Estimation and Inference of Quantile Impulse Response Functions by Local Projections: With Applications to VaR Dynamics

Heejoon Han* Whayoung Jung† Ji Hyung Lee‡

August 2020

Abstract

This article investigates the estimation and inference of quantile impulse response functions. We propose a new estimation method using the idea of local projections by Jordà (2005). We establish consistency and asymptotic normality of the estimator, thereby enabling asymptotic inference. We also consider the confidence interval construction based on the stationary bootstrap and prove its consistency. Confirmatory simulation results and empirical practices on Value-at-Risk dynamics are provided.

Keywords: Quantile Impulse Response, Local Projection, Stationary Bootstrap, Value-at-Risk.

JEL classification: C22

1 Introduction

Quantile methods have attracted increasing attention in economics and finance. Unlike the traditional mean-based analysis, quantile analysis enables researchers to study heterogeneous effects of independent variables on different quantiles of an outcome distribution. The effect of a shock associated with tail events or tail co-dependence among financial variables is important for risk analysis. Recent developments in time series quantile analysis include the quantilogram analysis by Linton and Whang (2007), Han et al. (2016) and Lee et al. (2019), CoVaR of Adrian and Brunnermeier (2016), the conditional autoregressive value at risk (CAViaR) of Engle and Manganelli (2004), and the quantile autoregression (QAR) of Koenker and Xiao (2006), to name a few. See also Baruník and Kley (2019), Davis and Mikosch (2009) and Li et al. (2015) for measuring quantile dependence and tail dependence.

Impulse response function (IRF) analysis has been a standard tool in macroeconomics and finance since the seminal paper by Sims (1980). In a variety of multivariate models, IRFs provide a comprehensive picture of shock-response mechanisms. Many vector autoregressive (VAR) studies have developed identification, estimation and econometric inference of IRFs.

*Department of Economics, Sungkyunkwan University, heejoonhan@skku.edu.

†Korea Capital Market Institute, junghy0103@kcmi.re.kr.

‡Department of Economics, University of Illinois, jihyung@illinois.edu.

In this article, we study the quantile impulse response function (QIRF) that can measure the effect of a shock on the response variables at different quantiles. Since a shock to one financial institution can propagate to other financial institutions or to the market, a proper tool would be a version of VAR models. However, the impulse response can be substantially different between downside and upside risks to the market. One unit of the market-wide shock would also have different implications for major financial institutions in good or bad times. The QIRF can capture these heterogeneous dynamic responses across different economic conditions, hence is a useful tool for the risk analysis conditional on the market information.

We propose a new easy-to-use QIRF estimation using the idea of local projections by Jordà (2005). We show that the QIRF estimation based on our method effectively describes the true QIRF, hence accommodating dynamic analysis of Value-at-Risk (VaR). We also provide valid econometric inferential tools based on both asymptotics and the stationary bootstrap (Politis and Romano, 1994). The bootstrap consistency in this framework is provided, which is of its own interest. Simulation evidence shows that both asymptotic and bootstrap confidence intervals have proper coverage probabilities. However, for relatively longer horizons, the bootstrap confidence interval works better because the bootstrap method avoids the estimation of a density-like nuisance parameter. In the quantile regression literature, estimation of this parameter has been difficult especially at tails.

Closely related to this paper is White, Kim, and Manganelli (2015, WKM henceforth), who proposed a so-called pseudo-QIRF. However, the pseudo-QIRF typically underestimates the effect of a shock on quantiles, since the dynamic evolution of volatility is not accounted for in its construction. We also choose to use the VAR for VaR model as WKM that directly estimates conditional quantile, instead of fully specifying the DGP, and adopt the idea of direct forecasting for multi-step prediction of conditional quantile. We confirm via simulations that the local projection QIRF estimation effectively approximates the true quantile responses, outperforming the pseudo-QIRF.

In the empirical application, we consider the stock return series of 61 US financial institutions and, using the local projection QIRF, examine a dynamic reaction of each financial institution's 1% and 5% VaR when there is a shock to the market. Following Acharya et al. (2017), we categorize 61 financial institutions into the following four sectors; Depositories, Other, Insurance and Broker-Dealers. The sectoral averages of QIRFs for 1% VaR show that the average response of Broker-Dealers tends to be the largest up to three weeks, whereas that of Depositories becomes the smallest after about two weeks. Moreover, the local projection QIRFs generally exhibit substantial fluctuations, whereas the pseudo-QIRFs monotonically converge to zero. For example, the average response of Broker-Dealers for 1% VaR reaches its maximum in two weeks instead of gradually decreasing.

Recently, there has been much attention on dynamic quantile analysis in economics literature. Chavleishvili and Manganelli (2019), Kim et al. (2019), and Montes-Rojas (2019)[1] consider

[1]Montes-Rojas (2019) investigated some benefits of local projections in his vector autoregressive quantile (VARQ) models.

QIRF based on their own multivariate quantile regression models with macroeconomic and financial variables. Moreover, the Growth-at-Risk (GaR) approach developed by IMF (2017) links current financial conditions to quantiles of future growth outcomes. Even though the current paper mainly focuses on financial return data and their VaR dynamics, this approach could be modified to accommodate macroeconomic time series, see Jung and Lee (2019). Please also see Loria et al. (2019) for related but different macroeconomics applications of QIRFs.

A recent statistical literature has also explored the idea of mutivariate conditional quantiles. See, e.g., Kong and Mizera (2012), Hallin et al. (2010), Paindaveine and Šiman (2011), and Carlier et al. (2016). Since there is no consensue on how to define multivariate quantiles, they propose new interesting definitions using a directional vector or optimal transport. In this paper, we construct multivariate quantiles by stacking the univariate conditional quantiles following WKM.

In Section 2, we introduce the definition and estimation of QIRF. In Section 3, we explore some prototypical financial volatility models, and study QIRFs of the models. In particular, we provide some intuition why the local projection can effectively estimate QIRFs of financial return data. Section 4 derives the asymptotic properties of QIRF estimators, and provides the asymptotic and bootstrap inferential methods. In Section 5, we evaluate the performance by a set of Monte Carlo simulations. Section 6 provides empirical applications to Value-at-Risk dynamics of financial time series, and Section 7 concludes.

2 QIRF Definition and Estimation

In this section, we introduce the definition of QIRF and propose a new method to estimate it. We first briefly explain the generalized impulse response function and the local projections in the literature of VAR models because these are closely related to our method. In VAR models, impulse responses are typically defined as the change in $\mathbf{y}_{t+s} = (y_{1t+s}, y_{2t+s}, ..., y_{nt+s})^\top$ caused by a shock at time t, which can be derived by the Wold decomposition. Alternatively, impulse responses can be defined as the difference between the forecast of $\mathbf{y}_{t+s}$ with a shock at time t and that without a shock at time t, which is the generalized impulse response function by Koop et al. (1996);

$$IRF^{(s)} := E\left[\mathbf{y}_{t+s}|\mathbf{u}_t = \boldsymbol{\delta}^0; \mathcal{F}_{t-1}\right] - E\left[\mathbf{y}_{t+s}|\mathbf{u}_t = \mathbf{0}; \mathcal{F}_{t-1}\right]$$

where $\mathbf{u}_t$ is the vector of reduced-form errors. Similarly, we define the QIRF as the difference between the forecast of conditional quantile with a shock at time t and that without a shock.

In estimating QIRF, we adopt the idea of direct forecasting to obtain multi-step ahead forecasts of the conditional quantile, which is similar to the local projections by Jordà (2005). Given a VAR(p) model, Jordà (2005) denotes the collection of the following regressions

$$\mathbf{y}_{t+s} = \mathbf{c}^{(s)} + \mathbf{B}_1^{(s)}\mathbf{y}_t + \mathbf{B}_2^{(s)}\mathbf{y}_{t-1} + \cdots + \mathbf{B}_p^{(s)}\mathbf{y}_{t-p+1} + \mathbf{u}_{t+s}^{(s)}, \quad \text{for} \quad s = 1, 2, \cdots, S \tag{1}$$

as local projections, from which the impulse responses are estimated by

$$\widehat{IRF}^{(s)} = \widehat{\mathbf{B}}_1^{(s)} \boldsymbol{\delta}^0.$$

Local projections are linear approximations and correspond to direct forecasting. For multi-step ahead forecasting, the direct forecasting procedure is known to be robust to model misspecification.

We consider the following set-up to describe quantiles of multiple variables. Given the natural filtration $\{\mathcal{F}_t\}_{t=-\infty}^{\infty}$ and $\tau_j \in (0,1)$, $q_{it}(\tau_j)$ represents the conditional τ_j quantile of a time series y_{it} for $i = 1, 2, ..., n$ such that $P(y_{it} \leq q_{it}(\tau_j) | \mathcal{F}_{t-1}) = \tau_j$. $q_{it}(\tau_j)$ is $\mathcal{F}_{t-1}$-measurable, and depends on the quantile level τ_j determined by the researcher. It should be noted that we can choose a different quantile level τ_j for each y_{it} with $q_{it}(\tau_j)$ for implementation. However, to make the presentation concise, we consider a single τ for all i's, suppressing $q_{it}(\tau) = q_{it}$ throughout the paper.

We let y_{it} be a demeaned returns series of an asset or a portfolio. While it can be a more general time series, we focus on financial time series and expect that our method could be useful for risk management in financial markets. There has been needs to analyze the dynamic response of the VaR, i.e., conditional quantile, of y_{it} when there is a shock to the market or another asset. The only available method has been the pseudo-QIRF by WKM, which was recently adopted by Bouri et al. (2018), Chuliá et al. (2017), Peng and Zeng (2019), Shen (2018), and Wen et al. (2019) among others. However, as WKM acknowledge, the pseudo-QIRF is based on the restrictive assumption that a shock at time t affects only $\mathbf{y}_t$ without generating any change in future values of $\mathbf{y}_t$. Consequently, the pseudo-QIRF underestimates the impact of a shock. See Section 3.2 for more details. This paper tries to overcome such a limitation. Our approach is based on the framework of WKM and, more specifically, we adopt the VAR for VaR model and the same shock formulation as in WKM.

The following definition of QIRF measures the impact of a shock on conditional quantile dynamics of financial time series.

Definition 2.1 *Let $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{nt})^\top$ and $\mathbf{q}_t = (q_{1t}, q_{2t}, ..., q_{nt})^\top$. We define the quantile impulse response function (QIRF) as*

$$\begin{aligned} QIRF^{(s)} &:= E\left[\mathbf{q}_{t+s} | \mathbf{y}_t = \boldsymbol{\delta}^0; \mathcal{F}_{t-1}\right] - E\left[\mathbf{q}_{t+s} | \mathbf{y}_t = \mathbf{0}; \mathcal{F}_{t-1}\right] \\ &= E_t\left[\mathbf{q}_{t+s} | \mathbf{y}_t = \boldsymbol{\delta}^0\right] - E_t\left[\mathbf{q}_{t+s} | \mathbf{y}_t = \mathbf{0}\right]. \end{aligned} \tag{2}$$

Remark 2.1 *The QIRF measures the change in $\mathbf{q}_{t+s}$ caused by a shock. Here, $\boldsymbol{\delta}^0$ represents a shock at time t, and we let $\mathbf{y}_t = \mathbf{0}$ in the absence of the shock. Given that our main application is to investigate VaR dynamics of financial asset returns, we consider the difference between the conditional quantiles under these two scenarios: $\mathbf{y}_t = \boldsymbol{\delta}^0$ and $\mathbf{y}_t = \mathbf{0}$. The case of $\mathbf{y}_t = \mathbf{0}$ considers the benchmark of asset return being centered around zero. The wide consensus on (near-) martingale difference assumption on asset returns $\mathbf{y}_t$ (conditional mean being zero) makes this formulation empirically relevant. This shock formulation for the QIRF construction is the same as* *that for WKM's pseudo-QIRF construction. We explain the reason why we do not impose a shock to innovations in Section 3.1 (see Equation (7) and (8) below).*

Remark 2.2 *There is another difference between the QIRF and the conventional IRF. Unlike the conventional IRF, the QIRF does not assume innovations between $t+1$ and $t+s$ to be zero. If y_{it} is a demeaned GARCH type process such that $y_{it} = \sigma_{it}\epsilon_{it}$ and $\epsilon_{it+1}, \cdots, \epsilon_{it+s}$ are assumed to be zero, $y_{it+1}, \cdots, y_{it+s}$ are restricted to be zero and the s-step ahead volatility forecast ($E_t[\sigma_{t+s}]$) becomes smaller accordingly.*[2] *Since the s-step ahead forecast of conditional quantile is also similarly affected, we do not assume innovations between $t+1$ and $t+s$ to be zero and the QIRF is defined as the change in the forecast of ($\mathcal{F}_{t+s-1}$-measurable) conditional quantile $\mathbf{q}_{t+s}$ attributable to an initial shock at time t.*

The definition of $QIRF^{(s)}$ in (2) shows that calculating quantile impulse responses is related to multi-step prediction of conditional quantile. We choose to use the VAR for VaR model that directly estimates conditional quantile, instead of fully specifying the whole distribution, and adopt the idea of direct forecasting for multi-step prediction. When one estimates the VaR of financial time series using a GARCH-type model, a typical approach is to fully specify the conditional variance and the distribution of innovations. For this parametric approach, there is inevitably a possibility of model misspecification. Instead of full specification of DGP, one can alternatively model the conditional quantile directly as the CAViaR model by Engle and Manganelli (2004). The CAViaR approach is known to perform well in estimating and forecasting the VaR of financial time series, and the VAR for VaR model is its multivariate extension, which is described as

$$\mathbf{q}_t = \mathbf{c} + \mathbf{A}|\mathbf{y}_{t-1}| + \mathbf{B}\mathbf{q}_{t-1}.$$

By repetitive substitutions, it can be expressed as

$$\mathbf{q}_t = \mathbf{C} + \mathbf{A}\,|\mathbf{y}_{t-1}| + \mathbf{B}\mathbf{A}\,|\mathbf{y}_{t-2}| + \mathbf{B}^2\mathbf{A}\,|\mathbf{y}_{t-3}| + \cdots,$$

where $\mathbf{C} = \mathbf{c} + \mathbf{B}\mathbf{c} + \mathbf{B}^2\mathbf{c} + \cdots$. By applying the idea of local projections by Jordà (2005), we can consider

$$\mathbf{q}_{t+s} = \mathbf{C}^{(s)} + \mathbf{A}^{(s)}\,|\mathbf{y}_t| + \mathbf{B}^{(s)}\mathbf{A}^{(s)}\,|\mathbf{y}_{t-1}| + (\mathbf{B}^{(s)})^2\mathbf{A}^{(s)}\,|\mathbf{y}_{t-2}| + \cdots \qquad \text{for} \quad s = 1, 2, \cdots, S, \tag{3}$$

similarly as in (1), and use the estimate of $\mathbf{A}^{(s)}$ to estimate QIRF.

Based on this idea, we propose to estimate $QIRF^{(s)}$ as the following:

$$\widehat{QIRF}^{(s)} = \widehat{\mathbf{A}}^{(s)}|\boldsymbol{\delta}^0|, \tag{4}$$

[2] For example, consider a univariate TS-GARCH(1,1) model in Section 3.1:

$$y_t = \sigma_t\epsilon_t, \quad \sigma_t = \omega + \alpha|y_{t-1}| + \beta\sigma_{t-1}, \qquad \epsilon_t \overset{\text{iid}}{\sim} (0,1).$$

Under the model, the forecast of σ_{t+s} is $E_t[\sigma_{t+s}] = \omega + E_t[\alpha|\epsilon_{t+s-1}| + \beta]E_t[\sigma_{t+s-1}]$. However, $E_t[\sigma_{t+s}]$ with an innovation turned off ($\epsilon_{t+s-1} = 0$) is smaller than that: $\omega + \beta E_t[\sigma_{t+s-1}]$.

where $\widehat{\mathbf{A}}^{(s)}$ is the estimate of $\mathbf{A}^{(s)}$ from

$$\mathbf{q}_t = \mathbf{c}^{(s)} + \mathbf{A}^{(s)}|\mathbf{y}_{t-s}| + \mathbf{B}^{(s)}\mathbf{q}_{t-1}, \quad \text{for} \;\; s = 1, 2, ..., S. \tag{5}$$

Remark 2.3 *It should be noted that (5) leads to (3) by repetitive substitutions. We interpret (5) as a conditional quantile version of local projections. Our local projection method effectively captures quantile responses to a shock, as will be explained in Section 3.3.*

Remark 2.4 *In the VAR for VaR model by WKM, the past returns have a symmetric effect on the VaR. Hence, we assume the conditional quantile as a function of $|\mathbf{y}_{t-s}|$, so the QIRF depends on $|\boldsymbol{\delta}^0|$. In the presence of an asymmetric effect of the past returns on the VaR, we can adjust the VAR for VaR model accordingly, as shown in Engle and Manganelli (2004), and the QIRF local projection may employ different response coefficients for positive and negative realizations of $\mathbf{y}_{t-s}$.*

3 Heuristics on QIRF and Its Estimator

In financial time series modeling, the volatility dynamics are of primary importance. Thus we study the relationship between volatility and quantiles in our QIRF framework. Even though QIRF and local projection estimation in Section 2 do not require a specific DGP or a model, we explore QIRFs of a few popular volatility models in this section. In particular, we investigate some prototypical GARCH-type models. To illustrate the idea, we first investigate a univariate GARCH model, and then discuss a multivariate case.

3.1 QIRF of GARCH Models

We first consider a simple univariate TS-GARCH(1,1) model of Taylor (1986) and Schwert (1989)[3]:

$$y_t = \epsilon_t \sigma_t, \quad \sigma_t = \omega + \alpha|y_{t-1}| + \beta\sigma_{t-1}, \qquad \epsilon_t \overset{\text{iid}}{\sim} (0, 1).$$

Since $\sigma_t = \omega + (\alpha|\epsilon_{t-1}| + \beta)\sigma_{t-1}$, for $s \geq 1$, iterating this equation yields[4]

$$\sigma_{t+s} = \omega + \sum_{j=1}^{s-1} \Big[\prod_{k=1}^{j} \big(\alpha|\epsilon_{t+s-k}| + \beta\big)\Big]\omega + \Big[\prod_{k=1}^{s-1} \big(\alpha|\epsilon_{t+s-k}| + \beta\big)\Big](\alpha|y_t| + \beta\sigma_t).$$

[3] By using the TS-GARCH model rather than the classical model $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2$ of Bollerslev (1986), the conditional quantile q_t becomes a linear function of $|y_{t-1}|$ and σ_{t-1}, providing intuitive analytical expressions of the QIRF.

[4] In this paper, we assign 0 or 1 to the following summation and product notations for notational simplicity. When the upper bound of summation is less than the lower bound of summation, we let the value of the summation be 0. When the upper bound of the product is less than the lower bound of the product, we let the value of the product be 1. For example, $\sum_{j=1}^{0}[\,\cdot\,] = 0$ and $\prod_{k=1}^{0}[\,\cdot\,] = 1$.

Note that $q_t = F_\epsilon^{-1}(\tau)\sigma_t$ where F_ϵ^{-1} is the inverse CDF of ϵ_t. Hence q_{t+s} is expressed as

$$q_{t+s} = F_\epsilon^{-1}(\tau)\Big[\omega + \sum_{j=1}^{s-1}\Big[\prod_{k=1}^{j}\big(\alpha|\epsilon_{t+s-k}|+\beta\big)\Big]\omega + \Big[\prod_{k=1}^{s-1}\big(\alpha|\epsilon_{t+s-k}|+\beta\big)\Big]\alpha|y_t|\Big] + \Big[\prod_{k=1}^{s-1}\big(\alpha|\epsilon_{t+s-k}|+\beta\big)\Big]\beta q_t.$$

The change in q_{t+s} attributable to a shock $y_t = \delta^0$ is then

$$F_\epsilon^{-1}(\tau)\Big[\prod_{k=1}^{s-1}\big(\alpha|\epsilon_{t+s-k}| + \beta\big)\Big]\alpha|\delta^0|. \tag{6}$$

The above expression illustrates the response of quantiles driven by the evolution of volatility. In GARCH models, the effect of a shock on volatility is transmitted via two channels, (i) the feed-through channel and (ii) the volatility-persistence channel. We can clearly see these channels by rewriting the dynamics of volatility as

$$\sigma_{t+1} = \omega + \alpha\big(|y_t| - E_t[|y_t|]\big) + \big(\alpha E_t[|\epsilon_t|] + \beta\big)\sigma_t.$$

The coefficient α measures the effect of a mean-zero volatility variation, $|y_t| - E_t[|y_t|]$, on the next period's volatility. The coefficient of autoregressive term, $\alpha E_t[|\epsilon_t|] + \beta$, measures volatility persistence.[5] Because the response of q_{t+s} to a shock $y_t = \delta^0$ arises from these two channels, the quantile response can be decomposed into the scale, volatility-persistence and feed-through components:

$$\underbrace{F_\epsilon^{-1}(\tau)}_{\text{scale}}\underbrace{\Big[\prod_{k=1}^{s-1}\big(\alpha|\epsilon_{t+s-k}| + \beta\big)\Big]}_{\text{volatility-persistence}} \times \underbrace{\alpha|\delta^0|}_{\text{feed-through}}.$$

Therefore, the QIRF of TS-GARCH(1,1) model, the conditional expectation of (6), accommodates the proper two channels of the shock transmission mechanism:

$$QIRF^{(s)} := E_t\left[q_{t+s}|y_t = \delta^0\right] - E_t\left[q_{t+s}|y_t = 0\right] = F_\epsilon^{-1}(\tau)\big(\alpha E[|\epsilon_t|] + \beta\big)^{s-1}\alpha|\delta^0|. \tag{7}$$

This example also explains the shock formulation in the QIRF definition. To employ ϵ_t as the shock, the QIRF could be defined as $E_t\left[q_{t+s}|\epsilon_t = \delta^0\right] - E_t\left[q_{t+s}|\epsilon_t = 0\right]$. Under this definition, however, the measurement of QIRF is infeasible. Since

$$E_t\left[q_{t+s}|\epsilon_t = \delta^0\right] - E_t\left[q_{t+s}|\epsilon_t = 0\right] = F_\epsilon^{-1}(\tau)\Big[\prod_{k=1}^{s-1}\big(\alpha|\epsilon_{t+s-k}| + \beta\big)\Big]\alpha|\delta^0|\sigma_t, \tag{8}$$

this definition employs a $\mathcal{F}_{t-1}$-measurable σ_t, for which a specific DGP is needed in practice. We therefore use the QIRF definition (2).

[5] See Chapter 12 of Campbell, Lo and MacKinlay (1997) for details about GARCH models.

We can extend the discussion to a multivariate case, which is a DGP example of WKM.[6] Adopting from their paper (p.173)[7]:

$$\mathbf{y}_t = \mathbf{e}_t \boldsymbol{\Sigma}_t,$$
$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \overset{\text{iid}}{\sim} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$
$$\boldsymbol{\Sigma}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}|\mathbf{y}_{t-1}| + \boldsymbol{\beta}\boldsymbol{\Sigma}_{t-1},$$

where $\mathbf{y}_t = (y_{1t}, y_{2t})^\top$, $|\mathbf{y}_t| = (|y_{1t}|, |y_{2t}|)^\top$, $\boldsymbol{\Sigma}_t = (\sigma_{1t}, \sigma_{2t})^\top$, $\boldsymbol{\omega} = (\omega_1, \omega_2)^\top$, $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}$, $\boldsymbol{\beta} = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix}$, $\mathbf{e}_t = \begin{bmatrix} e_{1t} & 0 \\ 0 & e_{2t} \end{bmatrix}$.

The marginal distributions of e_{it}, for $i = 1, 2$, are identical for simplicity. The change in $\mathbf{q}_{t+s} = (q_{1t+2}, q_{2t+2})^\top$ attributable to a shock $\mathbf{y}_t = \boldsymbol{\delta}^0$ is

$$F_e^{-1}(\tau)\Big[\prod_{k=1}^{s-1} \big(\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta}\big)\Big]\boldsymbol{\alpha}|\ \boldsymbol{\delta}^0|,$$

where F_e^{-1} is the inverse CDF of e_{it} and $|\mathbf{e}_t| = \begin{bmatrix} |e_{1t}| & 0 \\ 0 & |e_{2t}| \end{bmatrix}$. The QIRF is then:

$$QIRF^{(s)} := E_t\left[\mathbf{q}_{t+s}|\mathbf{y}_t = \boldsymbol{\delta}^0\right] - E_t\left[\mathbf{q}_{t+s}|\mathbf{y}_t = \mathbf{0}\right] = F_e^{-1}(\tau)\Big(\boldsymbol{\alpha}E\big[|\mathbf{e}_t|\big] + \boldsymbol{\beta}\Big)^{s-1}\boldsymbol{\alpha}|\boldsymbol{\delta}^0|. \quad (9)$$

3.2 Comparison to WKM's Pseudo-QIRF

WKM introduce a so-called pseudo-QIRF. Following their definition, the analytical expression of the pseudo-QIRF in GARCH models is

$$\textit{pseudo-}QIRF^{(s)} = F_e^{-1}(\tau)\boldsymbol{\beta}^{s-1}\boldsymbol{\alpha}|\boldsymbol{\delta}^0|,$$

whereas the true QIRF is (9). Thus pseudo-QIRF estimates the persistence of volatility to be $\boldsymbol{\beta}$ whereas the true persistence is $\boldsymbol{\alpha}E\big[|\mathbf{e}_t|\big] + \ \boldsymbol{\beta}$.

WKM define the pseudo-QIRF as the difference between the conditional quantiles from the following two time paths.

$$\{\ \dots, \quad \mathbf{y}_{t-2}, \quad \mathbf{y}_{t-1}, \quad \tilde{\mathbf{y}}_t = \boldsymbol{\delta}^0, \quad \mathbf{y}_{t+1} \quad \mathbf{y}_{t+2} \quad \mathbf{y}_{t+3} \quad \dots\ \}$$
$$\{\ \dots, \quad \mathbf{y}_{t-2}, \quad \mathbf{y}_{t-1}, \quad \mathbf{y}_t = \mathbf{0}, \quad \mathbf{y}_{t+1} \quad \mathbf{y}_{t+2} \quad \mathbf{y}_{t+3} \quad \dots\ \}$$

[6] The QIRF of a bivariate TS-GARCH(p,q) is discussed in detail in Appendix A.1.1.

[7] We assume structural shocks, $(\epsilon_{1t}, \epsilon_{2t})^\top$, can be derived by applying the Cholesky decomposition to reduced form shocks. That is,

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \overset{\text{iid}}{\sim} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

At time t, one of the time paths is hit by a shock ($\boldsymbol{\delta}^0$), but the other is not. Since the DGP is not fully specified, they assume the shock does not change $\mathbf{y}_{t+s}$ for $s \geq 1$ so the two time paths are identical except at time t. This scenario neglects the dynamic evolution in the second moment, as they acknowledge (p.173 of WKM). Consequently, the pseudo-QIRF underestimates the magnitude of the true QIRF, which is well illustrated in the simulation Section 5.

Compared to the pseudo-QIRF, our QIRF measures a more comprehensive effect. It accounts for the effect of a shock not only on the current return ($\mathbf{y}_t$), but also on subsequent returns ($\mathbf{y}_{t+s}$). Thus, it takes into account the evolution of moments such as volatility. Without fully specifying DGP, the QIRF is well approximated via the local projection.

3.3 Local Projection Estimation of QIRF in GARCH Models

As we can see from (7) and (9), QIRF estimation involves the expectation of an absolute value of the latent innovations. This estimation procedure is not trivial even with a specific multivariate DGP. In this section, we study the logic behind QIRF estimation by local projection and illustrate why the method can effectively estimate the QIRF.[8]

After some algebra, σ_{t+s} in the univariate TS-GARCH(1,1) is written as a function of $|y_t|$ and σ_{t+s-1} for $s \geq 2$,

$$\begin{aligned}\sigma_{t+s} = \omega + \sum_{j=1}^{s-1}\Big[\prod_{k=1}^{j}\big(\alpha|\epsilon_{t+s-k}| + \beta\big)\Big]\omega - \frac{(\alpha|\epsilon_{t+s-1}| + \beta)\beta}{\alpha|\epsilon_t| + \beta}\Big[\omega + \sum_{j=2}^{s-1}\Big[\prod_{k=2}^{j}(\alpha|\epsilon_{t+s-k}| + \beta)\Big]\omega\Big] \\ + \Big[\prod_{k=1}^{s-1}\big(\alpha|\epsilon_{t+s-k}| + \beta\big)\Big]\alpha|y_t| + \frac{(\alpha|\epsilon_{t+s-1}| + \beta)\beta}{\alpha|\epsilon_t| + \beta}\sigma_{t+s-1}.\end{aligned} \tag{10}$$

In the above expression, $\sigma_{t+s}(= \omega + \alpha|y_{t+s-1}| + \beta\sigma_{t+s-1})$ is a function of $|y_t|$ and σ_{t+s-1}. Accordingly, the expression includes innovations between t and $t+s-1$, and they have the same functional form (i.e., $\{\alpha|\epsilon_p| + \beta\}_{p=t}^{t+s-1}$). We use the following approximation to deliver the logic of local projection QIRF estimation.

Assume both σ_{t+s-1} and y_t are given but $\{|\epsilon_p|\}_{p=t}^{t+s-1}$ are random. Define $X_t := \alpha|\epsilon_t| + \beta$. From (10), we can find a function $g(\cdot)$ such that $\sigma_{t+s} = g(X_t, ..., X_{t+s-1})$; thus, $y_{t+s} = \epsilon_{t+s} \cdot g(X_t, ..., X_{t+s-1})$. Let $\mu_X = E[X_t] = \alpha E[|\epsilon_t|] + \beta$ and $\bar{g} = g(\mu_X, ..., \mu_X)$, then

$$\begin{aligned} Pr\Big(y_{t+s} \leq F_\epsilon^{-1}(\tau)\cdot\bar{g}\Big) &= Pr\Big(\epsilon_{t+s} \leq \frac{F_\epsilon^{-1}(\tau)\cdot\bar{g}}{g(X_t, ..., X_{t+s-1})}\Big) \\ &= \int\cdots\int F_\epsilon\Big(\frac{F_\epsilon^{-1}(\tau)\cdot\bar{g}}{g(X_t, ..., X_{t+s-1})}\Big)dF_{|\epsilon_t|}\cdots dF_{|\epsilon_{t+s-1}|} \\ &= E\Big[F_\epsilon\Big(\frac{F_\epsilon^{-1}(\tau)\cdot\bar{g}}{g(X_t, ..., X_{t+s-1})}\Big)\Big]. \end{aligned}$$

Using the first-order approximation of $F_\epsilon\Big(\frac{F_\epsilon^{-1}(\tau)\cdot\bar{g}}{g(X_t,...,X_{t+s-1})}\Big)$ evaluated at the mean effect of innova-

[8] Details of the derivation for the univariate and bivariate models are in Appendix A.1.2 and A.1.3, respectively.

tions ($X_t = ... = X_{t+s-1} = \mu_X$), we can show

$$Pr\Big(y_{t+s} \leq F_\epsilon^{-1}(\tau) \cdot \bar{g}\Big) \approx \tau.$$

That is, the τ-quantile of y_{t+s} approximates $F_\epsilon^{-1}(\tau) \cdot g(\mu_X, ..., \mu_X)$. Applying the approximation yields

$$q_{t+s} \approx F_\epsilon^{-1}(\tau)\Big[\omega + \sum_{j=1}^{s-1} \mu_X^j \omega - \beta\big[\omega + \sum_{j=2}^{s-1} \mu_X^{j-1}\omega\big]\Big] + F_\epsilon^{-1}(\tau)\mu_X^{s-1}\alpha|y_t| + \beta q_{t+s-1}, \quad \text{for } s \geq 2.$$

Thus, $\widehat{a}^{(s)}$ of the following CAViaR model

$$q_t = c^{(s)} + a^{(s)}|y_{t-s}| + b^{(s)}q_{t-1}, \quad \text{for } \ s = 1, 2, ..., S.$$

effectively estimates $F_\epsilon^{-1}(\tau)\mu_X^{s-1}\alpha$. From (7), $QIRF^{(s)}$ can be estimated as

$$\widehat{QIRF}^{(s)} = \widehat{a}^{(s)}|\delta^0|.$$

The similar approximation for a bivariate case leads to

$$\widehat{QIRF}^{(s)} = \widehat{\mathbf{A}}^{(s)}|\boldsymbol{\delta}^0|,$$

where $\widehat{\mathbf{A}}^{(s)}$ is an estimator for $\mathbf{A}^{(s)}$ from the following model,

$$\mathbf{q}_t = \mathbf{c}^{(s)} + \mathbf{A}^{(s)}|\mathbf{y}_{t-s}| + \mathbf{B}^{(s)}\mathbf{q}_{t-1}, \quad \text{for } \ s = 1, 2, ..., S.$$

The above estimation shares the local projection idea of Jordà (2005), who estimates the mean IRF. We therefore naturally label our procedure as local projection QIRF estimation as given in (5).

4 Asymptotic Theory and Stationary Bootstrap

In this section, we provide valid econometric inferential tools for the QIRF estimation. We first derive the asymptotic distribution of the local projection QIRF estimator that can be used to construct an asymptotic confidence interval (CI) with the estimated nuisance parameters. Second, we propose a stationary bootstrap based CI from Politis and Romano (1994), and confirm the bootstrap consistency. The proofs in this section are relegated to the Technical Appendix.

4.1 Assumptions

We adopt the assumptions from WKM and Engle and Manganelli (2004) but also combine them with those from Han et al. (2016) and Goncalves and de Jong (2003). In particular, we relax the moment condition using the L_2-NED condition from Goncalves and de Jong (2003) for Section 4.3. Hereafter in this section, we generalize the number of outcome variables to be n and allow quantiles of interest to differ across the variables.

Assumption 4.1 *The sequence $\{\mathbf{y}_t\}$ is stationary and strong mixing on the complete probability space $(\Omega, \mathcal{F}, P_0)$, where Ω is the sample space, $\mathcal{F}$ is a suitably chosen σ-field, and P_0 is the probability measure providing a complete description of the stochastic behavior for the sequence of $\{\mathbf{y}_t\}$.*

Remark 4.1 *A prototypical DGP from Section 3 is*

$$\mathbf{y}_t = \mathbf{e}_t \boldsymbol{\Sigma}_t,$$
$$\boldsymbol{\Sigma}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}|\mathbf{y}_{t-1}| + \boldsymbol{\beta}\boldsymbol{\Sigma}_{t-1},$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{nt})^\top$, $|\mathbf{y}_t| = (|y_{1t}|, |y_{2t}|, ..., |y_{nt}|)^\top$, $\boldsymbol{\Sigma}_t = (\sigma_{1t}, \sigma_{2t}, ..., \sigma_{nt})^\top$ and $\mathbf{e}_t = diag(e_{1t}, e_{2t}, ..., e_{nt})$ is an iid random matrix. $\mathbf{e}_t$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$. If (i) the largest eigenvalue in the absolute value of the matrix $\boldsymbol{\beta}$ is less than 1 and (ii) $E\|\boldsymbol{\alpha}|\mathbf{e}_t| + \boldsymbol{\beta}\|^s < 1$ for some integer $s \geq 1$ where $\|\mathbf{A}\|$ denotes some matrix norm such as the sup-norm defined as $\|\mathbf{A}\| = \sup_{\|x\|=1} \|\mathbf{A}x\|$, then stationarity and strong mixing in Assumption 4.1 is satisfied. Moreover, if $E\|\mathbf{e}_t\|^s < \infty$ then $E\|\mathbf{y}_t\|^s < \infty$.[9]

For $i = 1, 2, ..., n$, $F_{it}(y) = Pr[y_{it} \leq y|\mathcal{F}_{t-1}]$ denotes the conditional distribution function of y_{it} given $\mathcal{F}_{t-1}$ and $f_{it}(y)$ denotes its density function. The corresponding conditional quantile function is defined as $q_{it}(\tau_i) = \inf\{y : F_{it}(y) \geq \tau_i\}$ for $\tau_i \in (0, 1)$, $i = 1, 2, ..., n$. Define $\boldsymbol{\tau} = (\tau_1, \tau_2, .., \tau_n)^\top$ and $\mathbf{q}_t(\boldsymbol{\tau}) = \big(q_{1t}(\tau_1), q_{2t}(\tau_2), ..., , q_{nt}(\tau_n)\big)^\top$.

Assumption 4.2 *(1) $\{y_{it}\}$ is continuously distributed such that for each $\omega \in \Omega$, $F_{it}(\omega, \cdot)$ and $f_{it}(\omega, \cdot)$ are continuous on $\mathbb{R}$, $i = 1, 2, ..., n$, $t = 1, 2, ..., T$. (2) For the given $0 < \tau_i < 1$, we suppose the followings: (a) for each i, t and ω, $f_{it}(\omega, q_{it}(\tau_i)) > 0$; (b) there exists a real vector $\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}) = \mathrm{vec}\Big(\begin{bmatrix}\mathbf{c}_{\boldsymbol{\tau}}^{(s)} & \mathbf{A}_{\boldsymbol{\tau}}^{(s)} & \mathbf{B}_{\boldsymbol{\tau}}^{(s)}\end{bmatrix}\Big)$ such that for each t and s*

$$\mathbf{q}_t(\boldsymbol{\tau}) = \mathbf{c}_{\boldsymbol{\tau}}^{(s)} + \mathbf{A}_{\boldsymbol{\tau}}^{(s)}|\mathbf{y}_{t-s}| + \mathbf{B}_{\boldsymbol{\tau}}^{(s)}\mathbf{q}_{t-1}(\boldsymbol{\tau}), \qquad s = 1, 2, ..., S.$$

(3) There exists (a) a finite positive constant f_0 such that for each i, t, $\omega \in \Omega$ and each $y \in \mathbb{R}$, $f_{it}(\omega, y) \leq f_0 < \infty$; (b) a finite positive constant L_0 such that for each i, t, $\omega \in \Omega$ and each $x, y \in \mathbb{R}$, $|f_{it}(\omega, x) - f_{it}(\omega, y)| \leq L_0|x - y|$.

Let $d := n + 2n^2$ be the number of parameters in $\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})$.

Assumption 4.3 *(1) Let $\mathbb{A}$ be a compact subset of $\mathbb{R}^d$. For $i = 1, 2, ..., n$, we suppose the followings:*

[9] See Proposition 3, 4, and 5 of Carrasco and Chen (2002) for details.

(a) the sequence of functions $\{q^0_{it}:\Omega\times\mathbb{A}\to\mathbb{R}\}$ *is such that for each* t *and each* $\boldsymbol{\gamma}^{(s)}\in\mathbb{A}$, $q^0_{it}(\cdot,\boldsymbol{\gamma}^{(s)})$ *is* $\mathcal{F}_{t-1}$*-measurable; (b) for each* t *and each* $\omega\in\Omega$, $q^0_{it}(\omega,\cdot)$ *is continuous on* $\mathbb{A}$*; (c) for each* i *and* t, $q^0_{it}(\cdot,\boldsymbol{\gamma}^{(s)})$ *is specified as*

$$\begin{aligned}\mathbf{q}^0_t(\cdot,\boldsymbol{\gamma}^{(s)}) &= \Big(q^0_{1t}(\cdot,\boldsymbol{\gamma}^{(s)}),q^0_{2t}(\cdot,\boldsymbol{\gamma}^{(s)}),...,q^0_{nt}(\cdot,\boldsymbol{\gamma}^{(s)})\Big)^\top\\ &= \mathbf{c}^{(s)}+\mathbf{A}^{(s)}|\mathbf{y}_{t-s}|+\mathbf{B}^{(s)}\mathbf{q}^0_{t-1}(\cdot,\boldsymbol{\gamma}^{(s)}), \qquad s=1,2,...,S.\end{aligned}$$

where $\boldsymbol{\gamma}^{(s)}=\mathrm{vec}\Big(\Big[\mathbf{c}^{(s)}\quad\mathbf{A}^{(s)}\quad\mathbf{B}^{(s)}\Big]\Big)$. *(2) For each* t *and each* $\omega\in\Omega$, $\mathbf{q}^0_t(\omega,\cdot)$ *is twice continuously differentiable on* $\mathbb{A}$.

Define $\delta_{it}\big(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)=q^0_{it}(\cdot,\boldsymbol{\gamma}^{(s)})-q^0_{it}\big(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)$ and use the norm $\|\boldsymbol{\gamma}^{(s)}\|:=\max_{l=1,2,...,d}|\gamma^{(s)}_l|$ where $\boldsymbol{\gamma}^{(s)}=(\gamma^{(s)}_1,\gamma^{(s)}_2,...,\gamma^{(s)}_d)^\top$.

Assumption 4.4 *(1) There exists (a)* $\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\in\mathbb{A}$ *such that for all* i *and* t, $q^0_{it}\big(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)=q_{it}(\tau_i)$*; (b) a non-empty index set* $l\subset\{1,2,...,n\}$ *such that for each* $\epsilon>0$, *there exists* $\delta_\epsilon>0$ *such that for all* $\boldsymbol{\gamma}^{(s)}\in\mathbb{A}$ *with* $\|\boldsymbol{\gamma}^{(s)}-\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\|>\epsilon$,

$$Pr\Big[\cup_{i\in l}\Big\{\big|\delta_{it}\big(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big|>\delta_\epsilon\Big\}\Big]>0.$$

(2) $\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\in int(\mathbb{A})$, *where* $int(\cdot)$ *signifies the interior points of a given set.*

Assumption 4.5 *Define*

$$\begin{aligned}D_{0t}&:=\max_{i=1,2,...,n}\ \sup_{\boldsymbol{\gamma}^{(s)}\in\mathbb{A}}\big|q^0_{it}(\cdot,\boldsymbol{\gamma}^{(s)})\big|,\\ D_{1t}&:=\max_{i=1,2,...,n}\ \max_{j=1,2,...,d}\ \sup_{\boldsymbol{\gamma}^{(s)}\in\mathbb{A}}\left|\left(\frac{\partial}{\partial\gamma^{(s)}_j}\right)q^0_{it}(\cdot,\boldsymbol{\gamma}^{(s)})\right|,\\ D_{2t}&:=\max_{i=1,2,...,n}\ \max_{j=1,2,...,d}\ \max_{h=1,2,...,d}\ \sup_{\boldsymbol{\gamma}^{(s)}\in\mathbb{A}}\left|\left(\frac{\partial^2}{\partial\gamma^{(s)}_j\partial\gamma^{(s)}_h}\right)q^0_{it}(\cdot,\boldsymbol{\gamma}^{(s)})\right|.\end{aligned}$$

(1) For $i=1,2,...,n$, $E|y_{it}|^{r+\delta}<\infty$ *for some* $r>2$, *and* $\delta>0$. *(2)* $E[D_{0t}]<\infty$. *(3)* $E[D^3_{1t}]<\infty$. *(4)* $E[D_{2t}]<\infty$ *and* $E[D_{1t}D_{2t}]<\infty$.

Define

$$\begin{aligned}\mathbf{Q}^{(s)}_{\boldsymbol{\tau}}&=\sum_{i=1}^n E\big[f_{u_{it}}(0)\nabla q^0_{it}\big(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\nabla^\top q^0_{it}\big(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big],\\ \mathbf{V}^{(s)}_{\boldsymbol{\tau}}&=E\Big[\xi^{(s)}_t\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big(\xi^{(s)}_t\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)^\top\Big],\\ \xi^{(s)}_t(\boldsymbol{\gamma}^{(s)})&=\sum_{i=1}^n\nabla q^0_{it}(\cdot,\boldsymbol{\gamma}^{(s)})\psi_{\tau_i}\big(y_{it}-q^0_{it}(\cdot,\boldsymbol{\gamma}^{(s)})\big),\end{aligned}$$

where $u_{it} = y_{it} - q_{it}(\tau_i)$, $\psi_{\tau_i}(u) = \tau_i - 1[u < 0]$, $f_{u_{it}}$ is the density function of u_{it} conditional on $\mathcal{F}_{t-1}$, and $\nabla q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})$ is the gradient of $q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})$ with respect to $\boldsymbol{\gamma}^{(s)}$.

Assumption 4.6 *(1) $\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}$ is positive definite. (2) $\mathbf{V}_{\boldsymbol{\tau}}^{(s)}$ is positive definite.*

4.2 Asymptotic Theory for Local Projection QIRF Estimator

We estimate $\boldsymbol{\gamma}^{(s)}$ by the following minimization problem

$$\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau}) = \arg\min_{\boldsymbol{\gamma}^{(s)}} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} \rho_{\tau_i}\big(y_{it} - q_{it}^0(\cdot,\ \boldsymbol{\gamma}^{(s)})\big), \tag{11}$$

where $q_{it}^0(\cdot,\ \boldsymbol{\gamma}^{(s)})$ is defined in Assumption 4.3(1) and $\rho_{\tau_i}(u) = u(\tau_i - 1[u < 0])$.

We can show that $\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})$ is a consistent estimator satisfying a central limit theorem.

Lemma 4.1 *Suppose that Assumptions 4.1, 4.2(1, 2), 4.3(1), 4.4(1) and 4.5(1, 2) hold. Then,*

$$\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau}) \xrightarrow{p} \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}).$$

Lemma 4.2 *Suppose that Assumptions 4.1 - 4.6 hold. Then,*

$$\sqrt{T}\Big(\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau}) -\ \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\Big) = \big(\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big)^{-1}\mathbf{H}_T^{(s)} + o_p(1), \tag{12}$$

where $\mathbf{H}_T^{(s)} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_t^{(s)}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)$. The asymptotic distribution of the estimator $\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})$ is given by:

$$\sqrt{T}\Big(\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau}) -\ \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\Big) \xrightarrow{d} N\Big(\mathbf{0}, \big(\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big)^{-1}\mathbf{V}_{\boldsymbol{\tau}}^{(s)}\big(\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big)^{-1}\Big). \tag{13}$$

From Section 3.3, we approximate the QIRF at horizon s, $QIRF^{(s)}$, by $\mathbf{A}^{(s)}|\boldsymbol{\delta}^0|$ via the local projection. The QIRF local projection estimator $\widehat{QIRF}^{(s)}$ in (4) is a linear function of $\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})$, so the following theorem holds.

Theorem 4.1 *Suppose that Assumptions 4.1 - 4.6 hold. Then,*

$$\sqrt{T}\big(\widehat{\mathbf{A}}^{(s)}|\boldsymbol{\delta}^0| - \mathbf{A}^{(s)}|\boldsymbol{\delta}^0|\big) = \mathbf{G}\big(\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big)^{-1}\mathbf{H}_T^{(s)} + o_p(1),$$

where $\mathbf{G} = \Big[0 \quad |\boldsymbol{\delta}^{\mathbf{0}}|' \quad \mathbf{0}_{1\times n}\Big] \otimes \mathbf{I}_{n\times n}$, and the asymptotic distribution of $\widehat{QIRF}^{(s)}$ is given by:

$$\sqrt{T}\big(\widehat{\mathbf{A}}^{(s)}|\boldsymbol{\delta}^0| - \mathbf{A}^{(s)}|\boldsymbol{\delta}^0|\big) \xrightarrow{d} N\Big(\mathbf{0}, \mathbf{G}\big(\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big)^{-1}\mathbf{V}_{\boldsymbol{\tau}}^{(s)}\big(\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big)^{-1}\mathbf{G}^{\top}\Big).$$

Remark 4.2 *To construct the asymptotic confidence interval for $\widehat{QIRF}^{(s)}$, we use the following* *consistent estimator for* $\mathbf{V}_{\boldsymbol{\tau}}^{(s)}$ *and* $\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}$ *following WKM,*

$$
\begin{aligned}
\widehat{\mathbf{V}}_{\boldsymbol{\tau}}^{(s)} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t^{(s)}(\hat{\xi}_t^{(s)})^\top, \\
\hat{\xi}_t^{(s)} &= \sum_{i=1}^{n}\nabla q_{it}^0\big(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})\big)\psi_{\tau_i}(\hat{\epsilon}_{it}), \\
\hat{\epsilon}_{it} &= y_{it} - q_{it}^0\big(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})\big), \\
\widehat{\mathbf{Q}}_{\boldsymbol{\tau}}^{(s)} &= \frac{1}{2\hat{c}_T T}\sum_{i=1}^{n}\sum_{t=1}^{T}1[|\hat{\epsilon}_{it}| \le \hat{c}_T]\nabla q_{it}^0\big(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})\big)\Big(\nabla q_{it}^0\big(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})\big)\Big)^\top.
\end{aligned}
$$

4.3 Stationary Bootstrap Inference

The accurate estimation of $\mathbf{Q}_{\tau}^{(s)}$ in Theorem 4.1 can be challenging because of the nuisance parameter estimation, especially at tail quantiles where the number of observation is scarce. This difficulty has been well known in the quantile regression literature; see Koenker (1994, 2005) for instance. As a result, the performance of the asymptotic confidence intervals for $\widehat{QIRF}^{(s)}$ may not be satisfactory. In this section, we avoid the nuisance parameter estimation by providing a stationary bootstrap confidence interval by Politis and Romano (1994) and its validity.

We draw a sequence of *iid* random block lengths $\{L_i\}_{i\in\mathbb{N}}$ that has the following geometric distribution:

$$
Pr(L_i = x) = p(1-p)^{x-1}, \quad 0 < p < 1.
$$

A sequence of *iid* random variables $\{K_i\}_{i\in\mathbb{N}}$ has the discrete uniform distribution on $\{1, ..., T\}$, where $\{L_i\}_{i\in\mathbb{N}}$ and $\{K_i\}_{i\in\mathbb{N}}$ are independent. We build blocks $B_{K_i,L_i} = \{\mathbf{y}_t\}_{t=K_i}^{K_i+L_i-1}$ of length L_i starting with the K_i-th observation.[10] The stationary bootstrap generates bootstrap samples $\{\mathbf{y}_t^*\}_{t=1}^T$ by taking the first T observations from a sequence of blocks $\{B_{K_i,L_i}\}_{i\in\mathbb{N}}$. The following assumption from Goncalves and de Jong (2003) ensures the validity of the stationary bootstrap.

Assumption 4.7 *(1) For some* $r>2$ *and* $\delta>0$ *chosen as in Assumptions 4.5(1), (a)* y_{it} *(i=1,...,n) is* $L_{2+\delta}$*-NED on* $\{V_t\}$ *with NED coefficient* v_k *of size* -1*, i.e.* $v_k \equiv \sup_{i,t} E|\ y_{it} - E_{t-k}^{t+k}[y_{it}]\ |^2 \to 0$ *as* $k \to \infty$*, and* $v_k = O(k^{-1-\epsilon})$ *for some* $\epsilon > 0$*; (b)* $\{V_t\}$ *is an* α*-mixing sequence with* $\alpha(k)$ *of size* $-\frac{(2+\delta)(r+\delta)}{(r-2)}$*. (2)* $p = p_T \to 0$ *and* $Tp_T^2 \to \infty$ *as* $T \to \infty$*.*

Let $\{\mathbf{y}_t^* = (y_{1t}^*, y_{2t}^*, ..., , y_{nt}^*)^\top\}_{t=1}^T$ denote the stationary bootstrap sample. The local projection estimator with the stationary bootstrap sample solves the following minimization problem:

$$
\widehat{\boldsymbol{\gamma}}^{(s)*}(\boldsymbol{\tau}) = \arg\min_{\boldsymbol{\gamma}^{(s)}} \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{n}\rho_{\tau_i}\big(y_{it}^* - q_{it}^{0*}(\cdot,\ \boldsymbol{\gamma}^{(s)})\big),
$$

[10] In resampling, the first observation $\mathbf{y}_1$ is treated as the observation following the last observation $\mathbf{y}_T$. That is, for $t > T$, $\mathbf{y}_t$ is set to be $\mathbf{y}_{t-T}$.

where $q_{it}^{0*}(\cdot,\ \boldsymbol{\gamma}^{(s)})$ is defined as

$$\begin{aligned}\mathbf{q}_t^{0*}(\cdot, \boldsymbol{\gamma}^{(s)}) &= \Big(q_{1t}^{0*}(\cdot, \boldsymbol{\gamma}^{(s)}), q_{2t}^{0*}(\cdot, \boldsymbol{\gamma}^{(s)}), ..., q_{nt}^{0*}(\cdot, \boldsymbol{\gamma}^{(s)})\Big)^\top \\ &= \mathbf{c}^{(s)} + \mathbf{A}^{(s)}|\mathbf{y}_{t-s}^*| + \mathbf{B}^{(s)}\mathbf{q}_{t-1}^{0*}(\cdot, \boldsymbol{\gamma}^{(s)}), \qquad s = 1, 2, ..., S.\end{aligned}$$

The asymptotic distribution of $\widehat{\boldsymbol{\gamma}}^{(s)*}(\boldsymbol{\tau})$ and $\widehat{QIRF}^{(s)*} = \widehat{\mathbf{A}}^{(s)*}|\boldsymbol{\delta}^0|$ can be derived with the similar argument from Section 4.2, and the proofs can be found in the Technical Appendix.

Lemma 4.3 *Suppose that Assumptions 4.1 - 4.7 hold . Then,*

$$\sqrt{T}\Big(\widehat{\boldsymbol{\gamma}}^{(s)*}(\boldsymbol{\tau}) - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\Big) = \big(\mathbf{Q}_\tau^{(s)}\big)^{-1}\mathbf{H}_T^{(s)*} + o_p(1),$$

where $\mathbf{H}_T^{(s)*} = \frac{1}{\sqrt{T}}\sum_{t=1}^T \xi_t^{(s)*}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)$, $\xi_t^{(s)*}(\boldsymbol{\gamma}^{(s)}) = \sum_{i=1}^n \nabla q_{it}^{0*}(\cdot, \boldsymbol{\gamma}^{(s)})\psi_{\tau_i}\big(y_{it}^{0*} - q_{it}^{0*}(\cdot, \boldsymbol{\gamma}^{(s)})\big)$ *and* $\nabla q_{it}^{0*}(\cdot, \boldsymbol{\gamma}^{(s)})$ *is the gradient of* $q_{it}^{0*}(\cdot, \boldsymbol{\gamma}^{(s)})$ *with respect to* $\boldsymbol{\gamma}^{(s)}$.

Lemma 4.4 *Suppose that Assumptions 4.1 - 4.7 hold. Then,*

$$\sqrt{T}\big(\ \widehat{\mathbf{A}}^{(s)*}|\boldsymbol{\delta}^0| - \mathbf{A}^{(s)}|\boldsymbol{\delta}^0|\ \big) = \mathbf{G}\big(\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big)^{-1}\mathbf{H}_T^{(s)*} + o_p(1).$$

Define $\mathbf{B}_T^{(s)*} := \mathbf{H}_T^{(s)*} - \mathbf{H}_T^{(s)}$. Conditional on the original sample, we obtain the convergence in distribution as in the following lemma.

Lemma 4.5 *Suppose that Assumptions 4.1 - 4.7 hold. Then,*

$$\mathbf{B}_T^{(s)*} \ \xrightarrow{d}\ N\big(\mathbf{0}, \mathbf{V}_\tau^{(s)}\big),$$

conditional on the original sample, for almost every sequence.

We now obtain the asymptotic distribution of the bootstrapped QIRF estimator, confirming the stationary bootstrap consistency.

Theorem 4.2 *Suppose that Assumptions 4.1 - 4.7 hold. Then, in the sense of weak convergence conditional on the sample,*

$$\sqrt{T}\big(\ \widehat{\mathbf{A}}^{(s)*}|\boldsymbol{\delta}^0| - \widehat{\mathbf{A}}^{(s)}|\boldsymbol{\delta}^0|\ \big)\ \xrightarrow{*}\ N\Big(\mathbf{0}, \mathbf{G}\big(\mathbf{Q}_\tau^{(s)}\big)^{-1}\mathbf{V}_\tau^{(s)}\big(\mathbf{Q}_\tau^{(s)}\big)^{-1}\mathbf{G}^\top\Big).$$

Remark 4.3 *To construct the* $100\cdot(1-\alpha)\%$ *confidence interval for* $\widehat{QIRF}^{(s)}$, *(i) we draw* B *number of bootstrap samples, and (ii) for each bootstrap sample and a given* $s = 1, ..., S$, *we obtain* $\widehat{\mathbf{A}}^{(s)*}$. *Then, (iii) using* B *number of local projection estimator* $\widehat{\mathbf{A}}^{(s)*}$, $100\cdot(1-\frac{\alpha}{2})$ *and* $100\cdot(\frac{\alpha}{2})$ *empirical quantiles of* $\widehat{\mathbf{A}}^{(s)*}|\boldsymbol{\delta}^0|$ *provide the lower and upper bounds of the* $100\cdot(1-\alpha)\%$ *confidence interval of* $\widehat{QIRF}^{(s)}$.

5 Monte Carlo Simulations

We conduct several simulation studies to examine whether our local projection method provides reasonable approximations for true QIRFs and also investigate whether the two inferential procedures are valid in finite samples. First, we experiment with the bivariate TS-GARCH DGPs as in WKM and compare the local projection QIRF with the pseudo-QIRF by WKM. Specifically, we consider the following three DGPs.

DGP 1:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \sigma_{1t}e_{1t} \\ \sigma_{2t}e_{2t} \end{bmatrix},$$
$$\begin{bmatrix} \sigma_{1t} \\ \sigma_{2t} \end{bmatrix} = \begin{bmatrix} 0.02 \\ 0.02 \end{bmatrix} + \begin{bmatrix} 0.09 & 0.02 \\ 0.07 & 0.09 \end{bmatrix} \begin{bmatrix} |y_{1t-1}| \\ |y_{2t-1}| \end{bmatrix} + \begin{bmatrix} 0.89 & 0.01 \\ 0.06 & 0.85 \end{bmatrix} \begin{bmatrix} \sigma_{1t-1} \\ \sigma_{2t-1} \end{bmatrix},$$

for

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix},$$

where ϵ_{1t} and ϵ_{2t} are mutually independent and $\epsilon_{it} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$.

DGP 2: DGP 1 with

$$\epsilon_{it} \overset{\text{iid}}{\sim} (5/3)^{-\frac{1}{2}} \times t_5.$$

DGP 3:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \sigma_{1t}e_{1t} \\ \sigma_{2t}e_{2t} \end{bmatrix},$$
$$\begin{bmatrix} \sigma_{1t} \\ \sigma_{2t} \end{bmatrix} = \begin{bmatrix} 0.02 \\ 0.02 \end{bmatrix} + \begin{bmatrix} 0.05 & 0.01 \\ 0.03 & 0.04 \end{bmatrix} \begin{bmatrix} |y_{1t-1}| \\ |y_{2t-1}| \end{bmatrix} + \begin{bmatrix} 0.05 & 0.01 \\ 0.03 & 0.04 \end{bmatrix} \begin{bmatrix} |y_{1t-2}| \\ |y_{2t-2}| \end{bmatrix} + \begin{bmatrix} 0.89 & 0.01 \\ 0.06 & 0.85 \end{bmatrix} \begin{bmatrix} \sigma_{1t-1} \\ \sigma_{2t-1} \end{bmatrix},$$

for e_{1t} and e_{2t} in DGP 1.

For the parameter values in DGP 1 and DGP 2, we use the average estimates of 61 bivariate models for 5% VaR given in Table 4. The details of these estimates are explained in the next section. Here y_{1t} and y_{2t} correspond to a market return and an individual financial institution's stock return, respectively. The correlation ρ is set to be either 0 or 0.5. If $\rho = 0$, e_{1t} and e_{2t} are mutually independent. If $\rho \neq 0$, it is assumed that a shock to the market has a contemporaneous effect on the return of the individual financial institution, whereas the institution's specific shock has only a lagged effect on the market. We also consider DGP 3 particularly to investigate whether the local projection QIRF is robust to model misspecification. We can analytically calculate the true QIRF in each DGP and use it for comparison.

We generate data with sample size $4,000$ after removing the first 200 observations to get rid of the initial value effect. For each DGP, we estimate the model given in (5) for the 5% quantile.

For the QIRF, we assume that y_{1t} is hit by a negative shock with the magnitude of twice its unconditional standard deviation σ_1 and adopt a standard Cholesky decomposition as in WKM, which means $\boldsymbol{\delta}^0 = L\times(-2,0)'$ where a lower triangular matrix L is from the Cholesky decomposition of $cov(y_{1t}, y_{2t})$. The QIRFs are presented with horizons up to 30 and the number of repetition is $1,000$ for each DGP.

Figure 1 provides the simulation results, where each figure provides a true QIRF, an average local projection QIRF and an average pseudo-QIRF. Since the results for $\rho = 0$ are similar, we report only those for $\rho = 0.5$. It should be noted that the local projection QIRF is identical to the pseudo-QIRF for horizon $s = 1$, whereas they are different for horizon $s > 1$. In particular, the QIRF for longer horizons is of our interest. First of all, it is obvious that our local projection QIRF successfully approximates the true QIRF in each case, but the pseudo-QIRF does not. In Figures 1(B) and 1(D), the true QIRF decreases as the horizon increases, which means that the response of q_{2t} gets larger in absolute value as the horizon increases. Figures 1(B) and 1(D) show that the local projection QIRF successfully approximates the true QIRF. However, it is shown that the pseudo-QIRF monotonically converges to zero and gets further away from the true QIRF as the horizon increases.

Second, Figures 1(E) and 1(F) show that the local projection QIRF still well approximates the true QIRF even under model misspecification. Although the DGP 3 is a TS-GARCH(1,2) model, the local projection QIRF is based on the estimation of a TS-GARCH(1,1) model. Under DGP 3, the local projection QIRF is close to the true QIRF for longer horizons even if it is not that close for horizon $s = 1$. This implies that the local projection QIRF is robust to misspecification for longer horizons. On the other hand, the pseudo-QIRF is quite different from the true QIRF.

Next, we examine the validity of the two inferential procedures explained in the previous section. We consider the DGP 1 given above and adopt the same shock for QIRFs. For each data generated, we obtain two 95% CIs of the local projection QIRF. One is the asymptotic CI and the other is the stationary bootstrap CI based on $1,000$ bootstrapped replicates. The tuning parameter p is set to be 0.002. We count whether each CI includes the true QIRF for a given horizon. We consider only four horizons, $s = 1, 10, 20,$ and 30, because of the computational burden. We repeat this procedure $1,000$ times and calculate the effective coverage rates of two 95% CIs.

Table 1 reports the effective coverage rates. For shorter horizons $s = 1$ or 10, both asymptotic and bootstrap CIs exhibit similar coverage rates, between 0.90 and 0.94. However, for longer horizons $s = 20$ or 30, the coverage rate of the bootstrap CI is very close to 0.95, ranging between 0.94 and 0.96, whereas the coverage rate of the asymptotic CI is between 0.81 and 0.87. These results show that although both asymptotic and bootstrap CIs are valid inferential tools for relatively short horizons, the bootstrap CI is better for longer horizons.

6 Applications to Value-at-Risk Dynamics

What happens to each financial institution's VaR if the market crashes? This question is typically considered in a *stress test* of financial institutions, whereas the following question is also of interest. What happens to the market's VaR if a financial institution's stock return crashes? This is typically considered in an analysis of *systemic risk* of financial institutions. We use the QIRF to address these issues and examine a dynamic reaction of each financial institution's VaR when there is a shock to the market and a dynamic reaction of the market's VaR when there is a shock to a particular financial institution.

We use the CRSP market value weighted index return as the market index return and, at first, consider stock returns of three individual financial institutions: JP Morgan Chase (JPM), Morgan Stanley (MS), and AIG. We use daily observations from 24 Feb. 1993 to 29 Jun. 2018 with sample size 6,385.[11] We let y_{1t} and y_{2t} be the return series of the market and each financial institution, respectively, and first estimate the VAR for VaR model by WKM, which is the model given in (5) for $s = 1$. Table 2 reports the estimation results of the model for 5% VaR, where the quantiles of the market return and each financial institution's return are set to be 0.05. The autoregressive coefficients b_{11} and b_{22} are estimated to be mostly high (between 0.88 and 0.98), which indicate that the VaR processes are persistent. More importantly, some of the off-diagonal coefficients of the $\mathbf{A}^{(1)}$ or $\mathbf{B}^{(1)}$ matrices are significantly different from zero. The joint null hypothesis that all off-diagonal coefficients of the matrices $\mathbf{A}^{(1)}$ and $\mathbf{B}^{(1)}$ are equal to zero is rejected at the 1% significance level for all three cases. For JPM and MS, the estimates of a_{21} and b_{21} are relatively large in absolute value whereas a_{12} is close to zero. For AIG, the estimates of a_{12} and b_{12} are very close to zero.

Table 3 reports the estimation results of the model for 1% VaR, where the quantiles of the market return and each financial institution's return are set to be 0.01. The results are in general similar to those in Table 2 but there are a few differences. First, for 1% VaR, the joint null hypothesis that all off-diagonal coefficients of the matrices $\mathbf{A}^{(1)}$ and $\mathbf{B}^{(1)}$ are equal to zero is rejected for JPM and AIG but is not rejected for MS, for whom all off-diagonal coefficients of the $\mathbf{A}^{(1)}$ or $\mathbf{B}^{(1)}$ matrices are insignificant. Second, for 1% VaR of JPM, the estimates of a_{21} and b_{21} are larger in absolute value (-0.44 and -0.26, respectively) than those for 5% VaR (-0.18 and -0.14), which indicates that the effect of the market is larger for 1% VaR than for 5% VaR. For MS, the estimate of a_{21} for 1% VaR is also larger in absolute value (-0.37) than that for 5% VaR (-0.27) even if it is insignificant.

Next, we conduct a quantile impulse response analysis using the local projection method. Figures 2 and 3 provide the QIRFs of three individual financial institutions to a two standard deviation shock to the market index for 5% VaR and 1% VaR, respectively. As in WKM, the identification of the market shock relies on a Cholesky decomposition, which implicitly assumes that shocks to the market can contemporaneously affect each individual financial institution whereas shocks to the financial institution can affect the market only with a lag. In each figure, the horizontal axis indicates the time (expressed in days) and the vertical axis measures the change in the VaR of

[11] The stock return series are obtained from CRSP and Yahoo Finance. The stock return series of Morgan Stanley are available from 24 Feb. 1993.

the individual financial institution (expressed in percentage returns) as a reaction to the market shock. The solid line in each figure represents the local projection QIRF. The shaded area in each figure is the 95% confidence interval based on 1,000 bootstrapped replicates, for which we adopt the stationary bootstrap procedure explained in Section 4.3, and the tuning parameter p is chosen to be 0.002.[12] For comparison, we also provide the pseudo-QIRFs (dashed lines) with 95% confidence intervals (dotted lines).

In Figures 2 and 3, most importantly, the local projection QIRFs exhibit a non-monotonic trend. For example, in Figures 2(A) and 3(A), the local projection QIRFs of JPM exhibit the lowest value at horizon $s = 10$ and, in Figures 3(B), the local projection QIRF of MS reaches its minimum at horizon $s = 16$. That is, the reaction to the market shock reaches its maximum in two weeks for JPM's 5% and 1% VaRs and about three weeks for MS's 1% VaR. This feature is clearly different from the pseudo-QIRF. The pseudo-QIRFs monotonically converge to zero by construction, as is shown in Section 5. In Figures 2(A) and 3(A), the confidence intervals of the local projection QIRF at horizon $s = 10$ do not include the confidence intervals of the pseudo-QIRFs, which indicates that the local projection QIRFs are statistically different from the pseudo-QIRFs at horizon $s = 10$.

When we compare the QIRFs for 5% VaR with those for 1% VaR, the maximums in absolute value are larger for 1% VaR than for 5% VaR. It is not surprising that the reaction of 1% VaR is much larger than that of 5% VaR. For JPM, the market shock produces a 0.83% increase for 5% VaR and a 1.93% increase for 1% VaR at horizon $s = 10$. For MS, the maximum reaction is a 1.23% increase for 5% VaR at horizon $s = 2$ and a 2.81% increase for 1% VaR at horizon $s = 16$. For AIG, the market shock produces a 1.20% increase for 5% VaR and a 3.04% increase for 1% VaR at horizon $s = 2$.

Now we consider a situation where a shock is given to each financial institution instead of the market and investigate how the market VaR reacts in each case. Figures 4 and 5 provide the QIRFs of the market for 5% VaR and 1% VaR, respectively, when there is a '5% decrease shock' to each individual financial institution, which means that the stock return of each financial institution decreases by 5%. Since the value of two standard deviations is different for each institution's return, we instead impose the same shock to each institution to compare the responses of the market VaR. The QIRFs of the market to the shock to individual financial institutions are in general smaller in absolute value than the QIRFs of individual institutions to the market shock. This could be partly due to the identification assumption that an individual institution's shock has only a lagged effect on the market, whereas a shock to the market has a contemporaneous effect on individual institutions, as mentioned in WKM.

In Figures 4 and 5, the confidence intervals of the QIRF by local projection mostly include zero,

[12] In the stationary bootstrap, $1/p$ represents an average block length and $p = 0.002$ means the average block length is 500 in our application. When we choose p, we first tried the selection rule suggested by Politis and White (2004) and later corrected in Patton et al. (2009). However, the chosen average block length $1/p$ was too small and bootstrapped samples did not properly exhibit quantile dependence as the original samples did. We tried various tuning parameters, which showed that it would be desirable to have a large enough average block length in order for bootstrapped samples to exhibit as much of quantile dependence as the original samples did. We leave more rigorous investigation on this issue as future work.

which implies that the reaction of the market VaR is mostly insignificant. Nevertheless, the local projection QIRFs exhibit more substantial fluctuations with larger magnitudes than the pseudo-QIRFs. For example, when there is a shock to JPM or MS, the QIRF by local projection shows that the reaction of the 1% market VaR reaches its maximum (about -1%) at horizon $s = 20$, whereas the pseudo-QIRF is close to zero as shown in Figures 5(A) and 5(B). Meanwhile, when there is a shock to AIG, the response of the market VaR is the smallest. It is interesting to compare this result with that in Figures 2 and 3, where the response of AIG to a shock to the market is the largest at horizon $s = 1$.

The three financial institutions we consider belong to different financial sectors. JPM, MS and AIG belong to the Depositories group, the Broker-Dealers group and the Insurance group, respectively. In Figures 2 and 3, the minimums of the QIRFs of JPM are higher than those of MS or AIG, which implies that the reaction of JPM's VaR is the smallest. On the other hands, in Figures 4 and 5, the reaction of the market VaR is the smallest when there is a shock to AIG. One may ask whether these individual institutions' results could be generalized into sectoral characteristics. We examine this issue below.

Acharya et al. (2017) considered financial institutions in the U.S. that had a market cap in excess of 5 billion USD as of the end of June 2007 and categorized them into the following four groups: Depositories, Broker-Dealers, Insurance, and a group called Other, consisting of non-depository institutions, real estate, and so on. Following their categorization, we collect stock return series of individual financial institutions. Considering data availability, our analysis includes a total of 61 financial institutions[13] from 3 Jan. 2000 to 29 Jun. 2018 with sample size 4,653.[14]

Table 4 reports the summary statistics of the coefficient estimates of the 61 bivariate VAR for VaR models. Whereas the averages of a_{12} and b_{12} are close to zero, those of a_{21} and b_{21} quite different from zero. This implies that while the market VaR is marginally affected by each individual financial institution's return and VaR, individual institutions' VaRs are substantially influenced by the market's return and VaR. The cross-sectional standard deviation, minimum, and maximum also show that the estimates are heterogeneous.

Figure 6 provides the sectoral averages of the QIRFs to a two standard deviation shock to the market return for 5% VaR and 1% VaR, respectively. Figure 6(A1) is based on the local projection method for 5% VaR. In general, the average response of Broker-Dealers is the largest across the horizon, and the average response of Depositories becomes the smallest after about two weeks. In other words, when there is a shock to the market, the 5% VaR of Broker-Dealers increases by the largest amount across the horizon and that of Depositories does by the smallest amount after about two weeks. If the Broker-Dealers group tends to be involved in more risky investments than is the Depositories group, this result could reflect it. Figure 6(A2) is based on the local projection for 1% VaR. Similarly, the average response of Depositories becomes the smallest after about two

[13] The details are provided in Appendix C.

[14] If we stick to the same sample period from Feb. 1993, there are 51 companies whose stock return series are available. In that case, Goldman Sachs will be missing and there will be only four institutions in the Broker-Dealer group. Therefore, we instead consider the samples from Jan. 2000 and use 61 institutions.

weeks. The response of Broker-Dealers is more or less the largest up to three weeks, but Insurance shows the largest reaction at some horizons in about three weeks. Meanwhile, Figures 6(B1) and 6(B2) are based on the pseudo-QIRFs and the sectoral averages are quite close to each other even if it is relatively obvious that the response of Broker-Dealers is the largest for 5% VaR. Unlike our result, WKM's result (Fig. 3) shows that Insurance exhibits the largest quantile impulse response at horizon $s = 1$, apparently because the firms, categories, and sample period in their analysis are different from ours.

Next, we consider the response of the market when there is a 5% decrease shock to each individual financial institution as in Figures 4 and 5. Figure 7 provides the sectoral averages of the QIRFs of the market. Figure 7(A1) is for 5% VaR and it is not that any particular group exhibits a distinct feature. Figure 7(A2) is for 1% VaR and the average response of the 1% market VaR is the largest at horizon $s = 10$ and $s = 20$ for Broker-Dealers.

7 Conclusion

This paper studies the quantile impulse response function (QIRF) estimation, asymptotic theory, and statistical inference. The major application is for the financial market data whose stochastic property is largely determined by the persistent volatility dynamics. We provide a simple estimation method based on local projections, and valid inferential tools from asymptotics and stationary bootstrap. An extensive set of financial asset return data and their Value-at-Risk dynamics are examined to emphasize the benefit of the new local projection QIRF estimation and inferential methods.

An interesting future research agenda is how to apply this developed tool to a macroeconomic data set, whose stochastic property is substantially different from that of financial market data. Some recent studies by Chavleishvili and Manganelli (2019), Kim, Lee and Mizen (2019) and Montes-Rojas (2019) investigate this avenue using different approaches. We plan to extend the applicability of the current method to a macroeconomic environment, where the impulse response analysis has been the most popular.

A Technical Appendix

A.1 Derivations in Section 3

A.1.1 QIRF of a Bivariate TS-GARCH(p,q)

The underlying model is

$$\mathbf{y}_t = \mathbf{e}_t \boldsymbol{\Sigma}_t,$$
$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \overset{\text{iid}}{\sim} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right),$$
$$\boldsymbol{\Sigma}_t = \boldsymbol{\omega} + \sum_{i=1}^{q} \boldsymbol{\alpha}_i |\mathbf{y}_{t-i}| + \sum_{i=1}^{p} \boldsymbol{\beta}_i \boldsymbol{\Sigma}_{t-i}.$$

Arranging terms and multiplying $F_e^{-1}(\tau)$, $\mathbf{q}_{t+s}$ is expressed as

$$\mathbf{q}_{t+s} = F_e^{-1}(\tau)\boldsymbol{\omega} + F_e^{-1}(\tau) \sum_{i=1}^{q} \boldsymbol{\alpha}_i \Big[|\mathbf{y}_{t+s-i}| - E_t[|\mathbf{y}_{t+s-i}|] \Big] + \sum_{i=1}^{q} \boldsymbol{\alpha}_i E_t[|\mathbf{e}_{t+s-i}|] \mathbf{q}_{t+s-i} + \sum_{i=1}^{p} \boldsymbol{\beta}_i \mathbf{q}_{t+s-i}.$$

For $s = 1$, the impact of shock $\mathbf{y}_t = \boldsymbol{\delta}^0$ on $\mathbf{q}_{t+1}$ is

$$QIRF^{(1)} = F_e^{-1}(\tau)\boldsymbol{\alpha}_1 |\boldsymbol{\delta}^0|.$$

For $s = 2$, the impact of shock $\mathbf{y}_t = \boldsymbol{\delta}^0$ on $\mathbf{q}_{t+2}$ comes from two channels. The shock has an impact of $F^{-1}(\tau)\boldsymbol{\alpha}_2|\boldsymbol{\delta}^0|$ via the feed-through channel, and an impact of $(\boldsymbol{\alpha}_1|\mathbf{e}_{t+1}| + \boldsymbol{\beta}_1)QIRF^{(1)}$ via the volatility-persistence channel. Thus,

$$QIRF^{(2)} = F^{-1}(\tau)\boldsymbol{\alpha}_2|\boldsymbol{\delta}^0| + (\boldsymbol{\alpha}_1 E[|\mathbf{e}_t|] + \boldsymbol{\beta}_1)QIRF^{(1)}.$$

In the same way, we can derive $QIRF^{(s)}$ for $s > 2$,

In general forms, $QIRF^{(s)}$ can be expressed as the following depending on p and q:

(i) When $p = q$

$$QIRF^{(s)} = \begin{cases} F_e^{-1}(\tau)\boldsymbol{\alpha}_1|\boldsymbol{\delta}^0|, & \text{for } s = 1, \\ F_e^{-1}(\tau)\boldsymbol{\alpha}_s|\boldsymbol{\delta}^0| + \sum_{i=1}^{s-1} \Big[\big(\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big] + \boldsymbol{\beta}_i \big) QIRF^{(s-i)} \Big], & \text{for } 2 \le s \le p, \\ \sum_{i=1}^{p} \Big[\big(\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big] + \boldsymbol{\beta}_i \big) QIRF^{(s-i)} \Big], & \text{for } s \ge p+1. \end{cases}$$

(ii) When $p < q$

$$
QIRF^{(s)} = \begin{cases} F_e^{-1}(\tau)\boldsymbol{\alpha}_1|\boldsymbol{\delta}^0|, & \text{for } s = 1, \\ F_e^{-1}(\tau)\boldsymbol{\alpha}_s|\boldsymbol{\delta}^0| + \sum_{i=1}^{s-1}\Big[\big(\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big] + \boldsymbol{\beta}_i\big)QIRF^{(s-i)}\Big], & \text{for } 2 \le s \le p+1, \\ F_e^{-1}(\tau)\boldsymbol{\alpha}_s|\boldsymbol{\delta}^0| + \sum_{i=p+1}^{s-1}\Big[\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big]QIRF^{(s-i)}\Big] & \\ \qquad + \sum_{i=1}^{p}\Big[\big(\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big] + \boldsymbol{\beta}_i\big)QIRF^{(s-i)}\Big], & \text{for } p+2 \le s \le q, \\ \sum_{i=p+1}^{q}\Big[\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big]QIRF^{(s-i)}\Big] & \\ \qquad + \sum_{i=1}^{p}\Big[\big(\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big] + \boldsymbol{\beta}_i\big)QIRF^{(s-i)}\Big], & \text{for } s \ge q+1. \end{cases}
$$

(iii) When $p > q$

$$
QIRF^{(s)} = \begin{cases} F_e^{-1}(\tau)\boldsymbol{\alpha}_1|\boldsymbol{\delta}^0|, & \text{for } s = 1, \\ F_e^{-1}(\tau)\boldsymbol{\alpha}_s|\boldsymbol{\delta}^0| + \sum_{i=1}^{s-1}\Big[\big(\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big] + \boldsymbol{\beta}_i\big)QIRF^{(s-i)}\Big], & \text{for } 2 \le s \le q, \\ \sum_{i=q+1}^{s-1}\Big[\boldsymbol{\beta}_i QIRF^{(s-i)}\Big] & \\ \qquad + \sum_{i=1}^{q}\Big[\big(\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big] + \boldsymbol{\beta}_i\big)QIRF^{(s-i)}\Big], & \text{for } q+1 \le s \le p+1, \\ \sum_{i=q+1}^{p}\Big[\boldsymbol{\beta}_i QIRF^{(s-i)}\Big] & \\ \qquad + \sum_{i=1}^{q}\Big[\big(\boldsymbol{\alpha}_i E\big[|\mathbf{e}_t|\big] + \boldsymbol{\beta}_i\big)QIRF^{(s-i)}\Big], & \text{for } s \ge p+2. \end{cases}
$$

A.1.2 Local Projection for Univariate TS-GARCH(1,1)

We have

$$
\sigma_{t+s} = \omega + \sum_{j=1}^{s-1}\Big[\prod_{k=1}^{j}\big(\alpha|\epsilon_{t+s-k}| + \beta\big)\Big]\omega + \prod_{k=1}^{s-1}\big(\alpha|\epsilon_{t+s-k}| + \beta\big)\alpha|y_t| + \prod_{k=1}^{s-1}\big(\alpha|\epsilon_{t+s-k}| + \beta\big)\beta\sigma_t,
$$

for $s \ge 2$. With $\sigma_t = \omega + (\alpha|\epsilon_{t-1}| + \beta)\sigma_{t-1}$, σ_{t+s-1} is expressed as

$$
\sigma_{t+s-1} = \omega + \sum_{j=2}^{s-1}\Big[\prod_{k=2}^{j}(\alpha|\epsilon_{t+s-k}| + \beta)\Big]\omega + \Big[\prod_{k=2}^{s}(\alpha|\epsilon_{t+s-k}| + \beta)\Big]\sigma_t,
$$

which yields

$$
\sigma_t = \Big[\prod_{k=2}^{s}(\alpha|\epsilon_{t+s-k}| + \beta)\Big]^{-1}\Big[\sigma_{t+s-1} - \omega - \sum_{j=2}^{s-1}\Big[\prod_{k=2}^{j}(\alpha|\epsilon_{t+s-k}| + \beta)\Big]\omega\Big].
$$

Then σ_{t+s} can be rewritten as

$$\sigma_{t+s} = \omega + \sum_{j=1}^{s-1} \Big[\prod_{k=1}^{j} \big(\alpha|\epsilon_{t+s-k}| + \beta\big) \Big] \omega - \frac{\big(\alpha|\epsilon_{t+s-1}| + \beta\big)\beta}{\alpha|\epsilon_t| + \beta} \Big[\omega + \sum_{j=2}^{s-1} \Big[\prod_{k=2}^{j} (\alpha|\epsilon_{t+s-k}| + \beta) \Big] \omega \Big]$$
$$+ \Big[\prod_{k=1}^{s-1} \big(\alpha|\epsilon_{t+s-k}| + \beta\big) \Big] \alpha|y_t| + \frac{\big(\alpha|\epsilon_{t+s-1}| + \beta\big)\beta}{\alpha|\epsilon_t| + \beta} \sigma_{t+s-1}.$$

Assume both σ_{t+s-1} and y_t are given but $\{|\epsilon_p|\}_{p=t}^{t+s-1}$ are random. Define $X_t := \alpha|\epsilon_t| + \beta$. From the above expression, we can find function $g(\cdot)$ such that $\sigma_{t+s} = g(X_t, ..., X_{t+s-1})$, thus $y_{t+s} = \epsilon_{t+s} \cdot g(X_t, ..., X_{t+s-1})$. Let $\mu_X = E[X_t] = \alpha E[|\epsilon_t|] + \beta$ and $\bar{g} = g(\mu_X, ..., \mu_X)$, then

$$\begin{aligned} Pr\Big(y_{t+s} \le F_\epsilon^{-1}(\tau) \cdot \bar{g}\Big) &= Pr\Big(\epsilon_{t+s} \le \frac{F_\epsilon^{-1}(\tau) \cdot \bar{g}}{g(X_t, ..., X_{t+s-1})}\Big) \\ &= \int \cdots \int F_\epsilon\Big(\frac{F_\epsilon^{-1}(\tau) \cdot \bar{g}}{g(X_t, ..., X_{t+s-1})}\Big) dF_{|\epsilon_t|} \cdots dF_{|\epsilon_{t+s-1}|} \\ &= E\Big[F_\epsilon\Big(\frac{F_\epsilon^{-1}(\tau) \cdot \bar{g}}{g(X_t, ..., X_{t+s-1})}\Big)\Big]. \end{aligned}$$

The first-order approximation of $H(X_t, ..., X_{t+s-1}) := F_\epsilon\Big(\frac{F_\epsilon^{-1}(\tau)\cdot\bar{g}}{g(X_t,...,X_{t+s-1})}\Big)$ at the mean effect of innovations ($X_t = ... = X_{t+s-1} = \mu_X$) is

$$\begin{aligned} H(X_t, ..., X_{t+s-1}) &\approx H(\mu_X, ..., \mu_X) + \sum_{p=t}^{t+s-1} \frac{\partial H}{\partial X_p}(\mu_X, ..., \mu_X)(X_p - \mu_X) \\ &= F_\epsilon\Big(F_\epsilon^{-1}(\tau)\Big) + \sum_{p=t}^{t+s-1} \frac{\partial H}{\partial X_p}(\mu_X, ..., \mu_X)(|\epsilon_p| - E[|\epsilon_t|])\alpha \end{aligned} \quad (14)$$

Higher order terms are especially negligible when α is small because the l-th order terms has a factor α^l. In financial data, α is empirically known to be small enough that α^l is negligible for $l \ge 2$.[15] Thus, taking expectations on both sides of (14) yields the following approximation:

$$Pr\Big(y_{t+s} \le F_\epsilon^{-1}(\tau) \cdot \bar{g}\Big) \approx \tau.$$

That is, the τ-quantile of y_{t+s} approximates $F_\epsilon^{-1}(\tau) \cdot g(\mu_X, ..., \mu_X)$.

[15] Empirical evidence suggests that α is small, but β is large for financial returns because the major swing of volatility is attributable to the volatility-persistence channel. In GARCH models, the estimate for α is typically less than 0.1, and the estimate for β is larger than 0.9. For instance, Martin et al. (2013) estimate the GARCH(1,1) model for FTSE, DOW and NIKKEI, and their estimates of (α, β) are (0.08, 0.91), (0.05, 0.94) and (0.09, 0.90), respectively. For TS-GARCH(1,1) model given in Section 3.1, the estimates of (α, β) are (0.08, 0.92), (0.06, 0.95) and (0.09, 0.92), respectively.

i. Application of the Approximation to QIRF$^{(2)}$

$$
\begin{aligned}
q_{t+2} \approx F_\epsilon^{-1}(\tau) \cdot g(\mu_X, ..., \mu_X) &= F_\epsilon^{-1}(\tau)\Big[\omega + \mu_X\omega - \mu_X\beta\mu_X^{-1}\omega + \mu_X\alpha|y_t| + \mu_X\beta\mu_X^{-1}\sigma_{t+1}\Big] \\
&= F_\epsilon^{-1}(\tau)(\omega + \mu_X\omega - \beta\omega) + F_\epsilon^{-1}(\tau)\mu_X\alpha|y_t| + \beta q_{t+1}, \qquad (15)
\end{aligned}
$$

where

$$g(X_t, X_{t+1}) = \omega + X_{t+1}\omega - X_{t+1}\beta X_t^{-1}\omega + X_{t+1}\alpha|y_t| + X_{t+1}\beta X_t^{-1}\sigma_{t+1}.$$

$F_\epsilon^{-1}(\tau)\mu_X\alpha$, the coefficient of $|y_t|$ in (15), multiplied by $|\delta^0|$ is consistent with $QIRF^{(2)}$.

$$QIRF^{(2)} = F^{-1}(\tau)\underbrace{(\alpha E[|\epsilon_t|] + \beta)}_{=\mu_X}\alpha|\delta^0|.$$

ii. Application of the Approximation to QIRF$^{(s)}$ for $s \geq 3$

$$
\begin{aligned}
q_{t+s} &\approx F_\epsilon^{-1}(\tau) \cdot g(\mu_X, ..., \mu_X) \\
&= F_\epsilon^{-1}(\tau)\Bigg[\omega + \sum_{j=1}^{s-1}\mu_X^j\omega - \mu_X\beta\mu_X^{-1}\Big[\omega + \sum_{j=2}^{s-1}\mu_X^{j-1}\omega\Big] + \mu_X^{s-1}\alpha|y_t| + \mu_X\beta\mu_X^{-1}\sigma_{t+s-1}\Bigg] \\
&= F_\epsilon^{-1}(\tau)\Bigg[\omega + \sum_{j=1}^{s-1}\mu_X^j\omega - \beta\Big[\omega + \sum_{j=2}^{s-1}\mu_X^{j-1}\omega\Big]\Bigg] + F_\epsilon^{-1}(\tau)\mu_X^{s-1}\alpha|y_t| + \beta q_{t+s-1}, \qquad (16)
\end{aligned}
$$

where

$$
\begin{aligned}
g(X_t, ..., X_{t+s-1}) = \omega &+ \sum_{j=1}^{s-1}\Big(\prod_{k=1}^{j}X_{t+s-k}\Big)\omega - X_{t+s-1}\beta X_t^{-1}\Bigg[\omega + \sum_{j=2}^{s-1}\Big(\prod_{k=2}^{j}X_{t+s-k}\Big)\omega\Bigg] \\
&+ \Big(\prod_{k=1}^{s-1}X_{t+s-k}\Big)\alpha|y_t| + X_{t+s-1}\beta X_t^{-1}\sigma_{t+s-1}.
\end{aligned}
$$

$F_\epsilon^{-1}(\tau)k^{s-1}\alpha$, the coefficient of $|y_t|$ in (16), multiplied by $|\delta^0|$ is consistent with $QIRF^{(s)}$.

$$QIRF^{(s)} = F^{-1}(\tau)\underbrace{(\alpha E[|\epsilon_t|] + \beta)^{s-1}}_{=\mu_X^{s-1}}\alpha|\delta^0|.$$

As a result, $QIRF^{(s)}$ can be effectively constructed as

$$\widehat{QIRF}^{(s)} = \widehat{a}^{(s)}|\delta^0|,$$

where $\widehat{a}^{(s)}$ is the estimate for $a^{(s)}$ of the following slightly adjusted CAViaR model

$$q_t = c^{(s)} + a^{(s)}|y_{t-s}| + b^{(s)}q_{t-1}, \quad \text{for} \quad s = 1, 2, ..., S.$$

A.1.3 Local Projection for Bivariate TS-GARCH(1,1)

With $\mathbf{y}_t = \mathbf{e}_t \boldsymbol{\Sigma}_t$ and $\boldsymbol{\Sigma}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}|\mathbf{y}_{t-1}| + \boldsymbol{\beta}\boldsymbol{\Sigma}_{t-1}$ we have

$$\boldsymbol{\Sigma}_{t+s} = \boldsymbol{\omega} + \sum_{j=1}^{s-1} \Big[\prod_{k=1}^{j} \big(\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta}\big)\Big]\boldsymbol{\omega} + \Big[\prod_{k=1}^{s-1} \big(\ \boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta}\big)\Big](\boldsymbol{\alpha}|\mathbf{y}_t| + \boldsymbol{\beta}\boldsymbol{\Sigma}_t). \tag{17}$$

Since the dynamics of $\boldsymbol{\Sigma}_t$ can be rewritten as $\boldsymbol{\Sigma}_t = \boldsymbol{\omega} + (\boldsymbol{\alpha}|\mathbf{e}_{t-1}| + \boldsymbol{\beta})\boldsymbol{\Sigma}_{t-1}$, $\boldsymbol{\Sigma}_{t+s-1}$ is expressed as

$$\boldsymbol{\Sigma}_{t+s-1} = \boldsymbol{\omega} + \sum_{j=2}^{s-1} \Big[\prod_{k=2}^{j} \big(\ \boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta}\big)\Big]\boldsymbol{\omega} + \Big[\prod_{k=2}^{s} (\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta})\Big]\boldsymbol{\Sigma}_t,$$

which yields

$$\boldsymbol{\Sigma}_t = \Big[\prod_{k=2}^{s} (\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta})\Big]^{-1}\Big[\boldsymbol{\Sigma}_{t+s-1} - \boldsymbol{\omega} - \sum_{j=2}^{s-1}\Big[\prod_{k=2}^{j}\big(\ \boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta}\big)\Big]\boldsymbol{\omega}\Big]. \tag{18}$$

Combining (17) and (18), Σ_{t+s} for $s \geq 2$ can be rewritten as

$$\begin{aligned}
\boldsymbol{\Sigma}_{t+s} = {} & \boldsymbol{\omega} + \sum_{j=1}^{s-1} \Big[\prod_{k=1}^{j} \big(\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta}\big)\Big]\boldsymbol{\omega} \\
& - \Big[\prod_{k=1}^{s-1} \big(\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta}\big)\Big]\boldsymbol{\beta}\Big[\prod_{k=2}^{s} (\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta})\Big]^{-1}\Big[\boldsymbol{\omega} + \sum_{j=2}^{s-1}\Big[\prod_{k=2}^{j}(\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta})\Big]\boldsymbol{\omega}\Big] \\
& + \Big[\prod_{k=1}^{s-1} \big(\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta}\big)\Big]\boldsymbol{\alpha}|\mathbf{y}_t| + \Big[\prod_{k=1}^{s-1} \big(\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta}\big)\Big]\ \boldsymbol{\beta}\Big[\prod_{k=2}^{s}(\boldsymbol{\alpha}|\mathbf{e}_{t+s-k}| + \boldsymbol{\beta})\Big]^{-1}\boldsymbol{\Sigma}_{t+s-1}.
\end{aligned} \tag{19}$$

Like the univariate case, $\boldsymbol{\Sigma}_{t+s}(= \boldsymbol{\omega} + \boldsymbol{\alpha}|\mathbf{y}_{t+s-1}| + \boldsymbol{\beta}\boldsymbol{\Sigma}_{t+s-1})$ is a function of $|\mathbf{y}_t|$ and $\boldsymbol{\Sigma}_{t+s-1}$ in the above expression. Accordingly, the expression includes innovations between t and $t+s-1$, and they have the same functional form (i.e., $\{\boldsymbol{\alpha}|\mathbf{e}_p| + \boldsymbol{\beta}\}_{p=t}^{t+s-1}$).

Assume both $\boldsymbol{\Sigma}_{t+s-1}$ and $\mathbf{y}_t$ are given but $\{|\mathbf{e}_p|\}_{p=t}^{t+s}$ are random. Using *vec* operator which stacks the columns of a matrix, define $\mathbf{X}_t = (X_{1t}, X_{2t}, X_{3t}, X_{4t}) = vec\big(\boldsymbol{\alpha}|\mathbf{e}_t| + \boldsymbol{\beta}\big)^{\top}$. From (19), we can find function $\mathbf{g}(\cdot)$ such that

$$\boldsymbol{\Sigma}_{t+s} = \begin{bmatrix}\sigma_{1t+s} \\ \sigma_{2t+s}\end{bmatrix} = \mathbf{g}(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1}) = \begin{bmatrix} g_1(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1}) \\ g_2(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1})\end{bmatrix},$$

thus $y_{it+s} = e_{it+s} \cdot g_i(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1})$ for $i = 1, 2$. Let $\boldsymbol{\mu}_{\mathbf{X}} = (\mu_{X_1}, \mu_{X_2}, \mu_{X_3}, \mu_{X_4}) = E[\mathbf{X}_t]$ and

$\bar{\mathbf{g}} = (\bar{g}_1, \bar{g}_2)^\top = \mathbf{g}(\boldsymbol{\mu}_{\mathbf{X}}, ..., \boldsymbol{\mu}_{\mathbf{X}})$. For $i = 1, 2$, we have

$$
\begin{aligned}
Pr\Big(y_{it+s} \le F_e^{-1}(\tau)\cdot\bar{g}_i\Big) &= Pr\Bigg(e_{it+s} \le \frac{F_e^{-1}(\tau)\cdot\bar{g}_i}{g_i(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1})}\Bigg) \\
&= \int\int\cdots\int\int F_e\Big(\frac{F_e^{-1}(\tau)\cdot\bar{g}_i}{g_i(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1})}\Big) dF_{|e_{1t}|}dF_{|e_{2t}|}\cdots dF_{|e_{1t+s-1}|}dF_{|e_{2t+s-1}|} \\
&= E\Big[F_e\Big(\frac{F_e^{-1}(\tau)\cdot\bar{g}_i}{g_i(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1})}\Big)\Big].
\end{aligned}
$$

Let $H(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1}) = F_e\Big(\frac{F_e^{-1}(\tau)\cdot\bar{g}_i}{g_i(\mathbf{X}_t,...,\mathbf{X}_{t+s-1})}\Big)$. As $H(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1})$ can be expressed as a function of $X_{1t}, ..., X_{4t}, X_{1t+1}, ..., X_{4t+1}, ..., X_{1t+s-1}, ..., X_{4t+s-1}$, define $\tilde{H}(X_{1t}, ..., X_{4t+s-1}) = H(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1})$. Then, the first-order approximation of $H(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1})$ at the mean effect of innovations ($\mathbf{X}_t = ... = \mathbf{X}_{t+s-1} = \boldsymbol{\mu}_{\mathbf{X}}$) is

$$
\begin{aligned}
&H(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1}) \\
&\quad\approx H(\boldsymbol{\mu}_{\mathbf{X}}, ..., \boldsymbol{\mu}_{\mathbf{X}}) + \sum_{p=t}^{t+s-1}\sum_{n=1}^{4}\frac{\partial\tilde{H}}{\partial X_{np}}(\mu_{X_1}, ..., \mu_{X_4})(X_{np} - \mu_{X_n}) \\
&\quad= F_e\Big(F_e^{-1}(\tau)\Big) + \sum_{p=t}^{t+s-1}\sum_{n=1}^{4}\frac{\partial\tilde{H}}{\partial X_{np}}(\mu_{X_1}, ..., \mu_{X_4})(|\tilde{e}_{np}| - E[|\tilde{e}_{nt}|])\tilde{\alpha}_n \qquad (20)
\end{aligned}
$$

where

$$
\tilde{e}_{nt} = \begin{cases} e_{1t}, & \text{for } n = 1, \\ e_{1t}, & \text{for } n = 2, \\ e_{2t}, & \text{for } n = 3, \\ e_{2t}, & \text{for } n = 4, \end{cases}, \qquad \tilde{\alpha}_n = \begin{cases} \alpha_{11}, & \text{for } n = 1, \\ \alpha_{21}, & \text{for } n = 2, \\ \alpha_{12}, & \text{for } n = 3, \\ \alpha_{22}, & \text{for } n = 4. \end{cases}
$$

Higher order terms are especially negligible under small-valued elements of $\boldsymbol{\alpha}$. In financial data, elements of $\boldsymbol{\alpha}$ is empirically known to be small enough (less than 0.1) that $\prod_{j=1}^{l}\tilde{\alpha}_{n_j}$ is negligible for $l \ge 2$. Thus, taking expectations on both sides of (20) yields the following approximation:

$$
Pr\Big(y_{it+s} \le F_e^{-1}(\tau)\cdot\bar{g}_i\Big) \approx \tau. \qquad (21)
$$

That is, the τ-quantile of y_{it+s} approximates $F_e^{-1}(\tau)\cdot g_i(\mu_{X_1}, ..., \mu_{X_4})$.

i. Application of the Approximation to QIRF$^{(2)}$

With (19) and (21),

$$\mathbf{q}_{t+2} = \begin{bmatrix} q_{1t+2} \\ q_{2t+2} \end{bmatrix} \approx F_e^{-1}(\tau)\bar{\mathbf{g}} = F_e^{-1}(\tau)\big[\boldsymbol{\omega} + \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\omega} - \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\beta}\boldsymbol{\mu}_{\mathbf{X}}^{-1}\boldsymbol{\omega} + \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\alpha}|\mathbf{y}_t| + \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\beta}\boldsymbol{\mu}_{\mathbf{X}}^{-1}\boldsymbol{\Sigma}_{t+1}\big]$$
$$= F_e^{-1}(\tau)\big[\boldsymbol{\omega} + \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\omega} - \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\beta}\boldsymbol{\mu}_{\mathbf{X}}^{-1}\boldsymbol{\omega}\big] + F_e^{-1}\boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\alpha}|\mathbf{y}_t| + \boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\beta}\boldsymbol{\mu}_{\mathbf{X}}^{-1}\mathbf{q}_{t+1}, \tag{22}$$

where

$$\mathbf{g}(\mathbf{X}_t, \mathbf{X}_{t+1}) = \boldsymbol{\omega} + \mathbf{X}_{t+1}\boldsymbol{\omega} - \mathbf{X}_{t+1}\boldsymbol{\beta}\mathbf{X}_t^{-1}\boldsymbol{\omega} + \mathbf{X}_{t+1}\boldsymbol{\alpha}|\mathbf{y}_t| + \mathbf{X}_{t+1}\boldsymbol{\beta}\mathbf{X}_t^{-1}\boldsymbol{\Sigma}_{t+1}.$$

$F_e^{-1}\boldsymbol{\mu}_{\mathbf{X}}\boldsymbol{\alpha}$, the coefficient of $|\mathbf{y}_t|$ in (22), post-multiplied by $|\boldsymbol{\delta}^0|$ is consistent with $QIRF^{(2)}$.

$$QIRF^{(2)} = F_e^{-1}(\tau)\underbrace{(\boldsymbol{\alpha}E[|\mathbf{e}_t|] + \boldsymbol{\beta})}_{=\boldsymbol{\mu}_{\mathbf{X}}}\boldsymbol{\alpha}|\boldsymbol{\delta}^{\mathbf{0}}|.$$

ii. Application of the Approximation to QIRF$^{(s)}$ for $s \geq 3$

With (19) and (21),

$$\mathbf{q}_{t+s} = \begin{bmatrix} q_{1t+s} \\ q_{2t+s} \end{bmatrix} \approx F_e^{-1}(\tau)\mathbf{g}(\boldsymbol{\mu}_{\mathbf{X}}, ..., \boldsymbol{\mu}_{\mathbf{X}})$$
$$= F_e^{-1}(\tau)\Bigg[\boldsymbol{\omega} + \sum_{j=1}^{s-1}\boldsymbol{\mu}_{\mathbf{X}}^{j}\boldsymbol{\omega} - \boldsymbol{\mu}_{\mathbf{X}}^{s-1}\boldsymbol{\beta}\Big[\boldsymbol{\mu}_{\mathbf{X}}^{s-1}\Big]^{-1}\Big[\boldsymbol{\omega} + \sum_{j=2}^{s-1}\boldsymbol{\mu}_{\mathbf{X}}^{j-1}\boldsymbol{\omega}\Big]$$
$$+ \boldsymbol{\mu}_{\mathbf{X}}^{s-1}\boldsymbol{\alpha}|\mathbf{y}_t| + \boldsymbol{\mu}_{\mathbf{X}}^{s-1}\boldsymbol{\beta}\Big[\boldsymbol{\mu}_{\mathbf{X}}^{s-1}\Big]^{-1}\boldsymbol{\Sigma}_{t+s-1}\Bigg]$$
$$= F_e^{-1}(\tau)\Bigg[\boldsymbol{\omega} + \sum_{j=1}^{s-1}\boldsymbol{\mu}_{\mathbf{X}}^{j}\boldsymbol{\omega} - \boldsymbol{\mu}_{\mathbf{X}}^{s-1}\boldsymbol{\beta}\Big[\boldsymbol{\mu}_{\mathbf{X}}^{s-1}\Big]^{-1}\Big[\boldsymbol{\omega} + \sum_{j=2}^{s-1}\boldsymbol{\mu}_{\mathbf{X}}^{j-1}\boldsymbol{\omega}\Big]\Bigg]$$
$$+ F_e^{-1}(\tau)\boldsymbol{\mu}_{\mathbf{X}}^{s-1}\boldsymbol{\alpha}|\mathbf{y}_t| + \boldsymbol{\mu}_{\mathbf{X}}^{s-1}\boldsymbol{\beta}\Big[\boldsymbol{\mu}_{\mathbf{X}}^{s-1}\Big]^{-1}\mathbf{q}_{t+s-1}, \tag{23}$$

where

$$\mathbf{g}(\mathbf{X}_t, ..., \mathbf{X}_{t+s-1}) = \boldsymbol{\omega} + \sum_{j=1}^{s-1}\Big[\prod_{k=1}^{j}\mathbf{X}_{t+s-k}\Big]\boldsymbol{\omega} - \Big[\prod_{k=1}^{s-1}\mathbf{X}_{t+s-k}\Big]\boldsymbol{\beta}\Big[\prod_{k=2}^{s}\mathbf{X}_{t+s-k}\Big]^{-1}\Big[\boldsymbol{\omega} + \sum_{j=2}^{s-1}\Big[\prod_{k=2}^{j}\mathbf{X}_{t+s-k}\Big]\boldsymbol{\omega}\Big]$$
$$+ \Big[\prod_{k=1}^{s-1}\mathbf{X}_{t+s-k}\Big]\boldsymbol{\alpha}|\mathbf{y}_t| + \Big[\prod_{k=1}^{s-1}\mathbf{X}_{t+s-k}\Big]\boldsymbol{\beta}\Big[\prod_{k=2}^{s}\mathbf{X}_{t+s-k}\Big]^{-1}\boldsymbol{\Sigma}_{t+s-1}.$$

$F_e^{-1}(\tau)\boldsymbol{\mu}_{\mathbf{X}}^{s-1}\boldsymbol{\alpha}$, the coefficient of $|\mathbf{y}_t|$ in (23), post-multiplied by $|\boldsymbol{\delta}^0|$ is consistent with $QIRF^{(s)}$:

$$QIRF^{(s)} = F_e^{-1}(\tau)\underbrace{(\boldsymbol{\alpha}E[|\mathbf{e}_t|] + \boldsymbol{\beta})^{s-1}}_{=\boldsymbol{\mu}_{\mathbf{X}}^{s-1}}\boldsymbol{\alpha}|\boldsymbol{\delta}^{\mathbf{0}}|.$$

Thus, $QIRF^{(s)}$ can be effectively obtained as

$$\widehat{QIRF}^{(s)} = \widehat{\mathbf{A}}^{(s)}|\boldsymbol{\delta}^0|,$$

where $\widehat{\mathbf{A}}^{(s)}$ is the estimate for $\mathbf{A}^{(s)}$ of the following model

$$\mathbf{q}_t = \mathbf{c}^{(s)} + \mathbf{A}^{(s)}|\mathbf{y}_{t-s}| + \mathbf{B}^{(s)}\mathbf{q}_{t-1}, \quad \text{for } \ s = 1, 2, ..., S.$$

A.2 Proofs of Theorems and Lemmas in Section 4

The Proofs for theorems and lemmas in Section 4 adopt and extend from Engle and Manganelli (2004) and WKM.

Proof of Lemma 4.1. We can prove the lemma by verifying the conditions of Corollary 5.11 of White (1994). Assumption 4.1 and 4.3(1) ensure White's Assumption 2.1 and 5.1, respectively. By rewriting the linear program $\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})$ solves as

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau}) &= \underset{\boldsymbol{\gamma}^{(s)}\in\mathbb{A}}{\arg\min}\frac{1}{T}\sum_{t=1}^{T}\Big[\sum_{i=1}^{n}\rho_{\tau_i}\big(y_{it} - q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)})\big)\Big]\\ &= \underset{\boldsymbol{\gamma}^{(s)}\in\mathbb{A}}{\arg\max}\frac{1}{T}\sum_{t=1}^{T}\phi\big(\mathbf{Y}_t,\mathbf{q}_t(\cdot,\boldsymbol{\gamma}^{(s)})\big),\end{aligned}$$

where $\phi\big(\mathbf{Y}_t,\mathbf{q}_t(\cdot,\boldsymbol{\gamma}^{(s)})\big) = -\sum_{i=1}^{n}\rho_{\tau_i}\big(y_{it} - q_{it}^0(\cdot,\ \boldsymbol{\gamma}^{(s)})\big)$, White's Assumption 5.4 is obviously satisfied.

Next, we verify White's Assumption 3.1. We have

$$\begin{aligned}\big|\phi\big(\mathbf{Y}_t,\mathbf{q}_t(\cdot,\boldsymbol{\gamma}^{(s)})\big)\big| &\le \sum_{i=1}^{n}\Big|\big(y_{it} - q_{it}^0(\cdot,\ \boldsymbol{\gamma}^{(s)})\big)\Big(\tau_i - 1\big[y_{it} - q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)}) \le 0\big]\Big)\Big|\\ &\le \sum_{i=1}^{n}\big(|y_{it}| + |q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)})|\big)\\ &\le \sum_{i=1}^{n}|y_{it}| + nD_{0t},\end{aligned}$$

where the last inequality comes from Assumption 4.5. By Assumption 4.5(1, 2), $E\Big[\big|\phi\big(\mathbf{Y}_t,\mathbf{q}_t(\cdot,\boldsymbol{\gamma}^{(s)})\big)\big|\Big]$ is dominated ensuring White's Assumption 3.1(a). Since $q_{it}^0(\cdot,\ \boldsymbol{\gamma}^{(s)})$ is continuous in $\boldsymbol{\gamma}^{(s)}$, $\phi\big(\mathbf{Y}_t,\mathbf{q}_t(\cdot,\boldsymbol{\gamma}^{(s)})\big)$ is continuous, thus its expected value is also continuous (ensuring White's Assumption 3.1(b)). Assumption 4.1 and 4.3(1) ensure stationary and strong mixing (ensuring White's Assumption 3.1(c)).

White's Assumption 3.2 remains to be verified, which is the condition that $\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})$ is the unique maximizer of $E\big[\phi\big(\mathbf{Y}_t, \mathbf{q}_t(\cdot, \boldsymbol{\gamma}^{(s)})\big)\big]$. Consider $\boldsymbol{\gamma}^{(s)} \neq \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})$ such that $\|\boldsymbol{\gamma}^{(s)} - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\| > \epsilon$, and define

$$\Delta_{it}(\boldsymbol{\gamma}^{(s)}) := \rho_{\tau_i}\big(y_{it} - q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})\big) - \rho_{\tau_i}\big(y_{it} - q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big).$$

It will suffice to show that $E\big[\sum_{i=1}^n \Delta_{it}(\boldsymbol{\gamma}^{(s)})\big] > 0$. With $\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) = q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)}) - q_{it}^0\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)$ and $u_{it} = y_{it} - q_{it}(\tau_i)$, we have the following equation by Assumption 4.2(2.b) and 4.4(1);

$$\begin{aligned}
\Delta_{it}(\boldsymbol{\gamma}^{(s)}) &= \big(y_{it} - q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})\big)\big(\tau_i - 1[y_{it} < q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})]\big) - \Big(y_{it} - q_{it}^0\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)\Big(\tau_i - 1\big[y_{it} < q_{it}^0\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big]\Big) \\
&= \Big(u_{it} - \delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)\Big(\tau_i - 1\big[u_{it} < \delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big]\Big) - u_{it}\big(\tau_i - 1[u_{it} < 0]\big) \\
&= u_{it}\Big(1[u_{it} < 0] - 1\big[u_{it} < \delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big]\Big) - \delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big(\tau_i - 1\big[u_{it} < \delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big]\Big).
\end{aligned}$$

Since $E\big[1[u_{it} < 0]|\Omega_t\big] = \tau_i$, taking the conditional expectation yields

$$\begin{aligned}
E[\Delta_{it}(\boldsymbol{\gamma}^{(s)})|\Omega_t] &= 1\big[\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) < 0\big] \int_{\delta_{it}\left(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\right)}^{0} \Big(v - \delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big) f_{u_{it}}(v)dv \\
&\quad + 1\big[\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) > 0\big] \int_0^{\delta_{it}\left(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\right)} \Big(-v + \delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big) f_{u_{it}}(v)dv.
\end{aligned}$$

Reasoning following Powell (1984), Assumption 4.2(1, 2.a) implies there exists some $h > 0$ such that $f_{u_{it}}(v) > h$ whenever $|v| < h$. Hence, for any k sufficiently small such that $0 < k < h$, we have

$$\begin{aligned}
E\big[\Delta_{it}(\boldsymbol{\gamma}^{(s)})|\Omega_t\big] &\geq 1\big[\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) < -k\big] \int_{-k}^{0} (v + k) f_{u_{it}}(v)dv \\
&\qquad + 1\big[\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) > k\big] \int_0^k (-v + k) f_{u_{it}}(v)dv \\
&\geq 1\big[\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) < -k\big] \int_{-k}^{0} [(v + k)h]dv \\
&\qquad + 1\big[\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) > k\big] \int_0^k [(-v + k)h]dv \\
&= \frac{1}{2} h k^2 1\big[\big|\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big| > k\big].
\end{aligned}$$

Thus, taking the unconditional expectation yields

$$\begin{aligned}
E\Big[\sum_{i=1}^n \Delta_{it}(\boldsymbol{\gamma}^{(s)})\Big] &\geq \frac{1}{2} h k^2 E\Big[\sum_{i=1}^n 1\big[\big|\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big| > k\big]\Big] \\
&= \frac{1}{2} h k^2 \sum_{i=1}^n Pr\big[\big|\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big| > k\big] \\
&\geq \frac{1}{2} h k^2 Pr\Big[\bigcup_{i \in l} \big\{\big|\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\big| > k\big\}\Big] > 0,
\end{aligned}$$

where the last inequality follows from Assumption 4.4(1.b). ■

Proof of Lemma 4.2. The proof builds on Huber's (1967) theorem 3 as in Engle and Manganelli(2004) and WKM.

Step 1: Show the conditions for Huber's theorem hold. Assumptions (N-1) and (N-4) are obviously satisfied. We show

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{n}\nabla\hat{q}_{it}(\tau_i)\psi_{\tau_i}\big(y_{it}-\hat{q}_{it}(\tau_i)\big)=o_p(1), \tag{24}$$

where $\big(\hat{q}_{1t}(\tau_1),\hat{q}_{2t}(\tau_2),...,,\hat{q}_{nt}(\tau_n)\big)^\top=\mathbf{q}_t^0\big(\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})\big)$. The existence of $\nabla\hat{q}_{it}(\tau_i)$ is ensured by Assumption 4.3(2). Let e_i be the $d\times 1$ unit vector with the i-th element equal to one and the rest zero, and define

$$K_l(c):=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{n}\rho_{\tau_i}\big(y_{it}-q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+ce_l)\big),$$

for any real number c. By (11), $K_l(c)$ is minimized at $c=0$. Let $H_l(c)$ be the derivative of $K_l(c)$ with respect to c from the right:

$$H_l(c):=-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{n}\nabla_l q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+ce_l)\psi_{\tau_i}\big(y_{it}-q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+ce_l)\big),$$

where $\nabla_l q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+ce_l)$ is the l-th element of $\nabla q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+ce_l)$. Since $K_l(c)$ is continuous in c and achieves its minimum at $c=0$, for any $\epsilon>0$

$$\begin{aligned}
|H_l(0)| &\le H_l(\epsilon)-H_l(-\epsilon)\\
&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{n}\Big[-\nabla_l q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+\epsilon e_l)\psi_{\tau_i}\big(y_{it}-q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+\epsilon e_l)\big)\\
&\qquad\qquad+\nabla_l q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})-\epsilon e_l)\psi_{\tau_i}\big(y_{it}-q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})-\epsilon e_l)\big)\Big]\\
&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{n}\Big[\tau_i\big(\nabla_l q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})-\epsilon e_l)-\nabla_l q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+\epsilon e_l)\big)\\
&\qquad\qquad+\nabla_l q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+\epsilon e_l)1\big[y_{it}<q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})+\epsilon e_l)\big]\\
&\qquad\qquad-\nabla_l q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})-\epsilon e_l)1\big[y_{it}<q_{it}^0(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})-\epsilon e_l)\big]\Big].
\end{aligned}$$

Taking the limit for $\epsilon\to 0$ yields

$$|H_l(0)|\le\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{n}\big|\nabla_l q_{it}^0\big(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})\big)\big|1\big[y_{it}-q_{it}^0\big(\cdot,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})\big)=0\big].$$

By Lemma A.1 of Ruppert and Carroll (1980), with probability one there exists no vector $\boldsymbol{\gamma}^{(s)}$ such that $y_{it} = q^0_{it}\big(\cdot, \boldsymbol{\gamma}^{(s)}\big)$. Hence, $1\big[y_{it} - q^0_{it}\big(\cdot, \widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})\big) = 0\big] = o_p(1)$ for any given t and i. Therefore, $|H_l(0)| \xrightarrow{p} 0$. Since $H_l(0)$ is the l-th element of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[\sum_{i=1}^{n}\nabla\hat{q}_{it}(\tau_i)\psi_{\tau_i}\big(y_{it} - \hat{q}_{it}(\tau_i)\big)\Big]$, the claim in (24) is proven.

For each $\boldsymbol{\gamma}^{(s)} \in \mathbb{A}$, let us define the $d \times 1$ vector

$$\boldsymbol{\lambda}(\boldsymbol{\gamma}^{(s)}) := \sum_{i=1}^{n} E\Big[\nabla q^0_{it}(\cdot, \boldsymbol{\gamma}^{(s)})\psi_{\tau_i}\big(y_{it} - q^0_{it}(\cdot, \boldsymbol{\gamma}^{(s)})\big)\Big].$$

For Assumption (N-2), we show $\boldsymbol{\lambda}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) = 0$ from the followings:

$$\begin{aligned}
\boldsymbol{\lambda}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) &= \sum_{i=1}^{n} E\Big[E\Big[\nabla q^0_{it}(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}))\psi_{\tau_i}\Big(y_{it} - q^0_{it}\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)|\mathcal{F}_{t-1}\Big]\Big] \\
&= \sum_{i=1}^{n} E\Big[\nabla q^0_{it}(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}))E\Big[\psi_{\tau_i}\Big(y_{it} - q^0_{it}\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)|\mathcal{F}_{t-1}\Big]\Big] \\
&= \sum_{i=1}^{n} E\Big[\nabla q^0_{it}(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}))E\Big[\tau_i - 1[y_{it} < q^0_{it}\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)]|\mathcal{F}_{t-1}\Big]\Big] \\
&= 0.
\end{aligned}$$

$\boldsymbol{\lambda}(\boldsymbol{\gamma}^{(s)})$ can be rewritten as

$$\boldsymbol{\lambda}(\boldsymbol{\gamma}^{(s)}) = \sum_{i=1}^{n} E\Big[\nabla q^0_{it}(\cdot, \boldsymbol{\gamma}^{(s)})\int_{\delta_{it}\left(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\right)}^{0} f_{u_{it}}(v)dv\Big],$$

where $\delta_{it}\big(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) = q^0_{it}\big(\cdot, \boldsymbol{\gamma}^{(s)}\big) - q^0_{it}\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)$, $u_{it} = y_{it} - q_{it}(\tau_i)$ and $f_{u_{it}}(v) = \frac{d}{dv}F_{it}\Big(v + q^0_{it}\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)$ represents the conditional density of u_{it} with respect to Lebesgue measure. Assumption 4.3(2), 4.4(1) and 4.5(3) ensure the existence and finiteness of the above expression. The differentiability and domination conditions provided by Assumption 4.3(2) and 4.5(4) ensure the continuous differentiability of $\boldsymbol{\lambda}(\boldsymbol{\gamma}^{(s)})$ on $\mathbb{A}$:

$$\nabla\boldsymbol{\lambda}(\boldsymbol{\gamma}^{(s)}) := \sum_{i=1}^{n} E\Big[\nabla\big[\nabla^\top q^0_{it}(\cdot, \boldsymbol{\gamma}^{(s)})\int_{\delta_{it}\left(\boldsymbol{\gamma}^{(s)}, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\right)}^{0} f_{u_{it}}(v)dv\big]\Big].$$

Let $Q_l(\boldsymbol{\gamma}^{(s)})$ be the gradient of the l-th element of $\boldsymbol{\lambda}(\boldsymbol{\gamma}^{(s)})$ with respect to $\boldsymbol{\gamma}^{(s)}$. Since $\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})$ is interior to $\mathbb{A}$ by Assumption 4.4(2), the mean value theorem applies to each element of $\boldsymbol{\lambda}(\boldsymbol{\gamma}^{(s)})$ yielding

$$\boldsymbol{\lambda}(\boldsymbol{\gamma}^{(s)}) = Q_0(\boldsymbol{\gamma}^{(s)} - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})), \tag{25}$$

for $\boldsymbol{\gamma}^{(s)}$ in a convex compact neighborhood of $\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})$, where Q_0 is an $d \times d$ matrix with $1 \times d$ rows $Q_l(\bar{\boldsymbol{\gamma}}^{(s)}_{(l)}) = \nabla^\top\boldsymbol{\lambda}(\bar{\boldsymbol{\gamma}}^{(s)}_{(l)})$, $\bar{\boldsymbol{\gamma}}^{(s)}_{(l)}$ is a mean value (different for each l) lying on the segment connecting $\boldsymbol{\gamma}^{(s)}$ and $\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})$ with $l = 1, 2, .., d$.

Next, we show

$$\boldsymbol{\lambda}(\boldsymbol{\gamma}^{(s)}) = -\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big(\boldsymbol{\gamma}^{(s)} - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) + O(\|\boldsymbol{\gamma}^{(s)} - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\|^2), \tag{26}$$

for Assumption (N-3)(i). The chain rule and an application of Leibniz rule to $\int_{\delta_{it}\left(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\right)}^{0} f_{u_{it}}(v)dv$ then give

$$Q_l(\boldsymbol{\gamma}^{(s)}) = A_l(\boldsymbol{\gamma}^{(s)}) - B_l(\boldsymbol{\gamma}^{(s)}),$$

where

$$\begin{aligned}
A_l(\boldsymbol{\gamma}^{(s)}) &:= \sum_{i=1}^{n} E\Big[\nabla_l \nabla^\top q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)}) \int_{\delta_{it}\left(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\right)}^{0} f_{u_{it}}(v)dv\Big],\\
B_l(\boldsymbol{\gamma}^{(s)}) &:= \sum_{i=1}^{n} E\Big[f_{u_{it}}\Big(\delta_{it}\big(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)\nabla_l q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)})\nabla^\top q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)})\Big].
\end{aligned}$$

By Assumption 4.2(3) and 4.5, we have

$$\|A_l(\boldsymbol{\gamma}^{(s)})\| \le \sum_{i=1}^{n} E\Big[D_{2t}\Big|\int_{\delta_{it}\left(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\right)}^{0} f_0 dv\Big|\Big] \le \sum_{i=1}^{n} E\Big[D_{2t} f_0 \Big|\delta_{it}\big(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big|\Big].$$

Application of the mean value theorem to the last line and Assumption 4.5(4) yield

$$\|A_l(\boldsymbol{\gamma}^{(s)})\| \le \sum_{i=1}^{n} E\Big[D_{2t} f_0 D_{1t}\|\boldsymbol{\gamma}^{(s)} - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\|\Big] = O(\|\boldsymbol{\gamma}^{(s)} - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\|).$$

Let us define $Q_l^* := \sum_{i=1}^{n} E\Big[f_{u_{it}}(0)\nabla_l q_{it}^0\big(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\nabla^\top q_{it}^0\big(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big]$. By Assumption 4.2(3), 4.5(3, 4) and application of the mean value theorem, we have

$$\begin{aligned}
\|B_l(\boldsymbol{\gamma}^{(s)}) - Q_l^*\| = \Big\| \sum_{i=1}^{n} E\Big[& f_{u_{it}}\Big(\delta_{it}\big(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)\nabla_l q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)})\nabla^\top q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)})\\
&- f_{u_{it}}\Big(\delta_{it}\big(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)\nabla_l q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)})\nabla^\top q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}))\\
&+ f_{u_{it}}\Big(\delta_{it}\big(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)\nabla_l q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)})\nabla^\top q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}))\\
&- f_{u_{it}}\Big(\delta_{it}\big(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)\nabla_l q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}))\nabla^\top q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}))\\
&+ f_{u_{it}}\Big(\delta_{it}\big(\boldsymbol{\gamma}^{(s)},\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)\nabla_l q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}))\nabla^\top q_{it}^0(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau}))\\
&- f_{u_{it}}(0)\nabla_l q_{it}^0\big(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\nabla^\top q_{it}^0\big(\cdot,\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big]\Big\|\\
\le \sum_{i=1}^{n} E\Big[& 2 f_0 D_{1t} D_{2t}\|\boldsymbol{\gamma}^{(s)} - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\| + L_0 D_{1t}^3\|\boldsymbol{\gamma}^{(s)} - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\|\Big]\\
= O(\|\boldsymbol{\gamma}^{(s)} &- \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\|).
\end{aligned}$$

Hence, we have

$$Q_0 = -\mathbf{Q}_{\boldsymbol{\tau}}^{(s)} + O(\|\boldsymbol{\gamma}^{(s)} - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\|), \tag{27}$$

where

$$\mathbf{Q}_{\boldsymbol{\tau}}^{(s)} = \sum_{i=1}^{n} E\Big[f_{u_{it}}(0)\nabla q_{it}^{0}\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\nabla^{\top} q_{it}^{0}\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big].$$

Combining (25) and (27) yields (26). With (26), the condition that $\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}$ is positive-definite in Assumption 4.6(1) is sufficient for Assumption (N-3)(i).

Next, we define

$$u_t(\boldsymbol{\gamma}^{(s)}, \delta) := \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\xi_t^{(s)}(\boldsymbol{\beta}) - \xi_t^{(s)}(\boldsymbol{\gamma}^{(s)})\|.$$

For Assumption (N-3)(ii), we have that for the given small $\delta > 0$

$$\begin{aligned}
u_t(\boldsymbol{\gamma}^{(s)}, \delta) &\le \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \sum_{i=1}^{n} \|\nabla q_{it}^{0}(\cdot,\boldsymbol{\beta})\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big) - \nabla q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\big)\| \\
&= \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \sum_{i=1}^{n} \|\nabla q_{it}^{0}(\cdot,\boldsymbol{\beta})\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big) - \nabla q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big) \\
&\qquad\qquad + \nabla q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big) - \nabla q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\big)\| \\
&\le \sum_{i=1}^{n} \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\nabla q_{it}^{0}(\cdot,\boldsymbol{\beta}) - \nabla q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\| \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big)\| \\
&\quad + \sum_{i=1}^{n} \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\nabla q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\| \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\big) - \psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big)\|.
\end{aligned}$$

Using (i) $\|\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big)\| \le 1$, (ii) the mean value theorem applied to $\nabla q_{it}^{0}(\cdot,\boldsymbol{\beta})$ and $\nabla q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})$ and (iii) Assumption 4.5, we have

$$\sum_{i=1}^{n} \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\nabla q_{it}^{0}(\cdot,\boldsymbol{\beta}) - \nabla q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\| \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big)\| \le nD_{2t}\delta. \tag{28}$$

Using (i) $\|\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\big) - \psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big)\| \le 1\big[|y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})| < |q_{it}^{0}(\cdot,\boldsymbol{\beta}) - q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})|\big]$, (ii) the mean value theorem applied to $q_{it}^{0}(\cdot,\boldsymbol{\beta})$ and $q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})$ and (iii) Assumption 4.5, we have

$$\begin{aligned}
&\sum_{i=1}^{n} \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\nabla q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\| \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})\big) - \psi_{\tau_i}\big(y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\beta})\big)\| \\
&\qquad\qquad \le D_{1t} \sum_{i=1}^{n} 1\big[|y_{it} - q_{it}^{0}(\cdot,\boldsymbol{\gamma}^{(s)})| < D_{1t}\delta\big].
\end{aligned} \tag{29}$$

Combining (28) and (29) yields

$$u_t(\boldsymbol{\gamma}^{(s)}, \delta) \le nD_{2t}\delta + D_{1t}\sum_{i=1}^{n} 1\big[|y_{it} - q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})| < D_{1t}\delta\big]. \tag{30}$$

By Assumption 4.2(3.a) and 4.5(3, 4), we have $E[u_t(\boldsymbol{\gamma}^{(s)}, \delta)] \le nE[D_{2t}]\delta + 2nE[D_{1t}^2]f_0\delta$. Hence, Assumption (N-3)(ii) holds for $b = nE[D_{2t}] + 2nE[D_{1t}^2]f_0$, $d = \delta$ and $d_0 = 2\delta$.

For Assumption (N-3)(iii), we have

$$\begin{aligned}
u_t(\boldsymbol{\gamma}^{(s)}, \delta) &\le \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \sum_{i=1}^{n} \|\nabla q_{it}^0(\cdot, \boldsymbol{\beta})\psi_{\tau_i}\big(y_{it} - q_{it}^0(\cdot, \boldsymbol{\beta})\big) - \nabla q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})\psi_{\tau_i}\big(y_{it} - q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})\big)\| \\
&\le \sum_{i=1}^{n} \Bigg[\sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\nabla q_{it}^0(\cdot, \boldsymbol{\beta})\psi_{\tau_i}\big(y_{it} - q_{it}^0(\cdot, \boldsymbol{\beta})\big)\| \\
&\qquad\qquad + \sup_{\{\boldsymbol{\beta}:\|\boldsymbol{\beta}-\boldsymbol{\gamma}^{(s)}\|\le\delta\}} \|\nabla q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})\psi_{\tau_i}\big(y_{it} - q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})\big)\| \Bigg] \\
&\le 2nD_{1t}.
\end{aligned}$$

Combining the last inequality with (30) yields

$$u_t(\boldsymbol{\gamma}^{(s)}, \delta)^2 \le 2n^2 D_{1t}D_{2t}\delta + 2nD_{1t}^2 \sum_{i=1}^{n} 1\big[|y_{it} - q_{it}^0(\cdot, \boldsymbol{\gamma}^{(s)})| < D_{1t}\delta\big].$$

In a similar way that Assumption (N-3)(ii) was verified, it can be shown that $E[u_t(\boldsymbol{\gamma}^{(s)}, \delta)^2] \le 2n^2E[D_{1t}D_{2t}]\delta + 4n^2E[D_{1t}^3]f_0\delta$. Given Assumption 4.5(3, 4), Assumption (N-3)(iii) holds for $c = 2n^2E[D_{1t}D_{2t}] + 4n^2E[D_{1t}^3]f_0$, $d = \delta$ and $d_0 = 2\delta$.

Step 2: Apply Huber's theorem. As a result of the above, we apply Huber's theorem:

$$\sqrt{T}\boldsymbol{\lambda}(\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})) + \mathbf{H}_T^{(s)} = o_p(1). \tag{31}$$

Consistency of $\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})$ and application of Slutsky's theorem to (26) yields

$$\boldsymbol{\lambda}\big(\,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau})\big) = -\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big(\,\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau}) - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) + o_p(1).$$

Combining the above equation with (31), we have

$$\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\sqrt{T}\Big(\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau}) - \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\Big) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \xi_t^{(s)}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) + o_p(1),$$

or (12).

Step 3: Apply the central limit theorem. $\xi_t^{(s)}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)$ is $\mathcal{F}_t$-measurable. Thus, $\{\xi_t^{(s)}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)$,

$\mathcal{F}_{t-1}\}$ is a stationary strong mixing martingale difference sequence (MDS) since

$$\begin{aligned}
E\big[\xi_t^{(s)}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) \mid \mathcal{F}_{t-1}\big] &= E\Big[\sum_{i=1}^{n} \nabla q_{it}^0\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\psi_{\tau_i}\Big(y_{it} - q_{it}^0\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big) \mid \mathcal{F}_{t-1}\Big] \\
&= \sum_{i=1}^{n} \nabla q_{it}^0\big(\cdot, \boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) E\big[\psi_{\tau_i}(u_{it}) \mid \mathcal{F}_{t-1}\big] \\
&= 0.
\end{aligned}$$

By the ergodic theorem, $\frac{1}{T}\sum_{t=1}^{T} \xi_t^{(s)}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big(\xi_t^{(s)}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)\Big)^\top \xrightarrow{p} \mathbf{V}_{\boldsymbol{\tau}}^{(s)}$. Assumption 4.5(3) ensures $\mathbf{V}_{\boldsymbol{\tau}}^{(s)}$ is finite and Assumption 4.6(2) ensures $\mathbf{V}_{\boldsymbol{\tau}}^{(s)}$ is positive definite. Therefore, application of the MDS central limit theorem (Theorem 5.24 of White, 2001) to (12) yields (13). ■

Proof of Theorem 4.1. Since $\mathbf{A}_{\boldsymbol{\tau}}^{(s)}|\boldsymbol{\delta}^0| = \mathbf{G}\boldsymbol{\gamma}_{\boldsymbol{\tau}}^{(s)}$, it follows by Lemma 4.2 that

$$\begin{aligned}
\sqrt{T}\big(\widehat{\mathbf{A}}^{(s)}|\boldsymbol{\delta}^0| - \mathbf{A}^{(s)}|\boldsymbol{\delta}^0|\big) &= \sqrt{T}\big(\mathbf{G}\widehat{\boldsymbol{\gamma}}^{(s)}(\boldsymbol{\tau}) - \mathbf{G}\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big) \\
&= \mathbf{G}\big(\mathbf{Q}_{\boldsymbol{\tau}}^{(s)}\big)^{-1}\mathbf{H}_T^{(s)} + o_p(1).
\end{aligned}$$

Applying the MDS central limit theorem yields the asymptotic distribution. ■

Proof of Lemma 4.5. We use the Cramer-Wold device. For some $l \in \mathbb{R}^d$, let $v_t^* = l^\top \xi_t^{(s)*}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)$ and $v_t = l^\top \xi_t^{(s)}\big(\boldsymbol{\gamma}^{(s)}(\boldsymbol{\tau})\big)$. Then, we can write

$$l^\top \mathbf{B}_T^{(s)*} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(v_t^* - v_t).$$

The original time-series is a stationary sequence satisfying the strong mixing condition in Assumption 4.1, and a measurable functions of mixing processes involving finite lagged variables satisfies the same mixing condition. Thus, $\{v_t\}_{t\in\mathbb{N}}$ is a stationary time-series satisfying Assumption 4.1. By Theorems 1 and 2 of Goncalves and de Jong (2003), the bootstrap estimate of the variance convergences to $\sigma_l^2 = l^\top \mathbf{V}_{\tau}^{(s)}$ in probability, and we obtain the distribution convergence conditional on the original sample. The limiting distribution of $\mathbf{B}_T^{(s)*}$ is obtained by applying the MDS central limit theorem as in the proof of Lemma 4.2. ■

Proof of Theorem 4.2. By Theorem 4.1 and Lemma 4.4, we have

$$\sqrt{T}\big(\ \widehat{\mathbf{A}}^{(s)*}|\boldsymbol{\delta}^0| - \widehat{\mathbf{A}}^{(s)}|\boldsymbol{\delta}^0|\ \big) = \mathbf{G}\big(\mathbf{Q}_{\tau}^{(s)}\big)^{-1}\big(\mathbf{H}_T^{(s)*} - \mathbf{H}_T^{(s)}\big) + o_p(1).$$

By Lemma 4.5

$$\mathbf{H}_T^{(s)*} - \mathbf{H}_T^{(s)} = \mathbf{B}_T^{(s)*} \xrightarrow{d} N\big(\mathbf{0}, \mathbf{V}_{\tau}^{(s)}\big),$$

conditional on the original sample, for almost every sequence. By the continuous mapping theorem and Slutsky's theorem, we obtain the desired result. ■

B Tables and Figures

Table 1. Simulation results: Coverage rates

Horizon	$s = 1$	$s = 10$	$s = 20$	$s = 30$
Response of q_{1t}				
Asymptotic CI	0.93	0.91	0.87	0.83
Bootstrap CI	0.92	0.94	0.95	0.96
Response of q_{2t}				
Asymptotic CI	0.92	0.90	0.87	0.81
Bootstrap CI	0.91	0.93	0.95	0.94

Note: The table reports coverage rates of asymptotic confidence intervals and stationary bootstrap confidence intervals.

Table 2. Estimation results of VAR for 5% VaR

JPM									
c_1		a_{11}		a_{12}		b_{11}		b_{12}	
-0.03	***	-0.14	***	-0.03	**	0.88	***	0.00	
(0.01)		(0.02)		(0.01)		(0.04)		(0.02)	
c_2		a_{21}		a_{22}		b_{21}		b_{22}	
-0.05	***	-0.18	***	-0.06	***	-0.14	**	0.98	***
(0.02)		(0.04)		(0.02)		(0.06)		(0.03)	
MS									
c_1		a_{11}		a_{12}		b_{11}		b_{12}	
-0.03	**	-0.15	***	-0.03	**	0.90	***	-0.01	
(0.01)		(0.04)		(0.01)		(0.03)		(0.01)	
c_2		a_{21}		a_{22}		b_{21}		b_{22}	
-0.04	*	-0.27	***	-0.11	***	-0.15	**	0.95	***
(0.02)		(0.07)		(0.03)		(0.06)		(0.02)	
AIG									
c_1		a_{11}		a_{12}		b_{11}		b_{12}	
-0.03	***	-0.13	***	-0.01	***	0.90	***	0.00	*
(0.01)		(0.03)		(0.00)		(0.02)		(0.00)	
c_2		a_{21}		a_{22}		b_{21}		b_{22}	
0.00		-0.02		-0.25	**	0.00		0.88	***
(0.05)		(0.12)		(0.11)		(0.08)		(0.04)	

Note: The coefficients correspond to the VAR for VaR model, which is the model given in (5) for $s = 1$. c_i, a_{ij}, b_{ij} are the elements of $c^{(1)}$, $A^{(1)}$, $B^{(1)}$, respectively. The quantiles for market and each financial institution are set to be 0.05. Estimated coefficients are in the first row. Standard errors are reported in parentheses. Asterisks indicate coefficient significance at the * 10%, ** 5% and *** 1% level.

Table 3. Estimation results of VAR for 1% VaR

JPM									
c_1		a_{11}		a_{12}		b_{11}		b_{12}	
-0.12	***	-0.23	***	-0.03		0.87	***	0.00	
(0.04)		(0.05)		(0.04)		(0.04)		(0.01)	
c_2		a_{21}		a_{22}		b_{21}		b_{22}	
-0.22	**	-0.44	***	-0.09		-0.26	**	0.99	***
(0.10)		(0.11)		(0.10)		(0.11)		(0.02)	
MS									
c_1		a_{11}		a_{12}		b_{11}		b_{12}	
-0.07	**	-0.21	**	-0.04		0.92	***	-0.02	
(0.03)		(0.10)		(0.06)		(0.04)		(0.03)	
c_2		a_{21}		a_{22}		b_{21}		b_{22}	
-0.14	*	-0.37		-0.24		-0.10		0.90	***
(0.08)		(0.42)		(0.25)		(0.12)		(0.09)	
AIG									
c_1		a_{11}		a_{12}		b_{11}		b_{12}	
-0.10	***	-0.19	**	-0.02	***	0.89	***	0.00	
(0.04)		(0.08)		(0.00)		(0.04)		(0.00)	
c_2		a_{21}		a_{22}		b_{21}		b_{22}	
-0.06		0.00		-0.77	***	0.05		0.77	***
(0.20)		(0.55)		(0.21)		(0.20)		(0.08)	

Note: The quantiles for market and each financial institution are set to be 0.01. Same as Table 2.

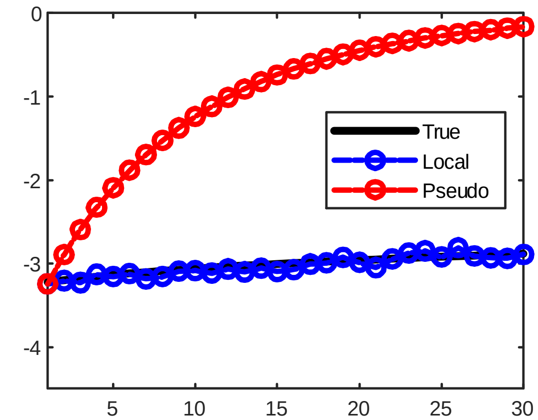
Table 4. Summary statistics of the full cross section of coefficients

5% VaR					
	$c1$	$a11$	$a12$	$b11$	$b12$
average	-0.03	-0.15	-0.03	0.89	0.01
std. dev.	0.01	0.02	0.02	0.04	0.03
min	-0.07	-0.21	-0.07	0.77	-0.04
max	0.00	-0.07	0.02	0.95	0.16
	$c2$	$a21$	$a22$	$b21$	$b22$
average	-0.03	-0.11	-0.14	0.06	0.85
std. dev.	0.10	0.08	0.08	0.46	0.26
min	-0.21	-0.28	-0.34	-0.27	-0.65
max	0.44	0.13	-0.01	2.83	1.06
1% VaR					
	$c1$	$a11$	$a12$	$b11$	$b12$
average	-0.11	-0.21	-0.03	0.86	0.01
std. dev.	0.04	0.06	0.03	0.07	0.05
min	-0.22	-0.34	-0.12	0.55	-0.08
max	0.00	-0.04	0.03	1.00	0.20
	$c2$	$a21$	$a22$	$b21$	$b22$
average	-0.03	-0.18	-0.30	0.22	0.75
std. dev.	0.29	0.23	0.19	0.57	0.30
min	-0.54	-0.67	-1.00	-0.29	-0.44
max	1.30	0.38	0.06	2.49	1.00

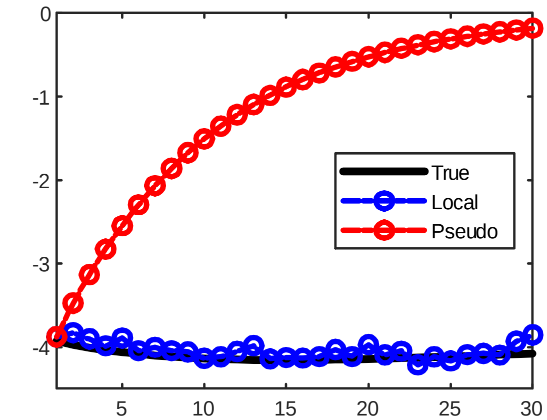
Note: The table reports the summary statistics of the coefficient estimates of the 61 bivariate models. Same as Table 2.

Figure 1. Simulation results of quantile impulse response

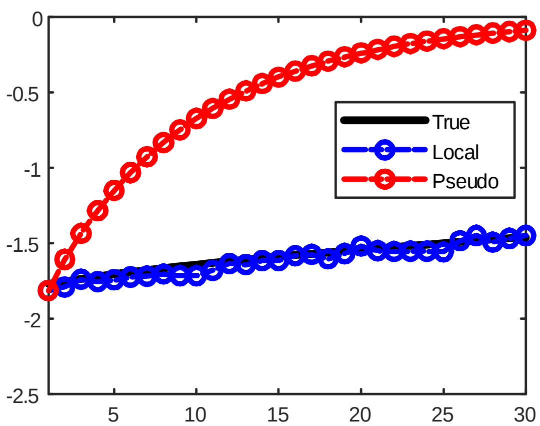
(a) Response of q_{1t} for DGP1



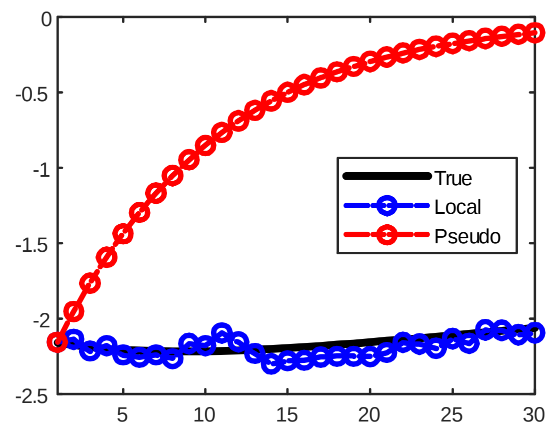
(b) Response of q_{2t} for DGP1



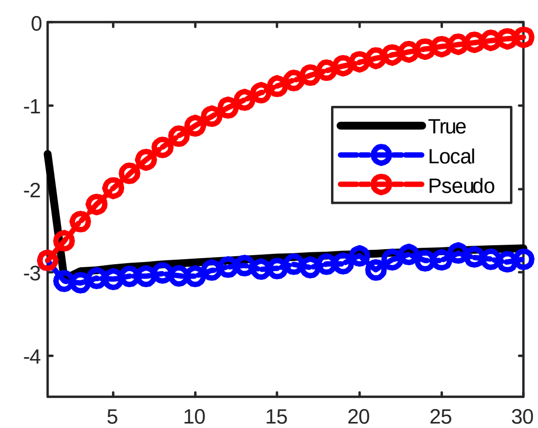
(c) Response of q_{1t} for DGP2



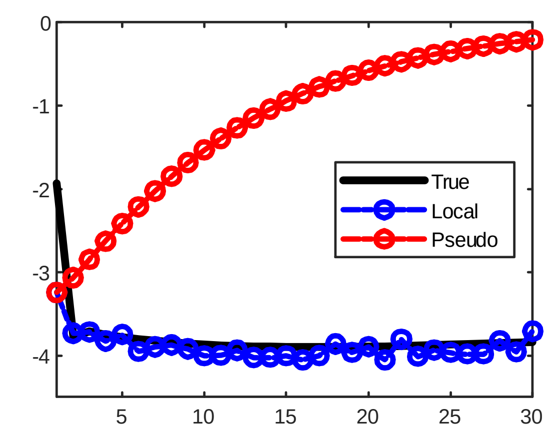
(d) Response of q_{2t} for DGP2



(e) Response of q_{1t} for DGP3



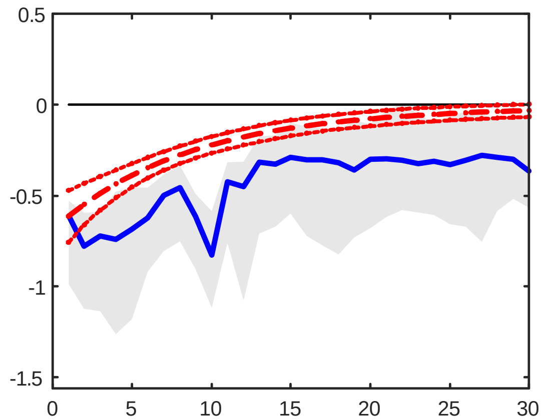
(f) Response of q_{2t} for DGP3



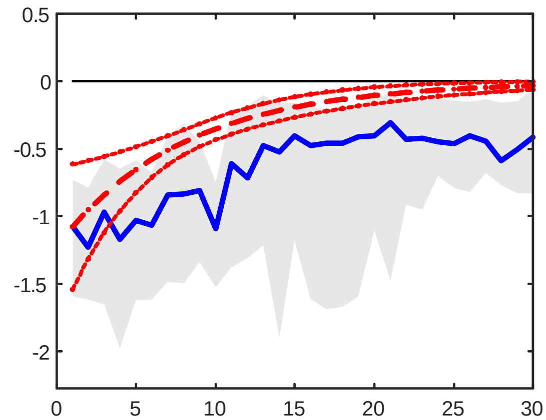
Note: Each figure presents a (analytically obtained) true QIRF, a local projection QIRF and a pseudo-QIRF for a given DGP. The local projection QIRF and pseudo-QIRF are averages based on 1,000 repetitions.

Figure 2. Quantile impulse response of each financial institution for 5% VaR

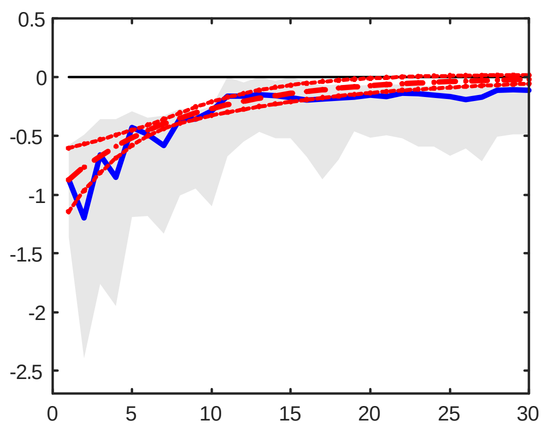
(a) Response of JPM



(b) Response of MS



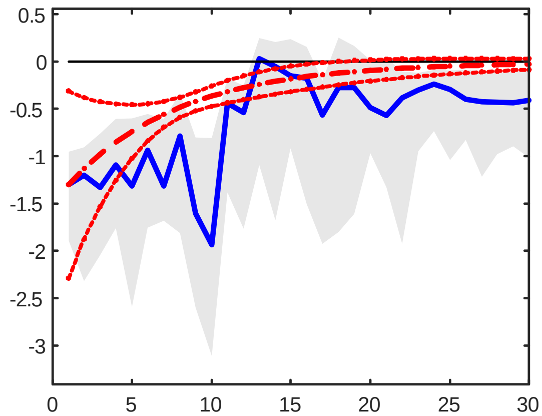
(c) Response of AIG



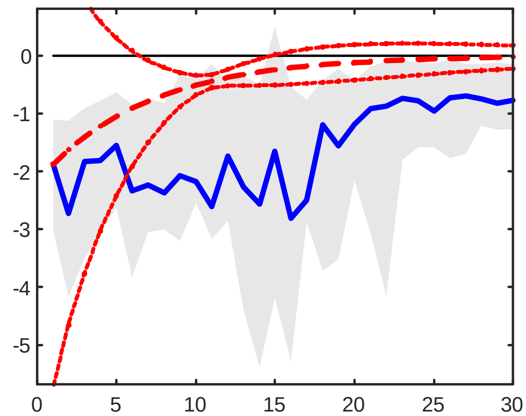
Note: The figures present QIRFs of individual financial institutions when there is a shock to the market. Blue solid lines are QIRFs from the local projection and shaded areas are their 95% confidence intervals based on the stationary bootstrap procedure. Red dashed lines are pseudo-QIRFs and dotted lines are their 95% confidence intervals.

Figure 3. Quantile impulse response of each financial institutions for 1% VaR

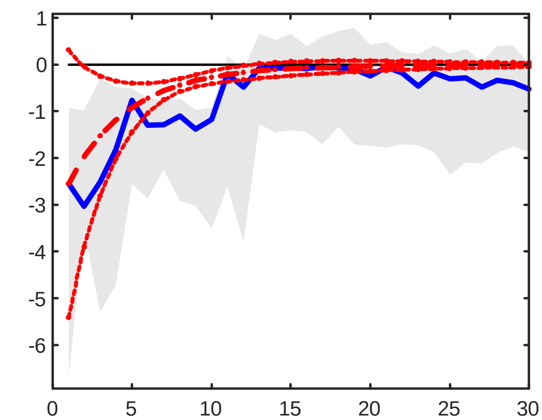
(a) Response of JPM



(b) Response of MS



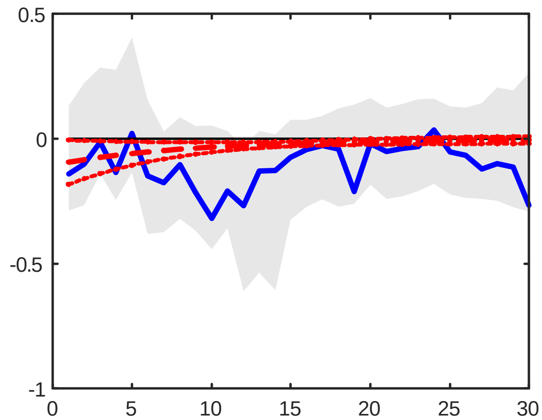
(c) Response of AIG



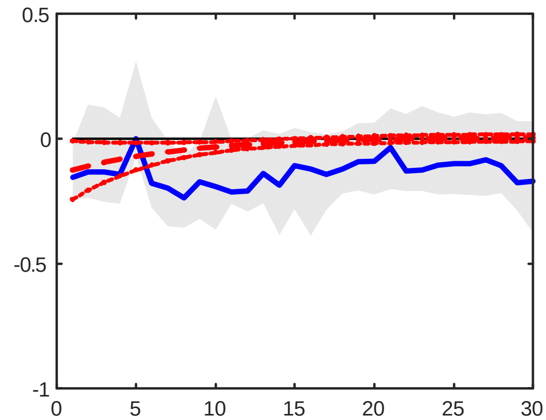
Note: Same as Figure 2.

Figure 4. [Systemic risk] Quantile impulse response of market for 5% VaR

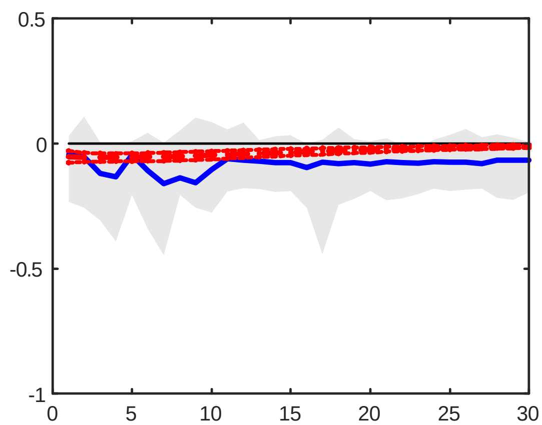
(a) Response to JPM



(b) Response to MS



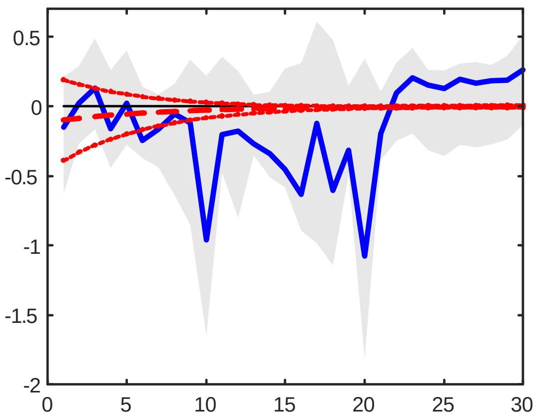
(c) Response to AIG



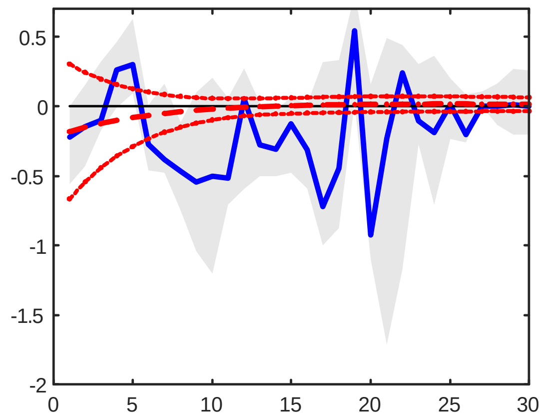
Note: The figures present QIRFs of the market when there is a shock to each individual financial institution. Same as Figure 2.

Figure 5. [Systemic risk] Quantile impulse response of market for 1% VaR

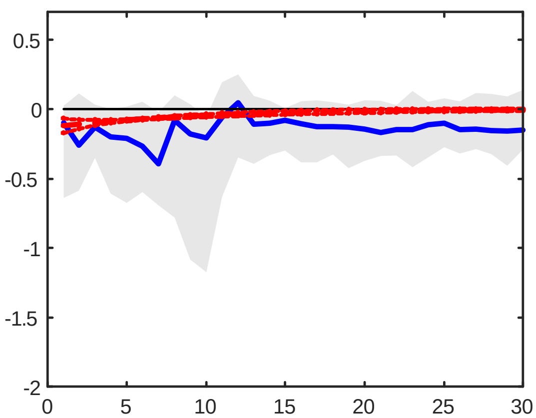
(a) Response to JPM



(b) Response to MS



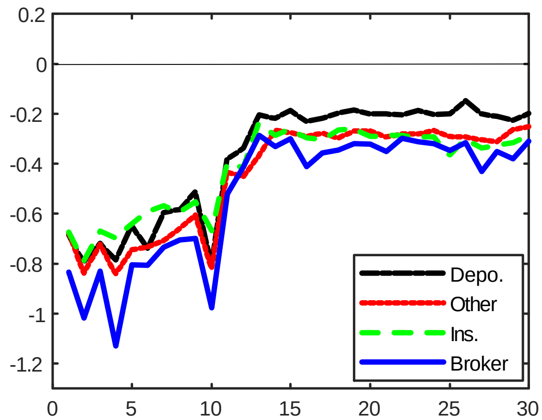
(c) Response to AIG



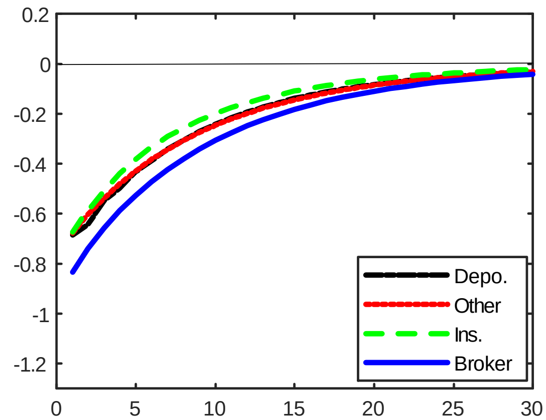
Note: The figures present QIRFs of the market when there is a shock to each individual financial institution. Same as Figure 2.

Figure 6. Sectoral averages of quantile impulse responses

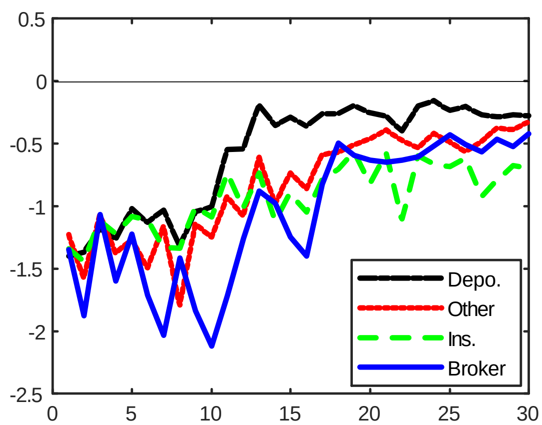
(A1) Local projection for 5% VaR



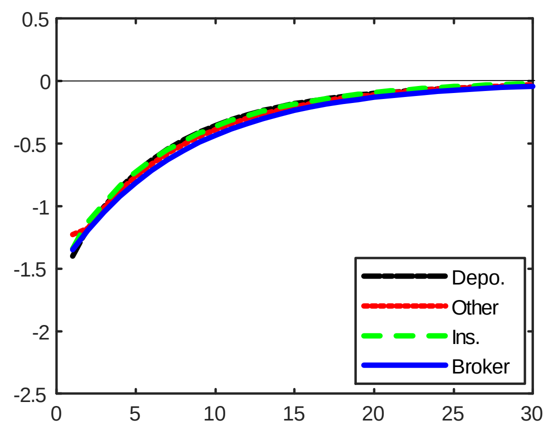
(B1) Pseudo for 5% VaR



(A2) Local projection for 1% VaR

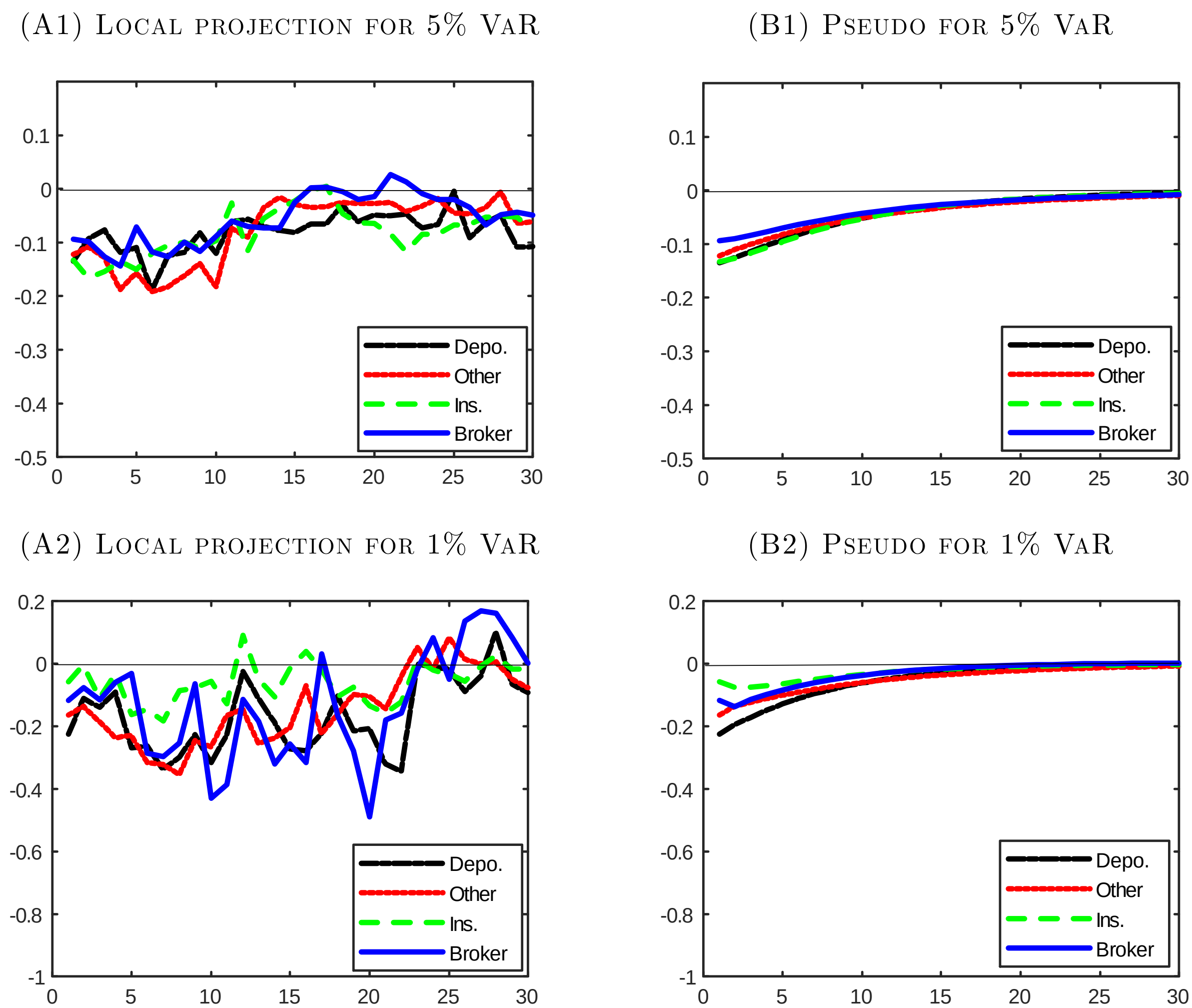


(B2) Pseudo for 1% VaR



Note: The figures present sectoral averages of the QIRFs of individual financial institutions when there is a shock to the market. Black dashdot line is for depositories. Red dotted line is for other. Green dashed line is for insurance. Blue solid line is for broker-dealers.

Figure 7. [Systemic risk] Sectoral averages of quantile impulse responses of market



Note: The figures present sectoral averages of the QIRFs of the market when there is a shock to each individual financial institution. Black dashdot line is for depositories. Red dotted line is for other. Green dashed line is for insurance. Blue solid line is for broker-dealers.

C List of Financial Institutions

This appendix contains the names of the U.S. financial institutions used in the analysis. As in Acharya et al. (2017), we consider financial institutions in the U.S. that had a market cap in excess of 5 billion USD as of end of June 2007 and categorize them into the following four groups: Depositories, Broker-Dealers, Insurance and a group called Other consisting of non-depository institutions, real estate, and so on. Considering data availability from 3 Jan. 2000 to 29 Jun. 2018, our analysis includes total 61 financial institutions. As in Acharya et al. (2017), we put Goldman Sachs in the group of Broker-Dealers. See Appendix B in Acharya et al. (2017) for more details. The list of institutions' names and tickers is given below.

Depositories: 20 companies, 2-digit SIC code=60: 1. BANK OF AMERICA CORP (BAC), 2. BB&T CORP (BBT), 3. BANK NEW YORK INC (BK), 4. CITIGROUP INC (C), 5. COMERICA INC (CMA), 6. HUNTINGTON BANCSHARES INC (HBAN), 7. JPMORGAN CHASE & CO (JPM), 8. KEYCORP NEW (KEY), 9. M&T BANK CORP (MTB), 10. NORTHERN TRUST CORP (NTRS), 11. NEW YORK COMMUNITY BANCORP INC (NYB), 12. PEOPLES UNITED FINANCIAL INC (PBCT), 13. PNC FINANCIAL SERVICIES GRP INC (PNC), 14. REGIONS FINANCIAL CORP NEW (RF), 15. SYNOVUS FINANCIAL CORP (SNV), 16. SUNTRUST BANKS INC (STI), 17. STATE STREET CORP (STT), 18. US BANCORP DEL (USB), 19. WELLS FARGO&CO NEW (WFC), 20. ZIONS BANCORP (ZION)

Other: Non-depository institutions etc: 13 companies, 2-digit SIC code=61, 62 (except 6211), 65, 67: 1. TD AMERITRADE HOLDING, CORP (AMTD), 2. AMERICAN EXPRESS CO (AXP), 3. FRANKLIN RESOURCES INC (BEN), 4. BLACKROCK INC (BLK), 5. CAPITAL ONE FINANCIAL CORP (COF), 6. EATON VANCE CORP (EV), 7. FIFTH THRID BANCORP (FITB), 8. FEDERAL HOME LOAN MORTGAGE CORP (FRE), 9. LEGG MASON INC (LM), 10. LEUCADIA NATIONAL CORP (LUK), 11. SEI INVESTMENTS COMPANY (SEIC), 12. SLM CORP (SLM), 13. UNION PACIFIC CORP (UNP)

Insurance: 23 companies, SIC code=63 and 64: 1. AETNA INC NEW (AET), 2. AFLAC INC (AFL), 3. AMERICAN INTERNATIONAL GROUP INC (AIG), 4. ALLSTATE CORP (ALL), 5. AON CORP (AOC), 6. BERKLEY WR CORP (BER), 7. BERKSHIRE HATHAWAY INC DEL(A) (BRK), 8. BERKSHIRE HATHAWAY INC DEL(B) (BRK), 9. CHUBB CORP (CB), 10. CIGNA CORP (CI), 11. CINCINNATI FINANCIAL CORP (CINF), 12. CNA FINANCIAL CORP (CNA), 13. HARTFORD FINANCIAL SVCS GROUP IN (HIG), 14. HUMANA INC (HUM), 15. LOEWS CORP1 (L), 16. LICOLN NATIONAL CORP IN (LNC), 17. MBIA INC (MBI), 18. MARSH & MCLENNAN COS INC (MMC), 19. PROGRESSIVE CORP OH (PGR), 20. TRAVELERS COMPANIES INC (STA), 21. TORCHMARK CORP (TMK), 22. UNITEDHEALTH GROUP INC (UNH), 23. UNUM GROUP (UNM)

Broker-Dealers: 5 companies, 4-digit SIC code=6211: 1. E TRADE FINANCIAL CORP

(ETFC), 2. GOLDMAN SACHS GROUP INC (GS), 3. MORGAN STANLEY DEAN WITTER & CO (MS), 4. SCHWAB CHARLES CORP NEW (SCHW), 5. T ROWE PRICE GROUP INC (TROW)

D Reference

Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies*, 30(1), 2–47.

Adrian, T, & Brunnermeier, M. K. (2016). CoVaR. *The American Economic Review*, 106(7), 1705–1741.

Baruník, J. & Kley, T. (2019). Quantile coherency: A general measure for dependence between cyclical economic variables. *The Econometrics Journal*, 22(2), 131–152.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307–327.

Bouri, E., Shahzad, S. J. H., Raza, N., & Roubaud, D. (2018) Oil volatiliy and sovereign risk of BRICS, *Energy Economics,* 70, 258–269.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Carlier, G., Chernozhukov, V., & Galichon, A. (2016). Vector quantile regression: An optimal transport approach. *The Annals of Statistics*, 44(3), 1165-1192.

Carrasco, M., & Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18(1), 17–39.

Chavleishvili, S., & Manganelli, S. (2019). Forecasting and stress testing with quantile vector autoregression. Working paper.

Chuliá, H., Gupta, R., Uribe, J. M., & Wohar, M. E. (2017) Impact of US uncertainties on emerging and mature markets: Evidence from a quantile-vector autoregressive approach, *Journal of International Financial Markets, Institutions & Money,* 48, 178–191.

Davis, R. A. & Mikosch, T. (2009) The extremogram: A correlogram for extreme events, *Bernoulli*, 15(4), 977–1009.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4), 367–381.

Gonçalves, S., & de Jong, R. (2003). Consistency of the stationary bootstrap under weak moment conditions. *Economics Letters*, 81(2), 273–278.

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

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(1), 251–270.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Volume: 1. University of California Press.

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

Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American economic review*, 95(1), 161–182.

Jung, W., & Lee, J.H., (2019). Quantile Impulse Response Analysis with Applications in Macroeconomics and Finance. UIUC Working paper.

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: Proceedings of the Fifth Prague Symposium, held from September 4, 1993*. Physica-Verlag, Heidelburg.

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

Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, 101(475), 980–990.

Kong, L., & Mizera, I. (2012). Quantile tomography: using quantiles with multivariate data. *Statistica Sinica*, 22, 1589-1610.

Lee, J. H., Linton, O., & Whang, Y. J. (2019). Quantilograms under Strong Dependence, forthcoming in *Econometric Theory*.

Li, G., Li, Y., & Tsai, C. L. (2015) Quantile correlations and quantile autoregressive modeling, *Journal of the American Statistical Association*, 110(509), 246–261.

Linton, O., & Whang, Y. J. (2007). The quantilogram: With an application to evaluating directional predictability. *Journal of Econometrics*, 141(1), 250–282.

Loria, Francesca and Matthes, Christian and Zhang, Donghai (2019). Assessing Macroeconomic Tail Risk. FRB Richmond Working Paper No. 19-10.

Martin, V., Hurn, S., & Harris, D. (2013). *Econometric Modelling with Time Series: Specification, Estimation and Testing*. Cambridge University Press.

Montes-Rojas, G. (2019). Multivariate Quantile Impulse Response Functions. *Journal of Time Series Analysis*, 40(5), 739–752.

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

Patton, A., Politis, D. N., & White, H. (2009). Correction to "Automatic block-length selection for the dependent bootstrap" by D. Politis and H. White. *Econometric Reviews*, 28(4), 372–375.

Peng, W. & Zeng, Y. (2019) Overnight exchange rate risk based on multi-quantile and joint-shock CAViaR models, *Economic Modelling*, 80, 392–399.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, 89(428), 1303–1313.

Politis, D. N., & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1), 53–70.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25(3), 303–325.

Ruppert, D., & Carroll, R. J. (1980). Trimmed least squares estimation in the linear model.

Journal of the American Statistical Association, 75(372), 828–838.

Schwert, G. W. (1989). Why does stock market volatility change over time?. *The journal of finance*, 44(5), 1115–1153.

Shen, Y. (2018) International risk transmission of stock market movements, *Economic Modelling*, 69, 220–236.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48, 1–48.

Taylor, S. J. (1986). *Modelling Financial Time Series*. John Wiley and Sons, London.

Wen, D., Wang, G. J., Ma, C. & Wang, Y. (2019) Risk spillovers between oil and stock markets: A VAR for VaR analysis, *Energy Economics*, 80, 524–535.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians*. Academic Press.

White, H., Kim, T. H., & Manganelli, S. (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics*, 187(1), 169–188.