Modeling Autoregressive Conditional Skewness and Kurtosis with Multi-Quantile CAViaR

Halbert White[a], Tae-Hwan Kim[b,c], and Simone Manganelli[d]

[a]Department of Economics, University of California, San Diego
[b]School of Economics, Yonsei University
[c]School of Economics, University of Nottingham
[d]European Central Bank, DG-Research

January 2008

Abstract: Engle and Manganelli (2004) propose CAViaR, a class of models suitable for estimating conditional quantiles in dynamic settings. Engle and Manganelli apply their approach to the estimation of Value at Risk, but this is only one of many possible applications. Here we extend CAViaR models to permit joint modeling of multiple quantiles, Multi-Quantile (MQ) CAViaR. We apply our new methods to estimate measures of conditional skewness and kurtosis defined in terms of conditional quantiles, analogous to the unconditional quantile-based measures of skewness and kurtosis studied by Kim and White (2004). We investigate the performance of our methods by simulation, and we apply MQ-CAViaR to study conditional skewness and kurtosis of S&P 500 daily returns.

Keywords: Asset returns; CAViaR; Conditional quantiles; Dynamic quantiles; Kurtosis; Skewness.

JEL classifications: C13, C32.

1 Introduction

It is widely recognized that the use of higher moments, such as skewness and kurtosis, can be important for improving the performance of various financial models. Responding to this recognition, researchers and practitioners have started to incorporate these higher moments into their models, mostly using the conventional measures, e.g. the sample skewness and/or the sample kurtosis. Models of conditional counterparts of the sample skewness and the sample kurtosis, based on extensions of the GARCH model, have also been developed and used; see, for example, Leon, Rubio, and Serna (2004). Nevertheless, Kim and White (2004) point out that because standard measures of skewness and kurtosis are essentially based on averages, they can be sensitive to one or a few outliers – a regular feature of financial returns data – making their reliability doubtful.

To deal with this, Kim and White (2004) propose the use of more stable and robust measures of skewness and kurtosis, based on quantiles rather than averages. Nevertheless, Kim and White (2004) only discuss unconditional skewness and kurtosis measures. In this paper, we extend the approach of Kim and White (2004) by proposing *conditional* quantile-based skewness and kurtosis measures. For this, we extend Engle and Manganelli's (2004) univariate CAViaR model to a multi-quantile version, MQ-CAViaR. This allows for both a general vector autoregressive structure in the conditional quantiles and the presence of exogenous variables. We then use the MQ-CAViaR model to specify conditional versions of the more robust skewness and kurtosis measures discussed in Kim and White (2004).

The paper is organized as follows. In Section 2, we develop the MQ-CAViaR data generating process (DGP). In Section 3, we propose a quasi-maximum likelihood estimator for the MQ-CAViaR process and prove its consistency and asymptotic normality. In Section 4, we show how to consistently estimate the asymptotic variance-covariance matrix of the MQ-CAViaR estimator. Section 5 specifies conditional quantile-based measures of skewness and kurtosis based on MQ-CAViaR estimates. Section 6 contains an empirical application of our methods to the S&P 500 index. We also report results of a simulation experiment designed to examine the finite sample behavior of our estimator. Section 7 contains a summary and concluding remarks. Mathematical proofs are gathered into the Mathematical Appendix.

2 The MQ-CAViaR Process and Model

We consider data generated as a realization of the following stochastic process.

Assumption 1 The sequence $\{(Y_t, X_t') : t = 0, \pm 1, \pm 2, ...,\}$ is a stationary and ergodic stochastic process on the complete probability space $(\Omega, \mathcal{F}, P_0)$, where Y_t is a scalar and X_t is a countably dimensioned vector whose first element is one.

Let $\mathcal{F}_{t-1}$ be the σ-algebra generated by $Z^{t-1} \equiv \{X_t, (Y_{t-1}, X_{t-1}), ...\}$, i.e. $\mathcal{F}_{t-1} \equiv \sigma(Z^{t-1})$. We let $F_t(y) \equiv P_0[Y_t < y|\mathcal{F}_{t-1}]$ define the cumulative distribution function (CDF) of Y_t conditional on $\mathcal{F}_{t-1}$.

Let $0 < \theta_1 < ... < \theta_p < 1$. For $j = 1, ..., p$, the $\theta_j th$ quantile of Y_t conditional on $\mathcal{F}_{t-1}$, denoted $q^*_{j,t}$, is

$$q^*_{j,t} \equiv \inf\{y : F_t(y) = \theta_j\}, \tag{1}$$

and if F_t is strictly increasing,

$$q^*_{j,t} = F_t^{-1}(\theta_j).$$

Alternatively, $q^*_{j,t}$ can be represented as

$$\int_{-\infty}^{q^*_{j,t}} dF_t(y) = E[1_{[Y_t \leq q^*_{j,t}]}|\mathcal{F}_{t-1}] = \theta_j, \tag{2}$$

where $dF_t(y)$ is the Lebesgue-Stieltjes probability density function (PDF) of Y_t conditional on $\mathcal{F}_{t-1}$, corresponding to $F_t(y)$.

Our objective is to jointly estimate the conditional quantile functions $q^*_{j,t}$, $j = 1, 2, ..., p$. For this we write $q^*_t \equiv (q^*_{1,t}, ..., q^*_{p,t})'$ and impose additional appropriate structure.

First, we ensure that the conditional distribution of Y_t is everywhere continuous, with positive density at each conditional quantile of interest, $q^*_{j,t}$. We let f_t denote the conditional probability density function (PDF) corresponding to F_t. In stating our next condition (and where helpful elsewhere), we make explicit the dependence of the conditional CDF F_t on ω by writing $F_t(\omega, y)$ in place of $F_t(y)$. Realized values of the conditional quantiles are correspondingly denoted $q^*_{j,t}(\omega)$. Similarly, we write $f_t(\omega, y)$ in place of $f_t(y)$.

After ensuring this continuity, we impose specific structure on the quantiles of interest.

Assumption 2 (i) Y_t is continuously distributed such that for each t and each $\omega \in \Omega, F_t(\omega, \cdot)$ and $f_t(\omega, \cdot)$ are continuous on $\mathbb{R}$; (ii) For given $0 < \theta_1 < ... <$

$\theta_p < 1$ and $\{q^*_{j,t}\}$ as defined above, suppose: (a) For each t and $j = 1, ..., p$, $f_t(\omega, q^*_{j,t}(\omega)) > 0$; (b) For given finite integers k and m, there exist a stationary ergodic sequence of random $k \times 1$ vectors $\{\Psi_t\}$, with Ψ_t measurable$-\mathcal{F}_{t-1}$, and real vectors $\beta^*_j \equiv (\beta^*_{j,1}, ..., \beta^*_{j,k})'$ and $\gamma^*_{j\tau} \equiv (\gamma^*_{j\tau 1}, ..., \gamma^*_{j\tau p})'$ such that for all t and $j = 1, ..., p$,

$$q^*_{j,t} = \Psi'_t \beta^*_j + \sum_{\tau=1}^{m} q^{*\prime}_{t-\tau} \gamma^*_{j\tau} \ . \tag{3}$$

The structure of eq. (3) is a multi-quantile version of the CAViaR process introduced by Engle and Manganelli (2004). When $\gamma^*_{j\tau i} = 0$ for $i \neq j$, we have the standard CAViaR process. Thus, we call processes satisfying our structure "Multi-Quantile CAViaR" (MQ-CAViaR) processes. For MQ-CAViaR, the number of relevant lags can differ across the conditional quantiles; this is reflected in the possibility that for given j, elements of $\gamma^*_{j\tau}$ may be zero for values of τ greater than some given integer. For notational simplicity, we do not represent m as depending on j. Nevertheless, by convention, for no $\tau \leq m$ do we have $\gamma^*_{j\tau}$ equal to the zero vector for all j.

The finitely dimensioned random vectors Ψ_t may contain lagged values of Y_t, as well as measurable functions of X_t and lagged X_t and Y_t. In particular, Ψ_t may contain Stinchcombe and White's (1998) GCR transformations, as discussed in White (2006).

For a particular quantile, say θ_j, the coefficients to be estimated are β^*_j and $\gamma^*_j \equiv (\gamma^{*\prime}_{j1}, ..., \gamma^{*\prime}_{jm})'$. Let $\alpha^{*\prime}_j \equiv (\beta^{*\prime}_j, \gamma^{*\prime}_j)$, and write $\alpha^* = (\alpha^{*\prime}_1, ..., \alpha^{*\prime}_p)'$, an $\ell \times 1$ vector, where $\ell \equiv p(k + mp)$. We will call α^* the "MQ-CAViaR coefficient vector." We estimate α^* using a correctly specified model of the MQ-CAViaR process.

First, we specify our model.

Assumption 3 Let $\mathbb{A}$ be a compact subset of $\mathbb{R}^\ell$. (i) The sequence of functions $\{q_t : \Omega \times \mathbb{A} \to \mathbb{R}^p\}$ is such that for each t and each $\alpha \in \mathbb{A}$, $q_t(\cdot, \alpha)$ is measurable$-\mathcal{F}_{t-1}$; for each t and each $\omega \in \Omega$, $q_t(\omega, \cdot)$ is continuous on $\mathbb{A}$; and for each t and $j = 1, ..., p$,

$$q_{j,t}(\cdot, \alpha) = \Psi'_t \beta_j + \sum_{\tau=1}^{m} q_{t-\tau}(\cdot, \alpha)' \gamma_{j\tau}.$$

Next, we impose correct specification and an identification condition. Assumption 4(i.a) delivers correct specification by ensuring that the MQ-CAViaR coefficient vector α^* belongs to the parameter space, $\mathbb{A}$. This ensures that α^* optimizes the estimation objective function asymptotically. Assumption 4(i.b) delivers identification by ensuring that α^* is the only such optimizer. In stating the identification condition, we define $\delta_{j,t}(\alpha, \alpha^*) \equiv q_{j,t}(\cdot, \alpha) - q_{j,t}(\cdot, \alpha^*)$ and use the norm $||\alpha|| \equiv \max_{i=1,...,\ell} |\alpha_i|$.

Assumption 4 (i)(a) There exists $\alpha^* \in \mathbb{A}$ such that for all t

$$q_t(\cdot, \alpha^*) = q_t^*; \tag{4}$$

(b) There exists a non-empty set $J \subseteq \{1, ..., p\}$ such that for each $\epsilon > 0$ there exists $\delta_\epsilon > 0$ such that for all $\alpha \in \mathbb{A}$ with $||\alpha - \alpha^*|| > \epsilon$,

$$P[\cup_{j \in J}\{|\delta_{j,t}(\alpha, \alpha^*)| > \delta_\epsilon\}] > 0.$$

Among other things, this identification condition ensures that there is sufficient variation in the shape of the conditional distribution to support estimation of a sufficient number ($\#J$) of variation-free conditional quantiles. In particular, distributions that depend on a given finite number of parameters, say k, will generally be able to support k variation-free quantiles. For example, the quantiles of the $N(\mu, 1)$ distribution all depend on μ alone, so there is only one "degree of freedom" for the quantile variation. Similarly the quantiles of scaled and shifted $t-$distributions depend on three parameters (location, scale, and kurtosis), so there are only three "degrees of freedom" for the quantile variation.

3 MQ-CAViaR Estimation: Consistency and Asymptotic Normality

We estimate α^* by the method of quasi-maximum likelihood. Specifically, we construct a quasi-maximum likelihood estimator (QMLE) $\hat{\alpha}_T$ as the solution to the following optimization problem:

$$\min_{\alpha \in} \bar{S}_T(\alpha) \equiv T^{-1} \sum_{t=1}^{T} \{\sum_{j=1}^{p} \rho_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha))\}, \tag{5}$$

where $\rho_\theta(e) = e\psi_\theta(e)$ is the standard "check function," defined using the usual quantile step function, $\psi_\theta(e) = \theta - 1_{[e \leq 0]}$. We thus view

$$S_t(\alpha) \equiv -\{\sum_{j=1}^{p} \rho_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha))\}$$

as the quasi log-likelihood for observation t. In particular, $S_t(\alpha)$ is the log-likelihood of a vector of p independent asymmetric double exponential random variables (see White, 1994, ch. 5.3; Kim and White, 2003; Komunjer, 2005). Because $Y_t - q_{j,t}(\cdot, \alpha^*), j = 1, ..., p$ need not actually have this distribution, the method is *quasi* maximum likelihood.

We can establish the consistency of $\hat{\alpha}_T$ by applying results of White (1994). For this we impose the following moment and domination conditions. In stating this next condition and where convenient elsewhere, we exploit stationarity to omit explicit reference to all values of t.

Assumption 5 (i) $E|Y_t| < \infty$; (ii) let $D_{0,t} \equiv \max_{j=1,\ldots,p} \sup_{\alpha \in} \; |q_{j,t}(\cdot, \alpha)|$, $t = 1, 2, \ldots$. Then $E(D_{0,t}) < \infty$.

We now have conditions sufficient to establish the consistency of $\hat{\alpha}_T$.

Theorem 1 *Suppose that Assumptions* $1, 2(i, ii), 3(i), 4(i)$, *and* $5(i, ii)$ *hold. Then* $\hat{\alpha}_T \overset{a.s.}{\rightarrow} \alpha^*$.

Next, we establish the asymptotic normality of $T^{1/2}(\hat{\alpha}_T - \alpha^*)$. We use a method originally proposed by Huber (1967) and later extended by Weiss (1991). We first sketch the method before providing formal conditions and results.

Huber's method applies to our estimator $\hat{\alpha}_T$, provided that $\hat{\alpha}_T$ satisfies the asymptotic first order conditions

$$T^{-1} \sum_{t=1}^{T} \{ \sum_{j=1}^{p} \nabla q_{j,t}(\cdot, \hat{\alpha}_T) \; \psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \hat{\alpha}_T)) \} = o_p(T^{1/2}), \tag{6}$$

where $\nabla q_{j,t}(\cdot, \alpha)$ is the $\ell \times 1$ gradient vector with elements $(\partial / \partial \alpha_i) q_{j,t}(\cdot, \alpha), i = 1, ..., \ell$, and $\psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \hat{\alpha}_T))$ is a generalized residual. Our first task is thus to ensure that eq. (6) holds.

Next, we define

$$\lambda(\alpha) \equiv \sum_{j=1}^{p} E[\nabla q_{j,t}(\cdot, \alpha) \psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha))].$$

With λ continuously differentiable at α^* interior to $\mathbb{A}$, we can apply the mean value theorem to obtain

$$\lambda(\alpha) = \lambda(\alpha^*) + Q_0(\alpha - \alpha^*), \tag{7}$$

where Q_0 is an $\ell \times \ell$ matrix with $(1 \times \ell)$ rows $Q_{0,i} = \nabla' \lambda(\bar{\alpha}_{(i)})$, where $\bar{\alpha}_{(i)}$ is a mean value (different for each i) lying on the segment connecting α and $\alpha^*, i = 1, ..., \ell$.

It is straightforward to show that correct specification ensures that $\lambda(\alpha^*)$ is zero. We will also show that

$$Q_0 = -Q^* + O(||\alpha - \alpha^*||), \tag{8}$$

where $Q^* \equiv \sum_{j=1}^p E[f_{j,t}(0)\nabla q_{j,t}(\cdot, \alpha^*)\nabla' q_t(\cdot, \alpha^*)]$ with $f_{j,t}(0)$ the value at zero of the density $f_{j,t}$ of $\varepsilon_{j,t} \equiv Y_t - q_{j,t}(\cdot, \alpha^*)$, conditional on $\mathcal{F}_{t-1}$. Combining eqs. (7) and (8) and putting $\lambda(\alpha^*) = 0$, we obtain

$$\lambda(\alpha) = -Q^*(\alpha - \alpha^*) + O(||\alpha - \alpha^*||^2). \tag{9}$$

The next step is to show that

$$T^{1/2}\lambda(\hat{\alpha}_T) + H_T = o_p(1), \tag{10}$$

where $H_T \equiv T^{-1/2}\sum_{t=1}^T \eta_t^*$, with $\eta_t^* \equiv \sum_{j=1}^p \nabla q_{j,t}(\cdot, \alpha^*)\psi_{\theta_j}(\varepsilon_{j,t})$. Eqs. (9) and (10) then yield the following asymptotic representation of our estimator $\hat{\alpha}_T$:

$$T^{1/2}(\hat{\alpha}_T - \alpha^*) = Q^{*-1}T^{-1/2}\sum_{t=1}^T \eta_t^* + o_p(1). \tag{11}$$

As we impose conditions sufficient to ensure that $\{\eta_t^*, \mathcal{F}_t\}$ is a martingale difference sequence (MDS), a suitable central limit theorem (e.g., theorem 5.24 in White, 2001) applies to (11) to yield the desired asymptotic normality of $\hat{\alpha}_T$:

$$T^{1/2}(\hat{\alpha}_T - \alpha^*) \xrightarrow{d} N(0, Q^{*-1}V^*Q^{*-1}), \tag{12}$$

where $V^* \equiv E(\eta_t^*\eta_t^{*\prime})$.

We now strengthen the conditions above to ensure that each step of the above argument is valid.

Assumption 2 (iii) (a) There exists a finite positive constant f_0 such that for each t, each $\omega \in \Omega$, and each $y \in \mathbb{R}$, $f_t(\omega, y) \leq f_0 < \infty$; (b) There exists a finite positive constant L_0 such that for each t, each $\omega \in \Omega$, and each $y_1, y_2 \in \mathbb{R}$, $|f_t(\omega, y_1) - f_t(\omega, y_2)| \leq L_0|y_1 - y_2|$.

Next we impose sufficient differentiability of q_t with respect to α.

Assumption 3 (ii) For each t and each $\omega \in \Omega$, $q_t(\omega, \cdot)$ is continuously differentiable on $\mathbb{A}$; (iii) For each t and each $\omega \in \Omega$, $q_t(\omega, \cdot)$ is twice continuously differentiable on $\mathbb{A}$;

To exploit the mean value theorem, we require that α^* belongs to the interior of $\mathbb{A}$, $int(\mathbb{A})$.

Assumption 4 (ii) $\alpha^* \in int(\mathbb{A})$.

Next, we place domination conditions on the derivatives of q_t.

Assumption 5 (iii) Let $D_{1,t} \equiv \max_{j=1,\ldots,p} \max_{i=1,\ldots,\ell} \sup_{\alpha\in} \ |(\partial/\partial\alpha_i)q_{j,t}(\cdot,\alpha)|, t = 1,2,\ldots$. Then (a) $E(D_{1,t}) < \infty$; (b) $E(D_{1,t}^2) < \infty$; (iv) Let $D_{2,t} \equiv \max_{j=1,\ldots,p} \max_{i=1,\ldots,\ell} \max_{h=1,\ldots,\ell} \sup_{\alpha\in} \ |(\partial^2/\partial\alpha_i\partial\alpha_h)q_{j,t}(\cdot,\alpha)|, t = 1,2,\ldots$. Then (a) $E(D_{2,t}) < \infty$; (b) $E(D_{2,t}^2) < \infty$.

Assumption 6 (i) $Q^* \equiv \sum_{j=1}^{p} E[f_{j,t}(0)\nabla q_{j,t}(\cdot,\alpha^*)\nabla' q_{j,t}(\cdot,\alpha^*)]$ is positive definite; (ii) $V^* \equiv E(\eta_t^*\eta_t^{*\prime})$ is positive definite.

Assumptions 3(ii) and 5(iii.a) are additional assumptions helping to ensure that eq. (6) holds. Further imposing Assumptions 2(iii), 3(iii.a), 4(ii), and 5(iv.a) suffices to ensure that eq. (9) holds. The additional regularity provided by Assumptions 5(iii.b), 5(iv.b), and 6(i) ensures that eq. (10) holds. Assumptions 5(iii.b) and 6(ii) help ensure the availability of the MDS central limit theorem.

We now have conditions sufficient to ensure asymptotic normality of our MQ-CAViaR estimator. Formally, we have

Theorem 2 *Suppose that Assumptions 1-6 hold. Then*

$$V^{*-1/2}Q^*T^{1/2}(\hat{\alpha}_T - \alpha^*) \xrightarrow{d} N(0,I).$$

Theorem 2 shows that our QML estimator $\hat{\alpha}_T$ is asymptotically normal with asymptotic covariance matrix $Q^{*-1}V^*Q^{*-1}$. There is, however, no guarantee that $\hat{\alpha}_T$ is asymptotically efficient. There is now a considerable literature investigating efficient estimation in quantile models; see, for example, Newey and Powell (1990), Otsu (2003), Komunjer and Vuong (2006, 2007a, 2007b). So far, this literature has only considered single quantile models. It is not obvious how the results for single quantile models extend to multi-quantile models such as ours. Nevertheless, Komunjer and Vuong (2007a) show that the class of QML estimators is not large enough to include an efficient estimator, and that the class of M-estimators, which strictly includes the QMLE class, yields an estimator that attains the efficiency bound. Specifically, they show that replacing the usual quantile check function ρ_{θ_j} appearing in eq.(5) with

$$\rho_{\theta_j}^*(Y_t - q_{j,t}(\cdot,\alpha)) = (\theta - 1_{[Y_t - q_{j,t}(\cdot,\alpha)\leq 0]})(F_t(Y_t) - F_t(q_{j,t}(\cdot,\alpha)))$$

will deliver an asymptotically efficient quantile estimator under the single quantile restriction. We conjecture that replacing ρ_{θ_j} with $\rho^*_{\theta_j}$ in eq.(5) will improve estimator efficiency. We leave the study of the asymptotically efficient multi-quantile estimator for future work.

4 Consistent Covariance Matrix Estimation

To test restrictions on α^* or to obtain confidence intervals, we require a consistent estimator of the asymptotic covariance matrix $C^* \equiv Q^{*-1}V^*Q^{*-1}$. First, we provide a consistent estimator $\hat{V}_T$ for V^*; then we give a consistent estimator $\hat{Q}_T$ for Q^*. It follows that $\hat{C}_T \equiv \hat{Q}_T^{-1}\hat{V}_T\hat{Q}_T^{-1}$ is a consistent estimator for C^*.

Recall that $V^* \equiv E(\eta_t^*\eta_t^{*\prime})$, with $\eta_t^* \equiv \sum_{j=1}^p \nabla q_{j,t}(\cdot,\alpha^*)\psi_{\theta_j}(\varepsilon_{j,t})$. A straightforward plug-in estimator of V^* is

$$\begin{aligned}
\hat{V}_T &\equiv T^{-1}\sum_{t=1}^T \hat{\eta}_t\hat{\eta}_t', \qquad \text{with} \\
\hat{\eta}_t &\equiv \sum_{j=1}^p \nabla q_{j,t}(\cdot,\hat{\alpha}_T)\psi_{\theta_j}(\hat{\varepsilon}_{j,t}) \\
\hat{\varepsilon}_{j,t} &\equiv Y_t - q_{j,t}(\cdot,\hat{\alpha}_T).
\end{aligned}$$

We already have conditions sufficient to deliver the consistency of $\hat{V}_T$ for V^*. Formally, we have

Theorem 3 *Suppose that Assumptions 1-6 hold. Then* $\hat{V}_T \xrightarrow{p} V^*$.

Next, we provide a consistent estimator of

$$Q^* \equiv \sum_{j=1}^p E[f_{j,t}(0)\nabla q_{j,t}(\cdot,\alpha^*)\nabla' q_{j,t}(\cdot,\alpha^*)].$$

We follow Powell's (1984) suggestion of estimating $f_{j,t}(0)$ with $1_{[-\hat{c}_T \le \hat{\varepsilon}_{j,t} \le \hat{c}_T]}/2\hat{c}_T$ for a suitably chosen sequence $\{\hat{c}_T\}$. This is also the approach taken in Kim and White (2003) and Engle and Manganelli (2004). Accordingly, our proposed estimator is

$$\hat{Q}_T = (2\hat{c}_T T)^{-1}\sum_{t=1}^T\sum_{j=1}^p 1_{[-\hat{c}_T \le \hat{\varepsilon}_{j,t} \le \hat{c}_T]}\nabla q_{j,t}(\cdot,\hat{\alpha}_T)\nabla' q_{j,t}(\cdot,\hat{\alpha}_T).$$

To establish consistency, we strengthen the domination condition on $\nabla q_{j,t}$ and impose conditions on $\{\hat{c}_T\}$.

Assumption 5 (iii.c) $E(D_{1,t}^3) < \infty$.

Assumption 7 $\{\hat{c}_T\}$ is a stochastic sequence and $\{c_T\}$ is a non-stochastic sequence such that (i) $\hat{c}_T/c_T \xrightarrow{p} 1$; (ii) $c_T = o(1)$; and (iii) $c_T^{-1} = o(T^{1/2})$.

Theorem 4 *Suppose that Assumptions 1-7 hold. Then* $\hat{Q}_T \xrightarrow{p} Q^*$.

5 Quantile-Based Measures of Conditional Skewness and Kurtosis

Moments of asset returns of order higher than two are important because these permit a recognition of the multi-dimensional nature of the concept of risk. Such higher order moments have thus proved useful for asset pricing, portfolio construction, and risk assessment. See, for example, Hwang and Satchell (1999) and Harvey and Siddique (2000). Higher order moments that have received particular attention are skewness and kurtosis, which involve moments of order three and four, respectively. Indeed, it is widely held as a "stylized fact" that the distribution of stock returns exhibits both left skewness and excess kurtosis (fat tails); there is a large amount of empirical evidence to this effect.

Recently, Kim and White (2004) have challenged this stylized fact and the conventional way of measuring skewness and kurtosis. As moments, skewness and kurtosis are computed using averages, specifically, averages of third and fourth powers of standardized random variables. Kim and White (2004) point out that averages are sensitive to outliers, and that taking third or fourth powers greatly enhances the influence of any outliers that may be present. Moreover, asset returns are particularly prone to containing outliers, as the result of crashes or rallies. According to Kim and White's simulation study, even a single outlier of a size comparable to the sharp drop in stock returns caused by the 1987 stock market crash can generate dramatic irregularities in the behavior of the traditional moment-based measures of skewness and kurtosis.

Kim and White (2004) propose using more robust measures instead, based on sample quantiles. For example, Bowley's (1920) coefficient of skewness is given by

$$SK_2 = \frac{q_3^* + q_1^* - 2q_2^*}{q_3^* - q_1^*},$$

where $q_1^* = F^{-1}(0.25), q_2^* = F^{-1}(0.5)$, and $q_3^* = F^{-1}(0.75)$, where $F(y) \equiv P_0[Y_t < y]$ is the unconditional CDF of Y_t. Similarly, Crow & Siddiqui's (1967) coefficient of kurtosis is given by

$$KR_4 = \frac{q_4^* - q_0^*}{q_3^* - q_1^*} - 2.91,$$

where $q_0^* = F^{-1}(0.025)$ and $q_4^* = F^{-1}(0.975)$. (The notations SK_2 and KR_4 correspond to those of Kim and White (2004).)

A limitation of these measures is that they are based on *unconditional* sample quantiles. Thus, in measuring skewness or kurtosis, these can neither incorporate useful information contained in relevant exogenous variables nor exploit the dynamic evolution of quantiles over time. To avoid these limitations, we propose constructing measures of conditional skewness and kurtosis using *conditional* quantiles $q_{j,t}^*$ in place of the unconditional quantiles q_j^*. In particular, the conditional Bowley coefficient of skewness and the conditional Crow & Siddiqui coefficient of kurtosis are given by

$$\begin{aligned} CSK_2 &= \frac{q_{3,t}^* + q_{1,t}^* - 2q_{2,t}^*}{q_{3,t}^* - q_{1,t}^*}, \\ CKR_4 &= \frac{q_{4,t}^* - q_{0,t}^*}{q_{3,t}^* - q_{1,t}^*} - 2.91. \end{aligned}$$

Another quantile-based kurtosis measure discussed in Kim and White (2004) is Moors's (1988) coefficient of kurtosis, which involves computing six quantiles. Because our approach requires joint estimation of all relevant quantiles, and, in our model, each quantile depends not only on its own lags, but also possibly on the lags of other quantiles, the number of parameters to be estimated can be quite large. Moreover, if the θ_j's are too close to each other, then the corresponding quantiles may be highly correlated, which can result in an analog of multicollinearity. For these reasons, in what follows we focus only on SK_2 and KR_4, as these require jointly estimating at most five quantiles.

6 Application and Simulation

6.1 Time-varying skewness and kurtosis for the S&P500

In this section we obtain estimates of time-varying skewness and kurtosis for the S&P 500 index daily returns. Figure 1 plots the S&P 500 daily returns series used for estimation. The sample ranges from January 1, 1999 to September 28, 2007, for a total of 2,280 observations.

First, we estimate time-varying skewness and kurtosis using the GARCH-type model of Leon, Rubio and Serna (2004), the LRS model for short. Letting r_t

denote the return for day t, we estimate the following specification of their model:

$$\begin{aligned}
r_t &= h_t^{1/2}\eta_t \\
h_t &= \beta_1 + \beta_2 r_{t-1}^2 + \beta_3 h_{t-1} \\
s_t &= \beta_4 + \beta_5 \eta_{t-1}^3 + \beta_6 s_{t-1} \\
k_t &= \beta_7 + \beta_8 \eta_{t-1}^4 + \beta_9 k_{t-1},
\end{aligned}$$

where we assume that $E_{t-1}(\eta_t) = 0$, $E_{t-1}(\eta_t^2) = 1$, $E_{t-1}(\eta_t^3) = s_t$, and $E_{t-1}(\eta_t^4) = k_t$, where E_{t-1} denotes the conditional expectation given $r_{t-1}, r_{t-2}, ...$ The likelihood is constructed using a Gram-Charlier series expansion of the normal density function for η_t, truncated at the fourth moment. We refer the interested reader to Leon, Rubio, and Serna (2004) for technical details.

The model is estimated via (quasi-) maximum likelihood. As starting values for the optimization, we use estimates of β_1, β_2, and β_3 from the standard GARCH model. We set initial values of β_4 and β_7 equal to the unconditional skewness and kurtosis values of the GARCH residuals. The remaining coefficients are initialized at zero. The point estimates for the model parameters are given in Table 1. Figures 3 and 5 display the time-series plots for s_t and k_t respectively.

Next, we estimate the MQ-CAViaR model. Given the expressions for CSK_2 and CKR_4, we require five quantiles, i.e. those for $\theta_j = 0.025, 0.25, 0.5, 0.75$, and 0.975. We thus estimate an MQ-CAViaR model for the following DGP:

$$\begin{aligned}
q_{0.025,t}^* &= \beta_{11}^* + \beta_{12}^* |r_{t-1}| + q_{t-1}^{*\prime}\gamma_1^* \\
q_{0.25,t}^* &= \beta_{21}^* + \beta_{22}^* |r_{t-1}| + q_{t-1}^{*\prime}\gamma_2^* \\
&\vdots \\
q_{0.975,t}^* &= \beta_{51}^* + \beta_{52}^* |r_{t-1}| + q_{t-1}^{*\prime}\gamma_5^*,
\end{aligned}$$

where $q_{t-1}^* \equiv (q_{0.025,t-1}^*, q_{0.25,t-1}^*, q_{0.5,t-1}^*, q_{0.75,t-1}^*, q_{0.975,t-1}^*)'$ and $\gamma_j^* \equiv (\gamma_{j1}^*, \gamma_{j2}^*, \gamma_{j3}^*, \gamma_{j4}^*, \gamma_{j5}^*)'$, $j = 1, ..., 5$. Hence, the coefficient vector α^* consists of all the coefficients β_{jk}^* and γ_{jk}^*, as above.

Estimating the full model is not trivial. We discuss this briefly before presenting the estimation results. We perform the computations in a step-wise fashion as follows. In the first step, we estimate the MQ-CAViaR model containing just the 2.5% and 25% quantiles. The starting values for optimization are the individual CAViaR estimates, and we initialize the remaining parameters at zero. We repeat this estimation procedure for the MQ-CAViaR model containing the 75% and 97.5% quantiles. In the second step, we use the estimated parameters of the first step as starting values for the optimization of the MQ-CAViaR model containing the 2.5%, 25%, 75%, and 97.5% quantiles, initializing the remaining parameters at zero. Third and finally, we use the estimates from the second step as starting values for the full MQ-CAViaR model optimization containing all five quantiles of interest, again setting to zero the remaining parameters.

The likelihood function appears quite flat around the optimum, making the optimization procedure sensitive to the choice of initial conditions. In particular, choosing a different combination of quantile couples in the first step of our estimation procedure tends to produce different parameter estimates for the full MQ-CAViaR model. Nevertheless, the likelihood values are similar, and there are no substantial differences in the dynamic behavior of the individual quantiles associated with these different estimates.

Table 2 presents our MQ-CAViaR estimation results. In calculating the standard errors, we have set the bandwidth to 1. Results are slightly sensitive to the choice of the bandwidth, with standard errors increasing for lower values of the bandwidth. We observe that there is interaction across quantile processes. This is particularly evident for the 75% quantile: the autoregressive coefficient associated with the lagged 75% quantile is only 0.04, while that associated with the lagged 97.5% quantile is 0.29. This implies that the autoregressive process of the 75% quantile is mostly driven by the lagged 97.5% quantile, although this is not statistically significant at the usual significance level. Figure 2 displays plots of the five individual quantiles for the time period under consideration.

Next, we use the estimates of the individual quantiles $q^*_{0.025,t}, ..., q^*_{0.975,t}$ to calculate the robust skewness and kurtosis measures CSK_2 and CKR_4. The resulting time-series plots are shown in Figures 4 and 6, respectively.

We observe that the LRS model estimates of both skewness and kurtosis do not vary much and are dwarfed by those for the end of February 2007. The market was doing well until February 27, when the S&P 500 index dropped by 3.5%, as the market worried about global economic growth. (The sub-prime mortgage fiasco was still not yet public knowledge.) Interestingly, this is not a particularly large negative return (there are larger negative returns in our sample between 2000 and 2001), but this one occurred in a period of relatively low volatility.

Our more robust MQ-CAViaR measures show more plausible variability and confirm that the February 2007 market correction was indeed a case of large negative conditional skewness and high conditional kurtosis. This episode appears to be substantially affecting the LRS model estimates for the entire sample, raising doubts about the reliability of LRS estimates in general, consistent with the findings of Sakata and White (1998).

6.2 Simulation

In this section we provide some Monte Carlo evidence illustrating the finite sample behavior of our methods. We consider the same MQ-CAViaR process estimated in the previous subsection,

$$\begin{aligned} q^*_{0.025,t} &= \beta^*_{11} + \beta^*_{12}|r_{t-1}| + q^{*\prime}_{t-1}\gamma^*_1 \\ q^*_{0.25,t} &= \beta^*_{21} + \beta^*_{22}|r_{t-1}| + q^{*\prime}_{t-1}\gamma^*_2 \\ &\vdots \\ q^*_{0.975,t} &= \beta^*_{51} + \beta^*_{52}|r_{t-1}| + q^{*\prime}_{t-1}\gamma^*_5. \end{aligned} \tag{13}$$

For the simulation exercise, we set the true coefficients equal to the estimates reported in Table 2. Using these values, we generate the above MQ-CAViaR process 100 times, and each time we estimate all the coefficients, using the procedure described in the previous subsection.

Data were generated as follows. We initialize the quantiles $q^*_{\theta_j,t}, j = 1, ..., 5$ at $t = 1$ using the empirical quantiles of the first 100 observations of our S&P 500 data. Given quantiles for time t, we generate a random variable r_t compatible with these using the following procedure. First, we draw a random variable U_t, uniform over the interval $[0, 1]$. Next, we find θ_j such that $\theta_{j-1} < U_t < \theta_j$. This determines the quantile range within which the random variable to be generated should fall. Finally, we generated the desired random variable r_t by drawing it from a uniform distribution within the interval $[q^*_{\theta_{j-1},t}, q^*_{\theta_j,t}]$. The procedure can be represented as follows:

$$r_t = \sum_{j=1}^{p+1} I(\theta_{j-1} < U_t < \theta_j)[q^*_{\theta_{j-1},t} + (q^*_{\theta_j,t} - q^*_{\theta_{j-1},t})V_t],$$

where U_t and V_t are i.i.d. U(0,1), $\theta_0 = 0$, $\theta_{p+1} = 1$, $q^*_{\theta_0,t} = q^*_{\theta_1,t} - 0.05$ and $q^*_{\theta_{p+1},t} = q^*_{\theta_p,t} + 0.05$. It is easy to check that the random variable r_t has the desired quantiles by construction. Further, it doesn't matter that the distribution within the quantiles is uniform, as that distribution has essentially no impact on the resulting parameter estimates. Using these values of r_t and q^*_t, we apply eq.(13) to generate conditional quantiles for the next period. The process iterates until $t = T$. Once we have a full sample, we perform the estimation procedure described in the previous subsection.

Tables 3 and 4 provide the sample means and standard deviations over 100 replications of each coefficient estimate for two different sample sizes, $T = 1,000$ and $T = 2,280$ (the sample size of the S&P 500 data), respectively. The mean estimates are fairly close to the values of Table 2, showing that the available sample sizes are sufficient to recover the true DGP parameters. (To obtain standard error estimates for the means, divide the reported standard deviations by 10.)

A potentially interesting experiment that one might consider is to generate data from the LRS process and see how the MQ-CAViaR model performs in revealing underlying patterns of conditional skewness and kurtosis. Nevertheless, we leave this aside here, as the LRS model depends on four distributional shape parameters, but we require five variation-free quantiles for the present exercise. As noted in Section 2, the MQ-CAViaR model will generally not satisfy the identification condition in such circumstances.

7 Conclusion

In this paper, we generalize Engle and Manganelli's (2004) single-quantile CAViaR process to its multi-quantile version. This allows for (i) joint modeling of multiple quantiles; (ii) dynamic interactions between quantiles; and (iii) the use of exogenous variables. We apply our MQ-CAViaR process to define conditional versions of existing unconditional quantile-based measures of skewness and kurtosis. Because of their use of quantiles, these measures may be much less sensitive than standard moment-based methods to the adverse impact of outliers that regularly appear in financial market data. An empirical analysis of the S&P 500 index demonstrates the use and utility of our new methods.

References

[1] Andrews, D.W.K. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* 4, 458-467.

[2] Bartle, R. (1966). *The Elements of Integration*. New York: Wiley.

[3] Bowley, A.L. (1920). *Elements of Statistics.* New York: Charles Scribner's Sons.

[4] Crow, E.L. and Siddiqui, M.M. (1967), Robust estimation of location. *Journal of the American Statistical Association* 62, 353-389.

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

[6] Harvey, C.R. and Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance* LV, 1263-1295.

[7] Huber, P.J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221–233.

[8] Hwang, S. and Satchell, S.E. (1999). Modelling emerging market risk premia using higher moments. *International Journal of Finance and Economics* 4, 271-296.

[9] Kim, T.-H. and White, H. (2003). Estimation, inference, and specification testing for possibly misspecified quantile regression. In T. Fomby and C. Hill, eds., *Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later*. New York: Elsevier, pp. 107-132.

[10] Kim, T.-H. and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1, 56-73.

[11] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* 46, 33–50.

[12] Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics* 128, 127-164.

[13] Komunjer, I. and Vuong, Q. (2006). Efficient conditional quantile estimation: the time series case. University of California, San Diego Department of Economics Discussion Paper 2006-10, .

[14] Komunjer, I. and Vuong, Q. (2007a). Semiparametric efficiency bound and M-estimation in time-series models for conditional quantiles. University of California, San Diego Department of Economics Discussion Paper

[15] Komunjer, I. and Vuong, Q. (2007b). Efficient estimation in dynamic conditional quantile models. University of California, San Diego Department of Economics Discussion Paper.

[16] Leon, A., G. Rubio and G. Serna (2004), Autoregressive Conditional Volatility, Skewness and Kurtosis, WP-AD 2004-13, Instituto Valenciano de Investigaciones Economicas.

[17] Moors, J.J.A. (1988), A quantile alternative for kurtosis. *The Statistician* 37, 25-32.

[18] Otsu, T. (2003). Empirical likelihood for quantile regression. University of Wisconsin, Madison Department of Economics Discussion Paper.

[19] Powell, J. (1984). Least absolute deviations estimators for the censored regression model. *Journal of Econometrics* 25, 303-325.

[20] Newey, W.K. and Powell, J.L. (1990). Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory* 6, 295-317.

[21] Sakata, S. and White, H. (1998). High breakdown point conditional dispersion estimation with application to S&P 500 daily returns volatility. *Econometrica* 66, 529-568.

[22] Stinchcombe, M. and White, H. (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory* 14, 295-324.

[23] Weiss, A. (1991). Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory* 7, 46-68.

[24] White, H. (1994). *Estimation, Inference and Specification Analysis.* New York: Cambridge University Press.

[25] White, H. (2001). *Asymptotic Theory for Econometricians.* San Diego: Academic Press.

[26] White, H. (2006). Approximate nonlinear forecasting methods. In G. Elliott, C.W.J. Granger, and A. Timmermann, eds., *Handbook of Economics Forecasting.* New York: Elsevier, pp. 460-512

Mathematical Appendix

Proof of Theorem 1: We verify the conditions of corollary 5.11 of White (1994), which delivers $\hat{\alpha}_T \to \alpha^*$, where

$$\hat{\alpha}_T \equiv \arg\max_{\alpha \in} T^{-1} \sum_{t=1}^{T} \varphi_t(Y_t, q_t(\cdot, \alpha)),$$

and $\varphi_t(Y_t, q_t(\cdot, \alpha)) \equiv -\sum_{j=1}^{p} \rho_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha))$. Assumption 1 ensures White's Assumption 2.1. Assumption 3(i) ensures White's Assumption 5.1. Our choice of ρ_{θ_j} satisfies White's Assumption 5.4. To verify White's Assumption 3.1, it suffices that $\varphi_t(Y_t, q_t(\cdot, \alpha))$ is dominated on $\mathbb{A}$ by an integrable function (ensuring White's Assumption 3.1(a,b)) and that for each α in $\mathbb{A}$, $\{\varphi_t(Y_t, q_t(\cdot, \alpha))\}$ is stationary and ergodic (ensuring White's Assumption 3.1(c), the strong uniform law of large numbers (ULLN)). Stationarity and ergodicity is ensured by Assumptions 1 and 3(i). To show domination, we write

$$\begin{aligned}
|\varphi_t(Y_t, q_t(\cdot, \alpha))| &\leq \sum_{j=1}^{p} |\rho_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha))| \\
&= \sum_{j=1}^{p} |(Y_t - q_{j,t}(\cdot, \alpha))(\theta_j - 1_{[Y_t - q_{j,t}(\cdot,\alpha) \leq 0]})| \\
&\leq 2\sum_{j=1}^{p} |Y_t| + |q_{j,t}(\cdot, \alpha)| \\
&\leq 2p(|Y_t| + |D_{0,t}|),
\end{aligned}$$

so that

$$\sup_{\alpha \in} |\varphi_t(Y_t, q_t(\cdot, \alpha))| \leq 2p(|Y_t| + |D_{0,t}|).$$

Thus, $2p(|Y_t| + |D_{0,t}|)$ dominates $|\varphi_t(Y_t, q_t(\cdot, \alpha))|$ and has finite expectation by Assumption 5(i,ii).

It remains to verify White's Assumption 3.2; here this is the condition that α^* is the unique maximizer of $E(\varphi_t(Y_t, q_t(\cdot, \alpha))$. Given Assumptions 2(ii.b) and 4(i), it follows by argument directly parallel to that in the proof of White (1994, corollary 5.11) that for all $\alpha \in \mathbb{A}$,

$$E(\varphi_t(Y_t, q_t(\cdot, \alpha)) \leq E(\varphi_t(Y_t, q_t(\cdot, \alpha^*)).$$

Thus, it suffices to show that the above inequality is strict for $\alpha \neq \alpha^*$. Letting $\Delta(\alpha) \equiv \sum_{j=1}^{p} E(\Delta_{j,t}(\alpha))$ with $\Delta_{j,t}(\alpha) \equiv \rho_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha)) - \rho_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha^*))$, it suffices to show that for each $\epsilon > 0$, $\Delta(\alpha) > 0$ for all $\alpha \in \mathbb{A}$ such that $||\alpha - \alpha^*|| > \epsilon$.

Pick $\epsilon > 0$ and $\alpha \in \mathbb{A}$ such that $||\alpha - \alpha^*|| > \epsilon$. With $\delta_{j,t}(\alpha, \alpha^*) \equiv q_t(\theta_j, \alpha) - q_t(\theta_j, \alpha^*)$, by Assumption 4(i.b), there exist $J \subseteq \{1, ..., p\}$ and $\delta_\epsilon > 0$ such that $P[\cup_{j \in J}\{|\delta_{j,t}(\alpha, \alpha^*)| > \delta_\epsilon\}] > 0$. For this δ_ϵ and all j, some algebra and Assumption

2(ii.a) ensure that

$$\begin{aligned}
E(\Delta_{j,t}(\alpha)) &= E[\int_0^{\delta_{j,t}(\alpha,\alpha^*)} (\delta_{j,t}(\alpha,\alpha^*) - s)\ f_{j,t}(s)ds] \\
&\geq E[\frac{1}{2}\delta_\epsilon^2 1_{[|\delta_{j,t}(\alpha,\alpha^*)|>\delta_\epsilon]} + \frac{1}{2}\delta_{j,t}(\alpha,\alpha^*)^2 1_{[|\delta_{j,t}(\alpha,\alpha^*)|\leq\delta_\epsilon]})] \\
&\geq \frac{1}{2}\delta_\epsilon^2 E[1_{[|\delta_{j,t}(\alpha,\alpha^*)|>\delta_\epsilon]}].
\end{aligned}$$

The first inequality above comes from the fact that Assumption 2(ii.a) implies that for any $\delta_\epsilon > 0$ sufficiently small, we have $f_{j,t}(s) > \delta_\epsilon$ for $|s| < \delta_\epsilon$. Thus,

$$\begin{aligned}
\Delta(\alpha) &\equiv \sum_{j=1}^{p} E(\Delta_{j,t}(\alpha)) \geq \frac{1}{2}\delta_\epsilon^2 \sum_{j=1}^{p} E[1_{[|\delta_{j,t}(\alpha,\alpha^*)|>\delta_\epsilon]}] \\
&= \frac{1}{2}\delta_\epsilon^2 \sum_{j=1}^{p} P[|\delta_{j,t}(\alpha,\alpha^*)| > \delta_\epsilon] \geq \frac{1}{2}\delta_\epsilon^2 \sum_{j\in J} P[|\delta_{j,t}(\alpha,\alpha^*)| > \delta_\epsilon] \\
&\geq \frac{1}{2}\delta_\epsilon^2 P[\cup_{j\in J}\{|\delta_{j,t}(\alpha,\alpha^*)| > \delta_\epsilon\}] \\
&> 0,
\end{aligned}$$

where the final inequality follows from Assumption 4(i.b). As $\epsilon > 0$ and α are arbitrary, the result follows.

Proof of Theorem 2: As outlined in the text, we first prove

$$T^{-1/2}\sum_{t=1}^{T}\sum_{j=1}^{p} \nabla q_{j,t}(\cdot,\hat{\alpha}_T)\ \psi_{\theta_j}(Y_t - q_{j,t}(\cdot,\hat{\alpha}_T)) = o_p(1). \tag{14}$$

The existence of $\nabla q_{j,t}$ is ensured by Assumption 3(ii). Let e_i be the $\ell \times 1$ unit vector with i^{th} element equal to one and the rest zero, and let

$$G_i(c) \equiv T^{-1/2}\sum_{t=1}^{T}\sum_{j=1}^{p} \rho_{\theta_j}(Y_t - q_{j,t}(\cdot,\hat{\alpha}_T + ce_i)),$$

for any real number c. Then by the definition of $\hat{\alpha}_T$, $G_i(c)$ is minimized at $c = 0$. Let $H_i(c)$ be the derivative of $G_i(c)$ with respect to c from the right. Then

$$H_i(c) = -T^{-1/2}\sum_{t=1}^{T}\sum_{j=1}^{p} \nabla_i q_{j,t}(\cdot,\hat{\alpha}_T + ce_i)\ \psi_{\theta_j}(Y_t - q_{j,t}(\cdot,\hat{\alpha}_T + ce_i)),$$

where $\nabla_i q_{j,t}(\cdot,\hat{\alpha}_T + ce_i)$ is the i^{th} element of $\nabla q_{j,t}(\cdot,\hat{\alpha}_T + ce_i)$. Using the facts that (i) $H_i(c)$ is non-decreasing in c and (ii) for any $\epsilon > 0$, $H_i(-\epsilon) \leq 0$ and $H_i(\epsilon) \geq 0$, we have

$$
\begin{aligned}
|H_i(0)| &\leq H_i(\epsilon) - H_i(-\epsilon) \\
&\leq T^{-1/2} \sum_{t=1}^{T} \sum_{j=1}^{p} |\nabla_i q_{j,t}(\cdot, \hat{\alpha}_T)| 1_{[Y_t - q_{j,t}(\cdot,\hat{\alpha}_T)=0]} \\
&\leq T^{-1/2} \max_{1 \leq t \leq T} D_{1,t} \sum_{t=1}^{T} \sum_{j=1}^{p} 1_{[Y_t - q_{j,t}(\cdot,\hat{\alpha}_T)=0]},
\end{aligned}
$$

where the last inequality follows by the domination condition imposed in Assumption 5(iii.a). Because $D_{1,t}$ is stationary, $T^{-1/2} \max_{1 \leq t \leq T} D_{1,t} = o_p(1)$. The second term is bounded in probability: $\sum_{t=1}^{T} \sum_{j=1}^{p} 1_{[Y_t - q_{j,t}(\cdot,\hat{\alpha}_T)=0]} = O_p(1)$ given Assumption 2(i,ii.a) (see Koenker and Bassett, 1978, for details). Since $H_i(0)$ is the i^{th} element of $T^{-1/2} \sum_{t=1}^{T} \sum_{j=1}^{p} \nabla q_{j,t}(\cdot, \hat{\alpha}_T)\ \psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \hat{\alpha}_T))$, the claim in (14) is proved.

Next, for each $\alpha \in \mathbb{A}$, Assumptions 3(ii) and 5(iii.a) ensure the existence and finiteness of the $\ell \times 1$ vector

$$
\begin{aligned}
\lambda(\alpha) &\equiv \sum_{j=1}^{p} E[\nabla q_{j,t}(\cdot, \alpha)\ \psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha))] \\
&= \sum_{j=1}^{p} E[\nabla q_{j,t}(\cdot, \alpha) \int_{\delta_{j,t}(\alpha,\alpha^*)}^{0} f_{j,t}(s) ds],
\end{aligned}
$$

where $\delta_{j,t}(\alpha, \alpha^*) \equiv q_{j,t}(\cdot, \alpha) - q_{j,t}(\cdot, \alpha^*)$ and $f_{j,t}(s) = (d/ds) F_t(s + q_{j,t}(\cdot, \alpha^*))$ represents the conditional density of $\varepsilon_{j,t} \equiv Y_t - q_{j,t}(\cdot, \alpha^*)$ with respect to Lebesgue measure. The differentiability and domination conditions provided by Assumptions 3(iii) and 5(iv.a) ensure (e.g., by Bartle, corollary 5.9??) the continuous differentiability of λ on $\mathbb{A}$, with

$$
\nabla \lambda(\alpha) = \sum_{j=1}^{p} E[\nabla \{\nabla' q_{j,t}(\cdot, \alpha) \int_{\delta_{j,t}(\alpha,\alpha^*)}^{0} f_{j,t}(s) ds\}].
$$

Since α^* is interior to $\mathbb{A}$ by Assumption 4(ii), the mean value theorem applies to each element of λ to yield

$$
\lambda(\alpha) = \lambda(\alpha^*) + Q_0(\alpha - \alpha^*), \tag{15}
$$

for α in a convex compact neighborhood of α^*,where Q_0 is an $\ell \times \ell$ matrix with $(1 \times \ell)$ rows $Q_i(\bar{\alpha}_{(i)}) = \nabla' \lambda_i(\bar{\alpha}_{(i)})$, where $\bar{\alpha}_{(i)}$ is a mean value (different for each i) lying on the segment connecting α and $\alpha^*, i = 1, ..., \ell$. The chain rule and an application of the Leibniz rule to $\int_{\delta_{j,t}(\alpha,\alpha^*)}^{0} f_{j,t}(s) ds$ then give

$$
Q_i(\alpha) = A_i(\alpha) - B_i(\alpha),
$$

where

$$\begin{aligned}
A_i(\alpha) &\equiv \sum_{j=1}^{p} E[\nabla_i \nabla' q_{j,t}(\cdot,\alpha) \int_{\delta_{j,t}(\alpha,\alpha^*)}^{0} f_{j,t}(s)ds] \\
B_i(\alpha) &\equiv \sum_{j=1}^{p} E[f_{j,t}(\delta_{j,t}(\alpha,\alpha^*))\nabla_i q_{j,t}(\cdot,\alpha)\nabla' q_{j,t}(\cdot,\alpha)].
\end{aligned}$$

Assumption 2(iii) and the other domination conditions (those of Assumption 5) then ensure that

$$\begin{aligned}
A_i(\bar{\alpha}_{(i)}) &= O(||\alpha-\alpha^*||) \\
B_i(\bar{\alpha}_{(i)}) &= Q_i^* + O(||\alpha-\alpha^*||),
\end{aligned}$$

where $Q_i^* \equiv \sum_{j=1}^{p} E[f_{j,t}(0)\nabla_i q_{j,t}(\cdot,\alpha^*)\nabla' q_{j,t}(\cdot,\alpha^*)]$. Letting $Q^* \equiv \sum_{j=1}^{p} E[f_{j,t}(0) \times \nabla q_{j,t}(\cdot,\alpha^*)\nabla' q_{j,t}(\cdot,\alpha^*)]$, we obtain

$$Q_0 = -Q^* + O(||\alpha-\alpha^*||). \tag{16}$$

Next, we have that $\lambda(\alpha^*) = 0$. To show this, we write

$$\begin{aligned}
\lambda(\alpha^*) &= \sum_{j=1}^{p} E[\nabla q_{j,t}(\cdot,\alpha^*)\ \psi_{\theta_j}(Y_t - q_{j,t}(\cdot,\alpha^*))] \\
&= \sum_{j=1}^{p} E(E[\nabla q_{j,t}(\cdot,\alpha^*)\ \psi_{\theta_j}(Y_t - q_{j,t}(\cdot,\alpha^*))|\mathcal{F}_{t-1}]) \\
&= \sum_{j=1}^{p} E(\nabla q_{j,t}(\cdot,\alpha^*)\ E[\psi_{\theta_j}(Y_t - q_{j,t}(\cdot,\alpha^*))|\mathcal{F}_{t-1}]) \\
&= \sum_{j=1}^{p} E(\nabla q_{j,t}(\cdot,\alpha^*)\ E[\psi_{\theta_j}(\varepsilon_{j,t})|\mathcal{F}_{t-1}]) \\
&= 0,
\end{aligned}$$

as $E[\psi_{\theta_j}(\varepsilon_{j,t})|\mathcal{F}_{t-1}] = \theta_j - E[1_{[Y_t \leq q_{j,t}^*]}|\mathcal{F}_{t-1}] = 0$, by definition of $q_{j,t}^*, j = 1, ..., p$ (see eq. (2)). Combining $\lambda(\alpha^*) = 0$ with eqs. (15) and (16), we obtain

$$\lambda(\alpha) = -Q^*(\alpha-\alpha^*) + O(||\alpha-\alpha^*||^2). \tag{17}$$

The next step is to show that

$$T^{1/2}\lambda(\hat{\alpha}) + H_T = o_p(1) \tag{18}$$

where $H_T \equiv T^{-1/2}\sum_{t=1}^{T}\eta_t^*$, with $\eta_t^* \equiv \eta_t(\alpha^*), \eta_t(\alpha) \equiv \sum_{j=1}^{p} \nabla q_{j,t}(\cdot,\alpha)\ \psi_{\theta_j}(Y_t - q_{j,t}(\cdot,\alpha))$. Let $u_t(\alpha,d) \equiv \sup_{\{\tau:||\tau-\alpha||\leq d\}} ||\eta_t(\tau) - \eta_t(\alpha)||$. By the results of Huber

(1967) and Weiss (1991), to prove (18) it suffices to show the following: (i) there exist $a > 0$ and $d_0 > 0$ such that $||\lambda(\alpha)|| \geq a||\alpha - \alpha^*||$ for $||\alpha - \alpha^*|| \leq d_0$; (ii) there exist $b > 0$, $d_0 > 0$, and $d \geq 0$ such that $E[u_t(\alpha, d)] \leq bd$ for $||\alpha - \alpha^*|| + d \leq d_0$; and (iii) there exist $c > 0, d_0 > 0$, and $d \geq 0$ such that $E[u_t(\alpha, d)^2] \leq cd$ for $||\alpha - \alpha^*|| + d \leq d_0$.

The condition that Q^* is positive-definite in Assumption 6(i) is sufficient for (i). For (ii), we have that for given (small) $d > 0$

$$\begin{aligned} u_t(\alpha, d) &\leq \sup_{\{\tau:||\tau-\alpha||\leq d\}} \sum_{j=1}^{p} ||\nabla q_{j,t}(\cdot, \tau)\psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \tau)) - \nabla q_{j,t}(\cdot, \alpha)\psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha))|| \\ &\leq \sum_{j=1}^{p} \sup_{\{\tau:||\tau-\alpha||\leq d\}} ||\psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \tau))|| \times \sup_{\{\tau:||\tau-\alpha||\leq d\}} ||\nabla q_{j,t}(\cdot, \tau) - \nabla q_{j,t}(\cdot, \alpha)|| \\ &\quad + \sum_{j=1}^{p} \sup_{\{\tau:||\tau-\alpha||\leq d\}} ||\psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha)) - \psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \tau))|| \times \sup_{\{\tau:||\tau-\alpha||\leq d\}} ||\nabla q_{j,t}(\cdot, \alpha)|| \\ &\leq pD_{2,t}d + D_{1,t} \sum_{j=1}^{p} 1_{[|Y_t - q_{j,t}(\cdot,\alpha)| < D_{2,t}d]} \end{aligned} \tag{19}$$

using the following; (i) $||\psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \tau))|| \leq 1$, (ii) $||\psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha)) - \psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \tau))|| \leq 1_{[|Y_t - q_{j,t}(\cdot,\alpha)| < |q_{j,t}(\cdot,\tau) - q_{j,t}(\cdot,\alpha)|]}$, and (iii) the mean value theorem applied to $\nabla q_{j,t}(\cdot, \tau)$ and $q_{j,t}(\cdot, \alpha)$. Hence, we have

$$E[u_t(\alpha, d)] \leq pC_0 d + 2pC_1 f_0 d$$

for some constants C_0 and C_1 given Assumptions 2(iii.a), 5(iii.a), and 5(iv.a). Hence, (ii) holds for $b = pC_0 + 2pC_1 f_0$ and $d_0 = 2d$. The last condition (iii) can be similarly verified by applying the c_r-inequality to eq. (19) with $d < 1$ (so that $d^2 < d$) and using Assumptions 2(iii.a), 5(iii.b), and 5(iv.b). Thus, eq. (18) is verified .

Combining eqs. (17) and (18) thus yields

$$Q^* T^{1/2}(\hat{\alpha}_T - \alpha^*) = T^{-1/2} \sum_{t=1}^{T} \eta_t^* + o_p(1)$$

But $\{\eta_t^*, \mathcal{F}_t\}$ is a stationary ergodic martingale difference sequence (MDS). In particular, η_t^* is measurable$-\mathcal{F}_t$, and $E(\eta_t^*|\mathcal{F}_{t-1}) = E(\sum_{j=1}^{p} \nabla q_{j,t}(\cdot, \alpha^*)\psi_{\theta_j}(\varepsilon_{j,t})|\mathcal{F}_{t-1}) = \sum_{j=1}^{p} \nabla q_{j,t}(\cdot, \alpha^*)E(\psi_{\theta_j}(\varepsilon_{j,t})|\mathcal{F}_{t-1}) = 0$, as $E[\psi_{\theta_j}(\varepsilon_{j,t})|\mathcal{F}_{t-1}] = 0$ for all $j = 1, ..., p$. Assumption 5(iii.b) ensures that $V^* \equiv E(\eta_t^* \eta_t^{*\prime})$ is finite. The MDS central limit theorem (e.g., theorem 5.24 of White, 2001) applies, provided V^* is positive definite (as ensured by Assumption 6(ii)) and that $T^{-1} \sum_{t=1}^{T} \eta_t^* \eta_t^{*\prime} = V^* + o_p(1)$, which is ensured by the ergodic theorem. The standard argument now gives

$$V^{*-1/2}Q^{*}T^{1/2}(\hat{\alpha}_T - \alpha^*) \xrightarrow{d} N(0, I),$$

which completes the proof.

Proof of Theorem 3: We have

$$\hat{V}_T - V^* = (T^{-1}\sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t' - T^{-1}\sum_{t=1}^{T}\eta_t^*\eta_t^{*\prime}) + (T^{-1}\sum_{t=1}^{T}\eta_t^*\eta_t^{*\prime} - E[\eta_t^*\eta_t^{*\prime}]),$$

where $\hat{\eta}_t \equiv \sum_{j=1}^{p}\nabla\hat{q}_{j,t}\hat{\psi}_{j,t}$ and $\eta_t^* \equiv \sum_{j=1}^{p}\nabla q_{j,t}^*\psi_{j,t}^*$, with $\nabla\hat{q}_{j,t} \equiv \nabla q_{j,t}(\cdot, \hat{\alpha}_T)$, $\hat{\psi}_{j,t} \equiv \psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \hat{\alpha}_T))$, $\nabla q_{j,t}^* \equiv \nabla q_{j,t}(\cdot, \alpha^*)$, and $\psi_{j,t}^* \equiv \psi_{\theta_j}(Y_t - q_{j,t}(\cdot, \alpha^*))$. Assumptions 1 and 2(i,ii) ensure that $\{\eta_t^*\eta_t^{*\prime}\}$ is a stationary ergodic sequence. Assumptions 3(i,ii), 4(i.a), and 5(iii) ensure that $E[\eta_t^*\eta_t^{*\prime}] < \infty$. It follows by the ergodic theorem that $T^{-1}\sum_{t=1}^{T}\eta_t^*\eta_t^{*\prime} - E[\eta_t^*\eta_t^{*\prime}] = o_p(1)$. Thus, it suffices to prove $T^{-1}\sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t' - T^{-1}\sum_{t=1}^{T}\eta_t^*\eta_t^{*\prime} = o_p(1)$.

The (h, i) element of $T^{-1}\sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t' - T^{-1}\sum_{t=1}^{T}\eta_t^*\eta_t^{*\prime}$ is

$$T^{-1}\sum_{t=1}^{T}\{\sum_{j=1}^{p}\sum_{k=1}^{p}\hat{\psi}_{j,t}\hat{\psi}_{k,t}\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t} - \psi_{j,t}^*\psi_{k,t}^*\nabla_h q_{j,t}^*\nabla_i q_{k,t}^*\}.$$

Thus, it will suffice to show that for each (h, i) and (j, k) we have

$$T^{-1}\sum_{t=1}^{T}\{\hat{\psi}_{j,t}\hat{\psi}_{k,t}\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t} - \psi_{j,t}^*\psi_{k,t}^*\nabla_h q_{j,t}^*\nabla_i q_{k,t}^*\} = o_p(1).$$

By the triangle inequality,

$$|T^{-1}\sum_{t=1}^{T}\{\hat{\psi}_{j,t}\hat{\psi}_{k,t}\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t} - \psi_{j,t}^*\psi_{k,t}^*\nabla_h q_{j,t}^*\nabla_i q_{k,t}^*\}| \leq A_T + B_T,$$

where

$$\begin{aligned} A_T &= |T^{-1}\sum_{t=1}^{T}\hat{\psi}_{j,t}\hat{\psi}_{k,t}\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t} - \psi_{j,t}^*\psi_{k,t}^*\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t}| \\ B_T &= |T^{-1}\sum_{t=1}^{T}\psi_{j,t}^*\psi_{k,t}^*\nabla_h q_{j,t}^*\nabla_i q_{k,t}^* - \psi_{j,t}^*\psi_{k,t}^*\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t}|. \end{aligned}$$

We now show that $A_T = o_p(1)$ and $B_T = o_p(1)$, delivering the desired result. For A_T, the triangle inequality gives

$$A_T \leq A_{1T} + A_{2T} + A_{3T},$$

where

$$
\begin{aligned}
A_{1T} &= T^{-1}\sum_{t=1}^{T}\theta_j|1_{[\varepsilon_{j,t}\le 0]}-1_{[\hat{\varepsilon}_{j,t}\le 0]}||\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t}| \\
A_{2T} &= T^{-1}\sum_{t=1}^{T}\theta_k|1_{[\varepsilon_{k,t}\le 0]}-1_{[\hat{\varepsilon}_{k,t}\le 0]}||\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t}| \\
A_{3T} &= T^{-1}\sum_{t=1}^{T}|1_{[\varepsilon_{j,t}\le 0]}1_{[\varepsilon_{k,t}\le 0]}-1_{[\hat{\varepsilon}_{j,t}\le 0]}1_{[\hat{\varepsilon}_{k,t}\le 0]}||\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t}|.
\end{aligned}
$$

Theorem 2, ensured by Assumptions $1-6$, implies that $T^{1/2}||\hat{\alpha}_T-\alpha^*||=O_p(1)$. This, together with Assumptions 2(iii,iv) and 5(iii.b), enables us to apply the same techniques used in Kim and White (2003) to show $A_{1T}=o_p(1)$, $A_{2T}=o_p(1)$, $A_{3T}=o_p(1)$, implying $A_T=o_p(1)$.

It remains to show $B_T=o_p(1)$. By the triangle inequality,

$$B_T\le B_{1T}+B_{2T},$$

where

$$
\begin{aligned}
B_{1T} &= |T^{-1}\sum_{t=1}^{T}\psi^*_{j,t}\psi^*_{k,t}\nabla_h q^*_{j,t}\nabla_i q^*_{k,t}-E[\psi^*_{j,t}\psi^*_{k,t}\nabla_h q^*_{j,t}\nabla_i q^*_{k,t}]| \\
B_{2T} &= |T^{-1}\sum_{t=1}^{T}\psi^*_{j,t}\psi^*_{k,t}\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{k,t}-E[\psi^*_{j,t}\psi^*_{k,t}\nabla_h q^*_{j,t}\nabla_i q^*_{k,t}]|.
\end{aligned}
$$

Assumptions 1, 2(i,ii), 3(i,ii), 4(i.a), and 5(iii) ensure that the ergodic theorem applies to $\{\psi^*_{j,t}\psi^*_{k,t}\nabla_h q^*_{j,t}\nabla_i q^*_{k,t}\}$, so $B_{1T}=o_p(1)$. Next, Assumptions 1, 3(i,ii), and 5(iii) ensure that the stationary ergodic ULLN applies to $\{\psi^*_{j,t}\psi^*_{k,t}\nabla_h q_{j,t}(\cdot,\alpha)\nabla_i q_{k,t}(\cdot,\alpha)\}$. This and the result of Theorem 1 ($\hat{\alpha}_T-\alpha^*=o_p(1)$) ensure that $B_{2T}=o_p(1)$ by e.g., White (1994, corollary 3.8), and the proof is complete.

Proof of Theorem 4: We begin by sketching the proof. We first define

$$Q_T\equiv(2c_TT)^{-1}\sum_{t=1}^{T}\sum_{j=1}^{p}1_{[-c_T\le\varepsilon_{j,t}\le c_T]}\nabla q^*_{j,t}\nabla' q^*_{j,t},$$

and then we will show the following:

$$Q^*-E(Q_T)\xrightarrow{p}0, \tag{20}$$

$$E(Q_T)-Q_T\xrightarrow{p}0, \tag{21}$$

$$Q_T-\hat{Q}_T\xrightarrow{p}0. \tag{22}$$

Combining the results above will deliver the desired outcome: $\hat{Q}_T - Q^* \xrightarrow{p} 0$.

For (20), one can show by applying the mean value theorem to $F_{j,t}(c_T) - F_{j,t}(-c_T)$, where $F_{j,t}(c) \equiv \int 1_{\{s\leq c\}} f_{j,t}(s)ds$, that

$$E(Q_T) = T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{p} E[f_{j,t}(\xi_{j,T})\nabla q_{j,t}^*\nabla' q_{j,t}^*] = \sum_{j=1}^{p} E[f_{j,t}(\xi_{j,T})\nabla q_{j,t}^*\nabla' q_{j,t}^*],$$

where $\xi_{j,T}$ is a mean value lying between $-c_T$ and c_T, and the second equality follows by stationarity. Therefore, the (h,i) element of $|E(Q_T) - Q^*|$ satisfies

$$\begin{aligned}
& |\sum_{j=1}^{p} E\left\{f_{j,t}(\xi_{j,T}) - f_{j,t}(0)\nabla_h q_{j,t}^*\nabla_i q_{j,t}^*\right\}| \\
\leq\ & \sum_{j=1}^{p} E\left\{|f_{j,t}(\xi_{j,T}) - f_{j,t}(0)||\nabla_h q_{j,t}^*\nabla_i q_{j,t}^*|\right\} \\
\leq\ & \sum_{j=1}^{p} L_0 E\left\{|\xi_{j,T}||\nabla_h q_{j,t}^*\nabla_i q_{j,t}^*|\right\} \\
\leq\ & pL_0 c_T E[D_{1,t}^2],
\end{aligned}$$

which converges to zero as $c_T \to 0$. The second inequality follows by Assumption 2(iii.b), and the last inequality follows by Assumption 5(iii.b). Therefore, we have the result in eq.(20).

To show (21), it suffices simply to apply a LLN for double arrays, e.g. theorem 2 in Andrews (1988).

Finally, for (22), we consider the (h,i) element of $|\hat{Q}_T - Q_T|$, which is given by

$$\begin{aligned}
& |\frac{1}{2\hat{c}_T T}\sum_{t=1}^{T}\sum_{j=1}^{p} 1_{[-\hat{c}_T\leq\hat{\varepsilon}_{j,t}\leq\hat{c}_T]}\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{j,t} - \frac{1}{2c_T T}\sum_{t=1}^{T}\sum_{j=1}^{p} 1_{[-c_T\leq\varepsilon_{j,t}\leq c_T]}\nabla_h q_{j,t}^*\nabla_i q_{j,t}^*| \\
=\ & \frac{c_T}{\hat{c}_T}|\frac{1}{2c_T T}\sum_{t=1}^{T}\sum_{j=1}^{p}(1_{[-\hat{c}_T\leq\hat{\varepsilon}_{j,t}\leq\hat{c}_T]} - 1_{[-c_T\leq\varepsilon_{j,t}\leq c_T]})\nabla_h\hat{q}_{j,t}\nabla_i\hat{q}_{j,t} \\
& +\frac{1}{2c_T T}\sum_{t=1}^{T}\sum_{j=1}^{p} 1_{[-c_T\leq\varepsilon_{j,t}\leq c_T]}(\nabla_h\hat{q}_{j,t} - \nabla_h q_{j,t}^*)\nabla_i\hat{q}_{j,t} \\
& +\frac{1}{2c_T T}\sum_{t=1}^{T}\sum_{j=1}^{p} 1_{[-c_T\leq\varepsilon_{j,t}\leq c_T]}\nabla_h q_{j,t}^*(\nabla_i\hat{q}_{j,t} - \nabla_i q_{j,t}^*) \\
& +\frac{1}{2c_T T}(1 - \frac{\hat{c}_T}{c_T})\sum_{t=1}^{T}\sum_{j=1}^{p} 1_{[-c_T\leq\varepsilon_{j,t}\leq c_T]}\nabla_h q_{j,t}^*\nabla_i q_{j,t}^*| \\
\leq\ & \frac{c_T}{\hat{c}_T}[A_{1T} + A_{2T} + A_{3T} + (1 - \frac{\hat{c}_T}{c_T})A_{4T}],
\end{aligned}$$

where

$$
\begin{aligned}
A_{1T} &\equiv \frac{1}{2c_T T}\sum_{t=1}^{T}\sum_{j=1}^{p}|1_{[-\hat{c}_T\le\hat{\varepsilon}_{j,t}\le\hat{c}_T]} - 1_{[-c_T\le\varepsilon_{j,t}\le c_T]}| \times |\nabla_h \hat{q}_{j,t}\nabla_i \hat{q}_{j,t}| \\
A_{2T} &\equiv \frac{1}{2c_T T}\sum_{t=1}^{T}\sum_{j=1}^{p}1_{[-c_T\le\varepsilon_{j,t}\le c_T]}|\nabla_h \hat{q}_{j,t} - \nabla_h q^*_{j,t}| \times |\nabla_i \hat{q}_{j,t}| \\
A_{3T} &\equiv \frac{1}{2c_T T}\sum_{t=1}^{T}\sum_{j=1}^{p}1_{[-c_T\le\varepsilon_{j,t}\le c_T]}|\nabla_h q^*_{j,t}| \times |\nabla_i \hat{q}_{j,t} - \nabla_i q^*_{j,t}| \\
A_{4T} &\equiv \frac{1}{2c_T T}\sum_{t=1}^{T}\sum_{j=1}^{p}1_{[-c_T\le\varepsilon_{j,t}\le c_T]}|\nabla_h q^*_{j,t}\nabla_i q^*_{j,t}|.
\end{aligned}
$$

It will suffice to show that $A_{1T} = o_p(1), A_{2T} = o_p(1), A_{3T} = o_p(1)$, and $A_{4T} = O_p(1)$. Then, because $\hat{c}_T/c_T \xrightarrow{p} 1$, we obtain the desired result: $\hat{Q}_T - Q^* \xrightarrow{p} 0$.

We first show $A_{1T} = o_p(1)$. It will suffice to show that for each j,

$$
\frac{1}{2c_T T}\sum_{t=1}^{T}|1_{[-\hat{c}_T\le\hat{\varepsilon}_{j,t}\le\hat{c}_T]} - 1_{[-c_T\le\varepsilon_{j,t}\le c_T]}| \times |\nabla_h \hat{q}_{j,t}\nabla_i \hat{q}_{j,t}| = o_p(1).
$$

Let α_T lie between $\hat{\alpha}_T$ and α^*, and put $d_{j,t,T} \equiv ||\nabla q_{j,t}(\cdot,\alpha_T)||\times||\hat{\alpha}_T-\alpha^*||+|\hat{c}_T-c_T|$. Then

$$
(2c_T T)^{-1}\sum_{t=1}^{T}|1_{[-\hat{c}_T\le\hat{\varepsilon}_{j,t}\le\hat{c}_T]} - 1_{[-c_T\le\varepsilon_{j,t}\le c_T]}| \times |\nabla_h \hat{q}_{j,t}\nabla_i \hat{q}_{j,t}| \;\le U_T + V_T,
$$

where

$$
\begin{aligned}
U_T &\equiv (2c_T T)^{-1}\sum_{t=1}^{T}1_{[|\varepsilon_{j,t}-c_T|<d_{j,t,T}]}|\nabla_h \hat{q}_{j,t}\nabla_i \hat{q}_{j,t}| \\
V_T &\equiv (2c_T T)^{-1}\sum_{t=1}^{T}1_{[|\varepsilon_{j,t}+c_T|<d_{j,t,T}]}|\nabla_h \hat{q}_{j,t}\nabla_i \hat{q}_{j,t}|.
\end{aligned}
$$

It will suffice to show that $U_T \xrightarrow{p} 0$ and $V_T \xrightarrow{p} 0$. Let $\eta > 0$ and let z be an arbitrary positive number. Then, using reasoning similar to that of Kim and White (2003, lemma 5), one can show that for any $\eta > 0$,

$$
\begin{aligned}
P(U_T &> \eta) \le P((2c_T T)^{-1}\sum_{t=1}^{T}1_{[|\varepsilon_{j,t}-c_T|<(||\nabla q_t(\theta_j,\alpha_0)||+1)zc_T]})|\nabla_h \hat{q}_{j,t}\nabla_i \hat{q}_{j,t}| > \eta) \\
&\le \frac{zf_0}{\eta T}\sum_{t=1}^{T}E\left\{(||\nabla q_t(\theta_j,\alpha_T)|| + 1)|\nabla_h \hat{q}_{j,t}\nabla_i \hat{q}_{j,t}|\right\} \\
&\le zf_0\{E|D^3_{1,t}| + E|D^2_{1,t}|\}/\eta,
\end{aligned}
$$

where the second inequality is due to the Markov inequality and Assumption 2(iii.a), and the third is due to Assumption 5(iii.c). As z can be chosen arbitrarily small and the remaining terms are finite by assumption, we have $U_T \xrightarrow{p} 0$. The same argument is used to show $V_T \xrightarrow{p} 0$. Hence, $A_{1T} = o_p(1)$ is proved.

Next, we show $A_{2T} = o_p(1)$. For this, it suffices to show $A_{2T,j} \equiv \frac{1}{2c_T T}\sum_{t=1}^{T} 1_{[-c_T \leq \varepsilon_{j,t} \leq c_T]} |\nabla_h \hat{q}_{j,t} - \nabla_h q^*_{j,t}| \times |\nabla_i \hat{q}_{j,t}| = o_p(1)$ for each j. Note that

$$\begin{aligned} A_{2T,j} &\leq \frac{1}{2c_T T}\sum_{t=1}^{T} |\nabla_h \hat{q}_{j,t} - \nabla_h q^*_{j,t}| \times |\nabla_i \hat{q}_{j,t}| \\ &\leq \frac{1}{2c_T T}\sum_{t=1}^{T} ||\nabla_h^2 q_{j,t}(\cdot, \tilde{\alpha})|| \times ||\hat{\alpha}_T - \alpha^*|| \times |\nabla_i \hat{q}_{j,t}| \\ &\leq \frac{1}{2c_T}||\hat{\alpha}_T - \alpha^*|| \frac{1}{T}\sum_{t=1}^{T} D_{2,t} D_{1,t} \\ &= \frac{1}{2c_T T^{1/2}} T^{1/2} ||\hat{\alpha}_T - \alpha^*|| \frac{1}{T}\sum_{t=1}^{T} D_{2,t} D_{1,t}, \end{aligned}$$

where $\tilde{\alpha}$ is between $\hat{\alpha}_T$ and α^*, and $\nabla_h^2 q_{j,t}(\cdot, \tilde{\alpha})$ is the first derivative of $\nabla_h \hat{q}_{j,t}$ with respect to α, which is evaluated at $\tilde{\alpha}$. The last expression above is $o_p(1)$ because (i) $T^{1/2}||\hat{\alpha}_T - \alpha^*|| = O_p(1)$ by Theorem 2, (ii) $T^{-1}\sum_{t=1}^{T} D_{2,t} D_{1,t} = O_p(1)$ by the ergodic theorem and (iii) $1/(c_T T^{1/2}) = o_p(1)$ by Assumption 7(iii). Hence, $A_{2T} = o_p(1)$. The other claims $A_{3T} = o_p(1)$ and $A_{4T} = O_p(1)$ can be analogously and more easily proven. Hence, they are omitted. Therefore, we finally have $Q_T - \hat{Q}_T \xrightarrow{p} 0$, which, together with (20) and (21), implies that $\hat{Q}_T - Q^* \xrightarrow{p} 0$. The proof is complete.

Table 1. S&P500 index: Estimation results for the LRS model

β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
0.01 (0.18)	0.05 (0.19)	0.94 (0.04)	-0.04 (0.15)	0.01 (0.01)	0.01 (0.02)	3.25 (0.04)	0.00 (0.00)	0.00 (0.00)
Note: standard errors are in parentheses.								

Table 2. S&P500 index : Estimation results for the MQ-CAViaR model

θ_j	β_{j1}	β_{j2}	γ_{j1}	γ_{j2}	γ_{j3}	γ_{j4}	γ_{j5}
0.025	-0.04 (0.05)	-0.11 (0.07)	0.93 (0.12)	0.02 (0.10)	0 (0.29)	0 (0.93)	0 (0.30)
0.25	0.001 (0.02)	-0.01 (0.05)	0 (0.06)	0.99 (0.03)	0 (0.04)	0 (0.63)	0 (0.20)
0.50	0.10 (0.02)	0.04 (0.04)	0.03 (0.04)	0 (0.02)	-0.32 (0.02)	0 (0.52)	-0.02 (0.17)
0.75	0.03 (0.31)	-0.01 (0.05)	0 (0.70)	0 (0.80)	0 (2.33)	0.04 (0.84)	0.29 (0.34)
0.975	0.03 (0.06)	0.24 (0.07)	0 (0.16)	0 (0.16)	0 (0.33)	0.03 (0.99)	0.89 (0.29)
Note: standard errors are in parentheses.							

Table 3. Means of point estimates through 100 replications (T = 1,000)

	True parameters						
θ_j	β_{j1}	β_{j2}	γ_{j1}	γ_{j2}	γ_{j3}	γ_{j4}	γ_{j5}
0.025	-0.05	-0.10	0.93	0.04	0.00	0.00	0.00
	(0.08)	(0.02)	(0.14)	(0.34)	(0.00)	(0.01)	(0.00)
0.25	-0.05	-0.01	0.04	0.81	0.00	0.00	0.00
	(0.40)	(0.05)	(0.17)	(0.47)	(0.01)	(0.00)	(0.00)
0.50	-0.08	0.02	0.00	0.00	-0.06	0.00	0.00
	(0.15)	(0.06)	(0.01)	(0.00)	(0.81)	(0.00)	(0.01)
0.75	0.20	0.05	0.00	0.00	0.00	0.38	0.13
	(0.42)	(0.11)	(0.02)	(0.02)	(0.00)	(0.63)	(0.19)
0.975	0.06	0.22	0.00	0.00	0.00	0.10	0.87
	(0.16)	(0.03)	(0.00)	(0.01)	(0.00)	(0.56)	(0.16)

Note: standard errors are in parentheses.

Table 4. Means of point estimates through 100 replications (T = 2,280)

	True parameters						
θ_j	β_{j1}	β_{j2}	γ_{j1}	γ_{j2}	γ_{j3}	γ_{j4}	γ_{j5}
0.025	-0.04	-0.10	0.93	0.03	0.00	0.00	0.00
	(0.03)	(0.01)	(0.07)	(0.21)	(0.00)	(0.00)	(0.00)
0.25	-0.04	-0.01	0.03	0.88	0.00	0.00	0.00
	(0.18)	(0.02)	(0.12)	(0.38)	(0.00)	(0.00)	(0.00)
0.50	-0.01	0.02	0.00	0.00	-0.03	0.00	0.00
	(0.11)	(0.04)	(0.00)	(0.01)	(0.75)	(0.00)	(0.02)
0.75	0.09	0.01	0.00	0.00	0.00	0.33	0.19
	(0.21)	(0.07)	(0.00	(0.01)	(0.00)	(0.58)	(0.18)
0.975	0.05	0.24	0.00	0.00	0.00	0.18	0.83
	(0.13)	(0.02)	(0.00)	(0.03)	(0.00)	(0.69)	(0.22)

Note: standard errors are in parentheses.

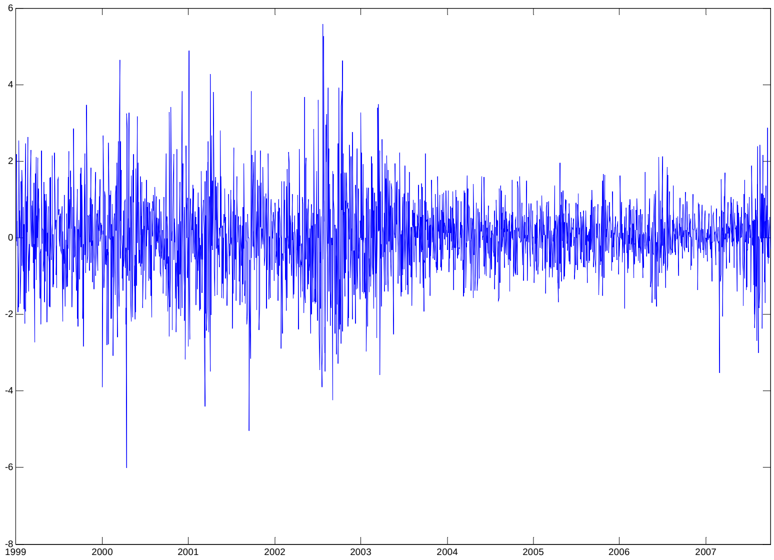

Figure 1: S&P500 daily returns: January 1, 1999 – September 30, 2007.

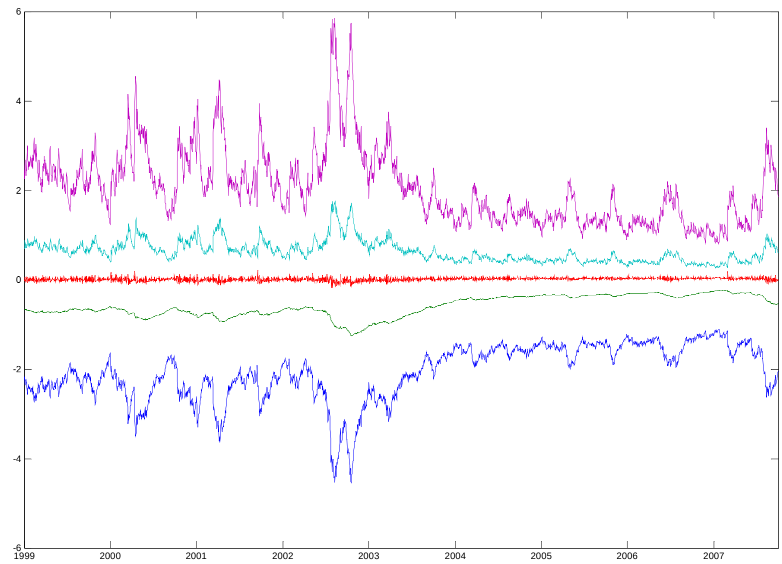

Figure 2: S&P500 conditional quantiles: January 1, 1999 - September 30, 2007.

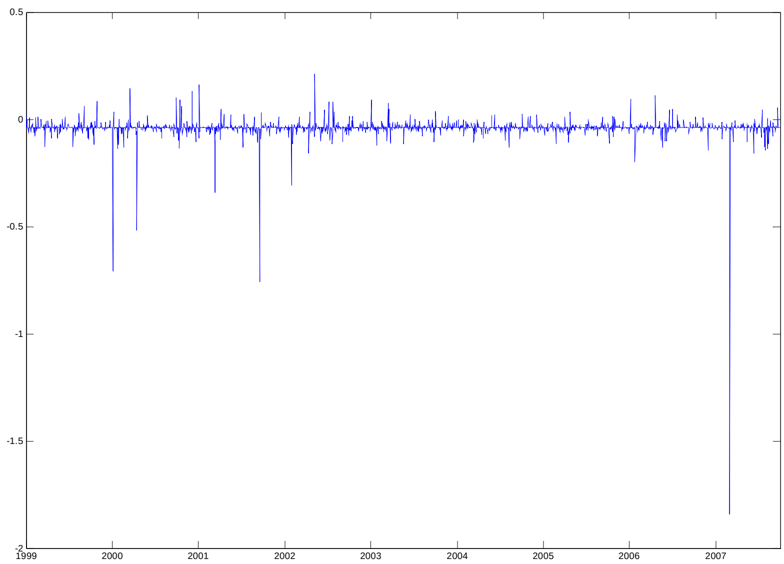

Figure 3: S&P 500: Estimated conditional skewness, LRS model.

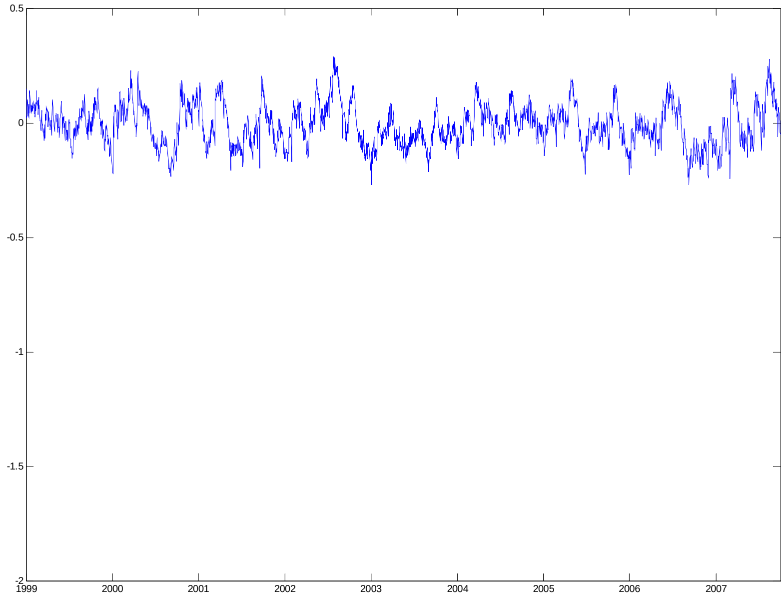

Figure 4: S&P 500: Estimated conditional skewness, MQ-CAViaR model.

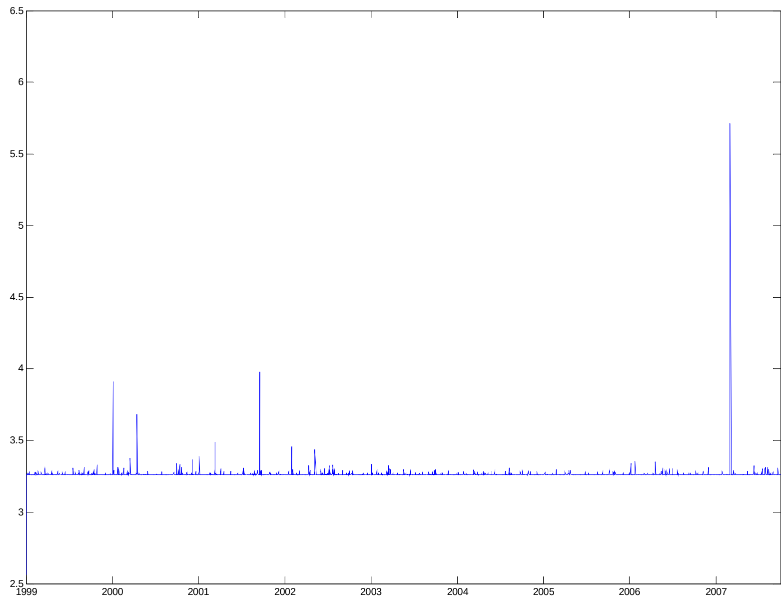

Figure 5: S&P 500: Estimated conditional kurtosis, LRS model.

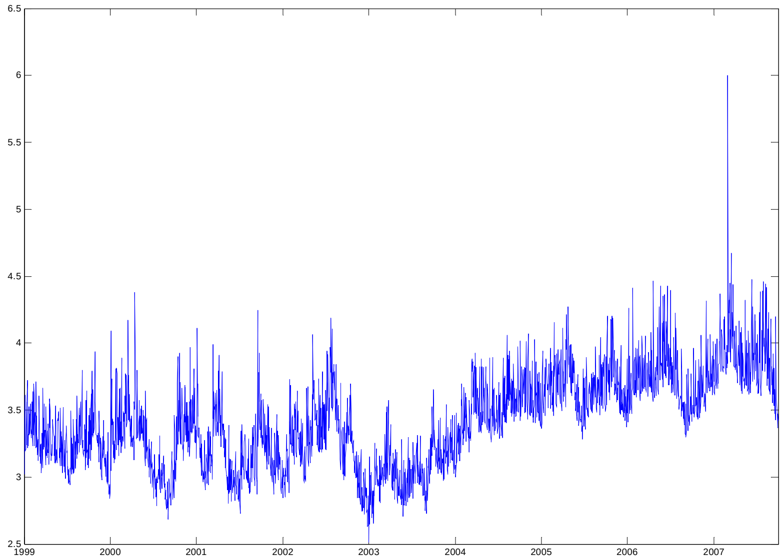

Figure 6: S&P 500: Estimated conditional kurtosis, MQ-CAViaR model.