishvili and Manganelli (2019) take a quite different approach to quantile responses. Their QIRF is derived applying what they call *the law of iterated quantiles*, and it measures the effect of a shock on the quantiles of future quantiles.

4 Estimation and Statistical Inference

In this section, we provide inferential tools for the QIRF estimation. We discuss inferential methods based on asymptotics and the residual-based moving block bootstrap (MBB). This paper mainly employs the statistical inference of Brüggemanna, Jentsch, and Trenkler (2016). While their inferential methods are for the mean impulse response in VAR models with conditional heteroskedasticity, this paper apply the methods for the quantile impulse response.

4.1 Assumption

First, we adopt Assumption 2.1 of Brüggemanna, Jentsch, and Trenkler (2016) who examine stable VAR models with conditional heteroskedasticity.

Assumption 4.1 *(1) Let* $A(L) = I - \sum_{k=1}^{p} A_k L^k$*.* $\det(A(z)) \neq 0$ *for all* $|z| \leq 1$*. (2) The white noise process* $\{\mathbf{u}_t\}$ *is strictly stationary and strong mixing. (3)* $\Sigma_{\mathbf{u}} = \mathbb{E}\left[\mathbf{u}_t\mathbf{u}_t^\top\right]$ *is positive definite.*

Assumption 4.2 *(1) Let* $\alpha_{\mathbf{u}}(k) = \sup_{A\in\mathcal{F}_{-\infty}^{0}, B\in\mathcal{F}_{k}^{\infty}} |P(A\cap B) - P(A)P(B)|$ *for* $k \in \mathbb{N}$ *denote the* α*-mixing coefficients of the process* $\{\mathbf{u}_t\}$*. For some* $\delta > 0$*,* $\sum_{k=1}^{\infty} (\alpha_{\mathbf{u}}(k))^{\delta/(2+\delta)} < 0$ *and* $\mathbb{E}\,|\mathbf{u}_t|_{4+2\delta}^{4+2\delta}$ *is bounded where* $|A|_p = \left(\sum_{i,j}|a_{ij}|^p\right)^{1/p}$ *for* $A = (a_{ij})$*. (2) For* $a, b, c \in \mathbb{Z}$*, define an* $(n^2 \times n^2)$ *matrix* $\kappa_{a,b,c} := \mathbb{E}\left(\operatorname{vec}\left(\mathbf{u}_t\mathbf{u}_{t-a}^\top\right)\operatorname{vec}\left(\mathbf{u}_{t-b}\mathbf{u}_{t-c}^\top\right)^\top\right)$ *and denote* $\tilde{n} = n(n+1)/2$*. For* $m \in \mathbb{N}$*, there*

[11] WKM acknowledge that the pseudo-QIRF ignores the dynamic evolution of distribution. Han, Jung, and Lee (2019) discuss and evaluate the performance of the pseudo-QIRF.

[12] In his empirical application, he estimates QIRFs with three variables: the output gap, inflation, and the federal funds rate. For the construction of the quantile response of output gap at $\tau = 0.1$, for example, he considers the following time path. After a shock, realizations of the output gap are assumed to be at its conditional 10% quantile continuously, but realizations of inflation and the federal funds rate are assumed to be their conditional median.

exists an $(n^2m+\tilde{n})\times(n^2m+\tilde{n})$ *positive definite matrix* Σ_m *defined as*

$$\Sigma_m := \begin{bmatrix} \Sigma_m^{(1,1)} & \Sigma_m^{(2,1)\top} \\ \Sigma_m^{(2,1)} & \Sigma^{(2,2)} \end{bmatrix},$$

where

$$\begin{aligned}
\Sigma_m^{(1,1)} &:= \left(\sum_{h=-\infty}^{\infty} \kappa_{i,h,h+j}\right)_{i,j,=1,2,\ldots,m}, \\
\Sigma_m^{(2,1)} &:= \sum_{h=-\infty}^{\infty} L_n\left(\kappa_{0,h,h+1},\ldots,\kappa_{0,h,h+m}\right), \\
\Sigma^{(2,2)} &:= \sum_{h=-\infty}^{\infty} L_n\left(\kappa_{0,h,h} - vec\left(\Sigma_{\mathbf{u}}\right) vec\left(\Sigma_{\mathbf{u}}\right)^\top\right) L_n^\top,
\end{aligned}$$

and L_n *is the* $(\tilde{n}\times n^2)$ *elimination matrix which is defined such that* $vech(A) = L_n vec(A)$ *holds for any* $(n\times n)$ *matrix* A.

For the innovation process, we assume it satisfies the mixing condition of Assumption 4.2 (1) instead of the *iid* assumption, and this allows for conditional heteroskedasticity. Their dynamics are described using the conditional quantile model of (3). Given Assumption 4.2 (2), the asymptotic covariance matrix of the VAR model is positive definite.

We also adopt the modeling assumptions in Section 2 and Appendix of WKM. As our QR model is not the same as their model, we adjust their assumptions.[13] With Assumption A.1–A.5, which are in Online Appendix A, the asymptotic distribution of the QR estimator is derived.

4.2 Estimation

Estimation of our QR model is not trivial because of the autoregressive conditional quantile specifications for multi quantiles. We use the following three-step estimation procedure.[14]

Step 1 Estimate the following VAR model using OLS:

$$\mathbf{y}_t = A_1\mathbf{y}_{t-1} + \ldots + A_p\mathbf{y}_{t-p} + \mathbf{u}_t, \qquad \mathbf{u}_t \sim (0, \Sigma_{\mathbf{u}}).$$

Let the estimator be $\{\widehat{A}_k\}_{k=1}^p$. We denote $\widehat{\mathbf{y}}_t = \sum_{k=1}^p \widehat{A}_k\mathbf{y}_{t-k}$ and $\widehat{\mathbf{u}}_t = [\widehat{u}_{1t}\ \widehat{u}_{2t}\ \ldots\ \widehat{u}_{nt}]^\top = \mathbf{y}_t - \widehat{\mathbf{y}}_t$.

[13] There are two main differences between the QR model of WKM and ours. First, we decompose $\{\mathbf{y}_t\}$ into its conditional mean and the innovations, then the QR is used to explain the conditional quantile of the latter. However, their QR model describes the conditional quantile of $\{\mathbf{y}_t\}$ directly without such decomposition. Second, as explained in Remark 2.2, our QR model does not allow the quantile dependence across variables. On the contrary, their model incorporates such codependence across variables.

[14] In this section, we assume the intercept of the VAR model is zero ($A_0 = 0$) for notational simplicity. The QIRF in (5) does not depend on A_0.

Step 2 Define a $(1+n+2l)\times 1$ vector $\boldsymbol{\theta}_{i,\tau} := [c_{i,\tau} \quad \mathbf{a}_{i,\tau}^\top \quad b_{1,i,\tau} \ \ldots \ b_{l,i,\tau} \ d_{1,i,\tau} \ \ldots \ d_{l,i,\tau}]^\top$ and estimate coefficients at *the shoulder quantiles*, τ_U and τ_L, by solving the following minimization problem:

$$\min_{\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}} \frac{1}{T}\sum_{t=1}^{T}\left[\rho_{\tau_U}\left(\widehat{u}_{it}-q_{i,t,\tau_U}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)\right)+\rho_{\tau_L}\left(\widehat{u}_{it}-q_{i,t,\tau_L}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)\right)\right], \tag{6}$$

where $q_{i,t,\tau}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right) = c_{i,\tau}+\mathbf{a}_{i,\tau}^\top\mathbf{y}_{t-1}+\sum_{k=1}^{l}[b_{k,i,\tau}\left(q_{i,t-k,\tau_U}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)-q_{i,t-k,\tau_L}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)\right)$ $+d_{k,i,\tau}q_{i,t-k,\tau}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)]$ for $\tau=\tau_U,\tau_L$, and $\rho_\tau(u)=u(\tau-1[u<0])$.

Step 3 Based on the estimator $\widehat{\boldsymbol{\theta}}_{i,\tau_U}$ and $\widehat{\boldsymbol{\theta}}_{i,\tau_L}$ from Step 2, estimate the τ-coefficient $\boldsymbol{\theta}_{i,\tau}$ for $\tau\neq\tau_U,\tau_L$, which solves the minimization problem below:

$$\min_{\boldsymbol{\theta}_{i,\tau}} \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(\widehat{u}_{it}-q_{i,t,\tau}\left(\boldsymbol{\theta}_{i,\tau}|\widehat{\boldsymbol{\theta}}_{i,\tau_U},\widehat{\boldsymbol{\theta}}_{i,\tau_L}\right)\right),$$

where $q_{i,t,\tau}\left(\boldsymbol{\theta}_{i,\tau}|\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right) = c_{i,\tau}+\mathbf{a}_{i,\tau}^\top\mathbf{y}_{t-1}+\sum_{k=1}^{l}[b_{k,i,\tau}\left(q_{i,t-k,\tau_U}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)-q_{i,t-k,\tau_L}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)\right)$ $+d_{k,i,\tau}q_{i,t-k,\tau}\left(\boldsymbol{\theta}_{i,\tau}|\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)]$ for $\tau\neq\tau_U,\tau_L$.

Remark 4.1 *Instead of Steps 2 and 3, $\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}$ and $\boldsymbol{\theta}_{i,\tau}$ could be estimated simultaneously. The simultaneous estimation solves the following minimization problem:*

$$\begin{aligned}\min_{\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L},\boldsymbol{\theta}_{i,\tau}} \frac{1}{T}\sum_{t=1}^{T}\Big[&\rho_{\tau_U}\left(\widehat{u}_{it}-q_{i,t,\tau_U}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)\right)\\ &+\rho_{\tau_L}\left(\widehat{u}_{it}-q_{i,t,\tau_L}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)\right)+\rho_\tau\left(\widehat{u}_{it}-q_{i,t,\tau}(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L},\boldsymbol{\theta}_{i,\tau})\right)\Big],\end{aligned}$$

where $q_{i,t,\tau}(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L},\boldsymbol{\theta}_{i,\tau}) = c_{i,\tau}+\mathbf{a}_{i,\tau}^\top\mathbf{y}_{t-1}+\sum_{k=1}^{l}[b_{k,i,\tau}\left(q_{i,t-k,\tau_U}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)-q_{i,t-k,\tau_L}\left(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L}\right)\right)$ $+d_{k,i,\tau}q_{i,t-k,\tau}(\boldsymbol{\theta}_{i,\tau_U},\boldsymbol{\theta}_{i,\tau_L},\boldsymbol{\theta}_{i,\tau})]$ for $\tau\neq\tau_U,\tau_L$. Under the simultaneous estimation, a different choice of τ leads to different estimates for $\boldsymbol{\theta}_{i,\tau_U}$ and $\boldsymbol{\theta}_{i,\tau_L}$ though the differences are not substantial. Since the distance between the τ_U and τ_L-quantiles serves as a measure of volatility in the QR model, the coefficients $\boldsymbol{\theta}_{i,\tau_U}$ and $\boldsymbol{\theta}_{i,\tau_L}$ play an important role in the construction of the QIRF. Hence, we adopt the three-step estimation strategy which yields robust estimates of the coefficients (thus the QIRF). We show the QR estimators are consistent in the following section.

In this paper, we assume a structural shock is identified by the Cholesky restriction: $\Sigma_{\mathbf{u}}=\Theta_0\Theta_0^\top$ and $\widehat{\Sigma}_{\mathbf{u}}=\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{u}}_t^\top=\widehat{\Theta}_0\widehat{\Theta}_0^\top$ where Θ_0 and $\widehat{\Theta}_0$ are lower triangular.[15] Let $IRF_i^{(s)}$ and $QIRF_{i,\tau}^{(s)}$ denote the i-th element of $IRF^{(s)}$ and $QIRF_\tau^{(s)}$, respectively. From (5), the estimator for $QIRF_{i,\tau}^{(s)}$ is recursively constructed using the OLS estimator ($\{\widehat{A}_k\}_{k=1}^p$ and $\widehat{\Theta}_0$) and QR estimator ($\widehat{\boldsymbol{\theta}}_{i,\tau_U},\widehat{\boldsymbol{\theta}}_{i,\tau_L}$

[15] Instead of the Cholesky restriction, an alternative identification strategy can be used with the identification restriction $C_{\Theta_0^{-1}}\mathrm{vec}\left(\Theta_0^{-1}\right)=c_{\Theta_0^{-1}}$ where $C_{\Theta_0^{-1}}$ is an $\tilde{n}\times n^2$ selection matrix and $c_{\Theta_0^{-1}}$ is a suitable $\tilde{n}\times 1$ fixed vector. For details, see Chapter 9.1 of Lütkepohl (2005).

and $\widehat{\boldsymbol{\theta}}_{i,\tau}$):

$$\begin{aligned}\widehat{QIRF}_{i,\tau}^{(s)} =& \widehat{IRF}_{i}^{(s)} + \widehat{\mathbf{a}}_{i\tau}^{\top}\widehat{IRF}^{(s-1)} \\ &+ \sum_{k=1}^{r}\left[\widehat{b}_{k,i,\tau}\left(\widehat{QIRF}_{i,\tau_U}^{(s-k)} - \widehat{QIRF}_{i,\tau_L}^{(s-k)}\right) + \widehat{d}_{k,i,\tau}\left(\widehat{QIRF}_{i,\tau}^{(s-k)} - \widehat{IRF}_{i}^{(s-k)}\right)\right],\end{aligned} \tag{7}$$

where $r = \min\{s-1, l\}$ and

$$\widehat{IRF}^{(s)} = \begin{cases}\widehat{\Theta}_0\boldsymbol{\epsilon}_t, & \text{for } \ s=0,\\ \sum_{k=1}^{\min\{s,p\}}\widehat{A}_k\widehat{IRF}^{(s-k)} & \text{for } \ s\geq 1.\end{cases}$$

4.3 Asymptotic Inference

Let $\boldsymbol{\beta} = \text{vec}\,(A_1, ..., A_p)$, $\sigma = \text{vech}\,(\Sigma_{\mathbf{u}})$, $\widehat{\boldsymbol{\beta}} = \text{vec}\left(\widehat{A}_1, ..., \widehat{A}_p\right)$ and $\widehat{\sigma} = \text{vech}\left(\widehat{\Sigma}_{\mathbf{u}}\right)$. Using the moving-average representation of $\mathbf{y}_t = \sum_{k=0}^{\infty}\Phi_k\mathbf{u}_{t-j}$, let $\Phi_0 = I$ and $\Phi_i = \sum_{k=1}^{min(i,p)}\Phi_{i-k}A_k$ for $i\in\mathbb{N}$. Define an $np\times n$ matrix $C_i = \left(\Phi_{i-1}^{\top}, ..., \Phi_{i-p}^{\top}\right)^{\top}$ and an $np\times np$ matrix $\Gamma = \sum_{k=1}^{\infty}C_k\Sigma_{\mathbf{u}}C_k^{\top}$.

Lemma 4.1 follows from Theorem 2.1 of Brüggemanna, Jentsch, and Trenkler (2016).

Lemma 4.1 *Under Assumption 4.1 and 4.2,*

$$\sqrt{T}\begin{bmatrix}\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\\ \widehat{\sigma}-\sigma\end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}V_{\widehat{\boldsymbol{\beta}}} & V_{\widehat{\sigma},\widehat{\boldsymbol{\beta}}}^{\top}\\ V_{\widehat{\sigma},\widehat{\boldsymbol{\beta}}} & V_{\widehat{\sigma}}\end{bmatrix}\right), \tag{8}$$

where

$$\begin{aligned}V_{\widehat{\boldsymbol{\beta}}} &= \left(\Gamma^{-1}\otimes I\right)\left(\sum_{i,j=1}^{\infty}(C_i\otimes I)\sum_{h=-\infty}^{\infty}\kappa_{i,h,h+j}\,(C_j\otimes I)^{\top}\right)\left(\Gamma^{-1}\otimes I\right)^{\top},\\ V_{\widehat{\sigma},\widehat{\boldsymbol{\beta}}} &= L_n\left(\sum_{j=1}^{\infty}\sum_{h=-\infty}^{\infty}\kappa_{0,h,h+j}\,(C_j\otimes I)^{\top}\right)\left(\Gamma^{-1}\otimes I\right)^{\top},\\ V_{\widehat{\sigma}} &= L_n\left(\sum_{h=-\infty}^{\infty}\left(\kappa_{0,h,h} - vec\,(\Sigma_{\mathbf{u}})\,vec\,(\Sigma_{\mathbf{u}})^{\top}\right)\right)L_n^{\top}.\end{aligned}$$

As $IRF^{(s)}$ is continuously differentiable functions of $\boldsymbol{\beta}$ and σ, the asymptotic distribution of the IRF estimator is obtained applying the Delta method to Lemma 4.1. Lemma 4.2 follows from Corollary 5.1 of Brüggemanna, Jentsch, and Trenkler (2016).

Lemma 4.2 *Under Assumption 4.1 and 4.2,*

$$\sqrt{T}\left(\widehat{IRF}_i^{(s)} - IRF_i^{(s)}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, V_{\widehat{IRF}_i^{(s)}}\right), \tag{9}$$

where $V_{\widehat{IRF}_i^{(s)}} = C_{i,\boldsymbol{\beta}}^{(s)} V_{\widehat{\boldsymbol{\beta}}} C_{i,\boldsymbol{\beta}}^{(s)\top} + C_{i,\sigma}^{(s)} V_{\widehat{\sigma}} C_{i,\sigma}^{(s)\top} + C_{i,\boldsymbol{\beta}}^{(s)} V_{\widehat{\sigma},\widehat{\boldsymbol{\beta}}}^{\top} C_{i,\sigma}^{(s)\top} + C_{i,\sigma}^{(s)} V_{\widehat{\sigma},\widehat{\boldsymbol{\beta}}} C_{i,\boldsymbol{\beta}}^{(s)\top}$, $C_{i,\boldsymbol{\beta}}^{(s)} = \frac{\partial IRF_i^{(s)}}{\partial \boldsymbol{\beta}^{\top}}\Big|_{\widehat{\boldsymbol{\beta}}}$ *and* $C_{i,\sigma}^{(s)} = \frac{\partial IRF_i^{(s)}}{\partial \sigma^{\top}}\Big|_{\widehat{\sigma}}$.

Define $\boldsymbol{\Theta}_{i,\tau} := \begin{bmatrix} \boldsymbol{\theta}_{i,\tau_U}^{\top} & \boldsymbol{\theta}_{i,\tau_L}^{\top} & \boldsymbol{\theta}_{i,\tau}^{\top} \end{bmatrix}^{\top}$ and its estimator $\widehat{\boldsymbol{\Theta}}_{i,\tau} := \begin{bmatrix} \widehat{\boldsymbol{\theta}}_{i,\tau_U}^{\top} & \widehat{\boldsymbol{\theta}}_{i,\tau_L}^{\top} & \widehat{\boldsymbol{\theta}}_{i,\tau}^{\top} \end{bmatrix}^{\top}$ which is estimated using the consistent estimator $\widehat{\boldsymbol{\beta}}$.[16] Consistency and asymptotic normality of the QR estimator follow from the asymptotic theories of Engle and Manganelli (2004) and WKM.

Lemma 4.3 *Under Assumptions 4.1, 4.2, and A.1–A.5,*

$$\sqrt{T}\left(\widehat{\boldsymbol{\Theta}}_{i,\tau} - \boldsymbol{\Theta}_{i,\tau}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{Q}_{i,\tau}^{-1} \mathbf{V}_{i,\tau} \mathbf{Q}_{i,\tau}^{-1}\right), \tag{10}$$

where $\mathbf{V}_{i,\tau}$ *and* $\mathbf{Q}_{i,\tau}$ *are defined in Online Appendix A.*

While the above lemmas derive the asymptotic distributions of the OLS, IRF and QR estimators, it is challenging to derive the asymptotic distribution of the QIRF estimator. Since the estimator is a function of both the OLS and QR estimators ($\widehat{\boldsymbol{\beta}}, \widehat{\sigma}$ and $\widehat{\boldsymbol{\Theta}}_{i,\tau}$) as in (7), its asymptotic distribution could be derived applying the Delta method to the joint asymptotic distribution of $\widehat{\boldsymbol{\beta}}, \widehat{\sigma}$ and $\widehat{\boldsymbol{\Theta}}_{i,\tau}$. However, the derivation is not easy because the QR estimator does not have an explicit expression.

Moreover, it is difficult to estimate the asymptotic covariance matrix of the QR estimator accurately because of a nuisance parameter.[17] The asymptotic inference is less satisfactory particularly at tail quantiles due to the small number of relevant observations. Accordingly, the performance of the QIRF estimator based on asymptotics might not be ideal. Thus, we provide inferential tools for the QIRF based on the residual-based moving block bootstrap.

4.4 Residual-Based Moving Block Bootstrap

This section describes the residual-based MBB procedure and provides the bootstrap consistency for the QIRF estimator. The procedure mainly follows the bootstrap algorithm in Brüggemanna, Jentsch, and Trenkler (2016).[18] We propose the following bootstrap procedure for the inference of the QIRF.

Step 1 Choose a block length $l_b < T$ and let $N = \lceil T/l_b \rceil$ be the number of blocks needed such that $l_b N \geq T$. Draw $i_1, ..., i_N$ from a random variable uniformly distributed on the set $\{1, 2, ..., T - l_b + 1\}$. Define $(n \times l_b)$-dimensional blocks $B_{i,l_b} = (\widehat{\mathbf{u}}_i, \widehat{\mathbf{u}}_{i+1}, ..., \widehat{\mathbf{u}}_{i+l_b-1})$ where $\widehat{\mathbf{u}}_t$ is defined in Section 4.2. Bootstrap residuals $\{\mathbf{u}_t^{(*)}\}_{t=1}^{T}$ are obtained laying blocks $B_{i_1,l_b}, B_{i_2,l_b}, ..., B_{i_N,l_b}$ end-to-end together with the last $Nl_b - T$ values discarded.

[16] Recall that $\widehat{\mathbf{u}}_t = \mathbf{y}_t - \sum_{k=1}^{p} \widehat{A}_k \mathbf{y}_{t-k}$ is used in Steps 2 and 3 of the estimation procedure.

[17] $\mathbf{Q}_{i,\tau}$ depends on the density function $f_{i,t,\tau}(0)$. See, e.g., Koenker (1994, 2005) for details about the asymptotic inference in quantile regressions.

[18] The residual-based MBB applies the block bootstrap to residuals after fitting a model to capture a weak dependence structure in time series data. See, e.g., Paparoditis and Politis (2003), Ioannidis (2005) and Jentsch, Politis and Paparoditis (2015).

Step 2 Define centered bootstrap residuals $\{\mathbf{u}_t^*\}_{t=1}^T$:

$$\mathbf{u}^*_{il_b+j} := \mathbf{u}^{(*)}_{il_b+j} - \frac{1}{T-l_b+1}\sum_{k=0}^{T-l_b}\widehat{\mathbf{u}}_{j+k},$$

for $j = 1, 2, ..., l_b$ and $i = 0, 1, ..., N-1$. Set bootstrap pre-sample values $\{\mathbf{y}_t^*\}_{t=-p+1}^0 = 0$ and generate the bootstrap sample $\{\mathbf{y}_t^*\}_{t=1}^T$ according to

$$\mathbf{y}_t^* = \sum_{k=1}^p \widehat{A}_k \mathbf{y}_{t-k}^* + \mathbf{u}_t^*.$$

Step 3 Compute the bootstrap OLS estimator $\widehat{\boldsymbol{\beta}}^* = \mathrm{vec}\left(\widehat{A}_1^*, ..., \widehat{A}_p^*\right)$ based on $\{\mathbf{y}_t^*\}_{t=-p+1}^T$. Denote the bootstrap residual from the VAR model $\widehat{\mathbf{u}}_t^* = \mathbf{y}_t^* - \sum_{k=1}^p \widehat{A}_k^* \mathbf{y}_{t-k}^*$, and define $\widehat{\Sigma}_{\mathbf{u}}^* = \frac{1}{T}\sum_{t=1}^T \widehat{\mathbf{u}}_t^* \widehat{\mathbf{u}}_t^{*\top}$ and $\widehat{\sigma}^* = \mathrm{vech}\left(\widehat{\Sigma}_{\mathbf{u}}^*\right)$.

Step 4 Based on $\{\widehat{\mathbf{u}}_t^*\}_{t=1}^T$ and $\{\mathbf{y}_t^*\}_{t=-p+1}^T$, estimate QR estimator $\widehat{\boldsymbol{\Theta}}_{i,\tau}^* = \begin{bmatrix}\widehat{\boldsymbol{\theta}}_{i,\tau_U}^{*\top} & \widehat{\boldsymbol{\theta}}_{i,\tau_L}^{*\top} & \widehat{\boldsymbol{\theta}}_{i,\tau}^{*\top}\end{bmatrix}^\top$ following Steps 2 and 3 of Section 4.2.

Step 5 Using $\widehat{\boldsymbol{\beta}}^*$, $\widehat{\sigma}^*$ and $\widehat{\boldsymbol{\Theta}}_{i,\tau}^*$, compute $\widehat{IRF}^{(s)*}$ and $\widehat{QIRF}_\tau^{(s)*}$, the bootstrap version of $\widehat{IRF}^{(s)}$ and $\widehat{QIRF}_\tau^{(s)}$.

Step 6 Let the number of bootstrap be B which is large. Repeating Steps 1 through 5 B times, empirical distributions of $\widehat{IRF}^{(s)*}$ and $\widehat{QIRF}_\tau^{(s)*}$ based on the repetition provide consistent approximation of the distributions of the IRF and QIRF estimators, respectively. The lower and upper bounds of the $100\cdot(1-\alpha)\%$ confidence interval for $\widehat{IRF}^{(s)}$ is constructed using $100\cdot(1-\frac{\alpha}{2})$ and $100\cdot(\frac{\alpha}{2})$ empirical quantiles of $\widehat{IRF}^{(s)*}$. The confidence interval for $\widehat{QIRF}_\tau^{(s)}$ is obtained in the same way.

As for the choice of the block length in our empirical applications, we use the rule $l_b = \kappa T^{1/4}$ following Jentsch and Lunsford (2019). The following assumption ensures the bootstrap consistency. The assumption is implied by the existent of all moments up to order eight of $\{\mathbf{u}_t\}$ and $\sum_{k=1}^\infty k^6 \left(\alpha_{\mathbf{u}}(k)\right)^{\delta/(14+\delta)} < 0$. See Remark A.1 of Künsch (1989).

Assumption 4.3 *The innovation process $\{\mathbf{u}_t\}$ has absolutely summable cumulants up to order eight. That is, for all $j = 2, ..., 8$ and $a_1, ..., a_j \in \{1, ..., n\}$,*

$$\sum_{h_2,...,h_j=-\infty}^{\infty} \left|\mathrm{cum}_{(a_1,...,a_j)}(0, h_2, ..., h_j)\right| < \infty$$

where $\mathrm{cum}_{(a_1,...,a_j)}(0, h_2, ..., h_j)$ denotes the j-th order joint cumulant of $(u_{a_1,0}, u_{a_2,h_2}, ..., u_{a_j,h_j})$.

Lemma 4.4 follows from Theorem 4.1 and Corollary 5.2 of Brüggemanna, Jentsch, and Trenkler (2016), and it shows the validity of the residual-based MBB for the OLS and the IRF estimators.

Lemma 4.4 *Suppose Assumptions 4.1–4.3 hold. If* $l_b \to \infty$ *such that* $l_b^3/T \to 0$ *as* $T \to \infty$, *then*

$$\sup_{x\in\mathbb{R}^{\widetilde{N}}} \left| Pr^* \left(\sqrt{T} \begin{bmatrix} \widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}} \\ \widehat{\sigma}^* - \widehat{\sigma} \end{bmatrix} \leq x \right) - Pr \left(\sqrt{T} \begin{bmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\sigma} - \sigma \end{bmatrix} \leq x \right) \right| \to 0$$

and

$$\sup_{x\in\mathbb{R}} \left| Pr^* \left(\sqrt{T} \left(\widehat{IRF}_i^{(s)*} - \widehat{IRF}_i^{(s)} \right) \leq x \right) - Pr \left(\sqrt{T} \left(\widehat{IRF}_i^{(s)} - IRF_i^{(s)} \right) \leq x \right) \right| \to 0$$

in probability, where Pr^* *is the probability measure induced by the residual-based MBB and* $\widetilde{N} = pn^2 + \tilde{n}$.

The following theorem provides the validity of the residual-based MBB for the QR estimator.

Theorem 4.1 *Suppose Assumptions 4.1–4.3 and A.1–A.5 hold. If* $l_b \to \infty$ *such that* $l_b^3/T \to 0$ *as* $T \to \infty$, *then*

$$\sup_{x\in\mathbb{R}^{\widetilde{M}}} \left| Pr^* \left(\sqrt{T} \left(\widehat{\boldsymbol{\Theta}}_{i,\tau}^* - \widehat{\boldsymbol{\Theta}}_{i,\tau} \right) \leq x \right) - Pr \left(\sqrt{T} \left(\widehat{\boldsymbol{\Theta}}_{i,\tau} - \boldsymbol{\Theta}_{i,\tau} \right) \leq x \right) \right| \to 0$$

in probability, where $\widetilde{M} = 2(1 + n + 2l)$.

As $QIRF_{i,\tau}^{(s)}$ is continuously differentiable functions of $\boldsymbol{\beta}, \sigma$ and $\boldsymbol{\Theta}_{i,\tau}$, the asymptotic validity of the bootstrap extends to the QIRF estimator from Lemma 4.4 and Theorem 4.1. The following corollary summarizes the result.

Corollary 4.1 *Suppose Assumptions 4.1–4.3 and A.1–A.5 hold. If* $l_b \to \infty$ *such that* $l_b^3/T \to 0$ *as* $T \to \infty$, *then*

$$\sup_{x\in\mathbb{R}} \left| Pr^* \left(\sqrt{T} \left(\widehat{QIRF}_{i,\tau}^{(s)*} - \widehat{QIRF}_{i,\tau}^{(s)} \right) \leq x \right) - Pr \left(\sqrt{T} \left(\widehat{QIRF}_{i,\tau}^{(s)} - QIRF_{i,\tau}^{(s)} \right) \leq x \right) \right| \to 0$$

in probability.

5 Quantile Impulse Response Analysis of the US Economy

Monetary policy has been one of the most heavily studied topics in macroeconomics, and the IRF is the main tool for evaluating its policy implications. Certain financial conditions indices have received much attention recently for explaining economic fluctuations.[19] The impact of financial shocks on the whole economy is now considered dominant after the 2007–2009 financial crisis.

[19]See, e.g., Brave and Butters (2011), Matheson (2012), and Koop and Korobilis (2014).

In this section, we apply the QIRF to US macroeconomic and financial data and investigate their dynamic quantile responses to monetary policy and financial shocks. In particular, we provide the dynamic responses of Growth-at-Risk (5% quantile of CFNAI) using the QIRF.[20] We also examine the quantile responses of macroeconomic variables in a distress scenario of financial instability.

5.1 Data

The variables under study are the CFNAI, the inflation rate (CPI), the federal funds rate (FFR), and the NFCI. The CFNAI is a monthly index for US economic activity, released by the Federal Reserve Bank of Chicago. The index is a weighted average of 85 indicators of national economic activity and captures movements of the GDP growth well. For our sample period (1971Q1–2019Q4), the correlation between GDP growth rate and CFNAI is 0.73. The NFCI is a weekly index describing US financial conditions in the money market, debt and equity markets, and traditional and shadow banking systems. The index, also released by the Federal Reserve Bank of Chicago, is a weighted average of 105 indicators of national financial activity.[21]

For the sample period from January 1971 to December 2019, we use monthly data of the four variables. We measure inflation rate as the log difference of CPI, multiplied by 100. For the federal funds rate between 2009 and 2015 during which it reached the zero lower bound, we use the shadow federal funds rate estimated by Wu and Xia (2016).[22] For the NFCI, we use its monthly average. All data are from Federal Reserve Economic Data (FRED).

We estimate model (1) and (3) with the four variables. Following the Bayesian information criterion (BIC), we first estimate a VAR(3) model. The VAR model is stable since the largest eigenvalue of the companion matrix is strictly less than one. Then, we estimate the QR model of lag order 1 for lagged conditional quantile terms (i.e. $l = 1$ in (2)). For the residual-based MBB, we use a block length of $l_b = 25$ following Jentsch and Lunsford (2019).

5.2 Estimated Conditional Quantiles

Prior to investigating QIRFs, we examine how the conditional quantiles of the four variables evolved over the sample period. Figure 1 illustrates the estimated conditional 5% and 95% quantiles over 2001–2015. While the two quantiles co-move in each variable, they do not fluctuate in the same way. Due to their heterogeneous movements, the distance between the lower and upper quantiles (the dispersion of distribution) changes over time. For instance, the conditional distribution was more dispersed during the global financial crisis compared to other periods, in all variables.

However, the degree of heterogeneity in the quantile movements varies among the variables. In particular, dissimilar dynamics of the downside and upside risks are pronounced in the CFNAI. Between February 2007 and 2009, for example, the conditional 5% quantile decreased by 4.2, but

[20]In the spirit of Value-at-Risk in finance, IMF (2017) introduced Growth-at-Risk to measure macrofinancial risks to economic activity.

[21]More details about the two indices are available at *https://www.chicagofed.org/publications/cfnai/index* (CFNAI) and *https://www.chicagofed.org/publications/nfci/index* (NFCI).

[22]The data is available at *https://sites.google.com/view/jingcynthiawu/shadow-rates*.

the 95% quantile decreased by 3.5, which illustrates heterogeneous tail risk dynamics in economic activity. For other variables, the movements of the tail quantiles are not heterogeneous as much as in the CFNAI.

The summary statics for the estimated conditional 5% and 95% quantiles in Table 1 also highlight that dynamics of the downside and upside risks are the most disparate in the CFNAI. The correlation coefficient between the two quantiles is smaller for the CFNAI. Moreover, its left tail shows much greater time variation than its right tail. Measured by standard deviation, the variation of the 5% quantile is 1.4 ($= \frac{0.97}{0.69}$) times as large as that of the 95% quantile. These results suggest substantial heterogeneity in the quantile response of the CFNAI.

For the CPI, the correlation coefficient between its 5% and 95% quantiles is smaller too: 0.63. But, their time variations are not as substantially different as they are in the CFNAI. The variation of the 95% quantile is 1.1 ($= \frac{0.26}{0.23}$) times that of the 5% quantile. Accordingly, a certain degree of heterogeneity is expected in the quantile response of CPI, but not as much as that of the CFNAI.

On the contrary, the quantile response of the financial variables (FFR and NFCI) is expected to be much less heterogeneous. As seen in Figure 1(c) and (d), the correlation between their conditional 5% and 95% quantiles is very strong: their correlation coefficients are close to one. Their time variations of the left and right tails are not substantially different as in the CFNAI. The standard deviation of the 95% quantile is 1.2 times as large as the 5% quantile standard deviation in the variables.

5.3 Quantile Impulse Response Analysis

We now construct QIRFs based on the model estimates as in (7). Since our QR model is a reduced form, we use a mean-based VAR model with Cholesky restrictions to identify a structural shock. Under the recursive identification, a variable is affected by the contemporaneous shocks to other variables if the variable is ordered after them, but not affected if ordered before them. Thus, slowly moving variables are ordered before fast-responding variables. The ordering of our variables (CFNAI, CPI, FFR, and NFCI, standard in the literature) implies that the NFCI instantly responds to all structural shocks. But economic activity (CFNAI) does not contemporaneously responds to shocks other than a shock to itself.

5.3.1 QIRF to a Monetary Policy Shock

First, we present how the conditional quantile of the variables responds to a monetary policy shock. Figure 2 is the QIRF to an expansionary monetary policy shock (-25bp) at five quantiles (5%, 16%, 50%, 84%, and 95%) as well as the IRF.[23]

Against the expansionary monetary policy shock, the median and mean of CFNAI show similar dynamics. However, the QIRF clearly illustrates that its tail responses are highly heterogeneous. The monetary policy shock significantly increases the 5% quantile (i.e. the shock effectively reduces

[23]Figures 7 and 8 provide bootstrap confidence intervals of the QIRF to monetary policy and financial shocks, respectively.

downside risk to growth). On the contrary, the response of the 95% quantile is close to zero implying upside risks are much less affected. These results highlight the monetary policy shock not only shifts the economic activity distribution but also significantly changes its shape, which can not be learned from the conventional IRF. Since quantiles at the left tail are more responsive, the volatility of economic activity decreases while its location shifts to the right. Loria, Matthes, and Zhang (2019) also find that a monetary policy shock affects lower quantiles of GDP growth more than its upper quantiles using local projections. Unlike their result, however, our QIRF does not show the reverse effect of the shock at longer horizons.[24]

In addition, the 5% quantile of economic activity displays a more persistent response to a monetary policy shock than the mean IRF. It takes 26 months for the 5% quantile response of CFNAI to dissipate by half from its peak, while it takes only 11 months for the mean response. These dynamics illustrate that the effect of a monetary policy shock is stronger on Growth-at-Risk in terms of persistence as well as magnitude.

For CPI, the quantile response, as in the mean response, exhibits the so-called price puzzle: a decrease (increase) in the price level in response to an expansionary (contractionary) monetary policy shock.[25] Though the QIRF reveals certain amount of heterogeneity across quantiles, the difference between the 5% and 95% quantiles is small (less than 0.1 percentage point in annualized rate).

An expansionary monetary policy shock leads to looser financial conditions (a decrease in the NFCI) and has persistent effects on the FFR. The shock reduces volatility of the FFR and the NFCI, as the upper quantiles decline slightly more than the lower quantiles. However, the financial variables show much more homogeneous quantile responses, especially compared to economic activity. We interpret their QIRFs as relatively stable responses of the Fed and financial markets: their response to a one-off monetary shock does not significantly increase tail risks to the FFR and financial conditions.

5.3.2 QIRF to a Financial Shock

Figure 3 plots the QIRF and IRF to a financial shock. As in the case of a monetary policy shock, the response of the CFNAI is highly heterogeneous across quantiles. An adverse financial shock shifts the conditional distribution of economic activity to the left, but the left tail quantiles decrease substantially more than the right tail. This empirical result is in line with Adrian, Boyarchenko, and Giannone (2019): economic growth is vulnerable to deteriorating financial conditions. They argue that downside GDP vulnerability (away from the steady state) can be explained by amplification mechanisms in the financial sector, such as the feedback loops mechanism by Brunnermeier and Sannikov (2014). While Adrian, Boyarchenko, and Giannone (2019) investigate the impact of

[24] The estimates of Loria, Matthes, and Zhang (2019) suggest the effect of a monetary policy shock reverse in the medium term. For example, they claim that a contractionary monetary policy shock has a positive effect on the GDP growth after 10–15 quarters.

[25] This price puzzle is attributable to identification of structural shocks. See, e.g., Sims (1992), Christiano, Eichenbaum, and Evans (1999), and Hanson (2004) for further discussion of the price puzzle.

financial conditions on the following period's Growth-at-Risk, our approach goes further describing the evolution of the quantile response over time.

Meanwhile, the impact of a financial shock on Growth-at-Risk is not as persistent as that of a monetary policy shock. The magnitudes of the IRF and 5% QIRF decay by half from its trough in 11 months and 10 months, respectively. This suggests a financial shock has a relatively acute effect on downside risks to economic activity.

The CPI also shows a certain degree of heterogeneity in its quantile response to a financial shock: its 95% quantiles increases more than the mean response, but its 5% quantile decreases. While the shock increases upside and downside risks to inflation, the magnitude of the quantile response is not so significant (between -0.4% and +0.3% in annualized rate) as in the case of a monetary policy shock.

For the FFR, the 95% quantile increases much more than other quantiles in response to a financial shock. The 95% quantile increases by up to 28 basis points and stays positive for 5 months, and the IRF also shows positive response for 4 months after the shock. Such responses are counterintuitive as accommodative monetary policies are expected against a financial shock. This upside risk to the FFR seems to be driven by observations in the 1970s. With a sample period from January 1981 to December 2019, the initial increases in both the 95% quantile and mean responses dramatically disappear. These empirical results may attribute to the change in monetary policy practice or the change in how the other variables respond to shocks.[26]

Figure 3(d) suggests that financial conditions deteriorate (increases in the NFCI) further for a few months after the intial financial shock. For all of the five quantiles considered, the response of NFCI is quite homogeneous along the mean response. That is, a financial shock shifts the distribution of the NFCI with little change in its volatility.

5.4 Growth-at-Risk Dynamics during the Global Financial Crisis

In this section, we examine how severely the conditional quantile of economic activity was affected by a series of shocks during a particular historical episode: the 2007–2009 GFC. We pay close attention to the conditional 5% quantile of the CFNAI, considered as *Growth-at-Risk* in this paper. The dynamic response of the quantile to a one-off shock is described by the QIRF as in Section 5.3. We now investigate the cumulative effects of a set of shocks on the quantile using QIRFs.

We provide an answer to the following question: what are the impacts of financial and monetary policy shocks concerning the GFC on the downside and upside risks to growth? As seen in the previous section, the lower quantiles of economic activity are much more responsive than the median or upper quantiles. Thus, quantitative assessment of the downside risks during the recession period is of great importance. In the following exercises, we first study the impacts of financial shocks on the risks during the financial distress period of August 2007–June 2009. We then examine the effects of ensuing unconventional monetary policy measures using the identified monetary policy

[26]See, e.g., Primiceri (2005) and Sims and Zha (2006) for a discussion of the change in monetary policy rules. We leave more rigorous investigation of these causal explanations to future research.

shocks from July 2009 to December 2015.

The assessment is carried out in a similar way to historical decomposition in the VAR analysis.[27] Suppose, for example, we want to quantify the effects of structural shocks from time 1 to T, $\{\boldsymbol{\epsilon}_t\}_{t=1}^T$, on the quantile response. Let $QIRF_{i,\tau}^{(s)}|\boldsymbol{\epsilon}_t$ denote the τ-quantile response of the i-th variable to $\boldsymbol{\epsilon}_t$ at horizon s. The QIRF estimates quantile dynamics against a shock at a specific time. To construct cumulative impacts of $\{\boldsymbol{\epsilon}_t\}_{t=1}^T$ on quantiles, we aggregate quantile responses of each of the shocks accounting for their dynamics. That is, the cumulative impacts of $\{\boldsymbol{\epsilon}_t\}_{t=1}^T$ on the τ-quantile of the variable i at time s is calculated as

$$\begin{cases} \sum_{t=1}^{s-1} QIRF_{i,\tau}^{(s-t)}|\boldsymbol{\epsilon}_t, & \text{for} \quad s \leq T, \\ \sum_{t=1}^{T} QIRF_{i,\tau}^{(s-t)}|\boldsymbol{\epsilon}_t, & \text{for} \quad s > T. \end{cases}$$

For estimation, we replace $QIRF_{i,\tau}^{(s-t)}$ and $\boldsymbol{\epsilon}_t$ with their respective estimators, $\widehat{QIRF}_{i,\tau}^{(s-t)}$ and $\widehat{\boldsymbol{\epsilon}}_t$.

The cumulative impacts of financial shocks from August 2007 to June 2009 on the conditional quantile of the CFNAI are illustrated in Figure 4(a).[28] Financial shocks during the period substantially increased the downside risk, but the upside risk was mildly affected. Over 2008–2010, the financial shocks decreased the 5% quantile of the CFNAI by 1.5 on average. However, the impacts of those shocks on the 95% quantile were much less; the conditional 95% quantile was lowered by 0.8 on average over the same period.

We then answer the following counterfactual question: what would have happened if the financial shocks were shut down? Figure 4(b) and (c) describe the counterfactual paths of the conditional 5% and 95% quantiles based on the cumulative impacts. The figures highlight that the increase in downside risks during the GFC was mainly attributable to financial shocks. The 5% quantile is much lower than the counterfactual 5% quantile. In the absence of the financial shocks, the downside risks would have been moderate. The average of the counterfactual 5% quantile over 2008–2009, -1.3, is a little less than the average of the 5% quantile over 1971–2007, -1.0. Without the financial shocks, the 95% quantile would have been higher. But the difference between the 95% quantile and its counterfactual is narrower compared to the case of the 5% quantile, suggesting asymmetric impact of the financial shocks on the economic activity distribution.

We perform the same exercise to quantify the effects of unconventional monetary policy measures implemented by the Fed in response to the financial crisis. Following Wu and Xia (2016), structural shocks to the FFR from July 2009 to December 2015 are used for assessment of the monetary policy.[29] Figure 5(a) demonstrates the cumulative impacts of those monetary policy shocks on the quantile of the CFNAI. The figure emphasizes the effectiveness of unconventional monetary policy for reducing the downside risks to growth. Over 2010–2015, the monetary policy increased its 5%

[27]See Kilian and Lütkepohl (2017) for details about historical decomposition in the VAR model.

[28]In August 2007, BNP Paribas halted redemptions on three investment funds because it could not value their holdings, which marked the start of the financial crisis. The National Bureau of Economic Research (NBER) identified June 2009 as the end of the recession associated with the financial crisis.

[29]While Wu and Xia (2016) study the effects of unconventional policy measures using monetary policy shocks from July 2009 to December 2013, we extend the period to December 2015, the end of the zero lower bound period.

quantile by 0.4, on average, but its 95% quantile was hardly affected.

The counterfactual paths in Figure 5(b) and (c) describe the heterogeneous effects of the unconventional measures on economic growth. Without the measures, downside risks would have been higher: the averages of the 5% quantile and its counterfactual over 2011–2015 are -0.9 and -1.3, respectively. The unconventional monetary policy consistently reduced the downside risks over the period. In contrast, the 95% quantile and its counterfactual are almost indistinguishable.

5.5 Quantile Responses under a Hypothetical Distress Scenario

Since extreme shocks leave long-lasting scars on the economy, policy makers require effective tools for testing macroeconomic resilience in a stress scenario. In this section, we conduct a hypothetical analysis to examine the quantile response in a distress scenario where a series of unfavorable tail events follow an initial shock.

The QIRF defined in Section 3.1 measures the impact of a shock assuming realizations of endogenous variables after the shock are their conditional mean. Here, we replace the conditional mean-time path with a stress scenario in which unfavorable tail events for endogenous variables materialize, then construct the quantile responses for a stress testing. This exercise is comparable to the stress testing of Chavleishvili and Manganelli (2019) and the QIRFs of Montes-Rojas (2019) in that the impacts of continuing realizations of tail events are studied.

In this exercise, we examine how the economy responds to a distress scenario initiated by a financial shock. In the scenario, an initial financial shock (one standard deviation shock) is followed by realizations of the NFCI at its conditional 95% quantile for six months. After that, realizations of its conditional mean are assumed for the NFCI. For variables other than the NFCI, realizations of the conditional mean are assumed after the initial shock. The distress scenario describes a rapid deterioration of financial conditions. As in the QIRF, the quantile response under this scenario can be expressed in a recursive manner, and its derivation is relegated to Online Appendix Section C.

Figure 6 describes the mean and quantile responses under the hypothetical scenario. First of all, NFCI displays non-stationary (locally explosive) behavior when it stays at its conditional 95% quantile. The distress scenario substantially shifts the distribution of NFCI to the right; financial conditions deteriorate rapidly. The mean of NFCI increases by up to 0.48. While a one-time financial shock increases its mean by up to 0.28 as in Figure 3(d), the continuous tail events lead to acute financial distress.

As a result of the financial instability, the economy suffers a severe downturn with substantial downside risk. The economic activity distribution shifts to the left greatly, and its left tail decreases much more than the median or upper quantiles. The 5% quantile of CFNAI decreases by up to 0.44, whereas its mean and 95% quantile decline by 0.28 and 0.24, respectively. These results may suggest that a financial crisis can develop in a short period of time.

Under the scenario, the initial response of the FFR displays certain amount of heterogeneity across quantiles. But, its response becomes more homogeneous at longer horizons; the considered five quantiles of FFR decrease by more than 30bp three years after the initial shock.

6 Conclusion

This article studies Quantile Impulse Response Function (QIRF) theory and its applications in macroeconomics and finance. Our QR model provides a multi-equation system with autoregressive specifications accounting for important dynamics of distributional evolution. The QIRF complements the conventional IRF providing a more complete shock-response mechanism.

The comprehensive QIRF analysis of the US economy provides evidence of a strong heterogeneity in the responses of economic activity across its distribution. Against monetary policy and financial shocks, the downside risks to growth are more responsive than the median or upside risks. We also quantitatively assessed the evolution of macroeconomic tail risks during the 2007–2009 global financial crisis and in a hypothetical scenario where financial conditions rapidly deteriorate. Considering tremendous implications of extreme events, such as market booms and crashes, our QR model and QIRF provide useful tools for dynamic risk management and policy analysis.

References

Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109, 1263–1289.

Adrian, T, & Brunnermeier, M. K. (2016). CoVaR. *The American Economic Review*, 106, 1705–1741.

Adrian, T., Grinberg, F., Liang, N., & Malik, S. (2018). The term structure of Growth-at-Risk. *IMF Working Paper.*

Brave, S. A., & Butters, R. A. (2011). Monitoring financial stability: A financial conditions index approach. *Economic Perspectives*, Federal Reserve Bank of Chicago, 35(1), 22–43.

Brunnermeier, M. K., & Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104, 379–421.

Brüggemanna, R., Jentsch, C., & Trenkler, C. (2016). Inference in VARs with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 191, 69-85.

Chang, Y., Kaufmann, R. K., Kim, C. S., Miller, J. I., Park, J. Y., & Park, S. (2020). Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. *Journal of Econometrics*, 214(1), 274-294.

Chang, Y., Kim, C. S., & Park, J. Y. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics*, 192(1), 152-167.

Chang, Y., Kim, S., & Park, J. Y. (2021). Effects of Macroeconomic Shocks on Income Distribution, *Working paper.*

Chang, Y., Miller, Z., & Park, J. Y. (2021). What Drives Temperature Anomalies? An Econometric Analysis Using Functional Autoregression, *Working paper.*

Chavleishvili, S., & Manganelli, S. (2019). Forecasting and stress testing with quantile vector autoregression. *ECB Working Paper Series No. 2330.*

Chevapatrakul, T., Kim, T.-H., & Mizen, P. (2009). The Taylor principle and monetary policy approaching a zero bound on nominal rates: Quantile regression results for the United States and Japan. *Journal of Money, Credit and Banking*, 41, 1705–1723.

Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end?. In *Handbook of Macroeconomics*, 1, pp. 65–148.

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

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

Han, H., Jung, W., & Lee, J. H. (2019). Estimation and inference of quantile impulse response functions by local projections: With applications to VaR dynamics. *Working paper.*

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

Hanson, M. S. (2004). The "price puzzle" reconsidered. *Journal of Monetary Economics*, 51, 1385–1413.

IMF (2017). Financial conditions and growth at risk. In *Global Financial Stability Report (October), Chapter 3.* Washington, DC: International Monetary Fund.

Ioannidis, E. E. (2005). Residual-based block bootstrap unit root testing in the presence of trend breaks. *The Econometrics Journal*, 8, 323-351.

Jentsch, C., Politis, D. N., & Paparoditis, E. (2015). Block bootstrap theory for multivariate integrated and cointegrated processes. *Journal of Time Series Analysis*, 36, 416-441.

Jentsch, C., & Lunsford, K. G. (2019). The dynamic effects of personal and corporate income tax changes in the United States: Comment. *American Economic Review*, 109(7), 2655-2678.

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

Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis.* Cambridge University Press.

Kim, T.-H., Lee, D. J., & Mizen, P. (2019). Impulse response analysis in conditional quantile models and an application to monetary policy, *Working paper.*

Koenker, R. (1994). Confidence intervals for regression quantiles. In *Asymptotic statistics.* Physica-Verlag, Heidelburg, pp. 349–359.

Koenker, R. (2005). *Quantile regression*. Cambridge University Press.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33–50.

Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, 101, 980–990.

Koenker, R., & Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, 12, 793–813.

Koop, G., & Korobilis, D. (2014). A new index of financial conditions. *European Economic Review*, 71, 101–116.

Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, 74, 119–147.

Künsch, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17, 1217-1241.

Li, G., Li, Y., & Tsai, C.-L. (2015). Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association*, 110, 246–261.

Linton, O. & Xiao, Z. (2018). Quantile regression applications in finance. In *Handbook of quantile regression*. CRC Press, pp. 381–407.

Loria, F., Matthes, C., & Zhang, D. (2019). Assessing Macroeconomic Tail Risk. *Federal Reserve Bank of Richmond Working Paper No. 19-10*.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin.

Matheson, T. D. (2012). Financial conditions indexes for the United States and euro area. *Economics Letters*, 115, 441–446.

Montes-Rojas, G. (2019). Multivariate quantile impulse response functions. *Journal of Time Series Analysis*, 40, 739–752.

Mumtaz, H., & Surico, P. (2015). The transmission mechanism in good and bad times. *International Economic Review*, 56, 1237–1259.

Paparoditis, E., & Politis, D. N. (2003). Residual-based block bootstrap for unit root testing. *Econometrica*, 71, 813-855.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72, 821–852.

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36, 975–1000.

Sims, C. A., & Zha, T. (2006). Were there regime switches in US monetary policy?. *American Economic Review*, 96, 54–81.

Stock, J. H., & Watson, M. W. (2002). Has the business cycle changed and why?. *NBER Macroeconomics Annual*, 17, 159–218.

Xiao, Z. (2009). Quantile cointegrating regression. *Journal of Econometrics*, 150, 248–260.

Xiao, Z. (2012). Time series quantile regressions. In *Time series analysis: methods and applications.* Elsevier, pp. 213–257.

Xiao, Z., & Koenker, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association*, 104, 1696–1712.

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

Wu, J. C., & Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48, 253–291.

7 Table and Figures

Table 1: Summary Statistics for the Estimated Conditional 5% and 95% Quantiles

	CFNAI	CPI	FFR	NFCI
Correlation coefficient between $\widehat{q}_{\mathbf{y}_t}(0.05)$ and $\widehat{q}_{\mathbf{y}_t}(0.95)$	0.71	0.63	0.99	0.998
Standard deviation of $\widehat{q}_{\mathbf{y}_t}(0.05)$	0.97	0.23	3.96	0.92
Standard deviation of $\widehat{q}_{\mathbf{y}_t}(0.95)$	0.69	0.26	4.57	1.10

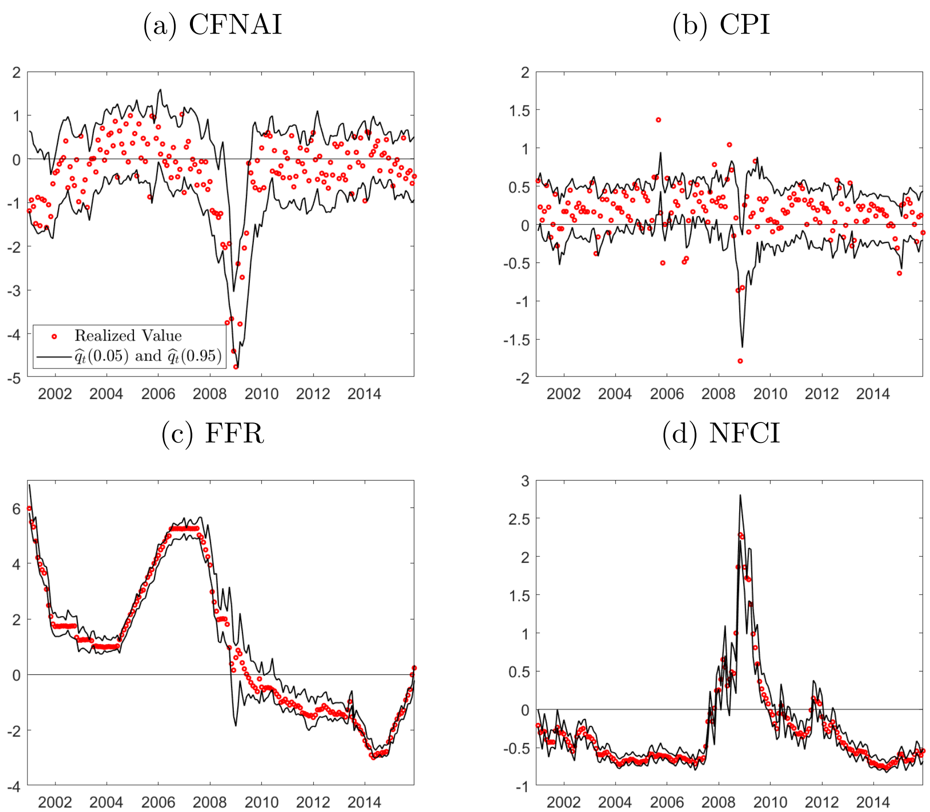


Figure 1: The Estimated Conditional 5% and 95% Quantiles over 2001–2015

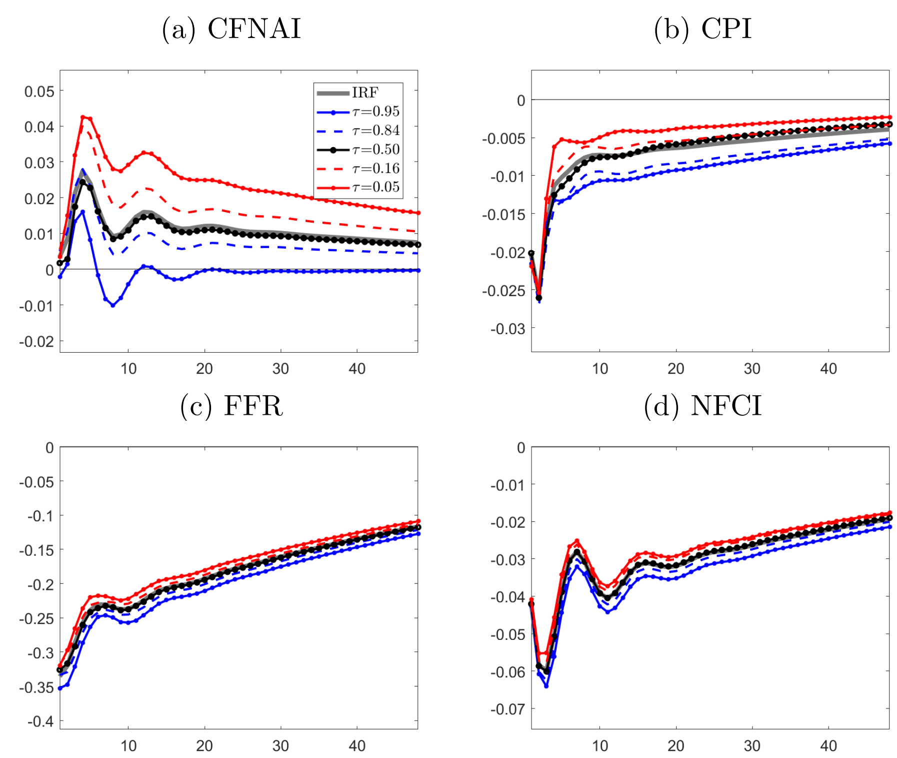


Figure 2: QIRF to an Expansionary Monetary Policy Shock (-25bp)

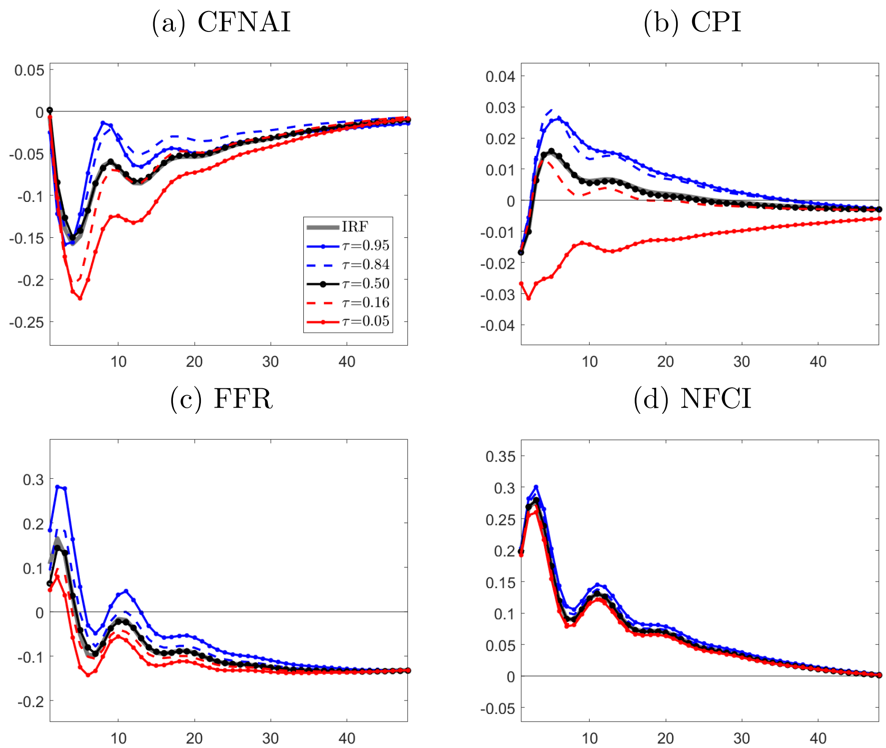


Figure 3: QIRF to a Financial Shock (one standard deviation shock)

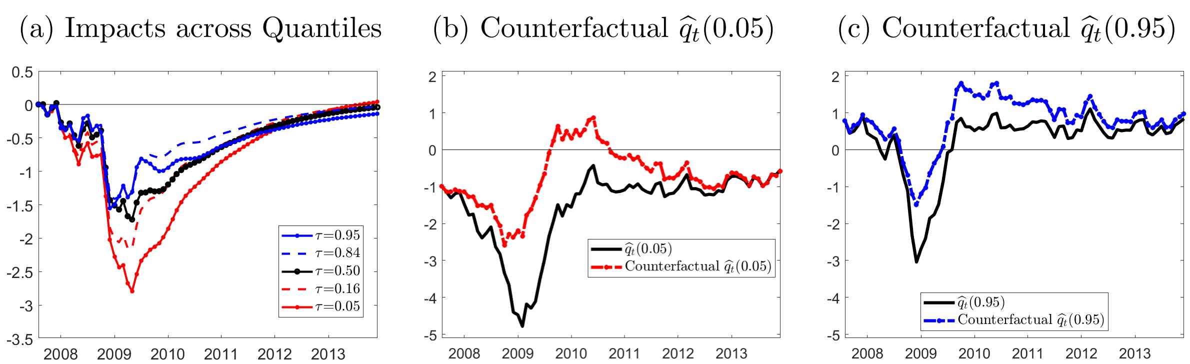


Figure 4: The Impacts of Financial Shocks (Aug.2007–Jun.2009) on CFNAI

Note: The counterfactual path describes the quantile path of CFNAI if the financial shocks were shut down.

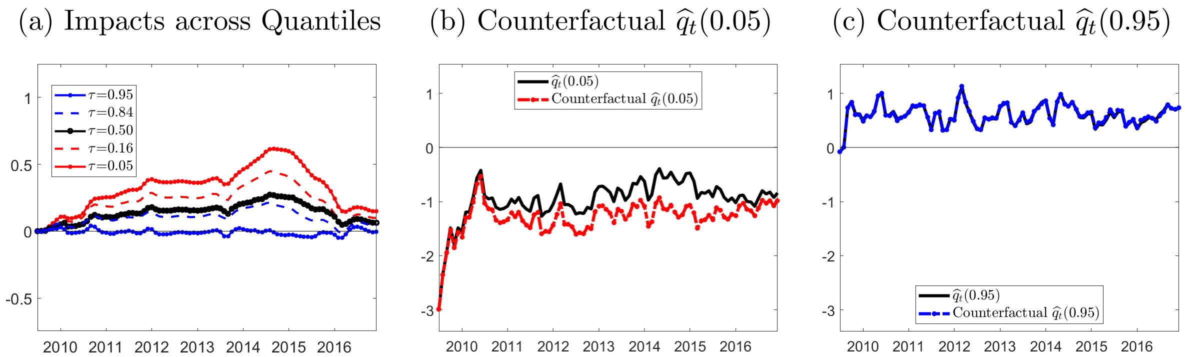


Figure 5: The Effects of Unconventional Monetary Policy (Jul.2009–Dec.2015) on CFNAI

Note: The counterfactual path describes the quantile path of CFNAI if the monetary policy shocks were shut down.

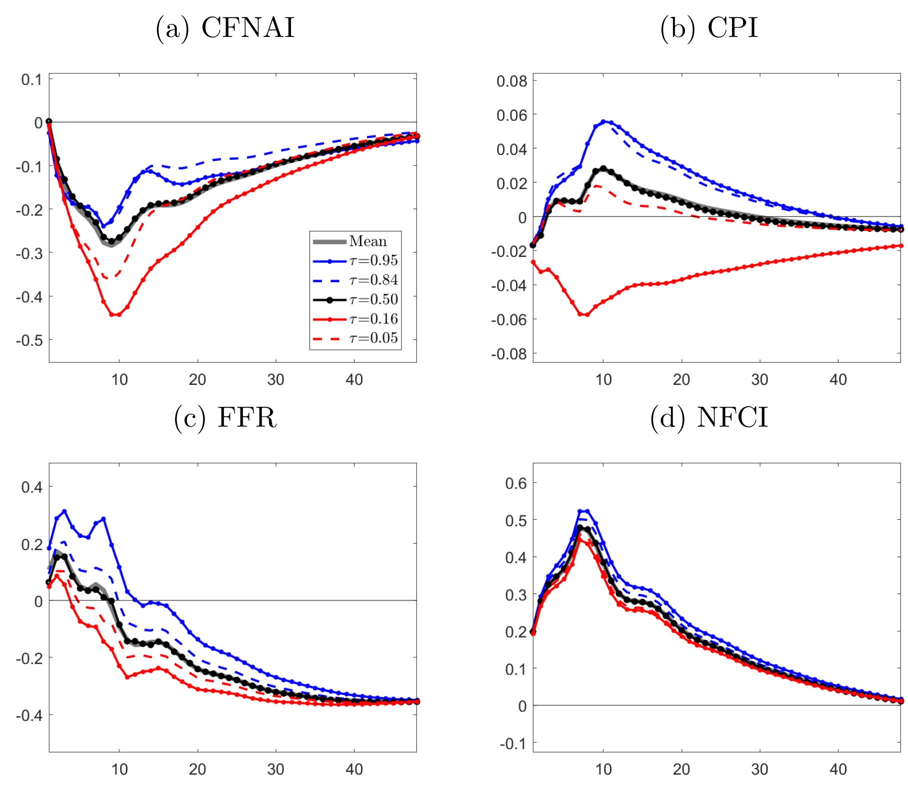
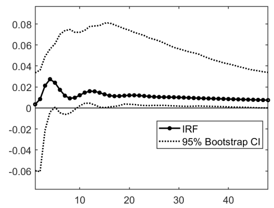


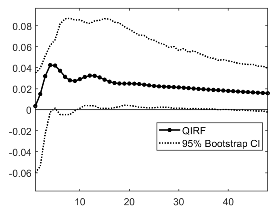
Figure 6: Responses of Variables Under the Distress Scenario

Note: After a shock of one standard deviation to the NFCI, realizations of the conditional mean are assumed for variables other than the NFCI. For the NFCI, realizations of the conditional 95% follow for the first 6 months after the shock. Afterwards, realizations of the conditional mean are assumed for the variable.

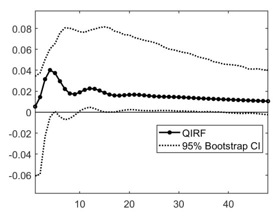
(a) CFNAI at the mean (b) CPI at the mean (c) FFR at the mean (d) NFCI at the mean



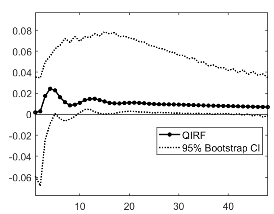
(e) CFNAI at $\tau = 0.05$ (f) CPI at $\tau = 0.05$ (g) FFR at $\tau = 0.05$ (h) NFCI at $\tau = 0.05$



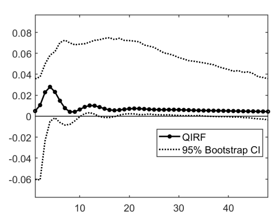
(i) CFNAI at $\tau = 0.16$ (j) CPI at $\tau = 0.16$ (k) FFR at $\tau = 0.16$ (l) NFCI at $\tau = 0.16$



(m) CFNAI at $\tau = 0.50$ (n) CPI at $\tau = 0.50$ (o) FFR at $\tau = 0.50$ (p) NFCI at $\tau = 0.50$



(q) CFNAI at $\tau = 0.84$ (r) CPI at $\tau = 0.84$ (s) FFR at $\tau = 0.84$ (t) NFCI at $\tau = 0.84$



(u) CFNAI at $\tau = 0.95$ (v) CPI at $\tau = 0.95$ (w) FFR at $\tau = 0.95$ (x) NFCI at $\tau = 0.95$

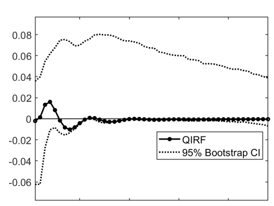


Figure 7: QIRF and IRF to a monetary policy shock (-25bp) with 95% bootstrap confidence interval using residual-based MBB, the number of bootstrap draws is 1,000 and the block length is 25.

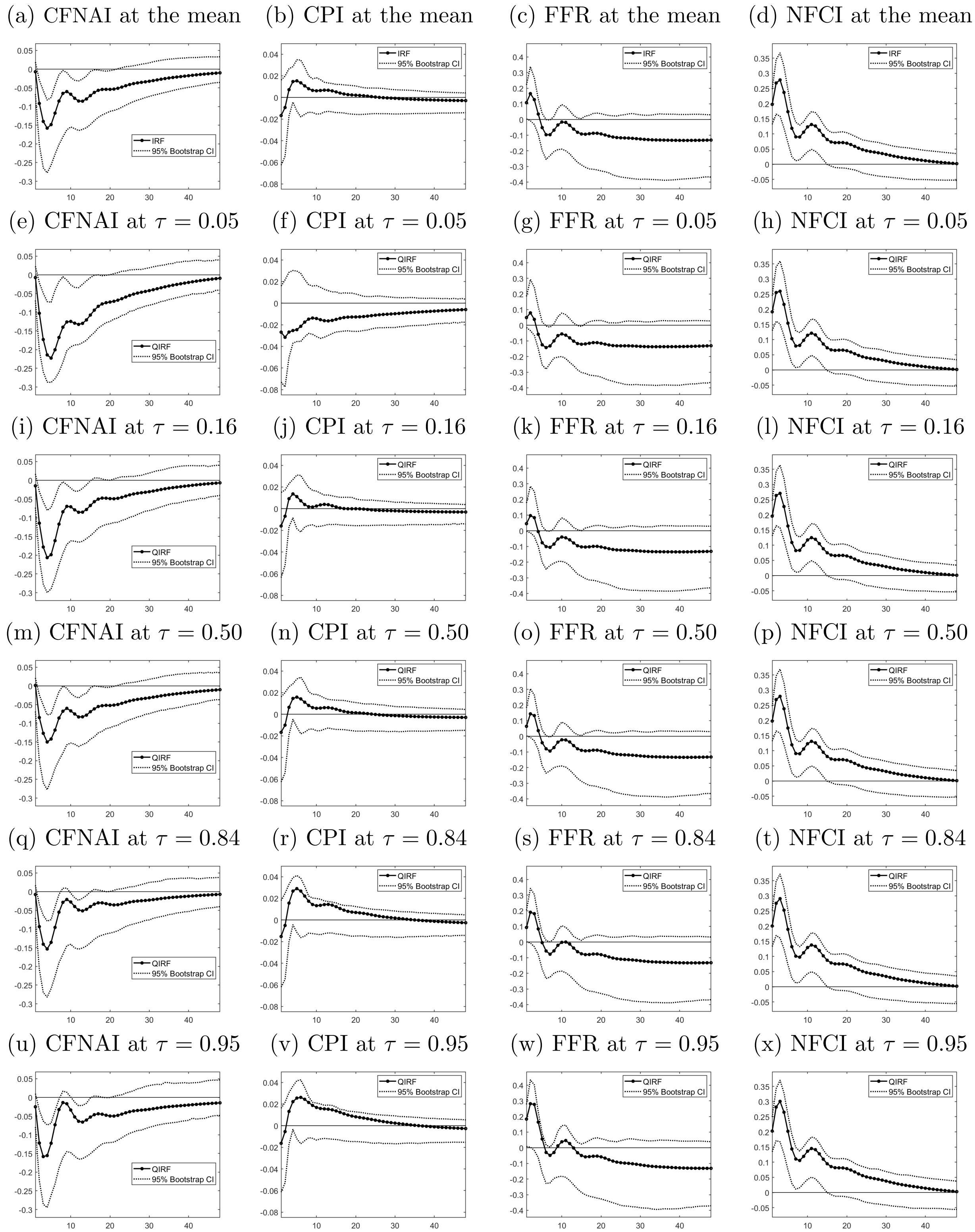


Figure 8: QIRF to a financial shock (one standard deviation shock) with 95% bootstrap confidence interval using residual-based MBB, the number of bootstrap draws is 1,000 and the block length is 25.