# Semiparametric Efficiency Bound and M-Estimation in Time-Series Models for Conditional Quantiles

**Ivana Komunjer and Quang Vuong**

*University of California, San Diego and Penn State University*

**Abstract:** In this paper we derive the semiparametric efficiency bound in time series models of conditional quantiles under a sole strong mixing assumption. We moreover provide an expression of Stein's (1956) least favorable parametric submodel. Our approach can be summarized as follows: first, we characterize the class of M–estimators that are consistent for the the conditional quantile parameter. We show that these estimators are asymptotically normal, and determine the minimum of their asymptotic covariance matrices. Second, we construct a fully parametric submodel that satisfies the conditional quantile restriction and contains the data generating process. Finally, we show that this submodel is the least favorable, i.e. the asymptotic covariance matrix of its maximum likelihood estimator is equal to the above minimum.

## 1. Introduction

In his seminal paper, Stein (1956) proposed a general method for obtaining semiparametric efficiency bounds that has been further studied by Begun, Hall, Huang, and Wellner (1983) and Bickel, Klaassen, Ritov, and Wellner (1993). In particular, the

**AMS 2000 subject classifications:** Primary 62G99, 62M10; secondary 62F12

**Keywords and phrases:** semiparametric efficiency, conditional quantiles, least favorable submodel, M-estimation

latter have developed an information calculus framework based on local asymptotic normality (LAN) and regularity of parameters of interest to derive Hajek-LeCam style convolution theorems for regular estimators. Most of the literature has focused on independent and identically distributed (iid) data as the study of sufficient conditions to satisfy the convolution theorem hypotheses has been largely successful in this setting. While dropping the iid assumption does not alter the validity of the convolution theorems, generalizing the sufficient conditions to satisfy them is still a challenge in time series contexts (Bickel and Kwon, 2001; Bickel, Klaassen, Ritov, and Wellner, 2005; Wellner, Klaassen, and Ritov, 2006).

This explains why the study of the non-iid case has first focused on time-series models in which some finitely parameterized transformation of the data leads back to the iid case. See e.g. Kreiss (1987) and Drost, Klaassen, and Werker (1997) who have derived general conditions under which LAN conditions hold (see also Hallin, Vermandeley, and Werker (2004)). Hence, the construction of semiparametric efficiency bounds is possible in such models, where the iid assumption of the innovation also allows for adaptive estimation in the sense of Bickel (1982). For example, this approach could be used to obtain the semiparametric efficiency bound in the following linear quantile regression model with ARCH errors (Koenker and Zhao, 1996):

$$Y_t = \theta' W_t + (1 + \gamma |U_{t-1}|) U_t, \tag{1}$$

where $W_t \equiv (1, Y_{t-1})'$, the process $\{(Y_t, W_t')'\}$ is strong mixing, the error sequence $\{U_t\}$ is independent of $\{W_t\}$ and iid with some absolutely continuous distribution function $H_0(\cdot)$ (continuous density $h_0(\cdot)$) such that $H_0^{-1}(\alpha) = 0$ for $\alpha \in (0,1)$. Here, $\theta$ is the parameter of interest and $\gamma$ and $h_0$ are two nuisance parameters where the latter is infinite dimensional.

More generally, consideration of information bounds and efficient estimation has been extended to nonparametric stationary Markov chains, where the parameter of interest satisfies a conditional moment restriction. See e.g. Greenwood, Müller, and Wefelmeyer (2004) for a current state of the art in this area. The time series model we study in this paper more closely falls in this category as the parameter of interest $\theta$ satisfies a conditional quantile restriction, though we do not restrict the data to be a stationary Markov Chain requiring it to be strong mixing only. In (1), this occurs when the sequence of errors $\{U_t\}$ is itself strong mixing. In this case, there exists no finitely parameterized transformation that would bring us back to the iid case. In other words, the time-dependence of the sequence $\{U_t\}$ introduces an additional nonparametric component that complicates the derivation of the semiparametric efficiency bound for $\theta$. In the words of Bickel, Klaassen, Ritov, and Wellner (2005), "none of these [LAN results] have been extended to honestly semiparametric contexts." Moreover, as the results by Pfanzagl (1976) and Fabian and Hannan (1982) indicate, adaptive estimation is not possible here thereby rendering nontrivial the derivation of the semiparametric efficiency bound.

In this paper, we use Stein's (1956) original definition and construct a "least favorable" submodel to derive the semiparametric efficiency bound in time-series models subject to a parametric conditional quantile restriction. To do so, we first characterize the class of M– (or extremal) estimators (Huber (1967)) that are consistent for the parameters $\theta$ of the conditional quantile. We show that these estimators are asymptotically normal under strong mixing, and we determine the minimum of their asymptotic covariance matrices. These results are of interest by themselves. In a second step we construct a fully parametric model that satisfies the conditional quantile restriction and contains the data generating process. We study the asymptotic properties of the maximum likelihood estimator (MLE) of such a parametric submodel.

In particular, we show that the asymptotic covariance matrix of this MLE, which is the inverse of the Fisher's information matrix, happens to be equal to the above minimum. Stein's (1956) argument then implies that the latter asymptotic covariance matrix is the semiparametric efficiency bound for $\theta$, while the constructed fully parametric submodel is a least favorable one.

As is known, least favorable models do not always exist. Examples when they do are the symmetric location model and the linear regression model where adaptive estimation is possible, and more recently, the normal bivariate copula model (Klaassen and Wellner (1997)), where adaptive estimation is not possible. When a least favorable model of same dimensionality as the parameter of interest is available, recent results by Murphy and der Vaart (2000) on profile likehoods (Severini and Wong (1992)) can be used to develop likelihood ratio type tests and confidence intervals in semiparametric models. Though our least favorable submodel has the same dimensionality as the parameter of interest, the absence of a convenient partitioned or variation-free parameterization $(\theta, \eta)$ with $\theta$ defined through a conditional moment restriction and $\eta$ infinitely dimensional complicates the application of Klaassen and Wellner (1997) results.

Of independent interest, Komunjer (2005) has characterized the class of quasi MLEs (QMLEs) that are consistent for the parameter $\theta$ in the conditional quantile restriction. Such a class is seen to be strictly included in the class of consistent M–estimators. In general, one can think of the class of GMM estimators as being the widest one. Then comes the class of M–estimators which can be viewed as just-identified GMM estimators. Finally comes the class of QMLEs which is the class of M–estimators whose objective functions satisfy an additional "integrability" condition and can thus be interpreted as quasi-likelihoods.

In models for conditional quantiles, efficient estimators are contained in the class of M–estimators. Hence, at least from a semiparametric efficiency viewpoint, no advantage is gained by considering GMM over M–estimators. Interestingly, our results show that the QMLE class is not sufficiently large to attain the semiparametric efficiency bound. On the other hand, as the latter is the minimum of the asymptotic covariance matrices of consistent M–estimators, a natural question is whether the bound can be attained. In Komunjer and Vuong (2007) under additional assumptions including smoothness ones, we propose a feasible M–type estimator that is semiparametrically efficient i.e. that achieves the semiparametric efficiency bound.

The remainder of the paper is as follows. In Section 2 we define our notation and introduce models for conditional quantiles. Section 3 characterizes the class of M–estimators that are consistent for the parameters of such models, provided they are correctly specified. In the same section we show that such estimators are also asymptotically normally distributed with an asymptotic covariance matrix whose expression depends on the form of the M–objective function being minimized. We then derive the minimum bound of the above family of matrices. Section 4 discusses Stein's (1956) definition of parametric submodels and proposes a parametric submodel of the conditional quantile semiparametric model. In the same section, we show that this model is the least favorable one thereby providing the semiparametric efficiency bound. We relegate all the proofs to the end of the paper.

## 2. Setup

***2.1. Notation and Definitions.*** Consider a stochastic sequence (a time series) $Z \equiv \{Z_t, t \in \mathbb{N}\}$ defined on a probability space $(\Omega, \mathcal{B}, P)$ where $Z : \Omega \to \mathbb{R}^{(n+1)\mathbb{N}}$ and $\mathbb{R}^{(n+1)\mathbb{N}}$ is the product space generated by taking a copy of $\mathbb{R}^{n+1}$ for each integer, i.e. $\mathbb{R}^{(n+1)\mathbb{N}} \equiv \times_{t=1}^{\infty}\mathbb{R}^{n+1}$, $n \in \mathbb{N}$. To simplify, we assume that for any $T \geqslant 1$, the joint distribution of $\underline{Z}_T \equiv (Z_1, \ldots, Z_T)$ has a positive continuous density $f_{\underline{Z}_T}(\cdot)$ on $\mathbb{R}^{(n+1)T}$, so that conditional densities are everywhere defined.

We partition the random vector $Z_t$ as $Z_t = (Y_t, X_t')'$ and are interested in the distribution of its endogenous (scalar) component, denoted $Y_t$, conditional on its own lags as well as the past and present values of the exogenous $n$-vector $X_t$. Specifically, we consider the family of subfields $\{\mathcal{W}_t, t \in \mathbb{N}\}$ where $\mathcal{W}_t \equiv \sigma(\underline{Z}_{t-1}, X_t)$ is the information set generated by the sequence of conditioning vectors up to time $t$. For every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, we let $F_t^0(\cdot)$ denote the conditional distribution function of $Y_t$ conditional upon $\mathcal{W}_t$, i.e. $F_t^0(y) \equiv P(Y_t \leqslant y_t|\mathcal{W}_t)$ for every $y \in \mathbb{R}$ with $f_t^0(\cdot)$ its conditional probability density. Letting $\underline{X}_T \equiv (X_1, \ldots, X_T)$, and using the fact that $X_t$'s are exogenous (so that their distribution is independent of the lagged values of $Y_t$), we then have that $f_{\underline{Z}_T}(\underline{z}_T) = [\prod_{t=1}^T f_t^0(y_t)] f_{\underline{X}_T}(\underline{x}_T)$, where lowercase letters are used to denote the realizations of the corresponding random variables.

Throughout the paper we assume that for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, $F_t^0(\cdot)$ is unknown and that it belongs to $\mathcal{F}$ which is the set of all absolutely continuous distribution functions on $\mathbb{R}$ with bounded density that is continuously differentiable, and whose derivative is bounded as well. In other words, there exist constants $M_0, M_1 > 0$ such that $\sup_{t\geqslant 1}\sup_{y\in\mathbb{R}} f_t^0(y) \leqslant M_0 < \infty$ and $\sup_{t\geqslant 1}\sup_{y\in\mathbb{R}} |df_t^0(y)/dy| \leqslant M_1 < \infty$.

The notation we use is standard: 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. $|M| \equiv \max_{1\leqslant i,j\leqslant n}|M_{ij}|$, and $M^+$ denotes a generalized inverse of $M$. If $A$ is a positive definite $n \times n$-matrix, then $A^{-1/2} = P$ where $P$ is invertible such that $PAP' = \mathrm{Id}$ where Id denotes the $n \times n$-identity matrix.

Let $f : E \rightarrow \mathbb{R}$, $V \mapsto f(V)$, with $E \subseteq \mathbb{R}^n$ and $V = (V_1, ..., V_n)'$, be continuously differentiable to order $R \geqslant 1$ on $E$. Let $r \equiv (r_1, ..., r_n) \in \mathbb{N}^n$: if $|r| \leqslant R$ then $D^r f(V) \equiv \partial^{|r|} f(V)/\partial V_1^{r_1} ... \partial V_n^{r_n}$ where $|r| \equiv r_1 + ... + r_n$ represents the order of derivation. If $r = 0$ then $D^0 f(V) = f(V)$. Further, let $r! \equiv r_1!...r_n!$ and $V^r \equiv V_1^{r_1}...V_n^{r_n}$. Then, for any $(V, V_0) \in E^2$ the (familiar) expression in a Taylor expansion of order $R$ can be written as $\sum_{|r| \leqslant R} \frac{D^r f(V_0)}{r!}(V - V_0)^r \equiv \sum_{k=0}^{R} \sum_{j_1,...,j_k \in (1,...,n)^k} \frac{1}{k!} \frac{\partial^k f(V_0)}{\partial V_{j_1} ... \partial V_{j_k}} (V_{j_1} - V_{0j_1}) ... (V_{j_k} - V_{0j_k})$, for $1 \leqslant l \leqslant R$. For example, when $R = 1$, we have $\sum_{|r| \leqslant 1} D^r f(V_0)(V - V_0)^r = f(V_0) + \sum_{i=1}^{n} [\partial f(V_0)/\partial V_i](V_i - V_{0i})$ (Schwartz, 1997). When $R \geqslant 2$, we let $\nabla_V f(V)$ denote the gradient of $f$, $\nabla_V f(V) \equiv (\partial f(V)/\partial V_i, ..., \partial f(V)/\partial V_n)'$, and use $\Delta_{VV} f(V)$ to denote its Hessian matrix, $\Delta_{VV} f(V) \equiv (\partial^2 f(V)/\partial V_i \partial V_j)_{1 \leqslant i,j \leqslant n}$.

Finally, the function $\mathbb{1} : \mathbb{R} \rightarrow [0, 1]$ denotes the Heaviside (or indicator) function: for any $x \in \mathbb{R}$, we have $\mathbb{1}(x) = 0$ if $x \leqslant 0$, and $\mathbb{1}(x) = 1$ if $x > 0$ (Bracewell, 2000). The Heaviside function is the indefinite integral of the Dirac delta function $\delta : \mathbb{R} \rightarrow \mathbb{R}$, with $\mathbb{1}(x) = \int_a^x d\delta$, where $a$ is an arbitrary (possibly infinite) negative constant, $a \leqslant 0$.

***2.2. Models for conditional quantiles.*** In this paper we do not consider the conditional distribution $F_t^0(\cdot)$ in its entirety but rather focus on a particular conditional quantile of $Y_t$. For a given probability $\alpha \in (0, 1)$, we denote by $\mathcal{M}$ a model for the conditional $\alpha$-quantile of $Y_t$, $\mathcal{M} \equiv \{q_\alpha(W_t, \theta)\}$, parametrized by $\theta$ in $\Theta$, where $\Theta$ is a compact subset of $\mathbb{R}^k$ with non-empty interior, $\mathring{\Theta} \neq \emptyset$, and in which $W_t$ is a random $m$-vector, $W_t \in \mathbb{R}^m$, that is $\mathcal{W}_t$-measurable. In what follows, we restrict our attention to conditional quantile models $\mathcal{M}$ in which the set of following conditions is satisfied:

**(A1)** *(i) the model $\mathcal{M}$ is identified on $\Theta$, i.e. for any $(\theta_1, \theta_2) \in \Theta^2$ we have: $q_\alpha(W_t, \theta_1)$ $= q_\alpha(W_t, \theta_2)$, a.s. $- P$, for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, if and only if $\theta_1 = \theta_2$; (ii)* *for every* $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, *the function* $q_\alpha(W_t, \cdot) : \Theta \to \mathbb{R}$ *is twice continuously differentiable on* $\Theta$ *a.s.* $-P$; *(iii) for every* $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, *the matrix* $\nabla_\theta q_\alpha(W_t, \theta) \nabla_\theta q_\alpha(W_t, \theta)'$ *is of full rank a.s.* $-P$ *for every* $\theta \in \Theta$.

The set of conditions in (A1) is fairly standard and generally verified for a wide variety of conditional quantile models. In what follows, for any given $\mathcal{M}$ we shall denote by $\mathcal{Q}$ the range of $q_\alpha$, i.e. $\mathcal{Q} \equiv \{q \in \mathbb{R} : q = q_\alpha(w, \theta), \theta \in \Theta, w \in \mathbb{R}^m\}$, $\mathcal{Q} \subseteq \mathbb{R}$.

One crucial assumption that we make in our analysis, and which is of different nature than the conditions above, is that the model $\mathcal{M}$ is correctly specified, so that there exists some true parameter value $\theta_0$ such that $F_t^0(q_\alpha(W_t, \theta_0)) = \alpha$, for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$. In other words, we assume the following:

**(A2)** *given* $\alpha \in (0,1)$, *there exists* $\theta_0 \in \mathring{\Theta}$ *such that* $E[\mathbb{1}(q_\alpha(W_t, \theta_0) - Y_t) | \mathcal{W}_t] = \alpha, a.s. - P$, *for every* $t$, $1 \leqslant t \leqslant T, T \geqslant 1$.

In other words, for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, the difference between the indicator variable above and $\alpha$ is assumed to be orthogonal to any $\mathcal{W}_t$-measurable random variable.

## 3. Semiparametric estimators for conditional quantiles

In this paper we consider a particular family of semiparametric conditional quantile estimators known as M– (or extremal) estimators (Huber, 1967). M–estimators for $\theta_0$, denoted $\theta_T$, are obtained by minimizing criterion functions of the form $\Psi_T(\theta) \equiv T^{-1} \sum_{t=1}^T \varphi(Y_t, q_\alpha(W_t, \theta), \xi_t)$ where for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, $\varphi$ is a real function of the variable of interest $Y_t$, the quantile $q_\alpha(W_t, \theta)$ and a (possibly inifinite-dimensional) random variable $\xi_t : \Omega \to E_t$, i.e. $\varphi : \mathbb{R} \times \mathcal{Q} \times E_t \to \mathbb{R}$. The variable $\xi_t$ can be thought of as a shape parameter of the objective function $\varphi$. We assume the following:

**(A3)** *(i) for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, $\xi_t$ is $\mathcal{W}_t$-measurable; (ii) for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, the function $\varphi(\cdot, \cdot, \cdot)$ is twice continuously differentiable a.s. $- P$ on $\mathbb{R} \times \mathcal{Q} \times E_t$ with respect to its second argument ($q$).*

By assumption (A3)(i), the random variable $\xi_t$ is allowed to depend only on variables contained in $\mathcal{W}_t$, i.e. $\xi_t$ is predictable (see e.g. Greenwood, Müller, and Wefelmeyer (2004)). In particular, if we consider objective functions $\varphi$ that depend on some estimator based on the observations of $Y_t$ and $W_t$ up to time $T$—kernel estimators of conditional distributions or densities are examples—then (A3)(i) fails to hold. The requirement (A3)(ii) allows for objective functions such as $|y - q|$ or $[\alpha - \mathbb{1}(q - y)](y - q)$, for example. Note that in those two cases the shape $\xi_t$ of $\varphi$ remains constant over time.

An important subfamily of the class of M–estimators defined above, is that of quasi-maximum likelihood estimators (QMLEs) (White, 1982; Gourieroux, Monfort, and Trognon, 1984; Komunjer, 2005). If in addition to (A3), we assume that there exists a real function $c : \mathbb{R} \times E_t \to \mathbb{R}$, $(y, \xi_t) \mapsto c(y, \xi_t) < \infty$, independent of $q$, and such that $\int_{\mathbb{R}} \exp[c(y, \xi_t) - \varphi(y, q, \xi_t)]dy = 1$ for all $(q, \xi_t) \in \mathcal{Q} \times E_t$, then we can let $\tilde{f}_t^0(Y_t, \theta) \equiv \exp[c(Y_t, \xi_t) - \varphi(Y_t, q_\alpha(W_t, \theta), \xi_t)]$, and $\tilde{f}_t^0(\cdot, \theta)$ can be interpreted as a (pseudo) density of $Y_t$ conditional on $\mathcal{W}_t$. Hence, any minimum $\theta_T$ of the function $\Psi_T(\theta)$ above, is also a maximum of the (quasi) log-likelihood function $L_T(\theta)$. Indeed, we then have $L_T(\theta) = [\sum_{t=1}^T \tilde{f}_t^0(Y_t, \theta)] + \ln f_{\underline{X}_T}(\underline{X}_T)$, and only the first term is a function of $\theta$. However, due to the above "integrability" constraint—that $\int_{\mathbb{R}} \exp[c(y, \xi_t) - \varphi(y, q, \xi_t)]dy = 1$ for all $(q, \xi_t) \in \mathcal{Q} \times E_t$—the class of QMLEs is included in that of M–estimators. We shall come back to this issue in our next section in which we focus on M–estimators for $\theta_0$ that are consistent.

***3.1. Class of consistent M–estimators.*** What are necessary conditions for the M–estimator $\theta_T$ satisfying (A3), to be consistent for the true conditional quantile parameter $\theta_0$ in (A2)? The key idea behind the answer to this question is fairly simple. Assume that the process $\{Z_t\}$ and the functions $\varphi(\cdot,\cdot,\xi_t)$ are such that $\theta_T - \theta_T^0 \xrightarrow{p} 0$, where $\theta_T^0$ is a unique minimum of $E[\Psi_T(\theta)] \equiv T^{-1}\sum_{t=1}^{T} E[\varphi(Y_t, q_\alpha(W_t,\theta),\xi_t)]$ on $\mathring{\Theta}$. Note that $\theta_T^0$ is also called the pseudo-true value of the parameter $\theta$. Then a necessary requirement for consistency of $\theta_T$ is that $\theta_T^0 - \theta_0 \to 0$ as $T$ becomes large. In what follows, we restrict our attention to estimators $\theta_T$ such that $\theta_T^0$ remains constant, i.e. $\forall T \geqslant 1$ we have $\theta_T^0 = \theta_\infty^0$. Then, the class of M–estimators that are consistent for $\theta_0$ is obtained by considering all the functions $\varphi(\cdot,\cdot,\xi_t)$ under which $\theta_\infty^0 = \theta_0$.

Note that the requirement of having $\theta_T^0 = \theta_0$ for all $T \geqslant 1$ is stronger than that of having $\theta_T^0 \to \theta_0$ (see White (1994, p.69-70) for a discussion). This implies that $\theta_0$ can be consistently estimated by minimizing objective functions that are different from the ones derived below, as long as the expected value of this difference converges uniformly to zero with $T$. An important example in which the condition $\theta_T^0 = \theta_0$ for all $T \geqslant 1$ fails is when the shape $\xi_t$ of the objective function $\varphi$ depends on observations up to time $T$ and is therefore not $\mathcal{W}_t$-measurable.

We now provide a more formal treatment of consistency. A set of sufficient assumptions for $\theta_T - \theta_\infty^0 \xrightarrow{p} 0$ is (see Theorem 2.1 in Newey and McFadden, 1994):

**(A4)** $\{Z_t\}$ *and* $\varphi(\cdot,\cdot,\xi_t)$ *are such that: (i) for every* $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, *and every* $\theta \in \Theta$, $|D^r\varphi(Y_t, q_\alpha(W_t,\theta),\xi_t)| \leqslant m_r(Y_t, W_t, \xi_t)$, *a.s.* $-$ $P$, *where* $E[m_r(Y_t, W_t, \xi_t)] < \infty$, *for* $r = 0, 1, 2$; *for any* $T \geqslant 1$, *(ii)* $E[\Psi_T(\theta)]$ *is uniquely minimized at* $\theta_\infty^0 \in \mathring{\Theta}$, *and (iii)* $\sup_{\theta\in\Theta} |\Psi_T(\theta) - E[\Psi_T(\theta)]| \xrightarrow{p} 0$.

Note that the above are not primitive conditions for consistency of $\theta_T$. For example, the dominance conditions in (A4)(i) are typically implied by more primitive assumptions on the existence of different moments of $Y_t$, $W_t$ and $\xi_t$. Condition (A4)(ii) states that $\theta^0_\infty$ is a minimum of $E[\Psi_T(\theta)]$ and that this minimum is moreover unique. The first requirement involves more primitive conditions on $\partial\varphi/\partial q$, $\partial^2\varphi/\partial q^2$ and $\nabla_\theta q_\alpha$, which depend on the shape $\xi_t$ of $\varphi$ and the functional form of $q_\alpha$. For example, a sufficient set of conditions for $\theta^0_\infty$ to be a minimum is that $T^{-1}\sum_{t=1}^T E[\nabla_\theta\varphi(Y_t, q_\alpha(W_t, \theta^0_\infty), \xi_t)] = 0$ and $T^{-1}\sum_{t=1}^T E[\Delta_{\theta\theta}\varphi(Y_t, q_\alpha(W_t, \theta^0_\infty), \xi_t)] \gg 0$. Finally, the uniform convergence condition (A4)(iii) can be obtained by applying an appropriate uniform law of large numbers to the sequence $\{\varphi(Y_t, q_\alpha(W_t, \theta), \xi_t)\}$. Implicit in (A4)(iii) are primitive assumptions on the dependence structure and heterogeneity of the process $\{Z_t\}$, and on the properties of $\varphi(Y_t, q_\alpha(W_t, \cdot), \xi_t)$. A simple example is one where $\{Z_t\}$ is iid and the functions $\varphi(Y_t, q_\alpha(W_t, \cdot), \xi_t)$ are Lipshitz-$\mathrm{L}_1$ $a.s. - P$ on $\Theta$.

The above pseudo-true value $\theta^0_\infty$ of the parameter $\theta$ equals the true value $\theta_0$ if and only if, for any $T \geqslant 1$, $\theta_0$ minimizes $E[\Psi_T(\theta)]$. A necessary and sufficient requirement for $\theta^0_\infty = \theta_0$ is given in the following theorem.

**Theorem 1 (Necessary and sufficient condition for consistency).** *Assume that (A1), (A3) and (A4) hold. If the true parameter $\theta_0$ satisfies the conditional moment condition in (A2), then the M-estimator $\theta_T$ is consistent for $\theta_0$, i.e. $\theta_T - \theta_0 \xrightarrow{p} 0$, if and only if there exist a real function $A(\cdot,\cdot) : \mathbb{R} \times E_t \rightarrow \mathbb{R}$ that is twice continuously differentiable and strictly increasing with respect to its first argument ($q$ or $y$) $a.s. - P$ on $\mathcal{Q} \times E_t$, and a real function $B(\cdot,\cdot) : \mathbb{R} \times E_t \rightarrow \mathbb{R}$, such that $\varphi(y, q, \xi_t) = [\alpha - 1\!\!I(q - y)][A(y, \xi_t) - A(q, \xi_t)] + B(y, \xi_t)$, $a.s. - P$ on $\mathbb{R} \times \mathcal{Q} \times E_t$, for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$.*

In other words, if for any given sample size $T \geqslant 1$ we are interested in consistently estimating the conditional quantile parameter of a continuously distributed random variable $Y_t$ by using an M–estimator $\theta_T$, then we must employ an objective function

$\Psi_T(\cdot) = T^{-1}\sum_{t=1}^{T}\varphi(Y_t, q_\alpha(W_t,\cdot), \xi_t)$ with

$$\begin{aligned}
&\varphi(Y_t, q_\alpha(W_t,\theta), \xi_t) \\
&\quad = [\alpha - \mathbb{I}(q_\alpha(W_t,\theta) - Y_t)][A(Y_t,\xi_t) - A(q_\alpha(W_t,\theta),\xi_t)] + B(Y_t,\xi_t), \qquad (2)
\end{aligned}$$

$a.s.{-}P$, for every $t$, $1 \leqslant t \leqslant T$. Using objective functions of this form is also a sufficient condition for $\theta_T$ to be consistent for the true parameter $\theta_0$ of a correctly specified model for the conditional $\alpha$-quantile. Note that the real functions $A(\cdot,\cdot)$ and $B(\cdot,\cdot)$ in Theorem 1 need not have the same shape parameter: we can let $\xi_t \equiv (\xi'_{At}, \xi'_{Bt})'$ where $\xi_{At}$ and $\xi_{Bt}$ are the shapes of $A(\cdot,\xi_{At})$ and $B(\cdot,\xi_{Bt})$, respectively. For simplicity, we write $A(\cdot,\xi_t)$ and $B(\cdot,\xi_t)$ with the understanding that changing the shape of $A(\cdot,\cdot)$ may not affect the shape of $B(\cdot,\cdot)$ and vice-versa.

Given that we restrict our attention to objective functions in which (A3)(ii) holds, the function $A(\cdot,\xi_t)$ in Theorem 1 needs to be twice continuously differentiable $a.s.{-}P$ on $\mathcal{Q}$. The continuity and differentiability of $A(\cdot,\xi_t)$ need not hold on $\mathbb{R}\backslash\mathcal{Q}$. The fact that there are no requirements on $A(\cdot,\xi_t)$ outside the range of $q_\alpha(W_t,\theta)$ is not surprising, given that changing the objective function outside $\mathcal{Q}$ does not affect the values of $\partial\varphi/\partial q$, and therefore has no effect on the optimum of $\Psi_T$. The fact that $A(\cdot,\xi_t)$ is necessarily strictly increasing $a.s. - P$ on $\mathcal{Q}$, comes from the requirement (A4)(ii) that $\theta^0_\infty$ be an interior minimum of $E[\Psi_T(\theta)]$ on $\Theta$. As previously, there are no requirements on the monotonicity of $A(\cdot,\xi_t)$ on $\mathbb{R}\backslash\mathcal{Q}$. Finally, note that there are no restrictions on the function $B(\cdot,\xi_t)$, as expected, since changing it does not affect the optimum of the objective function $\Psi_T$. In what follows we set $B(\cdot,\xi_t)$ identically equal to 0, which does not affect any of our results but has the benefit of simplifying the notation.

Well-known examples of conditional quantile estimators that satisfy Theorem 1 are: (1) Koenker and Bassett's (1978) unweighted quantile regression estimator for which $A(y, \xi_t) = y$, for all $y \in \mathbb{R}$; (2) Powell's (1984, 1986) left (right) censored quantile regression estimator obtained when, for all $y \in \mathbb{R}$, $A(y, \xi_t) = \max\{y, c_t\}$ ($A(y, \xi_t) = \min\{y, c_t\}$) with an observed censoring point $c_t$; (3) weighted quantile regression estimator, proposed by Newey and Powell (1990) and Zhao (2001), in which for all $y \in \mathbb{R}$, $A(y, \xi_t) = \omega_t y$ where $\omega_t$ is some nonnegative weight, as well as its censored version for which $A(y, \xi_t) = \omega_t \max\{y, c_t\}$. Note that $A(\cdot, \xi_t) = \max\{\cdot, c_t\}$ satisfies the strict monotonicity requirement $a.s. - P$ on $\mathcal{Q}$ because, in the censored quantile regression case, $q_\alpha(W_t, \theta_0) \geqslant c_t, a.s. - P$, as discussed by Powell (1984, p 4-6). The intuition behind this inequality is simple: suppose $Y_t = c_t, a.s. - P$ for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$. Then any value $\theta_0$ for which $q_\alpha(W_t, \theta_0) \leqslant c_t, a.s. - P$ for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, is a minimum of $E[\Psi_T(\theta)]$, which in that case equals 0. This violates the uniqueness assumption (A4)(ii), and hence affects the consistency of $\theta_T$. The latter is restored by requiring that $q_\alpha(W_t, \theta_0) \geqslant c_t, a.s. - P$ for a large enough portion of the sample (see Assumption R.1 in Powell, 1984). An analogous result holds for the right censored case.

We now show that the class of objective functions $\Psi_T(\theta)$ leading to consistent conditional quantile M–estimators is (strictly) larger than that leading to consistent QMLEs. To simplify the comparison, assume that $\mathcal{Q} = \mathbb{R}$. As pointed out previously, the main difference between the two classes of estimators lies in the "integrability" condition on the (pseudo) conditional densities $\tilde{f}_t^0(\cdot, \theta)$ of $Y_t$ conditional on $\mathcal{W}_t$. Compare the objective function in Theorem 1 with the family of tick-exponential (pseudo) densities which give consistent QMLEs for $\theta_0$ (Komunjer, 2005): $\bar{f}_t^0(Y_t, \theta) \equiv \alpha(1-\alpha)a(Y_t, \xi_t)\exp\{[\mathbb{1}(q_\alpha(W_t, \theta) - Y_t) - \alpha][A(Y_t, \xi_t) - A(q_\alpha(W_t, \theta), \xi_t)]\}$ with $A(\cdot, \xi_t)$ twice continuously differentiable and strictly increasing $a.s. - P$ on $\mathbb{R}$, with derivative $a(y, \xi_t) \equiv \partial A(y, \xi_t)/\partial y$.

For $\bar{f}_t^0(\cdot,\theta)$ to be a probability density on $\mathbb{R}$, we need $\lim_{y\to\pm\infty} A(y,\xi_t) = \pm\infty$, for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$. These limit conditions restrict the possible choice of functions $A(\cdot,\xi_t)$ in Theorem 1. They follow directly from the quantile restriction $\int_{-\infty}^{q_\alpha(W_t,\theta)} \bar{f}_t^0(y,\theta)dy = \alpha$, which is equivalent to $(1-\alpha)\exp[-(1-\alpha)A(q_\alpha(W_t,\theta),\xi_t)]\times \int_{-\infty}^{q_\alpha(W_t,\theta)} a(y,\xi_t)\exp[(1-\alpha)A(y,\xi_t)]dy = 1$, so that, upon the change of variable $u \equiv A(y,\xi_t)$, necessarily $A(q,\xi_t) \to -\infty$ as $q \to -\infty$. Combining the above quantile restriction with the condition $\int_{\mathbb{R}} \bar{f}_t^0(y,\theta)dy = 1$ yields the result for the limit in $+\infty$ by a similar reasoning.

For example, consider any distribution function $F_t(\cdot)$ in $\mathcal{F}$ having a density $f_t(\cdot)$ that is continuously differentiable $a.s. - P$, and let

$$A(y,\xi_t^F) \equiv F_t(y), \tag{3}$$

for any $y \in \mathbb{R}$. Note that the parameter $\xi_t^F$ in the objective function $A(\cdot,\xi_t^F)$ in Equation (3) corresponds to the conditional distribution $F_t(\cdot)$ which is stochastic and $\mathcal{W}_t$-measurable. Under the assumptions of Theorem 1, the M–estimator $\theta_T^F$, which minimizes $\Psi_T^F(\theta) \equiv T^{-1}\sum_{t=1}^T \varphi(Y_t, q_\alpha(W_t,\theta), \xi_t^F)$ with

$$\varphi(Y_t, q_\alpha(W_t,\theta), \xi_t^F) \equiv [\alpha - \mathbb{1}(q_\alpha(W_t,\theta) - Y_t)][F_t(Y_t) - F_t(q_\alpha(W_t,\theta))], \tag{4}$$

is consistent for $\theta_0$. However, the corresponding function $A(\cdot,\xi_t^F)$ in (3), bounded between 0 and 1, does not satisfy $\lim_{y\to\pm\infty} A(y,\xi_t) = \pm\infty$. As a consequence, the class of consistent QMLEs is strictly smaller than that of consistent M–estimators.

To resume, we have shown that an M–estimator $\theta_T$ that satisfies (A3) is consistent for $\theta_0$ if only if the objective functions $\varphi(\cdot,\cdot,\xi_t)$ are of the form given in Theorem 1. From the functional form of $\varphi(\cdot,\cdot,\xi_t)$ in (2), it follows that the asymptotic properties of $\theta_T$ only depend on the choice of $A(\cdot,\xi_t)$ since changing $B(\cdot,\xi_t)$ does not affect the minimum of $\Psi_T(\theta)$. Before considering a particular class of functions $A(\cdot,\xi_t)$, which makes the asymptotics of $\theta_T$ optimal, we need the asymptotic distribution of the latter. We derive the asymptotic distribution of $\theta_T$ in the next section.

***3.2. Asymptotic Distribution.*** We start by imposing the following assumptions, in addition to (A1)-(A3):

**(A5)** *for every* $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, *the functions* $A(\cdot, \xi_t) : \mathbb{R} \to \mathbb{R}$ *in Theorem 1 have bounded first and second derivatives, i.e. there exist constants* $K > 0$ *and* $L \geqslant 0$ *such that* $0 < \partial A(q, \xi_t)/\partial q \leqslant K$ *and* $|\partial^2 A(q, \xi_t)/\partial q^2| \leqslant L$, *a.s.* $- P$ *on* $\mathcal{Q} \times E_t$*;*

**(A6)** $\theta_0$ *is an interior point of* $\Theta$*;*

**(A7)** *the sequence* $\{(Y_t, W_t')'\}$ *is* $\alpha$*-mixing with* $\alpha$ *of size* $-r/(r-2)$, *with* $r > 2$*;*

**(A8)** *for some* $\epsilon > 0$ *(i)* $\sup_{1 \leqslant t \leqslant T, T \geqslant 1} E[\sup_{\theta \in \Theta} |\nabla_\theta q_\alpha(W_t, \theta)|^{2(r+\epsilon)}] < \infty$, $\sup_{1 \leqslant t \leqslant T, T \geqslant 1} E[\sup_{\theta \in \Theta} |\Delta_{\theta\theta} q_\alpha(W_t, \theta)|^{r+\epsilon}] < \infty$*; (ii)* $\sup_{1 \leqslant t \leqslant T, T \geqslant 1} E[\sup_{\theta \in \Theta} |A(q_\alpha(W_t, \theta), \xi_t)|^{r+\epsilon}] < \infty$, *and* $\sup_{1 \leqslant t \leqslant T, T \geqslant 1} E[|A(Y_t, \xi_t)|^{r+\epsilon}] < \infty$.

In addition to (A2) and (A3), we now require the functions $A(\cdot, \xi_t)$ to have bounded first and second derivatives (A5). The boundedness property is used to show that $\varphi(Y_t, q_\alpha(W_t, \cdot), \xi_t)$ are Lipshitz-$L_1$ on $\Theta$ *a.s.* $- P$. This implies that any pointwise convergence in $\theta$ becomes uniform on $\Theta$. Note that we can obtain a similar implication by an alternative argument, if the objective functions $\varphi(Y_t, q_\alpha(W_t, \cdot), \xi_t)$ are convex in the parameter $\theta$.

The convexity approach has been elegantly used to derive asymptotic normality of the standard Koenker and Bassett's (1978) quantile regression estimator by Pollard (1991), Hjort and Pollard (1993) and Knight (1998), for example. In the case of this estimator, the functions $A(\cdot, \xi_t)$ are linear and hence $\varphi(Y_t, q_\alpha(W_t, \cdot), \xi_t)$'s are convex in $\theta$, no matter which conditional quantile model $q_\alpha$ in (A1) we choose. Recall that $\varphi(Y_t, q_\alpha(W_t, \cdot), \xi_t)$ is convex in a neighborhood of $\theta_0$ if and only if the real function

$s \longmapsto [\varphi(Y_t, q_\alpha(W_t, \theta_0 + \nu s), \xi_t) - \varphi(Y_t, q_\alpha(W_t, \theta_0), \xi_t)]/s$ is increasing in $s \in \mathbb{R}$ ($\nu \in \mathbb{R}^k$). This condition holds for any conditional quantile model $\mathcal{M}$ in (A1), only if the functions $A(\cdot, \xi_t)$ have zero convexity, i.e. are linear.

Unfortunately, the convexity in $\theta$ of the objective functions $\varphi(Y_t, q_\alpha(W_t, \cdot), \xi_t)$ does not hold for general (nonlinear) $A(\cdot, \xi_t)$'s, such as the ones in (4). Therefore, we cannot rely on the convexity argument in our asymptotic normality proof. We are forced to abide by the classical approach which, though generally applicable, has the disadvantage of being more complicated and requires stronger regularity conditions, such as those in (A5).

Our assumptions on the heterogeneity and dependence structure of the data are, on the other hand, fairly weak. We allow the sequence $\{(Y_t, W_t')'\}$ to be nonstationary and our strong mixing (i.e. $\alpha$-mixing) assumption in (A7) allows for a wide variety of dependence structures (White, 2001). Assumption (A7) is further accompanied by a series of moment conditions in (A8) which guarantee that the appropriate law of large numbers and central limit theorem can be applied.

In the special case corresponding to Koenker and Bassett's (1978) quantile regression estimator for linear models $q_\alpha(W_t, \theta) = \theta' W_t$, the set of moment conditions (A8) reduces to: $\sup_{1 \leqslant t \leqslant T, T \geqslant 1} E[|W_t|^{2(r+\epsilon)}] < \infty$ and $\sup_{1 \leqslant t \leqslant T, T \geqslant 1} E[|Y_t|^{r+\epsilon}] < \infty$.

The asymptotic distribution of $\theta_T$ is given in the following theorem.

**Theorem 2** (**Asymptotic Distribution).** *Under (A1)-(A3) and (A5)-(A8), we have* $(\Sigma_T^0)^{-1/2} \Delta_T^0 \sqrt{T}(\theta_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$, *where* $\Delta_T^0 \equiv \frac{1}{T} \sum_{t=1}^T E[a(q_\alpha(W_t, \theta_0), \xi_t) \times f_t^0(q_\alpha(W_t, \theta_0)) \nabla_\theta q_\alpha(W_t, \theta_0) \nabla_\theta q_\alpha(W_t, \theta_0)']$, $\Sigma_T^0 \equiv \frac{1}{T} \sum_{t=1}^T \alpha(1-\alpha) E[a(q_\alpha(W_t, \theta_0), \xi_t)^2 \times \nabla_\theta q_\alpha(W_t, \theta_0) \nabla_\theta q_\alpha(W_t, \theta_0)']$, *and* $a(q_t, \xi_t) \equiv \partial A(q_t, \xi_t)/\partial q_t$ *a.s.* $- P$ *on* $\mathcal{Q} \times E_t$.

Theorem 2 shows that (A1)-(A3) and (A5)-(A8) are sufficient for the M–estimator $\theta_T$ to be consistent and asymptotically normally distributed. Note that these assumptions are primitive, unlike (A4). The asymptotic variance of $\theta_T$ has the usual "sandwich" form $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1}$: $\Delta_T^0$ is the Hessian matrix of the M–objective function $\Psi_T(\theta)$, i.e. $\Delta_T^0 = \frac{1}{T}\sum_{t=1}^T E[\Delta_{\theta\theta}\varphi(Y_t, q_\alpha(W_t,\theta_0),\xi_t)]$ with $\varphi(\cdot,\cdot,\cdot)$ as in (2); $\Sigma_T^0$ is the outer product of the gradient of $\Psi_T(\theta)$, i.e. $\Sigma_T^0 = E[\nabla_\theta\Psi_T(\theta)\nabla_\theta\Psi_T(\theta)']$. The expression for $\Sigma_T^0$ in Theorem 2 then follows from the fact that $\{\nabla_\theta\varphi(Y_t, q_\alpha(W_t,\theta_0),\xi_t),\mathcal{W}_t\}$ is a martingale difference sequence, which is itself implied by the correct specification assumption (A2) and the $\mathcal{W}_t$-measurability of the shape $\xi_t$ of $A(\cdot,\cdot)$ in (A3)(i).

In particular, the M–estimator $\theta_T^F$ proposed in (4) satisfies the conditions of Theorem 2, provided the conditional probability densities $f_t(\cdot)$ are differentiable $a.s.-P$ on $\mathbb{R}$ with bounded first derivatives, so that $|f_t'(y)| \leqslant L, a.s.-P$ on $\mathbb{R}$. Moreover, the moment conditions in (A8) are less stringent for $\theta_T^F$ than for Koenker and Bassett's (1978) estimator: they reduce to $E[|W_t|^{2(r+\epsilon)}] < \infty$, if the conditional quantile model is linear, for example. The fact that the moment conditions imposed on $Y_t$ disappear in the case of $\theta_T^F$ is simply due to the fact that—any conditional distribution function $F_t(\cdot)$ being bounded between 0 and 1—we always have $E[\sup_{\theta\in\Theta}|F_t(q_\alpha(W_t,\theta))|^{r+\epsilon}] \leqslant 1$ and $E[|F_t(Y_t)|^{r+\epsilon}] \leqslant 1$ so that (A8)(ii) is automatically satisfied. This difference is of particular importance in applications in which we have reason to believe that higher order moments of $Y_t$—order higher than 2—do not exist. In such applications, it is unclear what the asymptotic properties of Koenker and Bassett's (1978) estimator are. On the other hand, $\theta_T^F$ still converges in distribution at the usual $\sqrt{T}$ rate.

***3.3. Minimum Asymptotic Variance.*** Using Theorem 2, we can now rank all the consistent and asymptotically normal estimators constructed in the previous section by their asymptotic variances. Note that this ranking is useful, as we do not allow M–estimators to be superefficient. Superefficiency is ruled out by our continuity assumptions on $f_t^0(\cdot)$, $q_\alpha(W_t,\cdot)$ in (A1)(ii) and $a(\cdot,\xi_t)$ in Theorem 1. Typically, the asymptotic distribution of superefficient estimators is discontinuous in the true parameters, and our continuity assumptions rule out this discontinuity.

Using Theorem 2, we have $(\Sigma_T^{0,F})^{-1/2}\Delta_T^{0,F}\sqrt{T}(\theta_T^F - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$, with $\Delta_T^{0,F} \equiv T^{-1}\sum_{t=1}^T E[f_t(q_\alpha(W_t,\theta_0))f_t^0(q_\alpha(W_t,\theta_0))\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)']$, and $\Sigma_T^{0,F} \equiv T^{-1} \sum_{t=1}^T \alpha(1-\alpha)E[f_t(q_\alpha(W_t,\theta_0))^2\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)']$. Clearly, changing the distribution function $F_t(\cdot)$ in (3)—hence in (4)—affects the asymptotic covariance matrix of the corresponding M–estimator $\theta_T^F$, through the density term $f_t(\cdot)$ appearing in the expressions of $\Delta_T^{0,F}$ and $\Sigma_T^{0,F}$. In particular, this result suggests that appropriate choices of $F_t(\cdot)$ in (4) lead to efficiency improvements over Koenker and Bassett's (1978) conditional quantile estimator.

Specifically, when $f_t(\cdot)$ and the true conditional density $f_t^0(\cdot)$ coincide at the true quantile $q_\alpha(W_t,\theta_0)$, we have $\Sigma_T^{0,F}(\Delta_T^{0,F})^{-1} = \alpha(1-\alpha)\,\mathrm{Id}$. In other words, this particular choice of $f_t(\cdot)$ seems to lead to a conditional quantile M–estimator with the minimum asymptotic covariance matrix. The next theorem makes our heuristic argument more rigorous.

**Theorem 3** (**Minimum Asymptotic Variance**). *Assume that (A1)-(A3) and (A5)-(A8) hold. Then the set of matrices* $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1}$ *has a minimum* $V_T^0$ *given by* $V_T^0 \equiv \alpha(1-\alpha)\{T^{-1}\sum_{t=1}^T E[f_t^0(q_\alpha(W_t,\theta_0))^2\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)']\}^{-1}$, *attained by the M-estimator* $\theta_T^*$ *of* $\theta_0$ *which minimizes* $\Psi_T^*(\theta) \equiv T^{-1}\sum_{t=1}^T \varphi(Y_t, q_\alpha(W_t,\theta), \xi_t^*)$ *where* $\varphi(y,q,\xi_t^*) = [\alpha - 1\!\!1(q-y)][F_t^0(y) - F_t^0(q)]$, *a.s.* $-P$, *on* $\mathbb{R}\times\mathcal{Q}\times E_t$, *for every* $t$, $1 \leqslant t \leqslant T, T \geqslant 1$.

Theorem 3 shows two important results. Firstly, the matrix $V_T^0$ is the lower bound of the set of asymptotic variances of all the consistent and asymptotically normal M–estimators of $\theta_0$ that satisfy (A3): i.e. for any $\xi_t$ and $A(\cdot,\xi_t)$ in Theorem 1, the difference between the corresponding asymptotic covariance matrix $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1}$

and $V_T^0$ is always positive semidefinite. Secondly, Theorem 3 shows that this lower bound is attained, i.e. that there exists an optimal M–estimator $\theta_T^*$ whose asymptotic covariance matrix equals $V_T^0$. As a result, $V_T^0$ is the minimum asymptotic variance matrix of the class of M–estimators considered here.

The optimal M–estimator is obtained by minimizing the objective function $\Psi_T^*(\theta) = T^{-1}\sum_{t=1}^{T}\varphi(Y_t, q_\alpha(W_t,\theta), \xi_t^*)$, in which

$$\varphi(Y_t, q_\alpha(W_t,\theta), \xi_t^*) = [\alpha - \mathbb{1}(q_\alpha(W_t,\theta) - Y_t)][F_t^0(Y_t) - F_t^0(q_\alpha(W_t,\theta))], \tag{5}$$

$a.s. - P$, for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$. In particular, the shape $\xi_t^*$ of the optimal objective function in (5) is that of the true conditional distribution $F_t^0(\cdot)$, which is stochastic and $\mathcal{W}_t$ -measurable as required by (A3)(i). Even though the optimal M–estimator $\theta_T^*$ satisfies all the assumptions in (A3), note that its computation is not feasible, as the true conditional distribution $F_t^0(\cdot)$ is not known. Finally, it is worth pointing out that the optimal M–estimator $\theta_T^*$ is not a member of the QMLEs whose limit restrictions on $A(\cdot, \xi_t^F)$ exclude the possibility of choosing $A(\cdot, \xi_t^F)$ to be a distribution function.

What Theorem 3 does not show is whether $V_T^0$ is also the semiparametric efficiency bound for $\theta_0$, in addition to being the minimum of the set of asymptotic covariance matrices of consistent and asymptotically normal M–estimators. We provide an answer to this question in the next section.

## 4. Parametric estimators for conditional quantiles

The starting point of the previous section was the correct specification assumption (A2): it defined a conditional moment restriction which identified the conditional quantile parameter $\theta_0$. The question which we now pose is: what fully parametric estimators of $\theta_0$ can we construct that satisfy the conditional quantile restriction in (A2)?

***4.1. Stein's (1956) approach: an example.*** Stein's (1956) original concern was the possibility of adaptive estimation: can we estimate the parameter $\theta_0$ in the conditional quantile restriction (A2) as precisely as if we knew the set of true conditional densities $f^0 \equiv \{f_t^0(\cdot), 1 \leqslant t \leqslant T, T \geqslant 1\}$, up to some finite dimensional parameter? We illustrate it using the example given in the Introduction.

Recall that the model in (1) is pure time–series: there are no exogenous variables $X_t$ so $Z_t \equiv Y_t$. The conditioning set is $\mathcal{W}_t \equiv \sigma(\underline{Y}_{t-1})$ where $\underline{Y}_{t-1} \equiv (Y_1, ..., Y_{t-1})$, and the conditional quantile of $Y_t$ is assumed to be linear $\theta' W_t$ with $W_t \equiv (1, Y_{t-1})'$.

*4.1.1. Case 1: iid with finite dimensional nuisance parameter.* Assume that $U_t$'s are iid with know distribution function $H_0(\cdot)$ (and density $h_0(\cdot)$) such that $H_0^{-1}(\alpha) = 0$. The conditional density of $Y_t$ in the linear quantile regression model (1) then equals $f_t^0(y) = (1 + \gamma_0 |U_{t-1}|)^{-1} h_0([1 + \gamma_0 |U_{t-1}|]^{-1}[y - \theta_0' W_t])$, and its conditional $\alpha$-quantile is given by $\theta_0' W_t$, where $\theta_0$ and $\gamma_0$ denote the true values of the parameters $\theta \in \Theta \subseteq \mathbb{R}^k, k = 2$ and $\gamma \in \Gamma \subseteq \mathbb{R}_+$.

Here, the true set of conditional densities $f^0$ belongs to the parametric family $\mathcal{P}$, $\mathcal{P} \equiv \{f(\eta), \eta \in \Pi\}$ with $f(\eta) \equiv \{f_t(\cdot, \eta) : \mathbb{R} \to \mathbb{R}_*^+, 1 \leqslant t \leqslant T, T \geqslant 1\}$, indexed by a finite-dimensional parameter $\eta \in \Pi \subseteq \mathbb{R}^p$: $\eta \equiv (\theta', \gamma)' \in \Pi \equiv \Theta \times \Gamma$ and $p \equiv k + 1$. The members $f(\eta)$ of $\mathcal{P}$ are $f_t(y, \eta) = (1 + \gamma |U_{t-1}|)^{-1} h_0([1 + \gamma |U_{t-1}|]^{-1}[y - \theta' V_t])$, for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$. In this situation, the parameter of interest $\theta$ has a lower dimensionality than $\eta$: $\dim \theta = k$ and $\dim \eta = p = k + 1$. We can write $\theta = \theta(\eta)$, with $\theta : \Pi \to \Theta$ being some continuously differentiable function, and interpret the rest of $\eta$ as a nuisance parameter (Stein, 1956; Bickel, 1982).

Let $\eta_0$ index the true set of conditional densities of $Y_t$, i.e. $f(\eta_0) = f^0$, so that the true value of interest $\theta_0$ is now written as $\theta_0 = \theta(\eta_0)$ where $\eta_0 \equiv (\theta_0', \gamma_0)'$. The above parametric model $f(\eta)$ is regular (Bickel, 1982; Begun, Hall, Huang, and Wellner, 1983), all the conditional densities $f_t(\cdot, \eta_0)$ satisfy the conditional quantile restriction (A2) and are continuously differentiable on $\mathbb{R}$, and $f_t(Y_t, \cdot)$ is continuously differentiable on $\Pi$ $a.s. - P$. The true value $\theta_0$ of $\theta$ can be estimated by using a maximum likelihood approach. Given $T \geqslant 1$, denote by $L_T(\eta) \equiv \ln f_{\underline{Y}_T}(\underline{Y}_T)$ the log-likelihood function. Then $L_T(\eta) = \sum_{t=1}^{T} \ln f_t(Y_t, \eta)$. Also, let $I_T(\eta)$ denote the Fisher information matrix of the parametric model $\mathcal{P}$, $I_T(\eta) \equiv T^{-1}E[\nabla_\eta L_T(\eta)\nabla_\eta L_T(\eta)']$. Then, a regular estimator $\tilde{\theta}_T$ of $\theta_0$ is efficient if and only if $(C_T^0)^{-1/2}\sqrt{T}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$, with $C_T^0 \equiv \nabla_\eta\theta(\eta_0)(I_T(\eta_0))^{+}\nabla_\eta\theta(\eta_0)'$ (Bickel (1982)). In the special case where the sequence $\{(Y_t, W_t')'\}$ is iid, several authors have derived necessary and sufficient conditions for the MLE to be efficient.

*4.1.2. Case 2: iid with infinite dimensional nuisance parameter.* Now consider a more realistic situation in which the true density of $U_t$ in (1) is entirely unknown. Instead, $f^0$ are only known to belong to a class $\mathcal{S}$ which contains all parametric families such as $\mathcal{P}$. Unlike in $\mathcal{P}$, the sets of densities in $\mathcal{S}$ are indexed by an additional infinite dimensional parameter. In the case of our model (1) this infinite dimensional parameter is the unknown probability density $h_0(\cdot)$ of the error term $U_t$. The density $h_0(\cdot)$ could be any probability density in a set $\mathcal{H}$—set of all families $h$ of probability densities on $\mathbb{R}$, which satisfy $H^{-1}(\alpha) = 0$.

The set $\mathcal{S}$ is the union of all parametric sub-families $\mathcal{P}_h \equiv \{f_{h_\tau}(\eta), \eta \in \Pi\}$ obtained when $h$ is a smooth mapping $h : \Upsilon \to \mathcal{H}$ which associates a density $h_\tau(\cdot) \equiv h(\cdot, \tau)$ in $\mathcal{H}$ to a finite dimensional parameter $\tau \in \Upsilon \subseteq \mathbb{R}^r$ with $h_0(\cdot) = h(\cdot, \tau_0)$, $\tau_0 \in \Upsilon$. For any given $h$, the parametric submodel $f_{h_\tau}(\eta)$ is defined as $f_{h_\tau}(\eta) \equiv \{f_{h_\tau t}(\cdot, \eta) : \mathbb{R} \to \mathbb{R}_*^+, 1 \leqslant t \leqslant T, T \geqslant 1\}$. In addition, for any $\tau \in \Upsilon$, the conditional densities $f_{h_\tau t}(\cdot,\eta_0)$ satisfy the conditional quantile restriction (A2), are continuously differentiable on $\mathbb{R}$, and such that $f_{h_\tau t}(Y_t,\cdot)$ are continuously differentiable $a.s.-P$ on $\Pi$. As previously, consider the log-likelihood function under $\mathcal{P}_h$, $L_{hT}(\eta,\tau) = \sum_{t=1}^{T}\ln f_{h_\tau t}(Y_t,\eta)$, let $I_{hT}(\eta,\tau) \equiv T^{-1}E[\nabla_{(\eta',\tau')'}L_{hT}(\eta,\tau)\nabla_{(\eta',\tau')'}L_{hT}(\eta,\tau)']$ be the corresponding Fisher information matrix, and assume that the standard regularity conditions (Bickel, 1982; Begun, Hall, Huang, and Wellner, 1983) are satisfied. In particular, the matrix $I_{hT}(\eta_0,\tau_0)$ is such that its $k\times k$ upper left corner $C^0_{hT} \equiv \nabla_{(\eta',\tau')'}\theta(\eta_0)'(I_{hT}(\eta_0,\tau_0))^{+}\nabla_{(\eta',\tau')'}\theta(\eta_0)$ is nonsingular. Then, following Stein (1956), the semiparametric efficiency bound for the conditional quantile parameter $\theta_0$ is defined as the supremum of $C^0_{hT}$ over those $h$. If such a bound is attained by a particular family $h^*$, then $\mathcal{P}^* \equiv \mathcal{P}_{h^*}$ is called the least favorable parametric submodel.

*4.1.3. Case 3: non-iid with infinite dimensional nuisance parameter.* Finally, consider the case in which the errors $U_t$ are no longer iid, and let $\underline{h}^0_T$ denote the joint density of $\underline{U}_T \equiv (U_1,...,U_T)$, for $T \geqslant 1$. In the two previous cases $\underline{h}^0_T$ was uniquely determined by the marginal density $h_0$ of $U_t$; now, $\underline{h}^0_T$ contains an additional nonparametric component which is the time-dependence across different terms in the error sequence $\{U_t\}$. Still assuming the independence of $\{U_t\}$ and $\{W_t\}$, we have that $P(Y_t \leqslant \theta_0'W_t|\mathcal{W}_t) = P((1+\gamma_0|U_{t-1}|)U_t \leqslant 0|\mathcal{W}_t) = P(U_t \leqslant 0)$ where the last equality uses the fact that $\gamma_0 \in \mathbb{R}_+$. Hence, irrespective of the time-dependence structure across different $U_t$'s, the conditional quantile of $Y_t$ still equals $\theta_0'W_t$ provided $P(U_t \leqslant 0) = \alpha$. We therefore assume that $\underline{h}^0_T$ belongs to the set $\mathcal{H}_T$ of all joint probability densities on $\mathbb{R}^T$ whose marginals have zero $\alpha$-quantile.

Similar to previously, we let $\mathcal{S}$ be the union of all parametric sub-families $\mathcal{P}_{\underline{h}_T} \equiv \{f_{\underline{h}_{T\tau}}(\eta),\eta\in\Pi\}$ obtained when $\underline{h}_T$ is a smooth mapping $\underline{h}_T:\Upsilon\to\mathcal{H}_T$ which associates a joint probability density $\underline{h}_{T\tau}(\cdot)\equiv\underline{h}_T(\cdot,\tau)$ to $\tau\in\Upsilon\subseteq\mathbb{R}^r$, with $\underline{h}^0_T =$

$\underline{h}_T(\cdot, \tau_0)$ for some $\tau_0 \in \Upsilon$. The parameter $\tau$ now parameterizes all the marginal densities of $U_t$'s as well as their dependence. For any given $\underline{h}_T$, the parametric submodel $f_{\underline{h}_{T\tau}}(\eta)$ is defined as $f_{\underline{h}_{T\tau}}(\eta) \equiv \{f_{\underline{h}_{T\tau}t}(\cdot, \eta) : \mathbb{R} \to \mathbb{R}_*^+, 1 \leqslant t \leqslant T, T \geqslant 1\}$ with $f_{\underline{h}_{T\tau}t}(Y_t, \cdot)$ continuously differentiable $a.s. - P$ on $\Pi$, $f_{\underline{h}_T t}(\cdot, \eta_0)$ continuously differentiable on $\mathbb{R}$, and such that the conditional quantile restriction (A2) is satisfied. We denote by $I_{\underline{h}_T T}(\eta, \tau) \equiv T^{-1}E[\nabla_{(\eta',\tau')'}L_{\underline{h}_T T}(\eta, \tau)\nabla_{(\eta',\tau')'}L_{\underline{h}_T T}(\eta, \tau)']$ and $L_{\underline{h}_T T}(\eta, \tau) = \sum_{t=1}^T \ln f_{\underline{h}_{T\tau}t}(Y_t, \eta)$ the Fisher information matrix and the log-likelihood function of the parametric submodel $\mathcal{P}_{\underline{h}_T}$, respectively. Provided, again, the regularity of the model and nonsingularity of $C^0_{\underline{h}_T T} \equiv \nabla_{(\eta',\tau')'}\theta(\eta_0)'(I_{\underline{h}_T T}(\eta_0, \tau_0))^+\nabla_{(\eta',\tau')'}\theta(\eta_0)$, we can apply Stein's (1956) definition of the semiparametric efficiency bound for $\theta_0$ to be the supremum of $C^0_{\underline{h}_T T}$ over the mappings $\underline{h}_T$. Like in the previous case, if this supremum is attained by a particular family $\underline{h}^*_T$, then $\mathcal{P}^* \equiv \mathcal{P}_{\underline{h}^*_T}$ is called the least favorable parametric submodel.

***4.2. Parametric Submodel for the Conditional Quantile Restriction.*** We now go back to our general setup in which the conditioning set $\mathcal{W}_t$ not only contains lagged $Y_t$'s but also present and past values of the exogenous random $n$-vector $X_t$. The conditional quantile restriction in (A2) is equivalent to the following model:

$$Y_t = q_\alpha(\theta_0, W_t) + \epsilon_t, \tag{6}$$

in which $E[\mathbb{I}(\epsilon_t)|\mathcal{W}_t] + \alpha = 0, a.s. - P$, for any $t, 1 \leqslant t \leqslant T, T \geqslant 1$. The dependence assumption in (A7) implies that $\{(W_t', \epsilon_t)'\}$ is strong mixing. The model (6) contains our example (1) as a special case: $\epsilon = (1 + \gamma_0|U_{t-1}|)U_t$.

We now construct a parametric submodel of the conditional quantile model (6) by following the same steps as in the previous section. Recall that the log-likelihood function then equals $[\sum_{t=1}^T \ln f_t^0(y_t)] + \ln f_{\underline{X}_T}(\underline{x}_T)$. In other words, the log-likelihood function is the sum of two terms: the "conditional" log-likelihood of the $Y_t$'s and the log-likelihood of the $X_t$'s. Since the latter are exogenous, their log-likelihood does not depend on any parameters that enter the conditional distribution of the $Y_t$'s. Hence, the analysis in our previous section still applies provided we replace the log-likelihood function with the "conditional" log-likelihood function. The parameter $\eta$ now reduces to $\theta$; the parameter $\tau$ used to parameterize the joint densities is set to be $\theta$ itself.

Given $\mathcal{M}$ and the set of true conditional densities $f^0 \equiv \{f_t^0, 1 \leqslant t \leqslant T, T \geqslant 1\}$, consider the parametric submodel $\mathcal{P}^* \equiv \{f^*(\theta), \theta \in \Theta\}$ parameterized by the conditional quantile parameter $\theta$ in which $f^*(\theta) \equiv \{f_t^*(\cdot, \theta) : \mathbb{R} \to \mathbb{R}_*^+, 1 \leqslant t \leqslant T, T \geqslant 1\}$ with

$$f_t^*(y,\theta) \equiv f_t^0(y) \times \frac{\alpha(1-\alpha)\lambda(\theta)\exp\{\lambda(\theta)[F_t^0(y) - F_t^0(q_\alpha(W_t,\theta))][\mathbb{1}(q_\alpha(W_t,\theta)-y)-\alpha]\}}{1-\exp\{\lambda(\theta)[1-F_t^0(q_\alpha(W_t,\theta)) - \mathbb{1}(q_\alpha(W_t,\theta)-y)][\mathbb{1}(q_\alpha(W_t,\theta)-y)-\alpha]\}}, \tag{7}$$

for all $y \in \mathbb{R}$, where $\lambda(\theta) \equiv \Lambda(\theta - \theta_0)$ and $\Lambda : \mathbb{R}^k \to \mathbb{R}$ is at least twice continuously differentiable on $\mathbb{R}^k$ with $\Lambda(\cdot) > 0$ on $\mathbb{R}^k \backslash \{0\}$, $\Lambda(0) = 0$, $\nabla_\theta \Lambda(0) = 0$, $\Delta_{\theta\theta}\Lambda(0)$ nonsingular and $|\Delta_{\theta\theta}\Lambda(\cdot)| < \infty$ in a neighborhood of 0.

The following theorem shows that $\mathcal{P}^*$ is a parametric submodel in $\mathcal{S}$, i.e. that: (i) for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, $f_t^*(\cdot,\theta)$ is a probability density for all $\theta \in \Theta$; (ii) for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, $f_t^*(\cdot,\theta)$ satisfies the conditional quantile restriction $E_\theta[\mathbb{1}(q_\alpha(W_t,\theta) - Y_t) - \alpha | \mathcal{W}_t] = 0, a.s. - P$, for all $\theta \in \Theta$, where $E_\theta(\cdot|\mathcal{W}_t)$ denotes the conditional expectation under the density $f_t^*(\cdot,\theta)$ for $Y_t$ given $\mathcal{W}_t$; and (iii) that $\mathcal{P}^*$ contains the true data generating process $f^0 \in \mathcal{P}^*$.

**Theorem 4** (**Parametric Submodel**). *Under (A1)(ii) and (A2), the parametric submodel $\mathcal{P}^*$ defined by* (7) *is a submodel of $\mathcal{S}$.*

The analytic expression of the parametric submodel presented in Theorem 4 is new. In particula, the density $f_t^*(\cdot,\theta)$ is not of the 'tick-exponential' form derived by

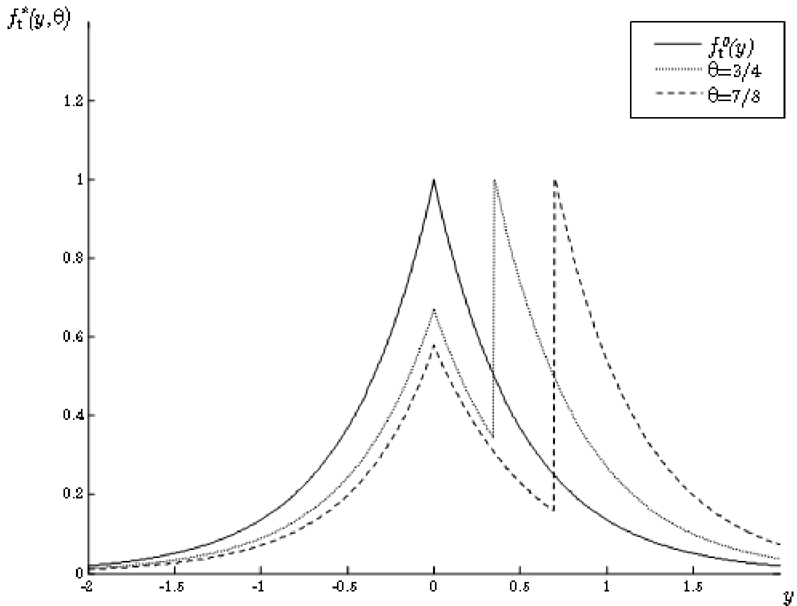


FIG 1. *Case* $\alpha = .5$, $q_\alpha(W_t, \theta) = \theta$ *and* $f_t^0(y) = \exp(-2|y|)$

Komunjer (2005): it depends on the true density $f_t^0(\cdot)$ as well as the true value $\theta_0$ and contains terms such as $\lambda(\theta)$. Note that a simple function $\Lambda(\cdot)$ which satisfies the conditions of Theorem 4 is $\Lambda(x) = x'x$.

In the parametric submodel $\mathcal{P}^*$, $\theta$ parameterizes both the conditional quantile model $\mathcal{M}$ and the shape of $f_t^*(\cdot, \theta)$—in other words, the shape of $f_t^*(\cdot, \theta)$ is now determined by $f_t^0(\cdot)$ and $\theta$ (see Figure 1 for a purely location model of a conditional median when the true density is double exponential). In particular, the density $f_t^*(\cdot, \theta)$ is discontinuous for all values of $\theta$ different from $\theta_0$; when $\theta = \theta_0$ the density $f_t^*(\cdot, \theta_0)$ equals the true density $f_t^0(y)$ which is continuous.

Because $\mathcal{P}^*$ is a parametric submodel of the set $\mathcal{S}$ of all densities satisfying the conditional quantile restriction in (A2), the semiparametric efficiency bound for $\theta_0$ is by Stein's (1956) definition at least as large as the asymptotic variance of the above MLE $\tilde{\theta}_T^*$. We derive the asymptotic distribution of $\tilde{\theta}_T^*$ in the following theorem.

**Theorem 5** (**Asymptotic Distribution of the MLE**). *Under (A1)-(A2) and (A6)-(A8)(i), the asymptotic distribution of the MLE $\tilde{\theta}_T^*$ associated with $\mathcal{P}^*$ is given by $(V_T^0)^{-1/2}\sqrt{T}(\tilde{\theta}_T^* - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$ where $V_T^0$ is as defined in Theorem 3.*

Recall from Theorem 3 that $V_T^0$ is the minimum of the asymptotic variances of the consistent and asymptotically normal M–estimators of $\theta_0$. On the other hand, Theorem 5 shows that there exists a parametric submodel $\mathcal{P}^*$ in which the MLE $\tilde{\theta}_T^*$ of the true parameter $\theta_0$ has the same asymptotic covariance matrix $V_T^0$. It follows, first, that the parametric model $\mathcal{P}^*$ is the least favorable parametric submodel in $\mathcal{S}$, and, second, that $V_T^0$ is the semiparametric efficiency bound.

**Corollary 6** (**Semiparametric Efficiency Bound**). *Under (A1)-(A2) and (A6)-(A8)(i), $V_T^0 = \alpha(1-\alpha)\{T^{-1}\sum_{t=1}^T E[f_t^0(q_\alpha(W_t, \theta_0))^2 \nabla_\theta q_\alpha(W_t, \theta_0) \nabla_\theta q_\alpha(W_t, \theta_0)']\}^{-1}$ is the semiparametric efficiency bound with efficient score $[\alpha(1-\alpha)]^{-1/2} f_t^0(q_\alpha(W_t, \theta_0)) \times \nabla_\theta q_\alpha(W_t, \theta_0)$.*

That $V_T^0$ is the semiparametric efficiency bound has the following interpretation: when the only thing we know about the model is that it satisfies the conditional quantile restriction (A2), then we cannot estimate the true conditional quantile parameter $\theta_0$ with precision higher than that given by $V_T^0$. Note that our result uses the moment restriction (A2) only; we do not make any additional assumptions regarding the properties of the error term $\epsilon_t$ in (6) (other than those contained in (A2) and (A7)). In particular, we allow for $\{\epsilon_t\}$ to be dependent and nonidentically distributed.

Perhaps the most important aspect of Corollary 6 is that it relaxes the independence assumption. So far as time series data are concerned, two leading situations in which the independence is violated come into mind. First is a quantile regression model in which $W_t$ contains serially dependent exogenous variables or/and lags of $Y_t$, but where the residuals are uncorrelated (though possibly conditionally heteroskedastic). There are some results on this case in Newey and Powell (1990), under the additional assumption that $\{(Y_t, W_t')'\}$ is iid. The authors derive the semiparametric efficiency bound for the parameters in the linear quantile regression $q_\alpha(W_t, \theta) = \theta' W_t$ by allowing for conditional heteroskedasticity (given $W_t$) in the error term $Y_t - \theta' W_t$. Corollary 6 generalizes Newey and Powell's (1990) results to the case where the sequence $\{(Y_t, W_t')'\}$ is weakly dependent and heterogeneous, as in (A7). Unsurprisingly, when the data is iid and $q_\alpha$ linear, the bound $V_T^0$ reduces to $V^0 \equiv \alpha(1-\alpha)\{E[(f_t^0(q_\alpha(W_t, \theta_0)))^2 W_t W_t']\}^{-1}$ derived by Newey and Powell (1990).

In the second time series situation of interest, the residuals themselves are correlated in addition to being heteroskedastic. Note that this situation is not covered in Newey and Powell's (1990) model; however, our assumption (A2) does not exclude the possiblity that $\{\epsilon_t\}$ be correlated. So far there exist no results on semiparametric efficiency bound which cover this dependent case. To the best of our knowledge, Corollary 6 provides the first result on semiparametric efficiency for nonlinear (and possibly censored) conditional quantile models when the data is dependent.

## 5. Proofs

*Proof of Theorem 1.* First, note that (A3)-(A4) together with the compactness of the parameter space $\Theta$, are sufficient conditions for $\theta_T$ to be consistent for $\theta_\infty^0 \in \mathring{\Theta}$ (see, e.g., Theorem 2.1 in Newey and McFadden, 1994). We now show that $\theta_\infty^0 = \theta_0$ for any $T \geqslant 1$ if and only if $\Psi_T(\cdot)$ is of the form defined by functions $A(\cdot,\cdot)$ and $B(\cdot,\cdot)$ in Theorem 1. We treat the two implications separately.

STEP 1 (sufficiency): By (A4)(i), $\nabla_\theta E[\Psi_T(\theta)] = T^{-1}\sum_{t=1}^T E[\nabla_\theta \varphi(Y_t, q_\alpha(W_t,\theta),\xi_t)]$. From the expression of $\varphi(\cdot,\cdot,\cdot)$, twice continuous differentiability of $A(\cdot,\xi_t)$ $a.s.-$on $\mathcal{Q}$ for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, and (A2) we have $E[\nabla_\theta \varphi(Y_t, q_\alpha(W_t,\theta_0),\xi_t)] = E\{\nabla_\theta q_\alpha(W_t,\theta_0) a(q_\alpha(W_t,\theta_0),\xi_t) E[\mathbb{1}(q_\alpha(W_t,\theta_0) - Y_t) - \alpha|\mathcal{W}_t]\} = 0, a.s. - P$, so that $\nabla_\theta E[\Psi_T(\theta_0)] = 0$. Similarly, $\Delta_{\theta\theta} E[\Psi_T(\theta_0)] = T^{-1}\sum_{t=1}^T E[\Delta_{\theta\theta}\varphi(Y_t, q_\alpha(W_t,\theta_0),\xi_t)]$ where for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$,

$$\begin{aligned}
&E[\Delta_{\theta\theta}\varphi(Y_t, q_\alpha(W_t,\theta_0),\xi_t)] \\
&= E\{\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)' a(q_\alpha(W_t,\theta),\xi_t) E[\delta(q_\alpha(W_t,\theta_0) - Y_t)|\mathcal{W}_t]\} \\
&= E[\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)' a(q_\alpha(W_t,\theta_0),\xi_t) f_t^0(q_\alpha(W_t,\theta_0))],
\end{aligned} \tag{8}$$

where $f_t^0(\cdot)$ is the true probability density function of $Y_t$ conditional on $\mathcal{W}_t$. We now show that $\Delta_{\theta\theta} E[\Psi_T(\theta_0)] \gg 0$. By using (8), $\chi'\Delta_{\theta\theta} E[\Psi_T(\theta_0)]\chi = 0$ for any $\chi \in \mathbb{R}^k$ only if $T^{-1}\sum_{t=1}^T E[\chi'\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'\chi a(q_\alpha(W_t,\theta_0),\xi_t) f_t^0(q_\alpha(W_t,\theta_0))] = 0$. Now, note that for any $t$, $1 \leqslant t \leqslant T$ and $T \geqslant 1$,

$$\begin{aligned}
&E[\chi'\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'\chi a(q_\alpha(W_t,\theta_0),\xi_t) f_t^0(q_\alpha(W_t,\theta_0))] \\
&= E[(\chi'\nabla_\theta q_\alpha(W_t,\theta_0))^2 a(q_\alpha(W_t,\theta_0),\xi_t) f_t^0(q_\alpha(W_t,\theta_0))] \geqslant 0,
\end{aligned} \tag{9}$$

for any $\chi \in \mathbb{R}^k$, since $a(q_\alpha(W_t,\theta_0),\xi_t) > 0, a.s. - P$ and $f_t^0(q_\alpha(W_t,\theta_0)) > 0, a.s. - P$. Taking into account (9) we have that $\chi'\Delta_{\theta\theta} E[\Psi_T(\theta_0)]\chi = 0$ for any $\chi \in \mathbb{R}^k$ only if $E[(\chi'\nabla_\theta q_\alpha(W_t,\theta_0))^2 a(q_\alpha(W_t,\theta_0),\xi_t) f_t^0(q_\alpha(W_t,\theta_0))] = 0$ for all $t$, $1 \leqslant t \leqslant T$, $T \geqslant 1$. Using again the strict positivity of $a(\cdot,\xi_t)$ and $f_t^0(\cdot)$ this last equality is true only if $\chi'\nabla_\theta q_\alpha(W_t,\theta_0) = 0, a.s. - P$, for every $t$, $1 \leqslant t \leqslant T$, $T \geqslant 1$. This, with (A1)(iii), implies that $\chi = 0$. From there we conclude that $\Delta_{\theta\theta} E[\Psi_T(\theta_0)] \gg 0$ and therefore $\theta_0$ is a minimizer $E[\Psi_T(\theta)]$ on $\mathring{\Theta}$. Since by (A4)(ii) this minimizer is unique, we have that for any $T \geqslant 1$, $\theta_\infty^0 = \theta_0$ which completes the sufficiency part of the proof.

STEP 2 (necessity): Given the differentiability of $E[\Psi_T(\theta)]$ on $\Theta$ by (A4)(i), a necessary requirement for $\theta^0_\infty = \theta_0$ is that the first order condition $\nabla_\theta E[\Psi_T(\theta_0)] = 0$ be satisfied, which is equivalent to

$$T^{-1}\sum_{t=1}^{T} E\{\nabla_\theta q_\alpha(W_t,\theta_0)E[\tfrac{\partial\varphi}{\partial q}(Y_t, q_\alpha(W_t,\theta_0),\xi_t)|\mathcal{W}_t]\} = 0.$$

Since the above equality needs to hold for any $T \geqslant 1$, any choice of conditional quantile model $\mathcal{M}$ and for any true parameter $\theta_0 \in \mathring{\Theta}$, we need to find a necessary condition for the implication

$$\begin{aligned} E[\mathbb{I}(q_\alpha(W_t,\theta_0) - Y_t) - \alpha|\mathcal{W}_t] &= 0, a.s. - P \\ &\Rightarrow E[\tfrac{\partial\varphi}{\partial q}(Y_t, q_\alpha(W_t,\theta_0),\xi_t)|\mathcal{W}_t] = 0, a.s. - P, \end{aligned} \tag{10}$$

to hold, for all $t$, $1 \leqslant t \leqslant T$, $T \geqslant 1$, and all absolutely continuous distribution function $F^0_t$ in $\mathcal{F}$. We now show that

$$\tfrac{\partial\varphi}{\partial q}(Y_t, q_\alpha(W_t,\theta_0),\xi_t) = a(q_\alpha(W_t,\theta_0),\xi_t)[\mathbb{I}(q_\alpha(W_t,\theta_0) - Y_t) - \alpha], a.s. - P, \tag{11}$$

for any $\theta_0 \in \mathring{\Theta}$ and any $t$, $1 \leqslant t \leqslant T$, $T \geqslant 1$, where $a(\cdot,\xi_t) : \mathbb{R} \to \mathbb{R}$ is strictly positive $a.s. - P$ on $\mathcal{Q}$, is a necessary condition for (10). Using a generalized Farkas lemma (Lemma 8.1, p 240, vol 1) in Gourieroux and Monfort (1995), (10) implies there exists a $\mathcal{W}_t$ -measurable random variable $a_t$ such that

$$\tfrac{\partial\varphi}{\partial q}(Y_t, q_\alpha(W_t,\theta_0),\xi_t) = a_t[\mathbb{I}(q_\alpha(W_t,\theta_0) - Y_t) - \alpha], a.s. - P.$$

Since the left-hand side only depends on $Y_t$, $q_\alpha(W_t,\theta_0)$ and $\xi_t$, the same must hold for the right-hand side. Hence, $a_t$ can only depend on $q_\alpha(W_t,\theta_0)$ and $\xi_t$ and we can write $a_t = a(q_\alpha(W_t,\theta_0),\xi_t)$; so (11) holds. We now need to show that $a(\cdot,\xi_t)$ is strictly positive $a.s. - P$ on $\mathcal{Q}$. A necessary condition for $\theta_0 \in \mathring{\Theta}$ to be a minimizer of $E[\Psi_T(\theta)]$ (in addition to the above first order condition) is that for every $\chi \in \mathbb{R}^k$ the quadratic form $\chi'\Delta_{\theta\theta}E[\Psi_T(\theta_0)]\chi \geqslant 0$ (existence of $\Delta_{\theta\theta}E[\Psi_T(\theta)]$ is ensured by (A4)(i)). Taking into account (11) and our previous computations leading to (8), we have

$$\chi'\Delta_{\theta\theta}E[\Psi_T(\theta_0)]\chi = T^{-1}\sum_{t=1}^{T} E[(\chi'\nabla_\theta q_\alpha(W_t,\theta_0))^2 a(q_\alpha(W_t,\theta_0),\xi_t) f_t^0(q_\alpha(W_t,\theta_0))].$$

Hence, the quadratic form $\chi'\Delta_{\theta\theta}E[\Psi_T(\theta_0)]\chi$ is nonnegative for any $T \geqslant 1$, any conditional quantile model $\mathcal{M}$, any true value $\theta_0 \in \mathring{\Theta}$ and any conditional density $f_t^0(\cdot)$, only if $a(q_\alpha(W_t,\theta_0),\xi_t) \geqslant 0, a.s. - P$, for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$. Note that the uniqueness of the solution $\theta_0$ implies that $a(q,\xi_t) > 0, a.s. - P$ for any $q \in \mathcal{Q}$ and for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$. Using the continuity of $\varphi(Y_t,\cdot,\xi_t)$ $a.s. - P$ on $\mathcal{Q}$ in (A3)(ii), the necessary condition (11) then integrates into

$$\varphi(Y_t, q_\alpha(W_t,\theta_0),\xi_t) = \begin{cases} (1-\alpha)[A(q_\alpha(W_t,\theta_0),\xi_t) - A(Y_t,\xi_t)], \text{ if } Y_t \leqslant q_\alpha(W_t,\theta_0), \\ -\alpha[A(q_\alpha(W_t,\theta_0),\xi_t) - A(Y_t,\xi_t)], \text{ if } Y_t > q_\alpha(W_t,\theta_0), \end{cases}$$

$+B(Y_t,\xi_t), a.s. - P$, where for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, $A(\cdot,\xi_t)$ is an indefinite integral of $a(\cdot,\xi_t)$, $A(q,\xi_t) \equiv \int_a^q a(r,\xi_t)dr$, $a \in \mathbb{R}$, and $B(\cdot,\xi_t) : \mathbb{R} \to \mathbb{R}$ is a real function. Note that the above equality has to hold for any $\theta_0 \in \mathring{\Theta}$ so that for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, $\varphi(y,q,\xi_t) = [\alpha - \mathbb{1}(q-y)][A(y,\xi_t) - A(q,\xi_t)] + B(y,\xi_t), a.s. - P$, on $\mathbb{R} \times \mathcal{Q} \times E_t$. $\square$

*Proof of Theorem 2.* To show that Theorem 2 holds, we first show that under primitive conditions given in (A1)-(A3) and (A5)-(A8), $\theta_T$ is consistent for $\theta_0$, i.e. $\theta_T - \theta_0 \xrightarrow{p} 0$. We proceed by checking that all the assumptions for consistency used by Komunjer (2005) in her Theorem 3 hold. Given that her proof of consistency for the family of tick-exponential QMLEs derived in Theorem 3 does not require any assumptions on the limits in $\pm\infty$ of the functions $A(\cdot,\xi_t)$, it applies directly to the M–estimator $\theta_T$ defined in (A3). Assumptions A2 and A3 in Komunjer (2005) are satisfied by imposing our (A6) and (A5), respectively. The $\alpha$-mixing condition A4 in

Komunjer (2005) and the assumption that $W_t$ is a function of some finite number of lags of $Z_t$ stated in A0.iv in Komunjer (2005) are used to ensure that $\{(Y_t, W_t')'\}$ is $\alpha$-mixing of with $\alpha$ of the same size $-r/(r-2)$, $r > 2$. Here, we directly impose the mixing of the sequence $\{(Y_t, W_t')'\}$ in our (A7), which is sufficient for the proof of Theorem 3 in Komunjer (2005) to go through. Finally, the moment conditions A5 in Komunjer (2005) directly follow from our (A8) and the fact that $E[\sup_{\theta\in\Theta} |\nabla_\theta q_\alpha(W_t, \theta)|] \leqslant \max\{1, E[\sup_{\theta\in\Theta} |\nabla_\theta q_\alpha(W_t, \theta)|^2]\} < \infty$. Hence we can use the results of Theorem 3 in Komunjer (2005)—corresponding to the case where the conditional quantile model is correctly specified (A2)—which proves the consistency of $\theta_T$. Similarly, we derive asymptotic normality by using the results of Corollary 5 in Komunjer (2005). The boundedness of the second derivative of $A(\cdot, \xi_t)$ contained in assumption A3' in Komunjer (2005) is directly implied by (A5). The moment condition in assumption A5' in Komunjer (2005) follows from our (A8). Finally in our setup we have assumed that the true conditional density $f_t^0(\cdot)$ of $Y_t$ is strictly positive and bounded on $\mathbb{R}$, which verifies assumption A6 in Komunjer (2005). Hence, from Corollary 5 in Komunjer (2005) we know that$\sqrt{T}(\Sigma_T^0)^{-1/2}\Delta_T^0(\theta_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$ where $\Delta_T^0$ and $\Sigma_T^0$ are as defined in Theorem 2. □

*Proof of Theorem 3.* The proof of this theorem is done in two steps: we first show that $V_T^0$ is the lower bound of the set of asymptotic matrices $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1}$ obtained with functions $A(\cdot, \xi_t)$ satisfying the conditions of Theorem 1. Then, we show that $V_T^0$ is attained by a particular M–estimator that satisfies (A2). It then follows that $V_T^0$ is the minimum asymptotic variance.

STEP 1: This part is inspired by a similar result by Gourieroux, Monfort, and Trognon (1984). Let $V_T^0$ be as defined in Theorem 3 and consider the difference $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} - V_T^0$. We show that this difference is positive definite for any $A(\cdot, \xi_t)$,

$1 \leqslant t \leqslant T, T \geqslant 1$, in Theorem 1, by writing $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} - V_T^0 = V_T^0(V_T^0)^{-1}V_T^0 - V_T^0\Delta_T^0(\Delta_T^0)^{-1} - (\Delta_T^0)^{-1}\Delta_T^0 V_T^0 + (\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1}$. Then we have $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} - V_T^0 = [\alpha(1-\alpha)]^{-1}T^{-1}\sum_{t=1}^{T} E[\chi_t\chi_t']$, where for every $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, we let $\chi_t \equiv [f_t^0(q_\alpha(W_t,\theta_0))V_T^0 - \alpha(1-\alpha)a(q_\alpha(W_t,\theta_0),\xi_t)(\Delta_T^0)^{-1}]\nabla_\theta q_\alpha(W_t,\theta_0)$, and $a(y,\xi_t) \equiv \partial A(y,\xi_t)/\partial y$ as previously. Hence, for any $A(\cdot,\xi_t)$, $1 \leqslant t \leqslant T, T \geqslant 1$, such that $a(\cdot,\xi_t) > 0, a.s. - P$ on $\mathcal{Q}$, the matrix $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} - V_T^0$ is positive semidefinite.

STEP 2: We now show that, under (A1)-(A2), (A6)-(A7) and (A8)(i), this lower bound $V_T^0$ is attained by the M–estimator $\theta_T^*$ whose objective function $\Psi_T^*(\theta) \equiv T^{-1}\sum_{t=1}^T \varphi(Y_t, q_\alpha(W_t,\theta),\xi_t^*)$ is such that $\varphi(Y_t,q_t,\xi_t^*) = [\alpha - \mathbb{1}(q_t - Y_t)][F_t^0(Y_t) - F_t^0(q_t)], a.s. - P$ on $\mathbb{R} \times \mathcal{Q} \times E_t$, for every $1 \leqslant t \leqslant T, T \geqslant 1$. Note that the shape $\xi_t^*$ corresponds to the true conditional distribution $F_t^0(\cdot)$ which is stochastic and $\mathcal{W}_t$-measurable thereby satisfying (A3)(i). Moreover, $F_t^0(\cdot)$ is twice continuously differentiable with bounded $f_t^0(y)$ and $|df_t^0(y)/dy|$, which satisfies (A3)(ii) and (A5). Since $F_t^0(\cdot)$ is bounded by 1 the moment conditions (A8)(ii) hold. Hence, we can apply Theorem 2 to show that, under (A1)-(A2), (A6)-(A7) and (A8)(i), $\theta_T^*$ with $A(\cdot,\xi_t^*) = F_t^0(\cdot)$, is asymptotically distributed as $\sqrt{T}(\Sigma_T^0)^{-1/2}\Delta_T^0(\theta_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$ with $\Sigma_T^0 = \alpha(1-\alpha)\Delta_T^0$ and $\Delta_T^0 = T^{-1}\sum_{t=1}^T E\{[f_t^0(q_\alpha(W_t,\theta_0))]^2\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'\}$, so $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} = V_T^0$. □

*Proof of Theorem 4.* To prove $(i)$ and $(iii)$, we start by showing that for any $\theta \in \Theta\backslash\{\theta_0\}$, the function $f_t^*(\cdot,\theta)$ in (7) is a probability density, for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$. First, note that for any $\theta \in \Theta\backslash\{\theta_0\}$, $f_t^*(\cdot,\theta)$ is continuous and $f_t^*(\cdot,\theta) > 0$ on $\mathbb{R}$. Thus it suffices to show that $\int_{\mathbb{R}} f_t^*(y,\theta)dy = 1$. Consider the change of variable $u \equiv \lambda(\theta)F_t^0(y)$, where $\lambda(\theta)F_t^0(\cdot)$ is strictly increasing in $y$ since $\lambda(\theta) = \Lambda(\theta - \theta_0) > 0$ and $f_t^0(\cdot)$ is strictly positive (so $du = \lambda(\theta)f_t^0(y)dy$). To simplify the notation, we let

$q_t(\theta) \equiv q_\alpha(W_t, \theta)$. Noting that $\mathbb{I}(q_t(\theta) - y) = \mathbb{I}[\lambda(\theta)F_t^0(q_t(\theta)) - u]$, we have

$$\begin{aligned}\int_{\mathbb{R}} f_t^*(y,\theta)dy &= \int_0^{\lambda(\theta)F_t^0(q_t(\theta))} \frac{\alpha(1-\alpha)\exp\{(1-\alpha)[u - \lambda(\theta)F_t^0(q_t(\theta))]\}}{1-\exp[-(1-\alpha)\lambda(\theta)F_t^0(q_t(\theta))]}du \\ &\quad + \int_{\lambda(\theta)F_t^0(q_t(\theta))}^{\lambda(\theta)} \frac{\alpha(1-\alpha)\exp\{-\alpha[u-\lambda(\theta)F_t^0(q_t(\theta))]\}}{1-\exp\{-\alpha\lambda(\theta)[1-F_t^0(q_t(\theta))]\}}du \\ &= \alpha + (1-\alpha) = 1,\end{aligned}$$

which shows that $f_t^*(\cdot,\theta)$ is a probability density for any $\theta \in \Theta\backslash\{\theta_0\}$. We now show that this is also true for $\theta_0$ and that $f_t^*(\cdot,\theta_0) = f_t^0(\cdot)$. For this, let $P_t(\theta) \equiv \alpha(1-\alpha)\lambda(\theta)\exp\{\lambda(\theta)[F_t^0(y) - F_t^0(q_t(\theta))][\mathbb{I}(q_t(\theta)-y)-\alpha]\}$ and $Q_t(\theta) \equiv 1 - \exp\{\lambda(\theta)[1 - F_t^0(q_t(\theta)) - \mathbb{I}(q_t(\theta)-y)][\mathbb{I}(q_t(\theta)-y)-\alpha]\}$, so that $f_t^*(y,\theta) = f_t^0(y)P_t(\theta)/Q_t(\theta)$. By (A1)(ii), the functions $P_t$ and $Q_t$ are at least twice continuously differentiable on $\Theta$ $a.s. - P$; thus for every $(\theta,\theta_0) \in \Theta^2$ we can write their respective Taylor developments of order two. Straightforward though lengthy computations show that, for any function $\lambda(\theta) = \Lambda(\theta-\theta_0)$ such that $\nabla_\theta\Lambda(0) = 0$ and $\Delta_{\theta\theta}\Lambda(0)$ nonsingular, we have $P_t(\theta_0) = 0$, $D^1P_t(\theta_0) = 0$, $D^2P_t(\theta_0) = \alpha(1-\alpha)D^2\lambda(\theta_0)$, and $Q_t(\theta_0) = 0$, $D^1Q_t(\theta_0) = 0$, $D^2Q_t(\theta_0) = \alpha(1-\alpha)D^2\lambda(\theta_0)$. Hence

$$P_t(\theta) = \tfrac{1}{2}\alpha(1-\alpha)D^2\lambda(\theta_0)(\theta-\theta_0)^2 + o(|\theta-\theta_0|^2), \tag{12}$$

$$Q_t(\theta) = \tfrac{1}{2}\alpha(1-\alpha)D^2\lambda(\theta_0)(\theta-\theta_0)^2 + o(|\theta-\theta_0|^2). \tag{13}$$

Given the nonsingularity of $\Delta_{\theta\theta}\Lambda(0)$, an immediate consequence of l'Hôpital's rule and (12) $-$ (13) is that $\lim_{\theta\to\theta_0} P_t(\theta)/Q_t(\theta) = 1$. Hence by $a.s. - P$ continuity of $f_t^*(y,\cdot)$ on $\Theta$, we have, for any $y \in \mathbb{R}$, $f_t^*(y,\theta_0) = \lim_{\theta\to\theta_0} f_t^*(y,\theta) = f_t^0(y)$. This shows that $f_t^*(\cdot,\theta)$ is a probability density for any $\theta \in \Theta$, and that $f_t^*(\cdot,\theta_0) = f_t^0(\cdot)$, so that $f^0 \in \mathcal{P}^*$, as desired. It remains to be shown that this parametric model $\mathcal{P}^*$ satisfies the conditional moment restriction in $(ii)$ for all $\theta \in \Theta$. This restriction is clearly satisfied when $\theta = \theta_0$ as $f_t^*(\cdot,\theta_0) = f_t^0(\cdot)$ and $[\theta_0, f_t^0(\cdot)]$ satisfies (A2) by assumption.

When $\theta \neq \theta_0$, we have

$$E_\theta[\mathbb{1}(q_t(\theta) - Y_t)|\mathcal{W}_t] = \int_{-\infty}^{q_t(\theta)} f_t^*(y, \theta)dy = \alpha,$$

where we have again used the change of variable $u \equiv \lambda(\theta)F_t^0(y)$. $\square$

*Proof of Theorem 5.* We now consider the MLE $\tilde{\theta}_T^*$ which maximizes the log-likelihood $L_T(\theta) \equiv T^{-1} \sum_{t=1}^T \ln f_t^*(Y_t, \theta)$.

STEP1: First, we establish the consistency of $\tilde{\theta}_T^*$ by checking that conditions (i)-(iv) of Theorem 2.1 in Newey and McFadden (1994) hold. Given (A1)(i) we know that $\ln f_t^*(Y_t, \theta) \neq \ln f_t^*(Y_t, \theta_0)$ $a.s. - P$, whenever $\theta \neq \theta_0$ (see Figure 1 for example); this verifies the uniqueness condition (i) of Theorem 2.1. The compactness condition (ii) of Theorem 2.1 follows by assumption. Using $q_t(\theta) = q_\alpha(W_t, \theta)$ we have

$$\begin{aligned}\ln f_t^*(Y_t, \theta) = {} & \ln[\alpha(1-\alpha)f_t^0(Y_t)] + \ln \lambda(\theta) + \lambda(\theta)[F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbb{1}(q_t(\theta) - Y_t) - \alpha] \\ & - \ln\left(1 - \exp\{\lambda(\theta)[\mathbb{1}(q_t(\theta) - Y_t) - \alpha][1 - \mathbb{1}(q_t(\theta) - Y_t) - F_t^0(q_t(\theta))]\}\right),\end{aligned}$$

showing that $E[\ln f_t^*(Y_t, \theta)]$ is continous on $\Theta$ and that $E[\sup_{\theta \in \Theta} |\ln f_t^*(Y_t, \theta)|^{r+\epsilon}] < \infty$ for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, and $\epsilon > 0$; this verifies condition (iii) of Theorem 2.1. We show the uniform convergence condition (iv) of Theorem 2.1 by following the same steps as in the proof of Theorem 3 in Komunjer (2005). To simplify the notation let

$$x(\theta) \equiv [\mathbb{1}(q_t(\theta) - Y_t) - \alpha][1 - \mathbb{1}(q_t(\theta) - Y_t) - F_t^0(q_t(\theta))] \text{ and } u(z) \equiv \frac{\exp z}{1 - \exp z},$$

for $\theta \in \Theta$ and $z \in \mathbb{R}_-$. Note that $-1 < x(\theta) < 0$ and $-\lambda(\theta) < \lambda(\theta)x(\theta) < 0$ on $\Theta$ $a.s. - P$. We have

$$\begin{aligned}\nabla_\theta \ln f_t^*(Y_t, \theta) = {} & \tfrac{\nabla_\theta \lambda(\theta)}{\lambda(\theta)} + \nabla_\theta \lambda(\theta)[F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbb{1}(q_t(\theta) - Y_t) - \alpha] \\ & - \lambda(\theta) f_t^0(q_t(\theta)) \nabla_\theta q_t(\theta)[\mathbb{1}(q_t(\theta) - Y_t) - \alpha] + u(\lambda(\theta)x(\theta))\nabla_\theta(\lambda(\theta)x(\theta)) \\ & + \lambda(\theta)[F_t^0(Y_t) - F_t^0(q_t(\theta))]\delta(q_t(\theta) - Y_t)\nabla_\theta q_t(\theta) \qquad (14)\end{aligned}$$

where $\nabla_\theta(\lambda(\theta)x(\theta)) = \nabla_\theta\lambda(\theta)x(\theta)+\lambda(\theta)\nabla_\theta x(\theta)$, and $\nabla_\theta x(\theta) = \{f_t^0(q_t(\theta))[\alpha-\mathbb{I}(q_t(\theta)-Y_t)] + \delta(q_t(\theta) - Y_t)[\alpha - F_t^0(q_t(\theta))]\}\nabla_\theta q_t(\theta)$, given that $[\mathbb{I}(\cdot)]^2 = \mathbb{I}(\cdot)$. Note that $u(z) = -1/z - 1/2 + o(1)$ in the neigborhood of 0, and that $\lambda(\theta)x(\theta) = o_p(1)$ in the neigborhood of $\theta_0$, so $u(\lambda(\theta)x(\theta))\nabla_\theta(\lambda(\theta)x(\theta)) = -\frac{\nabla_\theta\lambda(\theta)}{\lambda(\theta)} - \frac{\nabla_\theta x(\theta)}{x(\theta)} + o_p(1)$ in the neighborhood of $\theta_0$. In particular, combining the above results, we get

$$\begin{aligned}\nabla_\theta \ln f_t^*(Y_t, \theta_0) &= -\nabla_\theta q_t(\theta_0)\left\{\tfrac{f_t^0(q_t(\theta_0))[\alpha-\mathbb{I}(q_t(\theta_0)-Y_t)]+\delta(q_t(\theta_0)-Y_t)[\alpha-F_t^0(q_t(\theta_0))]}{[\mathbb{I}(q_t(\theta_0)-Y_t)-\alpha][1-\mathbb{I}(q_t(\theta_0)-Y_t)-F_t^0(q_t(\theta_0))]}\right\} \\ &= -\tfrac{1}{\alpha(1-\alpha)}\nabla_\theta q_t(\theta_0) f_t^0(q_t(\theta_0))[\mathbb{I}(q_t(\theta_0) - Y_t) - \alpha],\end{aligned} \tag{15}$$

where the second equality uses $x(\theta_0) = -\alpha(1-\alpha)$ and $F_t^0(q_t(\theta_0)) = \alpha$. Using $-1 < x(\theta) < 0$ on $\Theta$ $a.s. - P$ so that $\left|\frac{\nabla_\theta\lambda(\theta)}{\lambda(\theta)}\{1 + \lambda(\theta)x(\theta)u(\lambda(\theta)x(\theta))\}\right| \leqslant |x(\theta)\nabla_\theta\lambda(\theta)|$, we then have

$$\begin{aligned}\sup_{\theta\in\Theta}|\nabla_\theta \ln f_t^*(Y_t, \theta)| \;\leqslant\; & 2\sup_{\theta\in\Theta}|\nabla_\theta\lambda(\theta)| + \sup_{\theta\in\Theta}|\lambda(\theta)|M_0|\nabla_\theta q_t(\theta)| + \\ & +C_1\sup_{\theta\in\Theta}\left|\tfrac{f_t^0(q_t(\theta))\nabla_\theta q_t(\theta)}{1-\mathbb{I}(q_t(\theta)-Y_t)-F_t^0(q_t(\theta))}\right|, a.s. - P,\end{aligned} \tag{16}$$

where $C_1 \equiv \sup_{x\in[0,\sup_{\theta\in\Theta}\lambda(\theta)]}|\frac{x}{1-\exp(-x)}| < \infty$. We have $\sup_{t\geqslant1}\sup_{\theta\in\Theta} F_t^0(q_t(\theta)) \in (a, b)$, $a > 0$, $b < 1$, so $C_2 \equiv \sup_{t\geqslant1}\sup_{y\in\mathbb{R}}\sup_{\theta\in\Theta}(|1 - \mathbb{I}(q_t(\theta) - y) - F_t^0(q_t(\theta))|^{-1})$ $< \infty$, and the last term of the above inequality is bounded above by $C_1C_2M_0\sup_{\theta\in\Theta}$ $|\nabla_\theta q_t(\theta)|$. From (A8)(i) $E[\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)|] < \infty$, so $E[\sup_{\theta\in\Theta}|\nabla_\theta \ln f_t^*(Y_t, \theta)|] < \infty$ for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, which shows that equation (25) in Komunjer (2005) holds; together with (A7) and $E[\sup_{\theta\in\Theta}|\ln f_t^*(Y_t, \theta)|^{r+\epsilon}] < \infty$ for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, this establishes condition (iv) of Theorem 2.1 and completes the proof of consistency.

STEP2: We now show that the MLE $\tilde{\theta}_T^*$ is asymptotically normal by checking that conditions (i)-(v) of Theorem 7.2 in Newey and McFadden (1994) applied to $\nabla_\theta L_T(\theta)$ hold. We first establish the asymptotic first order condition $\sqrt{T}\nabla_\theta L_T(\tilde{\theta}_T^*) \xrightarrow{p} 0$ by following the same steps as in the proof of Lemma A1 in Komunjer (2005): for every

$j = 1, \ldots, k$, let $\tilde{G}^*_{T,j}(h)$ be the right-derivative of $\tilde{L}^*_{T,j}(h) \equiv T^{-1} \sum_{t=1}^T \ln f_t^*(Y_t, \tilde{\theta}_T^* + he_j)$, where $\{e_j\}_{j=1}^k$ is the standard basis of $\mathbb{R}^k$, and $h \in \mathbb{R}$ is such that for all $j = 1, \ldots, k$, $\tilde{\theta}_T^* + he_j \in \Theta$. Since for every $j = 1, \ldots, k$, $\tilde{L}^*_{T,j}(0) = L_T(\hat{\theta}_T)$ so that the functions $h \mapsto \tilde{L}^*_{T,j}(h)$ achieve their maximum at $h = 0$, we have, for $\varepsilon > 0$, $\tilde{G}^*_{T,j}(\varepsilon) \leqslant \tilde{G}^*_{T,j}(0) \leqslant \tilde{G}^*_{T,j}(-\varepsilon)$, with $\tilde{G}^*_{T,j}(\varepsilon) \leqslant 0$ and $\tilde{G}^*_{T,j}(-\varepsilon) \geqslant 0$. Therefore $|\tilde{G}^*_{T,j}(0)| \leqslant \tilde{G}^*_{T,j}(-\varepsilon) - \tilde{G}^*_{T,j}(\varepsilon)$. By taking the limit of this inequality as $\varepsilon \to 0$, we get $|\tilde{G}^*_{T,j}(0)| \leqslant T^{-1} \sum\limits_{t=1}^T [1 + 2C_1] \left[ \left| \frac{\partial \lambda(\tilde{\theta}_T^*)}{\partial \theta_j} \right| + \left| \lambda(\tilde{\theta}_T^*) f_t^0(q_t(\tilde{\theta}_T^*)) \frac{\partial q_t(\tilde{\theta}_T^*)}{\partial \theta_j} \right| \right] \mathbb{1}\{q_t(\tilde{\theta}_T^*) = Y_t\}$. Hence

$$\begin{aligned} &P\left(\sqrt{T} |\nabla_\theta L_T(\tilde{\theta}_T^*)| > \varepsilon\right) \leqslant P\left(\sqrt{T} \max_{1 \leqslant j \leqslant k} |\tilde{G}^*_{T,j}(0)| > \varepsilon\right) \\ &\leqslant P\left( \sum_{t=1}^T \left[ \left| \frac{\partial \lambda(\tilde{\theta}_T^*)}{\partial \theta_j} \right| + \left| \lambda(\tilde{\theta}_T^*) f_t^0(q_t(\tilde{\theta}_T^*)) \frac{\partial q_t(\tilde{\theta}_T^*)}{\partial \theta_j} \right| \right] \mathbb{1}\{q_t(\tilde{\theta}_T^*) = Y_t\} > \varepsilon \sqrt{T} (1 + 2C_1)^{-1} \right) \end{aligned}$$

That $P(\mathbb{1}\{q_t(\tilde{\theta}_T^*) = Y_t\} \neq 0) = 0$ and $E[\left| \frac{\partial \lambda(\tilde{\theta}_T^*)}{\partial \theta_j} \right| + \left| \lambda(\tilde{\theta}_T^*) f_t^0(q_t(\tilde{\theta}_T^*)) \frac{\partial q_t(\tilde{\theta}_T^*)}{\partial \theta_j} \right|]$ is bounded then ensure that $\lim_{T \to \infty} P\left(\sqrt{T} |\nabla_\theta L_T(\tilde{\theta}_T^*)| > \varepsilon\right) = 0$. Condition (i) of Theorem 7.2 follows from the correct specification of $f_t(\cdot)$ (see (iii) in Theorem 4). By (A6), $\theta_0$ is an interior point of $\Theta$ so that condition (iii) of Theorem 7.2 holds. We now check the differentiability of $E[\nabla_\theta L_T(\theta)]$ and the nonsingularity condition (ii) of Theorem 7.2. We have $E[\nabla_\theta L_T(\theta)] = T^{-1} \sum_{t=1}^T E[\nabla_\theta \ln f_t^*(Y_t, \theta)]$; using (14) the latter is easily shown to be differentiable at any $\theta \in \mathring{\Theta}$. We now show that $\nabla_\theta E[\nabla'_\theta L_T(\theta_0)] = T^{-1} \sum_{t=1}^T E[\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta_0)]$ and that the latter is nonsingular. We have $du(z)/dz =$

$u(z) + [u(z)]^2$, hence, for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$,

$$\begin{aligned}
&\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta) \\
&= \tfrac{\Delta_{\theta\theta}\lambda(\theta)}{\lambda(\theta)} - \tfrac{\nabla_\theta\lambda(\theta)\nabla_\theta\lambda(\theta)'}{[\lambda(\theta)]^2} + \Delta_{\theta\theta}\lambda(\theta)[F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbb{I}(q_t(\theta) - Y_t) - \alpha] \\
&+ 2\nabla_\theta\lambda(\theta)\nabla_\theta q_t(\theta)' \left\{ f_t^0(q_t(\theta))[\alpha - \mathbb{I}(q_t(\theta) - Y_t)] + \delta(q_t(\theta) - Y_t)[F_t^0(Y_t) - F_t^0(q_t(\theta))] \right\} \\
&+ \lambda(\theta)\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' \Big\{ \tfrac{df_t^0(q_t(\theta))}{dq}[\alpha - \mathbb{I}(q_t(\theta) - Y_t)] - 2f_t^0(q_t(\theta))\delta(q_t(\theta) - Y_t) \\
&+ [F_t^0(Y_t) - F_t^0(q_t(\theta))]\tfrac{d\delta(q_t(\theta)-Y_t)}{dq} \Big\} + u(\lambda(\theta)x(\theta))\Delta_{\theta\theta}(\lambda(\theta)x(\theta)) \\
&+ \lambda(\theta)\Delta_{\theta\theta}q_t(\theta) \left\{ f_t^0(q_t(\theta))[\alpha - \mathbb{I}(q_t(\theta) - Y_t)] + [F_t^0(Y_t) - F_t^0(q_t(\theta))]\delta(q_t(\theta) - Y_t) \right\} \\
&+ \left[ u(\lambda(\theta)x(\theta)) + (u(\lambda(\theta)x(\theta)))^2 \right] (\nabla_\theta(\lambda(\theta)x(\theta))) (\nabla_\theta(\lambda(\theta)x(\theta)))', \qquad (17)
\end{aligned}$$

where $\Delta_{\theta\theta}(\lambda(\theta)x(\theta)) = \Delta_{\theta\theta}\lambda(\theta)x(\theta) + 2\nabla_\theta\lambda(\theta)\nabla_\theta x(\theta)' + \lambda(\theta)\Delta_{\theta\theta}x(\theta)$ and

$$\begin{aligned}
\Delta_{\theta\theta}x(\theta) = &\Big\{ \tfrac{df_t^0(q_t(\theta))}{dq}[\alpha - \mathbb{I}(q_t(\theta) - Y_t)] - 2f_t^0(q_t(\theta))\delta(q_t(\theta) - Y_t) \\
&+ \tfrac{d\delta(q_t(\theta)-Y_t)}{dq}[\alpha - F_t^0(q_t(\theta))] \Big\} \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' \\
&+ \left\{ f_t^0(q_t(\theta))[\alpha - \mathbb{I}(q_t(\theta) - Y_t)] + \delta(q_t(\theta) - Y_t)[\alpha - F_t^0(q_t(\theta))] \right\} \Delta_{\theta\theta}q_t(\theta).
\end{aligned}$$

Now, note that $u(z) + [u(z)]^2 = 1/z^2 - 1/12 + o(1)$ in the neighborhood of 0 so that

$$\begin{aligned}
&[u(\lambda(\theta)x(\theta)) + u(\lambda(\theta)x(\theta))^2]\nabla_\theta(\lambda(\theta)x(\theta))\nabla_\theta(\lambda(\theta)x(\theta))' \\
&= \tfrac{\nabla_\theta\lambda(\theta)\nabla_\theta\lambda(\theta)'}{[\lambda(\theta)]^2} + 2\tfrac{\nabla_\theta\lambda(\theta)\nabla_\theta x(\theta)'}{\lambda(\theta)x(\theta)} \\
&+ \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'\{f_t^0(q_t(\theta))\tfrac{[\alpha - \mathbb{I}(q_t(\theta)-Y_t)]}{x(\theta)} + \delta(q_t(\theta) - Y_t)\tfrac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)}\}^2 + o_p(1), \qquad (18)
\end{aligned}$$

in the neighborhood of $\theta_0$. Similarly,

$$\begin{aligned}
&u(\lambda(\theta)x(\theta))\Delta_{\theta\theta}(\lambda(\theta)x(\theta)) \\
&= -\tfrac{\Delta_{\theta\theta}\lambda(\theta)}{\lambda(\theta)} - \tfrac{1}{2}\Delta_{\theta\theta}\lambda(\theta)x(\theta) - 2\tfrac{\nabla_\theta\lambda(\theta)\nabla_\theta x(\theta)'}{\lambda(\theta)x(\theta)} \\
&- \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' \left\{ \tfrac{df_t^0(q_t(\theta))}{dq}\tfrac{[\alpha - \mathbb{I}(q_t(\theta)-Y_t)]}{x(\theta)} - 2\tfrac{f_t^0(q_t(\theta))\delta(q_t(\theta)-Y_t)}{x(\theta)} + \tfrac{d\delta(q_t(\theta)-Y_t)}{dq}\tfrac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} \\
&- \Delta_{\theta\theta}q_t(\theta) \left\{ f_t^0(q_t(\theta))\tfrac{[\alpha - \mathbb{I}(q_t(\theta)-Y_t)]}{x(\theta)} + \delta(q_t(\theta) - Y_t)\tfrac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} + o_p(1), \qquad (19)
\end{aligned}$$

in the neighborhood of $\theta_0$. Combining (17) with (18) and (19), we then get that, for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$,

$$\begin{aligned}
&\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta) \\
&= \Delta_{\theta\theta}\lambda(\theta) \left\{[F_t^0(Y_t) - F_t^0(q_t(\theta))][1\!\!1(q_t(\theta) - Y_t) - \alpha] - \tfrac{1}{2}x(\theta)\right\} \\
&+ \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' \left\{ f_t^0(q_t(\theta)) \tfrac{[\alpha - 1\!\!1(q_t(\theta) - Y_t)]}{x(\theta)} + \delta(q_t(\theta) - Y_t) \tfrac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\}^2 \\
&- \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' \left\{ \tfrac{df_t^0(q_t(\theta))}{dq} \tfrac{[\alpha - 1\!\!1(q_t(\theta) - Y_t)]}{x(\theta)} - 2\tfrac{f_t^0(q_t(\theta))\delta(q_t(\theta) - Y_t)}{x(\theta)} + \tfrac{d\delta(q_t(\theta) - Y_t)}{dq} \tfrac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} \\
&- \Delta_{\theta\theta} q_t(\theta) \left\{ f_t^0(q_t(\theta)) \tfrac{[\alpha - 1\!\!1(q_t(\theta) - Y_t)]}{x(\theta)} + \delta(q_t(\theta) - Y_t) \tfrac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} + o_p(1), \qquad (20)
\end{aligned}$$

in the neigborhood of $\theta_0$. Using $\alpha = F_t^0(q_t(\theta_0))$ and $x(\theta_0) = -\alpha(1-\alpha)$ we have $|\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta_0)| \leqslant \frac{5}{2}|\Delta_{\theta\theta}\lambda(\theta_0)| + |\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'|(\frac{M_0^2}{[\alpha(1-\alpha)]^2} + \frac{M_1}{\alpha(1-\alpha)}) + |\Delta_{\theta\theta} q_t(\theta_0)| \times \frac{M_0}{\alpha(1-\alpha)} + o_p(1)$, with $|\Delta_{\theta\theta}\lambda(\theta_0)| < \infty$. From (A8)(i) $E[|\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'|] < \infty$ and $E[|\Delta_{\theta\theta} q_t(\theta_0)|] < \infty$, which shows that the expectation of the right hand side of the above inequality is finite; hence $\nabla_\theta E[\nabla'_\theta \ln f_t^*(Y_t, \theta_0)] = E[\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta_0)]$ for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$ and so $\nabla_\theta E[\nabla'_\theta L_T(\theta_0)] = T^{-1} \sum_{t=1}^T E[\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta_0)]$ as desired. Now consider $E[\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta_0)]$; for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, we have

$$\begin{aligned}
&E\left(\Delta_{\theta\theta}\lambda(\theta_0) \left\{[F_t^0(Y_t) - F_t^0(q_t(\theta_0))][1\!\!1(q_t(\theta_0) - Y_t) - \alpha] - \tfrac{1}{2}x(\theta_0)\right\}\right) \\
&\quad = \Delta_{\theta\theta}\lambda(\theta_0) \left[E\left([F_t^0(Y_t) - \alpha][1\!\!1(q_t(\theta_0) - Y_t) - \alpha]\right) + \tfrac{1}{2}\alpha(1-\alpha)\right] \\
&\quad = \Delta_{\theta\theta}\lambda(\theta_0) \left[-\tfrac{1}{2}\alpha(1-\alpha) + \tfrac{1}{2}\alpha(1-\alpha)\right] = 0,
\end{aligned}$$

since

$$\begin{aligned}
&E_t\left([F_t^0(Y_t) - \alpha][1\!\!1(q_t(\theta_0) - Y_t) - \alpha]\right) \\
&= (1-\alpha) \int_{-\infty}^{q_t(\theta_0)} [F_t^0(y) - \alpha] f_t^0(y) dy - \alpha \int_{q_t(\theta_0)}^{+\infty} [F_t^0(y) - \alpha] f_t^0(y) dy \\
&= (1-\alpha) \left[\tfrac{1}{2}[F_t^0(y) - \alpha]^2\right]_{-\infty}^{q_t(\theta_0)} - \alpha \left[\tfrac{1}{2}[F_t^0(y) - \alpha]^2\right]_{q_t(\theta_0)}^{+\infty} = -\tfrac{1}{2}\alpha(1-\alpha).
\end{aligned}$$

In addition, $\alpha = F_t^0(q_t(\theta_0))$ and $x(\theta_0) = -\alpha(1-\alpha)$ so

$$E\left(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\left\{f_t^0(q_t(\theta_0))\frac{[\alpha-\mathbb{I}(q_t(\theta_0)-Y_t)]}{x(\theta_0)}+\delta(q_t(\theta_0)-Y_t)\frac{[\alpha-F_t^0(q_t(\theta_0))]}{x(\theta_0)}\right\}^2\right)$$
$$=E\left(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'E_t\left\{\frac{[f_t^0(q_t(\theta_0))]^2[\alpha-\mathbb{I}(q_t(\theta_0)-Y_t)]^2}{\alpha^2(1-\alpha)^2}\right\}\right)$$
$$=E\left(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)}\right),$$

where the last equality uses $E_t\left([\mathbb{I}(q_t(\theta_0)-Y_t)-\alpha]^2\right)=\alpha(1-\alpha)$, $a.s.-P$. Similarly,

$$E\left(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\left\{\frac{df_t^0(q_t(\theta_0))}{dq}\frac{[\alpha-\mathbb{I}(q_t(\theta_0)-Y_t)]}{x(\theta_0)}-2\frac{f_t^0(q_t(\theta_0))\delta(q_t(\theta_0)-Y_t)}{x(\theta_0)}+\frac{d\delta(q_t(\theta_0)-Y_t)}{dq}\frac{[\alpha-F_t^0(q_t(\theta_0))]}{x(\theta_0)}\right\}\right)$$
$$=E\left(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'E_t\left\{\frac{df_t^0(q_t(\theta_0))}{dq}\frac{[\mathbb{I}(q_t(\theta_0)-Y_t)-\alpha]}{\alpha(1-\alpha)}+2\frac{f_t^0(q_t(\theta_0))\delta(q_t(\theta_0)-Y_t)}{\alpha(1-\alpha)}\right\}\right)$$
$$=2E\left(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)}\right),$$

where the last equality uses $E_t\left(\mathbb{I}(q_t(\theta_0)-Y_t)-\alpha\right)=0$, $a.s.-P$ and $E_t(\delta(q_t(\theta_0)-Y_t))=f_t^0(q_t(\theta_0))$, $a.s.-P$. Finally, by the same reasoning $E(\Delta_{\theta\theta}q_t(\theta_0)\{f_t^0(q_t(\theta_0))\times\frac{[\alpha-\mathbb{I}(q_t(\theta_0)-Y_t)]}{x(\theta_0)}+\delta(q_t(\theta_0)-Y_t)\frac{[\alpha-F_t^0(q_t(\theta_0))]}{x(\theta_0)}\})=0$. Combining the above results then yields, by (20), $E[\Delta_{\theta\theta}\ln f_t^*(Y_t,\theta_0)]=-E(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)})$, for all $t$, $1\leqslant t\leqslant T, T\geqslant 1$. So for any $\chi\in\mathbb{R}^k$, $\chi'\nabla_\theta E[\nabla_\theta' L_T(\theta_0)]\chi=-T^{-1}\sum_{t=1}^T E(|\nabla_\theta q_t(\theta_0)'\chi|^2\times\frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)})\leqslant 0$, with equality if and only if $\chi=0$. Hence $\nabla_\theta E[\nabla_\theta' L_T(\theta_0)]$ is negative definite (therefore nonsingular). We now check condition (iv) of Theorem 7.2 by using a CLT for $\alpha$-mixing sequences (e.g. Theorem 5.20 in White, 2001, p.130). By (A7), for any $\theta\in\mathring{\Theta}$, the sequence $\{\nabla_\theta\ln f_t^*(Y_t,\theta)\}$ is strong mixing (i.e. $\alpha$-mixing) with $\alpha$ of size $-r/(r-2)$, $r>2$ (see, e.g., Theorem 3.49 in White, 2001, p.50). Moreover, using (14) and (A2), $E[\nabla_\theta\ln f_t^*(Y_t,\theta_0)]=0$ and using (A8)(i), $E[|\nabla_\theta\ln f_t^*(Y_t,\theta_0)|^r]\leqslant$

$\{M_0/[\alpha(1-\alpha)]\}^r E[\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)|^r] < \infty$, for all $t$, $1 \leqslant t \leqslant T, T \geqslant 1$. Now,

$$\begin{aligned}
&\mathrm{Var}\left(T^{-1}\sum_{t=1}^{T}\nabla_\theta \ln f_t^*(Y_t,\theta_0)\right)\\
&= E\left(T^{-1}\sum_{t=1}^{T}\nabla_\theta \ln f_t^*(Y_t,\theta_0)\nabla_\theta \ln f_t^*(Y_t,\theta_0)'\right)\\
&= E\left(T^{-1}\sum_{t=1}^{T}\tfrac{[f_t^0(q_t(\theta_0))]^2[\mathbb{1}(q_t(\theta_0)-Y_t)-\alpha]^2}{[\alpha(1-\alpha)]^2}\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\right) = V_T^0
\end{aligned}$$

where the first equality uses $E_t\left(\nabla_\theta \ln f_t^*(Y_t,\theta_0)\right) = 0$, $a.s. - P$, implied by (A2), and the last equality uses $E_t\left([\mathbb{1}(q_t(\theta_0)-Y_t)-\alpha]^2\right) = \alpha(1-\alpha)$, $a.s. - P$. Applying Theorem 5.20 in White (2001) we then have $(V_T^0)^{-1/2}\sqrt{T}\nabla_\theta L_T(\theta_0) \xrightarrow{d} \mathcal{N}(0,\mathrm{Id})$ with $V_T^0$ as defined in Theorem 3. Finally, we check the stochastic equicontinuity condition (v) of Theorem 7.2 by veryfing that all the assumptions in Theorem 7.3 in Newey and McFadden (1994) hold. (The main reason for using Theorem 7.3 is that it does not put any restrictions on the dependence structure of $\{(Y_t, W_t')'\}$.) For any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$, let $r_t(\theta) = |\nabla_\theta \ln f_t^*(Y_t,\theta) - \nabla_\theta \ln f_t^*(Y_t,\theta_0) - \Delta_{\theta\theta}\ln f_t^*(Y_t,\theta)'(\theta-\theta_0)|/|\theta-\theta_0|$, for $\theta \in \mathring{\Theta}$. Using $u(z) = -1/z - 1/2 + o(1)$ in the neigborhood of 0 and $\lambda(\theta)x(\theta) = o_p(|\theta-\theta_0|)$ in the neigborhood of $\theta_0$, we have, from (14), (15) and (17), $r_t(\theta) \leqslant r_t^{(1)}(\theta) + r_t^{(2)}(\theta) + r_t^{(3)}(\theta) + o_p(1)$, where

$$\begin{aligned}
r_t^{(1)}(\theta) &= \left|[F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbb{1}(q_t(\theta)-Y_t)-\alpha] - \tfrac{x(\theta)}{2}\right| \tfrac{|\nabla_\theta\lambda(\theta)-\Delta_{\theta\theta}\lambda(\theta)'(\theta-\theta_0)|}{|\theta-\theta_0|}\\
r_t^{(2)}(\theta) &= \left|\tfrac{f_t^0(q_t(\theta))}{2}[\mathbb{1}(q_t(\theta)-Y_t)-\alpha] + \delta(q_t(\theta)-Y_t)[\tfrac{F_t^0(q_t(\theta))-2F_t^0(Y_t)+\alpha}{2}]\right| \tfrac{|\lambda(\theta)\nabla_\theta q_t(\theta)|}{|\theta-\theta_0|}\\
r_t^{(3)}(\theta) &= \left|\tfrac{\nabla_\theta x(\theta)}{x(\theta)} - \tfrac{\nabla_\theta x(\theta_0)}{x(\theta_0)} - \tfrac{\Delta_{\theta\theta}x(\theta)'(\theta-\theta_0)}{x(\theta)} + \tfrac{\nabla_\theta x(\theta)\nabla_\theta x(\theta)'(\theta-\theta_0)}{[x(\theta)]^2}\right| /|\theta-\theta_0|.
\end{aligned}$$

With probability one, $r_t^{(1)}(\theta) \leqslant 2|\nabla_\theta\lambda(\theta) - \Delta_{\theta\theta}\lambda(\theta)'(\theta-\theta_0)|/|\theta-\theta_0|$ for any $\theta \in \mathring{\Theta}$. Given that $\lambda(\cdot)$ is twice continously differentiable on $\mathbb{R}^k$, with probability one $r_t^{(1)}(\theta) \to 0$ as $\theta \to \theta_0$ and there exists $\varepsilon_1 > 0$ such that $E(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} r_t^{(1)}(\theta)) <$

$\infty$. Now, note that $|f_t^0(q_t(\theta))[\mathbb{I}(q_t(\theta) - Y_t) - \alpha]| \leqslant M_0$ for any $\theta \in \mathring{\Theta}$, so

$$\begin{aligned} r_t^{(2)}(\theta) &\leqslant \tfrac{1}{2}\Big\{M_0 + \delta(q_t(\theta) - Y_t)[F_t^0(q_t(\theta)) - 2F_t^0(Y_t) + \alpha]\Big\}\tfrac{|\lambda(\theta)\nabla_\theta q_t(\theta)|}{|\theta-\theta_0|} \\ &\leqslant \tfrac{1}{2}\Big\{M_0 + \delta(q_t(\theta) - Y_t)[F_t^0(q_t(\theta)) - 2F_t^0(Y_t) + \alpha]\Big\}|\nabla_\theta \lambda(\theta_c)| \cdot |\nabla_\theta q_t(\theta)| \end{aligned}$$

for some $\theta_c \equiv c\theta_0 + (1-c)\theta$ with $c \in (0,1)$. Hence, using the fact that $\nabla_\theta \lambda(\cdot)$ is continuous on $\mathbb{R}^k$, that $\nabla_\theta \lambda(\theta_0) = 0$ and that $\delta(q_t(\theta_0) - Y_t)[F_t^0(q_t(\theta_0)) - 2F_t^0(Y_t) + \alpha] = 0$, with probability one $r_t^{(2)}(\theta) \to 0$ as $\theta \to \theta_0$. Moreover, for some $\theta_d \equiv d\theta_0 + (1-d)\theta$, $d \in (0,1)$,

$$\begin{aligned} &E\left(\sup\nolimits_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} r_t^{(2)}(\theta)\right) \\ &\leqslant E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} \{\tfrac{M_0}{2} + E_t(\delta(q_t(\theta) - Y_t)|\tfrac{F_t^0(q_t(\theta))}{2} - F_t^0(Y_t) + \tfrac{\alpha}{2}|)\}|\nabla_\theta \lambda(\theta_c)||\nabla_\theta q_t(\theta)|\right) \\ &\leqslant \tfrac{M_0}{2} \sup\nolimits_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} |\nabla_\theta \lambda(\theta_c)| \left[E\left(\sup\nolimits_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} |\nabla_\theta q_t(\theta)|\right)\right. \\ &\quad \left. + M_0 E\left(\sup\nolimits_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} |\nabla_\theta q_t(\theta)||\nabla_\theta q_t(\theta_d)|\right)\right] < \infty, \end{aligned}$$

where the last inequality uses the continuity of $\nabla_\theta \lambda(\cdot)$ on $\mathbb{R}^k$, (A8)(i) and the Cauchy-Schwarz inequality. Finally, let $r_x(\theta) = [x(\theta_0) - x(\theta) - \nabla_\theta x(\theta)'(\theta_0 - \theta)]/|\theta_0 - \theta|$ and $R_x(\theta) = [\nabla_\theta x(\theta_0) - \nabla_\theta x(\theta) - \Delta_{\theta\theta} x(\theta)'(\theta_0 - \theta)]/|\theta_0 - \theta|$ and note that with probability one $\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} |r_x(\theta)| \to 0$ and $\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} |R_x(\theta)| \to 0$ as $\theta \to \theta_0$. This implies that with probability one $r_t^{(3)}(\theta) \to 0$ as $\theta \to \theta_0$. Moreover

$$\begin{aligned} &E\left(\sup\nolimits_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} r_t^{(3)}(\theta)\right) \\ &\leqslant E\left(\sup\nolimits_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}[|r_x(\theta)| + |R_x(\theta)|]/|x(\theta)|\right) \\ &\leqslant E\left(\sup\nolimits_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}[1/|x(\theta)|]\left(\sup\nolimits_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} |r_x(\theta)| + \sup\nolimits_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} |R_x(\theta)|\right)\right) < \infty \end{aligned}$$

where the last inequality uses $\sup_{t\geqslant 1}\sup_{\theta\in\Theta} F_t^0(q_t(\theta)) \in (a,b)$ with $a > 0$ and $b < 1$, so $C_3 \equiv \sup_{t\geqslant 1}\sup_{y\in\mathbb{R}}\sup_{\theta\in\Theta}\left(|[\mathbb{I}(q_t(\theta) - Y_t) - \alpha][1 - \mathbb{I}(q_t(\theta) - y) - F_t^0(q_t(\theta))]|^{-1}\right) <$

$\infty$. Combining previous results then gives that with probability one $r_t(\theta) \to 0$ as $\theta \to \theta_0$ and that $E\left(\sup_{\theta \in \mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} r_t(\theta)\right) < \infty$. It remains to be shown that for all $\theta$ in a neighborhood of $\theta_0$ we have $T^{-1}\sum_{t=1}^T \Delta_{\theta\theta} \ln f_t^*(Y_t,\theta) \xrightarrow{p} \nabla_\theta E[\nabla'_\theta L_T(\theta)]$. By (A7), for any $\theta \in \mathring{\Theta}$, the sequence $\{\Delta_{\theta\theta} \ln f_t^*(Y_t,\theta)\}$ is strong mixing (i.e. $\alpha$-mixing) with $\alpha$ of size $-r/(r-2)$, $r > 2$ (see, e.g. Theorem 3.49 in White, 2001, p.50). Now, given $\theta \in \mathring{\Theta}$, there exists $\theta_a = a\theta_0 + (1-a)\theta$, $a \in (0,1)$, such that for any $\eta > 0$, we have $P(\delta(q_t(\theta)-Y_t)|\alpha - F_t^0(q_t(\theta))| > \eta) \leqslant E(|\alpha - F_t^0(q_t(\theta))| f_t^0(q_t(\theta)))/\eta \leqslant |\theta-\theta_0| E(|\nabla_\theta q_t(\theta_a)| f_t^0(q_t(\theta_a)) f_t^0(q_t(\theta)))/\eta \leqslant |\theta-\theta_0| M_0^2 E[\sup_{\theta\in\Theta} |\nabla_\theta q_t(\theta)|]/\eta$, so that in a neighborhood of $\theta_0$, $\delta(q_t(\theta)-Y_t)\,|\alpha - F_t^0(q_t(\theta))| = o_p(1)$. Similarly, $P([d\delta(q_t(\theta)-Y_t)/dq]|\alpha - F_t^0(q_t(\theta))| > \eta) \leqslant E(|\alpha - F_t^0(q_t(\theta))| df_t^0(q_t(\theta))/dq)/\eta \leqslant |\theta-\theta_0| M_0 M_1 \times E[\sup_{\theta\in\Theta} |\nabla_\theta q_t(\theta)|]/\eta$, where the first inequality uses the fact that $E_t(d\delta(q_t(\theta)-Y_t)/dq) = df_t^0(q_t(\theta))/dq$, $a.s.-P$. From (20) we have that for any $t$, $1 \leqslant t \leqslant T, T \geqslant 1$,

$$
\begin{aligned}
&\Delta_{\theta\theta} \ln f_t^*(Y_t,\theta) \\
&= \Delta_{\theta\theta}\lambda(\theta)\left\{[F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbb{1}(q_t(\theta)-Y_t) - \alpha] - \tfrac{1}{2}x(\theta)\right\} \\
&+ \tfrac{\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'}{[x(\theta)]^2}\Big\{\big(f_t^0(q_t(\theta))[\alpha - \mathbb{1}(q_t(\theta)-Y_t)]\big)^2 + \big(\delta(q_t(\theta)-Y_t)[\alpha - F_t^0(q_t(\theta))]\big)^2 \\
&- x(\theta)\tfrac{df_t^0(q_t(\theta))}{dq}[\alpha - \mathbb{1}(q_t(\theta)-Y_t)] + x(\theta)\tfrac{d\delta(q_t(\theta)-Y_t)}{dq}[\alpha - F_t^0(q_t(\theta))]\Big\} \\
&- \tfrac{\Delta_{\theta\theta} q_t(\theta)}{x(\theta)}\left\{f_t^0(q_t(\theta))[\alpha - \mathbb{1}(q_t(\theta)-Y_t)] + \delta(q_t(\theta)-Y_t)[\alpha - F_t^0(q_t(\theta))]\right\} + o_p(1),
\end{aligned}
$$

in a neighborhood of $\theta_0$, which then gives

$$
\begin{aligned}
&\Delta_{\theta\theta} \ln f_t^*(Y_t,\theta) \\
&= \Delta_{\theta\theta}\lambda(\theta)\left\{[F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbb{1}(q_t(\theta)-Y_t) - \alpha] - \tfrac{1}{2}x(\theta)\right\} \\
&+ \tfrac{\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'}{[x(\theta)]^2}\Big\{\big(f_t^0(q_t(\theta))[\alpha - \mathbb{1}(q_t(\theta)-Y_t)]\big)^2 - x(\theta)\tfrac{df_t^0(q_t(\theta))}{dq}[\alpha - \mathbb{1}(q_t(\theta)-Y_t)]\Big\} \\
&- \tfrac{\Delta_{\theta\theta} q_t(\theta)}{x(\theta)}\left\{f_t^0(q_t(\theta))[\alpha - \mathbb{1}(q_t(\theta)-Y_t)]\right\} + o_p(1).
\end{aligned}
$$

Hence, for a given $\varepsilon > 0$, there is a positive constant $n_{r,\varepsilon}$ such that $|\Delta_{\theta\theta} \ln f_t^*(Y_t,\theta)|^{r+\varepsilon} \leqslant$

$n_{r,\varepsilon}\{|\Delta_{\theta\theta}\lambda(\theta)|^{r+\varepsilon}(5/2)^{r+\varepsilon}+|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|^{r+\varepsilon}C_3^{2(r+\varepsilon)}(M_0^2+M_1)^{r+\varepsilon}+|\Delta_{\theta\theta}q_t(\theta)|^{r+\varepsilon}\times C_3^{r+\varepsilon}M_0^{r+\varepsilon}\}+o_p(1)$, in a neigborhood of $\theta_0$, and so using (A8)(i) and the fact that $|\Delta_{\theta\theta}\lambda(\theta)|<\infty$ in a neighborhood of $\theta_0$, we have $E[|\Delta_{\theta\theta}\ln f_t^*(Y_t,\theta)|^{r+\varepsilon}]<\infty$. The weak LLN then follows from Corollary 3.48 in White (2001). This completes the proof of asymptotic normality of the MLE $\tilde{\theta}_T^*$. □

## References

Begun, J., W. Hall, W. Huang, and J. Wellner (1983): "Information and Asymptotic Efficiency in Parametric-Nonparametric Models," *Annals of Statistics*, 11, 432–452.

Bickel, P. J. (1982): "On Adaptive Estimation," *Annals of Statistics*, 10, 647–671.

Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1993): *Efficient and Adaptive Inference in Semiparametric Models*. Johns Hopkins University Press, Baltimore.

——— (2005): "Semiparametric Inference and Models," University of California, Berkeley.

Bickel, P. J., and J. Kwon (2001): "Inference for Semiparametric Models: Some Questions and an Answer," *Statistica Sinica*, 11, 863–960.

Bracewell, R. N. (2000): *The Fourier Transform and Its Applications*. McGraw–Hill, 3rd edn.

Drost, F. C., C. Klaassen, and B. J. Werker (1997): "Adaptive Estimation in Time-Series Models," *Annals of Statistics*, 25, 786–818.

Fabian, V., and J. Hannan (1982): "On Estimation and Adaptive Estimation for Locally Asymptotically Normal Families," *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 59, 459–478.

Gourieroux, C., and A. Monfort (1995): *Statistics and Econometric Models.* Cambridge University Press.

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

Greenwood, P. E., U. U. Müller, and W. Wefelmeyer (2004): "An Introduction to Effcient Estimation for Semiparametric Time Series," in *Parametric and Semiparametric Models with Applications to Reliablilty, Survival Analysis, and Quality of Life*, ed. by M. S. Mikulin, N. Balakrishnan, M. Mesbah, and N. Limnios, pp. 253–272. Birkhäuser, Basel.

Hallin, M., C. Vermandeley, and B. J. Werker (2004): "Semiparametrically Efficient Inference Based on Signs and Ranks for Median Restricted Models," Université Libre de Bruxelles and Tilburg University.

Hjort, N. L., and D. Pollard (1993): "Asymptotics for Minimizers of Convex Processes," Yale University.

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*, Berkeley. University of California Press.

Klaassen, C. A. J., and J. A. Wellner (1997): "Efficient Estimation in the Bivariate Normal Copula Model: Normal Margins Are Least Favourable," *Bernoulli*, 3, 55–77.

Knight, K. (1998): "Limiting Distributions for $L_1$-Regression Estimators under General Conditions," *Annals of Statistics*, 26, 755–770.

Koenker, R., and G. Bassett, Jr. (1978): "Regression Quantiles," *Econometrica*, 46(1), 33–50.

Koenker, R., and Q. Zhao (1996): "Conditional Quantile Estimation and Inference for ARCH Models," *Econometric Theory*, 12, 793–813.

Komunjer, I. (2005): "Quasi-Maximum Likelihood Estimation for Conditional Quantiles," *Journal of Econometrics*, 128, 137–164.

Komunjer, I., and Q. Vuong (2007): "Efficient Estimation in Dynamic Conditional Quantile Models," University of California and Penn State University.

Kreiss, J. P. (1987): "On Adaptive Estimation in Stationary ARMA Processes," *Annals of Statistics*, 15, 112–133.

Murphy, S. A., and A. W. V. der Vaart (2000): "On Profile Likelihood," *Journal of the American Statistical Association*, 95, 449–465.

Newey, W. K., and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, pp. 2113–2247. Elsevier Science.

Newey, W. K., and J. L. Powell (1990): "Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions," *Econometric Theory*, 6, 295–317.

Pfanzagl, J. (1976): "Investigating the Quantile of an Unknown Distribution," in *Contributions to Applied Statistics*, ed. by J. W. Zeigler. Birkhäuser, Basel.

POLLARD, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186–199.

SCHWARTZ, L. (1997): *Analyse*. Hermann, Paris.

SEVERINI, T. A., AND W. H. WONG (1992): "Profile Likelihood and Conditionally Parametric Models," *Annals of Statistics*, 20, 1768–1802.

STEIN, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium in Mathematical Statistics and Probability*, vol. 1, pp. 187–196, Berkeley. University of California Press.

WELLNER, J. A., C. A. J. KLAASSEN, AND Y. RITOV (2006): "Semiparametric Models: A Review of Progress Since BKRW (1993)," in *Frontiers in Statistics: In Honor of Peter J. Bickel's 65th Birthday*, ed. by J. Fan, and H. Koul, pp. 25–35. Imperial College Press, London.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

——— (2001): *Asymptotic Theory for Econometricians*. Academic Press, San Diego.

ZHAO, Q. (2001): "Asymptotically Efficient Median Regression in the Presence of Heteroskedasticity of Unknown Form," *Econometric Theory*, 17, 765–784.