DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

QUASI-MAXIMUM LIKELIHOOD ESTIMATION FOR CONDITIONAL QUANTILES

Ivana Komunjer

SOCIAL SCIENCE WORKING PAPER 1139

July 2002

QUASI-MAXIMUM LIKELIHOOD ESTIMATION FOR CONDITIONAL QUANTILES*

By Ivana Komunjer†

In this paper we derive the asymptotic distribution of a new class of quasi-maximum likelihood estimators (QMLE) based on a 'tick-exponential' family of densities. We show that the 'tick-exponential' assumption is a necessary and sufficient condition for a QMLE to be consistent for the parameters of a correctly specified model of a given conditional quantile. Hence, the role of this family of densities in the conditional quantile estimation is analog to the role of the linear-exponential family in the conditional mean estimation. The 'tick-exponential' QMLEs are shown to be asymptotically normal with an asymptotic covariance matrix that has a novel form, not seen in earlier work, and which accounts for possible model misspecification. For practical purposes, we show that the maximization of the 'tick-exponential' (quasi) log-likelihood can conveniently be carried out by using standard gradient-based optimization techniques. More importantly, we provide a consistent estimator for the asymptotic covariance matrix based on the "scores" of the log-likelihood, which allows us to compute the conditional quantile confidence intervals.

Keywords: 'tick-exponential' densities, conditional quantiles, quasi-maximum likelihood estimation, misspecification, asymptotic distribution.

JEL classification: C13, C20, C51, C63.

---

*Paper presented at the 2002 Summer Meeting of the Econometric Society, UCLA.

†I am indebted to Graham Elliott, Philip Gill, Alain Monfort, Christopher Sims, Gary Solon and Hal White for their suggestions and comments.

1. INTRODUCTION

The vast majority of the empirical literature in economics has traditionally focused on conditional mean models, often using the quasi-maximum likelihood framework for estimation and inference. The cornerstones of the asymptotic theory for quasi-maximum likelihood estimators (QMLEs) have been set by White (1982), Gourieroux, Monfort and Trognon (1984), Bollerslev and Wooldridge (1992), White (1994) and Newey and Steigerwald (1997). Virtually all quasi-maximum likelihood literature has since focused on the problem of conditional mean and variance estimation, leaving behind other potentially interesting distribution characteristics, such as its quantiles. Over the last decade however, there has been a growing interest in the problem of conditional quantile estimation and inference, prompt by a rapid growth in the empirical quantile regression literature in various applied areas of economics (see e.g., Koenker and Hallock, 2000, for a review). Hence, the focus of this paper is the asymptotic theory for conditional quantile QMLEs.

Since the seminal work by Koenker and Bassett (1978), several authors have provided asymptotic distribution results for conditional quantile estimators obtained by quantile regression (QR) under various dependence structures. In the context of linear models, for example, results for independent random variables have been derived by Koenker and Bassett (1978), for special cases of conditionally heteroskedastic processes by Koenker and Bassett (1982) and Koenker and Zhao (1996), while a rather complete asymptotic theory for linear quantile regression estimators with dependent data is due to Portnoy (1991). Treatment of possibly misspecified linear quantile regression models with independent observations has been recently proposed by Kim and White (2002). Fewer results are available in the context of nonlinear models, with exception of censored quantile regression models, as treated by Powell (1986) for example.

In this paper we consider possibly misspecified nonlinear conditional quantile models that we estimate by using a quasi-maximum likelihood approach. We show that there is only one class of QMLEs - class that we call 'tick-exponential' - which is consistent for the parameters of a correctly specified model of a given conditional quantile. Moreover, the 'tick-exponential' QMLEs are shown to be asymptotically normal with an asymptotic covariance matrix that has a novel form, which takes into account possible model misspecification. Hence, we propose a rather complete asymptotic theory for QMLEs of possibly misspecified nonlinear conditional quantile models with dependent random variables, which, to the best of our knowledge, has not yet been derived in the literature.

When compared with the existing QR results, there are several advantages in using the QMLEs for conditional quantile estimation. Firstly, the asymptotic distribution results developed here cover all previously obtained quantile regression results as special cases. They are therefore applicable to both linear and nonlinear models, under a wide range of dependence structures and they take into account model misspecification effects. Secondly, a quasi-maximum likelihood approach provides a "directly" computable consistent estimator of the asymptotic covariance matrix. This point is particularly relevant for empirical applications, in which the computation of confidence intervals for QR estimators typically involves different simulation or bootstrap techniques that considerably increase the computational costs (see e.g., Buchinsky, 1995, Fitzenberger, 1997). Finally, the computation of the QMLE can be easily carried out by transforming the initial maximization of a non-differentiable (quasi) log-likelihood into a "minimax" problem involving continuously differentiable functions allowing the use of standard gradient-based optimization techniques. This computational advantage is particularly important in nonlinear conditional quantile models in which linear programming techniques need to be replaced by more cumbersome interior-point optimization methods (see e.g., Koenker and Park, 1996).

The remainder of the paper is organized as follows: Section 2 is an overview in which we introduce the 'tick-exponential' family of densities. In Section 3 we show that for a QMLE to be consistent for the parameters of a given conditional quantile, it is necessary and sufficient condition to be built upon a density which belongs to the 'tick-exponential' family. In Section 4 we show asymptotic normality of 'tick-exponential' QMLEs and derive a novel form for the asymptotic covariance matrix which takes into account model misspecification. The practical implementation issues - optimization and consistent covariance matrix estimation - are treated in Section 5, which concludes the paper. All technicalities regarding the proofs as well as the assumptions used in the paper are relegated to the Appendix.

2. OVERVIEW

Consider a stochastic process $X \equiv \{X_t : \Omega \longrightarrow \mathbb{R}^{n+1}, n \in \mathbb{N}, t = 1, \ldots, T\}$ defined on a complete probability space $(\Omega, \mathcal{F}, P_0)$ where $\mathcal{F} = \{\mathcal{F}_t, t = 1, \ldots, T\}$ and $\mathcal{F}_t$ is the σ-field $\mathcal{F}_t \equiv \sigma\{X_s, s \leqslant t\}$. In what follows, we partition the observed vector X_t as $X_t \equiv (Y_{t-1}, Z_t')'$, where $Y_{t-1} \in \mathbb{R}$ is the scalar variable of interest and $Z_t \in \mathbb{R}^n$ a vector of exogenous variables.[1] We denote by y_{t-1} and z_t the observations of the variables Y_{t-1} and Z_t, respectively. The variable Y_t is assumed to be continuous and we denote by $F_{0,t}$ (resp. $f_{0,t}$) its true conditional distribution (resp. density with respect to a given measure ν), which is unknown. By convention, the subscript t denotes conditioning on the information set $\mathcal{F}_t$.

Let $Q_\alpha(Y_t|\mathcal{F}_t)$ denote the α-quantile of Y_t conditional on the information set $\mathcal{F}_t$, where for a given value of probability level $\alpha \in (0,1)$, $\alpha = P_0(Y_t \leqslant Q_\alpha(Y_t|\mathcal{F}_t)|\mathcal{F}_t)$. The aim of this paper is to study the asymptotic properties of a large class of estimators for $Q_\alpha(Y_t|\mathcal{F}_t)$ - the quasi-maximum likelihood estimators (QMLE).

The approach used to construct the QMLE of the conditional α-quantile of Y_t is analog to those which are employed to estimate the conditional expectation of Y_t. Let $\mathcal{M}$ denote a model for the conditional α-quantile of Y_t, $\mathcal{M} = \{q_t^\alpha\}$, with $q_t^\alpha(W_t, \cdot) : \Theta \to \mathbb{R}$ and where W_t is a vector of variables that are $\mathcal{F}_t$-measurable. In practice, W_t consists of different functions of (i) a subset of Z_t, n-vector of exogenous variables, and (ii) lags of Y_{t-1}. In finance, examples of different specifications for q_t^α are: Koenker and Zhao's (1996) conditional quantile model: $q_t^\alpha(W_t, \theta) \equiv \beta_0 + \sum_{i=1}^p \beta_i Y_{t-i} + \sigma_t q_\alpha$ for a particular class of ARCH processes where the conditional scale, σ_t, satisfies $\sigma_t = \gamma_0 + \sum_{j=1}^q \gamma_j |Y_{t-j} - \beta_0 - \sum_{i=1}^p \beta_i Y_{t-j-i}|$; Engle and Manganelli's (1999) CAViaR model with first order representation: $q_t^\alpha(W_t, \theta) \equiv \beta_0 + \beta_1 q_{t-1}^\alpha(W_{t-1}, \theta) + l(\beta_2, Y_{t-1}, q_{t-1}^\alpha(W_{t-1}, \theta))$, in which l corresponds to some loss function; Taylor's (1999) and Chernozhukov and Umanstev's (2000) linear VaR: $q_t^\alpha(W_t, \theta) \equiv W_t'\theta$ and quadratic VaR models: $q_t^\alpha(W_t, \theta) \equiv W_t'\beta + W_t B W_t'$.

In what follows, we treat two types of situations, depending on whether or not the model $\mathcal{M}$ is correctly specified. We say that $\mathcal{M}$ is correctly specified for the parameters of the conditional α-quantile of Y_t if there exists a k-vector θ_0 in Θ, $\Theta \subset \mathbb{R}^k$, such that $q_t^\alpha(W_t, \theta_0) = Q_\alpha(Y_t|\mathcal{F}_t)$, where W_t is a vector of variables that are $\mathcal{F}_t$-measurable.

We want to estimate θ_0 by considering the class of quasi-maximum likelihood estimators, $\hat{\theta}_T$, obtained by solving

$$(1) \qquad \max_{\theta \in \Theta} L_T(\theta) \equiv T^{-1} \sum_{t=1}^{T} \ln l_t(y_t, q_t^\alpha(w_t, \theta)),$$

where l_t is a period-t conditional (quasi) log-likelihood of Y_t given $\mathcal{F}_t$. It is a well know result that different choices of l_t affect the asymptotic properties of the QMLE $\hat{\theta}_T$ when the object of interest is the conditional mean of Y_t. Specifically, and under standard regularity assumptions, the QMLE provides a consistent estimate of the true parameters of a correctly specified model of the conditional mean, $\{\mu_t\}$, if and only if l_t belongs to the linear-exponential family, i.e. for $y \in \mathbb{R}$,

$$(2) \qquad l_t(y, \eta) = \exp[a_t(\eta) + b_t(y) + yc_t(\eta)],$$

where the functions $a_t : M_t \to \mathbb{R}$ and $c_t : M_t \to \mathbb{R}$ are continuous, $M_t \subset \mathbb{R}$, the function $b_t : \mathbb{R} \to \mathbb{R}$ is $\mathcal{F}_t$-measurable, and a_t, b_t, c_t are such that l_t is a probability density with mean η. The QMLE obtained by solving $\max_{\theta \in \Theta} T^{-1} \sum_{t=1}^{T} \ln l_t(y_t, \mu_t(w_t, \theta))$ in which l_t is given by (2) is consistent for the true value θ_0 of a correctly specified model for the conditional mean even if other aspects of the conditional distribution of Y_t are misspecified, i.e. the true density $f_{0,t}$ is *not* equal to $l_t(\cdot, \mu_t(w_t, \theta))$. This property was derived by White (1994), as a generalization of the result proposed by Gourieroux, Monfort, and Trognon (1984). In this paper, we derive an analog result for the case where the object of interest is the conditional α-quantile of Y_t. We start by defining the tick-exponential family of densities - family whose role in the conditional quantile estimation is analog to the role of the linear-exponential family in the conditional mean estimation.

Definition 1 (tick-exponential family) A family of probability measures on $\mathbb{R}$ admitting a density φ_t^α indexed by a parameter η, $\eta \in M_t$, $M_t \subset \mathbb{R}$, is called tick-exponential of order α, $\alpha \in (0,1)$, if and only if: *(i)* for $y \in \mathbb{R}$,

$$\varphi_t^\alpha(y, \eta) = \exp\{-(1-\alpha)[a_t(\eta) - b_t(y)]1(y \leqslant \eta) + \alpha[a_t(\eta) - c_t(y)]1(y > \eta)\},$$

where $a_t : M_t \rightarrow \mathbb{R}$ is continuously differentiable and $b_t : \mathbb{R} \rightarrow \mathbb{R}$ and $c_t : \mathbb{R} \rightarrow \mathbb{R}$ are $\mathcal{F}_t$-measurables; the functions a_t, b_t and c_t are such that for $\eta \in M_t$: *(ii)* φ_t^α is a probability density, i.e. $\int_{\mathbb{R}} \varphi_t^\alpha(y,\eta)dy = 1$; *(iii)* η is the α-quantile of φ_t^α, i.e. $\int_{-\infty}^{\eta} \varphi_t^\alpha(y,\eta)dy = \alpha$.

In other words, for a given value of probability α, φ_t^α is linear-exponential "by parts" where the two parts have different slopes, proportional to $1-\alpha$ and α, respectively. Note that by setting $d_t(y) \equiv (1-2\alpha)^{-1} \cdot [(1-\alpha)b_t(y) - \alpha c_t(y)]$ and $g_t(y) \equiv \alpha(1-\alpha)(2\alpha-1)^{-1} \cdot [b_t(y) - c_t(y)]$, we obtain an alternative expression for φ_t^α, given by $\varphi_t^\alpha(y,\eta) = \exp\{g_t(y) - (1-\alpha)[a_t(\eta) - d_t(y)]1(y \leqslant \eta) + \alpha[a_t(\eta) - d_t(y)]1(y > \eta)\}$, which has separately been studied by Gourieroux, Monfort and Renault (1987) in the M-estimation context.[2] In a special case when $a_t(\eta) = [1/(\alpha(1-\alpha))]\eta$ and $b_t(y) = c_t(y) = [1/(\alpha(1-\alpha))]y$, the function $\ln \varphi_t^\alpha$ is proportional to $t_\alpha(y,\eta) \equiv (\alpha - 1(y \leqslant \eta)) \cdot (y - \eta)$, which corresponds to the 'tick' function, also known as 'asymmetrical slope' or 'check' function in the literature. This is why we call 'tick-exponential' the family of functions defined in Definition 1.

PROPERTY 2 Let $\varphi_t^\alpha : \mathbb{R} \times M_t \rightarrow \mathbb{R}$ be a tick-exponential density of order α, $\alpha \in (0,1)$, as defined in Definition 1. For every $\eta \in M_t$, the functions $a_t : M_t \rightarrow \mathbb{R}$, $b_t : \mathbb{R} \rightarrow \mathbb{R}$ and $c_t : \mathbb{R} \rightarrow \mathbb{R}$ then satisfy:

(i) $a_t'(\eta) > 0$;

(ii) $\exp\{-(1-\alpha)[a_t(\eta) - b_t(\eta)]\} = \alpha(1-\alpha)a_t'(\eta)$;

(iii) $\exp\{\alpha[a_t(\eta) - c_t(\eta)]\} = \alpha(1-\alpha)a_t'(\eta)$;

(iv) $(1-\alpha)b_t(\eta) + \alpha c_t(\eta) = a_t(\eta)$.

Note that the last equality *(iv)* in particular implies that $\varphi_t^\alpha(\cdot,\eta)$ is continuous on $\mathbb{R}$. In cases where the argument η corresponds to a function of a random variable W_t and of the k-vector of parameters θ, such as $q_t^\alpha(W_t,\theta)$ for example, we further assume that for all θ we have $q_t^\alpha(W_t,\theta) \in M_t$ a.s.-P_0, and that the conditions of Definition 1 are satisfied for $\varphi_t^\alpha(Y_t, q_t^\alpha(W_t,\theta))$.

3. CONSISTENCY OF THE TICK-EXPONENTIAL QMLE

Let us now turn to the asymptotic properties of the QMLE based on tick-exponential family of order α. Let φ_t^α be a tick-exponential density of order α, as defined in Definition 1 and $\hat{\theta}_T$ the corresponding QMLE, solution to

$$(3) \quad \max_{\theta \in \Theta} L_T(\theta) \equiv T^{-1} \sum_{t=1}^{T} \ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \theta)).$$

The following theorem establishes consistency of $\hat{\theta}_T$. For the sake of clarity, all the assumptions used in this and the following theorems are grouped in the Appendix.

Theorem 3 (Sufficient Conditions for Consistency) *Let $\hat{\theta}_T$ be a tick-exponential QMLE, obtained by solving the maximization problem* (3)*. Under assumptions (A0)-(A5) and (A7), $\hat{\theta}_T \xrightarrow{p} \theta_0$.*

In other words, if l_t belongs to the tick-exponential family of densities, the QMLE provides a consistent estimate of the true parameters of a correctly specified model of the conditional α-quantile of Y_t despite distributional misspecification, i.e. even if the true conditional distribution of Y_t is *not* tick-exponential. Hence, we need not know the true distribution of neither Y_t nor the exogenous variables Z_t in order to obtain consistent estimates for the parameters of the conditional α-quantile of Y_t. Even though the consistency result in Theorem 3 is robust to distributional misspecification, it is only valid if the conditional quantile model $\mathcal{M} = \{q_t^\alpha\}$ is correctly specified (assumption (A0)), which may not always be true. Under model misspecification we have the following result.

Corollary 4 (Sufficient Conditions for Consistency under Misspecification) *Let $\theta^* \equiv \arg\max_{\theta \in \Theta} L_0(\theta)$ be the pseudo-true value of the parameter θ when the model $\mathcal{M} = \{q_t^\alpha\}$ for the conditional α-quantile of Y_t is not correct, and let $\hat{\theta}_T$ be a tick-exponential QMLE, solution to the maximization problem* (3)*. Under assumptions (A1)-(A5) and (A7), $\hat{\theta}_T \xrightarrow{p} \theta^*$.*

Both consistency results are valid under standard regularity assumptions, which in general can be classified in three groups: compactness, uniqueness and uniform convergence assumptions. The parameter space Θ is compact by assumption (A1). The purpose of the uniqueness assumptions is to ensure that θ_0 (or θ^*) is the unique maximizer of the expected log-likelihood $L_0(\theta) \equiv E[\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))]$. While this requirement is trivially verified for θ^*, it needs to be checked for θ_0, under correct specification of the conditional α-quantile of Y_t (assumption (A0)). The uniqueness is achieved by imposing an identification assumption on the conditional quantile model $\mathcal{M} = \{q_t^\alpha\}$ (assumption (A3)). The most delicate part of the consistency proof relies on the uniform convergence assumptions. These need to ensure *(i)* that the function $\ln \varphi_t^\alpha$ is uniformly continuous in θ and *(ii)* that the stochastic process $\{\ln \varphi_t^\alpha\}$ has certain dependence structure so that a uniform law of large numbers (ULLN) can be applied. The first requirement is achieved by considering functions that satisfy the Lipshitz condition, implied in the paper by (A2) and (A4). The second requirement is met by imposing heterogeneity restrictions on the process $\{\ln \varphi_t^\alpha\}$. More specifically, we use Andrews' (1988) concept of L_1-mixingales (assumption (A7)), which covers a wide range of dependence structures for X_t. Assumption (A5) ensures that $\{\ln \varphi_t^\alpha\}$ is moreover uniformly integrable, which allows us to use the ULLN.

We now derive conditions which are necessary if we want a QMLE to be consistent for the parameters of a given conditional α-quantile.

Theorem 5 (Necessary Condition for Consistency) *Let $\hat{\theta}_T$ be the QMLE obtained by solving the maximization problem* (1) *in which $q_t^\alpha(W_t, \cdot)$ and $\ln l_t(Y_t, q_t^\alpha(W_t, \cdot))$ are continuously differentiable on Θ a.s.-P_0. Assume that (i) Θ is compact; (ii) $L_T(\theta)$ converges uniformly in probability to $L_0(\theta) \equiv E[\ln l_t(Y_t, q_t^\alpha(W_t, \theta))]$; (iii) $L_0(\theta)$ is continuous; (iv) $L_0(\theta)$ is uniquely maximized at θ_0. Then, a necessary condition for $\hat{\theta}_T$ to be consistent for the parameters of the conditional α-quantile of Y_t, $q_t^\alpha(W_t, \theta_0)$, is that l_t be a tick-exponential density of order α.*

In other words, if we want the QMLE to be consistent for the true parameters of a correctly specified model of the conditional α-quantile of Y_t then we must choose a member l_t of the tick-exponential family. Note that Theorem 5 is not exactly the converse of the result given in Theorem 3. In order to derive the necessary condition for consistency we assume that $q_t^\alpha(W_t,\cdot)$ and $\ln l_t(Y_t, q_t^\alpha(W_t,\cdot))$ are continuously differentiable on Θ a.s.-P_0, i.e. that for all $(y_t, w_t')'$ in some set A_t of measure one, $P_0(A_t) = 1$, we have $q_t^\alpha(w_t,\cdot)$ and $\ln l_t(y_t, q_t^\alpha(w_t,\cdot))$ continuously differentiable on Θ. This property is for example satisfied when for every $\theta \in \Theta$, $\partial \ln l_t(y_t, q_t^\alpha(w_t,\theta))/\partial\theta$ exists and is continuous for almost all $(y_t, w_t')'$, or when $\partial \ln l_t(y_t, q_t^\alpha(w_t,\theta))/\partial\theta$ has a finite set of discontinuities $\{\theta_j(y_t)\}$ where each $d\theta_j/dy_t$ exists and is not zero. The remainder of the assumptions in Theorem 5 are the standard consistency assumptions *(i)* - *(iv)*. Note that under the additional differentiability assumptions we obtain the continuous differentiability of the expected log-likelihood $L_0(\theta)$ and can therefore write the first order condition, $\nabla_\theta L_0(\theta_0) = 0$, upon which is based the proof of Theorem 5.

4. ASYMPTOTIC NORMALITY OF THE TICK-EXPONENTIAL QMLE

Let us now turn to the asymptotic normality of the tick-exponential QMLE $\hat\theta_T$, solution to the maximization problem (3). The classical asymptotic normality results for QMLE require that the log-likelihood function $L_T(\theta)$ be twice continuously differentiable. The main idea is to then use the first-order Taylor expansion of the gradient $\nabla_\theta L_T(\theta)$ around the QMLE $\hat\theta_T$, which satisfies the first order condition $\nabla_\theta L_T(\hat\theta_T) = 0$. This approach requires $L_T(\theta)$ to be sufficiently smooth, which is not the case with the tick-exponential family of densities due to the presence of indicator functions in Definition 1. Indeed, under tick-exponential assumption,

$$\begin{aligned}(4)\qquad L_T(\theta) = T^{-1}\sum_{t=1}^{T}\{&-(1-\alpha)[a_t(q_t^\alpha(w_t,\theta)) - b_t(y_t)]1(y_t \leqslant q_t^\alpha(w_t,\theta)) \\ &+\alpha[a_t(q_t^\alpha(w_t,\theta)) - c_t(y_t)]1(y_t > q_t^\alpha(w_t,\theta))\},\end{aligned}$$

where the functions $a_t(\cdot)$, $b_t(\cdot)$ and $c_t(\cdot)$ are as defined in Definition 1. The non-differentiability problem has prompt several authors to develop asymptotic normality results under a weaker set of assumptions, generally requiring that $\nabla_\theta L_T(\theta)$ exist with probability one. Examples include Daniels (1961), Huber (1967), Pollard (1985), Pakes and Pollard (1989), Newey and McFadden (1994). In the particular case of this paper, we assume that the function $q_t^\alpha(W_t, \cdot) : \Theta \to \mathbb{R}$ is continuously differentiable on Θ a.s.-P_0 (assumption (A4)), so that the log-likelihood function $L_T(\theta)$ is continuously differentiable on Θ with probability one, i.e. for every $\theta \in \Theta$,

$$(5) \quad \nabla_\theta L_T(\theta) \equiv T^{-1} \sum_{t=1}^{T} [\alpha - H(q_t^\alpha(w_t, \theta) - y_t)] a_t'(q_t^\alpha(w_t, \theta)) \nabla_\theta q_t^\alpha(w_t, \theta),$$

exists and is continuous for almost all $(y_t, w_t')'_{t=1,\dots,T}$. The function $H : \mathbb{R}^* \to \{0, 1\}$ is the Heaviside function, i.e. $H(x) = 1$ if $x > 0$ and 0 if $x < 0$.

We now derive the asymptotic distribution of $\hat{\theta}_T$.

Theorem 6 (Asymptotic Normality) *Let $\hat{\theta}_T$ be the tick-exponential QMLE, i.e. $\hat{\theta}_T = \arg\max_{\theta \in \Theta} T^{-1} \sum_{t=1}^{T} \ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \theta))$. Under assumptions (A0), (A1'), (A2'), (A3), (A4') and (A5)-(A8)*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \to \mathcal{N}(0, \Delta_0^{-1} \Sigma_0 \Delta_0^{-1}),$$

where

$$\Delta_0 = -E[f_{0,t}(q_t^\alpha(W_t, \theta_0)) \cdot a_t'(q_t^\alpha(W_t, \theta_0)) \cdot \nabla_\theta q_t^\alpha(W_t, \theta_0) \nabla_\theta q_t^\alpha(W_t, \theta_0)']$$

and

$$\Sigma_0 = -\alpha(1 - \alpha) E[(a_t'(q_t^\alpha(W_t, \theta_0)))^2 \cdot \nabla_\theta q_t^\alpha(W_t, \theta_0) \nabla_\theta q_t^\alpha(W_t, \theta_0)'].$$

Note that the assumptions imposed in Theorem 6 are stronger than the ones used for the consistency of $\hat{\theta}_T$ in Theorem 3. We now require θ_0 to be an interior point of Θ (assumption (A1')). The functions $a_t(\cdot)$ and $q_t^\alpha(W_t, \cdot)$ are assumed to be twice continuously differentiable (assumptions (A2') and (A4')), so that the gradient of the log-likelihood function, $\nabla_\theta L_T$, is stochastically equicontinuous. Similarly to the assumption (A7), assumption (A8) consists of dependence constraints on the process X_t, while by (A6) we constrain the true conditional density of Y_t, $f_{0,t}$, to be bounded and non-zero at the true conditional α-quantile $q_t^\alpha(W_t, \theta_0)$.

The result of Theorem 6 can easily be generalized by relaxing the correct specification assumption (A0). As previously, we denote by θ^* the pseudo-true value of the parameter θ, $\theta^* \equiv \arg\max_{\theta \in \Theta} L_0(\theta)$, when the model $\{q_t^\alpha\}$ is *not* correctly specified. Under model misspecification, θ^* is a solution to the first order condition $E[(\alpha - H(q_t^\alpha(W_t, \theta^*) - Y_t)) \cdot a_t'(q_t^\alpha(W_t, \theta^*)) \cdot \nabla_\theta q_t^\alpha(W_t, \theta^*)] = 0$, but we no longer have $E_t[\alpha - H(q_t^\alpha(W_t, \theta^*) - Y_t)] = 0$, i.e. $\alpha = F_{0,t}(q_t^\alpha(W_t, \theta^*))$, which is a stronger requirement. Hence, the asymptotic distribution of $\hat{\theta}_T$ changes, as stated by the following Corollary.

Corollary 7 (Asymptotic Normality under Misspecification) *Let* $\theta^* \equiv \arg\max_{\theta \in \Theta} L_0(\theta)$ *be the pseudo-true value of the parameter* θ *when the model* $\{q_t^\alpha\}$ *for the conditional* α*-quantile of* Y_t *is not correct. Under assumptions (A1'), (A2'), (A3), (A4') and (A5)-(A8),*

$$\sqrt{T}(\hat{\theta}_T - \theta^*) \to \mathcal{N}(0, \Delta^{*-1}\Sigma^*\Delta^{*-1}),$$

where

$$\begin{aligned}\Delta^* = \; & -E[f_{0,t}(q_t^\alpha(W_t, \theta^*)) \cdot a_t'(q_t^\alpha(W_t, \theta^*)) \cdot \nabla_\theta q_t^\alpha(W_t, \theta^*) \nabla_\theta q_t^\alpha(W_t, \theta^*)'] \\ & + E[(\alpha - F_{0,t}(q_t^\alpha(W_t, \theta^*))) \\ & \cdot [a_t''(q_t^\alpha(W_t, \theta^*)) \cdot \nabla_\theta q_t^\alpha(W_t, \theta^*) \nabla_\theta q_t^\alpha(W_t, \theta^*)' + a_t'(q_t^\alpha(W_t, \theta^*)) \cdot \nabla_{\theta\theta} q_t^\alpha(W_t, \theta^*)]]\end{aligned}$$

and

$$\Sigma^* = E[(\alpha^2 - (2\alpha - 1)F_{0,t}(q_t^\alpha(W_t, \theta^*))) \cdot (a_t'(q_t^\alpha(W_t, \theta^*)))^2 \cdot \nabla_\theta q_t^\alpha(W_t, \theta^*) \nabla_\theta q_t^\alpha(W_t, \theta^*)'].$$

The asymptotic distribution derived in Corollary 7 is a generalization of all the results obtained by the previous literature on conditional quantile estimation. For example, let $a_t(\eta) = [1/(\alpha(1-\alpha))]\eta$ and $b_t(y) = c_t(y) = [1/(\alpha(1-\alpha))]y$, so that the function $\ln \varphi_t^\alpha$ is proportional to the "tick" function, $t_\alpha(y, \eta)$. Hence, θ^* is a solution to a standard non-linear quantile regression problem, $\min_{\theta \in \Theta} E[(\alpha - 1(Y_t \leqslant q_t^\alpha(W_t, \theta))) \cdot (Y_t - q_t^\alpha(W_t, \theta))]$. Assuming a linear model for the conditional α-quantile of Y_t, $q_t^\alpha(W_t, \theta) = \theta' W_t$, it can be shown that $\Delta^* = -1/(\alpha(1-\alpha)) \cdot E[f_{0,t}(\theta^{*\prime} W_t) W_t W_t']$ and $\Sigma^* = 1/(\alpha(1-\alpha))^2 \cdot E[(\alpha^2 - (2\alpha - 1)F_{0,t}(\theta^{*\prime} W_t)) W_t W_t']$. This case, under an additional iid assumption on X_t, corresponds to Kim and White (2002): $\sqrt{T}(\hat{\theta}_T - \theta^*) \to \mathcal{N}(0, E[f_{0,t}(\theta^{*\prime} W_t) W_t W_t']^{-1} \cdot E[(\alpha^2 - (2\alpha - 1)F_{0,t}(\theta^{*\prime} W_t)) W_t W_t'] \cdot E[f_{0,t}(\theta^{*\prime} W_t) W_t W_t']^{-1})$. In a more restricted case where the linear conditional quantile model is correctly specified, it can be shown that $\sqrt{T}(\hat{\theta}_T - \theta_0) \to \mathcal{N}(0, \alpha(1-\alpha)E[f_{0,t}(\theta_0' W_t) W_t W_t']^{-1} \cdot E[W_t W_t'] \cdot E[f_{0,t}(\theta_0' W_t) W_t W_t']^{-1})$, which was derived by Powell (1986). Finally, if the true conditional density of Y_t is $\mathcal{F}_t$-independent, i.e. $f_{0,t} = f_0$, we obtain the original result by Koenker and Bassett (1978): $\sqrt{T}(\hat{\theta}_T - \theta_0) \to \mathcal{N}(0, \alpha(1-\alpha)[f_0^2(\theta_0' W_t) \cdot E(W_t W_t')]^{-1})$.

5. PRACTICAL IMPLEMENTATION

The quasi-maximum likelihood approach based on tick-exponential family of densities provides consistent and asymptotically normal estimators for conditional quantiles under a relatively weak set of assumptions. In practice, however, solving the maximization problem (3) seems difficult a task. The objective function $L_T(\theta)$ is not everywhere differentiable, which prevents us from using the traditional gradient-based optimization techniques in order to determine the maximum θ^*. When the conditional quantile model is linear, the optimization can easily be carried out by linear programming. In the nonlinear case, however, the optimization relies on interior-point methods and is substantially less effective to carry out, as shown by Koenker and Park (1996) for example. The optimization algorithm that we design in this paper applies to the general case and can be used to optimize the log-likelihood function $L_T(\theta)$ in spite of its non-everywhere differentiability.

We describe the intuition behind our approach by first considering the case $T = 1$, i.e. only observations (y_1, z_1) are available. The problem of maximizing $L_1(\theta)$ becomes in that case $\max_{\theta \in \Theta} \ln \varphi_1^\alpha(y_1, q_t^\alpha(w_{1t}, \theta))$, i.e.

(6) $$\max_{\theta \in \Theta} \min \left\{ \ln \psi_1^\alpha(y_1, q_t^\alpha(w_{1t}, \theta)), \ln \phi_1^\alpha(y_1, q_t^\alpha(w_{1t}, \theta)) \right\}$$

where we have defined

$$\begin{aligned} \psi_t^\alpha(y, \eta) &\equiv \exp\{\alpha[a_t(\eta) - c_t(y)]\}, \\ \phi_t^\alpha(y, \eta) &\equiv \exp\{-(1-\alpha)[a_t(\eta) - b_t(y)]\}, \end{aligned}$$

for all $t > 0$, $y \in \mathbb{R}$ and $\eta \in M_t$. Recall that $a_t : M_t \to \mathbb{R}$ is continuously differentiable and $b_t : \mathbb{R} \to \mathbb{R}$ and $c_t : \mathbb{R} \to \mathbb{R}$ are $\mathcal{F}_t$-measurables. Hence, the functions $\psi_t^\alpha(y, \cdot) : M_t \to \mathbb{R}$

and $\phi_t^\alpha(y,\cdot) : M_t \to \mathbb{R}$ are continuously differentiable. By noting that for all $(x,y) \in \mathbb{R}^2$ we have $\min\{x,y\} = -\max\{-x,-y\}$, the maximization problem (6) is equivalent to $\max_{\theta\in\Theta}[-\max\{-\ln\psi_1^\alpha(y_1, q_t^\alpha(w_{1t},\theta)), -\ln\phi_1^\alpha(y_1, q_t^\alpha(w_{1t},\theta))\}]$. Moreover, if Θ is compact (assumption (A1)), the previous maximization problem is equivalent to the minimization problem

$$(7) \qquad -\min_{\theta\in\Theta}[\max\{-\ln\psi_1^\alpha(y_1, q_t^\alpha(w_{1t},\theta)), -\ln\phi_1^\alpha(y_1, q_t^\alpha(w_{1t},\theta))\}].$$

We have thus transformed the initial maximization problem, $\max_{\theta\in\Theta}\ln\varphi_1^\alpha(y_1, q_t^\alpha(w_{1t},\theta))$, into a "minimax" problem (7), which involves only continuously differentiable functions ψ_1^α and ϕ_1^α. Similar reasoning applies when $T > 1$ and the corresponding equivalence is provided by the following Theorem.

Theorem 8 (Optimization) *Let $\varepsilon_\theta \equiv (\varepsilon_{\theta,1}, \varepsilon_{\theta,2}, \ldots, \varepsilon_{\theta,T})'$ be a T-vector of order statistics, $\varepsilon_{\theta,1} \leqslant \varepsilon_{\theta,2} \leqslant \ldots \leqslant \varepsilon_{\theta,T}$, of an "error" term $\varepsilon_t \equiv y_t - q_t^\alpha(w_t,\theta)$, and let $y_\theta \equiv (y_{\theta,1}, y_{\theta,2}, \ldots, y_{\theta,T})'$ and $w_\theta \equiv (w_{\theta,1}, w_{\theta,2}, \ldots, w_{\theta,T})'$ be T-vectors of corresponding observations. Under assumption (A1), the QMLE $\hat\theta_T$ is a solution to the "minimax" problem*

$$\min_{\theta\in\Theta}[\max_{0\leqslant k\leqslant T}\{P_k(y_\theta, w_\theta, \theta)\}]$$

where $P_k(y_\theta, w_\theta, \theta)$ is defined as

$$P_k(y_\theta, w_\theta, \theta) \equiv \begin{cases} T^{-1}\sum_{t=1}^{T} -\ln\psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)), & \text{if } k = 0, \\ T^{-1}[\sum_{t=1}^{k} -\ln\phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) & \\ \quad + \sum_{s=k+1}^{T} -\ln\psi_s^\alpha(y_{\theta,s}, q_s^\alpha(w_{\theta,s},\theta))], & \text{if } 1 \leqslant k \leqslant T-1, \\ T^{-1}\sum_{t=1}^{T} -\ln\phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)), & \text{if } k = T. \end{cases}$$

The tick-exponential QMLE $\hat\theta_T$ can thus be obtained as a solution to the classical "minimax" problem. Moreover, for all k, $0 \leqslant k \leqslant T$, the function $P_k(y_\theta, w_\theta, \cdot)$ is continuously differentiable on Θ. We can therefore use the standard gradient-based optimization techniques to determine the optimum $\hat\theta_T$.

We now turn to the problem of asymptotic covariance matrix estimation. One approach to estimating the asymptotic covariance matrix of $\hat{\theta}_T$ is to use the formulas for Δ^* and Σ^*, derived in Corollary 7. The main difficulty of this approach however is that it requires estimating conditional density, $f_{0,t}$, and distribution, $F_{0,t}$, of Y_t, which is a difficult problem in itself. An alternative approach is to estimate Δ^* and Σ^* by numerical differentiation. Recall that Δ^* corresponds to expected value of the "Hessian" of $\ln \varphi_t^\alpha$, while Σ^* is the asymptotic covariance matrix of the "scores" of $\ln \varphi_t^\alpha$. This second-moment matrix can be estimated by the sample second moment of the scores $\{s_t(\hat{\theta}_T)\}_{1 \leqslant t \leqslant T}$, $\hat{\Sigma} \equiv T^{-1} \sum_{t=1}^T s_t(\hat{\theta}_T) s_t(\hat{\theta}_T)'$. The jth row of s_t, $s_{t,j}$, is obtained by numerical differentiation,

$$(8) \qquad s_{t,j} \equiv [\ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \hat{\theta}_T + e_j \epsilon_T)) - \ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \hat{\theta}_T - e_j \epsilon_T))]/2\epsilon_T,$$

where e_j the jth unit vector and ϵ_T a small positive constant that depends on the sample size. Similarly, the second-order numerical derivative estimator of Δ^*, $\hat{\Delta}$, has (i,j)th element given by

$$(9) \qquad \begin{aligned} \hat{\Delta}_{i,j} \equiv\; & [L_T(\hat{\theta}_T + e_i \epsilon_T + e_j \epsilon_T) - L_T(\hat{\theta}_T - e_i \epsilon_T + e_j \epsilon_T) \\ & - L_T(\hat{\theta}_T + e_i \epsilon_T - e_j \epsilon_T) + L_T(\hat{\theta}_T - e_i \epsilon_T - e_j \epsilon_T)]/4\epsilon_T^2. \end{aligned}$$

If the step size ϵ_T is such that $\epsilon_T \to 0$ and $T^{1/2}\epsilon_T \to \infty$, then $\hat{\Sigma} - \Sigma^* \xrightarrow{p} 0$ and $\hat{\Delta} - \Delta^* \xrightarrow{p} 0$. Hence the asymptotic covariance matrix of $\hat{\theta}_T$ can be consistently estimated by $\hat{\Delta}^{-1}\hat{\Sigma}\hat{\Delta}^{-1}$ (see, e.g. Theorem 7.4 in Newey and McFadden 1994).

6. CONCLUSION

In this paper we have defined a new family of densities, called 'tick-exponential', whose role in the conditional quantile estimation is analog to the role of the linear-exponential family in the conditional mean estimation. Our first result is that the class of 'tick-exponential' QMLEs is consistent for the parameters of a correctly specified conditional quantile model. Our second result is that the 'tick-exponential' assumption is also a necessary condition for consistency. Our third result is that the class of 'tick-exponential' QMLEs is asymptotically normal with the asymptotic covariance matrix which accounts for possible model misspecification. A natural extension of our results is to derive a specification test for conditional quantile models, topic which we leave for future research.

For practical purposes, we have provided an easy-to-implement algorithm for the maximization of the 'tick-exponential' (quasi) log-likelihood as well as a consistent covariance matrix estimator based on the scores. More generally, the estimation method proposed in this paper can be seen as an alternative to the computationally expensive nonlinear quantile regression methods. A more detailed comparison of conditional quantile confidence intervals obtained through our approach with different bootstrap methods, traditionally used in quantile regression, is an interesting empirical topic that we leave for future research.

Notes

[1]The choice of different time subscripts for Y_{t-1} and Z_t depends on whether we want to condition on contemporaneous Z_t.

[2]The author wishes to thank Alain Monfort for pointing out this analogy, which she was unaware of prior to the writing of this paper.

References

Andrews, D. W. K. (1988), "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables", *Econometric Theory*, 4,458-467.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, 31(3), 307–328.

Bollerslev, T., and Wooldridge, J. M. (1992), "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances", *Econometric Reviews*, 11, 143-179.

Buchinsky, M. (1995), "Estimating the asymptotic covariance matrix for quantile regression models - A Monte Carlo study", *Journal of Econometrics*, 68, 303-338.

Chernozhukov, V. and Umantsev, L. (2000), "Conditional Value-at-Risk: Aspects of Modeling and Estimation", *MIT Department of Economics Working Paper*, 01-19.

Daniels, H.E. (1961), "The Asymptotic Efficiency of a Maximum Likelihood Estimator", in *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Vol. 4, Berkeley: University of California Press.

Engle, R.F., and Manganelli, S. (1999), "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles", *UCSD Department of Economics Discussion Paper*, 1999-20.

Fitzenberger, B. (1997), "The Moving Blocks Bootstrap and Robust Inference for Linear Least Squares and Quantile Regressions", *Journal of Econometrics*, 82, 235-287.

Gourieroux, C., Monfort, A. and Trognon, A. (1984), "Pseudo Maximum Likelihood Methods: Theory", *Econometrica*, 52, 681-700.

Gourieroux, C., Monfort, A. and Renault, E. (1987), "Consistent M-Estimators in a Semi-Parametric Model", *Document de Travail de l'INSEE, 8706.*

Huber, P.J. (1967), "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions", in *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Vol. 1, Berkeley: University of California Press.

Kim, T-H. and White, H. (2002), "Estimation, Inference, and Specification Analysis for Possibly Misspecified Quantile Regression", *UCSD Department of Economics Discussion Paper*, 2002-09.

Koenker, R. and Bassett, G. (1978), "Regression Quantiles", *Econometrica*, 46, 33-50.

Koenker, R. and Bassett, G. (1982), "Robust Tests for Heteroskedasticity Based on Regression Quantiles", *Econometrica*, 50, 43-61.

Koenker, R. and Park, B. J. (1996), "An interior point algorithm for nonlinear quantile regression", *Journal of Econometrics*, 71, 265-283.

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

Koenker, R. and Hallock, K.F. (2000), "Quantile Regression: An Introduction", *Presented at the Journal of Economic Perspectives Symposium on Econometric Tools.*

Laplace, P.S. (1774), "Mémoire sur la probabilité des causes par les événements", *Mémoire de l'Academie Royale des Sciences de Paris*, 6, 621-656 (reprinted in *Oeuvres Complètes*, 8, 325-377).

Newey, W. K. and McFadden, D. L. (1994), "Large sample estimation and hypothesis testing", in R. F. Engle and D. L. Newey (eds.), *Handbook of Econometrics*, 4, 2113-2247, Elsevier Science.

Newey, W. K. and Steigerwald, D. G. (1997), "Asymptotic Bias for Quasi-Maximum Likelihood Estimators in Conditional Heteroskedasticity Models", *Econometrica*, 65, 587-599.

Pakes, A. and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators", *Econometrica*, 57, 1027-1057.

Pollard, D. (1985), "New Ways to Prove Central Limit Theorems", *Econometric Theory*, 1, 295-314.

Portnoy, S. (1991), "Behavior of regression quantiles in non-stationary, dependent cases", *Journal of Multivariate Analysis*, 38, 100-113.

Powell, J. L. (1986), "Censored Regression Quantiles", *Journal of Econometrics*, 32, 143-155.

Taylor, J. (1999), "A Quantile Regression Approach to Estimating the Distribution of Multi-Period Returns", *Journal of Derivatives*, Fall, 64-78.

White, H., (1982), "Maximum likelihood estimation of misspecified models", *Econometrica*, 50, 1-25.

White, H. (1994), "Estimation, Inference and Specification Analysis", in J.-M. Grandmont and A. Monfort (eds.), Econometric Society Monographs No. 22, Cambridge University Press.

7. APPENDIX

ASSUMPTIONS:

(A0) there exists a k-vector θ_0 in Θ, $\Theta \subset \mathbb{R}^k$, such that $q_t^\alpha(W_t, \theta_0) = Q_\alpha(Y_t|\mathcal{F}_t)$, where W_t is a vector of variables that are $\mathcal{F}_t$-measurable;

(A1) the parameter space Θ is compact, $\Theta \subset \mathbb{R}^k$;

(A1') the parameter space Θ is compact, $\Theta \subset \mathbb{R}^k$, and θ_0 and θ^* are interior points of Θ;

(A2) the function $a_t : M_t \to \mathbb{R}$ has bounded derivative, i.e. there exists a constant $K > 0$ such that for all $\eta \in M_t$ we have $0 < a_t'(\eta) \leqslant K$;

(A2') the function $a_t : M_t \to \mathbb{R}$ is twice continuously differentiable with bounded derivatives, i.e. there exist constants $K > 0$ and $M \geqslant 0$ such that for all $\eta \in M_t$ we have $0 < a_t'(\eta) \leqslant K$ and $|a_t''(\eta)| \leqslant M$;

(A3) the model $M = \{q_t^\alpha\}$ is such that $\theta_0 \in \Theta$ is identified, i.e. for any $\theta \in \Theta$, $q_t^\alpha(W_t, \theta) = q_t^\alpha(W_t, \theta_0)$ a.s.-P_0, implies $\theta = \theta_0$;

(A4) the function $q_t^\alpha(W_t, \cdot) : \Theta \to \mathbb{R}$ is continuously differentiable on Θ a.s.-P_0, and for each $\theta \in \Theta$, $E[||\nabla_\theta q_t^\alpha(W_t, \theta)||] < \infty$;

(A4') the function $q_t^\alpha(W_t, \cdot) : \Theta \to \mathbb{R}$ is twice continuously differentiable on Θ a.s.-P_0, and there exist some $\delta > 0$ and $\epsilon > 0$ such that, for each $\theta \in \Theta$, $E[||\nabla_\theta q_t^\alpha(W_t, \theta)||^{2+\delta}] < \infty$ and $E[||\nabla_{\theta\theta} q_t^\alpha(W_t, \theta)||^{1+\epsilon}] < \infty$. Moreover, $E[\nabla_\theta q_t^\alpha(W_t, \theta_0) \nabla_\theta q_t^\alpha(W_t, \theta_0)']$ is nonsingular;

(A5) for some $\delta > 0$, $E[|b_t(Y_t)|^{1+\delta}] < \infty$, $E[|c_t(Y_t)|^{1+\delta}] < \infty$ and $E[(\sup_{\theta\in\Theta} |a_t(q_t^\alpha(W_t, \theta))|)^{1+\delta}] < \infty$;

(A6) the true density of Y_t conditional on the information set $\mathcal{F}_t$, $f_{0,t}$, is bounded, i.e. there exists some $C \geqslant 0$ such that $\sup_{y\in\mathbb{R}} f_{0.t}(y) = C < \infty$, and nonzero at $q_t^\alpha(W_t, \theta_0)$ and $q_t^\alpha(W_t, \theta^*)$;

(A7) $\{\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)), \mathcal{F}_t\}$ is an L_1-mixingale;

(A8) $\{s_t^2(Y_t, W_t, \theta), \mathcal{F}_t\}$ and $\{\Delta(Y_t, W_t, \theta), \mathcal{F}_t\}$ are L_1-mixingales, where s_t is the gradient of the tick-exponential log-likelihood $\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))$,

$$s_t(Y_t, W_t, \theta) \equiv [\alpha - H(q_t^\alpha(W_t, \theta) - Y_t)] \cdot a_t'(q_t^\alpha(W_t, \theta)) \cdot \nabla_\theta q_t^\alpha(W_t, \theta),$$

and Δ the Hessian matrix of second derivatives

$$\begin{aligned}\Delta(Y_t, W_t, \theta) \equiv\ & [(\alpha - H(q_t^\alpha(W_t, \theta) - Y_t)) \cdot a_t''(q_t^\alpha(W_t, \theta)) - \delta(q_t^\alpha(W_t, \theta) - Y_t) \cdot a_t'(q_t^\alpha(w_t, \theta))] \\ & \cdot \nabla_\theta q_t^\alpha(w_t, \theta) \cdot \nabla_\theta q_t^\alpha(W_t, \theta)' + a_t'(q_t^\alpha(W_t, \theta)) \cdot \nabla_{\theta\theta} q_t^\alpha(W_t, \theta);\end{aligned}$$

Proof. (Property 2) Differentiating the property *(iii)* in Definition 1 with respect to the variable η, we have, for every $\eta \in M_t$,

$$\exp\{-(1-\alpha)[a_t(\eta) - b_t(\eta)]\} - (1-\alpha)a_t'(\eta)\int_{-\infty}^{\eta} \varphi_t^\alpha(y, \eta)dy = 0,$$

which implies that

$$\exp\{-(1-\alpha)[a_t(\eta) - b_t(\eta)]\} = \alpha(1-\alpha)a_t'(\eta),$$

for every $\eta \in M_t$. Hence the equality *(ii)* in Property 2 is satisfied. Similarly, by combining *(ii)* and *(iii)* in Definition 1 and then differentiating the resulting equation with respect to η, we show that the equality *(iii)* in Property 2 holds, i.e.

$$\exp\{\alpha[a_t(\eta) - c_t(\eta)]\} = \alpha(1-\alpha)a_t'(\eta),$$

for every $\eta \in M_t$. These two equalities in particular imply that $\varphi_t^\alpha(\cdot, \eta)$ is continuous at $y = \eta$, i.e. $(1-\alpha)b_t(\eta) + \alpha c_t(\eta) = a_t(\eta)$, for every $\eta \in M_t$, and that the function a_t is strictly increasing on M_t, i.e. for every $\eta \in M_t$, $a_t'(\eta) > 0$,.which shows that *(i)* and *(iv)* in Property 2 hold. ■

Proof. (Theorem 3) To show that Theorem 3 holds we use the fundamental consistency result for extremum estimators:
The basic consistency theorem: If there is a function $L_0(\theta)$ such that (i) $L_0(\theta)$ is uniquely maximized at θ_0; (ii) Θ is compact; (iii) $L_0(\theta)$ is continuous; (iv) $L_T(\theta)$ converges uniformly in probability to $L_0(\theta)$, then $\hat{\theta}_T \xrightarrow{p} \theta_0$ (see e.g., Theorem 2.1 in Newey and McFadden, 1994, p 2121).
Let $L_T(\theta) \equiv T^{-1}\sum_{t=1}^{T} \ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \theta))$ and $L_0(\theta) \equiv E[\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))]$.

Compactness: the compactness condition *(ii)* is satisfied by imposing (A1).

Uniqueness: we first show that under correct specification of the conditional α-quantile of Y_t (assumption (A0)), θ_0 maximizes $L_0(\theta)$, i.e. $L_0(\theta) \leqslant L_0(\theta_0)$ for all $\theta \in \Theta$.

Recall that we have

$$\begin{aligned}
L_0(\theta) = \;& E[-(1-\alpha)(a_t(q_t^\alpha(W_t,\theta)) - b_t(Y_t)) \cdot 1(Y_t \leqslant q_t^\alpha(W_t,\theta)) \\
& +\alpha(a_t(q_t^\alpha(W_t,\theta)) - c_t(Y_t)) \cdot 1(Y_t > q_t^\alpha(W_t,\theta))] \\
= \;& E[a_t(q_t^\alpha(W_t,\theta)) \cdot (\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta))) \\
& +(1-\alpha)b_t(Y_t) \cdot 1(Y_t \leqslant q_t^\alpha(W_t,\theta)) \\
& -\alpha c_t(Y_t) \cdot 1(Y_t > q_t^\alpha(W_t,\theta))] \\
= \;& E\{a_t(q_t^\alpha(W_t,\theta)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\
& +(1-\alpha)E_t[b_t(Y_t) \cdot 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\
& -\alpha E_t[c_t(Y_t) \cdot 1(Y_t > q_t^\alpha(W_t,\theta))]\},
\end{aligned}$$

so that we need to show

$$\begin{gathered}
a_t(q_t^\alpha(W_t,\theta)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\
+(1-\alpha)E_t[b_t(Y_t) \cdot 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\
-\alpha E_t[c_t(Y_t) \cdot 1(Y_t > q_t^\alpha(W_t,\theta))] \\
\leqslant \\
(1-\alpha)E_t[b_t(Y_t) \cdot 1(Y_t \leqslant q_t^\alpha(W_t,\theta_0))] \\
-\alpha E_t[c_t(Y_t) \cdot 1(Y_t > q_t^\alpha(W_t,\theta_0))],\; a.s - P_0,
\end{gathered}$$

i.e.

$$\begin{gathered}
a_t(q_t^\alpha(W_t,\theta)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\
\leqslant \\
(1-\alpha)E_t[b_t(Y_t) \cdot (1(Y_t \leqslant q_t^\alpha(W_t,\theta_0)) - 1(Y_t \leqslant q_t^\alpha(W_t,\theta)))] \\
-\alpha E_t[c_t(Y_t) \cdot (1(Y_t > q_t^\alpha(W_t,\theta_0)) - 1(Y_t > q_t^\alpha(W_t,\theta)))],\; a.s. - P_0.
\end{gathered}$$

Under correct specification $q_t^\alpha(W_t, \theta_0) = Q_\alpha(Y_t|\mathcal{F}_t)$ (assumption (A0)) and $E_t[1(Y_t \leqslant q_t^\alpha(W_t, \theta_0))] = \alpha$ so that the previous inequality becomes

$$
\begin{aligned}
(10) \qquad & a_t(q_t^\alpha(W_t, \theta)) \cdot E_t[d_t(Y_t, W_t, \theta, \theta_0)] \leqslant \\
& E_t[((1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)) \cdot d_t(Y_t, W_t, \theta, \theta_0)], \ a.s. - P_0,
\end{aligned}
$$

where $d_t(Y_t, W_t, \theta, \theta_0) \equiv 1(Y_t \leqslant q_t^\alpha(W_t, \theta_0)) - 1(Y_t \leqslant q_t^\alpha(W_t, \theta))$. We now show that inequality (10) holds: first, consider the sets $A_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega), \theta_0) \leqslant q_t^\alpha(W_t(\omega), \theta)\}$ and $B_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega), \theta_0) < Y_t(\omega) \leqslant q_t^\alpha(W_t(\omega), \theta)\}$. We have $d_t(Y_t, W_t, \theta, \theta_0) = 1(q_t^\alpha(W_t, \theta_0) < Y_t \leqslant q_t^\alpha(W_t, \theta)) = -1$ on $A_t \cap B_t$, and $= 0$ on $A_t \cap B_t^c$. Moreover, by continuity of $\varphi_t^\alpha(\cdot, \eta)$, we have $(1-\alpha)b_t(Y_t) + \alpha c_t(Y_t) = a_t(Y_t) \leqslant a_t(q_t^\alpha(W_t, \theta))$ on $A_t \cap B_t$ (recall that $a_t' > 0$) so that

$$
\begin{aligned}
a_t(q_t^\alpha(W_t, \theta)) \cdot d_t(Y_t, W_t, \theta, \theta_0) &\leqslant ((1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)) \cdot d_t(Y_t, W_t, \theta, \theta_0), \text{ on } A_t \cap B_t, \\
a_t(q_t^\alpha(W_t, \theta)) \cdot d_t(Y_t, W_t, \theta, \theta_0) &= ((1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)) \cdot d_t(Y_t, W_t, \theta, \theta_0) = 0, \text{ on } A_t \cap B_t^c.
\end{aligned}
$$

Next, consider $A_t^c = \{\omega \in \Omega : q_t^\alpha(W_t(\omega), \theta_0) > q_t^\alpha(W_t(\omega), \theta)\}$ and $C_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega), \theta) < Y_t(\omega) \leqslant q_t^\alpha(W_t(\omega), \theta_0)\}$. Similarly, $d_t(Y_t, W_t, \theta, \theta_0) = 1(q_t^\alpha(W_t, \theta) < Y_t \leqslant q_t^\alpha(W_t, \theta_0)) = 1$ on $A_t^c \cap C_t$, and $= 0$ on $A_t^c \cap C_t^c$. Now we have $(1-\alpha)b_t(Y_t) + \alpha c_t(Y_t) = a_t(Y_t) > a_t(q_t^\alpha(W_t, \theta))$ on $A_t^c \cap C_t$ so that

$$
\begin{aligned}
a_t(q_t^\alpha(W_t, \theta)) \cdot d_t(Y_t, W_t, \theta, \theta_0) &< ((1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)) \cdot d_t(Y_t, W_t, \theta, \theta_0), \text{ on } A_t^c \cap C_t, \\
a_t(q_t^\alpha(W_t, \theta)) \cdot d_t(Y_t, W_t, \theta, \theta_0) &= ((1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)) \cdot d_t(Y_t, W_t, \theta, \theta_0) = 0, \text{ on } A_t^c \cap C_t^c.
\end{aligned}
$$

We conclude that

$$
a_t(q_t^\alpha(W_t, \theta)) \cdot d_t(Y_t, W_t, \theta, \theta_0) \leqslant ((1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)) \cdot d_t(Y_t, W_t, \theta, \theta_0), \ a.s. - P_0,
$$

which in turn implies that inequality (10) holds and that θ_0 is a maximizer of $L_0(\theta)$.
We now show that it is unique: note that the previous inequality becomes an equality if and only if $A_t \cap B_t = \emptyset$ and $A_t^c \cap C_t = \emptyset$, i.e. if and only if $B_t = C_t = \emptyset$. The variable Y_t

being continuous, we have that $B_t = C_t = \emptyset$ if and only if $q_t^\alpha(W_t, \theta) = q_t^\alpha(W_t, \theta_0)$ a.s.-P_0. Assumption (A3) implies that $q_t^\alpha(W_t, \theta) = q_t^\alpha(W_t, \theta_0)$ a.s.-P_0 if and only if $\theta = \theta_0$, so that we have uniqueness.

Uniform convergence: we next show that both the continuity condition *(iii)* and the uniform convergence condition *(iv)* hold by using a weak form of the uniform law of large numbers (ULLN).

The weak ULLN theorem: If the function $\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))$ is Lipshitz-L_1 a.s. on Θ and $\{\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)), F_t\}$ is a uniformly integrable L_1-mixingale, then $L_T(\theta)$ converges uniformly in probability to $L_0(\theta)$ and $L_0(\theta)$ is continuous on Θ (see e.g., Theorem A.2.9 in White, 1994, p 355).

We start by showing that $\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))$ is Lipshitz-L_1 a.s. on Θ. First, recall the definition of a function that is Lipshitz-L_1 a.s. on Θ: for each $\theta_0 \in \Theta$, there exists a constant $\delta_0 > 0$ and an $\mathcal{F}_t$-measurable function $L_t^0 : \Omega \to \mathbb{R}^+$ such that for all θ such that $||\theta - \theta_0|| \leqslant \delta_0$, we have

$$(11) \quad |\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) - \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta_0))| \leqslant L_t^0 ||\theta - \theta_0||, \ a.s. - P_0.$$

Moreover, L_t^0 has to satisfy: $T^{-1}\sum_{t=1}^T E[L_t^0] < \infty$ (see e.g., Definition A.2.3 in White, 1994, p 352). By assumption (A4), $q_t^\alpha(W_t, \cdot)$ is continuous a.s. on Θ, i.e. for each $\theta_0 \in \Theta$ and for each $\varepsilon > 0$ there exists $\delta_{\varepsilon,0} > 0$ such that for $||\theta - \theta_0|| < \delta_{\varepsilon,0}$, $q_t^\alpha(W_t, \theta_0)$ and $q_t^\alpha(W_t, \theta)$ are 'sufficiently' close, meaning that if $Y_t \gtrless q_t^\alpha(W_t, \theta_0)$ a.s.-P_0 then $Y_t \gtrless q_t^\alpha(W_t, \theta)$ a.s.-P_0.
CASE 1: if $Y_t < q_t^\alpha(W_t, \theta_0)$ a.s.-P_0, then for $||\theta - \theta_0|| < \delta_{\varepsilon,0}$ we have

$$\begin{aligned}
&|\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) - \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta_0))| \\
&\quad = (1-\alpha) \cdot |a_t(q_t^\alpha(W_t, \theta_0)) - a_t(q_t^\alpha(W_t, \theta))|, \ a.s. - P_0 \\
&\quad = (1-\alpha) \cdot a_t'(q_t^\alpha(W_t, \bar{\theta}_0)) \cdot |\nabla_\theta q_t^\alpha(W_t, \bar{\theta}_0)'(\theta - \theta_0)|, \ a.s. - P_0
\end{aligned}$$

for some $\bar{\theta}_0 \equiv c\theta + (1-c)\theta_0$, $c \in (0,1)$. Note that by assumption (A2) a_t' is bounded on M_t, i.e. there exists a constant $K > 0$ such that $a_t' \leqslant K$ on M_t, and that by assumption (A4) $E[||\nabla_\theta q_t^\alpha(W_t, \theta)||] < \infty$. Therefore

$$|\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) - \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta_0))| \leqslant K \cdot ||\nabla_\theta q_t^\alpha(W_t, \bar{\theta}_0)|| \cdot ||\theta - \theta_0||, \ a.s. - P_0.$$

We have thus constructed an $\mathcal{F}_t$-measurable function L_t^0, $L_t^0 \equiv K||\nabla_\theta q_t^\alpha(W_t, \bar{\theta}_0)||$, such that the inequality (11) holds, and satisfying $T^{-1}\sum_{t=1}^T E[L_t^0] < \infty$. CASE 2: if $Y_t > q_t^\alpha(W_t, \theta_0)$ a.s.-P_0, then by a similar reasoning we can show that the same result holds. Thus, we conclude that $\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))$ is Lipshitz-L_1 a.s. on Θ.

Next, the requirement that $\{\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)), \mathcal{F}_t\}$ be a uniformly integrable L_1-mixingale, is treated in two steps. First, by assumption (A7) we know that $\{\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)), \mathcal{F}_t\}$ is a L_1-mixingale. Examples of dependent processes that are L_1-mixingales can be found in Andrews (1988) and White (1994). We second need to show that the uniform integrability condition holds (see e.g., Definition A.2.8 in White, 1994, p 354). According to White (1994), the uniform integrability condition will be satisfied under mild domination conditions, such as $E[(\sup_{\theta\in\Theta} |\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))|)^{1+\delta}] < \infty$ for some $\delta > 0$. This domination condition is implied by assumption (A5) since

$$|\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))| \leqslant |a_t(q_t^\alpha(W_t, \theta))| + (1-\alpha)|b_t(Y_t)| + \alpha|c_t(Y_t)|, \ a.s. - P_0,$$

so that

$$\sup_{\theta\in\Theta} |\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))| \leqslant (1-\alpha)|b_t(Y_t)| + \alpha|c_t(Y_t)| + \sup_{\theta\in\Theta} |a_t(q_t^\alpha(W_t, \theta))|, \ a.s. - P_0.$$

For a given $\delta > 0$, there exists a constant $n_\delta > 1$ such that

$$\begin{aligned} &(\sup_{\theta\in\Theta} |\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))|)^{1+\delta} \\ &\leqslant \max(1, n_\delta[(1-\alpha)^{1+\delta}|b_t(Y_t)|^{1+\delta} + \alpha^{1+\delta}|c_t(Y_t)|^{1+\delta} + (\sup_{\theta\in\Theta} |a_t(q_t^\alpha(W_t, \theta))|)^{1+\delta}]), \ a.s. - P_0. \end{aligned}$$

Thus, $E[|b_t(Y_t)|^{1+\delta}] < \infty$, $E[|c_t(Y_t)|^{1+\delta}] < \infty$ and $E[(\sup_{\theta\in\Theta} |a_t(q_t^\alpha(W_t, \theta))|)^{1+\delta}] < \infty$ (assumption (A5)) imply $E[(\sup_{\theta\in\Theta} |\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))|)^{1+\delta}] < \infty$, so that the uniform integrability condition holds.

Finally, we can apply the basic theorem for consistency to show that $\hat{\theta}_T \xrightarrow{p} \theta_0$, which completes the proof of Theorem 3. ■

PROOF. (COROLLARY 4) To show that Corollary 4 holds we need to check that $L_0(\theta)$

is uniquely maximized at θ^*. The remainder of the proof is then identical to the one for Theorem 3.

Recall that we have

$$\begin{aligned} L_0(\theta) = \;& E\{a_t(q_t^\alpha(W_t,\theta)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\ & +(1-\alpha)E_t[b_t(Y_t) \cdot 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\ & -\alpha E_t[c_t(Y_t) \cdot 1(Y_t > q_t^\alpha(W_t,\theta))]\}. \end{aligned}$$

Hence, we need to show that

$$\begin{gathered} a_t(q_t^\alpha(W_t,\theta)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\ +(1-\alpha)E_t[b_t(Y_t) \cdot 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\ -\alpha E_t[c_t(Y_t) \cdot 1(Y_t > q_t^\alpha(W_t,\theta))] \\ = \\ a_t(q_t^\alpha(W_t,\theta^*)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta^*))] \\ (1-\alpha)E_t[b_t(Y_t) \cdot 1(Y_t \leqslant q_t^\alpha(W_t,\theta^*))] \\ -\alpha E_t[c_t(Y_t) \cdot 1(Y_t > q_t^\alpha(W_t,\theta^*))],\; a.s - P_0, \end{gathered}$$

implies $\theta = \theta^*$. Note that the previous equality can be written

$$\begin{gathered} a_t(q_t^\alpha(W_t,\theta)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\ -a_t(q_t^\alpha(W_t,\theta^*)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta^*))] \\ = \\ (1-\alpha)E_t[b_t(Y_t) \cdot (1(Y_t \leqslant q_t^\alpha(W_t,\theta^*)) - 1(Y_t \leqslant q_t^\alpha(W_t,\theta)))] \\ -\alpha E_t[c_t(Y_t) \cdot (1(Y_t > q_t^\alpha(W_t,\theta^*)) - 1(Y_t > q_t^\alpha(W_t,\theta)))],\; a.s. - P_0. \end{gathered} \tag{12}$$

As previously, let $d_t(Y_t, W_t, \theta, \theta^*) \equiv 1(Y_t \leqslant q_t^\alpha(W_t, \theta^*)) - 1(Y_t \leqslant q_t^\alpha(W_t, \theta))$. We then have

$$
\begin{gathered}
a_t(q_t^\alpha(W_t,\theta)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta))] \\
-a_t(q_t^\alpha(W_t,\theta^*)) \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta^*))] \\
= \\
\alpha[a_t(q_t^\alpha(W_t,\theta)) - a_t(q_t^\alpha(W_t,\theta^*))] \\
+a_t(q_t^\alpha(W_t,\theta)) \cdot E_t[d_t(Y_t,W_t,\theta,\theta^*)] \\
-[a_t(q_t^\alpha(W_t,\theta)) - a_t(q_t^\alpha(W_t,\theta^*))] \cdot E_t[1(Y_t \leqslant q_t^\alpha(W_t,\theta^*))],
\end{gathered}
$$

so that (12) becomes

$$
\begin{gathered}
[a_t(q_t^\alpha(W_t,\theta)) - a_t(q_t^\alpha(W_t,\theta^*))] \cdot E_t[\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta^*))] \\
+a_t(q_t^\alpha(W_t,\theta)) \cdot E_t[d_t(Y_t,W_t,\theta,\theta^*)] \\
= \\
E_t[((1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)) \cdot d_t(Y_t,W_t,\theta,\theta^*)],\ a.s. - P_0.
\end{gathered} \tag{13}
$$

First, consider the sets $A_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega),\theta^*) \leqslant q_t^\alpha(W_t(\omega),\theta)\}$ and $B_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega),\theta^*) < Y_t(\omega) \leqslant q_t^\alpha(W_t(\omega),\theta)\}$. We have $d_t(Y_t,W_t,\theta,\theta^*) = 1(q_t^\alpha(W_t,\theta^*) < Y_t \leqslant q_t^\alpha(W_t,\theta)) = -1$ on $A_t \cap B_t$, and $= 0$ on $A_t \cap B_t^c$. Hence, (13) becomes

$$
\begin{aligned}
&[a_t(q_t^\alpha(W_t,\theta)) - a_t(q_t^\alpha(W_t,\theta^*))] \cdot \alpha \\
&\qquad\qquad = a_t(q_t^\alpha(W_t,\theta)) - [(1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)],\ \text{on } A_t \cap B_t,
\end{aligned}
$$

and

$$
[a_t(q_t^\alpha(W_t,\theta)) - a_t(q_t^\alpha(W_t,\theta^*))] \cdot [\alpha - 1(Y_t \leqslant q_t^\alpha(W_t,\theta^*))] = 0,\ \text{on } A_t \cap B_t^c.
$$

Next, consider $A_t^c = \{\omega \in \Omega : q_t^\alpha(W_t(\omega),\theta^*) > q_t^\alpha(W_t(\omega),\theta)\}$ and $C_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega),\theta) < Y_t(\omega) \leqslant q_t^\alpha(W_t(\omega),\theta^*)\}$. Similarly, $d_t(Y_t,W_t,\theta,\theta^*) = 1(q_t^\alpha(W_t,\theta) < Y_t \leqslant$

$q_t^\alpha(W_t, \theta^*)) = 1$ on $A_t^c \cap C_t$, and $= 0$ on $A_t^c \cap C_t^c$, so that (13) becomes

$$[a_t(q_t^\alpha(W_t, \theta)) - a_t(q_t^\alpha(W_t, \theta^*))] \cdot [\alpha - 1]$$
$$= [(1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)] - a_t(q_t^\alpha(W_t, \theta)), \text{ on } A_t^c \cap C_t,$$

and

$$[a_t(q_t^\alpha(W_t, \theta)) - a_t(q_t^\alpha(W_t, \theta^*))] \cdot [\alpha - 1(Y_t \leqslant q_t^\alpha(W_t, \theta^*))] = 0, \text{ on } A_t^c \cap C_t^c.$$

In particular, these equalities imply that

$$a_t(q_t^\alpha(W_t, \theta)) - a_t(q_t^\alpha(W_t, \theta^*)) = 0, \text{ on } A_t \cap B_t^c \text{ and } A_t^c \cap C_t^c,$$

which in turn implies

(14) $q_t^\alpha(W_t, \theta) = q_t^\alpha(W_t, \theta^*)$, on $A_t \cap B_t^c$ and $A_t^c \cap C_t^c$,

since $a_t' > 0$. The equality (14) implies that $A_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega), \theta^*) = q_t^\alpha(W_t(\omega), \theta)\}$ and $B_t = A_t^c = \emptyset$, so that $q_t^\alpha(W_t, \theta) = q_t^\alpha(W_t, \theta^*)$, $a.s. - P_0$. Assumption (A3) implies that $q_t^\alpha(W_t, \theta) = q_t^\alpha(W_t, \theta^*)$ a.s.-P_0 if and only if $\theta = \theta^*$, so that θ^* is the unique maximizer of $L_0(\theta)$. ■

Proof. (Theorem 5) The following proof is inspired by Gourieroux, Monfort, and Trognon (1984). We show that the Theorem 5 already holds for P_0 such that Y_t is iid and $q_t^\alpha(W_t, \theta_0) = \theta_0$. In this case $l_t(Y_t, q_t^a(W_t, \theta)) = l(Y_t, \theta)$ and $\Theta \subset \mathbb{R}$. The log-likelihood function $\ln l(Y_t, \cdot)$ being continuously differentiable on Θ a.s.-P_0, we know that for every $y \in A$, $P_0(A) = 1$, $\ln l(y, \cdot)$ is continuously differentiable for all $\theta \in \Theta$. Since $L_0(\theta) \equiv E[\ln l(Y_t, \theta)] = \int_R \ln l(y, \theta) dP_0(y) = \int_A \ln l(y, \theta) dP_0(y)$, the expected log-likelihood $L_0(\theta)$ is continuously differentiable on Θ, and the first order condition (FOC) holds, i.e. $dL_0/d\theta|_{\theta=\theta_0} = 0$.
Suppose that the support of Y_t consists of y_1 and y_2, such that $-\infty < y_1 \leqslant \theta_0 < y_2 < +\infty$. Let $p_0 \equiv P_0(Y_t = y_1)$. Since $\alpha = P_0(Y_t \leqslant \theta_0)$ we have $p_0 = \alpha$. The expected log-likelihood, $L_0(\theta)$, is then $L_0(\theta) = [\alpha \ln l(y_1, \theta) + (1-\alpha) \ln l(y_2, \theta)]$. The FOC can be written

(15) $$\alpha \left. \frac{\partial \ln l(y_1, \theta)}{\partial \theta} \right|_{\theta=\theta_0} + (1-\alpha) \left. \frac{\partial \ln l(y_2, \theta)}{\partial \theta} \right|_{\theta=\theta_0} = 0.$$

First, consider y_1 as fixed: then there exists a constant $\phi_1(\theta_0) \in \mathbb{R}$ such that for every $y_2 > \theta_0$, we have $\partial \ln l(y_2, \theta)/\partial\theta|_{\theta=\theta_0} = -(\alpha/(1-\alpha))\phi_1(\theta_0)$. Similarly, by fixing y_2 and varying y_1, we conclude that there exists a constant $\phi_2(\theta_0)$ such that for every $y_1 \leqslant \theta_0$, we have $\partial \ln l(y_1, \theta)/\partial\theta|_{\theta=\theta_0} = -((1-\alpha)/\alpha)\phi_2(\theta_0)$. Thus, the FOC (15) becomes $(1-\alpha)\phi_2(\theta_0) + \alpha\phi_1(\theta_0) = 0$. Let $\phi(\theta_0) \equiv \phi_2(\theta_0)/\alpha = -\phi_1(\theta_0)/(1-\alpha)$. We then have

$$(16)\quad \partial \ln l(y, \theta)/\partial\theta|_{\theta=\theta_0} = \begin{cases} -(1-\alpha)\phi(\theta_0), & \text{if } y \leqslant \theta_0, \\ \alpha\phi(\theta_0), & \text{if } y > \theta_0. \end{cases}$$

By integrating both parts of (16) with respect to θ we obtain

$$(17)\quad \ln l(y, \theta_0) = \begin{cases} -(1-\alpha)[a(\theta_0) - b(y)], & \text{if } y \leqslant \theta_0, \\ \alpha[a(\theta_0) - c(y)], & \text{if } y > \theta_0, \end{cases}$$

where the continuity of $\ln l(\cdot, \theta_0)$ at $y = \theta_0$ implies $(1-\alpha)b(\theta_0) + \alpha c(\theta_0) = a(\theta_0)$. ■

Proof. (Theorem 6) To show that Theorem 6 holds we use the following result adopted from Newey and McFadden (1994):

The basic asymptotic normality theorem: Let φ_t^α be a tick-exponential density of order α, as defined in Definition 1 and $\hat{\theta}_T$ the corresponding QMLE,

$$\hat{\theta}_T = \arg\max_{\theta\in\Theta} T^{-1} \sum_{t=1}^{T} \ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \theta)).$$

Assume that $\hat{\theta}_T \xrightarrow{p} \theta_0$ and that $\sqrt{T}\nabla_\theta L_T(\hat{\theta}_T) \xrightarrow{p} 0$. Suppose that (i) θ_0 is an interior point of Θ; (ii) $E[\nabla_\theta \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta_0))] = 0$; (iii) there is $\Delta(Y_t, W_t, \theta_0)$ such that with probability one, $r(Y_t, W_t, \theta) \equiv ||\nabla_\theta \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) - \nabla_\theta \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta_0)) - \Delta(Y_t, W_t, \theta_0)(\theta - \theta_0)||/||\theta - \theta_0|| \to 0$ as $\theta \to \theta_0$; (iv) $T^{-1}\sum_{t=1}^T \Delta(Y_t, W_t, \theta_0) \xrightarrow{p} E[\Delta(Y_t, W_t, \theta_0)] \equiv \Delta_0$; (v) there is $\varepsilon > 0$ such that $E[\sup_{||\theta-\theta_0||<\varepsilon} r(Y_t, W_t, \theta)] < \infty$; (vi) $\sqrt{T}\nabla_\theta L_T(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_0)$. Then $\sqrt{T}(\hat{\theta}_T - \theta_0) \to \mathcal{N}(0, \Delta_0^{-1}\Sigma_0\Delta_0^{-1})$ (see Theorems 7.2 and 7.3 in Newey and McFadden, 1994, p 2186-2188).

As previously, $L_T(\theta) = T^{-1}\sum_{t=1}^T \ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \theta))$ and $L_0(\theta) = E[\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))]$.

Asymptotic first order condition: we start by showing that the asymptotic first order condition, $\sqrt{T}\nabla_\theta L_T(\hat{\theta}_T) \xrightarrow{p} 0$, holds. From equation (5) we have $\nabla_\theta L_T(\theta) = T^{-1}\sum_{t=1}^T s_t(y_t, w_t, \theta)$, where $s_t(y_t, w_t, \theta)$ is given by

$$s_t(y_t, w_t, \theta) \equiv [\alpha - H(q_t^\alpha(w_t, \theta) - y_t)] \cdot a_t'(q_t^\alpha(w_t, \theta)) \cdot \nabla_\theta q_t^\alpha(w_t, \theta).$$

We use the approach by Ruppert and Carroll (1980) (see their proof of Lemma A.2). Let

$$L_{T,\theta,j}(a) \equiv T^{-1} \sum\nolimits_{t=1}^{T} \ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \hat{\theta}_T + ae_j),$$

where $\{e_j\}_{j=1}^k$ is the standard basis of $\mathbb{R}^k$, and $a \in \mathbb{R}$ is such that for all $j = 1, \ldots, k$, $\hat{\theta}_T + ae_j \in \Theta$. Also, let $G_{T,\theta,j}(a)$ be the derivative form right of $L_{T,\theta,j}(a)$, so that

$$G_{T,\theta,j}(a) = T^{-1} \sum\nolimits_{t=1}^{T} [\alpha - H(q_t^\alpha(w_t, \hat{\theta}_T + ae_j) - y_t)] \cdot a_t'(q_t^\alpha(w_t, \hat{\theta}_T + ae_j)) \cdot \frac{\partial}{\partial \theta_j} q_t^\alpha(w_t, \hat{\theta}_T + ae_j).$$

By the same argument as in Ruppert and Carroll (1980), we have

$$|G_{T,\theta,j}(0)| \leqslant T^{-1} \sum\nolimits_{t=1}^{T} |a_t'(q_t^\alpha(w_t, \hat{\theta}_T))| \cdot |\frac{\partial}{\partial \theta_j} q_t^\alpha(w_t, \hat{\theta}_T)| \cdot 1(y_t = q_t^\alpha(w_t, \hat{\theta}_T)).$$

By assumption (A2') a_t' is bounded, $0 < a_t' \leqslant K$, so that

$$\sqrt{T} |G_{T,\theta,j}(0)| \leqslant K T^{-1/2} \max_{1 \leqslant t \leqslant T} |\frac{\partial}{\partial \theta_j} q_t^\alpha(w_t, \hat{\theta}_T)| \cdot \sum\nolimits_{t=1}^{T} 1(y_t = q_t^\alpha(w_t, \hat{\theta}_T)).$$

Furthermore, $T^{-1/2} \max_{1 \leqslant t \leqslant T} |\frac{\partial}{\partial \theta_j} q_t^\alpha(W_t, \hat{\theta}_T)| = o_p(1)$ by assumption (A4'). Now, note that $\sqrt{T} \nabla_\theta L_T(\hat{\theta}_T) = \sqrt{T} G_{T,\theta}(0)$. Since Y_t is a continuous random variable, we have $\sum_{t=1}^T 1(Y_t = q_t^\alpha(W_t, \hat{\theta}_T)) = O_p(1)$, which ensures that $\sqrt{T} \nabla_\theta L_T(\hat{\theta}_T) \xrightarrow{p} 0$, and completes the proof of the asymptotic first order condition.

We now show that all other conditions of the basic asymptotic normality theorem are verified. Condition (i) is an assumption of Theorem 6. Condition (ii) holds since $L_T(\theta)$ is continuously differentiable on Θ with probability one. We now check for conditions (iii)-(v), which ensure the stochastic equicontinuity of the gradient $\nabla_\theta L_T(\theta)$.

Stochastic equicontinuity: note that the gradient of the tick-exponential log-likelihood can be written $s_t(y_t, w_t, \theta) = [\alpha - H(q_t^\alpha(w_t, \theta) - y_t)] \cdot g_t(w_t, \theta)$, where

$$g_t(w_t, \theta) \equiv a_t'(q_t^\alpha(w_t, \theta)) \cdot \nabla_\theta q_t^\alpha(w_t, \theta).$$

The function $g_t(W_t, \cdot) : \Theta \to \mathbb{R}^k$ is continuously differentiable a.s. on Θ with "derivative" $dg_t(W_t, \cdot) : \Theta \to \mathbb{R}^{k \times k}$ given by

$$dg_t(W_t, \theta) \equiv a_t''(q_t^\alpha(W_t, \theta)) \cdot \nabla_\theta q_t^\alpha(W_t, \theta) \nabla_\theta q_t^\alpha(W_t, \theta)' + a_t'(q_t^\alpha(W_t, \theta)) \cdot \nabla_{\theta\theta} q_t^\alpha(W_t, \theta).$$

Note that the existence of a_t'' is guaranteed by assumption (A2'). Similarly, by (A4') we assume that the Hessian $\nabla_{\theta\theta} q_t^\alpha(W_t, \theta)$ exists for every $\theta \in \Theta$, a.s.-P_0. The "gradient" of $s_t(Y_t, W_t, \cdot)$, i.e. the Hessian of $\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \cdot))$, is the function $\Delta(Y_t, W_t, \cdot) : \Theta \to \mathbb{R}^{k \times k}$ such that

$$\begin{aligned} \Delta(Y_t, W_t, \theta) \quad \equiv \quad & -\delta(q_t^\alpha(W_t, \theta) - Y_t) \cdot g_t(W_t, \theta) \cdot \nabla_\theta q_t^\alpha(W_t, \theta)' \\ & + [\alpha - H(q_t^\alpha(W_t, \theta) - Y_t)] \cdot dg_t(W_t, \theta), \end{aligned}$$

where $\delta(\cdot)$ represents the Dirac function, i.e. $\delta(x) = 0$ if $x \neq 0$ and $\int_{\mathbb{R}} \delta(x) dx = 1$. Alternatively, for any integrable function $\phi : \mathbb{R} \to \mathbb{R}$, the Dirac function verifies $\int_{\mathbb{R}} \delta(x) \phi(x) dx = \phi(0)$. The function $\delta(\cdot)$ is the derivative of $H(\cdot)$, so that we have $|H(x+\varepsilon) - H(x) - \varepsilon\delta(x)| = o(|\varepsilon|)$ for all $x \in \mathbb{R}$. We will show that $\Delta(Y_t, W_t, \cdot)$ is the "gradient" of $s_t(Y_t, W_t, \cdot)$ in a neighborhood of θ_0, in a sense that $||s_t(Y_t, W_t, \theta) - s_t(Y_t, W_t, \theta_0) - \Delta(Y_t, W_t, \theta_0)'(\theta - \theta_0)|| = o_p(||\theta - \theta_0||)$. Let

$$r(Y_t, W_t, \theta) \equiv ||s_t(Y_t, W_t, \theta) - s_t(Y_t, W_t, \theta_0) - \Delta(Y_t, W_t, \theta_0)'(\theta - \theta_0)|| / ||\theta - \theta_0||.$$

In order to simplify the notation, let $X_t \equiv q_t^\alpha(W_t, \theta_0) - Y_t$ and $\varepsilon_t \equiv q_t^\alpha(W_t, \theta) - q_t^\alpha(W_t, \theta_0)$.

Thus

$$
\begin{aligned}
r(Y_t, W_t, \theta_0) &= ||[\alpha - H(X_t + \varepsilon_t)] \cdot g_t(W_t, \theta) - [\alpha - H(X_t)] \cdot g_t(W_t, \theta_0) \\
&\quad + \{\delta(X_t) \cdot g_t(W_t, \theta_0) \cdot \nabla_\theta q_t^\alpha(W_t, \theta_0)' \\
&\quad - [\alpha - H(X_t)] \cdot dg_t(W_t, \theta_0)\}'(\theta - \theta_0)||/||\theta - \theta_0|| \\
&= ||[\alpha - H(X_t)] \cdot [g_t(W_t, \theta) - g_t(W_t, \theta_0) - dg_t(W_t, \theta_0)'(\theta - \theta_0)] \\
&\quad + [\alpha - H(X_t + \varepsilon_t)] \cdot g_t(W_t, \theta) - [\alpha - H(X_t)] \cdot g_t(W_t, \theta) + \varepsilon_t \cdot \delta(X_t) \cdot g_t(W_t, \theta) \\
&\quad + \delta(X_t) \cdot \nabla_\theta q_t^\alpha(W_t, \theta_0) \cdot g_t(W_t, \theta_0)'(\theta - \theta_0) \\
&\quad - \delta(X_t) \cdot [q_t^\alpha(W_t, \theta) - q_t^\alpha(W_t, \theta_0)] \cdot g_t(W_t, \theta)||/||\theta - \theta_0||,
\end{aligned}
$$

so that

$$
\begin{aligned}
&(18) \quad r(Y_t, W_t, \theta_0) \leqslant ||g_t(W_t, \theta) - g_t(W_t, \theta_0) - dg_t(W_t, \theta_0)'(\theta - \theta_0)||/||\theta - \theta_0|| \\
&(19) \quad\quad + |H(X_t + \varepsilon_t) - H(X_t) - \varepsilon_t \cdot \delta(X_t)| \cdot ||g_t(W_t, \theta)||/||\theta - \theta_0|| \\
&(20) \quad\quad + \delta(X_t) \cdot ||\nabla_\theta q_t^\alpha(W_t, \theta_0) \cdot g_t(W_t, \theta_0)'(\theta - \theta_0)||/||\theta - \theta_0|| \\
&(21) \quad\quad + \delta(X_t) \cdot ||[q_t^\alpha(W_t, \theta) - q_t^\alpha(W_t, \theta_0)] \cdot g_t(W_t, \theta)||/||\theta - \theta_0||
\end{aligned}
$$

Since the function $g_t(W_t, \cdot)$ is continuously differentiable a.s. on Θ with gradient $dg_t(W_t, \cdot)$, the first term (18) of the right hand side of the previous inequality is $o_p(1)$. For the second term note that,

$$
\begin{aligned}
(19) &= |a_t'(q_t^\alpha(W_t, \theta))| \cdot ||\nabla_\theta q_t^\alpha(W_t, \theta)|| \cdot |q_t^\alpha(W_t, \theta_0) - q_t^\alpha(W_t, \theta)|/||\theta - \theta_0|| \\
&\quad \cdot |H(X_t + \varepsilon_t) - H(X_t) - \varepsilon_t \cdot \delta(X_t)|/|\varepsilon_t|.
\end{aligned}
$$

By assumption (A2'), a_t' is bounded by some positive constant K so that $a_t'(q_t^\alpha(W_t, \theta)) \leqslant K$, a.s-$P_0$, and by assumption (A4'), $q_t^\alpha(W_t, \cdot)$ is continuously differentiable a.s. on Θ so that $|q_t^\alpha(W_t, \theta_0) - q_t^\alpha(W_t, \theta)|/||\theta - \theta_0|| = O_p(1)$, which implies that the first term of the previous equality is $O_p(1)$. Moreover, for all $x \in \mathbb{R}$ we have $|H(x + \varepsilon) - H(x) - \varepsilon \cdot \delta(x)|/|\varepsilon| = o(1)$ so that the term (19) is $o_p(1)$ as well. For the remaining terms note that

$$\begin{aligned}(20) \quad &\leqslant \delta(X_t) \cdot ||\nabla_\theta q_t^\alpha(W_t, \theta_0)|| \cdot |g_t(W_t, \theta_0)'(\theta - \theta_0)|/||\theta - \theta_0|| \\ &\leqslant \delta(X_t) \cdot ||\nabla_\theta q_t^\alpha(W_t, \theta_0)|| \cdot ||g_t(W_t, \theta_0)||,\end{aligned}$$

and

$$(21) \leqslant \delta(X_t) \cdot ||g_t(W_t, \theta)|| \cdot ||q_t^\alpha(W_t, \theta) - q_t^\alpha(W_t, \theta_0)]||/||\theta - \theta_0||,$$

so that (by abuse of notation) both (20) and (21) can be written $\delta(X_t) \cdot O_p(1)$. Now, note that $P_0(X_t = 0) = 0$, so that $P_0(\delta(X_t) \neq 0) = 0$. By combining all those results we conclude that $r(Y_t, W_t, \theta_0) \xrightarrow{p} 0$ as $\theta \to \theta_0$ so that condition (iii) is verified.
Moreover, both (18) and (19) are bounded by some positive constants so that $E[\sup_{||\theta - \theta^*|| < \epsilon} r(Y_t, W_t, \theta_0)] < \infty$, which corresponds to condition (iv) of the basic theorem.
Also, we have $\sum_{t=0}^{T-1} \Delta(Y_t, W_t, \theta)/T \xrightarrow{p} \Delta \equiv E[\Delta(Y_t, W_t, \theta)]$ with

$$\begin{aligned}\Delta &= E[-\delta(q_t^\alpha(W_t, \theta) - Y_t) \cdot g_t(W_t, \theta) \cdot \nabla_\theta q_t^\alpha(W_t, \theta)' \\ &\quad +[\alpha - H(q_t^\alpha(W_t, \theta) - Y_t)] \cdot dg_t(W_t, \theta)] \\ &= E[g_t(W_t, \theta) \cdot \nabla_\theta q_t^\alpha(W_t, \theta)' \cdot E_t[-\delta(q_t^\alpha(W_t, \theta) - Y_t)]] \\ &\quad +E[dg_t(W_t, \theta) E_t[\alpha - H(q_t^\alpha(W_t, \theta) - Y_t)]] \\ (22) \qquad &= -E[f_{0,t}(q_t^\alpha(W_t, \theta)) \cdot g_t(W_t, \theta) \cdot \nabla_\theta q_t^\alpha(W_t, \theta)'] \\ &\quad +E[(\alpha - F_{0,t}(q_t^\alpha(W_t, \theta))) \cdot dg_t(W_t, \theta)],\end{aligned}$$

where $F_{0,t}$ and $f_{0,t}$ are the true distribution and density of Y_t conditional on the information set $\mathcal{F}_t$. Under correct specification (assumption (A0)), $\alpha = F_{0,t}(q_t^\alpha(W_t, \theta_0))$ and the matrix $\Delta_0 \equiv E[\Delta(Y_t, W_t, \theta_0)]$,

$$\Delta_0 = -E[f_{0,t}(q_t^\alpha(W_t, \theta_0)) \cdot a_t'(q_t^\alpha(w_t, \theta_0)) \cdot \nabla_\theta q_t^\alpha(w_t, \theta_0) \nabla_\theta q_t^\alpha(W_t, \theta_0)'],$$

is nonsingular at θ_0 since $f_{0,t}(q_t^\alpha(W_t, \theta_0)) \neq 0$ (assumption (A6)), $a_t' > 0$ and $E[\nabla_\theta q_t^\alpha(W_t, \theta_0) \nabla_\theta q_t^\alpha(W_t, \theta_0)']$ nonsingular (assumption (A4')). Note that the convergence in probability of the sample mean of Δ holds since: by assumption (A6) $f_{0,t}$ is bounded, i.e. $\sup_{y \in \mathbb{R}} f_{0.t}(y) = C < \infty$, and by

(A2') a_t' is bounded, so that

$$
\begin{aligned}
E[||\Delta(Y_t, W_t, \theta)||^{1+\delta}] &\leqslant n_\delta\{\max(1, C^{1+\delta}) \cdot \max(1, K^{1+\delta}) \cdot E[||\nabla_\theta q_t^\alpha(W_t, \theta)\nabla_\theta q_t^\alpha(W_t, \theta)'||^{1+\delta}] \\
&\quad + E[||dg_t(W_t, \theta)||^{1+\delta}]\},
\end{aligned}
$$

where n_δ is some positive constant. By assumption (A4'), $E[||\nabla_\theta q_t^\alpha(W_t, \theta)||^{2+\epsilon}] < \infty$ for some $\epsilon > 0$, which implies that $E[||\nabla_\theta q_t^\alpha(W_t, \theta)\nabla_\theta q_t^\alpha(W_t, \theta)'||^{1+\delta}] < \infty$. Moreover, by assumption (A2')

$$
\begin{aligned}
E[||dg_t(W_t, \theta)||^{1+\delta}] &\leqslant m_\delta\{\max(1, M^{1+\delta}) \cdot E[||\nabla_\theta q_t^\alpha(W_t, \theta)\nabla_\theta q_t^\alpha(W_t, \theta)'||^{1+\delta}] \\
&\quad + \max(1, K^{1+\delta}) \cdot E[||\nabla_{\theta\theta} q_t^\alpha(W_t, \theta)||^{1+\delta}]\},
\end{aligned}
$$

m_δ being a positive constant, so that by (A4') we obtain $E[||dg_t(W_t, \theta)||^{1+\delta}] < \infty$. Therefore, $E[||\Delta(Y_t, W_t, \theta)||^{1+\delta}] < \infty$. Finally, by using assumption (A8) and a law of large numbers (e.g. Theorem A.2.7 in White 1994) we obtain the convergence of the sample mean of Δ, which shows that condition (v) of the basic theorem holds.

At last, we show that condition (vi) holds, i.e. that $\sqrt{T}\nabla_\theta L_T(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_0)$. For this, we use a central limit theorem (e.g. Theorem A.3.4 in White 1994): by assumptions (A2') and (A4'), we have

$$
E[||s_t(Y_t, W_t, \theta)||^{2+\delta}] \leqslant \max(1, K^{2+\delta}) \cdot E[||\nabla_\theta q_t^\alpha(W_t, \theta)||^{2+\delta}] < \infty.
$$

Moreover, by (A8) $\{s_t^2(Y_t, W_t, \theta), \mathcal{F}_t\}$ is an L_1-mixingale so that $\sqrt{T}\nabla_\theta L_T(\theta) = \sum_{t=1}^T s_t(y_t, w_t, \theta)/\sqrt{T} \xrightarrow{d} \mathcal{N}(0, \Sigma)$ where

$$
\begin{aligned}
\Sigma &= E[(a_t'(q_t^\alpha(W_t, \theta))^2 \cdot (\alpha - H(q_t^\alpha(W_t, \theta) - Y_t)))^2 \cdot \nabla_\theta q_t^\alpha(W_t, \theta)\nabla_\theta q_t^\alpha(W_t, \theta)'] \\
&= E[(a_t'(q_t^\alpha(W_t, \theta)))^2 \cdot \nabla_\theta q_t^\alpha(W_t, \theta)\nabla_\theta q_t^\alpha(W_t, \theta)' \cdot E_t[(\alpha - H(q_t^\alpha(W_t, \theta) - Y_t))^2]].
\end{aligned}
$$

Note that

$$\begin{aligned} E_t[(\alpha - H(q_t^\alpha(W_t,\theta) - Y_t))^2] &= E_t[\alpha^2 - 2\alpha H(q_t^\alpha(W_t,\theta) - Y_t) + H^2(q_t^\alpha(W_t,\theta) - Y_t)] \\ &= E_t[\alpha^2 - (2\alpha - 1)H(q_t^\alpha(W_t,\theta) - Y_t)] \\ &= \alpha^2 - (2\alpha - 1)E_t[H(q_t^\alpha(W_t,\theta) - Y_t)] \\ &= \alpha^2 - (2\alpha - 1)F_{0,t}(q_t^\alpha(W_t,\theta)), \end{aligned}$$

since for all $x \in \mathbb{R}$ we have $H^2(x) = H(x)$. Thus,

$$\Sigma = E[(\alpha^2 - (2\alpha - 1)F_{0,t}(q_t^\alpha(W_t,\theta))) \cdot (a_t'(q_t^\alpha(W_t,\theta)))^2 \cdot \nabla_\theta q_t^\alpha(W_t,\theta)\nabla_\theta q_t^\alpha(W_t,\theta)']. \tag{23}$$

Under correct specification of the conditional α-quantile of Y_t (assumption (A0)), we have $\alpha^2 - (2\alpha - 1)F_{0,t}(q_t^\alpha(W_t,\theta)) = -\alpha(1-\alpha)$ so that

$$\Sigma_0 = -\alpha(1-\alpha)E[(a_t'(q_t^\alpha(W_t,\theta_0)))^2 \cdot \nabla_\theta q_t^\alpha(W_t,\theta_0)\nabla_\theta q_t^\alpha(W_t,\theta_0)'].$$

We can now use the basic asymptotic normality result adapted from Newey and McFadden (1994, Theorems 7.2 and 7.3) to show that $\sqrt{T}(\hat{\theta}_T - \theta_0) \to \mathcal{N}(0, \Delta_0^{-1}\Sigma_0\Delta_0^{-1})$, which completes the proof of Theorem 6. ■

Proof. (Corollary 7) The proof of Corollary 7 is identical to the one for Theorem 6. Note however that Δ^* and Σ^* are now obtained from (22) and (23) without using the equality $E_t[H(q_t^\alpha(W_t,\theta^*) - Y_t)] = \alpha$, which, in Theorem 6, was implied by the correct specification assumption (A0). ■

Proof. (Theorem 8) When $T = 1$ then note that $P_0(y_\theta, w_\theta, \theta) = -\ln \psi_1^\alpha(y_{\theta,1}, q_1^\alpha(w_{\theta,1},\theta)) = -\ln \psi_1^\alpha(y_1, q_1^\alpha(w_1,\theta))$ and $P_1(y_\theta, w_\theta, \theta) = -\ln \phi_1^\alpha(y_{\theta,1}, q_1^\alpha(w_{\theta,1},\theta)) = -\ln \phi_1^\alpha(y_1, q_1^\alpha(w_1,\theta))$ so that the "minimax" problem $\min_{\theta\in\Theta}[\max_{0\leqslant k\leqslant 1}\{P_k(y_\theta, w_\theta, \theta)\}]$ is equivalent to (7).
Let us now consider the case where $T > 1$. We denote by ε_t the "error" term $\varepsilon_t = y_t - q_t^\alpha(w_t,\theta)$. In what follows we assume that $\varepsilon_\theta \equiv (\varepsilon_{\theta,1}, \varepsilon_{\theta,2}, \ldots, \varepsilon_{\theta,T})'$ is a vector of order statistics so that we have $\varepsilon_{\theta,1} \leqslant \varepsilon_{\theta,2} \leqslant \ldots \leqslant \varepsilon_{\theta,T}$. We denote by y_θ and w_θ,

$y_\theta \equiv (y_{\theta,1}, y_{\theta,2}, \ldots, y_{\theta,T})'$ and $w_\theta \equiv (w_{\theta,1}, w_{\theta,2}, \ldots, w_{\theta,T})'$, the vectors of corresponding observations. For given observations $(y_1, w_1, \ldots, y_T, w_T)$ the problem of maximizing the tick-exponential log-likelihood can be restated as

$$(24) \quad -\min_{\theta\in\Theta}[T^{-1}\sum_{t=1}^{T}\max\{-\ln\psi_t^\alpha(y_t, q_t^\alpha(w_t,\theta)), -\ln\phi_t^\alpha(y_t, q_t^\alpha(w_t,\theta))\}],$$

by similar reasoning as for $T = 1$. In order to transform (24) into the standard "minimax" problem we proceed as follows: for θ fixed, let k_θ, $0 \leqslant k_\theta \leqslant T$, denote the order such that $\varepsilon_{\theta,k_\theta} < 0 \leqslant \varepsilon_{\theta,k_\theta+1}$. By convention $\varepsilon_{\theta,0} \equiv -\infty$ and $\varepsilon_{\theta,T+1} \equiv +\infty$. First consider all the indices t such that $1 \leqslant t \leqslant k_\theta$. In that case we have $\varepsilon_{\theta,t} \leqslant \varepsilon_{\theta,k_\theta} < 0$ so that

$$\max\{-\ln\psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)), -\ln\phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta))\} = -\ln\phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)), \text{ for } 1 \leqslant t \leqslant k_\theta.$$

In the same way, for t such that $k_\theta + 1 \leqslant t \leqslant T$, we know that $0 \leqslant \varepsilon_{\theta,k_\theta+1} \leqslant \varepsilon_{\theta,t}$ and so

$$\max\{-\ln\psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)), -\ln\phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta))\} = -\ln\psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)), \text{ for } k_\theta+1 \leqslant t \leqslant T.$$

Thus

$$(25) \quad \begin{aligned} & T^{-1}\sum_{t=1}^{T}\max\{-\ln\psi_t^\alpha(y_t, q_t^\alpha(w_t,\theta)), -\ln\phi_t^\alpha(y_t, q_t^\alpha(w_t,\theta))\} \\ = \; & T^{-1}[\sum_{t=1}^{k_\theta} -\ln\phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k_\theta+1}^{T} -\ln\psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta))], \end{aligned}$$

where by convention $\sum_t^s \equiv 0$ if $s < t$. We now show that the right hand side of (25) is a maximum over k of $P_k(y_\theta, w_\theta, \theta)$, i.e. that for every $0 \leqslant k \leqslant T$ we have

$$(26) \quad \begin{aligned} & T^{-1}[\sum_{t=1}^{k} -\ln\phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k+1}^{T} -\ln\psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta))] \\ \leqslant \; & T^{-1}[\sum_{t=1}^{k_\theta} -\ln\phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k_\theta+1}^{T} -\ln\psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta))]. \end{aligned}$$

First consider $k < k_\theta$. We then have

$$
\begin{aligned}
&\sum_{t=1}^{k} -\ln \phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k+1}^{T} -\ln \psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) \\
&= \sum_{t=1}^{k} -\ln \phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k+1}^{k_\theta} -\ln \psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k_\theta+1}^{T} -\ln \psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) \\
&\leqslant \sum_{t=1}^{k} -\ln \phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k+1}^{k_\theta} \max\{-\ln \psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)), -\ln \phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta))\} \\
&+ \sum_{t=k_\theta+1}^{T} -\ln \psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) \\
&= \sum_{t=1}^{k} -\ln \phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k+1}^{k_\theta} -\ln \phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k_\theta+1}^{T} -\ln \psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) \\
&= \sum_{t=1}^{k_\theta} -\ln \phi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)) + \sum_{t=k_\theta+1}^{T} -\ln \psi_t^\alpha(y_{\theta,t}, q_t^\alpha(w_{\theta,t},\theta)).
\end{aligned}
$$

Similarly we can show that the same result holds for $k > k_\theta$, which proves (26) for every k. Taking into account (25), we have just shown that

$$
T^{-1} \sum_{t=1}^{T} \max\{-\ln \psi_t^\alpha(y_t, q_t^\alpha(w_t,\theta)), -\ln \phi_t^\alpha(y_t, q_t^\alpha(w_t,\theta))\} = \max_{0 \leqslant k \leqslant T}\{P_k(y_\theta, w_\theta, \theta)\},
$$

which completes the proof of Theorem 8. ∎