Econometric Theory, **26**, 2010, 383–405.
doi:10.1017/S0266466609100038

SEMIPARAMETRIC EFFICIENCY BOUND IN TIME-SERIES MODELS FOR CONDITIONAL QUANTILES

IVANA KOMUNJER
University of California, San Diego

QUANG VUONG
Pennsylvania State University

We derive the semiparametric efficiency bound in dynamic models of conditional quantiles under a sole strong mixing assumption. We also provide an expression of Stein's (1956) least favorable parametric submodel. Our approach is as follows: First, we construct a fully parametric submodel of the semiparametric model defined by the conditional quantile restriction that contains the data generating process. We then compare the asymptotic covariance matrix of the MLE obtained in this submodel with those of the M-estimators for the conditional quantile parameter that are consistent and asymptotically normal. Finally, we show that the minimum asymptotic covariance matrix of this class of M-estimators equals the asymptotic covariance matrix of the parametric submodel MLE. Thus, (i) this parametric submodel is a least favorable one, and (ii) the expression of the semiparametric efficiency bound for the conditional quantile parameter follows.

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 latter have developed an information calculus framework based on local asymptotic normality (LAN) and regularity of parameters to derive Hajek-LeCam convolution theorems for regular estimators. Most of the literature has focused on independent and identically distributed (i.i.d) data as the study of sufficient conditions to satisfy the convolution theorem hypotheses has been largely successful in this setting. While dropping the i.i.d. 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-i.i.d. case has first focused on time-series models in which some finitely parameterized transformation of the data

Address correspondence to Ivana Komunjer, University of California San Diego, 9500 Gilman Drive MC 0508, La Jolla, CA 92093; e-mail: komunjer@ucsd.edu.

© Cambridge University Press, 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

leads back to the i.i.d. case. See 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). For example, this approach can 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 error sequence $\{U_t\}$ is independent of $\{W_t\}$ and i.i.d. with some absolutely continuous distribution function $H_0(\cdot)$ (continuous density $h_0(\cdot)$) such that $H_0^{-1}(\alpha) = 0$ for some $\alpha \in (0, 1)$. Here, θ is the parameter of interest and γ and h_0 are two nuisance parameters where the latter is infinite dimensional.

More generally—when no finite parameterizations leading back to the i.i.d. case are available—information bounds and efficient estimation have been extended to nonparametric stationary Markov chains, where the parameter of interest satisfies a conditional moment restriction. See Greenwood, Müller, and Wefelmeyer (2004) for a 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 θ 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.[1] In (1), this occurs when the sequence of errors $\{U_t\}$ is strong mixing. In this case, there exists no finitely parameterized transformation that would bring us back to the i.i.d. case as the time-dependence of the sequence $\{U_t\}$ introduces an additional nonparametric component. 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 exhibit 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 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 equals the minimum of a set of asymptotic covariance matrices of M- (or extremal) estimators (Huber, 1967) for the conditional quantile parameter θ that are consistent and asymptotically normal under strong mixing. Stein's (1956) argument then implies that the latter asymptotic covariance matrix is the semiparametric efficiency bound for θ, while the constructed fully parametric submodel is a least favorable one.[2] Our results provide what we believe to be the first formal proof of Ogaki's (1988) conjecture in the conditional quantile context that the infimum asymptotic covariance of our class of M-estimators is equal to the semiparametric efficiency bound.

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

Least favorable models do not always exist. Examples of 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 the same dimensionality as the parameter of interest is available, recent results by Murphy and van der Vaart (2000) on profile likelihoods (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 conveniently partitioned or variation-free parameterization (θ, η) with θ defined through a conditional moment restriction and η infinitely dimensional complicates the application of these results.

The paper is organized as follows. In Section 2 we introduce models for conditional quantiles. Section 3 discusses Stein's (1956) definition of parametric submodels and proposes a parametric submodel of the conditional quantile model. Section 4 studies a class of M-estimators for the parameters of such models that are consistent, asymptotically normally distributed with a minimum asymptotic covariance matrix that equals that of the parametric submodel MLE. As a consequence, the parametric submodel is a least favorable one, thereby providing the semiparametric efficiency bound. We relegate all the proofs to the Appendix.

2. SETUP

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 $\mathbb{R}^{(n+1)\mathbb{N}}$ with $n \in \mathbb{N}$ is the product space generated by taking a copy of $\mathbb{R}^{n+1}$ for each integer. 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 Y_t conditional on its own lags as well as past and present values of the n-vector X_t assumed strongly exogenous in the sense of Engle, Hendry, and Richard (1983). 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 strongly exogenous, we have $f_{\underline{Z}_T}(\underline{z}_T) = [\prod_{t=1}^T f_t^0(y_t)] f_{\underline{X}_T}(\underline{x}_T)$, where lowercase letters are used to denote the realizations of the corresponding random variables.[3] Throughout the paper we assume that for any $T \geqslant 1$ and every t, $1 \leqslant t \leqslant T$, $F_t^0(\cdot)$ is unknown and belongs to $\mathcal{F}$, which is the set of all absolutely continuous distribution functions on $\mathbb{R}$ with bounded densities that are continuously differentiable, and whose derivatives are bounded.[4]

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

In this paper we do not consider the conditional distribution $F_t^0(\cdot)$ in its entirety but 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 α-quantile of Y_t, $\mathcal{M} \equiv \{q_\alpha(W_t,\theta)\}$, parameterized by θ in Θ, where Θ is a compact subset of $\mathbb{R}^k$ with nonempty 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. Hereafter we make the following dependence assumption.

Assumption A1. The sequence $\{(Y_t, W_t')'\}$ is α-mixing with α of size $-r/(r-2)$, with $r > 2$.

Assumption A1 is fairly weak. We allow the sequence $\{(Y_t, W_t')'\}$ to be nonstationary and our strong mixing (i.e., α-mixing) assumption allows for a wide variety of dependence structures (White, 2001). Hereafter we restrict our attention to conditional quantile models $\mathcal{M}$ in which the following set of conditions is satisfied.

Assumption A2. The model $\mathcal{M}$ is identified on Θ, 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 $t \geqslant 1$, if and only if $\theta_1 = \theta_2$; (ii) for every $t \geqslant 1$, the function $q_\alpha(W_t,\cdot) : \Theta \to \mathbb{R}$ is twice continuously differentiable on Θ a.s. $-P$; (iii) for every $t \geqslant 1$, 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 Assumption A2 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_α, 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 a different nature than Assumptions A1–A2, is that the model $\mathcal{M}$ is correctly specified, so that there exists some true parameter value θ_0 such that $F_t^0(q_\alpha(W_t,\theta_0)) = \alpha$, for every $t \geqslant 1$.

Assumption A3. Given $\alpha \in (0,1)$, there exists $\theta_0 \in \mathring{\Theta}$ such that $\mathrm{E}[\mathbb{1}(q_\alpha(W_t, \theta_0) - Y_t)|\mathcal{W}_t] = \alpha, a.s. - P$, for every $t \geqslant 1$.

The starting point of all the estimators considered in this paper is the correct specification assumption: it defines a conditional moment restriction that identifies the conditional quantile parameter θ_0. The question we now pose is: what fully parametric estimators of θ_0 can we construct that satisfy the conditional quantile restriction in Assumption A3?

3. PARAMETRIC ESTIMATORS FOR CONDITIONAL QUANTILES

3.1. Stein's (1956) Approach: An Example

We illustrate Stein's (1956) approach using the example in Section 1. 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 $\underline{Y}_{t-1} \equiv (Y_1,\ldots,Y_{t-1})$, and the conditional quantile of Y_t is assumed to be linear $\theta' W_t$, with $W_t \equiv (1, Y_{t-1})'$.

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

We are concerned with the situations in which the true density of U_t in (1) is entirely unknown and the errors U_t are not necessarily i.i.d. Given $T \geqslant 1$, let $\underline{h}_T^0$ denote the joint density of $\underline{U}_T \equiv (U_1, \ldots, U_T)$. In the i.i.d. case, $\underline{h}_T^0$ is uniquely determined by the marginal density h_0 of U_t; now, $\underline{h}_T^0$ contains an additional nonparametric component which is the time dependence across different terms in the error sequence $\{U_t\}$.

Assuming the independence of $\{U_t\}$ and $\{W_t\}$, we have that $P(Y_t \leqslant \theta_0' W_t | \mathcal{W}_t) = P((1+\gamma|U_{t-1}|)U_t \leqslant 0 | \mathcal{W}_t) = P(U_t \leqslant 0)$, where the last equality uses the fact that the nuisance parameter γ in (1) is in $\mathbb{R}_+$. Hence, irrespective of the time-dependence structure across different U_t's, the conditional quantile of Y_t equals $\theta_0' W_t$, provided $P(U_t \leqslant 0) = \alpha$. We therefore assume that $\underline{h}_T^0$ belongs to the set $\mathcal{H}_T$ of all joint probability densities on $\mathbb{R}^T$ whose marginals have zero α-quantile.

Let $\mathcal{S}$ be the union of all parametric subfamilies $\mathcal{P}_{\underline{h}_T} \equiv \{f_{\underline{h}_{T\tau}}(\eta), \eta \in \Pi\}$ obtained when $\underline{h}_T$ is a smooth mapping $\underline{h}_T : \Upsilon \to \mathcal{H}_T$ associating a joint probability density $\underline{h}_{T\tau}(\cdot) \equiv \underline{h}_T(\cdot, \tau)$ to $\tau \in \Upsilon \subseteq \mathbb{R}^r$, with $\underline{h}_T^0 = \underline{h}_T(\cdot, \tau_0)$ for some $\tau_0 \in \Upsilon$. The parameter τ parameterizes 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 $f_{\underline{h}_{T\tau}}(\eta) \equiv \{f_{\underline{h}_{T\tau t}}(\cdot, \eta) : \mathbb{R} \to \mathbb{R}_*^+, 1 \leqslant t \leqslant T\}$ with $f_{\underline{h}_{T\tau t}}(Y_t, \cdot)$ continuously differentiable $a.s. - P$ on Π, $f_{\underline{h}_T t}(\cdot, \eta_0)$ continuously differentiable on $\mathbb{R}$, and such that the conditional quantile restriction (A3) is satisfied. Let $I_{\underline{h}_T T}(\eta, \tau) \equiv T^{-1}\mathrm{E}[\nabla_{(\eta',\tau')'} L_{\underline{h}_T T}(\eta, \tau) \nabla_{(\eta',\tau')'} L_{\underline{h}_T T}(\eta, \tau)']$ and $L_{\underline{h}_T T}(\eta, \tau) = \sum_{t=1}^T \ln f_{\underline{h}_{T\tau t}}(Y_t, \eta)$ be the Fisher information matrix and the log-likelihood function of the parametric submodel $\mathcal{P}_{\underline{h}_T}$, respectively. Provided the model is regular (Bickel, 1982; Begun, Hall, Huang, and Wellner, 1983) and $C^0_{\underline{h}_T T} \equiv \nabla_{(\eta',\tau')'}\theta(\eta_0, \tau_0)'(I_{\underline{h}_T T}(\eta_0, \tau_0))^+ \nabla_{(\eta',\tau')'}\theta(\eta_0, \tau_0)$ is nonsingular, Stein's (1956) definition of the semiparametric efficiency bound for θ_0 is the supremum of $C^0_{\underline{h}_T T}$ over the mappings $\underline{h}_T$. If this supremum is attained by a particular family $\underline{h}_T^*$, then $\mathcal{P}^* \equiv \mathcal{P}_{\underline{h}_T^*}$ is a least favorable parametric submodel.

3.2. Parametric Submodel for the Conditional Quantile Restriction

We now return 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 Assumption A3 is equivalent to the following model:

$$Y_t = q_\alpha(\theta_0, W_t) + \epsilon_t, \tag{2}$$

in which $\mathrm{E}[\mathbb{1}(-\epsilon_t)|\mathcal{W}_t] = \alpha$, a.s. $- P$, for every $t \geqslant 1$. The dependence assumption in Assumption (A1) implies that $\{(W_t', \epsilon_t)'\}$ is strong mixing. Model (2) contains example (1) as a special case: $\epsilon_t = (1+\gamma|U_{t-1}|)U_t$.

We now construct a parametric submodel of the conditional quantile model (2) by following the steps of the previous section. Recall that the log-likelihood function 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

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

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 the previous section applies with the log-likelihood function replaced by the "conditional" one. The parameter η now reduces to θ; the parameter τ used to parameterize the joint densities is set to be θ itself.

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

$$f_t^*(y, \theta) \equiv f_t^0(y) \times \frac{\alpha(1-\alpha)\lambda(\theta)\exp\{\lambda(\theta)[F_t^0(y) - F_t^0(q_\alpha(W_t, \theta))][\mathbb{1}(q_\alpha(W_t, \theta) - y) - \alpha]\}}{1 - \exp\{\lambda(\theta)[1 - F_t^0(q_\alpha(W_t, \theta)) - \mathbb{1}(q_\alpha(W_t, \theta) - y)][\mathbb{1}(q_\alpha(W_t, \theta) - y) - \alpha]\}}, \tag{3}$$

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.

Theorem 1 shows that $\mathcal{P}^*$ is a parametric submodel in $\mathcal{S}$, i.e., that: (i) for every $t \geqslant 1$, $f_t^*(\cdot, \theta)$ is a probability density for all $\theta \in \Theta$; (ii) for every $t \geqslant 1$, $f_t^*(\cdot, \theta)$ satisfies the conditional quantile restriction $\mathrm{E}_\theta[\mathbb{1}(q_\alpha(W_t, \theta) - Y_t) - \alpha | \mathcal{W}_t] = 0$, a.s. $- P$, for all $\theta \in \Theta$, where $\mathrm{E}_\theta(\cdot|\mathcal{W}_t)$ 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 1 (Parametric submodel). *Under Assumptions A2(ii) and A3, the parametric submodel $\mathcal{P}^*$ defined by* (3) *is a submodel of $\mathcal{S}$.*

The expression of the parametric submodel in Theorem 1 is new. In particular, the density $f_t^*(\cdot, \theta)$ is not of the "tick-exponential" form derived by Komunjer (2005): It depends on the true density $f_t^0(\cdot)$ as well as the true value θ_0 and contains terms such as $\lambda(\theta)$. A simple function $\Lambda(\cdot)$ that satisfies the conditions of Theorem 1 is $\Lambda(x) = x'x$. In the parametric submodel $\mathcal{P}^*$, θ 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 θ (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 θ different from θ_0; when $\theta = \theta_0$, the density $f_t^*(\cdot, \theta_0)$ equals the true density $f_t^0(\cdot)$, which is continuous.

3.3. Asymptotic Distribution of the Parametric Submodel MLE

Because $\mathcal{P}^*$ is a parametric submodel of the set $\mathcal{S}$ of all densities satisfying the conditional quantile restriction in Assumption A3, the semiparametric efficiency bound for θ_0 is by Stein's (1956) definition at least as large as the asymptotic

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

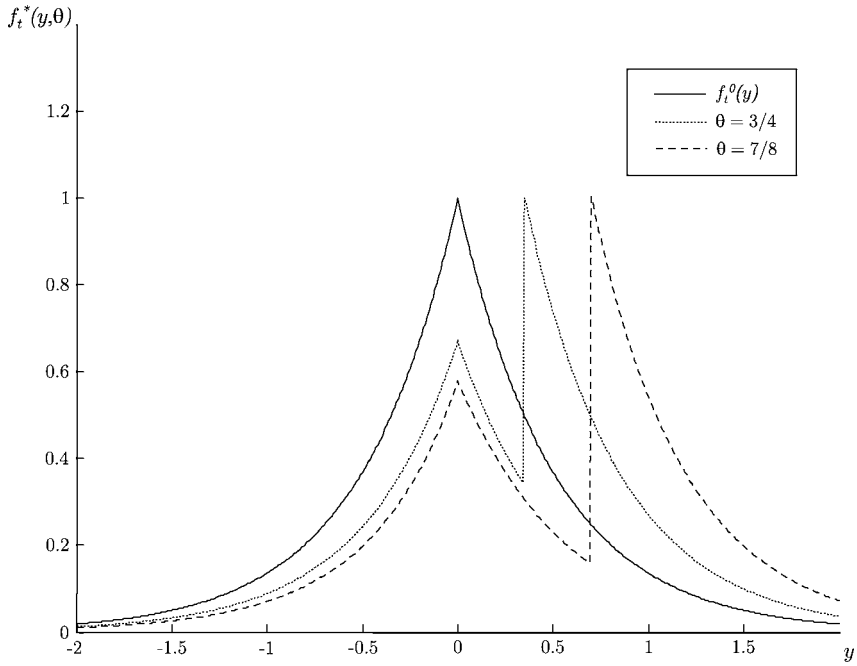


FIGURE 1. Case $\alpha = .5$, $q_\alpha(W_t, \theta) = \theta$ and $f_t^0(y) = \exp(-2|y|)$.

variance of the above MLE $\tilde{\theta}_T^*$. We derive the asymptotic distribution of $\tilde{\theta}_T^*$ under Assumptions 4 and 5.

Assumption A4. θ_0 is an interior point of Θ.

Assumption A5. $\sup_{t \geqslant 1} \mathrm{E}[\sup_{\theta \in \Theta} |\nabla_\theta q_\alpha(W_t, \theta)|^{2(r+\epsilon)}] < \infty$ and $\sup_{t \geqslant 1} \mathrm{E}[\sup_{\theta \in \Theta} |\Delta_{\theta\theta} q_\alpha(W_t, \theta)|^{r+\epsilon}] < \infty$, for some $\epsilon > 0$.

When combined with Assumption A1, the moment conditions in Assumption A5 guarantee that the appropriate law of large numbers and central limit theorem can be applied. In the special case corresponding to linear quantile regression models $q_\alpha(W_t, \theta) = \theta' W_t$, the set of moment conditions in Assumption A5 reduces to $\sup_{t \geqslant 1} \mathrm{E}[|W_t|^{2(r+\epsilon)}] < \infty$. The asymptotic distribution of $\tilde{\theta}_T^*$ is given in the following theorem.

THEOREM 2 (Asymptotic distribution of the MLE). *Under Assumptions A1–A5, 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 defined as*

$$V_T^0 \equiv \alpha(1-\alpha)\left\{ T^{-1} \sum_{t=1}^{T} E\left[f_t^0(q_\alpha(W_t, \theta_0))^2 \nabla_\theta q_\alpha(W_t, \theta_0) \nabla_\theta q_\alpha(W_t, \theta_0)' \right] \right\}^{-1}.$$

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

We first comment on the assumptions used in Theorem 2. Note that our result uses the moment restriction in Assumption A3 only; we do not make any additional assumptions regarding the properties of the error term ϵ_t in (2) other than those contained in Assumptions A3 and A1. In particular, we allow for $\{\epsilon_t\}$ to be dependent and nonidentically distributed.

Theorem 2 shows that Assumptions A1–A5 are sufficient primitive conditions for the MLE $\tilde{\theta}_T^*$ to be consistent and asymptotically normally distributed with asymptotic covariance matrix V_T^0. Since $\mathcal{P}^*$ is a parametric submodel compatible with (A3) and the estimator $\tilde{\theta}_T^*$ is fully parametric, the semiparametric efficiency bound for θ_0 cannot be smaller than V_T^0 (Stein, 1956). In other words, V_T^0 is a lower bound. In order to show that V_T^0 is also an upper bound, we now consider semiparametric estimators for θ_0.

4. SEMIPARAMETRIC ESTIMATORS FOR CONDITIONAL QUANTILES

In this section we consider a particular family of semiparametric conditional quantile estimators known as M- (or extremal) estimators (Huber, 1967). M-estimators for θ_0 are obtained by minimizing criterion functions of the form $\Psi_T(\theta) \equiv T^{-1}\sum_{t=1}^T \varphi_t(Y_t, q_\alpha(W_t,\theta))$, where for every $t \geqslant 1$, φ_t is a real function of the dependent variable Y_t and the quantile $q_\alpha(W_t,\theta)$. We shall focus on the class of M-estimators obtained when

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

where $F_t(\cdot)$ is a distribution function in $\mathcal{F}$. Note that to each choice of $F_t(\cdot)$ in $\mathcal{F}$ corresponds to a different M-estimator; hence we denote the latter as θ_T^F. We assume the following.

Assumption A6. For every $t \geqslant 1$, the distribution $F_t(\cdot)$ is: (i) $\mathcal{W}_t$-measurable; and (ii) twice continuously differentiable a.s. $-P$ on $\mathbb{R}$ with density $f_t(\cdot)$ satisfying: $\sup_{t\geqslant 1}\sup_{y\in\mathbb{R}} f_t(y) \leqslant M_0 < \infty$, and $\sup_{t\geqslant 1}\sup_{y\in\mathbb{R}} |df_t(y)/dy| \leqslant M_1 < \infty$ with $M_0, M_1 > 0$.

By Assumption (A6)(i), the conditional distribution $F_t(\cdot)$ in (4) is allowed to depend only on variables contained in $\mathcal{W}_t$, i.e., $F_t(\cdot)$ is predictable (see, e.g., Greenwood, Müller, and Wefelmeyer, 2004). By Assumption A6(ii) we require $F_t(\cdot)$ to be absolutely continuous with probability densities $f_t(\cdot)$ that are differentiable $a.s. - P$ on $\mathbb{R}$, bounded and with bounded derivatives.[5] The boundedness property is used to show that $\varphi_t(Y_t, q_\alpha(W_t,\cdot))$ are Lipshitz-L_1 on Θ $a.s. - P$. This implies that any pointwise convergence in θ becomes uniform on Θ. Alternatively, a similar implication can be obtained if the objective functions $\varphi_t(Y_t, q_\alpha(W_t,\cdot))$ are convex in the parameter θ. See, e.g., Pollard (1991).

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

4.1. Asymptotic Distribution

The asymptotic properties of θ_T^F depend on the choice of $F_t(\cdot)$. Before considering a particular choice of $F_t(\cdot)$, that makes the asymptotics of θ_T^F optimal, we need the asymptotic distribution of the latter.

THEOREM 3 (Asymptotic distribution). *Under Assumptions A1–A6,* $(\Sigma_T^0)^{-1/2}\Delta_T^0\sqrt{T}(\theta_T^F - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$, *where* $\Delta_T^0 \equiv \frac{1}{T}\sum_{t=1}^T \mathrm{E}[f_t(q_\alpha(W_t, \theta_0))\ f_t^0(q_\alpha(W_t,\theta_0))\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)']$, *and* $\Sigma_T^0 \equiv \frac{1}{T}\sum_{t=1}^T \alpha(1-\alpha)\,\mathrm{E}[f_t(q_\alpha\ (W_t,\theta_0))^2\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)']$.

Theorem 3 shows that Assumptions A1–A6 are sufficient for the M-estimator θ_T^F to be consistent and asymptotically normally distributed. The asymptotic variance of θ_T^F has the usual "sandwich" form $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1}$: Δ_T^0 is the Hessian matrix of the M-objective function $\Psi_T(\theta)$, i.e., $\Delta_T^0 = \frac{1}{T}\sum_{t=1}^T \mathrm{E}[\Delta_{\theta\theta}\varphi_t(Y_t, q_\alpha(W_t,\theta_0))]$; Σ_T^0 is the outer product of the gradient of $\Psi_T(\theta)$, i.e., $\Sigma_T^0 = \mathrm{E}[\nabla_\theta\Psi_T(\theta)\nabla_\theta\Psi_T(\theta)']$. The expression for Σ_T^0 in Theorem 3 then follows from the fact that $\{\nabla_\theta\varphi_t(Y_t, q_\alpha(W_t,\theta_0)), \mathcal{W}_t\}$ is a martingale difference sequence, which is implied by the correct specification in Assumption A3 and the $\mathcal{W}_t$-measurability of the distribution $F_t(\cdot)$.

4.2. Minimum Asymptotic Variance

Using Theorem 3, we can now rank all the M-estimators 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 $f_t^0(\cdot)$ and $q_\alpha(W_t,\cdot)$ in Assumption A2(ii). Typically, the asymptotic distribution of superefficient estimators is discontinuous in the true parameters, and our continuity assumptions rule out this discontinuity.

From Theorem 3 it is clear that changing the distribution function $F_t(\cdot)$ in (4) affects the asymptotic covariance matrix of the corresponding M-estimator θ_T^F, through the density term $f_t(\cdot)$ appearing in the expressions of Δ_T^0 and Σ_T^0. 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 $\Sigma_T^0(\Delta_T^0)^{-1} = \alpha(1-\alpha)\mathrm{Id}$. 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 rigorous.

THEOREM 4 (Minimum asymptotic variance). *Under Assumptions A1–A6,* V_T^0 *from Theorem 2 is the minimum of the set of matrices* $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1}$. *This minimum is attained by the M-estimator* θ_T^* *of* θ_0, *which minimizes* $\Psi_T^*(\theta) \equiv T^{-1}\sum_{t=1}^T \varphi_t^*(Y_t, q_\alpha(W_t,\theta))$ *where* $\varphi_t^*(y,q) = [\alpha - 1\ I(q-y)][F_t^0(y) - F_t^0(q)]$, $a.s.-P$, *on* $\mathbb{R}\times\mathcal{Q}$, *for every* $t \geqslant 1$.

Theorem 4 shows two important results. First, the matrix V_T^0 is the lower bound of the set of asymptotic variances of the class of M-estimators θ_T^F of θ_0. Thus, for any $F_t(\cdot)$, the difference between the corresponding asymptotic covariance matrix

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

$(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1}$ and V_T^0 is always positive semidefinite. Second, Theorem 4 shows that there exists an optimal M-estimator θ_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. This implies that the semiparametric efficiency bound needs to be lower than V_T^0.[6]

Recall from Theorem 2 that V_T^0 is the asymptotic covariance matrix of the MLE $\tilde{\theta}_T^*$ of the true parameter θ_0 in the parametric submodel $\mathcal{P}^*$. Thus, the semiparametric efficiency bound must be at least as large as V_T^0. It follows, first, that the parametric model $\mathcal{P}^*$ is a least favorable parametric submodel in $\mathcal{S}$, and, second, that V_T^0 is the semiparametric efficiency bound.

COROLLARY 5 (Semiparametric efficiency bound). *Under Assumptions A1–A6, V_T^0 in Theorems 2 and 4 is the semiparametric efficiency bound, and $[\alpha(1-\alpha)]^{-1/2} f_t^0(q_\alpha(W_t,\theta_0))\nabla_\theta q_\alpha(W_t,\theta_0)$ is the efficient score.*

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 in Assumption A3, then we cannot estimate the true conditional quantile parameter θ_0 with precision higher than that given by V_T^0.

Perhaps the most important aspect of Corollary 5 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 to mind. First is a quantile regression model in which W_t contains serially dependent exogenous variables/and or lags of Y_t, but where the residuals are uncorrelated (though possibly conditionally heteroskedastic). There are some results in this case in Newey and Powell (1990), under the additional assumption that $\{(Y_t, W_t')'\}$ is i.i.d. 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 5 generalizes Newey and Powell's (1990) results to the case where the sequence $\{(Y_t, W_t')'\}$ is weakly dependent and heterogeneous, as in Assumption A1. Unsurprisingly, when the data are i.i.d. and q_α linear, our bound V_T^0 reduces to $V^0 \equiv \alpha(1-\alpha)\{\mathrm{E}[(f_t^0(q_\alpha(W_t,\theta_0)))^2 W_t W_t']\}^{-1}$ derived by Newey and Powell (1990).

In the second time-series situation of interest, the residuals themselves are correlated in addition to being heteroskedastic. This situation is not covered in Newey and Powell (1990); however, Assumption A3 does not exclude that $\{\epsilon_t\}$ be correlated. So far there exist no results on semiparametric efficiency bound that cover this dependent case. To the best of our knowledge, Corollary 5 provides the first result on semiparametric efficiency for nonlinear (and possibly censored) conditional quantile models when the data is weakly dependent and heterogeneous.

5. CONCLUSION

The contributions of this paper are twofold: First, we derive the semiparametric efficiency bound V_T^0 for parameters of conditional quantiles in time-series models

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

with weakly dependent and/or heterogeneous data. Our bound V_T^0 generalizes expressions previously derived in the literature on efficient conditional quantile estimation. In particular we allow the data to exhibit dependence and/or conditional heteroskedasticity. The second contribution of the paper is to exhibit a least favorable submodel for the conditional quantile restriction.

The main limitation of working with parametric submodels comes from the difficulty of finding the least favorable one. In order to demonstrate that a given parametric submodel is the least favorable, we need to find a semiparametric estimator whose asymptotic variance equals that of the MLE under the submodel. Depending on the nature of the problem, this task may or may not be easy (see, e.g., Komunjer and Ragusa, 2009). The main advantage of our parametric submodel approach, however, is that it does not suffer from independence assumptions traditionally imposed in the literature on efficient estimation. Our construction of the least favorable parametric submodel and the corresponding MLE does not depend on any particular dependence or heterogeneity structure of the data. Its extension to general moment restrictions—other than those involving the conditional quantiles—is an interesting venue for future research.

NOTES

1. Carrasco and Florens (2004) and Carrasco, Chernov, Florens, and Ghysels (2007) study semiparametric efficiency using the assumption of strictly stationary and strong mixing Markov chains.

2. Our least favorable submodel approach is close in spirit to empirical likelihood-based methods. Kitamura (2006) presents empirical likelihood-based approaches that are suitable for estimating the parameters defined through a set of unconditional moment restrictions with time-series data.

3. Strong exogeneity implies that the conditional process generating $\{Y_t|\mathcal{W}_t\}$ and the marginal process generating $\{X_t\}$ can be separated from each other so that X_t can be treated as "fixed" in the conditional model $f_t^0(y_t)$ since lagged values of Y_t do not affect the process generating X_t.

4. For most of our conditions and statements that hold for any $T \geqslant 1$ and every t, $1 \leqslant t \leqslant T$, we shall often use a shortcut notation "for every $t \geqslant 1$" when doing so does not introduce any ambiguity. For example, the above condition assumes 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 remaining notation is standard.

5. Recall that all distributions in $\mathcal{F}$ satisfy this requirement.

6. The optimal M-estimator is not a member of the QMLEs; see Komunjer (2005). It is obtained by minimizing an objective function that depends on the true conditional distribution $F_t^0(\cdot)$. Hence, in principle the optimal M-estimator is "infeasible." On the other hand, all other members of our considered class of M-estimators are feasible. It is then easy to see that V_T^0 is also the infimum of the asymptotic variances of the latter. Though θ_T^* satisfies all the assumptions in (A6), its computation requires knowing or estimating the true conditional distribution $F_t^0(\cdot)$. Komunjer and Vuong (2007) develop this idea and show that the semiparametric efficiency bound is attained by a feasible version of θ_T^*.

REFERENCES

Begun, J., W. Hall, W. Huang, & J. Wellner (1983) Information and asymptotic efficiency in parametric-nonparametric models. *Annals of Statistics* 11, 432–452.

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

Bickel, P.J. (1982) On adaptive estimation. *Annals of Statistics* 10, 647–671.

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

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, & J.A. Wellner (2005) Semiparametric Inference and Models. Mimeo, University of California, Berkeley.

Bickel, P.J. & J. Kwon (2001) Inference for semiparametric models: Some questions and an answer. *Statistica Sinica* 11, 863–960.

Carrasco, M., M. Chernov, J.-P. Florens, & E. Ghysels (2007) Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics* 140, 529–573.

Carrasco, M. & J.-P. Florens (2004) On the Asymptotic Efficiency of GMM. Mimeo, University of Montreal and University of Toulouse I.

Drost, F.C., C. Klaassen, & B.J. Werker (1997) Adaptive estimation in time-series models. *Annals of Statistics* 25, 786–818.

Engle, R.F., D.F. Hendry, & J.-F. Richard (1983) Exogeneity. *Econometrica* 51, 277–304.

Fabian, V. & J. Hannan (1982) On estimation and adaptive estimation for locally asymptotically normal families. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* 59, 459–478.

Gourieroux, C., A. Monfort, & A. Trognon (1984) Pseudo maximum likelihood methods: Theory. *Econometrica* 52, 681–700.

Greenwood, P.E., U.U. Müller, & W. Wefelmeyer (2004) An introduction to effcient estimation for semiparametric time series. In M.S. Mikulin, N. Balakrishnan, M. Mesbah, & N. Limnios (eds.), *Parametric and Semiparametric Models with Applications to Reliablilty, Survival Analysis, and Quality of Life*, pp. 253–272. Birkhäuser.

Hallin, M. & I. Mizera (2001) Sample heterogeneity and the asymptotics of M-estimators. *Journal of Statistical Planning and Inference* 93, 139–160.

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

Hansen, L.P., J. Heaton, & M. Ogaki (1988) Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association* 83, 863–871.

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*. University of California Press.

Kitamura, Y. (2006) Empirical Likelihood Methods in Econometrics: Theory and Practice. Cowles Foundation discussion paper No. 1569.

Klaassen, C.A.J. & J.A. Wellner (1997) Efficient estimation in the bivariate normal copula model: Normal margins are least favourable. *Bernoulli* 3, 55–77.

Koenker, R. & 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. & G. Ragusa (2009) Existence and Uniqueness of Semiparametric Projections. Mimeo, University of California San Diego and University of California Irvine.

Komunjer, I. & Q. Vuong (2007) Efficient Estimation in Dynamic Conditional Quantile Models. Mimeo, 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. & A.W. van der Vaart (2000) On profile likelihood. *Journal of the American Statistical Association* 95, 449–465.

Newey, W.K. & D.L. McFadden (1994) Large sample estimation and hypothesis testing. In R. F. Engle, & D. L. McFadden (eds.), *Handbook of Econometrics*, pp. 2113–2247. Elsevier Science.

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

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

Pfanzagl, J. (1976) Investigating the quantile of an unknown distribution. In J.W. Zeigler (ed.), *Contributions to Applied Statistics*. Birkhäuser.

Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.

Severini, T.A. & 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. University of California Press.

Wellner, J.A., C.A.J. Klaassen, & Y. Ritov (2006) Semiparametric models: A review of progress since BKRW (1993). In J. Fan & H. Koul (eds.), *Frontiers in Statistics: In Honor of Peter J. Bickel's 65th Birthday*, pp. 25–35. Imperial College Press.

White, H. (2001) *Asymptotic Theory for Econometricians*. Academic Press.

APPENDIX: Proofs

Proof of Theorem 1. 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 (3) is a probability density, for every $t \geqslant 1$. 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 $\mathbf{I}(q_t(\theta) - y) = \mathbf{I}[\lambda(\theta)F_t^0(q_t(\theta)) - u]$, we have

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

$$+\int_{\lambda(\theta)F_t^0(q_t(\theta))}^{\lambda(\theta)} \frac{\alpha(1-\alpha)\exp\{-\alpha[u-\lambda(\theta)F_t^0(q_t(\theta))]\}}{1-\exp\{-\alpha\lambda(\theta)[1-F_t^0(q_t(\theta))]\}}du$$

$$= \alpha + (1-\alpha) = 1,$$

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 θ_0 and that $f_t^*(\cdot,\theta_0) = f_t^0(\cdot)$. For this, let $P_t(\theta) \equiv \alpha(1-\alpha)\lambda(\theta)\exp\{\lambda(\theta)[F_t^0(y) - F_t^0(q_t(\theta))][\mathbf{I}(q_t(\theta)-y)-\alpha]\}$ and $Q_t(\theta) \equiv 1-\exp\{\lambda(\theta)[1-F_t^0(q_t(\theta)) - \mathbf{I}(q_t(\theta)-y)][\mathbf{I}(q_t(\theta)-y)-\alpha]\}$, so that $f_t^*(y,\theta) = f_t^0(y)P_t(\theta)/Q_t(\theta)$. By Assumption A2(ii), the functions P_t and Q_t are at least twice continuously differentiable on Θ $a.s.-P$; 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,

$$P_t(\theta) = \frac{1}{2}\alpha(1-\alpha)D^2\lambda(\theta_0)(\theta-\theta_0)^2 + o(|\theta-\theta_0|^2), \quad \text{and} \tag{A.1}$$

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

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

Given the nonsingularity of $\Delta_{\theta\theta}\Lambda(0)$, an immediate consequence of l'Hôpital's rule and (A.1)–(A.2) 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 Θ, 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 show 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 (A3) by assumption. When $\theta \neq \theta_0$, we have

$$\mathrm{E}_\theta[\mathbf{I}(q_t(\theta) - Y_t)|\mathcal{W}_t] = \int_{-\infty}^{q_t(\theta)} f_t^*(y,\theta)dy = \alpha,$$

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

Proof of Theorem 2. We now consider the MLE $\tilde{\theta}_T^*$ that maximizes the log-likelihood $L_T(\theta) \equiv T^{-1}\sum_{t=1}^T \ln f_t^*(Y_t,\theta)$.

STEP 1: 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. Hallin and Mizera (2001) offer an interesting alternative to proving consistency of M-estimators. Given Assumption A2(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

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

showing that $\mathrm{E}[\ln f_t^*(Y_t,\theta)]$ is continuous on Θ and that $\mathrm{E}[\sup_{\theta\in\Theta}|\ln f_t^*(Y_t,\theta)|^{r+\epsilon}] < \infty$ for every $t \geqslant 1$, 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

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

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

$$\begin{aligned}\nabla_\theta \ln f_t^*(Y_t,\theta) = {} & \frac{\nabla_\theta\lambda(\theta)}{\lambda(\theta)} + \nabla_\theta\lambda(\theta)[F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbf{I}(q_t(\theta) - Y_t) - \alpha] \\ & - \lambda(\theta)f_t^0(q_t(\theta))\nabla_\theta q_t(\theta)[\mathbf{I}(q_t(\theta) - Y_t) - \alpha] + u(\lambda(\theta)x(\theta))\nabla_\theta(\lambda(\theta)x(\theta)) \\ & + \lambda(\theta)[F_t^0(Y_t) - F_t^0(q_t(\theta))]\delta(q_t(\theta) - Y_t)\nabla_\theta q_t(\theta), \qquad \textbf{(A.3)}\end{aligned}$$

where $\nabla_\theta(\lambda(\theta)x(\theta)) = \nabla_\theta\lambda(\theta)x(\theta) + \lambda(\theta)\nabla_\theta x(\theta)$, and $\nabla_\theta x(\theta) = \{f_t^0(q_t(\theta))[\alpha - \mathbf{I}(q_t(\theta) - Y_t)] + \delta(q_t(\theta) - Y_t)[\alpha - F_t^0(q_t(\theta))]\}\nabla_\theta q_t(\theta)$, given that $[\mathbf{I}(\cdot)]^2 = \mathbf{I}(\cdot)$. Note

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

that $u(z) = -1/z - 1/2 + o(1)$ in the neighborhood of 0, and that $\lambda(\theta)x(\theta) = o_p(1)$ in the neighborhood of θ_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 θ_0. In particular, combining the above results, we get

$$\begin{aligned}
&\nabla_\theta \ln f_t^*(Y_t, \theta_0) \\
&= -\nabla_\theta q_t(\theta_0) \left\{ \frac{f_t^0(q_t(\theta_0))[\alpha - \mathbf{I}(q_t(\theta_0) - Y_t)] + \delta(q_t(\theta_0) - Y_t)[\alpha - F_t^0(q_t(\theta_0))]}{[\mathbf{I}(q_t(\theta_0) - Y_t) - \alpha][1 - \mathbf{I}(q_t(\theta_0) - Y_t) - F_t^0(q_t(\theta_0))]} \right\} \\
&= -\frac{1}{\alpha(1-\alpha)} \nabla_\theta q_t(\theta_0) f_t^0(q_t(\theta_0))[\mathbf{I}(q_t(\theta_0) - Y_t) - \alpha], \qquad \textbf{(A.4)}
\end{aligned}$$

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 Θ $a.s.-P$ 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

$$\begin{aligned}
\sup_{\theta\in\Theta} |\nabla_\theta \ln f_t^*(Y_t, \theta)| &\leqslant 2 \sup_{\theta\in\Theta} |\nabla_\theta\lambda(\theta)| + \sup_{\theta\in\Theta} |\lambda(\theta)| M_0 |\nabla_\theta q_t(\theta)| \\
&\quad + C_1 \sup_{\theta\in\Theta} \left| \frac{f_t^0(q_t(\theta))\nabla_\theta q_t(\theta)}{1 - \mathbf{I}(q_t(\theta) - Y_t) - F_t^0(q_t(\theta))} \right|, \text{a.s.} - P, \qquad \textbf{(A.5)}
\end{aligned}$$

where $C_1 \equiv \sup_{x\in[0,\sup_{\theta\in\Theta}\lambda(\theta)]} \left|\frac{x}{1-\exp(-x)}\right| < \infty$. We have $\sup_{t\geqslant 1}\sup_{\theta\in\Theta} F_t^0(q_t(\theta)) \in (a, b)$, $a > 0$, $b < 1$, so $C_2 \equiv \sup_{t\geqslant 1}\sup_{y\in\mathbb{R}}\sup_{\theta\in\Theta}\left(\left|1 - \mathbf{I}(q_t(\theta) - y) - F_t^0(q_t(\theta))\right|^{-1}\right) < \infty$, 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 Assumption A5, $\mathrm{E}[\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)|] < \infty$, so $\mathrm{E}[\sup_{\theta\in\Theta}|\nabla_\theta \ln f_t^*(Y_t, \theta)|] < \infty$ for every $t \geqslant 1$, which shows that (25) in Komunjer (2005) holds; together with (A1) and $\mathrm{E}[\sup_{\theta\in\Theta}|\ln f_t^*(Y_t, \theta)|^{r+\epsilon}] < \infty$ for every $t \geqslant 1$, this establishes condition (iv) of Theorem 2.1 and completes the proof of consistency.

STEP 2: 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 $|\tilde{G}_{T,j}^*(0)| \leqslant T^{-1}\sum_{t=1}^T[1 + 2C_1]\left[\left|\frac{\partial\lambda(\tilde{\theta}_T^*)}{\partial\theta_j}\right| + \left|\lambda(\tilde{\theta}_T^*) f_t^0(q_t(\tilde{\theta}_T^*))\frac{\partial q_t(\tilde{\theta}_T^*)}{\partial\theta_j}\right|\right] \mathbf{I}\{q_t(\tilde{\theta}_T^*) = Y_t\}$. Hence,

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

$$P\big(\sqrt{T}|\nabla_\theta L_T(\tilde\theta_T^*)|>\varepsilon\big)\leqslant P\Big(\sqrt{T}\max_{1\leqslant j\leqslant k}|\tilde G_{T,j}^*(0)|>\varepsilon\Big)$$
$$\leqslant P\left(\sum_{t=1}^T\left[\left|\frac{\partial\lambda(\tilde\theta_T^*)}{\partial\theta_j}\right|+\left|\lambda(\tilde\theta_T^*)f_t^0(q_t(\tilde\theta_T^*))\frac{\partial q_t(\tilde\theta_T^*)}{\partial\theta_j}\right|\right]\mathbf{I}\{q_t(\tilde\theta_T^*)=Y_t\}>\varepsilon\sqrt{T}(1+2C_1)^{-1}\right).$$

That $P(\mathbf{I}\{q_t(\tilde\theta_T^*)=Y_t\}\neq 0)=0$ and $\mathrm{E}\left[\left|\frac{\partial\lambda(\tilde\theta_T^*)}{\partial\theta_j}\right|+\left|\lambda(\tilde\theta_T^*)f_t^0(q_t(\tilde\theta_T^*))\frac{\partial q_t(\tilde\theta_T^*)}{\partial\theta_j}\right|\right]$ is bounded then ensure that $\lim_{T\to\infty}P\big(\sqrt{T}\big|\nabla_\theta L_T(\tilde\theta_T^*)\big|>\varepsilon\big)=0$. Condition (i) of Theorem 7.2 follows from the correct specification of $f_t(\cdot)$ (see (iii) in Theorem 1). By Assumption A4, θ_0 is an interior point of Θ so that condition (iii) of Theorem 7.2 holds. We now check the differentiability of $\mathrm{E}[\nabla_\theta L_T(\theta)]$ and the nonsingularity condition (ii) of Theorem 7.2. We have $\mathrm{E}[\nabla_\theta L_T(\theta)]=T^{-1}\sum_{t=1}^T\mathrm{E}[\nabla_\theta\ln f_t^*(Y_t,\theta)]$; using $(A.3)$, the latter is easily shown to be differentiable at any $\theta\in\mathring\Theta$. We now show that $\nabla_\theta\mathrm{E}[\nabla_\theta' L_T(\theta_0)]=T^{-1}\sum_{t=1}^T\mathrm{E}[\Delta_{\theta\theta}\ln f_t^*(Y_t,\theta_0)]$ and that the latter is nonsingular. We have $du(z)/dz=u(z)+[u(z)]^2$, hence, for every $t\geqslant 1$,

$$\begin{aligned}\Delta_{\theta\theta}\ln f_t^*(Y_t,\theta)&=\frac{\Delta_{\theta\theta}\lambda(\theta)}{\lambda(\theta)}-\frac{\nabla_\theta\lambda(\theta)\nabla_\theta\lambda(\theta)'}{[\lambda(\theta)]^2}\\&+\Delta_{\theta\theta}\lambda(\theta)[F_t^0(Y_t)-F_t^0(q_t(\theta))][\mathbf{I}(q_t(\theta)-Y_t)-\alpha]\\&+2\nabla_\theta\lambda(\theta)\nabla_\theta q_t(\theta)'\big\{f_t^0(q_t(\theta))[\alpha-\mathbf{I}(q_t(\theta)-Y_t)]\\&+\delta(q_t(\theta)-Y_t)[F_t^0(Y_t)-F_t^0(q_t(\theta))]\big\}\\&+\lambda(\theta)\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'\left\{\frac{df_t^0(q_t(\theta))}{dq}[\alpha-\mathbf{I}(q_t(\theta)-Y_t)]-2f_t^0(q_t(\theta))\delta(q_t(\theta)-Y_t)\right.\\&\qquad\left.+[F_t^0(Y_t)-F_t^0(q_t(\theta))]\frac{d\delta(q_t(\theta)-Y_t)}{dq}\right\}\\&+u(\lambda(\theta)x(\theta))\Delta_{\theta\theta}(\lambda(\theta)x(\theta))+\lambda(\theta)\Delta_{\theta\theta}q_t(\theta)\big\{f_t^0(q_t(\theta))[\alpha-\mathbf{I}(q_t(\theta)-Y_t)]\\&+[F_t^0(Y_t)-F_t^0(q_t(\theta))]\delta(q_t(\theta)-Y_t)\big\}\\&+\big[u(\lambda(\theta)x(\theta))+(u(\lambda(\theta)x(\theta)))^2\big](\nabla_\theta(\lambda(\theta)x(\theta)))(\nabla_\theta(\lambda(\theta)x(\theta)))',\end{aligned}\tag{A.6}$$

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

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

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

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

$$
\begin{aligned}
&[u(\lambda(\theta)x(\theta)) + u(\lambda(\theta)x(\theta))^2]\nabla_\theta(\lambda(\theta)x(\theta))\nabla_\theta(\lambda(\theta)x(\theta))' \\
&\quad = \frac{\nabla_\theta\lambda(\theta)\nabla_\theta\lambda(\theta)'}{[\lambda(\theta)]^2} + 2\frac{\nabla_\theta\lambda(\theta)\nabla_\theta x(\theta)'}{\lambda(\theta)x(\theta)} \\
&\qquad + \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'\left\{ f_t^0(q_t(\theta))\frac{[\alpha - \mathbf{I}(q_t(\theta) - Y_t)]}{x(\theta)} \right. \\
&\qquad\qquad \left. + \delta(q_t(\theta) - Y_t)\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\}^2 + o_p(1), \qquad \textbf{(A.7)}
\end{aligned}
$$

in the neighborhood of θ_0. Similarly,

$$
\begin{aligned}
&u(\lambda(\theta)x(\theta))\Delta_{\theta\theta}(\lambda(\theta)x(\theta)) \\
&\quad = -\frac{\Delta_{\theta\theta}\lambda(\theta)}{\lambda(\theta)} - \frac{1}{2}\Delta_{\theta\theta}\lambda(\theta)x(\theta) - 2\frac{\nabla_\theta\lambda(\theta)\nabla_\theta x(\theta)'}{\lambda(\theta)x(\theta)} \\
&\qquad - \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'\left\{ \frac{df_t^0(q_t(\theta))}{dq}\frac{[\alpha - \mathbf{I}(q_t(\theta) - Y_t)]}{x(\theta)} - 2\frac{f_t^0(q_t(\theta))\delta(q_t(\theta) - Y_t)}{x(\theta)} \right. \\
&\qquad\qquad \left. + \frac{d\delta(q_t(\theta) - Y_t)}{dq}\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} \\
&\qquad - \Delta_{\theta\theta}q_t(\theta)\left\{ f_t^0(q_t(\theta))\frac{[\alpha - \mathbf{I}(q_t(\theta) - Y_t)]}{x(\theta)} \right. \\
&\qquad\qquad \left. + \delta(q_t(\theta) - Y_t)\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} + o_p(1), \qquad \textbf{(A.8)}
\end{aligned}
$$

in the neighborhood of θ_0. Combining (A.6), (A.7), and (A.8), we get that, for every $t \geqslant 1$,

$$
\begin{aligned}
&\Delta_{\theta\theta}\ln f_t^*(Y_t,\theta) \\
&\quad = \Delta_{\theta\theta}\lambda(\theta)\left\{ [F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbf{I}(q_t(\theta) - Y_t) - \alpha] - \frac{1}{2}x(\theta) \right\} \\
&\qquad + \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'\left\{ f_t^0(q_t(\theta))\frac{[\alpha - \mathbf{I}(q_t(\theta) - Y_t)]}{x(\theta)} \right. \\
&\qquad\qquad \left. + \delta(q_t(\theta) - Y_t)\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\}^2 \\
&\qquad - \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'\left\{ \frac{df_t^0(q_t(\theta))}{dq}\frac{[\alpha - \mathbf{I}(q_t(\theta) - Y_t)]}{x(\theta)} - 2\frac{f_t^0(q_t(\theta))\delta(q_t(\theta) - Y_t)}{x(\theta)} \right. \\
&\qquad\qquad \left. + \frac{d\delta(q_t(\theta) - Y_t)}{dq}\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} \\
&\qquad - \Delta_{\theta\theta}q_t(\theta)\left\{ f_t^0(q_t(\theta))\frac{[\alpha - \mathbf{I}(q_t(\theta) - Y_t)]}{x(\theta)} \right. \\
&\qquad\qquad \left. + \delta(q_t(\theta) - Y_t)\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} + o_p(1), \qquad \textbf{(A.9)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

in the neighborhood of θ_0. Using $\alpha = F_t^0(q_t(\theta_0))$ and $x(\theta_0) = -\alpha(1-\alpha)$, we have $|\Delta_{\theta\theta} \ln f_t^*(Y_t,\theta_0)| \leqslant \frac{5}{2}|\Delta_{\theta\theta}\lambda(\theta_0)| + |\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'|\left(\frac{M_0^2}{[\alpha(1-\alpha)]^2} + \frac{M_1}{\alpha(1-\alpha)}\right) + |\Delta_{\theta\theta} q_t(\theta_0)| \times \frac{M_0}{\alpha(1-\alpha)} + o_p(1)$, with $|\Delta_{\theta\theta}\lambda(\theta_0)| < \infty$. From Assumption A5, $\mathrm{E}[|\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'|] < \infty$ and $\mathrm{E}[|\Delta_{\theta\theta} q_t(\theta_0)|] < \infty$, which shows that the expectation of the right-hand side of the above inequality is finite; hence, $\nabla_\theta \mathrm{E}[\nabla_\theta' \ln f_t^*(Y_t,\theta_0)] = \mathrm{E}[\Delta_{\theta\theta} \ln f_t^*(Y_t,\theta_0)]$ for every $t \geqslant 1$ and so $\nabla_\theta \mathrm{E}[\nabla_\theta' L_T(\theta_0)] = T^{-1}\sum_{t=1}^T \mathrm{E}[\Delta_{\theta\theta} \ln f_t^*(Y_t,\theta_0)]$ as desired. Now consider $\mathrm{E}[\Delta_{\theta\theta} \ln f_t^*(Y_t,\theta_0)]$; for every $t \geqslant 1$, we have

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

since

$$
\begin{aligned}
&\mathrm{E}_t\left([F_t^0(Y_t) - \alpha][\mathbf{I}(q_t(\theta_0) - Y_t) - \alpha]\right) \\
&\quad = (1-\alpha)\int_{-\infty}^{q_t(\theta_0)} [F_t^0(y) - \alpha] f_t^0(y)dy - \alpha\int_{q_t(\theta_0)}^{+\infty} [F_t^0(y) - \alpha] f_t^0(y)dy \\
&\quad = (1-\alpha)\left[\frac{1}{2}[F_t^0(y) - \alpha]^2\right]_{-\infty}^{q_t(\theta_0)} - \alpha\left[\frac{1}{2}[F_t^0(y) - \alpha]^2\right]_{q_t(\theta_0)}^{+\infty} = -\frac{1}{2}\alpha(1-\alpha).
\end{aligned}
$$

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

$$
\begin{aligned}
&\mathrm{E}\Bigg(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\Bigg\{f_t^0(q_t(\theta_0))\frac{[\alpha - \mathbf{I}(q_t(\theta_0) - Y_t)]}{x(\theta_0)} \\
&\qquad\qquad + \delta(q_t(\theta_0) - Y_t)\frac{[\alpha - F_t^0(q_t(\theta_0))]}{x(\theta_0)}\Bigg\}^2\Bigg) \\
&\quad = \mathrm{E}\left(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\mathrm{E}_t\left\{\frac{[f_t^0(q_t(\theta_0))]^2[\alpha - \mathbf{I}(q_t(\theta_0) - Y_t)]^2}{\alpha^2(1-\alpha)^2}\right\}\right) \\
&\quad = \mathrm{E}\left(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)}\right),
\end{aligned}
$$

where the last equality uses $\mathrm{E}_t([\mathbf{I}(q_t(\theta_0) - Y_t) - \alpha]^2) = \alpha(1-\alpha)$, $a.s.-P$. Similarly,

$$
\begin{aligned}
&\mathrm{E}\Bigg(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'\Bigg\{\frac{df_t^0(q_t(\theta_0))}{dq}\frac{[\alpha - \mathbf{I}(q_t(\theta_0) - Y_t)]}{x(\theta_0)} - 2\frac{f_t^0(q_t(\theta_0))\delta(q_t(\theta_0) - Y_t)}{x(\theta_0)} \\
&\qquad\qquad + \frac{d\delta(q_t(\theta_0) - Y_t)}{dq}\frac{[\alpha - F_t^0(q_t(\theta_0))]}{x(\theta_0)}\Bigg\}\Bigg)
\end{aligned}
$$

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

$$= \mathrm{E}\Bigg(\nabla_\theta q_t(\theta_0) \nabla_\theta q_t(\theta_0)' \mathrm{E}_t \Bigg\{ \frac{df_t^0(q_t(\theta_0))}{dq} \frac{[\mathbf{I}(q_t(\theta_0) - Y_t) - \alpha]}{\alpha(1-\alpha)} + 2 \frac{f_t^0(q_t(\theta_0))\delta(q_t(\theta_0) - Y_t)}{\alpha(1-\alpha)} \Bigg\} \Bigg)$$

$$= 2\mathrm{E}\left(\nabla_\theta q_t(\theta_0) \nabla_\theta q_t(\theta_0)' \frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)} \right),$$

where the last equality uses $\mathrm{E}_t(\mathbf{I}(q_t(\theta_0) - Y_t) - \alpha) = 0$, a.s. $- P$ and $\mathrm{E}_t(\delta(q_t(\theta_0) - Y_t)) = f_t^0(q_t(\theta_0))$, a.s. $- P$. Finally, by the same reasoning, $\mathrm{E}\left(\Delta_{\theta\theta} q_t(\theta_0) \left\{ f_t^0(q_t(\theta_0)) \times \frac{[\alpha - \mathbf{I}(q_t(\theta_0) - Y_t)]}{x(\theta_0)} + \delta(q_t(\theta_0) - Y_t) \frac{[\alpha - F_t^0(q_t(\theta_0))]}{x(\theta_0)} \right\} \right) = 0$. Combining the above results then yields, by (A.9), $\mathrm{E}[\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta_0)] = -\mathrm{E}\left(\nabla_\theta q_t(\theta_0) \nabla_\theta q_t(\theta_0)' \frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)} \right)$, for every $t \geqslant 1$. So for any $\chi \in \mathbb{R}^k$, $\chi' \nabla_\theta \mathrm{E}[\nabla_\theta' L_T(\theta_0)] \chi = -T^{-1} \sum_{t=1}^T \mathrm{E}\left(|\nabla_\theta q_t(\theta_0)' \chi|^2 \times \frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)} \right) \leqslant 0$, with equality if and only if $\chi = 0$. Hence, $\nabla_\theta \mathrm{E}[\nabla_\theta' L_T(\theta_0)]$ is negative definite (therefore nonsingular). We now check condition (iv) of Theorem 7.2 by using a CLT for α-mixing sequences (e.g., White, 2001, Thm. 5.20, p. 130). By (A1), for any $\theta \in \mathring{\Theta}$, the sequence $\{\nabla_\theta \ln f_t^*(Y_t, \theta)\}$ is strong mixing (i.e., α-mixing) with α of size $-r/(r-2)$, $r > 2$ (see, e.g., White, 2001, Thm. 3.49, p. 50). Moreover, using (A.3) and Assumption A3, $\mathrm{E}[\nabla_\theta \ln f_t^*(Y_t, \theta_0)] = 0$ and using Assumption A5, $\mathrm{E}[|\nabla_\theta \ln f_t^*(Y_t, \theta_0)|^r] \leqslant \{M_0/[\alpha(1-\alpha)]\}^r \mathrm{E}[\sup_{\theta \in \Theta} |\nabla_\theta q_t(\theta)|^r] < \infty$, for every $t \geqslant 1$. Now,

$$\mathrm{Var}\left(T^{-1} \sum_{t=1}^T \nabla_\theta \ln f_t^*(Y_t, \theta_0) \right)$$

$$= \mathrm{E}\left(T^{-1} \sum_{t=1}^T \nabla_\theta \ln f_t^*(Y_t, \theta_0) \nabla_\theta \ln f_t^*(Y_t, \theta_0)' \right)$$

$$= \mathrm{E}\left(T^{-1} \sum_{t=1}^T \frac{[f_t^0(q_t(\theta_0))]^2 [\mathbf{I}(q_t(\theta_0) - Y_t) - \alpha]^2}{[\alpha(1-\alpha)]^2} \nabla_\theta q_t(\theta_0) \nabla_\theta q_t(\theta_0)' \right) = V_T^0,$$

where the first equality uses $\mathrm{E}_t(\nabla_\theta \ln f_t^*(Y_t, \theta_0)) = 0$, *a.s.* $- P$, implied by (A3), and the last equality uses $\mathrm{E}_t([\mathbf{I}(q_t(\theta_0) - Y_t) - \alpha]^2) = \alpha(1-\alpha)$, *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 4. Finally, we check the stochastic equicontinuity condition (v) of Theorem 7.2 by verifying 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 every $t \geqslant 1$, 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 neighborhood of 0 and $\lambda(\theta) x(\theta) = o_p(|\theta - \theta_0|)$ in the neighborhood of θ_0, we have, from (A.3), (A.4), and (A.6), $r_t(\theta) \leqslant r_t^{(1)}(\theta) + r_t^{(2)}(\theta) + r_t^{(3)}(\theta) + o_p(1)$,

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

where

$$r_t^{(1)}(\theta) = \left| [F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbf{I}(q_t(\theta) - Y_t) - \alpha] - \frac{x(\theta)}{2} \right| \frac{|\nabla_\theta \lambda(\theta) - \Delta_{\theta\theta}\lambda(\theta)'(\theta - \theta_0)|}{|\theta - \theta_0|},$$

$$r_t^{(2)}(\theta) = \left| \frac{f_t^0(q_t(\theta))}{2} [\mathbf{I}(q_t(\theta) - Y_t) - \alpha] + \delta(q_t(\theta) - Y_t) \right.$$
$$\left. \times \left[\frac{F_t^0(q_t(\theta)) - 2F_t^0(Y_t) + \alpha}{2} \right] \right| \frac{|\lambda(\theta)\nabla_\theta q_t(\theta)|}{|\theta - \theta_0|}, \quad \text{and}$$

$$r_t^{(3)}(\theta) = \left| \frac{\nabla_\theta x(\theta)}{x(\theta)} - \frac{\nabla_\theta x(\theta_0)}{x(\theta_0)} - \frac{\Delta_{\theta\theta} x(\theta)'(\theta - \theta_0)}{x(\theta)} + \frac{\nabla_\theta x(\theta) \nabla_\theta x(\theta)'(\theta - \theta_0)}{[x(\theta)]^2} \right| / |\theta - \theta_0|.$$

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 continuously 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 $\mathrm{E}(\sup_{\theta \in \mathring{\Theta}: |\theta - \theta_0| < \varepsilon_1} r_t^{(1)}(\theta)) < \infty$. Now, note that $|f_t^0(q_t(\theta))[\mathbf{I}(q_t(\theta) - Y_t) - \alpha]| \leqslant M_0$ for any $\theta \in \mathring{\Theta}$, so

$$r_t^{(2)}(\theta) \leqslant \frac{1}{2} \left\{ M_0 + \delta(q_t(\theta) - Y_t) \left[F_t^0(q_t(\theta)) - 2F_t^0(Y_t) + \alpha \right] \right\} \frac{|\lambda(\theta)\nabla_\theta q_t(\theta)|}{|\theta - \theta_0|}$$
$$\leqslant \frac{1}{2} \left\{ M_0 + \delta(q_t(\theta) - Y_t)[F_t^0(q_t(\theta)) - 2F_t^0(Y_t) + \alpha] \right\} |\nabla_\theta \lambda(\theta_c)| \cdot |\nabla_\theta q_t(\theta)|$$

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 $\theta_d \equiv d\theta_0 + (1-d)\theta$, $d \in (0,1)$,

$$\mathrm{E}\left(\sup_{\theta \in \mathring{\Theta}: |\theta - \theta_0| < \varepsilon_1} r_t^{(2)}(\theta) \right)$$
$$\leqslant \mathrm{E}\left(\sup_{\theta \in \mathring{\Theta}: |\theta - \theta_0| < \varepsilon_1} \left\{ \frac{M_0}{2} + \mathrm{E}_t\left(\delta(q_t(\theta) - Y_t) \left| \frac{F_t^0(q_t(\theta))}{2} \right.\right.\right.\right.$$
$$\left.\left.\left.\left. - F_t^0(Y_t) + \frac{\alpha}{2} \right| \right) \right\} |\nabla_\theta \lambda(\theta_c)| |\nabla_\theta q_t(\theta)| \right)$$
$$\leqslant \frac{M_0}{2} \sup_{\theta \in \mathring{\Theta}: |\theta - \theta_0| < \varepsilon_1} |\nabla_\theta \lambda(\theta_c)| \left[\mathrm{E}\left(\sup_{\theta \in \mathring{\Theta}: |\theta - \theta_0| < \varepsilon_1} |\nabla_\theta q_t(\theta)| \right) \right.$$
$$\left. + M_0 \mathrm{E}\left(\sup_{\theta \in \mathring{\Theta}: |\theta - \theta_0| < \varepsilon_1} |\nabla_\theta q_t(\theta)| |\nabla_\theta q_t(\theta_d)| \right) \right] < \infty,$$

where the last inequality uses the continuity of $\nabla_\theta \lambda(\cdot)$ on $\mathbb{R}^k$, (A5), 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

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

$R_x(\theta) = [\nabla_\theta x(\theta_0) - \nabla_\theta x(\theta) - \Delta_{\theta\theta} x(\theta)'(\theta_0 - \theta)]/|\theta_0 - \theta|$ 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,

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

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 $C_3 \equiv \sup_{t\geqslant 1}\sup_{y\in\mathbb{R}}\sup_{\theta\in\Theta}\left(\left|\left[\mathbf{I}(q_t(\theta) - Y_t) - \alpha\right][1 - \mathbf{I}(q_t(\theta) - y) - F_t^0(q_t(\theta))\right]\right|^{-1}\right) < \infty$. Combining previous results then gives that with probability one $r_t(\theta) \to 0$ as $\theta \to \theta_0$ and that $\mathrm{E}\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} r_t(\theta)\right) < \infty$. It remains to show that for all θ in a neighborhood of θ_0 we have $T^{-1}\sum_{t=1}^T \Delta_{\theta\theta}\ln f_t^*(Y_t,\theta) \xrightarrow{p} \nabla_\theta \mathrm{E}[\nabla_\theta' L_T(\theta)]$. By Assumption A1, for any $\theta \in \mathring{\Theta}$, the sequence $\{\Delta_{\theta\theta}\ln f_t^*(Y_t,\theta)\}$ is strong mixing (i.e., α-mixing) with α of size $-r/(r-2)$, $r > 2$ (see, e.g., White, 2001, Thm. 3.49, 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 $P(\delta(q_t(\theta) - Y_t)|\alpha - F_t^0(q_t(\theta))| > \eta) \leqslant \mathrm{E}(|\alpha - F_t^0(q_t(\theta))|f_t^0(q_t(\theta)))/\eta \leqslant |\theta - \theta_0|\mathrm{E}(|\nabla_\theta q_t(\theta_a)|f_t^0(q_t(\theta_a))f_t^0(q_t(\theta)))/\eta \leqslant |\theta - \theta_0|M_0^2\mathrm{E}[\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)|]/\eta$, so that in a neighborhood of θ_0, $\delta(q_t(\theta) - Y_t)|\alpha - F_t^0(q_t(\theta))| = o_p(1)$. Similarly, $P([d\delta(q_t(\theta) - Y_t)/dq]|\alpha - F_t^0(q_t(\theta))| > \eta) \leqslant \mathrm{E}(|\alpha - F_t^0(q_t(\theta))|df_t^0(q_t(\theta))/dq)/\eta \leqslant |\theta - \theta_0|M_0M_1\mathrm{E}[\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)|]/\eta$, where the first inequality uses the fact that $\mathrm{E}_t(d\delta(q_t(\theta) - Y_t)/dq) = df_t^0(q_t(\theta))/dq$, a.s. $-P$. From (A.9) we have that for every $t \geqslant 1$,

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

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

$$
\begin{aligned}
&- \frac{\Delta_{\theta\theta} q_t(\theta)}{x(\theta)} \left\{ f_t^0(q_t(\theta))[\alpha - \mathbf{I}(q_t(\theta) - Y_t)] + \delta(q_t(\theta) - Y_t)[\alpha - F_t^0(q_t(\theta))] \right\} \\
&+ o_p(1),
\end{aligned}
$$

in a neighborhood of θ_0, which then gives

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

Hence, for a given $\varepsilon > 0$, there is a positive constant $n_{r,\varepsilon}$ such that $|\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta)|^{r+\varepsilon} \leqslant n_{r,\varepsilon} \left\{ |\Delta_{\theta\theta}\lambda(\theta)|^{r+\varepsilon} (5/2)^{r+\varepsilon} + |\nabla_\theta q_t(\theta) \nabla_\theta q_t(\theta)'|^{r+\varepsilon} C_3^{2(r+\varepsilon)} (M_0^2 + M_1)^{r+\varepsilon} + |\Delta_{\theta\theta} q_t(\theta)|^{r+\varepsilon} C_3^{r+\varepsilon} M_0^{r+\varepsilon} \right\} + o_p(1)$, in a neighborhood of θ_0, and so using Assumption A5 and the fact that $|\Delta_{\theta\theta}\lambda(\theta)| < \infty$ in a neighborhood of θ_0, we have $\mathrm{E}[|\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta)|^{r+\varepsilon}] < \infty$. The weak LLN then follows from Corollary 3.48 in White (2001). This completes the proof of asymptotic normality of the MLE $\tilde{\theta}_T^*$. ■

Proof of Theorem 3. We first show that θ_T^F is consistent for θ_0, i.e., $\theta_T^F - \theta_0 \xrightarrow{p} 0$. We proceed by checking that all the assumptions for consistency used by Komunjer (2005) in her Theorem 3 hold. Assumptions A2 and A3 in Komunjer (2005) are satisfied by imposing our Assumptions 4 and 6(ii), respectively. The α-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 α-mixing with α of the same size $-r/(r-2)$, $r > 2$. Here, we directly impose the mixing of the sequence $\{(Y_t, W_t')'\}$ in our (A1), 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 Assumption 5 and the fact that $\mathrm{E}[\sup_{\theta\in\Theta} |\nabla_\theta q_\alpha(W_t, \theta)|] \leqslant \max\{1, \mathrm{E}[\sup_{\theta\in\Theta} |\nabla_\theta q_\alpha(W_t, \theta)|^2]\} < \infty$. Hence, we can use Theorem 3 in Komunjer (2005), corresponding to the case where the conditional quantile model is correctly specified in our Assumption 3, which proves the consistency of θ_T. Similarly, we derive asymptotic normality by using Corollary 5 in Komunjer (2005). The boundedness of the second derivative of $A(\cdot, \xi_t)$ in Assumption A3′ in Komunjer (2005) is directly implied by our Assumption 6(ii). The moment condition in Assumption A5′ in Komunjer (2005) follows from our Assumption 5. 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 Δ_T^0 and Σ_T^0 are as defined in Theorem 3. ■

Proof of Theorem 4. The proof is in two steps: We first show that V_T^0 is the lower bound of the set of asymptotic matrices $(\Delta_T^0)^{-1} \Sigma_T^0 (\Delta_T^0)^{-1}$. We then show that V_T^0

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press

is attained by a particular M-estimator. It 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 4 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 $F_t \in \mathcal{F}$, $t \geqslant 1$, 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 $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} - V_T^0 = [\alpha(1-\alpha)]^{-1}T^{-1}\sum_{t=1}^T \mathrm{E}[\chi_t\chi_t']$, where for every $t \geqslant 1$, $\chi_t \equiv [f_t^0(q_\alpha(W_t,\theta_0))V_T^0 - \alpha(1-\alpha)f_t(q_\alpha(W_t,\theta_0))(\Delta_T^0)^{-1}]\nabla_\theta q_\alpha(W_t,\theta_0)$. Hence, for any $F_t \in \mathcal{F}$, $t \geqslant 1$, such that $f_t > 0, a.s. - P$ 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 this lower bound V_T^0 is attained by the M-estimator θ_T^*. Note that the true conditional distribution $F_t^0(\cdot)$ is by definition $\mathcal{W}_t$-measurable, thereby satisfying (A6)(i). Moreover, $F_t^0(\cdot)$ is twice continuously differentiable with bounded $f_t^0(y)$ and $|df_t^0(y)/dy|$, which satisfies Assumption A6(ii). Hence, we can apply Theorem 3 to show that, under Assumptions A1–A6, θ_T^* 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 $\Delta_T^0 = T^{-1}\sum_{t=1}^T \mathrm{E}\{[f_t^0(q_\alpha(W_t,\theta_0))]^2\nabla_\theta q_\alpha\ (W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'\}$, so $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} = V_T^0$. ■

https://doi.org/10.1017/S0266466609100038 Published online by Cambridge University Press