# Quasi-Maximum Likelihood Estimation for Conditional Quantiles

Ivana Komunjer*,†

Caltech

**Abstract**

In this paper we derive a new class of conditional quantile estimators, which contain Koenker and Bassett's (1978) nonlinear quantile regression estimator as a special case. The latter belong to the family of quasi-maximum likelihood estimators (QMLEs) and are based on a new family of densities which we call 'tick-exponential'. Analogously to the linear-exponential family, the tick-exponential assumption is a necessary condition for a QMLE to be consistent for the parameters of a correctly specified conditional quantile model. We show that the tick-exponential QMLEs are moreover asymptotically normally distributed with an asymptotic covariance matrix that accounts for possible conditional quantile model misspecification and which can be consistently estimated by using the tick-exponential quasi-likelihood scores and its hessian. The practical problem of likelihood maximization is easily solved by using a 'minimax' representation not seen in the earlier work on conditional quantile estimation.

*Author's address: Division of the Humanities and Social Sciences 228-77, California Institute of Technology, Pasadena, CA 91125, U.S.A., phone: (626) 395-3531, fax: (626) 793-4681 *(komunjer@hss.caltech.edu)*

†I would like to thank Moshe Buchinsky, Bruno Crepon, Graham Elliott, Denis Fougere, Philip Gill, Jinyong Hahn, Alain Monfort, Roger Moon, Geert Ridder, Christopher Sims, Gary Solon, Quang Vuong, Hal White and all the participants at UCLA and USC Econometrics seminars.

**Keywords**: tick-exponential family, minimax representation, QMLE, quantiles, asymptotic distribution, misspecification.

**JEL codes**: C22, C51, C61

# 1 Introduction

A vast majority of the empirical work in economics and finance has traditionally focused on models for conditional means. Over the last decade, however, applied literature has devoted increasing attention to other aspects of conditional distributions, such as their quantiles (see, e.g., Koenker and Hallock, 2000). This important empirical work has in turn rekindled the interest of the academic community in the theoretical problem of conditional quantile estimation and inference, a problem which we address in this paper.

Since the seminal work by Koenker and Bassett (1978), numerous authors have used a quantile regression framework for conditional quantile estimation (see, e.g., Koenker and Bassett, 1982, Powell, 1986, Portnoy, 1991, Koenker and Zhao, 1996, Kim and White, 2002) and specification testing (see, e.g., Koenker and Bassett, 1982, Zheng, 1998, Bierens and Ginther, 2000, Horowitz and Spokoiny, 2002, Kim and White, 2002, Koenker and Xiao, 2002). Common finding of the extant literature is that the quantile regression estimator has nice asymptotic properties under various data dependence structures and for a wide variety of conditional quantile models. While many efforts have been made in generalizing the existing results to new models and data structures, surprisingly little attention has been devoted to finding alternative semi-parametric estimators for conditional quantiles. There are yet important theoretical and practical benefits in having different conditional quantile estimators available. For example, it is difficult, if not impossible, to address questions such as finding a minimal covariance estimator or constructing a Hausman-type model specification test, if we only have one conditional quantile estimator available.

In contrast to the prior literature, our approach to conditional quantile estimation is based on a quasi-maximum likelihood. As already demonstrated in the context of conditional mean estimation, the quasi-maximum likelihood framework allows one to simply determine the class of all consistent estimators (see, e.g., Gourieroux, Monfort and Trognon, 1984, White 1994). It is a well known result that there exist a variety of non-Gaussian quasi-maximum likelihood estimators (QMLEs) which, under standard regularity conditions and provided that they belong a linear-exponential family, are consistent for the parameters of a correctly specified conditional mean model.[1] In this paper, we derive an analog result valid in the context of conditional quantile estimation. In other words, we show that there exist an entire 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. In the particular case where the tick-exponential density equals an asymmetric Laplace (or double exponential) density, the tick-exponential QMLE reduces to the standard Koenker and Bassett (1978) quantile regression estimator.

The QMLE generalization is not only of theoretical interest but also has a substantial practical contribution. For example, it provides an alternative approach to the conditional quantile covariance matrix estimation. While in the Gaussian QMLE case, it has become a standard approach to use the scores and the hessian of the Gaussian quasi-likelihood to construct an estimator for the asymptotic covariance matrix, there exist, to the best of our knowledge, no similar interpretation in the quantile regression case. The key advantage of using the tick-exponential scores and hessian to estimate the conditional quantile covariance matrix is that it avoids estimating conditional densities. While in linear homoskedastic models, the latter is easy to carry out by using kernel based methods (see, e.g., Kim and White, 2002, Koenker and Xiao, 2002), its implementation is far more cumbersome in a more general context of nonlinear heteroskedastic models (see, e.g., Zheng, 1998). The (quasi) likelihood based approach also offers an alternative to various bootstrap schemes used to estimate the conditional quantile covariance matrix (see, e.g., Hahn, 1995, Horowitz, 1998), which usually suffer from high computational costs (see, e.g., Fitzenberger, 1997).

[1]Examples of non-Guassian quasi-likelihoods used in the empirical work include: Gamma, Bernoulli, Poisson and Student-$t$ (see, e.g., Bollerslev, 1987).

In practice, computation of tick-exponential QMLEs is a challenge since it involves maximizing objective functions which are neither convex nor differentiable. In order to allow the practitioners to easily implement the proposed estimation method, we suggest a new 'minimax' approach to optimization. Based on a simple transformation of the tick-exponential objective function, we are able to transform the initial maximization problem into a saddle-point search problem, which involves only continuously differentiable functions. Even though the new characterization does not affect the convexity of the objective function, it does recover the differentiability property. The main advantages of our minimax approach to conditional quantile estimation are: (1) it is applicable for both linear and nonlinear conditional quantile models, unlike linear programming techniques; (2) its convergence properties are well established in the literature (see, e.g., Brayton, Director, Hachtel and Vidigal, 1979); and (3) unlike the interior-point methods (see, e.g., Koenker and Park, 1996), it is easy to implement by using the standard gradient-based optimization techniques, such as Sequential Quadratic Programming (SQP) methods, for example.

The remainder of the paper is organized as follows: in Section 2 we define the notation and give the basic properties of conditional quantile models considered in our setup. Section 3 introduces the tick-exponential family of densities and shows that its role in the conditional quantile estimation is analog to the one of the linear-exponential family in the conditional mean estimation. In Section 4 we study the asymptotic properties of the tick-exponential QMLEs. In particular we derive primitive conditions for consistency and show asymptotic normality of the QMLE. The practical implementation of the tick-exponential quasi-likelihood maximization problem is treated in Section 5, in which we introduce the minimax representation. All technicalities regarding the proofs are relegated to the Appendix.

## 2 Notation and Setup

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 probability space $(\Omega, \mathcal{F}, P_0)$, where $\mathcal{F} \equiv \{\mathcal{F}_t, t = 1, \ldots, T\}$ and $\mathcal{F}_t$ denotes the smallest $\sigma$-algebra that $X_t$ is adapted to, i.e. $\mathcal{F}_t \equiv \sigma\{X_1, \ldots, X_t\}$. Assume that the joint distribution of $(X_1, \ldots, X_T)$ has a strictly positive continuous density $p_T$ so that conditional densities are everywhere defined. In particular, we will be interested in the conditional distribution of the first (scalar) component of the random vector $X_t$. We therefore partition $X_t$ as $X_t \equiv (Y_t, Z_t')'$, where $Y_t \in \mathbb{R}$ is the scalar variable of interest and $Z_t \in \mathbb{R}^n$ a vector of exogenous variables. The choice of conditioning variables depends upon the nature of applications in hand: in a time series context, we are primarily interested in using various lags of $Y_t$, while in a cross section analysis the emphasis is put on the exogenous variables $Z_t$. In order to keep our analysis unified, we introduce the following notation: let $\mathcal{G}_t \equiv \sigma\{X_1, \ldots, X_{t-1}, Z_t\}$, so that the information set $\mathcal{G}_t$ contains different functions of various lags of the variable of interest $Y_t$ and possibly contemporaneous values of $Z_t$. We use the standard notation and let $P_0(Y_t \in A|\mathcal{G}_t)$ denote the conditional distribution of $Y_t$, with $A$ an element of the Borel $\sigma$-algebra on $\mathbb{R}$, $A \in \mathcal{B}(\mathbb{R})$. Further, we let $F_{0,t}$ denote the true conditional distribution function of $Y_t$, i.e. $F_{0,t}(y) \equiv P_0(Y_t \leqslant y|\mathcal{G}_t)$ for every $y \in \mathbb{R}$ and every $t$, $1 \leqslant t \leqslant T$, and we call $f_{0,t}$ the corresponding conditional probability density. Of course, $F_{0,t}$ and $f_{0,t}$ are unknown and we assume that for every $t$, $1 \leqslant t \leqslant T$, $F_{0,t}$ belongs to $\tilde{F}$, which is the set of all absolutely continuous distribution functions on $\mathbb{R}$, and that $f_{0,t}$ is strictly positive on the support of $Y_t$. Hereafter, lower case letters (e.g., $x_t$) will be used to denote the observations of the corresponding random variables (e.g., $X_t$).

As motivated in the Introduction, the focus of this paper is on different conditional quantiles of $Y_t$. For a given value of probability $\alpha$, $\alpha \in (0, 1)$, let then $Q_\alpha(Y_t|\mathcal{G}_t)$ denote the $\alpha$-quantile of $Y_t$ conditional on the information set $\mathcal{G}_t$, i.e.

$$Q_\alpha(Y_t|\mathcal{G}_t) \equiv \inf_{v \in \mathbb{R}}\{v : F_{0,t}(v) > \alpha\}.$$

In the case where the conditional distribution of $Y_t$ is continuous, such as the one examined here, the above definition is equivalent to $\alpha = F_{0,t}(Q_\alpha(Y_t|\mathcal{G}_t))$ or $Q_\alpha(Y_t|\mathcal{G}_t) = F_{0,t}^{-1}(\alpha)$.

The approach used to estimate the conditional $\alpha$-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 $\alpha$-quantile of $Y_t$, $\mathcal{M} \equiv \{q_t^\alpha(W_t, \theta)\}$, in which $W_t$ is a $\mathcal{G}_t$-measurable random vector, $W_t : \Omega \to \mathbb{R}^m$, $\theta$ is an unknown parameter in $\Theta$, where $\Theta$ is a subset of $\mathbb{R}^k$ with non-empty interior, $\mathring{\Theta} \neq \emptyset$, and $q_t^\alpha(W_t, \cdot) : \Theta \to \mathbb{R}$ some real function. Note that through $W_t$ we allow both different lags of the variable of interest $Y_t$ and the exogenous variables $Z_t$ to appear in $q_t^\alpha$. In what follows we restrict our attention to conditional quantile models $\mathcal{M}$ in which the following conditions are satisfied:

(A0) *(i)* the model $\mathcal{M}$ is identified on $\Theta$, i.e. for any $(\theta_1, \theta_2) \in \Theta^2$ we have: $q_t^\alpha(W_t, \theta_1) = q_t^\alpha(W_t, \theta_2), a.s. - P_0$, for all $t$, $1 \leqslant t \leqslant T$, if only if $\theta_1 = \theta_2$; *(ii)* for every $t$, $1 \leqslant t \leqslant T$, the function $q_t^\alpha(W_t, \cdot) : \Theta \to \mathbb{R}$ is twice continuously differentiable *a.s.* on $\Theta$; *(iii)* for every $t$, $1 \leqslant t \leqslant T$, the matrix $\nabla_\theta q_t^\alpha(W_t, \theta) \nabla_\theta q_t^\alpha(W_t, \theta)'$ is of full rank $a.s. - P_0$ for all $\theta \in \Theta$; *(iv)* the $m$-vector $W_t$ is function of $Z_t$ and of some finite number $\tau$ of lags of $X_t$, i.e. $W_t \equiv h(Z_t, X_{t-1}, \ldots, X_{t-\tau})$ where $h$ is a function into $\mathbb{R}^m$.

Conditions *(i)-(iv)* in (A0) are fairly standard and verified for a variety of conditional quantile models. Examples of such models include: Koenker and Zhao's (1996) conditional quantile model, $q_t^\alpha(W_t, \theta) \equiv \beta_0 + \sum_{i=1}^p \beta_i Y_{t-i} + \sigma_t \delta$ with $\sigma_t = \gamma_0 + \sum_{j=1}^q \gamma_j |Y_{t-j} - \beta_0 - \sum_{i=1}^p \beta_i Y_{t-j-i}|$, $p, q \geqslant 1$, in which $W_t \equiv (Y_{t-1}, \ldots, Y_{t-p-q})'$ and $\theta \equiv (\beta_0, \ldots \beta_p, \delta, \gamma_0, \ldots, \gamma_q)'$; Engle and Manganelli's (1999) CAViaR model, $q_t^\alpha(W_t, \theta) \equiv \beta_0 + \beta_1 q_{t-1}^\alpha(W_{t-1}, \theta) + l(\beta_2, Y_{t-1}, q_{t-1}^\alpha(W_{t-1}, \theta))$, in which $l$ corresponds to some loss function and we have $W_t \equiv (Y_{t-1}, \ldots, Y_1)'$ and $\theta \equiv (\beta_0, \beta_1, \beta_2)'$; 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 $\alpha$-quantile of $Y_t$ if and only if there exists some true parameter value $\theta_0$ such that $F_{0,t}(q_t^\alpha(W_t, \theta_0)) = \alpha$, for every $t$, $1 \leqslant t \leqslant T$. More formally, under correct model specification we assume the following:

(A1) given $\alpha \in (0,1)$, there exists $\theta_0 \in \mathring{\Theta}$ such that $E[1(q_t^\alpha(W_t, \theta_0) - Y_t)|\mathcal{G}_t] = \alpha, a.s. - P_0$, for every $t$, $1 \leqslant t \leqslant T$.[2]

In other words, for any $t$, $1 \leqslant t \leqslant T$, the difference between the indicator variable above and $\alpha$ is assumed to be orthogonal to any $\mathcal{G}_t$-measurable random variable. Unfortunately, and in most practical applications, the correct specification assumption (A1) is unlikely to hold, implying that the conditional quantile model $\mathcal{M}$ is misspecified. Given the importance of the misspecified case in practice, we devote particular attention, in our theoretical results, to cases where the assumption (A1) fails.

# 3 Tick-Exponential Family of Densities

We consider a class of quasi-maximum likelihood estimators (QMLEs), $\hat{\theta}_T$, obtained by solving

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

where $l_t$ is a period-$t$ conditional quasi-likelihood. It is a well known result that different choices of $l_t$ affect the asymptotic properties of the QMLE when the object of interest is the conditional mean of $Y_t$. Specifically, let $\{\mu_t\}$ denote a model for the conditional mean of $Y_t$, which is correctly specified, and consider $\hat{\theta}_T$ which solves $\max_{\theta \in \Theta} T^{-1} \sum_{t=1}^{T} \ln l_t(y_t, \mu_t(w_t, \theta))$. Under standard regularity conditions, $\hat{\theta}_T$ is consistent for the true parameter $\theta_0$ of $\{\mu_t\}$ *only if* the quasi-likelihood $l_t$ belongs to the linear-exponential family of densities, i.e. only if we

[2]The function $1 : \mathbb{R} \to [0,1]$ is the Heaviside function, i.e. for any $x \in \mathbb{R}$, we have $1(x) = 0$ if $x < 0$ and $1(x) = 1$ if $x \geqslant 0$, so that $1(\cdot)$ is right-continuous, i.e. $\lim_{h \to 0^+} 1(x+h) = 1(x)$ for all $x \in \mathbb{R}$. The Heaviside function is the indefinite integral of the Dirac delta function $\delta$, with $1(x) = \int_a^x d\delta$, where $a$ is an arbitrary (possibly infinite) negative constant, $a \leqslant 0$.

have $l_t(y,\eta) = \varphi_t(y,\eta)$ with

$$\varphi_t(y,\eta) = \exp[a_t(\eta) + b_t(y) + yc_t(\eta)], \tag{2}$$

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 $\varphi_t$ is a probability density with mean $\eta$. In other words, the linear-exponential QMLE is consistent for the true value 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 $\varphi_t(\cdot, \mu_t(w_t, \theta))$. 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, which is valid in the case when the object of interest is no longer the conditional mean of $Y_t$ but rather its conditional $\alpha$-quantile. We start by defining a family of densities whose role in the conditional quantile estimation is analog to the one of the linear-exponential family (2) in the conditional mean estimation. We call such family 'tick-exponential'.

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

$$\varphi_t^\alpha(y,\eta) = \exp(-(1-\alpha)[a_t(\eta) - b_t(y)]1\{y \leqslant \eta\} + \alpha[a_t(\eta) - c_t(y)]1\{y > \eta\}), \tag{3}$$

*where $a_t : M_t \to \mathbb{R}$ is continuously differentiable, $a_t \in \mathcal{C}^1(M_t, \mathbb{R})$, and $b_t : \mathbb{R} \to \mathbb{R}$ and $c_t : \mathbb{R} \to \mathbb{R}$ are $\mathcal{F}_t$-measurable;*[3] *the functions $a_t$, $b_t$ and $c_t$ are such that for $\eta \in M_t$: (ii) $\varphi_t^\alpha$ is a probability density, i.e. $\int_{\mathbb{R}} \varphi_t^\alpha(y,\eta)dy = 1$; (iii) $\eta$ is the $\alpha$-quantile of $\varphi_t^\alpha$, i.e. $\int_{-\infty}^{\eta} \varphi_t^\alpha(y,\eta)dy = \alpha$.*

[3] The function $1 : \mathcal{F}_t \to \{0,1\}$ is a standard indicator function, i.e. for any event $A \in \mathcal{F}_t$ we have $1\{A\} = 1$ if $A$ is true and $= 0$ otherwise. Note that in the footnote 1 we use a slightly different notation for the Heaviside function $x \mapsto 1(x)$ which is a real function, i.e. $1 : \mathbb{R} \to [0,1]$.

In other words, for a given value of probability $\alpha$, the density function $\varphi_t^\alpha$ in (3) is exponential by parts where the two parts have different slopes, proportional to $1-\alpha$ and $\alpha$, respectively. Note that by letting $d_t(y) \equiv (1-\alpha)b_t(y) + \alpha c_t(y)$ and $g_t(y) \equiv \alpha[(1-\alpha)b_t(y) - \alpha c_t(y)]$, we obtain an alternative expression for $\varphi_t^\alpha$, 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 been studied separately by Gourieroux, Monfort and Renault (1987) in a context of M-estimation.[4] In the 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 the 'tick' function $t_\alpha(y,\eta) \equiv [\alpha - 1\{y \leqslant \eta\}](y-\eta)$, also known in the literature as the asymmetrical slope or check function. This is why we call 'tick-exponential' the family of functions defined in Definition 1.

Following are some interesting properties of the tick-exponential density of order $\alpha$, $\varphi_t^\alpha$, as defined in (3). For every $\eta \in M_t$, the functions $a_t$, $b_t$ and $c_t$ in Definition 1 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)* $(1-\alpha)b_t(\eta) + \alpha c_t(\eta) = a_t(\eta)$.[5]

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

As pointed out previously, the role played by the tick-exponential family of densities in the context of conditional quantile estimation is perfectly analog to the one played by the

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

[5] The property $(ii)$ is obtained by differentiating the third condition of Definition 1 with respect to $\eta$. Similarly, by combining conditions two and three in Definition 1 and differentiating the resulting equation with respect to $\eta$, we show that property $(iii)$ holds as well. Finally, combining $(ii)$ and $(iii)$ implies that $(i)$ and $(iv)$ hold.

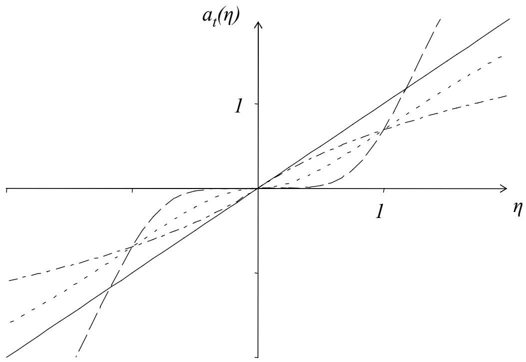
linear-exponential family in the context of conditional mean estimation. In other words, if we can consistently estimate the true parameter $\theta_0$ of a correctly specified conditional quantile model $\mathcal{M}$ by maximizing a quasi-likelihood, then this quasi-likelihood necessarily belongs to the tick-exponential family. Put differently, the tick-exponential assumption is a necessary condition for consistency of any QMLE $\hat{\theta}_T$ and we can obtain an entire class of consistent QMLEs by considering different functions $a_t$, $b_t$ and $c_t$ of the tick-exponential family in Definition 1. We give a more formal statement of this result in the following theorem.

**Theorem 2 (necessary condition for consistency)** *Consider a correctly specified conditional quantile model $\mathcal{M}$ with true parameter $\theta_0$ which satisfies the conditional moment restriction in (A1). Let $\hat{\theta}_T$ be the QMLE obtained by solving the maximization problem* (1)*. Assume: (i) $\Theta$ is compact, and $\{X_t\}$ and $l_t$ are such that (ii) for every $t$, $1 \leqslant t \leqslant T$, $E[\ln l_t(Y_t, q_t^\alpha(W_t, \theta))] < \infty$ and is continuous for any $\theta \in \Theta$, (iii) $E[L_T(\theta)]$ is uniquely maximized at $\theta^* \in \mathring{\Theta}$, and (iv) $L_T(\theta)$ converges uniformly in probability to $E[L_T(\theta)]$.*
*Then, the QMLE $\hat{\theta}_T$ is consistent for $\theta_0$ only if, for every $t$, $1 \leqslant t \leqslant T$, $y \in \mathbb{R}$ and $\eta \in M_t$, we have*

$$l_t(y, \eta) = \varphi_t^\alpha(y, \eta),$$

*where $\varphi_t^\alpha$ is a tick-exponential density of order $\alpha$ as in Definition 1.*

What Theorem 2 shows is, if we want the QMLE to be consistent for the true parameter of a correctly specified model of the conditional $\alpha$-quantile of $Y_t$, then we must choose a member $l_t$ of the tick-exponential family. There are at least two important implications of this result: first, Theorem 2 shows that a consistent QMLE for the conditional $\alpha$-quantile is not necessarily the one obtained by a standard non-linear quantile regression: $\min_{\theta \in \Theta} T^{-1} \sum_{t=1}^{T} [\alpha - 1(q_t^\alpha(w_t, \theta) - y_t)][y_t - q_t^\alpha(w_t, \theta)]$. Indeed, Koenker and Bassett's (1978) quantile regression estimator is a special case of our family of tick-exponential QMLEs which is obtained by letting $a_t(\eta) = [1/(\alpha(1-\alpha))]\eta$ and $b_t(y) = c_t(y) = [1/(\alpha(1-\alpha))]y$, so that the function $\ln \varphi_t^\alpha$ in (3) is proportional to the aforementioned tick function $t_\alpha$. In other words,



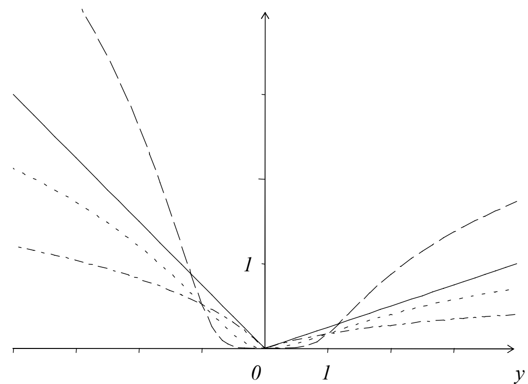




Figure 1: Functions $\eta \mapsto a_t(\eta)$ and the corresponding criteria plotted at $\eta = 0$: $\frac{1}{\alpha(1-\alpha)}\eta$ and $[\alpha - 1\{y \leqslant \eta\}](y-\eta)$ (solid line), $\frac{1}{\alpha(1-\alpha)} \operatorname{sgn}(\eta) \ln(1+|\eta|^p)$ and $[\alpha - 1\{y \leqslant \eta\}][\operatorname{sgn}(y) \ln(1+|y|^p) - \operatorname{sgn}(\eta) \ln(1+|\eta|^p)]$, with $p = 1$ (dot-dashed line), $p = 2$ (dotted line) and $p = 5$ (dashed line).

there exist alternative estimators that one can use to consistently estimate conditional quantile models and the second important implication of Theorem 2 is to show that those are necessarily of the tick-exponential form, with functions $a_t$, $b_t$ and $c_t$ as in Definition 1.

Figure 1 provides examples of functions $a_t$ that satisfy the requirements of Definition 1. For each choice of $a_t$ we derive a corresponding criterion which is minimized by the QMLE. For example, in the standard Koenker and Bassett (1978) case, we have $a_t(\eta) = [1/(\alpha(1-\alpha))]\eta$ and the criterion is the tick function $t_\alpha$. Similarly, if we define $\operatorname{sgn}(x) \equiv 1\{x \geqslant 0\} - 1\{x \leqslant 0\}$ and let

$$a_t(\eta) = \tfrac{1}{\alpha(1-\alpha)} \operatorname{sgn}(\eta) \ln(1+|\eta|^p), \; p \in \mathbb{N}^*, \tag{4}$$

then we obtain a whole new class of conditional quantile QMLEs, $\hat{\theta}_T$, which solve

$$\begin{aligned} \max_{\theta\in\Theta} T^{-1} \textstyle\sum_{t=1}^T \tfrac{1}{\alpha} &[\operatorname{sgn}(y_t) \ln(1+|y_t|^p) - \operatorname{sgn}(q_t^\alpha(w_t,\theta)) \ln(1+|q_t^\alpha(w_t,\theta)|^p)] 1\{y_t \leqslant q_t^\alpha(w_t,\theta)\} \\ &- \tfrac{1}{1-\alpha}[\operatorname{sgn}(y_t) \ln(1+|y_t|^p) - \operatorname{sgn}(q_t^\alpha(w_t,\theta)) \ln(1+|q_t^\alpha(w_t,\theta)|^p)] 1\{y > q_t^\alpha(w_t,\theta)\}. \end{aligned} \tag{5}$$

It is straightforward to show that the function $a_t$ in (4) satisfies the requirements in Definition 1, which under conditions of Theorem 2, ensures that $\hat{\theta}_T$ is consistent for the true conditional quantile parameter $\theta_0$. Conditional quantile estimators such as the QMLEs obtained by maximizing (5) have not yet been seen in the literature. Also, note that the above maximization problem is equivalent to a novel non-linear quantile regression problem: $\min_{\theta\in\Theta} T^{-1}\sum_{t=1}^{T}[\alpha - 1(q_t^\alpha(w_t,\theta)-y_t)][\mathrm{sgn}(y_t)\ln(1+|y_t|^p) - \mathrm{sgn}(q_t^\alpha(w_t,\theta))\ln(1+|q_t^\alpha(w_t,\theta)|^p)]$, with $p \in \mathbb{N}^*$.

The key intuition behind the proof of Theorem 2 is as follows: first, we assume that the process $\{X_t\}$ and the quasi-likelihood $l_t$ are such that the QMLE converges in probability to the pseudo-true value $\theta^*$ of the parameter $\theta$ of interest. The pseudo-true parameter $\theta^*$ is defined as a unique maximizer of $E[L_T(\theta)]$ on $\mathring{\Theta}$. The conditions *(i) - (iv)* are sufficient for $\hat{\theta}_T \xrightarrow{p} \theta^*$ to hold.[6] In the second step and once we know that the QMLE is consistent for $\theta^*$, we show that the equality $\theta^* = \theta_0$ can only hold if $l_t(y,\eta) = \varphi_t^\alpha(y,\eta)$ with $\varphi_t^\alpha$ as defined in (3). In other words, the tick-exponential assumption is a necessary condition for the QMLE $\hat{\theta}_T$ to be consistent for the true value $\theta_0$ of a correctly specified conditional quantile model $\mathcal{M}$.

We now turn to the study of sufficient conditions for consistency of tick-exponential QMLEs. The focus of the following section is twofold: first, we propose a set of conditions for consistency of $\hat{\theta}_T$ which are more primitive than the ones used in Theorem 2, and ,second, we derive the asymptotic distribution of the QMLE.

[6]Note that these are not exactly the primitive conditions for consistency of $\hat{\theta}_T$. The integrability of $\ln l_t$ with respect to $P_0$ required by *(ii)* involves more primitive conditions on the existence of different moments of $Y_t$ and $W_t$. The condition *(iii)* states that $\theta^*$ is a minimum of $E[L_T(\theta)]$ and that this minimum is moreover unique. The first requirement involves more primitive conditions on $\partial \ln l_t/\partial\eta$, $\partial^2 \ln l_t/\partial\eta^2$ and $\nabla_\theta q_t^\alpha$, which depend on the functional form of both $l_t$ and $q_t^\alpha$. Finally, the uniform convergence condition *(iv)* can be obtained by applying an appropriate uniform law of large numbers, such as the one corresponding to the simplest case where $\{X_t\}$ is independent and identically distributed and the function $\ln l_t$ is Lipshitz.

# 4 Asymptotic Properties of Tick-Exponential QMLEs

Let $\varphi_t^\alpha$ be a member of the tick-exponential family of order $\alpha$, as in Definition 1, and let $\hat{\theta}_T$ be the corresponding QMLE, solution to

$$\max_{\theta\in\Theta} L_T(\theta) \equiv T^{-1}\sum_{t=1}^{T} \ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t,\theta)). \tag{6}$$

The following theorem provides sufficient conditions for the consistency of $\hat{\theta}_T$.

**Theorem 3 (consistency)** *Let $\hat{\theta}_T$ be a tick-exponential QMLE obtained by solving* (6) *and let $\theta^*$ be the pseudo-true value of the parameter $\theta$, $\theta^* \equiv \arg\max_{\theta\in\Theta} E[L_T(\theta)]$. Assume:*
*(A2) $\Theta$ is compact and $\theta_0$ and $\theta^*$ are interior points of $\Theta$;*
*(A3) there exists $K > 0$ such that for every $t$, $1 \leqslant t \leqslant T$, and $\eta \in M_t$, $0 < a_t'(\eta) \leqslant K$;*
*(A4) the sequence $\{X_t\}$ is strong mixing with $\alpha$ of size $-r/(r-2)$, with $r > 2$;*
*(A5) $E[\sup_{\theta\in\Theta} |\nabla_\theta q_t^\alpha(W_t,\theta)|] < \infty$ and for some $\epsilon > 0$, $E[\sup_{\theta\in\Theta} |a_t(q_t^\alpha(W_t,\theta))|^{r+\epsilon}] < \infty$, $E[|b_t(Y_t)|^{r+\epsilon}] < \infty$, $E[|c_t(Y_t)|^{r+\epsilon}] < \infty$, for all $t$, $1 \leqslant t \leqslant T$;*
*Under assumptions (A0), (A2)-(A5), we have $\hat{\theta}_T \xrightarrow{p} \theta^*$. If, in addition, the conditional quantile model $\mathcal{M}$ is correctly specified (A1), then $\hat{\theta}_T \xrightarrow{p} \theta_0$.*

In other words, if $l_t$ belongs to the tick-exponential family of densities, then the QMLE is consistent for the true value $\theta_0$ of a correctly specified model $\mathcal{M}$ despite distributional misspecification, i.e. even if the true conditional density of $Y_t$, $f_{0,t}$, 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 $\alpha$-quantile of $Y_t$. Even though the consistency result in Theorem 3 is robust to distributional misspecification, the convergence of $\hat{\theta}_T$ to the true $\theta_0$ is only valid if the conditional quantile model $\mathcal{M}$ is correctly specified (A1), which may not hold. Under model misspecification, the QMLE converges to the pseudo-true value $\theta^*$ which maximizes the expected quasi log-likelihood. The pseudo-true value of the parameter $\theta$ is of particular interest in applications such as forecasting, for example. If forecasting is the goal then one is interested in finding parameter values that minimize the expected loss of the forecast error, irrespective of whether or not the forecasting model is correctly specified. In our case, the pseudo-true value $\theta^*$ minimizes $T^{-1}\sum_{t=1}^{T} E[-\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))]$, so the quantity $q_t^\alpha(W_t, \theta^*)$ corresponds to the best forecast of $Y_t$ under the loss $-\ln \varphi_t^\alpha$. What Theorem 3 then shows is that the QMLE $\hat{\theta}_T$ converges in probability to the optimal forecast parameter $\theta^*$ as the sample size increases.

Assumptions used to derive the results of Theorem 3 are fairly standard and can be classified in three groups: compactness (A2), uniqueness and uniform convergence assumptions. The purpose of uniqueness assumptions is to ensure that $\theta^*$ (or $\theta_0$) is a unique maximizer of the expected log-likelihood $E[L_T(\theta)]$. While this requirement is easily verified for $\theta^*$, it needs to be checked for $\theta_0$, under correct specification of the conditional $\alpha$-quantile of $Y_t$ (A1). Uniform convergence assumptions are used to ensure that the function $\ln \varphi_t^\alpha$ is uniformly continuous in $\theta$ and 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 are Lipshitz, implied by (A3) and (A5). An alternative way to obtain uniform convergence out of pointwise convergence would be to use the convexity of $\ln \varphi_t^\alpha$ (see, e.g., Pollard, 1991, Hjort and Pollard, 1993, Knight, 1998). Despite its convenience (and elegance), the convexity approach only works if for every $t$, the quasi log-likelihood $\ln \varphi_t^\alpha$ is convex in $\theta$, which is not the case for a general choice of $a_t$ in Definition 1. In the case of a standard non-linear quantile regression à la Koenker and Bassett (1978), however, the function $a_t$ is linear in $\eta$ and $\ln \varphi_t^\alpha$ is convex in $\theta$, so that the convexity argument can be applied. The second requirement for the ULLN to be applicable is that the process $\{\ln \varphi_t^\alpha\}$ obeys certain heterogeneity restrictions. More specifically, we assume that the sequence $\{X_t\}$ is strong or $\alpha$-mixing (A4), which, by imposing an additional constraint on $W_t$ (A0), ensures that $\{\ln \varphi_t^\alpha\}$ is $\alpha$-mixing of the same size. It is important to note that we do not require $\{X_t\}$ to be a stationary sequence.

Finally, note that Theorem 2 is not exactly the converse of the result given in Theorem 3. When deriving the necessary condition for consistency, we have assumed that $q_t^\alpha$ and $\ln l_t$

were continuously differentiable with probability 1, i.e. that for all $t$, $1 \leqslant t \leqslant T$, we had $q_t^\alpha(W_t, \cdot)$ and $\ln l_t(Y_t, q_t^\alpha(W_t, \cdot))$ continuously differentiable on $\Theta$, $a.s. - P_0$. 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.

Let us now turn to the asymptotic normality of the tick-exponential QMLE $\hat{\theta}_T$, solution to the maximization problem in (6). The classical asymptotic normality results for QMLEs require that the log-likelihood function $L_T$ be twice continuously differentiable. The main idea is to then use the first-order Taylor expansion of the gradient $\nabla_\theta L_T$ around the QMLE $\hat{\theta}_T$, which satisfies the first order condition $\nabla_\theta L_T(\hat{\theta}_T) = 0$. This approach requires $L_T$ to be sufficiently smooth, which is not the case with the tick-exponential family of densities due to the presence of indicator functions in (3). Indeed, under tick-exponential assumption,

$$\begin{aligned} L_T(\theta) = T^{-1} \sum\nolimits_{t=1}^{T} & -(1-\alpha)[a_t(q_t^\alpha(w_t, \theta)) - b_t(y_t)]1\{y_t \leqslant q_t^\alpha(w_t, \theta)\} \\ & +\alpha[a_t(q_t^\alpha(w_t, \theta)) - c_t(y_t)]1\{y_t > q_t^\alpha(w_t, \theta)\}, \end{aligned} \tag{7}$$

where the functions $a_t$, $b_t$ and $c_t$ are as in Definition 1. The non-differentiability problem has prompted 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), Chen, Linton and van Keilegom (2003). In this paper, we focus on conditional quantile models that are continuously differentiable on $\Theta$ (A0), so that the log-likelihood function $L_T(\theta)$ is continuously differentiable $a.s.$ on $\Theta$. In this case, for every $\theta \in \Theta$, $\nabla_\theta L_T(\theta)$ exists and is continuous with probability 1, and we have $\nabla_\theta L_T(\theta) \equiv T^{-1} \sum_{t=1}^{T} s_t(y_t, w_t, \theta)$, where

$$s_t(Y_t, W_t, \theta) \equiv a_t'(q_t^\alpha(W_t, \theta)) \nabla_\theta q_t^\alpha(W_t, \theta)[\alpha - 1(q_t^\alpha(W_t, \theta) - Y_t)], \tag{8}$$

for every $t$, $1 \leqslant t \leqslant T$. In the following theorem, we derive the asymptotic distribution of $\hat{\theta}_T$.

**Theorem 4 (asymptotic normality)** *Let the conditions (A0), (A2)-(A5) from Theorem 3 hold. In addition, assume:*

*(A1') $E[s_t(Y_t, W_t, \theta^*)] = 0$, for all $t$, $1 \leqslant t \leqslant T$, and $\{s_t(Y_t, W_t, \theta^*)\}$ is uncorrelated;*

*(A3') $a_t \in \mathcal{C}^2(M_t, \mathbb{R})$ and for every $t$, $1 \leqslant t \leqslant T$, and $\eta \in M_t$, $|a_t''(\eta)| \leqslant L$, with $L > 0$;*

*(A5') for some $\epsilon > 0$, $E[\sup_{\theta\in\Theta} |\nabla_\theta q_t^\alpha(W_t, \theta)|^{2(r+\epsilon)}] < \infty$ and $E[\sup_{\theta\in\Theta} |\Delta_{\theta\theta} q_t^\alpha(W_t, \theta)|^{r+\epsilon}] < \infty$, for all $t$, $1 \leqslant t \leqslant T$;*

*(A6) there exists $C > 0$ such that $\sup_{y\in\mathbb{R}} f_{0,t}(y) = C < \infty$, for every $t$, $1 \leqslant t \leqslant T$.*

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

$$\begin{aligned}\Delta(\theta^*) \quad = \quad & -T^{-1}\sum_{t=1}^{T} E\{[f_{0,t}(q_t^\alpha(W_t,\theta^*))a_t'(q_t^\alpha(W_t,\theta^*))\nabla_\theta q_t^\alpha(W_t,\theta^*)\nabla_\theta q_t^\alpha(W_t,\theta^*)'] \\ & +[F_{0,t}(q_t^\alpha(W_t,\theta^*)) - \alpha][a_t''(q_t^\alpha(W_t,\theta^*))\nabla_\theta q_t^\alpha(W_t,\theta^*)\nabla_\theta q_t^\alpha(W_t,\theta^*)' \\ & +a_t'(q_t^\alpha(W_t,\theta^*))\Delta_{\theta\theta} q_t^\alpha(W_t,\theta^*)]\}.\end{aligned} \tag{9}$$

*and*

$$\begin{aligned}\Sigma(\theta^*) \quad = \quad & T^{-1}\sum_{t=1}^{T} E\{[\alpha^2 - (2\alpha-1)F_{0,t}(q_t^\alpha(W_t,\theta^*))] \cdot \\ & [a_t'(q_t^\alpha(W_t,\theta^*))]^2\nabla_\theta q_t^\alpha(W_t,\theta^*)\nabla_\theta q_t^\alpha(W_t,\theta^*)'\}.\end{aligned} \tag{10}$$

Note that the assumptions imposed in Theorem 4 are stronger than the ones used for the consistency of $\hat{\theta}_T$ in Theorem 3. We now require (A1') that the sequence $\{s_t(Y_t, W_t, \theta^*)\}$ be uncorrelated with $E[s_t(Y_t, W_t, \theta^*)] = 0$, for all $t$, $1 \leqslant t \leqslant T$, and (A5') that further moment conditions hold, so that an appropriate Central Limit Theorem (CLT) applies. Note that in the standard version of the Lindeberg-Feller CLT (for independent random variables) the condition (A1') is not necessary. The reason behind is that in the independent case, the sample covariance $T^{-1}\sum_{t=1}^{T} s_t(y_t, w_t, \theta^*)s_t(y_t, w_t, \theta^*)'$ converges in probability to $\Sigma(\theta^*)$ above. However, in a more general case in which we allow for mixing (A4), this convergence is no longer implied by the Lindeberg condition, which is why we need to impose (A1').

The function $a_t$ is now required to be twice continuously differentiable with bounded second derivative (A3'), which together with (A6) ensures that the gradient of the log-likelihood function, $\nabla_\theta L_T$, is stochastically equicontinuous. In general, primitive conditions for stochastic equicontinuity can be viewed as a trade-off between restrictions on the functional form of $\ln \varphi_t^\alpha$ and assumptions on the dependence of $\{X_t\}$. For example, the conditions described by Pollard (1985) and Andrews (1994) which repose on elegant entropy results, allow for a wide variety of non-differentiable objective functions $\ln \varphi_t^\alpha$, but hold for independent or $m$-independent sequences only, which excludes mixing.[7] An alternative approach, proposed by Newey and McFadden (1994), requires a stronger Lipshitz condition on $\ln \varphi_t^\alpha$, but on the other hand, places no restrictions on the dependence of $\{X_t\}$. Given its wide applicability in the time-series context, our proof of Theorem 4 uses the latter result.

The result of Theorem 4 can easily be adapted to the case where the conditional quantile model $\mathcal{M}$ is correctly specified (A1). In the correctly specified case, $\{s_t(Y_t, W_t, \theta_0), \mathcal{G}_t\}$ is a martingale difference sequence, i.e. $E[s_t(Y_t, W_t, \theta_0)|\mathcal{G}_t] = 0$, $1 \leqslant t \leqslant T$. This is a stronger property than the one in (A1'), which we no longer need to impose. Moreover, under (A1) we know that for every $t$, $1 \leqslant t \leqslant T$, $F_{0,t}(q_t^\alpha(W_t, \theta_0)) = \alpha$, $a.s. - P_0$, which simplifies the expressions of $\Delta(\theta_0)$ and $\Sigma(\theta_0)$, as shown in the following corollary to Theorem 4.

**Corollary 5** *Let all the conditions from Theorem 4, except (A1'), hold. In addition, assume that the model $M$ is correctly specified (A1). Then* $\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow \mathcal{N}(0, \Delta(\theta_0)^{-1}\Sigma(\theta_0)\Delta^{-1}(\theta_0))$, *with*

$$\Delta(\theta_0) = -T^{-1} \sum_{t=1}^{T} E[f_{0,t}(q_t^\alpha(W_t, \theta_0)) a_t'(q_t^\alpha(W_t, \theta_0)) \nabla_\theta q_t^\alpha(W_t, \theta_0) \nabla_\theta q_t^\alpha(W_t, \theta_0)'], \tag{11}$$

*and*

$$\Sigma(\theta_0) = T^{-1} \sum_{t=1}^{T} \alpha(1-\alpha) E[(a_t'(q_t^\alpha(W_t, \theta_0))^2 \nabla_\theta q_t^\alpha(W_t, \theta_0) \nabla_\theta q_t^\alpha(W_t, \theta_0)']. \tag{12}$$

Results of Theorem 4 and its Corollary 5 have important theoretical and practical implications. From a theoretical viewpoint, by varying the functions $a_t$, $b_t$ and $c_t$ in Definition 1,

[7] In principle, and as noted by Pollard (1985) in his proof of Lemma 4, these results could be extended to mixing sequences whose correlations decay rapidly enough.

we obtain an entirely new class of conditional quantile estimators which are asymptotically normally distributed and whose asymptotic covariance matrix depends on derivatives of $a_t$. Hence, by choosing $a_t$ appropriately and according to the requirements of Theorem 4, one can derive QMLEs with different asymptotic covariances. From the practical viewpoint, results of Theorem 4 and its Corollary 5 suggest an interesting approach to estimation of the asymptotic covariance matrix of $\hat{\theta}_T$. We now discuss both these issues in more details.

As an illustrative example, consider estimating the parameter $\theta$ of a linear model for conditional quantiles, $q_t^\alpha(W_t, \theta) = \theta' W_t$, with independent and identically distributed observations $\{X_t\}$. The standard approach (see, e.g., Koenker and Bassett, 1978) is to let $a_t(\eta) = [1/(\alpha(1-\alpha))]\eta$ and $b_t(y) = c_t(y) = [1/(\alpha(1-\alpha))]y$, in which case the asymptotic covariance matrix $\Delta(\theta^*)^{-1}\Sigma(\theta^*)\Delta(\theta^*)^{-1}$ in Theorem 4 is derived from $\Delta(\theta^*) = -E[f_{0,t}(\theta^{*\prime}W_t)\frac{1}{\alpha(1-\alpha)}W_t W_t']$ and $\Sigma(\theta^*) = E[(\alpha^2 - (2\alpha-1)F_{0,t}(\theta^{*\prime}W_t))\frac{1}{\alpha^2(1-\alpha)^2}W_t W_t']$ (see, e.g., Kim and White, 2002). Note that in the case where we further restrict the linear conditional quantile model to be correctly specified, $\theta^* = \theta_0$ and the asymptotic covariance matrix in Theorem 4 reduces to $\alpha(1-\alpha)E[f_{0,t}(\theta_0' W_t)W_t W_t']^{-1}E[W_t W_t']\ E[f_{0,t}(\theta_0' W_t)W_t W_t']^{-1}$ (see, e.g., Powell, 1986). Moreover,if we assume that the true conditional density of $Y_t$ is $\mathcal{G}_t$-independent, i.e. that for all $t$, $1 \leqslant t \leqslant T$, $f_{0,t} = f_0$, the previous expression further reduces to $\alpha(1-\alpha)[f_0^2(\theta_0' W_t)E(W_t W_t')]^{-1}$, which is the original result by Koenker and Bassett (1978).

As pointed out in Section 3, an entirely new set of QMLEs is obtained by using the function $a_t$ in (4), for which it is straightforward to show that it satisfies all the requirements in Definition 1 and Theorem 4. The first two derivatives of $a_t$ in (4) are given by $a_t'(\eta) = [1/(\alpha(1-\alpha))][p|\eta|^{p-1}/(1+|\eta|^p)]$ and $a_t''(\eta) = [1/(\alpha(1-\alpha))]\,\mathrm{sgn}(\eta)p|\eta|^{p-2}[p-1-|\eta|^p]/[1+|\eta|^p]^2$, hence $0 < a_t'(\eta) \leqslant p/[\alpha(1-\alpha)]$ and $|a_t''(\eta)| \leqslant 2p/[\alpha(1-\alpha)]$, for all $\eta \in M_t$ and $p \in \mathbb{N}^*$. The tick-exponential QMLE which corresponds to this choice of $a_t$ solves the maximization problem (5) and the components of its asymptotic covariance matrix $\Delta(\theta^*)^{-1}\Sigma(\theta^*)\Delta(\theta^*)^{-1}$ in Theorem 4 are given by

$$\Delta(\theta^*) = -E\{\tfrac{p}{\alpha(1-\alpha)}\tfrac{|\theta^{*\prime}W_t|^{p-2}}{1+|\theta^{*\prime}W_t|^p}[f_{0,t}(\theta^{*\prime}W_t)|\theta^{*\prime}W_t| + (F_{0,t}(\theta^{*\prime}W_t)-\alpha)\tfrac{p-1-|\theta^{*\prime}W_t|^p}{1+|\theta^{*\prime}W_t|^p}]W_t W_t'\}, \quad (13)$$

and

$$\Sigma(\theta^*) = E[(\alpha^2 - (2\alpha - 1)F_{0,t}(\theta^{*\prime}W_t))\frac{1}{\alpha^2(1-\alpha)^2}\frac{p^2|\theta^{*\prime}W_t|^{2p-2}}{(1+|\theta^{*\prime}W_t|^p)^2}W_tW_t'], \tag{14}$$

for any $p \in \mathbb{N}^*$. The matrices $\Delta(\theta^*)$ and $\Sigma(\theta^*)$ in (13) and (14) above have an entirely novel form, not seen in the previous work. Their expressions can further be used to study the behavior of the asymptotic covariance matrix of $\hat{\theta}_T$ as $p$ changes, for example. Depending on the true conditional distribution of the data in hand, different values of $p$ in the above expressions will lead to estimators with different asymptotic covariances, some of which may be lower than in the standard Koenker and Bassett (1978) case.

We now turn to the problem of asymptotic covariance matrix estimation, based on the formulas for $\Delta(\theta^*)$ and $\Sigma(\theta^*)$, derived in Theorem 4. The main difficulty is that the latter 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 $\Delta(\theta^*)$ and $\Sigma(\theta^*)$ by numerical differentiation. Recall that $\Delta(\theta^*)$ corresponds to expected value of the Hessian matrix of $\ln\varphi_t^\alpha$, while $\Sigma(\theta^*)$ 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 $\{s_t(y_t, w_t, \hat{\theta}_T)\}_{1\leqslant t\leqslant T}$, $\hat{\Sigma}_T(\hat{\theta}_T) \equiv T^{-1}\sum_{t=1}^T s_t(y_t, w_t, \hat{\theta}_T)s_t(y_t, w_t, \hat{\theta}_T)'$. The $j$th row of $s_t$, $s_{t,j}$, is obtained by numerical differentiation,

$$s_{t,j}(y_t, w_t, \hat{\theta}_T) \equiv [\ln\varphi_t^\alpha(y_t, q_t^\alpha(w_t, \hat{\theta}_T + e_j\epsilon_T)) - \ln\varphi_t^\alpha(y_t, q_t^\alpha(w_t, \hat{\theta}_T - e_j\epsilon_T))]/2\epsilon_T, \tag{15}$$

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

$$\begin{aligned} \hat{\Delta}_T(\hat{\theta}_T)_{i,j} \equiv\ & [L_T(\hat{\theta}_T + e_i\epsilon_T + e_j\epsilon_T) - L_T(\hat{\theta}_T - e_i\epsilon_T + e_j\epsilon_T) \\ & - L_T(\hat{\theta}_T + e_i\epsilon_T - e_j\epsilon_T) + L_T(\hat{\theta}_T - e_i\epsilon_T - e_j\epsilon_T)]/4\epsilon_T^2. \end{aligned} \tag{16}$$

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

# 5 Minimax Representation

Our previous theoretical developments have shown that the quasi-maximum likelihood approach based on tick-exponential family of densities provides consistent and asymptotically normal estimators for conditional quantiles. In practice, however, solving the maximization problem (6) is made difficult by the properties of the objective function $L_T$, which is (i) not linear, (ii) not convex and (iii) not everywhere differentiable. We illustrate the implications of (i)-(iii) on the optimization of $L_T(\theta)$, by considering the most favorable situation in which the conditional quantile $q_t^\alpha$ is linear in $\theta$, $q_t^\alpha(w_t, \theta) = \theta' w_t$, and the function $a_t$ is linear in $\eta$.

In the linear case, (6) can be formulated as a linear program: the main idea is to consider the dual problem in $d \equiv (d_1, \ldots, d_T)'$, which can be written $\max_d T^{-1} \sum_{t=1}^T y_t d_t$ subject to $d_t \in [\alpha - 1, \alpha]$ for all $t$, $1 \leqslant t \leqslant T$, and $T^{-1} \sum_{t=1}^T w_t d_t = 0$. Questions of uniqueness of the solution to the dual problem and the practical implementation of the linear programming algorithm have already been studied in the literature (see, e.g., Buchinsky, 1992, Koenker and Park, 1996). When we relax the linearity of $q_t^\alpha$, the objective function $\ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \theta))$ is not linear in $\theta$ and the initial optimization problem no longer has a linear programming representation. However, due to the linearity of $a_t$, the function to maximize is still convex (though non-differentiable) in $\theta$ and its optimization can be carried out by using algorithms based on the computation of sub-gradients, such as cutting plane methods (see, e.g. Frenk, Gromicho and Zhang, 1994) or can be solved directly by interior-point methods (see, e.g., Koenker and Park, 1996). Once we consider functions $a_t$ which are not linear in $\eta$, we are left with objective functions $L_T$ which are neither convex nor differentiable in $\theta$. Hence, standard optimization techniques for convex non-differentiable functions no longer apply and one is left with non-gradient based methods such as simulated annealing, genetic algorithm or Markov Chain Monte Carlo methods (see, e.g., Chernozhukov and Hong, 2003).

The optimization algorithm that we propose in this paper is based on the following simple idea: the function $\varphi_t^\alpha$ in (3) is exponential by parts and can therefore be represented as a maximum of two exponential branches which are twice continuously differentiable. For example, consider a simple case where $T = 1$, 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. $\max_{\theta\in\Theta} \min \{\ln \psi_1^\alpha(y_1, q_t^\alpha(w_{1t}, \theta)), \ln \phi_1^\alpha(y_1, q_t^\alpha(w_{1t}, \theta))\}$, where we have defined

$$\psi_t^\alpha(y, \eta) \equiv \exp\{\alpha[a_t(\eta) - c_t(y)]\} \text{ and } \phi_t^\alpha(y, \eta) \equiv \exp\{-(1-\alpha)[a_t(\eta) - b_t(y)]\}, \tag{17}$$

for all $t > 0$, $y \in \mathbb{R}$ and $\eta \in M_t$. Provided (A3) and (A3') hold, the functions $\psi_t^\alpha(y, \cdot) : M_t \to \mathbb{R}$ and $\phi_t^\alpha(y, \cdot) : M_t \to \mathbb{R}$ in (17) are twice continuously differentiable. By using the fact that for all $(x, y) \in \mathbb{R}^2$ we have $\min \{x, y\} = -\max \{-x, -y\}$, and that the parameter space $\Theta$ is compact (A2), the previous maximization problem is equivalent to the minimization problem

$$-\min_{\theta\in\Theta}[\max \{-\ln \psi_1^\alpha(y_1, q_t^\alpha(w_{1t}, \theta)), -\ln \phi_1^\alpha(y_1, q_t^\alpha(w_{1t}, \theta))\}]. \tag{18}$$

We have thus transformed the initial maximization problem into a minimax problem (18), which involves only functions $\psi_1^\alpha$ and $\phi_1^\alpha$ which are twice continuously differentiable in $\theta$. In other words, even though we cannot change the convexity of the objective function $L_T$, we can recover the differentiability property by using the minimax representation. Note that this transformation makes minimal assumptions on $\Theta$ and is applicable for both linear and nonlinear conditional quantile models $q_t^\alpha$ as well as for functions $a_t$ which may or may not be linear. We show, in the following theorem, that a result similar to the one above applies for $T > 1$.

**Theorem 6 (minimax representation)** *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 (A2), the QMLE $\hat{\theta}_T$ is a solution to a minimax problem*

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

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

The tick-exponential QMLE $\hat{\theta}_T$ can thus be obtained as a solution to the classical minimax problem, in which the function $P_k(y_\theta, w_\theta, \cdot)$ is twice continuously differentiable on $\Theta$, for all $k$, $0\leqslant k\leqslant T$. The minimax problem in Theorem 6 can further be transformed into a constrained minimization problem with quadratic objective function, whose solution converges to the solution of the initial problem. Note that solving $\min_{\theta\in\Theta}[\max_{0\leqslant k\leqslant T}\{P_k(y_\theta, w_\theta, \theta)\}]$ is equivalent to solving: $\min\gamma$ subject to: $P_k(y_\theta, w_\theta, \theta)\leqslant\gamma$, for all $k$, $0\leqslant k\leqslant T$. The Kuhn-Tucker equations relative to this constrained minimization problem are the same as if we were searching the step $\Delta\theta\neq 0$ which solves the quadratic problem: $\min_{\Delta\theta}\frac{1}{2}\Delta\theta'\tilde{Q}\Delta\theta+\Delta\gamma$ subject to: $P_k(y_\theta, w_\theta, \theta)+\nabla_\theta P_k(y_\theta, w_\theta, \theta)'\Delta\theta\leqslant\gamma+\Delta\gamma$, in which $\tilde{Q}$ is some positive definite matrix. In other words, the step $\Delta\theta$ is in a direction of descent for the function $\max_{0\leqslant k\leqslant T}\{P_k(y_\theta, w_\theta, \theta)\}$, which ensures the convergence of the algorithm (see, e.g., Theorem 1 in Brayton, Director, Hachtel and Vidigal, 1979). The practical implementation of a minimax algorithm can, for example, be done by using a Sequential Quadratic Programming (SQP) method.[8] SQP methods represent the state of the art in nonlinear programming methods and their overview can be found in Gill, Murray and Wright (1981).

# 6 Conclusion

In this paper we have defined a new family of densities, called the tick-exponential family, 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 to show that if

[8] This is the approach used to implement the function `fminimax` in Matlab statistical software.

one can consistently estimate the true parameter of a correctly specified conditional quantile model by maximizing a quasi-likelihood, then this quasi-likelihood necessarily belongs to the tick-exponential family. In other words, we show that the tick-exponential assumption is a necessary condition for consistency. As a second result of the paper, we provide a set of primitive conditions which are sufficient for the class of tick-exponential QMLEs to be consistent. Thirdly, we show that the latter are also asymptotically normally distributed with the asymptotic covariance matrix which accounts for possible model misspecification. As a natural application of our results we propose a consistent covariance matrix estimator based on the tick-exponential scores and hessian thus providing an alternative to extant kernel- or bootstrap based methods. For practical purposes and as a final result of our paper, we provide an easy-to-implement algorithm for the maximization of the tick-exponential (quasi) log-likelihood based on the minimax representation.

# References


[1] Andrews, D.W.K., (1994), "Empirical Process Methods in Econometrics", in R. F. Engle and D. L. Newey (eds.), *Handbook of Econometrics*, 4, 2247-2294, Elsevier Science.

[2] Bierens, H.J., and Ginther, D. (2001), "Integrated Conditional Moment Testing of Quantile Regression Models", *Empirical Economics*, 26, 307-324.

[3] Bollerslev, T. (1987), "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return", *Review of Economics and Statistics*, 69, 542–547.

[4] 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.

[5] Brayton, R.K., Director, S.W., Hachtel, G.D. and Vidigal, L.M. (1979), "A New Algorithm for Statistical Circuit Design Based on Quasi-Newton Methods and Function Splitting", *IEEE Transactions on Circuits and Systems*, 26, 784-794.

[6] Buchinsky, M. (1992), "Methodological Issues in Quantile Regression", *manuscript*, Brown University.

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

[8] Chen, X., Linton, O., and van Keilegom, I. (2003), "Estimation of Semi-parametric Models when the Criterion Function Is Not Smooth", *Econometrica*, 71, 1591-1608.

[9] Chernozhukov, V. and Hong, H. (2003), "An MCMC approach to classical estimation", *Journal of Econometrics*, 115, 293-346.

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

[11] 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.

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

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

[14] Frenk, J.B.G., Gromicho, J. and Zhang, S. (1994), "A deep cut ellipsoid algorithm for convex programming: theory and applications", *Mathematical Programming*, 163, 83-108.

[15] Gill, P.E., Murray, W. and Wright, M.H. (1981), *Practical Optimization*, London, Academic Press.

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

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

[18] Hahn, J. (1995), "Bootstrapping Quantile Regression Models", *Econometric Theory*, 11, 105-121.

[19] Hjort, N. L. and D. Pollard (1993), "Asymptotics for minimizers of convex processes", *preprint (http://www.stat.yale.edu/Preprints).*

[20] Horowitz, J. (1998), "Bootstrap Methods for Median Regression Models", *Econometrica*, 66, 1327-1352.

[21] Horowitz, J. and Spokoiny, V.G. (2002), "An Adaptive Rate-Optimal Test of Linearity for Median Regression Models", *Journal of the American Statistical Association*, 97, 822-835.

[22] 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.

[23] 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.

[24] Knight, K. (1998), "Limiting distributions for $L_1$-regression estimators under general conditions", *Annals of Statistics*, 26, 755-770.

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

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

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

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

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

[30] Koenker, R. and Xiao, Z. (2002), "Inference on the Quantile Regression Process", *Econometrica*, 70, 1583-1612.

[31] 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.

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

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

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

[35] Pollard, D. (1991), "Asymptotics for Least Absolute Deviation Regression Estimators", *Econometric Theory*, 7, 186-199.

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

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

[38] Ruppert, D. and Caroll, R. J. (1980), "Trimmed Least Squares Estimation in the Linear Model", *Journal of the American Statistical Association*, 75, 828-838.

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

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

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

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

# Appendix

**Notation:**

if $V$ is a real $n$-vector, $V \equiv (V_1, \ldots, V_n)'$, then $|V|$ denotes the $L_2$-norm of $V$, i.e. $|V|^2 \equiv V'V = \sum_{i=1}^n V_i^2$. If $M$ is a real $n \times n$-matrix, $M \equiv (M_{ij})_{1 \leqslant i,j \leqslant n}$, then $|M|$ denotes the $L_\infty$-norm of $M$, i.e. $|M| \equiv \max_{1 \leqslant i,j \leqslant n} |M_{ij}|$. The function $1 : \mathcal{F}_t \to \{0,1\}$ is a standard indicator function, i.e. for any event $A \in \mathcal{F}_t$ we have $1\{A\} = 1$ if $A$ is true and $= 0$ otherwise. Note that we use a slightly different notation for the Heaviside function $1 : \mathbb{R} \to [0,1]$ which is a real function, i.e. for any $x \in \mathbb{R}$, we have $1(x) = 0$ if $x < 0$ and $1(x) = 1$ if $x \geqslant 0$.

**Proof of Theorem 2.** First, note that under *(i) - (iv)* $\hat{\theta}_T$ is consistent for $\theta^* \in \mathring{\Theta}$ (see, e.g., Theorem 2.1 in Newey and McFadden, 1994). If $\mathcal{M}$ is correctly specified, then $\theta^* = \theta_0$ only if $\nabla_\theta E[L_T(\theta_0)] = 0$, i.e. $T^{-1} \sum_{t=1}^T E\{\nabla_\theta q_t^\alpha(W_t, \theta_0) E[\partial \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta_0))/\partial\eta | \mathcal{G}_t]\} = 0$. Since this first order condition needs to hold for any sample size $T$, any choice of $\mathcal{M}$ and any true value $\theta_0 \in \mathring{\Theta}$, we have that $\theta^* = \theta_0$ only if

$$\begin{aligned} E[1(q_t^\alpha(W_t, \theta_0) - Y_t) - \alpha | \mathcal{G}_t] &= 0, a.s. - P_0 \\ &\Rightarrow E[\partial \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta_0))/\partial\eta | \mathcal{G}_t] = 0, a.s. - P_0, \end{aligned} \tag{19}$$

for all $t$, $1 \leqslant t \leqslant T$, and all absolutely continuous distribution function $F_{0,t}$ in $\tilde{F}$. Let then

$$\tilde{a}_t(Y_t, q_t^\alpha(W_t, \theta_0)) \equiv \frac{\partial \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta_0))/\partial\eta}{\alpha - 1(q_t^\alpha(W_t, \theta_0) - Y_t)}. \tag{20}$$

We now show that (19) implies: (1) $\tilde{a}_t$ is $\mathcal{G}_t$-measurable, and (2) $\tilde{a}_t > 0$, for any $t$, $1 \leqslant t \leqslant T$. Consider a decomposition $\tilde{a}_t(Y_t, q_t^\alpha(W_t, \theta_0)) = E[\tilde{a}_t(Y_t, q_t^\alpha(W_t, \theta_0)) | \mathcal{G}_t] + \varepsilon_t$, where $E[\varepsilon_t | \mathcal{G}_t] = 0$. Then $E[\frac{\partial \ln \varphi_t^\alpha}{\partial\eta}(Y_t, q_t^\alpha(W_t, \theta_0)) | \mathcal{G}_t] = E[\tilde{a}_t(Y_t, q_t^\alpha(W_t, \theta_0)) | \mathcal{G}_t] E[\alpha - 1(q_t^\alpha(W_t, \theta_0) - Y_t) | \mathcal{G}_t]$ $+ E\{\varepsilon_t[\alpha - 1(q_t^\alpha(W_t, \theta_0) - Y_t)] | \mathcal{G}_t\}$, so that (19) implies that for every $F_{0,t} \in \tilde{F}$ we have $E\{\varepsilon_t[\alpha - 1(q_t^\alpha(W_t, \theta_0) - Y_t)] | \mathcal{G}_t\} = 0, a.s. - P_0$, for all $t$, $1 \leqslant t \leqslant T$. Hence $\varepsilon_t = 0, a.s. - P_0$, for all $t$, $1 \leqslant t \leqslant T$, so that $\tilde{a}_t$ is $\mathcal{G}_t$-measurable in which case it cannot depend on $Y_t$ and we can write

$$\tilde{a}_t(Y_t, q_t^\alpha(W_t, \theta_0)) = \tilde{a}_t(q_t^\alpha(W_t, \theta_0)). \tag{21}$$

Moreover, $\tilde{a}_t > 0$: a necessary condition for $\theta_0 \in \mathring{\Theta}$ to be a maximizer of $E[L_T(\theta)]$ is that for every $\xi \in \mathbb{R}^k$, $\xi' \Delta_{\theta\theta} E[L_T(\theta_0)]\xi \leqslant 0$. Combining (20), (21) and (A1), we have $\xi' \Delta_{\theta\theta} E[L_T(\theta_0)]\xi = -T^{-1} \sum_{t=1}^{T} Q_t$, where for any $t$, $1 \leqslant t \leqslant T$,

$$Q_t \equiv E[(\xi' \nabla_\theta q_t^\alpha(W_t, \theta_0))^2 \tilde{a}_t(q_t^\alpha(W_t, \theta_0)) f_{0,t}(q_t^\alpha(W_t, \theta_0))].$$

Hence, $\xi' \Delta_{\theta\theta} E[L_T(\theta_0)]\xi$ is negative for any sample size $T$, only if $Q_t \geqslant 0$, for all $t$, $1 \leqslant t \leqslant T$. Since this inequality needs to hold for any conditional pdf $f_{0,t}$, a necessary condition is that $\tilde{a}_t(q_t^\alpha(W_t, \theta_0)) \geqslant 0, a.s. - P_0$, for all $t$, $1 \leqslant t \leqslant T$. The interiority of the solution $\theta_0$ moreover implies that $\tilde{a}_t \neq 0$ so that $\tilde{a}_t > 0$, for all $t$, $1 \leqslant t \leqslant T$. Therefore, (19) implies that for any $\theta_0 \in \mathring{\Theta}$ and any $t$, $1 \leqslant t \leqslant T$,

$$\partial \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta_0))/\partial \eta = \tilde{a}_t(q_t^\alpha(W_t, \theta_0))[\alpha - 1(q_t^\alpha(W_t, \theta_0) - Y_t)], a.s. - P_0, \tag{22}$$

with $\tilde{a}_t > 0$. The remainder of the proof is straightforward: we need to integrate (22) with respect to $\eta$. Taking into account the $a.s. - P_0$ continuity of $\ln \varphi_t^\alpha$ and the fact that (22) needs to hold for any $\theta_0 \in \mathring{\Theta}$, the above integrates into

$$\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) = \begin{cases} -(1-\alpha)[a_t(q_t^\alpha(W_t, \theta)) - b_t(Y_t)], \text{ if } Y_t \leqslant q_t^\alpha(W_t, \theta), \\ \alpha[a_t(q_t^\alpha(W_t, \theta)) - c_t(Y_t)], \text{ if } Y_t > q_t^\alpha(W_t, \theta), \end{cases} , a.s. - P_0,$$

where for every $t$, $1 \leqslant t \leqslant T$, $a_t$ is an indefinite integral of $\tilde{a}_t$ so that $a_t' = \tilde{a}_t > 0$, and $b_t : \mathbb{R} \to \mathbb{R}$ and $c_t : \mathbb{R} \to \mathbb{R}$ are $\mathcal{F}_t$-measurable. This completes the proof of Theorem 2. ■

**Proof of Theorem 3.** As previously, we consider $L_T(\theta) = T^{-1} \sum_{t=1}^{T} \ln \varphi_t^\alpha(y_t, q_t^\alpha(w_t, \theta))$ and

$$\begin{aligned} E[L_T(\theta)] &= T^{-1} \sum_{t=1}^{T} E\{a_t(q_t^\alpha(W_t, \theta)) E_t[\alpha - 1(q_t^\alpha(W_t, \theta) - Y_t)] \\ &\quad + (1-\alpha) E_t[b_t(Y_t) \cdot 1(q_t^\alpha(W_t, \theta) - Y_t)] - \alpha E_t[c_t(Y_t) \cdot (1 - 1(q_t^\alpha(W_t, \theta) - Y_t))]\}, \end{aligned} \tag{23}$$

and check that all the conditions of Theorem 2.1 in Newey and McFadden (1994, p 2121) hold. First, note that their condition *(ii)* is satisfied by imposing (A2). Next, we show that their uniqueness condition *(i)* holds: assume there exists $\tilde{\theta} \in \Theta$ such that $E[L_T(\tilde{\theta})] = E[L_T(\theta^*)]$. Since for any $\tilde{\theta} \in \Theta$, $E[L_T(\theta^*) - L_T(\tilde{\theta})] \geqslant 0$, the previous equality implies

$$\begin{aligned}
&[a_t(q_t^\alpha(W_t, \tilde{\theta})) - a_t(q_t^\alpha(W_t, \theta^*))]E_t[\alpha - 1(q_t^\alpha(W_t, \theta^*) - Y_t)] \\
&+ a_t(q_t^\alpha(W_t, \tilde{\theta}))E_t[d_t(Y_t, W_t, \tilde{\theta}, \theta^*)] \\
&= E_t\{[(1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)]d_t(Y_t, W_t, \tilde{\theta}, \theta^*)\}, \ a.s. - P_0,
\end{aligned} \tag{24}$$

for every $t$, $1 \leqslant t \leqslant T$, where $d_t(Y_t, W_t, \tilde{\theta}, \theta^*) \equiv 1(q_t^\alpha(W_t, \theta^*) - Y_t) - 1(q_t^\alpha(W_t, \tilde{\theta}) - Y_t)$. Let $A_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega), \theta^*) \leqslant q_t^\alpha(W_t(\omega), \tilde{\theta})\}$ and $B_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega), \theta^*) < Y_t(\omega) \leqslant q_t^\alpha(W_t(\omega), \tilde{\theta})\}$. We have $d_t(Y_t, W_t, \tilde{\theta}, \theta^*) = -1$ on $A_t \cap B_t$, and $= 0$ on $A_t \cap B_t^c$. Hence, (24) becomes

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

Next, let $C_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega), \tilde{\theta}) < Y_t(\omega) \leqslant q_t^\alpha(W_t(\omega), \theta^*)\}$. Similarly, $d_t(Y_t, W_t, \tilde{\theta}, \theta^*) = 1$ on $A_t^c \cap C_t$, and $= 0$ on $A_t^c \cap C_t^c$, so that (24) becomes

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

Hence, for every $t$, $1 \leqslant t \leqslant T$, $a_t(q_t^\alpha(W_t, \tilde{\theta})) = a_t(q_t^\alpha(W_t, \theta^*))$, on $A_t \cap B_t^c$ and $A_t^c \cap C_t^c$, which by continuity and monotonicity of $a_t$ on $M_t$ in turn implies $q_t^\alpha(W_t, \tilde{\theta}) = q_t^\alpha(W_t, \theta^*)$, on $A_t \cap B_t^c$ and $A_t^c \cap C_t^c$, which can only hold if $A_t \equiv \{\omega \in \Omega : q_t^\alpha(W_t(\omega), \theta^*) = q_t^\alpha(W_t(\omega), \theta)\}$, i.e. $A_t^c = \emptyset$ and $B_t = C_t = \emptyset$. The variable $Y_t$ being continuously distributed, this implies $q_t^\alpha(W_t, \tilde{\theta}) = q_t^\alpha(W_t, \theta^*)$, $a.s. - P_0$, for all $t$, $1 \leqslant t \leqslant T$, which by using the identification condition *(i)* in (A0) gives $\tilde{\theta} = \theta^*$.

We next show that the continuity condition *(iii)* and the uniform convergence condition *(iv)* of Theorem 2.1 hold, by using a weak uniform law of large numbers (ULLN) in Theorem A.2.5 in White (1994, p 353). We start by showing that for every $t$, $1 \leqslant t \leqslant T$, the function $\theta \mapsto \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))$ is Lipshitz-$L_1$ $a.s.$ on $\Theta$ (see Definition A.2.3 in White, 1994, p 352):

recall that for every $t$, $1 \leqslant t \leqslant T$, $q_t^\alpha(W_t, \cdot)$ is continuous $a.s.$ on $\Theta$, i.e. for each $\bar{\theta} \in \Theta$ and for each $\varepsilon > 0$ there exists $\bar{\delta}_\varepsilon > 0$ such that for $|\theta - \bar{\theta}| < \bar{\delta}_\varepsilon$, $q_t^\alpha(W_t, \bar{\theta})$ and $q_t^\alpha(W_t, \theta)$ are sufficiently close, meaning that if $Y_t \gtrless q_t^\alpha(W_t, \bar{\theta})$, $a.s. - P_0$, then $Y_t \gtrless q_t^\alpha(W_t, \theta)$, $a.s. - P_0$. Now fix some $t$, $1 \leqslant t \leqslant T$, and consider the following two cases:

CASE 1: if $Y_t < q_t^\alpha(W_t, \bar{\theta})$, $a.s. - P_0$, then for $|\theta - \bar{\theta}| < \bar{\delta}_\varepsilon$ we have $|\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) - \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \bar{\theta}))| = (1-\alpha) a_t'(q_t^\alpha(W_t, \bar{\theta}^c)) |\nabla_\theta q_t^\alpha(W_t, \bar{\theta}^c)'(\theta - \bar{\theta})|$, $a.s. - P_0$, for some $\bar{\theta}^c \equiv c\theta + (1-c)\bar{\theta}$, $c \in (0,1)$. Using (A3) we then have $|\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) - \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \bar{\theta}))| \leqslant \bar{L}_t |\theta - \bar{\theta}|$, $a.s. - P_0$, where $\bar{L}_t \equiv K |\nabla_\theta q_t^\alpha(W_t, \bar{\theta}^c)|$. From (A5) we know that $E[|\nabla_\theta q_t^\alpha(W_t, \bar{\theta}^c)|] \leqslant E[\sup_{\theta \in \Theta} |\nabla_\theta q_t^\alpha(W_t, \theta)|] < \infty$, so that $E[|\bar{L}_t|] < \infty$, for any $t$, $1 \leqslant t \leqslant T$. Hence, we have an $\mathcal{F}_t$-measurable random variable $\bar{L}_t : \Omega \to \mathbb{R}^+$ such that for all $\theta : |\theta - \bar{\theta}| \leqslant \bar{\delta}$,

$$|\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) - \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \bar{\theta}))| \leqslant \bar{L}_t |\theta - \bar{\theta}|, \ a.s. - P_0, \tag{25}$$

and which satisfies $T^{-1} \sum_{t=1}^T E(\bar{L}_t) < \infty$.

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 (25) holds, so that for every $t$, $1 \leqslant t \leqslant T$, the function $\theta \mapsto \ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))$ is Lipshitz-$L_1$ a.s. on $\Theta$. Next, we need to show that the sequences $\{\bar{\varphi}_t(\delta)\}$ and $\{\underline{\varphi}_t(\delta)\}$ obey the weak law of large numbers (LLN) locally at $\bar{\theta}$ for all $\bar{\theta} \in \mathring{\Theta}$, where for every $t$, $1 \leqslant t \leqslant T$, we have defined $\bar{\varphi}_t(\delta) \equiv \sup_{\theta \in \Theta}\{\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) : |\theta - \bar{\theta}| < \delta\}$ and $\underline{\varphi}_t(\delta) \equiv \inf_{\theta \in \Theta}\{\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta)) : |\theta - \bar{\theta}| < \delta\}$ (see Definition A.2.4 and Theorem A.2.5 in White, 1994). We know from (A4) that $\{X_t\}$ is $\alpha$-mixing of size $-r/(r-2)$, $r > 2$, so that by condition *(iv)* in (A0) $W_t$ is $\alpha$-mixing of same size (see Theorem 3.49 in White, 2001, p 50). Similarly, for every $t$, $1 \leqslant t \leqslant T$, $\ln \varphi_t^\alpha$ is an $\mathcal{F}_t$-measurable function of $Y_t$ and $W_t$. Hence, for every $\theta \in \Theta$ the sequence $\{\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))\}$ is $\alpha$-mixing of size $-r/(r-2)$, $r > 2$, then so must be $\{\bar{\varphi}_t(\delta)\}$ and $\{\underline{\varphi}_t(\delta)\}$. Now fix some $\bar{\theta} \in \mathring{\Theta}$ and let $\bar{\delta}_\theta > 0$ be such that $\{\theta : |\theta - \bar{\theta}| < \bar{\delta}_\theta\} \subset \Theta$. Note that for every $0 < \delta \leqslant \bar{\delta}_\theta$ and every $t$, $1 \leqslant t \leqslant T$, we have $|\bar{\varphi}_t(\delta)| \leqslant \sup_{\theta \in \Theta} |\ln \varphi_t^\alpha(Y_t, q_t^\alpha(W_t, \theta))|$, $a.s. - P_0$, so that by triangular inequality $|\bar{\varphi}_t(\delta)| \leqslant \sup_{\theta \in \Theta} |a_t(q_t^\alpha(W_t, \theta))| + |b_t(Y_t)| + |c_t(Y_t)|, a.s. - P_0$. The same inequality holds for $\underline{\varphi}_t(\delta)$. Now, for a given $r > 2$ and $\epsilon > 0$, there exists a constant $n_{r,\epsilon} > 1$ such that, for every

$0 < \delta \leqslant \bar{\delta}_\theta$ and every $t$, $1 \leqslant t \leqslant T$, we have

$$\begin{aligned} E(|\bar{\varphi}_t(\delta)|^{r+\epsilon}) &\leqslant n_{r,\epsilon} E\{[\sup_{\theta\in\Theta} |a_t(q_t^\alpha(W_t,\theta))|]^{r+\epsilon} + |b_t(Y_t)|^{r+\epsilon} + |c_t(Y_t)|^{r+\epsilon}\} \\ &\leqslant n_{r,\epsilon}[\max\{1, E(\sup_{\theta\in\Theta} |a_t(q_t^\alpha(W_t,\theta))|^{r+\epsilon})\} + E(|b_t(Y_t)|^{r+\epsilon}) + E(|c_t(Y_t)|^{r+\epsilon})], \end{aligned}$$

where the same inequality holds for $\underline{\varphi}_t(\delta)$. Thus, by (A5), $E(|\bar{\varphi}_t(\delta)|^{r+\epsilon}) < \infty$ and $E(|\underline{\varphi}_t(\delta)|^{r+\epsilon}) < \infty$, for all $t$, $1 \leqslant t \leqslant T$. The weak LLN for $\alpha$-mixing sequences provided by Corollary 3.48 in White (2001, p 49) then ensures $T^{-1}\sum_{t=1}^T \{\bar{\varphi}_t(\delta) - E[\bar{\varphi}_t(\delta)]\} \xrightarrow{p} 0$ and, similarly, $T^{-1}\sum_{t=1}^T \{\underline{\varphi}_t(\delta) - E[\underline{\varphi}_t(\delta)]\} \xrightarrow{p} 0$. (Note that the results of Corollary 3.48 in White, 2001, p 49, apply to mixing sequences with $\alpha$ of size $-r/(r-1)$ with $r > 1$. We know however that, for $r > 2$, mixing with $\alpha$ of size $-r/(r-2)$ implies mixing with $\alpha$ of size $-r/(r-1)$).

We can now use the results from Theorem A.2.5 in White (1994, p 353): the above convergence of $\{\bar{\varphi}_t(\delta)\}$ and $\{\underline{\varphi}_t(\delta)\}$ together with the Lipshitz-$L_1$ property of $\{\ln \varphi_t^\alpha\}$ imply that the function $\theta \mapsto E[L_T(\theta)]$ is continuous on $\Theta$, and that $L_T(\theta) - E[L_T(\theta)] \xrightarrow{p} 0$ uniformly on $\Theta$. This shows that the conditions *(iii)* and *(iv)* of Newey and McFadden's (1994) Theorem 2.1 hold. We can therefore apply their result to show that $\hat{\theta} \xrightarrow{p} \theta^*$, which completes the first part of the proof.

It remains to show that when $\mathcal{M}$ is correctly specified so that (A1) holds, we have $\theta^* = \theta_0$ where $\theta_0$ corresponds to the true value of the parameter $\theta$. For this it is sufficient to show that $\theta_0$ maximizes $E[L_T(\theta)]$. Combining (23) with (A1), we know that $E[L_T(\theta_0)] \geqslant E[L_T(\theta)]$ if and only if for every $t$, $1 \leqslant t \leqslant T$, we have

$$a_t(q_t^\alpha(W_t,\theta))E_t[d_t(Y_t,W_t,\theta,\theta_0)] \leqslant E_t\{[(1-\alpha)b_t(Y_t)+\alpha c_t(Y_t)]d_t(Y_t,W_t,\theta,\theta_0)\}, a.s.-P_0. \quad (26)$$

As previously, $d_t(Y_t,W_t,\theta,\theta_0) = -1$ on $A_t \cap B_t$, $= 0$ on $A_t \cap B_t^c$, $= 1$ on $A_t^c \cap C_t$, and $= 0$ on $A_t^c \cap C_t^c$. Moreover, by continuity of $\varphi_t^\alpha(\cdot,\eta)$, $(1-\alpha)b_t(Y_t)+\alpha c_t(Y_t) = a_t(Y_t) \leqslant a_t(q_t^\alpha(W_t,\theta))$ on $A_t \cap B_t$ (recall that $a_t' > 0$) and $(1-\alpha)b_t(Y_t) + \alpha c_t(Y_t) = a_t(Y_t) > a_t(q_t^\alpha(W_t,\theta))$ on $A_t^c \cap C_t$. By using a similar reasoning to the one above we show that $a_t(q_t^\alpha(W_t,\theta))d_t(Y_t,W_t,\theta,\theta_0)$ $\leqslant [(1-\alpha)b_t(Y_t) + \alpha c_t(Y_t)]d_t(Y_t,W_t,\theta,\theta_0)$, $a.s. - P_0$, which implies (26) and ensures that $\theta_0$ is a maximizer of $E[L_T(\theta)]$. Combined with the previous uniqueness result, this implies that

$\theta^* = \theta_0$ so that under correct model specification we have $\hat{\theta}_T \xrightarrow{p} \theta_0$, which completes the proof of Theorem 3. ■

**Proof of Theorem 4 and Corollary 5.** In this proof, we check that all the conditions of Theorem 7.2 in Newey and McFadden (1994, p 2186) hold. We first show that $\hat{\theta}_T$ satisfies an asymptotic first order condition:

**Lemma 7** *Under assumptions of Theorem 4,* $\sqrt{T}\nabla_\theta L_T(\hat{\theta}_T) \xrightarrow{p} 0$.

As shown in the proof of Theorem 3, we have $\nabla_\theta E[L_T(\theta^*)] = 0$ and $\nabla_\theta E[L_T(\theta_0)] = 0$ when $\mathcal{M}$ is correctly specified (condition *(i)* of Theorem 7.2). By (A2), $\theta^*$ and $\theta_0$ are interior points of $\Theta$ so that condition *(ii)* holds. We now check the nonsingularity condition *(iii)*: by (A3') we know that, for every $t$, $1 \leqslant t \leqslant T$, $a_t'(q_t^\alpha(W_t,\theta))\nabla_\theta q_t^\alpha(W_t,\theta)$ is continuously differentiable $a.s. - P_0$ on $\Theta$ with derivative $a_t''(q_t^\alpha(W_t,\theta))\nabla_\theta q_t^\alpha(W_t,\theta)\nabla_\theta q_t^\alpha(W_t,\theta)'$ $+a_t'(q_t^\alpha(W_t,\theta))\Delta_{\theta\theta} q_t^\alpha(W_t,\theta)$. Let then

$$\Delta_t(Y_t, W_t, \theta) \equiv -\Delta_{1,t}(\theta) + \Delta_{2,t}(\theta), \tag{27}$$

where for all $t$, $1 \leqslant t \leqslant T$,

$$\begin{aligned}
\Delta_{1,t}(\theta) &\equiv a_t'(q_t^\alpha(W_t,\theta))\nabla_\theta q_t^\alpha(W_t,\theta)\nabla_\theta q_t^\alpha(W_t,\theta)'\delta(q_t^\alpha(W_t,\theta) - Y_t), \\
\Delta_{2,t}(\theta) &\equiv [a_t''(q_t^\alpha(W_t,\theta))\nabla_\theta q_t^\alpha(W_t,\theta)\nabla_\theta q_t^\alpha(W_t,\theta)' \\
&\quad +a_t'(q_t^\alpha(W_t,\theta))\Delta_{\theta\theta} q_t^\alpha(W_t,\theta)][\alpha - 1(q_t^\alpha(W_t,\theta) - Y_t)],
\end{aligned}$$

and $\delta$ is the Dirac delta function (see footnote 1). We first show that locally at any $\theta \in \Theta$, the sample mean of $\{\Delta_t(Y_t, W_t, \theta)\}$ converges in probability to its expected value. By (iv) in (A0) and (A4) we know that for every $\theta \in \Theta$, $\{\Delta_t(Y_t, W_t, \theta)\}$ is $\alpha$-mixing with $\alpha$ of size $-r/(r-2)$ with $r > 2$ (see Theorem 3.49 in White, 2001, p 50). From (A5') we know that for $r > 2$ and some $\epsilon > 0$, $E(\sup_{\theta\in\Theta} |\nabla_\theta q_t^\alpha(W_t,\theta)|^{2(r+\epsilon)}) < \infty$ and $E(\sup_{\theta\in\Theta} |\Delta_{\theta\theta} q_t^\alpha(W_t,\theta)|^{r+\epsilon}) < \infty$, for all $t$, $1 \leqslant t \leqslant T$. Moreover, using (A3), (A3') and triangle inequality, there exist some some constant $n_{r,\epsilon} > 1$ such that $|\Delta_t(Y_t, W_t, \theta)|^{r+\epsilon} \leqslant n_{r,\epsilon}[L^{r+\epsilon}|\nabla_\theta q_t^\alpha(W_t,\theta)\nabla_\theta q_t^\alpha(W_t,\theta)'|^{r+\epsilon}$

$+K^{r+\epsilon}|\Delta_{\theta\theta}q_t^{\alpha}(W_t,\theta)|^{r+\epsilon}]$, $a.s. - P_0$. The norm equivalence, $|\nabla_\theta q_t^{\alpha}(W_t,\theta)\nabla_\theta q_t^{\alpha}(W_t,\theta)'|^{r+\epsilon} \leqslant c^2|\nabla_\theta q_t^{\alpha}(W_t,\theta)|^{2(r+\epsilon)}$, $a.s. - P_0$, for some $c > 0$ and all $\theta \in \Theta$, (A5') and the previous inequality then imply $E(\sup_{\theta\in\Theta}|\Delta_t(Y_t,W_t,\theta)|^{r+\epsilon}) < \infty$, for all $t$, $1 \leqslant t \leqslant T$. Applying the weak LLN for $\alpha$-mixing sequences (see Corollary 3.48 in White, 2001, p 49) then gives $T^{-1}\sum_{t=1}^{T}\Delta_t(Y_t,W_t,\theta) \xrightarrow{p} \Delta(\theta) \equiv T^{-1}\sum_{t=1}^{T}E[\Delta_t(Y_t,W_t,\theta)]$ locally at $\theta$, for all $\theta \in \Theta$. We can then say that $\Delta(\theta)$ is the derivative of $\nabla_\theta E[L_T(\theta)]$ in a neighborhood of any $\theta \in \Theta$. In particular, for $\theta^*$ we have $\Delta(\theta^*)$ as in (9). For any $\xi \in \mathbb{R}^k$, we know that $\xi'\Delta(\theta^*)\xi \leqslant 0$ so that $\xi'\Delta(\theta^*)\xi = 0$ implies that for all $t$, $1 \leqslant t \leqslant T$, $\xi' E[\Delta_{1,t}(\theta^*)|\mathcal{G}_t]\xi = \xi' E[\Delta_{2,t}(\theta^*)|\mathcal{G}_t]\xi$, $a.s. - P_0$. We now show (by contradiction) that $\xi = 0$: assume $\xi \neq 0$. Given positive definiteness of $E[\Delta_{1,t}(\theta^*)|\mathcal{G}_t]$, for all $t$, $1 \leqslant t \leqslant T$, (since $a_t' > 0$, $f_{0,t} > 0$ and $\nabla_\theta q_t \nabla_\theta q_t'$ of full rank) we then have $\xi' E[\Delta_{2,t}(\theta^*)|\mathcal{G}_t]\xi > 0$, $a.s. - P_0$. Consider for example $q_t^{\alpha}(W_t,\theta) = \theta$ so that $\xi' E[\Delta_{2,t}(\theta^*)|\mathcal{G}_t]\xi = |\xi|^2 a_t''(\theta^*)[\alpha - F_{0,t}(\theta^*)]$. We can always find $a_t$ such that $a_t''(\theta^*)$ and $[\alpha - F_{0,t}(\theta^*)]$ have opposite signs which leads to contradiction. Hence $\xi'\Delta(\theta^*)\xi = 0$ implies $\xi = 0$ so that $\Delta(\theta^*)$ is negative definite, therefore nonsingular. Same result for $\Delta(\theta_0)$ in (11) is an immediate consequence of *(iii)* in (A0), (A1), $a_t' > 0$ and $f_{0,t} > 0$, $1 \leqslant t \leqslant T$. Hence, $\Delta(\theta^*)$ and $\Delta(\theta_0)$ are nonsingular, which verifies condition *(iii)* of Theorem 7.2.

Next, we use a central limit theorem (CLT) $\alpha$-mixing sequences (see e.g. Theorem 5.20 in White, 2001, p 130) to show that condition *(iv)* of Theorem 7.2 holds. We need the following lemma:

**Lemma 8** *Under assumptions of Theorem 4,* $T^{-1}\sum_{t=1}^{T}s_t(y_t,w_t,\theta^*)s_t(y_t,w_t,\theta^*)' \xrightarrow{p} \Sigma(\theta^*)$, *with* $s_t$ *and* $\Sigma(\theta^*)$ *as in* (8) *and* (10), *respectively. If, in addition, (A1) holds, then we have* $T^{-1}\sum_{t=1}^{T}s_t(y_t,w_t,\theta_0)s_t(y_t,w_t,\theta_0)' \xrightarrow{p} \Sigma(\theta_0)$ *with* $\Sigma(\theta_0)$ *as in* (12).

By *(iv)* in (A0) and (A4), we know that the sequence $\{s_t(Y_t,W_t,\theta)\}$ is $\alpha$-mixing with $\alpha$ of size $-r/(r-2)$, $r > 2$, for every $\theta \in \Theta$ (see Theorem 3.49 in White, 2001, p 50). Using (A5') and (A3) we have: for $r > 2$ and some $\epsilon > 0$ we have $E(\sup_{\theta\in\Theta}|\nabla_\theta q_t^{\alpha}(W_t,\theta)|^{2(r+\epsilon)}) < \infty$, so that $E(\sup_{\theta\in\Theta}|s_{i,t}(Y_t,W_t,\theta)|^{2+\epsilon}) \leqslant K_i^{2+\epsilon}\max\{1, E(\sup_{\theta\in\Theta}|\nabla_\theta q_t^{\alpha}(W_t,\theta)|^{2(r+\epsilon)})\} < \infty$, for every $t$, $1 \leqslant t \leqslant T$. Taking into account results from Lemma 8, $a_t' > 0$, $0 < \alpha < 1$

and *(iii)* in (A0), we know that $\Sigma(\theta^*)$ and $\Sigma(\theta_0)$ are positive definite and can therefore apply the CLT for $\alpha$-mixing sequences (Theorem 5.20 in White, 2001, p 130) to show that $\sqrt{T}\nabla_\theta L_T(\theta^*) \xrightarrow{d} \mathcal{N}(0, \Sigma(\theta^*))$ and $\sqrt{T}\nabla_\theta L_T(\theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma(\theta_0))$.

Finally, the stochastic equicontinuity condition *(v)* of Theorem 7.2 is shown in the following lemma:

**Lemma 9** *Under assumptions of Theorem 4, for any $\delta_T \to 0$, $\sup_{\theta:|\theta-\theta^*|\leqslant\delta_T} \sqrt{T}|\nabla_\theta L_T(\theta) - \nabla_\theta L_T(\theta^*) - E[\nabla_\theta L_T(\theta)]|/[1+\sqrt{T}|\theta-\theta^*|] \xrightarrow{p} 0$, and the same holds if we replace $\theta^*$ with $\theta_0$.*

Application of the result by Newey and McFadden (1994) then gives $\sqrt{T}(\hat{\theta}_T - \theta^*) \to \mathcal{N}(0, \Delta(\theta^*)^{-1}\Sigma(\theta^*)\Delta(\theta^*)^{-1})$ with $\Sigma(\theta^*)$ as in (10) and $\Delta(\theta^*)$ as in (9). Equivalently, under (A1), $\sqrt{T}(\hat{\theta}_T - \theta_0) \to \mathcal{N}(0, \Delta(\theta_0)^{-1}\Sigma(\theta_0)\Delta(\theta_0)^{-1})$ with $\Sigma(\theta_0)$ as in (12) and $\Delta(\theta_0)$ as in (11), which completes the proof of Theorem 4. ■

**Proof of Lemma 7.** Our approach is similar to that of Ruppert and Carroll (1980) (see their proof of Lemma A.2). Let $\hat{L}_{T,j}(h) \equiv T^{-1}\sum_{t=1}^T \ln\varphi_t^\alpha(y_t, q_t^\alpha(w_t, \hat{\theta}_T + he_j)$, where $\{e_j\}_{j=1}^k$ is the standard basis of $\mathbb{R}^k$, and $h \in \mathbb{R}$ is such that for all $j = 1, \ldots, k$, $\hat{\theta}_T + he_j \in \Theta$. Note that $\hat{L}_{T,j}(0) = L_T(\hat{\theta}_T)$, for every $j = 1, \ldots, k$. Also, let $\hat{G}_{T,j}(h)$ be the derivative form right of $\hat{L}_{T,j}(h)$, so that

$$\hat{G}_{T,j}(h) = T^{-1}\sum_{t=1}^T a_t'(q_t^\alpha(w_t, \hat{\theta}_T + he_j))[\alpha - 1(q_t^\alpha(w_t, \hat{\theta}_T + he_j) - y_t)]\partial q_t^\alpha(w_t, \hat{\theta}_T + he_j)/\partial\theta_j.$$

Since the function $h \mapsto \hat{L}_{T,j}(h)$ achieves its minimum at $h = 0$ we have, for $\varepsilon > 0$, $\hat{G}_{T,j}(-\varepsilon) \leqslant \hat{G}_{T,j}(0) \leqslant \hat{G}_{T,j}(\varepsilon)$, with $\hat{G}_{T,j}(-\varepsilon) \leqslant 0$ and $\hat{G}_{T,j}(\varepsilon) \geqslant 0$. Therefore $|\hat{G}_{T,j}(0)| \leqslant \hat{G}_{T,j}(\varepsilon) - \hat{G}_{T,j}(-\varepsilon)$. By taking the limit of this inequality as $\varepsilon \to 0$, we get $|\hat{G}_{T,j}(0)| \leqslant T^{-1}\sum_{t=1}^T |a_t'(q_t^\alpha(w_t, \hat{\theta}_T))| \cdot 1\{y_t = q_t^\alpha(w_t, \hat{\theta}_T)\} \cdot |\partial q_t^\alpha(w_t, \hat{\theta}_T)/\partial\theta_j|$. Now let $\epsilon > 0$ and note that

$$\begin{aligned}
&P_0(\sqrt{T}|\nabla_\theta L_T(\hat{\theta}_T)| > \epsilon) \\
&\leqslant P_0(\sqrt{T}\max_{1\leqslant j\leqslant k}|\hat{G}_{T,j}(0)| > \epsilon) \\
&\leqslant P_0(\max_{1\leqslant j\leqslant k}\textstyle\sum_{t=1}^T |a_t'(q_t^\alpha(W_t, \hat{\theta}_T))| \cdot 1\{Y_t = q_t^\alpha(W_t, \hat{\theta}_T)\} \cdot |\partial q_t^\alpha(W_t, \hat{\theta}_T)/\partial\theta_j| > \epsilon\sqrt{T}).
\end{aligned}$$

Since $Y_t$ is continuously distributed, we have $P_0(1\{Y_t = q_t^\alpha(W_t, \hat{\theta}_T)\} \neq 0) = 0$, which combined with the fact that $a_t'$ is bounded (A3) ensures (by dominated convergence) that, for any $\epsilon > 0$, $\lim_{T\to\infty} P_0(\sqrt{T}\nabla_\theta L_T(\hat{\theta}_T) > \epsilon) = 0$, and completes the proof of Lemma 7. ■

**Proof of Lemma 8.** Let $\sigma^{m,n}(\theta)$, $1 \leqslant m, n \leqslant k$, denote the element of the matrix $\Sigma(\theta)$ which lies in the $m$th row and $n$th column, i.e.

$$\sigma^{m,n}(\theta) \equiv cov[T^{-1/2}\sum\nolimits_{t=1}^{T} s_t^m(Y_t, W_t, \theta), T^{-1/2}\sum\nolimits_{t=1}^{T} s_t^n(Y_t, W_t, \theta)].$$

Given that the sequence $\{s_t(Y_t, W_t, \theta^*)\}$ is uncorrelated and that $E[s_t(Y_t, W_t, \theta^*)] = 0$, for all $t$, $1 \leqslant t \leqslant T$, (A1') we have that $\sigma^{m,n}(\theta^*) = T^{-1}\sum_{t=1}^{T} E[s_t^m(Y_t, W_t, \theta^*)s_t^n(Y_t, W_t, \theta^*)]$, for all $m, n$. Similarly by using the martingale difference property of $\{s_t(Y_t, W_t, \theta_0), \mathcal{G}_t\}$ implied by (A1), we have $\sigma^{m,n}(\theta_0) = T^{-1}\sum_{t=1}^{T} E[s_t^m(Y_t, W_t, \theta_0)s_t^n(Y_t, W_t, \theta_0)]$. We now show that the sample mean of $\{s_t^m(Y_t, W_t, \theta^*)s_t^n(Y_t, W_t, \theta^*)\}$ and $\{s_t^m(Y_t, W_t, \theta_0)s_t^n(Y_t, W_t, \theta_0)\}$ converge to their expected values: first note that by *(iv)* in (A0) and (A4), the sequence $\{s_t(Y_t, W_t, \theta)\}$ is $\alpha$-mixing with $\alpha$ of size $-r/(r-2)$, $r > 2$, for every $\theta \in \Theta$ (see Theorem 3.49 in White, 2001, p 50), and so are $\{s_t^m(Y_t, W_t, \theta)s_t^n(Y_t, W_t, \theta)\}$ for any couple $(m, n)$, $1 \leqslant m, n \leqslant k$. Next, by (A5') we know that for $r > 2$ and some $\epsilon > 0$, $E(\sup_{\theta\in\Theta} |\nabla_\theta q_t^\alpha(W_t, \theta)|^{2(r+\epsilon)}) < \infty$, for all $t$, $1 \leqslant t \leqslant T$. Using (A3) and norm equivalence, there exists a positive constant $c$ such that $E(\sup_{\theta\in\Theta} |s_t^m(Y_t, W_t, \theta)s_t^n(Y_t, W_t, \theta)|^{r+\epsilon}) \leqslant \max\{1, K_i^{2(r+\epsilon)}\}E(\sup_{\theta\in\Theta} |\partial q_t^\alpha(W_t, \theta)/\partial\theta_m \cdot \partial q_t^\alpha(W_t, \theta)/\partial\theta_n|^{r+\epsilon}) \leqslant \max\{1, K_i^{2(r+\epsilon)}\} \max\{1, c^2 E(\sup_{\theta\in\Theta} |\nabla_\theta q_t^\alpha(W_t, \theta)|^{2(r+\epsilon)})\}$, $a.s.-P_0$, for any $\theta \in \Theta$. Hence, for any $(m, n)$, $1 \leqslant m, n \leqslant k$, $E(\sup_{\theta\in\Theta} |s_t^m(Y_t, W_t, \theta)s_t^n(Y_t, W_t, \theta)|^{r+\epsilon}) < \infty$, for all $t$, $1 \leqslant t \leqslant T$, and the weak LLN for $\alpha$-mixing sequences (see Corollary 3.48 in White, 2001, p 49) ensures that $T^{-1}\sum_{t=1}^{T} s_t^m(Y_t, W_t, \theta^*)s_t^n(Y_t, W_t, \theta^*) \xrightarrow{p} \sigma^{m,n}(\theta^*)$ and $T^{-1}\sum_{t=1}^{T} s_t^m(Y_t, W_t, \theta_0)s_t^n(Y_t, W_t, \theta_0) \xrightarrow{p} \sigma^{m,n}(\theta_0)$. Finally, note that $E_t[(\alpha - 1(q_t^\alpha(W_t, \theta^*) - Y_t)))^2] = \alpha^2 - (2\alpha - 1)F_{0,t}(q_t^\alpha(W_t, \theta^*))$ which implies (10). Under (A1), we have $\alpha^2 - (2\alpha - 1)F_{0,t}(q_t^\alpha(W_t, \theta_0)) = \alpha(1-\alpha)$, which shows (12) and completes the proof of Lemma 8. ■

**Proof of Lemma 9.** We use primitive conditions for stochastic equicontinuity given in Theorem 7.3 in Newey and McFadden (1994, p 2188): by (A1') we have $E[s_t(Y_t, W_t, \theta^*)] = 0$, for all $t$, $1 \leqslant t \leqslant T$, and the same holds for $\theta_0$ under (A1). Let then

$$r_t(Y_t, W_t, \theta) \equiv |s_t(Y_t, W_t, \theta) - s_t(Y_t, W_t, \theta^*) - \Delta_t(Y_t, W_t, \theta^*)(\theta - \theta^*)|/|\theta - \theta^*|, \tag{28}$$

where $s_t$ and $\Delta_t$ are as defined in (8) and (27), respectively. We simplify the notation and let $g_t(W_t, \theta) \equiv a_t'(q_t^\alpha(W_t, \theta))\nabla_\theta q_t^\alpha(W_t, \theta)$, $U_t \equiv q_t^\alpha(W_t, \theta^*) - Y_t$ and $\varepsilon_t \equiv q_t^\alpha(W_t, \theta) - q_t^\alpha(W_t, \theta^*)$, $1 \leqslant t \leqslant T$. Note that by (A0) and (A3') $g_t(W_t, \cdot)$, $1 \leqslant t \leqslant T$, is continuously differentiable $a.s. - P_0$ on $\Theta$ with derivative $dg_t(W_t, \cdot)$, and that $\lim_{\theta \to \theta^*} P_0(\varepsilon_t = 0) = 1$ and $P_0(U_t = 0) = 0$, for any $\theta$ in $\mathring{\Theta}$. We have

$$\begin{aligned} r_t(Y_t, W_t, \theta) \leqslant\; & |g_t(W_t, \theta) - g_t(W_t, \theta^*) - dg_t(W_t, \theta^*)'(\theta - \theta^*)|/|\theta - \theta^*| \\ & + |1(U_t + \varepsilon_t) - 1(U_t) - \varepsilon_t \delta(U_t)| \cdot |g_t(W_t, \theta)|/|\theta - \theta^*| \\ & + \delta(U_t)|\nabla_\theta q_t^\alpha(W_t, \theta^*) g_t(W_t, \theta^*)'(\theta - \theta^*)|/|\theta - \theta^*| \\ & + \delta(U_t)|[q_t^\alpha(W_t, \theta) - q_t^\alpha(W_t, \theta^*)] g_t(W_t, \theta)|/|\theta - \theta^*|, \end{aligned} \tag{29}$$

for all $t$, $1 \leqslant t \leqslant T$, and show that the four terms on the right hand side of (29), denoted $r_{i,t}(Y_t, W_t, \theta)$, $1 \leqslant i \leqslant 4$, converge to zero with probability one as $\theta \to \theta^*$: note that the mean value expansion of $g_t$ around $\theta^*$ and the continuity of $dg_t$ imply (by dominated convergence) the result for the first term. By multiplying $r_{2,t}(Y_t, W_t, \theta)$ above and below by $|\varepsilon_t|$, using (A3) and writing the mean value expansion of $q_t^\alpha$ around $\theta^*$, we show that the second term of the right hand side of (29) is bounded above by $KS_t|1(U_t + \varepsilon_t) - 1(U_t) - \varepsilon_t \delta(U_t)|/|\varepsilon_t|$, where, for any $t$, $1 \leqslant t \leqslant T$, we let $S_t \equiv \max\{1, \sup_{\theta \in \Theta} |\nabla_\theta q_t^\alpha(W_t, \theta)|^2\}$. Note that by (A5') we have $E(S_t^2) < \infty$, for every $t$, $1 \leqslant t \leqslant T$. For any $\eta > 0$ and $\varepsilon > 0$ let then $S \equiv [2E(S_t^2)/\eta]^{1/2} < \infty$, $\varepsilon' \equiv \varepsilon/K > 0$ and $\varepsilon'' \equiv \varepsilon'/S > 0$: we have $P_0(r_{2,t}(Y_t, W_t, \theta) > \varepsilon) \leqslant P_0(|1(U_t + \varepsilon_t) - 1(U_t) - \varepsilon_t \delta(U_t)|/|\varepsilon_t| > \varepsilon'/S) + P_0(S_t > S)$, so that by Chebyshev's inequality $P_0(r_{2,t}(Y_t, W_t, \theta) > \varepsilon) \leqslant P_0(|1(U_t + \varepsilon_t) - 1(U_t) - \varepsilon_t \delta(U_t)|/|\varepsilon_t| > \varepsilon'') + \eta/2$. Given $\varepsilon'' > 0$ and $\eta/3 > 0$, there exist some $e > 0$ such that $|\varepsilon_t| < e$ implies $P_0(|1(U_t + \varepsilon_t) - 1(U_t) - \varepsilon_t \delta(U_t)|/|\varepsilon_t| > \varepsilon'') < \eta/3$. By continuity of $q_t^\alpha(W_t, \cdot)$ $a.s.$ on $\Theta$, we know that for this $e > 0$, there exist some $\rho > 0$ such that $|\theta - \theta^*| < \rho$ implies $|\varepsilon_t| < e$, and therefore implies the previous inequality. Hence, for any $\eta > 0$ and any $\varepsilon > 0$ we have found $\rho > 0$ such that $|\theta - \theta^*| < \rho$ implies $P_0(r_{2,t}(Y_t, W_t, \theta) > \varepsilon) < \eta$, i.e. $P_0(\lim_{\theta\to\theta^*} r_{2,t}(Y_t,W_t,\theta)=0)=1$. Finally, for $r_{3,t}(Y_t,W_t,\theta)$ and $r_{4,t}(Y_t,W_t,\theta)$, note that by (A3) and the mean value expansion of $g_t$ around $\theta^*$, they are bounded above by $\delta(U_t)KS_t$, for every $t$, $1\leqslant t\leqslant T$. Same as previously, we show that for any $\theta\in\Theta$, $\eta>0$ and $\varepsilon>0$ we have $P_0(r_{3,t}(Y_t,W_t,\theta)>\varepsilon)\leqslant P_0(\delta(U_t)>\varepsilon'')+\eta/2$, so that by using $P_0(\delta(U_t)>\varepsilon'')=0$ we get $P_0(r_{3,t}(Y_t,W_t,\theta)>\varepsilon)<\eta$, i.e. $P_0(r_{3,t}(Y_t,W_t,\theta)=0)=1$. Same conclusion holds for $r_{4,t}(Y_t,W_t,\theta)$. Combining previous results yields $r_t(Y_t,W_t,\theta)\to 0$ as $\theta\to\theta^*$ with probability one, for every $t$, $1\leqslant t\leqslant T$. We now show that there exists some $\epsilon>0$ such that $E[\sup_{\theta\in\Theta:|\theta-\theta^*|<\epsilon} r_t(Y_t,W_t,\theta)]<\infty$, for all $t$, $1\leqslant t\leqslant T$. Using the same notation as above, we have, by (A3), (A3') and norm equivalence, that for every $\epsilon_1>0$ such that $\{\theta:|\theta-\theta^*|<\epsilon_1\}\subset\Theta$, $E[\sup_{\theta\in\Theta:|\theta-\theta^*|<\epsilon_1} r_{1,t}(Y_t,W_t,\theta)]\leqslant 2Lc^2E[\sup_{\theta\in\Theta}|\nabla_\theta q_t^\alpha(W_t,\theta)|^2]+2K\cdot E[\sup_{\theta\in\Theta}|\Delta_{\theta\theta}q_t^\alpha(W_t,\theta)|]$, where $c$ is some positive constant. (A5') then implies finiteness. For $r_{2,t}(Y_t,W_t,\theta)$, note that by continuity of $q_t^\alpha(W_t,\cdot)$, we can always choose an $\epsilon_2>0$ such that for every $\theta\in\{\theta:|\theta-\theta^*|<\epsilon_2\}\subset\Theta$, $q_t^\alpha(W_t,\theta^*)-Y_t$ and $q_t^\alpha(W_t,\theta)-Y_t$ are of same sign. In that case, using (A3), we have $\sup_{\theta\in\Theta:|\theta-\theta^*|<\epsilon_2} r_{2,t}(Y_t,W_t,\theta)\leqslant KS_t\delta(U_t)$, $a.s.-P_0$, since $U_t$ does not depend on $\theta$. Hence, $E[\sup_{\theta\in\Theta:|\theta-\theta^*|<\epsilon_2} r_{2,t}(Y_t,W_t,\theta)]\leqslant K\cdot E[S_tf_{0,t}(q_t^\alpha(W_t,\theta^*))]<\infty$, for all $t$, $1\leqslant t\leqslant T$, by using (A5') and (A6). Similarly, we show that there exist $\epsilon_3>0$ and $\epsilon_4>0$ such that the same inequality holds for $r_{3,t}(Y_t,W_t,\theta)$ and $r_{4,t}(Y_t,W_t,\theta)$. Letting $\epsilon\equiv\min\{\epsilon_i:1\leqslant i\leqslant 4\}>0$ gives the result. Note that equivalent reasoning applies if we replace $\theta^*$ with $\theta_0$. Combining above results with the established local convergence in probability of the sample mean of $\{\Delta_t(Y_t,W_t,\theta)\}$ for all $\theta\in\Theta$, then shows (by Theorem 7.3 in Newey and McFadden, 1994) that Lemma 9 holds. ■

**Proof of Theorem 6.** When $T=1$, the two optimization problems have been shown to be equivalent. Now consider $T>1$: given $(y_1,w_1,\ldots,y_T,w_T)'$, maximizing the tick-exponential log-likelihood is equivalent to minimizing $T^{-1}\sum_{t=1}^T\max\{-\ln\psi_t^\alpha(y_t,q_t^\alpha(w_t,\theta)),$ $-\ln\phi_t^\alpha(y_t,q_t^\alpha(w_t,\theta))\}$, where the functions $\psi_t^\alpha$ and $\phi_t^\alpha$, $1\leqslant t\leqslant T$, are as defined in (17). For $\theta$ and $T$ fixed, let $k_\theta$, $1\leqslant k_\theta\leqslant T$, denote the order such that $\varepsilon_{\theta,k_\theta}<0\leqslant\varepsilon_{\theta,k_\theta+1}$, and let $\varepsilon_{\theta,0}\equiv-\infty$ and $\varepsilon_{\theta,T+1}\equiv+\infty$. First, consider $t$ such that $1\leqslant t\leqslant k_\theta$: then $\varepsilon_{\theta,t}\leqslant\varepsilon_{\theta,k_\theta}<0$

so that $\max\{-\ln\psi_t^\alpha(y_{\theta,t},q_t^\alpha(w_{\theta,t},\theta)),-\ln\phi_t^\alpha(y_{\theta,t},q_t^\alpha(w_{\theta,t},\theta))\}=-\ln\phi_t^\alpha(y_{\theta,t},q_t^\alpha(w_{\theta,t},\theta))$, for $1\leqslant t\leqslant k_\theta$. Similarly, for $t$ such that $k_\theta+1\leqslant t\leqslant T$, we have $0\leqslant\varepsilon_{\theta,k_\theta+1}\leqslant\varepsilon_{\theta,t}$ and so $\max\{-\ln\psi_t^\alpha(y_{\theta,t},q_t^\alpha(w_{\theta,t},\theta)),-\ln\phi_t^\alpha(y_{\theta,t},q_t^\alpha(w_{\theta,t},\theta))\}=-\ln\psi_t^\alpha(y_{\theta,t},q_t^\alpha(w_{\theta,t},\theta))$, for $k_\theta+1\leqslant t\leqslant T$. Hence

$$\begin{aligned}
&T^{-1}\sum\nolimits_{t=1}^{T}\max\{-\ln\psi_t^\alpha(y_t,q_t^\alpha(w_t,\theta)),-\ln\phi_t^\alpha(y_t,q_t^\alpha(w_t,\theta))\} \\
=\;&T^{-1}[\sum\nolimits_{t=1}^{k_\theta}-\ln\phi_t^\alpha(y_{\theta,t},q_t^\alpha(w_{\theta,t},\theta))+\sum\nolimits_{t=k_\theta+1}^{T}-\ln\psi_t^\alpha(y_{\theta,t},q_t^\alpha(w_{\theta,t},\theta))],
\end{aligned}\tag{30}$$

where $\sum_t^s\equiv 0$ if $s<t$. First consider $k$ such that $k<k_\theta$:

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

Similarly we can show that the same result holds for all $k$ such that $k>k_\theta$. Hence, the right hand side of the above inequality is a maximum over $k$ of $P_k(y_\theta,w_\theta,\theta)$, which combined with (30) completes the proof of Theorem 6. ■