The Econometrics Journal

Econometrics Journal (2014), volume **17**, pp. 1–23.
doi: 10.1111/ectj.12023

Weighted composite quantile regression estimation of DTARCH models

JIANCHENG JIANG[†], XUEJUN JIANG[‡,§] AND XINYUAN SONG[‖]

[†]*University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223, USA.*
E-mail: jjiang1@uncc.edu

[‡]*South University of Science and Technology, 1088 Xueyuan Ave, Nanshan, Shenzhen, Guangdong, China*

[§]*Zhongnan University of Economics and Law, 114 Wuluo Rd, Wuchang, Wuhan, Hubei, China.*
E-mail: jiang.xj@sustc.edu.cn

[‖]*The Chinese University of Hong Kong, Shatin, NT, Hong Kong.*
E-mail: xysong@sta.cuhk.edu.hk

First version received: June 2012; final version accepted: October 2013

Summary In modelling volatility in financial time series, the double-threshold autoregressive conditional heteroscedastic (DTARCH) model has been demonstrated as a useful variant of the autoregressive conditional heteroscedastic (ARCH) models. In this paper, we propose a weighted composite quantile regression method for simultaneously estimating the autoregressive parameters and the ARCH parameters in the DTARCH model. This method involves a sequence of weights and takes a data-driven weighting scheme to maximize the asymptotic efficiency of the estimators. Under regularity conditions, we establish asymptotic distributions of the proposed estimators for a variety of heavy- or light-tailed error distributions. Simulations are conducted to compare the performance of different estimators, and the proposed approach is used to analyse the daily S&P 500 Composite index, both of which endorse our theoretical results.

Keywords: *Conditional heteroscedasticity, Double-threshold, Weighted composite quantile regression.*

1. INTRODUCTION

Modelling volatility lies at the core of activity in financial markets. Because volatility is fundamental for asset pricing, monetary policymaking, proprietary trading, portfolio management and risk analysis, it is especially important to accurately forecast volatility. There are two main volatility modelling methods leading to different concepts of volatility. The first is the discrete time method in the ARCH/generalized ARCH models (see Engle, 1982, Bollerslev, 1986, Bollerslev et al., 1992, 1994, Bera and Higgins, 1993, Fan and Yao, 2003, and Fan et al., 2012), where the volatility is thought of as conditional variance (or standard deviation) of the return. The second is the continuous time method in the diffusion models (see Merton, 1969, 1971, 1973, Ait-Sahalia, 1998, and Sundaresan, 2000, among others), where the volatility refers to either the instantaneous diffusion coefficient or the quadratic variation over a given time period (integrated volatility). One advantage of the discrete time method is that it provides conditional volatility, which is unknown in the continuous time method. Ait-Sahalia and Hansen (2009) have written a comprehensive and modern handbook about modelling volatility.

In this paper, we employ the discrete time method and focus on the double-threshold ARCH (DTARCH) model in Li and Li (1996). This model has some advantages. On the one hand, it is useful for detecting non-linear structures such as asymmetric behaviour in the mean and the volatility of an asset return, and heteroscedasticity with clustering in the volatility. On the other hand, when the mean and volatility parameters in different regions are the same, the DTARCH model reduces to the celebrated ARCH model of Engle (1982), so it can be used to check if a specific ARCH model is appropriate for modelling a given data set.

For the DTARCH model, Li and Li (1996) studied model identification, estimation and diagnostic checks based on the maximum likelihood (ML) principle, and Hui and Jiang (2005) investigated similar problems using L_1 regression. Although they are useful, both estimation approaches have limitations. The former is efficient when the error is normal, but it is sensitive to outliers and not robust against the error distribution, whereas the latter is resistant to outliers and robust against the error distribution, but not efficient when the error is normal. Hence, there is a genuine need for us to study robust and efficient estimation of the DTARCH model without specifying the error distribution. In particular, it is important to achieve an estimation method that is robust like the L_1 regression and efficient like the maximum likelihood estimation (MLE). This is the target we are aiming at.

Quantile regression (QR) is designed to estimate and to conduct inference about conditional quartile functions. The basic motivation for using quantiles rather than simple mean regression is that the stochastic relationship between random variables can be better portrayed and with much more accuracy; see, e.g., Chaudhuri et al. (1997). The QR provides more robust and efficient estimates than the mean regression when the error is non-normal (Koenker and Bassett, 1978, and Koenker and Zhao, 1996). This approach has been widely used in time series analysis (see, e.g., Koenker and Zhao, 1996, Davis and Dunsmuir, 1997, Hall and Yao, 2003, and Peng and Yao, 2003), but not for the DTARCH model. However, a straightforward use of QR leads to biased estimators of parameters in the DTARCH model (see Section 3.1.1). This motivates us to propose a weighted composite quantile regression (WCQR) method based on transformed residuals, with the aim of providing a new robust and efficient estimation method in finance and econometrics.

The proposed WCQR includes the QR as a special case and is considerably more efficient than the QR, while inheriting robustness. The WCQR with a data-driven weighting scheme is adaptive, in the sense that it performs as well as if the optimal weights were known, and hence it achieves maximum asymptotic efficiency among all WCQRs. For the ARCH model with an AR part, Engle (1982) and Koenker and Zhao (1996) studied a residual-based approach to the estimation of the volatility parameters. This approach can be extended to the DTARCH model. However, the residual-based QR approach depends on the initial estimates of the AR parameters (see also Section 3.3). This motivates us to study the simultaneous WCQR estimation of the AR and ARCH parameters. Yet, it involves minimizing a non-linear and non-convex function and leads to non-uniqueness of the solution, which makes derivation of theoretical results infeasible. To surmount this difficulty, we introduce one-step simultaneous estimation of the parameters to facilitate derivation of the asymptotics. In general, the one-step simultaneous estimator depends on the initial estimator, but the fully iterative estimator does not (see Theorems 3.3 and 3.4). If the innovation distribution is symmetrical, the one-step simultaneous estimator is shown to share the same asymptotic normality as the fully iterative simultaneous estimation while reducing its computational burden (see Corollary 3.1). Simulations are conducted to compare different estimators. Application of the proposed methodology to the daily S&P Composite index indicates that the DTARCH model can detect asymmetric volatility effect and the proposed fully iterative simultaneous WCQR estimator has best accuracy.

This paper is organized as follows. In Section 2, we review DTARCH models. In Section 3, we introduce the WCQR for DTARCH models and establish asymptotic properties of the proposed estimators, where data-driven estimators are developed to maximize the asymptotic efficiency. In Section 4, we conduct simulations and real data analysis. In Section 5, we present some concluding remarks. We provide all technical proofs in the Appendix.

2. REVIEW ON DTARCH MODELS

Given a time series y_t, $t = 1, \ldots, n$, let $\mathcal{F}_t$ be the σ-field generated by the realized value $\{y_t, y_{t-1}, \ldots\}$ at time t. Assume that y_t is generated from

$$y_t = \mathbf{X}'_{t,j}\boldsymbol{\alpha}^{(j)} + \varepsilon_t \quad \text{if} \quad r_{j-1} < y_{t-d} \leq r_j, \tag{2.1}$$

where $j = 1, \ldots, m$, the delay parameter d is a positive integer, the threshold parameters r_j satisfy $-\infty = r_0 < r_1 < r_2 < \cdots < r_m = \infty$, $\mathbf{X}_{t,j} = (1, y_{t-1}, \ldots, y_{t-p_j})^T$ is a $(p_j + 1) \times 1$ vector of lagged variables, and $\boldsymbol{\alpha}^{(j)} = (\alpha_0^{(j)}, \alpha_1^{(j)}, \ldots, \alpha_{p_j}^{(j)})^T$ is a $(p_j + 1) \times 1$ parameter vector. The stochastic error satisfies $\varepsilon_t = h_t(\boldsymbol{\beta})u_t$ with

$$h_t(\boldsymbol{\beta}) = \sum_{j=1}^{m} I_{t,j}\left(\beta_0^{(j)} + \beta_1^{(j)}|\varepsilon_{t-1}| + \cdots + \beta_{q_j}^{(j)}|\varepsilon_{t-q_j}|\right) \equiv \sum_{j=1}^{m} I_{t,j}\mathbf{Z}_{t,j}^T\boldsymbol{\beta}^{(j)}, \tag{2.2}$$

where $I_{t,j} = I(r_{j-1} < y_{t-d} \leq r_j)$, $\mathbf{Z}_{t,j} = (1, |\varepsilon_{t-1}|, \ldots, |\varepsilon_{t-q_j}|)^T$, the parameters in the conditional scales satisfy $\beta_0^{(j)} > 0$, $\beta_i^{(j)} \geq 0$ $(i = 1, \ldots, q_j)$, and the innovations $\{u_t\}$ are independent and identically distributed (i.i.d.) with an unknown distribution $F(x)$ and a density function $f(u)$. For convenience, as in Tsay (1989) and Li and Li (1996), we refer to the model in (2.1) and (2.2) as a DTARCH$(p_1, \ldots, p_m; q_1, \ldots, q_m)$ model, where the first m integers $p_1, \ldots, p_m$ represent the AR orders in the m regimes and the last m integers $q_1, \ldots, q_m$ denote the ARCH orders. The interval $r_{j-1} < y_{t-d} \leq r_j$ is the jth regime of y_{t-d}. The proposed model is similar to that in Li and Li (1996), where the conditional variance instead of the conditional scale is specified as the ARCH structure.

The model in (2.1) and (2.2) has some nice features. It depicts the clustering of deviations at different regions of the lagged variable y_{t-d}. As a natural extension to the Tong threshold model (see Tong and Lim, 1980), the double-threshold structure allows us to capture non-linear phenomena, such as asymmetric cycles, jump resonance and amplitude-frequency dependence. The fact that there is no assumption on the form of error distribution enables robust inference for the model.

Modelling the conditional scale is important. As noted by Bickel and Lehmann (1976), scale provides a more natural dispersion concept than variance, and also offers substantial advantages from the robustness point of view; see Bickel (1978), and Carroll and Ruppert (1988). Therefore, model (2.2) is especially appropriate for robust modelling. The advantage of such an approach with conditional scale instead of conditional variance was discussed for ARCH models by Koenker and Zhao (1996) and Jiang et al. (2001).

3. WCQR ESTIMATION OF THE DTARCH MODEL

In this section, we introduce the WCQR estimators of the parameters in the DTARCH model and establish their asymptotic properties. For explicit exposure of the methodology, we first consider the DTARCH model without the AR part.

3.1. Purely double-threshold conditional heteroscedastic model

The purely double-threshold conditional heteroscedastic model is the DTARCH model without the AR part, which takes the form

$$y_t = \varepsilon_t = h_t(\boldsymbol{\beta})u_t, \tag{3.1}$$

where $h_t(\boldsymbol{\beta})$ is defined in (2.2). Let $\mathbf{Z}_t = \mathrm{vec}(I_{t,1}\mathbf{Z}_{t,1}, \ldots, I_{t,m}\mathbf{Z}_{t,m})$, and let $\boldsymbol{\beta} = \mathrm{vec}(\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(m)})$. Then

$$h_t(\boldsymbol{\beta}) = \mathbf{Z}_t'\boldsymbol{\beta}. \tag{3.2}$$

3.1.1. Quantile regression. Note that the τth conditional quantile of ε_t given $\mathcal{F}_{t-1}$ is $Q_\tau(\varepsilon_t|\mathcal{F}_{t-1}) = h_t(\boldsymbol{\beta})F^{-1}(\tau)$, where $F^{-1}(\tau)$ denotes the τth quantile of u_t. Following the idea in Koenker and Bassett (1978), the τth QR estimator of $\boldsymbol{\beta}$ can be obtained by minimizing

$$\sum_{t=s+1}^{n} \rho_\tau\left(\varepsilon_t/h_t(\boldsymbol{\beta}) - b_\tau\right), \tag{3.3}$$

over b_τ and $\boldsymbol{\beta}$ for $0 < \tau < 1$, where $s = \max(q_1, \ldots, q_m)$ and $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function with derivative $\psi_\tau(u) = \tau - I(u < 0)$ for $u \neq 0$. Let the resulting estimator of $\boldsymbol{\beta}$ be $\hat{\boldsymbol{\beta}}_0$.

However, the resulting estimator $\hat{\boldsymbol{\beta}}_0$ is, unfortunately, asymptotically biased (see Theorem A.1 in the Appendix). To overcome this shortcoming, we define a modified form of the QR estimator. Note that $\log(|\varepsilon_t|) = \log(h_t(\beta)) + e_t$, where $e_t = \log(|u_t|)$. The τth QR estimate of $\boldsymbol{\beta}$ can be obtained by minimizing

$$\sum_{t=s+1}^{n} \rho_\tau(\log|\varepsilon_t| - \log(h_t(\boldsymbol{\beta})) - c_\tau) \tag{3.4}$$

over $\boldsymbol{\beta}$ and c_τ.

The distribution of $|\varepsilon_t|$ is confined to the non-negative half-axis and is typically skewed. The log-transform is a natural way to make the distribution less skewed. Peng and Yao (2003) advocated the log-transform and studied the L_1 regression for the ARCH/GARCH models.

3.1.2. Weighted composite quantile regression. For linear regression models,

$$y_i = x_i'\boldsymbol{\theta} + \epsilon_i, \quad i = 1, \ldots, n,$$

where ϵ_i are i.i.d. noises. Zou and Yuan (2008) studied the composite quantile regression (CQR) estimation,

$$\min_{b_{\tau_k},\boldsymbol{\theta}} \sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k}\left(y_i - \boldsymbol{x}_i'\boldsymbol{\theta} - b_{\tau_k}\right),$$

where different QR models receive the same weight, and $\{\tau_k\}_{k=1}^{K}$ are predetermined over (0, 1). Intuitively, it is more effective if different weights are used for different quantile regression models. Combined with (3.4), this motivates us to estimate the model parameters in (3.1) by minimizing

$$\sum_{k=1}^{K} \omega_k \sum_{t=s+1}^{n} \rho_{\tau_k}(\log|\varepsilon_t| - \log(h_t(\boldsymbol{\beta})) - c_{\tau_k}) \tag{3.5}$$

over c_{τ_k} and $\boldsymbol{\beta}$. Here, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)^T$ is a vector of weights such that $\|\boldsymbol{\omega}\| = 1$ with $\|\cdot\|$ denoting the Euclidean norm, and c_{τ_k} is the τ_kth quantile of e_t. We denote by $\hat{\boldsymbol{\beta}}_1$ the resulting solution for $\boldsymbol{\beta}$. For convenience, we refer to it as the WCQR estimator. Because c_{τ_k} is monotonic, we can minimize (3.5) under this monotonic constraint to improve the finite sample performance.

The idea of the WCQR method has appeared in the statistical literature. Using this idea, Koenker (1984) and Bai et al. (1992) studied L-estimates and M-estimation for linear models, respectively, Bradic et al. (2011) investigated model selection for sparse linear models with non-polynomial dimensionality, and Jiang et al. (2012) considered a similar problem for non-linear models with a diverging number of parameters. However, these works focused on i.i.d. data. Our current study concentrates on the non-linear time series model in (2.1) and (2.2).

For $\omega_i = 1/\sqrt{K}$, the above method can be regarded as an extension of the CQR estimation to the DTARCH model. Typically, we can use the equally spaced quantiles: $\tau_k = k/(K+1)$ for $k = 1, 2, \ldots, K$. The weight ω_k controls the amount of contribution of the τ_kth QR. In order to derive the asymptotic property of the estimator, we introduce some notations and conditions. Let $\boldsymbol{\beta}^*$ be the true value of $\boldsymbol{\beta}$. Denote by $b_\tau^* = F^{-1}(\tau)$, $h_t = h_t(\boldsymbol{\beta}^*)$, and $\boldsymbol{\mu} = E[\mathbf{Z}_t h_t^{-1}]$.

ASSUMPTION 3.1. *$E|y_t|^{2+\delta} < +\infty$ for some $\delta > 0$, and y_t is strictly stationary and ergodic.*

ASSUMPTION 3.2. *$\boldsymbol{\Gamma} = var(\mathbf{Z}_t h_t^{-1})$ is positive definitive.*

ASSUMPTION 3.3. *The density of u_t, $f(u)$, is positive and continuous at b_τ^*.*

ASSUMPTION 3.4. *The error e_t has an unknown distribution function $G(\cdot)$ with density $g(\cdot)$ being positive and continuous at its τ_kth quantiles $c_{\tau_k}^*$.*

Assumption 3.1 is used to ensure the ergodicity of $\mathbf{Z}_t$ so that the ergodic theorem for $\mathbf{Z}_t h_t^{-1}$ holds, that is

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=s+1}^{n} \mathbf{Z}_t h_t^{-1} \xrightarrow{p} E[\mathbf{Z}_t h_t^{-1}].$$

Assumption 3.2 is used to ensure non-singularity of the asymptotic variance–covariance matrix of the proposed estimators. Assumptions 3.3 and 3.4 are required for the availability of observations to run the τth QR in (3.3) and (3.5), respectively.

THEOREM 3.1. *Suppose that the threshold and delay parameters are known. Under Assumptions 3.1, 3.2 and 3.4, with probability tending to one, there is a local minimizer $\hat{\boldsymbol{\beta}}_1$ in (3.5) such that*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}^*) = \Big(\boldsymbol{\Gamma}\sum_{k=1}^{K}\omega_k g(c^*_{\tau_k})\Big)^{-1}\mathbf{z}_n + o_p(1),$$

where

$$z_n = n^{-1/2}\sum_{k=1}^{K}\omega_k\sum_{t=s'+1}^{n}(\mathbf{Z}_t h_t^{-1} - \boldsymbol{\mu})\psi_{\tau_k}(e_t - c^*_{\tau_k}).$$

Furthermore,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \sigma^2(\boldsymbol{\omega})\boldsymbol{\Gamma}^{-1}),$$

where

$$\sigma^2(\boldsymbol{\omega}) = \frac{\sum_{k,k'=1}^{K}\omega_k\omega_{k'}\min(\tau_k, \tau_{k'})(1 - \max(\tau_k, \tau_{k'}))}{\big(\sum_{k=1}^{K}\omega_k g(c^*_{\tau_k})\big)^2}.$$

Compared with $\hat{\boldsymbol{\beta}}_0$, $\hat{\boldsymbol{\beta}}_1$ is asymptotically unbiased. This confirms that the logarithm transformation of $|\varepsilon_t|$ in (3.4) is reasonable.

3.2. Choice of weights

Because $\boldsymbol{\Gamma}$ does not involve $\boldsymbol{\omega}$, the weights should be selected to minimize $\sigma^2(\boldsymbol{\omega})$. Let $\boldsymbol{g} = (g(c^*_{\tau_1}), \ldots, g(c^*_{\tau_K}))^T$, and let $\boldsymbol{A}$ be a $K \times K$ matrix with the (k, k') element being $A_{kk'} = \min(\tau_k, \tau_{k'})(1 - \max(\tau_k, \tau_{k'}))$. Then, the optimal weight $\boldsymbol{\omega}_{\text{opt}}$, which minimizes $\sigma(\boldsymbol{\omega})$, can be shown as

$$\boldsymbol{\omega}_{\text{opt}} = (\mathbf{g}'\mathbf{A}^{-2}\mathbf{g})^{-1/2}\mathbf{A}^{-1}\mathbf{g},$$

under the condition of $\|\boldsymbol{\omega}\| = 1$. The optimal weight components can be very different and some might even be negative. In fact, in our simulations, we also experience such a scenario. This reflects the necessity to use a data-driven weighting scheme. The density function $g(\cdot)$ of e_t can be estimated by running the kernel smoother over residuals from the CQR or L_1-estimator, and the τ_kth quantile $c^*_{\tau_k}$ of e_t can be estimated by the empirical estimate $\hat{c}_{\tau_k}$ from the residuals. Let the resulting estimate of $\mathbf{g}$ be $\hat{\mathbf{g}} = (\hat{g}(\hat{c}^*_{\tau_1}), \ldots, \hat{g}(\hat{c}^*_{\tau_k}))^T$. Then, $\hat{\boldsymbol{\omega}} = (\hat{\mathbf{g}}'\mathbf{A}^{-2}\hat{\mathbf{g}})^{-1/2}\mathbf{A}^{-1}\hat{\mathbf{g}}$ provides a non-parametric estimator of $\boldsymbol{\omega}_{\text{opt}}$ such that $\hat{\omega}_k = \omega_{k,\text{opt}}(1 + o_p(1))$ uniformly for $k = 1, \ldots, K$. This leads to an adaptive estimator of $\boldsymbol{\beta}$ by minimizing

$$\sum_{k=1}^{K}\hat{\omega}_k\sum_{t=s+1}^{n}\rho_{\tau_k}(\log|\varepsilon_t| - \log(h_t(\boldsymbol{\beta})) - c_{\tau_k}) \tag{3.6}$$

over c_{τ_k} and $\boldsymbol{\beta}$. Let the resulting estimator of $\boldsymbol{\beta}$ be $\tilde{\boldsymbol{\beta}}_1$. Then, $\tilde{\boldsymbol{\beta}}_1$ is asymptotically normal from the following theorem.

THEOREM 3.2. *Under Assumptions 3.1, 3.2 and 3.4, with probability tending to one, there exists a local minimizer* $\tilde{\boldsymbol{\beta}}_1$ *such that*

$$\sqrt{n}(\tilde{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}^*) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\mathbf{g}'\mathbf{A}^{-1}\mathbf{g})^{-1}\boldsymbol{\Gamma}^{-1}).$$

Because $\sigma^2(\boldsymbol{\omega}_{\text{opt}}) = (\mathbf{g}'\mathbf{A}^{-1}\mathbf{g})^{-1}$, $\tilde{\boldsymbol{\beta}}_1$ has the same asymptotic variance–covariance matrix as $\hat{\boldsymbol{\beta}}_1$ if $\boldsymbol{\omega}_{\text{opt}}$ were known. That is, the estimator $\tilde{\boldsymbol{\beta}}_1$ is adaptive. Therefore, $\hat{\boldsymbol{\omega}}$ is called the adaptive weight vector. In practice, the innovation distribution is unknown. Suppose it is known in advance. Then, the MLE of $\boldsymbol{\beta}$ can be obtained. We call it the oracle MLE. It is easy to show that the oracle MLE has asymptotic variance matrix $I_g^{-1}\boldsymbol{\Gamma}^{-1}$, where $I_g = \int (g'(t))^2/g(t)\,dt$ is the Fisher information. The asymptotic relative efficiency (ARE) of $\tilde{\boldsymbol{\beta}}_1$ with respect to the MLE is

$$e(\tilde{\boldsymbol{\beta}}_1, \text{MLE}) = I_g^{-1}\mathbf{g}'\mathbf{A}^{-1}\mathbf{g}.$$

Assume that the derivative $g'(\cdot)$ of $g(\cdot)$ is uniformly continuous. Then, for equally spaced $\{\tau_k\}_{k=1}^K$ over (0, 1), it can be shown that (see Jiang et al., 2012)

$$\lim_{K\to\infty} e(\tilde{\boldsymbol{\beta}}_1, \text{MLE}) = 1, \tag{3.7}$$

which means that $\tilde{\boldsymbol{\beta}}_1$ is nearly as efficient as the oracle MLE for various error distributions, including the normal, mixed-normal and Student-T distributions. This is a great advantage of the proposed methodology. Note that the error distribution is not specified in the WCQR approach, but the oracle MLE is applicable only when the distribution is known. As noted by the associate editor, it is interesting and challenging to investigate if (3.7) holds when K is replaced by an appropriate K_n depending on n. We assume that it is possible if the idea of regularization estimation is used.

The optimal ω_{opt} and its estimates in (3.6) might be negative, because it is obtained under the condition $||\omega|| = 1$. As discussed above, negative weights are used to improve efficiency; for details, see Jiang et al. (2012). Because $\hat{\omega}$ can be calculated and is a consistent estimator of ω_{opt}, with probability tending to one, from the value of $\hat{\omega}$ we know if ω_{opt} is non-negative. The objective function is not convex and hence there can exist multiple minimizers, which is the price we have to pay for using negative weights. This problem was alleviated, because we used a root-n consistent estimate as an initial estimate for solving the minimization problem, which leads to root-n consistent one-step and full-iteration estimators. In our experience, we always obtained a unique minimizer close to the true parameter. It can be shown (though it is very tedious) that, under more restrictions, our minimizer is the global one.

3.3. *Estimation of the DTARCH model with the AR part*

When there is an AR part in the model, a typical way of estimation is to estimate the ARCH parameters based on residuals. This method was studied for ARCH models by Engle (1982) and Koenker and Zhao (1996) in two steps. In the first step, they estimated the autoregressive parameters and computed the residuals. In the second step, they estimated the ARCH parameters by regressing the (squared) residuals on the lagged (squared) residuals. However, the residual-based estimate of $\boldsymbol{\beta}$ generally depends on the initial estimate $\boldsymbol{\alpha}_n$ of $\boldsymbol{\alpha}$. In general, $\boldsymbol{\alpha}_n$ inflates the asymptotic variance of the estimator of $\boldsymbol{\beta}$ in the second step, which is not a desired property (see Koenker and Zhao, 1996). This motivates us to study the simultaneous estimation of the model parameters based on the WCQR. Our methodology is also useful for deriving mathematical properties of the simultaneous estimation for ARCH models in Koenker and Zhao (1996) and others, where the asymptotic properties of the simultaneous estimation are still an open problem.

We rewrite the DTARCH model in (2.1) and (2.2) as

$$y_t = \mathbf{X}'_{t,j}\boldsymbol{\alpha}^{(j)} + h_t(\boldsymbol{\alpha},\boldsymbol{\beta})u_t, \quad \text{if} \quad r_{j-1} < y_{t-d} \le r_j,$$

which is equivalent to

$$y_t = \mathbf{X}'_t\boldsymbol{\alpha} + h_t(\boldsymbol{\alpha},\boldsymbol{\beta})u_t, \tag{3.8}$$

where $\boldsymbol{\alpha} = \mathrm{vec}(\boldsymbol{\alpha}^{(1)}, \ldots, \boldsymbol{\alpha}^{(m)})$, $\mathbf{X}_t = \mathrm{vec}(I_{t,1}\mathbf{X}_{t,1}, \ldots, I_{t,m}\mathbf{X}_{t,m})$, and

$$h_t(\boldsymbol{\alpha},\boldsymbol{\beta}) = \sum_{j=1}^{m} I_{t,j}\big(\beta_0^{(j)} + \beta_1^{(j)}|\varepsilon_{t-1}(\boldsymbol{\alpha})| + \cdots + \beta_{q_j}^{(j)}|\varepsilon_{t-q_j}(\boldsymbol{\alpha})|\big).$$

Let $\varepsilon_t(\boldsymbol{\alpha}) = y_t - \mathbf{X}'_t\boldsymbol{\alpha}$. From model (3.8), the τ_kth quantile of $\log|\varepsilon_t(\boldsymbol{\alpha})|$ given $\mathcal{F}_{t-1}$ is $Q_{\tau_k}(\log|\varepsilon_t(\boldsymbol{\alpha})|\,\big|\mathcal{F}_{t-1}) = \log(h_t(\boldsymbol{\alpha},\boldsymbol{\beta})) + c_{\tau_k}$. Applying the WCQR scheme, we can simultaneously estimate the AR and ARCH parameters by minimizing

$$\sum_{k=1}^{K}\omega_k \sum_{t=s'+1}^{n} \rho_{\tau_k}(\log|\varepsilon_t(\boldsymbol{\alpha})| - \log(h_t(\boldsymbol{\alpha},\boldsymbol{\beta})) - c_{\tau_k}) \tag{3.9}$$

over c_{τ_k}, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

This minimization problem has no closed-form solution, but it can easily be solved numerically. Minimization problems (3.4) and (3.5) are special cases of (3.9). We can use the following algorithm to solve (3.9).

STEP 1. Given a consistent estimate of $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_0$, for computing the pseudo-data $\varepsilon_t(\boldsymbol{\alpha}_0)$, use the extended interior point algorithm (see Jiang et al., 2012) to find the solution $(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1)$ to

$$\min_{\alpha,\beta,c_{\tau_k}} \sum_{k=1}^{K}\omega_k \sum_{t=s'+1}^{n} \rho_{\tau_k}(\log|\varepsilon_t(\boldsymbol{\alpha}_0)| - \log(h_t(\boldsymbol{\alpha},\boldsymbol{\beta})) - c_{\tau_k}).$$

STEP 2. Use $\boldsymbol{\alpha}_1$ to replace the above $\boldsymbol{\alpha}_0$, and then solve the above optimization problem to obtain an updated estimate of $\boldsymbol{\alpha}$. Repeat this procedure until convergence.

The above minimization involves two iteration loops: one is for the the extended interior point algorithm in Step 1 and the other is in Step 2, which is still a heavy burden in computation. In addition, there is no guarantee for convergence, because the objective function is non-linear and non-convex. To obtain a convergent solution, in the following we first propose one-step simultaneous estimators of the parameters and then iteratively update the one-step estimator until convergence.

With an initially consistent estimator $\boldsymbol{\alpha}_n$ of $\boldsymbol{\alpha}$, applying the WCQR scheme, we can estimate the AR and ARCH parameters simultaneously by minimizing

$$\sum_{k=1}^{K}\omega_k \sum_{t=s'+1}^{n} \rho_{\tau_k}(\log|\varepsilon_t(\boldsymbol{\alpha}_n)| - \log(h_t(\boldsymbol{\alpha},\boldsymbol{\beta})) - c_{\tau_k}) \tag{3.10}$$

over c_{τ_k}, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

We use $(\hat{\boldsymbol{\alpha}}^{(1)}, \hat{\boldsymbol{\beta}}^{(1)})$ to denote the resulting estimator. Because only Step 1 in the algorithm is involved in the estimation, we call the above estimation approach the one-step simultaneous

WCQR estimation. This is different from the existing one-step estimation in Bickel (1975) and Fan and Jiang (2000), where the estimators are explicitly defined and have closed forms. There are many choices for the initial estimator $\boldsymbol{\alpha}_n$ with root-n consistency. For robustness, we use the least absolute deviation (LAD) estimator.

Let

$$\mathbf{B}_{t,j} = (\mathbf{0}, \mathbf{X}_{t-1}\mathrm{sgn}(\varepsilon_{t-1}), \ldots, \mathbf{X}_{t-q_j}\mathrm{sgn}(\varepsilon_{t-q_j})), \quad \mathbf{B}_t' = (I_{t,1}\mathbf{B}_{t,1}, \ldots, I_{t,m}\mathbf{B}_{t,m}),$$

$$\mathbf{w_t} = (-\{\mathbf{B_t'}\boldsymbol{\beta}^*\mathbf{h_t^{-1}}\}', \mathbf{Z_t'h_t^{-1}})', \quad \boldsymbol{\eta}_\mathbf{n} = \mathbf{n}^{-(1/2)} \sum_{\mathbf{k=1}}^{\mathbf{K}} \omega_\mathbf{k} \sum_{\mathbf{t=s'+1}}^{\mathbf{n}} (\mathbf{w_t} - \mathbf{E}[\mathbf{w_t}])\psi_{\tau_\mathbf{k}}(\mathbf{e_t} - \mathbf{c^*}_{\tau_\mathbf{k}}),$$

$$\mathbf{D} = \mathrm{Cov}(\mathbf{w}_t, \mathbf{X}_t h_t^{-1}), \quad \boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{w}_t) \quad \text{and} \quad \tilde{f}(c^*_{\tau_k}) = f(e^{c^*_{\tau_k}}) - f(-e^{c^*_{\tau_k}}).$$

Then, it is easy to see that $\boldsymbol{\eta}_n$ is asymptotically normal with mean zero and covariance matrix $\boldsymbol{\omega}'\mathbf{A}\boldsymbol{\omega}\boldsymbol{\Sigma}$. Similar to Assumption 3.2, we need the following assumption.

ASSUMPTION 3.2′. *$\boldsymbol{\Sigma}$ is positive definitive.*

The following result establishes the relationship between the one-step simultaneous estimators and the initial estimator.

THEOREM 3.3. *Under Assumptions 3.1, 3.2′ and 3.4, if the initial estimator $\boldsymbol{\alpha}_n$ satisfies $\sqrt{n}(\boldsymbol{\alpha}_n - \boldsymbol{\alpha}^*) = O_p(1)$, then with probability tending to one there is a local minimizer $(\hat{\boldsymbol{\alpha}}^{(1)}, \hat{\boldsymbol{\beta}}^{(1)})$ in (3.10) such that*

$$\sqrt{n}((\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha}^*)', (\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^*)')' = \Big(\sum_{k=1}^{K} \omega_k g(c^*_{\tau_k})\Big)^{-1} \boldsymbol{\Sigma}^{-1} \times \Big(\boldsymbol{\eta}_n - \sum_{k=1}^{K} \omega_k \tilde{f}(c^*_{\tau_k})\mathbf{D}\sqrt{n}(\boldsymbol{\alpha}_n - \boldsymbol{\alpha}^*)\Big) + o_p(1).$$

COROLLARY 3.1. *Under Assumptions 3.1, 3.2′ and 3.4, if the innovation density $f(u)$ of u_t is symmetrical about zero, then $\sqrt{n}(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^*)$ is asymptotically normal with mean zero and variance matrix being $\sigma^2(\boldsymbol{\omega})\boldsymbol{\Gamma}^{-1}$.*

From Corollary 3.1, we can see that, when the innovation u_t is symmetrical, $\hat{\boldsymbol{\beta}}^{(1)}$ shares the same asymptotic variance as $\hat{\boldsymbol{\beta}}_1$ in Theorem 3.1, which means $\boldsymbol{\beta}$ can be estimated as well as if the true value of $\boldsymbol{\alpha}$ were known. In general, the asymptotic behaviour of $(\hat{\boldsymbol{\alpha}}^{(1)}, \hat{\boldsymbol{\beta}}^{(1)})$ depends on the initial estimator $\boldsymbol{\alpha}_n$. Using the one-step estimator as the initial estimator, we can obtain a two-step estimator $(\hat{\boldsymbol{\alpha}}^{(2)}, \hat{\boldsymbol{\beta}}^{(2)})$, which is also $\sqrt{n}$-consistent. We continue to use this procedure until convergence and obtain the final fully iterative estimator as a solution to the problem (3.9). We call it the fully iterative simultaneous WCQR estimator and denote it by $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$. Because, at each step of the iterations, the estimator is $\sqrt{n}$-consistent and is very close to the true value, it usually takes a few steps to achieve convergence.

Let $\boldsymbol{\Sigma}_k = g(c^*_{\tau_k})\boldsymbol{\Sigma} + \tilde{f}(c^*_{\tau_k})(\mathbf{D}, \mathbf{0})$ and $\mathbf{B}_k = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}_k$. The following theorem reports the joint asymptotic property of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$.

THEOREM 3.4. *Under Assumptions 3.1, 3.2′ and 3.4, the fully iterative estimator* $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ *satisfies that*

$$\sqrt{n}((\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)', (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)')' = \Big(\sum_{k=1}^{K} \omega_k \boldsymbol{\Sigma}_k \Big)^{-1} \boldsymbol{\eta}_n + o_p(1), \tag{3.11}$$

which is jointly asymptotically normal with mean zero and covariance matrix

$$\boldsymbol{\omega}^T \mathbf{A} \boldsymbol{\omega} \Big(\sum_{k=1}^{K} \sum_{k'=1}^{K} \omega_k \omega_{k'} \mathbf{B}_k^T \mathbf{B}_{k'} \Big)^{-1}.$$

Minimizing the trace of the above covariance matrix, under the condition of $\|\boldsymbol{\omega}\| = 1$, leads to an optimal weight. Theorem 3.4 demonstrates that the fully iterative estimator asymptotically does not depend on the initial estimator. This is its advantage over the one-step estimation. By Theorems 3.3 and 3.4, the one-step estimator is asymptotically equivalent to the full iterative estimator if $f(u)$ is symmetrical.

The proposed WCQR estimators are $\sqrt{n}$-consistent and asymptotically normal under Assumption 3.4, which enables efficient and robust inference for the model parameters. In particular, the tail-weight of the innovation is irrelevant because there is no condition on the moments of the innovation beyond $E|y_t|^2 < \infty$. In contrast, the asymptotic normality for the MLE in ARCH models requires a finite fourth moment.

3.4. *Model identification*

In practice, implementation of the WCQR estimation involves the choice of threshold r_j, delay parameter d and the orders (p, q) of the DTARCH models. This was noted by the associate editor and anonymous referees of this paper. Here, we extend the generalized Akaike information criterion (GAIC) and the generalized Bayesian information criterion GBIC) methods; see Section 4.1.2 in Fan and Yao (2003). For simplicity, we set $d \leq d_{\max} \equiv \max_{1 \leq j \leq m} \max(p_j, q_j)$. The threshold r_j are searched within certain inner sample ranges.

For the partition $\{A_j\} = \{-\infty = r_0 < r_1 < \cdots < r_m = \infty\}$ in (2.1), theoretically we can determine the partition in the manner of an exhausting search as follows. For a given collection of partitions $\{A_j\}$, let the minimum value of (3.7) be

$$L(A_j, \hat{\boldsymbol{\alpha}}^{(j)}, \hat{\boldsymbol{\beta}}^{(j)}) = \sum_{k=1}^{K} \omega_k \sum_{t=s'+1}^{n} \rho_{\tau_k}\Big(\log |\varepsilon_t(\hat{\boldsymbol{\alpha}}^{(j)})| - \log \big(h_t(\hat{\boldsymbol{\alpha}}^{(j)}, \hat{\boldsymbol{\beta}}^{(j)}) \big) - c_{\tau_k} \Big),$$

where $\hat{\boldsymbol{\alpha}}^{(j)}$ and $\hat{\boldsymbol{\beta}}^{(j)}$ are the WCQR estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively. We search for the partition $\{\hat{A}_j\}$ that minimizes $L(\{A_j\})$.

For the choice of the delay parameters d and the orders p, q, we take a two-step procedure. In the first step, for each fixed $d \in \{1, 2, \ldots, d_{\max}\}$, we minimize the WCQR objection function in (3.7) and then choose d so that the objective function has the smallest value. In the second step, we choose (p, q) by the GAIC/GBIC criterion, where GAIC and GBIC are defined as

$$\text{GAIC}(p, q) = (n - s') \log(\hat{\sigma}^2) + 4(p + q + 1)$$

and

$$\mathrm{GBIC}(p, q) = (n - s') \log(\hat{\sigma}^2) + 2(p + q + 1) \log(n - s'),$$

with

$$\hat{\sigma}^2 = \frac{1}{n - s'} \sum_{s'+1}^{n} \left(\log(|\hat{\varepsilon}_t^2|) - \log(h_t(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})) \right)^2.$$

4. NUMERICAL RESULTS

4.1. *Real data analysis*

We study the daily S&P 500 Composite index from 3 January 2000 to 27 July 2011. This index represents the bulk of the daily value in the US equity market. The return series y_t is defined as the difference of the log-price. We are interested in the asymmetry of the conditional mean and conditional variance.

The proposed WCQR estimation approach with $K = 7$ and equally spaced τ_k on (0, 1) is applied. Our experience has shown that increasing K does not change the results significantly in the current settings. To fit the return series with 2908 observations, we need to identify the values of d, m and r_j. Because m usually takes a small value in practice, we set $m = 2, 3, 4$ in this example. Using the model identification method in Section 3.4, we identify $d = 1$ and $m = 2$ with $-\infty = r_0 < r_1 = 0.0 < r_2 = \infty$. This is consistent with the observations in the stock market and agrees with our aim to investigate the asymmetry of the conditional mean and variance. Therefore, we consider the following DTARCH model for the return series y_t

$$y_t = \begin{cases} \alpha_1^{(1)} y_{t-1} + \alpha_2^{(1)} y_{t-2} + \cdots + \alpha_p^{(1)} y_{t-p} + \varepsilon_t, & \text{if } \ y_{t-1} \le 0, \\ \alpha_1^{(2)} y_{t-1} + \alpha_2^{(2)} y_{t-2} + \cdots + \alpha_p^{(2)} y_{t-p} + \varepsilon_t, & \text{if } \ y_{t-1} > 0, \end{cases}$$

where $\varepsilon_t = h_t u_t$, and

$$h_t = \begin{cases} \beta_0^{(1)} + \beta_1^{(1)} |\varepsilon_{t-1}| + \beta_2^{(1)} |\varepsilon_{t-2}| + \cdots + \beta_q^{(1)} |\varepsilon_{t-q}|, & \text{if } \ y_{t-1} \le 0, \\ \beta_0^{(2)} + \beta_1^{(2)} |\varepsilon_{t-1}| + \beta_2^{(2)} |\varepsilon_{t-2}| + \cdots + \beta_q^{(2)} |\varepsilon_{t-q}|, & \text{if } \ y_{t-1} > 0. \end{cases}$$

We set $p, q = 1, \ldots, 7$ in DTARCH models and search the values of (p, q) in the range. When $p = 2$ and $q = 4$, both the GAIC and GBIC statistics achieve the minimum values at 0.2294 and 0.2582. This identifies a DTARCH(2,4) model for the return series.

For comparison, we calculate the WCQR estimate with equal weights ω_k (labelled as CQR estimate) and QR estimates (i.e., $K = 1$) at $\tau_K = 0.5$ and 0.75. For each of the fully iterative and one-step QR, CQR and WCQR estimation approaches, we calculate the estimated parameters and their standard errors. For $\tau_K = 0.5$, it corresponds to the L_1-estimate. Because $\beta_1^{(2)}$ is not significant at the 1% level, we re-estimate the model with $\beta_1^{(2)} = 0$. Tables 1 and 2 report the estimates with estimated standard deviations. All coefficients are now significant at the 1% level. Because the L_1-estimation performs badly, we omit it here. It is obvious that the fully iterative simultaneous estimation is best because of its nearly smallest standard errors, and the one-step simultaneous estimation is comparable to the fully iterative simultaneous estimation. Negative

Table 1. Estimates of parameters.

Method	OS-QR ($K = 1$, $\tau_K = 0.75$)	OS-CQR	OS-WCQR
$\alpha_1^{(1)}$	−0.0995 (0.0011)	−0.1144 (0.0004)	−0.1039 (0.0002)
$\alpha_2^{(1)}$	−0.0879 (0.0014)	−0.0853 (0.0004)	−0.0875 (0.0004)
$\alpha_1^{(2)}$	−0.0584 (0.0004)	−0.0352 (0.0001)	−0.0552 (0.0001)
$\alpha_2^{(2)}$	−0.0320 (0.0001)	−0.0290 (0.0004)	−0.0317 (0.0001)
$\beta_0^{(1)}$	0.0025 (0.0005)	0.0025 (0.0005)	0.0026 (0.0004)
$\beta_1^{(1)}$	0.1376 (0.0370)	0.1388 (0.0350)	0.1393 (0.0314)
$\beta_2^{(1)}$	0.2326 (0.0397)	0.2436 (0.0376)	0.2441 (0.0338)
$\beta_3^{(1)}$	0.1713 (0.0385)	0.1719 (0.0363)	0.1724 (0.0327)
$\beta_4^{(1)}$	0.2287 (0.0397)	0.2292 (0.0374)	0.2302 (0.0337)
$\beta_0^{(2)}$	0.0037 (0.0006)	0.0037 (0.0005)	0.0038 (0.0004)
$\beta_1^{(2)}$	0 (–)	0 (–)	0 (–)
$\beta_2^{(2)}$	0.2232 (0.0358)	0.2259 (0.0336)	0.2251 (0.0304)
$\beta_3^{(2)}$	0.1979 (0.0362)	0.1958 (0.0340)	0.1994 (0.0308)
$\beta_4^{(2)}$	0.1346 (0.0377)	0.1329 (0.0353)	0.1355 (0.0320)

Note: Estimated standard errors are given in parentheses. OS denotes one-step.

Table 2. Estimates of parameters.

Method	FI-QR ($K = 1$, $\tau_K = 0.75$)	FI-CQR	FI-WCQR
$\alpha_1^{(1)}$	−0.0992 (0.0013)	−0.1174 (0.0006)	−0.1018 (0.0004)
$\alpha_2^{(1)}$	−0.0881 (0.0015)	−0.0848 (0.0011)	−0.0880 (0.0004)
$\alpha_1^{(2)}$	−0.0583 (0.0004)	−0.0302 (0.0003)	−0.0587 (0.0002)
$\alpha_2^{(2)}$	−0.0319 (0.0001)	−0.0284 (0.0002)	−0.0316 (0.0001)
$\beta_0^{(1)}$	0.0025 (0.0005)	0.0025 (0.0005)	0.0026 (0.0004)
$\beta_1^{(1)}$	0.1373 (0.0369)	0.1386 (0.0350)	0.1395 (0.0315)
$\beta_2^{(1)}$	0.2421 (0.0396)	0.2434 (0.0376)	0.2443 (0.0338)
$\beta_3^{(1)}$	0.1709 (0.0384)	0.1716 (0.0362)	0.1726 (0.0327)
$\beta_4^{(1)}$	0.2282 (0.0395)	0.2288 (0.0374)	0.2305 (0.0337)
$\beta_0^{(2)}$	0.0037 (0.0006)	0.0037 (0.0005)	0.0038 (0.0005)
$\beta_1^{(2)}$	0 (–)	0 (–)	0 (–)
$\beta_2^{(2)}$	0.2227 (0.0338)	0.2263 (0.0318)	0.2249 (0.0291)
$\beta_3^{(2)}$	0.1976 (0.0334)	0.1944 (0.0315)	0.2002 (0.0287)
$\beta_4^{(2)}$	0.1342 (0.0344)	0.1326 (0.0324)	0.1359 (0.0296)

Note: Estimated standard errors are given in parentheses. FI denotes fully iterative.

AR-part parameters indicate that the future mean return will be forecasted as positive (negative) if the current and its past return are negative (positive). This is expected in an efficient market. Figure 1 displays the estimated volatility $\hat{h}_t$ where blue circles are used to denote negative return and red crosses denote positive return. Overall, this indicates that the clustered volatility is higher

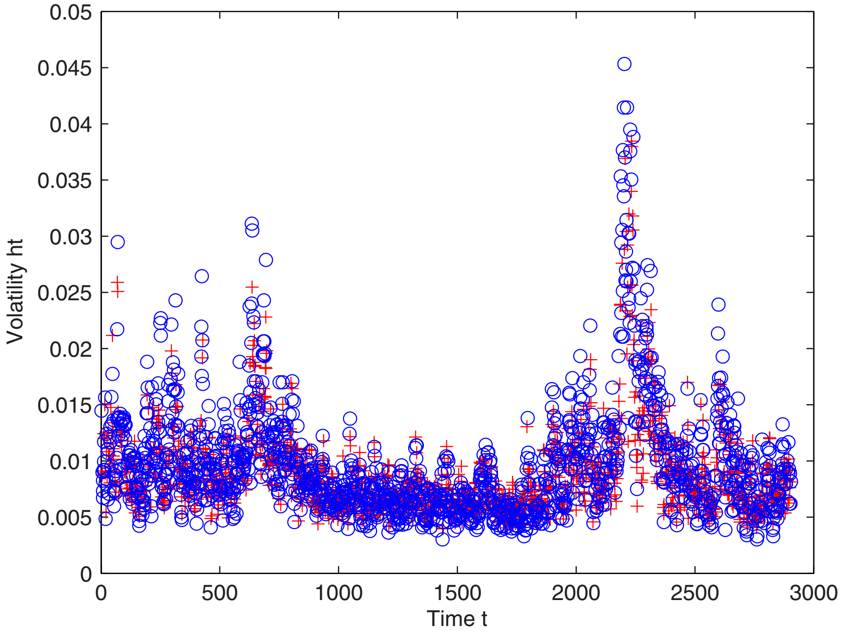

Figure 1. Estimated volatility.

when prices are falling. That is, volatility tends to be higher in bear markets, an asymmetric volatility effect described by the EGARCH model of Nelson (1991).

4.2. Simulation results

In this section, simulations are conducted to investigate the advantages of the WCQR estimation. Data are sampled from the following DTARCH model:

$$y_t = \begin{cases} \alpha_1^{(1)} y_{t-1} + \varepsilon_t & \text{if } y_{t-1} \le 0, \\ \alpha_1^{(2)} y_{t-1} + \varepsilon_t & \text{if } y_{t-1} > 0. \end{cases}$$

Here, $(\alpha_1^{(1)}, \alpha_1^{(2)}) = (0.20, 0.35)$, and $\varepsilon_t = h_t u_t$, with

$$h_t = \begin{cases} \beta_0^{(1)} + \beta_1^{(1)} |\varepsilon_{t-1}| & \text{if } y_{t-1} \le 0, \\ \beta_0^{(2)} + \beta_1^{(2)} |\varepsilon_{t-1}| & \text{if } y_{t-1} > 0, \end{cases}$$

where $\beta_0^{(1)} = 0.02$, $\beta_1^{(1)} = 0.3$ and $(\beta_0^{(2)}, \beta_1^{(2)}) = (0.04, 0.25)$.

We employ four sets of innovation variables for u_t, $N(0, 1)$, $0.9N(0, 1) + 0.1N(0, 3^2)$, $t(5)$ and $\chi^2(4)$, which are centralized and normalized so that the medians of the absolute innovations are ones. We conduct 400 simulations. In each simulation, a sample of size $n = 1600$ is drawn. As before, we used $K = 7$ and equally spaced τ_k on $(0, 1)$.

Table 3. RMSE comparison: scaled normal innovation.

Estimate	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\beta_0^{(1)}$	$\beta_1^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(2)}$	Σ
OS-L_1	8.07	10.07	0.15	5.64	0.43	5.46	29.81
FI-L_1	7.95	9.83	0.16	6.03	0.46	5.54	29.97
OS-CQR	6.67	8.18	0.12	4.31	0.35	4.50	24.14
FI-CQR	6.60	8.10	0.12	4.37	0.34	4.56	24.09
OS-WCQR	3.89	6.25	0.10	3.39	0.26	3.16	17.05
FI-WCQR	3.89	6.25	0.10	3.39	0.26	3.17	17.06
OMLE	3.92	4.19	0.11	2.59	0.18	3.54	14.52

Notes: RMSE is multiplied by 10^2. OMLE denotes oracle MLE, OS denotes one-step and FI denotes fully iterative.

Table 4. RMSE comparison: scaled $t(5)$ innovation.

Estimate	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\beta_0^{(1)}$	$\beta_1^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(2)}$	Σ
OS-L_1	6.43	8.63	0.16	5.49	0.40	5.25	26.35
FI-L_1	6.58	8.62	0.16	5.58	0.40	5.25	26.60
OS-CQR	4.90	6.08	0.11	3.64	0.27	3.46	18.46
FI-CQR	4.84	6.18	0.11	3.65	0.28	3.47	18.54
OS-WCQR	4.51	5.63	0.12	3.99	0.29	3.47	18.01
FI-WCQR	4.51	5.62	0.13	3.99	0.29	3.46	18.01
OMLE	4.07	4.10	0.14	3.01	0.23	3.77	15.31

Notes: RMSE is multiplied by 10^2. OMLE denotes oracle MLE, OS denotes one-step and FI denotes fully iterative.

Table 5. RMSE comparison: scaled mixed normal innovation.

Estimate	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\beta_0^{(1)}$	$\beta_1^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(2)}$	Σ
OS-L_1	8.28	9.94	0.16	6.01	0.44	5.27	30.09
FI-L_1	7.80	9.75	0.16	6.10	0.45	5.40	29.67
OS-CQR	6.04	7.99	0.12	4.33	0.34	4.38	23.20
FI-CQR	6.01	7.86	0.12	4.39	0.34	4.38	23.10
OS-WCQR	3.85	5.83	0.09	3.31	0.25	3.22	16.55
FI-WCQR	3.85	5.83	0.09	3.31	0.25	3.23	16.57
OMLE	3.92	4.19	0.11	2.59	0.18	3.54	15.05

Notes: RMSE is multiplied by 10^2. OMLE denotes oracle MLE, OS denotes one-step and FI denotes fully iterative.

We compare the proposed approach with the oracle MLE of known innovations and several other methods. In each simulation, the root mean squared error (RMSE) for different estimators is calculated, and their averages over simulations are reported in Tables 3–6, where Σ denotes the sum of RMSE for all components in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Therefore, better estimators should have smaller Σ values. As expected, the oracle MLE performs the best, the WCQR performs better than the CQR and L_1, and L_1 is the worst. In terms of overall performance (the value of Σ), WCQR estimators uniformly dominate the L_1 and CQR estimators and are comparable to the oracle MLE

Table 6. RMSE comparison: scaled $\chi^2(4)$ innovation.

Estimate	$\alpha_1^{(1)}$	$\alpha_1^{(2)}$	$\beta_0^{(1)}$	$\beta_1^{(1)}$	$\beta_0^{(2)}$	$\beta_1^{(2)}$	Σ
OS-L_1	8.28	7.59	0.13	5.40	0.33	4.55	26.28
FI-L_1	9.17	6.69	0.13	5.86	0.33	5.32	27.50
OS-CQR	6.99	6.47	0.11	4.30	0.27	3.66	21.80
FI-CQR	7.56	5.97	0.11	4.79	0.26	4.09	22.78
OS-WCQR	4.28	4.45	0.09	3.19	0.18	2.53	14.72
FI-WCQR	4.42	4.35	0.09	3.30	0.18	2.67	15.01
OMLE	3.46	3.98	0.05	1.78	0.09	2.50	11.84

Notes: RMSE is multiplied by 10^2. OMLE denotes oracle MLE, OS denotes one-step and FI denotes fully iterative.

under all types of innovations considered. As we have observed, one-step estimators perform almost the same as fully iterative estimators. This is in accordance with our previous theoretical results. Overall, the WCQR estimators are not only robust but also efficient, which exemplifies the statement in (3.7) and the value of the proposed methodology.

5. CONCLUDING REMARKS

The WCQR estimation for the DTARCH model has been proposed to estimate the AR and ARCH parameters of the DTARCH model. Asymptotic properties of the proposed estimators have been established, which enriches the estimation theory of the DTARCH model. Theoretical and computational results all support our findings that the data-driven WCQR estimation uniformly dominates the L_1 and CQR estimations, and nearly reaches the efficiency of the oracle MLE with known innovations. The obtained results provide new insights into quantile regression.

The idea of WCQR can be extended to other models, such as non-parametric or semi-parametric regression models, and time-varying or functional coefficient models, with local WCQR modelling techniques. Such an extension will greatly enlarge the scope of application of the WCQR in financial and econometric models. Further topics can also include hypothesis testing based on the WCQR fitting.

ACKNOWLEDGEMENTS

The authors are grateful to an associate editor and the referees for their valuable comments and suggestions, which have improved the article substantially. This research was supported in part by GRF grant 403109 from the Research Grant Council of the Hong Kong Special Administration Region and NSFC grant 71361010. Additional partial support was also provided by the NSF Grant DMS-09-06482 and the FRG grant of the University of North Carolina at Charlotte for Jiancheng Jiang and by the NSF Grant 11101432 for Xuejun Jiang.

REFERENCES

Ait-Sahalia, Y. (1998). Dynamic equilibrium and volatility in financial asset markets. *Journal of Econometrics 84*, 93–127.

Ait-Sahalia, Y. and L. P. Hansen (2009). *Handbook of Financial Econometrics*. Amsterdam: North-Holland.
Bai, Z. D., C. R. Rao, and Y. Wu (1992). M-estimation of multivariate linear regression parameters under a convex discrepancy function. *Statistica Sinica 2*, 237–54.
Bera, A. K. and M. L. Higgins (1993). On ARCH models: properties, estimation and testing. *Journal of Economic Surveys 7*, 305–66.
Bickel, P. J. (1975). One-step Huber estimates in linear models. *Journal of the American Statistics Association 70*, 428–33.
Bickel, P. J. (1978). Tests for heteroscedasticity, nonlinearity. *Annals of Statistics 6*, 266–91.
Bickel, P. J. and E. L. Lehmann (1976). Descriptive statistics for non-parametric models. III. Dispersion. *Annals of Statistics 4*, 1139–58.
Bollerslev, T. (1986). ARCH modelling in finance. *Journal of Econometrics 31*, 307–27.
Bollerslev, T., R. Chou, and K. Kroner (1992). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 50*, 5–59.
Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics, Volume 4*, 2959–3038. Amsterdam: North-Holland.
Bradic, J., J. Fan, and W. Wang (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *Journal of the Royal Statistical Society*, Series B *73*, 325–49.
Carroll, R. J. and D. Ruppert (1988). *Transformations and Weighting in Regression*. London: Chapman and Hall.
Chaudhuri, P., K. Doksum, and A. Samarov (1997). On average derivative quantile regression. *Annals of Statistics 25*, 715–44.
Davis, R. A. and W. T. M. Dunsmuir (1997). Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability 10*, 481–97.
Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the UK inflation. *Econometrica 50*, 987–1008.
Fan, J. and J. Jiang (2000). Variable bandwidth and one-step local M-estimator. *Science in China*, Series A *43*, 65–81.
Fan, J. and Q. Yao (2003). *Nonlinear Time Series: Non-Parametric and Parametric Methods*. New York, NY: Springer.
Fan, J., L. Qi, and D. Xiu (2012). Quasi maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. Working paper, Princeton University.
Hall, P. and Q. Yao (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica 71*, 285–317.
Hui, Y. V. and J. Jiang (2005). Robust modelling of DTARCH models. *Econometrics Journal 8*, 143–58.
Jiang, J., Q. Zhao, and Y. V. Hui (2001). Robust modelling of ARCH models. *Journal of Forecasting 20*, 111–33.
Jiang, X., J. Jiang, and X. Song (2012). Oracle model selection for nonlinear models based on weighted composite quantile regression. *Statistica Sinica 22*, 1479–506.
Koenker, R. (1984). A note on L-estimates for linear models. *Statistics and Probability Letters 2*, 323–25.
Koenker, R. (2005). *Quantile Regression*. Cambridge: Cambridge University Press.
Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.
Koenker, R. and Q. Zhao (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory 12*, 793–813.
Li, C. W. and W. K. Li (1996). On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics 11*, 253–74.
Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous time case. *Review of Economics and Statistics 51*, 247–57.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous time model. *Journal of Economic Theory 3*, 373–413.

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics 4*, 141–83.

Nelson, D. (1991). Conditional heteroscedasticity in asset returns: a new approach. *Econometrica 59*, 347–70.

Peng, L. and Q. Yao (2003). Least absolute deviation estimation for ARCH and GARCH models. *Biometrika 90*, 967–75.

Sundaresan, S. M. (2000). Continuous-time methods in finance: a review and an assessment. *Journal of Finance 55*, 1569–622.

Tong, H. and K. S. Lim (1980). Threshold autoregressive, limit cycles and cyclical data for ARCH and GARCH models. *Journal of the Royal Statistical Society*, Series B *42*, 245–92.

Tsay, R. S. (1989). Testing and modelling threshold autoregressive processes for ARCH and GARCH models. *Journal of the American Statistical Association 84*, 231–40.

Zou, H. and M. Yuan (2008). Composite quantile regression and the oracle model selection theory for ARCH and GARCH models. *Annals of Statistics 3*, 1108–26.

APPENDIX A: PROOFS OF RESULTS

In the following, we give the proofs of Theorems 3.1, 3.3 and 3.4. Theorem 3.2 can be proven along the same lines as for Theorem 3.1. For convenience, we adopt the notations in Jiang et al. (2001). Let M be a generic positive number, and let $E_{t-1}[\cdot]$ and $\mathrm{Var}_{t-1}(\cdot)$ denote the expectation and variance conditional on $\mathcal{F}_{t-1}$. Denote by $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)'$, $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1', \boldsymbol{\Delta}_2')'$, $\tilde{\boldsymbol{\Delta}}_{1n} = \sqrt{n}(\boldsymbol{\alpha}_n - \boldsymbol{\alpha}^*)$, $\alpha(\boldsymbol{\Delta}_1) = \alpha^* + n^{-(1/2)}\boldsymbol{\Delta}_1$, $\boldsymbol{\beta}(\boldsymbol{\Delta}_2) = \boldsymbol{\beta}^* + n^{-(1/2)}\boldsymbol{\Delta}_2$ and $c_{\tau_k}(\nu_k) = c^*_{\tau_k} + n^{-1/2}\nu_k$. Let $\mathcal{C}_M = ((\boldsymbol{\nu}, \boldsymbol{\Delta}) : \|\boldsymbol{\nu}\| \le M, \|\boldsymbol{\Delta}\| \le M)$. Put $\varepsilon_t(\boldsymbol{\Delta}_1) = \varepsilon_t - n^{-(1/2)}\mathbf{X}_t'\boldsymbol{\Delta}_1$, $\hat{\varepsilon}_t = \varepsilon_t(\tilde{\boldsymbol{\Delta}}_{1n})$ and $h_t(\boldsymbol{\Delta}) = \mathbf{Z}_t'(\boldsymbol{\Delta}_1)\boldsymbol{\beta}(\boldsymbol{\Delta}_2)$, where

$$\mathbf{Z}_t(\boldsymbol{\Delta}_1) = \mathrm{vec}\left(I_{t,1}\mathbf{Z}_{t,1}(\boldsymbol{\Delta}_1), \ldots, I_{t,m}\mathbf{Z}_{t,m}(\boldsymbol{\Delta}_1)\right),$$
$$\mathbf{Z}_{t,j}(\boldsymbol{\Delta}_1) = \left(1, |\varepsilon_{t-1}(\boldsymbol{\Delta}_1)|, \ldots, |\varepsilon_{t-q_j}(\boldsymbol{\Delta}_1)|\right)'.$$

Without ambiguity, we use $h_t(\boldsymbol{\Delta}_2)$ to denote $h_t(\boldsymbol{\Delta})$ when $\boldsymbol{\Delta}_1 = 0$. Let

$$\mathbf{B}_t'(\boldsymbol{\Delta}_1) = (I_{t,1}\mathbf{B}_{t,1}(\boldsymbol{\Delta}_1), \ldots, I_{t,m}\mathbf{B}_{t,m}(\boldsymbol{\Delta}_1)),$$

where $\mathbf{B}_{t,j}(\boldsymbol{\Delta}_1) = (\mathbf{0}, \mathbf{X}_{t-1}\mathrm{sgn}(\varepsilon_{t-1}(\boldsymbol{\Delta}_1)), \ldots, \mathbf{X}_{t-q_j}\mathrm{sgn}(\varepsilon_{t-q_j}(\boldsymbol{\Delta}_1)))$.

The following technical lemmata are introduced to streamline the proofs of the theorems. To save space, we have relegated the proofs of lemmata to the supplementary material.

LEMMA A.1. *Let*

$$\mathbf{V_1}(\boldsymbol{\nu}, \boldsymbol{\Delta}_2) = n^{-(1/2)} \sum_{t=s'+1}^{n} (\omega_1\psi_{\tau_1}(\xi_t(\nu_1, \boldsymbol{\Delta}_2)), \ldots, \omega_K\psi_{\tau_K}(\xi_t(\nu_K, \boldsymbol{\Delta}_2)))',$$

$$\mathbf{V_2}(\boldsymbol{\nu}, \boldsymbol{\Delta}_2) = n^{-(1/2)} \sum_{k=1}^{K} \omega_k \sum_{t=s'+1}^{n} \mathbf{Z_t}\mathbf{h_t^{-1}}\psi_{\tau_\mathbf{k}}(\xi_\mathbf{t}(\nu_\mathbf{k}, \boldsymbol{\Delta}_\mathbf{2})),$$

$$\mathbf{V}(\boldsymbol{\nu}, \boldsymbol{\Delta}_2) = (\{\mathbf{V_1}(\boldsymbol{\nu}, \boldsymbol{\Delta}_2)\}', \{\mathbf{V_2}(\boldsymbol{\nu}, \boldsymbol{\Delta}_2)\}')',$$

where $\xi_t(\nu_k, \boldsymbol{\Delta}_2) = e_t - c^*_{\tau_k} - n^{-(1/2)}(\nu_k + \mathbf{Z_t'}\mathbf{h_t^{-1}}\boldsymbol{\Delta}_2)$. *Under Assumptions 3.1, 3.2 and 3.4, (a) the kth component* $V_{1k}(\boldsymbol{\nu}, \boldsymbol{\Delta}_2)$ *of* $\mathbf{V_1}(\boldsymbol{\nu}, \boldsymbol{\Delta}_2)$ *satisfies that*

$$\sup_{\mathcal{C}_M} \|V_{1k}(\boldsymbol{\nu}, \boldsymbol{\Delta}_2) - V_{1k}(\mathbf{0}) + \omega_k g(c^*_{\tau_k})(\nu_k + \boldsymbol{\mu}'\boldsymbol{\Delta}_2)\| = o_p(1),$$

where $V_{1k}(\mathbf{0}) = \omega_k q_{n,k}$ *and* $q_{n,k} = n^{-(1/2)} \sum_{t=s'+1}^{n} \psi_{\tau_k}(e_t - c^*_{\tau_k})$, *and (b)* $\mathbf{V}_2(\boldsymbol{\nu}, \boldsymbol{\Delta}_2)$ *satisfies that*

$$\sup_{\mathcal{C}_M} \left\| \mathbf{V}_2(\boldsymbol{\nu}, \boldsymbol{\Delta}_2) - \mathbf{V}_2(\mathbf{0}) + \sum_{k=1}^{K} \omega_k g(c^*_{\tau_k})(\boldsymbol{\mu}\nu_k + \mathbf{G}_2\boldsymbol{\Delta}_2) \right\| = o_p(1),$$

where

$$\mathbf{V}_2(\mathbf{0}) = \sum_{k=1}^{K} \omega_k \mathbf{z}_{n,k},$$

$$\mathbf{z}_{n,k} = n^{-(1/2)} \sum_{t=s'+1}^{n} \mathbf{Z}_t h_t^{-1} \psi_{\tau_k}(e_t - c^*_{\tau_k})$$

and

$$\mathbf{G}_2 = E[\mathbf{Z}_t \mathbf{Z}_t' h_t^{-2}].$$

LEMMA A.2. *Let* $\mathbf{c}_{t-1}(\boldsymbol{\Delta}) = c^*_{\tau_k} + n^{-(1/2)}(\nu_k - (\mathbf{B}_t'\boldsymbol{\beta}^*)' h_t^{-1}\boldsymbol{\Delta}_1 + \mathbf{Z}_t' h_t^{-1}\boldsymbol{\Delta}_2)$, $\tilde{\xi}_t(\nu_k, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) = e_t - \mathbf{c}_{t-1}(\boldsymbol{\Delta}) - n^{-(1/2)}\mathbf{X}_t'\varepsilon_t^{-1}\tilde{\boldsymbol{\Delta}}_{1n}$ *and* $\tilde{\mathbf{U}}(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) = ((\tilde{\mathbf{U}}_1(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}))', (\tilde{\mathbf{U}}_2(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}))')'$, *where* $\tilde{\mathbf{U}}_1(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) = n^{-(1/2)} \sum_{t=s'+1}^{n} (\omega_1 \psi_{\tau_1}(\tilde{\xi}_t(\nu_1, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n})), \ldots, \omega_K \psi_{\tau_K}(\tilde{\xi}_t\ (\nu_K, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n})))'$, *and* $\tilde{\mathbf{U}}_2(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) = n^{-(1/2)} \sum_{k=1}^{K} \omega_k \sum_{t=s'+1}^{n} \mathbf{w}_t \psi_{\tau_k}(\tilde{\xi}_t(\nu_k, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}))$. *Under Assumptions 3.1, 3.2 and 3.4, (a)* $\tilde{U}_{1k}(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n})$, *the kth component of* $\tilde{\mathbf{U}}_1(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n})$, *satisfies that*

$$\sup_{\mathcal{C}_M} \left\| \tilde{U}_{1k}(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) - \tilde{U}_{1k}(\mathbf{0}) + \omega_k g(c^*_{\tau_k})(\nu_k + \boldsymbol{\mu}'\boldsymbol{\Delta}_2) + \omega_k \phi(\boldsymbol{\Delta}_1, \tilde{\boldsymbol{\Delta}}_{1n}) \right\| = o_p(1),$$

where $\phi(\boldsymbol{\Delta}_1, \tilde{\boldsymbol{\Delta}}_{1n}) = \tilde{f}(c^*_{\tau_k}) E[\mathbf{X}_t' h_t^{-1}] \tilde{\boldsymbol{\Delta}}_{1n} - g(c^*_{\tau_k}) E[\mathbf{B}_t' \boldsymbol{\beta}^* h_t^{-1}] \boldsymbol{\Delta}_1$, $\tilde{U}_{1k}(\mathbf{0}) = \omega_k q_{n,k}$, *and* $\tilde{f}(c^*_{\tau_k}) = f(e^{c^*_{\tau_k}}) - f(-e^{c^*_{\tau_k}})$, *and (b)* $\tilde{\mathbf{U}}_2(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n})$ *satisfies that*

$$\sup_{\mathcal{C}_M} \|\tilde{\mathbf{U}}_2(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) - \tilde{\mathbf{U}}_2(\mathbf{0}) + \sum_{k=1}^{K} \omega_k g(c^*_{\tau_k}) \boldsymbol{\varphi}(\nu_k, \boldsymbol{\Delta}) + \sum_{k=1}^{K} \omega_k \Omega_k^* \tilde{\boldsymbol{\Delta}}_{1n}\| = o_p(1),$$

where $\tilde{\mathbf{U}}_2(\mathbf{0}) = \sum_{k=1}^{K} \omega_k \boldsymbol{\eta}_{n,k}$, $\boldsymbol{\eta}_{n,k} = n^{-(1/2)} \sum_{t=s'+1}^{n} \mathbf{w}_t \psi_{\tau_k}(e_t - c^*_{\tau_k})$, $\boldsymbol{\varphi}(\nu_k, \boldsymbol{\Delta}) = \nu_k E[\mathbf{w}_t] + E[\mathbf{w}_t \mathbf{w}_t']\boldsymbol{\Delta}$, *and* $\Omega_k^* = \tilde{f}(c^*_{\tau_k}) E[\mathbf{w}_t \mathbf{X}_t' h_t^{-1}]$.

LEMMA A.3. *Let* $\xi_t^0(\nu_k, \boldsymbol{\Delta}_2) = \log|\varepsilon_t| - \log(h_t(\boldsymbol{\Delta}_2)) - (c^*_{\tau_k} + n^{-(1/2)}\nu_k)$ *and* $\hat{\xi}_t(\nu_k, \boldsymbol{\Delta}) = \log|\hat{\varepsilon}_t| - \log(h_t(\boldsymbol{\Delta})) - (c^*_{\tau_k} + n^{-(1/2)}\nu_k)$. *Under Assumptions 3.1, 3.2 and 3.4, we have (a)* $n^{-1/2} \sum_{t=s'+1}^{n} (I(\xi_t(\nu_k, \boldsymbol{\Delta}_2) < 0) - I(\xi_t^0(\nu_k, \boldsymbol{\Delta}_2) < 0)) = o_p(1)$, *(b)* $n^{-1/2} \sum_{t=s'+1}^{n} \mathbf{Z}_t h_t^{-1}\ (\boldsymbol{\Delta}_2)(I(\xi_t(\nu_k, \boldsymbol{\Delta}_2) < 0) - I(\xi_t^0(\nu_k, \boldsymbol{\Delta}_2) < 0)) = o_p(1)$, *(c)* $n^{-1/2} \sum_{t=s'+1}^{n} (I(\hat{\xi}_t(\nu_k, \boldsymbol{\Delta}) < 0) - I(\tilde{\xi}_t(\nu_k, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) < 0)) = o_p(1)$, *and (d)* $n^{-1/2} \sum_{t=s'+1}^{n} \mathbf{w}_t (I(\hat{\xi}_t(\nu_k, \boldsymbol{\Delta}) < 0) - I(\tilde{\xi}_t(\nu_k, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) < 0)) = o_p(1)$, *uniformly for* $(\nu_k, \boldsymbol{\Delta}) \in \mathcal{C}_M$.

LEMMA A.4. *Let* $(\hat{\boldsymbol{\nu}}_n, \hat{\boldsymbol{\Delta}}_n)$ *be a minimizer of the objective function* $L_n(\mathbf{v}, \boldsymbol{\Delta})$ *in (3.10), where* $\hat{\nu}_{n,k} = \sqrt{n}(\hat{c}_{\tau_k} - c^*_{\tau_k})$, $\hat{\boldsymbol{\nu}}_n = (\hat{\nu}_{n,1}, \ldots, \hat{\nu}_{n,k})'$, $\hat{\boldsymbol{\Delta}}_{1n} = \sqrt{n}(\hat{\boldsymbol{\alpha}}^{(1)} - \boldsymbol{\alpha}^*)$, $\hat{\boldsymbol{\Delta}}_{2n} = \sqrt{n}(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^*)$, *and* $\hat{\boldsymbol{\Delta}}_n = (\hat{\boldsymbol{\Delta}}_{1n}', \hat{\boldsymbol{\Delta}}_{2n}')'$. *The objective function in (3.10) can be written as*

$$L_n(\boldsymbol{\nu}, \boldsymbol{\Delta}) = \sum_{k=1}^{K} \omega_k \sum_{t=s'+1}^{n} \rho_{\tau_k}(\hat{\xi}_t(\nu_k, \boldsymbol{\Delta})).$$

Define the score function of $L_n(\mathbf{v}, \boldsymbol{\Delta})$ *as*

$$\mathbf{U}_n(\boldsymbol{\nu}, \boldsymbol{\Delta}) = ((\mathbf{U}_{n1}(\boldsymbol{\nu}, \boldsymbol{\Delta}))', (\mathbf{U}_{n2}(\boldsymbol{\nu}, \boldsymbol{\Delta}))')',$$

where $\mathbf{U}_{n1}(\boldsymbol{\nu}, \boldsymbol{\Delta}) = n^{-(1/2)} \sum_{t=s'+1}^{n} (\omega_1 \psi_{\tau_1}(\hat{\xi}_t(\nu_1, \boldsymbol{\Delta})), \ldots, \omega_K \psi_{\tau_K}(\hat{\xi}_t(\nu_K, \boldsymbol{\Delta})))'$, $\mathbf{U}_{n2}(\boldsymbol{\nu}, \boldsymbol{\Delta}) = n^{-(1/2)} \sum_{k=1}^{K} \omega_k \sum_{t=s'+1}^{n} \mathbf{w}_t(\boldsymbol{\Delta}) \psi_{\tau_k}(\hat{\xi}_t(\nu_k, \boldsymbol{\Delta}))$ *with*

$$\mathbf{w}_t(\boldsymbol{\Delta}) = \begin{pmatrix} \mathbf{w}_{t1}(\boldsymbol{\Delta}) \\ \mathbf{w}_{t2}(\boldsymbol{\Delta}) \end{pmatrix} \equiv \begin{pmatrix} -\mathbf{B}_t'(\boldsymbol{\Delta}_1)\boldsymbol{\beta}(\boldsymbol{\Delta}_2)/h_t(\boldsymbol{\Delta}) \\ \mathbf{Z}_t(\boldsymbol{\Delta}_1)/h_t(\boldsymbol{\Delta}) \end{pmatrix}.$$

Then, under Assumptions 3.1, 3.2 and 3.4, $\|\mathbf{U}_n(\hat{\boldsymbol{\nu}}_n, \hat{\boldsymbol{\Delta}}_n)\| = o_p(1)$.

LEMMA A.5. *Let* $\tilde{\mathbf{U}}_n(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) = (\tilde{\mathbf{U}}'_{n1}(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}), \tilde{\mathbf{U}}'_{n2}(\boldsymbol{\nu}, \boldsymbol{\Delta}))'$, *where* $\tilde{\mathbf{U}}_{n1}(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) = \tilde{\mathbf{U}}_1(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n})$ *and* $\tilde{\mathbf{U}}_{n2}(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) = n^{-(1/2)} \sum_{k=1}^{K} \omega_k \sum_{t=s'+1}^{n} \mathbf{w}_t(\boldsymbol{\Delta}) \psi_{\tau_k}(\tilde{\xi}_t(\nu_k, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}))$. *Then, under Assumptions 3.1, 3.2 and 3.4,*

$$\sup_{\mathcal{C}_M} \|\tilde{\mathbf{U}}_n(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n}) - \tilde{\mathbf{U}}(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n})\| = o_p(1). \tag{A.1}$$

LEMMA A.6. *Under Assumptions 3.1, 3.2 and 3.4, with probability tending to one, there exist root-n consistent minimizers in (3.5) and (3.10).*

THEOREM A.1. *Suppose that the threshold and the delay parameters are known. Under Assumptions 3.1–3.3, with probability tending to one, there is a local minimizer* $\hat{\boldsymbol{\beta}}_0$ *admitting the Bahadur representation,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}^* + c_2) = (b^*_\tau f(b^*_\tau))^{-1} \boldsymbol{\Gamma}^{-1}(\boldsymbol{\eta}_{n,\tau} - b^*_\tau \boldsymbol{\mu} q_{n,\tau}) + o_p(1),$$

where $q_{n,\tau} = n^{-(1/2)} \sum_{t=s+1}^{n} \psi_\tau(u_t - b^*_\tau)$, $\boldsymbol{\eta}_{n,\tau} = n^{-(1/2)} \sum_{t=s+1}^{n} \mathbf{Z}_t h_t^{-1}(u_t \psi_\tau(u_t - b^*_\tau) - E[u_t \psi_\tau(u_t - b^*_\tau)])$, *and* $c_2 = E[u_t \psi_\tau(u_t - b^*_\tau)](b^{*2}_\tau f(b^*_\tau))^{-1} \boldsymbol{\Gamma}^{-1} \boldsymbol{\mu}$.

Proof: Define $v = \sqrt{n}(b_\tau - b^*_\tau)$ and $\boldsymbol{\Delta}_2 = \sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}^*)$. Then, the score function of the objective function (3.3) can be defined as

$$\mathbf{P}_n(v, \boldsymbol{\Delta}_2) = n^{-(1/2)} \sum_{t=s+1}^{n} \left(1, \mathbf{Z}'_t h_t^{-2}(\boldsymbol{\Delta}_2)\varepsilon_t\right)' \psi_\tau \left(\varepsilon_t h_t^{-1}(\boldsymbol{\Delta}_2) - b_\tau\right).$$

Let $\mathbf{P}(v, \boldsymbol{\Delta}_2) = n^{-(1/2)} \sum_{t=s+1}^{n} (1, \mathbf{Z}'_t h_t^{-1} u_t)' \psi_\tau(\zeta(v, \boldsymbol{\Delta}_2))$, where

$$\zeta_t(v, \boldsymbol{\Delta}_2) = u_t - b^*_\tau - n^{-(1/2)}(u_t \mathbf{Z}'_t h_t^{-1} \boldsymbol{\Delta}_2 + v).$$

Then, $\mathbf{P}(\mathbf{0}) = (q'_{n,\tau}, \mathbf{z}'_{n,\tau})'$, where $q_{n,\tau}$ is defined in Theorem A.1, and

$$\mathbf{z}_{n,\tau} = n^{-(1/2)} \sum_{t=s+1}^{n} \mathbf{Z}_t h_t^{-1} u_t \psi_\tau(u_t - b^*_\tau).$$

Note that, under Assumption 3.3,

$$\begin{aligned} &E_{t-1}[\psi_\tau(\zeta_t(v, \boldsymbol{\Delta}_2)) - \psi_\tau(u_t - b^*_\tau)] \\ &= -E_{t-1}\left[I\left(u_t < (1 - n^{-(1/2)} \mathbf{Z}'_t h_t^{-1} \boldsymbol{\Delta}_2)^{-1}(b^*_\tau + n^{-(1/2)} v)\right) - I(u_t - b^*_\tau)\right] \\ &= F((1 - n^{-(1/2)} \mathbf{Z}'_t h_t^{-1} \boldsymbol{\Delta}_2)^{-1}(b^*_\tau + n^{-(1/2)} v)) - F(b^*_\tau) \\ &= n^{-(1/2)} f(b^*_\tau)(v + b^*_\tau \mathbf{Z}'_t h_t^{-1} \boldsymbol{\Delta}_2) + o_p(n^{-(1/2)}). \end{aligned}$$

It follows that

$$E_{t-1}\left[n^{-(1/2)} \sum_{s+1}^{n} (\psi_\tau(\zeta_t(v, \boldsymbol{\Delta}_2)) - \psi_\tau(u_t - b^*_\tau))\right] = -f(b^*_\tau)(v + b^*_\tau \boldsymbol{\mu}' \boldsymbol{\Delta}_2) + o_p(1). \tag{A.2}$$

Moreover,

$$\mathrm{Var}_{t-1}\left(n^{-(1/2)} \sum_{s+1}^{n} (\psi_\tau(\zeta_t(v, \boldsymbol{\Delta}_2)) - \psi_\tau(u_t - b^*_\tau))\right) = o_p(1). \tag{A.3}$$

A combination of (A.2) and (A.3) leads to

$$n^{-(1/2)}\sum_{s+1}^{n}(\psi_\tau(\zeta_t(v,\mathbf{\Delta}_2))-\psi_\tau(u_t-b_\tau^*))=-f(b_\tau^*)(v+b_\tau^*\boldsymbol{\mu}'\mathbf{\Delta}_2)+o_p(1). \tag{A.4}$$

Similarly,

$$n^{-(1/2)}\sum_{s+1}^{n}\mathbf{Z}_t h_t^{-1}(u_t\psi_\tau(\zeta_t(v,\mathbf{\Delta}_2))-u_t\psi_\tau(u_t-b_\tau^*))=-b_\tau^*f(b_\tau^*)(\boldsymbol{\mu}\boldsymbol{\nu}+b_\tau^*\mathbf{G}_2\mathbf{\Delta}_2)+o_p(1). \tag{A.5}$$

Combining (A.4) and (A.5) leads to

$$\mathbf{P}(v,\mathbf{\Delta}_2)-\mathbf{P}(\mathbf{0})+f(b_\tau^*)\mathbf{B}(v,\mathbf{\Delta}_2')'=o_p(1), \tag{A.6}$$

where

$$\mathbf{B}=\begin{pmatrix}1 & b_\tau^*\boldsymbol{\mu}'\\ b_\tau^*\boldsymbol{\mu} & (b_\tau^*)^2\mathbf{G}_2\end{pmatrix}.$$

Note that $E[q_{n,\tau}]=0$ and $E[\mathbf{z}_{n,\tau}]=\sqrt{n}\boldsymbol{\mu}E[u_t\psi_\tau(u_t-b_\tau^*)]$. Let $\boldsymbol{\eta}_{n,\tau}=\mathbf{z}_{n,\tau}-E[\mathbf{z}_{n,\tau}]$. Then, the centred $\mathbf{P}(\mathbf{0})$ is $\mathbf{P}^*(\mathbf{0})=(q_{n,\tau}',\boldsymbol{\eta}_{n,\tau}')'$, which is jointly normal. Now, we distribute the above mean part to the third term on the left-hand side of (A.6) and reparametrize the parameters.

Let

$$f(b_\tau^*)\mathbf{B}(v,\mathbf{\Delta}_2')'-\sqrt{n}(\mathbf{0}',\boldsymbol{\mu}'E[u_t\psi_\tau(u_t-b_\tau^*)])'=f(b_\tau^*)\mathbf{B}(v^*,\mathbf{\Delta}_2^{*'})'.$$

Then

$$(v^*,\mathbf{\Delta}_2^{*'})'=(v,\mathbf{\Delta}_2')'+\sqrt{n}(c_1,c_2')'=(\sqrt{n}(b_\tau-b_\tau^*+c_1),\sqrt{n}(\boldsymbol{\beta}-\boldsymbol{\beta}^*+c_2)')',$$

where $(c_1,c_2')'=-f^{-1}(b_\tau^*)\mathbf{B}^{-1}(\mathbf{0}',\boldsymbol{\mu}'E[u_t\psi_\tau(u_t-b_\tau^*)])'$. It is easy to verify that $c_2=E[u_t\psi_\tau(u_t-b_\tau^*)](b_\tau^{*2}f(b_\tau^*))^{-1}\mathbf{\Gamma}^{-1}\boldsymbol{\mu}$. Now (A.6) becomes

$$\mathbf{P}(v^*,\mathbf{\Delta}_2^*)-\mathbf{P}^*(\mathbf{0})+f(b_\tau^*)\mathbf{B}(v^*,\mathbf{\Delta}_2^{*'})'=o_p(1), \tag{A.7}$$

where $\mathbf{P}(v^*,\mathbf{\Delta}_2^*)=\mathbf{P}(v,\mathbf{\Delta}_2)$. Using the chaining argument in Bickel (1975), we establish that

$$\sup_{\|v^*\|\le M,\|\mathbf{\Delta}_2^*\|\le M}\|\mathbf{P}(v^*,\mathbf{\Delta}_2^*)-\mathbf{P}^*(\mathbf{0})+f(b_\tau^*)\mathbf{B}(v^*,\mathbf{\Delta}_2^{*'})'\|=o_p(1). \tag{A.8}$$

Using the same argument as that for Lemma A.5, we can obtain that

$$\sup_{\|v^*\|\le M,\|\mathbf{\Delta}_2^*\|\le M}\|\mathbf{P}_n(v^*,\mathbf{\Delta}_2^*)-\mathbf{P}(v^*,\mathbf{\Delta}_2^*)\|=o_p(1),$$

which combined with (A.8) leads to

$$\sup_{\|v^*\|\le M,\|\mathbf{\Delta}_2^*\|\le M}\|\mathbf{P}_n(v^*,\mathbf{\Delta}_2^*)-\mathbf{P}^*(\mathbf{0})+f(b_\tau^*)\mathbf{B}(v^*,\mathbf{\Delta}_2^{*'})'\|=o_p(1). \tag{A.9}$$

Simple algebra gives $\mathbf{P}^*(\mathbf{0})=O_p(1)$. Let

$$(\hat{v}_n^*,\hat{\mathbf{\Delta}}_{2n}^{*'})'=\left(\sqrt{n}(\hat{b}_\tau-b_\tau^*+c_1),\sqrt{n}(\hat{\boldsymbol{\beta}}_0-\boldsymbol{\beta}^*+c_2)'\right)',$$

Applying Lemma A.2 of Ruppert and Carroll (1980), we obtain

$$\mathbf{P}_n(\hat{v}_n^*,\hat{\mathbf{\Delta}}_{2n}^*)=\mathbf{P}_n(\hat{v}_n,\hat{\mathbf{\Delta}}_{2n})=o_p(1).$$

Using an argument similar to that for Lemma A.6, there exist $\hat{\upsilon}_n^*$ and $\hat{\mathbf{\Delta}}_{2n}^*$ such that $\|\hat{\upsilon}_n^*\| = O_p(1)$ and $\|\hat{\mathbf{\Delta}}_{2n}^*\| = O_p(1)$, and that $(\hat{\upsilon}_n^*, \hat{\mathbf{\Delta}}_{2n}^*)$ satisfies the following score equation

$$\begin{cases} f(b_\tau^*)(\hat{\upsilon}_n^* + b_\tau^* \boldsymbol{\mu}' \hat{\mathbf{\Delta}}_{2n}^*) = q_{n,\tau} + o_p(1) \\ f(b_\tau^*)(b_\tau^* \boldsymbol{\mu} \hat{\upsilon}_n^* + (b_\tau^*)^2 \mathbf{G}_2 \hat{\mathbf{\Delta}}_{2n}^*) = \boldsymbol{\eta}_{n,\tau} + o_p(1). \end{cases}$$

Solving the above equation, we obtain

$$\hat{\mathbf{\Delta}}_{2n}^* = (b_\tau^{*2} f(b_\tau^*))^{-1} \mathbf{\Gamma}^{-1} (\boldsymbol{\eta}_{n,\tau} - b_\tau^* \boldsymbol{\mu} q_{n,\tau}) + o_p(1).$$

Equivalently,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_0 - \boldsymbol{\beta}^* + c_2) = (b_\tau^* f(b_\tau^*))^{-1} \mathbf{\Gamma}^{-1} (\boldsymbol{\eta}_{n,\tau} - b_\tau^* \boldsymbol{\mu} q_{n,\tau}) + o_p(1). \qquad \square$$

Proof of Theorem 3.1: Let $\boldsymbol{\nu}_k = \sqrt{n}(c_{\tau_k} - c_{\tau_k}^*)$, $\mathbf{\Delta}_2 = \sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}^*)$, $\hat{\upsilon}_{n,k} = \sqrt{n}(\hat{c}_{\tau_k} - c_{\tau_k}^*)$, $\hat{\boldsymbol{\nu}}_n = (\hat{\upsilon}_{n,1}, \ldots, \hat{\upsilon}_{n,k})'$, and $\hat{\mathbf{\Delta}}_{2n} = \sqrt{n}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}^*)$. By Lemma A.6, with probability tending to one, there exist $(\hat{\boldsymbol{\nu}}_n, \hat{\mathbf{\Delta}}_{2n})$ such that $\|(\hat{\boldsymbol{\nu}}_n', \hat{\mathbf{\Delta}}_{2n}')'\| = O_p(1)$. Define the score function of the objective function in (3.5) as

$$\mathbf{V}_n(\boldsymbol{\nu}, \mathbf{\Delta}_2) = (\mathbf{V}_{n1}'(\boldsymbol{\nu}, \mathbf{\Delta}_2), \mathbf{V}_{n2}'(\boldsymbol{\nu}, \mathbf{\Delta}_2))',$$

where $\mathbf{V}_{n1}(\boldsymbol{\nu}, \mathbf{\Delta}_2) = (V_{n11}, \ldots, V_{n1K})'$ with

$$V_{n1k}(\nu_k, \mathbf{\Delta}_2) = n^{-(1/2)} \sum_{t=s'+1}^{n} \omega_k \psi_{\tau_k} \left(\xi_t^0(\nu_k, \mathbf{\Delta}_2) \right)$$

and $\mathbf{V}_{n2}(\boldsymbol{\nu}, \mathbf{\Delta}_2) = \sum_{k=1}^{K} \omega_k V_{n2k}(\nu_k, \mathbf{\Delta}_2)$ with

$$V_{n2k}(\nu_k, \mathbf{\Delta}_2) = n^{-(1/2)} \sum_{t=s'+1}^{n} \mathbf{Z}_t h_t(\mathbf{\Delta}_2)^{-1} \psi_{\tau_k} \left(\xi_t^0(\nu_k, \mathbf{\Delta}_2) \right).$$

By Lemma A.3(a), we have

$$\begin{aligned} V_{n1k}(\nu_k, \mathbf{\Delta}_2) &= n^{-(1/2)} \sum_{t=s'+1}^{n} \omega_k \psi_{\tau_k}(\xi_t(\nu_k, \mathbf{\Delta}_2)) + n^{-(1/2)} \sum_{t=s'+1}^{n} \\ &\quad \times \omega_k \left(\psi_{\tau_k}(\xi_t^0(\nu_k, \mathbf{\Delta}_2)) - \psi_{\tau_k}(\xi_t(\nu_k, \mathbf{\Delta}_2)) \right) \\ &\equiv V_{1k}(\nu_k, \mathbf{\Delta}_2) + o_p(1) \end{aligned}$$

uniformly for $(\nu_k, \mathbf{\Delta}_2) \in \mathcal{C}_M$. Similarly, by Lemma A.3(b),

$$V_{n2k}(\nu_k, \mathbf{\Delta}_2) = n^{-(1/2)} \sum_{t=s'+1}^{n} \mathbf{Z}_t h_t(\mathbf{\Delta}_2)^{-1} \psi_{\tau_k}(\xi_t(\nu_k, \mathbf{\Delta}_2)) + o_p(1)$$

uniformly for $(\nu_k, \mathbf{\Delta}) \in \mathcal{C}_M$. The first term on the right-hand side of the above equation can be decomposed as

$$\begin{aligned} & n^{-(1/2)} \sum_{t=s'+1}^{n} \mathbf{Z}_t h_t^{-1} \psi_{\tau_k}(\xi_t(\nu_k, \mathbf{\Delta}_2)) + n^{-(1/2)} \sum_{t=s'+1}^{n} \mathbf{Z}_t (h_t(\mathbf{\Delta}_2)^{-1} - h_t^{-1}) \psi_{\tau_k}(\xi_t(\nu_k, \mathbf{\Delta}_2)) \\ & \quad + o_p(1) = V_{2k}(\nu_k, \mathbf{\Delta}_2) + V_{2k}^*(\nu_k, \mathbf{\Delta}_2) + o_p(1). \end{aligned}$$

Note that $h_t(\boldsymbol{\Delta}_2) - h_t = n^{-1/2}\mathbf{Z}_t'\boldsymbol{\Delta}_2$. It can be shown that $V_{2k}^*(\nu_k, \boldsymbol{\Delta}_2) = o_p(1)$ uniformly for $(\nu_k, \boldsymbol{\Delta}) \in \mathcal{C}_M$. Then,

$$V_{n2k}(\nu_k, \boldsymbol{\Delta}_2) = V_{2k}(\nu_k, \boldsymbol{\Delta}_2) + o_p(1)$$

uniformly for $(\nu_k, \boldsymbol{\Delta}) \in \mathcal{C}_M$, and hence

$$\sup_{\mathcal{C}_M} \|\mathbf{V}_n(\boldsymbol{\nu}, \boldsymbol{\Delta}_2) - \mathbf{V}(\boldsymbol{\nu}, \boldsymbol{\Delta}_2)\| = o_p(1). \tag{A.10}$$

Applying Lemma A.5 of Koenker and Zhao (1996), we obtain

$$\|\mathbf{V}_n(\hat{\boldsymbol{\nu}}_n, \hat{\boldsymbol{\Delta}}_{2n})\| = o_p(1). \tag{A.11}$$

Combining (A.10) and (A.11) leads to

$$\|\mathbf{V}(\hat{\boldsymbol{\nu}}_n, \hat{\boldsymbol{\Delta}}_{2n})\| = o_p(1). \tag{A.12}$$

Simple algebra gives $V(\mathbf{0}) = O_p(1)$. By (A.12) and Lemma A.1, we have

$$\begin{cases} \omega_k g(c_{\tau_k}^*)(\hat{\nu}_{nk} + \boldsymbol{\mu}'\hat{\boldsymbol{\Delta}}_{2n}) = \omega_k q_{n,k} + o_p(1) \\ \sum_{k=1}^{K} \omega_k g(c_{\tau_k}^*)(\boldsymbol{\mu}\hat{\nu}_{nk} + \mathbf{G}_2\hat{\boldsymbol{\Delta}}_{2n}) = \sum_{k=1}^K \omega_k \mathbf{z}_{n,k} + o_p(1), \end{cases}$$

with $\mathbf{G}_2 = E[h_t^{-2}\mathbf{Z}_t\mathbf{Z}_t']$. Solving the above equations, we obtain

$$\hat{\boldsymbol{\Delta}}_{2n} = \left(\boldsymbol{\Gamma}\sum_{k=1}^{K}\omega_k g(c_{\tau_k}^*)\right)^{-1}\mathbf{z}_n + o_p(1). \qquad \square$$

Proof of Theorem 3.2: It is easy to see that $\hat{\omega}_k = \omega_k + o_p(1)$ uniformly for $k = 1, \ldots, K$. The difference between the score function $\mathbf{V}_n(\boldsymbol{\nu}, \boldsymbol{\Delta}_2)$ with $\hat{\omega}_k$ and that with $\omega_{k,\text{opt}}$ is $o_p(1)$. Equations (A.10)–(A.12) still hold. This, together with a result similar to Lemma A.1, leads to the result of the theorem. $\square$

Proof of Theorem 3.3: By Lemma A.4, the score function $\mathbf{U}_n(\boldsymbol{\nu}, \boldsymbol{\Delta})$ of $L_n(\boldsymbol{\nu}, \boldsymbol{\Delta})$ in (3.10) satisfies

$$\|\mathbf{U}_n(\hat{\boldsymbol{\nu}}_n, \hat{\boldsymbol{\Delta}}_n)\| = o_p(1). \tag{A.13}$$

Note that $\psi_{\tau_k}(u) = \tau_k - I(u < 0)$. Under the condition $\|\tilde{\boldsymbol{\Delta}}_{1n}\| = O_p(1)$, by Lemma A.3(c) and (d), we have

$$\sup_{\mathcal{C}_M}\left\|\mathbf{U}_n(\boldsymbol{\nu}, \boldsymbol{\Delta}) - \tilde{\mathbf{U}}_n(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n})\right\| = o_p(1),$$

which, combined with Lemma A.5, yields

$$\sup_{\mathcal{C}_M}\left\|\mathbf{U}_n(\boldsymbol{\nu}, \boldsymbol{\Delta}) - \tilde{\mathbf{U}}(\boldsymbol{\nu}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\Delta}}_{1n})\right\| = o_p(1). \tag{A.14}$$

Applying Lemma A.2 and (A.14), we obtain that

$$\sup_{\mathcal{C}_M}\Big\| U_{n1k}(\boldsymbol{\nu}, \boldsymbol{\Delta}) - U_{1k}(0) + \omega_k g(c_{\tau_k}^*)(\nu_k + \boldsymbol{\mu}'\boldsymbol{\Delta}_2 - E[\mathbf{B}_t'\boldsymbol{\beta}^* h_t^{-1}]\boldsymbol{\Delta}_1)$$
$$+ \omega_k \tilde{f}(c_\tau^*)E[\mathbf{X}_t' h_t^{-1}]\tilde{\boldsymbol{\Delta}}_{1n}\Big\| = o_p(1) \tag{A.15}$$

and

$$\sup_{\mathcal{C}_M}\left\| \mathbf{U}_{n2}(\boldsymbol{\nu},\boldsymbol{\Delta}) - \mathbf{U}_2(0) + \sum_{k=1}^{K}\omega_k g(c^*_{\tau_k})\boldsymbol{\varphi}(\nu_k,\boldsymbol{\Delta}) + \sum_{k=1}^{K}\omega_k \Omega^*_k \tilde{\boldsymbol{\Delta}}_{1n}\right\| = o_p(1). \tag{A.16}$$

By Lemma A.6, with probability tending to one, there exist $(\hat{\boldsymbol{\nu}}_n, \hat{\boldsymbol{\Delta}}_n)$ such that $\|(\hat{\boldsymbol{\nu}}'_n, \hat{\boldsymbol{\Delta}}'_n)'\| = O_p(1)$. From (A.13), (A.15), and (A.16) , we have

$$\left(\sum_{k=1}^{K}\omega_k g(c^*_{\tau_k})\right)\boldsymbol{\Sigma}\hat{\boldsymbol{\Delta}}_n = \boldsymbol{\eta}_n - \sum_{k=1}^{K}\omega_k \tilde{f}(c^*_{\tau_k})\mathbf{D}\tilde{\boldsymbol{\Delta}}_{1n} + o_p(1),$$

where $\mathbf{D} = \mathrm{Cov}(\mathbf{w}_t, \mathbf{X}_t h_t^{-1})$ and $\boldsymbol{\eta}_n \overset{d}{\sim} \mathcal{N}(0, \boldsymbol{\omega}'\mathbf{A}\boldsymbol{\omega}\boldsymbol{\Sigma})$. Therefore, under Assumption 3.2′

$$\hat{\boldsymbol{\Delta}}_n = \left(\sum_{k=1}^{K}\omega_k g(c^*_{\tau_k})\boldsymbol{\Sigma}\right)^{-1}\left(\boldsymbol{\eta}_n - \sum_{k=1}^{K}\omega_k \tilde{f}(c^*_{\tau_k})\mathbf{D}\tilde{\boldsymbol{\Delta}}_{1n}\right) + o_p(1). \tag{A.17}$$

□

Proof of Theorem 3.4: Using the one-step estimator as an initial estimator, by Theorem 3.3, we can obtain a refined consistent one-step estimator labelled as $(\hat{\boldsymbol{\alpha}}^{(2)}, \hat{\boldsymbol{\beta}}^{(2)})$. We continue to use this procedure until convergence and obtain the final fully iterative estimator as a solution to the problem (3.9). Then, by (A.17), the final fully iterative estimator satisfies

$$\hat{\boldsymbol{\Delta}}_n = \left(\sum_{k=1}^{K}\omega_k g(c^*_{\tau_k})\boldsymbol{\Sigma}\right)^{-1}\left(\boldsymbol{\eta}_n - \sum_{k=1}^{K}\omega_k \tilde{f}(c^*_{\tau_k})\mathbf{D}\hat{\boldsymbol{\Delta}}_{1n}\right) + o_p(1). \tag{A.18}$$

Obviously, equation (A.18) is equivalent to

$$\sum_{k=1}^{K}\omega_k\left(g(c^*_{\tau_k})\boldsymbol{\Sigma} + \tilde{f}(c^*_{\tau_k})(\mathbf{D},\mathbf{0})\right)\hat{\boldsymbol{\Delta}}_n = \boldsymbol{\eta}_n + o_p(1).$$

Then (3.11) holds. This completes the proof of the theorem. □

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Supplementary Material