Econometric Theory, **38**, 2022, 793–839.
doi:10.1017/S026646662100030X

QUANTILE DOUBLE AUTOREGRESSION

QIANQIAN ZHU
Shanghai University of Finance and Economics

GUODONG LI
University of Hong Kong

Many financial time series have varying structures at different quantile levels, and also exhibit the phenomenon of conditional heteroskedasticity at the same time. However, there is presently no time series model that accommodates both of these features. This paper fills the gap by proposing a novel conditional heteroskedastic model called "quantile double autoregression". The strict stationarity of the new model is derived, and self-weighted conditional quantile estimation is suggested. Two promising properties of the original double autoregressive model are shown to be preserved. Based on the quantile autocorrelation function and self-weighting concept, three portmanteau tests are constructed to check the adequacy of the fitted conditional quantiles. The finite sample performance of the proposed inferential tools is examined by simulation studies, and the need for use of the new model is further demonstrated by analyzing the S&P500 Index.

1. INTRODUCTION

Conditional heteroskedastic models have become a standard family of nonlinear time series models since the introduction of Engle's (1982) autoregressive conditional heteroskedastic (ARCH) and Bollerslev's (1986) generalized autoregressive conditional heteroskedastic (GARCH) models. Among the existing conditional heteroskedastic models, the double autoregressive (AR) model has recently attracted increasing attention (see Ling, 2004, 2007; Ling and Li, 2008; Zhu and Ling, 2013; Li, Ling, and Zhang, 2016; Li et al., 2017; Zhu, Zheng, and Li, 2018 and the references therein). The double AR model takes the form of

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \sqrt{\omega + \sum_{j=1}^{p} \beta_j y_{t-j}^2}, \tag{1.1}$$

We are grateful to the Editor, Co-Editor, Associate Editor, and three anonymous referees for their valuable comments and suggestions, which led to substantial improvement of this article. Zhu's research was supported by an NSFC grant 12001355, Shanghai Pujiang Program 2019PJC051, and Shanghai Chenguang Program 19CG44. Li's research was partially supported by a Hong Kong RGC grant 17304617 and an NSFC grant 72033002. Address correspondence to Qianqian Zhu, School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China; e-mail: zhu.qianqian@mail.shufe.edu.cn.

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

where $\omega > 0$, $\beta_j \geq 0$ with $1 \leq j \leq p$, and $\{\varepsilon_t\}$ are identically and independently distributed (i.i.d.) random variables with mean zero and variance one. It is a special case of the AR–ARCH model in Weiss (1986), and reduces to Engle's (1982) ARCH model when all ϕ_i's are zero. The double AR model has two novel properties. First, it has a larger parameter space than that of the commonly used AR model. For example, when $p = 1$, the double AR model can still be stationary even as $|\phi_1| \geq 1$ (Ling, 2004), which is impossible for AR–ARCH models. Second, no moment condition on y_t is needed to derive the asymptotic normality of the Gaussian quasi-maximum likelihood estimator (QMLE; Ling, 2007). This is in contrast to the autoregressive moving average (ARMA)–GARCH model, for which the finite fourth moment of the process is unavoidable in deriving the asymptotic distribution of the Gaussian QMLE (Francq and Zakoian, 2004), resulting in a much narrower parameter space (Li and Li, 2009).

Meanwhile, conditional heteroskedastic models are used primarily to model volatility and financial risk. Some quantile-based measures, such as the value-at-risk (VaR), expected shortfall, and limited expected loss, are intimately related to quantile estimation (see, e.g., Artzner et al., 1999; Wu and Xiao, 2002; Bassett, Koenker, and Kordas, 2004; Francq and Zakoian, 2015). Therefore, it is natural to consider quantile estimation for conditional heteroskedastic models. Many researchers have investigated conditional quantile estimation (CQE) for the (G)ARCH models (see Koenker and Zhao, 1996 for linear ARCH models, Xiao and Koenker, 2009 for linear GARCH models, and Lee and Noh, 2013; Zheng et al., 2018 for quadratic GARCH models). Chan and Peng (2005) investigated weighted least absolute deviation (LAD) estimation for double AR models with the order of $p = 1$, and Zhu and Ling (2013) studied quasi-maximum exponential likelihood estimation for a general double AR model. Cai, Montes-Rojas, and Olmo (2013) explored Bayesian estimation of model (1.1) with ε_t following a generalized lambda distribution. However, performing CQE for double AR models remains an open problem.

In addition, for the double AR process generated by model (1.1), its τth conditional quantile has the form of

$$Q_\tau(y_t|\mathcal{F}_{t-1}) = \sum_{i=1}^{p} \phi_i y_{t-i} + b_\tau \sqrt{\omega + \sum_{j=1}^{p} \beta_j y_{t-j}^2}, \quad 0 < \tau < 1, \tag{1.2}$$

and the coefficients of the ϕ_i's are all τ-independent, where $\mathcal{F}_t$ is the σ-field generated by $\{y_s, s \leq t\}$, and b_τ is the τth quantile of ε_t (see also Cai et al., 2013). However, when modeling the closing prices of the S&P500 Index by CQE, we found the estimated coefficients of the ϕ_i's to be significantly dependent on the quantile level, whereas those of the β_j's were slightly dependent on that level (see Figures 6 and 7 in Section 6 for empirical evidence). This phenomenon can also be found in many other stock indices. Koenker and Xiao (2006) proposed a quantile AR model by extending the common autoregression, with the corresponding coefficients defined as functions of quantile levels. By adapting the method in

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Koenker and Xiao (2006), this paper introduces a new conditional heteroskedastic model called "quantile double autoregression" to better interpret financial time series with the aforementioned phenomenon. We state our first main contribution below in detail.

(a) A direct extension of Koenker and Xiao's method results in a strong constraint: the coefficients of the y_{t-j}^2's are simultaneously zero at a certain quantile level (see Section 2). A novel transformation of $S_{\mathrm{Q}}(x) = \sqrt{|x|}\,\mathrm{sgn}(x)$ is first introduced to the conditional scale structure in Section 2, thereby removing that constraint from the derived model. Moreover, there is no nonnegative constraint on the coefficient functions, which renders numerical optimization more flexible. Section 2 also establishes the strict stationarity and ergodicity of the newly proposed model, and shows that the first novel property of double AR models is preserved.

Many financial time series are heavy-tailed, and the LAD estimator of infinite variance AR models usually has a faster convergence rate, but a more complicated asymptotic distribution, than the commonly encountered asymptotic normality (see, e.g., Gross and Steiger, 1979; An and Chen, 1982; Davis, Knight, and Liu, 1992). Ling (2005) proposed self-weighted LAD estimation, where the weights are used to reduce the moment condition on the process. As a result, asymptotic normality can be obtained, although superconsistency is sacrificed. The self-weighting approach has also been applied to quasi-maximum likelihood estimation for ARMA–GARCH models (Zhu and Ling, 2011) and augmented double AR models (Jiang, Li, and Zhu, 2020), as well as to CQE for linear double AR models (Zhu et al., 2018). The CQE may suffer from a similar problem. Actually, the finite third moment on $\{y_t\}$ is inevitable for the asymptotic normality of CQE (Zhu and Ling, 2011), which makes the resulting parameter space narrower (see Section 2).

(b) Motivated by Ling (2005), this paper considers self-weighted CQE for the newly proposed model in Section 3. Only a fractional moment on $\{y_t\}$ is needed for asymptotic normality, and the second novel property of double AR models is thus preserved. As a result, the proposed methodology can be applied to very heavy-tailed data. Moreover, the objective function in this paper is nondifferentiable and nonconvex, which causes a great difficulty in asymptotic derivations. Section 3 overcomes that difficulty by adopting the bracketing method in Pollard (1985), which is another important contribution of the paper.

According to the Box–Jenkins three-stage modeling strategy, it is important to check the adequacy of the fitted conditional quantiles, and the autocorrelation function (ACF) of residuals is supposed to play an important role in the literature (see Box and Jenkins, 1970; Li, 2004). Li, Li, and Tsai (2015) introduced a quantile correlation to measure the linear relationship between any two random variables for a given quantile in quantile regression settings, and proposed a quantile autocorrelation function (QACF) to assess the adequacy of the fitted

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

quantile AR models (Koenker and Xiao, 2006). Meanwhile, diagnostic checking for the conditional location usually uses fitted residuals (Ljung and Box, 1978), whereas that for the conditional scale employs absolute residuals (Li and Li, 2005). It is noteworthy that the proposed quantile double AR model consists of both the conditional location and scale parts.

(c) In line with the estimating procedure in Section 3, Section 4 first introduces a self-weighted QACF of residuals to measure the quantile correlation, and then constructs two portmanteau tests to check the adequacy of the fitted conditional quantiles in the conditional location and scale components, respectively. A combined test is also considered. This is the third main contribution of this paper.

In addition, Section 5 conducts simulation studies to evaluate the finite sample performance of the self-weighted estimators and portmanteau tests, and Section 6 presents an empirical example to illustrate the usefulness of the new model and its inferential tools. The conclusion and the discussion appear in Section 7. All technical details are relegated to the Appendix. The Supplementary Material is also available online. Throughout the paper, $\to_d$ denotes convergence in distribution, and $o_p(1)$ denotes a sequence of random variables converging to zero in probability. We denote by $\|\cdot\|$ the norm of a matrix or column vector, defined as $\|A\| = \sqrt{\mathrm{tr}(AA')} = \sqrt{\sum_{i,j} a_{ij}^2}$.

2. QUANTILE DOUBLE AUTOREGRESSION

This section introduces a new conditional heteroskedastic model by generalizing the double AR model at (1.1) or (1.2). The new model can also be used to accommodate the phenomenon that the financial time series may have varying structures at different quantile levels.

We first consider a direct extension of model (1.2), which, as in Zhu, Zheng, and Li (2018), can be reparameterized into $Q_\tau(y_t|\mathcal{F}_{t-1}) = \sum_{i=1}^p \phi_i y_{t-i} + b_\tau^* \sqrt{1 + \sum_{j=1}^p \beta_j^* y_{t-j}^2}$, where $b_\tau^* = b_\tau\sqrt{\omega}$ and $\beta_j^* = \beta_j/\omega$ with $1 \le j \le p$. Along the lines of quantile autoregression (Koenker and Xiao, 2006), by letting the corresponding coefficients depend on the quantile level τ, we have

$$Q_\tau(y_t|\mathcal{F}_{t-1}) = \sum_{i=1}^p \phi_i(\tau) y_{t-i} + b^*(\tau)\sqrt{1 + \sum_{j=1}^p \beta_j^*(\tau) y_{t-j}^2}, \quad 0 < \tau < 1. \tag{2.1}$$

However, if there exists τ such that $b^*(\tau) = 0$, then all of the y_{t-j}^2's disappear from the quantile structure at that level; i.e., the contributions of all y_{t-j}^2's are simultaneously zero at a certain quantile level. In addition, both coefficient functions, $b^*(\tau)$ and $\beta_j^*(\tau)$, are related to the same term of y_{t-j}^2, and it may not be a good idea to separate the influence of y_{t-j}^2 on $Q_\tau(y_t|\mathcal{F}_{t-1})$ into two τ-dependent functions. As a result, the extension at (2.1) may not be a good choice.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

This paper attempts to tackle the drawback of model (2.1) by looking for a way to move $b_\tau \in \mathbb{R}$ at (1.2) inside the square root. Doing so then leads to a transformation of $S_{\mathrm{Q}}(x) = \sqrt{|x|}\,\mathrm{sgn}(x)$, which is an extension of the square root function with support from $\mathbb{R}_+ = [0, +\infty)$ to $\mathbb{R}$, where $\mathrm{sgn}(\cdot)$ is the sign function. More specifically, model (1.2) can be reparameterized into $Q_\tau(y_t|\mathcal{F}_{t-1}) = \sum_{i=1}^p \phi_i y_{t-i} + S_{\mathrm{Q}}(b_\tau^\star + \sum_{j=1}^p \beta_j^\star y_{t-j}^2)$ alternatively, where $b_\tau^\star = b_\tau^2\omega$ and $\beta_j^\star = b_\tau^2\beta_j$ with $1 \le j \le p$. By letting the ϕ_i's, $b_\tau^\star$, and $\beta_j^\star$'s depend on the quantile level τ, we then propose the following new conditional heteroskedastic model:

$$Q_\tau(y_t|\mathcal{F}_{t-1}) = \sum_{i=1}^p \phi_i(\tau) y_{t-i} + S_{\mathrm{Q}}\left(b(\tau) + \sum_{j=1}^p \beta_j(\tau) y_{t-j}^2\right), \quad \text{for all } 0 < \tau < 1, \tag{2.2}$$

where $b(\cdot)$, $\phi_i(\cdot)$'s, and $\beta_j(\cdot)$'s with $1 \le i,j \le p$ are continuous functions $(0,1) \to \mathbb{R}$, and the model will reduce to model (1.2) when $b(\tau) = b_\tau^\star$, $\phi_j(\tau) = \phi_j$, and $\beta_j(\tau) = \beta_j^\star$ with $1 \le j \le p$. It is not necessary to assume that $b(\cdot)$ and $\beta_j(\cdot)$'s are all equal to zero at a certain quantile level, and the drawback of the definition at (2.1) is thus removed.

Assume that, as in Koenker and Xiao (2006), the right-hand side of (2.2) is an increasing function with respect to τ. Let $\{u_t\}$ be a sequence of i.i.d. standard uniform random variables, and model (2.2) then has an equivalent form:

$$y_t = \sum_{i=1}^p \phi_i(u_t) y_{t-i} + S_{\mathrm{Q}}\left(b(u_t) + \sum_{j=1}^p \beta_j(u_t) y_{t-j}^2\right). \tag{2.3}$$

It will reduce to the quantile AR model in Koenker and Xiao (2006) when the $\beta_j(u_t)$'s are zero with probability one. We call model (2.2) or (2.3) quantile double AR for simplicity.

Remark 1. The proposed quantile double AR is a generalization of the double AR model at (1.1). Specifically, let $\omega = E(|b(u_t)|) = \int_0^1 |b(\tau)|d\tau$ and $\varepsilon_t = S_Q[b(u_t)]/\sqrt{\omega}$. Assume that $\phi_i(u_t) = \phi_i$ with $1 \le i \le p$, and $\beta_j(u_t) = b(u_t)\beta_j/\omega$ with $1 \le j \le p$ with probability one, and then (2.3) has the same form as in (1.1). Moreover, we may further assume that $E(\varepsilon_t) = \int_0^1 S_Q[b(\tau)]d\tau/\sqrt{\omega} = 0$.

Remark 2. For model (2.2), consider the case with $p = 1$ and $\beta_1(\tau) = 0$. Phillips (2015) showed that there exists a positive probability such that the right-hand side is not an increasing function with respect to τ, so that the model is misspecified. Koenker and Xiao (2006) discussed the quantile crossing problem and showed that a reparameterization is possible for which co-monotonicity applies at selected points. We follow that idea in the discussion. Suppose that all $b(\cdot)$ and $\beta_j(\cdot)$'s with $1 \le j \le p$ are increasing functions. Note that $S_{\mathrm{Q}}(\cdot)$ is continuous and monotonically increasing, and hence that the second term on the right-hand side of (2.2) is increasing with respect to τ. If $\sum_{i=1}^p \phi_i(\tau) y_{t-i}$ is also an increasing function with

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

respect to τ (Koenker and Xiao, 2006), then it is guaranteed that (2.2) defines a qualified conditional quantile function.

Remark 3. The AR–ARCH model is another popular conditional heteroskedastic model in the literature, and it is certainly of interest to consider a similar generalization for it. The AR–ARCH model takes the form of

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + e_t, \quad e_t = \varepsilon_t \sigma_t, \quad \sigma_t^2 = \omega + \sum_{j=1}^{p} \beta_j e_{t-j}^2,$$

where the coefficients and $\{\varepsilon_t\}$ are defined as in (1.1), and its τth conditional quantile is given as

$$Q_\tau(y_t|\mathcal{F}_{t-1}) = \sum_{i=1}^{p} \phi_i y_{t-i} + b_\tau \sqrt{\omega + \sum_{j=1}^{p} \beta_j [y_{t-j} - \sum_{i=1}^{p} \phi_i y_{t-j-i}]^2}, \quad 0 < \tau < 1,$$

which, however, has a more complicated form than model (1.2). This makes it impossible to reach an easy-to-interpret quantile model. In addition, compared with AR–ARCH models, the double AR has many other advantages (see Section 1).

Let $\boldsymbol{y}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$, where $\{y_t\}$ is a quantile double AR process generated by model (2.3). It can be verified that $\{\boldsymbol{y}_t\}$ is a homogeneous Markov chain on the state space $(\mathbb{R}^p, \mathcal{B}^p, \nu^p)$, where $\mathcal{B}^p$ is the class of Borel sets of $\mathbb{R}^p$, and ν^p is the Lebesgue measure on $(\mathbb{R}^p, \mathcal{B}^p)$. Denote by $F(\cdot)$ and $f(\cdot)$ the distribution and density functions of $b(u_t)$, respectively. It then holds that $F(\cdot) = b^{-1}(\cdot)$ and $E|b(u_t)|^\delta = \int_0^1 |b(\tau)|^\delta d\tau$, for any $\delta > 0$.

Assumption 1. The density function $f(\cdot)$ is positive and continuous on $\{x \in \mathbb{R} : 0 < F(x) < 1\}$, and $\int_0^1 |b(\tau)|^{\kappa/2} d\tau < \infty$, for some $0 < \kappa \leq 1$.

THEOREM 1. *Under Assumption 1, if*

$$\sum_{i=1}^{p} \max\left\{ \int_0^1 \left|\phi_i(\tau) - \sqrt{|\beta_i(\tau)|}\right|^\kappa d\tau, \int_0^1 \left|\phi_i(\tau) + \sqrt{|\beta_i(\tau)|}\right|^\kappa d\tau \right\} < 1,$$

then there exists a strictly stationary solution $\{y_t\}$ to the quantile double AR model at (2.3), and that solution is unique and geometrically ergodic with $E|y_t|^\kappa < \infty$.

Since u_t is a standard uniform random variable, the condition in the above theorem is equivalent to $\sum_{i=1}^{p} \max\left\{E\left|\phi_i(u_t) - \sqrt{|\beta_i(u_t)|}\right|^\kappa, E\left|\phi_i(u_t) + \sqrt{|\beta_i(u_t)|}\right|^\kappa\right\} < 1$. Moreover, for the double AR model at (1.1), Ling (2007) derived its stationarity condition for the case of ε_t being a standard normal random variable, although it remains unknown for the other distributions of ε_t. From Theorem 1 and Remark 1, we can obtain a stationarity condition for a general double AR model as follows.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

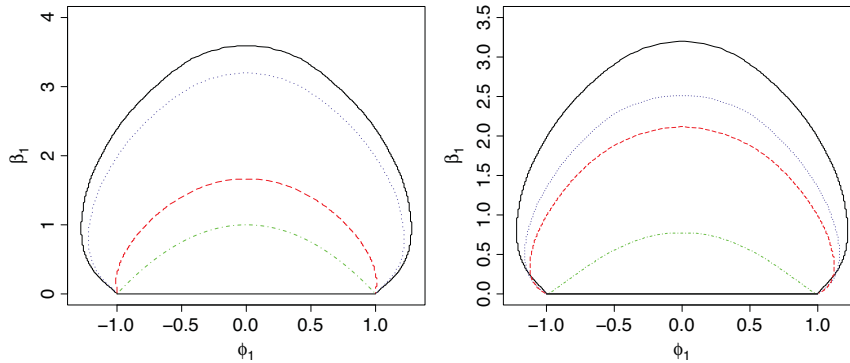


FIGURE 1. Left panel: stationarity regions of $E\ln|\phi_1+\varepsilon_t\sqrt{\beta_1}|<0$ (solid line), $E|\phi_1+\varepsilon_t\sqrt{\beta_1}|^\kappa<1$ with $\kappa=0.1$ (dotted line), $\kappa=0.9$ (long-dashed line), or $\kappa=2$ (dot-dashed line), respectively, where ε_t follows a standard normal distribution. Right panel: stationarity regions of $E|\phi_1+\varepsilon_t\sqrt{\beta_1}|^\kappa<1$ with $\kappa=0.1$ and ε_t being standard normal (solid line), Student's t_5 (dotted line), t_3 (long-dashed line), or standard Cauchy (dot-dashed line) random variable, respectively.

COROLLARY 1. *Suppose that ε_t has a positive and continuous density on its support, and $E|\varepsilon_t|^\kappa<\infty$, for some $0<\kappa\le 1$. If $\sum_{i=1}^p \max\left\{E|\phi_i-\varepsilon_t\sqrt{\beta_i}|^\kappa, E|\phi_i+\varepsilon_t\sqrt{\beta_i}|^\kappa\right\}<1$, then there exists a strictly stationary solution $\{y_t\}$ to the double AR model at (1.1), and that solution is unique and geometrically ergodic with $E|y_t|^\kappa<\infty$.*

For the case with the order of $p=1$, when ε_t is symmetrically distributed, the above condition can be simplified to $E|\phi_1+\varepsilon_t\sqrt{\beta_1}|^\kappa<1$. Moreover, if the normality of ε_t is further assumed, the necessary and sufficient condition for strict stationarity is then $E\ln|\phi_1+\varepsilon_t\sqrt{\beta_1}|<0$ (Ling, 2007). The comparison of the foregoing stationarity regions is illustrated in the left panel of Figure 1. It can be seen that a larger value of κ in Corollary 1 leads to a higher moment of y_t, thus resulting in a narrower stationarity region. Figure 1 also gives the stationarity regions with different distributions of ε_t. As expected, the parameter space of model (1.1) becomes smaller as ε_t becomes more heavy-tailed.

3. SELF-WEIGHTED CONDITIONAL QUANTILE ESTIMATION

Let $\boldsymbol{y}_{1,t}=(y_t,\ldots,y_{t-p+1})'$ and $\boldsymbol{y}_{2,t}=(y_t^2,\ldots,y_{t-p+1}^2)'$. For the proposed quantile double AR model at (2.2), denote by $\boldsymbol{\theta}_\tau=(\boldsymbol{\phi}'(\tau),b(\tau),\boldsymbol{\beta}'(\tau))'$ the parameter vector, where $\boldsymbol{\phi}(\tau)=(\phi_1(\tau),\ldots,\phi_p(\tau))'$ and $\boldsymbol{\beta}(\tau)=(\beta_1(\tau),\ldots,\beta_p(\tau))'$. We can then define the conditional quantile function as follows:

$$q_t(\boldsymbol{\theta}_\tau)=\boldsymbol{y}_{1,t-1}'\boldsymbol{\phi}(\tau)+S_{\mathrm{Q}}\left(b(\tau)+\boldsymbol{y}_{2,t-1}'\boldsymbol{\beta}(\tau)\right).$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

This paper also considers self-weighted CQE for model (2.2):

$$\widehat{\boldsymbol{\theta}}_{\tau n} = (\widehat{\boldsymbol{\phi}}_n'(\tau), \widehat{b}_n(\tau), \widehat{\boldsymbol{\beta}}_n'(\tau))' = \underset{\boldsymbol{\theta}_\tau}{\operatorname{argmin}} \sum_{t=p+1}^{n} w_t \rho_\tau \left(y_t - q_t(\boldsymbol{\theta}_\tau)\right), \tag{3.1}$$

where $\rho_\tau(x) = x[\tau - I(x < 0)]$ is the check function, and $\{w_t\}$ are nonnegative random weights (see also Ling, 2005; Zhu and Ling, 2011).

Remark 4. When $w_t = 1$, for all t, with probability one, self-weighted CQE becomes the common form of CQE. Third-order moments are needed for asymptotic normality, which leads to a much narrower stationarity region (see Figure 1 for an illustration). For heavy-tailed data, the LAD estimator of infinite variance AR models is shown to have a faster convergence rate but a more complicated asymptotic distribution (An and Chen, 1982; Davis et al., 1992), and a similar situation may be expected for CQE. By adopting the self-weighting method in Ling (2005), this paper suggests self-weighted CQE at (3.1) to maintain asymptotic normality even for heavy-tailed data with only a finite fractional moment.

Denote the true parameter vector by $\boldsymbol{\theta}_{\tau 0} = (\boldsymbol{\phi}_0'(\tau), b_0(\tau), \boldsymbol{\beta}_0'(\tau))'$. It is assumed to be an interior point of the parameter space $\Theta \subset \mathbb{R}^{2p+1}$, which is a compact set. Moreover, let $F_{t-1}(\cdot)$ and $f_{t-1}(\cdot)$ be the distribution and density functions of y_t conditional on $\mathcal{F}_{t-1}$, respectively.

Assumption 2. $\{y_t\}$ is strictly stationary and ergodic with $E|y_t|^\kappa < \infty$, for some $0 < \kappa \leq 1$.

Assumption 3. $\{w_t\}$ is strictly stationary and ergodic, and w_t is nonnegative and measurable with respect to $\mathcal{F}_{t-1}$ with $E[w_t \|\boldsymbol{y}_{1,t-1}\|^3] < \infty$.

Assumption 4. With probability one, $f_{t-1}(\cdot)$ and its derivative function $\dot{f}_{t-1}(\cdot)$ are uniformly bounded, and $f_{t-1}(\cdot)$ is positive on the support $\{x : 0 < F_{t-1}(x) < 1\}$.

Theorem 1 in Section 2 provides a sufficient condition for Assumption 2, and the proposed self-weighted CQE can handle very heavy-tailed data, since only a fractional moment is needed. Discussion of the random weights $\{w_t\}$ is delayed until Remark 7.

Remark 5. Assumption 4 is commonly used in the literature (see, e.g., Assumption 3 in Ling, 2005 for the self-weighted LAD estimator of infinite variance AR models, Assumption (A2) in Lee and Noh, 2013 for CQE of GARCH models, and Assumption 4 in Zhu, Li, and Xiao, 2021 for CQE of linear models with GARCH-X errors). More specifically, the strong consistency requires the positiveness and continuity of $f_{t-1}(\cdot)$, whereas the boundedness of $f_{t-1}(\cdot)$ and $\dot{f}_{t-1}(\cdot)$ is needed for $\sqrt{n}$-consistency and asymptotic normality. The primary purpose of Assumption 4 is to simplify the technical proofs, and it can certainly be weakened. For example, we could replace the boundedness of $\dot{f}_{t-1}(\cdot)$ by assuming the uniform continuity of $f_{t-1}(\cdot)$ on a closed interval. However, a lengthy technical proof would be required.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

THEOREM 2. *Suppose that* $\min\{|b(\tau)|, |\beta_1(\tau)|, \ldots, |\beta_p(\tau)|\} \geq C$ *for a constant* $C > 0$. *If Assumptions 2–4 hold, then* $\widehat{\boldsymbol{\theta}}_{\tau n} \to \boldsymbol{\theta}_{\tau 0}$ *almost surely as* $n \to \infty$.

Remark 6. The nonzero restriction on $b(\tau)$ and $\beta_j(\tau)$ with $1 \leq j \leq p$ serves two purposes: (i) the first-order derivative of $S_Q(x)$ does not exist at $x = 0$, and we need to bound the term of $b(\tau) + \boldsymbol{y}'_{2,t-1}\boldsymbol{\beta}(\tau)$ away from zero such that the first derivative of $q_t(\boldsymbol{\theta}_\tau)$ exists, and the asymptotic variance in Theorem 3 is then well defined; and (ii) this term will also be used to reduce the moment requirement on $\boldsymbol{y}_{1,t-1}$ or $\boldsymbol{y}_{2,t-1}$ in the technical proofs, and, without it, a moment condition on y_t would be required. For many financial time series, at quantile levels around $\tau = 0.5$, the values of $b(\tau)$ and $\beta_j(\tau)$'s are all close to zero, and hence we may fail to provide a reliable estimating result (see Figure 7 for details). This is actually a common problem for all CQE, and it can be solved by assuming a parametric form for these coefficient functions. We leave the issue for future research.

Denote $h_t(\boldsymbol{\theta}_\tau) = b(\tau) + \boldsymbol{y}'_{2,t-1}\boldsymbol{\beta}(\tau)$. Let $\dot{q}_t(\boldsymbol{\theta}_\tau) = (\boldsymbol{y}'_{1,t-1}, 0.5|h_t(\boldsymbol{\theta}_\tau)|^{-1/2}, 0.5|h_t(\boldsymbol{\theta}_\tau)|^{-1/2}\boldsymbol{y}'_{2,t-1})'$ be the first derivative of $q_t(\boldsymbol{\theta}_\tau)$. Define $(2p+1) \times (2p+1)$ symmetric matrices

$$\Omega_0(\tau) = E\big[w_t^2 \dot{q}_t(\boldsymbol{\theta}_{\tau 0})\dot{q}'_t(\boldsymbol{\theta}_{\tau 0})\big], \quad \Omega_1(\tau) = E\big[f_{t-1}(F_{t-1}^{-1}(\tau))w_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})\dot{q}'_t(\boldsymbol{\theta}_{\tau 0})\big]$$

and $\Sigma(\tau) = \tau(1-\tau)\Omega_1^{-1}(\tau)\Omega_0(\tau)\Omega_1^{-1}(\tau)$.

THEOREM 3. *Suppose that the conditions of Theorem 2 hold. If* $\Omega_1(\tau)$ *is positive definite, then* $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) \to_d N(\mathbf{0}, \Sigma(\tau))$ *as* $n \to \infty$.

The technical proof of Theorem 3 is nontrivial, since the objective function of self-weighted CQE is nonconvex and nondifferentiable, and the main difficulty is to prove $\sqrt{n}$-consistency, i.e., $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) = O_p(1)$. We overcome it by adopting the bracketing method (Pollard, 1985), as in Zhu and Ling (2011).

Remark 7. For random weights $\{w_t\}$, there are many choices that satisfy Assumption 3. For the double AR model at (1.2) and Remark 1, by a method similar to that for proving Theorem 3 in Zhu, Zheng, and Li (2018), we can show that the asymptotic variance $\Sigma(\tau)$ is minimized when $w_t = |b_0(\tau) + \boldsymbol{y}'_{2,t-1}\boldsymbol{\beta}_0(\tau)|^{-1/2}$ and $Ey_t^2 < \infty$. However, for a general quantile double AR, the asymptotic variance depends on the weights in a complicated way, and the optimal weights do not have an explicit form. In fact, the selection of optimal weights is still an open problem for AR models (Ling, 2005), and it becomes more complicated for a quantile model with the additional conditional scale structure in (2.2). Note that the key step of using the weights in technical proofs is to bound the term of $w_t y_{t-j}$ by $O(|y_{t-j}|^\delta)$ for some fractional value $\delta > 0$. As a result, Ling (2005) heuristically used $w_t = I(a_t = 0) + C^3 a_t^{-3} I(a_t \neq 0)$, where $a_t = \sum_{i=1}^p |y_{t-i}| I(|y_{t-i}| > C)$, for some constant $C > 0$. Following Zhu, Zheng, and Li (2018) and Jiang, Li, and Zhu (2020) for AR-type conditional heteroskedastic models, we simply employ the weights with $w_t = (1 + \sum_{i=1}^p |y_{t-i}|^3)^{-1}$.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Remark 8. To estimate the quantity of $f_{t-1}(F_{t-1}^{-1}(\tau))$ in the asymptotic variance in Theorem 3, we consider the difference quotient method (Koenker, 2005), i.e.,

$$\widehat{f}_{t-1}(F_{t-1}^{-1}(\tau)) = \frac{2h}{\widehat{Q}_{\tau+h}(y_t|\mathcal{F}_{t-1}) - \widehat{Q}_{\tau-h}(y_t|\mathcal{F}_{t-1})},$$

for some appropriate choices of bandwidth h, where $\widehat{Q}_\tau(y_t|\mathcal{F}_{t-1}) = q_t(\widehat{\boldsymbol{\theta}}_{\tau n})$ is the fitted τth conditional quantile. As in Koenker and Xiao (2006) and Li, Li, and Tsai (2015), we employ two commonly used choices for bandwidth h as follows:

$$h_B = n^{-1/5}\left\{\frac{4.5f_N^4(F_N^{-1}(\tau))}{[2(F_N^{-1}(\tau))^2+1]^2}\right\}^{1/5} \quad \text{and} \quad h_{HS} = n^{-1/3}z_\alpha^{2/3}\left\{\frac{1.5f_N^2(F_N^{-1}(\tau))}{2(F_N^{-1}(\tau))^2+1}\right\}^{1/3}, \tag{3.2}$$

where $f_N(\cdot)$ and $F_N(\cdot)$ are the standard normal density and distribution functions, respectively, and $z_\alpha = F_N^{-1}(1-\alpha/2)$ with α set to 0.05 (see Bofinger, 1975; Hall and Sheather, 1988). Bandwidth h_B is selected by minimizing the mean square error of Gaussian density estimation, whereas h_{HS} is obtained based on the Edgeworth expansion for studentized quantiles. The simulation results in Section 5 indicate that these two bandwidths have similar performance. As a result, the two matrices $\Omega_0(\tau)$ and $\Omega_1(\tau)$ can then be approximated by the sample averages:

$$\widehat{\Omega}_0(\tau) = \frac{1}{n}\sum_{t=p+1}^{n} w_t^2 \dot{q}_t(\widehat{\boldsymbol{\theta}}_{\tau n})\dot{q}_t'(\widehat{\boldsymbol{\theta}}_{\tau n}) \quad \text{and}$$

$$\widehat{\Omega}_1(\tau) = \frac{1}{n}\sum_{t=p+1}^{n} \widehat{f}_{t-1}(F_{t-1}^{-1}(\tau))w_t\dot{q}_t(\widehat{\boldsymbol{\theta}}_{\tau n})\dot{q}_t'(\widehat{\boldsymbol{\theta}}_{\tau n}),$$

where $\dot{q}_t(\widehat{\boldsymbol{\theta}}_{\tau n}) = (\boldsymbol{y}_{1,t-1}', 0.5|h_t(\widehat{\boldsymbol{\theta}}_{\tau n})|^{-1/2}, 0.5|h_t(\widehat{\boldsymbol{\theta}}_{\tau n})|^{-1/2}\boldsymbol{y}_{2,t-1}')'$. Consequently, a consistent estimator $\widehat{\Sigma}(\tau)$ of the asymptotic variance matrix $\Sigma(\tau)$ can be constructed.

From self-weighted CQE $\widehat{\boldsymbol{\theta}}_{\tau n}$, the τth quantile of y_t conditional on $\mathcal{F}_{t-1}$ can be estimated by $q_t(\widehat{\boldsymbol{\theta}}_{\tau n})$. The following corollary provides the theoretical justification for one-step-ahead forecasting, and is a direct result of the Taylor expansion and Theorem 3.

COROLLARY 2. *Under the conditions of Theorem 3 and* $E(|y_t|) < \infty$, *it holds that*

$$\widehat{Q}_\tau(y_{n+1}|\mathcal{F}_n) - Q_\tau(y_{n+1}|\mathcal{F}_n) = \dot{q}_{n+1}'(\boldsymbol{\theta}_{\tau 0})(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) + o_p(n^{-1/2}).$$

Remark 9. In practice, we may consider multiple quantile levels simultaneously, say $\tau_1 < \tau_2 < \cdots < \tau_K$, and the quantile crossing problem will happen frequently. Specifically, it is possible that neither $\{Q_{\tau_k}(y_{n+1}|\mathcal{F}_n)\}_{k=1}^K$ nor $\{\widehat{Q}_{\tau_k}(y_{n+1}|\mathcal{F}_n)\}_{k=1}^K$ is a monotonic increasing sequence (see Remark 2). This paper

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

adopts the rearranging method (Chernozhukov, Fernández-Val, and Galichon, 2010) to enforce the monotonicity of quantile estimates $\{\widehat{Q}_{\tau_k}(y_{n+1}|\mathcal{F}_n)\}_{k=1}^K$.

Remark 10. In real applications, we do not know the order of p at model (2.2), and an information criterion is thus introduced. At each quantile level of τ, we can consider the following Bayesian information criterion (BIC):

$$\mathrm{BIC}_\tau(p) = 2(n-p_{\max})\log(L_n(\widehat{\boldsymbol{\theta}}^p_{\tau n})) + (2p+1)\log(n-p_{\max}),$$

where $p_{\max}$ is a predetermined positive integer, $\widehat{\boldsymbol{\theta}}^p_{\tau n}$ is the self-weighted CQE defined in (3.1) with order p, and $L_n(\widehat{\boldsymbol{\theta}}^p_{\tau n}) = (n-p_{\max})^{-1}\sum_{t=p_{\max}+1}^n w_t\rho_\tau(y_t - q_t(\widehat{\boldsymbol{\theta}}^p_{\tau n}))$ with $w_t = (1+\sum_{i=1}^{p_{\max}}|y_{t-i}|^3)^{-1}$ (see also Machado, 1993; Zhu, Zheng, and Li, 2018). Note that the selected orders depend on τ, and we may need a uniform order for all quantile levels. As a result, we suggest the following combined version:

$$\begin{aligned}\mathrm{BIC}(p) &= \frac{1}{K}\sum_{k=1}^K \mathrm{BIC}_{\tau_k}(p)\\ &= 2(n-p_{\max})\frac{1}{K}\sum_{k=1}^K \log(L_n(\widehat{\boldsymbol{\theta}}^p_{\tau n})) + (2p+1)\log(n-p_{\max}),\end{aligned}$$

where $\tau_k = k/(K+1)$, with K being a fixed integer. Let $\widehat{p}_n = \arg\min_{1\le p\le p_{\max}}\mathrm{BIC}(p)$ be the selected order. From the simulation results in Section 5, we know that the proposed BIC has satisfactory performance. Note that the likelihood function of model (2.2) or (2.3) involves the inverse function of $Q_\tau(y_t|\mathcal{F}_{t-1})$ with respect to τ, which, however, has no closed form. Hence, it is not feasible to use the likelihood function to design an information criterion.

Remark 11. It is of interest to consider a quantile double autoregression with orders different for the conditional location and scale parts:

$$Q_\tau(y_t|\mathcal{F}_{t-1}) = \sum_{i=1}^{p_1}\phi_i(\tau)y_{t-i} + S_{\mathrm{Q}}\left(b(\tau) + \sum_{j=1}^{p_2}\beta_j(\tau)y_{t-j}^2\right),$$

for all $0<\tau<1$. The counterpart of model (2.3) can be given similarly. Let $p = \max\{p_1,p_2\}$ with $\phi_i(\cdot) = 0$, for $i > p_1$, or $\beta_j(\cdot) = 0$, for $j > p_2$. Theorem 1 and Corollary 1 can then be used to establish strict stationarity. For the case with $p_1 < p_2$, all of the theoretical results in this section still hold. However, when $p_1 > p_2$, the conditional scale of $S_Q(b(u_t) + \sum_{j=1}^{p_2}\beta_j(u_t)y_{t-j}^2)$ is not sufficient to reduce the moment requirement on y_t, and hence a higher-order moment of y_t is needed for theoretical justification. We leave this general setting for future research.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

4. DIAGNOSTIC CHECKING FOR CONDITIONAL QUANTILES

To check the adequacy of the fitted conditional quantiles, we construct two portmanteau tests to detect possible misspecifications in the conditional location and scale, respectively, and a combined test is also considered.

Let $\eta_{t,\tau} = y_t - Q_\tau(y_t|\mathcal{F}_{t-1}) = y_t - q_t(\boldsymbol{\theta}_{\tau 0})$ be the conditional quantile error. Instead of applying the QACF to quantile errors $\{\eta_{t,\tau}\}$ directly, we introduce a self-weighted QACF that naturally combines the QACF in Li, Li, and Tsai (2015) with the idea of self-weighting in Ling (2005). Specifically, the self-weighted QACF of $\{\eta_{t,\tau}\}$ at lag k is defined as

$$\rho_{k,\tau} = \mathrm{qcor}_\tau\{\eta_{t,\tau}, \eta_{t-k,\tau}\} = \frac{E[w_t\psi_\tau(\eta_{t,\tau})(\eta_{t-k,\tau} - \mu_{1,\tau})]}{\sqrt{(\tau-\tau^2)\sigma_{1,\tau}^2}}, \quad k = 1, 2, \ldots,$$

where $\psi_\tau(x) = \tau - I(x < 0)$, $\{w_t\}$ are the random weights used in Section 3, $\mu_{1,\tau} = E(\eta_{t,\tau})$, and $\sigma_{1,\tau}^2 = \mathrm{var}(\eta_{t,\tau})$. By replacing $\eta_{t-k,\tau}$ with $|\eta_{t-k,\tau}|$, a variant of $\rho_{k,\tau}$ can be defined as

$$r_{k,\tau} = \mathrm{qcor}_\tau\{\eta_{t,\tau}, |\eta_{t-k,\tau}|\} = \frac{E[w_t\psi_\tau(\eta_{t,\tau})(|\eta_{t-k,\tau}| - \mu_{2,\tau})]}{\sqrt{(\tau-\tau^2)\sigma_{2,\tau}^2}}, \quad k = 1, 2, \ldots,$$

where $\mu_{2,\tau} = E(|\eta_{t,\tau}|)$ and $\sigma_{2,\tau}^2 = \mathrm{var}(|\eta_{t,\tau}|)$. Note that if $Q_\tau(y_t|\mathcal{F}_{t-1})$ is correctly specified by model (2.2), then $\rho_{k,\tau} = 0$ and $r_{k,\tau} = 0$, for all $k \geq 1$.

Accordingly, denote by $\{\widehat{\eta}_{t,\tau}\}$ the conditional quantile residuals, where $\widehat{\eta}_{t,\tau} = y_t - \widehat{Q}_\tau(y_t|\mathcal{F}_{t-1}) = y_t - q_t(\widehat{\boldsymbol{\theta}}_{\tau n})$. The self-weighted residual QACFs at lag k can then be defined as

$$\widehat{\rho}_{k,\tau} = \frac{1}{\sqrt{(\tau-\tau^2)\widehat{\sigma}_{1,\tau}^2}} \frac{1}{n-p} \sum_{t=p+k+1}^{n} w_t\psi_\tau(\widehat{\eta}_{t,\tau})(\widehat{\eta}_{t-k,\tau} - \widehat{\mu}_{1,\tau})$$

and

$$\widehat{r}_{k,\tau} = \frac{1}{\sqrt{(\tau-\tau^2)\widehat{\sigma}_{2,\tau}^2}} \frac{1}{n-p} \sum_{t=p+k+1}^{n} w_t\psi_\tau(\widehat{\eta}_{t,\tau})(|\widehat{\eta}_{t-k,\tau}| - \widehat{\mu}_{2,\tau}),$$

where $\widehat{\mu}_{1,\tau} = (n-p)^{-1}\sum_{t=p+1}^{n}\widehat{\eta}_{t,\tau}$, $\widehat{\mu}_{2,\tau} = (n-p)^{-1}\sum_{t=p+1}^{n}|\widehat{\eta}_{t,\tau}|$, $\widehat{\sigma}_{1,\tau}^2 = (n-p)^{-1}\sum_{t=p+1}^{n}(\widehat{\eta}_{t,\tau} - \widehat{\mu}_{1,\tau})^2$, and $\widehat{\sigma}_{2,\tau}^2 = (n-p)^{-1}\sum_{t=p+1}^{n}(|\widehat{\eta}_{t,\tau}| - \widehat{\mu}_{2,\tau})^2$. For a predetermined positive integer K, let $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_{1,\tau}, \ldots, \widehat{\rho}_{K,\tau})'$ and $\widehat{\boldsymbol{r}} = (\widehat{r}_{1,\tau}, \ldots, \widehat{r}_{K,\tau})'$. We first derive the asymptotic distribution of $\widehat{\boldsymbol{\rho}}$ and $\widehat{\boldsymbol{r}}$.

Denote $H_{1k} = E[w_t f_{t-1}(F_{t-1}^{-1}(\tau))\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\eta_{t-k,\tau}]$ and $H_{2k} = E[w_t f_{t-1}(F_{t-1}^{-1}(\tau))\dot{q}_t(\boldsymbol{\theta}_{\tau 0})|\eta_{t-k,\tau}|]$. Let $\boldsymbol{\epsilon}_{1,t} = (\eta_{t,\tau}, \ldots, \eta_{t-K+1,\tau})'$ and $\boldsymbol{\epsilon}_{2,t} = (|\eta_{t,\tau}|, \ldots, |\eta_{t-K+1,\tau}|)'$. For $i = 1$ and 2, denote the $K \times (2p+1)$ matrices $H_i(\tau) = (H_{i1}, \ldots, H_{iK})'$ and $M_i(\tau) = E[w_t^2\boldsymbol{\epsilon}_{i,t-1}\dot{q}_t'(\boldsymbol{\theta}_{\tau 0})]$, and the $K \times K$ matrices $\Psi_i(\tau) = E(w_t^2\boldsymbol{\epsilon}_{i,t-1}\boldsymbol{\epsilon}_{i,t-1}')$ and

$$\begin{aligned}\Pi_i(\tau) = \sigma_{i,\tau}^{-2}[&\Psi_i(\tau) + H_i(\tau)\Xi(\tau)H_i'(\tau) - M_i(\tau)\Omega_1^{-1}(\tau)H_i'(\tau)\\ &- H_i(\tau)\Omega_1^{-1}(\tau)M_i'(\tau)],\end{aligned}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

where $\Xi(\tau) = \Omega_1^{-1}(\tau)\Omega_0(\tau)\Omega_1^{-1}(\tau)$. In addition, denote the $2K \times (2p+1)$ matrices $H(\tau) = (\sigma_{1,\tau}^{-1}H_1'(\tau), \sigma_{2,\tau}^{-1}H_2'(\tau))'$ and $M(\tau) = E[w_t^2\boldsymbol{\epsilon}_{t-1}\dot{q}_t'(\boldsymbol{\theta}_{\tau 0})]$, where $\boldsymbol{\epsilon}_t = (\sigma_{1,\tau}^{-1}\boldsymbol{\epsilon}_{1,t}', \sigma_{2,\tau}^{-1}\boldsymbol{\epsilon}_{2,t}')'$. Define the $2K \times 2K$ matrices $\Psi(\tau) = E(w_t^2\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}_{t-1}')$ and

$$\begin{aligned}\Pi(\tau) &= \Psi(\tau) + H(\tau)\Xi(\tau)H'(\tau) - M(\tau)\Omega_1^{-1}(\tau)H'(\tau) - H(\tau)\Omega_1^{-1}(\tau)M'(\tau)\\ &:= \begin{pmatrix} \Pi_1(\tau) & \Pi_3'(\tau) \\ \Pi_3(\tau) & \Pi_2(\tau) \end{pmatrix}.\end{aligned}$$

THEOREM 4. *Suppose that the conditions of Theorem 3 hold and* $E(y_t^2) < \infty$. *It then holds that* $\sqrt{n}(\widehat{\boldsymbol{\rho}}', \widehat{\boldsymbol{r}}')' \to_d N(0, \Pi(\tau))$ *as* $n \to \infty$.

The finite second moment of y_t is required by Theorem 4, since this condition is needed to show the consistency of $\widehat{\sigma}_{i,\tau}^2$, for $i = 1$ and 2. We have tried many other approaches, such as transforming the residuals by a bounded and strictly increasing function (Zhu, Zheng, and Li, 2018) and then applying the self-weighted QACF to the transformed residuals. However, the condition of $E(y_t^2) < \infty$ is still unavoidable. As a comparison, if we alternatively use the unweighted QACFs of residuals with $w_t = 1$, for all t, then the condition of $E(|y_t|^3) < \infty$ is unavoidable for asymptotic normality.

As for estimating $\Sigma(\tau)$ in Section 3, we can employ the difference quotient method to estimate the quantity of $f_{t-1}(F_{t-1}^{-1}(\tau))$, and then approximate the expectations in $\Pi(\tau)$ by sample averages with $\eta_{t,\tau}$ being replaced by $\widehat{\eta}_{t,\tau}$. Consequently, a consistent estimator, denoted by $\widehat{\Pi}(\tau)$, for the asymptotic covariance in Theorem 4 can be constructed. We can then check the significance of the $\widehat{\rho}_{k,\tau}$'s and $\widehat{r}_{k,\tau}$'s individually by establishing their confidence intervals based on $\widehat{\Pi}(\tau)$.

Let $\boldsymbol{z} = (\boldsymbol{z}_1', \boldsymbol{z}_2')' \in \mathbb{R}^{2K}$ be a multivariate normal random vector with zero-mean vector and covariance matrix $\Pi(\tau)$, where $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^K$. To check the first K lags of the $\widehat{\rho}_{k,\tau}$'s (or $\widehat{r}_{k,\tau}$'s) jointly, by Theorem 4, the Box–Pierce type test statistics can be designed as follows:

$$Q_1(K) = n\sum_{k=1}^{K}\widehat{\rho}_{k,\tau}^2 \to_d \boldsymbol{z}_1'\boldsymbol{z}_1 \text{ and } Q_2(K) = n\sum_{k=1}^{K}\widehat{r}_{k,\tau}^2 \to_d \boldsymbol{z}_2'\boldsymbol{z}_2,$$

as $n \to \infty$. It is also of interest to consider a combined test, i.e., $Q(K) = Q_1(K) + Q_2(K) \to_d \boldsymbol{z}'\boldsymbol{z}$ as $n \to \infty$. To calculate the critical value or p-value of the three tests, we generate a sequence of, say, $B = 10,000$, multivariate random vectors with the same distribution of $\boldsymbol{z}$, and then use the empirical distributions to approximate the corresponding null distributions. As expected, the simulation results in Section 5 suggest that $Q_1(K)$ is more powerful in detecting the misspecification in the conditional location, $Q_2(K)$ has better performance in detecting the misspecification in the conditional scale, and the performance of $Q(K)$ falls between that of $Q_1(K)$ and $Q_2(K)$. Therefore, we can first use $Q(K)$ to check the overall adequacy of the fitted conditional quantiles, and then employ $Q_1(K)$ and $Q_2(K)$ to look for more details (see also Li and Li, 2008; Zhu, Zheng, and Li, 2018).

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Remark 12. There are two advantages to introducing the self-weights $\{w_t\}$ into the QACF. First, in line with the estimating procedure in Section 3, self-weights can reduce the moment restriction on y_t in establishing the asymptotic normality of $\widehat{\rho}_{k,\tau}$ and $\widehat{r}_{k,\tau}$ (see Theorem 4), thereby making the diagnostic tools applicable to heavy-tailed data. Second, as validated by the simulation results in Section 5, the self-weighted residual QACFs $\widehat{\rho}_{k,\tau}$ and $\widehat{r}_{k,\tau}$ are more efficient than their unweighted counterparts with $w_t = 1$, for all t. In contrast to the efficiency loss in estimation due to self-weights, interestingly, self-weights can improve the efficiency of the QACFs in terms of standard deviations.

5. SIMULATION STUDIES

5.1. Self-Weighted CQE and Model Selection

The first experiment is to evaluate the self-weighted CQE $\widehat{\boldsymbol{\theta}}_{\tau n}$ in Section 3. The data generating process (DGP) is

$$y_t = \phi(u_t)y_{t-1} + S_{\mathrm{Q}}(b(u_t) + \beta(u_t)y_{t-1}^2), \tag{5.1}$$

where $\{u_t\}$ are i.i.d. standard uniform random variables, and we also consider two sets of coefficient functions:

$$\phi(\tau) = -0.2,\ b(\tau) = S_{\mathrm{Q}}^{-1}(F_b^{-1}(\tau)) \text{ and } \beta(\tau) = 0.4b(\tau) \tag{5.2}$$

and

$$\phi(\tau) = 0.5\tau,\ b(\tau) = S_{\mathrm{Q}}^{-1}(F_b^{-1}(\tau)) \text{ and } \beta(\tau) = 0.5\tau b(\tau), \tag{5.3}$$

where $S_{\mathrm{Q}}^{-1}(\tau) = \tau^2 \operatorname{sgn}(\tau)$ is the inverse function of $S_{\mathrm{Q}}(\cdot)$, and $F_b(\cdot)$ is the distribution function of the standard normal, the Student's t_5, or Student's t_3 random variable. It holds that $S_{\mathrm{Q}}(b(u_t) + \beta(u_t)y_{t-1}^2) = S_{\mathrm{Q}}(b(u_t))\sqrt{1 + cy_{t-1}^2}$ when $\beta(\tau) = cb(\tau)$, with $c > 0$ being a constant. Let $\varepsilon_t = S_{\mathrm{Q}}(b(u_t))$, and then the coefficient functions at (5.2) and (5.3) correspond to

$$y_t = -0.2y_{t-1} + \varepsilon_t\sqrt{1 + 0.4y_{t-1}^2} \quad \text{and} \quad y_t = 0.5u_t y_{t-1} + \varepsilon_t\sqrt{1 + 0.5u_t y_{t-1}^2},$$

respectively. We consider two sample sizes, $n = 500$ and 1,000, and there are 1,000 replications for each. The self-weighted CQE in (3.1) is employed to fit the data, and the quasi-Newton algorithm can be used to obtain the estimate for each replication. To estimate the asymptotic standard deviation (ASD) of $\widehat{\boldsymbol{\theta}}_{\tau n}$, two bandwidths, h_B and h_{HS} at (3.2), are used to estimate $f_{t-1}(F_{t-1}^{-1}(\tau))$ in the difference quotient method, and the resulting ASDs are denoted as ASD_1 and ASD_2, respectively.

Tables 1 and 2 present the bias, empirical standard deviation (ESD), and ASD of $\widehat{\boldsymbol{\theta}}_{\tau n}$ at quantile level $\tau = 0.05$ or 0.25 for settings (5.2) and (5.3), respectively, and the corresponding values of $\boldsymbol{\theta}_{\tau 0}$ are also given. We report the results for the standard normal and Student's t_5 cases, only with that for the Student's t_3 case

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

TABLE 1. Biases, ESDs, and ASDs of $\widehat{\boldsymbol{\theta}}_{\tau n}$ at quantile level $\tau = 0.05$ or 0.25 for model (5.1) with coefficient functions (5.2), where ASD_1 and ASD_2 correspond to the bandwidths h_B and h_{HS}, respectively. $F(\cdot)$ is the normal or Student's t_5 distribution function.

		$\tau = 0.05$					$\tau = 0.25$				
n		True	Bias	ESD	ASD_1	ASD_2	True	Bias	ESD	ASD_1	ASD_2
					Normal distribution						
500	b	−2.706	−0.015	0.489	0.743	0.529	−0.455	−0.004	0.134	0.137	0.133
	ϕ	−0.200	−0.003	0.143	0.215	0.155	−0.200	−0.003	0.085	0.095	0.093
	β	−1.082	−0.052	0.525	1.066	0.578	−0.182	−0.011	0.134	0.140	0.136
1,000	b	−2.706	−0.003	0.350	0.379	0.409	−0.455	−0.007	0.094	0.097	0.095
	ϕ	−0.200	−0.002	0.098	0.109	0.116	−0.200	0.000	0.064	0.066	0.065
	β	−1.082	−0.029	0.370	0.398	0.495	−0.182	−0.004	0.096	0.099	0.096
					Student's t_5 distribution						
500	b	−4.060	−0.067	1.030	1.189	1.184	−0.528	−0.008	0.169	0.179	0.172
	ϕ	−0.200	−0.004	0.235	0.279	0.274	−0.200	−0.001	0.097	0.111	0.107
	β	−1.624	−0.163	0.996	1.146	1.211	−0.211	−0.017	0.151	0.162	0.157
1,000	b	−4.060	−0.020	0.726	0.857	0.847	−0.528	−0.009	0.119	0.126	0.121
	ϕ	−0.200	−0.005	0.162	0.193	0.194	−0.200	0.000	0.073	0.077	0.074
	β	−1.624	−0.099	0.683	0.855	0.879	−0.211	−0.008	0.108	0.113	0.109

relegated to the Supplementary Material to save space. It can be seen that, as the sample size increases, the biases, ESDs, and ASDs decrease, and the ESDs move closer to their corresponding ASDs. Moreover, the biases, ESDs, and ASDs become smaller, as the quantile level increases from $\tau = 0.05$ to $\tau = 0.25$. This is expected since there are more observations when the quantile level nears the center, although the true values of $b(\tau)$ and $\beta(\tau)$ also become smaller as τ approaches 0.5. Finally, most of the biases, ESDs, and ASDs increase, as the distribution of $F_b(\cdot)$ becomes heavy-tailed, and bandwidth h_{HS} slightly outperforms h_B.

The second experiment is to evaluate the proposed BIC in Remark 10. The DGP is

$$y_t = \phi_1(u_t)y_{t-1} + \phi_2(u_t)y_{t-2} + S_{\text{Q}}(b(u_t) + \beta_1(u_t)y_{t-1}^2 + \beta_2(u_t)y_{t-2}^2) \tag{5.4}$$

with three sets of coefficient functions:

(i) $\phi_1(\tau) = 0.1$, $\phi_2(\tau) = 0.3$, $b(\tau) = S_{\text{Q}}^{-1}(F_b^{-1}(\tau))$, $\beta_1(\tau) = 0.1b(\tau)$ and $\beta_2(\tau) = 0.4b(\tau)$,

(ii) $\phi_1(\tau) = 0.1\tau$, $\phi_2(\tau) = 0.3$, $b(\tau) = S_{\text{Q}}^{-1}(F_b^{-1}(\tau))$, $\beta_1(\tau) = 0.1\tau b(\tau)$ and $\beta_2(\tau) = 0.4b(\tau)$,

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

TABLE 2. Biases, ESDs, and ASDs of $\widehat{\boldsymbol{\theta}}_{\tau n}$ at quantile level $\tau = 0.05$ or 0.25 for model (5.1) with coefficient functions (5.3), where ASD_1 and ASD_2 correspond to the bandwidths h_B and h_{HS}, respectively. $F(\cdot)$ is the normal or Student's t_5 distribution function.

		$\tau = 0.05$					$\tau = 0.25$				
n		True	Bias	ESD	ASD_1	ASD_2	True	Bias	ESD	ASD_1	ASD_2
						Normal distribution					
500	b	−2.706	0.011	0.425	0.558	0.676	−0.455	−0.001	0.122	0.124	0.123
	ϕ	0.025	−0.006	0.117	0.146	0.166	0.125	−0.004	0.074	0.081	0.080
	β	−0.068	−0.090	0.294	0.520	0.751	−0.057	−0.013	0.085	0.095	0.098
1,000	b	−2.706	0.014	0.291	0.349	0.326	−0.455	−0.004	0.082	0.089	0.085
	ϕ	0.025	−0.005	0.081	0.093	0.089	0.125	−0.003	0.054	0.057	0.055
	β	−0.068	−0.046	0.185	0.275	0.245	−0.057	−0.006	0.060	0.068	0.064
						Student's t_5 distribution					
500	b	−4.060	0.010	0.869	1.316	1.028	−0.528	−0.003	0.153	0.157	0.155
	ϕ	0.025	−0.010	0.194	0.268	0.230	0.125	−0.004	0.082	0.091	0.090
	β	−0.102	−0.238	0.538	1.222	0.748	−0.066	−0.018	0.091	0.099	0.105
1,000	b	−4.060	0.025	0.588	0.759	0.729	−0.528	−0.005	0.102	0.112	0.116
	ϕ	0.025	−0.009	0.131	0.164	0.158	0.125	−0.002	0.060	0.063	0.068
	β	−0.102	−0.128	0.311	0.543	0.530	−0.066	−0.008	0.062	0.072	0.085

(iii) $\phi_1(\tau) = 0.1\tau$, $\phi_2(\tau) = 0.3$, $b(\tau) = S_{\text{Q}}^{-1}(F_b^{-1}(\tau))$, $\beta_1(\tau) = 0.1\tau b(\tau)$ and $\beta_2(\tau) = 0.4\tau b(\tau)$,

where the other settings are preserved as in the first experiment. The proposed BIC is used to select order p with $p_{\max} = 5$. Since the true order is two, the cases of underfitting, correct selection, and overfitting correspond to $\widehat{p}_n$ being 1, 2, and greater than 2, respectively. Table 3 gives the percentages of underfitted, correctly selected, and overfitted models. It can be seen that the BIC performs well in general. Its performance becomes better when the sample size increases, and becomes slightly worse when the distribution of $F_b(\cdot)$ becomes more heavy-tailed.

5.2. Three Portmanteau Tests

The third experiment considers the self-weighted residual QACFs, $\widehat{\rho}_{k,\tau}$ and $\widehat{r}_{k,\tau}$, and the approximation of their asymptotic distributions. The number of lags is set to $K = 6$, and all of the other settings are the same as those in the first experiment. For model (5.1) with coefficient functions in (5.2), Tables 4 and 5 give the biases, ESDs, and ASDs of $\widehat{\rho}_{k,\tau}$ and $\widehat{r}_{k,\tau}$, respectively, with lags $k = 2, 4$, and 6. The results

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

TABLE 3. Percentages of underfitted, correctly selected, and overfitted models by BIC based on 1,000 replications for three sets of coefficient functions. $F(\cdot)$ is the normal, Student's t_5, or Student's t_3 distribution function.

		Set (i)			Set (ii)			Set (iii)		
	n	Under	Exact	Over	Under	Exact	Over	Under	Exact	Over
Normal	500	3.7	95.3	1.0	2.9	95.3	1.8	4.3	93.7	2.0
	1,000	0.0	99.3	0.7	0.0	99.7	0.3	0.0	98.9	1.1
t_5	500	6.9	90.0	3.1	5.3	92.1	2.6	11.7	86.0	2.3
	1,000	0.3	98.1	1.6	0.1	99.1	0.8	0.1	98.5	1.4
t_3	500	11.6	81.7	6.7	9.5	84.6	5.9	17.8	80.4	1.8
	1,000	0.5	96.1	3.4	1.2	96.2	2.6	1.1	97.6	1.3

TABLE 4. Biases ($\times 100$), ESDs ($\times 100$), and ASDs ($\times 100$) of $\widehat{\rho}_{k,\tau}$ with $k = 2, 4$, or 6 for model (5.2). The quantile level is $\tau = 0.05$ or 0.25, and ASD_1 and ASD_2 correspond to the bandwidths h_B and h_{HS}, respectively. $F(\cdot)$ is the normal or Student's t_5 distribution function.

		$\tau = 0.05$				$\tau = 0.25$			
n	Lag	Bias	ESD	ASD_1	ASD_2	Bias	ESD	ASD_1	ASD_2
					Normal distribution				
500	2	0.01	2.64	3.04	2.70	−0.20	2.71	2.65	2.65
	4	0.04	3.03	5.60	3.01	−0.11	3.06	3.00	3.00
	6	−0.01	3.07	3.33	3.06	−0.17	3.10	3.04	3.04
1,000	2	−0.06	1.94	1.89	2.07	−0.13	1.94	1.87	1.87
	4	0.05	2.10	2.12	2.14	−0.04	2.15	2.12	2.12
	6	−0.08	2.13	2.15	2.22	−0.10	2.25	2.15	2.15
					Student's t_5 distribution				
500	2	−0.01	2.15	2.13	2.17	−0.12	2.11	2.04	2.05
	4	0.08	2.63	2.67	2.68	−0.13	2.75	2.64	2.65
	6	0.00	2.84	2.81	2.83	−0.14	2.89	2.80	2.81
1,000	2	−0.01	1.50	1.53	1.50	−0.06	1.44	1.42	1.42
	4	0.05	1.84	1.97	1.91	−0.04	1.90	1.86	1.86
	6	−0.04	1.94	2.07	2.00	−0.12	2.03	1.97	1.97

for the Student's t_3 case are relegated to the Supplementary Material to save space. We have four findings for residual QACFs $\widehat{\rho}_{k,\tau}$ and $\widehat{r}_{k,\tau}$: (1) the biases, ESDs, and ASDs decrease as the sample size increases or as the quantile level increases from $\tau = 0.05$ to 0.25; (2) the ESDs and ASDs are very close to each other; (3)

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

TABLE 5. Biases ($\times 100$), ESDs ($\times 100$), and ASDs ($\times 100$) of $\widehat{r}_{k,\tau}$ with $k = 2$, 4, or 6 for model (5.2). The quantile level is $\tau = 0.05$ or 0.25, and ASD_1 and ASD_2 correspond to the bandwidths h_B and h_{HS}, respectively. $F(\cdot)$ is the normal or Student's t_5 distribution function.

		$\tau = 0.05$				$\tau = 0.25$			
n	Lag	Bias	ESD	ASD_1	ASD_2	Bias	ESD	ASD_1	ASD_2
					Normal distribution				
500	2	0.02	2.67	3.09	2.73	−0.10	2.72	2.69	2.69
	4	0.07	3.03	5.69	3.01	−0.06	3.06	3.00	3.00
	6	0.03	3.09	3.33	3.06	−0.04	3.03	3.03	3.03
1,000	2	−0.06	1.96	1.91	2.09	−0.07	1.99	1.90	1.90
	4	0.05	2.09	2.13	2.14	−0.02	2.17	2.12	2.12
	6	−0.06	2.12	2.15	2.23	−0.04	2.20	2.15	2.15
					Student's t_5 distribution				
500	2	−0.01	2.22	2.17	2.21	−0.06	2.06	2.03	2.04
	4	0.10	2.63	2.68	2.69	−0.04	2.71	2.62	2.62
	6	0.04	2.85	2.81	2.83	0.00	2.78	2.79	2.79
1,000	2	−0.01	1.52	1.56	1.53	−0.03	1.46	1.40	1.40
	4	0.04	1.84	1.97	1.91	0.00	1.93	1.84	1.84
	6	−0.02	1.93	2.00	2.01	−0.07	1.97	1.96	1.96

the biases, ESDs, and ASDs become smaller as $F_b(\cdot)$ becomes heavy-tailed; and (4) the two bandwidths, h_{HS} and h_B, perform similarly. Similar observations can be found for the coefficient functions in (5.3), and the results are relegated to the Supplementary Material to save space.

The fourth experiment is to evaluate the size and power of the three portmanteau tests in Section 4, and the DGP is

$$y_t = c_1 y_{t-2} + S_{\text{Q}}\left(b(u_t) + 0.1 b(u_t) y_{t-1}^2 + c_2 b(u_t) y_{t-2}^2\right), \tag{5.5}$$

with $b(\cdot)$ defined in previous experiments, while a quantile double AR model with order one is fitted to the generated sequences. As a result, the case of $c_1 = c_2 = 0$ corresponds to the size, the case of $c_1 \neq 0$ to the misspecification in the conditional location, and the case of $c_2 \neq 0$ to the misspecification in the conditional scale. Two departure levels, 0.1 and 0.3, are considered for both c_1 and c_2, and we calculate the critical values by generating $B = 10{,}000$ random vectors. Table 6 gives the rejection rates of $Q_1(6)$, $Q_2(6)$, and $Q(6)$, where the critical values of the three tests are calculated based on estimated covariance matrix $\widehat{\Pi}(\tau)$ using bandwidth h_{HS}. It can be seen that the size grows closer to the nominal rate as sample size n increases to 1,000 or quantile level τ increases to 0.25, and almost all of the powers increase as the sample size or departure level increases. Moreover,

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

TABLE 6. Rejection rate (%) of test statistics $Q_1(6)$, $Q_2(6)$, and $Q(6)$. The significance level is 5%, and the quantile level is $\tau = 0.05$ or 0.25. $F(\cdot)$ is the normal, Student's t_5, or Student's t_3 distribution function.

			$\tau = 0.05$			$\tau = 0.25$		
n	c_1	c_2	Q_1	Q_2	Q	Q_1	Q_2	Q
				Normal distribution				
500	0.0	0.0	3.5	3.7	3.6	5.5	4.4	5.8
	0.1	0.0	8.3	7.3	7.4	15.7	9.0	13.3
	0.3	0.0	47.3	41.8	45.6	92.4	53.4	89.5
	0.0	0.1	6.3	6.2	6.6	6.5	6.4	6.5
	0.0	0.3	13.7	14.1	13.7	7.4	11.2	9.5
1,000	0.0	0.0	5.2	5.2	5.3	4.7	6.2	5.7
	0.1	0.0	12.9	11.4	12.2	28.1	16.0	25.1
	0.3	0.0	83.6	78.3	81.4	99.9	90.2	99.8
	0.0	0.1	7.4	7.2	7.4	5.7	6.5	5.7
	0.0	0.3	16.5	19.2	18.1	6.2	16.3	11.8
				Student's t_5 distribution				
500	0.0	0.0	5.6	5.5	5.6	5.7	5.2	5.3
	0.1	0.0	7.8	6.1	6.8	15.7	6.6	11.5
	0.3	0.0	32.4	19.1	26.9	93.4	27.8	86.3
	0.0	0.1	11.8	12.3	12.2	6.8	8.2	7.7
	0.0	0.3	31.1	32.0	32.9	7.6	17.9	16.0
1,000	0.0	0.0	5.6	5.6	5.6	4.6	6.5	5.2
	0.1	0.0	10.3	7.0	7.8	30.0	11.0	23.4
	0.3	0.0	60.7	37.9	53.5	99.8	56.8	99.7
	0.0	0.1	13.9	15.2	14.1	6.4	10.5	9.1
	0.0	0.3	42.7	52.0	49.9	8.6	30.0	23.2
				Student's t_3 distribution				
500	0.0	0.0	6.9	6.3	6.6	5.3	4.6	5.2
	0.1	0.0	8.9	7.4	8.3	14.8	6.1	10.7
	0.3	0.0	28.1	14.1	22.1	91.9	14.4	82.8
	0.0	0.1	16.8	17.2	18.1	7.5	9.0	8.5
	0.0	0.3	41.8	43.0	43.3	7.2	20.6	17.5
1,000	0.0	0.0	5.8	6.3	6.5	4.5	5.4	5.2
	0.1	0.0	10.0	7.2	9.0	29.3	6.3	21.0
	0.3	0.0	50.5	22.3	39.8	99.7	25.0	99.3
	0.0	0.1	24.5	27.4	27.0	7.7	13.2	12.2
	0.0	0.3	60.3	63.8	63.6	8.3	34.7	28.4

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

in general, $Q_1(K)$ is more powerful than $Q_2(K)$ in detecting the misspecification in the conditional location, $Q_2(K)$ is more powerful in detecting the misspecification in the conditional scale, and $Q(K)$ falls between $Q_1(K)$ and $Q_2(K)$. Finally, when the data are more heavy-tailed, $Q_1(K)$ becomes less powerful in detecting the misspecification in the conditional location, and $Q_2(K)$ becomes more powerful in detecting the misspecification in the conditional scale, which may be due to the mixture of two effects: the worse performance of the estimation and the larger value of $|c_2 b(\tau)|$ in the conditional scale. We also calculate the rejection rates of the three tests using bandwidth h_B and another DGP, with similar findings observed (see the Supplementary Material for details).

5.3. Random Weights

The last experiment is to compare the proposed inferential tools with and without random weights in term of efficiency. Note that $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widehat{\rho}_{k,\tau}$, and $\widehat{r}_{k,\tau}$ are the self-weighted CQE and two self-weighted QACFs, respectively; we denote their unweighted counterparts by $\widetilde{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\rho}_{k,\tau}$, and $\widetilde{r}_{k,\tau}$.

The data are generated from the process in (5.1), and two sets of coefficient functions are considered:

(a) $\phi(\tau) = 0.5\tau$, $b(\tau) = S_Q^{-1}(F_b^{-1}(\tau))$, $\beta(\tau) = 0.5\tau b(\tau)$, and

(b) $\phi(\tau) = 0.5\tau$, $b(\tau) = S_Q^{-1}(F_b^{-1}(\tau))$, $\beta(\tau) = 0.8(\tau - 0.5)$,

where $F_b(\cdot)$ is defined in the first experiment, and set (a) is the same as that in (5.3). We generate data of size 2,000 with 1,000 replications, and fit a quantile double AR model with order one to each replication using weighted and unweighted CQE at quantile levels $\tau = 0.05$ and 0.25. Figure 2 presents box plots for the weighted and unweighted CQE estimators, and Figures 3 and 4 give those for the residual QACFs of the weighted and unweighted versions.

On the one hand, the interquartile range of weighted estimator $\widehat{\boldsymbol{\theta}}_{\tau n}$ is larger than that of its unweighted counterpart, $\widetilde{\boldsymbol{\theta}}_{\tau n}$, although the efficiency loss due to the random weights seems smaller as τ increases from 0.05 to 0.25 or the distribution of $F_b(\tau)$ becomes less heavy-tailed. On the other hand, the interquartile range of weighted residual QACF $\widehat{\rho}_{k,\tau}$ (or $\widehat{r}_{k,\tau}$) is smaller than that of its unweighted counterpart $\widetilde{\rho}_{k,\tau}$ (or $\widetilde{r}_{k,\tau}$), and the efficiency gain of the residual QACFs owing to the random weights grows larger as the distribution of $F_b(\tau)$ becomes more heavy-tailed. In other words, the random weights play the opposite role for the estimators and residual QACFs in finite samples. The case with the sample size of 1,000 is similar, and its results can be found in the Supplementary Material.

In sum, both self-weighted CQE and the diagnostic tools can be used to handle heavy-tailed time series by introducing self-weights at the expense of efficiency in estimation, and the weights can even lead to an efficiency gain for the residual QACFs. In terms of the portmanteau tests, the combined test, $Q(K)$, is preferable for checking the overall adequacy of the fitted conditional quantiles, whereas

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

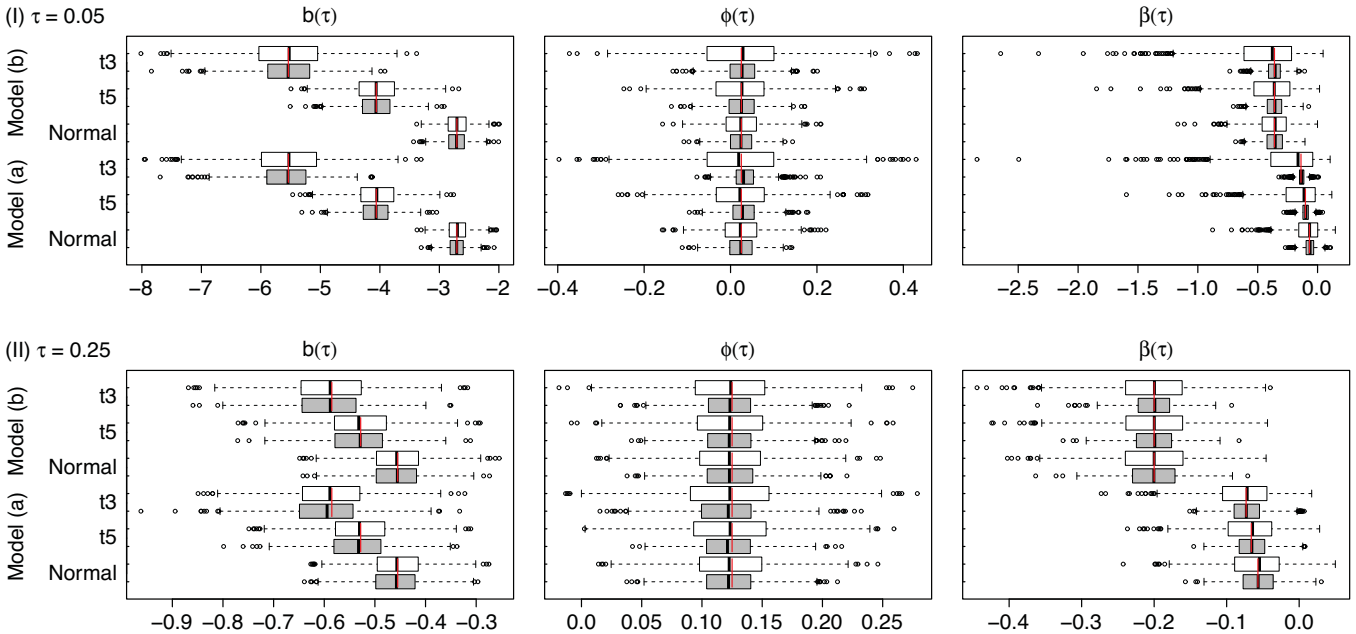


FIGURE 2. Box plots for the self-weighted estimator $\widehat{\boldsymbol{\theta}}_{\tau n}$ (white boxes) and unweighted estimator $\widetilde{\boldsymbol{\theta}}_{\tau n}$ (gray boxes), at $\tau = 0.05$ or 0.25, for the two models with $F_b(\cdot)$ being the normal, Student's t_5, or Student's t_3 distribution function. Model (a): $b(\tau) = S_Q^{-1}(F_b^{-1}(\tau)), \phi(\tau) = 0.5\tau, \beta(\tau) = 0.5\tau b(\tau)$; Model (b): $b(\tau) = S_Q^{-1}(F_b^{-1}(\tau)), \phi(\tau) = 0.5\tau, \beta(\tau) = 0.8(\tau - 0.5)$. The thick black line in the center of the box indicates the sample median, and the thin red line indicates the value of the corresponding element of the true parameter vector $\boldsymbol{\theta}_{\tau 0}$. The notations $b(\tau)$, $\phi(\tau)$, and $\beta(\tau)$ represent the corresponding elements of $\widehat{\boldsymbol{\theta}}_{\tau n}$ and $\widetilde{\boldsymbol{\theta}}_{\tau n}$.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

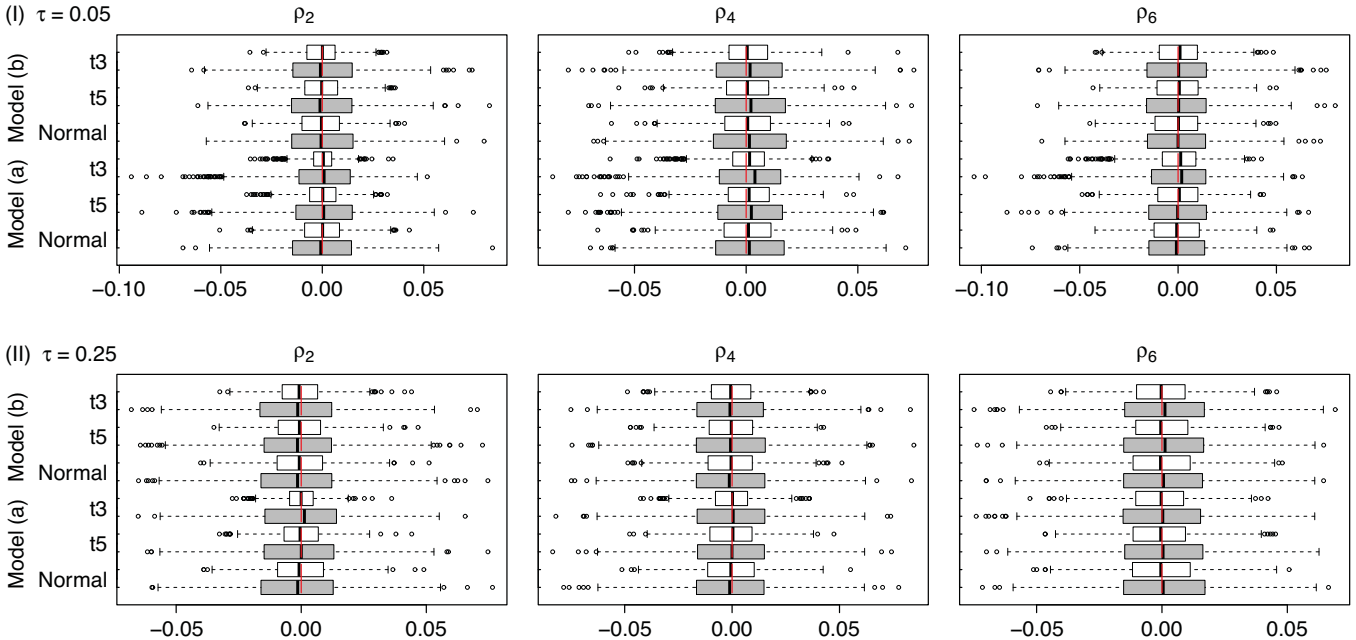


FIGURE 3. Box plots for the self-weighted residual QACFs $\widehat{\rho}_{k,\tau}$ (white boxes) and unweighted residual QACFs $\widetilde{\rho}_{k,\tau}$ (gray boxes), at $\tau = 0.05$ or 0.25, $k = 2$, 4, or 6, for the two models with $F_b(\cdot)$ being the normal, Student's t_5, or Student's t_3 distribution function. Model (a): $b(\tau) = S_Q^{-1}(F_b^{-1}(\tau)), \phi(\tau) = 0.5\tau, \beta(\tau) = 0.5\tau b(\tau)$; Model (b): $b(\tau) = S_Q^{-1}(F_b^{-1}(\tau)), \phi(\tau) = 0.5\tau, \beta(\tau) = 0.8(\tau - 0.5)$. The thick black line in the center of the box indicates the sample median, and the thin red line indicates the true value zero of $\rho_{k,\tau}$ if $Q_\tau(y_t|\mathcal{F}_{t-1})$ is correctly specified. The notations ρ_2, ρ_4, and ρ_6 represent $\widehat{\rho}_{k,\tau}$ and $\widetilde{\rho}_{k,\tau}$ at $k = 2$, 4, and 6.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

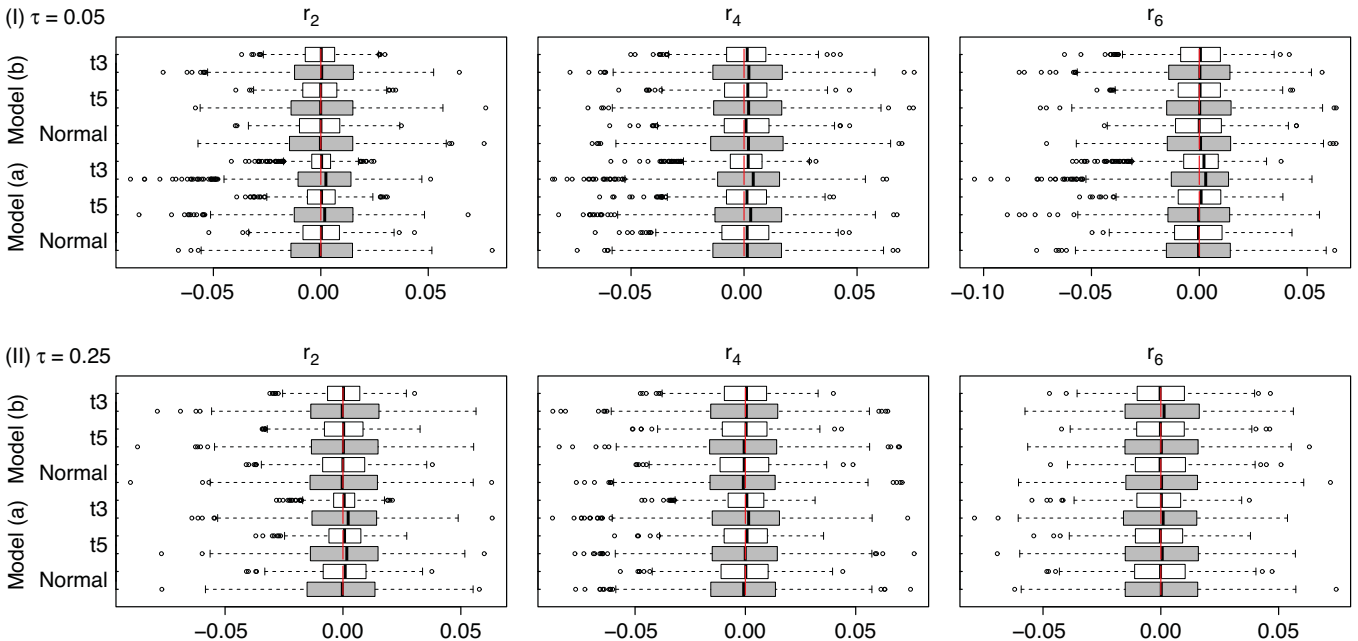


FIGURE 4. Box plots for the self-weighted residual QACFs $\widehat{r}_{k,\tau}$ (white boxes) and unweighted residual QACFs $\widetilde{r}_{k,\tau}$ (gray boxes), at $\tau = 0.05$ or 0.25, $k = 2$, 4, or 6, for the two models with $F_b(\cdot)$ being the normal, Student's t_5, or Student's t_3 distribution function. Model (a): $b(\tau) = S_Q^{-1}(F_b^{-1}(\tau)), \phi(\tau) = 0.5\tau, \beta(\tau) = 0.5\tau b(\tau)$; Model (b): $b(\tau) = S_Q^{-1}(F_b^{-1}(\tau)), \phi(\tau) = 0.5\tau, \beta(\tau) = 0.8(\tau - 0.5)$. The thick black line in the center of the box indicates the sample median, and the thin red line indicates the true value zero of $r_{k,\tau}$ if $Q_\tau(y_t|\mathcal{F}_{t-1})$ is correctly specified. The notations r_2, r_4, and r_6 represent $\widehat{r}_{k,\tau}$ and $\widetilde{r}_{k,\tau}$ at $k = 2$, 4, and 6.

TABLE 7. Summary statistics of log returns

Min	Max	Mean	Median	Std. dev.	Skewness	Kurtosis
−20.189	11.251	0.000	0.097	2.488	−0.746	6.119

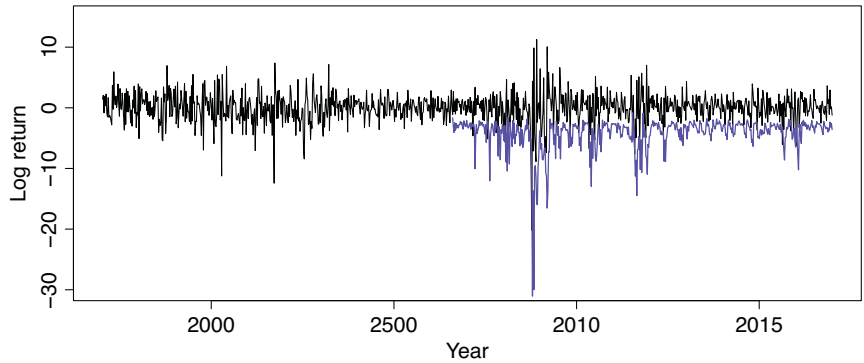


FIGURE 5. Time plot of weekly log returns (black line) of the S&P500 Index from January 10, 1997, to December 30, 2016, and negative 5% VaR forecasts (blue line) from August 11, 2006, to December 30, 2016.

tests $Q_1(K)$ and $Q_2(K)$ can be used to detect the possible misspecification in the conditional location and scale, respectively. With respect to the selection of bandwidth to estimate the quantity of $f_{t-1}(F_{t-1}^{-1}(\tau))$, although h_{HS} and h_B perform very similarly in diagnostic checking, we recommend h_{HS}, since it exhibits more stable performance in estimation. Therefore, h_{HS} is used to estimate the covariance matrices of the self-weighted estimator and residual QACFs in the next section.

6. AN EMPIRICAL EXAMPLE

This section analyzes the weekly closing prices of the S&P500 Index from January 10, 1997, to December 30, 2016. Figure 5 presents the time plot of log returns in percentage form, denoted by $\{y_t\}$. There are 1,043 observations in total. The summary statistics for $\{y_t\}$ are listed in Table 7, from which it can be seen that the data are negatively skewed and heavy-tailed.

Autocorrelation can be seen in the sample ACFs of both $\{y_t\}$ and $\{y_t^2\}$. We first consider a fitted double AR model:

$$y_t = -0.091_{0.040}y_{t-1} + 0.051_{0.036}y_{t-2} - 0.013_{0.037}y_{t-3} + \varepsilon_t\sqrt{2.583_{0.193} + 0.255_{0.031}y_{t-1}^2 + 0.132_{0.039}y_{t-2}^2 + 0.197_{0.035}y_{t-3}^2}, \quad \textbf{(6.1)}$$

where the Gaussian QMLE is employed, standard errors are given in the corresponding subscripts, and the order is selected by the BIC with $p_{\max} = 10$. It can be seen that, at the 5% significance level, all of the fitted coefficients in the conditional

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

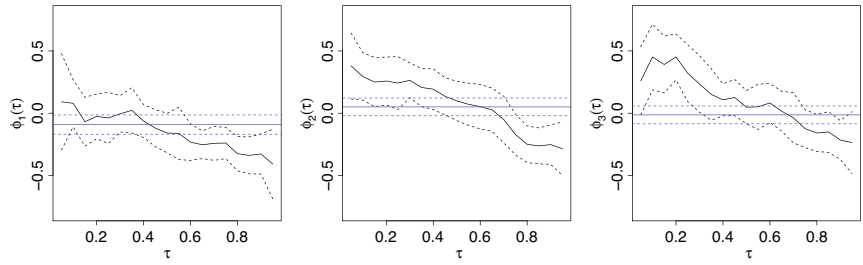


FIGURE 6. The estimates of $\phi_i(\tau_k)$ (black solid) from the fitted quantile double AR model, together with their 95% confidence band (black dotted), at $\tau_k = k/20$ with $1 \le k \le 19$, and estimates of ϕ_i (blue solid) from the fitted double AR model, together with their 95% confidence interval (blue dotted).

mean are insignificant or marginally significant, while those in the conditional variance are significant.

We apply the quantile double AR model to the sequence, with the quantile levels set to $\tau_k = k/20$ with $1 \le k \le 19$. The proposed BIC in Section 3 is employed with $p_{\max} = 10$, and the selected order is $\widehat{p}_n = 3$. The estimates of $\phi_i(\tau)$, for $1 \le i \le 3$, together with their 95% confidence bands, are plotted against the quantile level in Figure 6, and those of ϕ_i in the fitted double AR model (6.1) are also given for the sake of comparison. The confidence bands of $\phi_i(\tau)$ and ϕ_i at lags 1 and 2 do not overlap at the quantile levels around $\tau = 0.8$, and those at lag 3 are significantly separated from each other around $\tau = 0.2$. We can conclude the τ-dependence of the $\phi_i(\tau)$'s, and the commonly used double AR model is limited in interpreting this type of financial time series.

We next attempt to compare the fitted coefficients in the conditional scale of our model with those of model (6.1). Note that, from Remark 1, the quantity of $\beta_j(\tau)/b(\tau)$, for each $1 \le j \le p$, in the quantile double AR model corresponds to β_j/ω in the double AR model. Moreover, when the quantile level τ is close to 0.5, the estimate of $b(\tau)$ is very small, which makes the value of $\widehat{\beta}_j(\tau)/\widehat{b}(\tau)$ abnormally large (see also Remark 6). As a result, Figure 7 plots the fitted values of $\beta_j(\tau)/b(\tau)$, for $1 \le j \le 3$, together with their 95% confidence bands, against the quantile level with $|\tau - 0.5| \ge 0.15$. The fitted values of β_j/ω from model (6.1) are also reported. The confidence bands of $\beta_1(\tau)/b(\tau)$ and β_1/ω are separated from each other for the quantile levels around $\tau = 0.25$. Note that, for the double AR model (1.2), the coefficients in the conditional scale include $S_Q(b_\tau \beta_j)$ with $1 \le j \le p$, which already depends on the quantile level. It could thus be argued that the model is still insufficiently flexible to interpret this series.

Since 5% VaR is usually of interest to the practitioner, we now provide more details about the fitted conditional quantile at $\tau = 0.05$:

$$\begin{aligned}\widehat{Q}_{0.05}(y_t|\mathcal{F}_{t-1}) &= 0.091_{0.198} y_{t-1} + 0.379_{0.135} y_{t-2} + 0.260_{0.139} y_{t-3} \\ &\quad + S_{\mathrm{Q}}\left(-6.951_{1.773} - 0.261_{0.698} y_{t-1}^2 - 0.367_{0.413} y_{t-2}^2 - 1.346_{0.501} y_{t-3}^2\right),\end{aligned}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

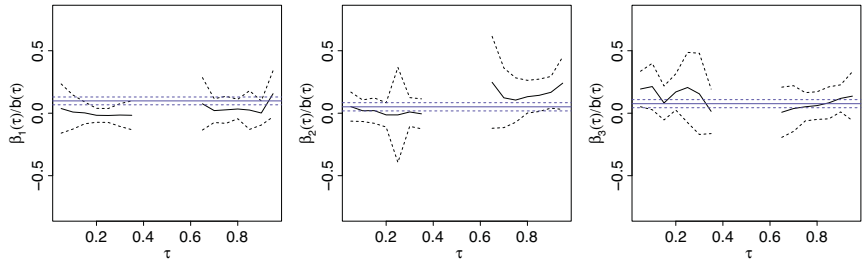


FIGURE 7. The estimates of $\beta_j(\tau_k)/b(\tau_k)$ (black solid) from the fitted quantile double AR model, together with their 95% confidence band (black dotted), at $\tau_k = k/20$ with $1 \le k \le 7$ and $13 \le k \le 19$, and estimates of β_j/ω (blue solid) from the fitted double AR model, together with their 95% confidence interval (blue dotted).

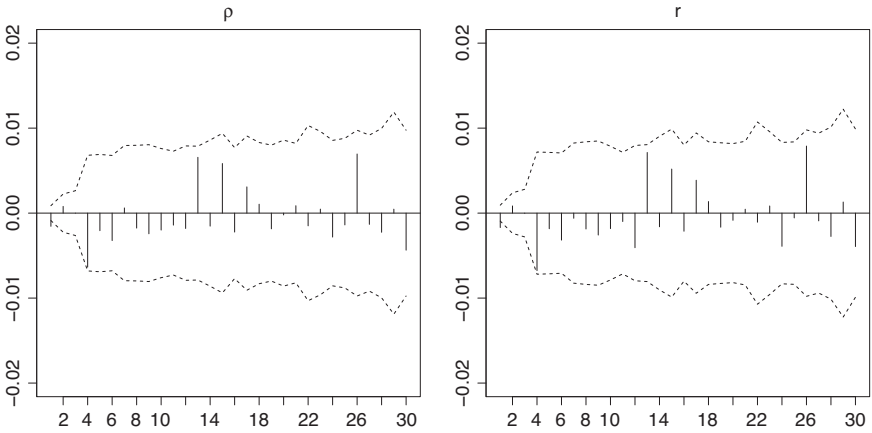


FIGURE 8. Residual QACFs of the fitted quantile double AR model at $\tau = 0.05$, with 95% confidence bands.

where the standard errors are given in the corresponding subscripts of the estimated coefficients. Figure 8 plots the residual QACFs $\widehat{\rho}_{k,\tau}$ and $\widehat{r}_{k,\tau}$, which slightly stand out from the 95% confidence bands only at lags 1 and 4. The p-values of $Q_1(K)$, $Q_2(K)$, and $Q(K)$ are all larger than 0.717 for $K = 10$, 20, and 30. We can thus conclude that both residual QACFs are insignificant both individually and jointly, and that the fitted conditional quantile is adequate.

We now consider one-step-ahead conditional quantile prediction at level $\tau = 0.05$, which is the negative value of the 5% VaR forecast, with a rolling forecast procedure performed. We begin with the forecast origin $n = 501$, which corresponds to the date of August 11, 2006, and obtain the estimated coefficients of the quantile double AR model with order three using the data from the beginning to the forecast origin (exclusive). For each fitted model, we calculate the one-step-ahead conditional quantile prediction for the next trading week by $\widehat{Q}_{0.05}(y_n|\mathcal{F}_{n-1}) =$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

$\sum_{i=1}^{3}\widehat{\phi}_i(0.05)y_{n-i}+S_{\mathrm{Q}}\left(\widehat{b}(0.05)+\sum_{j=1}^{3}\widehat{\beta}_j(0.05)y_{n-j}^2\right)$. We then advance the forecast origin by one and repeat the previous estimation and prediction until all data are utilized. These predicted values are displayed against the time plot in Figure 5. The magnitudes of the VaRs become larger as the return becomes more volatile, and the returns fall below their calculated negative 5% VaRs occasionally.

Furthermore, we compare the forecast performance of the proposed quantile double autoregressive (QDAR) model with two commonly used models in the literature: the quantile autoregressive (QAR) model (Koenker and Xiao, 2006) and the two-regime threshold quantile autoregressive (TQAR) model (Galvao Jr., Montes-Rojas, and Olmo, 2011). The order is fixed at three for all three models. In estimating TQAR models, the delay parameter d is searched among $\{1,2,3\}$, and the threshold parameter r is searched among a compact grid with the empirical percentiles of y_t from the 10th to 90th quantiles. The rolling forecast procedure is employed again, and we consider the VaRs at four levels: $\tau = 5\%$, 10%, 90%, and 95%. To evaluate the forecasting performance, the empirical coverage rate (ECR) is calculated as the percentage of last 543 observations that fall below the corresponding fitted conditional quantiles. In addition to ECRs, we also conduct two VaR backtests: the likelihood ratio test for correct conditional coverage (CC) in Christoffersen (1998) and the dynamic quantile (DQ) test in Engle and Manganelli (2004). More specifically, the null hypothesis of CC tests is that, conditional on $\mathcal{F}_{t-1}$, $\{H_t\}$ are i.i.d. Bernoulli random variables with the success probability being τ, where the hit is $H_t = I(y_t < Q_\tau(y_t|\mathcal{F}_{t-1}))$. For DQ tests, following Engle and Manganelli (2004), we regress H_t on constant, four lagged hits $H_{t-\ell}$ with $1 \le \ell \le 4$, and the contemporaneous VaR forecast. The null hypothesis of DQ tests is that the intercept is equal to the quantile level τ and the four regression coefficients are zero. If the null hypothesis of each VaR backtest cannot be rejected, then the VaR forecasts are satisfactory.

Table 8 gives the ECRs and p-values of two VaR backtests for the one-step-ahead forecasts. It can be seen that ECRs of three models are all close to the corresponding nominal levels. In terms of two backtests, the QDAR model performs satisfactorily with all p-values greater than 0.1, whereas the QAR model

Table 8. ECRs (%) and p-values of the two VaR backtests of the three models at the 5%, 10%, 90%, and 95% conditional quantiles. M1, M2, and M3 represent the QDAR, QAR, and TQAR models, respectively.

	$\tau = 5\%$			$\tau = 10\%$			$\tau = 90\%$			$\tau = 95\%$		
	ECR	CC	DQ	ECR	CC	DQ	ECR	CC	DQ	ECR	CC	DQ
M1	5.34	0.88	0.33	9.02	0.34	0.22	91.53	0.25	0.11	95.95	0.23	0.51
M2	5.16	0.17	0.00	9.58	0.03	0.00	92.45	0.08	0.00	95.95	0.33	0.00
M3	6.45	0.20	0.01	10.13	0.81	0.02	91.34	0.19	0.12	95.03	0.42	0.02

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

performs poorly at all four levels, and the TQAR model performs well only at the level of 90%. This may be due to the fact that both QAR and TQAR models ignore the conditional heteroskedastic structure of stock prices. In addition, we also conduct the quantile rearrangement (Chernozhukov, Fernández-Val, and Galichon, 2010) to ensure monotonicity, and almost the same results can be observed for backtests and empirical coverage (see the Supplementary Material for details).

In sum, it is necessary to consider using the proposed QDAR model to interpret these stock indices. The proposed inferential tools can also provide reliable results.

7. CONCLUSION AND DISCUSSION

This paper proposes a new conditional heteroskedastic model, which has varying structures at different quantile levels, and its necessity is illustrated by analyzing the weekly S&P500 Index. A novel transformation, $S_Q(\cdot)$, is introduced to the conditional scale, which renders the AR coefficients free from nonnegative constraints. The strict stationarity of the new model is derived, and inferential tools, including a self-weighted CQE and self-weighted residual QACF-based portmanteau tests, are also constructed.

Our model can be extended in two directions. First, as far as we know, it is the first quantile conditional heteroskedastic model in the literature, and it would thus certainly be of interest to consider other types of quantile conditional heteroskedastic time series models (Francq and Zakoian, 2010). Second, based on the transformation of $S_Q(\cdot)$, the proposed quantile double AR model (2.3) has a seemingly linear structure. As a result, we may also consider a multivariate quantile double AR model, say with order one,

$$y_{lt} = \sum_{m=1}^{N} \phi_{lm}(u_{lt}) y_{m,t-1} + S_{\mathrm{Q}}\left(b_l(u_{lt}) + \sum_{m=1}^{N} \beta_{lm}(u_{lt}) y_{m,t-1}^2 \right), \quad 1 \le l \le N,$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$ is an N-dimensional time series, $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{Nt})'$, $\boldsymbol{\phi}(\boldsymbol{u}_t) = (\phi_{lm}(u_{lt}))$, and $\boldsymbol{\beta}(\boldsymbol{u}_t) = (\beta_{lm}(u_{lt}))$ are $N \times N$ coefficient matrices, $\boldsymbol{b}(\boldsymbol{u}_t) = (b_1(u_{1t}), \ldots, b_N(u_{Nt}))'$, and the marginal distributions of $\boldsymbol{u}_t$ are all standard uniform (see Tsay, 2014). It may even be possible to handle high-dimensional time series if the coefficient matrices $\boldsymbol{\phi}(\boldsymbol{u}_t)$ and $\boldsymbol{\beta}(\boldsymbol{u}_t)$ are sparsed or have a low-rank structure (Zhu et al., 2017). We leave it for future research.

APPENDIX

Technical Proofs

This appendix gives the technical proofs of Theorems 1–4 and Corollary 2. Moreover, Lemmas 1 and 2 with proofs are also included, and they provide some preliminary results for proving Theorem 3. Throughout the appendix, the notation C is a generic constant which may take different values in different locations.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Proof of Theorem 1. Let $\boldsymbol{y}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$, $\mathcal{B}^p$ be the class of Borel sets of $\mathbb{R}^p$, and ν_p be the Lebesgue measure on $(\mathbb{R}^p, \mathcal{B}^p)$. Denote by $m : \mathbb{R}^p \to \mathbb{R}$ the projection map onto the first coordinate, i.e., $m(\boldsymbol{x}) = x_1$, for $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$. Define the function $G(u; \boldsymbol{v}_1, \boldsymbol{v}_2) = \boldsymbol{v}_1' \boldsymbol{\phi}(u) + S_{\mathrm{Q}}\left(b(u) + \boldsymbol{v}_2' \boldsymbol{\beta}(u)\right)$, for $u \in (0,1)$, and vectors $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^p$, where $b(\cdot)$, $\boldsymbol{\phi}(\cdot)$, and $\boldsymbol{\beta}(\cdot)$ are defined as in Section 3. As a result, $\{\boldsymbol{y}_t\}$ is a homogeneous Markov chain on the state space $(\mathbb{R}^p, \mathcal{B}^p, \nu_p)$, with transition probability

$$P(\boldsymbol{x}, A) = \int_{m(A)} 2|b(\widetilde{u}) + \boldsymbol{x}_{\mathrm{Q}}' \boldsymbol{\beta}(\widetilde{u})|^{1/2} f\left(S_{\mathrm{Q}}^{-1}\left(z - \boldsymbol{x}' \boldsymbol{\phi}(\widetilde{u})\right) - \boldsymbol{x}_{\mathrm{Q}}' \boldsymbol{\beta}(\widetilde{u})\right) dz,$$

for $\boldsymbol{x} \in \mathbb{R}^p$ and $A \in \mathcal{B}^p$, where $\boldsymbol{x}_{\mathrm{Q}} = (x_1^2, x_2^2, \ldots, x_p^2)'$, $\widetilde{u} = G^{-1}(z; \boldsymbol{x}, \boldsymbol{x}_{\mathrm{Q}})$ with G^{-1} being the inverse function of $G(u; \boldsymbol{x}, \boldsymbol{x}_{\mathrm{Q}})$, f is the density function of $b(u_t)$, and $S_{\mathrm{Q}}^{-1}(x) = x^2 \operatorname{sgn}(x)$ is the inverse function of $S_{\mathrm{Q}}(x)$.

Let $\boldsymbol{X}_{1,i} = (z_i, \ldots, z_1, x_1, \ldots, x_{p-i})'$ and $\boldsymbol{X}_{2,i} = (z_i^2, \ldots, z_1^2, x_1^2, \ldots, x_{p-i}^2)'$. We can further show that the p-step transition probability of the Markov chain $\{\boldsymbol{y}_t\}$ is

$$\begin{aligned} P^p(\boldsymbol{x}, A) = \int_A \prod_{i=1}^{p} & 2|b(\widetilde{u}_i) + \boldsymbol{X}_{2,i-1}' \boldsymbol{\beta}(\widetilde{u}_i)|^{1/2} \\ & \times f\left(S_{\mathrm{Q}}^{-1}\left(z_i - \boldsymbol{X}_{1,i-1}' \boldsymbol{\phi}(\widetilde{u}_i)\right) - \boldsymbol{X}_{2,i-1}' \boldsymbol{\beta}(\widetilde{u}_i)\right) dz_1 \ldots dz_p, \end{aligned} \tag{A.1}$$

where $\widetilde{u}_i = G^{-1}(z_i; \boldsymbol{X}_{1,i-1}, \boldsymbol{X}_{2,i-1})$. Observe that, by Assumption 1, $P^p(\boldsymbol{x}, A) > 0$, for any $\boldsymbol{x} \in \mathbb{R}^p$ and $A \in \mathcal{B}^p$ with $\nu_p(A) > 0$, i.e., $\{\boldsymbol{y}_t\}$ is ν_p-irreducible.

To show that $\{\boldsymbol{y}_t\}$ is geometrically ergodic, next, we verify Tweedie's drift criterion (Tweedie, 1983, Thm. 4). Note that $(a+b)^\kappa \le a^\kappa + b^\kappa$, for $a, b > 0$ and $0 < \kappa \le 1$. Moreover, since $-|c| - |d| \le c + d \le |c| + |d|$ for any constants c and d holds. Then, for any $u \in (0,1)$, it follows that

$$-\sqrt{|b(u)|} - \sum_{i=1}^{p} \sqrt{|\beta_i(u)|}|x_i| \le \left| b(u) + \sum_{i=1}^{p} \beta_i(u) x_i^2 \right|^{1/2} \le \sqrt{|b(u)|} + \sum_{i=1}^{p} \sqrt{|\beta_i(u)|}|x_i|.$$

As a result, by Assumption 1, it can be verified that, for some $0 < \kappa \le 1$,

$$\begin{aligned} E(|y_{t+1}|^\kappa | \boldsymbol{y}_t = \boldsymbol{x}) &= E\left| \sum_{i=1}^{p} \phi_i(u_{t+1}) x_i + S_{\mathrm{Q}}\left(b(u_{t+1}) + \sum_{i=1}^{p} \beta_i(u_{t+1}) x_i^2 \right) \right|^\kappa \\ &\le \sum_{i=1}^{p} \left[E|\phi_i(u_{t+1}) \operatorname{sgn}(x_i) + \sqrt{|\beta_i(u_{t+1})|}|^\kappa \right] |x_i|^\kappa + E|b(u_{t+1})|^{\kappa/2} \\ &\le \sum_{i=1}^{p} a_i |x_i|^\kappa + E|b(u_{t+1})|^{\kappa/2}, \end{aligned}$$

where $a_i = \max\{E|\phi_i(u_{t+1}) - \sqrt{|\beta_i(u_{t+1})|}|^\kappa, E|\phi_i(u_{t+1}) + \sqrt{|\beta_i(u_{t+1})|}|^\kappa\}$, for $1 \le i \le p$. Note that $\sum_{i=1}^{p} a_i < 1$, and then we can find positive values $\{r_1, \ldots, r_{p-1}\}$ such that

$$a_p < r_{p-1} < 1 - \sum_{i=1}^{p-1} a_i \quad \text{and} \quad a_{i+1} + r_{i+1} < r_i < 1 - \sum_{k=1}^{i} a_k \quad \text{for } 1 \le i \le p-2. \tag{A.2}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Consider the test function $g(\boldsymbol{x}) = 1 + |x_1|^\kappa + \sum_{i=1}^{p-1} r_i |x_{i+1}|^\kappa$, and we have that

$$\begin{aligned} &E[g(\boldsymbol{y}_{t+1})|\boldsymbol{y}_t = \boldsymbol{x}] \\ &\quad \le 1 + \sum_{i=1}^{p} a_i |x_i|^\kappa + \sum_{i=1}^{p-1} r_i |x_i|^\kappa + E|b(u_{t+1})|^{\kappa/2} \\ &\quad = 1 + (a_1 + r_1)|x_1|^\kappa + \sum_{i=2}^{p-1} \frac{a_i + r_i}{r_{i-1}} r_{i-1} |x_i|^\kappa + \frac{a_p}{r_{p-1}} r_{p-1} |x_p|^\kappa + E|b(u_{t+1})|^{\kappa/2} \\ &\quad \le \rho g(\boldsymbol{x}) + 1 - \rho + E|b(u_{t+1})|^{\kappa/2}, \end{aligned}$$

where, from (A.2),

$$\rho = \max\left\{ a_1 + r_1, \frac{a_2 + r_2}{r_1}, \ldots, \frac{a_{p-1} + r_{p-1}}{r_{p-2}}, \frac{a_p}{r_{p-1}} \right\} < 1.$$

Denote $\epsilon = 1 - \rho - (1 - \rho + E|b(u_{t+1})|^{\kappa/2})/g(\boldsymbol{x})$, and $K = \{\boldsymbol{x} : \|\boldsymbol{x}\| \le L\}$, where L is a positive constant such that $g(\boldsymbol{x}) > 1 + E|b(u_{t+1})|^{\kappa/2}/(1 - \rho)$ as $\|\boldsymbol{x}\| > L$. We can verify that

$$E[g(\boldsymbol{y}_{t+1})|\boldsymbol{y}_t = \boldsymbol{x}] \le (1 - \epsilon) g(\boldsymbol{x}), \quad \boldsymbol{x} \notin K,$$

and

$$E[g(\boldsymbol{y}_{t+1})|\boldsymbol{y}_t = \boldsymbol{x}] \le C < \infty, \quad \boldsymbol{x} \in K,$$

i.e., Tweedie's drift criterion (Tweedie, 1983, Thm. 4) holds. Moreover, $\{\boldsymbol{y}_t\}$ is a Feller chain, since, for each bounded continuous function $g^*(\cdot)$, $E[g^*(\boldsymbol{y}_t)|\boldsymbol{y}_{t-1} = \boldsymbol{x}]$ is continuous with respect to $\boldsymbol{x}$, and then K is a small set. As a result, from Theorem 4(ii) in Tweedie (1983) and Theorems 1 and 2 in Feigin and Tweedie (1985), $\{\boldsymbol{y}_t\}$ is geometrically ergodic with a unique stationary distribution $\pi(\cdot)$, and

$$\int_{\mathbb{R}^p} g(\boldsymbol{x}) \pi(d\boldsymbol{x}) = 1 + \left(1 + \sum_{i=1}^{p-1} r_i\right) E|y_t|^\kappa < \infty,$$

which implies that $E|y_t|^\kappa < \infty$. This accomplishes the proof. ■

Proof of Theorem 2. Recall that $q_t(\boldsymbol{\theta}_\tau) = \boldsymbol{y}'_{1,t-1}\boldsymbol{\phi}(\tau) + S_{\mathrm{Q}}\left(b(\tau) + \boldsymbol{y}'_{2,t-1}\boldsymbol{\beta}(\tau)\right)$, where $\boldsymbol{\theta}_\tau = (\boldsymbol{\phi}'(\tau), b(\tau), \boldsymbol{\beta}'(\tau))'$. Define $L_n(\boldsymbol{\theta}_\tau) = n^{-1} \sum_{t=p+1}^{n} \omega_t \ell_t(\boldsymbol{\theta}_\tau)$, where $\ell_t(\boldsymbol{\theta}_\tau) = \rho_\tau(y_t - q_t(\boldsymbol{\theta}_\tau))$. To show the consistency, it suffices to verify the following claims:

(i) $E[\sup_{\Theta_\tau} \omega_t \ell_t(\boldsymbol{\theta}_\tau)] < \infty$;
(ii) $E[\omega_t \ell_t(\boldsymbol{\theta}_\tau)]$ has a unique minimum at $\boldsymbol{\theta}_{\tau 0}$;
(iii) For any $\boldsymbol{\theta}^\dagger_\tau \in \Theta_\tau$, $E[\sup_{\boldsymbol{\theta}_\tau \in B_\eta(\boldsymbol{\theta}^\dagger_\tau)} \omega_t |\ell_t(\boldsymbol{\theta}_\tau) - \ell_t(\boldsymbol{\theta}^\dagger_\tau)|] \to 0$ as $\eta \to 0$, where $B_\eta(\boldsymbol{\theta}^\dagger_\tau) = \{\boldsymbol{\theta}_\tau \in \Theta_\tau : \|\boldsymbol{\theta}^\dagger_\tau - \boldsymbol{\theta}_\tau\| < \eta\}$ is an open neighborhood of $\boldsymbol{\theta}^\dagger_\tau$ with radius $\eta > 0$.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

We first prove Claim (i). By Assumptions 2 and 3, the boundedness of $b(\tau)$, $\boldsymbol{\phi}(\tau)$, and $\boldsymbol{\beta}(\tau)$, and the fact that $|\rho_\tau(x)| \le |x|$, it holds that

$$E[\sup_{\Theta_\tau} \omega_t \ell_t(\boldsymbol{\theta}_\tau)] \le E[\sup_{\Theta_\tau} |\omega_t y_t|] + E[\sup_{\Theta_\tau} |\omega_t q_t(\boldsymbol{\theta}_\tau)|] < \infty.$$

Hence, (i) is verified.

We next prove (ii). For $x \neq 0$, it holds that

$$\begin{aligned}\rho_\tau(x-y) - \rho_\tau(x) &= -y\psi_\tau(x) + y\int_0^1 [I(x \le ys) - I(x \le 0)]ds \\ &= -y\psi_\tau(x) + (x-y)[I(0 > x > y) - I(0 < x < y)], \end{aligned} \tag{A.3}$$

where $\psi_\tau(x) = \tau - I(x < 0)$ (see Knight, 1998). Let $v_t(\boldsymbol{\theta}_\tau) = q_t(\boldsymbol{\theta}_\tau) - q_t(\boldsymbol{\theta}_{\tau 0})$ and $\eta_{t,\tau} = y_t - q_t(\boldsymbol{\theta}_{\tau 0})$. By (A.3), it follows that

$$\begin{aligned}&\ell_t(\boldsymbol{\theta}_\tau) - \ell_t(\boldsymbol{\theta}_{\tau 0}) \\ &= -v_t(\boldsymbol{\theta}_\tau)\psi_\tau(\eta_{t,\tau}) + [\eta_{t,\tau} - v_t(\boldsymbol{\theta}_\tau)]\big[I(0 > \eta_{t,\tau} > v_t(\boldsymbol{\theta}_\tau)) - I(0 < \eta_{t,\tau} < v_t(\boldsymbol{\theta}_\tau))\big].\end{aligned}$$

This, together with $\omega_t \ge 0$ by Assumption 3 and $E[\psi_\tau(\eta_{t,\tau})] = 0$, implies that

$$\begin{aligned}&E[\omega_t \ell_t(\boldsymbol{\theta}_\tau)] - E[\omega_t \ell_t(\boldsymbol{\theta}_{\tau 0})] \\ &\quad = E\left\{\omega_t[\eta_{t,\tau} - v_t(\boldsymbol{\theta}_\tau)]\big[I(0 > \eta_{t,\tau} > v_t(\boldsymbol{\theta}_\tau)) - I(0 < \eta_{t,\tau} < v_t(\boldsymbol{\theta}_\tau))\big]\right\} \ge 0. \end{aligned} \tag{A.4}$$

By Assumption 4, $f_{t-1}(x)$ is continuous at a neighborhood of $q_t(\boldsymbol{\theta}_{\tau 0})$, then the above equality holds if and only if $v_t(\boldsymbol{\theta}_\tau) = 0$ a.s., for some $t \in \mathbb{Z}$. Then, we have

$$\boldsymbol{y}'_{1,t-1}[\boldsymbol{\phi}_0(\tau) - \boldsymbol{\phi}(\tau)] = S_{\mathrm{Q}}\left(b(\tau) + \boldsymbol{y}'_{2,t-1}\boldsymbol{\beta}(\tau)\right) - S_{\mathrm{Q}}\left(b_0(\tau) + \boldsymbol{y}'_{2,t-1}\boldsymbol{\beta}_0(\tau)\right).$$

Note that $b_0(\tau) \neq 0$ and, given $\mathcal{F}_{t-2}$, y_{t-1} is independent of all the others. As a result, it holds that $\phi_1(\tau) = \phi_{10}(\tau)$ and $\beta_1(\tau) = \beta_{10}(\tau)$. Sequentially, we can show that $\phi_i(\tau) = \phi_{i0}(\tau)$ and $\beta_i(\tau) = \beta_{i0}(\tau)$, for $i \ge 2$, and hence $b(\tau) = b_0(\tau)$. Therefore, $\boldsymbol{\theta}_\tau = \boldsymbol{\theta}_{\tau 0}$ and (ii) is verified.

Finally, we show (iii). By the Taylor expansion, it holds that

$$|q_t(\boldsymbol{\theta}_\tau) - q_t(\boldsymbol{\theta}^\dagger_\tau)| \le \|\boldsymbol{\theta}_\tau - \boldsymbol{\theta}^\dagger_\tau\| \left\|\dot{q}_t(\boldsymbol{\theta}^*_\tau)\right\|,$$

where $\boldsymbol{\theta}^*_\tau$ is between $\boldsymbol{\theta}_\tau$ and $\boldsymbol{\theta}^\dagger_\tau$. Thus, together with the Lipschitz continuity of $\rho_\tau(x)$, by Assumption 3, we have

$$E[\sup_{\boldsymbol{\theta}_\tau \in B_\eta(\boldsymbol{\theta}^\dagger_\tau)} \omega_t |\ell_t(\boldsymbol{\theta}_\tau) - \ell_t(\boldsymbol{\theta}^\dagger_\tau)|] \le C\eta E[\omega_t \dot{q}_t(\boldsymbol{\theta}^*_\tau)]$$

tends to 0 as $\eta \to 0$. Hence, Claim (iii) holds.

Based on Claims (i) and (iii), by a method similar to that in Huber (1973), we next verify the consistency. Let V be any open neighborhood of $\boldsymbol{\theta}_{\tau 0} \in \Theta_\tau$. By Claim (iii), for any $\boldsymbol{\theta}^\dagger_\tau \in V^c = \Theta_\tau / V$ and $\epsilon > 0$, there exists an $\eta_0 > 0$ such that

$$E[\inf_{\boldsymbol{\theta}_\tau \in B_{\eta_0}(\boldsymbol{\theta}^\dagger_\tau)} \omega_t \ell_t(\boldsymbol{\theta}_\tau)] \ge E[\omega_t \ell_t(\boldsymbol{\theta}^\dagger_\tau)] - \epsilon. \tag{A.5}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

From Claim (i), by the ergodic theorem, it follows that

$$\frac{1}{n}\sum_{t=p+1}^{n}\inf_{\boldsymbol{\theta}_\tau\in B_{\eta_0}(\boldsymbol{\theta}_\tau^\dagger)}\omega_t\ell_t(\boldsymbol{\theta}_\tau)\geq E[\inf_{\boldsymbol{\theta}_\tau\in B_{\eta_0}(\boldsymbol{\theta}_\tau^\dagger)}\omega_t\ell_t(\boldsymbol{\theta}_\tau)]-\epsilon, \tag{A.6}$$

as n is large enough. Since V^c is compact, we can choose $\{B_{\eta_0}(\boldsymbol{\theta}_{\tau i}):\boldsymbol{\theta}_{\tau i}\in V^c, i=1,\ldots,k\}$ to be a finite covering of V^c. Then, by (A.5) and (A.6), we have

$$\begin{aligned}\inf_{\boldsymbol{\theta}_\tau\in V^c}L_n(\boldsymbol{\theta}_\tau)&=\min_{1\leq i\leq k}\inf_{\boldsymbol{\theta}_\tau\in B_{\eta_0}(\boldsymbol{\theta}_{\tau i})}L_n(\boldsymbol{\theta}_\tau)\\&\geq\min_{1\leq i\leq k}\frac{1}{n}\sum_{t=p+1}^{n}\inf_{\boldsymbol{\theta}_\tau\in B_{\eta_0}(\boldsymbol{\theta}_{\tau i})}\omega_t\ell_t(\boldsymbol{\theta}_\tau)\\&\geq\min_{1\leq i\leq k}E[\inf_{\boldsymbol{\theta}_\tau\in B_{\eta_0}(\boldsymbol{\theta}_{\tau i})}\omega_t\ell_t(\boldsymbol{\theta}_\tau)]-\epsilon,\end{aligned} \tag{A.7}$$

as n is large enough. Moreover, for each $\boldsymbol{\theta}_{\tau i}\in V^c$, by Claim (ii), there exists an $\epsilon_0>0$ such that

$$E[\inf_{\boldsymbol{\theta}_\tau\in B_{\eta_0}(\boldsymbol{\theta}_{\tau i})}\omega_t\ell_t(\boldsymbol{\theta}_\tau)]\geq E[\omega_t\ell_t(\boldsymbol{\theta}_{\tau 0})]+3\epsilon_0. \tag{A.8}$$

Therefore, by (A.7) and (A.8), taking $\epsilon=\epsilon_0$, it holds that

$$\inf_{\boldsymbol{\theta}_\tau\in V^c}L_n(\boldsymbol{\theta}_\tau)\geq E[\omega_t\ell_t(\boldsymbol{\theta}_{\tau 0})]+2\epsilon_0. \tag{A.9}$$

Furthermore, by the ergodic theorem, it follows that

$$\inf_{\boldsymbol{\theta}_\tau\in V}L_n(\boldsymbol{\theta}_\tau)\leq L_n(\boldsymbol{\theta}_{\tau 0})=\frac{1}{n}\sum_{t=p+1}^{n}\omega_t\ell_t(\boldsymbol{\theta}_{\tau 0})\leq E[\omega_t\ell_t(\boldsymbol{\theta}_{\tau 0})]+\epsilon_0. \tag{A.10}$$

Combining (A.9) and (A.10), we have

$$\inf_{\boldsymbol{\theta}_\tau\in V^c}L_n(\boldsymbol{\theta}_\tau)\geq E[\omega_t\ell_t(\boldsymbol{\theta}_{\tau 0})]+2\epsilon_0>E[\omega_t\ell_t(\boldsymbol{\theta}_{\tau 0})]+\epsilon_0\geq\inf_{\boldsymbol{\theta}_\tau\in V}L_n(\boldsymbol{\theta}_\tau), \tag{A.11}$$

which implies that

$$\widehat{\boldsymbol{\theta}}_{\tau n}\in V\quad\text{a.s. for }\forall V,\text{ as } n\text{ is large enough.}$$

By the arbitrariness of V, it implies that $\widehat{\boldsymbol{\theta}}_{\tau n}\to\boldsymbol{\theta}_{\tau 0}$ a.s. The proof of this theorem is complete. ■

Proof of Theorem 3. For $\boldsymbol{u}\in\mathbb{R}^{2p+1}$, define $H_n(\boldsymbol{u})=n[L_n(\boldsymbol{\theta}_{\tau 0}+\boldsymbol{u})-L_n(\boldsymbol{\theta}_{\tau 0})]$, where $L_n(\boldsymbol{\theta}_\tau)=n^{-1}\sum_{t=p+1}^{n}\omega_t\rho_\tau(y_t-q_t(\boldsymbol{\theta}_\tau))$. Denote $\widehat{\boldsymbol{u}}_n=\widehat{\boldsymbol{\theta}}_{\tau n}-\boldsymbol{\theta}_{\tau 0}$. By Theorem 2, it holds that $\widehat{\boldsymbol{u}}_n=o_p(1)$. Note that $\widehat{\boldsymbol{u}}_n$ is the minimizer of $H_n(\boldsymbol{u})$, since $\widehat{\boldsymbol{\theta}}_{\tau n}$ minimizes $L_n(\boldsymbol{\theta}_\tau)$. Define $J=\Omega_1(\tau)/2$. By Assumption 3 and the ergodic theorem, $J_n=J+o_p(1)$, where J_n is defined in Lemma 2. Moreover, from Lemma 2, it follows that

$$\begin{aligned}H_n(\widehat{\boldsymbol{u}}_n)&=-\sqrt{n}\widehat{\boldsymbol{u}}_n'\boldsymbol{T}_n+\sqrt{n}\widehat{\boldsymbol{u}}_n'J\sqrt{n}\widehat{\boldsymbol{u}}_n+o_p(\sqrt{n}\|\widehat{\boldsymbol{u}}_n\|+n\|\widehat{\boldsymbol{u}}_n\|^2)\\&\geq-\sqrt{n}\|\widehat{\boldsymbol{u}}_n\|[\|\boldsymbol{T}_n\|+o_p(1)]+n\|\widehat{\boldsymbol{u}}_n\|^2[\lambda_{\min}+o_p(1)],\end{aligned} \tag{A.12}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

where $\lambda_{\min}$ is the smallest eigenvalue of J, and $\boldsymbol{T}_n$ is defined in Lemma 2. Note that, as $n \to \infty$, $\boldsymbol{T}_n$ converges in distribution to a normal random variable with mean zero and variance matrix $\tau(1-\tau)\Omega_0(\tau)$.

Since $H_n(\widehat{\boldsymbol{u}}_n) \le 0$, by Assumption 3, it holds that

$$\sqrt{n}\|\widehat{\boldsymbol{u}}_n\| \le [\lambda_{\min} + o_p(1)]^{-1}[\|\boldsymbol{T}_n\| + o_p(1)] = O_p(1). \tag{A.13}$$

This together with Theorem 2 verifies the $\sqrt{n}$-consistency. Hence, Statement (i) holds.

Let $\sqrt{n}\boldsymbol{u}_n^* = J^{-1}\boldsymbol{T}_n/2 = \Omega_1^{-1}(\tau)\boldsymbol{T}_n$, then we have

$$\sqrt{n}\boldsymbol{u}_n^* \to N\left(\boldsymbol{0}, \tau(1-\tau)\Omega_1^{-1}(\tau)\Omega_0(\tau)\Omega_1^{-1}(\tau)\right)$$

in distribution as $n \to \infty$. Therefore, it suffices to show that $\sqrt{n}\boldsymbol{u}_n^* - \sqrt{n}\widehat{\boldsymbol{u}}_n = o_p(1)$. By (A.12) and (A.13), we have

$$\begin{aligned} H_n(\widehat{\boldsymbol{u}}_n) =& -\sqrt{n}\widehat{\boldsymbol{u}}_n'\boldsymbol{T}_n + \sqrt{n}\widehat{\boldsymbol{u}}_n' J\sqrt{n}\widehat{\boldsymbol{u}}_n + o_p(1) \\ =& -2\sqrt{n}\widehat{\boldsymbol{u}}_n' J\sqrt{n}\boldsymbol{u}_n^* + \sqrt{n}\widehat{\boldsymbol{u}}_n' J\sqrt{n}\widehat{\boldsymbol{u}}_n + o_p(1), \end{aligned}$$

and

$$H_n(\boldsymbol{u}_n^*) = -\sqrt{n}\boldsymbol{u}_n^{*\prime}\boldsymbol{T}_n + \sqrt{n}\boldsymbol{u}_n^{*\prime} J\sqrt{n}\boldsymbol{u}_n^* + o_p(1) = -\sqrt{n}\boldsymbol{u}_n^{*\prime} J\sqrt{n}\boldsymbol{u}_n^* + o_p(1).$$

It follows that

$$\begin{aligned} H_n(\widehat{\boldsymbol{u}}_n) - H_n(\boldsymbol{u}_n^*) =& (\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n^*)' J(\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n^*) + o_p(1) \\ \ge& \lambda_{\min}\|\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n^*\|^2 + o_p(1). \end{aligned} \tag{A.14}$$

Since $H_n(\widehat{\boldsymbol{u}}_n) - H_n(\boldsymbol{u}_n^*) = n[L_n(\boldsymbol{\theta}_{\tau 0} + \widehat{\boldsymbol{u}}_n) - L_n(\boldsymbol{\theta}_{\tau 0} + \boldsymbol{u}_n^*)] \le 0$ a.s., then (A.14) implies that $\|\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n^*\| = o_p(1)$. We verify the asymptotic normality in Statement (ii); the proof is hence accomplished. ■

Proof of Corollary 2. Recall that $q_t(\boldsymbol{\theta}_\tau) = \boldsymbol{y}_{1,t-1}'\boldsymbol{\phi}(\tau) + S_{\mathrm{Q}}\left(b(\tau) + \boldsymbol{y}_{2,t-1}'\boldsymbol{\beta}(\tau)\right)$, $Q_\tau(y_{n+1}|\mathcal{F}_n) = q_{t+1}(\boldsymbol{\theta}_{\tau 0})$, and $\widehat{Q}_\tau(y_{n+1}|\mathcal{F}_n) = q_{t+1}(\widehat{\boldsymbol{\theta}}_{\tau n})$. Since $\min\{|b(\tau)|, |\beta_1(\tau)|, \ldots, |\beta_p(\tau)|\} \ge C$ for a constant $C > 0$ is assumed for $\boldsymbol{\theta}_\tau$, then the first and second derivative functions of $q_t(\boldsymbol{\theta}_\tau)$ are well defined. Specifically, the first derivative function is $\dot{q}_t(\boldsymbol{\theta}_\tau) = (\boldsymbol{y}_{1,t-1}', 0.5|h_t(\boldsymbol{\theta}_\tau)|^{-1/2}, 0.5|h_t(\boldsymbol{\theta}_\tau)|^{-1/2}\boldsymbol{y}_{2,t-1}')'$, and the second derivative function is

$$\ddot{q}_t(\boldsymbol{\theta}_\tau) = \begin{pmatrix} \boldsymbol{0} & -\dfrac{\operatorname{sgn}(h_t(\boldsymbol{\theta}_\tau))}{|h_t(\boldsymbol{\theta}_\tau)|^{3/2}} & -\dfrac{\operatorname{sgn}(h_t(\boldsymbol{\theta}_\tau))}{|h_t(\boldsymbol{\theta}_\tau)|^{3/2}}\boldsymbol{y}_{2,t-1}' \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\dfrac{\operatorname{sgn}(h_t(\boldsymbol{\theta}_\tau))}{|h_t(\boldsymbol{\theta}_\tau)|^{3/2}}\boldsymbol{y}_{2,t-1} & -\dfrac{\operatorname{sgn}(h_t(\boldsymbol{\theta}_\tau))}{|h_t(\boldsymbol{\theta}_\tau)|^{3/2}}\boldsymbol{y}_{2,t-1}\boldsymbol{y}_{2,t-1}' \end{pmatrix} \tag{A.15}$$

with $h_t(\boldsymbol{\theta}_\tau) = b(\tau) + \boldsymbol{y}_{2,t-1}'\boldsymbol{\beta}s(\tau)$. By the Taylor expansion, we have

$$q_{t+1}(\widehat{\boldsymbol{\theta}}_{\tau n}) - q_{t+1}(\boldsymbol{\theta}_{\tau 0}) = \dot{q}_{t+1}'(\boldsymbol{\theta}_{\tau 0})(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0})'\ddot{q}_{t+1}(\boldsymbol{\theta}_\tau^*)(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}), \tag{A.16}$$

where $\boldsymbol{\theta}_\tau^*$ is between $\boldsymbol{\theta}_{\tau 0}$ and $\widehat{\boldsymbol{\theta}}_{\tau n}$. By Theorem 3, we have $\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0} = O_p(n^{-1/2})$. Then, both $\dot{q}_t(\boldsymbol{\theta}_{\tau 0})$ and $\ddot{q}_t(\boldsymbol{\theta}_\tau^*)$ are well defined. To show the representation of Corollary 2, it is

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

sufficient to show that $\ddot{q}_{t+1}(\boldsymbol{\theta}^*_\tau) = O_p(1)$. The definition of $\ddot{q}_t(\cdot)$ in (A.15) together with $E(|y_t|) < \infty$ implies that $\ddot{q}_{t+1}(\boldsymbol{\theta}^*_\tau) = O_p(1)$. The proof of this corollary is complete. ■

Proof of Theorem 4. We first show that $\widehat{\mu}_{i,\tau} = \mu_{i,\tau} + o_p(1)$ and $\widehat{\sigma}^2_{i,\tau} = \sigma^2_{i,\tau} + o_p(1)$, for $i = 1$ and 2. Recall that $\eta_{t,\tau} = y_t - Q_\tau(y_t|\mathcal{F}_{t-1}) = y_t - q_t(\boldsymbol{\theta}_{\tau 0})$ and $\widehat{\eta}_{t,\tau} = y_t - \widehat{Q}_\tau(y_t|\mathcal{F}_{t-1}) = y_t - q_t(\widehat{\boldsymbol{\theta}}_{\tau n})$. By the Taylor expansion, we have

$$\widehat{\eta}_{t,\tau} - \eta_{t,\tau} = -[q_t(\widehat{\boldsymbol{\theta}}_{\tau n}) - q_t(\boldsymbol{\theta}_{\tau 0})] = -\dot{q}'_t(\boldsymbol{\theta}^*_\tau)(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}),$$

and

$$\widehat{\eta}^2_{t,\tau} - \eta^2_{t,\tau} = -[2y_t - q_t(\boldsymbol{\theta}_{\tau 0}) - q_t(\widehat{\boldsymbol{\theta}}_{\tau n})]\dot{q}'_t(\boldsymbol{\theta}^*_\tau)(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}),$$

where $\boldsymbol{\theta}^*_\tau$ is between $\boldsymbol{\theta}_{\tau 0}$ and $\widehat{\boldsymbol{\theta}}_{\tau n}$. Then, by the law of large numbers, $E(y_t^2) < \infty$, and the fact that $\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0} = O_p(n^{-1/2})$, it holds that

$$\widehat{\mu}_{1,\tau} = \frac{1}{n-p}\sum_{t=p+1}^{n} \eta_{t,\tau} + \frac{1}{n-p}\sum_{t=p+1}^{n} (\widehat{\eta}_{t,\tau} - \eta_{t,\tau}) = \mu_{1,\tau} + o_p(1), \tag{A.17}$$

and

$$\widehat{\sigma}^2_{1,\tau} = \frac{1}{n-p}\sum_{t=p+1}^{n} (\widehat{\eta}^2_{t,\tau} - \eta^2_{t,\tau}) + \frac{1}{n-p}\sum_{t=p+1}^{n} \eta^2_{t,\tau} - \widehat{\mu}^2_{1,\tau} = \sigma^2_{1,\tau} + o_p(1). \tag{A.18}$$

Similarly, we can show that

$$\begin{aligned}\widehat{\mu}_{2,\tau} &= \frac{1}{n-p}\sum_{t=p+1}^{n} |\eta_{t,\tau}| + \frac{1}{n-p}\sum_{t=p+1}^{n} (|\widehat{\eta}_{t,\tau}| - |\eta_{t,\tau}|) \\ &\le \frac{1}{n-p}\sum_{t=p+1}^{n} |\eta_{t,\tau}| + \frac{1}{n-p}\sum_{t=p+1}^{n} |\widehat{\eta}_{t,\tau} - \eta_{t,\tau}| = \mu_{2,\tau} + o_p(1),\end{aligned}$$

and

$$\widehat{\sigma}^2_{2,\tau} = \frac{1}{n-p}\sum_{t=p+1}^{n} (|\widehat{\eta}_{t,\tau}|^2 - |\eta_{t,\tau}|^2) + \frac{1}{n-p}\sum_{t=p+1}^{n} |\eta_{t,\tau}|^2 - \widehat{\mu}^2_{2,\tau} = \sigma^2_{2,\tau} + o_p(1).$$

Since $|\sum_{t=p+1}^{n} \psi_\tau(\widehat{\eta}_{t,\tau})| < 1$, by an elementary calculation, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} w_t\psi_\tau(\widehat{\eta}_{t,\tau})[\widehat{\eta}_{t-k,\tau} - \widehat{\mu}_{1,\tau}] \\ =&\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} w_t\psi_\tau(\widehat{\eta}_{t,\tau})\widehat{\eta}_{t-k,\tau} + O_p(n^{-1/2}) \\ =&\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} w_t\psi_\tau(\eta_{t,\tau})\eta_{t-k,\tau} + A_{n1} + A_{n2} + A_{n3} + O_p(n^{-1/2}),\end{aligned} \tag{A.19}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

where

$$A_{n1} = \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} w_t[\psi_\tau(\widehat{\eta}_{t,\tau}) - \psi_\tau(\eta_{t,\tau})]\eta_{t-k,\tau},$$

$$A_{n2} = \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} w_t\psi_\tau(\eta_{t,\tau})(\widehat{\eta}_{t-k,\tau} - \eta_{t-k,\tau}),$$

$$A_{n3} = \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} w_t[\psi_\tau(\widehat{\eta}_{t,\tau}) - \psi_\tau(\eta_{t,\tau})](\widehat{\eta}_{t-k,\tau} - \eta_{t-k,\tau}).$$

First, we consider A_{n1}. For any $\boldsymbol{\nu} \in \mathbb{R}^{2p+1}$, denote

$$\zeta_t(\boldsymbol{\nu}) = w_t[\psi_\tau(y_t - q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu})) - \psi_\tau(y_t - q_t(\boldsymbol{\theta}_{\tau 0}))]\eta_{t-k,\tau} \quad \text{and}$$

$$\phi_n(\boldsymbol{\nu}) = \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} \{\zeta_t(\boldsymbol{\nu}) - E[\zeta_t(\boldsymbol{\nu})|\mathcal{F}_{t-1}]\}.$$

Note that $w_t \in \mathcal{F}_{t-1}$. Then, by the Taylor expansion and the Cauchy–Schwarz inequality, together with $E[w_t\|\boldsymbol{y}_{1,t-1}\|^3] < \infty$ by Assumption 3 and the fact that $f_{t-1}(x)$ is bounded by Assumption 4, it holds that

$$\begin{aligned}
E[\zeta_t^2(\boldsymbol{\nu})] &= E\{w_t|I(y_t < q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu})) - I(y_t < q_t(\boldsymbol{\theta}_{\tau 0}))| \cdot w_t\eta_{t-k,\tau}^2\} \\
&\le \left\{E[w_t|F_{t-1}(q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu})) - F_{t-1}(q_t(\boldsymbol{\theta}_{\tau 0}))|]^2\right\}^{1/2} [E(w_t^2\eta_{t-k,\tau}^4)]^{1/2} \\
&\le C\left\{E[f_{t-1}^2(q_t(\boldsymbol{\theta}_\tau^*))|n^{-1/2}\boldsymbol{\nu}'w_t^2\dot{q}_t(\boldsymbol{\theta}_\tau^*)\dot{q}_t'(\boldsymbol{\theta}_\tau^*)n^{-1/2}\boldsymbol{\nu}|]\right\}^{1/2} \\
&\le Cn^{-1/2}\|\boldsymbol{\nu}\|\left\{\|E[f_{t-1}^2(\boldsymbol{\theta}_\tau^*)w_t^2\dot{q}_t(\boldsymbol{\theta}_\tau^*)\dot{q}_t'(\boldsymbol{\theta}_\tau^*)]\|\right\}^{1/2} = o(1),
\end{aligned}$$

where $\boldsymbol{\theta}_\tau^*$ is between $\boldsymbol{\theta}_{\tau 0}$ and $\widehat{\boldsymbol{\theta}}_{\tau n}$. This, together with $E[\phi_n(\boldsymbol{\nu})] = 0$, implies that

$$E[\phi_n^2(\boldsymbol{\nu})] \le \frac{1}{n} \sum_{t=p+k+1}^{n} E[\zeta_t^2(\boldsymbol{\nu})] = o(1). \tag{A.20}$$

For any $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2 \in \mathbb{R}^{2p+1}$ and $\delta > 0$, it holds that

$$\begin{aligned}
&I(y_t < q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_2)) - I(y_t < q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_1)) \\
&\quad = I(q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_1) < y_t < q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_2)) \\
&\qquad - I(q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_1) > y_t > q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_2)).
\end{aligned}$$

Then, by the Taylor expansion, for any $\boldsymbol{\nu}_1, \boldsymbol{\nu}_2 \in \mathbb{R}^{2p+1}$ and $\delta > 0$, it can be verified that

$$\begin{aligned}
&E \sup_{\|\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\| \le \delta} |\zeta_t(\boldsymbol{\nu}_1) - \zeta_t(\boldsymbol{\nu}_2)| \\
&\le \sup_{\|\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\| \le \delta} |w_t[I(y_t < q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_2)) - I(y_t < q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_1))]\eta_{t-k,\tau}| \\
&\le E\left\{w_t I(|y_t| \le \sup_{\|\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2\| \le \delta} |q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_2) - q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu}_1)|)|\eta_{t-k,\tau}|\right\}
\end{aligned}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

$$\leq E\left\{w_t \Pr\left(|y_t| \leq \sup_{\|\boldsymbol{\nu}_1-\boldsymbol{\nu}_2\|\leq\delta} n^{-1/2}\|\boldsymbol{\nu}_1-\boldsymbol{\nu}_2\|\|\dot{q}_t(\boldsymbol{\theta}^*_\tau)\|\right)|\eta_{t-k,\tau}|\right\}$$
$$\leq 2n^{-1/2}\delta E[\|w_t \sup_x f_{t-1}(x)\dot{q}_t(\boldsymbol{\theta}^*_\tau)\eta_{t-k,\tau}\|],$$

where $\boldsymbol{\theta}^*_\tau$ is between $\boldsymbol{\theta}_{\tau 0}+n^{-1/2}\boldsymbol{\nu}_1$ and $\boldsymbol{\theta}_{\tau 0}+n^{-1/2}\boldsymbol{\nu}_2$. This, together with $E[w_t\|\boldsymbol{y}_{1,t-1}\|^3] < \infty$ by Assumption 3, implies that

$$E\sup_{\|\boldsymbol{\nu}_1-\boldsymbol{\nu}_2\|\leq\delta}|\phi_n(\boldsymbol{\nu}_1)-\phi_n(\boldsymbol{\nu}_2)| \leq \frac{2}{\sqrt{n}}\sum_{t=p+k+1}^{n} E\sup_{\|\boldsymbol{\nu}_1-\boldsymbol{\nu}_2\|\leq\delta}|\zeta_t(\boldsymbol{\nu}_1)-\zeta_t(\boldsymbol{\nu}_2)| \leq C\delta. \tag{A.21}$$

Therefore, it follows from (A.20), (A.21), and the finite covering theorem that

$$\sup_{\|\boldsymbol{\nu}\|\leq M}|\phi_n(\boldsymbol{\nu})| = o_p(1). \tag{A.22}$$

Note that $E[\zeta_t(\boldsymbol{\nu})|\mathcal{F}_{t-1}] = w_t[F_{t-1}(q_t(\boldsymbol{\theta}_{\tau 0}))-F_{t-1}(q_t(\boldsymbol{\theta}_{\tau 0}+n^{-1/2}\boldsymbol{\nu}))]\eta_{t-k,\tau}$. Moreover, by the Taylor expansion and $E[w_t\|\boldsymbol{y}_{1,t-1}\|^3] < \infty$ by Assumption 3, we can show that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} E[\zeta_t(\boldsymbol{\nu})|\mathcal{F}_{t-1}]\\
=&-\frac{1}{n}\sum_{t=p+k+1}^{n} w_t f_{t-1}(F^{-1}_{t-1}(\tau))\dot{q}'_t(\boldsymbol{\theta}_{\tau 0})\eta_{t-k,\tau}\cdot\boldsymbol{\nu}\\
&-\boldsymbol{\nu}'\cdot\frac{1}{n}\sum_{t=p+k+1}^{n} w_t f_{t-1}(F^{-1}_{t-1}(\tau))\ddot{q}_t(\boldsymbol{\theta}^*_\tau)\eta_{t-k,\tau}\cdot n^{-1/2}\boldsymbol{\nu}^*\\
&-\boldsymbol{\nu}'\cdot\frac{1}{n}\sum_{t=p+k+1}^{n} w_t f_{t-1}(q_t(\boldsymbol{\theta}^{**}_\tau))\dot{q}_t(\boldsymbol{\theta}^{**}_\tau)\dot{q}'_t(\boldsymbol{\theta}^{\ddagger}_\tau)\eta_{t-k,\tau}\cdot n^{-1/2}\boldsymbol{\nu}^\dagger,
\end{aligned}$$

where $\boldsymbol{\nu}^*$ and $\boldsymbol{\nu}^\dagger$ are between $\mathbf{0}$ and $\boldsymbol{\nu}$, and $\boldsymbol{\theta}^*_\tau$, $\boldsymbol{\theta}^{**}_\tau$, and $\boldsymbol{\theta}^{\ddagger}_\tau$ are between $\boldsymbol{\theta}_{\tau 0}$ and $\boldsymbol{\theta}_{\tau 0}+n^{-1/2}\boldsymbol{\nu}$. Then, by the law of large numbers, Assumptions 3 and 4, it follows that

$$\sup_{\|\boldsymbol{\nu}\|\leq\delta}\left|\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} E[\zeta_t(\boldsymbol{\nu})|\mathcal{F}_{t-1}]+H'_{1k}\boldsymbol{\nu}\right| \leq n^{-1/2}C\delta+o_p(1) = o_p(1).$$

This, together with (A.22) and the fact that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n}-\boldsymbol{\theta}_{\tau 0}) = O_p(1)$, implies that

$$A_{n1} = \frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} E[\zeta_t(\boldsymbol{\nu})|\mathcal{F}_{t-1}]+o_p(1) = -H'_{1k}\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n}-\boldsymbol{\theta}_{\tau 0})+o_p(1). \tag{A.23}$$

Next, we consider A_{n2}. Since $E[\psi_\tau(\eta_{t,\tau})] = 0$, it holds that $E(A_{n2}) = 0$. By the Taylor expansion, the law of large numbers, and the fact that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n}-\boldsymbol{\theta}_{\tau 0}) = O_p(1)$, we can verify that $E(A^2_{n2}) = o_p(1)$ under Assumption 3. Hence,

$$A_{n2} = o_p(1). \tag{A.24}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Finally, we consider A_{n3}. For any $\boldsymbol{\nu} \in \mathbb{R}^{2p+1}$, denote $\eta_{t,\tau}(\boldsymbol{\nu}) = y_t - q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu})$ and

$$\varsigma_t(\boldsymbol{\nu}) = w_t[\psi_\tau(y_t - q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu})) - \psi_\tau(y_t - q_t(\boldsymbol{\theta}_{\tau 0}))][\eta_{t-k,\tau}(\boldsymbol{\nu}) - \eta_{t-k,\tau}].$$

By a method similar to the proof of (A.20) and (A.22), we can show that, for any $\delta > 0$,

$$\sup_{\|\boldsymbol{\nu}\| \le \delta} \left| \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} \{\varsigma_t(\boldsymbol{\nu}) - E[\varsigma_t(\boldsymbol{\nu}) | \mathcal{F}_{t-1}]\} \right| = o_p(1)$$

and

$$\sup_{\|\boldsymbol{\nu}\| \le \delta} \left| \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} E[\varsigma_t(\boldsymbol{\nu}) | \mathcal{F}_{t-1}] \right| = o_p(1).$$

As a result,

$$\sup_{\|\boldsymbol{\nu}\| \le \delta} \left| \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} w_t[\psi_\tau(y_t - q_t(\boldsymbol{\theta}_{\tau 0} + n^{-1/2}\boldsymbol{\nu})) - \psi_\tau(y_t - q_t(\boldsymbol{\theta}_{\tau 0}))] \right.$$
$$\left. [\eta_{t-k,\tau}(\boldsymbol{\nu}) - \eta_{t-k,\tau}] \right| = o_p(1),$$

which together with $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) = O_p(1)$ implies that

$$A_{n3} = o_p(1). \tag{A.25}$$

Combining (A.17)-(A.19), (A.23)-(A.25), and Theorem 3, we have

$$\widehat{\rho}_{k,\tau} = \frac{1}{\sqrt{(\tau - \tau^2)\sigma_{1,\tau}^2}} \frac{1}{n} \sum_{t=p+k+1}^{n} [w_t \eta_{t-k,\tau} - H_{1k}' \Omega_1^{-1}(\tau) w_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})] \psi_\tau(\eta_{t,\tau}) + o_p(n^{-1/2}).$$

Therefore, for $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_{1,\tau}, \ldots, \widehat{\rho}_{K,\tau})'$, we have

$$\widehat{\boldsymbol{\rho}} = \frac{1}{\sqrt{(\tau - \tau^2)\sigma_{1,\tau}^2}} \frac{1}{n} \sum_{t=p+k+1}^{n} [w_t \boldsymbol{\epsilon}_{1,t-1} - H_1(\tau) \Omega_1^{-1}(\tau) w_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})] \psi_\tau(\eta_{t,\tau}) + o_p(n^{-1/2}),$$

where $\boldsymbol{\epsilon}_{1,t} = (\eta_{t,\tau}, \ldots, \eta_{t-K+1,\tau})'$ and $H_1(\tau) = (H_{11}, \ldots, H_{1K})'$.

Let $\widehat{\boldsymbol{r}} = (\widehat{r}_{1,\tau}, \ldots, \widehat{r}_{K,\tau})'$. Similar to the proof of $\widehat{\boldsymbol{\rho}}$, we can show that

$$\widehat{\boldsymbol{r}} = \frac{1}{\sqrt{(\tau - \tau^2)\sigma_{2,\tau}^2}} \frac{1}{n} \sum_{t=p+k+1}^{n} [w_t \boldsymbol{\epsilon}_{2,t-1} - H_2(\tau) \Omega_1^{-1}(\tau) w_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})] \psi_\tau(\eta_{t,\tau}) + o_p(n^{-1/2}),$$

where $\boldsymbol{\epsilon}_{2,t} = (|\eta_{t,\tau}|, \ldots, |\eta_{t-K+1,\tau}|)'$ and $H_2(\tau) = (H_{21}, \ldots, H_{2K})'$. Therefore, it follows that

$$\sqrt{n}(\widehat{\boldsymbol{\rho}}', \widehat{\boldsymbol{r}}')'$$
$$= \frac{1}{\sqrt{\tau - \tau^2}} \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} [w_t \boldsymbol{\epsilon}_{t-1} - H(\tau) \Omega_1^{-1}(\tau) w_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})] \psi_\tau(\eta_{t,\tau}) + o_p(n^{-1/2}),$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

where $\boldsymbol{\epsilon}_t = (\sigma_{1,\tau}^{-1}\boldsymbol{\epsilon}'_{1,t}, \sigma_{2,\tau}^{-1}\boldsymbol{\epsilon}'_{2,t})'$ and $H(\tau) = (\sigma_{1,\tau}^{-1}H'_1(\tau), \sigma_{2,\tau}^{-1}H'_2(\tau))'$. Note that $E[\psi_\tau(\eta_{t,\tau})] = 0$ and $\text{var}[\psi_\tau(\eta_{t,\tau})] = \tau - \tau^2$. These together with the law of iterated expectations, we complete the proof by the central limit theorem and the Cramér–Wold device. ■

The following two preliminary lemmas are used to prove Theorem 3. Specifically, Lemma 1 verifies the stochastic differentiability condition defined by Pollard (1985), and the bracketing method in Pollard (1985) is used for its proof. Lemma 2 is used to obtain the $\sqrt{n}$-consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\tau n}$, and its proof needs Lemma 1.

LEMMA 1. *Under Assumptions 2–4, then for $\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0} = o_p(1)$, it holds that*

$$\zeta_n(\boldsymbol{\theta}_\tau) = o_p(\sqrt{n}\|\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0}\| + n\|\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0}\|^2),$$

where $\zeta_n(\boldsymbol{\theta}_\tau) = \sum_{t=p+1}^n \omega_t q_{1t}(\boldsymbol{\theta}_\tau)\left\{\xi_{1t}(\boldsymbol{\theta}_\tau) - E[\xi_{1t}(\boldsymbol{\theta}_\tau)|\mathcal{F}_{t-1}]\right\}$ *with* $q_{1t}(\boldsymbol{\theta}_\tau) = (\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0})'\dot{q}_t(\boldsymbol{\theta}_{\tau 0})$, *and*

$$\xi_{1t}(\boldsymbol{\theta}_\tau) = \int_0^1 \left[I(y_t \le F_{t-1}^{-1}(\tau) + q_{1t}(\boldsymbol{\theta}_\tau)s) - I(y_t \le F_{t-1}^{-1}(\tau))\right] ds.$$

Proof. Note that

$$|\zeta_n(\boldsymbol{\theta}_\tau)| \le \sqrt{n}\|\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0}\| \sum_{j=1}^{2p+1} \left| \frac{1}{\sqrt{n}} \sum_{t=p+1}^n m_{t,j}\left\{\xi_{1t}(\boldsymbol{\theta}_\tau) - E[\xi_{1t}(\boldsymbol{\theta}_\tau)|\mathcal{F}_{t-1}]\right\}\right|,$$

where $m_{t,j} = \omega_t \partial q_t(\boldsymbol{\theta}_{\tau 0})/\partial\theta_{\tau,j}$ with $\theta_{\tau,j}$ being the jth element of $\boldsymbol{\theta}_\tau$. For $1 \le j \le 2p+1$, define $g_t = \max_j\{m_{t,j}, 0\}$ or $g_t = \max_j\{-m_{t,j}, 0\}$. Denote $\boldsymbol{u} = \boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0}$. Let $f_t(\boldsymbol{u}) = g_t\xi_{1t}(\boldsymbol{\theta}_\tau)$ and define

$$D_n(\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \left\{f_t(\boldsymbol{u}) - E\left[f_t(\boldsymbol{u})|\mathcal{F}_{t-1}\right]\right\}.$$

To establish Lemma 1, it suffices to show that, for any $\delta > 0$,

$$\sup_{\|\boldsymbol{u}\| \le \delta} \frac{|D_n(\boldsymbol{u})|}{1 + \sqrt{n}\|\boldsymbol{u}\|} = o_p(1). \tag{A.26}$$

We follow the method in Lemma 4 of Pollard (1985) to verify (A.26). Let $\mathfrak{F} = \{f_t(\boldsymbol{u}) : \|\boldsymbol{u}\| \le \delta\}$ be a collection of functions indexed by $\boldsymbol{u}$. First, we verify that $\mathfrak{F}$ satisfies the bracketing condition defined on page 304 of Pollard (1985). Let $B_r(\boldsymbol{v})$ be an open neighborhood of $\boldsymbol{v}$ with radius $r > 0$, and define a constant C_0 to be selected later. For any $\epsilon > 0$ and $0 < r \le \delta$, there exists a sequence of small cubes $\{B_{\epsilon r/C_0}(\boldsymbol{u}_i)\}_{i=1}^{K(\epsilon)}$ to cover $B_r(\boldsymbol{0})$, where $K(\epsilon)$ is an integer less than $C\epsilon^{-(2p+1)}$, and the constant C is not depending on ϵ and r (see Huber, 1967, p. 227). Denote $V_i(r) = B_{\epsilon r/C_0}(\boldsymbol{u}_i) \bigcap B_r(\boldsymbol{0})$, and let $U_1(r) = V_1(r)$ and $U_i(r) = V_i(r) - \bigcup_{j=1}^{i-1} V_j(r)$ for $i \ge 2$. Note that $\{U_i(r)\}_{i=1}^{K(\epsilon)}$ is a partition of $B_r(\boldsymbol{0})$. For each $\boldsymbol{u}_i \in U_i(r)$ with $1 \le i \le K(\epsilon)$, define the following bracketing functions:

$$f_t^L(\boldsymbol{u}_i) = g_t \int_0^1 \left[I\left(y_t \le F_{t-1}^{-1}(\tau) + \boldsymbol{u}_i'\dot{q}_t(\boldsymbol{\theta}_{\tau 0})s - \frac{\epsilon r}{C_0}\|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|\right) - I(y_t \le F_{t-1}^{-1}(\tau))\right] ds,$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

$$f_t^U(\boldsymbol{u}_i) = g_t \int_0^1 \left[I\left(y_t \le F_{t-1}^{-1}(\tau) + \boldsymbol{u}_i' \dot{q}_t(\boldsymbol{\theta}_{\tau 0}) s + \frac{\epsilon r}{C_0} \|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\| \right) - I(y_t \le F_{t-1}^{-1}(\tau)) \right] ds.$$

Since the indicator function $I(\cdot)$ is nondecreasing and $g_t \ge 0$, for any $\boldsymbol{u} \in U_i(r)$, we have

$$f_t^L(\boldsymbol{u}_i) \le f_t(\boldsymbol{u}) \le f_t^U(\boldsymbol{u}_i). \tag{A.27}$$

Furthermore, by the Taylor expansion, it holds that

$$E\left[f_t^U(\boldsymbol{u}_i) - f_i^L(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \le \frac{\epsilon r}{C_0} \cdot 2 \sup_x f_{t-1}(x) \omega_t \|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|^2. \tag{A.28}$$

Denote $\Delta_t = 2 \sup_x f_{t-1}(x) \omega_t \|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|^2$. By Assumption 4, we have $\sup_x f_{t-1}(x) < \infty$. Choose $C_0 = E(\Delta_t)$. Then, by iterated expectation and Assumption 3, it follows that

$$E\left[f_t^U(\boldsymbol{u}_i) - f_t^L(\boldsymbol{u}_i) \right] = E\left\{ E\left[f_t^U(\boldsymbol{\theta}_i) - f_t^L(\boldsymbol{\theta}_i) | \mathcal{F}_{t-1} \right] \right\} \le \epsilon r.$$

This, together with (A.27), implies that the family $\mathfrak{F}$ satisfies the bracketing condition.

Put $r_k = 2^{-k}\delta$. Let $B(k) = B_{r_k}(\mathbf{0})$ and $A(k)$ be the annulus $B(k) \setminus B(k+1)$. From the bracketing condition, for fixed $\epsilon > 0$, there is a partition $U_1(r_k), U_2(r_k), \ldots, U_{K(\epsilon)}(r_k)$ of $B(k)$. First, consider the upper tail case. For $\boldsymbol{u} \in U_i(r_k)$, by (A.28), it holds that

$$\begin{aligned} D_n(\boldsymbol{u}) \le & \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \left\{ f_t^U(\boldsymbol{u}_i) - E\left[f_t^U(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \right\} + \frac{1}{\sqrt{n}} \sum_{t=p+1}^n E\left[f_t^U(\boldsymbol{u}_i) - f_t^L(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \\ \le & D_n^U(\boldsymbol{u}_i) + \sqrt{n} \epsilon r_k \frac{1}{n C_0} \sum_{t=p+1}^n \Delta_t, \end{aligned} \tag{A.29}$$

where

$$D_n^U(\boldsymbol{u}_i) = \frac{1}{\sqrt{n}} \sum_{t=p+1}^n \left\{ f_t^U(\boldsymbol{u}_i) - E\left[f_t^U(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \right\}.$$

Define the event

$$E_n = \left\{ \omega : \frac{1}{n C_0} \sum_{t=p+1}^n \Delta_t(\omega) < 2 \right\}.$$

For $\boldsymbol{u} \in A(k)$, $1 + \sqrt{n}\|\boldsymbol{u}\| > \sqrt{n} r_{k+1} = \sqrt{n} r_k / 2$. Then, by (A.29) and Chebyshev's inequality, we have

$$\begin{aligned} \mathrm{P}\left(\sup_{\boldsymbol{u} \in A(k)} \frac{D_n(\boldsymbol{u})}{1 + \sqrt{n}\|\boldsymbol{u}\|} > 6\epsilon, E_n \right) \le & \mathrm{P}\left(\max_{1 \le i \le K(\epsilon)} \sup_{\boldsymbol{u} \in U_i(r_k) \cap A(k)} D_n(\boldsymbol{u}) > 3\sqrt{n}\epsilon r_k, E_n \right) \\ \le & K(\epsilon) \max_{1 \le i \le K(\epsilon)} \mathrm{P}\left(D_n^U(\boldsymbol{u}_i) > \sqrt{n}\epsilon r_k \right) \\ \le & K(\epsilon) \max_{1 \le i \le K(\epsilon)} \frac{E\{[D_n^U(\boldsymbol{u}_i)]^2\}}{n\epsilon^2 r_k^2}. \end{aligned} \tag{A.30}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Moreover, by iterated expectation, the Taylor expansion, Assumption 4, and $\|\boldsymbol{u}_i\| \le r_k$, for $\boldsymbol{u}_i \in U_i(r_k)$, we have

$$\begin{aligned}
&E\left\{[f_t^U(\boldsymbol{u}_i)]^2\right\} = E\left\{E\left\{[f_t^U(\boldsymbol{u}_i)]^2|\mathcal{F}_{t-1}\right\}\right\}\\
\le& 2E\left\{g_t^2\left|\int_0^1\left[F_{t-1}\left(F_{t-1}^{-1}(\tau)+\boldsymbol{u}_i'\dot{q}_t(\boldsymbol{\theta}_{\tau 0})s+\frac{\epsilon r_k}{C_0}\|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|\right)-F_{t-1}\left(F_{t-1}^{-1}(\tau)\right)\right]ds\right|\right\}\\
\le& C\sup_x f_{t-1}(x) r_k E\left[\omega_t^2\|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|^3\right].
\end{aligned}$$

This, together with $E[w_t\|\boldsymbol{y}_{1,t-1}\|^3] < \infty$ by Assumption 3, $\sup_x f_{t-1}(x) < \infty$ by Assumption 4, and the fact that $f_t^U(\boldsymbol{u}_i) - E[f_t^U(\boldsymbol{u}_i)|\mathcal{F}_{t-1}]$ is a martingale difference sequence, implies that

$$\begin{aligned}
E\{[D_n^U(\boldsymbol{u}_i)]^2\} &= \frac{1}{n}\sum_{t=p+1}^{n} E\{\{f_t^U(\boldsymbol{u}_i) - E[f_t^U(\boldsymbol{u}_i)|\mathcal{F}_{t-1}]\}^2\}\\
&\le \frac{1}{n}\sum_{t=p+1}^{n} E\{[f_t^U(\boldsymbol{u}_i)]^2\}\\
&\le \frac{Cr_k}{n}\sum_{t=p+1}^{n} E\left[\omega_t^2\|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|^3\right] := \Delta(r_k). \qquad \textbf{(A.31)}
\end{aligned}$$

Combining (A.30) and (A.31), we have

$$\mathrm{P}\left(\sup_{\boldsymbol{u}\in A(k)}\frac{D_n(\boldsymbol{u})}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\epsilon, E_n\right) \le \frac{K(\epsilon)\Delta(r_k)}{n\epsilon^2 r_k^2}.$$

Similar to the proof of the upper tail case, we can obtain the same bound for the lower tail case. Therefore,

$$\mathrm{P}\left(\sup_{\boldsymbol{u}\in A(k)}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\epsilon, E_n\right) \le \frac{2K(\epsilon)\Delta(r_k)}{n\epsilon^2 r_k^2}. \qquad \textbf{(A.32)}$$

Note that since $\Delta(r_k) \to 0$ as $k \to \infty$, we can choose k_ϵ such that $2K(\epsilon)\Delta(r_k)/(\epsilon^2\delta^2) < \epsilon$, for $k \ge k_\epsilon$. Let k_n be the integer such that $n^{-1/2}\delta \le r_{k_n} \le 2n^{-1/2}\delta$, and split $B_\delta(\mathbf{0})$ into two events $B := B(k_n+1)$ and $B^c := B(0) - B(k_n+1)$. Note that $B^c = \bigcup_{k=0}^{k_n} A(k)$ and $\Delta(r_k)$ is bounded by Assumption 3. Then, by (A.32), it holds that

$$\begin{aligned}
\mathrm{P}\left(\sup_{\boldsymbol{u}\in B^c}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\epsilon\right) &\le \sum_{k=0}^{k_n}\mathrm{P}\left(\sup_{\boldsymbol{u}\in A(k)}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\epsilon, E_n\right) + \mathrm{P}(E_n^c)\\
&\le \frac{1}{n}\sum_{k=0}^{k_\epsilon-1}\frac{CK(\epsilon)}{\epsilon^2\delta^2}2^{2k} + \frac{\epsilon}{n}\sum_{k=k_\epsilon}^{k_n}2^{2k} + \mathrm{P}(E_n^c)\\
&\le O\left(\frac{1}{n}\right) + 4\epsilon + \mathrm{P}(E_n^c). \qquad \textbf{(A.33)}
\end{aligned}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Furthermore, for $\boldsymbol{u} \in B$, we have $1+\sqrt{n}\|\boldsymbol{u}\| \geq 1$ and $r_{k_n+1} \leq n^{-1/2}\delta < n^{-1/2}$. Similar to the proof of (A.30) and (A.31), we can show that

$$\mathrm{P}\left(\sup_{\boldsymbol{u}\in B}\frac{D_n(\boldsymbol{u})}{1+\sqrt{n}\|\boldsymbol{u}\|} > 3\epsilon, E_n\right) \leq \mathrm{P}\left(\max_{1\leq i\leq K(\epsilon)} D_n^U(\boldsymbol{u}_i) > \epsilon, E_n\right) \leq \frac{K(\epsilon)\Delta(r_{k_n+1})}{\epsilon^2}.$$

We can obtain the same bound for the lower tail. Therefore, we have

$$\begin{aligned}\mathrm{P}\left(\sup_{\boldsymbol{u}\in B}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 3\epsilon\right) =& \mathrm{P}\left(\sup_{\boldsymbol{u}\in B}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 3\epsilon, E_n\right) + \mathrm{P}(E_n^c)\\ &\leq \frac{2K(\epsilon)\Delta(r_{k_n+1})}{\epsilon^2} + \mathrm{P}(E_n^c). \end{aligned} \tag{A.34}$$

Note that $\Delta(r_{k_n+1}) \to 0$ as $n \to \infty$. Moreover, by the ergodic theorem, $\mathrm{P}(E_n) \to 1$ and thus $\mathrm{P}(E_n^c) \to 0$ as $n \to \infty$. Equation (A.34) together with (A.33) asserts (A.26). The proof of this lemma is accomplished. ■

LEMMA 2. *Suppose that Assumptions 2–4 hold, then*

$$\begin{aligned}n[L_n(\boldsymbol{\theta}_\tau) - L_n(\boldsymbol{\theta}_{\tau 0})] =& -\sqrt{n}(\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0})'\boldsymbol{T}_n + \sqrt{n}(\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0})' J_n \sqrt{n}(\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0})\\ &+ o_p(\sqrt{n}\|\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0}\| + n\|\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0}\|^2),\end{aligned}$$

for $\boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0} = o_p(1)$, *where* $L_n(\boldsymbol{\theta}_\tau) = n^{-1}\sum_{t=p+1}^n \omega_t \rho_\tau(y_t - q_t(\boldsymbol{\theta}_\tau))$, *and*

$$\boldsymbol{T}_n = \frac{1}{\sqrt{n}}\sum_{t=p+1}^n \omega_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})\psi_\tau(\eta_{t,\tau}) \text{ and } J_n = \frac{1}{2n}\sum_{t=p+1}^n f_{t-1}(F_{t-1}^{-1}(\tau))\omega_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})\dot{q}_t'(\boldsymbol{\theta}_{\tau 0})$$

with $\eta_{t,\tau} = y_t - q_t(\boldsymbol{\theta}_{\tau 0})$.

Proof. Denote $\boldsymbol{u} = \boldsymbol{\theta}_\tau - \boldsymbol{\theta}_{\tau 0}$. Let $v_t(\boldsymbol{u}) = q_t(\boldsymbol{\theta}_\tau) - q_t(\boldsymbol{\theta}_{\tau 0})$, and define the function $\xi_t(\boldsymbol{u}) = \int_0^1 \left[I(y_t \leq F_{t-1}^{-1}(\tau) + v_t(\boldsymbol{u})s) - I(y_t \leq F_{t-1}^{-1}(\tau))\right] ds$. Recall that $L_n(\boldsymbol{\theta}_\tau) = n^{-1}\sum_{t=p+1}^n \omega_t \rho_\tau(y_t - q_t(\boldsymbol{\theta}_\tau))$ and $q_t(\boldsymbol{\theta}_{\tau 0}) = F_{t-1}^{-1}(\tau)$. By the Knight identity (A.3), it can be verified that

$$\begin{aligned}n[L_n(\boldsymbol{\theta}_\tau) - L_n(\boldsymbol{\theta}_{\tau 0})] =& \sum_{t=p+1}^n \omega_t\left[\rho_\tau\left(\eta_{t,\tau} - v_t(\boldsymbol{u})\right) - \rho_\tau\left(\eta_{t,\tau}\right)\right]\\ =& K_{1n}(\boldsymbol{u}) + K_{2n}(\boldsymbol{u}),\end{aligned} \tag{A.35}$$

where

$$K_{1n}(\boldsymbol{u}) = -\sum_{t=p+1}^n \omega_t v_t(\boldsymbol{u})\psi_\tau(\eta_{t,\tau}) \text{ and } K_{2n}(\boldsymbol{u}) = \sum_{t=p+1}^n \omega_t v_t(\boldsymbol{u})\xi_t(\boldsymbol{u}).$$

By the Taylor expansion, we have $v_t(\boldsymbol{u}) = q_{1t}(\boldsymbol{u}) + q_{2t}(\boldsymbol{u})$, where $q_{1t}(\boldsymbol{u}) = \boldsymbol{u}'\dot{q}_t(\boldsymbol{\theta}_{\tau 0})$ and $q_{2t}(\boldsymbol{u}) = \boldsymbol{u}'\ddot{q}_t(\boldsymbol{\theta}_\tau^*)\boldsymbol{u}/2$ for $\boldsymbol{\theta}_\tau^*$ between $\boldsymbol{\theta}_\tau$ and $\boldsymbol{\theta}_{\tau 0}$, and $\ddot{q}_t(\boldsymbol{\theta}_\tau)$ is defined as (A.15). Then, it

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

follows that

$$K_{1n}(\boldsymbol{u}) = -\sum_{t=p+1}^{n} \omega_t q_{1t}(\boldsymbol{u})\psi_\tau(\eta_{t,\tau}) - \sum_{t=p+1}^{n} \omega_t q_{2t}(\boldsymbol{u})\psi_\tau(\eta_{t,\tau})$$
$$= -\sqrt{n}\boldsymbol{u}'\boldsymbol{T}_n - \sqrt{n}\boldsymbol{u}' R_{1n}(\boldsymbol{\theta}_\tau^*)\sqrt{n}\boldsymbol{u}, \tag{A.36}$$

where

$$\boldsymbol{T}_n = \frac{1}{\sqrt{n}}\sum_{t=p+1}^{n} \omega_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})\psi_\tau(\eta_{t,\tau}) \text{ and } R_{1n}(\boldsymbol{\theta}_\tau^*) = \frac{1}{2n}\sum_{t=p+1}^{n} \omega_t \ddot{q}_t(\boldsymbol{\theta}_\tau^*)\psi_\tau(\eta_{t,\tau}).$$

From $E[w_t\|\boldsymbol{y}_{1,t-1}\|^3] < \infty$ by Assumption 3 and the fact that $|\psi_\tau(\eta_{t,\tau})| \le 1$, we have

$$E[\sup_{\boldsymbol{\theta}_\tau^* \in \Theta_\tau} \left\|\omega_t \ddot{q}_t(\boldsymbol{\theta}_\tau^*)\psi_\tau(\eta_{t,\tau})\right\|] \le CE[\sup_{\boldsymbol{\theta}_\tau^* \in \Theta_\tau} \left\|\omega_t \ddot{q}_t(\boldsymbol{\theta}_\tau^*)\right\|] < \infty.$$

Moreover, by iterated expectation and the fact that $E[\psi_\tau(\eta_{t,\tau})] = 0$, it follows that

$$E\left[\omega_t \ddot{q}_t(\boldsymbol{\theta}_\tau^*)\psi_\tau(\eta_{t,\tau})\right] = 0.$$

Then, by Theorem 3.1 in Ling and McAleer (2003), we can show that

$$\sup_{\boldsymbol{\theta}_\tau^* \in \Theta_\tau} \|R_{1n}(\boldsymbol{\theta}_\tau^*)\| = o_p(1).$$

This, together with (A.36), implies that

$$K_{1n}(\boldsymbol{u}) = -\sqrt{n}\boldsymbol{u}'\boldsymbol{T}_n + o_p(n\|\boldsymbol{u}\|^2). \tag{A.37}$$

Denote $\xi_t(\boldsymbol{u}) = \xi_{1t}(\boldsymbol{u}) + \xi_{2t}(\boldsymbol{u})$, where

$$\xi_{1t}(\boldsymbol{u}) = \int_0^1 \left[I(y_t \le F_{t-1}^{-1}(\tau) + q_{1t}(\boldsymbol{u})s) - I(y_t \le F_{t-1}^{-1}(\tau))\right] ds \text{ and}$$
$$\xi_{2t}(\boldsymbol{u}) = \int_0^1 \left[I(y_t \le F_{t-1}^{-1}(\tau) + v_t(\boldsymbol{u})s) - I(y_t \le F_{t-1}^{-1}(\tau) + q_{1t}(\boldsymbol{u})s)\right] ds.$$

Then, for $K_{2n}(\boldsymbol{\theta}_\tau)$, it holds that

$$K_{2n}(\boldsymbol{u}) = R_{2n}(\boldsymbol{u}) + R_{3n}(\boldsymbol{u}) + R_{4n}(\boldsymbol{u}) + R_{5n}(\boldsymbol{u}), \tag{A.38}$$

where

$$R_{2n}(\boldsymbol{u}) = \boldsymbol{u}' \sum_{t=p+1}^{n} \omega_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0}) E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}],$$
$$R_{3n}(\boldsymbol{u}) = \boldsymbol{u}' \sum_{t=p+1}^{n} \omega_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})\{\xi_{1t}(\boldsymbol{u}) - E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}]\},$$
$$R_{4n}(\boldsymbol{u}) = \boldsymbol{u}' \sum_{t=p+1}^{n} \omega_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0})\xi_{2t}(\boldsymbol{u}) \text{ and } R_{5n}(\boldsymbol{u}) = \frac{\boldsymbol{u}'}{2}\sum_{t=p+1}^{n} \omega_t \ddot{q}_t(\boldsymbol{\theta}_\tau^*)\xi_t(\boldsymbol{u})\boldsymbol{u}.$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Note that

$$E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}] = \int_0^1 [F_{t-1}(F_{t-1}^{-1}(\tau) + q_{1t}(\boldsymbol{u})s) - F_{t-1}(F_{t-1}^{-1}(\tau))]ds. \tag{A.39}$$

Then, by the Taylor expansion, together with Assumption 4, it follows that

$$\begin{aligned} E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}] =& \frac{1}{2} f_{t-1}(F_{t-1}^{-1}(\tau)) q_{1t}(\boldsymbol{u}) \\ &+ q_{1t}(\boldsymbol{u}) \int_0^1 [f_{t-1}(F_{t-1}^{-1}(\tau) + q_{1t}(\boldsymbol{u})s^*) - f_{t-1}(F_{t-1}^{-1}(\tau))]sds, \end{aligned}$$

where s^* is between 0 and s. Therefore, it follows that

$$R_{2n}(\boldsymbol{u}) = \sqrt{n}\boldsymbol{u}' J_n \sqrt{n}\boldsymbol{u} + \sqrt{n}\boldsymbol{u}' \Pi_{1n}(\boldsymbol{u}) \sqrt{n}\boldsymbol{u}, \tag{A.40}$$

where $J_n = (2n)^{-1} \sum_{t=p+1}^n f_{t-1}(F_{t-1}^{-1}(\tau)) \omega_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0}) \dot{q}_t'(\boldsymbol{\theta}_{\tau 0})$ and

$$\Pi_{1n}(\boldsymbol{u}) = \frac{1}{n} \sum_{t=p+1}^n \omega_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0}) \dot{q}_t'(\boldsymbol{\theta}_{\tau 0}) \int_0^1 [f_{t-1}(F_{t-1}^{-1}(\tau) + q_{1t}(\boldsymbol{u})s^*) - f_{t-1}(F_{t-1}^{-1}(\tau))]sds.$$

By the Taylor expansion and Assumptions 3 and 4, for any $\eta > 0$, it holds that

$$\begin{aligned} E\left(\sup_{\|\boldsymbol{u}\| \le \eta} \|\Pi_{1n}(\boldsymbol{u})\| \right) &\le \frac{1}{n} \sum_{t=p+1}^n E\left[\sup_{\|\boldsymbol{u}\| \le \eta} \|\omega_t \dot{q}_t(\boldsymbol{\theta}_{\tau 0}) \dot{q}_t'(\boldsymbol{\theta}_{\tau 0})\| \sup_x |\dot{f}_{t-1}(x)| |\boldsymbol{u}' \dot{q}_t(\boldsymbol{\theta}_{\tau 0})\| \right] \\ &\le C\eta \sup_x |\dot{f}_{t-1}(x)| E[\omega_t \|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|^3] \end{aligned}$$

tends to 0 as $\eta \to 0$. Therefore, for any $\epsilon, \delta > 0$, there exists $\eta_0 = \eta_0(\epsilon) > 0$ such that

$$\Pr\left(\sup_{\|\boldsymbol{u}\| \le \eta_0} \|\Pi_{1n}(\boldsymbol{u})\| > \delta \right) < \frac{\epsilon}{2}, \tag{A.41}$$

for all $n \ge 1$. Since $\boldsymbol{u} = o_p(1)$, it follows that

$$\Pr(\|\boldsymbol{u}\| > \eta_0) < \frac{\epsilon}{2}, \tag{A.42}$$

as n is large enough. From (A.41) and (A.42), we have

$$\begin{aligned} \Pr(\|\Pi_{1n}(\boldsymbol{u})\| > \delta) &\le \Pr(\|\Pi_{1n}(\boldsymbol{u})\| > \delta, \|\boldsymbol{u}\| \le \eta_0) + \Pr(\|\boldsymbol{u}\| > \eta_0) \\ &\le \Pr\left(\sup_{\|\boldsymbol{u}\| \le \eta_0} \|\Pi_{1n}(\boldsymbol{u})\| > \delta \right) + \frac{\epsilon}{2} < \epsilon, \end{aligned}$$

as n is large enough. Therefore, $\Pi_{1n}(\boldsymbol{u}) = o_p(1)$. This, together with (A.40), implies that

$$R_{2n}(\boldsymbol{u}) = \sqrt{n}\boldsymbol{u}' J_n \sqrt{n}\boldsymbol{u} + o_p(n\|\boldsymbol{u}\|^2). \tag{A.43}$$

For $R_{3n}(\boldsymbol{u})$, by Lemma 1, it holds that

$$R_{3n}(\boldsymbol{u}) = o_p(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2). \tag{A.44}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Note that

$$E[\xi_{2t}(\boldsymbol{u})|\mathcal{F}_{t-1}] = \int_0^1 \left[F_{t-1}(F_{t-1}^{-1}(\tau) + v_t(\boldsymbol{u})s) - F_{t-1}(F_{t-1}^{-1}(\tau) + q_{1t}(\boldsymbol{u})s)\right] ds. \quad \textbf{(A.45)}$$

Then, by iterated expectation, the Taylor expansion, and the Cauchy–Schwarz inequality, together with $E[w_t\|\boldsymbol{y}_{1,t-1}\|^3] < \infty$ by Assumption 3 and $\sup_x |\dot{f}_{t-1}(x)| < \infty$ by Assumption 4, for any $\eta > 0$, it holds that

$$\begin{aligned} E\left(\sup_{\|\boldsymbol{u}\|\le\eta} \frac{|R_{4n}(\boldsymbol{u})|}{n\|\boldsymbol{u}\|^2}\right) \le& \frac{\eta}{n}\sum_{t=p+1}^{n} E\left\{\omega_t \|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\| \frac{1}{2}\sup_x f_{t-1}(x) \sup_{\Theta_\tau}\|\ddot{q}_t(\boldsymbol{\theta}_\tau)\|\right\} \\ \le& C\eta E\left\{\|\sqrt{\omega_t}\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\| \sup_{\Theta_\tau}\|\sqrt{\omega_t}\ddot{q}_t(\boldsymbol{\theta}_\tau)\|\right\} \\ \le& C\eta\left[E\left(\omega_t\|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|^2\right)\right]^{1/2}\left[E\left(\sup_{\Theta_\tau}\omega_t\|\ddot{q}_t(\boldsymbol{\theta}_\tau)\|^2\right)\right]^{1/2} \end{aligned}$$

tends to 0 as $\eta \to 0$. Similar to (A.41) and (A.42), we can show that

$$R_{4n}(\boldsymbol{u}) = o_p(n\|\boldsymbol{u}\|^2). \quad \textbf{(A.46)}$$

For $R_{5n}(\boldsymbol{u})$, it follows that

$$R_{5n}(\boldsymbol{u}) = \sqrt{n}\boldsymbol{u}'\Pi_{2n}(\boldsymbol{u})\sqrt{n}\boldsymbol{u} + \sqrt{n}\boldsymbol{u}'\Pi_{3n}(\boldsymbol{u})\sqrt{n}\boldsymbol{u}, \quad \textbf{(A.47)}$$

where

$$\Pi_{2n}(\boldsymbol{u}) = \frac{1}{2n}\sum_{t=p+1}^{n}\omega_t\ddot{q}_t(\boldsymbol{\theta}_\tau^*)\xi_{1t}(\boldsymbol{u}) \text{ and } \Pi_{3n}(\boldsymbol{u}) = \frac{1}{2n}\sum_{t=p+1}^{n}\omega_t\ddot{q}_t(\boldsymbol{\theta}_\tau^*)\xi_{2t}(\boldsymbol{u}).$$

Similar to the proof of $R_{4n}(\boldsymbol{u})$, by (A.39), we can show that, for any $\eta > 0$,

$$\begin{aligned} E\left(\sup_{\|\boldsymbol{u}\|\le\eta}\|\Pi_{2n}(\boldsymbol{u})\|\right) \le& \frac{\eta}{n}\sum_{t=p+1}^{n} E\left\{\omega_t \sup_{\Theta_\tau}\|\ddot{q}_t(\boldsymbol{\theta}_\tau)\| \frac{1}{2}\sup_x f_{t-1}(x)\|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|\right\} \\ \le& C\eta\left[E\left(\sup_{\Theta_\tau}\omega_t\|\ddot{q}_t(\boldsymbol{\theta}_\tau)\|^2\right)\right]^{1/2}\left[E\left(\omega_t\|\dot{q}_t(\boldsymbol{\theta}_{\tau 0})\|^2\right)\right]^{1/2} \end{aligned}$$

tends to 0 as $\eta \to 0$. And for $\Pi_{3n}(\boldsymbol{u})$, by (A.45), we have

$$\begin{aligned} E\left(\sup_{\|\boldsymbol{u}\|\le\eta}\|\Pi_{3n}(\boldsymbol{u})\|\right) \le& \eta^2 E\left\{\frac{1}{n}\sum_{t=p+1}^{n}\omega_t \sup_{\Theta_\tau}\|\ddot{q}_t(\boldsymbol{\theta}_\tau)\| \frac{1}{2}\sup_{x\in\mathbb{R}} f_{t-1}(x)\sup_{\Theta_\tau}\|\ddot{q}_t(\boldsymbol{\theta}_\tau)\|\right\} \\ \le& C\eta^2 E\left(\sup_{\Theta_\tau}\omega_t\|\ddot{q}_t(\boldsymbol{\theta}_\tau)\|^2\right) \end{aligned}$$

tends to 0 as $\eta \to 0$. Similar to (A.41) and (A.42), we can show that $\Pi_{2n}(\boldsymbol{u}) = o_p(1)$ and $\Pi_{3n}(\boldsymbol{u}) = o_p(1)$. Therefore, together with (A.47), it follows that

$$R_{5n}(\boldsymbol{u}) = o_p(n\|\boldsymbol{u}\|^2). \quad \textbf{(A.48)}$$

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

From (A.38), (A.43)-(A.46), and (A.48), we have

$$K_{2n}(\boldsymbol{u}) = \sqrt{n}\boldsymbol{u}' J_n \sqrt{n}\boldsymbol{u} + o_p(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2). \quad \textbf{(A.49)}$$

In view of (A.35), (A.37), and (A.49), we accomplish the proof of this lemma. ■

SUPPLEMENTARY MATERIAL

To view the supplementary material for this article, please visit https://doi.org/10.1017/S026646662100030X

REFERENCES

An, H. Z. & Z. G. Chen (1982) On convergence of LAD estimates in autoregresion with infinite variance. *Journal of Multivariate Analysis* 12, 335–345.

Artzner, P., F. Delbaen, J. M. Eber, & D. Heath (1999) Coherent measures of risk. *Mathematical Finance* 9, 203–228.

Bassett, G., R. Koenker, & G. Kordas (2004) Pessimistic portfolio allocation and Choquet expected utility. *Journal of Financial Econometrics* 4, 477–492.

Bofinger, E. (1975) Estimation of a density function using order statistics. *Australian Journal of Statistics* 17, 1–7.

Bollerslev, T. (1986) Generalized autoregression conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.

Box, G. E. P. & G. M. Jenkins (1970) *Time Series Analysis, Forecasting and Control*. Holden-Day.

Cai, Y., G. Montes-Rojas, & J. Olmo (2013) Quantile double AR time series models for financial returns. *Journal of Forecasting* 32, 551–560.

Chan, N. H. & L. Peng (2005) Weighted least absolute deviations estimation for an AR(1) process with ARCH(1) errors. *Biometrika* 92, 477–484.

Chernozhukov, V., I. Fernández-Val, & A. Galichon (2010) Quantile and probability curves without crossing. *Econometrica* 78, 1093–1125.

Christoffersen, P. F. (1998) Evaluating interval forecasts. *International Economic Review* 39, 841–862.

Davis, R. A., K. Knight, & J. Liu (1992) M-estimation for autoregressions with infinite variances. *Stochastic Processes and Their Applications* 40, 145–180.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

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

Feigin, P. D. & R. L. Tweedie (1985) Random coefficient autoregressive processes: A Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis* 6, 1–14.

Francq, C. & J.-M. Zakoian (2004) Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes. *Bernoulli* 10, 605–637.

Francq, C. & J.-M. Zakoian (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.

Francq, C. & J.-M. Zakoian (2015) Risk-parameter estimation in volatility models. *Journal of Econometrics* 184, 158–174.

Galvao Jr., A. F., G. Montes-Rojas, & J. Olmo (2011) Threshold quantile autoregressive models. *Journal of Time Series Analysis* 32, 253–267.

Gross, S. & W. L. Steiger (1979) Least absolute deviation estimates in autoregression with infinite variance. *Journal of Applied Probability* 16, 104–116.

Hall, P. & S. Sheather (1988) On the distribution of a studentized quantile. *Journal of the Royal Statistical Society, Series B* 50, 381–391.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

Huber, P. (1967) The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 221–233.

Huber, P. (1973) Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics* 1, 799–821.

Jiang, F., D. Li, & K. Zhu (2020) Non-standard inference for augmented double autoregressive models with null volatility coefficients. *Journal of Econometrics* 215, 165–183.

Knight, K. (1998) Limiting distributions for L_1 regression estimators under general conditions. *Annals of Statistics* 26, 755–770.

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

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

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

Lee, S. & J. Noh (2013) Quantile regression estimator for GARCH models. *Scandinavian Journal of Statistics* 40, 2–20.

Li, D., S. Ling, & R. Zhang (2016) On a threshold double autoregressive model. *Journal of Business & Economic Statistics* 34, 68–80.

Li, G. & W. K. Li (2005) Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika* 92, 691–701.

Li, G. & W. K. Li (2008) Least absolute deviation estimation for fractionally integrated autoregressive moving average time series models with conditional heteroscedasticity. *Biometrika* 95, 399–414.

Li, G. & W. K. Li (2009) Least absolute deviation estimation for unit root processes with GARCH errors. *Econometric Theory* 25, 1208–1227.

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

Li, G., Q. Zhu, Z. Liu, & W. K. Li (2017) On mixture double autoregressive time series models. *Journal of Business and Economic Statistics* 35, 306–317.

Li, W. K. (2004) *Diagnostic Checks in Time Series*. Chapman and Hall.

Ling, S. (2004) Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society, Series B* 66, 63–78.

Ling, S. (2005) Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society, Series B* 67, 381–393.

Ling, S. (2007) A double AR(p) model: Structure and estimation. *Statistica Sinica* 17, 161–175.

Ling, S. & D. Li (2008) Asymptotic inference for a nonstationary double AR(1) model. *Biometrika* 95, 257–263.

Ling, S. & M. McAleer (2003) Asymptotic theory for a vector ARMA–GARCH model. *Econometric Theory* 19, 280–310.

Ljung, G. M. & G. E. P. Box (1978) On a measure of lack of fit in time series models. *Biometrika* 65, 297–303.

Machado, J. A. (1993) Robust model selection and M-estimation. *Econometric Theory* 9, 478–493.

Phillips, P. C. B. (2015) Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics* 13, 521–555.

Pollard, D. (1985) New ways to prove central limit theorems. *Econometric Theory* 1, 295–314.

Tsay, R. S. (2014) *Multivariate Time Series Analysis: With R and Financial Applications*. Wiley.

Tweedie, R. L. (1983) Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In J. F. C. Kingman & G. E. H. Reuter (eds.), *Probability, Statistics and Analysis*, pp. 260–276. Cambridge University Press.

Weiss, A. (1986) Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* 2, 107–131.

Wu, G. & Z. Xiao (2002) An analysis of risk measures. *Journal of Risk* 4, 53–75.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press

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

Zheng, Y., Q. Zhu, G. Li, & Z. Xiao (2018) Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *Journal of the Royal Statistical Society, Series B* 80, 975–993.

Zhu, K. & S. Ling (2011) Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA–GARCH/IGARCH models. *Annals of Statistics* 39, 2131–2163.

Zhu, K. & S. Ling (2013) Quasi-maximum exponential likelihood estimators for a double AR(p) model. *Statistica Sinica* 23, 251–270.

Zhu, Q., G. Li, & Z. Xiao (2021) Quantile estimation of regression models with GARCH-X errors. *Statistica Sinica*, to appear. doi: 10.5705/ss.202019.0003.

Zhu, Q., Y. Zheng, & G. Li (2018) Linear double autoregression. *Journal of Econometrics* 207, 162–174.

Zhu, X., R. Pan, G. Li, Y. Liu, & H. Wang (2017) Network vector autoregression. *Annals of Statistics* 45, 1096–1123.

https://doi.org/10.1017/S026646662100030X Published online by Cambridge University Press