Essays in quantile regression models and their applications to financial time series

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Download date: 06 May. 2021

Doctoral thesis

ESSAYS IN QUANTILE REGRESSION MODELS AND THEIR APPLICATIONS TO FINANCIAL TIME SERIES

Li Sun

2021

ESSAYS IN QUANTILE REGRESSION MODELS AND THEIR APPLICATIONS TO FINANCIAL TIME SERIES

Dissertation

To obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. R.M. Letschert,
in accordance with the decision of the Board of Deans,
to be defended in public
on Thursday 25th of February 2021, at 16.00 hours

by

Li Sun

Promotor
   Prof. dr. Alain W. Hecq

Copromotor
   Dr. Stefan T.M. Straetmans

Assessment Committee
   Prof. dr. Franz Palm (chair)
   Prof. dr. Bertrand Candelon (Université Catholique de Louvain)
   Prof. Christophe Hurlin (University of Orleans)
   Dr. Ines Wilms (Maastricht University)

Acknowledgments

This dissertation is the result of my PhD journey that started in September 2017, at Maastricht University. Looking backwards, I appreciate this journey and the people guiding/accompanying me along the way. I would like to take this opportunity to express my gratitude to some of them in particular.

First and foremost, I would like to thank my PhD supervisor Alain Hecq who actually anchored the starting point of this dissertation for me. Thank you for offering me this PhD position. I still remember the first time that we met in your office for this position interview. My English was poor, and I even answered fast because I was nervous and more nervous in speaking English. When you offered me the position, I believe firmly in your decision-making from that time on. During my PhD, you are generous in time to discuss with me, patient and open-minded in listening to my ideas some of which I even could only explain ambiguously. And you sophisticatedly pointed me time to time when I struggled. I appreciate all your comments, insights and quick responses.

Then, I would like to thank my co-supervisor Stefan Straetmans. Thank you for stepping in for my third PhD project. I benefited a lot from the papers that you sent me, which indeed inspired me to form my third paper. And I appreciate your comments, approval for this dissertation and helps with my defence.

Further, I would also like to thank all the members of my assessment committee, Franz Palm, Bertrand Candelon, Christophe Hurlin, Ines Wilms, for taking the time to read my dissertation before the Christmas.

Also, I would like to thank for all my colleagues at the Quantitative Economics department. My past three years not only shape me into a researcher but also open a lovely life to me. Certainly, the lovely life is filled with lovely people. I would like to thank some of them in particular and preserve some beautiful memory pieces here. Caterina, my PhD-officemate. I got to know Moka pot because of you, and have been using it every morning since then. You are caring and genuine. I feel like we appreciate the same virtue of an office. And I enjoy working with you in our office. Benoit, thank you for organizing many fun events. You always shared you studio with us to cook/eat/play. Because of you, I know how to play many board games and card games, and got close with the group. Luca, we went to conferences and took some trainings together in our first-year PhD. You are helpful always. I do like your English, and mimicked yours at some time back. Aditya, we both were tenants in the Student Hotel Maastricht in our first-year PhD. You are light-hearted also inspiring. I like listening to your talks at drinks or on our way back home. Niels, thanks for organizing many hiking events. You interchange your role from a local guide to a translator each time

of our hikes. Yicong, thank you for many walks with me after lunch. You are sharing and motivated. You shared many of your experiences with me. And I got to reflect on myself and to be prepared.

At last but the most, I want to thank my fiance, Chunming Sui. You applaud me when I manage to reach a point in life, and pat me on my back each time when I feel low. We have been through all the happiness and setbacks together for 9 years around. Because of you, I got to control my temper and soften my heart. Because of you, I become a better person and want to make you happy all the time. I feel truly blessed to have you in my life.

<div align="right">

Li Sun
Maastricht
January 2021

</div>

Contents

1

Introduction

"Finance is defined as the management of money and includes activities such as investing, borrowing, lending, budgeting, saving, and forecasting." (CFI, 2020) Financial modelling is all about capturing financial data patterns and utilizing the knowledge for financial benefits. By now, there is a consensus on a set of empirical stylized facts about financial time series (see, for example, Pagan, 1996; Shephard, 1996; Cont, 2001; Gourieroux and Jasiak, 2001). For examples, it is well known that financial return time series exhibit unconditional and conditional heavy tails, volatility clustering and time-varying cross-correlation which will be discussed in Chapter 2, 3 and 4 respectively and some practical economic/financial problems are addressed therein correspondingly.

Quantile regression is the major econometric tool that we work on across this dissertation as it has been gradually emerging as a unified statistical methodology for estimating models of conditional quantile functions. Classical least-squares regressions are widely used for estimating models of conditional mean functions which is the exclusive focus on response variables by least-squares regressions. By complementing this exclusive focus of classical least-squares regressions, quantile regressions offer a systematic way for examining how covariates influence statistical characteristics, such as the location, the scale or the shape, of the entire response distribution so as to reveal information which, for instance, can distinguish model performances, see Chapter 2. A brief introduction on quantile regression is given below.

1.1 Quantile Regression

We know by name that quantile regression is to estimate a quantile of a random variable (say X) or quantiles of random variable(s). The definition of the τ-th $(0 < \tau < 1)$ quantile of random variable X is defined as follows:

Definition 1 (Quantile).
Suppose random variable X on a probability space $(\Omega, \mathscr{F}, P)$ is characterized by a (right-continuous) distribution function denoted by $F(\cdot)$, i.e.,

$$F(x) = P\{X \leq x\}.$$

For any $0 < \tau < 1$, the τ-th quantile of X is given by

$$F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}. \tag{1.1}$$

Suppose we have $n > 0$ observations on X and would like to estimate the τ-th quantile of X as a constant value. Denote the sample of these observations as $\{x_i\}_{i=1}^n$. We can apply a simple quantile regression to estimate $F^{-1}(\tau)$ by minimizing

$$\sum_{i=1}^n \rho_\tau(x_i - u), \tag{1.2}$$

where $\rho_\tau(u) := u(\tau - \mathbb{1}_{\{u<0\}})$ is called check function in quantile regressions (Koenker, 2005), and the indicator function $\mathbb{1}_{\{\mathscr{A}\}}$ on a set $\mathscr{A}$ or a logical statement $\mathscr{A}$ equals 1 if $\mathscr{A}$ is non-empty or true, otherwise $\mathbb{1}_{\{\mathscr{A}\}} = 0$. So the quantile regression estimator of the τ-th quantile of X denoted by $\widehat{u}$ is given by

$$\widehat{u} := \operatorname*{argmin}_{u \in \mathbb{R}} \sum_{i=1}^n \rho_\tau(x_i - u). \tag{1.3}$$

Quantile regression is a numerically efficient tool when particular quantiles of random variables are of interest. For instance, value-at-risk (VaR) is a standard risk measure for market risk management. The VaR of a company's return variable X at $\alpha \in (0,1)$ level is defined as

$$\mathrm{VaR}_\alpha = -F^{-1}(\alpha), \tag{1.4}$$

where $F^{-1}(\alpha)$ is the α-th quantile of the return variable X, and we usually take $\alpha = 5\%$ or $\alpha = 1\%$. VaR_α can be seen as the maximum possible loss on a fixed asset over a fixed time horizon within confidence level $(1 - \alpha)$. VaR is widely employed in the financial industry for both internal control and regulatory reporting. Among many popular approaches for VaR estimation, quantile

regressions stand out for the advantages in semi-parametric specification and numerical efficiency. The quantile regression family working for this measure has been extended from static quantile regression models (QR, see Koenker and Bassett Jr, 1978) to quantile autoregression models (QAR, see Koenker and Xiao, 2006), to conditional autoregressive value-at-risk models (CAViaR, see Engle and Manganelli, 2004), to multivariate multi-quantile CAViaR models (MVMQ-CAViaR, see White et al., 2015). In Chapter 3, we investigate performances of CAViaR models from estimation to inference testing, and aim to address any practical problems in CAViaR applications. Chapter 4 are based on MVMQ-CAViaR models to measure systemic risks of big financial institutions.

1.2 Outline of This Dissertation

The remainder of this dissertation is structured as follows:

In Chapter 2, we introduce mixed causal and noncausal models and our research background. We propose quantile autoregressions in the time reverse version called quantile noncausal autoregressions (QNCAR) along with a generalized asymptotic theorem in a stable law for both QCAR and QNCAR. A common issue in the model selection through SRAR comparison is brought out. The use of the aggregate SRAR over all quantiles as a new model selection criterion is then proposed with the shape of SRAR curves being analysed. Furthermore, we illustrate our analysis using hyperinflation episodes of four Latin American countries.

In Chapter 3, conditions for CAViaR data generating processes (DGPs) to be non-explosive are derived. We investigate the size performance of Wald tests for CAViaR models and find large size distortions by the usual estimation strategy. So we develop up a method called *adaptive random bandwidth*. An empirical study is performed on stock returns.

In Chapter 4, we introduce the generic MVMQ-CAViaR model specification first and propose to generalize it to the vector autoregressive model of VaR and CoVaR. We also prove the estimation consistency and asymptotic normality in this generalized model regression. We call this generalized model as *systemic MVMQ-CAViaR model*. It follows that some inference tests are given in order to infer the significance of contemporaneous terms in the CoVaR specification. We illustrate on applying CoVaR returned by our generalized model to measure the systemic risk of financial institutions. We construct quantile impulse response functions in correspondence to our model and apply the local projection with use of expansion of estimated terms for QIRF estimation. Some results of

Monte Carlo simulations regarding systemic MVMQ-CAViaR models are presented. An empirical application is performed on big banks with the market index *S&P* 500.

In Chapter 5, we conclude this dissertation by providing a short summary of the research results presented in this dissertation and an outlook of future research based on my presented work.

2

Selecting between Causal and Noncausal Models with Quantile Autoregressions

---

This chapter is based on the article '*Selecting between causal and noncausal models with quantile autoregressions*' by Hecq and Sun (2020), Studies in Non-linear Dynamics & Econometrics, 1 (ahead-of-print).

Abstract

In Chapter 2, we propose a model selection criterion to detect purely causal from purely noncausal models in the framework of quantile autoregressions (QAR). We also derive asymptotic properties for the i.i.d. case with regularly varying distributed innovations in QAR. This new modelling perspective is appealing for investigating the presence of bubbles in economic and financial time series, and is an alternative to approximate maximum likelihood methods. We illustrate our analysis using hyperinflation episodes in Latin American countries.

2.1 Motivation

Mixed causal and noncausal time series models have been recently used in order (i) to obtain a stationary solution to explosive autoregressive processes, (ii) to improve forecast accuracy, (iii) to model expectation mechanisms implied by economic theory, (iv) to interpret non-fundamental shocks resulting from the asymmetric information between economic agents and econometricians, (v) to generate non-linear features from simple linear models with non-Gaussian disturbances, (vi) to test for time reversibility. When the distribution of innovations is known, a non-Gaussian likelihood approach can be used to discriminate between lag and lead polynomials of the dependent variable. For instance, the R package `MARX` developed by Hecq, Lieb and Telg (2017) estimates univariate mixed models under the assumption of a Student's $t-$distribution with v degrees of freedom (see also Lanne and Saikkonen, 2011, 2013) as well as the Cauchy distribution as a special case of the Student's t when $v = 1$. Gouriéroux and Zakoian (2016) privilege the latter distribution to derive analytical results. Gouriéroux and Zakoian (2015), Fries and Zakoian (2017) provide an additional flexibility to involve some skewness by using the family of alpha-stable distributions. However, all those aforementioned results require the estimation of a parametric distributional form. In this article we take another route.

The objective of this chapter is to select between causal and noncausal models without using parametric distributional assumptions. To achieve that, we adopt a quantile regression (QR) framework and apply quantile autoregressions (QCAR hereafter) (Koenker and Xiao, 2006) on candidate models. Although we obviously also require non-Gaussian innovations in time series, we do not make any parametric distributional assumption about the innovations. By using quantile regressions, we consider a statistic called the sum of rescaled absolute residuals (SRAR hereafter) to measure model performances and reveal properties of time series. Remarkably we find that SRAR cannot always favour a model uniformly along quantiles. This issue is common for time series of asymmetric distributed innovations, which causes confusion in model selection and calls for a robust statistic to meet the goal. Considering that, we propose to aggregate the SRAR information over quantiles. It is worth mentioning that when coefficients are constant in the underlying model with a symmetrically i.i.d. error term, the aggregate SRAR criterion is equivalently to select between forward and backward conditional mean models (termed by Gourieroux and Zakoian (2017)). However, the aggregate SRAR is a measure based on the whole dynamics of the underlying process, which is not dominated by the conditional mean information any more. This characteristic of the aggregate SRAR criterion indeed makes it robust in model selection even for some general situations such as with asymmetric distributed innovations. Another remark on this chapter is that our method is restricted to the model framework of purely causal or noncausal autoregres-

sions without other explanatory variables, thereby this method can be used to questions like asset pricing of exchange rate where current exchange rate is associated with future exchange rates. However, it cannot be used to questions like the Taylor (1993) rule which associates the dynamics of the nominal interest rate with dynamics of some endogenous variables (eg., inflation).

The rest of this chapter is constructed as follows. Section 2.2 introduces mixed causal and noncausal models and our research background. In Section 2.3, we propose quantile autoregressions in the time reverse version called quantile noncausal autoregressions (QNCAR) along with a generalized asymptotic theorem in a stable law for both QCAR and QNCAR. Section 2.4 brings out a common issue in the model selection through SRAR comparison. The use of the aggregate SRAR over all quantiles as a new model selection criterion is then proposed with the shape of SRAR curves being analysed. Furthermore, we illustrate our analysis using hyperinflation episodes of four Latin American countries in Section 2.6. Section 2.7 concludes this chapter.

2.2 Causal and Noncausal Time Series Models

Brockwell and Davis introduce in their texbooks (1991, 2002) a univariate noncausal specification as a way to rewrite an autoregressive process with explosive roots into a process in reverse time with roots outside the unit circle. This noncausal process possesses a stable forward looking solution whereas the explosive autoregressive process in direct time does not. This approach can be generalized to allow for both lead and lag polynomials. This is the so called mixed causal-noncausal univariate autoregressive process for y_t that we denote MAR(r, s)

$$\pi(L)\phi(L^{-1})y_t = \varepsilon_t, \tag{2.1}$$

where $\pi(L) = 1 - \pi_1 L - ... - \pi_r L^r$, $\phi(L^{-1}) = 1 - \phi_1 L^{-1} - ... - \phi_s L^{-s}$. L is the usual backshift operator that creates lags when raised to positive powers and leads when raised to negative powers, i.e., $L^j y_t = y_{t-j}$ and $L^{-j} y_t = y_{t+j}$. The roots of both polynomials are assumed to lie strictly outside the unit circle, that is $\pi(z) = 0$ and $\phi(z) = 0$ for $|z| > 1$ and therefore

$$y_t = \pi(L)^{-1}\phi(L^{-1})^{-1}\varepsilon_t = \sum_{i=-\infty}^{\infty} a_i \varepsilon_{t-i} \tag{2.2}$$

has an infinite two sided moving average representation. We also have that $E(|\varepsilon_t|^\delta) < \infty$ for $\delta > 0$[1] and the Laurent expansion parameters are such that

---

[1] The errors do not necessarily have finite second order moments. For $\delta \geq 2$ the second order moment exists, for $\delta \in [1, 2)$ the errors have infinite variance but finite first order moment, for $\delta \in (0, 1)$ the errors do not have finite order moments.

$\sum\limits_{i=-\infty}^{\infty} |a_i|^\delta < \infty$. The representation (2.2) is sometimes clearer than (2.1) to motivate the terminology "causal/noncausal". Indeed those terms refer to as the fact that y_t depends on a causal (resp. noncausal) component $\sum\limits_{i=0}^{\infty} a_i \varepsilon_{t-i}$ (resp. noncausal $\sum\limits_{i=-\infty}^{-1} a_i \varepsilon_{t-i}$). With this in mind, it is obvious that an autoregressive process with explosive roots will be defined as noncausal.

Note that in (2.1), the process y_t is a purely causal MAR$(r,0)$, also known as the conventional causal AR(r) process, when $\phi_1 = ... = \phi_s = 0$,

$$\pi(L)y_t = \varepsilon_t, \tag{2.3}$$

while the process is a purely noncausal MAR$(0,s)$

$$\phi(L^{-1})y_t = \varepsilon_t, \tag{2.4}$$

when $\pi_1 = ... = \pi_r = 0$.

A crucial point of this literature is that innovation terms ε_t must be i.i.d. non-Gaussian to ensure the identifiability of a causal from a noncausal specification (Breidt, Davis, Lii and Rosenblatt, 1991). The departure from Gaussianity is not as such an ineptitude as a large part of macroeconomic and financial time series display nonlinear and non-normal features.

We have already talked in Section 2.1 about the reasons for looking at models with a lead component. Our main motivation in this chapter lies in the fact that MAR(r,s) models with non-Gaussian disturbances are able to replicate nonlinear features (e.g., bubbles, asymmetric cycles) that previously were usually obtained by highly nonlinear models. As an example, we simulate in Figure 2.1 an MAR(1,1) of $(1 - 0.8L)(1 - 0.6L^{-1})y_t = \varepsilon_t$ with $\varepsilon_t \overset{d}{\sim} t(3)$ for 200 observations.[2] One can observe asymmetric cycles and multiple bubbles.[3]

Once a distribution or a group of distributions is chosen, the parameters in $\pi(L)\phi(L^{-1})$ can be estimated. Assuming for instance a non-standardized $t-$distribution for the innovation process, the parameters of mixed causal-noncausal autoregressive models of the form (2.1) can be consistently estimated by the approximate maximum likelihood (AML) method. Let $(\varepsilon_1, ..., \varepsilon_T)$ be a sequence of i.i.d. zero mean $t-$distributed random variables,

---

[2]We use the package MARX develop in R by Hecq, Lieb and Telg (2017).

[3]MAR(r,s) models can be generated in two steps (see Gourieroux and Jasiak, 2016; Hecq, Lieb and Telg, 2016). We propose in the Appendix an alternative method based on matrix representation that is very compact in code writing and intuitive in understanding.
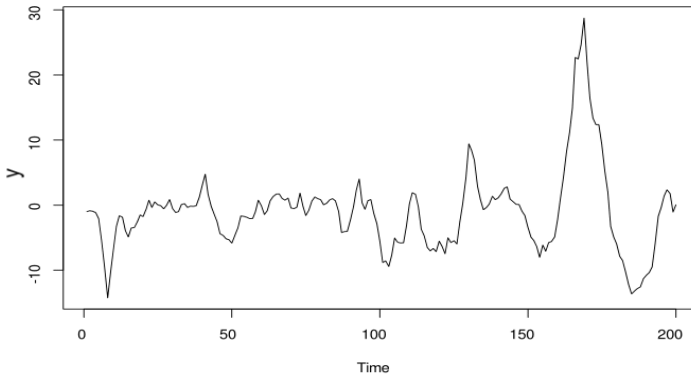
Figure 2.1: Simulation of a MAR(1,1) model, T=200

then its joint probability density function can be characterized as

$$f_\varepsilon(\varepsilon_1,...,\varepsilon_T|\sigma,\nu) = \prod_{t=1}^{T} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma} \left(1 + \frac{1}{\nu}\left(\frac{\varepsilon_t}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}},$$

where $\Gamma(\cdot)$ denotes the gamma function. The corresponding (approximate) log-likelihood function conditional on the observed data $\boldsymbol{y} = (y_1,...,y_T)$ can be formulated as

$$l_y(\boldsymbol{\phi}, \boldsymbol{\varphi}, \boldsymbol{\lambda}, \alpha|\boldsymbol{y}) = (T-p)\left[\ln(\Gamma((\nu+1)/2)) - \ln(\sqrt{\nu\pi}) - \ln(\Gamma(\nu/2)) - \ln(\sigma)\right]$$
$$- (\nu+1)/2 \sum_{t=r+1}^{T-s} \ln(1 + ((\pi(L)\phi(L^{-1})y_t\alpha)/\sigma)^2/\nu),$$

(2.5)

where $p = r + s$ and $\varepsilon_t = \pi(L)\phi(L^{-1})y_t - \alpha$ is replaced by a nonlinear function of the parameters when expanding the product of polynomials. The distributional parameters are collected in $\boldsymbol{\lambda} = [\sigma, \nu]'$, with σ representing the scale parameter and ν the degrees of freedom. α denotes an intercept that can be introduced in model (2.1). Thus, the AML estimator corresponds to the solution $\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}\in\Theta} l_y(\boldsymbol{\theta}|\boldsymbol{y})$, with $\boldsymbol{\theta} = [\boldsymbol{\phi}', \boldsymbol{\varphi}', \boldsymbol{\lambda}']'$ and Θ is a permissible parameter space containing the true value of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}_0$, as an interior point. Since an analytical solution of the score function is not directly available, gradient based numerical procedures can be used to find $\hat{\boldsymbol{\theta}}_{ML}$. If $\nu > 2$, and hence $E(|\varepsilon_t|^2) < \infty$, the AML estimator is $\sqrt{T}$-consistent and asymptotically normal. Lanne and Saikonen (2011) also show that a consistent estimator of the limiting covariance matrix is obtained from the standardized Hessian of the log-likelihood. For the estimation of the parameters and the standard innovations as well as for the selection of mixed causal-noncausal models we can also follow the procedure

proposed by Hecq, Lieb and Telg (2016).

However, the AML estimation is based on a parametric form of the innovation term in (2.1), which makes this method not flexible enough to adapt uncommon distributions as complex in reality. To be more practical and get rid of strong distribution assumptions on innovations, in next section we adopt quantile regression methods with some properties discussed there. This chapter only focuses on purely causal and noncausal models.

2.3 QCAR & QNCAR

Koenker and Xiao (2006) have introduced a quantile autoregressive model of order p denoted as QAR(p) which is formulated as the following form:

$$y_t = \theta_0(u_t) + \theta_1(u_t)y_{t-1} + ... + \theta_p(u_t)y_{t-p}, \qquad t = p+1, ..., T, \qquad (2.6)$$

where u_t is a sequence of i.i.d. standard uniform random variables. In order to emphasize the causal characteristic of this kind of autoregressive models, we refer to (2.6) as QCAR(p) hereafter. Provided that the right-hand side of (2.6) is monotone increasing in u_t, the $\tau-$th conditional quantile function of y_t can be written as

$$Q_{y_t}(\tau|y_{t-1},...y_{t-p}) = \theta_0(\tau) + \theta_1(\tau)y_{t-1} + ... + \theta_p(\tau)y_{t-p}. \qquad (2.7)$$

If an observed time series $\{y_t\}_{t=1}^T$ can be written into a QCAR(p) process, its parameters as in (2.7) can be obtained from the following minimization problem.

$$\hat{\boldsymbol{\theta}}(\tau) = \underset{\boldsymbol{\theta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{t=1}^T \rho_\tau(y_t - \boldsymbol{x}_t'\boldsymbol{\theta}), \qquad (2.8)$$

where $\rho_\tau(u) := u(\tau - \mathbb{1}_{\{u<0\}})$ is called check function in quantile regressions (Koenker, 2005), and $\boldsymbol{x}_t' := [1, y_{t-1}, ..., y_{t-p}]$, $\boldsymbol{\theta}' := [\theta_0, \theta_1, ..., \theta_{t-p}]$ with the indicator function $\mathbb{1}_{\{\cdot\}}$. We define the sum of rescaled absolute residuals (SRAR) for each pair of $(\tau, \boldsymbol{\theta})$ as

$$\text{SRAR}(\tau, \boldsymbol{\theta}) := \sum_{t=1}^T \rho_\tau(y_t - \boldsymbol{x}_t'\boldsymbol{\theta}). \qquad (2.9)$$

Substitute (2.9) into (2.8) and write the minimization problem (2.8) as

$$\hat{\boldsymbol{\theta}}(\tau) = \underset{\boldsymbol{\theta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \text{ SRAR}(\tau, \boldsymbol{\theta}). \qquad (2.10)$$

The estimation consistency and asymptotic normality in the minimization problem (2.8) have been provided by Koenker and Xiao (2006). A modified simplex

algorithm proposed by Barrodale and Roberts (1973) can be used to solve the minimization, and in practice parameters for each $\tau-$th quantile can be obtained, for instance, through the `rq()` function from the **quantreg** package in R or in EViews.

2.3.1 QNCAR

A QNCAR(p) specification is introduced here as the noncausal counterpart of the QCAR(p) model by reversing time, explicitly as follows:

$$Q_{y_t}(\tau|y_{t+1},...y_{t+p}) = \phi_0(\tau) + \phi_1(\tau)y_{t+1} + ... + \phi_p(\tau)y_{t+p}. \tag{2.11}$$

Analogously to the QCAR(p), the estimation of the QNCAR(p) goes through solving

$$\hat{\boldsymbol{\theta}}(\tau) = \underset{\boldsymbol{\theta}\in\mathbb{R}^{p+1}}{\operatorname{argmin}}\ \mathrm{SRAR}(\tau,\boldsymbol{\theta})$$

with

$$\boldsymbol{x}_t' = [1, y_{t+1}, \ldots, y_{t+p}],$$

where for the simplicity of notations, we use $\hat{\boldsymbol{\theta}}(\tau)$ to denote the estimate in quantile noncausal autoregression. Drawing on the asymptotics derived by Koenker and Xiao (2006), we present the following theorem for QNCAR(p) based on three assumptions which are made to ensure covariance stationarity of the time series (by (A1) and (A2)) and the existence of quantile estimates (by (A3)).

Remark. *There is an issue in the estimation consistency of QCAR(p) as reported by Fan and Fan (2010). This is due to the violation on the monotonicity requirement of the right side of (2.6) in u_t but not exclusively the monotonicity of $\theta_i(u_t)$ in u_t. So to recover an AR(p) process of coefficients $\theta_i(u_t)$ $(i = 0,\ldots,p)$ monotonic in u_t, quantile autoregression is not a 100% match tool unless the monotonicity requirement is met beforehand. This issue is also illustrated in Section 2.4.1.*

Theorem 2. *A QNCAR(p) model can be written in the following vectorized companion form:*

$$\tilde{\boldsymbol{x}}_t = \mathbf{A}_t\tilde{\boldsymbol{x}}_{t+1} + \boldsymbol{v}_t, \tag{2.12}$$

where $\tilde{\boldsymbol{x}}_t' := [y_t, y_{t+1},\ldots,y_{t+p-1}]$, $\boldsymbol{x}_t' := [1,\tilde{\boldsymbol{x}}_t']$, $\mathbf{A}_t := \begin{bmatrix} \phi_{1,t} & \phi_{2,t} & \cdots & \phi_{p,t} \\ & \mathbf{I}_{p-1} & & \mathbf{0}_{(p-1)\times 1} \end{bmatrix}$ and $\boldsymbol{v}_t := \begin{bmatrix} \varepsilon_t \\ \mathbf{0}_{(p-1)\times 1} \end{bmatrix}$, satisfying the following assumptions:

(A1) : $\{\varepsilon_t\}_{t=1}^n$ *are i.i.d. innovations with mean* 0 *and variance* $\sigma^2 < \infty$. *The distribution function of* ε_t, *denoted as* $F(\cdot)$, *has a continuous density* $f(\cdot)$ *with* $f(\varepsilon) > 0$ *on* $\mathscr{U} := \{\varepsilon : 0 < F(\varepsilon) < 1\}$.

(A2) : *The eigenvalues of* $E\left[\mathbf{A}_t \otimes \mathbf{A}_t\right]$ *have moduli less than one.*

(A3) : $F_{y_t|\tilde{\mathbf{x}}_{t+1}}(\cdot) := \mathbf{P}\left[y_t < \cdot \mid y_{t+1}, y_{t+2}, \ldots, y_{t+p}\right]$ *has derivative* $f_{y_t|\tilde{\mathbf{x}}_{t+1}}(\cdot)$ *which is uniformly integrable on* $\mathscr{U}$ *and non-zero with probability one.*

Then,

$$\boldsymbol{\Sigma}^{-\frac{1}{2}}\sqrt{T}\left(\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}(\tau)\right) \overset{d}{\sim} \mathbb{B}_{p+1}(\tau), \qquad (2.13)$$

where $\boldsymbol{\Sigma} := \boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1}$, $\boldsymbol{\Sigma}_0 := E\left[\boldsymbol{x}_t\boldsymbol{x}_t'\right]$, $\boldsymbol{\Sigma}_1 := \lim T^{-1}\sum_{t=1}^T f_{y_t|\tilde{\mathbf{x}}_{t+1}}\left(F_{y_t|\tilde{\mathbf{x}}_{t+1}}^{-1}(\tau)\right)\boldsymbol{x}_t\boldsymbol{x}_t'$, $\boldsymbol{\phi}(\tau)' := \left[F^{-1}(\tau), \phi_1(\tau), \ldots, \phi_p(\tau)\right]$, $\mathbb{B}_{p+1}(\tau) := \mathcal{N}\left(\mathbf{0}, \tau(1-\tau)\mathbf{I}_{p+1}\right)$ *with sample size* T *and* $\mathbf{I}_{p+1}$ *being the* $(p+1) \times (p+1)$ *identity matrix.*

The above result can be further simplified into Corollary 3 by adding the following assumption:

(A4): *The coefficient matrix* $\mathbf{A}_t$ *in* (2.12) *is constant over time. (We denote* $\mathbf{A} := \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ \mathbf{I}_{p-1} & & \mathbf{0}_{(p-1)\times 1} \end{bmatrix}$ *for* $\mathbf{A}_t$ *under this assumption.)*

Corollary 3. *Under assumptions (A1), (A2), (A3) and (A4),*

$$\sqrt{T}f\left(F^{-1}(\tau)\right)\boldsymbol{\Sigma}_0^{\frac{1}{2}}\left(\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}_\tau\right) \overset{d}{\sim} \mathbb{B}_{p+1}(\tau), \qquad (2.14)$$

where $\boldsymbol{\phi}_\tau := \left[F^{-1}(\tau), \phi_1, \ldots, \phi_p\right]$.

As can be seen, QCAR(p) and QNCAR(p) generalize the classical purely causal and purely noncausal models respectively by allowing random coefficients on lag or lead regressors over time. Corollary 3 provides additional results when the same coefficients except the intercept are used to generate each quantile. However, the moment requirement in (A1) is very strict for heavy tailed time series. In order to study noncausality by QAR in heavy tailed distributions, we have to show its applicability when weakening the assumption (A1). This goal is achieved by Theorem 4 which presents the asymptotic behaviour of the QAR estimator for a classical purely noncausal model. Similarly, the asymptotics in a classical purely causal model follows right after reversing time.

Theorem 4 (Asymptotics in regularly varying distributed innovations)**.**
Under Assumption (A4), a purely noncausal AR(p) of the following form

$$\phi(L^{-1})y_t = \varepsilon_t,$$

where $\phi(L^{-1}) = 1 - \phi_1 L^{-1} - \ldots - \phi_p L^{-p}$, also satisfies the following assumptions:

(A5) : $\{\varepsilon_t\}_{t=1}^n$ are i.i.d. innovation variables with regularly vary tails defined as

$$P(|\varepsilon_t| > x) = x^{-\alpha} L(x), \qquad (2.15)$$

where $L(x)$ is slowly varying at ∞ and $0 < \alpha < 2$. There is a sequence $\{a_T\}$ satisfying

$$T \cdot P\{|\varepsilon_t| > a_T x\} \to x^{-\alpha} \qquad \text{for all } x > 0. \qquad (2.16)$$

with $b_T = \mathbb{E}[\varepsilon_t I[|\varepsilon_t| \le a_T]] = 0.$[4] The distribution function of ε_t, denoted as $F(\cdot)$, has continuous density $f(\cdot)$ with $f(\varepsilon) > 0$ on $\{\varepsilon : 0 < F(\varepsilon) < 1\}$ in probability one;

(A6) : The roots of the polynomial $\phi(z)$ are greater than one, such that y_t can be written into

$$y_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t+j}, \qquad (2.17)$$

where $\sum_{j=0}^{\infty} j |c_j|^{\delta} < \infty$ for some $\delta < \alpha, \delta \le 1$.

Then

$$\frac{f(F^{-1}(\tau)) \cdot a_T \sqrt{T}}{\sqrt{\tau(1-\tau)}} \left(\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}_{\tau} \right) \overset{d}{\sim}$$

$$\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \left(\int_0^1 S_{\alpha}^2(s)\, ds\, \boldsymbol{\Omega} \right)^{-1} \end{bmatrix} \begin{bmatrix} W(1), \sum_{j=0}^{\infty} c_j \int_0^1 S_{\alpha}(s)\, dW(s), \ldots, \sum_{j=0}^{\infty} c_j \int_0^1 S_{\alpha}(s)\, dW(s) \end{bmatrix}_{(p+1) \times 1}.$$
$$(2.18)$$

where $\boldsymbol{\phi}_{\tau} := \left[\frac{F^{-1}(\tau)}{a_T}, \phi_1, \ldots, \phi_p \right]$, $\boldsymbol{\Omega} := [\omega_{ik}]$ being a $p \times p$ matrix with entry $\omega_{ik} := \sum_{j=0}^{\infty} c_j c_{j+|k-i|}$ at the i-th row and the k-th column, $\{S_{\alpha}(s)\}$ being a process of stable distributions with index α which are independent of Brownian motions $\{W(s)\}$. In this theorem the intercept regressor in QNCAR(p) is changed to a_T such that $\mathbf{x}_t' := [a_T, y_t, y_{t+1}, \ldots, y_{t+p-1}]$.

Proof. See Appendix 2.A. □

Heuristically, next we restrict our focus on the classical models and explore consequences of causality misspecification in quantile regressions.

---

[4] Without loss of generality, we assume b_T to be zero in the derivation for the simplicity.

2.3.2 Causal and noncausal models with Gaussian i.i.d. disturbances

Suppose a causal AR(1) process $\{y_t\}_{t=1}^T$, $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$, with for instance $[\alpha, \beta] = [1, 0.5]$, i.i.d. standard normal $\{\varepsilon_t\}$ and $T = 200$. Figure 2.2 displays a simulated sample following this data generating process (DGP).

$$y_t = 1 + 0.5y_{t-1} + \varepsilon_t$$
$$\varepsilon_t \sim N(0,1)$$
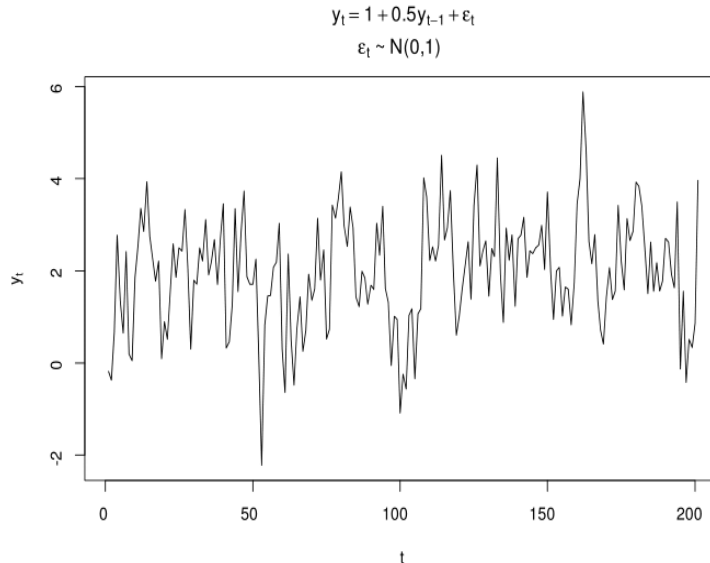
Figure 2.2: Simulation of a one-regime process with $N(0,1)$ innovations, $T = 200$

The information displayed in Figure 2.3 is the SRAR(τ) of each candidate model along quantiles, indicating their goodness of fit. The two SRAR curves almost overlap at every quantile, which implies no discrimination between QCAR and QNCAR in Gaussian innovations, in line with results in the OLS case. The Gaussian distribution is indeed time reversible, weak and strict stationary. Its first two moments characterize the whole distribution and consequently every quantile. Note that we obtain similar results for a stationary noncausal AR(p) process with i.i.d. Gaussian $\{\varepsilon_t\}$. The results are not reported to save space.

2.3.3 Causal and noncausal models with Student's t distributed innovations

Things become different if we depart from Gaussianity. Suppose now a causal AR(1) process $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$ with again $[\alpha, \beta] = [1, 0.5]$ but where $\{\varepsilon_t\}$ are i.i.d. Student's $t-$distributed with 2 degrees of freedom (hereafter using shorthand notation: $t(2)$). Figure 2.4 depicts a simulation in this AR(1) with $T = 200$.
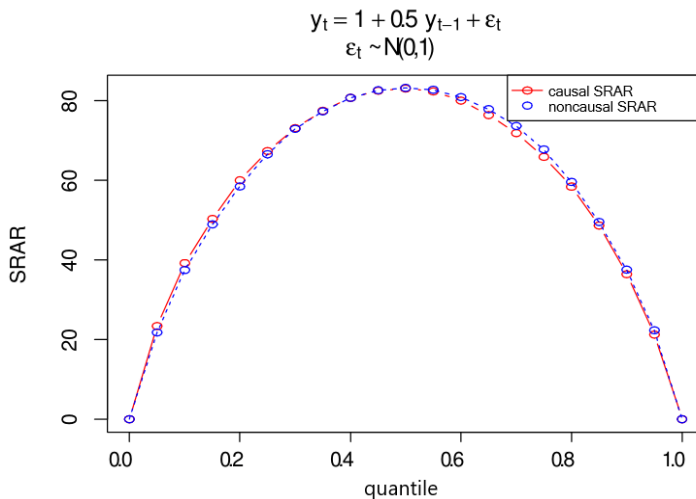
Figure 2.3: SRAR plot under an AR(1) with $N(0,1)$ innovations, $T = 200$

Applying QCAR and QNCAR respectively on this series results in the SRAR curves displayed in Figure 2.5. The distance between the two curves is obvious compared to the Gaussian case, favouring the causal specification at almost all quantiles. Figure 2.6 is the SRAR plot of a purely noncausal process with i.i.d. Cauchy innovations. The noncausal specification is preferred in the SRAR comparison.

It seems now that applying the SRAR comparison at one quantile, such as the median, is sufficient for model identification, but it is not true in general. In Section 2.4, we will spot an identification issue in the SRAR plots, the true model even having higher SRAR values at certain quantiles than the misspecified model.

So far we have applied QCAR and its counterpart QNCAR on the classical purely causal or noncausal models with symmetrically i.i.d. innovations. Within this restricted scope, the conditional mean models of those data generating processes only differ from their conditional quantile models in the intercept term. And we see that model selection by the SRAR comparison gives uniform decisions along quantiles. However, such a model selection is not always that clear in practice. For example, in the empirical study later, we will encounter an identification issue which can be seen by checking the SRAR plots. In the next section, we will present this identification issue with some possible reasons, and propose a robust model selection criterion called the aggregate SRAR to cope with this issue.
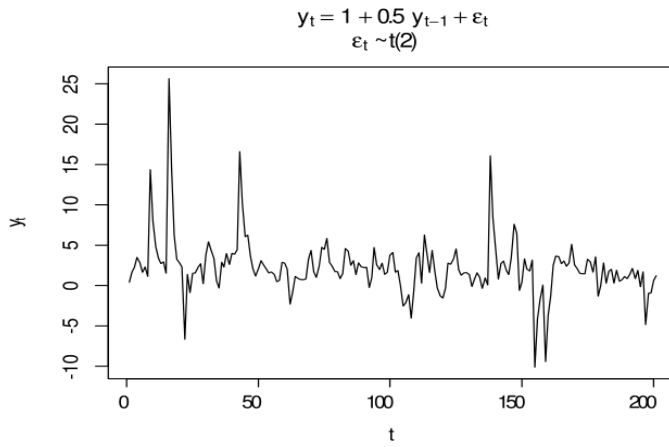
Figure 2.4: Simulation of an AR(1) with $t(2)$ innovations, $T = 200$
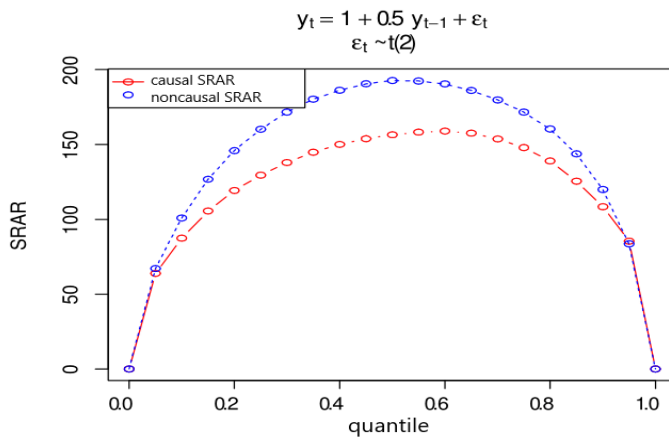
Figure 2.5: SRAR plot under an AR(1) with $t(2)$, $T = 200$
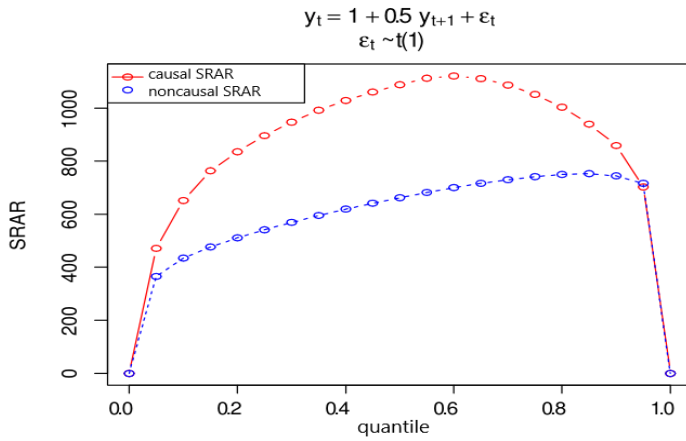
Figure 2.6: SRAR plot under a noncausal model with Cauchy innovations, $T = 200$
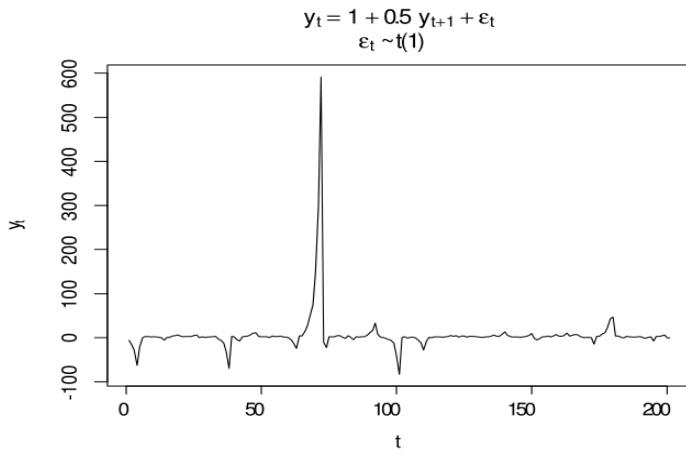
Figure 2.7: Simulation of a noncausal model with Cauchy innovations, $T = 200$

2.4 SRAR as a Model Selection Criterion

It is natural to think about SRAR as a model selection criterion since a lower SRAR means a better goodness of fit in quantile regressions. However, SRAR is a function of quantile, which raises a question on which quantile to be considered for model selection. It is empirically common to see an identification problem by checking the SRAR plots, which gives different model selections at certain quantiles and makes a selection unreliable if only one quantile is considered. In this section, we discuss this issue and propose a more robust model selection criterion based on aggregating SRARs.

2.4.1 Identification issue spotted from the SRAR plots

First let us see some possible model settings causing the identification problem in SRAR plots. The first case is linked to the existence of multi-regimes in coefficients.

Suppose a regime-shift model is specified as follows:

$$y_t = \beta_t \, y_{t+1} + \varepsilon_t \,, \tag{2.19}$$

where $\{\varepsilon_t\}$ is an i.i.d. innovation process with cumulative probability function $F(\cdot)$, and β_t is defined as follows:

$$\beta_t = \begin{cases} \beta_1, & \text{if } 0 < F(\varepsilon_t) \leq \tau^*; \\ \beta_2, & \text{if } \tau^* < F(\varepsilon_t) \leq 1, \end{cases} \tag{2.20}$$

with $\tau^* \in (0,1)$ and $\beta_1 < \beta_2$. In essence, the regime shift of β_t depends on the quantile occurrence of ε_t which is indexed by $\tau_t := F(\varepsilon_t)$ with $\{\tau_t\}$ being i.i.d. in the standard uniform distribution.

If $\{y_t\}$ can be negative, then there is a problem in using QNCAR to recover the coefficients in the underlying model (2.19) because the τ-th regime is not necessary to produce the τ-th conditional quantile of y_t. So the comonotonicity condition of the linear quantile regressions (2.19) is not satisfied. However, by restricting to the non-negative region of the covariate y_{t+1} (also see Fan and Fan, 2010) we force the regression model to satisfy the comonotonicity requirement without losing its association with $\{\tau_t\}$. The obtained estimator is also consistent to the true coefficients in (2.19).[5] QCAR (or QNCAR) with such a

---

[5]An alternative is to use simultaneous linear quantile regressions (see Tokdar and Kadane, 2011; Liu, 2019) which assumes a base quantile function for all coefficients in a (reparametrized) regression model. The dependence structure of these coefficients and priors on parameters in the base quantile function are further assumed for simulation, model fitting and posterior distribution summary. We do not include this method here as this chapter aims to avoid assumptions on parametrizing underlying conditional distributions in the model setting. Additionally, re-

restriction, hereafter denoted as RQCAR (or RQNCAR) shorthand for restricted quantile causal autoregression (or restricted quantile noncausal autoregression), is formulated as follows:

$$\hat{\boldsymbol{\theta}}(\tau) = \underset{\boldsymbol{\theta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{t=1}^{T} \mathbb{1}_{\{t \in \mathfrak{T}\}} \rho_{\tau}(y_t - \boldsymbol{x}_t' \boldsymbol{\theta}) \tag{2.21}$$

where $\mathfrak{T}$ is the set restricting the quantile regression on a particular information set. The restriction is usually imposed in order for quantile regressions to meet the comonotonicity condition. To study the regime-shift model (2.19), we restrict the QNCAR on non-negative covariates, i.e., $\mathfrak{T} = \{t : \boldsymbol{x}_t \geq \boldsymbol{0}\}$.

Figure 2.8 shows four SRAR curves estimated from QCAR, QNCAR, RQCAR and RQNCAR. We consider a time series $\{y_t\}_{t=1}^{600}$ simulated from the model (2.19) with $\tau^* = 0.7, \beta_1 = 0.2, \beta_2 = 0.8$ and i.i.d. innovation process in $t(3)$, i.e., $F^{-1}(\cdot) = F_{t(3)}^{-1}(\cdot)$. Figure 2.8 illustrates such an identification issue in
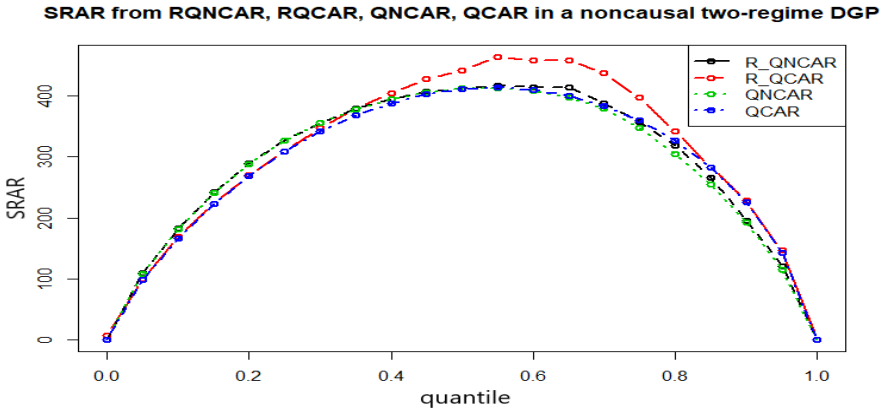
Figure 2.8: Identification problem spotted in the SRAR plot for restricted quantile autoregressions

which the SRAR curve from a true model is not always lower than one from misspecification. Applying restriction helps to enlarge the SRAR difference between a true model and a misspecified time direction.

The second case we investigate is the presence of skewed distributed innovations.

Let us consider a time series $\{y_t\}$ following a purely noncausal AR(1): $y_t = 0.8 y_{t+1} + \varepsilon_t$ with $\{\varepsilon_t\}$ i.i.d. in a demeaned skewed t distribution with skewing

---

garding the data in our empirical study, very few negative points are observed. So the concern on losing data by such a restriction can be addressed.

parameter $\gamma = 2$ and $v = 3$ degrees of freedom (hereafter $t(v, \gamma)$ is the shorthand notation for a skewed t-distribution). The probability density function of $t(v, \gamma)$ (see Francq and Zakoïan 2007) is defined as

$$f(x) = \begin{cases} \dfrac{2}{\gamma + \frac{1}{\gamma}} f_t(\gamma x) & \text{for } x < 0 \\[2ex] \dfrac{2}{\gamma + \frac{1}{\gamma}} f_t(\dfrac{x}{\gamma}) & \text{for } x \geq 0 \end{cases} \tag{2.22}$$

where $f_t(\cdot)$ is the probability density function of the symmetric $t(v)$ distribution. Figure 2.9 shows four SRAR curves obtained from the estimation of the QCAR, the QNCAR, the RQCAR and the RQNCAR respectively. The curves from the QNCAR and the RQNCAR almost overlap each other, which confirms our understanding that the monotonicity requirement is met in the true model. The estimations and the corresponding SRAR curves should be the same unless many observations are omitted by the restriction. On the other hand, the SRAR curve gets enlarged from the QCAR to the RQCAR, which is very reasonable as the feasible set in the QCAR is larger and the misspecification is not ensured to satisfy the monotonicity requirement. Again we see this identification problem from the SRAR plot. Remarkably, the SRAR curve from a true model can be higher at certain quantiles than the one from a misspecified model. Consequently the SRAR comparison relying only on particular quantiles, such as the least absolute deviation (LAD) method for the median only, is not robust in general. Therefore, we propose a new model selection criterion in next subsection by including the information over all quantiles.
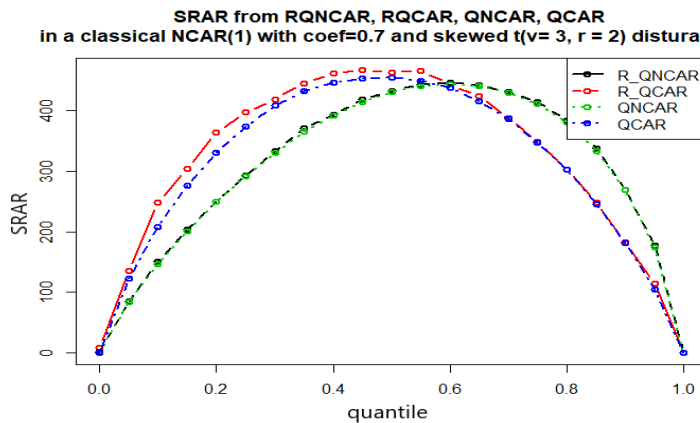
Figure 2.9: Identification issue spotted from the SRAR plot for a skewed distribution

2.4.2 The aggregate SRAR criterion

Based on the same number of explanatory variables in QCAR and QNCAR with a fixed sample size in quantile regressions, the best model is supposed to exhibit the highest goodness of fit among candidate models. Similarly to the R-squared criterion in the OLS, when turning to quantile regressions, we are led to use the SRAR criterion for model selection. The aggregate SRAR is regarded as an overall performance of a candidate model over all quantiles such as:

$$\text{aggregate SRAR} := \int_0^1 SRAR(\tau)d\tau.$$

There are many ways to calculate this integral. One way is to approximate the integral by the trapezoidal rule. Another way is to sum up SRARs over a fine enough quantile grid with equal weights. In other words, this aggregation is regarded as an average of performances ($SRAR(\tau)$, $\tau \in (0,1)$) of a candidate model. In practice, there is almost no difference in model selection between the two aggregation methods.

As equal weights are used on all quantiles in the aggregate SRAR above, people may argue to use a different weighting scheme. The weighting scheme indeed can be different as when weight being one for one quantile and zero for others is to select a conditional quantile model. We agree that the weighting scheme can be customized in justice of users' purpose. The equal-weight scheme proposed here is inspired to calculate the area under the SRAR curve over quantiles when we check the SRAR plot. Intuitively, such areas are directly linked to model performance when we concerns the whole dynamics of the underlying process. And we compare models by viewing the gap between their SRAR curves, which is the difference between the areas under their SRAR curves. This leads us to the aggregate SRAR measure.

Performances of the SRAR model selection criteria in Monte Carlo simulations are reported in Table 2.1. It shows the average frequencies with which we find the correct model based on the SRAR criterion per quantile and the aggregate SRAR criterion. The sample size T is 200 and each reported number is based on 2000 Monte Carlo simulations. Columns of Table 2.1 refer to as a particular distribution previously illustrated in this chapter. As observed, the aggregate SRAR criterion performs very well even in situations with the identification issue. The Gaussian distribution being weakly and strictly stationary we cannot obviously discriminate between causal and noncausal specifications leading to an average frequency of around 50% to detect the correct model.

Table 2.1: Frequencies of selecting the correct model using the SRAR criteria

Quantiles	Gaussian (Fig. 2.2)	$t(2)$ (Fig. 2.4)	$t(1)$ (Fig. 2.7)	two-regime (Fig. 2.8)	$t(v=3,\gamma=2)$ (Fig. 2.9)
0	0.698	0.678	0.601	0.787	0.476
0.05	0.516	0.416	0.653	0.044	1.000
0.10	0.51	0.677	0.763	0.059	1.000
0.15	0.519	0.858	0.841	0.095	1.000
0.20	0.512	0.948	0.907	0.167	1.000
0.25	0.513	0.981	0.947	0.305	1.000
0.30	0.488	0.992	0.978	0.487	1.000
0.35	0.487	0.998	0.996	0.654	1.000
0.40	0.486	0.999	0.996	0.798	1.000
0.45	0.487	1.000	0.996	0.892	1.000
0.50	0.5	1.000	0.995	0.950	1.000
0.55	0.499	0.999	0.995	0.974	0.994
0.60	0.492	0.999	0.995	0.988	0.533
0.65	0.478	0.997	0.995	0.991	0.018
0.70	0.467	0.994	0.979	0.996	0.001
0.75	0.49	0.984	0.951	0.998	0.000
0.80	0.493	0.954	0.903	0.999	0.000
0.85	0.481	0.862	0.858	1.000	0.000
0.90	0.469	0.72	0.791	1.000	0.000
0.95	0.484	0.454	0.668	0.997	0.000
1	0.653	0.58	0.595	0.780	0.420
aggregate SRAR	0.483	0.998	0.995	0.995	0.999

2.4.3 Shape of SRAR curves

By observing SRAR plots, we see that SRAR curves vary when the underling distribution varies. It is interesting to investigate the reasons. In this subsection, we will provide some insights on the slope and concavity of $\text{SRAR}_{y_t}(\tau, \hat{\boldsymbol{\theta}}(\tau))$ curves under assumptions (A1), (A2), (A3) and (A4). Since $\rho_\tau(y_t - \boldsymbol{x}_t'\boldsymbol{\theta})$ is a continuous function in $\boldsymbol{\theta} \in \mathbb{R}^{(p+1)}$, by the continuous mapping theorem and $\hat{\boldsymbol{\theta}}(\tau)) \xrightarrow{p} \boldsymbol{\phi}_\tau$, we know that

$$\rho_\tau(y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\theta}}) \xrightarrow{p} \rho_\tau(y_t - \boldsymbol{x}_t'\boldsymbol{\phi}_\tau).$$

We also know that

$$\rho_\tau(y_t - \boldsymbol{x}_t'\boldsymbol{\phi}_\tau) = \rho_\tau(\varepsilon_t - F^{-1}(\tau)).$$

Therefore instead of directly deriving the shape of a $\text{SRAR}_{y_t}(\tau, \hat{\boldsymbol{\theta}}(\tau))$ curve, we look at the properties of its intrinsic curve $\text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))$. We derive the first and second order derivatives of $\text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))$ with respect to τ in

order to determine the shape of $\text{SRAR}_{y_t}(\tau, \hat{\boldsymbol{\theta}}(\tau))$.

The slope property

One major difference between SRAR curves in a plot is their slopes. We can compute the first-order derivative of SRAR with respect to τ if the derivative exists. Under the following assumption:

(A7): The inverse distribution function $F^{-1}(\cdot)$ of innovation ε_t is continuous and differentiable on $(0,1)$ to the second order;

we can then take the first-order derivative of $\text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))$ with respect to τ.

Suppose $0 < \tau < \tau + \Delta\tau < 1, \Delta\tau > 0$ and denote $\Delta F^{-1}(\tau) := F^{-1}(\tau + \Delta\tau) - F^{-1}(\tau)$.

$$\text{SRAR}_{\varepsilon_t}(\tau + \Delta\tau, F^{-1}(\tau + \Delta\tau)) - \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))$$

$$= \sum_{t=1}^{T} \left(\rho_{\tau+\Delta\tau}\left(\varepsilon_t - F^{-1}(\tau + \Delta\tau)\right) - \rho_\tau\left(\varepsilon_t - F^{-1}(\tau)\right) \right)$$

$$= \sum_{t=1}^{T} \left(\left(\varepsilon_t - F^{-1}(\tau + \Delta\tau)\right)\left(\tau + \Delta\tau - \mathbb{1}_{\{\varepsilon_t - F^{-1}(\tau+\Delta\tau) \leq 0\}}\right) - \left(\varepsilon_t - F^{-1}(\tau)\right)\left(\tau - \mathbb{1}_{\{\varepsilon_t - F^{-1}(\tau) \leq 0\}}\right) \right)$$

$$= \sum_{t=1}^{T} \left(\varepsilon_t \left(\Delta\tau - \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau+\Delta\tau)\}}\right) + \tau\left(F^{-1}(\tau) - F^{-1}(\tau + \Delta\tau)\right) - \Delta\tau F^{-1}(\tau + \Delta\tau) \right.$$

$$\left. + F^{-1}(\tau + \Delta\tau)\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau+\Delta\tau)\}} - F^{-1}(\tau)\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau)\}} \right)$$

$$= \sum_{t=1}^{T} \left(\Delta\tau\left(\varepsilon_t - F^{-1}(\tau + \Delta\tau)\right) + \left(F^{-1}(\tau+\Delta\tau) - F^{-1}(\tau)\right)\left(\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau+\Delta\tau)\}} - \tau\right) \right.$$

$$\left. + \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau+\Delta\tau)\}}\left(F^{-1}(\tau) - \varepsilon_t\right) \right).$$

$$(2.23)$$

Divide the above difference by $\Delta\tau$, and take the limit $\Delta\tau \downarrow 0$. It gives us

$$\lim_{\Delta\tau\downarrow 0} \frac{\text{SRAR}_{\varepsilon_t}(\tau + \Delta\tau, F^{-1}(\tau + \Delta\tau)) - \text{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{\Delta\tau}$$

$$= \sum_{t=1}^{T} \left(\varepsilon_t - F^{-1}(\tau) + \frac{dF^{-1}(\tau)}{d\tau}\left(\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau)\}} - \tau\right) \right),$$

$$(2.24)$$

because

$$\lim_{\Delta\tau\downarrow 0} \frac{\Delta\tau\left(\varepsilon_t - F^{-1}(\tau+\Delta\tau)\right)}{\Delta\tau} = \varepsilon_t - F^{-1}(\tau),$$

$$\lim_{\Delta\tau\downarrow 0} \frac{\left(F^{-1}(\tau+\Delta\tau) - F^{-1}(\tau)\right)\left(\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau+\Delta\tau)\}} - \tau\right)}{\Delta\tau} = \frac{dF^{-1}(\tau)}{d\tau}\left(\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau)\}} - \tau\right), \quad (2.25)$$

$$\lim_{\Delta\tau\downarrow 0} \frac{\mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau+\Delta\tau)\}}\left(F^{-1}(\tau) - \varepsilon_t\right)}{\Delta\tau} = 0.$$

The last line is from

$$\begin{cases} \mathbb{1}_{\{F^{-1}(\tau)<\varepsilon_t\leq F^{-1}(\tau+\Delta\tau)\}}\left(F^{-1}(\tau)-\varepsilon_t\right)=0, & \text{when} \quad \varepsilon_t\notin\left(F^{-1}(\tau),F^{-1}(\tau+\Delta\tau)\right]; \\ \left(F^{-1}(\tau)-F^{-1}(\tau+\Delta\tau)\right)\leq\left(F^{-1}(\tau)-\varepsilon_t\right)<0, & \text{when} \quad \varepsilon_t\in\left(F^{-1}(\tau),F^{-1}(\tau+\Delta\tau)\right]; \end{cases}$$

(2.26)

and

$$0=\mathbb{1}_{\{F^{-1}(\tau)<\varepsilon_t\leq F^{-1}(\tau)\}}\frac{dF^{-1}(\tau)}{d\tau}\leq\lim_{\Delta\tau\downarrow0}\frac{\mathbb{1}_{\{F^{-1}(\tau)<\varepsilon_t\leq F^{-1}(\tau+\Delta\tau)\}}\left(F^{-1}(\tau)-\varepsilon_t\right)}{\Delta\tau}\leq0. \qquad (2.27)$$

In analogue, the left-hand limit $\lim_{\Delta\tau\uparrow0}\frac{\mathrm{SRAR}_{\varepsilon_t}(\tau+\Delta\tau,F^{-1}(\tau+\Delta\tau))-\mathrm{SRAR}_{\varepsilon_t}(\tau,F^{-1}(\tau))}{\Delta\tau}$ gives the same result. Therefore, we have the first-order derivative as below.

$$\frac{d\,\mathrm{SRAR}_{\varepsilon_t}(\tau,F^{-1}(\tau))}{d\tau}=\sum_{t=1}^{T}\left(\varepsilon_t-F^{-1}(\tau)+\frac{dF^{-1}(\tau)}{d\tau}\left(\mathbb{1}_{\{\varepsilon_t\leq F^{-1}(\tau)\}}-\tau\right)\right).$$

(2.28)

To manifest this result with a fixed T, we take expectation and obtain

$$\mathbb{E}\left[\frac{d\,\mathrm{SRAR}_{\varepsilon_t}(\tau,F^{-1}(\tau))}{d\tau}\right]=T\left(\mathbb{E}\left[\varepsilon_t\right]-F^{-1}(\tau)\right), \qquad (2.29)$$

when $\mathbb{E}\left[\varepsilon_t\right]$ exists. In practice, we are not strict with $\mathbb{E}\left[\varepsilon_t\right]<\infty$ since the mean of an i.i.d. $\{\varepsilon_t\}_{t=1}^{T}$ can be estimated empirically to replace $\mathbb{E}\left[\varepsilon_t\right]$ in (2.29) without affecting other terms.

Now we have the expectation of $\frac{d\,\mathrm{SRAR}_{\varepsilon_t}(\tau,F^{-1}(\tau))}{d\tau}$ which can be regarded as the underlying guideline for the slope of a SRAR curve. Before interpreting this result, let us derive the second-order derivative of $\mathrm{SRAR}_{\varepsilon_t}(\tau,F^{-1}(\tau))$ with respect to τ and make an interpretation together.

The concave property

One empirically observed property of SRAR curves is their concavity which can be explained through the second-order derivative of $\mathrm{SRAR}_{\varepsilon_t}(\tau,F^{-1}(\tau))$ with respect to τ under assumptions (A1), (A2), (A3), (A4) and (A7). Suppose $0<$

$\tau < \tau + \Delta\tau < 1, \Delta\tau > 0.$

$$\Delta^2 \operatorname{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau)) :=$$
$$\operatorname{SRAR}_{\varepsilon_t}(\tau + \Delta\tau, F^{-1}(\tau + \Delta\tau)) - 2\operatorname{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau)) + \operatorname{SRAR}_{\varepsilon_t}(\tau - \Delta\tau, F^{-1}(\tau - \Delta\tau))$$
$$= \sum_{t=1}^{T} \Bigg(\left(\varepsilon_t - F^{-1}(\tau)\right) \left(\mathbb{1}_{\{F^{-1}(\tau - \Delta\tau) < \varepsilon_t \leq F^{-1}(\tau)\}} - \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} \right)$$
$$+ \tau \left(2F^{-1}(\tau) - F^{-1}(\tau + \Delta\tau) - F^{-1}(\tau - \Delta\tau)\right) + \Delta\tau \left(F^{-1}(\tau - \Delta\tau) - F^{-1}(\tau + \Delta\tau)\right)$$
$$+ \left(F^{-1}(\tau + \Delta\tau) + F^{-1}(\tau - \Delta\tau) - 2F^{-1}(\tau)\right) \mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau - \Delta\tau)\}}$$
$$+ \left(F^{-1}(\tau + \Delta\tau) - F^{-1}(\tau)\right) \mathbb{1}_{\{F^{-1}(\tau) < \varepsilon_t \leq F^{-1}(\tau + \Delta\tau)\}} \Bigg).$$

$$(2.30)$$

Divide the above second order central difference by $\Delta\tau^2$, and take the limit $\Delta\tau \downarrow 0$. It gives us

$$\frac{d^2 \operatorname{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{d\tau^2} = \lim_{\Delta\tau \downarrow 0} \frac{\Delta^2 \operatorname{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{\Delta\tau^2}$$
$$= \sum_{t=1}^{T} \left(\frac{d^2 F^{-1}(\tau)}{d\tau^2} \left(\mathbb{1}_{\{\varepsilon_t \leq F^{-1}(\tau)\}} - \tau \right) - 2 \frac{dF^{-1}(\tau)}{d\tau} \right),$$

$$(2.31)$$

the last line of which is obtained similarly to (2.25). To interpret this result, we take expectation and get the following:

$$\mathbb{E}\left[\frac{d^2 \operatorname{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{d\tau^2} \right] = -2 \frac{dF^{-1}(\tau)}{d\tau} T < 0. \qquad (2.32)$$

where the inequality holds with probability one since $f(\varepsilon) > 0$ with probability one in the assumption (A1). Now we have the expectation of $\frac{d^2 \operatorname{SRAR}_{\varepsilon_t}(\tau, F^{-1}(\tau))}{d\tau^2}$ which can be regarded as the underlying guideline for the concavity of a SRAR curve. Together with the slope information, it implies that SRAR curves are always in arch shapes, going upward and then downward, with a peak point at $\mathbb{E}[\varepsilon_t] = F^{-1}(\tau)$. We can also know the skewness of ε_t from the location of the peak point: ε_t is left-skewed when the SRAR curve reaches its peak in the region $\tau < 0.5$, or right-skewed when the peak in $\tau > 0.5$. If ε_t is symmetrically distributed, its SRAR curve is symmetric, and vice versa.

2.5 Binding Functions

Plotting SRAR is a way to present the goodness of fit in quantile regressions for each candidate model. Quantile regressions are the path to get residuals for SRAR calculation. As we know and provide unbiased consistent estimation for true models. To study the estimation in misspecification we adopt the concept of binding function (Dhaene, Gourieroux and Scaillet, 1998). Binding function is

defined as a mapping from coefficients in the true model to pseudo-true coefficients in a misspecified model.

The estimator of a pseudo-true coefficient in quantile regression for a misspecified QCAR(p) or QNCAR(p) converges to a limiting value which is characterized into the binding function. It is difficult to derive the binding functions explicitly in a general case so that they are studied by means of simulations (see Gouriéroux and Jasiak, 2017). Suppose a noncausal AR(1): $y_t = \pi_1 y_{t+1} + \varepsilon_t$, with $\{\varepsilon_t\}$ i.i.d. $t(v)$ for $v = 1, 3, 5$ and 10. It is observed that the binding function in the misspecified QCAR(1) varies with two factors: (i) the distribution of ε_t and (ii) the distance function in regression which is the check function $\rho_\tau(\cdot)$ in quantile regression. Figure 2.10, Figure 2.11 and Figure 2.12 illustrate the effect of those factors. Each point is an average value of estimates based on 1000 simulations and 600 observations. Since $t(v)$ is symmetric, the estimation results are in the same pattern for negative true coefficient region and $(1 - \tau)$th-quantile regression as in these three figures. Sometimes the binding function is not injective, which is evidenced in Figure 2.10 and Figure 2.11 for small absolute true coefficients. The non-injectivity of the binding function for Cauchy distributed innovations is also illustrated in Gouriéroux and Jasiak (2017) result that disables encompassing tests. On the other hand, we see that on Figure 2.12 the injectivity of binding functions seems recovered at $\tau = 10\%$. In the case of Cauchy distributed innovations, there are no binding functions from extreme quantile regressions like 0.1th- or 0.9th-quantile regression because the estimate is not convergent. Although a value for $\pi_1 \in (0, 1)$ is plotted in Figure 2.12, it is just the average of binding function estimates for π_1 for illustration.
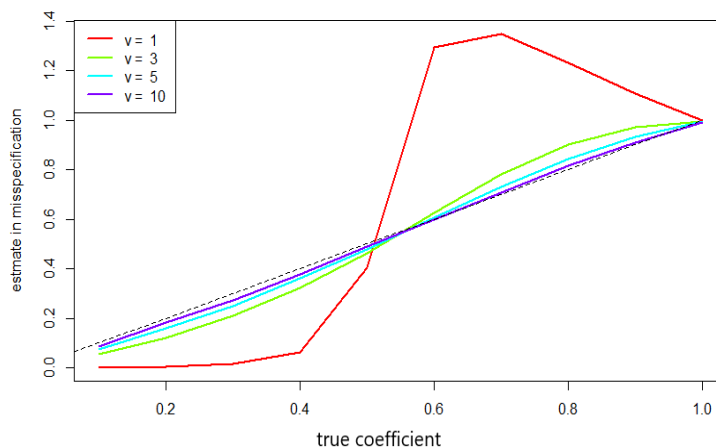
Figure 2.10: Binding function for a misspecified QCAR(1) in 0.5th-quantile regression
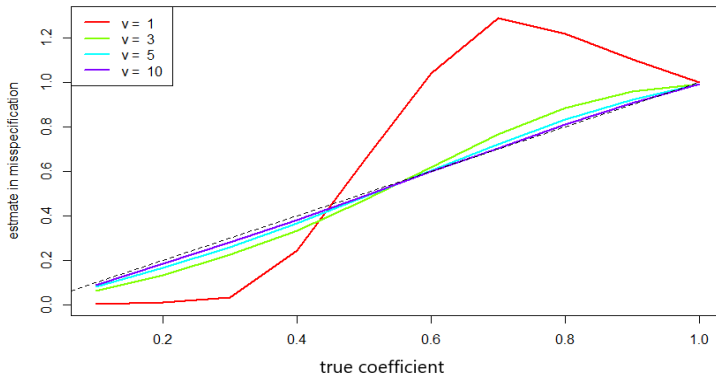
Figure 2.11: Binding function for a misspecified QCAR(1) in 0.3th-quantile regression
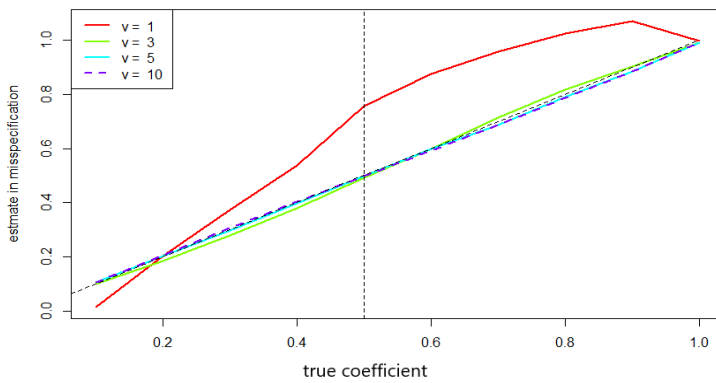
Figure 2.12: Binding function for a misspecified QCAR(1) in 0.1th-quantile regression

2.6 Modelling Hyperinflation in Latin America

2.6.1 The model specification

The motivation of our empirical analysis comes from the rational expectation (RE) hyperinflation model originally proposed by Cagan (1956) and investigated by several authors (see e.g. Adam and Szafarz, 1992; Broze and Szafarz, 1985). We follow Broze and Szafarz (1985) notations with

$$m_t^d = \alpha p_t + \beta E(p_{t+1}|I_t) + x_t. \tag{2.33}$$

In (2.33), m_t^d and p_t respectively denote the logarithms of money demand and price, x_t is the disturbance term summarizing the impact of exogenous factors. $E(p_{t+1}|I_t)$ is the rational expectation, when it is equal to conditional expectation, of p_{t+1} at time t based on the information set I_t. Assuming that the money supply $m_t^s = z_t$ is exogenous, the equilibrium $m_t^d = m_t^s$ provides the following equation for prices

$$\begin{aligned} p_t &= -\frac{\beta}{\alpha}[E(p_{t+1}|I_t)] + \frac{z_t - x_t}{\alpha}, \\ &= \phi[E(p_{t+1}|I_t)] + u_t. \end{aligned}$$

Broze and Szafarz (1985) show that a forward-looking recursive solution of this model exists when x_t is stationary and $|\phi| < 1$. The deviation from that solution is called the bubble B_t with $p_t = \sum_{i=0}^{\infty} \phi^i E(u_{t+i}|I_t)] + B_t$. Finding conditions under which this process has rational expectation equilibria (forward and or backward looking) is out of the scope of our chapter. We only use this framework to illustrate the interest of economists for models with leads components. Under a perfect foresight scheme $E(p_{t+1}|I_t) = p_{t+1}$ we obtain the purely noncausal model

$$p_t = \phi p_{t+1} + \tilde{\varepsilon}_t, \tag{2.34}$$

with $\tilde{\varepsilon}_t = u_t$. In the more general setting, for instance when $E(p_{t+1}|I_t) = p_{t+1} + v_t$ with v_t a martingale difference, the new disturbance term is $\tilde{\varepsilon}_t = v_t + u_t$. Empirically, a specification with one lead only might be too restrictive to capture the underlying dynamics of the observed variables. We consequently depart from the theoretical model proposed above and we consider empirical specifications with more leads or lags. Lanne and Luoto (2013, 2017) and Hecq et al. (2017) in the context of the new hybrid Keynesian Phillips curve assume for instance that $\tilde{\varepsilon}_t$ is a MAR$(r-1, s-1)$ process such as

$$\rho(L)\pi(L^{-1})\tilde{\varepsilon}_t = c + \varepsilon_t, \tag{2.35}$$

where ε_t is *iid* and c an intercept term. Inserting (2.35) in (2.34) we observe that if $\tilde{\varepsilon}_t$ is a purely noncausal model (i.e. a MAR$(0, s-1)$ with $\rho(L) = 1$), we obtain

a noncausal MAR$(0,s)$ motion for prices

$$(1 - \phi L^{-1})p_t = \pi(L^{-1})^{-1}(c + \varepsilon_t),$$

$$(1 - \phi L^{-1})(1 - \pi_1 L^{-1} - ... - \pi_{s-1}L^{-(s-1)})p_t = c + \varepsilon_t,$$

We would obtain a mixed causal and noncausal model if $\rho(L) \neq 1$. Our guess that the same specification might in some circumstances empirically (although not mathematically as the lag polynomial does not annihilate the lead polynomial) gives rise to a purely causal model in small samples when the autoregressive part dominates the lead component.

The above illustration presents a context of pure causal and noncausal models so that we can apply our approach to give an empirical analysis. It would be interesting to extend our modelling to investigate theoretical models with both forward and backward behaviours such as backward- and forward-looking Taylor rule for instance. To do that however we have to introduce additional regressors and extend the approach of Hecq, Issler and Telg (2020) to quantile regressions, which can be further investigated by future research and is out of the scope of this chapter.

2.6.2 The data and unit root testing

We consider seasonally unadjusted quarterly Consumer Price Index (CPI) series for four Latin American countries: Brazil, Mexico, Costa Rica and Chile. Monthly raw price series are downloaded at the OECD database for the largest span available (in September 2018). Despite the fact that quarterly data are directly available at OECD, we do not consider those series as they are computed from the unweighted average over three months of the corresponding quarters. Hence, these data are constructed using a linear filter, leading to undesirable properties for the detection of mixed causal and noncausal models (see Hecq, Telg and Lieb, 2017 on this specific issue). As a consequence, we use quarterly data computed by point-in-time sampling from monthly variables. The first observation is 1969Q1 for Mexico, 1970Q1 for Chile, 1976Q1 for Costa Rica and 1979Q4 for Brazil. Our last observation is 2018Q2 for every series. We do not use monthly data in this chapter as monthly inflation series required a very large number of lags to capture their dynamic feature. Moreover, the detection of seasonal unit roots in the level of monthly price series was quite difficult.

Applying seasonal unit root tests (here HEGY tests, see Hylleberg et al., 1990) with a constant, a linear trend and deterministic seasonal dummies, we reject (see Table 2.2 in which a * denotes a rejection of the null unit root hypothesis at a specific frequency corresponding to 5% significance level) the null of seasonal unit roots in each series whereas we do not reject the null of a unit root at the

zero frequency. The implementation of the unit root tests here is concerned with conditional mean models of the raw data to ensure that we process the data and use its weakly stationary time series in quantile regressions for analysis. The unit root testing can also been done per quantile (Koenker and Xiao, 2004) to relate short-term explosiveness of time series to unit-root quantile models, which is an interesting perspective to treat explosive time series and alternative to causal and noncausal modelling. We do not go deeper in the unit-root direction for this chapter but with its outlook on future research.

The number of lags of the dependent variable used to whiten for the presence of autocorrelation is chosen by AIC. From these results we compute quarterly inflation rates for the four countries in annualized rate, i.e. $\Delta \ln P_t^i \times 400$. Next we carry out a regression of $\Delta \ln P_t^i \times 400$ on seasonal dummies to capture the potential presence of deterministic seasonality. The null of no deterministic seasonality is not rejected for the four series. Figure 2.13 displays quarterly inflation rates and it illustrates the huge inflation episodes that the countries had faced. Among the four inflation rates, Brazil and Mexico show the typical pattern closer to the intuitive notion of what a speculative bubble is, namely a rapid increase of the series until a maximum value is reached before the bubble bursts.
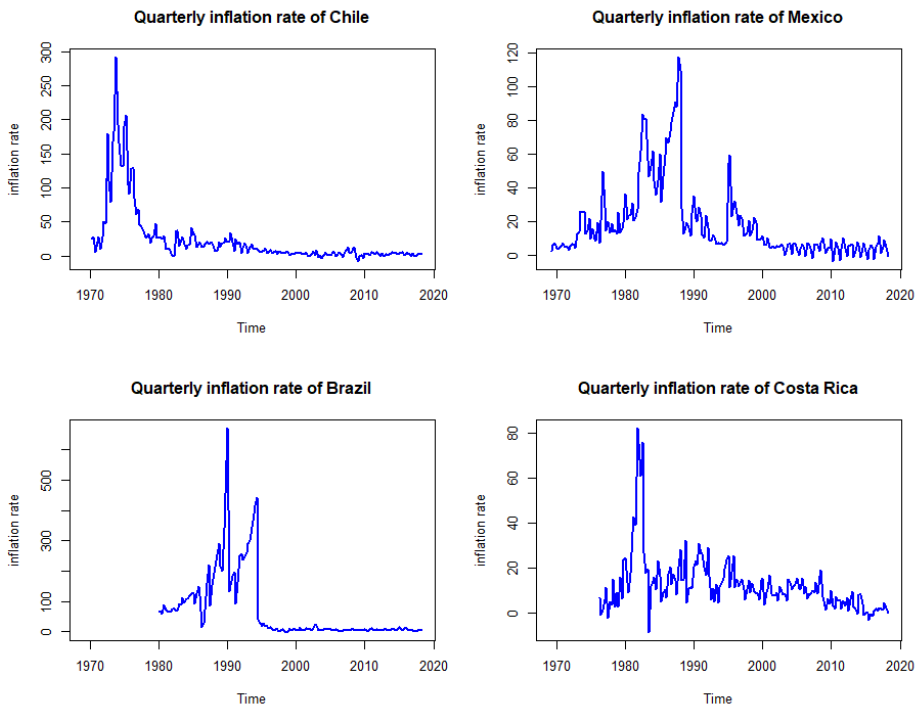
Figure 2.13: Quarterly inflation rate series plot for 4 Latin American countries

Table 2.2: Seasonal HEGY unit root tests in the log levels of prices

Country	$H_0 : \pi_1 = 0$	$H_0 : \pi_2 = 0$	$H_0 : \pi_3 = \pi_4 = 0$	Sample
$\ln P_t^{Bra}$	-1.39	$-5.75*$	$48.28*$	$1979Q4 - 2018Q2$
$\ln P_t^{Chi}$	-2.98	$-6.32*$	$20.13*$	$1970Q1 - 2018Q2$
$\ln P_t^{Costa}$	-1.80	$-4.23*$	$7.81*$	$1976Q1 - 2018Q2$
$\ln P_t^{Mex}$	-0.88	$-11.92*$	$60.10*$	$1969Q1 - 2018Q2$

2.6.3 Empirical findings and identification of noncausal models

Table 2.3 reports for each quarterly inflation rates the autoregressive model obtained using the Hannan-Quinn information criterion. Given our results on the binding function (see also Gouriéroux and Jasiak, 2017) it is safer to determine the pseudo-true autoregressive lag length using such an OLS approach than using quantile regressions or using maximum likelihood method. Indeed there is the risk that a regression in direct time from a noncausal DGP provides an underestimation of the lag order for some distributions (e.g. the Cauchy) and some values of the parameters.

Estimating autoregressive univariate models gives the lag length range from $p = 1$ for Brazil to $p = 7$ for the Chilean inflation rate. The $p - values$ of the Breush-Pagan LM test (see column labeled $LM[1 - 2]$) for the null of no-autocorrelation after having included those lags show that we do not reject the null in every four cases. On the other hand, we reject the null of normality (Jarque-Bera test) in the disturbances of each series. We should consequently be able to identify causal from noncausal models. From columns *skew.* and *kurt.* it emerges that the residuals are skewed to the left for Brazil and Mexico and skewed to the right for Chile and Costa Rica. Heavy tails are present in each series. At a 5% significance level we reject the null of no ARCH (see column $ARCH[1 - 2]$) for Brazil and Mexico. Gouriéroux and Zakoian (2017) have derived the closed form conditional moments of a misspecified causal model obtained from a purely noncausal process with alpha stable disturbances. They show that the conditional mean (in direct time) is a random walk with a time varying conditional variance in the Cauchy case. This result would maybe favour the presence of a purely noncausal specification for Brazil and Mexico as the null of no ARCH is rejected. But this assertion must be carefully evaluated and tested, for instance using our comparison of quantile autoregressions in direct and reverse time. The results by the Q(N)CAR are reported in Table 2.4, and the RQ(N)CAR produces the same results. Each cell of Table 2.4 provides the selection frequency of MAR$(0, p)$ or MAR$(p, 0)$ identified by the SRAR at quantiles 0.1, 0.3, 0.5, 0.7, 0.9 as well as the aggregated SRAR. Figure 2.14 displays the SRAR curves from 0.05th-quantile to 0.95th-quantile by the Q(N)CAR for the four economies respectively, similarly to the ones by the RQ(N)CAR with

Table 2.3: Descriptive statistics for quarterly inflation rates

Country	HQ	BJ	$skew.$	$kurt.$	$LM[1-2]$	$ARCH[1-2]$
$\Delta \ln P_t^{Bra}$	1	< 0.001	-2.54	56.96	0.19	< 0.001
$\Delta \ln P_t^{Chi}$	7	< 0.001	2.84	22.45	0.09	0.09
$\Delta \ln P_t^{Costa}$	4	< 0.001	1.01	8.73	0.47	0.30
$\Delta \ln P_t^{Mex}$	3	< 0.001	-0.40	13.81	0.20	< 0.001

restriction on non-negative regressors. As observed, the identification problem is raised in the SRAR plots. Especially in the SRAR plot for Brazil, it is hard to trust a model from evidence at single quantiles. However, the aggregate SRAR criterion comes to help in this situation from an overall perspective. We conclude that Brazil, Mexico and Costa Rica are better characterized as being purely noncausal while Chile being purely causal according to the aggregate SRAR criterion.
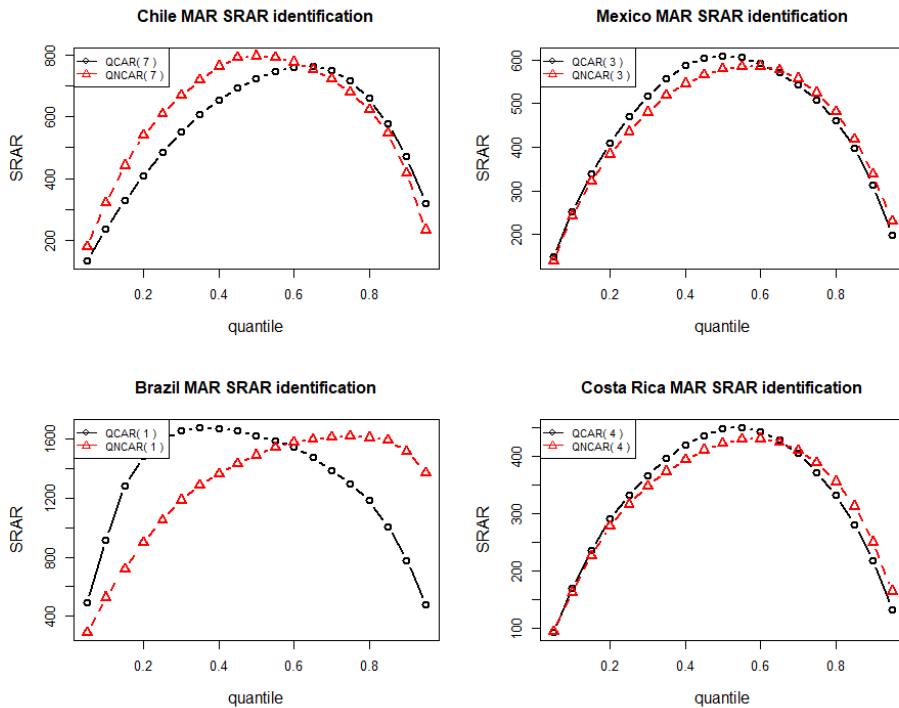
Figure 2.14: SRAR plots of the inflation rates of four Latin American countries respectively

Table 2.4: SRAR identification results

Country	$SRAR_{\tau=0.1}$	$SRAR_{\tau=0.3}$	$SRAR_{\tau=0.5}$	$SRAR_{\tau=0.7}$	$SRAR_{\tau=0.9}$	$SRAR_{total}$
$\Delta \ln P_t^{Bra}$	$MAR(0,1)$	$MAR(0,1)$	$MAR(0,1)$	$MAR(1,0)$	$MAR(1,0)$	$MAR(0,1)$
$\Delta \ln P_t^{Chi}$	$MAR(7,0)$	$MAR(7,0)$	$MAR(7,0)$	$MAR(0,7)$	$MAR(0,7)$	$MAR(7,0)$
$\Delta \ln P_t^{Costa}$	$MAR(0,4)$	$MAR(0,4)$	$MAR(0,4)$	$MAR(4,0)$	$MAR(4,0)$	$MAR(0,4)$
$\Delta \ln P_t^{Mex}$	$MAR(0,3)$	$MAR(0,3)$	$MAR(0,3)$	$MAR(3,0)$	$MAR(3,0)$	$MAR(0,3)$

2.7 Conclusions

This chapter introduces a new way to select between causal and noncausal models by comparing residuals from quantile autoregressions developed by Koenker and Xiao (2006) and from the time-reverse specifications. To adapt to heavy tailed distributions, we generalize the quantile autoregression theory for regularly varying distributions. This also confirms the validity of quantile autoregressions in analysing heavy tailed time series, such as explosive or bubble-type dynamics. It is natural to consider SRAR as a model selection criterion in the quantile regression framework. However due to the identification problem spotted in the SRAR plots as presented in this chapter, we propose to use the aggregate SRAR criterion for model selection. The robustness in its performance has been seen from all the results in this chapter. It is worth mentioning that when coefficients are constant in the underlying model with a symmetrically i.i.d. error term, the aggregate SRAR criterion is equivalently to select between forward and backward conditional mean models (termed by Gourieroux and Zakoian (2017)). However, the aggregate SRAR is a measure based on the whole dynamics of the underlying process, which is not dominated by the conditional mean information any more. This characteristic of the aggregate SRAR criterion indeed makes it robust in model selection even for some general situations such as with asymmetric distributed innovations. In the empirical study on the inflation rates of four Latin American countries, we found that the purely noncausal specification is favoured in three cases.

Finally some possible extensions of our approach can be to the identification of mixed models in addition to purely causal and noncausal specifications, to enhancing QCAR and QNCAR with some explanatory variables in order to investigate the Taylor (1993) rule, and to investigating the unit-root testing per quantile for QCAR as well as QNCAR. Also, a formal testing on SRAR differences would require the application of a bootstrap approach which is beyond the scope of this chapter but in our outlook for the future research.

2.A Appendix

Alternative Way to Simulate MAR Models

Suppose that the DGP is a MAR(r,s) as in (2.1). First, we rewrite (2.1) into a matrix representation as follows:

$$\mathbf{M}\boldsymbol{y} = \boldsymbol{\varepsilon},$$

$$\mathbf{M} := \begin{bmatrix} \pi(L)\phi(L^{-1}) & 0 & \dots & 0 \\ 0 & \pi(L)\phi(L^{-1}) & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \pi(L)\phi(L^{-1}) \end{bmatrix},$$

$$\boldsymbol{y} := \begin{bmatrix} y_1 & y_2 & \dots & y_T \end{bmatrix}',$$

$$\boldsymbol{\varepsilon} := \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \dots & \varepsilon_T \end{bmatrix}', \tag{2.36}$$

where $\mathbf{M}$ is $T \times T$ matrix and T is the sample size. The equivalence to (2.1) holds by assuming $y_{1-r}, y_{2-r}, \dots, y_0$ and $y_{T+1}, y_{T+2}, \dots, y_{T+s}$ are all zeros. This assumption effect can be neglected by deleting enough observations from the beginning and the end of a simulated sample, for instance, $\{y_t\}_{t=201}^{T-200}$ kept for analysis from a first simulated $\{y_t\}_{t=1}^{T}$. Next, $\mathbf{M}$ can be decomposed into a product of two diagonal matrices, denoted as $\mathbf{L}$ and $\mathbf{U}$, of main diagonal entries being $\pi(L)$ and $\phi(L^{-1})$ respectively as follows.

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & & 0 \\ -\pi_1 & 1 & 0 & 0 & \dots & & 0 \\ -\pi_2 & -\pi_1 & 1 & 0 & \dots & & 0 \\ & \dots & & & & \dots & \\ 0 & & \dots & -\pi_r & \dots & -\pi_1 & 1 \end{bmatrix},$$

$$\mathbf{U} = \begin{bmatrix} 1 & -\psi_1 & \dots & -\psi_s & 0 & & \dots & 0 \\ 0 & 1 & -\psi_1 & \dots & -\psi_s & 0 & \dots & 0 \\ \dots & & \dots & & & & \dots & \\ 0 & & \dots & & & 0 & \dots & 1 \end{bmatrix} \tag{2.37}$$

Substitute (2.37) into (2.36). We get

$$\mathbf{L}\mathbf{U}\boldsymbol{y} = \boldsymbol{\varepsilon},$$

such as

$$\boldsymbol{y} = \mathbf{U}^{-1}\mathbf{L}^{-1}\boldsymbol{\varepsilon}. \tag{2.38}$$

Given $\boldsymbol{\varepsilon}$, $\boldsymbol{y}$ can be obtained directly since $\mathbf{L}$ and $\mathbf{U}$ are positive definite triangular matrices. This *MAR*(r,s) simulating method can easily be generalized, for instance, for an *MAR*(r,s) involving some exogenous independent variables presented by Hecq, Issler and Telg (2020). In practice this vector-wise simulation

method is slower than the element-wise method because of the matrix creation
and storage in simulation.

Proof of Theorem 4

Proof.
First, we rewrite $\text{SRAR}(\tau, \hat{\boldsymbol{\theta}}(\tau))$ as follows:

$$
\begin{aligned}
\text{SRAR}(\tau, \hat{\boldsymbol{\theta}}(\tau)) &= \sum_{t=1}^{T} \rho_\tau(y_t - \boldsymbol{x}_t' \hat{\boldsymbol{\theta}}(\tau)) \\
&= \sum_{t=1}^{T} \rho_\tau(y_t - \boldsymbol{x}_t' \boldsymbol{\phi}_\tau + \boldsymbol{x}_t' \boldsymbol{\phi}_\tau - \boldsymbol{x}_t' \hat{\boldsymbol{\theta}}(\tau)) \qquad (2.39) \\
&= \sum_{t=1}^{T} \rho_\tau \left(u_{t\tau} - \frac{1}{a_T \sqrt{T}} \boldsymbol{v}' \boldsymbol{x}_t \right),
\end{aligned}
$$

where $\boldsymbol{x}_t' := [a_T, y_{t+1}, \ldots, y_{t+p}]$, $u_{t\tau} := y_t - \boldsymbol{x}_t' \boldsymbol{\phi}_\tau = \varepsilon_t - F^{-1}(\tau)$,
$\boldsymbol{v} := a_T \sqrt{T} \left(\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}_\tau \right)$. We know from Davis and Resnick (1985) and Knight
(1989, 1991) that

$$
\frac{1}{a_T} \left(\sum_{t=1}^{\lfloor T \cdot s \rfloor} (\varepsilon_t - b_T) \right) \overset{d}{\sim} \text{S}_\alpha(s),
$$

$$
\frac{1}{a_T \sqrt{T}} \sum_{t=1}^{T} \left(y_t - \lfloor T \cdot s \rfloor \sum_{j=0}^{\infty} c_j b_T \right) \overset{d}{\sim} \sum_{j=0}^{\infty} c_j \int_0^1 \text{S}_\alpha(s) \, ds, \qquad (2.40)
$$

$$
\frac{1}{a_T^2 T} \sum_{t=1}^{T} \left(y_t \cdot y_{t+h} - \lfloor T \cdot s \rfloor \sum_{j=0}^{\infty} c_j c_{j+h} b_T^2 \right) \overset{d}{\sim} \sum_{j=0}^{\infty} c_j c_{j+h} \int_0^1 \text{S}_\alpha^2(s) \, ds,
$$

where $t = \lfloor T \cdot s \rfloor$, and $\{\text{S}_\alpha(s)\}$ is a process of stable distributions with index α.
Without loss of generality, we assume $b_T = 0$ for the proof below. In use of the
limiting behaviour information presented in (2.40), we get that

$$
\frac{1}{a_T^2 T} \sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t' \overset{d}{\sim} \begin{bmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & \int_0^1 \text{S}_\alpha^2(s) \, ds \, \boldsymbol{\Omega} \end{bmatrix}_{(p+1) \times (p+1)} \qquad (2.41)
$$

where

$$
\boldsymbol{\Omega} := \left[\omega_{ik} \right]_{p \times p},
$$

$$
\omega_{ik} := \sum_{j=0}^{\infty} c_j c_{j+|k-i|},
$$

with ω_{ik} being the entry at $\mathbf{\Omega}$'s i-th row and k-th column. $\mathbf{\Omega}$ is positive definite symmetric. Note that $\hat{\boldsymbol{\theta}}(\tau) = \underset{\boldsymbol{\theta} \in \mathbb{R}^{p+1}}{\text{argmin}} \ \text{SRAR}(\tau, \boldsymbol{\theta})$ which also minimizes

$$Z_T(\mathbf{v}) := \sum_{t=1}^{T} \left[\rho_\tau \left(u_{t\tau} - \frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t \right) - \rho_\tau \left(u_{t\tau} \right) \right]. \tag{2.42}$$

$Z_T(\mathbf{v})$ is a convex random function. Knight (1989) showed that if $Z_T(\mathbf{v})$ converges in distribution to $Z(\mathbf{v})$ and $Z(\mathbf{v})$ has unique minimum, then the convexity of $Z_T(\mathbf{v})$ ensures $\hat{\mathbf{v}} = \underset{\mathbf{v} \in \mathbb{R}^{p+1}}{\text{argmin}} Z_T(\mathbf{v})$ converging in distribution to $\underset{\mathbf{v} \in \mathbb{R}^{p+1}}{\text{argmin}} Z(\mathbf{v})$.
By using the following check function identity:

$$\rho_\tau(v_1 - v_2) - \rho_\tau(v_1) = -v_2 \xi_\tau(v_1) + (v_1 - v_2) \left(\mathbb{1}_{\{0 > v_1 > v_2\}} - \mathbb{1}_{\{0 < v_1 < v_2\}} \right)$$

$$= -v_2 \xi_\tau(v_1) + \int_0^{v_2} \left(\mathbb{1}_{\{v_1 \leq s\}} - \mathbb{1}_{\{v_1 < 0\}} \right) ds, \tag{2.43}$$

where $\xi_\tau(v) := \tau - \mathbb{1}_{\{v < 0\}}$, we can rewrite $Z_T(\mathbf{v})$ into

$$Z_T(\mathbf{v}) = -\sum_{t=1}^{T} \frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t \, \xi_\tau(u_{t\tau}) + \sum_{t=1}^{T} \int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} \left(\mathbb{1}_{\{u_{t\tau} \leq s\}} - \mathbb{1}_{\{u_{t\tau} < 0\}} \right) ds$$

$$= Z_T^{(1)}(\mathbf{v}) + Z_T^{(2)}(\mathbf{v}), \tag{2.44}$$

where

$$\begin{cases} Z_T^{(2)}(\mathbf{v}) := \sum_{t=1}^{T} \int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} \left(\mathbb{1}_{\{u_{t\tau} \leq s\}} - \mathbb{1}_{\{u_{t\tau} < 0\}} \right) ds \\[2mm] Z_T^{(1)}(\mathbf{v}) := -\sum_{t=1}^{T} \frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t \, \xi_\tau(u_{t\tau}) \end{cases}$$

Further denote

$$\eta_t(\mathbf{v}) := \int_0^{\frac{1}{a_T \sqrt{T}} \mathbf{v}' \mathbf{x}_t} \left(\mathbb{1}_{\{u_{t\tau} \leq s\}} - \mathbb{1}_{\{u_{t\tau} < 0\}} \right) ds,$$

$$\bar{\eta}_t(\mathbf{v}) := E\left[\eta_t(\mathbf{v}) | \mathbf{x}_t \right],$$

$$\overline{Z}_T^{(2)}(\mathbf{v}) := \sum_{t=1}^{T} \bar{\eta}_t(\mathbf{v}).$$

By Assumption (A5) and small enough $\frac{1}{a_T\sqrt{T}}\mathbf{v}'\mathbf{x}_t$, we further rewrite $\overline{Z}_T^{(2)}(\mathbf{v})$ as follows:

$$
\begin{aligned}
\overline{Z}_T^{(2)}(\mathbf{v}) &= \sum_{t=1}^{T} E\left[\int_0^{\frac{1}{a_T\sqrt{T}}\mathbf{v}'\mathbf{x}_t}\left(\mathbb{1}_{\{u_{t\tau}\le s\}} - \mathbb{1}_{\{u_{t\tau}<0\}}\right)ds\,\middle|\,\mathbf{x}_t\right] \\
&= \sum_{t=1}^{T}\int_0^{\frac{1}{a_T\sqrt{T}}\mathbf{v}'\mathbf{x}_t}\left[\int_{F^{-1}(\tau)}^{s+F^{-1}(\tau)}f(r)dr\right]ds \\
&= \sum_{t=1}^{T}\int_0^{\frac{1}{a_T\sqrt{T}}\mathbf{v}'\mathbf{x}_t}\frac{F\left(s+F^{-1}(\tau)\right) - F\left(F^{-1}(\tau)\right)}{s}\,s\,ds \\
&= \sum_{t=1}^{T}\int_0^{\frac{1}{a_T\sqrt{T}}\mathbf{v}'\mathbf{x}_t}f\left(F^{-1}(\tau)\right)s\,ds \\
&= \frac{1}{2a_T^2 T}\sum_{t=1}^{T}f\left(F^{-1}(\tau)\right)\mathbf{v}'\mathbf{x}_t\mathbf{x}_t'\mathbf{v} + o_p(1) \\
&= \frac{1}{2a_T^2 T}f\left(F^{-1}(\tau)\right)\mathbf{v}'\left(\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)\mathbf{v} + o_p(1)
\end{aligned}
\tag{2.45}
$$

Using the limiting behaviour information presented in (2.40), we get the limiting distribution for $\overline{Z}_T^{(2)}(\mathbf{v})$ so as for $Z_T^{(2)}(\mathbf{v})$ as follows:

$$
Z_T^{(2)}(\mathbf{v}) \overset{d}{\sim} \frac{1}{2}f\left(F^{-1}(\tau)\right)\mathbf{v}'\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \int_0^1 S_\alpha^2(s)\,ds\,\mathbf{\Omega}\end{bmatrix}\mathbf{v},
\tag{2.46}
$$

by the fact that $Z_T^{(2)}(\mathbf{v}) - \overline{Z}_T^{(2)}(\mathbf{v}) \overset{p}{\sim} 0$ which can be proved by following the arguments of Knight (1989).

The limiting distribution of $Z_T^{(1)}(\mathbf{v})$ can also be deduced in using (2.40) as follows.

$$
-\sum_{t=1}^{T}\frac{1}{a_T\sqrt{T}}\mathbf{v}'\mathbf{x}_t\,\xi_\tau(u_{t\tau}) \overset{d}{\sim}
$$

$$
\mathbf{v}'\left[\sigma_\xi W(1), \sum_{j=0}^{\infty}c_j\sigma_\xi\int_0^1 S_\alpha(s)\,dW(s), \dots, \sum_{j=0}^{\infty}c_j\sigma_\xi\int_0^1 S_\alpha(s)\,dW(s)\right]_{(p+1)\times 1},
\tag{2.47}
$$

where $[\dots]_{(p+1)\times 1}$ is a column vector of $(p+1)$ elements, $\int dW(s)$ is a stochastic integral with Brownian motion $\{W(s)\}$ independent of $\{S_\alpha(s)\}$ (see Knight (1991)), and σ_ξ is the standard deviation of $\xi_\tau(u_{t\tau})$ which equals $\sqrt{\tau(1-\tau)}$. Therefore by Davis and Resnick (1985) and Knight (1989, 1991),

$$
Z_T^{(1)}(\mathbf{v}) \overset{d}{\sim} \mathbf{v}'\sqrt{\tau(1-\tau)}\left[W(1), \sum_{j=0}^{\infty}c_j\int_0^1 S_\alpha(s)\,dW(s), \dots, \sum_{j=0}^{\infty}c_j\int_0^1 S_\alpha(s)\,dW(s)\right]_{(p+1)\times 1}.
\tag{2.48}
$$

Thus,

$$Z_T(\boldsymbol{v}) \overset{d}{\sim} Z(\boldsymbol{v}) :=$$

$$\boldsymbol{v}' \sqrt{\tau(1-\tau)} \left[W(1), \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) \, dW(s), \dots, \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) \, dW(s) \right]_{(p+1) \times 1}$$

$$+ \frac{1}{2} f\left(F^{-1}(\tau) \right) \boldsymbol{v}' \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \int_0^1 S_\alpha^2(s) \, ds \, \boldsymbol{\Omega} \end{bmatrix} \boldsymbol{v}.$$

$$(2.49)$$

and so

$$\frac{f\left(F^{-1}(\tau) \right) \cdot a_T \sqrt{T}}{\sqrt{\tau(1-\tau)}} \left(\hat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\phi}_\tau \right) \overset{d}{\sim}$$

$$\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \left(\int_0^1 S_\alpha^2(s) \, ds \, \boldsymbol{\Omega} \right)^{-1} \end{bmatrix} \left[W(1), \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) \, dW(s), \dots, \sum_{j=0}^{\infty} c_j \int_0^1 S_\alpha(s) \, dW(s) \right]_{(p+1) \times 1}.$$

follows by setting the derivative of $Z(\boldsymbol{v})$ to 0 and solving for $\boldsymbol{v}$. $\qquad \square$

3

Adaptive Random Bandwidth for Inference in CAViaR Models

Abstract

Chapter 3 investigates the size performance of Wald tests for CAViaR models (Engle and Manganelli, 2004). We find that the usual estimation strategy on test statistics yields inaccuracies. Indeed, we show that existing density estimation methods cannot adapt to the time-variation in the conditional probability densities of CAViaR models. Consequently, we develop a method called *adaptive random bandwidth* which can approximate time-varying conditional probability densities robustly for inference testing on CAViaR models based on the asymptotic normality of the model parameter estimator. This proposed method also avoids the problem of choosing an optimal bandwidth in estimating probability densities, and can be extended to multivariate quantile regressions straightforward.

3.1 Introduction

Financial risk management is at the heart of banks' and financial institutions' activities to guide them in their investment plans, supervisory decisions, risk capital allocations and for external regulations. The use of quantitative risk measures has become essential in financial risk management. One of the most popular risk measures associated with financial portfolios is the value at risk (VaR hereafter). The VaR at probability $\tau \in (0, 1)$ of a portfolio is defined as the minimum potential loss that the portfolio may suffer in the worst τ portion of all possible outcomes over a given time horizon. VaR is very intuitive (Duffie and Pan, 1997) and has for instance been incorporated into *the 1996 Amendment to the Capital Accord* for measuring the market risk in financial positions of each financial institution. Therefore, VaR is still a widely used risk measure even though many approaches to measuring market and credit risks have been proposed in the literature.

Generally, there are three ways to estimate VaR: *(i)* historical simulations, *(ii)* semi-parametric approaches and *(iii)* fully parametric frameworks. Within the class of semi-parametric approaches, it typically includes extreme value theory analyses and quantile regression techniques. In this chapter, we focus on quantile regressions for the VaR estimation as quantile regressions are straightforward in studying one quantile of interest and numerically efficient without imposing parametric distributional assumptions.

Despite that the VaR is just a particular quantile of future portfolio losses conditional on present information, it is essentially a part of the underlying conditional distribution. VaR models are supposed to embrace features of the empirical conditional distributions of returns, such as time-variation and conditional heteroskedasticity. Drawing on (G)ARCH specifications which capture the presence of time-varying conditional heteroskedasticity in time series, Engle and Manganelli (2004) have proposed to estimate conditional autoregressive value at risk by regression quantiles (CAViaR). It is appealing to consider CAViaR models for estimating VaR as CAViaR models associate the conditional quantile of interest with observable variables as well as the implicit information on lagged conditional quantiles.

This chapter carefully investigates the size performance of Wald tests for CAViaR models. Having an accurate test statistic is important to obtain reliable models in financial applications. Several specifications are nested within a CAViaR specification, such as static quantile regressive models and quantile autoregressive models (see Koenker and Xiao, 2006; Hecq and Sun, 2020). Moreover, there exists several models nested within the general CAViaR specification that have been proposed in the literature. For instance, asymmetric slope CAViaR models (Engle and Manganelli, 2004) that split the effect of

positive and negative yesterday's news shocks. Wald tests are used to test the null of a symmetric news impact. However, we find that the usual estimation strategy yields inaccuracies. Indeed, we show that existing density estimation methods cannot adapt to the time-variation in the conditional probability densities of CAViaR models. The method that we develop in this chapter is able to adapt to time-varying conditional probability densities and produces much more reliable results than the existing ones for inference testing on CAViaR models based on the asymptotic normality of the model parameter estimator. This proposed method also avoids the haunting problem of choosing an optimal bandwidth in estimating probability densities, and can be extended to multivariate quantile regressions straightforward in theory.

The remainder of this chapter is structured as follows. In Section 3.2, stability conditions for CAViaR data generating processes (DGPs) to be non-explosive are derived. In Section 3.3, we investigate the size performance of Wald tests for CAViaR models and find large size distortions by the usual estimation strategy. So we introduce a method called *adaptive random bandwidth*. An empirical study on stock returns is performed in Section 3.4. Finally Section 3.5 concludes this chapter.

3.2 CAViaR Models

Let us consider a stationary time series process $\{y_t\}_{t=1}^T$ for instance the return of an asset or a portfolio, and denote $\boldsymbol{x}_t$ a vector of observable variables at time t and $\mathscr{F}_t$ the information set up to time t which is the σ-algebra generated by $\{\boldsymbol{x}_t, y_t, \boldsymbol{x}_{t-1}, y_{t-1}, \ldots\}$. The τ-th quantile ($\tau \in (0,1)$) or the opposite VaR$_\tau$ of y_t conditional on $\mathscr{F}_{t-1}$ is denoted as $f_t(\boldsymbol{\beta}_\tau, \boldsymbol{x}_{t-1})$ (or simply $f_t(\boldsymbol{\beta}_\tau)$ when $\boldsymbol{x}_{t-1}$ is taken in obviously). A generic CAViaR specification proposed by Engle and Manganelli (2004) is

$$f_t(\boldsymbol{\beta}_\tau) = \beta_0 + \sum_{i=1}^q \beta_i f_{t-i}(\boldsymbol{\beta}_\tau) + \sum_{j=1}^r \beta_{q+j} l(\boldsymbol{x}_{t-j}), \qquad (3.1)$$

where $\boldsymbol{\beta}'_\tau := [\beta_0, \beta_1, \ldots, \beta_p]$ collects the $p = q + r$ slope parameters, and l is a function of a finite number of lagged observable variables, for instance the lagged returns entering potentially with different weights for positive and negative past lagged returns. As described in Engle and Manganelli (2004) the autoregressive terms $\beta_i f_{t-i}(\boldsymbol{\beta}_\tau)$ can ensure that the quantile changes smoothly over time. The quantile autoregressive model (QAR) of Koenker and Xiao (2006) is nested in the CAViaR specification by restricting $\beta_1 = \ldots = \beta_q = 0$ in CAViaR. The role of $l(\boldsymbol{x}_{t-j})$ is to account for the association of $f_t(\boldsymbol{\beta}_\tau)$ with observable variables in $\mathscr{F}_{t-1}$. CAViaR models as a generalization of QAR models are able to capture the

time-variation in the conditional quantile in a way similar to GARCH models in explaining time-varying volatility and volatility clustering in financial time series in addition to ARCH models.

The CAViaR model (3.1) is nonlinear in parameters as long as there exists a nonzero $\beta_i, i \in \{1, \ldots, q\}$ which leads to $\frac{\partial f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_i} = f_{t-i}(\boldsymbol{\beta}_\tau) + \beta_i \frac{\partial f_{t-i}(\boldsymbol{\beta}_\tau)}{\partial \beta_i}$ not independent of β_i.[1] The algorithm to estimate CAViaR models is given in Section 3.2.2.

For illustration, we simulate samples from the following three CAViaR DGPs in (3.2) and plot Figure 3.1 (a). [2] In Figure 3.1 (a), we see a decreasing trend in CAViaR DGP 1.a mainly due to the negative term $-0.5|y_{t-1}|$ in $f_t(\boldsymbol{\beta}_\tau)$ compared with CAViaR DGP 1.b. Comparing CAViaR DGP 1.b with 1.c, we find that CAViaR DGP 1.b has a larger spread due to a higher slope of $f_{t-1}(\tau)$ in $f_t(\boldsymbol{\beta}_\tau)$. A similar finding further applies on Figure 3.1 (b) which plots simulated samples of CAViaR DGP 2.a, 2.b and 2.c in (3.3) respectively.

$$
\begin{cases}
\text{CAViaR DGP 1.a:} & f_t(\boldsymbol{\beta}_{u_t}) = F_{t(3)}^{-1}(u_t) + 0.5 f_{t-1}(\boldsymbol{\beta}_{u_t}) - 0.5|y_{t-1}|, \\
\text{CAViaR DGP 1.b:} & f_t(\boldsymbol{\beta}_{u_t}) = F_{t(3)}^{-1}(u_t) + 0.5 f_{t-1}(\boldsymbol{\beta}_{u_t}) - 0.5 y_{t-1}, \\
\text{CAViaR DGP 1.c:} & f_t(\boldsymbol{\beta}_{u_t}) = F_{t(3)}^{-1}(u_t) - 0.5 y_{t-1},
\end{cases} \quad (3.2)
$$

where $\{u_t\}$ is i.i.d. in the standard uniform distribution (denoted as $\mathcal{U}(0,1)$) and $F_{t(3)}^{-1}(\cdot)$ is the inverse function of Student's t-distribution with 3 degrees of freedom ($t(3)$ hereafter).

$$
\begin{cases}
\text{CAViaR DGP 2.a:} & f_t(\boldsymbol{\beta}_{u_t}) = F_{t(3)}^{-1}(u_t) - 0.5 f_{t-1}(\boldsymbol{\beta}_{u_t}) + 0.5|y_{t-1}| \\
\text{CAViaR DGP 2.b:} & f_t(\boldsymbol{\beta}_{u_t}) = F_{t(3)}^{-1}(\tau) - 0.5 f_{t-1}(\boldsymbol{\beta}_{u_t}) + 0.5 y_{t-1} \\
\text{CAViaR DGP 2.c:} & f_t(\boldsymbol{\beta}_{u_t}) = F_{t(3)}^{-1}(u_t) + 0.5 y_{t-1},
\end{cases} \quad (3.3)
$$

where $u_t \overset{i.i.d.}{\sim} \mathcal{U}(0,1)$, $t = 1, 2, \ldots, T$.

3.2.1 Stability conditions for CAViaR models

The stationarity of CAViaR time series is required for the model estimation consistency (Engle and Manganelli, 2004). After simulating a CAViaR DGP, we can

---

[1] In Appendix 3.A, the gradient and the Hessian matrix of CAViaR models are illustrated to emphasize that the nonlinearity of model parameters makes CAViaR models different from other linear quantile regression models.

[2] All the simulations of CAViaR DGPs in this chapter follow the procedure given in Appendix 3.B.
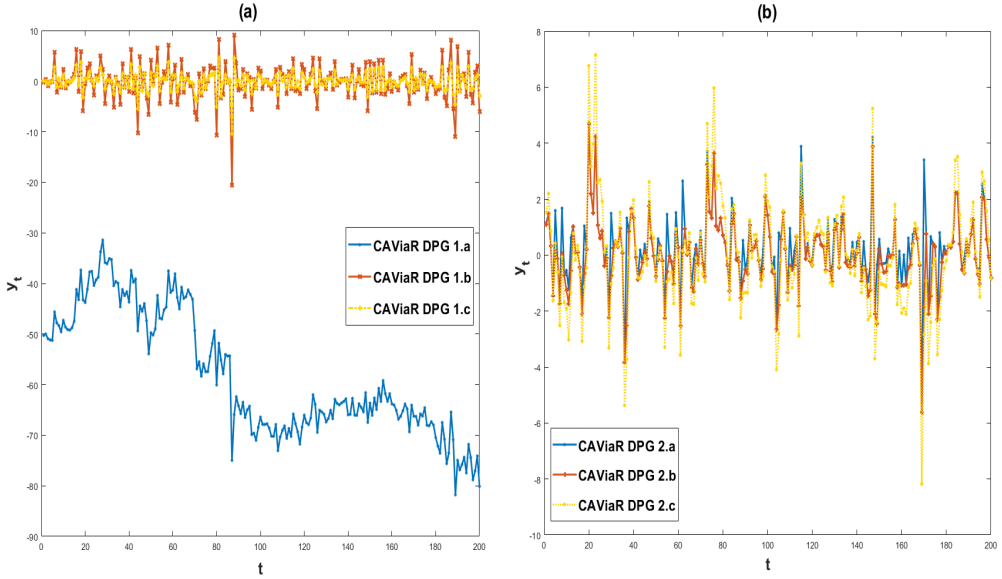
Figure 3.1: Time series plots of CAViaR DGP samples

view its behaviour such as explosiveness in the long run. We know that a time series is explosive if and only if at least one conditional quantile of the time series with nonzero probability density to occur is explosive. So we derive stability conditions for the conditional τ-th ($\tau \in (0,1)$) quantile of a CAViaR DGP $\{y_t\}$ specified as follows:

$$y_t = f_t(\boldsymbol{\beta}_{u_t}) = \beta_0(u_t) + \sum_{i=1}^{q} \beta_i(u_t) f_{t-i}(\boldsymbol{\beta}_{u_t}) + \sum_{j=1}^{r} \beta_{q+j}(u_t) y_{t-j}, \qquad (3.4)$$

where $u_t \overset{i.i.d.}{\sim} \mathscr{U}(0,1)$, and $\boldsymbol{\beta}'_{u_t} := [\beta_0(u_t), \beta_1(u_t), \ldots, \beta_p(u_t)]$ with $p = q + r$. There is a monotonicity requirement on this model which is that $f_t(\boldsymbol{\beta}_{u_t})$ is monotonically increasing in u_t so that the τ-th quantile ($\tau \in (0,1)$) of y_t conditional on $\mathscr{F}_{t-1}$ can be expressed as $f_t(\boldsymbol{\beta}_\tau)$.

Assume the conditional τ-th quantile of $\{y_t\}$ follows the model (3.4) with nonzero probability density to occur at each time. Without loss of generality, there is a time $t \in \{1, \ldots, T\}$ such that

$$y_t = f_t(\boldsymbol{\beta}_\tau) = \beta_0 + \sum_{i=1}^{q} \beta_i f_{t-i}(\boldsymbol{\beta}_\tau) + \sum_{j=1}^{r} \beta_{q+j} y_{t-j}.$$

Now let us derive the value of y_t. First we have the following equation

from (3.4).

$$\left(1 - \sum_{j=1}^{r} \beta_{q+j} L^j\right) y_t = \beta_0 + \sum_{i=1}^{q} \beta_i f_{t-i}(\boldsymbol{\beta}_\tau)$$

$$= \beta_0 + \sum_{i=1}^{q} \beta_i \left(\beta_0 + \sum_{i_1=1}^{q} \beta_{i_1} f_{t-i-i_1}(\boldsymbol{\beta}_\tau) + \sum_{j=1}^{r} \beta_{q+j} y_{t-i-j} \right)$$

$$= \beta_0 + \sum_{i=1}^{q} \beta_i \left(\beta_0 + \sum_{i_1=1}^{q} \beta_{i_1} f_{t-i-i_1}(\boldsymbol{\beta}_\tau) + \sum_{j=1}^{r} \beta_{q+j} L^{i+j} y_t \right)$$

$$= \left(1 + \sum_{i=1}^{q} \beta_i\right) \beta_0 + \sum_{i=1}^{q} \sum_{i_1=1}^{q} \beta_i \beta_{i_1} f_{t-i-i_1}(\boldsymbol{\beta}_\tau) + \sum_{i=1}^{q} \sum_{j=1}^{r} \beta_i \beta_{q+j} L^{i+j} y_t,$$

where the second line is obtained by substituting the specification (3.4) of $f_{t-i-i_1}(\boldsymbol{\beta}_\tau)$ into the first line, and L is the lag operator. Further rewrite the above equation, and we have

$$\left(1 - \sum_{j=1}^{r} \beta_j L^j - \sum_{i=1}^{q} \sum_{j=1}^{r} \beta_i \beta_{q+j} L^{i+j}\right) y_t = \left(1 + \sum_{i=1}^{q} \beta_i\right) \beta_0 + \sum_{i=1}^{q} \sum_{i_1=1}^{q} \beta_i \beta_{i_1} f_{t-i-i_1}(\boldsymbol{\beta}_\tau)$$

$$= \left(1 + \sum_{i=1}^{q} \beta_i\right) \beta_0 + \sum_{i=1}^{q} \sum_{i_1=1}^{q} \beta_i \beta_{i_1} f_{t-i-i_1}(\boldsymbol{\beta}_\tau).$$

We continue to rewrite the lagged terms of $f_t(\boldsymbol{\beta}_\tau)$ on the right-hand side of the above equation, and then organize the equation such that only the left-hand side contains terms of y_t. Therefore, we obtain that

$$\left(1 - \sum_{j=1}^{r} \beta_{q+j} L^j - \sum_{i=1}^{q} \sum_{j=1}^{r} \beta_i \beta_{q+j} L^{i+j} - \ldots - \sum_{i=1}^{q} \sum_{i_1=1}^{q} \cdots \sum_{i_n=1}^{q} \sum_{j=1}^{r} \beta_i \beta_{i_1} \ldots \beta_{i_n} \beta_{q+j} L^{i+i_1\ldots+i_n+j}\right) y_t$$

$$= \left(1 + \sum_{i=1}^{q} \beta_i + \left(\sum_{i=1}^{q} \beta_i\right)^2 + \ldots + \left(\sum_{i=1}^{q} \beta_i\right)^n\right) \beta_0$$

$$+ \sum_{i=1}^{q} \sum_{i_1=1}^{q} \cdots \sum_{i_{n+1}=1}^{q} \sum_{j=1}^{r} \beta_i \beta_{i_1} \ldots \beta_{i_{n+1}} f_{t-i-i_1-\ldots-i_{n+1}}(\boldsymbol{\beta}_\tau).$$

$$(3.5)$$

Now we can get the first necessary condition for $\{y_t\}$ to be nonexplosive, which is

$$\left| \sum_{i=1}^{q} \beta_i \right| < 1. \tag{3.6}$$

Under the condition (3.6), we can simplify the equation (3.5) when letting $n \to \infty$ as follows:

$$\left(1 - \sum_{j=1}^{r} \beta_{q+j} L^j \sum_{m=0}^{\infty} \left(\sum_{i=1}^{q} \beta_i L^i\right)^m\right) y_t = \frac{1}{1 - \sum_{i=1}^{q} \beta_i} \beta_0.$$

Now we obtain the autoregressive polynomial $g(x)$ of y_t which is

$$g(x) := 1 - \sum_{j=1}^{r} \beta_{q+j} x^j \sum_{m=0}^{\infty} \left(\sum_{i=1}^{q} \beta_i x^i \right)^m . \tag{3.7}$$

So the second necessary condition for $\{y_t\}$ to be nonexplosive is that the roots of $g(x)$ are outside the unit circle. When there exists at least one $\beta_i \neq 0, i \in \{1, \ldots, q\}$, this second condition is equivalent to require that the roots of $g_1(x) := 1 - \sum_{i=1}^{q} \beta_i x^i - \sum_{j=1}^{r} \beta_{q+j} x^j$ and the common roots of $g_2(x) := 1 - \sum_{i=1}^{q} \beta_i x^i$ and $g_3(x) := 1 - \sum_{j=1}^{r} \beta_{q+j} x^j$ all are outside the unit circle. More examples of CAViaR DGPs are illustrated in Appendix 3.D, among which we can find explosive DGPs which break the condition on the roots of $g_1(x)$ but meet the condition on the common roots of $g_2(x)$ and $g_3(x)$.

We can review Figure 3.1 with the above stability conditions. CAViaR DGP 1.b, 1.c, 2.b and 2.c meet the above conditions and we also see their nonexplosive behaviours in the plots. The nonexplosiveness of CAViaR DGP 2.a can also be ensured since it has a narrower spread in theory in comparison with CAViaR DGP 2.b. On the other hand, we know that CAViaR DGP 1.a has a downward trend due to the negative term $-0.5|y_{t-1}|$ and hence is explosive.

3.2.2 Estimation algorithm

The estimation for CAViaR models can be achieved by the differential evolutionary genetic algorithm (Storn and Price, 1997) used by Engle and Manganelli (2004). Suppose the model $f_t(\boldsymbol{\beta})$ is specified as (3.1) for data $\{y_t\}_{t=1}^{T}$. We want to obtain the parameter estimator $\widehat{\boldsymbol{\beta}}$ by the following optimization:

$$\begin{cases} \widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\arg\min} \, S_T(\boldsymbol{\beta}) \\ S_T(\boldsymbol{\beta}) := \sum_{t=1}^{T} \rho_\tau (y_t - f_t(\boldsymbol{\beta})) \end{cases} \tag{3.8}$$

where $S_T(\boldsymbol{\beta})$ is the objective function in quantile regressions, and $\rho_\tau(x) := x \left(\tau - \mathbb{1}_{\{x<0\}} \right)$ is called check function (Koenker, 2005) with the indicator function $\mathbb{1}_{\{\cdot\}}$.

Following the steps below, we can obtain $\widehat{\boldsymbol{\beta}}$ in (3.8).

Step 1: Generate n (say 10^4) trial vectors independently from a uniform distribution $\mathscr{U}(\boldsymbol{b}_L, \boldsymbol{b}_p)$ as n parameter initial trials, where $\boldsymbol{b}_L$ and $\boldsymbol{b}_p$ are $(p+1) \times 1$

vectors roughly covering the lower and upper bounds of the true parameter vector $\boldsymbol{\beta}_\tau^o$ of the underlying process in our belief. It is worth mentioning that the values of $\{f_{1-i}(\boldsymbol{\beta}_\tau), i = 1, \ldots, q\}$ and $\{y_{1-j}, j = 1, \ldots, r\}$ acting as initial conditions are also input-demanded in order to calculate $\{f_t(\boldsymbol{\beta})\}_{t=1}^T$ for any $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$. For instance, as used by Engle and Manganelli (2004) $f_0(\boldsymbol{\beta}_\tau)$ is given as the estimated τ-th quantile of $\{y_t\}_{t=1}^{\lfloor 0.1T \rfloor}$ and is fixed in the optimization.[3]

Step 2: Each parameter initial is used to kick off a minimization routine[4] on the objective function $S_T(\boldsymbol{\beta})$, and the returned value of $\widehat{\boldsymbol{\beta}}$ from the routine and its objective function value are stored.

Step 3: Select m (say 10) returned vectors of $\widehat{\boldsymbol{\beta}}$ which result in the lowest m values among the n stored objective function values.

Step 4: Denote the m selected vectors as $\widehat{\boldsymbol{\beta}}^{(1)}, \ldots, \widehat{\boldsymbol{\beta}}^{(m)}$ and use them as initials to restart the minimization routine individually, and update $\widehat{\boldsymbol{\beta}}^{(1)}, \ldots, \widehat{\boldsymbol{\beta}}^{(m)}$ with the newly returned vectors respectively.

Step 5: Repeat Step 4 a (say 5) times.

Step 6: Calculate $S_T(\widehat{\boldsymbol{\beta}}^{(i)}), i = 1, \ldots, m$. And set the solution to be $\widehat{\boldsymbol{\beta}} = \underset{i=1,\ldots,m}{\operatorname{argmin}} S_T(\widehat{\boldsymbol{\beta}}^{(i)})$.

We implement the above estimation algorithm throughout this chapter for CAViaR model parameter estimations. There might be a concern if the artificial input of the initial values $\{f_{1-i}(\boldsymbol{\beta}_\tau), i = 1, \ldots, q\}$ and $\{y_{1-j}, j = 1, \ldots, r\}$ affects the parameter estimator. In fact, the effect usually is small and can be neglected when the sample size is large enough because the fitted conditional quantiles $\{f_t(\widehat{\boldsymbol{\beta}})\}_{t=1}^T$ are kept close to the true ones $\{f_t(\boldsymbol{\beta})\}_{t=1}^T$ such that it can minimize the objective function despite some burn-in period.

3.3 Adaptive Random Bandwidth Method for CAViaR Covariance Matrix Estimation

Consistency and asymptotic normality of CAViaR model parameters have been proved by Engle and Manganelli (2004). After regressing data onto a CAViaR model, we would like to implement an inference testing on whether the model is correctly specified. In this section we first investigate how we result in the

---

[3] $\lfloor \cdot \rfloor$ is known as the floor function (or the greatest integer function) and $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{Z}$ of a real number x denotes the greatest integer less than or equal to x.

[4] The Nelder Mead simplex algorithm is used in our minimization routine.

asymptotic normality of CAViaR model parameter estimators. We focus on the elements of the asymptotic covariance matrix to highlight their roles in connecting sample elements with the corresponding limit behaviours. Next, we check whether existing estimation strategies can perform robustly and satisfactorily for Wald tests on CAViaR models. Finally, we propose a new method called *adaptive random bandwidth* for CAViaR models.

3.3.1 Asymptotics of CAViaR

Consider a time series $\{y_t\}$ of random variables y_t on a complete prbability space $(\Omega, \mathscr{F}, P)$ [5]. For applying a generic CAViaR model (3.1) on $\{y_t\}$, the consistency and asymptotic normality of the estimator $\widehat{\boldsymbol{\beta}} := \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_\tau (y_t - f_t(\boldsymbol{\beta}))$ has been derived out by Engle and Manganelli (2004):

Theorem 5 (Asymptotics given by Engle and Manganelli (2004))**.**
For a data generating process $\{y_t\}$ with its time t conditional τ-th quantile following a generic CAViaR model as (3.1) *parametrized by $\boldsymbol{\beta}^o$, it satisfies the regularity conditions (C0,..., C7, AN1,..., AN7) in the proof of Engle and Manganelli (2004). Then*

$$\sqrt{T} A_T^{-1/2} D_T \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o \right) \stackrel{d}{\sim} N(\mathbf{0}, \boldsymbol{I}_{(p+1)\times(p+1)}), \tag{3.9}$$

where

$$\widehat{\boldsymbol{\beta}} := \underset{\boldsymbol{\beta} \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{t=1}^{T} \rho_\tau (y_t - f_t(\boldsymbol{\beta})),$$

$$A_T := \mathbb{E}\left[T^{-1} \tau(1-\tau) \sum_{t=1}^{T} \nabla' f_t(\boldsymbol{\beta}^o) \nabla f_t(\boldsymbol{\beta}^o) \right],$$

$$D_T := \mathbb{E}\left[T^{-1} \sum_{t=1}^{T} h_t\left(0|\mathscr{F}_{t-1}\right) \nabla' f_t(\boldsymbol{\beta}^o) \nabla f_t(\boldsymbol{\beta}^o) \right], \tag{3.10}$$

$$\varepsilon_{\tau t} := y_t - f_t(\boldsymbol{\beta}^o),$$

and $h_t(0|\mathscr{F}_{t-1})$ is denoted as the probability density of $\varepsilon_{\tau t}$ evaluated at 0 conditional on the information set $\mathscr{F}_{t-1}$. $\boldsymbol{I}_{(p+1)\times(p+1)}$ is the $(p+1) \times (p+1)$ identity matrix.

The above theorem is useful for quantile model (mis)specification tests. For instance, Wald tests can be used to check whether the current model is correctly

---

[5]See the assumption C0 of Engle and Manganelli (2004). We also apply this assumption throughout this chapter. That is to say, all the random variables considered in this chapter are assumed on a complete prbability space $(\Omega, \mathscr{F}, P)$.

specified by testing the validity of a more parsimonious nested model. To perform such a quantile model specification test, it often requires to estimate A_T, $h_t(0|\mathscr{F}_{t-1})$ and D_T. When using traditional estimates $\widehat{A}_T$, $\{\widehat{h}_t(0|\mathscr{F}_{t-1})\}$, $\widehat{D}_T$ of A_T, $\{h_t(0|\mathscr{F}_{t-1})\}$ and D_T respectively, we found considerable size distortions in inference tests on CAViaR models in general. We will show that the reason lies in the inaccuracy of $\{\widehat{h}_t(0|\mathscr{F}_{t-1})\}$ in the next subsection. In order to spot the discrepancy in approximating $\{h_t(0|\mathscr{F}_{t-1})\}$, we need a clear picture on how $\{h_t(0|\mathscr{F}_{t-1})\}$ comes up into the asymptotic normality of the model parameter estimator. Doing so, we can see the role of $\{h_t(0|\mathscr{F}_{t-1})\}$ and whether a sequence $\left\{\widehat{h}_t(0|\mathscr{F}_{t-1})\right\}$ is capable to achieve the same role in practice. Let us review the proof of Engle and Manganelli (2004) for Theorem 5 below.

The proof of Engle and Manganelli (2004) is obtained by applying Theorem 3 of Huber et al. (1967) onto $T^{-1/2}\sum_{t=1}^{T}\left(\mathbb{1}_{\left\{y_t\leq f_t(\widehat{\boldsymbol{\beta}})\right\}}-\tau\right)\nabla'f_t(\widehat{\boldsymbol{\beta}})$ and the central limit theorem onto $T^{-1/2}\sum_{t=1}^{T}\left(\mathbb{1}_{\left\{y_t\leq f_t(\boldsymbol{\beta}^o)\right\}}-\tau\right)\nabla'f_t(\boldsymbol{\beta}^o)$. Huber's conditions are verified in the proof before applying Huber's theorem. Denote

$$\begin{aligned}\text{Hit}_t(\boldsymbol{\beta})&:=\mathbb{1}_{\{y_t\leq f_t(\boldsymbol{\beta})\}}-\tau,\\g_t(\boldsymbol{\beta})&:=\nabla f_t(\boldsymbol{\beta}).\end{aligned}\tag{3.11}$$

$\text{Hit}_t(\boldsymbol{\beta})$ gives value $-\tau$ every time y_t exceeds $f_t(\boldsymbol{\beta})$ and $1-\tau$ otherwise. With the true underlying parameter $\boldsymbol{\beta}^o$, $\{\text{Hit}_t(\boldsymbol{\beta}^o)\}$ is a martingale difference sequence with respect to $\{\mathscr{F}_{t-1}\}$. It is easy to get that $T^{-1/2}\sum_{t=1}^{T}\text{Hit}_t(\boldsymbol{\beta}^o)g_t(\boldsymbol{\beta}^o)$ follows the central limit theorem because $\left\{\text{Hit}_t(\boldsymbol{\beta}^o)g_t(\boldsymbol{\beta}^o)\right\}$ is a martingale difference sequence with the assumption AN1 of Engle and Manganelli (2004) on its uniformly bounded second moment. So we get that

$$T^{-1/2}\sum_{t=1}^{T}\text{Hit}_t(\boldsymbol{\beta}^o)g_t(\boldsymbol{\beta}^o)\overset{d}{\sim}N(0,A_T).\tag{3.12}$$

It has also been proved by Engle and Manganelli (2004) that

$$T^{-1/2}\sum_{t=1}^{T}\text{Hit}_t(\widehat{\boldsymbol{\beta}})g_t(\widehat{\boldsymbol{\beta}})=o_p(1).\tag{3.13}$$

Next, we are going to manifest $\{h_t(0|\mathscr{F}_{t-1})\}$ in the proof in a way which makes the appearance of $\{h_t(0|\mathscr{F}_{t-1})\}$ more intuitive. We rewrite $\text{Hit}_t(\widehat{\boldsymbol{\beta}})\,g_t(\widehat{\boldsymbol{\beta}})$ as fol-

lows:

$$\text{Hit}_t(\widehat{\boldsymbol{\beta}}) g_t(\widehat{\boldsymbol{\beta}}) = \left(\text{Hit}_t(\boldsymbol{\beta}^o) + \text{Hit}_t(\widehat{\boldsymbol{\beta}}) - \text{Hit}_t(\boldsymbol{\beta}^o) \right) g_t(\widehat{\boldsymbol{\beta}})$$
$$= \text{Hit}_t(\boldsymbol{\beta}^o) g_t(\widehat{\boldsymbol{\beta}}) + \left(\text{Hit}_t(\widehat{\boldsymbol{\beta}}) - \text{Hit}_t(\boldsymbol{\beta}^o) \right) g_t(\widehat{\boldsymbol{\beta}}). \tag{3.14}$$

Take expectation on the both sides of Equation (3.14) and get

$$T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[\text{Hit}_t(\widehat{\boldsymbol{\beta}}) g_t(\widehat{\boldsymbol{\beta}}) \right]$$

$$= T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[\text{Hit}_t(\boldsymbol{\beta}^o) g_t(\widehat{\boldsymbol{\beta}}) + \left(\text{Hit}_t(\widehat{\boldsymbol{\beta}}) - \text{Hit}_t(\boldsymbol{\beta}^o) \right) g_t(\widehat{\boldsymbol{\beta}}) \right]$$

$$= T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[\mathbb{E} \left[\text{Hit}_t(\boldsymbol{\beta}^o) | \mathscr{F}_{t-1} \right] g_t(\widehat{\boldsymbol{\beta}}) \right] + T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[\left(\text{Hit}_t(\widehat{\boldsymbol{\beta}}) - \text{Hit}_t(\boldsymbol{\beta}^o) \right) g_t(\widehat{\boldsymbol{\beta}}) \right]$$

$$= T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[\mathbb{E} \left[\left(\text{Hit}_t(\widehat{\boldsymbol{\beta}}) - \text{Hit}_t(\boldsymbol{\beta}^o) \right) \Big| \mathscr{F}_{t-1} \right] g_t(\boldsymbol{\beta}^o) \right]$$

$$+ T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[\mathbb{E} \left[\left(\text{Hit}_t(\widehat{\boldsymbol{\beta}}) - \text{Hit}_t(\boldsymbol{\beta}^o) \right) \Big| \mathscr{F}_{t-1} \right] \left(g_t(\widehat{\boldsymbol{\beta}}) - g_t(\boldsymbol{\beta}^o) \right) \right]$$

$$= T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[\mathbb{E} \left[\left(\text{Hit}_t(\widehat{\boldsymbol{\beta}}) - \text{Hit}_t(\boldsymbol{\beta}^o) \right) \Big| \mathscr{F}_{t-1} \right] g_t(\boldsymbol{\beta}^o) \right]$$

$$+ T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[\mathbb{E} \left[\left(\text{Hit}_t(\widehat{\boldsymbol{\beta}}) - \text{Hit}_t(\boldsymbol{\beta}^o) \right) \Big| \mathscr{F}_{t-1} \right] \right] \mathscr{O}_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_\infty), \tag{3.15}$$

where $\|\cdot\|_\infty$ is the supremum norm of vectors. And

$$\mathbb{E} \left[\left(\text{Hit}_t(\widehat{\boldsymbol{\beta}}) - \text{Hit}_t(\boldsymbol{\beta}^o) \right) \Big| \mathscr{F}_{t-1} \right] = P \left\{ y_t \leq f_t(\widehat{\boldsymbol{\beta}}) \Big| \mathscr{F}_{t-1} \right\} - P \left\{ y_t \leq f_t(\boldsymbol{\beta}^o) | \mathscr{F}_{t-1} \right\}$$

$$= F_t \left(f_t(\widehat{\boldsymbol{\beta}}) \Big| \mathscr{F}_{t-1} \right) - F_t \left(f_t(\boldsymbol{\beta}^o) | \mathscr{F}_{t-1} \right)$$

$$= F_t' \left(f_t(\boldsymbol{\beta}^o) | \mathscr{F}_{t-1} \right) \left(f_t(\widehat{\boldsymbol{\beta}}) - f_t(\boldsymbol{\beta}^o) \right) + \mathscr{O}_p \left(f_t(\widehat{\boldsymbol{\beta}}) - f_t(\boldsymbol{\beta}^o) \right)^2$$

$$= h_t(0 | \mathscr{F}_{t-1}) \left(\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o \right)' \nabla f_t(\boldsymbol{\beta}^o) + \mathscr{O}_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_\infty^2) \right) + \mathscr{O}_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_\infty^2)$$

$$= \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o \right)' h_t(0 | \mathscr{F}_{t-1}) \nabla f_t(\boldsymbol{\beta}^o) + \mathscr{O}_p(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_\infty^2), \tag{3.16}$$

where $F_t(\cdot | \mathscr{F}_{t-1})$ is the probability density function of y_t conditional on $\mathscr{F}_{t-1}$, and $h_t(0 | \mathscr{F}_{t-1}) = F_t' \left(f_t(\boldsymbol{\beta}^o) | \mathscr{F}_{t-1} \right)$. Substituting (3.16) into (3.15) gives

$$T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[\text{Hit}_t(\widehat{\boldsymbol{\beta}}) g_t(\widehat{\boldsymbol{\beta}}) \right]$$

$$= \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o \right)' \cdot T^{-1/2} \sum_{t=1}^{T} \mathbb{E} \left[h_t(0 | \mathscr{F}_{t-1}) \nabla f_t(\boldsymbol{\beta}^o) \nabla' f_t(\boldsymbol{\beta}^o) \right] \tag{3.17}$$

$$+ \mathscr{O}_p(T^{1/2} \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_\infty^2).$$

Success in applying Huber's theorem gives

$$T^{-1/2} \sum_{t=1}^{T} \mathbb{E}\left[\text{Hit}_t(\widehat{\boldsymbol{\beta}}) g_t(\widehat{\boldsymbol{\beta}}) \right] = -T^{-1/2} \sum_{t=1}^{T} \left(\mathbb{1}_{\{y_t \le f_t(\boldsymbol{\beta}^o)\}} - \tau \right) \nabla' f_t(\boldsymbol{\beta}^o) + o_p(1) \tag{3.18}$$

Therefore, the asymptotic normality of $T^{1/2}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\right)$ is obtained by substituting (3.12) and (3.17) into (3.18).

From the above derivation, it is clear that the role of $h_t(0|\mathscr{F}_{t-1})$ is actually an approximation to $F_t'\left(f_t(\bar{\boldsymbol{\beta}})\Big|\mathscr{F}_{t-1}\right)$ in which $\bar{\boldsymbol{\beta}}$ is between $\boldsymbol{\beta}^o$ and $\widehat{\boldsymbol{\beta}}$. This role comes to the surface of (3.16) using the fact that

$$F_t\left(f_t(\widehat{\boldsymbol{\beta}})\Big|\mathscr{F}_{t-1}\right) - F_t\left(f_t(\boldsymbol{\beta}^o)|\mathscr{F}_{t-1}\right) = F_t'\left(f_t(\bar{\boldsymbol{\beta}})\Big|\mathscr{F}_{t-1}\right)\left(\nabla' f_t(\boldsymbol{\beta}^o)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\right)\right) \tag{3.19}$$

by the Mean Value Theorem. This approximating role of $h_t(0|\mathscr{F}_{t-1})$ sets a clear mission of any $\widehat{h}_t(0|\mathscr{F}_{t-1})$ supposed to achieve, which can be used to examine an estimator for $h_t(0|\mathscr{F}_{t-1})$ as well as to propose an improved estimation method. In next subsection, we are going to examine the performances of some existing methods for estimating $h_t(0|\mathscr{F}_{t-1})$ and the role of $h_t(0|\mathscr{F}_{t-1})$ will help to find out the intrinsic defects of those methods.

3.3.2 Existing methods for CAViaR covariance matrix estimation

Based on the literature on quantile regressions, in general there are two ways to estimate $\{h_t(0|\mathscr{F}_{t-1})\}$ in D_T with $\{\varepsilon_{\tau t}\}$ being potentially non-i.i.d.. One is referred to as the *Hendricks Koenker Sandwich Approach* (Hendricks and Koenker, 1992; Koenker, 2005) analogous to the *finite difference* idea resulting in the estimator $\widehat{h}_t^{fd}(0|\mathscr{F}_{t-1})$ for $h_t(0|\mathscr{F}_{t-1})$ as follows:

$$\widehat{h}_t^{fd}(0|\mathscr{F}_{t-1}) = \frac{2\Delta\tau_T}{f_t(\boldsymbol{\beta}_{\tau+\Delta\tau_T}) - f_t(\boldsymbol{\beta}_{\tau-\Delta\tau_T})}, \tag{3.20}$$

where $\Delta\tau_T$ is subject to $0 < \tau \pm \Delta\tau_T < 1$ with $\Delta\tau_T \to 0$ as $T \to \infty$. The other one is referred to as the *Powell Sandwich* (Powell, 1991; Koenker, 2005) based on the kernel density estimation idea resulting in the estimator $\widehat{h}_t^{kernel}(0|\mathscr{F}_{t-1})$ for $h_t(0|\mathscr{F}_{t-1})$ as follows:

$$\begin{aligned} \widehat{h}_t^{kernel}(0|\mathscr{F}_{t-1}) &= \frac{\mathrm{P}\{y_t \le f_t(\boldsymbol{\beta}_\tau) + c_T|\mathscr{F}_{t-1}\} - \mathrm{P}\{y_t \le f_t(\boldsymbol{\beta}_\tau) - c_T|\mathscr{F}_{t-1}\}}{2c_T} \\ &\approx \frac{1}{2c_T} K\left(\frac{y_t - f_t(\boldsymbol{\beta}_\tau)}{2c_T} \right) \end{aligned} \tag{3.21}$$

where $K(\cdot)$ is a suitable kernel function with bandwidth $2c_T$ and $c_T \to 0$ as $T \to \infty$. As we can see in (3.21), one kernel function is applied throughout $\{y_t\}$ with $y_t - f_t(\boldsymbol{\beta}_\tau)$ being the only distinguishable information for $\widehat{h}_t^{kernel}(0|\mathscr{F}_{t-1})$. Therefore, this kernel method does not capture sufficient information to distinguish time-varying conditional distributions of $\{y_t\}$, and consequently cannot fully adapt to the time-variations. Additionally, the choice of the kernel function $K(\cdot)$ and the bandwidth parameter c_T are still in a lot of nettlesome questions in practice. A similar issue in the *Hendricks Koenker Sandwich Approach* is on choosing $\Delta\tau_T$ and extra error resulted from estimating $f_t(\boldsymbol{\beta}_{\tau+\Delta\tau_T})$ and $f_t(\boldsymbol{\beta}_{\tau-\Delta\tau_T})$.

The estimation method adopted by Engle and Manganelli (2004) is a form of the *Powell Sandwich* as follows:

$$\widehat{h}_t^{ker}(0|\mathscr{F}_{t-1}) = \frac{\mathbb{1}_{\{|y_t - f_t(\widehat{\boldsymbol{\beta}}_\tau)| < \widehat{c}_T\}}}{2c_T} \tag{3.22}$$

As suggested by Koenker (2005) and Machado and Silva (2013), the bandwidth $\widehat{c}_T$ generally adopted is defined as follows:

$$\widehat{c}_T = \widehat{k}_T \left[\Phi^{-1}(\tau + m_T) - \Phi^{-1}(\tau - m_T) \right], \tag{3.23}$$

where m_T is defined as

$$\widehat{m}_T = T^{-\frac{1}{3}} \left(\Phi^{-1}(1 - \frac{0.05}{2}) \right)^{\frac{2}{3}} \left(\frac{1.5(\phi(\Phi^{-1}(\tau)))^2}{2(\Phi^{-1}(\tau))^2 + 1} \right)^{\frac{1}{3}}, \tag{3.24}$$

with $\Phi(\cdot)$ and $\phi(\cdot)$ being the cumulative distribution and probability density functions of $N(0,1)$ respectively. And $\widehat{k}_T$ is defined as the median absolute deviation of the conditional τ-th quantile regression residuals.

Wald tests are applied in this subsection to check the performances of the above estimation methods for CAViaR models.

First, we consider the following candidate model specifications for the conditional τ-th ($\tau \in (0,1)$) quantile of a time series $\{y_t\}$ with $f_t(\boldsymbol{\beta}_\tau)$ denoted as the τ-th quantile of y_t conditional on the information set $\mathscr{F}_{t-1}$.

- full specification:

$$f_t(\boldsymbol{\beta}_\tau^{FM}) = \beta_0^{FM}(\tau) + \beta_1^{FM} f_{t-1}(\boldsymbol{\beta}_\tau^{FM}) + \beta_2^{FM}(y_{t-1})^+ + \beta_3^{FM}(y_{t-1})^-, \tag{3.25}$$

  where the operators $(\cdot)^+$ and $(\cdot)^-$ are defined as $(x)^+ = \max(x,0), (x)^- = -\min(x,0)$.

- restrictive model 1:

$$
\begin{aligned}
f_t(\boldsymbol{\beta}_\tau^{R1}) &= \beta_0(\tau)^{R1} + \beta_1^{R1} f_{t-1}(\boldsymbol{\beta}_\tau^{R1}) + \beta_2^{R1} |y_{t-1}| \\
&= \beta_0(\tau)^{R1} + \beta_1^{R1} f_{t-1}(\boldsymbol{\beta}_\tau^{R1}) + \beta_2^{R1} (y_{t-1})^+ + \beta_2^{R1} (y_{t-1})^- .
\end{aligned}
\tag{3.26}
$$

- restrictive model 2:

$$
\begin{aligned}
f_t(\boldsymbol{\beta}_\tau^{R2}) &= \beta_0^{R2}(\tau) + \beta_1^{R2} f_{t-1}(\boldsymbol{\beta}_\tau^{R2}) + \beta_2^{R2} y_{t-1} \\
&= \beta_0^{R2}(\tau) + \beta_1^{R2} f_{t-1}(\boldsymbol{\beta}_\tau^{R2}) + \beta_2^{R2} (y_{t-1})^+ - \beta_2^{R2} (y_{t-1})^- .
\end{aligned}
\tag{3.27}
$$

The models (3.26) and (3.27) are nested within model (3.25). Now let us consider the Wald test on models (3.25) and (3.26) first. Simulate a time series $\{y_t\}$ with its DGP specified as the model (3.26) with the underlying parameter vector $\boldsymbol{\beta}_{u_t}^{R1} = [F_{N(0,1)}^{-1}(u_t), 0.2, 0.3]'$, where $\{u_t\} \overset{i.i.d.}{\sim} \mathcal{U}(0,1)$ and $F_{N(0,1)}^{-1}(\cdot)$ is the inverse standard normal probability distribution function. The sample size of each simulated sample is 4000. Conditional 50%-th quantiles are estimated for each of total 1000 simulated samples in this DGP by regressing the sample onto the full model (3.25). The Wald test implemented here consists of the null hypothesis of the form $H_0 : R\boldsymbol{\beta}_\tau^{FM} = \gamma$, where $R = [0, 0, 1, -1]$, $\gamma = 0$, and $\widehat{\boldsymbol{\beta}}_\tau$ is the estimator of the full model parameter vector in (3.25). The Wald test statistic denoted by W_T is formulated (Weiss, 1991) as follows:

$$
W_T = T \left(R\widehat{\boldsymbol{\beta}}_\tau - \gamma \right)' \left[R\widehat{D}_T^{-1}\widehat{A}_T\widehat{D}_T^{-1}R' \right]^{-1} \left(R\widehat{\boldsymbol{\beta}}_\tau - \gamma \right),
\tag{3.28}
$$

where $\widehat{A}_T$ and $\widehat{D}_T$ are estimates for A_T and D_T in (3.10) respectively. It is straightforward to obtain $\widehat{A}_T$ and $\widehat{D}_T$ by plugging in $\widehat{\boldsymbol{\beta}}_\tau$ and $\left\{\widehat{h}_t(0|\mathscr{F}_{t-1})\right\}$, i.e.,

$$
\begin{cases}
\widehat{A}_T = T^{-1} \tau(1-\tau) \sum_{t=1}^{T} \nabla' f_t(\widehat{\boldsymbol{\beta}}_\tau) \nabla f_t(\widehat{\boldsymbol{\beta}}_\tau), \\
\widehat{D}_T = T^{-1} \sum_{t=1}^{T} \widehat{h}_t(0|\mathscr{F}_{t-1}) \nabla' f_t(\widehat{\boldsymbol{\beta}}_\tau) \nabla f_t(\widehat{\boldsymbol{\beta}}_\tau).
\end{cases}
$$

Notations on $\widehat{D}_T$ to distinguish different estimators used for $\{h_t(0|\mathscr{F}_{t-1})\}$ are given by

$$
\widehat{D}_T^{ker} = (2T\widehat{c}_T)^{-1} \sum_{t=1}^{T} \mathbb{1}_{\{|y_t - f_t(\widehat{\boldsymbol{\beta}}_\tau)| < \widehat{c}_T\}} \nabla' f_t(\widehat{\boldsymbol{\beta}}_\tau) \nabla f_t(\widehat{\boldsymbol{\beta}}_\tau),
\tag{3.29}
$$

$$
\widehat{D}_T^{fd} = (T)^{-1} \sum_{t=1}^{T} \frac{2\Delta\tau_T}{f_t(\boldsymbol{\beta}_{\tau+\Delta\tau_T}) - f_t(\boldsymbol{\beta}_{\tau-\Delta\tau_T})} \nabla' f_t(\widehat{\boldsymbol{\beta}}_\tau) \nabla f_t(\widehat{\boldsymbol{\beta}}_\tau),
\tag{3.30}
$$

where $\widehat{c}_T$ is determined as (3.23).

We are going to examine each element in the estimation of D_T. The analytic solution to $h_t(0|\mathscr{F}_{t-1})$ can be obtained as follows:

$$
\begin{aligned}
h_t(0|\mathscr{F}_{t-1}) &= \frac{\partial \tau}{\partial f_t(\boldsymbol{\beta}_\tau)} = \frac{1}{\frac{\partial \beta_0(\tau)}{\partial \tau} + \beta_1 \frac{\partial f_{t-1}(\boldsymbol{\beta}_\tau)}{\partial \tau}} \\
&= \frac{1}{\frac{\partial \beta_0(\tau)}{\partial \tau} \sum_{i=0}^{n} \beta_1^i + \beta_1^{n+1} \frac{\partial f_{t-n-1}(\boldsymbol{\beta}_\tau)}{\partial \tau}} \\
&= (1 - \beta_1) \frac{1}{\beta_0'(\tau)}
\end{aligned}
\tag{3.31}
$$

where $\beta_0'(\tau) := \frac{\partial \beta_0(\tau)}{\partial \tau}$. The last line is obtained by knowing $|\beta_1| < 1$. The analytic solution to $h_t(0|\mathscr{F}_{t-1})$ is used to help identify inaccurate elements in $\widehat{D}_T$ by comparing the test performances of using $\widehat{D}_T^{ker}$, $\widehat{D}_T^{fd}$ and the following

$$
\widehat{D}_T^{h0} = (T)^{-1} \sum_{t=1}^{T} (1 - \beta_1) \frac{1}{\beta_0'(\tau)} \nabla' f_t(\widehat{\boldsymbol{\beta}}_\tau) \nabla f_t(\widehat{\boldsymbol{\beta}}_\tau).
\tag{3.32}
$$

The test performances of using $\widehat{D}_T^{ker}$, $\widehat{D}_T^{fd}$ and $\widehat{D}_T^{h0}$ are shown in Table 3.1 and 3.2, which are compared together with the Wald test result using the true underlying parameter vector $\boldsymbol{\beta}_\tau^{FM} = [F_{N(0,1)}^{-1}(\tau), 0.2, 0.3]'$, $\tau \in (0,1)$ into

$$
\begin{aligned}
\widehat{D}_T^0 &= (T)^{-1} \sum_{t=1}^{T} (1 - \beta_1) \frac{1}{\beta_0'(\tau)} \nabla' f_t(\boldsymbol{\beta}_\tau) \nabla f_t(\boldsymbol{\beta}_\tau) \\
&= \frac{(1 - 0.2)\, \phi(F_{N(0,1)}^{-1}(\tau))}{T} \sum_{t=1}^{T} \nabla' f_t(\boldsymbol{\beta}_\tau^o) \nabla f_t(\boldsymbol{\beta}_\tau^o),
\end{aligned}
\tag{3.33}
$$

where $\phi(\cdot)$ is the probability density function of $N(0,1)$.

The size performances of the Wald tests on the models (3.25) and (3.26) using different D_T estimators are listed in Table 3.1 in which each estimated size is obtained by the percentage rejection rate among the 1000 samples of $T = 4000$ in the DGP (3.26). Analogously, we implement the Wald test on models (3.25) and (3.27) with the underlying DGP $\{y_t\}$ specified as the model (3.27) with the underlying parameter vector $\boldsymbol{\beta}_{u_t}^{R2} = [F_{N(0,1)}^{-1}(u_t), 0.2, 0.3]'$, where $\{u_t\} \overset{i.i.d.}{\sim} \mathscr{U}(0,1)$. The number of observations in each stimulated sample from this DGP is 4000. Conditional 50%-th quantiles are estimated for each of 1000 simulated samples by regressing the sample onto the full model (3.25). The Wald test implemented in this case consists of the null hypothesis of the form $H_0 : R\boldsymbol{\beta}_\tau^{FM} = \gamma$, where $R = [0,0,1,1]$, $\gamma = 0$, and $\widehat{\boldsymbol{\beta}}_\tau$ is the estimator of the full model regression (3.25).

In result, the size performances of the Wald tests on (3.25) and (3.27) are listed in Table 3.2.

From Table 3.1 and 3.2, we can see large size distortions with $\widehat{D}_T^{fd}$, unlike $\widehat{D}_T^{ker}$, $\widehat{D}_T^{h_0}$ or $\widehat{D}_T^0$ that are performing in line with the nominal size. This comparison points out the crucial element estimation to the accuracy of $\widehat{D}_T$ which is $\left\{ \widehat{h}_t \left(0 | \mathscr{F}_{t-1} \right) \right\}$. To check whether $\{\widehat{h}_t^{ker} \left(0 | \mathscr{F}_{t-1} \right)\}$ is capable to achieve the role of $\{h_t \left(0 | \mathscr{F}_{t-1} \right)\}$ robustly for time-varying conditional probability densities, we consider the following DGP:

$$y_t = f_t(\boldsymbol{\beta}_{u_t}^{R3}) = \beta_0^{R3}(u_t)\sqrt{(y_{t-1})^+} + \beta_1^{R3} f_{t-1}(\boldsymbol{\beta}_\tau^{R3}) + \beta_2^{R3} |y_{t-1}| \tag{3.34}$$

$$= \beta_0^{R3}(u_t)\sqrt{(y_{t-1})^+} + \beta_1^{R3} f_{t-1}(\boldsymbol{\beta}_\tau^{R3}) + \beta_2^{R3} (y_{t-1})^+ + \beta_2^{R3} (y_{t-1})^-, \tag{3.35}$$

where $\{u_t\} \overset{i.i.d.}{\sim} \mathscr{U}(0,1)$ and the underlying parameters are given as $\boldsymbol{\beta}_{u_t}^{R3} = [F_{N(0,1)}^{-1}(u_t), 0.2, 0.3]'$. The analytic form of the corresponding conditional probability density $h_t \left(0 | \mathscr{F}_{t-1} \right)$ of y_t at its τ-th quantile $f_t(\boldsymbol{\beta}_\tau)$ given $\mathscr{F}_{t-1}$ can be derived out as follows:

$$h_t \left(0 | \mathscr{F}_{t-1} \right) = \left(\frac{\partial f_t(\boldsymbol{\beta}_\tau)}{\partial \tau} \right)^{-1} = \left(\frac{\partial \beta_0(\tau)}{\partial \tau} \sqrt{(y_{t-1})^+} + \beta_1 \frac{\partial f_{t-1}(\boldsymbol{\beta}_{u_t})}{\partial \tau} \right)^{-1}$$

$$= \left(\frac{\partial \beta_0(\tau)}{\partial \tau} \sum_{i=1}^{\infty} \beta_1^{i-1} \sqrt{(y_{t-i})^+} \right)^{-1}, \tag{3.36}$$

where the first equation is obtained by iteratively rewriting $\frac{\partial f_{t-i}(\boldsymbol{\beta}_\tau)}{\partial \tau}$ at each i and knowing $|\beta_1| < 1$. This analytic form of $h_t \left(0 | \mathscr{F}_{t-1} \right)$ in (3.34) shows that $\{h_t \left(0 | \mathscr{F}_{t-1} \right)\}$ indeed is time-varying and nonzero with probability one.

We simulate 1000 samples from the DGP (3.34) with $T = 5000$, and estimate the conditional 50%-th quantiles of each sample by regressing the sample onto the full model specification (3.35). The Wald test described as (3.28) with $R = [0, 0, 1, -1]$ is performed on these 1000 samples and the size performance is presented in Table 3.3. We see a large size distortion with the kernel method $\widehat{D}_T^{ker}$ in Table 3.3. More tests are conducted for different DGPs and together with the results are presented in Appendix 3.E. Based on our test results, we see that the kernel method for estimating $\{h_t \left(0 | \mathscr{F}_{t-1} \right)\}$ is not robust and cannot fully adapt to time-varying conditional probability densities.

Estimating $\{h_t \left(0 | \mathscr{F}_{t-1} \right)\}$ robustly has to be achieved in order to ensure the reliability of CAViaR analysis based on the asymptotic properties of CAViaR model

parameter estimators. In seeking for improving the accuracy of $\left\{\widehat{h}_t\left(0|\mathscr{F}_{t-1}\right)\right\}$, we bear in mind two guidances. One is the role of $\{h_t\left(0|\mathscr{F}_{t-1}\right)\}$ on how it links sample elements with the corresponding limit behaviours, see Section 3.3.1. The other guidance is the fundamental flaws of $\{\widehat{h}_t^{ker}\left(0|\mathscr{F}_{t-1}\right)\}$ and $\{\widehat{h}_t^{fd}\left(0|\mathscr{F}_{t-1}\right)\}$ in their accuracy. In terms of $\{\widehat{h}_t^{fd}\left(0|\mathscr{F}_{t-1}\right)\}$, $\Delta\tau_T$ needs to be determined properly and two more quantile regressions need to be preformed in order to obtain $\widehat{\boldsymbol{\beta}}_{\tau+\Delta\tau_T}$ and $\widehat{\boldsymbol{\beta}}_{\tau-\Delta\tau_T}$. The effect of this extra estimation error is crucial to the performance of $\{\widehat{h}_t^{fd}\left(0|\mathscr{F}_{t-1}\right)\}$. Although $\{\widehat{h}_t^{ker}\left(0|\mathscr{F}_{t-1}\right)\}$ does not need extra quantile regressions, it still requires a proper choice on the kernel function $K(\cdot)$ and the bandwidth $\widehat{c}_T$. Remarkably, $\{\widehat{h}_t^{ker}\left(0|\mathscr{F}_{t-1}\right)\}$ does not differentiate the observations within the bandwidth regardless of the number of the observations in the bandwidth while using the kernel function $\mathbb{1}_{\{|y_t-f_t(\widehat{\boldsymbol{\beta}}_\tau)|<\widehat{c}_T\}}$. Therefore, it is desirable to get rid of choosing bandwidth $\Delta\tau_T$ or c_T and the kernel function $K(\cdot)$ in the estimation. In the next subsection, a robust estimation method for $\{h_t\left(0|\mathscr{F}_{t-1}\right)\}$ is developed up without the need in choosing a bandwidth or a kernel function.

3.3.3 Adaptive random bandwidth method

We have noticed that the accuracy of the $\{h_t(0|\mathscr{F}_{t-1})\}$ estimation is crucial to the performance of inference tests based on the asymptotic normality of CAViaR model parameter estimators. It is also well known that $\{\widehat{h}_t^{fd}\left(0|\mathscr{F}_{t-1}\right)\}$ suffers both from the error in estimating $f_t(\boldsymbol{\beta}_{\tau+\Delta\tau})$ and $f_t(\boldsymbol{\beta}_{\tau-\Delta\tau})$ and from choosing a proper $\Delta\tau_T$. On the other hand, $\{\widehat{h}_t^{ker}\left(0|\mathscr{F}_{t-1}\right)\}$ has some fundamental problems. First of all, $\{\widehat{h}_t^{ker}\left(0|\mathscr{F}_{t-1}\right)\}$ cannot fully adapt to time-varying conditional distributions of time series due to the fact that the same kernel function $K(\cdot)$ and only timely information (y_t-f_t) are used in estimating $h_t(0|\mathscr{F}_{t-1})$ for all t. Second, finding a proper kernel function $K(\cdot)$ with a proper bandwidth c_T still faces a lot nettlesome problems in practice. Neither of these two methods is practically robust. The goal in this subsection is to develop an estimation method for $\{h_t(0|\mathscr{F}_{t-1})\}$ which can adapt to time-variation characteristics of CAViaR DGPs and is robust in practice without the need to determine a proper bandwidth. We name this estimation method as the *adaptive random bandwidth* (ARB) method which can reliably bridge asymptotic properties of CAViaR models in theory with CAViaR applications.

The idea of this method is inspired by viewing the role of $\{h_t\left(0|\mathscr{F}_{t-1}\right)\}$ on how it links sample elements with the corresponding limit behaviours, see Section 3.3.1. Reviewing equation (3.16), we can explicitly formulate $\{h_t\left(0|\mathscr{F}_{t-1}\right)\}$

as follows:

$$h_t\left(0|\mathscr{F}_{t-1}\right) = \mathbb{E}_{y_t,\widehat{\boldsymbol{\beta}}}\left[\left.\frac{\mathrm{Hit}_t(\widehat{\boldsymbol{\beta}}) - \mathrm{Hit}_t(\boldsymbol{\beta}^o)}{\nabla' f_t(\boldsymbol{\beta}^o)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\right)}\right| \mathscr{F}_{t-1}\right], \qquad (3.37)$$

which actually is a conditional expectation taken with respect to random variables y_t and $\widehat{\boldsymbol{\beta}}$. We use the subscript in $\mathbb{E}$ to clarify the expectation is taken with respect to specific random variable(s) hereafter. Considering this role of $\{h_t\left(0|\mathscr{F}_{t-1}\right)\}$ as well as equation (3.19), we are enlightened to use random bandwidth $\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)$ with $\sqrt{T}\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right) \overset{d}{\sim} N(\boldsymbol{0}, \boldsymbol{V_d})$ and $i = 1, 2, \ldots, n$. We can set $\boldsymbol{V_d} = I_{(p+1)\times(p+1)}$ to start with. After sufficient n times Monte Carlo simulating $\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}$ from $N(\boldsymbol{0}, \boldsymbol{V_d})$, an estimator of $h_t\left(0|\mathscr{F}_{t-1}\right)$ can be achieved as follows:

$$\widehat{h}_t(0|\mathscr{F}_{t-1}) := n^{-1}\sum_{i=1}^{n}\frac{\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})\right\}} - \mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}}) + \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)\right\}}}{-\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)}. \qquad (3.38)$$

After achieving the above $\widehat{h}_t(0|\mathscr{F}_{t-1})$, we can estimate $\widehat{D}_T$ so as to update $\boldsymbol{V_d} = \widehat{D}_T^{-1}\widehat{A}_T\widehat{D}_T^{-1}$. Redo the simulation of $\{\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\}_{i=1}^{n}$ with the updated $\boldsymbol{V_d}$. We can estimate $\widehat{h}_t(0|\mathscr{F}_{t-1})$ and $\widehat{D}_T$ again. This estimation repetition can mitigate the influence of an arbitrary chosen $\widehat{h}_t(0|\mathscr{F}_{t-1})$ in ARB.

Compared to the Powell Sandwich estimation (3.21) with c_T, our proposed method uses random bandwidth $\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)$ and Monte Carlo simulations such that it can adapt to time-varying conditional distributions of CAViaR DGPs by approaching to the role of $\{h_t\left(0|\mathscr{F}_{t-1}\right)\}$ as in (3.19) and in (3.37). The adaptive random bandwidth method can remarkably outperform the Powell Sandwich method in the applications on DGPs of time-varying conditional distributions, as shown in Table 3.3. In theory, the adaptive random bandwidth method is valid as long as $\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^o - \widehat{\boldsymbol{\beta}}$ have the same order of magnitude. We formally establish this adaptive random bandwidth method in Theorem 6.

Theorem 6 (Adaptive Random Bandwidth Method).
Assume the conditions and the asymptotic normality result in Theorem 5. Choose an arbitrary positive definite symmetric matrix $\boldsymbol{V_d}$. Under the condition that

$$\sqrt{T}\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right) \overset{i.i.d.}{\sim} N(\boldsymbol{0}, \boldsymbol{V_d}), \qquad i = 1, \ldots, n, \qquad (3.39)$$

and

$$\left|\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)\right| \neq 0,$$

the adaptive random bandwidth estimator for $h_t(0|\mathscr{F}_{t-1})$ is formulated as follows:

$$\widehat{h}_t(0|\mathscr{F}_{t-1}) = \begin{cases} n^{-1} \sum_{i=1}^{n} \dfrac{\mathbb{1}\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}}) + \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)\right\} - \mathbb{1}\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})\right\}}{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)}, & \text{when } y_t \neq f_t(\widehat{\boldsymbol{\beta}}), \\[4mm] 0, & \text{when } y_t = f_t(\widehat{\boldsymbol{\beta}}), \end{cases}$$

(3.40)

such that

$$\mathbb{E}_{y_t, \widehat{\boldsymbol{\beta}}}\left[\widehat{h}_t(0|\mathscr{F}_{t-1}) \,\middle|\, \mathscr{F}_{t-1}\right] \xrightarrow{p} h_t(0|\mathscr{F}_{t-1})$$

as $n \to \infty$. [6]

Proof. See Appendix 3.C. □

We separate the case of $y_t = f_t(\widehat{\boldsymbol{\beta}})$ from others to maintain the convergence of the ARB estimator due to $\lim_{x \to 0} \frac{1}{x} = \infty$. Zero given to $\widehat{h}_t(0|\mathscr{F}_{t-1})$ at $y_t = f_t(\widehat{\boldsymbol{\beta}})$ also enables the ARB estimator to approximate $h_t(0|\mathscr{F}_{t-1})$ from the left and from the right in half weights respectively in expectation, see the proof of Theorem 6. The convergence property of the partial sum in the sequence $\{\widehat{h}_t(0|\mathscr{F}_{t-1})\}$ by ARB is given in Corollary 7.

Corollary 7. *Under the conditions of Theorem 6, the adaptive random bandwidth estimator $\{\widehat{h}_t(0|\mathscr{F}_{t-1})\}$ has the following property:*

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{h}_t(0|\mathscr{F}_{t-1}) \xrightarrow{m.s.} \frac{1}{T}\sum_{t=1}^{T} h_t(0|\mathscr{F}_{t-1}),$$

(3.41)

as $T, n \to \infty$.

Proof. See Appendix 3.C. □

It is clear that both $\widehat{\varepsilon}_t := y_t - f_t(\widehat{\boldsymbol{\beta}})$ and $\nabla' f_t(\widehat{\boldsymbol{\beta}})$ are taken into account by ARB to approximate $h_t(0|\mathscr{F}_{t-1})$. In order to identify how $\widehat{\varepsilon}_t$ and $\nabla' f_t(\widehat{\boldsymbol{\beta}})$ jointly shape $\widehat{h}_t(0|\mathscr{F}_{t-1})$, we would like to formulate $\widehat{h}_t(0|\mathscr{F}_{t-1})$ in Theorem 6 into an analytic

---

[6]We regard the least $(p+1)$ absolute residuals in $\{|y_t - f_t(\widehat{\boldsymbol{\beta}})|\}_{t=1}^{T}$ as zeros. In fact, iterations of a simplex-based direct search method like the Nelder–Mead method for optimizing $(p+1)$ parameters terminates at the vertices of a simplex in the parameter space (Lagarias et al., 1998). That is to say, the iterations in optimizing the τ-th quantile regression objective function terminate with $(p+1)$ elements of $\{(\tau - \mathbb{1}_{\{y_t - f_t(\boldsymbol{\beta}) < 0\}})(y_t - f_t(\boldsymbol{\beta}))\}$ solved to be zeros. Therefore, we set $\widehat{h}_t(0|\mathscr{F}_{t-1}) = 0$ at the least $(p+1)$ absolute residuals in $\{|y_t - f_t(\widehat{\boldsymbol{\beta}})|\}_{t=1}^{T}$ in all the tests throughout this chapter.

expression in terms of $\widehat{\varepsilon}_t$ and $\nabla' f_t(\widehat{\boldsymbol{\beta}})$ so as to manifest the relationship. The analytic form of $\widehat{h}_t(0|\mathscr{F}_{t-1})$ by ARB described in Theorem 6 is presented in Corollary 8.

Corollary 8. *Under the conditions of Theorem 6, we can get the analytic form of $\widehat{h}_t(0|\mathscr{F}_{t-1})$ as follows:*

$$
\widehat{h}_t(0|\mathscr{F}_{t-1}) = \begin{cases} \dfrac{1}{2\delta_{\nabla_t}\sqrt{2\pi}} E_1\left(\dfrac{\widehat{\varepsilon}_t^2}{2\delta_{\nabla_t}^2}\right), & \text{when } \widehat{\varepsilon}_t \neq 0, \\ 0, & \text{when } \widehat{\varepsilon}_t = 0, \end{cases} \tag{3.42}
$$

where $\widehat{\varepsilon}_t := y_t - f_t(\widehat{\boldsymbol{\beta}})$, $\delta_{\nabla} := \sqrt{\dfrac{\nabla' f_t(\widehat{\boldsymbol{\beta}}) \boldsymbol{V_d} \nabla f_t(\widehat{\boldsymbol{\beta}})}{T}} = T^{-\frac{1}{2}} \|\nabla f_t\|_2$, *and* $E_1(s) := \int_s^\infty x^{-1} e^{-x} dx$ *is a special integral known as the exponential integral or the incomplete gamma function* $\Gamma(0,s)$.

Proof. See Appendix 3.C. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

For visually checking the roles of $\widehat{\varepsilon}_t$ and δ_{∇_t} in the analytic $\widehat{h}_t(0|\mathscr{F}_{t-1})$ in Corollary 8, we present a level plot of the analytic $\widehat{h}_t(0|\mathscr{F}_{t-1})$ over $\widehat{\varepsilon}_t$ and δ_{∇_t} in Figure 3.2 which uses colors to differentiate different ranges of $\widehat{h}_t(0|\mathscr{F}_{t-1})$. It is straightforward to get that the analytic $\widehat{h}_t(0|\mathscr{F}_{t-1})$ is decreasing in $|\widehat{\varepsilon}_t|$ as also shown in Figure 3.2. However, δ_{∇_t}, or say $T^{-\frac{1}{2}} \|\nabla f_t\|_2$, can shift $\widehat{h}_t(0|\mathscr{F}_{t-1})$ by reflecting on how rare an $\widehat{\varepsilon}_t$ is observed given the information set $\mathscr{F}_{t-1}$ and the model specification. That is how the information of δ_{∇_t} in ARB shapes $\widehat{h}_t(0|\mathscr{F}_{t-1})$ adaptively to time-varying conditional probability densities.

The ARB estimator $\widehat{h}_t(0|\mathscr{F}_{t-1})$ via simulations in Theorem 6 performs as robustly as the analytic ARB estimator $\widehat{h}_t(0|\mathscr{F}_{t-1})$ in Corollary 8, as shown in Table 3.1, 3.2 and 3.3. The analytic way is faster than the simulation one. However, the ARB estimator via simulations is more intuitive and more flexible to adapt to a very different distribution for simulating $\{\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\}_{i=1}^n$.

D_T need to be estimated consistently for inference tests on CAViaR models based on the asymptotic normality of the model parameter estimator. $\{\widehat{h}_t(0|\mathscr{F}_{t-1})\}$ by ARB facilitates our estimation on D_T by just plugging in $\widehat{\boldsymbol{\beta}}$ and $\{\widehat{h}_t(0|\mathscr{F}_{t-1})\}$. The resulted estimator $\widehat{D}_T^{arb}$ has the consistency property presented in Theorem 9.

Theorem 9. *Under the conditions of Theorem 6, we can get that*

$$
\widehat{D}_T^{arb} \xrightarrow{p} D_T, \tag{3.43}
$$

Level plot of $\frac{1}{2\delta_{\nabla_t}\sqrt{2\pi}}E_1\left(\frac{\hat{\epsilon}_t^2}{2\delta_{\nabla_t}^2}\right)$
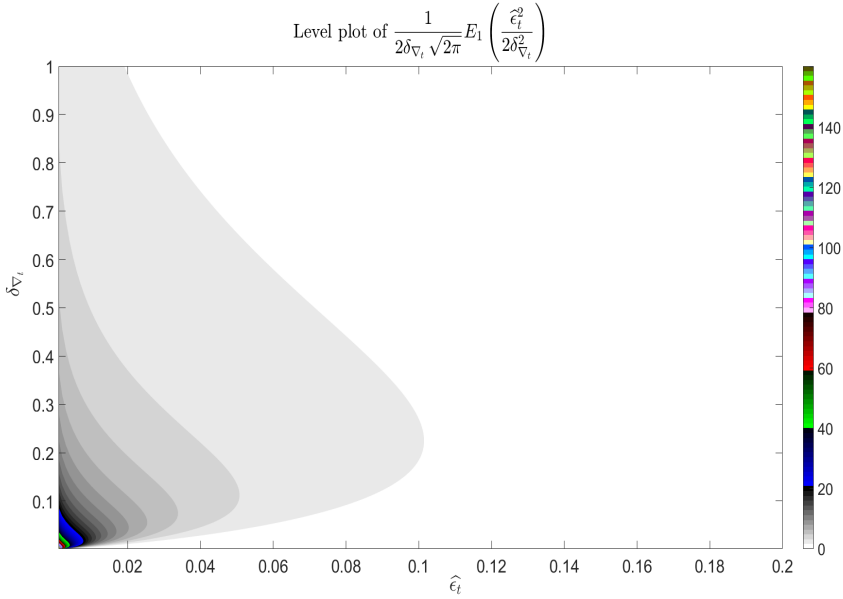
Figure 3.2: Level plot for the analytic form of $\widehat{h}_t(0|\mathscr{F}_{t-1})$ by ARB in Corollary 8.

as $T \to \infty$ *and* $n \to \infty$, *where* $\widehat{D}_T^{arb} := T^{-1}\sum_{t=1}^{T}\widehat{h}_t(0|\mathscr{F}_{t-1})\nabla'f_t(\widehat{\boldsymbol{\beta}}_\tau)\nabla f_t(\widehat{\boldsymbol{\beta}}_\tau)$ *and*

$\widehat{h}_t(0|\mathscr{F}_{t-1})$ *is the adaptive random bandwidth estimator shown in* (3.40).

Proof. See Appendix 3.C. □

The *adaptive random bandwidth* (ARB) method is intuitive, robust and simple in practice, which can adapt to time-varying conditional distributions without a specific bandwidth or kernel function. A comparison of size performances of Wald tests using ARB with other competing methods are presented in Tables 3.1, 3.2 and 3.3. We also find that updating $\boldsymbol{V_d}$ improves the size performance with use of α levels in the interquartile range around but not much for α levels like $1\%, 5\%$. More test results are presented in Appendix 3.E with changing sample size, quantile index and varying DGPs. The performance of ARB is robust. ARB can also be easily generalized to apply on multivariate quantile regressions, which is beyond the scope of this chapter but in the interest of multivariate quantile regressions for future research. ARB also has the potential to achieve the second-order accuracy to Wald tests of nonlinear restrictions (Phillips and Park, 1988; de Paula Ferrari and Cribari-Neto, 1993) in quantile regressions, which we would like to leave for future research.

Table 3.1: The size performances of the Wald test on the restricted model (3.26) to (3.25) with different estimation methods for $\{h_t(0|\mathscr{F}_{t-1})\}$ ($\boldsymbol{\beta}_{0.5}^{R1} = [0, 0.2, 0.3, 0.3], R = [0, 0, 1, -1], T = 4000$)

Tests	size: $\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
Using $\widehat{D}_T^0$	0.017	0.063	0.127	0.215
Using $\widehat{D}_T^{h_0}$	0.016	0.066	0.131	0.215
Using $\widehat{D}_T^{arb}$ ($n = 10^4$, $\boldsymbol{V_d} = \boldsymbol{I}_{4\times4}$ with no update)	0.012	0.052	0.098	0.196
Using $\widehat{D}_T^{arb}$ (analytic, $\boldsymbol{V_d} = \boldsymbol{I}_{4\times4}$ with no update)	0.012	0.052	0.102	0.198
Using $\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.014	0.062	0.126	0.221
Using $\widehat{D}_T^{arb}$ (analytic, 2 times updating $\boldsymbol{V_d}$)	0.014	0.061	0.125	0.219
Using $\widehat{D}_T^{fd}$ ($\Delta\tau_T = \frac{10}{T}$)	0.080	0.150	0.201	0.272
Using $\widehat{D}_T^{ker}$	0.017	0.069	0.129	0.223

Table 3.2: The size performances of the Wald test on the restricted model (3.27) to (3.25) with different estimation methods for $\{h_t(0|\mathscr{F}_{t-1})\}$ ($\boldsymbol{\beta}_{0.5}^{R2} = [0, 0.2, 0.3, -0.3], R = [0, 0, 1, 1], T = 4000$)

Tests	size: $\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
Using $\widehat{D}_T^0$	0.008	0.050	0.104	0.206
Using $\widehat{D}_T^{h_0}$	0.007	0.050	0.105	0.207
Using $\widehat{D}_T^{arb}$ ($n = 10^4$, $\boldsymbol{V_d} = \boldsymbol{I}_{4\times4}$ with no update)	0.01	0.046	0.084	0.168
Using $\widehat{D}_T^{arb}$ (analytic, $\boldsymbol{V_d} = \boldsymbol{I}_{4\times4}$ with no update)	0.009	0.044	0.083	0.168
Using $\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.011	0.049	0.098	0.192
Using $\widehat{D}_T^{arb}$ (analytic, 2 times updating $\boldsymbol{V_d}$)	0.01	0.048	0.097	0.19
Using $\widehat{D}_T^{fd}$ ($\Delta\tau_T = \frac{10}{T}$)	0.049	0.104	0.153	0.229
Using $\widehat{D}_T^{ker}$	0.011	0.05	0.094	0.203

Table 3.3: The size performances of the Wald test on the restricted model (3.34) to (3.35) with different estimation methods for $\{h_t(0|\mathscr{F}_{t-1})\}$ ($\boldsymbol{\beta}_{0.5}^{R3} = [0, 0.2, 0.3, 0.3], R = [0, 0, 1, -1], T = 2000$)

Tests	size: $\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
Using $\widehat{D}_T^{arb}$ ($n = 10^4$, $\boldsymbol{V_d} = \boldsymbol{I}_{4\times4}$ with no update)	0.024	0.052	0.095	0.169
Using $\widehat{D}_T^{arb}$ (analytic, $\boldsymbol{V_d} = \boldsymbol{I}_{4\times4}$ with no update)	0.023	0.054	0.093	0.168
Using $\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.021	0.055	0.095	0.188
Using $\widehat{D}_T^{arb}$ (analytic, 2 times updating $\boldsymbol{V_d}$)	0.022	0.055	0.098	0.186
Using $\widehat{D}_T^{ker}$	0.067	0.118	0.16	0.256

3.4 Empirical Results

We study four US stock prices which are the Dow Jones Composite Average (DJCA), the NASDAQ 100 Index (NASDAQ100), the S&P 500, and the Wilshire 5000 Total Market Index (Will5000ind). We implement inference tests using the *adaptive random bandwidth* method with $n = 1000$ and $\boldsymbol{V_d} = \boldsymbol{I}_{(p+1)\times(p+1)}$ which is not updated in simulations in this section. Each stock price time series has 2448 daily prices, ranging from 8th April 2010 to 30th December 2019. The price data were converted to return rates by multiplying 100 with the difference of the natural logarithm of the daily prices. The obtained return time series of each stock contains 2447 observations which of the last 400 observations are used for the out-of-sample testing after the first 2047 observations are used to estimate the model.

The 5% 1-day VaRs of a return time series are the opposite conditional 5% 1-day quantiles of this time series. There are four different CAViaR models considered in this section to model the conditional quantiles of the stock return time series. The 5% 1-day VaRs are estimated via the four different CAViaR specifications and the estimation results are shown in Table 3.4, 3.5, 3.6 and 3.7 respectively. Each table contains the estimated parameters in a specified model, the corresponding standard errors obtained by the *adaptive random bandwidth* method with $n = 1000$ and $\boldsymbol{V_d} = \boldsymbol{I}_{(p+1)\times(p+1)}$, the resulted two-sided p-values on parameter significance, the optimized value of the quantile regression objective function (RQ), the percentage of times the VaR is exceeded, and the p-values of dynamic quantile (DQ) tests, both in-sample and out-of-sample. The model estimations, the in-sample DQ tests as well as the out-of-sample DQ tests in this empirical study are set up in the same way of Section 6 of Engle and Manganelli (2004).

- Adaptive CAViaR:

$$f_t(\beta_1) = f_{t-1}(\beta_1) + \beta_1 \left\{ [1 + exp(G[y_{t-1} - f_{t-1}(\beta_1))])]^{-1} - \tau \right\}, \quad (3.44)$$

  where τ is the quantile index of interest.

- Symmetric absolute value CAViaR:

$$f_t(\boldsymbol{\beta}) = \beta_0 + \beta_1 f_{t-1}(\boldsymbol{\beta}) + \beta_2 |y_{t-1}|. \qquad (3.45)$$

- Asymmetric slope CAViaR:

$$f_t(\boldsymbol{\beta}) = \beta_0 + \beta_1 f_{t-1}(\boldsymbol{\beta}) + \beta_2 (y_{t-1})^+ + \beta_3 (y_{t-1})^-. \qquad (3.46)$$

- Indirect GARCH$(1,1)$:

$$f_t(\boldsymbol{\beta}) = -\sqrt{\beta_0 + \beta_1 f_{t-1}(\boldsymbol{\beta})^2 + \beta_2 y_{t-1}^2}. \qquad (3.47)$$

The above four CAViaR specifications have been defined as the adaptive CAViaR, the symmetric absolute value CAViaR, the asymmetric slope CAViaR, and the indirect GARCH$(1,1)$ respectively in the Section 3 of Engle and Manganelli (2004). In the implementation of the adaptive model in this emprical study, we follow Engle and Manganelli (2004) and set $G = 10$.

Table 3.4: The Asymmetric Slope Model ($\tau = 0.05$)

Stock Name	DJCA	NASDAQ100	S&P500	Will5000ind
$\hat{\beta}_0$	-0.0538	-0.1366	-0.0772	-0.0803
s.e.($\hat{\beta}_0$)	0.0192	0.0324	0.0283	0.0259
p-value($\hat{\beta}_0$)	0.0051*	0.0000*	0.0063*	0.0019*
$\hat{\beta}_1$	0.8913	0.8536	0.8651	0.8613
s.e.($\hat{\beta}_1$)	0.0276	0.0356	0.0344	0.0344
p-value($\hat{\beta}_1$)	0.0000*	0.0000*	0.0000*	0.0000*
$\hat{\beta}_2$	-0.0175	0.0381	0.0264	0.0158
s.e.($\hat{\beta}_2$)	0.0325	0.0717	0.0732	0.0831
p-value($\hat{\beta}_2$)	0.5918	0.5950	0.7179	0.8487
$\hat{\beta}_3$	-0.3069	-0.3626	-0.4249	-0.4226
s.e.($\hat{\beta}_3$)	0.0667	0.0673	0.1214	0.1153
p-value($\hat{\beta}_3$)	0.0000*	0.0000*	0.0005*	0.0002*
RQ	205.1100	253.9000	215.3800	219.3200
Exceedance in-sample (%)	5.0166	5.0639	5.0166	5.0166
Exceedance out-of-sample %	4.7326	4.8746	4.5906	4.5433
DQ in-sample (p value)	0.4306	0.5140	0.3094	0.4425
DQ out-of-sample (p value)	1.0000	1.0000	1.0000	1.0000

Comparing with the results in Section 6 of Engle and Manganelli (2004), we can see the standard errors obtained by the *adaptive random bandwidth* method is much smaller relatively to the size of estimated parameters. We use significance level 5% to reject a parameter equal to zero as well as DQ tests. " * " denotes the rejections in Table 3.4, 3.5, 3.6 and 3.7. Each of the four models shows almost the same rejection results for the stock return time series. Remarkably, it is observed that the coefficient β_1 of the VaR autoregressive term is highly significant from zero in all the four models for each stock return time series. This further supports the standpoint of CAViaR specifications, confirming that the phenomenon of volatility clustering can be associated with the autoregressive VaR behaviour. The VaR exceedance in percentage indicates the realized risk

Table 3.5: The Symmetric Absolute Value Model ($\tau = 0.05$)

Stock Name	DJCA	NASDAQ100	S&P500	Will5000ind
$\hat{\beta}_0$	-0.0507	-0.1310	-0.0544	-0.0521
s.e.($\hat{\beta}_0$)	0.0405	0.0641	0.0430	0.0315
p-value($\hat{\beta}_0$)	0.2103	0.0410*	0.2064	0.0984
$\hat{\beta}_1$	0.8546	0.8127	0.8495	0.8676
s.e.($\hat{\beta}_1$)	0.0418	0.0629	0.0544	0.0324
p-value($\hat{\beta}_1$)	0.0000*	0.0000*	0.0000*	0.0000*
$\hat{\beta}_2$	-0.2375	-0.2492	-0.2485	-0.2161
s.e.($\hat{\beta}_2$)	0.0266	0.0785	0.0775	0.0311
p-value($\hat{\beta}_2$)	0.0000*	0.0015*	0.0013*	0.0000*
RQ	210.7300	263.0400	223.5300	227.3200
Exceedance in-sample (%)	5.0166	5.0166	5.0166	5.0166
Exceedance out-of-sample (%)	5.3952	5.2532	4.9219	4.9692
DQ in-sample (p value)	0.2306	0.3548	0.0470*	0.1537
DQ out-of-sample (p value)	1.0000	1.0000	1.0000	1.0000

Table 3.6: The indirect GARCH$(1, 1)$ ($\tau = 0.05$)

Stock Name	DJCA	NASDAQ100	S&P500	Will5000ind
$\hat{\beta}_0$	0.0651	0.2152	0.0878	0.0758
s.e.($\hat{\beta}_0$)	0.0325	0.1069	0.0384	0.0414
p-value($\hat{\beta}_0$)	0.0450*	0.0442*	0.0223*	0.0670
$\hat{\beta}_1$	0.8741	0.7930	0.8290	0.8566
s.e.($\hat{\beta}_1$)	0.0247	0.0444	0.0261	0.0258
p-value($\hat{\beta}_1$)	0.0000*	0.0000*	0.0000*	0.0000*
$\hat{\beta}_2$	0.2551	0.3775	0.3638	0.2964
s.e.($\hat{\beta}_2$)	0.2169	0.2031	0.2096	0.2041
p-value($\hat{\beta}_2$)	0.2395	0.0631	0.0826	0.1465
RQ	209.4600	262.4600	222.1100	226.5200
Exceedance in-sample (%)	4.9692	5.0166	5.0639	5.0639
Exceedance out-of-sample (%)	5.3005	5.2059	4.6853	4.8273
DQ in-sample (p value)	0.3678	0.4108	0.2887	0.4216
DQ out-of-sample (p value)	1.0000	1.0000	1.0000	1.0000

Table 3.7: The Adaptive Model ($\tau = 0.05$)

Stock Name	DJCA	NASDAQ100	S&P500	Will5000ind
$\hat{\beta}_1$	-0.6980	-0.7027	-0.9827	-1.5480
s.e.($\hat{\beta}_1$)	0.0768	0.0760	0.0520	0.0014
p-value($\hat{\beta}_1$)	0.0000*	0.0000*	0.0000*	0.0000*
RQ	213.4500	272.7100	226.9600	231.9700
Exceedance in-sample (%)	4.4487	4.8746	4.6380	4.3067
Exceedance out-of-sample %	4.7799	5.1585	4.8746	4.4960
DQ in-sample (p value)	0.6518	0.9802	0.9545	0.2118
DQ out-of-sample (p value)	1.0000	1.0000	1.0000	1.0000

level in applications. Dynamic quantile (DQ) tests based on the independence information regarding $\{\text{Hit}_t\}$ are used to test model misspecification. We see a rejection in the in-sample DQ test on the symmetric absolute value model for the S&P500 but the realized VaR exceedances (in-sample and out-of-sample) are much close to 5% in Table 3.5. So it can be complementary to judge CAViaR model specifications by looking at both VaR exceedances and inference tests like DQ tests.

In contract to the significance of β_1, the coefficient β_2 of $(y_{t-1})^2$ is insignificant in the indirect GARCH(1,1) model for all the stock return time series, see Table 3.6. And the coefficient β_2 of $(y_{t-1})^+$ is insignificant in the asymmetric slope model, see Table 3.4. Although the coefficient of y_{t-1} is significant in the symmetric absolute model for all the stock return time series (see Table 3.5), it is mainly due to the significant explanatory role of $(y_{t-1})^-$ based on the results of the asymmetric slope model which the symmetric absolute model is nested in. The significance results of β_1 in the adaptive model for each stock return time series suggest that the 5% 1-day VaR can be associated with its 1-day lagged VaR violation which equals one if $y_{t-1} \le f_{t-1}$ and zero otherwise. The significance results together implies that negative movements of a stock is significantly influential on its 5% 1-day VaR in the next day.

In terms of the model goodness of fit, we look at the RQ results. The asymmetric slope model presents the lowest RQ result for each stock return time series among the four models despite that it has the most coefficients.

Overall, all the four stock return time series present the same strong associations with the lagged 5% 1-day VaR in interpreting the present 5% 1-day VaR. The asymmetric slope model and the adaptive CAViaR are satisfying for all the four stock returns in terms of data interpretation and model performance concerns.

3.5 Conclusions

We found that the inference test performance in CAViaR models is not robust and unsatisfying due to the estimation of the conditional probability densities of time series. We found that the existing density estimation methods cannot fully adapt to time-varying conditional probability densities of CAViaR time series. So in this chapter we have developed a method called *adaptive random bandwidth* which can robustly approximate the time-varying conditional probability densities of CAViaR time series by Monte Carlo simulations. This method not only avoids the haunting problem of choosing an optimal bandwidth but also ensures the reliability of CAViaR analysis based on the asymptotic normality of the model parameter estimator. In theory, our proposed method can be extended to general quantile regressions including multivariate cases easily and robustly. This method also has the potential to achieve the second-order accuracy to Wald tests of nonlinear restrictions (Phillips and Park, 1988; de Paula Ferrari and Cribari-Neto, 1993) in quantile regressions.

3.A Appendix: Nonlinearity of Parameters in CAViaR Models

Nonlinearity of parameters in CAViaR models differentiates CAViaR from linear quantile regressive models. In this appendix, we would like to illustrate the non-linearity explicitly by showing the gradient, and the Hessian matrix of a CAViaR model.

$$f_t(\boldsymbol{\beta}_\tau) = \beta_0(\tau) + \beta_1 f_{t-1}(\boldsymbol{\beta}_\tau) + \beta_2 \, (y_{t-1})^+ + \beta_3 \, (y_{t-1})^-, \tag{3.48}$$

where $\tau \in (0,1)$, and the operators $(\cdot)^+$ and $(\cdot)^-$ are defined as $(x)^+ = \max(x,0), (x)^- = -\min(x,0)$. This model can be rewritten by continuously substituting lagged conditional quantiles such as

$$
\begin{aligned}
f_t(\boldsymbol{\beta}_\tau) &= \beta_0(\tau) + \beta_1 f_{t-1}(\boldsymbol{\beta}_\tau) + \beta_2 \, (y_{t-1})^+ + \beta_3 \, (y_{t-1})^- \\
&= \beta_0(\tau) + \beta_1 \left(\beta_0(\tau) + \beta_1 f_{t-2}(\boldsymbol{\beta}_\tau) + \beta_2 \, (y_{t-2})^+ + \beta_3 \, (y_{t-2})^- \right) \\
&\quad + \beta_2 \, (y_{t-1})^+ + \beta_3 \, (y_{t-1})^- \\
&= \frac{\beta_0(\tau)}{1 - \beta_1} + \beta_2 \sum_{j=1}^{\infty} \beta_1^{j-1} \, (y_{t-j})^+ + \beta_3 \sum_{j=1}^{\infty} \beta_1^{j-1} \, (y_{t-j})^-,
\end{aligned}
\tag{3.49}
$$

where the last line comes from $|\beta_1| < 1$. If $\beta_1 \neq 0$, (3.49) reveals explicitly the nonlinear pattern of parameters in this CAViaR model. From this explicit form, we can further get the gradient and the Hessian matrix of the CAViaR model (3.48) to emphasize the roles of the parameters.

3.A.1 Gradient ∇f_t

The gradient of $f_t(\boldsymbol{\beta}_\tau)$ at a conditional quantile index $\tau \in (0,1)$ of interest can be derived as follows:

$$
\begin{aligned}
\nabla f_t(\boldsymbol{\beta}_\tau) &=
\begin{bmatrix}
\frac{\partial f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_0} \\
\frac{\partial f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_1} \\
\frac{\partial f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_2} \\
\frac{\partial f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_3}
\end{bmatrix}
=
\begin{bmatrix}
1 + \beta_1 \frac{\partial f_{t-1}(\boldsymbol{\beta}_\tau)}{\partial \beta_0} \\
f_{t-1}(\boldsymbol{\beta}_\tau) + \beta_1 \frac{\partial f_{t-1}(\boldsymbol{\beta}_\tau)}{\partial \beta_1} \\
(y_{t-1})^+ + \beta_1 \frac{\partial f_{t-1}(\boldsymbol{\beta}_\tau)}{\partial \beta_2} \\
(y_{t-1})^- + \beta_1 \frac{\partial f_{t-1}(\boldsymbol{\beta}_\tau)}{\partial \beta_3}
\end{bmatrix} \\[2ex]
&=
\begin{bmatrix}
1 + \beta_1 \left(1 + \beta_1 \frac{\partial f_{t-1}(\boldsymbol{\beta}_\tau)}{\partial \beta_0} \right) \\
f_{t-1}(\boldsymbol{\beta}_\tau) + \beta_1 \left(f_{t-2}(\boldsymbol{\beta}_\tau) + \beta_1 \frac{\partial f_{t-2}(\boldsymbol{\beta}_\tau)}{\partial \beta_1} \right) \\
(y_{t-1})^+ + \beta_1 \left((y_{t-2})^+ + \beta_1 \frac{\partial f_{t-2}(\boldsymbol{\beta}_\tau)}{\partial \beta_2} \right) \\
(y_{t-1})^- + \beta_1 \left((y_{t-2})^- + \beta_1 \frac{\partial f_{t-2}(\boldsymbol{\beta}_\tau)}{\partial \beta_3} \right)
\end{bmatrix}
=
\begin{bmatrix}
\frac{1}{1 - \beta_1} \\
\sum_{i=1}^{\infty} \beta_1^{i-1} f_{t-i}(\boldsymbol{\beta}_\tau) \\
\sum_{i=1}^{\infty} \beta_1^{i-1} (y_{t-i})^+ \\
\sum_{i=1}^{\infty} \beta_1^{i-i} (y_{t-i})^-
\end{bmatrix}.
\end{aligned}
\tag{3.50}
$$

By knowing (3.49), we substitute

$$f_{t-i}(\boldsymbol{\beta}_\tau) = \frac{\beta_0(\tau)}{1-\beta_1} + \beta_2 \sum_{j=1}^\infty \beta_1^{j-1}(y_{t-i-j})^+ + \beta_3 \sum_{j=1}^\infty \beta_1^{j-1}(y_{t-i-j})^-,$$

into $\nabla f_t(\boldsymbol{\beta}_\tau)$ in (3.50) and get

$$\nabla f_t(\boldsymbol{\beta}_\tau) = \begin{bmatrix} \frac{1}{1-\beta_1} \\ \sum_{i=1}^\infty \beta_1^{i-1}\left(\frac{\beta_0(\tau)}{1-\beta_1} + \beta_2 \sum_{j=1}^\infty \beta_1^{j-1}(y_{t-i-j})^+ + \beta_3 \sum_{j=1}^\infty \beta_1^{j-1}(y_{t-i-j})^- \right) \\ \sum_{i=1}^\infty \beta_1^{i-1}(y_{t-i})^+ \\ \sum_{i=1}^\infty \beta_1^{i-i}(y_{t-i})^- \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{1-\beta_1} \\ \frac{\beta_0(\tau)}{(1-\beta_1)^2} + \beta_2 \sum_{h=2}^\infty (h-1)\beta_1^{h-2}(y_{t-h})^+ + \beta_3 \sum_{h=2}^\infty (h-1)\beta_1^{h-2}(y_{t-h})^- \\ \sum_{i=1}^\infty \beta_1^{i-1}(y_{t-i})^+ \\ \sum_{i=1}^\infty \beta_1^{i-1}(y_{t-i})^- \end{bmatrix}.$$

$$(3.51)$$

Now we can see the role of the parameters $\boldsymbol{\beta}_\tau$ explicitly. $\boldsymbol{\beta}_\tau$ shows up in all the elements of the gradient in a nonlinear form which makes it doubtless that the Hessian matrix does not fade out with $\boldsymbol{\beta}_\tau$ either.

3.A.2 Hessian matrix

The second partial derivatives of $f_t(\boldsymbol{\beta}_\tau)$ exist as $\nabla f_t(\boldsymbol{\beta}_\tau)$ does, which can be seen from the derivation of the Hessian matrix $\boldsymbol{H}(\boldsymbol{\beta}_\tau)$ of $f_t(\boldsymbol{\beta}_\tau)$ as follows:

$$\boldsymbol{H}(\boldsymbol{\beta}_\tau) = \begin{bmatrix} \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_0^2} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_0 \partial \beta_2} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_0 \partial \beta_3} \\ \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_1 \partial \beta_0} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_1^2} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_1 \partial \beta_2} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_1 \partial \beta_3} \\ \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_2 \partial \beta_0} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_2^2} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_2 \partial \beta_3} \\ \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_3 \partial \beta_0} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_3 \partial \beta_1} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_3 \partial \beta_2} & \frac{\partial^2 f_t(\boldsymbol{\beta}_\tau)}{\partial \beta_3^2} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & (1-\beta_1)^{-2} & 0 & 0 \\ \sum_{i=1}^\infty \beta_1^{i-1}\frac{\partial f_{t-i}(\boldsymbol{\beta}_\tau)}{\partial \beta_0} & \sum_{i=2}^\infty (i-1)\beta_1^{i-2}f_{t-i}(\boldsymbol{\beta}_\tau) + \sum_{i=1}^\infty \beta_1^{i-1}\frac{\partial f_{t-i}(\boldsymbol{\beta}_\tau)}{\partial \beta_1} & \sum_{i=1}^\infty \beta_1^{i-1}\frac{\partial f_{t-i}(\boldsymbol{\beta}_\tau)}{\partial \beta_2} & \sum_{i=1}^\infty \beta_1^{i-1}\frac{\partial f_{t-i}(\boldsymbol{\beta}_\tau)}{\partial \beta_3} \\ 0 & \sum_{i=2}^\infty (i-1)\beta_1^{i-2}(y_{t-i})^+ & 0 & 0 \\ 0 & \sum_{i=2}^\infty (i-1)\beta_1^{i-2}(y_{t-i})^- & 0 & 0 \end{bmatrix}.$$

$$(3.52)$$

Considering the rewritten form of, the gradient of, and the Hessian matrix of this CAViaR model, it might raise a caution of estimating those variables by using estimated parameters because the persistent appearance of the parameters can give a slow convergence rate. That is how in essence the nonlinearity of parameters in CAViaR models differentiates CAViaR from linear quantile regressive models.

3.B Appendix: How to Simulate CAViaR Data Generating Processes

Before estimating CAViaR models, we would like to provide a general way to simulate a time series $\{y_t\}$ of all conditional quantiles following a CAViaR specification. To generate such a CAViaR data generating process (DGP), it is required to get the information on the parameter specification for every possible quantile so that the conditional distribution of $\{y_t\}$ at each time can be constructed no matter which quantile is realized. Indeed, when studying a data set, we might be interested in the 1%-th, 5%-th, 50%-th or 95%-th conditional quantiles. For instance in the climate change literature, extreme positive events are also of interest.

This requirement also applies when generating QAR DGPs. However, simulating CAViaR models is more tedious than QAR simulations because the past conditional distributions also need to be stored over time as they serve for the CAViaR DGP simulation through the model VaR autoregressive terms each time. Let us illustrate the simulation process through an example. First, we need to specify a CAViaR DGP at all quantiles for instance of (3.4) as follows:

$$y_t = f_t(\boldsymbol{\beta}_{u_t}) = \beta_0(u_t) + \sum_{i=1}^{q} \beta_i(u_t) f_{t-i}(\boldsymbol{\beta}_{u_t}) + \sum_{j=1}^{r} \beta_{q+j}(u_t) y_{t-j},$$

where $\boldsymbol{\beta}'_{u_t} := [\beta_0(u_t), \beta_1(u_t), \ldots, \beta_p(u_t)]$ with $p = q + r$, and $\{u_t\}$ is i.i.d. in the standard uniform distribution (denoted as $\mathscr{U}(0,1)$). There is a monotonicity requirement on this model which is that $f_t(\boldsymbol{\beta}_{u_t})$ is monotonically increasing in u_t so that the τ-th quantile ($\tau \in (0,1)$) of y_t conditional on $\mathscr{F}_{t-1}$ can be expressed as $f_t(\boldsymbol{\beta}_\tau)$. The additional step before simulating $\{y_t\}_{t=1}^{T}$ is to specify the initial conditional distributions and the initial observations, i.e., $\{f_{1-i}(\boldsymbol{\beta}_\tau), \tau \in (0,1), i = 1, \ldots, q\}$ and $\{y_{1-j}, j = 1, \ldots, r\}$. For example, we can take $f_{1-i}(\boldsymbol{\beta}_\tau) = F_{N(0,1)}^{-1}(\tau)$ for any $\tau \in (0,1), i = 1, \ldots, q$ and $y_{1-j} = 0$ for $j = 1, \ldots, r$, where $F_{N(0,1)}^{-1}(\tau)$ is denoted as the inverse function of the standard normal distribution.

With the above set-up, we can start the simulation by following the steps below.

Step 1: Simulate a sequence of $\{u_t\}_{t=1}^{T}$ independently and identically distributed (i.i.d.) in $\mathscr{U}(0,1)$. u_t indicates that y_t is realized as its conditional u_t-th quantile.

Step 2: At time $t = 1$, y_t is realized as its u_t-th quantile which is equal to

$$f_t(\boldsymbol{\beta}_{u_t}) = \beta_0(u_t) + \sum_{i=1}^{q} \beta_i(u_t) f_{t-i}(\boldsymbol{\beta}_{u_t}) + \sum_{j=1}^{r} \beta_{q+j}(u_t) y_{t-j}.$$

Step 3: Store $\{f_t(\boldsymbol{\beta}_{u_{t+k}})\}_{k=1}^{T}$ by

$$f_t(\boldsymbol{\beta}_{u_{t+k}}) = \beta_0(u_{t+k}) + \sum_{i=1}^{q} \beta_i(u_{t+k}) f_{t-i}(\boldsymbol{\beta}_{u_{t+k}}) + \sum_{j=1}^{r} \beta_{q+j}(u_{t+k}) y_{t-j}.$$

This step serves for generating $\{y_{t+k}\}_{k=1}^{T}$ later. For instance, $y_{t+k} = f_{t+k}(\boldsymbol{\beta}_{u_{t+k}})$ is generated via the information on $f_{t+k-i}(\boldsymbol{\beta}_{u_{t+k}}), i = 1,\ldots,q$. Iteratively, it requires the conditional u_{t+k}-th quantiles of $\{y_{t+k-i}\}_{i=1}^{t+k-1}$ to be stored for generating y_{t+k}.

Step 4: Repeat Step 2 and 3 for $t = 2,3,\ldots,T$ until we get $\{y_t\}_{t=1}^{T}$.

Step 5: In order to leave out the influence of the given initial values in this simulation, we have to delete the observations in the burn-in period. We delete the first 200 observations and keep the rest $\{y_t\}_{t=201}^{T}$ as a suitable sample for studying the DGP (3.4).

The above simulation procedure can be easily adapted to other CAViaR DGPs of which model equations of $f_t(\boldsymbol{\beta}_\tau)$ can be substituted into Step 2 with observed values of any involved predetermined variables.

3.C Appendix: Proofs

3.C.1 Proof of Theorem 6

Proof.
First, since expectation is a linear function, we can rewrite

$\mathbb{E}_{y_t,\widehat{\boldsymbol{\beta}}}\left[\widehat{h}_t(0|\mathscr{F}_{t-1})\Big|\mathscr{F}_{t-1}\right]$ as follows:

$$
\mathbb{E}_{y_t,\widehat{\boldsymbol{\beta}}}\left[\widehat{h}_t(0|\mathscr{F}_{t-1})\Big|\mathscr{F}_{t-1}\right]
$$

$$
= n^{-1}\sum_{i=1}^{n}\mathbb{E}_{y_t,\widehat{\boldsymbol{\beta}}}\left[\frac{\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})+\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)\right\}}-\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})\right\}}}{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)}\Bigg|\mathscr{F}_{t-1}, y_t \neq f_t(\widehat{\boldsymbol{\beta}})\right]
$$

$$
= n^{-1}\sum_{i=1}^{n}\mathbb{E}_{y_t,\widehat{\boldsymbol{\beta}}}\left[\frac{\mathbb{1}_{\left\{0<y_t-f_t(\widehat{\boldsymbol{\beta}})\leq \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)\right\}}}{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)}\Bigg|\mathscr{F}_{t-1}\right]
$$

$$
+ n^{-1}\sum_{i=1}^{n}\mathbb{E}_{y_t,\widehat{\boldsymbol{\beta}}}\left[\frac{\mathbb{1}_{\left\{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)<y_t-f_t(\widehat{\boldsymbol{\beta}})<0\right\}}}{-\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)}\Bigg|\mathscr{F}_{t-1}\right].
$$

$$(3.53)$$

This equality holds when n goes to infinity by applying the dominated convergence theorem as we regard the least $(p+1)$ absolute residuals in $\{|y_t - f_t(\widehat{\boldsymbol{\beta}})|\}_{t=1}^{T}$ as zeros. Denote $\widehat{\varepsilon}_t := y_t - f_t(\widehat{\boldsymbol{\beta}})$. We rank $\{|\widehat{\varepsilon}_t|\}_{t=1}^{T}$ from the smallest to largest into $\{|\widehat{\varepsilon}|_{(1)}, \cdot, |\widehat{\varepsilon}|_{(T)}\}$. In fact, iterations of a simplex-based direct search method like the Nelder–Mead method for optimizing $(p+1)$ parameters terminates at the vertices of a simplex in the parameter space (Lagarias et al., 1998). That is to say, the iterations in optimizing the τ-th quantile regression objective function terminate with $(p+1)$ elements of $\{(\tau - \mathbb{1}_{\{y_t-f_t(\boldsymbol{\beta})<0\}})(y_t - f_t(\boldsymbol{\beta}))\}$ solved to be zeros. Therefore, we set $\widehat{h}_t(0|\mathscr{F}_{t-1})$ at $\widehat{\varepsilon}|_{(1)},\ldots,|\widehat{\varepsilon}|_{(p+1)}$. And

$$
|\widehat{h}_t(0|\mathscr{F}_{t-1})| \leq \frac{1}{|\widehat{\varepsilon}|_{(p+2)}} < \infty, \tag{3.54}
$$

where $|\widehat{\varepsilon}|_{(p+2)} \neq 0$ for a well-defined convex function minimization.

Since $\{\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\}_{i=1}^{n}$ is i.i.d in $N(\boldsymbol{0},\boldsymbol{V_d})$ with restriction to $\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right) \neq 0$, we can get that for each $t \in \{1,\ldots,T\}$,

$$
\left\{\mathbb{E}_{y_t,\widehat{\boldsymbol{\beta}}}\left[\frac{\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})+\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)\right\}}-\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})\right\}}}{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)}\Bigg|\mathscr{F}_{t-1}, y_t \neq f_t(\widehat{\boldsymbol{\beta}}), y_t \neq f_t(\widehat{\boldsymbol{\beta}})\right]\right\}_{i=1}^{n}
$$

is a sequence of independent random variables with finite second moments by the assumption of $\|\nabla' f_t(\widehat{\boldsymbol{\beta}})\| \leq \boldsymbol{F_0} < \infty$ (see Assumption AN1(a) of Engle and Manganelli (2004)). Then we can use Kolmogorov's strong Law of Large Number(see e.g. White, 2014, Corollary 3.9) and get that

$$
\mathbb{E}_{y_t,\widehat{\boldsymbol{\beta}}}\left[\widehat{h}_t(0|\mathscr{F}_{t-1})\Big|\mathscr{F}_{t-1}\right] \xrightarrow{a.s.} \mathbb{E}_{y_t,\widehat{\boldsymbol{\beta}},b_i}\left[\frac{\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})+\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)\right\}}-\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})\right\}}}{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i-\widehat{\boldsymbol{\beta}}\right)}\Bigg|\mathscr{F}_{t-1}, y_t \neq f_t(\widehat{\boldsymbol{\beta}})\right],
$$

$$(3.55)$$

as $n \to \infty$ conditionally on $\mathscr{F}_{t-1}$. And we can further get that

$$
\mathbb{E}_{y_t, \widehat{\boldsymbol{\beta}}, b_i} \left[\frac{\mathbb{1}_{\left\{ y_t \le f_t(\widehat{\boldsymbol{\beta}}) + \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right) \right\}} - \mathbb{1}_{\left\{ y_t \le f_t(\widehat{\boldsymbol{\beta}}) \right\}}}{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right)} \middle| \mathscr{F}_{t-1}, y_t \ne f_t(\widehat{\boldsymbol{\beta}}) \right]
$$

$$
= \mathbb{E}_{\widehat{\boldsymbol{\beta}}, b_i} \left[\mathbb{E}_{y_t} \left[\frac{\mathbb{1}_{\left\{ 0 < y_t - f_t(\widehat{\boldsymbol{\beta}}) \le \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right) \right\}}}{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right)} \middle| \mathscr{F}_{t-1} \right] \middle| \mathscr{F}_{t-1} \right]
$$

$$
+ \mathbb{E}_{\widehat{\boldsymbol{\beta}}, b_i} \left[\mathbb{E}_{y_t} \left[\frac{\mathbb{1}_{\left\{ \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right) < y_t - f_t(\widehat{\boldsymbol{\beta}}) < 0 \right\}}}{-\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right)} \middle| \mathscr{F}_{t-1} \right] \middle| \mathscr{F}_{t-1} \right]
$$

$$
= \mathbb{E}_{\widehat{\boldsymbol{\beta}}, b_i} \left[\frac{F_t\left(f_t(\widehat{\boldsymbol{\beta}}) + \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right)\right) - F_t\left(f_t(\widehat{\boldsymbol{\beta}})\right)}{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right)} \middle| \mathscr{F}_{t-1}, \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right) > 0 \right]
$$

$$
+ \mathbb{E}_{\widehat{\boldsymbol{\beta}}, b_i} \left[\frac{F_t\left(f_t(\widehat{\boldsymbol{\beta}})\right) - F_t\left(f_t(\widehat{\boldsymbol{\beta}}) + \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right)\right)}{-\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right)} \middle| \mathscr{F}_{t-1}, \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right) < 0 \right]
$$

$$
= \mathbb{E}_{\widehat{\boldsymbol{\beta}}, b_i} \left[\frac{F_t'\left(f_t(\widehat{\boldsymbol{\beta}})\right) \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right) + \mathscr{O}_p\left(\left(b_i - \widehat{\boldsymbol{\beta}}\right)' \nabla f_t(\widehat{\boldsymbol{\beta}}) \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right) \right)}{\nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right)} \middle| \mathscr{F}_{t-1} \right]
$$

$$
= \mathbb{E}_{\widehat{\boldsymbol{\beta}}, b_i} \left[F_t'\left(f_t(\widehat{\boldsymbol{\beta}})\right) + \mathscr{O}_p\left(\nabla' f_t(\widehat{\boldsymbol{\beta}})(b_i - \widehat{\boldsymbol{\beta}}) \right) \middle| \mathscr{F}_{t-1} \right]
$$

$$
= \mathbb{E}_{\widehat{\boldsymbol{\beta}}, b_i} \left[F_t'\left(f_t(\widehat{\boldsymbol{\beta}})\right) \middle| \mathscr{F}_{t-1} \right] \xrightarrow{T \to \infty} h_t(0|\mathscr{F}_{t-1}),
$$

$$
\tag{3.56}
$$

where the last two lines are obtained by Taylor's expansion for $F_t\left(f_t(\widehat{\boldsymbol{\beta}}) + \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(b_i - \widehat{\boldsymbol{\beta}}\right)\right)$ at $f_t(\widehat{\boldsymbol{\beta}})$ and by knowing $b_i - \widehat{\boldsymbol{\beta}} = o_p(1)$ and $\lim_{T \to \infty} \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^o$ with $F_t'(f_t(\cdot))$ being a continuous function (see AN1 and AN2 of (Engle and Manganelli, 2004)) respectively.

Therefore, we have $\mathbb{E}_{y_t}\left[\widehat{h}_t(0|\mathscr{F}_{t-1}) \middle| \mathscr{F}_{t-1}) \right] - h_t(0|\mathscr{F}_{t-1}) = o_p(1)$ and conclude this proof. $\square$

3.C.2 Proof of Corollary 7

Proof.
From Theorem 6, we can obtain that

$$
\mathbb{E}\left[\frac{1}{T} \sum_{t=1}^{T} \left(\widehat{h}_t(0|\mathscr{F}_{t-1}) - h_t(0|\mathscr{F}_{t-1}) \right) \right] = o_p(1) \tag{3.57}
$$

because $\mathbb{E}\left[\widehat{h}_t(0|\mathscr{F}_{t-1}) - h_t(0|\mathscr{F}_{t-1})\right] = o_p(1)$ when $n \to \infty$. And

$$
\frac{1}{T^2} \sum_{t=1}^{T} \mathbb{E}\left[\left(\widehat{h}_t(0|\mathscr{F}_{t-1}) - h_t(0|\mathscr{F}_{t-1})\right)^2\right]
$$
$$
= \frac{1}{T^2} \sum_{t=1}^{T} \mathbb{E}\left[\widehat{h}_t^2(0|\mathscr{F}_{t-1}) - h_t^2(0|\mathscr{F}_{t-1})\right] + o_p(1). \tag{3.58}
$$

Denote $\widehat{h}_{t,i} := \dfrac{\mathbb{1}_{\left\{0 < y_t - f_t(\widehat{\boldsymbol{\beta}}) \le \nabla' f_t(\widehat{\boldsymbol{\beta}})(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}})\right\}}}{\nabla' f_t(\widehat{\boldsymbol{\beta}})(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}})} + \dfrac{\mathbb{1}_{\left\{\nabla' f_t(\widehat{\boldsymbol{\beta}})(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}) < y_t - f_t(\widehat{\boldsymbol{\beta}}) < 0\right\}}}{-\nabla' f_t(\widehat{\boldsymbol{\beta}})(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}})}, i = 1,\ldots,n.$ We can derive that

$$
\mathbb{E}\left[\widehat{h}_t^2(0|\mathscr{F}_{t-1})\right] = \mathbb{E}\left[\frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \widehat{h}_{t,i}\widehat{h}_{t,j}\right] + \mathbb{E}\left[\frac{1}{n^2} \sum_{i=1}^{n} \widehat{h}_{t,i}^2\right]
$$
$$
= \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}\left[\widehat{h}_{t,j}\widehat{h}_{t,i}\right] + \mathbb{E}\left[\frac{1}{n^2} \sum_{i=1}^{n} \widehat{h}_{t,i}^2\right]
$$
$$
= \frac{2}{n^2} \frac{n(n-1)}{2} \mathbb{E}\left[F_t'^2\left(f_t(\widehat{\boldsymbol{\beta}})\right) + \mathscr{O}_p\left(\nabla' f_t(\widehat{\boldsymbol{\beta}})(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}})\right)\right] + \mathbb{E}\left[\frac{1}{n^2} \sum_{i=1}^{n} \widehat{h}_{t,i}^2\right]
$$
$$
\xrightarrow{T \to \infty} \mathbb{E}\left[h_t^2(0|\mathscr{F}_{t-1})\right] + \frac{1}{n}\mathbb{E}\left[h_t^2(0|\mathscr{F}_{t-1})\right] + o_p(1) + \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}\left[\widehat{h}_{t,i}^2\right]
$$
$$
= \mathbb{E}\left[h_t^2(0|\mathscr{F}_{t-1})\right] + \mathscr{O}_p(\frac{1}{n}), \tag{3.59}
$$

which follows the reasoning of (3.56), and herein the last line is obtained by knowing $\{h_t(0|\mathscr{F}_t)\}$ and $\{\widehat{h}_t(0|\mathscr{F}_t)\}$ is uniformly bounded by a finite constant according to Assumption AN2 of Engle and Manganelli (2004) and (3.54) respectively. Now substitute (3.59) back to (3.58) and get that

$$
\frac{1}{T^2} \sum_{t=1}^{T} \mathbb{E}\left[\left(\widehat{h}_t(0|\mathscr{F}_{t-1}) - h_t(0|\mathscr{F}_{t-1})\right)^2\right]
$$
$$
= \frac{1}{T^2} \sum_{t=1}^{T} \mathbb{E}\left[h_t^2(0|\mathscr{F}_{t-1}) - h_t^2(0|\mathscr{F}_{t-1})\right] + \mathscr{O}_p(\frac{1}{Tn}) \tag{3.60}
$$
$$
= \mathscr{O}_p(\frac{1}{Tn}),
$$

which leads to

$$
\frac{1}{T^2} \sum_{t=1}^{T} \mathbb{E}\left[\left(\widehat{h}_t(0|\mathscr{F}_{t-1}) - h_t(0|\mathscr{F}_{t-1})\right)^2\right] \xrightarrow{P} 0 \tag{3.61}
$$

when $T, n \to \infty$. Therefore, we obtain the mean square convergence (3.41) for the mean of the adaptive random bandwidth estimator sequence $\{\widehat{h}_t(0|\mathscr{F}_{t-1})\}$.

$\square$

3.C.3 Proof of Corollary 8

Proof.
From the condition (3.39), we can know that

$$\nabla' f_t(\widehat{\boldsymbol{\beta}}) \left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right) \overset{i.i.d.}{\sim} N(0, \frac{1}{T} \nabla' f_t(\widehat{\boldsymbol{\beta}}) \boldsymbol{V}_d \nabla f_t(\widehat{\boldsymbol{\beta}})), \qquad i = 1, \ldots, n,$$

and

$$\left| \nabla' f_t(\widehat{\boldsymbol{\beta}}) \left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right) \right| \neq 0.$$

Denote the probability distribution function of $\nabla' f_t(\widehat{\boldsymbol{\beta}}) \left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)$ as $F_\nabla(\cdot)$, $\delta_\nabla :=$ $\sqrt{\frac{\nabla' f_t(\widehat{\boldsymbol{\beta}}) \boldsymbol{V}_d \nabla f_t(\widehat{\boldsymbol{\beta}})}{T}}$ and $\widehat{\varepsilon}_t := y_t - f_t(\widehat{\boldsymbol{\beta}})$.

From (3.55) in the proof of Theorem 6, we know that

$$\widehat{h}_t(0|\mathscr{F}_{t-1}) \overset{n\to\infty}{\longrightarrow} \mathbb{E}_{\boldsymbol{b}_i} \left[\frac{\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}}) + \nabla' f_t(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)\right\}} - \mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})\right\}}}{\nabla' f_t(\widehat{\boldsymbol{\beta}}) \left(\boldsymbol{b}_i - \widehat{\boldsymbol{\beta}}\right)} \Bigg| \mathscr{F}_{t-1}, y_t \neq f_t(\widehat{\boldsymbol{\beta}}) \right]$$

$$= \int_{\mathbb{R}\setminus[-|\widehat{\varepsilon}_t|, |\widehat{\varepsilon}_t|)} \frac{\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}}) + x\right\}} - \mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})\right\}}}{x} \, dF_\nu(x).$$

$$(3.62)$$

We can further rewrite (3.62) based on two cases in $\widehat{\varepsilon}_t$, namely $\widehat{\varepsilon}_t > 0$, $\widehat{\varepsilon}_t < 0$ since $\widehat{h}_t(0|\mathscr{F}_{t-1})$ is set to be zero in ARB when $\widehat{\varepsilon}_t = 0$.

When $\widehat{\varepsilon}_t > 0$, we get

$$\widehat{h}_t(0|\mathscr{F}_{t-1}) = \int_{\mathbb{R}\setminus[-|\widehat{\varepsilon}_t|, |\widehat{\varepsilon}_t|)} \frac{\mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}}) + x\right\}} - \mathbb{1}_{\left\{y_t \leq f_t(\widehat{\boldsymbol{\beta}})\right\}}}{x} \, dF_\nu(x)$$

$$= \int_{\mathbb{R}\setminus[-|\widehat{\varepsilon}_t|, |\widehat{\varepsilon}_t|)} \frac{\mathbb{1}_{\{\widehat{\varepsilon}_t \leq x\}}}{x} \, dF_\nu(x) \qquad (3.63)$$

$$= \int_{\widehat{\varepsilon}_t}^{\infty} \frac{1}{x} \frac{1}{\delta_{\nabla_t} \sqrt{2\pi}} e^{-\frac{x^2}{2\delta_\nabla^2}} \, dx.$$

Substitute $u := \frac{x^2}{2\delta_{\hat{\nabla}_t}^2}$ into (3.63) and get

$$
\begin{aligned}
\widehat{h}_t(0|\mathscr{F}_{t-1}) &= \frac{1}{\delta_{\hat{\nabla}_t}\sqrt{2\pi}} \frac{1}{2} \int_{\frac{\hat{\varepsilon}_t^2}{2\delta_{\hat{\nabla}_t}^2}}^{\infty} \frac{e^{-u}}{u} \, du \\
&= \frac{1}{2\delta_{\hat{\nabla}_t}\sqrt{2\pi}} E_1\left(\frac{\hat{\varepsilon}_t^2}{2\delta_{\hat{\nabla}_t}^2}\right),
\end{aligned}
\tag{3.64}
$$

where $E_1(s) := \int_s^\infty x^{-1}e^{-x}dx$ is a special integral known as the exponential integral or the incomplete gamma function $\Gamma(0,s)$.

Analogously, when $\hat{\varepsilon}_t < 0$, we can also get

$$
\widehat{h}_t(0|\mathscr{F}_{t-1}) = \frac{1}{2\delta_{\hat{\nabla}_t}\sqrt{2\pi}} E_1\left(\frac{\hat{\varepsilon}_t^2}{2\delta_{\hat{\nabla}_t}^2}\right).
\tag{3.65}
$$

Therefore, we conclude this proof. □

3.C.4 Proof of Theorem 9

Proof.
Denote that

$$
\bar{D}_T := T^{-1}\sum_{t=1}^{T} h_t(0|\mathscr{F}_{t-1})\nabla' f_t(\widehat{\boldsymbol{\beta}}_\tau)\nabla f_t(\widehat{\boldsymbol{\beta}}_\tau).
\tag{3.66}
$$

Note that

$$
\widehat{D}_T^{arb} - D_T = \widehat{D}_T^{arb} - \bar{D}_T + \bar{D}_T - D_T.
\tag{3.67}
$$

It is straightforward to get that

$$
\widehat{D}_T^{arb} - \bar{D}_T = o_p(1),
\tag{3.68}
$$

since we know that

$$
\frac{1}{T}\sum_{t=1}^{T}\widehat{h}_t(0|\mathscr{F}_{t-1}) - \frac{1}{T}\sum_{t=1}^{T} h_t(0|\mathscr{F}_{t-1}) = o_p(1)
$$

from $\frac{1}{T}\sum_{t=1}^{T}\widehat{h}_t(0|\mathscr{F}_{t-1}) \xrightarrow{m.s.} \frac{1}{T}\sum_{t=1}^{T} h_t(0|\mathscr{F}_{t-1})$ given in Corollary 7 with $\{\nabla f_t(\boldsymbol{\beta})\nabla' f_t(\boldsymbol{\beta})\}$ being uniformly bounded in $\mathbb{R}^{p+1}$ by Assumption AN1 of Engle and Manganelli (2004). And

$$
\bar{D}_T - D_T = o_p(1),
\tag{3.69}
$$

since that $\widehat{A}_T - A_T \xrightarrow{p} 0$ which has been proved in Theorem 3 of Engle and Manganelli (2004) and $\{h_t(\cdot|\mathscr{F}_{t-1})\}$ is uniformly bounded by a finite constant according to Assumption AN2 of Engle and Manganelli (2004).

Therefore, we have that $\widehat{D}_T^{arb} - D_T = o_p(1)$ and conclude this proof. $\square$
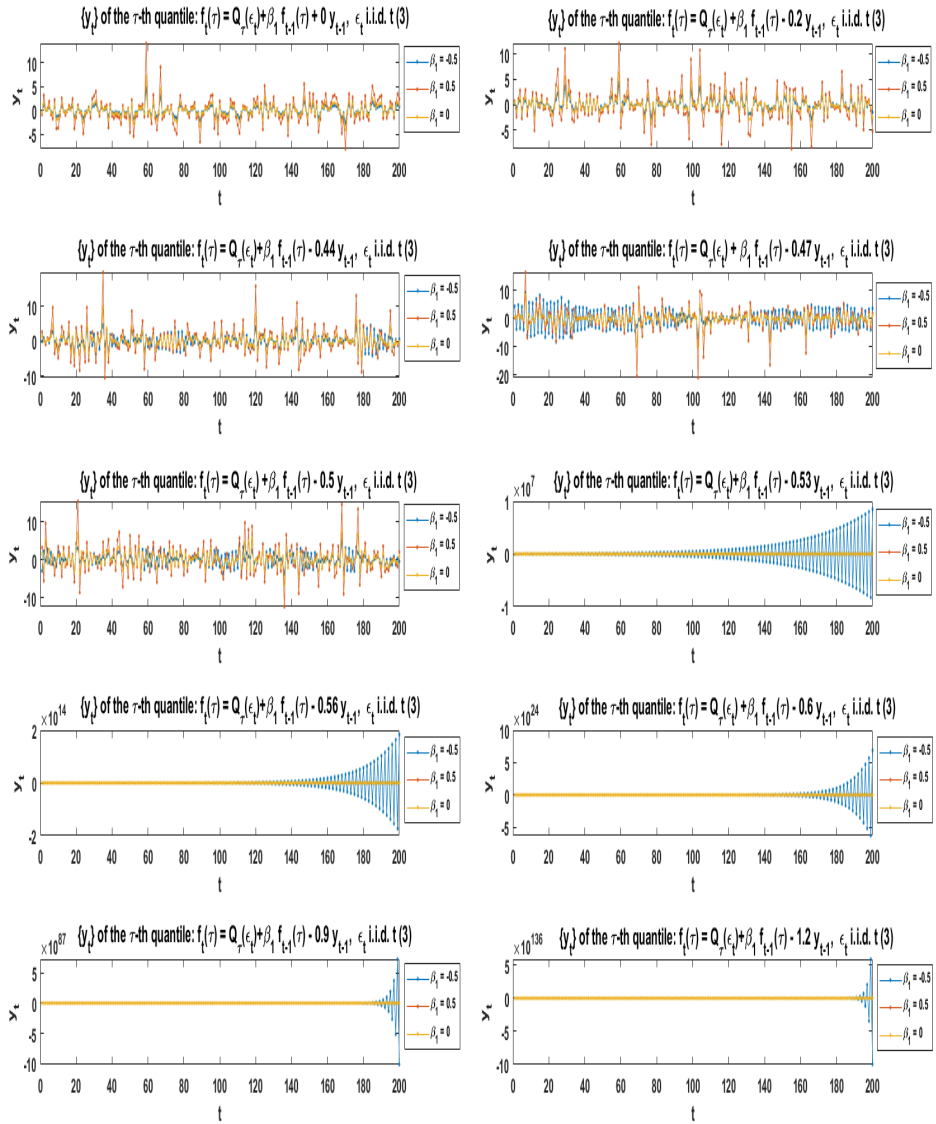
3.D Appendix: Extra Figures

Figure 3.3: Time series plots of CAViaR DGP samples for illustration
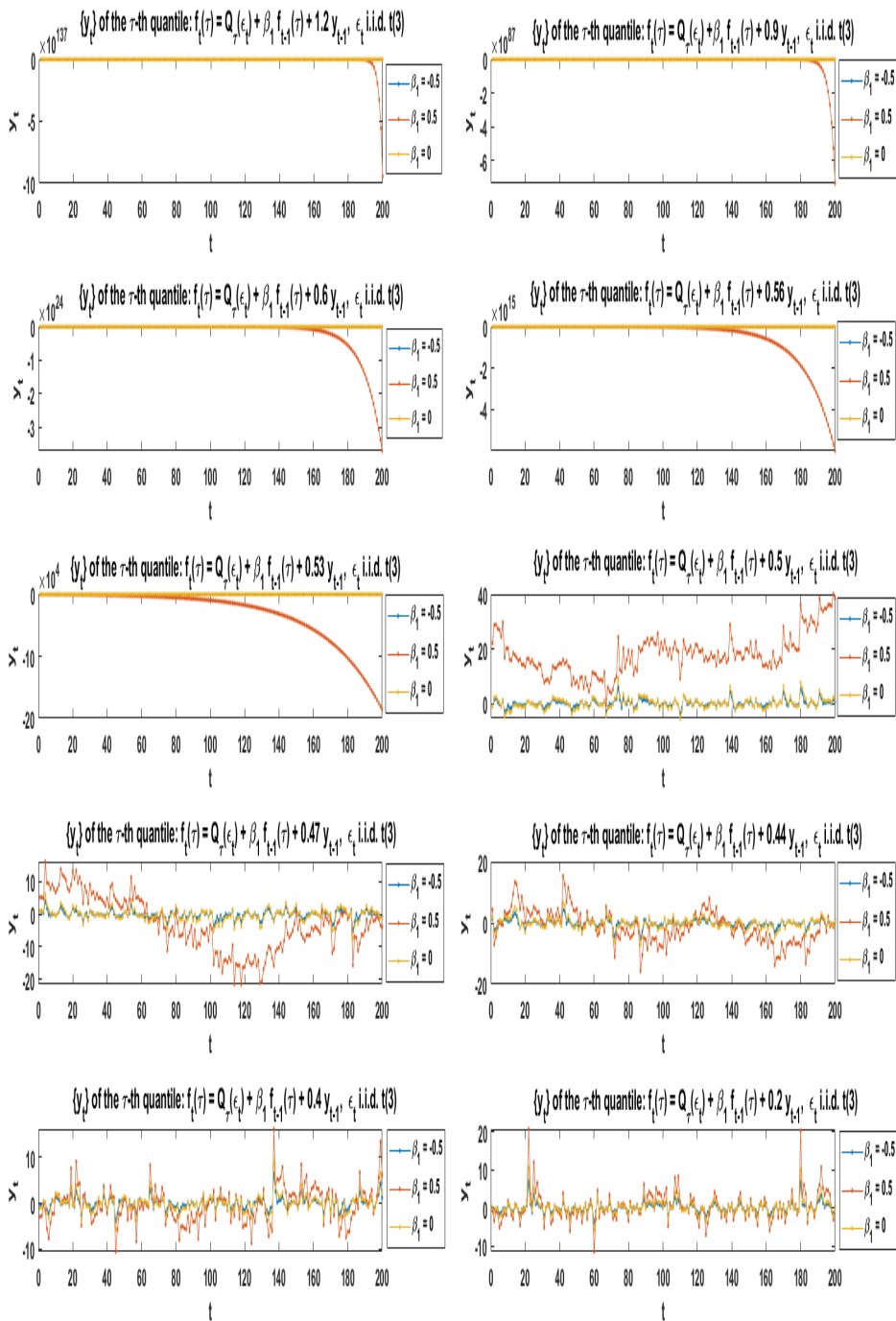
Figure 3.4: Time series plots of CAViaR DGP samples for illustration

3.E Appendix: Extra Test Results

- Simulate 1000 samples from the following DGP:

$$
\begin{aligned}
y_t = f_t(\boldsymbol{\beta}_{u_t}^{R4}) &= \beta_0^{R4}(u_t) + \beta_1^{R4}(u_t) f_{t-1}(\boldsymbol{\beta}_\tau^{R4}) + \beta_2^{R4}(u_t)\,|y_{t-1}| \\
&= \beta_0^{R4}(u_t) + \beta_1^{R4}(u_t) f_{t-1}(\boldsymbol{\beta}_{u_t}^{R4}) + \beta_2^{R4}(u_t)\,(y_{t-1})^+ + \beta_2^{R4}(u_t)\,(y_{t-1})^- ,
\end{aligned}
\tag{3.70}
$$

where $\{u_t\} \overset{i.i.d.}{\sim} \mathscr{U}(0,1)$ and the underlying parameters change over u_t as follows:

$$
\begin{cases}
\beta_0^{R4}(u_t) = \begin{cases}
3\,F_{N(0,1)}^{-1}, & 0 < u_t \le 0.4; \\[4pt]
F_{N(0,1)}^{-1}, & 0.4 < u_t \le 0.6; \\[4pt]
2\,F_{N(0,1)}^{-1}, & 0.6 < u_t < 1,
\end{cases} \\[12pt]
\beta_1^{R4}(u_t) = 0.2, & 0 < u_t < 1, \\[4pt]
\beta_2^{R4}(u_t) = 0.3, & 0 < u_t < 1,
\end{cases}
\tag{3.71}
$$

where $F_{N(0,1)}^{-1}(\cdot)$ is the inverse standard normal probability distribution function. Conditional 5%-th, 30%-th, 50%-th quantiles are estimated for each of the total 1000 simulated samples of sample size T by regressing the sample onto the full model (3.25). The results of the Wald test using the adaptive random bandwidth method and the kernel method (3.22) are listed in Table 3.8 in which each estimated size is obtained by the percentage rejection rate among the 1000 samples of sample size T.

Table 3.8: The size performances of the Wald test on the restricted model (3.26) to (3.25) ($\boldsymbol{\beta}_{u_t}^{R4} = [F_{N(0,1)}^{-1}(u_t), 0.2, 0.3]'$, $R = [0,0,1,-1]$)

quantile index τ & sample size T	methods	size: $\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
$\tau = 0.05, T = 5000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.016	0.054	0.091	0.179
	$\widehat{D}_T^{ker}$	0.026	0.066	0.126	0.21
$\tau = 0.05, T = 2000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.024	0.08	0.134	0.228
	$\widehat{D}_T^{ker}$	0.036	0.107	0.176	0.288
$\tau = 0.3, T = 5000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.01	0.045	0.085	0.168
	$\widehat{D}_T^{ker}$	0.011	0.053	0.095	0.182
$\tau = 0.3, T = 2000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.015	0.049	0.085	0.192
	$\widehat{D}_T^{ker}$	0.009	0.036	0.091	0.197
$\tau = 0.5, T = 5000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.014	0.056	0.087	0.18
	$\widehat{D}_T^{ker}$	0	0	0.001	0.026
$\tau = 0.5, T = 2000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.007	0.041	0.076	0.157
	$\widehat{D}_T^{ker}$	0	0	0	0.006

- Simulate 1000 samples of the DGP $\{y_t\}$ specified as the model (3.26) with

the underlying parameters are given as $\boldsymbol{\beta}^{R1}_{u_t} = [F^{-1}_{N(0,1)}(u_t), 0.2, 0.3]'$, where $\{u_t\} \overset{i.i.d.}{\sim} \mathscr{U}(0,1)$ and $F^{-1}_{N(0,1)}(\cdot)$ is the inverse standard normal probability distribution function. Conditional 5%-th, 30%-th, 50%-th quantiles are estimated for each of the total 1000 simulated samples of sample size T by regressing the sample onto the full model (3.25). The results of the Wald test using the adaptive random bandwidth method and the kernel method (3.22) are listed in Table 3.9 in which each estimated size is obtained by the percentage rejection rate among the 1000 samples of sample size T.

Table 3.9: The size performances of the Wald test on the restricted model (3.26) to (3.25) ($\boldsymbol{\beta}^{R1}_{u_t} = [F^{-1}_{N(0,1)}(u_t), 0.2, 0.3]'$, $R = [0,0,1,-1]$)

quantile index τ & sample size T	methods	size: $\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
$\tau = 0.05, T = 4000$	$\widehat{D}^{arb}_T$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.018	0.06	0.101	0.187
	$\widehat{D}^{ker}_T$	0.017	0.064	0.131	0.235
$\tau = 0.05, T = 2000$	$\widehat{D}^{arb}_T$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.019	0.054	0.107	0.19
	$\widehat{D}^{ker}_T$	0.035	0.085	0.14	0.245
$\tau = 0.3, T = 4000$	$\widehat{D}^{arb}_T$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.01	0.058	0.103	0.187
	$\widehat{D}^{ker}_T$	0.013	0.053	0.103	0.202
$\tau = 0.3, T = 2000$	$\widehat{D}^{arb}_T$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.021	0.061	0.11	0.184
	$\widehat{D}^{ker}_T$	0.014	0.058	0.111	0.2
$\tau = 0.5, T = 4000$	$\widehat{D}^{arb}_T$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.014	0.062	0.126	0.221
	$\widehat{D}^{ker}_T$	0.017	0.069	0.129	0.223
$\tau = 0.5, T = 2000$	$\widehat{D}^{arb}_T$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.025	0.064	0.1	0.194
	$\widehat{D}^{ker}_T$	0.02	0.065	0.118	0.206

- Simulate 1000 samples of the DGP $\{y_t\}$ specified as the model (3.34) with the underlying parameters are given as $\boldsymbol{\beta}^{R3}_{u_t} = [F^{-1}_{N(0,1)}(u_t), 0.2, 0.3]'$, where $\{u_t\} \overset{i.i.d.}{\sim} \mathscr{U}(0,1)$ and $F^{-1}_{N(0,1)}(\cdot)$ is the inverse standard normal probability distribution function. Conditional 5%-th, 30%-th, 50%-th quantiles are estimated for each of the total 1000 simulated samples of sample size T by regressing the sample onto the full model (3.25). The results of the Wald test using the adaptive random bandwidth method and the kernel method (3.22) are listed in Table 3.10 in which each estimated size is obtained by the percentage rejection rate among the 1000 samples of sample size T.

Table 3.10: The size performances of the Wald test on the restricted model (3.34) to (3.25) $(\boldsymbol{\beta}_{u_t}^{R3} = [F_{N(0,1)}^{-1}(u_t), 0.2, 0.3]', R = [0,0,1,-1])$

quantile index τ & sample size T	methods	size: $\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.20$
$\tau = 0.05, T = 5000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.032	0.069	0.096	0.17
	$\widehat{D}_T^{ker}$	0.082	0.137	0.199	0.287
$\tau = 0.05, T = 2000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.052	0.093	0.127	0.19
	$\widehat{D}_T^{ker}$	0.143	0.221	0.271	0.341
$\tau = 0.3, T = 5000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.032	0.071	0.121	0.207
	$\widehat{D}_T^{ker}$	0.073	0.137	0.207	0.3
$\tau = 0.3, T = 2000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.031	0.063	0.123	0.204
	$\widehat{D}_T^{ker}$	0.092	0.156	0.216	0.308
$\tau = 0.5, T = 5000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.021	0.055	0.095	0.188
	$\widehat{D}_T^{ker}$	0.067	0.118	0.16	0.256
$\tau = 0.5, T = 2000$	$\widehat{D}_T^{arb}$ ($n = 10^4$, 2 times updating $\boldsymbol{V_d}$)	0.034	0.069	0.118	0.208
	$\widehat{D}_T^{ker}$	0.088	0.158	0.212	0.311

4

VAR for VaR and CoVaR

Abstract

In Chapter 4, we generalize multivariate multi-quantile CAViaR models (MVMQ-CAViaR, see White et al., 2015) by incorporating CoVaR specification (see Adrian and Brunnermeier, 2011) into the model specification. The proposed model presents a vector-autoregression (VAR) of financial institutions' value-at-risk (VaR) as well as their CoVaR. This model generalization is able to capture contemporaneous tail dependence of financial institutions and market indexes so that we can interpret the systemic risks of the institutions over time. We provide consistency and asymptotic normality of the generalized model estimator as well as some relevant inference tests. For tracing the transmission of a single shock to a financial institution in the financial system, we also construct quantile impulse response functions (QIRF) accordingly using the local projection idea (Jordà, 2005) and expansion of estimated terms. Applications to real data shows strong evidence of contemporaneous effects of big banks on the market index S&P 500, and supports this methodology.

4.1 Introduction

Value-at-Risk (VaR) is a standard risk measure for market risk management with defining risk as loss on a fixed asset over a fixed time horizon. It is widely employed in the financial industry for both internal control and regulatory reporting. Among many popular approaches for VaR estimation, quantile regressions stand out for the advantages in semi-parametric specification and numerical efficiency. The quantile regression family working for this measure has been extended from static quantile regression models (QR, see Koenker and Bassett Jr, 1978) to quantile autoregression models (QAR, see Koenker and Xiao, 2006), to conditional autoregressive value-at-risk models (CAViaR, see Engle and Manganelli, 2004), to multivariate multi-quantile CAViaR models (MVMQ-CAViaR, see White et al., 2015). This evolution can be hinted by the specification testing summarized by Chernozhukov and Umantsev (2001) with regard to checking the conditionality and the functional form validity of quantile regression models, which enlightens us to see that each model generation in the quantile regression evolution can be viewed as a consequence from the inability of its predecessors in the conditional information sufficiency or in the functional form validity.

MVMQ-CAViaR is capable of measuring the individual tail risk as well as the tail dependence of financial institutions by modelling the VaR of their stock returns in a vector-autoregressive way. We notice that only predetermined (i.e., lagged and exogenous) information is accounted in the MVMQ-CAViaR specification (White et al., 2015) which cannot measure the contemporaneous tail dependence among financial assets or cover the CoVaR specification (Adrian and Brunnermeier, 2011) for measuring systemic risk of financial institutions. As we always see clear contemporaneous comovement patterns between market portfolios and some big financial institutions, we question if MVMQ-CAViaRs are sufficient to explain this pattern.

In this chapter we propose to generalize the vector-autoregressive VaR in the generic MVMQ-CAViaR specification by incorporating the CoVaR specification of a financial market portfolio. In this chapter, we would like to see if we can find significant contemporaneous tail dependence between the financial market portfolio and some financial institutions so as to measure their systemic risk. We will also provide the estimation consistency and asymptotic normality proofs of this generalized model along with some testing methods to infer the significance of the contemporaneous tail dependence between a market portfolio/index and some big financial institutions. This model generalization also can allow us to study the links across the whole financial market network and to trace the transmission of a single shock to a financial institution in the financial system by using quantile impulse response functions (QIRF) which we will construct in this chapter accordingly with the use of the local projection idea (Jordà, 2005) and expansion of estimated terms.

The remainder of this chapter is organized as follows. In Section 4.2, we introduce the generic MVMQ-CAViaR model specification first and propose to generalize it to the vector autoregressive model of VaR and CoVaR along with the proofs on the estimation consistency and asymptotic normality of this generalized model. We also call this generalized model as *systemic MVMQ-CAViaR model*. It follows that in Section 4.2.3 some inference tests are given in order to infer the significance of contemporaneous terms in the CoVaR specification. In Section 4.3, we illustrate on applying CoVaR returned by our generalized model to measure the systemic risk of financial institutions. In Section 4.4, we construct quantile impulse response functions in correspondence to our model and apply the local projection with use of expansion of estimated terms for QIRF estimation. Some results of Monte Carlo simulations regarding systemic MVMQ-CAViaR models are presented in Section 4.5. Section 4.6 presents an empirical application to real data. Section 4.7 concludes this chapter.

4.2 Model Generalization to Systemic MVMQ-CAViaR

In Section 4.2.1, we first introduce the generic MVMQ-CAViaR model specification which is also refered to as VAR for VaR (see White et al., 2015). Being motivated to measure the systemic risk of some financial institutions and to measure the contemporaneous tail dependence among financial assets, we propose to generalize MVMQ-CAViaR models to the vector autoregressive model of VaR and CoVaR in Section 4.2.2. In Section 4.2.3 some inference tests are proposed in order to infer the significance of the contemporaneous tail dependence.

4.2.1 MVMQ-CAViaRs (VAR for VaR)

White et al. (2015) proposed the framework of multivariate multi-quantile CAViaR (MVMQ-CAViaR) models and established the theoretical validity in its application by proving its estimation consistency and asymptotic normality under some regularity conditions. In essense, this framework is based on autoregressing the VaR of multiple random variables onto their lags so as to measure their tail dependence, which can be regarded as an extension to CAViaR models which are autoregressions of univariate VaR. The generic MVMQ-CAViAR specification given by White et al. (2015) is shown below.

Suppose $\boldsymbol{Y} := [Y_1, Y_2, \ldots, Y_n]'$ is a vector of n random variables of interest, with its multivariate time series $\{\boldsymbol{Y}_t := [y_{1,t}, \ldots, y_{n,t}]'\}_{t=1}^T$. We consider a vector of explanatory exogenous variables denoted by $\boldsymbol{X}$ whose first element is one, with time series $\{\boldsymbol{X}_t\}_{t=1}^T$. We consider p quantile indexes denoted by $\theta_{i,1}, \ldots, \theta_{i,p}$ and $0 < \theta_{i,1} < \ldots, < \theta_{i,p} < 1$ for each $Y_i, i = 1, \ldots, n$. Define the information set $\mathscr{F}_t$ until time t to be the σ-algebra generated

by $Z^{(t)} := \{\boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{Y}_t, \boldsymbol{X}_{t-1}, \boldsymbol{Y}_{t-1}, \ldots\}$, i.e., $\mathscr{F}_t := \sigma(Z^{(t)}), t = 1, \ldots, T$. Denote the cumulative distribution function of $y_{i,t}$ conditional on $\mathscr{F}_{t-1}$ by $F_{it}(\cdot)$[1]. The θ_{ij}-th quantile of $y_{i,t}$ conditional on $\mathscr{F}_{t-1}$, denoted by $q_{i,j,t} := \inf\{y : F_{it}(y) \geq \theta_{ij}\}$, can be specified by a generic MVMQ-CAViaR as follows:

$$q_{i,j,t} = \Psi_t' \boldsymbol{\beta}_{ij} + \sum_{\tau=1}^{m} \boldsymbol{q}_{t-\tau}' \boldsymbol{\gamma}_{i,j,\tau}, \qquad i = 1, \ldots, n; \; j = 1, \ldots, p, \qquad (4.1)$$

where $\{\Psi_t\}$ is a sequence of $k \times 1$ variables with Ψ_t predetermined to $\boldsymbol{Y}_t$ and being $\mathscr{F}_{t-1}$-measurable. And we use the notations that $\boldsymbol{q}_{t-\tau} := [q_{1,1,t-\tau}, \ldots, q_{1,p,t-\tau}, \ldots, q_{n,1,t-\tau}, \ldots, q_{n,p,t-\tau}]'$, $\boldsymbol{\beta}_{ij} := [\beta_{i,j,1}, \ldots, \beta_{i,j,k}]'$ being a $k \times 1$ real vector, and $\boldsymbol{\gamma}_{i,j,\tau} := [\boldsymbol{\gamma}_{i,j,\tau,1}', \ldots, \boldsymbol{\gamma}_{i,j,\tau,n}']'$ with each $\boldsymbol{\gamma}_{i,j,\tau,k}$ being a $p \times 1$ real vector. Let $\boldsymbol{\gamma}_{ij} := [\boldsymbol{\gamma}_{i,j,1}', \ldots, \boldsymbol{\gamma}_{i,j,m}']'$, $\boldsymbol{\alpha}_{ij} := [\boldsymbol{\beta}_{ij}', \boldsymbol{\gamma}_{ij}']'$ and $\boldsymbol{\alpha} := [\boldsymbol{\alpha}_{11}', \ldots, \boldsymbol{\alpha}_{1p}', \ldots, \boldsymbol{\alpha}_{n1}', \ldots, \boldsymbol{\alpha}_{np}']'$, where $\boldsymbol{\alpha} \in \mathbb{A}$ and $\mathbb{A}$ is a compact set of $\mathbb{R}^{l_r}$ with $l_r := np(k + npm)$.

MVMQ-CAViaR is capable of measuring the individual tail risk as well as the tail dependence of financial institutions. However, due to the conditional limitation in MVMQ-CAViaR, MVMQ-CAViaR models cannot measure the contemporaneous tail dependence among financial assets or cover the CoVaR specification (Adrian and Brunnermeier, 2011) for measuring systemic risk of financial institutions. As we always see clear contemporaneous comovement patterns between market portfolios and some big financial institutions, we question if MVMQ-CAViaRs are sufficient to explain this pattern. This question makes sense as it is often believed that information is rapidly reflected into stock prices. Neglecting contemporaneous return spillovers probably leads to underestimation of systemic risk contributions or exposures. Appealed to the systemic feature of financial systems, in the following we are going to generalize MVMQ-CAViaR models by incorporating the CoVaR specification on a financial market portfolio return.

4.2.2 Systemic MVMQ-CAViaR (VAR for VaR and CoVaR)

The financial market can react to news rapidly and extensively. One type of news deemed to be influential on the market is regarding 'too big to fail' financial institutions due to their systemic risk. Systemic risk in the financial market is defined as the risk that an event at the company level triggers severe instability

---

[1]White et al. (2015) specified the dependence of $F_{it}(\cdot)$ on each $\omega \in \Omega$ by denoting $F_{it}(\cdot)$ more specifically as $F_{it}(\omega, \cdot)$. In this chapter, we do not specify explicitly the role of each $\omega \in \Omega$ as $\omega \in \Omega$ cannot be extracted explicitly to formulate its role and its influence can be exerted partially or fully through the conditional set $\mathscr{F}_{t-1}$ each time.

or collapse of an entire industry or even the economy. The 'too big to fail' financial institutions are highly interconnected with the financial market both directly and indirectly. The direct links can happen through contractual commitments and counterparty credit risk exposures. The indirect links are, for instance, price effects and liquidity spirals. Such interconnection the 'too big to fail' financial institutions possess in the financial market is also accompanied with their systemic risk to the whole market. With the objective to measure the systemic risk and contemporaneous tail dependence of financial institutions, we would like to generalize MVMQ-CAViaR models by incorporating the CoVaR specification (see Adrian and Brunnermeier, 2011) into the autoregressive model (4.1) as follows which presents to us a vector-autoregressive model of financial institutions' VaR as well as their CoVaR.

A generic *systemic MVMQ-CAViaR* model specification given below in (4.2) is based on the set-up of the generic MVMQ-CAViaR model specification (4.1) with generalization to measure the contemporaneous tail-dependence of response variables. It is worth mentioning that the contemporaneous dependence direction is predetermined by our expertise and belief.[2]

$$
\begin{cases}
q_{i,j,t} = \Phi_t'\boldsymbol{\beta}_{ij} + \sum_{\tau=1}^{m} \boldsymbol{q}_{t-\tau}' \boldsymbol{\gamma}_{i,j,\tau} & i=1, j=1,\ldots,p, \\[2mm]
q_{i,j,t} = \Phi_t'\boldsymbol{\beta}_{ij} + \sum_{\tau=1}^{m} \boldsymbol{q}_{t-\tau}' \boldsymbol{\gamma}_{i,j,\tau} + g_{i,j}\left(\boldsymbol{s}_{i,j}, \boldsymbol{u}_{i,t}\right), & i=2,\ldots,n, \ j=1,\ldots,p,
\end{cases}
\tag{4.2}
$$

where

$$
\boldsymbol{u}_{i,t} :=
\begin{cases}
[y_{1,t}-q_{1,1,t}, \ldots, y_{1,t}-q_{1,p,t}]', & \\
& \text{when} \quad i=2; \\
[y_{1,t}-q_{1,1,t}, \ldots, y_{1,t}-q_{1,p,t}, \ldots, y_{i-1,t}-q_{i-1,1,t}, \ldots, y_{i-1,t}-q_{i-1,p,t}]', & \\
& \text{when} \quad i=3,\ldots,n.
\end{cases}
$$

And parameters $\boldsymbol{s}_{i,j} \in \mathbb{S}_{ij}$ which are compact sets of $\mathbb{R}^{d_{ij}}$ for $i=2,\ldots,n$ and $j = 1,\ldots,p$ with d_{ij} being nonnegative integers. The mapping $g_{i,j} : \mathbb{S}_{ij} \times \mathbb{R}^{(i-1)p} \to \mathbb{R}$, accounts for the contemporaneous effect on $q_{i,j,t}$ due to the realization $\boldsymbol{u}_{i,t}$. Intuitively speaking, $g_{i,j}$ intends to capture the influence of contemporaneous news due to $y_{1,t}, \ldots, y_{i-1,t}$ onto the conditional quantiles of $y_{i,t}$ so as to measure their contemporaneous tail dependence and the systemic risk of $y_{1,t}, \ldots, y_{i-1,t}$ if Y_i is a market index return variable. Some functional form examples of $g_{i,j}$ are given below. For instance,

$$
g_{i,j}\left(\boldsymbol{s}_{i,j}, \boldsymbol{u}_{i,t}\right) = \boldsymbol{s}_{i,j}' \boldsymbol{u}_{i,t} = \sum_{k_i,k_j=1}^{k_i=i-1, k_j=p} s_{k_i,k_j}\left(y_{k_i,t}-q_{k_i,k_j,t}\right)
$$

---

[2]Identifying the contemporaneous dependence direction is beyond the scope of this chapter.

is intended to explain the comovement of $q_{i,j,t}$ with $y_{1,t}, \ldots, y_{i-1,t}$ and $\boldsymbol{u}_{i,t}$, where $\boldsymbol{s}_{i,j} = [s_{1,1}, \ldots, s_{i-1,p}] \in \mathbb{R}^{(i-1)p}$. Another example is that

$$g_{i,j}(\boldsymbol{s}_{i,j}, \boldsymbol{u}_{i,t}) = \sum_{k_i, k_j=1}^{k_i=i-1, k_j=p} s_{k_i,k_j,1} \left\{ \left[1 + \exp(s_{k_i,k_j,2}[y_{k_i,t} - q_{k_i,k_j,t}]) \right]^{-1} - \theta_{k_i,k_j} \right\}$$

is intended to explain the θ_{k_i,k_j}-th quantile violation of $y_{k_i,t}$ shifting $q_{i,j,t}$ so as to affect $y_{i,t}$.

For the ease of notations, we stack the parameters of (4.2) into $\boldsymbol{\alpha} := [\boldsymbol{\alpha}'_{11}, \ldots, \boldsymbol{\alpha}'_{1p}, \ldots, \boldsymbol{\alpha}'_{n1}, \ldots, \boldsymbol{\alpha}'_{np}]$ where $\boldsymbol{\alpha} \in \Theta := \mathbb{A} \times \mathbb{S}_{21} \times \ldots \times \mathbb{S}_{np}$, a compact set of $\mathbb{R}^{l_s}$ with $\boldsymbol{\gamma}_{ij} := [\boldsymbol{\gamma}'_{i,j,1}, \ldots, \boldsymbol{\gamma}'_{i,j,m}]', \boldsymbol{\alpha}_{ij} := [\boldsymbol{\beta}'_{ij}, \boldsymbol{\gamma}'_{ij}, \boldsymbol{s}'_{ij}]'$, $l_s := np(k+npm) + \sum_{j=1}^{p} \sum_{i=2}^{n} d_{ij}$. To estimate true parameters $\boldsymbol{\alpha}^o$, we apply the quasi-maximum likelihood method by optimizing the objective function $\bar{S}_T(\boldsymbol{\alpha})$ and obtain the quasi-maximum likelihood estimator (QMLE) $\widehat{\boldsymbol{\alpha}}$ as shown below.

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha} \in \Theta}{\operatorname{argmin}} \, \bar{S}_T(\boldsymbol{\alpha}),$$
$$\bar{S}_T(\boldsymbol{\alpha}) := T^{-1} \sum_{t=1}^{T} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{p} \rho_{\theta_{ij}}(y_{i,t} - q_{i,j,t}(\boldsymbol{\alpha})) \right\}, \tag{4.3}$$

where $\rho_\theta(\varepsilon) = \varepsilon(\theta - \mathbb{1}_{\{\varepsilon<0\}})$ is known as check function in quantile regressions. Denote $\psi_\theta := (\theta - \mathbb{1}_{\{\varepsilon<0\}})$.

To prove the consistency and asymptotic normality of the generic systemic MVMQ-CAviaR (4.2), we impose some assumptions below on the contemporaneous terms $g_{i,j}(\boldsymbol{s}_{i,j}, \boldsymbol{u}_{i,t})$ in addition to all the assumptions on MVMQ-CAViaRs given by White et al. (2015).

Assumption 10 (Contemporaneous terms).

G1: For each $\boldsymbol{s}_{i,j} \in \mathbb{S}_{ij}$, a compact sets of $\mathbb{R}^{d_{ij}}$ ($i = 2, \ldots, n$, $j = 1, \ldots, p$), $g_{i,j}(\boldsymbol{s}_{i,j}, \cdot)$ is measurable with respect to an updated information set $\mathscr{F}_{t,i-1}$ which is the σ-algebra generated by $Z^{(t,i-1)} := \{\boldsymbol{Y}_t[1:(i-1)], \boldsymbol{X}_{t+1}, \boldsymbol{X}_t, \boldsymbol{Y}_t, \boldsymbol{X}_{t-1}, \boldsymbol{Y}_{t-1}, \ldots\}$, i.e., $\mathscr{F}_{t,i-1} := \sigma(Z^{(t,i-1)}), t = 1, \ldots, T$. [3]

G2: For any $\omega \in \mathscr{F}_{t,i-1}$, $g_{i,j}(\cdot, \omega)$ is continuous on $\mathbb{S}_{ij}$, $i = 2, \ldots, n$, $j = 1, \ldots, p$.

---

[3] $\boldsymbol{Y}_t[1:(i-1)]$ are the vector containing the first $(i-1)$ elements of $\boldsymbol{Y}_t$, $i = 1, 2, \ldots, n$. And when $i = 1$, $\boldsymbol{Y}_t[1:(i-1)] = 0$.

G3: For any $1 \leq t \leq T$, $\mathbb{E}\left[\sup_{\boldsymbol{s}_{ij} \in \mathbb{S}_{ij}} g_{i,j}\left(\boldsymbol{s}_{i,j}, \boldsymbol{u}_{i,j,t}\right)\right] < \infty$, $i = 2, \ldots, n$, $j = 1, \ldots, p$.

G4: For any $\omega \in \mathscr{F}_{t,i-1}$, $g_{i,j}(\cdot, \omega)$ *is differentiable in* $\mathbb{S}_{ij}$, $i = 2, \ldots, n$; $j = 1, \ldots, p$.

Theorem 11 (Consistency). *Suppose that the assumptions of Theorem 1 of White et al. (2015) and Assumptions G1-3 hold. Then we have*

$$\widehat{\boldsymbol{\alpha}} \overset{a.s.}{\to} \boldsymbol{\alpha}^o, \tag{4.4}$$

where $\widehat{\boldsymbol{\alpha}}$ *is the quasi-likelihood estimator (QML) obtained in (4.3) for estimating the true parameter vector* $\boldsymbol{\alpha}^o$ *in the underlying systemic MVMQ-CAViaR process* $\{\boldsymbol{Y}_t\}$.

Proof. See Appendix 4.B. □

Theorem 12 (Asymptotic normality). *Suppose that the assumptions of Theorem 2 of White et al. (2015) and Assumptions G1-4 hold. Then the asymptotic distribution of the QML estimator* $\widehat{\boldsymbol{\alpha}}$ *obtained from (4.3) is as follows:*

$$\sqrt{T}\left(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^o\right) \overset{d}{\sim} N\left(\boldsymbol{0}, Q^{-1} V Q^{-1}\right), \tag{4.5}$$

where

$$\begin{aligned}
Q &:= \sum_{i=1}^{n} \sum_{j=1}^{p} \mathbb{E}\left[f_{i,j,t}(0) \nabla q_{i,j,t}(\boldsymbol{\alpha}^o) \nabla' q_{i,j,t}(\boldsymbol{\alpha}^o)\right], \\
V &:= \mathbb{E}\left[\boldsymbol{\eta}_t^o \left(\boldsymbol{\eta}_t^o\right)'\right] \\
\boldsymbol{\eta}_t^o &:= \sum_{i=1}^{n} \sum_{j=1}^{p} \mathbb{E}\left[\nabla q_{i,j,t}(\boldsymbol{\alpha}^o) \psi(\varepsilon_{i,j,t})\right], \\
\varepsilon_{i,j,t} &:= y_{i,t} - q_{i,j,t}(\boldsymbol{\alpha}^o),
\end{aligned} \tag{4.6}$$

and $f_{i,j,t}(\cdot)$ *is the continuous density of* $\varepsilon_{i,j,t}$ *conditional on* $\mathscr{F}_{t,i-1}$, *and* $\boldsymbol{\alpha}^o$ *is the true parameter vector in the underlying systemic MVMQ-CAViaR process* $\{\boldsymbol{Y}_t\}$.

Proof. See Appendix 4.B. □

Theorem 13. *Suppose that the assumptions of Theorem 3 of White et al. (2015) and Assumptions G1-4 hold. To estimate V in Theorem 12,* $\widehat{V}_T$ *is obtained by*

plugging the QMLE $\widehat{\boldsymbol{\alpha}}$ into (4.6) as follows:

$$\widehat{V}_T = \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{\eta}}_t\widehat{\boldsymbol{\eta}}_t'$$

$$\widehat{\boldsymbol{\eta}}_t = \sum_{i=1}^{n}\sum_{j=1}^{p}\nabla q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\psi(\widehat{\varepsilon}_{i,j,t}), \tag{4.7}$$

$$\widehat{\varepsilon}_{i,j,t} = y_{i,t} - q_{i,j,t}(\widehat{\boldsymbol{\alpha}}).$$

Then we have

$$\widehat{V}_T \xrightarrow{p} V. \tag{4.8}$$

Proof. See Appendix 4.B. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We can also know that the conditional probability density of $y_{i,t}$ at its conditional j-th quantile $q_{i,j,t}$ is $f_{i,j,t}(0)$, $i = 1,\ldots,n$, $j = 1,\ldots,p$, $t = 1,\ldots,T$. We apply the adaptive random bandwidth method (Hecq and Sun, 2021) to estimate $f_{i,j,t}(0)$ $(i = 1,\ldots,n, j = 1,\ldots,p, t = 1,\ldots,T)$ as follows.

Theorem 14 (Adaptive random bandwidth method).
Given the conditions and the asymptotic normality result in Theorem 12 and assuming the condition that

$$\sqrt{T}\left(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}}\right) \overset{d}{\sim} N(\mathbf{0}, \boldsymbol{I}_{l_s}), \qquad z = 1,\ldots,N,$$

with $l_s := np(k+npm) + \sum_{j=1}^{p}\sum_{i=2}^{n}d_{ij}$ and the exclusion of $\boldsymbol{\alpha}_z$ such that

$$\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\left(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}}\right) = 0,$$

we can get the following estimator of $f_{i,j,t}(0)$:

$$\widehat{f}_{i,j,t}(0) = \begin{cases} \dfrac{1}{N}\sum_{z=1}^{N}\dfrac{\mathbb{1}_{\left\{y_{i,t}\le q_{i,j,t}(\widehat{\boldsymbol{\alpha}})+\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z-\widehat{\boldsymbol{\alpha}})\right\}} - \mathbb{1}_{\left\{y_{i,t}\le q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\right\}}}{\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\left(\boldsymbol{\alpha}_z-\widehat{\boldsymbol{\alpha}}\right)}, & \text{when } y_{i,t} \ne q_{i,j,t}(\widehat{\boldsymbol{\alpha}}), \\ 0, & \text{when } y_{i,t} = q_{i,j,t}(\widehat{\boldsymbol{\alpha}}), \end{cases} \tag{4.9}$$

and

$$\mathbb{E}_{y_{i,t},\widehat{\boldsymbol{\alpha}}}\left[\widehat{f}_{i,j,t}(0)\Big|\mathscr{F}_{t,i-1}\right] \xrightarrow{p} f_{i,j,t}(0), \tag{4.10}$$

as $N \to \infty$ for $i = 1,\ldots,n$, $j = 1,\ldots,p$, $t = 1,\ldots,T$. [4]

---

[4]We regard the least l_s absolute residual elements in $\{|\widehat{\varepsilon}_{i,j,t}|, i = 1,\ldots,n; j = 1,\ldots,p; t = 1,\ldots,T.\}$ as zeros, see Theorem 6. Therefore, we set $\widehat{f}_{i,j,t}(0) = 0$ at the least $(p+1)$ absolute residual elements in $\{|\widehat{\varepsilon}_{i,j,t}|, i = 1,\ldots,n; j = 1,\ldots,p; t = 1,\ldots,T.\}$ in all the tests throughout

Theorem 15. *Suppose that the assumptions of Theorem 3 of White et al. (2015) and Assumptions G1-4 hold. To estimate Q in Theorem 12, $\widehat{Q}_T$ is obtained by plugging the QMLE $\widehat{\boldsymbol{\alpha}}$ into (4.6) as follows:*

$$\widehat{Q}_T := \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{p} \widehat{f}_{i,j,t}(0) \nabla q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) \nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}}), \tag{4.11}$$

where $\widehat{f}_{i,j,t}(0)$ is obtained by the adaptive random bandwidth method (Hecq and Sun, 2021) in Theorem 14. Then we have

$$\widehat{Q}_T \xrightarrow{p} Q. \tag{4.12}$$

Proof. See Appendix 4.B. □

Here we would like to show a very basic data generating process (DGP) example on the proposed systemic MVMQ-CAViaR (4.2). We want to use this example to compare MVMQ-CAViaR with the systemic one in their conditions on ensuring the estimation consistency as well as to see their linkage. Suppose a bivariate time series $\{\boldsymbol{Y}_t := [y_{1,t}, y_{2,t}]'\}_{t=1}^{T}$ follows a systemic bivariate CAViaR DGP specified as follows:

$$\begin{bmatrix} 1 & 0 \\ s_{2,1}^o & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \beta_{10}^o \\ \beta_{20}^o \end{bmatrix} + \begin{bmatrix} \beta_{11}^o & \beta_{12}^o \\ \beta_{21}^o & \beta_{22}^o \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11}^o & \gamma_{12}^o \\ \gamma_{21}^o & \gamma_{22}^o \end{bmatrix} \begin{bmatrix} q_{1,t-1} \\ q_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \tag{4.13}$$

where $\{\varepsilon_{1,t}\}$ and $\{\varepsilon_{2,t}\}$ are independently distributed to each other as well as within their own processes with $[P\{\varepsilon_{1,t} \leq 0|\mathscr{F}_{t,0}\}, P\{\varepsilon_{2,t} \leq 0|\mathscr{F}_{t,1}\}] = [\theta, \theta]$. $q_{1,t}$ and $q_{2,t}$ are the θ-th quantiles of $y_{1,t}$ and $y_{2,t}$ conditional on $\mathscr{F}_{t,0}$ and $\mathscr{F}_{t,1}$ respectively, i.e., $[P\{y_{1,t} \leq q_{1,t}|\mathscr{F}_{t,0}\}, P\{y_{2,t} \leq q_{2,t}|\mathscr{F}_{t,1}\}] = [\theta, \theta]$. Under the assumptions of Theorem 11, we can get its QML estimator $\widehat{\boldsymbol{\alpha}}^{(s)}$ consistent to its true parameter vector denoted as $\boldsymbol{\alpha}^{o(s)}$.[5]

If we ignore the contemporaneous effect and just regress $\{\boldsymbol{Y}_t := [y_{1,t}, y_{2,t}]'\}_{t=1}^{T}$ onto the bivariate CAViaR without contemporaneous terms, a correct quantile specification condition has to be imposed further as follows in order to ensure the estimation consistency in this regression.

$$\begin{bmatrix} P\{\varepsilon_{1,t} \leq 0|\mathscr{F}_{t,0}\} \\ P\{\varepsilon_{2,t} - s_{2,1} \cdot \varepsilon_{1,t} \leq 0|\mathscr{F}_{t,0}\} \end{bmatrix} = \begin{bmatrix} \theta \\ \theta \end{bmatrix} \tag{4.14}$$

---

this chapter.

[5]The subscript '(s)' is used to distinguish the parameters for systemic MVMQ-CAViaR models with the ones for MVMQ-CAViaR models which can be deemed in the reduced-form so use subscript '(r)'.

With this correct quantile specification condition, we also have $[P\{y_{1,t} \leq q_{1,t}|\mathscr{F}_{t,0}\}, P\{y_{2,t} \leq q_{2,t}|\mathscr{F}_{t,0}\}] = [\theta, \theta]$ as well as get consistent estimator $\widehat{\boldsymbol{\alpha}}^{(r)}$ to the true reduced-form one denoted as $\boldsymbol{\alpha}^{o\,(r)}$ consisting of the following elements.

$$
\begin{aligned}
\boldsymbol{\beta}^{o\,(r)} &:= \begin{bmatrix} 1 & 0 \\ -s_{2,1}^o & 1 \end{bmatrix} \boldsymbol{\beta}^{o\,(s)} = \begin{bmatrix} 1 & 0 \\ -s_{2,1}^o & 1 \end{bmatrix} \begin{bmatrix} \beta_{10}^o & \beta_{11}^o & \beta_{12}^o \\ \beta_{20}^o & \beta_{21}^o & \beta_{22}^o \end{bmatrix}, \\
\boldsymbol{\gamma}^{o\,(r)} &= \begin{bmatrix} 1 & 0 \\ -s_{2,1}^o & 1 \end{bmatrix} \boldsymbol{\gamma}^{o\,(s)} = \begin{bmatrix} 1 & 0 \\ -s_{2,1}^o & 1 \end{bmatrix} \begin{bmatrix} \gamma_{11}^o & \gamma_{12}^o \\ \gamma_{21}^o & \gamma_{22}^o \end{bmatrix}.
\end{aligned}
\tag{4.15}
$$

From this example, we can see the condition (4.14) is a linkage which realizes the transition between the bivariate CAViaR estimator and the systemic one. On the other hand, such transitions are not always applicable to ensure the estimation consistency of MVMQ CAViaR regressions on systemic MVMQ CAViaR processes. And the benefits of the systemic estimator are that we can check if there is significant contemporaneous effect to be accounted for and measure the systemic risk of Y_1 if Y_2 is a market index return variable. For inferring the significance of the contemporaneous terms in systemic MVMQ CAViaR models, we are going to provide some testing tools in the following.

4.2.3 Inference Testing

In this subsection, some inference tests are proposed in order to indicate if the contemporaneous terms in systemic MVMQ CAViaR models are significant enough so as to favour MVMQ CAViaR models or the systemic ones. There are some ways to design such tests. If we run MVMQ CAViaR regression first, it is rigours to consider if one response variable has some contemporaneous explanatory power on the conditional quantile of another response variable. Considering that, the first test proposed in Section 4.2.3 takes the way to test if the quantile coverage of the latter response variable has a significant difference between with conditioning on the contemporaneous quantile violation of the first response variable and without the contemporaneous conditioning. We can also test if the first response variable or its functional forms, like its disturbance term, has significant contemporaneous explanatory power on the latter response variable. This path is taken by the second test proposed in Section 4.2.3 based on dynamic quantile (DQ) tests developed by Engle and Manganelli (2004). DQ tests play the same role as misspecification Lagrange multiplier (LM) tests in regressions models. So we apply DQ tests as quantile model misspecification tests to check whether other variables should be considered into the model. If we run systemic MVMQ CAViaR regression first, it is also rigours to see if involved contemporaneous terms are significant enough. Wald tests typically can fulfil this testing role and are presented at last in Section 4.2.3.

Conditional Coverage Method

The conditional coverage method is commonly used for VaR backtesting. Unlike the unconditional coverage method which only focuses on the the frequency of VaR exceedance, the conditional coverage method (Christoffersen, 1998) is to perform tests on the conditional VaR coverage which is estimated by the conditional frequency of the VaR exceedance of a univariate time series. The conditional coverage in the tests is set to equal the unconditional one under the null hypothesis so that we can see if the conditional variables can significantly affect the probability of the VaR exceedance.

We adopt this method to test if the occurrence of one asset loss rate $(-y_{1t})$ exceeding its VaR_α $(-q_{1t})$ significantly affects the probability of a financial market index loss rate $(-y_{2t})$ exceeding its VaR_α $(-q_{2t})$ at the same time. To serve this purpose, the testing hypothesis is stated as follows:

$$\begin{cases} H_o: & \mathbb{E}\left[\mathbb{1}_{\{y_{2t} \leq q_{2t}\}} \big| \mathbb{1}_{\{y_{1t} \leq q_{1t}\}=1}\right] = \alpha, & \text{for all } t; \\ H_A: & \mathbb{E}\left[\mathbb{1}_{\{y_{2t} \leq q_{2t}\}} \big| \mathbb{1}_{\{y_{1t} \leq q_{1t}\}=1}\right] \neq \alpha, & \text{for all } t. \end{cases} \tag{4.16}$$

Count the frequencies of four possible outcomes of $\mathbb{1}_{\{y_{2t} \leq q_{2t}\}} \times \mathbb{1}_{\{y_{1t} \leq q_{1t}\}}$ over $t = 1, 2, \ldots, T$, and summarize it into the following table:

frequency in $\{y_t\}_{t=1}^T$	$\mathbb{1}_{\{y_{1t} \leq q_{1t}\}} = 0$	$\mathbb{1}_{\{y_{1t} \leq q_{1t}\}} = 1$
$\mathbb{1}_{\{y_{2t} \leq q_{2t}\}} = 0$	n_{00}	n_{10}
$\mathbb{1}_{\{y_{2t} \leq q_{2t}\}} = 1$	n_{01}	n_{11}

For testing the above hypothesis in VaR backtesting, the conditional coverage method provides a likelihood ratio statistic as follows:

$$\begin{aligned} LR_{cc} &= -2 \log\left(\frac{L(\alpha, T)}{L(n_{00}, n_{01}, n_{10}, n_{11}, T)}\right) \\ &= -2 \log\left(\frac{(1-\alpha)^{n_{00}+n_{10}} \alpha^{n_{01}+n_{11}}}{(1 - \frac{n_{01}}{n_{00}+n_{01}})^{n_{00}} (\frac{n_{01}}{n_{00}+n_{01}})^{n_{01}} (1 - \frac{n_{11}}{n_{10}+n_{11}})^{n_{10}} (\frac{n_{11}}{n_{10}+n_{11}})^{n_{11}}}\right) \\ &\overset{d}{\sim} \chi^2(2), \end{aligned}$$

$$\tag{4.17}$$

where $\chi^2(v)$ is denoted as the chi-square distribution with v degree(s) of freedom.

DQ Tests

Dynamic quantile (DQ) tests proposed by Engle and Manganelli (2004) are intended to test if there is significant explanatory power of some omitted variables on the conditional quantile of a response variable over time. It can also serve for testing the hypothesis (4.16). DQ tests are analogous to the specification test illustrated by Chernozhukov and Umantsev (2001) for quantile regressions. Compared with the conditional coverage method, DQ tests are more flexible in testing omitted variables and can also perform in-sample tests by taking the asymptotic distribution of QML estimator $\widehat{\boldsymbol{\alpha}}$ into account.

To test the contemporaneous explanatory power of $y_{i_c,t}$ on the $\theta_{i,j}$-th quantile $q_{i,j,t}$ of $y_{i,t}$ conditional on $\mathscr{F}_{t,0}$ with $1 \leq i_c < i < n$ and $1 \leq j \leq p$, the testing hypothesis in our DQ tests is stated as follows:

$$
\begin{cases}
H_o: & P\{\varepsilon_{i,j,t} \leq 0 | y_{i_c,t}\} = \theta \quad \text{for all } t, \\
H_A: & P\{\varepsilon_{i,j,t} \leq 0 | y_{i_c,t}\} \neq \theta \quad \text{for all } t,
\end{cases}
\tag{4.18}
$$

where $\varepsilon_{i,j,t} := y_{i,t} - q_{i,j,t}$.

For testing the above hypothesis, the in-sample and the out-of-sample DQ test statistics are given in Theorem 16 and 17 respectively along with their asymptotic distributions.

Theorem 16 (In-sample DQ test statistic). *Suppose that the assumptions of Theorem 3 of White et al. (2015) and the H_o hypothesis in* (4.18) *hold. The in-sample DQ test statistic denoted as DQ_{IS} is given below with its asymptotic distribution.*

$$
DQ_{IS} =
$$
$$
\frac{1}{T\theta(1-\theta)} \left(\sum_{t=1}^{T} y_{i_c,t} \left(\mathbb{1}_{\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\}} - \theta_{i,j} \right) \right)^2 \left(\widehat{G}_T' \widehat{Q}_T^{-1} \sum_{i=1,i\neq 2}^{n} \frac{\partial \widehat{q}_{i,j,t}}{\partial \widehat{\boldsymbol{\alpha}}} \frac{\partial \widehat{q}_{i,j,t}}{\partial \widehat{\boldsymbol{\alpha}}'} \widehat{Q}_T^{-1} \widehat{G}_T + \frac{1}{T} \sum_{t=1}^{T} \left(y_{i_c,t} - \frac{\partial q_{i,j,t}(\widehat{\boldsymbol{\alpha}})}{\partial \widehat{\boldsymbol{\alpha}}'} \widehat{Q}_T^{-1} \widehat{G}_T \right)^2 \right)^{-1}
$$
$$
\overset{d}{\sim} \chi^2(1),
\tag{4.19}
$$

where

$$
\widehat{G}_T := \frac{1}{T} \sum_{t=1}^{T} y_{i_c,t} \, \widehat{f}_{i,j,t}(0) \frac{\partial \widehat{q}_{i,j,t}}{\partial \widehat{\boldsymbol{\alpha}}},
\tag{4.20}
$$

$\widehat{f}_{i,j,t}(0)$ *is the estimate of the probability density function $f_{i,j,t}$ of $\varepsilon_{i,j,t}$ at 0 conditional on $\mathscr{F}_{t,0}$, obtained by the adaptive bandwidth method, and $\widehat{\boldsymbol{\alpha}}$, $\frac{\partial \widehat{q}_{i,j,t}}{\partial \widehat{\boldsymbol{\alpha}}}$ and $\widehat{Q}_T$ are obtained as instructed in the above theorems.*

Proof. See Appendix 4.B. □

Theorem 17 (Out-of-sample DQ test statistic). *Suppose that the assumptions of Theorem 3 of White et al. (2015) and the H_o hypothesis in (4.18) hold. Denote the number of in-sample observations as T_R and the number of out-of-sample observations as N_R. The subscript R indicates the dependence of T_R and N_R on R with the following properties:*

$$\begin{cases} \lim\limits_{R \to \infty} \dfrac{N_R}{T_R} = 0, \\ \lim\limits_{R \to \infty} T_R = \infty, \\ \lim\limits_{R \to \infty} N_R = \infty. \end{cases} \tag{4.21}$$

Then the out-of-sample DQ test statistic denoted as DQ_{OOS} is given below with its asymptotic distribution.

$$DQ_{OOS}$$
$$= \frac{1}{N_R \theta (1-\theta)} \left(\sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} \left(\mathbb{1}_{\{y_{i,t} \le q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\}} - \theta \right) \right)^2 \left(\frac{1}{N_R} \sum_{t=1}^{N_R} y_{i_c,t}^2 \right)^{-1} \tag{4.22}$$
$$\overset{d}{\sim} \chi^2(1),$$

where $q_{i,j,t}(\widehat{\boldsymbol{\alpha}})$ is obtained by plugging the QMLE $\widehat{\boldsymbol{\alpha}}$ of the in-sample $\{\boldsymbol{Y}_t\}_{t=1}^{T_R}$ into $q_{i,j,t}(\cdot)$.

Proof. See Appendix 4.B. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Wald Tests

If we run systemic MVMQ CAViaR regression first and have the concern on the significance of the involved contemporaneous terms, Wald tests are adopted here to check if the contemporaneous terms are significant enough to remain in the model.

To test the significance of the contemporaneous term $g_{i,j}(\boldsymbol{s}_{i,j}, \boldsymbol{u}_{i,t})$ in the generic model (4.2), the testing hypothesis in our Wald tests is stated as follows:

$$\begin{cases} H_o: & \boldsymbol{s}_{i,j} = \boldsymbol{0}, \\ H_A: & \boldsymbol{s}_{i,j} \ne \boldsymbol{0}. \end{cases} \tag{4.23}$$

For testing the above hypothesis, the Wald test statistic denoted by W_T is given as follows:

$$W_T = T \widehat{\boldsymbol{s}}_{i,j}' \left[R \widehat{Q}_T^{-1} \widehat{V}_T \widehat{Q}_T^{-1} R' \right]^{-1} \widehat{\boldsymbol{s}}_{i,j} \overset{d}{\sim} \chi^2(d_{ij}) \tag{4.24}$$

where R is a $d_{ij} \times l_s$ matrix indicating the location of each element of $s_{i,j}$ in $\boldsymbol{\alpha}$ such that $R\boldsymbol{\alpha} = s_{i,j}$, and $\widehat{s}_{i,j}$, $\widehat{Q}_T$ and $\widehat{V}_T$ are the estimates of $s_{i,j}$, Q and V in (4.6) respectively.

4.3 Measuring Systemic Risk via CoVaR

The European Central Bank (ECB) (2010) defines systemic risk in the financial market as a risk of financial instability so widespread that it impairs the functioning of a financial system to the point where economic growth and welfare suffer materially. CoVaR, a systemic risk measure proposed by Adrian and Brunnermeier (2011), is defined as the VaR of the financial system conditional on institutions being under distress. We generalized MVMQ-CAViaR models into the proposed systemic specification in order to incorporate the CoVaR specification on financial index return variables. Therefore, we can use systemic MVMQ-CAViaR models to measure the systemic risk of financial institutions of interest, and this application is elaborated below.

Given two asset return variables Y_1 and Y_2, $\text{CoVaR}_{\theta}^{2|\mathbb{C}(Y_1)}$ is formulated by Adrian and Brunnermeier (2011) as the θ-th quantile of the institution 2 (or the financial system) conditional on some event $\mathbb{C}(Y_1)$ of the institution 1, which fits in the following property:

$$P\left\{Y_2 \leq \text{CoVaR}_{\theta}^{2|\mathbb{C}(Y_1)} \middle| \mathbb{C}(Y_1)\right\} = \theta. \tag{4.25}$$

Suppose the bivariate time series $\{Y_t = [y_{1,t}, y_{2,t}]'\}_{t=1}^{T}$ follows the DGP as (4.13) and we run the following systemic bivariate CAViaR regression:

$$Y_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 Y_{t-1} + \boldsymbol{\gamma} q_{t-1} + \begin{bmatrix} 0 \\ s \end{bmatrix} y_{1,t}, \tag{4.26}$$

where $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\gamma}$ are 2×1, 2×2 and 2×2 parameter matrices. After the regression, we estimate $\text{CoVaR}_{\theta}^{2|Y_1}$ in use of the estimates $\widehat{\boldsymbol{\beta}}_0, \widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\gamma}}, \widehat{s}$ as follows:

$$\widehat{\text{CoVaR}}_{\theta,t}^{2|Y_1} = \widehat{\beta}_{20} + \widehat{\beta}_{21} y_{1,t-1} + \widehat{\beta}_{22} y_{2,t-1} + \widehat{\gamma}_{21} q_{1,t-1} + \widehat{\gamma}_{22} q_{2,t-1} + \widehat{s} y_{1,t}. \tag{4.27}$$

And

$$\widehat{\text{CoVaR}}_{\theta,t}^{2|y_{1,t}=\widehat{q}_{1,t}}$$
$$= \left(\widehat{\beta}_{20} + \widehat{s}\widehat{\beta}_{10}\right) + \left(\widehat{\beta}_{21} + \widehat{s}\widehat{\beta}_{11}\right) y_{1,t-1} + \left(\widehat{\beta}_{22} + \widehat{s}\widehat{\beta}_{12}\right) y_{2,t-1} + \left(\widehat{\gamma}_{21} + \widehat{s}\widehat{\gamma}_{11}\right) q_{1,t-1}$$
$$+ \left(\widehat{\gamma}_{22} + \widehat{s}\widehat{\gamma}_{12}\right) q_{2,t-1}.$$
$$\tag{4.28}$$

If we also model the conditional 50%-th quantile $q_{1,t}^{(50\%)}$ of $y_{1,t}$ into the above systemic bivariate CAViaR model and get estimate $\widehat{q}_{1,t}^{(50\%)}$. $\Delta\text{CoVaR}_{\theta,t}^{2|1}$ defined by Adrian and Brunnermeier (2011) as the part of institution 1's systemic risk that can be attributed to Y_2 can be estimated as follows:

$$\Delta\text{CoVaR}_{\theta,t}^{2|1} := \widehat{\text{CoVaR}}_{\theta,t}^{2|y_{1,t}=\widehat{q}_{1,t}} - \widehat{\text{CoVaR}}_{\theta,t}^{2|y_{1,t}=\widehat{q}_{1,t}^{(50\%)}}. \tag{4.29}$$

Unlike the situation above which we know everything for certain, in reality we do not know the true model specification on an underlying DGP or precise contemporaneous terms to be involved. Considering that, we would like to give a rigorous application procedure here for studying contemporaneous tail dependence and CoVaR in use of systemic MVMQ-CAViaRs. In general, at first we would like to run an MVMQ-CAViaR regression based upon our knowledge on the multivariate time series of our interest. If we consider over some possible contemporaneous terms on explaining the conditional quantile of a response variate, DQ tests are implemented to check the significance of their explanatory power so that we can convincingly implement the systemic MVMQ-CAViaR regression when seeing the significance. After running the systemic MVMQ-CAViaR regression, Wald tests are applied to check if all the explanatory terms in the model are significant enough to be kept. After the Wald tests, the confirmed model specification is used to measure the systemic risk of involved financial institutions via their CoVaR as well as to measure contemporaneous tail dependence of involved financial assets. We will implement the above procedure in Section 4.5 with results presented correspondingly to each step.

4.4 Quantile Impulse Response Functions

The literature on quantile impulse response functions (QIRF) is scarce. We have a brief review here. White et al. (2015) presented a concept called pseudo quantile impulse response function in order to study how a shock to the present variable y_t influences the quantile (denoted as $q_{t+h|t}$) of its future variable y_{t+h} at h-th $(h \geq 1)$ step ahead given the current information set $\mathscr{F}_t$. Actually, pseudo quantile impulse functions derived by White et al. (2015) strongly assume that the intermediate future variables $(y_{t+1},\dots,y_{t+h-1})$ right before the h-th step are fixed and not affected by the shock. Instead of fixing the intermediate future values $(y_{t+1},\dots,y_{t+h-1})$, Montes-Rojas (2019) considered quantile paths of $y_{t+1},\dots,y_{t+h-1}$ for forecasting q_{t+h}. However, the way that Montes-Rojas (2019) tackles the randomness of future quantile paths in forecasting $q_{t+h|t}$ is by fixing a specific future quantile path, such as assuming all median occurrences in the path. Although a future quantile path can be freely chosen to match some senario, the way of Montes-Rojas (2019) in

forecasting $q_{t+h|t}$ can still not adapt to distributional characteristics of $y_{t+h}|\mathscr{F}_t$ (short for y_{t+h} conditional on $\mathscr{F}_t$), let alone $q_{t+h|t}$. The local projection method proposed by Jordà (2005) for estimating mean impulse response functions is also touched upon by Montes-Rojas (2019) to linearly regress $q_{t+h|t+h-1}$ on a specific quantile path of $y_{t+1}, \ldots, y_{t+h-1}$ and variables measurable to $\mathscr{F}_t$. Chavleishvili and Manganelli (2019) still used a quantile specification of $y_t|\mathscr{F}_{t-1}$ to represent the specification of the quantile of $y_{t+h}|\mathscr{F}_t$, and obtained the QIRF by manipulating the part of intermediate disturbances into zeros. Analogously to the fixed-intermediate (White et al., 2015) or specific future quantile path (Montes-Rojas, 2019) ideas, Han et al. (2019) and Jung and Lee (2019) used expectation of intermediate variables to define quantile impulse response functions, and adopted the local projection Jordà (2005) for estimation.

We aim in this section to define quantile impulse response function in a general way which can adapt to distributional characteristics of $y_{t+h}|\mathscr{F}_t$, and then to adopt the local projection idea with expansion of estimated terms to estimate quantile impulse response functions.

Considering a multivariate time series $\{\boldsymbol{Y}_t\}$ in a DGP as (4.2), let us discuss on how to forecast $\boldsymbol{Y}_{t+h}|\mathscr{F}_t$. Without loss of generality, we take a bivariate time series $\{\boldsymbol{Y}_t := [y_{1,t}, y_{2,t}]'\}_{t=1}^T$ with its model specification as follows:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} q_{1,t-1} \\ q_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}, \quad (4.30)$$

where

$$\boldsymbol{q}_t := \begin{bmatrix} q_{1,t} \\ q_{2,t} \end{bmatrix}$$

is the $\theta-$th quantile of $\boldsymbol{Y}_t := \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}$ conditional on $\mathscr{F}_{t-1}$ which is the σ-algebra generated by $\{\boldsymbol{Y}_{t-1}, \boldsymbol{q}_{t-1}, \boldsymbol{Y}_{t-2}, \boldsymbol{q}_{t-2}, \ldots\}$ with

$$\begin{bmatrix} P\{\varepsilon_{1,t} \le 0|\mathscr{F}_{t-1}\} \\ P\{\varepsilon_{2,t} \le 0|\mathscr{F}_{t-1}\} \end{bmatrix} = \begin{bmatrix} \theta \\ \theta \end{bmatrix}. \quad (4.31)$$

Therefore,

$$\begin{bmatrix} q_{1,t} \\ q_{2,t} \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} q_{1,t-1} \\ q_{2,t-1} \end{bmatrix}. \quad (4.32)$$

Denote

$$\boldsymbol{\beta}_0 := \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} \quad \boldsymbol{\beta}_1 := \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \quad \boldsymbol{\gamma} := \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}. \quad (4.33)$$

Suppose the above model specification is known for $\{Y_t\}$. And we want to forecast the quantile (denoted as $q_{t+h|t}$) of y_{t+h} ($h \geq 1$) given $\mathscr{F}_t$ now. At the first step, we need to rewrite the specification (4.30) of Y_{t+h} by iteratively substitutions until manifesting ε_t and q_t as follows:

$$
\begin{aligned}
Y_{t+h} &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 Y_{t+h-1} + \boldsymbol{\gamma} q_{t+h-1} + \boldsymbol{\varepsilon}_{t+h} \\
&= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 (q_{t+h-1} + \boldsymbol{\varepsilon}_{t+h-1}) + \boldsymbol{\gamma} q_{t+h-1} + \boldsymbol{\varepsilon}_{t+h} \\
&= \boldsymbol{\beta}_0 + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma}) q_{t+h-1} + (\boldsymbol{\varepsilon}_{t+h} + \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t+h-1}) \\
&= \boldsymbol{\beta}_0 + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma}) (\boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 Y_{t+h-2} + \boldsymbol{\gamma} q_{t+h-2}) + (\boldsymbol{\varepsilon}_{t+h} + \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t+h-1}) \\
&= \sum_{i=1}^{h} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{i-1} \boldsymbol{\beta}_0 + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{h-1} \boldsymbol{\beta}_1 Y_t + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{h-1} \boldsymbol{\gamma} q_t + \boldsymbol{\varepsilon}_{t+h} + \sum_{i=1}^{\max\{h-1,1\}} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{i-1} \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t+h-i} \\
&= \sum_{i=1}^{h} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{i-1} \boldsymbol{\beta}_0 + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{h} q_t + \boldsymbol{\varepsilon}_{t+h} + \sum_{i=1}^{h} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{i-1} \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t+h-i},
\end{aligned}
$$
(4.34)

where $h \in \{1, 2, \ldots\}$.

It is worth mentioning that based on (4.34), there is an alternative way to rewrite Y_{t+h} into a function of $\{\varepsilon_{t+h}, \varepsilon_{t+h-1}, \ldots\}$:

$$
(I - (\boldsymbol{\beta}_1 + \boldsymbol{\gamma}) L) Y_{t+h} = \boldsymbol{\beta}_0 + (I - \boldsymbol{\gamma} L) \boldsymbol{\varepsilon}_{t+h},
$$

$$
\iff Y_{t+h} = (I - (\boldsymbol{\beta}_1 + \boldsymbol{\gamma}))^{-1} \boldsymbol{\beta}_0 + \left(\sum_{i=0}^{\infty} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^i L^i \right) (I - \boldsymbol{\gamma} L) \boldsymbol{\varepsilon}_{t+h}
$$

$$
= (I - (\boldsymbol{\beta}_1 + \boldsymbol{\gamma}))^{-1} \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_{t+h} + \sum_{i=1}^{\infty} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{i-1} \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t+h-i}
$$
(4.35)

where L is the lag operator, and it holds under the condition that the spectral radius of $(\boldsymbol{\beta}_1 + \boldsymbol{\gamma})$ is less than one . However, the first rewriting way (4.34) is more generally applicable to systemic MVMQ-CAViaR DGPs, which also reduces the number of explanatory variables in the consideration for forecasting $q_{t+h|t}$.

Following the result (4.34) from the first step, we now can get the preliminary predetermined part in $q_{t+h|t}$ as follows:

$$
\begin{aligned}
q_{t+h|t} &:= \text{Quant}_\theta (Y_{t+h} | \mathscr{F}_t) \\
&= \sum_{i=1}^{h-1} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{i-1} \boldsymbol{\beta}_0 + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{h} q_t + (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{h-1} \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_t \\
&\quad + \text{Quant}_\theta \left(\boldsymbol{\varepsilon}_{t+h} + \sum_{i=1}^{h-1} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{i-1} \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t+i} \,\Big|\, \mathscr{F}_t \right).
\end{aligned}
$$
(4.36)

Based on our assumptions before, $\left(\boldsymbol{\varepsilon}_{t+h} + \sum_{i=1}^{h-1} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{i-1} \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t+i} \right)$ is not neces-

sary to be independent of $\boldsymbol{q}_t$ or $\boldsymbol{\varepsilon}_t$ unless $\{\boldsymbol{\varepsilon}_t\}$ is independently distributed. The distribution of $\left(\boldsymbol{\varepsilon}_{t+h} + \sum_{i=1}^{h-1} (\boldsymbol{\beta}_1 + \boldsymbol{\gamma})^{i-1} \boldsymbol{\beta}_1 \boldsymbol{\varepsilon}_{t+i} \right)$ can vary with the information on $\{\boldsymbol{\varepsilon}_t, \boldsymbol{p}_t, \boldsymbol{\varepsilon}_{t-1}, \ \boldsymbol{p}_{t-1}, \ldots, \}$ so as to influence the distributional characteristic $q_{t+h|t}$. In order to forecast $q_{t+h|t}$ and study the effect of $\boldsymbol{\varepsilon}_t$, we draw on the local projection idea of Jordà (2005) and consider to run the θ-th quantile regression of $\boldsymbol{Y}_{t+h}$ onto some explanatory variables measurable to $\mathscr{F}_t$. Based on the result (4.36), we know at least we should use explanatory variables $\boldsymbol{q}_t$ and $\boldsymbol{Y}_t$. We do not use $\boldsymbol{\varepsilon}_t$ directly because it is not observed and has the same role as $\boldsymbol{Y}_t$ in $q_{t+h|t}$ when we include $\boldsymbol{q}_t$ as another explanatory variable. However, we do not observe $\boldsymbol{q}_t$ either. Using estimated $\boldsymbol{q}_t$ brings its estimation error in the quantile regression and can make the regression result unreliable. So to mitigate the effect of its estimation error on the regression, we expand $\boldsymbol{q}_t$ to have more observed terms and use its expansion terms along with $\boldsymbol{Y}_t$ into the local quantile regression. Specifically, we expand $\boldsymbol{q}_t$ as follows:

$$\boldsymbol{q}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\gamma} \boldsymbol{q}_{t-1}, \tag{4.37}$$

$$= (\boldsymbol{\beta}_0 + \boldsymbol{\gamma}\boldsymbol{\beta}_0) + \boldsymbol{\beta}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\gamma}\boldsymbol{\beta}_1 \boldsymbol{Y}_{t-2} + \boldsymbol{\gamma}^2 \boldsymbol{q}_{t-2}. \tag{4.38}$$

So we can use explanatory variables $\{\boldsymbol{Y}_{t-1}, \boldsymbol{q}_{t-1}\}$ in replacement of $\boldsymbol{q}_t$ to mitigate the estimation error effect of $\boldsymbol{q}_t$, or use variables $\{\boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \boldsymbol{q}_{t-2}\}$ to further mitigate the estimation error effect as long as the spectral radius of $\boldsymbol{\gamma}$ is smaller than one because any v (denoted as an estimation error) will get vanished by $\lim_{n \to \infty} \boldsymbol{\gamma}^n v = \boldsymbol{0}$. Simulation results in next section show that the local quantile regression result on the coefficient of $\boldsymbol{Y}_t$ become much more reliable when we replace the explanatory variable $\boldsymbol{q}_t$ with $\{\boldsymbol{Y}_{t-1}, \boldsymbol{q}_{t-1}\}$ or $\{\boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \boldsymbol{q}_{t-2}\}$.[6]

Now we come to defining θ-*th quantile response function* of $\{\boldsymbol{Y}_t\}$ given a shock $\boldsymbol{\delta}$ to $\boldsymbol{\varepsilon}_t$ by taking the difference between $\mathrm{Quant}_\theta \left(\boldsymbol{Y}^*_{t+h} | \boldsymbol{\varepsilon}^*_t := \boldsymbol{\varepsilon}_t + \boldsymbol{\delta}, \mathscr{F}_{t-1} \right)$ and $\mathrm{Quant}_\theta \left(\boldsymbol{Y}_{t+h} | \boldsymbol{\varepsilon}_t, \mathscr{F}_{t-1} \right)$ as follows:

$$
\begin{aligned}
&\mathrm{QIRF}_h \left(\theta, \boldsymbol{\delta} | \boldsymbol{\varepsilon}_t, \mathscr{F}_{t-1} \right) \\
&= \mathrm{Quant}_\theta \left(\boldsymbol{Y}^*_{t+h} | \boldsymbol{\varepsilon}^*_t := \boldsymbol{\varepsilon}_t + \boldsymbol{\delta}, \mathscr{F}_{t-1} \right) - \mathrm{Quant}_\theta \left(\boldsymbol{Y}_{t+h} | \boldsymbol{\varepsilon}_t, \mathscr{F}_{t-1} \right) \\
&= \mathrm{Quant}_\theta \left(\boldsymbol{Y}^*_{t+h} | \boldsymbol{Y}^*_t := \boldsymbol{Y}_t + \boldsymbol{\delta}, \mathscr{F}_{t-1} \right) - \mathrm{Quant}_\theta \left(\boldsymbol{Y}_{t+h} | \boldsymbol{Y}_t, \mathscr{F}_{t-1} \right),
\end{aligned} \tag{4.39}
$$

where we can notice that

$$\mathrm{Quant}_\theta \left(\boldsymbol{\varepsilon}^*_t | \mathscr{F}_{t-1} \right) = \boldsymbol{\delta}$$

due to the shock, but $\mathrm{Quant}_\theta \left(\boldsymbol{\varepsilon}^*_{t+i} | \boldsymbol{\varepsilon}^*_{t+i-1}, \ldots, \boldsymbol{\varepsilon}^*_t, \mathscr{F}_{t-1} \right) = \boldsymbol{0}$ ($i = 1, \ldots, h$) according to the correct specification assumption (4.31).

---

[6]The optimal number of explanatory variables in replacing q_t is out of scope of this chapter.

$\mathrm{QIRF}_h\left(\theta, \boldsymbol{\delta} | \boldsymbol{\varepsilon}_t, \mathscr{F}_{t-1}\right)$ can be obtained by the local θ-quantile regression of Y_{t+h} onto $\left\{\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \boldsymbol{q}_{t-1}\right\}$ or onto $\left\{\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \boldsymbol{q}_{t-2}\right\}$ as follows:

$$\mathrm{QIRF}_h\left(\theta, \boldsymbol{\delta} | \boldsymbol{\varepsilon}_t, \mathscr{F}_{t-1}\right) = \lambda \boldsymbol{\delta}, \tag{4.40}$$

where λ is the coefficient of Y_t in the local θ-quantile regression of Y_{t+h} onto $\left\{\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \boldsymbol{q}_{t-1}\right\}$ or onto $\left\{\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \boldsymbol{q}_{t-2}\right\}$. We can also use higher moments of $\boldsymbol{Y}_t$ as explantory variables in the local quantile regression, and $\mathrm{QIRF}_h\left(\theta, \boldsymbol{\delta} | \boldsymbol{\varepsilon}_t, \mathscr{F}_{t-1}\right)$ can easily obtained by plugging the local quantile regression result into the definition (4.39).

4.5 Simulations

In this section, we are going to generate multivariate time series in an MVMQ-CAViaR DGP and a systemic MVMQ-CAViaR DGP, and implement the application procedure proposed in Section 4.3 on both DGPs to study their performances.

The MVMQ-CAViaR DGP that we simulate in this section is the bivariate CAViaR DGP specified below in which a bivariate time series sample is denoted as $\{\boldsymbol{Y}_t^{(r)}\}$.

$$\boldsymbol{Y}_t^{(r)} := \begin{bmatrix} y_{1,t}^{(r)} \\ y_{2,t}^{(r)} \end{bmatrix} = \begin{bmatrix} F_{t(3)}^{-1}(0.3) \\ F_{t(3)}^{-1}(0.3) \end{bmatrix} + \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} y_{1,t-1}^{(r)} \\ y_{2,t-1}^{(r)} \end{bmatrix} + \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} q_{1,t-1}^{(r)} \\ q_{2,t-1}^{(r)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t}^{(r)} \\ \varepsilon_{2,t}^{(r)} \end{bmatrix}, \tag{4.41}$$

or equivalently

$$\boldsymbol{Y}_t^{(r)} = A_i + A_y \boldsymbol{Y}_{t-1}^{(r)} + A_q \boldsymbol{q}_{t-1}^{(r)} + \boldsymbol{\varepsilon}_t^{(r)}, \tag{4.42}$$

where

$$A_i := \begin{bmatrix} F_{t(3)}^{-1}(0.3) \\ F_{t(3)}^{-1}(0.3) \end{bmatrix}, \qquad A_y := \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix}, \qquad A_q := \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix}, \tag{4.43}$$

$\boldsymbol{q}_t^{(r)} := \mathrm{Quant}_{0.3}(\boldsymbol{Y}_t^{(r)} | \mathscr{F}_{t-1})$, and $\{\boldsymbol{\varepsilon}_t^{(r)} - A_i\}$ is i.i.d. in Student's t-distribution with 3 degrees of freedom ($t(3)$ as the shorthand notation thereafter) with $\mathrm{Quant}_{0.3}(\boldsymbol{\varepsilon}_t^{(r)} - A_i | \mathscr{F}_{t-1}) = [0,0]'$ for all t and $F_{t(3)}^{-1}(\cdot)$ denoted as the inverse probability distribution function of $t(3)$.

The systemic MVMQ-CAViaR DGP that we simulate in this section is the systemic bivariate CAViaR DGP specified below in which a bivariate time series

sample is denoted as $\{\boldsymbol{Y}_t^{(s)}\}$.

$$\begin{bmatrix} 1 & 0 \\ -0.5 & 1 \end{bmatrix} \begin{bmatrix} y_{1,t}^{(s)} \\ y_{2,t}^{(s)} \end{bmatrix} = \begin{bmatrix} F_{t(3)}^{-1}(0.3) \\ F_{t(3)}^{-1}(0.3) \end{bmatrix} + \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix} \begin{bmatrix} y_{1,t-1}^{(s)} \\ y_{2,t-1}^{(s)} \end{bmatrix} + \begin{bmatrix} 0.2 & 0 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} q_{1,t-1}^{(s)} \\ q_{2,t-1}^{(s)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t}^{(s)} \\ \varepsilon_{2,t}^{(s)} \end{bmatrix},$$

$$\tag{4.44}$$

or equivalently

$$A_s \boldsymbol{Y}_t^{(s)} = A_i + A_y \boldsymbol{Y}_{t-1}^{(s)} + A_q \boldsymbol{q}_{t-1}^{(s)} + \boldsymbol{\varepsilon}_t^{(s)}, \tag{4.45}$$

where

$$A_s = \begin{bmatrix} 1 & 0 \\ -0.5 & 1 \end{bmatrix}, \tag{4.46}$$

with A_i, A_y, A_q as defined in (4.43), $\boldsymbol{q}_t^{(s)} := [\text{Quant}_{0.3}(y_{1,t}^{(s)}|\mathscr{F}_{t,0}), \text{Quant}_{0.3}(y_{2,t}^{(s)}|\mathscr{F}_{t,1})]'$, and $\{\boldsymbol{\varepsilon}_t^{(s)} - A_i\}$ is i.i.d. in $t(3)$ with $\text{Quant}_{0.3}(\boldsymbol{\varepsilon}_t^{(s)} - A_i|\mathscr{F}_{t-1}) = [0,0]'$ for all t. It is easy to simulate samples from these two DGPs. Specifically on simulating a sample in the bivariate DGP (4.42) with its sample size denoted as T, we first generate a sample of $\{\boldsymbol{\varepsilon}_t^{(r)} - A_i\} \overset{d}{\sim} t(3)$ in sample size $T + 200$. And simulate $\left\{\boldsymbol{Y}_t^{(r)}\right\}_{t=1}^{T+200}$ based on the following equation:

$$\boldsymbol{Y}_t^{(r)} = \sum_{i=1}^{t-1}(A_y + A_q)^{i-1}A_i + (A_y + A_q)^{t-1}\boldsymbol{q}_1 + \boldsymbol{\varepsilon}_t + \sum_{i=1}^{t-1}(A_y + A_q)^{i-1}A_y\boldsymbol{\varepsilon}_{t-i}^{(r)}, \tag{4.47}$$

with setting the initial value $\boldsymbol{q}_1^{(r)} = [0,0]'$. Delete the first 200 observations due to the burn-in effect of the initial value $\boldsymbol{q}_1^{(r)} = [0,0]'$ in the simulation, and return $\left\{\boldsymbol{Y}_t^{(r)}\right\}_{t=201}^{T+200}$ as the generated sample in the DGP (4.42). Analogously, we simulate a sample in the systemic bivariate DGP (4.45) by the following equation:

$$\boldsymbol{Y}_t^{(s)} = \sum_{i=1}^{t-1}(A_s^{-1}A_y + A_s^{-1}A_q)^{i-1}A_s^{-1}A_i + (A_s^{-1}A_y + A_s^{-1}A_q)^{t-1}\boldsymbol{q}_1^{(s)}$$

$$+ \boldsymbol{\varepsilon}_t + \sum_{i=1}^{t-1}(A_s^{-1}A_y + A_s^{-1}A_q)^{i-1}A_s^{-1}A_yA_s^{-1}\boldsymbol{\varepsilon}_{t-i}^{(s)} \tag{4.48}$$

with the same set-up that $\boldsymbol{q}_1^{(s)} = [0,0]'$ and the burn-in period of 200 observations.[7]

We can visually compare $\boldsymbol{Y}^{(r)}$ and $\boldsymbol{Y}^{(s)}$ by a plot of their samples as shown in Figure 4.1. As can be seen in Figure 4.1, $\{\boldsymbol{Y}_t^{(r)}\}$ is quite smooth , not as bumpy or comovement-like as $\{\boldsymbol{Y}^{(s)}\}$ which is due to the fact that in the systemic DGP the movement of $y_{1,t}$ immediately influences the conditional distribution of $y_{2,t}$

---

[7]The length of burn-in periods chosen in this chapter is based on our experience, and it can be adjusted for each specific DGP based on readers' expertise.
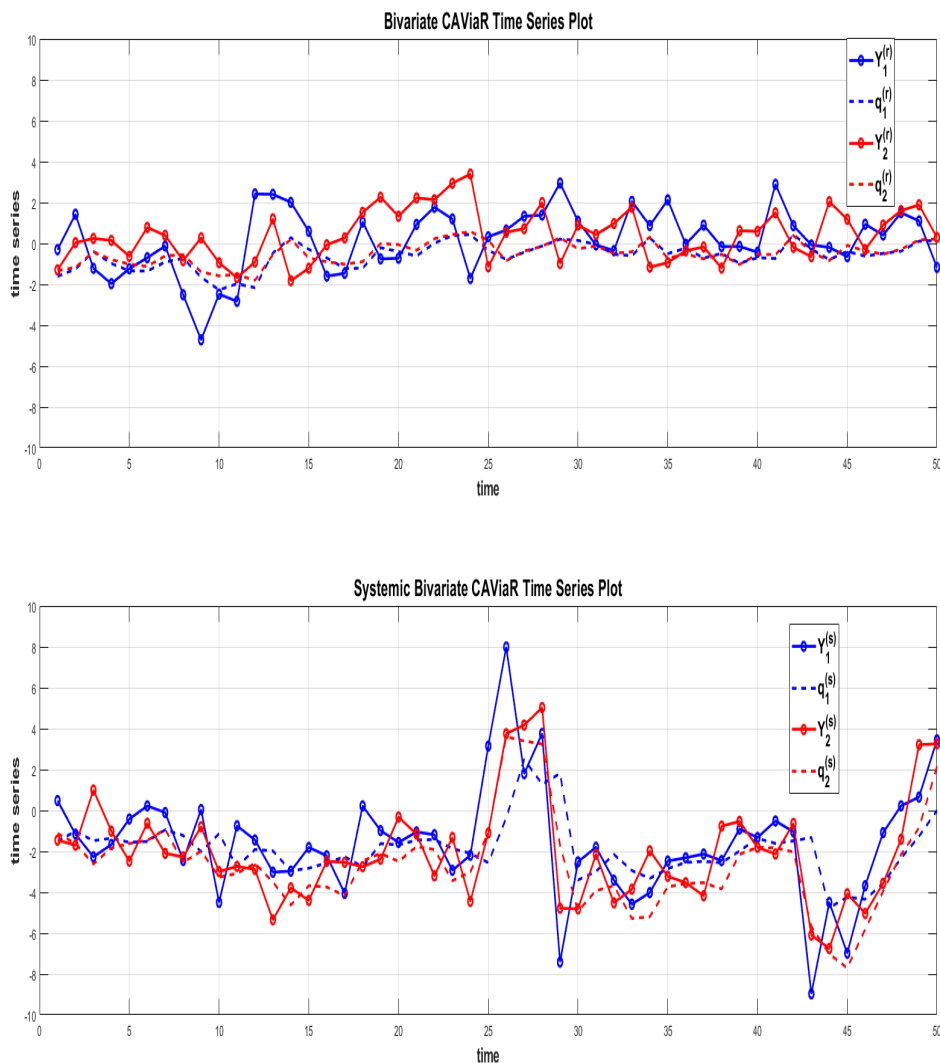
so as to be reflected in $y_{2,t}$.

Figure 4.1: Compare $\boldsymbol{Y}^{(r)}$ and $\boldsymbol{Y}^{(s)}$.

After obtaining samples $\{\boldsymbol{Y}_t^{(r)}\}$ and $\{\boldsymbol{Y}_t^{(s)}\}$ of sample size $T = 5000$ from the DGPs (4.42) and (4.45) respectively, we regress both $\{\boldsymbol{Y}_t^{(r)}\}$ and $\{\boldsymbol{Y}_t^{(s)}\}$ onto the bivariate CAViaR model specification (4.42) of quantile index 0.3. After regressions, we run the DQ tests to check if $y_{1,t}$ still has significant contemporaneous explanatory power on the conditional 0.3-th quantile of $y_{2,t}$ with the hypothesis statement as in (4.18). We use two methods to estimate the asymptotic covariance matrix of the bivariate CAViaR model for $\{\boldsymbol{Y}_t^{(r)}\}$, namely the adaptive random bandwidth method (Hecq and Sun, 2021) and the kernel method with the

optimal bandwidth used by White et al. (2015). From the size performances of these two methods in the DQ test of $\{\boldsymbol{Y}_t^{(r)}\}$ (see Table 4.1), we found that the adaptive random bandwidth (ARB) can well adapt to multivariate CAViaR models and robust in estimating the model asymptotic covariance matrix for various inference tests. Hereafter we only show test results in use of the adaptive random bandwidth method. The size performances of the DQ tests of $\{\boldsymbol{Y}_t^{(r)}\}$ and $\{\boldsymbol{Y}_t^{(s)}\}$ are shown in Table 4.1. We can see that the DQ test works robustly in indicating if some contemporaneous terms are significant to be involved into the modelling.

Table 4.1: Rejection rates (Size performances) of DQ tests (4.18) after regressing both $\boldsymbol{Y}^{(r)}$ and $\boldsymbol{Y}^{(s)}$ onto the bivariate CAViaR model specification (4.42)

DQ Tests	significance level: $\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$
$\boldsymbol{Y}^{(r)}$ (ARB)	0.010	0.0470	0.0790	0.1900	0.2980
$\boldsymbol{Y}^{(r)}$ (kernel)	0.001	0.0340	0.0700	0.1580	0.2670
$\boldsymbol{Y}^{(s)}$ (ARB)	1.000	1.000	1.000	1.000	1.000

Now we regress both $\{\boldsymbol{Y}_t^{(r)}\}$ and $\{\boldsymbol{Y}_t^{(s)}\}$ onto the systemic bivariate CAViaR model specification (4.45) of quantile index 0.3. After regressions, we run the Wald tests to check if $y_{1,t}$ is significant enough in explaining the conditional 0.3-th quantile of $y_{2,t}$ with the hypothesis statement as in (4.23). The Wald test performances are shown in Table 4.2. We can see that the Wald test works robustly on confirming if some contemporaneous terms are significant enough to be kept in the model.

After confirming the systemic model specification for $\boldsymbol{Y}^{(s)}$, we can use the systemic model regression result to measure the systemic risk of Y_1 by estimating $\text{CoVaR}_{0.3,t}^{2|y_{1,t}=q_{1,t}}$ as instructed in Section 4.3. Figure 4.2 shows a sample of $\boldsymbol{Y}^{(s)}$ with its $\widehat{\boldsymbol{q}}^{(s)}$, $-\widehat{\text{CoVaR}}_{0.3,t}^{2|y_{1,t}=\widehat{q}_{1,t}}$. We plot $-\widehat{\text{CoVaR}}_{0.3,t}^{2|y_{1,t}=\widehat{q}_{1,t}}$ not $\widehat{\text{CoVaR}}_{0.3,t}^{2|y_{1,t}=\widehat{q}_{1,t}}$ with $\boldsymbol{Y}^{(s)}$ because we regard $\boldsymbol{Y}^{(s)}$ as return variables so that we can comparatively view $\boldsymbol{Y}^{(s)}$ with its $\widehat{\boldsymbol{q}}^{(s)}$ and $-\widehat{\text{CoVaR}}_{0.3,t}^{2|y_{1,t}=\widehat{q}_{1,t}}$ in one plot.

Now we confirm the bivariate CAViaR model for $\boldsymbol{Y}^{(r)}$ and the systemic model for $\boldsymbol{Y}^{(s)}$. And we study their 0.3-th quantile impulse response functions (QIRFs) by the local 0.3-th quantile regression of $\boldsymbol{Y}_{t+h}$ $(h \geq 1)$ onto vector regressors

Table 4.2: Rejection rates (Size performances) of Wald tests (4.23) after regressing both $\boldsymbol{Y}^{(r)}$ and $\boldsymbol{Y}^{(s)}$ onto the systemic bivariate CAViaR model specification (4.45)

Wald Tests (4.23)	significance level: $\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$
$\boldsymbol{Y}^{(r)}$ (ARB)	0.0260	0.0680	0.1120	0.1980	0.2860
$\boldsymbol{Y}^{(s)}$ (ARB)	1.000	1.000	1.000	1.000	1.000

$$Y_t^{(s)}, \ q_t(s), \ \text{-CoVaR}_{0.1,t}^{2|y_{1,t}=q_{1,t}}$$
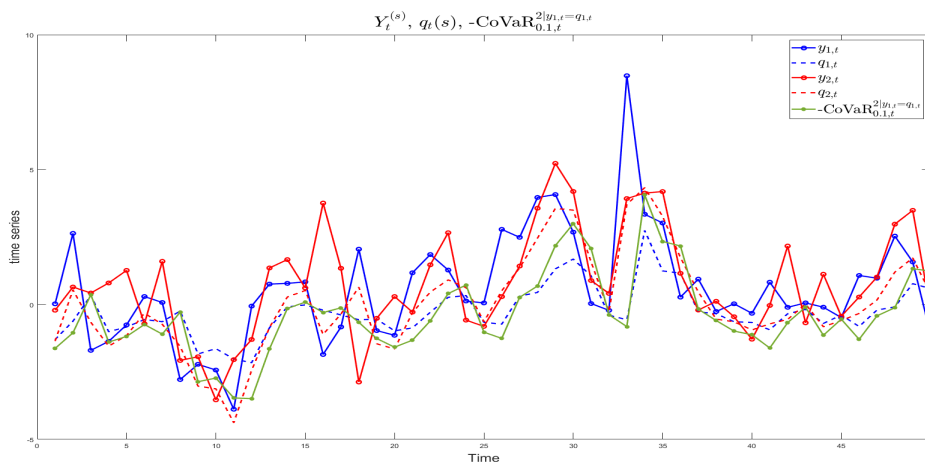
Figure 4.2: Time series plot of sample $\boldsymbol{Y}^{(s)}$ with its $\widehat{\boldsymbol{q}}^{(s)}$, $-\widehat{\text{CoVaR}}_{0.3,t}^{2|y_{1,t}=\widehat{q}_{1,t}}$

$\left\{ \boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \widehat{\boldsymbol{q}}_{t-2} \right\}$. Figure 4.3 and 4.4 compare the QIRF results among using explanatory variables $\{ \boldsymbol{Y}_t, \widehat{\boldsymbol{q}}_t \}$ and $\left\{ \boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \boldsymbol{Y}_{t-2}, \widehat{\boldsymbol{q}}_{t-2} \right\}$ with the true QIRF of $\boldsymbol{Y}$, in which we can see that using the expansion terms of $\widehat{\boldsymbol{q}}_t$ is more robust over using $\widehat{\boldsymbol{q}}_t$ directly. In fact, the outperformance of using $\widehat{\boldsymbol{q}}_t$ is more obvious for tail quantile indexes like 0.1 with the sample size being relatively large enough compared to the number of coefficients to be estimated in a local quantile regression.

In next section, we are going to implement the above application procedure on some empirical data and analyse the results for our empirical knowledge.
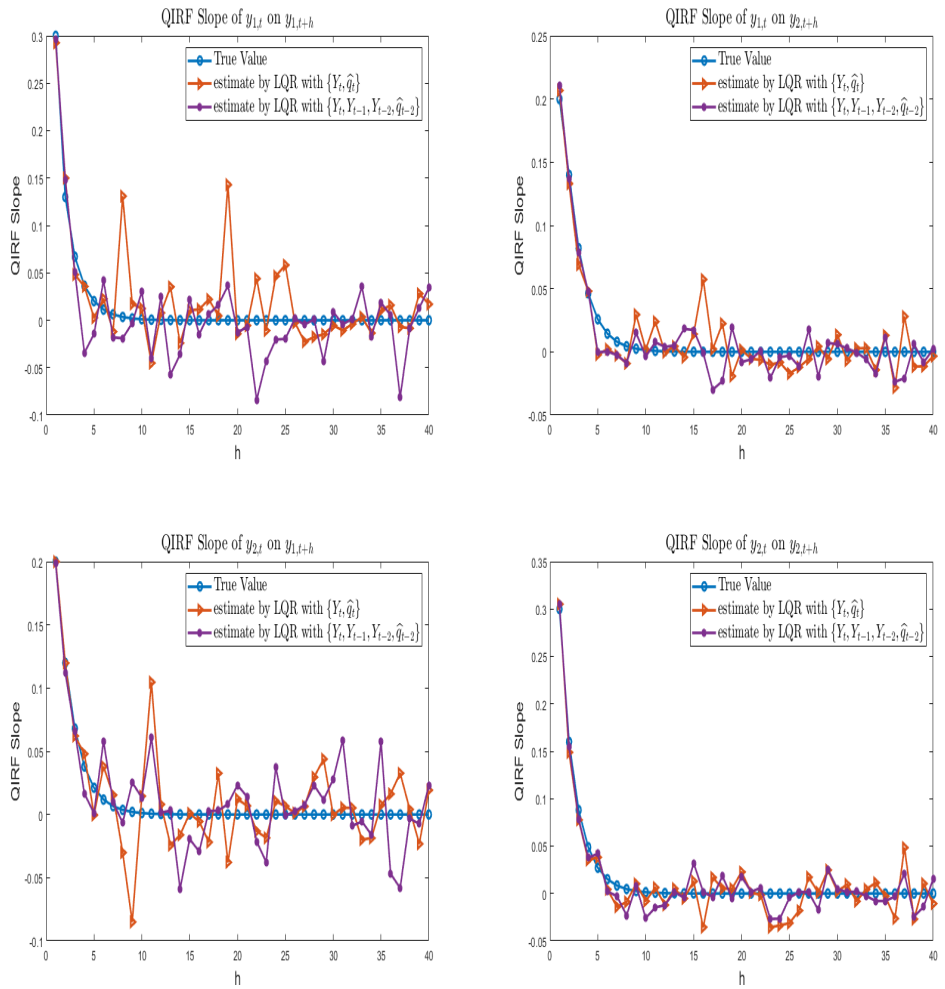
Figure 4.3: QIRF of $\boldsymbol{Y}^{(r)}$ estimated by the local 0.1-th quantile regression
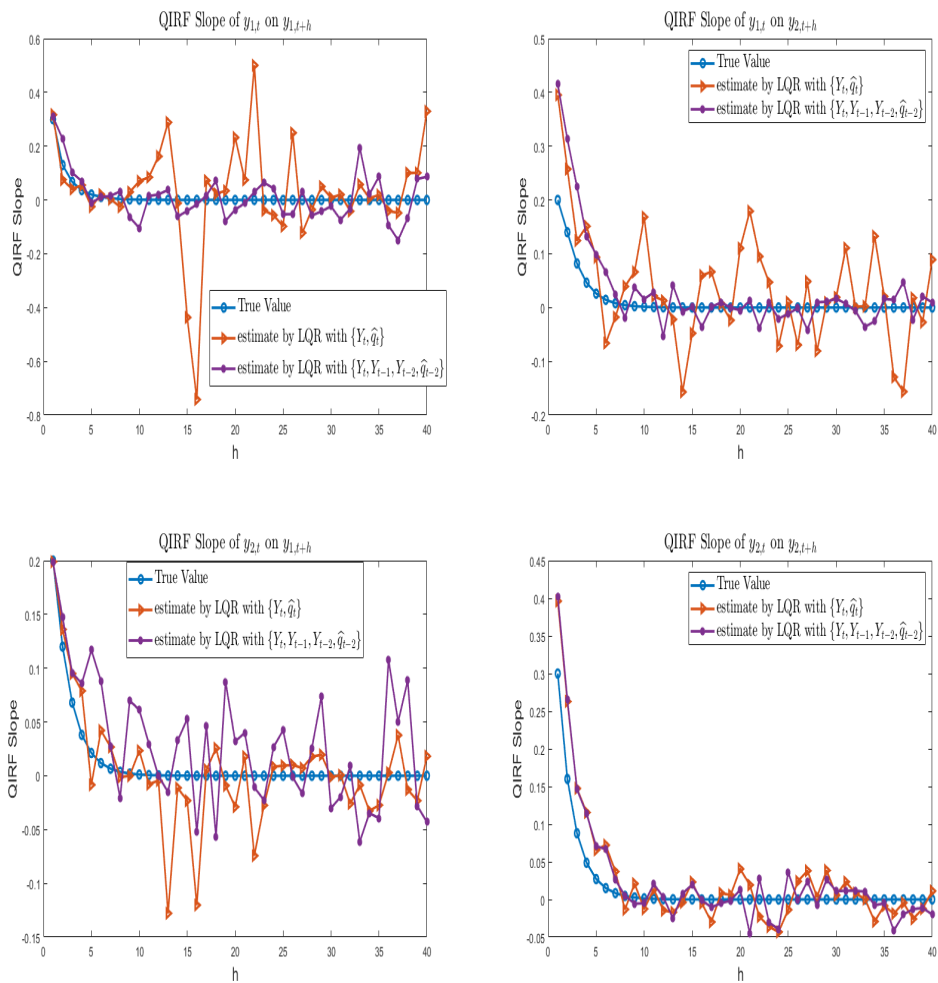
Figure 4.4: QIRF of $\boldsymbol{Y}^{(s)}$ estimated by the local 0.1-th quantile regression

4.6 Empirical Applications

We apply the above systemic modelling procedure to study the systemic risks of JPMorgan Chase & Co., the Bank of America Corporation, Wells Fargo & Company, the Goldman Sachs Group, Inc., Citigroup Inc. and Morgan Stanley which are the six largest bank holding companies in the United States ranked by total assets of March 31, 2020 per the Federal Financial Institutions Examination Council. Systemic risk of large financial intuitions can be as scary as the Great Recession that occurred between 2007 – 2009 in national economies globally. Before Lehman Brothers Holdings Inc. filed for bankruptcy in 2008, it was the fourth-largest investment bank in the United States. After Lehman Brothers filed for bankruptcy, global markets immediately plummeted and investors lost confidence, which caused bank runs, funding liquidity shortage, high haircuts, fire sales of assets and high counterpart credit risk in financial markets. The distress was spreading over financial institutions globally, and triggered the financial crisis of 2007—2008 which sparked the Great Recession (2007 – 2009), the most severe global recession since the Great Depression (1929-1933). Therefore, we are concerned about systemic risk of big financial intuitions and would like to measure the systemic risk of the six largest bank holding companies by our proposed method.

We use the S&P500 as the market index of interest which is deemed vulnerable to the systemic risks of those big banks. We downloaded daily adjusted closing stock prices of these six banks and the S&P500 from *Yahoo! Finance*, and each stock price time series has 3189 daily prices, ranging from 31-Dec-2006 to 01-Sep-2019. The price data were converted to percentage returns by multiplying 100 with the difference of the natural logarithm of the daily prices. The obtained return time series of each stock contains 3188 observations in the period of 01-Jan-2007 to 01-Sep-2019. In each return time series, the last 300 observations are used for the out-of-sample testing after the first 2888 observations are used to estimate the model.

We measure the systemic risk of each of these six banks individually by bivariate CAViaR models with the S&P500 daily returns. The bivariate CAViaR models used for estimating the conditional 5% quantiles of a big bank and the S&P500 daily returns are (4.30) and (4.13) for bivariate CAViaR and systemic bivariate CAViaR regressions respectively.

We set up inference tests in the same way as we did in the preceding sections. The inference testing results based on our systemic modelling procedure for the conditional 5% quantiles of the six banks' and the S&P500 daily returns are presented in Table 4.3 for in-sample tests and Table 4.4 for out-of-sample tests. The DQ p-value (MVMQ) columns of Table 4.3 and 4.4 provide strong evidence of contemporaneous (daily) spillovers of financial distress at individual financial

Table 4.3: In-sample test results on the empirical data (quantile index=0.05)

Banks	VaR exceedance rates (MVMQ) [bank, the S&P500]	DQ p-value (MVMQ)	VaR exceedance rates (SMVMQ) [bank, the S&P500]	Wald p-value (SMVMQ)
BAC	[0.0502, 0.0499]	0	[0.0495, 0.0495]	0
C	[0.0506, 0.0502]	0	[0.0502, 0.0506]	0
GS	[0.0516, 0.0509]	0	[0.0519, 0.0492]	0
JPM	[0.0502, 0.0533]	0	[0.0499, 0.0488]	0
MS	[0.0512, 0.0506]	0	[0.0509, 0.0519]	0
WFC	[0.0499, 0.0561]	0	[0.0495, 0.0495]	0

institutions to the S&P500 index return. There is further evidence that the involved contemporaneous terms are significant in the systemic bivariate CAViaR models according to the Wald p-value (SMVMQ) columns. From the results of the out-of-sample tests in Table 4.4 for $VaR_{5\%}$ backtesting, we see that we can not reject the systemic model of Goldman Sachs and the S&P500 which even has $VaR_{5\%}$ exceedance rates close to the risk level 5%. Other systemic models are rejected by the out-of-sample DQ tests which means that those banks still have significant explanatory power on conditional 0.5-th quantiles of the market index which is not revealed by the in-sample estimation. There are many possible reasons behind those model rejections. One reason can be the inappropriate functional form of the contemporaneous terms in banks' returns we considered in those models. Another possible reason is that those rejected models omitted some other significant contemporaneous terms such as Goldman Sachs. The results in Table 4.3 and 4.4 let us confirm the systemic model of Goldman Sachs and the S&P500 first so as to measure the systemic risk of Goldman Sachs, and also gives us some clues to explore the functional forms of the contemporaneous effect of the banks on the market index. For example, we can run the systemic model on one of the five banks in model rejections with Goldman Sachs so that we count out the common contemporaneous effect of that bank on the market index with Goldman Sachs and focus on their idiosyncratic parts for the contemporaneous effects and the proper functional forms. We are not going to measure the systemic risks of these six banks together by a seven-variate CAViaR model in this chapter. Since systemic MVMQ-CAViaR models are directional, we have to decide on the contemporaneous influence direction among these seven stocks. However, the results above of the (systemic) bivariate CAViaR models can be the starting point to build a proper seven-variate CAViaR model and then study their systemic risks in a whole system, which involves an enumeration of model estimations and inference tests and is left for future research.

After confirming the systemic model specification for the conditional 5%-th quantiles of Goldman Sachs and the S&P500, we can use the systemic model regression result to measure the systemic risk of Goldman Sachs by estimating

Table 4.4: Out-of-sample test results on the empirical data (quantile index=0.05)

Banks	VaR exceedance rates (MVMQ) [bank, the S&P500]	DQ p-value (MVMQ)	VaR exceedance rates (SMVMQ) [bank, the S&P500]	DQ p-value (SMVMQ)
BAC	[0.0733, 0.0367]	0	[0.0833, 0.0733]	0
C	[0.0667, 0.0367]	0	[0.0867, 0.0800]	0
GS	[0.0500, 0.0400]	0	[0.0567, 0.0700]	0.4576
JPM	[0.0167, 0.0400]	0	[0.0167, 0.0767]	0.0005
MS	[0.0267, 0.0433]	0	[0.0467, 0.0700]	0
WFC	[0.0367, 0.0433]	0	[0.0467, 0.0833]	0.0015

$\text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}=q_{GS,t}(5\%)}$ as instructed in Section 4.3.

$$\Delta\text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}} := \text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}=q_{GS,t}(5\%)} - \text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}=q_{GS,t}(50\%)}$$

is defined by Adrian and Brunnermeier (2011) as the part of Goldman Sachs' systemic risk that can be attributed to the S&P500. To view $\Delta\text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}}$, we also need to model and estimate the conditional 50%-th quantiles of Goldman Sachs' returns. Analogously, we regress the returns of Goldman Sachs and the S&P500 onto the bivariate CAViaR (4.30) and the systemic bivariate CAViaR model (4.13) of quantile index 0.50 respectively, and perform the inference tests as we did before which results in Table 4.5 and 4.6.

Table 4.5: In-sample test results on the returns of Goldman Sachs and the S&P500 (quantile index=0.50)

Bank	VaR exceedance rates (MVMQ) [bank, the S&P500]	DQ p-value (MVMQ)	VaR exceedance rates (SMVMQ) [bank, the S&P500]	Wald p-value (SMVMQ)
GS	[0.5000, 0.4997]	0	[0.4997, 0.4997]	0

Table 4.6: Out-of-sample test results on the returns of Goldman Sachs and the S&P500 (quantile index=0.05)

Bank	VaR exceedance rates (MVMQ) [bank, the S&P500]	DQ p-value (MVMQ)	VaR exceedance rates (SMVMQ) [bank, the S&P500]	DQ p-value (SMVMQ)
GS	[0.5133, 0.5033]	0	[0.5233, 0.4700]	0.5607

As we can see that the systemic model is not rejected for the conditional 0.5-th quantiles of Goldman Sachs' and the S&P500 returns so that we can use its estimated conditional 0.5-th quantiles of Goldman Sachs' returns to calculate $\text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}=q_{GS,t}(50\%)}$ so for us to view $\Delta\text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}}$. The estimated conditional 0.05-th and 0.5-th quantiles of Goldman Sachs' and the S&P500 returns are plotted in Figure 4.8 and 4.9 respectively, see Appendix 4.A. Figure 4.5 shows the returns of the S&P500 index, $-\text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}=q_{GS,t}(5\%)}$ and $-\text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}=q_{GS,t}(50\%)}$. We plot $-\text{CoVaR}_{0.05,t}^{SP500|y_{GS,t}=q_{GS,t}}$ not

$\text{CoVaR}_{0.05,t}^{SP500|y_{GS,t}=q_{GS,t}}$ with the returns so that we can comparatively view the returns of the S&P500 index with $-\text{CoVaR}_{0.05,t}^{SP500|y_{GS,t}=q_{GS,t}}$ in one plot. Figure 4.6 reflects the part of Goldman Sachs' systemic risk attributed to the S&P500 index over time by plotting $\{\Delta\text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}}\}$. We can see from Figure 4.6 that the part of Goldman Sachs' systemic risk attributed to the S&P500 index was cumulating after Lehman Brothers filed for bankruptcy on September 15, 2008, and reached unprecedentedly high until mid-2009. Figure 4.6 is quite story-telling and links almost every peak in the figure to a distress event of Goldman Sachs, which can be informative for financial market regulators in systemic risk management.

The 0.05-th QIRF coefficients of Goldman Sachs on the S&P500 is obtained by the local 0.05-th quantile regressions (see Section 4.4) in use of the expansion terms up to 3-step lagged return vector variables, which is plotted in Figure 4.7. The 0.05-th QIRF coefficient drawn at $h = 0$ in Figure 4.7 is the estimated coefficient of the contemporaneous term in Goldman Sachs' return on the conditional 0.05-th quantile of the S&P500 return by the systemic model regression (4.13).[8] Figure 4.7 says that a shock to Goldman Sachs' return at $h = 0$ contemporaneously shifts the conditional 0.05-th quantile of the S&P500 return in tandem considerably, and in the rest of days the conditional 0.05-th quantile of the S&P500 returns are less memorable of this shock.

From the empirical application above, we have seen that the contemporaneous effects of the big banks' returns are significant on conditional quantiles of the S&P500 returns, and it is informative to use systemic MVMQ CAViaR models proposed in this chapter with the systemic modelling procedure (see Section 4.3) to analyse and monitor the systemic risks of big financial institutions.

---

[8]When we used the expansion terms up to 4-step lagged return vector variables in the local 0.05-th quantile regression, we got a similar result to the 0.05-th QIRF coefficients in Figure 4.7.
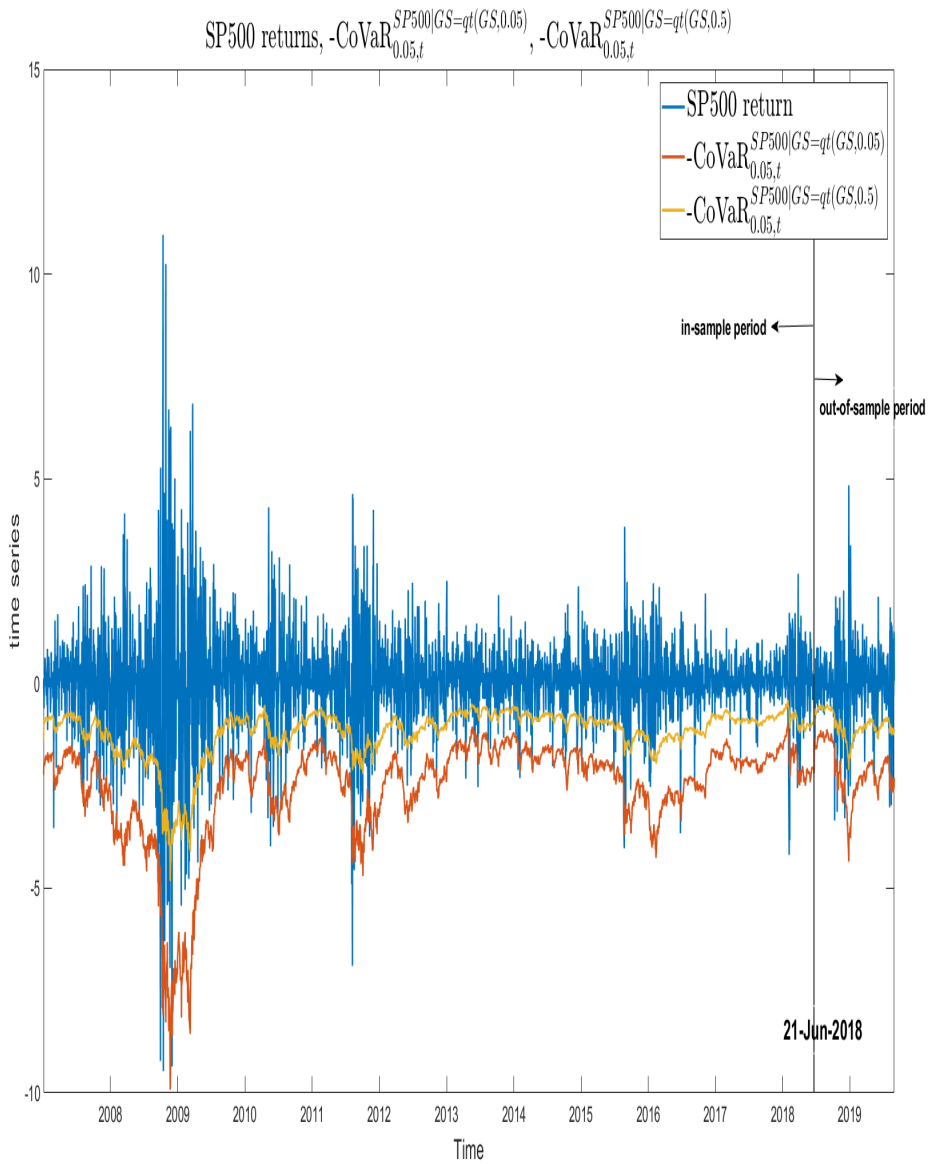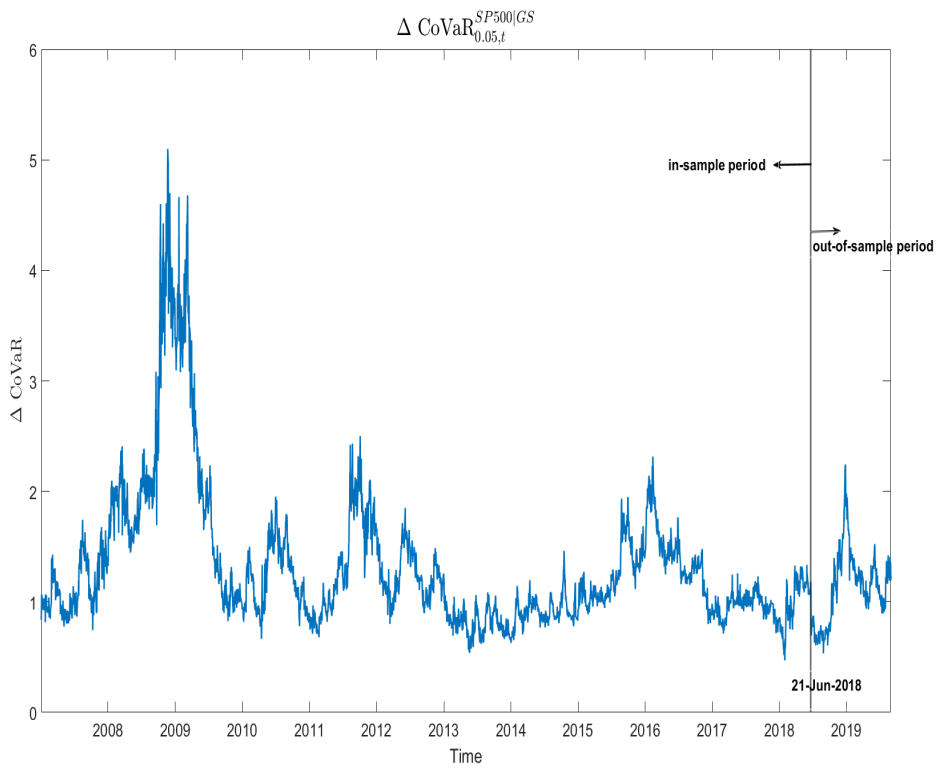
Figure 4.5: CoVaR plot of Goldman Sachs on the S&P500

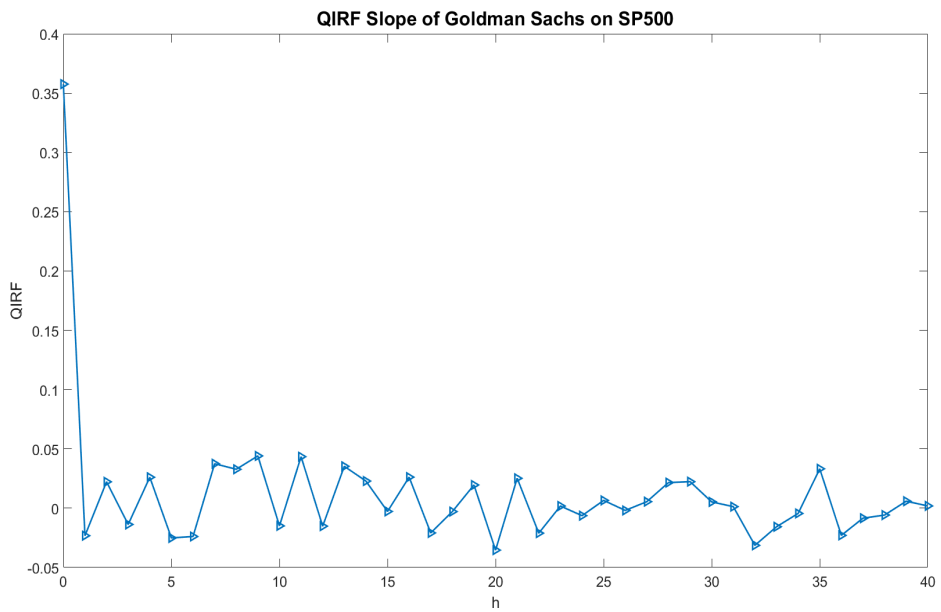Figure 4.6: $\Delta\text{CoVaR}_{5\%,t}^{SP500|y_{GS,t}}$ plot over time

Figure 4.7: 0.05-th QIRF coefficient of Goldman Sachs on the S&P500

4.7 Conclusions

We generalized multivariate multi-quantile CAViaR models (MVMQ-CAViaR, see White et al., 2015) by incorporating CoVaR specification (see Adrian and Brunnermeier, 2011) into the model specification in this chapter. The proposed systemic MVMQ-CAViaR model presents a vector-autoregressive (VAR) specification of financial institutions' value-at-risk (VaR) as well as their CoVaR. This model generalization is able to capture contemporaneous tail dependence of financial institutions and market indexes so that we can interpret the systemic risks of the institutions over time. The consistency and asymptotic normality proofs of this generalized model are provided in this chapter along with some relevant inference tests, for which we implemented simulation tests and showed robust model performances. For tracing the transmission of a single shock to a financial institution in the financial system, we also constructed quantile impulse response functions (QIRF) accordingly in use of the local projection idea (Jordà, 2005) and expansion of estimated terms. Based on our simulation results, we can see that using the expansion terms of $\widehat{q}_t$ is more robust than directly using $\widehat{q}_t$ in the local quantile regression for the QIRF estimation. Applications to real data provided empirical support to this methodology.
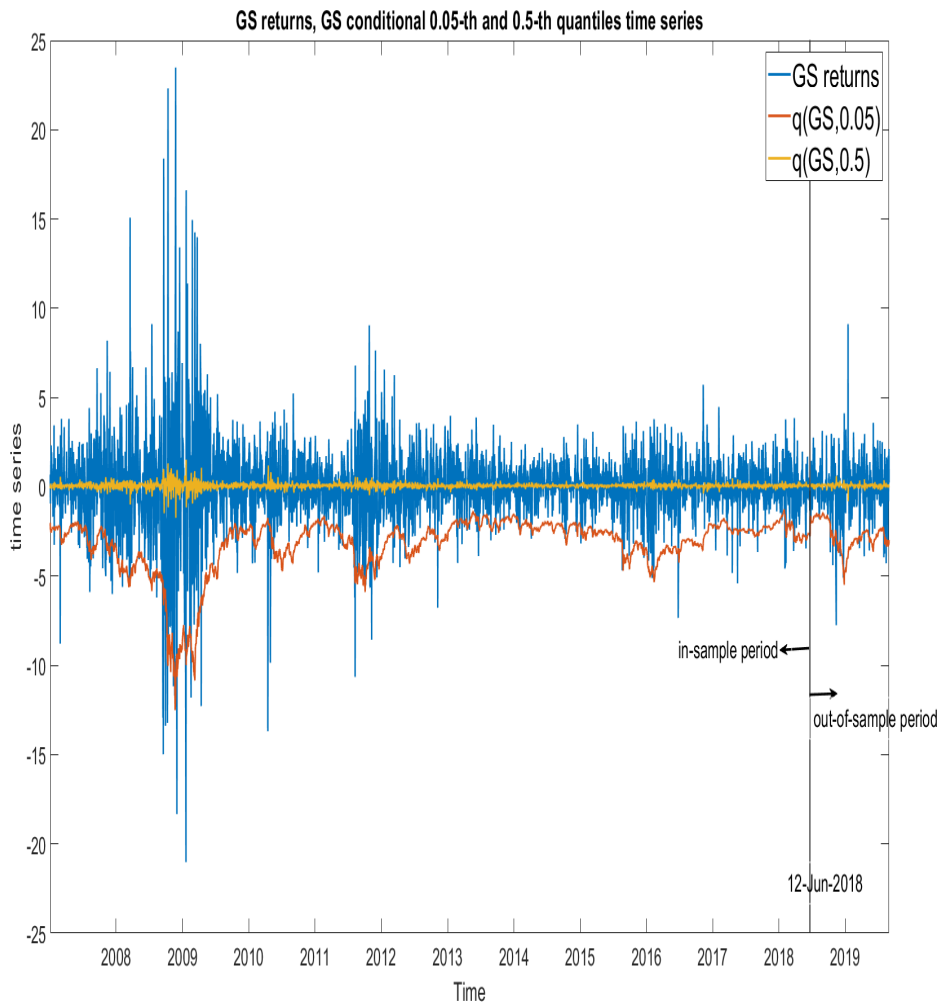
4.A Appendix: Extra Figures

Figure 4.8: Time series plot of the returns of Goldman Sachs with its fitted conditional 0.05-th and 0.5-th quantiles by the systemic MVMQ CAViaR regression with the S&P500
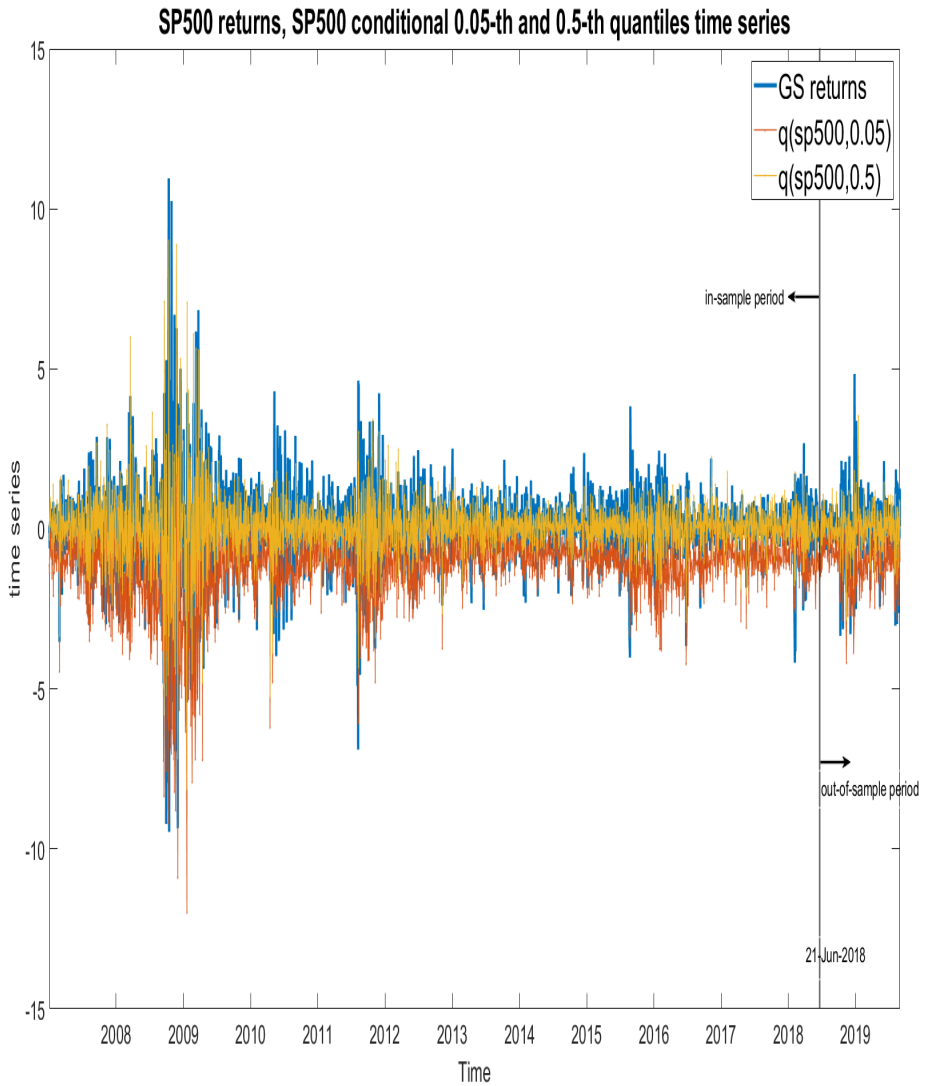
Figure 4.9: Time series plot of the returns of the S&P500 with its fitted conditional 0.05-th and 0.5-th quantiles by the systemic MVMQ CAViaR regression with Goldman Sachs

4.B Appendix: Proofs

Proof of Theorem 11.
The proof builds on Engle and Manganelli (2004)'s Thoerem 1 and White et al. (2015)'s Theorem 1, and can be obtained immediately by following the proof of White et al. (2015)'s Theorem 1 on pp. 184. □

Proof of Theorem 12.
The proof builds on Engle and Manganelli (2004)'s Thoerem 2 and White et al. (2015)'s Theorem 2, and can be obtained by following the proof of White et al. (2015)'s Theorem 2 on pp. 186. □

Proof of Theorem 13.
The proof builds on Engle and Manganelli (2004)'s Thoerem 3 and White et al. (2015)'s Theorem 3, and can be obtained immediately by following the proof of White et al. (2015)'s Theorem 3 on pp. 186. □

Proof of Theorem 14.
First, since expectation is a linear function, we rewrite $\mathbb{E}_{y_{i,t},\widehat{\boldsymbol{\alpha}}}\left[\widehat{f}_{i,j,t}(0)\Big|\mathscr{F}_{t,i-1}\right]$ as follows:

$$
\begin{aligned}
&\mathbb{E}_{y_{i,t},\widehat{\boldsymbol{\alpha}}}\left[\widehat{f}_{i,j,t}(0)\Big|\mathscr{F}_{t,i-1}\right]\\
=&N^{-1}\sum_{z=1}^{N}\mathbb{E}_{y_{i,t},\widehat{\boldsymbol{\alpha}}}\left[\frac{\mathbb{1}_{\left\{y_{i,t}\leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})+\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z-\widehat{\boldsymbol{\alpha}})\right\}}-\mathbb{1}_{\left\{y_{i,t}\leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\right\}}}{\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z-\widehat{\boldsymbol{\alpha}})}\Bigg|\mathscr{F}_{t-1,i-1},y_{i,t}\neq q_{i,j,t}\right]
\end{aligned}
$$
(4.49)

This equality holds when N goes to infinity by applying the dominated convergence theorem as we regard the least l_s absolute residual elements in $\{|\widehat{\varepsilon}_{i,j,t}|, i=1,\ldots,n; j=1,\ldots,p; t=1,\ldots,T.\}$ as zeros, see Theorem 6. We rank $\{|\widehat{\varepsilon}_{i,j,t}|, i=1,\ldots,n; j=1,\ldots,p; t=1,\ldots,T.\}$ from the smallest to largest into $\{|\widehat{\varepsilon}|_{(1)},\cdot,|\widehat{\varepsilon}|_{(npT)}\}$. In fact, iterations of a simplex-based direct search method like the Nelder–Mead method for optimizing l_s parameters terminates at the vertices of a simplex in the parameter space (Lagarias et al., 1998). That is to say, the iterations terminate with l_s elements of $\{|\widehat{\varepsilon}_{i,j,t}|, i=1,\ldots,n; j=1,\ldots,p; t=1,\ldots,T.\}$ solved to be zeros. Therefore, we get

$$
|\widehat{f}_{i,j,t}(0)|\leq\frac{1}{|\widehat{\varepsilon}|_{(l_s+1)}}<\infty,
$$
(4.50)

where $|\widehat{\varepsilon}|_{(l_s+1)}\neq 0$ for a well-defined convex function minimization.

Since $\{(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})\}_{z=1}^{N}$ is i.i.d in $N(\mathbf{0}, \boldsymbol{I}_{l_s})$, we can get that for each $t \in \{1, \ldots, T\}$,

$$\left\{ \mathbb{E}_{y_{i,t}, \widehat{\boldsymbol{\alpha}}} \left[\left. \frac{\mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) + \nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})\right\}} - \mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\right\}}}{\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})} \right| \mathscr{F}_{t-1,i-1}, y_{i,t} \neq q_{i,j,t} \right] \right\}_{z=1}^{N}$$

is a sequence of independent random variables with finite second moments by Assumption 5(iii) of White et al. (2015) that

$$
\begin{cases}
D_{1,t} := \max\limits_{i=1,\ldots,n} \max\limits_{j=1,\ldots,p} \max\limits_{s=1,\ldots,l_s} \sup\limits_{\boldsymbol{\alpha} \in \Theta} \left| \frac{\partial q_{i,j,t}(\boldsymbol{\alpha})}{\partial \alpha_s} \right|, \\
\mathbb{E}[D_{1,t}] < \infty, \\
\mathbb{E}[D_{1,t}^2] < \infty.
\end{cases}
\tag{4.51}
$$

Hence, we can use the strong Law of Large Number to get that

$$N^{-1} \sum_{z=1}^{N} \mathbb{E}_{y_{i,t}, \widehat{\boldsymbol{\alpha}}} \left[\left. \frac{\mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) + \nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})\right\}} - \mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\right\}}}{\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})} \right| \mathscr{F}_{t-1,i-1}, y_{i,t} \neq q_{i,j,t} \right]$$

$$\xrightarrow{a.s.} \mathbb{E}_{y_{i,t}, \widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}_z} \left[\left. \frac{\mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) + \nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})\right\}} - \mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\right\}}}{\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})} \right| \mathscr{F}_{t-1,i-1}, y_{i,t} \neq q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) \right],$$

$$\tag{4.52}$$

as $N \to \infty$. Denote $F_{i,t}(\cdot)$ and $f_{i,t}(\cdot)$ as the probability distribution function and the probability density function of $y_{i,t}$ conditional on $\mathscr{F}_{t,i-1}$ respectively. By Assumption 2(i) and 3(ii) of White et al. (2015) that $F_{i,t}(\cdot)$ and $f_{i,t}(\cdot)$ are continuous in $\mathbb{R}$ and $q_{i,j,t}(\cdot)$ is continuously differential on Θ with the conditional probability density of $y_{i,t}$ at its conditional $\theta_{i,j}$-th quantile $q_{i,j,t}(\boldsymbol{\alpha}^o)$ being $f_{i,j,t}(0)$, we can get that

$$\mathbb{E}_{y_{i,t}, \widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}_z} \left[\left. \frac{\mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) + \nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})\right\}} - \mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\right\}}}{\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})} \right| \mathscr{F}_{t-1,i-1}, y_{i,t} \neq q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) \right]$$

$$= \mathbb{E}_{\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}_z} \left[\left. \mathbb{E}_{y_{i,t}} \left[\left. \frac{\mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) + \nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})\right\}} - \mathbb{1}_{\left\{y_{i,t} \leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\right\}}}{\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})} \right| \mathscr{F}_{t-1,i-1}, y_{i,t} \neq q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) \right] \right| \mathscr{F}_{t-1,i-1} \right]$$

$$= \mathbb{E}_{\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}_z} \left[\left. \frac{F_{i,t}\left(q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) + \nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})\right) - F_{i,t}\left(q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\right)}{\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})} \right| \mathscr{F}_{t-1,i-1} \right]$$

$$= \mathbb{E}_{\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}_z} \left[\left. \frac{f_{i,t}(q_{i,j,t}(\widehat{\boldsymbol{\alpha}})) \nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}}) + O_p\left((\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})' \nabla q_{i,j,t}(\widehat{\boldsymbol{\alpha}}) \nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})\right)}{\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})} \right| \mathscr{F}_{t-1,i-1} \right]$$

$$= \mathbb{E}_{\widehat{\boldsymbol{\alpha}}, \boldsymbol{\alpha}_z} \left[f_{i,t}(q_{i,j,t}(\widehat{\boldsymbol{\alpha}})) + O_p\left(\nabla' q_{i,j,t}(\widehat{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}})\right) \middle| \mathscr{F}_{t-1,i-1} \right]$$

$$= f_{i,j,t}(0),$$

$$\tag{4.53}$$

where the last line is obtained by applying the dominated convergence theorem as $\boldsymbol{\alpha}_z - \widehat{\boldsymbol{\alpha}} = o_p(1)$ and $\lim_{T \to \infty} \widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^o$ with assuming $f_{i,t}(\cdot)$ being a continuous function and $\mathbb{E}[D_{1,t}] < \infty$.

Therefore, we have that $\mathbb{E}_{y_{i,t}, \widehat{\boldsymbol{\alpha}}} \left[\widehat{f}_{i,j,t}(0) \middle| \mathscr{F}_{t,i-1} \right] \xrightarrow{p} f_{i,j,t}(0)$ for $t \in \{1, 2, \ldots, T\}$

as $N \to \infty$, and conclude this proof. □

Proof of Theorem 15.
Denote that

$$\bar{Q}_T := \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{p} \widehat{f}_{i,j,t}(0) \nabla q_{i,j,t}(\boldsymbol{\alpha}^o) \nabla' q_{i,j,t}(\boldsymbol{\alpha}^o),. \qquad (4.54)$$

Note that

$$\widehat{Q}_T - Q = \widehat{Q}_T - \bar{Q}_T + \bar{Q}_T - Q. \qquad (4.55)$$

It is straightforward to get that

$$\widehat{Q}_T - \bar{Q}_T = o_p(1), \qquad (4.56)$$

because it is easy to get $\frac{1}{T} \sum_{t=1}^{T} \widehat{f}_{i,j,t}(0)$ converges in probability to $\frac{1}{T} \sum_{t=1}^{T} f_{i,j,t}(0)$,

i.e, $\frac{1}{T} \sum_{t=1}^{T} \widehat{f}_{i,j,t}(0) - \frac{1}{T} \sum_{t=1}^{T} f_{i,j,t}(0) = o_p(1)$ by following the proof of Corollary 7
with Assumption 5(iii) of White et al. (2015) as shown in (4.51).
And we can get that

$$\bar{Q}_T - Q = o_p(1), \qquad (4.57)$$

since that $\widehat{V}_T \overset{p}{\longrightarrow} V$ in Theorem 13 and $q_{i,j,t}(\cdot)$ is bounded so as for continuous
function $f_{i,t}(q_{i,j,t}(\cdot))$.

Therefore, we have that $\widehat{Q}_T - Q = o_p(1)$ and conclude this proof. □

Proof of Theorem 16.
To prove this theorem, we need to find out the limiting distribution of

$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \left(\mathbb{1}_{\{y_{i,t} \le q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\}} - \theta_{i,j} \right)$. First we define

$$\begin{cases} \text{Hit}_{i,j,t}(\boldsymbol{\alpha}) := \mathbb{1}_{\{y_{i,t} \le q_{i,j,t}(\boldsymbol{\alpha})\}} - \theta_{i,j}, \\ \mathbf{Hit}_{i,t}(\boldsymbol{\alpha}) := [\text{Hit}_{i,1,t}(\boldsymbol{\alpha}), \ldots, \text{Hit}_{i,p,t}(\boldsymbol{\alpha})]', \\ \mathbf{Hit}_t(\boldsymbol{\alpha}) := [\mathbf{Hit}'_{1,t}(\boldsymbol{\alpha}), \ldots, \mathbf{Hit}'_{n,t}(\boldsymbol{\alpha})]', \\ \boldsymbol{q}_{i,t}(\boldsymbol{\alpha}) := [q_{i,1,t}(\boldsymbol{\alpha}), \ldots, q_{i,p,t}(\boldsymbol{\alpha})]', \\ \boldsymbol{q}_t(\boldsymbol{\alpha}) := [\boldsymbol{q}'_{1,t}(\boldsymbol{\alpha}), \ldots, \boldsymbol{q}'_{n,t}(\boldsymbol{\alpha})]', \end{cases} \qquad (4.58)$$

and start to derive the the limiting behaviour of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \left(\mathbb{1}_{\{y_{i,t} \leq q_{i,j,t}(\widehat{\alpha})\}} - \theta_{i,j} \right)$ as follows:

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \left(\mathbb{1}_{\{y_{i,t} \leq q_{i,j,t}(\widehat{\alpha})\}} - \theta_{i,j} \right)
$$

$$
= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \left(\text{Hit}_{i,j,t}(\widehat{\alpha}) - \text{Hit}_{i,j,t}(\alpha^o) + \text{Hit}_{i,j,t}(\alpha^o) \right) \tag{4.59}
$$

$$
= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \left(\text{Hit}_{i,j,t}(\widehat{\alpha}) - \text{Hit}_{i,j,t}(\alpha^o) \right) + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \text{Hit}_{i,j,t}(\alpha^o).
$$

Use the result proved by Engle and Manganelli (2004) that

$$
\left(
\begin{aligned}
& \text{Hit}_{i,j,t}^{\oplus}(\alpha) := \left(1 + exp\left(\frac{y_{i,t} - q_{i,j,t}(\alpha)}{c_T} \right) \right)^{-1} - \theta_{i,j}, \\
& \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \text{Hit}_{i,j,t}^{\oplus}(\alpha^o) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \text{Hit}_{i,j,t}(\alpha^o) = o_p(1), \\
& \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \text{Hit}_{i,j,t}^{\oplus}(\widehat{\alpha}) - \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \text{Hit}_{i,j,t}(\widehat{\alpha}) = o_p(1),
\end{aligned}
\right. \tag{4.60}
$$

where c_T is a nonstochastic sequence such that $\lim_{T \to \infty} c_T = 0$. We can approximate the first term in the last line of (4.59) as follows:

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \left(\text{Hit}_{i,j,t}(\widehat{\alpha}) - \text{Hit}_{i,j,t}(\alpha^o) \right)
$$

$$
= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} \left(\text{Hit}_{i,j,t}^{\oplus}(\widehat{\alpha}) - \text{Hit}_{i,j,t}^{\oplus}(\alpha^o) \right) + o_p(1)
$$

$$
= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_{i_c,t} f_{i,j,t}(0) \frac{\partial q_{i,j,t}(\alpha^o)}{\partial \alpha^{o\prime}} (\widehat{\alpha} - \alpha^o) + o_p(1) \tag{4.61}
$$

$$
= \sqrt{T}(\widehat{\alpha} - \alpha^o)' \frac{1}{T} \sum_{t=1}^{T} y_{i_c,t} f_{i,j,t}(0) \frac{\partial q_{i,j,t}(\alpha^o)}{\partial \alpha^{o\prime}} + o_p(1)
$$

$$
= -\frac{1}{\sqrt{T}} \left(\sum_{t=1}^{T} \sum_{i=1}^{n} \text{Hit}_t'(\alpha^o) \frac{\partial q_t(\alpha^o)}{\partial \alpha^{o\prime}} \right) Q^{-1} G + o_p(1),
$$

where $G := \frac{1}{T} \sum_{t=1}^{T} y_{i_c,t} f_{i,j,t}(0) \frac{\partial q_{i,j,t}(\alpha^o)}{\partial \alpha^{o\prime}}$, and the third and last lines are obtained

by respectively applying (B.5) of Engle and Manganelli (2004) on pp. 379 and

$$\sqrt{T}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^o)' = -\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T}\sum_{i=1}^{n}\mathbf{Hit}'_t(\boldsymbol{\alpha}^o)\frac{\partial \boldsymbol{q}_t(\boldsymbol{\alpha}^o)}{\partial \boldsymbol{\alpha}^{o'}}\right)Q^{-1} + o_p(1)$$

$$\overset{d}{\sim} N(\mathbf{0}, Q^{-1}VQ^{-1}) \tag{4.62}$$

which is in the proof of White et al. (2015) on pp.187. So we substitute(4.61) back into (4.59) and get

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_{i_c,t}\left(\mathbb{1}_{\{y_{i,t}\leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\}} - \theta_{i,j}\right)$$

$$= -\frac{1}{\sqrt{T}}\left(\sum_{t=1}^{T}\sum_{i=1}^{n}\mathbf{Hit}'_t(\boldsymbol{\alpha}^o)\frac{\partial \boldsymbol{q}_t(\boldsymbol{\alpha}^o)}{\partial \boldsymbol{\alpha}^{o'}}\right)Q^{-1}G + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_{i_c,t}\mathrm{Hit}_{i,j,t}(\boldsymbol{\alpha}^o) + o_p(1). \tag{4.63}$$

Apply Assumption 5(i)-(iii) of White et al. (2015), the ergodic theorem and the martingale difference central limit theorem (see Theorem 3.35 and 5.24 of White (2001)) on (4.63) and obtain that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}y_{i_c,t}\left(\mathbb{1}_{\{y_{i,t}\leq q_{i,j,t}(\widehat{\boldsymbol{\alpha}})\}} - \theta_{i,j}\right)$$

$$\overset{d}{\sim} N\left(\mathbf{0}, \frac{1}{T}\theta_{i,j}(1 - \theta_{i,j})\right)\left[G'Q^{-1'}\sum_{t=1}^{T}\sum_{i\neq i_c}^{N}\frac{\partial q_{i,j,t}(\boldsymbol{\alpha}^o)}{\boldsymbol{\alpha}^o}\frac{\partial q_{i,j,t}(\boldsymbol{\alpha}^o)}{\boldsymbol{\alpha}^{o'}}Q^{-1}G\right]. \tag{4.64}$$

And we know that

$$\widehat{Q}_T \overset{p}{\to} Q,$$

$$\widehat{f}_{i,j,t}(0) - f_{i,j,t}(0) = o_p(1), \tag{4.65}$$

$$\frac{\partial q_{i,j,t}(\widehat{\boldsymbol{\alpha}})}{\partial \widehat{\boldsymbol{\alpha}}'} \overset{p}{\to} \frac{\partial q_{i,j,t}(\boldsymbol{\alpha}^o)}{\partial \boldsymbol{\alpha}^{o'}},$$

where $\widehat{Q}_T$ is the estimator given in Theorem 15, $\widehat{f}_{i,j,t}(0)$ is the estimator given in Theorem 14, and the last equality is obtained because that $\frac{\partial q_{i,j,t}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}$ is continuous in Θ and $\widehat{\boldsymbol{\alpha}} \overset{p}{\to} \boldsymbol{\alpha}^o$. So we also have that

$$\widehat{G}_T := \frac{1}{T}\sum_{t=1}^{T}y_{i_c,t}\widehat{f}_{i,j,t}(0)\frac{\partial q_{i,j,t}(\widehat{\boldsymbol{\alpha}})}{\partial \widehat{\boldsymbol{\alpha}}'} \overset{p}{\to} G, \tag{4.66}$$

and

$$\mathrm{DQ_{IS}} \overset{d}{\sim} \chi^2(1),$$

which concludes this proof. $\qquad\square$

Proof of Theorem 17.

As in the proof of Theorem 16, we apply the approximation result (4.60) to derive the limiting distribution of $\frac{1}{\sqrt{N_R}} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} \left(\mathbb{1}_{\left\{ y_{i,t} \leq q_{i,j,t}(\hat{\boldsymbol{\alpha}}) \right\}} - \theta \right)$ as follows:

$$
\begin{aligned}
&\lim_{R \to \infty} \frac{1}{\sqrt{N_R}} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} \left(\mathbb{1}_{\left\{ y_{i,t} \leq q_{i,j,t}(\hat{\boldsymbol{\alpha}}) \right\}} - \theta \right) \\
&= \lim_{R \to \infty} \frac{1}{\sqrt{N_R}} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} \left(\text{Hit}_{i,j,t}(\hat{\boldsymbol{\alpha}}) - \text{Hit}_{i,j,t}(\boldsymbol{\alpha}^o) + \text{Hit}_{i,j,t}(\boldsymbol{\alpha}^o) \right) \\
&= \lim_{R \to \infty} \frac{1}{\sqrt{N_R}} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} \left(\text{Hit}_{i,j,t}^{\oplus}(\hat{\boldsymbol{\alpha}}) - \text{Hit}_{i,j,t}^{\oplus}(\boldsymbol{\alpha}^o) + \text{Hit}_{i,j,t}(\boldsymbol{\alpha}^o) \right) \\
&= \lim_{R \to \infty} \frac{1}{\sqrt{N_R}} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} f_{i,j,t}(0) \frac{\partial q_{i,j,t}(\boldsymbol{\alpha}^o)}{\partial \boldsymbol{\alpha}^{o\prime}} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^o) + \lim_{R \to \infty} \frac{1}{\sqrt{N_R}} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} \text{Hit}_{i,j,t}(\boldsymbol{\alpha}^o), \\
&= \lim_{R \to \infty} \sqrt{\frac{N_R}{T_R}} \sqrt{T_R} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^o)' \frac{1}{N_R} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} f_{i,j,t}(0) \frac{\partial q_{i,j,t}(\boldsymbol{\alpha}^o)}{\partial \boldsymbol{\alpha}^o} + \lim_{R \to \infty} \frac{1}{\sqrt{N_R}} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} \text{Hit}_{i,j,t}(\boldsymbol{\alpha}^o), \\
&= \lim_{R \to \infty} \frac{1}{\sqrt{N_R}} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} \text{Hit}_{i,j,t}(\boldsymbol{\alpha}^o).
\end{aligned}
$$

$$(4.67)$$

Apply Assumption 5(i)-(iii) of White et al. (2015), the ergodic theorem and the martingale difference central limit theorem (see Theorem 3.35 and 5.24 of White (2001)) on (4.67) and obtain that

$$
\begin{aligned}
&\lim_{R \to \infty} \frac{1}{\sqrt{N_R}} \sum_{t=T_R+1}^{T_R+N_R} y_{i_c,t} \left(\mathbb{1}_{\left\{ y_{i,t} \leq q_{i,j,t}(\hat{\boldsymbol{\alpha}}) \right\}} - \theta \right) \\
&\stackrel{d}{\sim} N \left(\mathbf{0}, \frac{\theta_{i,j}(1 - \theta_{i,j})}{N_R} \sum_{t=1}^{N_R} y_{i_c,t}^2 \right)
\end{aligned}
$$

$$(4.68)$$

which leads to

$$
\text{DQ}_{\text{OOS}} \quad \stackrel{d}{\sim} \quad \chi^2(1), \tag{4.69}
$$

and concludes this proof. $\qquad \square$

5

Summary and Outlook

This chapter provides a brief summary of this dissertation. Following up this summary of our research work so far, an outlook of possible extensions to our work is also provided in the end.

This dissertation studies financial time series by quantile regressions. Quantile autoregressions (Koenker and Xiao, 2006) are applied in Chapter 2. Specifically, Chapter 2 introduced a new way to select between causal and noncausal models by comparing residuals from quantile autoregressions developed by Koenker and Xiao (2006) and from the time-reverse specifications. To adapt to heavy tailed distributions, we generalize the quantile autoregression theory for regularly varying distributions. This also confirms the validity of quantile autoregressions in analysing heavy tailed time series, such as explosive or bubble-type dynamics. It is natural to consider SRAR as a model selection criterion in the quantile regression framework. However due to the identification problem spotted in the SRAR plots as presented in this chapter, we propose to use the aggregate SRAR criterion for model selection. The robustness in its performance has been seen from all the results in this chapter. It is worth mentioning that when coefficients are constant in the underlying model with a symmetrically i.i.d. error term, the aggregate SRAR criterion is equivalently to select between forward and backward conditional mean models (termed by Gourieroux and Zakoian, 2017). However, the aggregate SRAR is a measure based on the whole dynamics of the underlying process, which is not dominated by the conditional mean information any more. This characteristic of the aggregate SRAR criterion indeed makes it robust in model selection even for some general situations such as with asymmetric distributed innovations.

In Chapter 3, we found that the inference test performance in CAViaR models is not robust and unsatisfying due to the estimation of the conditional probability

densities of time series. We found that the existing density estimation methods cannot fully adapt to time-varying conditional probability densities of CAViaR time series. So in this chapter we have developed a method called *adaptive random bandwidth* which can robustly approximate the time-varying conditional probability densities of CAViaR time series by Monte Carlo simulations. This method not only avoids the haunting problem of choosing an optimal bandwidth but also ensures the reliability of CAViaR analysis based on the asymptotic normality of the model parameter estimator.

In Chapter 4, we generalized multivariate multi-quantile CAViaR models (MVMQ-CAViaR, see White et al., 2015) by incorporating CoVaR specification (see Adrian and Brunnermeier, 2011) into the model specification in this chapter. The proposed systemic MVMQ-CAViaR model presents a vector-autoregressive (VAR) specification of financial institutions' value-at-risk (VaR) as well as their CoVaR. This model generalization is able to capture contemporaneous tail dependence of financial institutions and market indexes so that we can interpret the systemic risks of the institutions over time. The consistency and asymptotic normality proofs of this generalized model are provided in this chapter along with some relevant inference tests, for which we implemented simulation tests and showed robust model performances. For tracing the transmission of a single shock to a financial institution in the financial system, we also constructed quantile impulse response functions (QIRF) accordingly in use of the local projection idea (Jordà, 2005) and expansion of estimated terms. Based on our simulation results, we can see that using the expansion terms of $\widehat{q}_t$ is more robust than directly using $\widehat{q}_t$ in the local quantile regression for QIRF estimation. An empirical application performed on big banks with the market index *S&P* 500 shows the significant contemporaneous effects of the big banks on *S&P* 500 so supports our methodology.

I would like to point out some possible extensions to our work for future research. Regarding Chapter 2, the aggregate SRAR is proposed as a new model criterion but no asymptotic behaviours or theoretical validity of this criterion has been established yet. Bootstrapping can be a potential tool to estimate the underlying noncausal model selection rate given a time series. So a hypothesis testing can be preformed accordingly with the null hypothesis being that the model selection result is false. In addition, mixed causal and noncausal models are still underdeveloped because of model estimators' consistency issue. It is worth finding out a way to obtain consistent estimators for mix causal noncausal models so that model selection in mixed causal and noncausal models can be investigated as a generalization. Regarding Chapter 3, the adaptive random bandwidth method can be extended to general quantile regressions including multivariate cases in theory. Test performances can be checked for this method in high-dimensional cases to see its robustness. This method also has the potential to achieve the second-order accuracy to Wald tests of nonlinear restrictions (Phillips and Park,

1988; de Paula Ferrari and Cribari-Neto, 1993) in quantile regressions. Regarding Chapter 4, the proposed systemic MVMQ-CAViaR model requires a predetermined contemporaneous effecting order. An automatic way of determining the contemporaneous dependence direction can be considered to develop up in theory. Moreover, the functional form of a contemporaneous term has not been commonly on consensus or developed enough to pass most model specification tests. An investigation on the functional form of contemporaneous effects in financial markets can be helpful.

Bibliography

What is finance? - definition, overview, types of finance. (CFI, 2020).

Adam, M. C. and Szafarz, A. (1992). Speculative bubbles and financial markets. *Oxford Economic Papers*, 44(4):626–640.

Adrian, T. and Brunnermeier, M. K. (2011). Covar. Technical report, National Bureau of Economic Research.

Alessi, L., Barigozzi, M., and Capasso, M. (2011). Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, 79(1):16–47.

Andrews, B., Davis, R. A., and Breidt, F. J. (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis*, 97(7):1638–1659.

Barrodale, I. and Roberts, F. D. (1973). An improved algorithm for discrete l_1 linear approximation. *SIAM Journal on Numerical Analysis*, 10(5):839–848.

Breid, F. J., Davis, R. A., Lh, K.-S., and Rosenblatt, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, 36(2):175–198.

Brockwell, P. J. and Davis, R. A. (2016). *Introduction to time series and forecasting*. springer.

Brockwell, P. J., Davis, R. A., and Fienberg, S. E. (1991). *Time series: theory and methods*. Springer Science & Business Media.

Broze, L. and Szafarz, A. (1985). Solutions des modèles linéaires à anticipations rationnelles. In *Annales de l'INSEE*, pages 99–118. JSTOR.

Buchinsky, M. (1995). Estimating the asymptotic covariance matrix for quantile regression models a monte carlo study. *Journal of Econometrics*, 68(2):303–338.

Cagan, P. (1956). The monetary dynamics of hyperinflation. *Studies in the Quantity Theory if Money*.

Cavaliere, G., Nielsen, H. B., and Rahbek, A. (2020). Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling. *Journal of Business & Economic Statistics*, 38(1):55–67.

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

Chernozhukov, V. and Umantsev, L. (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics*, 26(1):271–292.

Choi, C.-Y. and Chudik, A. (2019). Estimating impulse response functions when the shock series is observed. *Economics Letters*, 180:71–75.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, pages 841–862.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues.

Davis, R. and Resnick, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *The Annals of Probability*, pages 179–195.

de Paula Ferrari, S. L. and Cribari-Neto, F. (1993). On the corrections to the wald test of non-linear restrictions. *Economics Letters*, 42(4):321–326.

Dhaene, G., Gourieroux, C., and Scaillet, O. (1998). Instrumental models and indirect encompassing. *Econometrica*, pages 673–688.

Duffie, D. and Pan, J. (1997). An overview of value at risk. *Journal of derivatives*, 4(3):7–49.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, pages 569–593. Springer.

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

Fan, J. and Fan, Y. (2010). Issues on quantile autoregression.

Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the american statistical association*, 93(441):359–371.

Francq, C. and Zakoïan, J.-M. (2007). Hac estimation and strong linearity testing in weak arma models. *Journal of Multivariate Analysis*, 98(1):114–144.

Fries, S. and Zakoian, J.-M. (2017). Mixed causal-noncausal ar processes and the modelling of explosive bubbles.

Gourieroux, C. and Hencic, A. (2014). Noncausal autoregressive model in application to bitcoin/usd exchange rate. *Econometrics of Risk", Series: Studies in Computational Intelligence, Springer*.

Gourieroux, C. and Jasiak, J. (2001). *Financial econometrics: Problems, models, and methods*, volume 2. Princeton University Press.

Gourieroux, C. and Jasiak, J. (2016). Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, 37(3):405–430.

Gourieroux, C. and Jasiak, J. (2018). Misspecification of noncausal order in autoregressive processes. *Journal of Econometrics*, 205(1):226–248.

Gourieroux, C., Jasiak, J., et al. (2015). *Semi-Parametric Estimation of Noncausal Vector Autoregression*. CREST.

Gouriéroux, C. and Zakoïan, J.-M. (2015). On uniqueness of moving average representations of heavy-tailed stationary processes. *Journal of Time Series Analysis*, 36(6):876–887.

Gouriéroux, C. and Zakoïan, J.-M. (2017). Local explosion modelling by noncausal process series b statistical methodology.

Han, H., Jung, W., and Lee, J. H. (2019). Estimation and inference of quantile impulse response functions by local projections: With applications to var dynamics. *Available at SSRN 3466198*.

Hecq, A., Issler, J. V., and Telg, S. (2020). Mixed causal–noncausal autoregressions with exogenous regressors. *Journal of Applied Econometrics*, 35(3):328–343.

Hecq, A., Lieb, L., and Telg, S. (2016). Identification of mixed causal-noncausal models in finite samples. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, (123/124):307–331.

Hecq, A., Lieb, L., and Telg, S. (2017a). Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The marx package. *Available at SSRN 3015797*.

Hecq, A. and Sun, L. (2020). Selecting between causal and noncausal models with quantile autoregressions. *Studies in Nonlinear Dynamics & Econometrics*, 1(ahead-of-print).

Hecq, A. and Sun, L. (2021). Adaptive random bandwidth for inference in caviar models. *arXiv preprint arXiv:2102.01636*.

Hecq, A., Telg, S., and Lieb, L. (2017b). Do seasonal adjustments induce noncausal dynamics in inflation rates? *Econometrics*, 5(4):48.

Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American statistical Association*, 87(417):58–68.

Herce, M. A. (1996). Asymptotic theory of" lad" estimation in a unit root process with finite variance errors. *Econometric Theory*, pages 129–153.

Homm, U. and Breitung, J. (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics*, 10(1):198–231.

Huber, P. J. et al. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233. University of California Press.

Hylleberg, S., Engle, R. F., Granger, C. W., and Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of econometrics*, 44(1-2):215–238.

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

Jung, W. and Lee, J. H. (2019). Quantile impulse response analysis with applications in macroeconomics and finance.

Kim, T.-H. and White, H. (2003). Estimation, inference, and specification testing for possibly misspecified quantile regression. *Advances in Econometrics*, 17:107–132.

Knight, K. (1989). Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pages 261–278.

Knight, K. (1991). Limit theory for m-estimates in an integrated infinite variance process. *Econometric Theory*, pages 200–212.

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

Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50.

Koenker, R. and Machado, J. A. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448):1296–1310.

Koenker, R. and Xiao, Z. (2004). Unit root quantile autoregression inference. *Journal of the American Statistical Association*, 99(467):775–787.

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

Koenker, R. W. and d'Orey, V. (1987). Algorithm as 229: Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 36(3):383–393.

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147.

Lanne, M. and Luoto, J. (2013). Autoregression-based estimation of the new keynesian phillips curve. *Journal of Economic Dynamics and Control*, 37(3):561–570.

Lanne, M. and Luoto, J. (2017). A new time-varying parameter autoregressive model for us inflation expectations. *Journal of Money, Credit and Banking*, 49(5):969–995.

Lanne, M., Luoto, J., and Saikkonen, P. (2012). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, 28(3):623–631.

Lanne, M. and Saikkonen, P. (2011a). Gmm estimation with non-causal instruments. *Oxford Bulletin of Economics and Statistics*, 73(5):581–592.

Lanne, M. and Saikkonen, P. (2011b). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, 3(3).

Lanne, M. and Saikkonen, P. (2013). Noncausal vector autoregression. *Econometric Theory*, pages 447–481.

Lii, K.-S. and Rosenblatt, M. (1993). Non-gaussian autoregressive moving average processes. *Proceedings of the National Academy of Sciences*, 90(19):9168–9170.

Lippi, M. and Reichlin, L. (1994). Var analysis, nonfundamental representations, blaschke matrices. *Journal of Econometrics*, 63(1):307–325.

Liu, X. (2019). Quantile-based asymmetric dynamics of real gdp growth. *Macroeconomic Dynamics*, pages 1–29.

Machado, J. A. and Silva, J. (2013). Quantile regression and heteroskedasticity. *https://jmcss. som. surrey. ac. uk/JM_JSS. pdf. Accessed*, 5(7):2015.

Montes-Rojas, G. (2019). Multivariate quantile impulse response functions. *Journal of Time Series Analysis*, 40(5):739–752.

Pagan, A. (1996). The econometrics of financial markets. *Journal of empirical finance*, 3(1):15–102.

Phillips, P. C. and Park, J. Y. (1988). On the formulation of wald tests of non-linear restrictions. *Econometrica: Journal of the Econometric Society*, pages 1065–1083.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, 7(2):186–199.

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. *Nonparametric and semiparametric methods in Econometrics*, pages 357–384.

Shephard, N. (1996). Statistical aspects of arch and stochastic volatility. *Monographs on Statistics and Applied Probability*, 65:1–68.

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359.

Taylor, J. B. (1993). Discretion versus policy rules in practice. In *Carnegie-Rochester conference series on public policy*, volume 39, pages 195–214. Elsevier.

Tokdar, S. T., Kadane, J. B., et al. (2012). Simultaneous linear quantile regression: a semiparametric bayesian approach. *Bayesian Analysis*, 7(1):51–72.

Weiss, A. A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory*, 7(1):46–68.

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

White, H. (2014). *Asymptotic theory for econometricians*. Academic press.

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

Impact Paragraph

The value of research is in why and how it can benefit our society. Realizing the value of a research result can be direct through applications in the research context or be potential by researchers' insights on the work. In this chapter, we discuss the values of the research results in this dissertation.

Quantile regressions are applied throughout this dissertation and used to studying a particular quantile of our interest. You might ask that when we care about a particular quantile. The answer is that when we care about an outcome and know the outcome is in uncertainty so we want to find a threshold of this random outcome for securing our position when it happens. For individuals, we can concern about how much cash to be kept during a trip. If you are careless between cash and bank cards, you still somehow concern about your financial positions. You might think about how much you can invest in an asset based on your household balance sheets in the future. If you are at the standing point of a group, it is necessary to secure the financial position of the whole group so that the group can be still functioning at an optimal chance and cost while an adverse outcome happen. Besides financial positions, groups might also concern about other welfare positions, such as capacity for urgent patients in a hospital. We can see that quantiles are common in our concerns. And the tool of quantile regressions is straightforward for achieving a quantile of our interest.

Next, we are going to discuss about what we have researched with regard to quantile regressions in this dissertation, what information we can convey for benefiting our society, and what tools we provide for the society to reliably apply quantile regressions in their applications.

Chapter 2 introduces a new way to select between causal and noncausal models by comparing residuals from quantile autoregressions developed by Koenker and Xiao (2006) and from the time-reverse specifications. When a noncausal model is selected for a time series, it seems that the future of the time series leads its movement. Actually, this noncausal model selection can be interpreted as the time series moved recklessly due to some external force. And it got dragged back by its intrinsic force until the intrinsic force surpasses the external one as the time series deviated far enough or long enough from its intrinsic level. This rule is not hard to understand with elastic bands. After identifying this noncausal characteristic and the external force in the time series, we can better cooperate with the time series at our positions. For instance, when a noncausal model is selected for a country's inflation rate time series, the local policy-maker should regulate and censor carefully such external forces in order to keep the inflation stable.

In Chapter 3, we found that the inference test performance in CAViaR models is not robust and unsatisfying always due to the inaccuracy in estimating the conditional probability densities of time series. We found that the existing density estimation methods cannot fully adapt to time-varying conditional probability densities of CAViaR time series. Implementing probabilistic modelling but without reliable inference testing, such as Wald tests for model specification testing, can be misleading. So we have developed up a method called *adaptive random bandwidth* (ARB) which can robustly approximate the time-varying conditional probability densities of CAViaR time series by Monte Carlo simulations. This method is a success in avoiding the haunting problem of choosing an optimal bandwidth and ensures the reliability of any CAViaR analysis based on the asymptotic normality of the model parameter estimator. This proposed method can be extended to general quantile regressions including multivariate cases easily and robustly. Having an accurate test statistic is important to obtain reliable models in financial applications. This ARB tool indeed helps our society to ensure reliability in analysing quantile regressions so as in the obtained quantiles for applications.

In Chapter 4, we generalized multivariate multi-quantile CAViaR models (MVMQ-CAViaR, see White et al., 2015) by incorporating CoVaR specification (see Adrian and Brunnermeier, 2011) into the model specification. This model generalization is able to capture contemporaneous tail dependence of financial institutions and market indexes so that we can interpret the systemic risks of the institutions over time. Systemic risk should be carefully managed and censored over time as we know that it is a major contributor to the financial crisis of 2008. The consistency and asymptotic normality proofs of this generalized model are provided along with some relevant inference tests, for which we implemented simulation tests and showed robust model performances. For tracing the transmission of a single shock to a financial institution in the financial system, we also constructed quantile impulse response functions (QIRF) accordingly in use of the local projection idea (Jordà, 2005) and expansion of estimated terms. Based on our simulation results, we can see that using the expansion terms of $\widehat{q}_t$ is more robust than directly using $\widehat{q}_t$ in the local quantile regression for QIRF estimation. The research work in this chapter helps policy-makers to take the contemporaneous effect in account for measuring the systemic risk of a financial institution, and also provides complementary statistical tools for them to better supervise the financial institution based on the systemic risk to the financial market therein.

About the Author

Li Sun was born in Dalian, China on April 6th, 1991. In September 2010, she started a Bachelor of Science program on Applied Mathematics in Dalian University of Technology. Four years later, with achieving the BSc degree in August 2014, she was admitted into Delft University of Technology for further studying Applied Mathematics in a Master of Science program. Under the supervision of Prof.dr.ir. C.W. Oosterlee for her master thesis, she accomplished the MSc program and achieved the MSc degree on 31st August, 2016. During her MSc program, she did a six-month internship on mortgage offer options for Marke& ALM Treasury Risk Modelling Department in ABN AMRO Bank, the Netherlands.

From September 2017 until February 2021, she was a PhD student at Maastricht University, under the supervision of Prof.dr. Alain W. Hecq. During her PhD, she worked on econometrics and mainly researched in quantile regressions and financial risk management. The findings of her PhD research are presented in this dissertation. Li also presented her research work at various international conferences, seminars and workshops. She also refereed for Journal of Econometrics and Statistics.

In March 2021, she will start a post-doctoral fellowship at University of Liege, Belgium.