Title
Efficientt Conditional Quantile Estimation: The Time Series Case

Authors
Komunjer, Ivana
Vuong, Quang

Publication Date
2006-10-01

2006-10

UNIVERSITY OF CALIFORNIA, SAN DIEGO

DEPARTMENT OF ECONOMICS

Efficient Conditional Quantile Estimation: The Time Series Case

By

Ivana Komunjer
Department of Economics
University of California, San Diego

and

Quang Vuong
Department of Economics
The Pennsylvannia State University

DISCUSSION PAPER 2006-10
October 2006

EFFICIENT CONDITIONAL QUANTILE ESTIMATION: THE TIME SERIES CASE

IVANA KOMUNJER AND QUANG VUONG

ABSTRACT. In this paper we consider the problem of efficient estimation in conditional quantile models with time series data. Our first result is to derive the semiparametric efficiency bound in time series models of conditional quantiles; this is a nontrivial extension of a large body of work on efficient estimation, which has traditionally focused on models with independent and identically distributed data. In particular, we generalize the bound derived by Newey and Powell (1990) to the case where the data is weakly dependent and heterogeneous. We then proceed by constructing an M-estimator which achieves the semiparametric efficiency bound. Our efficient M-estimator is obtained by minimizing an objective function which depends on a nonparametric estimator of the conditional distribution of the variable of interest rather than its density.

Keywords: semiparametric efficiency, time series models, dependence, parametric submodels, conditional quantiles.

Affiliations and Contact information. Komunjer: Department of Economics, University of California San Diego (komunjer@ucsd.edu). Vuong: Department of Economics, Penn State University (qvuong@psu.edu).

Acknowledgments: Earlier versions of this paper were presented at EEA/EESM 2003 meetings in Stockholm, NSF/NBER 2004 Time Series conference at SMU and CEME 2005 conference at MIT. Many thanks to Rob Engle, Clive Granger, Jin Hahn, Bruce Hansen, Cheng Hsiao, Guido Imbens, Guido Kuersteiner, Guy Laroque, Essie Maasoumie, Roger Moon, Whitney Newey, James Powell, Bernard Salanié, Ruey Tsay, Hal White and to all the participants at Ohio State, Penn State, UC Berkeley, UC San Diego, CREST/INSEE and USC econometric seminars.

1. Introduction

The purpose of this paper is to study the problem of asymptotically efficient estimation in models for conditional quantiles. We provide answers to the following closely related questions: what is the semiparametric efficiency bound for the parameters of a given conditional quantile, when the data is weakly dependent and heterogeneous? Is efficient estimation possible in such models, and if so, what is an efficient conditional quantile estimator?

The computation of semiparametric efficiency bounds in models with conditional moment restrictions—which include the one studied here—has been considered by numerous authors (Chamberlain, 1986, 1987; Robinson, 1987; Hansen, Heaton, and Ogaki, 1988; Newey, 1990a,b, 1993; Hahn, 1997; Bickel, Klaassen, Ritov, and Wellner, 1998; Brown and Newey, 1998; Ai and Chen, 2003; Cosslett, 2004; Newey, 2004). Our contribution to this large literature is twofold. First, we derive the semiparametric efficiency bound in models with a conditional quantile restriction allowing the data to be weakly dependent and/or heterogeneous. Second, we propose a new estimator for conditional quantiles which actually attains the semiparametric efficiency bound. Our results are important because they do not require independence nor identical distribution of the data.

The first of those assumptions—independence—has been prevalent in the existing literature on efficient estimation, for reasons which pertain to the very definition of the semiparametric efficiency bound. Depending on how we characterize the bound—as an "infimum" or as a "supremum"—there are two approaches to its computation. Most of the above literature, with the exception of Chamberlain (1987), has used the "infimum" approach, which can be summarized as follows.

Consider a model in which the parameter vector of interest θ is identified via a conditional moment restriction. Assume that the model is regular in the sense of Begun, Hall, Huang, and Wellner (1983) and Newey (1990b). A familiar approach to estimating θ is by using semiparametric estimators such as GMM (Hansen, 1982), M– (Huber, 1967) or instrumental variable estimators. Associated with the choice of a particular semiparametric estimator is its covariance matrix. Hence, to the set of all semiparametric estimators corresponds a set of positive semidefinite matrices. The crucial property of this set is its orthogonal structure (Bickel, 1982; Begun, Hall, Huang, and Wellner, 1983; Chamberlain, 1986; Newey, 1990b): any matrix Ω in this set can be written as a covariance matrix of a Gaussian random

variable—with a positive semidefinite matrix V—plus an independent noise. The matrix V which is the infimum of this set, is the semiparametric efficiency bound for θ.

This characterization of the semiparametric efficiency bound is the starting point of the "infimum" approach to its computation. Essentially geometric, the "infimum" approach uses projection arguments to find V. As such, it requires certain orthogonality conditions, which in econometric terms correspond to the requirement that the random variables involved be independent (Bickel, 1982). Hence, most of the "infimum" approach literature has exclusively focused on models with independent observations.[1] In models in which we relax the independence assumption, the projection arguments are difficult to implement, which makes dealing with time series data difficult. Consideration such as those have lead Ai and Chen (2003), for example, to conclude: "although our results [...] can be easily extended to weakly dependent time series data, the problem of semiparametric efficiency bound with time series data is nontrivial."

In this paper, we use the alternative—"supremum"—approach pioneered by Chamberlain (1987). In his seminal paper on semiparametric efficiency bounds in models with conditional moment restrictions, Chamberlain (1987) compares the asymptotic distribution of an efficient GMM estimator—efficient in the sense of Hansen (1982)—with that of a maximum likelihood estimator (MLE). The key property of the MLE is that it is efficient, when correctly specified. Hence any MLE in which the specified likelihood is consistent with the conditional moment restriction and which contains the data generating process, needs to have its asymptotic covariance matrix smaller than the semiparametric efficiency bound. In other words, the semiparametric efficiency bound can be defined as the supremum of asymptotic covariance matrices of all parametric submodels which satisfy the conditional moment restrictions and contain the data generating process—this is the key insight behind Stein's (1956) characterization of semiparametric efficiency bounds and the starting point of the "supremum" approach.

[1]Hansen, Heaton, and Ogaki (1988) is an important exception. Their approach however is based on the assumption that some transformation—forward filter—of the moment function used in the conditional moment restriction is serially uncorrelated (see their equation (4.2) and the discussion thereof). Hence, unless the parameters involved in the forward filter transformation are known, the approach of Hansen, Heaton, and Ogaki (1988) is not applicable. For example, in models with conditional moment restrictions in which the variables follow an ARMA(p, q) process—with lags p and q known—one needs to know the q MA parameters in order to construct the forward filter.

Chamberlain (1987) implements the "supremum" approach in the case where the random variables involved in the conditional moment restriction are independent and identically distributed (iid). In the iid case, the efficient (in the sense of Hansen, 1982) GMM estimator and the MLE obtained when the data is generated from a multinomial distribution are both asymptotically normally distributed with asymptotic covariance matrices respectively equal to Ω and I^{-1}, where I is the Fisher information matrix of the multinomial model. When the data has finite support, Chamberlain (1987) shows that Ω and I^{-1} are the same. Hence, they must be equal to the semiparametric efficiency bound V. Given that any distribution can be approximated arbitrarily well by a multinomial distribution, the general expression for the bound follows. The iid assumption plays an important role in Chamberlain's (1987) construction of the semiparametric bounds; without it the multinomial approximation is no longer valid, making the extension of Chamberlain's (1987) results to time series data difficult.

The first contribution of this paper is to extend Chamberlain's (1987) results to weakly dependent data, by using the "supremum" characterization of the semiparametric efficiency bound, initially due to Stein (1956). In particular, we focus on models with conditional quantile restrictions. In such models, there is no published work prior to ours on asymptotically efficient estimation which would allow for the data to be weakly dependent. Hence, our first contribution is to fill the gaps in the extant literature on efficient conditional quantile estimation (Newey and Powell, 1990; Koenker and Zhao, 1996; Zhao, 2001) and derive the semiparametric efficiency bound in weakly dependent time series models with conditional quantile restrictions.

Our "supremum" approach is somewhat different from that used by Chamberlain (1987). We start by constructing a matrix V which is a potential candidate for the semiparametric efficiency bound. Such candidate matrix is obtained as a minimum within a family of asymptotic covariance matrices of conditional quantile M–estimators that are consistent for the parameters of a correctly specified conditional quantile model. Once the candidate matrix V in hand, we follow the insightful approach by Stein (1956), and look for a parametric submodel that is "as difficult" as the semiparametric model. In other words, we construct a fully parametric model that satisfies the conditional quantile restriction, contains the data generating process and in which the inverse of the Fisher's information matrix equals V.

This second step is what distinguishes our work from the rest of the literature on asymptotically efficient estimation—specifically, we are able to analytically derive the least favorable parametric submodel.

Our result on the semiparametric efficiency bound is general: we derive it under the sole assumption that the model satisfies the conditional quantile restriction. In particular, when constructing V, we do not make any additional assumptions regarding the properties of the residuals from the (nonlinear) quantile regression: they can be dependent and nonidentically distributed. Hence, for the first time in the literature on efficient estimation, we are able to derive the semiparametric efficiency bound in conditional quantile models with time series data that are dependent and conditionally heteroskedastic.

The second contribution of this paper is to propose a new conditional quantile estimator that is efficient. We note that the problem of constructing an efficient estimator is even more difficult than that of computing the semiparametric efficiency bound. Though to some extent applicable to time series data, the projection methods used in the "infimum" approach shed no light on how to construct efficient estimators. As already pointed out by Hansen, Heaton, and Ogaki (1988), "although [they] delineate the sense of approximation required for the sequences of GMM estimators to get arbitrarily close to the efficiency bound, [they] do not show how to construct estimators that actually attain the efficiency bound." It is an open question whether the procedures along the lines of Newey (1990a,b, 1993, 2004) can be extended to models with time series data. Our second contribution to the literature on efficient estimation is to show how—at least in models with conditional quantile restrictions—the "supremum" approach naturally leads to estimators that are efficient.

Standard approaches to constructing an efficient estimator are as follows: given a consistent estimator of the parameter of interest θ, take a step away from it in a direction predicted by the efficient score; the resulting estimator is then efficient. An example of this construction method is Newey and Powell's (1990) "one-step" estimator for the parameters of a quantile regression. Alternatively, instead of taking a step away from an initial consistent estimator of θ, we can use it to construct a set of weights—functions of the efficient score—and compute the corresponding weighted estimator; the weighted estimator is also efficient. An example of this method is Zhao's (2001) weighted conditional quantile estimator. More recently, extending the conditional empirical likelihood (CEL) approach by Kitamura, Tripathi, and Ahn (2004), Otsu (2003) constructs an efficient estimator in the quantile regression model in the iid case.

We propose an efficient conditional quantile MINPIN-type estimator (Andrews, 1994a) whose construction differs from the previous ones, in two ways. First, our efficient estimator does not require a preliminary consistent estimate of the parameter of interest, hence it is similar to the estimator proposed by Otsu (2003). While Otsu's (2003) efficient estimator is based on the empirical likelihood principle, our efficient estimator is obtained by minimizing an efficient M–objective function. Second, our efficient estimator depends on a nonparametric estimate of the true conditional distribution, unlike Newey and Powell's (1990) and Zhao's (2001) efficient estimators which depend on nonparametric estimates of the true conditional density. For these two reasons, we can expect our efficient estimator to behave better in small samples than the efficient estimators proposed by Newey and Powell (1990) and Zhao (2001). In particular, whenever it is easier to estimate the conditional distributions than densities (Hansen, 2004a,b), we would expect our efficient estimator to perform better than the existing ones.

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. In Section 4, we derive the minimum bound of the above family of matrices and show that it corresponds to the semiparametric efficiency bound. An efficient conditional quantile estimator is constructed in Section 5, which concludes the paper. We relegate all the proofs to the end of the paper.

2. Setup

2.1. **Notation.** Consider a stochastic sequence (a time series) $X \equiv \{X_t, t \in \mathbb{N}\}$ defined on a probability space $(\Omega, \mathcal{B}, P)$ where $X : \Omega \to \mathbb{R}^{(m+1)\mathbb{N}}$ and $\mathbb{R}^{(m+1)\mathbb{N}}$ is the product space generated by taking a copy of $\mathbb{R}^{m+1}$ for each integer, i.e. $\mathbb{R}^{(m+1)\mathbb{N}} \equiv \times_{t=1}^{\infty}\mathbb{R}^{m+1}$, $m \in \mathbb{N}$. We partition the random vector X_t as $X_t = (Y_t, W_t')'$ and are interested in the distribution of its first (scalar) component, denoted Y_t, conditional on the random m-vector W_t. In particular, we allow W_t to contain lagged values of Y_t—particularly interesting for time series applications—together with other (exogenous) components. The family of subfields $\{\mathcal{W}_t, t \in \mathbb{N}\}$ with $\mathcal{W}_t \equiv \sigma(W_1, \ldots, W_t)$ corresponds to the information set generated by the sequence of conditioning vectors up to time t.

We use standard notations and let $P(Y_t \in A|\mathcal{W}_t)$ denote the conditional distribution of Y_t, with A an element of the Borel σ-algebra on $\mathbb{R}$. To simplify, we assume that for any $T \geqslant 1$, the joint distribution of $(Y_1, W_1, \ldots, Y_T, W_T)$ has a strictly positive continuous density p_T on $\mathbb{R}^{(m+1)T}$ so that conditional densities are everywhere defined.[2] Then, 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. $F_t^0(y) \equiv P(Y_t \leqslant y|\mathcal{W}_t)$ for every $y \in \mathbb{R}$, and we call $f_t^0(\cdot)$ the corresponding conditional probability density. Of course, $F_t^0(\cdot)$ (like $f_t^0(\cdot)$) is unknown and we assume that it belongs to $\mathcal{F}$ which is the set of all absolutely continuous distribution functions with continuously differentiable densities on $\mathbb{R}$. Throughout the paper we assume that for every t, $1 \leqslant t \leqslant T, T \geqslant 1$, $f_t^0(\cdot)$ and its derivative are bounded so that 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 $\sup_{t \geqslant 1} \sup_{y \in \mathbb{R}} |df_t^0(y)/dy| \leqslant M_1 < \infty$.

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_∞-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 $f : E \to \mathbb{R}, V \mapsto f(V)$, 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 $\sum_{|r| \leqslant R} \frac{D^r f(V_0)}{r!}(V - V_0)^r \equiv \sum_{k=0}^R \sum_{j_1, \ldots, j_k \in (1, \ldots, n)^k} \frac{1}{k!} \frac{\partial^k f(V_0)}{\partial V_{j_1} ... \partial V_{j_k}} (V_{j_1} - V_{0j_1}) ... (V_{j_k} - V_{0j_k})$, 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 $\mathbb{1} : \mathbb{R} \to [0, 1]$ 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 $\delta : \mathbb{R} \to \mathbb{R}$, with $\mathbb{1}(x) = \int_a^x d\delta$, where a is an arbitrary (possibly infinite) negative constant, $a \leqslant 0$.

[2]This excludes the possibility that W_t contains indicator functions of lags of Y_t or other variables, for example.

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. In recent years, conditional quantiles have been of particular interest in both applied and theoretical work in economics in which numerous choices for the conditioning variables have been proposed.[3] In order to keep our analysis both simple and general, we introduce the following notation: for a given $\alpha \in (0,1)$, let $\mathcal{M}$ denote a model for the conditional α-quantile of Y_t, $\mathcal{M} \equiv \{q_\alpha(W_t,\theta)\}$, with an unknown parameter θ in Θ, where Θ is a compact subset of $\mathbb{R}^k$ with non-empty interior, $\mathring{\Theta} \neq \emptyset$. In what follows we restrict our attention to conditional quantile models $\mathcal{M}$ in which the set of following conditions is satisfied:

(A0) *(i) the model 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, $1 \leqslant t \leqslant T, T \geqslant 1$, 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 Θ $a.s.-P$; (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 (A0) is fairly standard and generally verified for a wide variety of conditional quantile models. In what follows, we shall always assume that $\mathcal{M}$ is a conditional quantile model in which properties (A0)(i)-(iii) above hold. Further, for any given $\mathcal{M}$ we shall denote by $\mathcal{Q}$ the range of q_α, i.e. $\mathcal{Q} \equiv \{q_t \in \mathbb{R} : q_t = q_\alpha(W_t,\theta), \theta \in \Theta, W_t \in \mathbb{R}^m\}$, $\mathcal{Q} \subseteq \mathbb{R}$.

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 θ_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:

(A1) *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, $1 \leqslant t \leqslant T, T \geqslant 1$.*

[3]Since the seminal work by Koenker and Bassett (1978), numerous authors have studied the problems of conditional quantile estimation (Koenker and Bassett, 1978; Powell, 1984, 1986; Newey and Powell, 1990; Pollard, 1991; Portnoy, 1991; Koenker and Zhao, 1996; Buchinsky and Hahn, 1998; Khan, 2001; Cai, 2002; Kim and White, 2003; Komunjer, 2005b) and specification testing (Koenker and Bassett, 1982; Zheng, 1998; Bierens and Ginther, 2001; Horowitz and Spokoiny, 2002; Koenker and Xiao, 2002; Kim and White, 2003; Angrist, Chernozhukov, and Fernandez-Val, 2006). An excellent review of applications of quantile regressions in economics (Buchinsky, 1994; Chernozhukov and Hong, 2002; Angrist, Chernozhukov, and Fernandez-Val, 2006) can be found in Koenker and Hallock (2001).

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

3. M–estimators for conditional quantiles

In this paper we consider a particular family of conditional quantile estimators known as M–(or extremal) estimators (Huber, 1967). M–estimators for θ_0, denoted θ_T, are obtained by minimizing criterion functions $\Psi_T(\theta)$ 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, $1 \leqslant t \leqslant T, T \geqslant 1$, φ 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 ξ_t can be thought of as a shape parameter of the objective function φ. We assume the following:

(A2) *(i) for every t, $1 \leqslant t \leqslant T, T \geqslant 1$, ξ_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_t).*

By assumption (A2)(i), the random variable ξ_t is allowed to depend only on variables contained in $\mathcal{W}_t$. In other words, the functional form (or shape) of φ cannot depend on any variable that is observed after time t. We shall see in subsequent sections that the $\mathcal{W}_t$-measurability of ξ_t is not trivially satisfied. In particular, if we consider objective functions φ 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 an example—then (A2)(i) fails to hold. The requirement (A2)(ii) that, for given realizations of Y_t and ξ_t, φ be twice continuously differentiable with respect to q_t on $\mathcal{Q}$ $a.s. - P$, allows for objective functions such as $|Y_t - q_t|$ or $[\alpha - \mathbb{1}(q_t - Y_t)](Y_t - q_t)$, for example. Note that in those two cases the shape ξ_t of φ 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). If in addition to (A2), 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_t, and such that $\int_{\mathbb{R}} \exp[c(y, \xi_t) - \varphi(y, q_t, \xi_t)] dy = 1$ for all $(q_t, \xi_t) \in \mathcal{Q} \times E_t$, then we can let $l_t(\cdot, q_t) \equiv \exp[c(\cdot, \xi_t) - \varphi(\cdot, q_t, \xi_t)]$, and $l_t(\cdot, q_t)$ can be interpreted as the (quasi) likelihood of Y_t conditional on $\mathcal{W}_t$. Hence, any minimum θ_T of the function $\Psi_T(\theta)$ above, is also a maximum of the (quasi) log-likelihood function $L_T(\theta)$, $L_T(\theta) \equiv T^{-1} \sum_{t=1}^{T} \ln l_t(Y_t, q_\alpha(W_t, \theta))$ (Komunjer, 2005b). However, due to the above

"integrability" constraint on $\varphi(\cdot, q_t, \xi_t)$, the class of QMLEs is smaller than that of M–estimators.[4] We shall see in subsequent sections that this difference plays a greatly important role for efficient conditional quantile estimation. We now focus on M–estimators for θ_0 that are consistent.

3.1. **Class of consistent M–estimators.** What are necessary conditions for the M–estimator θ_T satisfying (A2), to be consistent for the true conditional quantile parameter θ_0 in (A1)? The key idea behind the answer to this question is fairly simple. Assume that the process $\{X_t\}$ and the functions $\varphi(\cdot, \cdot, \xi_t)$ are such that $\theta_T - \theta_T^0 \xrightarrow{p} 0$, where θ_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 $\mathring{\Theta}$.[5] Then a necessary requirement for consistency of θ_T is that $\theta_T^0 - \theta_0 \to 0$ as T becomes large. In what follows, we restrict our attention to estimators θ_T such that θ_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 θ_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 $\theta_T^0 \to \theta_0$.[6] This implies that θ_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 ξ_t of the objective function φ depends on observations up to time T—hence is not $\mathcal{W}_t$-measurable—as in the case of the estimator $\hat{\theta}_T$ proposed in Section 5. In that case, $\hat{\theta}_T$ is consistent provided the difference between its (M–) objective function $\hat{\Psi}_T$ and an (M–) objective function Ψ_T^* derived in Theorem 3, converges uniformly to zero with T.

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

(A3) *$\{X_t\}$ 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) $\sup_{\theta\in\Theta} |\Psi_T(\theta) - E[\Psi_T(\theta)]| \xrightarrow{p} 0$.*

[4] We call $\int_{\mathbb{R}} \exp[c(y, \xi_t) - \varphi(y, q_t, \xi_t)]dy = 1$ for all $(q_t, \xi_t) \in \mathcal{Q} \times E_t$ the "integrability" constraint. This requirement is stronger than $\exp[-\varphi(\cdot, q_t, \xi_t)]$ being integrable with respect to the Lebesgue measure on $\mathbb{R}$.

[5] θ_T^0 is also called the pseudo-true value of the parameter θ.

[6] See White (1994, p.69-70) for a discussion of the requirement $\theta_T^0 = \theta_0$.

Note that the above are not primitive conditions for consistency of θ_T. For example, the integrability of $D^r\varphi(Y_t, q_\alpha(W_t,\theta), \xi_t)$ with respect to the probability P implied by (A3)(i) involves more primitive conditions on the existence of different moments of Y_t, W_t and ξ_t. Condition (A3)(ii) states that θ^0_∞ is a minimum of $E[\Psi_T(\theta)]$ and that this minimum is moreover unique. The first requirement involves more primitive conditions on $\partial\varphi/\partial q_t$, $\partial^2\varphi/\partial q_t^2$ and $\nabla_\theta q_\alpha$, which depend on the shape ξ_t of φ and the functional form of q_α. For example, a sufficient set of conditions for θ^0_∞ 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 $T^{-1}\sum_{t=1}^T E[\Delta_{\theta\theta}\varphi(Y_t, q_\alpha(W_t,\theta^0_\infty),\xi_t)] \gg 0$ (Schwartz, 1997). Finally, the uniform convergence condition (A3)(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 (A3)(iii) are primitive assumptions on the dependence structure and heterogeneity of the process $\{X_t\}$, and on the properties of $\varphi(Y_t, q_\alpha(W_t,\cdot),\xi_t)$. A simple example is one where $\{X_t\}$ is iid and the functions $\varphi(Y_t, q_\alpha(W_t,\cdot),\xi_t)$ are Lipshitz-L_1 $a.s.-P$ on Θ (see, e.g., Definition A.2.3 in White, 1994).

The above pseudo-true value θ^0_∞ of the parameter θ equals the true value θ_0 if and only if, for any $T \geqslant 1$, θ_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 (A0), (A2) and (A3) hold. If the true parameter θ_0 satisfies the conditional moment condition in (A1), then the M-estimator θ_T is consistent for θ_0, i.e. $\theta_T - \theta_0 \xrightarrow{p} 0$, if and only if there exist a real function $A(\cdot,\cdot) : \mathbb{R}\times E_t \to \mathbb{R}$ that is twice continuously differentiable and strictly increasing with respect to its first argument (q_t or Y_t) $a.s.-P$ on $\mathcal{Q}\times E_t$, and a real function $B(\cdot,\cdot) : \mathbb{R}\times E_t \to \mathbb{R}$, such that $\varphi(Y_t, q_t, \xi_t) = [\alpha - \mathbb{1}(q_t - Y_t)][A(Y_t,\xi_t) - A(q_t,\xi_t)] + B(Y_t,\xi_t)$, $a.s.-P$ on $\mathbb{R}\times\mathcal{Q}\times E_t$, for every t, $1 \leqslant t \leqslant T, T \geqslant 1$.*[7]

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 θ_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

$$(1)\quad \varphi(Y_t, q_\alpha(W_t,\theta),\xi_t) = [\alpha - \mathbb{1}(q_\alpha(W_t,\theta) - Y_t)][A(Y_t,\xi_t) - A(q_\alpha(W_t,\theta),\xi_t)] + B(Y_t,\xi_t),$$

[7]The real functions A and B in Theorem 1 need not have the same shape parameter: we can let $\xi_t \equiv (\xi'_{At}, \xi'_{Bt})'$ where ξ_{At} and ξ_{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 may not affect the shape of B and vice-versa.

$a.s.-P$, for every t, $1 \leqslant t \leqslant T$. Using objective functions of this form is also a sufficient condition for θ_T to be consistent for the true parameter θ_0 of a correctly specified model for the conditional α-quantile.

Given that we restrict our attention to objective functions in which (A2)(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 $\partial\varphi/\partial q_t$, and therefore has no effect on the optimum of Ψ_T. The fact that $A(\cdot, \xi_t)$ is necessarily strictly increasing $a.s.-P$ on $\mathcal{Q}$, comes from the requirement (A3)(ii) that θ_∞^0 be an interior minimum of $E[\Psi_T(\theta)]$ on Θ. 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 Ψ_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;[8] (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 ω_t is some nonnegative weight, as well as its censored version for which $A(y, \xi_t) = \omega_t \max\{y, c_t\}$.

In particular, the class of objective functions Ψ_T leading to consistent conditional quantile M–estimators is larger than that leading to consistent QMLEs. In order to simplify the comparison between M–estimators and QMLEs, assume that at any point in time t, $1 \leqslant t \leqslant T, T \geqslant 1$, the conditional α-quantile of Y_t can take any real value, so $\mathcal{Q} = \mathbb{R}$. As

[8]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 elegantly 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 θ_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 (A3)(ii), and hence affects the consistency of θ_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.

pointed out previously, the main difference between the two classes of estimators lies in the "integrability" condition on the pseudo-densities. Compare the objective function in Theorem 1 with the family of tick-exponential pseudo-densities which give consistent QMLEs for θ_0 (Komunjer, 2005b): $f_\alpha(Y_t, q_t, \xi_t) \equiv \alpha(1-\alpha)a(Y_t,\xi_t)\exp\{[\mathbb{1}(q_t - Y_t) - \alpha][A(Y_t,\xi_t) - A(q_t,\xi_t)]\}$ with $A(\cdot,\xi_t)$ twice continuously differentiable and strictly increasing $a.s.-P$ on $\mathbb{R}$, with derivative $a(y,\xi_t) \equiv \partial A(y,\xi_t)/\partial y$.[9] For $f_\alpha(\cdot, q_t, \xi_t)$ to be a probability density on $\mathbb{R}$, we need $\lim_{y\to\pm\infty} A(y,\xi_t) = \pm\infty$, for any t, $1 \leqslant t \leqslant T, T \geqslant 1$.[10] This limit condition restricts the possible choice of functions $A(\cdot,\xi_t)$ in Theorem 1.

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

$$A(y,\xi_t^F) \equiv F_t(y), \tag{2}$$

for any $y \in \mathbb{R}$. Note that the parameter ξ_t^F in the objective function $A(\cdot,\xi_t^F)$ in Equation (2) 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 θ_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

$$\varphi(Y_t, q_\alpha(W_t,\theta),\xi_t^F) \equiv [\alpha - \mathbb{1}(q_\alpha(W_t,\theta) - Y_t)][F_t(Y_t) - F_t(q_\alpha(W_t,\theta))], \tag{3}$$

is consistent for θ_0; however, the corresponding function $A(\cdot,\xi_t^F)$ in Equation (2), bounded between 0 and 1, does not satisfy the above limit condition. As a consequence, the class of consistent QMLEs is strictly smaller than that of consistent M–estimators. In subsequent sections we show that the limit restrictions on $A(\cdot,\xi_t^F)$ play a particularly important role for efficient conditional quantile estimation, by constructing an efficient M–estimator whose objective function is of the form (3).

To resume, we have shown that an M–estimator θ_T that satisfies (A2) is consistent for θ_0, only if the objective functions $\varphi(\cdot,\cdot,\xi_t)$ are of the form given in Theorem 1. The conditions provided in Theorem 1 are not only necessary but also sufficient for consistency. From the

[9]It is straightforward to see that $\varphi(Y_t, q_t, \xi_t)$ in Theorem 1 and $f_\alpha(Y_t, q_t, \xi_t)$ in Komunjer (2005b) have the same optimum.

[10]The limit conditions on $A(\cdot,\xi_t)$ directly follow from the quantile restriction $\int_{-\infty}^{q_t} f_\alpha(y, q_t, \xi_t)dy = \alpha$, which is equivalent to $(1-\alpha)\exp[-(1-\alpha)A(q_t,\xi_t)]\int_{-\infty}^{q_t} a(y,\xi_t)\exp[(1-\alpha)A(y,\xi_t)]dy = 1$, so that, upon the change of variable $u \equiv A(y,\xi_t)$, necessarily $A(q_t,\xi_t) \to -\infty$ as $q_t \to -\infty$. Combining the above quantile restriction with the condition $\int_{\mathbb{R}} f_\alpha(y, q_t, \xi_t)dy = 1$ yields the result for the limit in $+\infty$ by a similar reasoning.

functional form of $\varphi(\cdot,\cdot,\xi_t)$ in Equation (1), it follows that the asymptotic properties of θ_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 θ_T optimal, we need the asymptotic distribution of the latter. We derive the asymptotic distribution of θ_T in the next section.

3.2. **Asymptotic Distribution.** We start by imposing the following assumptions, in addition to (A0)-(A2):

(A4) *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* $0 < \partial A(q_t,\xi_t)/\partial q_t \leqslant K$ *and* $|\partial^2 A(q_t,\xi_t)/\partial q_t^2| \leqslant L$, *a.s.* $- P$ *on* $\mathcal{Q} \times E_t$;
(A5) θ_0 *is an interior point of* Θ;
(A6) *the sequence* $\{(Y_t, W_t')'\}$ *is* α*-mixing with* α *of size* $-r/(r-2)$, *with* $r > 2$;
(A7) *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* $\sup_{1\leqslant t\leqslant T,T\geqslant 1} E[|A(Y_t,\xi_t)|^{r+\epsilon}] < \infty$.

The above assumptions provide a set of sufficient conditions for the asymptotic normality of θ_T that are primitive, unlike the ones for consistency in (A3). In addition to (A1) and (A2), we now require the functions $A(\cdot,\xi_t)$ to have bounded first and second derivatives (A4). The boundedness property is used to show that $\varphi(Y_t, q_\alpha(W_t,\cdot),\xi_t)$ are Lipshitz-L_1 on Θ $a.s. - P$. This implies that any pointwise convergence in θ becomes uniform on Θ. 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 θ. This elegant convexity approach has, for example, been used by Pollard (1991), Hjort and Pollard (1993) and Knight (1998) to derive asymptotic normality of the standard Koenker and Bassett's (1978) quantile regression estimator. 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 θ, no matter which conditional quantile model q_α in (A0) we choose.[11] Unfortunately, the convexity in θ 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 proposed in Equation (3). Therefore, we cannot rely on the convexity argument in our asymptotic normality proof.

[11]Recall that $\varphi(Y_t, q_\alpha(W_t,\cdot),\xi_t)$ is convex in a neighborhood of θ_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 model q_α in (A0), only if the functions $A(\cdot,\xi_t)$ have zero convexity, i.e. are linear.

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 the ones in (A4).

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. α-mixing) assumption in (A6) allows for a wide variety of dependence structures (White, 2001). Assumption (A6) is further accompanied by a series of moment conditions in (A7) 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 (A7) 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 θ_T is given in the following theorem.

Theorem 2 (Asymptotic Distribution). *Under (A0)-(A2) and (A4)-(A7), we have* $(\Sigma_T^0)^{-1/2} \Delta_T^0 \sqrt{T}(\theta_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$*, where* $\Delta_T^0 \equiv T^{-1} \sum_{t=1}^T E[a(q_\alpha(W_t, \theta_0), \xi_t) f_t^0(q_\alpha(W_t, \theta_0)) \times \nabla_\theta q_\alpha(W_t, \theta_0) \nabla_\theta q_\alpha(W_t, \theta_0)']$ *and* $\Sigma_T^0 \equiv T^{-1} \sum_{t=1}^T \alpha(1-\alpha) E[(a(q_\alpha(W_t, \theta_0), \xi_t))^2 \nabla_\theta q_\alpha(W_t, \theta_0) \times \nabla_\theta q_\alpha(W_t, \theta_0)']$*, where* $a(q_t, \xi_t) \equiv \partial A(q_t, \xi_t)/\partial q_t$ *a.s.* $- P$ *on* $\mathcal{Q} \times E_t$*.*

In particular, the M–estimator θ_T^F proposed in Equation (3) 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 (A7) are less stringent for θ_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 θ_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 (A7)(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, θ_T^F still converges in distribution at the usual $\sqrt{T}$ rate.

Using the results of Theorem 2, the asymptotic distribution of θ_T^F is: $(\Sigma_T^{0,F})^{-1/2} \Delta_T^{0,F} \times \sqrt{T}(\theta_T^F - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$, with $\Delta_T^{0,F} \equiv T^{-1} \sum_{t=1}^T E[f_t(q_\alpha(W_t, \theta_0)) f_t^0(q_\alpha(W_t, \theta_0)) \nabla_\theta q_\alpha(W_t, \theta_0) \times \nabla_\theta q_\alpha(W_t, \theta_0)']$ 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 Equation (2)—hence in Equation (3)—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 $\Delta_T^{0,F}$ and $\Sigma_T^{0,F}$. In particular, this result suggests that appropriate choices of $F_t(\cdot)$ in Equation (3) lead to efficiency improvements over Koenker and Bassett's (1978) conditional quantile estimator. Specifically, when the values of $f_t(\cdot)$ and of 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. In the next section we make our heuristic argument more rigorous by exploring the questions of minimum variance and efficient estimation in more details.

4. Semiparametric Efficiency Bound

Our first step in discussing the asymptotic efficiency of conditional quantile estimators is to 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, i.e. to have asymptotic variances which for some true parameter value are smaller than that of the maximum-likelihood estimator. Superefficiency is ruled out by our continuity assumptions on $f_t^0(\cdot)$, $q_\alpha(W_t, \cdot)$ in (A0)(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.

Theorem 3 (**Minimum Asymptotic Variance**). *Assume that (A0)-(A2) and (A4)-(A7) 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*

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

Moreover, an M-estimator θ_T^* *of the parameter* θ_0 *obtained by minimizing* $\Psi_T^*(\theta) \equiv T^{-1}\sum_{t=1}^T \varphi(Y_t, q_\alpha(W_t,\theta), \xi_t^*)$ *attains* V_T^0, $(V_T^0)^{-1/2}\sqrt{T}(\theta_T^* - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$, *if and only if* $\varphi(Y_t, q_t, \xi_t^*) = [\alpha - \mathbb{1}(q_t - Y_t)][F_t^0(Y_t) - F_t^0(q_t)], a.s.-P$, *on* $\mathbb{R}\times\mathcal{Q}\times E_t$, *for every* t, $1 \leqslant t \leqslant T, T \geqslant 1$.

Theorem 3 shows two important results. Firstly, the matrix V_T^0 is the minimum of the asymptotic variances of all the consistent and asymptotically normal M–estimators of θ_0 that satisfy (A2). In other words, for any ξ_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, there exists a unique M–estimator θ_T^* whose asymptotic

covariance matrix equals V_T^0. This 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

$$\varphi(Y_t, q_\alpha(W_t,\theta), \xi_t^*) = [\alpha - \mathbb{1}(q_\alpha(W_t,\theta) - Y_t)][F_t^0(Y_t) - F_t^0(q_\alpha(W_t,\theta))], \tag{4}$$

$a.s.-P$, for every t, $1 \leqslant t \leqslant T, T \geqslant 1$. In particular, the shape ξ_t^* of the optimal objective function in Equation (4) is that of the true conditional distribution $F_t^0(\cdot)$, which is stochastic and $\mathcal{W}_t$-measurable as required by Assumption (A0)(i). Even though our estimator θ_T^* satisfies all the assumptions in (A2), its computation is not feasible in reality. In order to construct θ_T^* we would need to know the true conditional distribution $F_t^0(\cdot)$ whose inverse—the conditional α-quantile—is the very object that we are trying to estimate. We come back to this important feasibility issue in Section 5.

What Theorem 3 does not show is whether V_T^0 is also the semiparametric efficiency bound for θ_0, in addition to being the minimum of the set of asymptotic covariance matrices of consistent and asymptotically normal M–estimators.

4.1. **Stein's (1956) approach: an example.** In order to show that V_T^0 in Theorem 3 is the semiparametric efficiency bound in the time series models satisfying the conditional quantile restriction (A1), we follow the ingenious approach by Stein (1956). Stein's (1956) original concern was the possibility of estimating the true parameter adaptively: can we estimate the parameter θ_0 in the conditional quantile restriction (A1) 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?

If the set of true conditional densities $f^0 \equiv \{f_t^0(\cdot), 1 \leqslant t \leqslant T, T \geqslant 1\}$ in the conditional quantile restriction (A1) were known up to a finite dimensional parameter, then we could easily construct an estimate of θ_0 whose asymptotic covariance matrix attains the classical Cramer-Rao bound. As an illustration, consider the following conditionally heteroskedastic (CH) model with linear heteroskedasticity

$$Y_t = \beta_0' V_t + (1 + |\gamma_0' R_t|)U_t, \tag{5}$$

where $W_t \equiv (V_t', R_t')'$, the process $\{(Y_t, W_t')'\}$ is α-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 $E(U_t) = 0$ and $E(U_t^2) = 1$, and where β_0 and γ_0 denote the true values of the parameters $\beta \in B \subseteq \mathbb{R}^b$ and $\gamma \in \Gamma \subseteq \mathbb{R}^c$. Letting $V_t \equiv (1, Y_{t-1})'$ and $R_t \equiv U_{t-1}$

the above equation reduces to a well-known AR(1)-ARCH model, for example (Koenker and Zhao, 1996).[12]

4.1.1. *Case 1: no nuisance parameter.* Assume that the distribution function $H_0(\cdot)$ is known. In financial applications $h_0(\cdot)$ is typically chosen to be a standardized Gaussian or Student-t density. The conditional density of Y_t in the CH model (5) then equals $f_t^0(y) = (1 + |\gamma_0' R_t|)^{-1} h_0([1 + |\gamma_0' R_t|]^{-1}[y - \beta_0' V_t])$, and its conditional α-quantile is given by: $\beta_0' V_t + H_0^{-1}(\alpha)(1+|\gamma_0' R_t|)$. Here, the parameter of interest is $\theta \equiv (\beta', \gamma')' \in \Theta \equiv B \times \Gamma$, $\Theta \subseteq \mathbb{R}^k$ with $k \equiv b+c$. Note that θ is the only unknown parameter of the conditional density $f_t^0(\cdot)$. Hence, we are in the case where the true conditional density is known up to a finite dimensional parameter. The true value $\theta_0 \equiv (\beta_0', \gamma_0')'$ of θ can be estimated by using a maximum likelihood approach. Under standard regularity conditions (Bickel, 1982; Newey, 2004), the maximum-likelihood estimator (MLE) $\tilde{\theta}_T$ of θ_0 is known to be efficient: $(I_T^0)^{1/2}\sqrt{T}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$, where I_T^0 is the Fisher information matrix, $I_T^0 \equiv T^{-1} \sum_{t=1}^T E[(\nabla_\theta \ln f_t^0(Y_t))(\nabla_\theta \ln f_t^0(Y_t))']$, in which the gradient is evaluated at θ_0.[13]

4.1.2. *Case 2: finite dimensional nuisance parameter.* In many interesting situations, the true density $h_0(\cdot)$ of U_t in the CH model (5) is not entirely known and this uncertainty adversely affects the precision of the M–estimates of θ_0. A familiar case is the one where the error U_t belongs to some parametric family of distributions, indexed by a finite dimensional parameter τ. For example, instead of being standardized Gaussian we can assume $H_0(\cdot)$ to be a standardized Asymmetric Power Distribution (APD), with unknown exponent and asymmetry parameters (Komunjer, 2005a). In other words, the true distribution function of U_t is of the form $H_0(\cdot, \tau_0)$ where $\tau_0 \in \Upsilon \subset \mathbb{R}_*^+ \times (0,1)$ is the unknown parameter of the APD family. 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 $\eta \in \Pi$, $\Pi \subseteq \mathbb{R}^p$: $\eta \equiv (\beta', \gamma', \tau')' \in \Pi \equiv B \times \Gamma \times \Upsilon$ and $p \equiv b + c + 2$. The members $f(\eta)$ of $\mathcal{P}$ are such that $f_t(y, \eta) = (1 + |\gamma' R_t|)^{-1} h_0([1 + |\gamma' R_t|]^{-1}[y - \beta' V_t], \tau)$, for all t, $1 \leqslant t \leqslant T, T \geqslant 1$, and the conditional quantile parameter θ is now given by

[12] In that case we moreover assume that the parameter spaces B and Γ are such that the standard stationarity and invertibility conditions hold.

[13] Following Bickel (1982) and Newey (2004), the regularity conditions imposed are: $[f_t^0(\cdot)]^{1/2}$ is mean-square differentiable with respect to θ_0, the Fisher information matrix I_T^0 is nonsingular and continuous in θ_0 on Θ.

$\theta \equiv (\beta', \gamma', q)' \in \Theta \equiv B \times \Gamma \times Q$, $\Theta \subseteq \mathbb{R}^k$ with $k \equiv b+c+1$.[14] In this interesting situation, the parameter of interest θ has a lower dimensionality than η: $\dim \theta = k$ and $\dim \eta = p = k+1$. We write $\theta = \theta(\eta)$, with $\theta : \Pi \to \Theta$ being some continuously differentiable function, and interpret the rest of η as a nuisance parameter (Stein, 1956; Bickel, 1982).

Similar to the previous case, we assume that the above parametric model $f(\eta)$ is regular (Bickel, 1982; Newey, 2004), that all the conditional densities $f_t(\cdot, \eta)$ satisfy the conditional quantile restriction (A1) and are continuously differentiable on $\mathbb{R}$ for each $\eta \in \Pi$, and that $f_t(Y_t, \cdot)$ is continuously differentiable on Π $a.s.-P$. Let η_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 θ_0 is now written as $\theta_0 = \theta(\eta_0)$ where $\eta_0 \equiv (\beta_0', \gamma_0', \tau_0')'$. Also, let $I_T(\eta)$ denote the Fisher information matrix of the parametric model $\mathcal{P}$, $I_T(\eta) \equiv T^{-1} \sum_{t=1}^{T} E[(\nabla_\eta \ln f_t(Y_t, \eta))(\nabla_\eta \ln f_t(Y_t, \eta))']$. Then, an estimator $\tilde{\theta}_T$ of θ_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 $C_T^0 \equiv \nabla_\eta \theta(\eta_0)(I_T(\eta_0))^+ \nabla_\eta \theta(\eta_0)'$. 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 (see, e.g., Conditions S and S^* in Stein, 1956; Bickel, 1982; Manski, 1984); those are typically expressed as orthogonality conditions on the gradient of the log-likelihood $\nabla_\eta \ln f_t(Y_t, \eta_0)$.

4.1.3. *Case 3: infinite dimensional nuisance parameter.* Now consider the more realistic situation in which the true density of U_t in Equation (5) 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 CH model (5) 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 for example Gaussian, Student-t, Gamma or any other probability density in a set $\mathcal{H}$—set of all families h of probability densities, which are parametrized by τ and satisfy some appropriate conditions, such as being standardized.

The set $\mathcal{S}$ is the union of all parametric sub-families $\mathcal{P}_h \equiv \{f_h(\eta), \eta \in \Pi\}$ obtained when h varies across $\mathcal{H}$. For any given $h \in \mathcal{H}$, the parametric submodel $f_h(\eta)$ is defined as $f_h(\eta) \equiv \{f_{ht}(\cdot, \eta) : \mathbb{R} \to \mathbb{R}_*^+, 1 \leqslant t \leqslant T, T \geqslant 1\}$ and is assumed to satisfy standard regularity conditions (Bickel, 1982; Newey, 2004). We let $I_{hT}(\eta) \equiv T^{-1} \sum_{t=1}^{T} E[(\nabla_\eta \ln f_{ht}(Y_t, \eta))(\nabla_\eta \ln f_{ht}(Y_t, \eta))']$ be the Fisher information matrix of the parametric submodel $\mathcal{P}_h$. In particular, the

[14] The set Q corresponds to the range of α-quantiles of U_t when the parameter τ of its distribution function $H_0(\cdot, \tau)$ varies in Υ.

matrix $I_{hT}(\eta_0)$, in which $f_h(\eta_0) = f^0$, is such that $C^0_{hT} \equiv \nabla_\eta \theta(\eta_0)'(I_{hT}(\eta_0))^+ \nabla_\eta \theta(\eta_0)$ is nonsingular.

In addition, we assume that for any $\eta \in \Pi$ and $h \in \mathcal{H}$, the conditional densities $f_{ht}(\cdot, \eta)$ satisfy the conditional quantile restriction (A1) and are continuously differentiable on $\mathbb{R}$, and that for any $h \in \mathcal{H}$, $f_{ht}(Y_t, \cdot)$ are continuously differentiable $a.s. - P$ on Π. Then, the semiparametric efficiency bound for the conditional quantile parameter θ_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.

4.2. **Least favorable parametric submodel.** Following Stein's (1956) ingenious definition, V^0_T in Theorem 3 is the semiparametric efficiency bound, if and only if, there exists a parametric submodel $\mathcal{P}_{h^*}$ in which the MLE $\tilde{\theta}^*_T$ of the true parameter θ_0 has the same asymptotic covariance matrix V^0_T. The following theorem exhibits the least favorable parametric submodel which satisfies the conditional quantile restriction (A1).

Theorem 4 (Least Favorable Parametric Submodel). *Given $\mathcal{M}$ and the set of true conditional densities $f^0 \equiv \{f^0_t, 1 \leqslant t \leqslant T, T \geqslant 1\}$, consider the parametric submodel $\mathcal{P}^* \equiv \{f^*(\theta), \theta \in \Theta\}$ parametrized by the conditional quantile parameter θ in which $f^*(\theta) \equiv \{f^*_t(\cdot, \theta) : \mathbb{R} \to \mathbb{R}^+_*, 1 \leqslant t \leqslant T, T \geqslant 1\}$ with*

$$f^*_t(y, \theta) \equiv$$

$$f^0_t(y) \frac{\alpha(1-\alpha)\lambda(\theta)\exp\{\lambda(\theta)[F^0_t(y) - F^0_t(q_\alpha(W_t, \theta))][\mathbb{1}(q_\alpha(W_t, \theta) - y) - \alpha]\}}{1 - \exp\{\lambda(\theta)[1 - F^0_t(q_\alpha(W_t, \theta)) - \mathbb{1}(q_\alpha(W_t, \theta) - y)][\mathbb{1}(q_\alpha(W_t, \theta) - y) - \alpha]\}}, \tag{6}$$

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.[15] *Then, under (A0)(ii) and (A1), $\mathcal{P}^*$ is a parametric submodel in $\mathcal{S}$, i.e.:*

(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 $\mathcal{W}_t$;

(iii) $f^0 \in \mathcal{P}^*$.

Moreover, under (A0)-(A1) and (A5)-(A7)(i), $\mathcal{P}^*$ is the least favorable submodel in $\mathcal{S}$, i.e.

[15] A simple function $\Lambda(\cdot)$ in Equation (6) which satisfies the conditions of Theorem 4 is $\Lambda(x) = x'x$.

the asymptotic distribution of the MLE $\tilde{\theta}_T^*$ *associated with* $\mathcal{P}^*$ *is* $(V_T^0)^{-1/2}\sqrt{T}(\tilde{\theta}_T^* - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$ *where* V_T^0 *is as defined in Theorem 3.*

Because $\mathcal{P}^*$ is a parametric submodel of the set $\mathcal{S}$ of all densities satisfying the conditional quantile restriction in (A1), the semiparametric efficiency bound for θ_0 is by Stein's (1956) definition at least as large as the asymptotic variance of the above MLE $\tilde{\theta}_T^*$; Theorem 4 shows that the latter equals V_T^0. On the other hand, in Theorem 3 we have shown that V_T^0 is also the minimum of the asymptotic variances of the consistent and asymptotically normal M–estimators of θ_0. It follows, first, that the semiparametric efficiency bound is V_T^0, and, second, that the parametric model $\mathcal{P}^*$ is the least favorable parametric submodel in $\mathcal{S}$.

The first result—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 (A1), then we cannot estimate the true conditional quantile parameter θ_0 with precision higher than that given by V_T^0. Note that our result uses the moment restriction (A1) only; we do not make any additional assumptions regarding the properties of the "error" term $Y_t - q_\alpha(W_t, \theta)$ (other than those contained in (A1) and (A6)). In particular, we allow for $Y_t - q_\alpha(W_t, \theta)$ to be dependent and nonidentically distributed.

Perhaps the most important aspect of Theorem 4 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 the CH model (5): W_t contains serially dependent exogenous variables or/and lags of Y_t, residuals are uncorrelated and conditionally heteroskedastic.[16] 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 $Y_t - \theta' W_t$. The first part of Theorem 4 generalizes Newey and Powell's (1990) results to the case where the sequence $\{(Y_t, W_t')'\}$ is weakly dependent and heterogeneous, as in (A6). Unsurprisingly, when the data is iid and q_α 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).[17] In the second time series situation of interest, the residuals themselves are correlated in addition to

[16] In the CH model (5) we have: $Y_t - q_\alpha(W_t, \theta_0) = (1 + |\gamma_0' R_t|)[U_t - \mu_0 - \sigma_0 H_0^{-1}(\alpha)]$.

[17] This result is a special case of the result derived by Chamberlain (1987) for models with conditional moment restrictions.

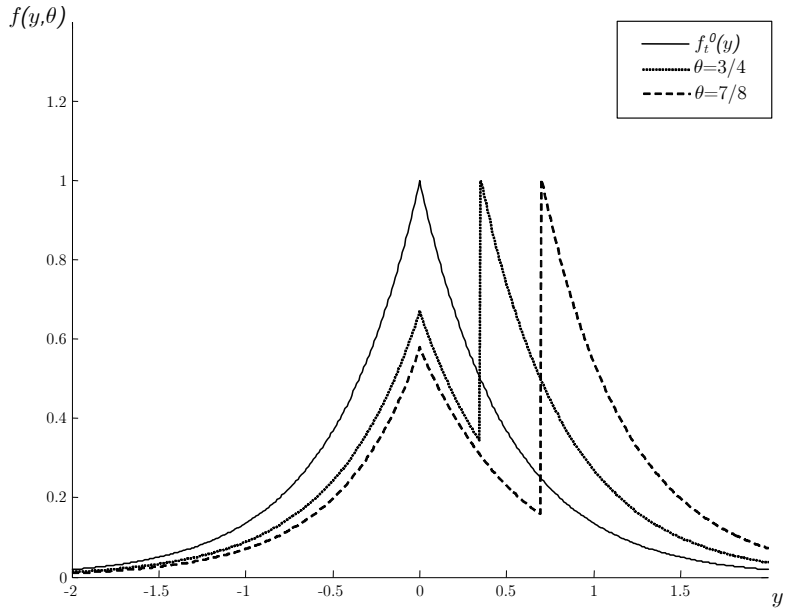

FIGURE 1. Case $\alpha = .5$, $q_\alpha(W_t, \theta) = \theta$ and $f_t^0(y) = \exp(-2|y|)$.

being heteroskedastic. Note that this situation is not covered in the CH model (5); however, our assumption (A1) does not exclude the possiblity that $Y_t - q_\alpha(W_t, \theta)$ be correlated. So far there exist no results on semiparametric efficiency bound which cover this dependent case. To the best of our knowledge, Theorem 4 provides the first result on attainable asymptotic efficiency for nonlinear (and possibly censored) conditional quantile models when the data is dependent.

The second result of Theorem 4—an analytic expression of the least favorable parametric submodel—is entirely new and not yet seen in the literature on efficient estimation under conditional moment restrictions. The density $f_t^*(\cdot, \theta)$ in Equation (6) is not of the 'tick-exponential' form derived by Komunjer (2005b): it depends on the true density $f_t^0(\cdot)$ as well as the true value θ_0 and contains terms such as $\lambda(\theta)$. In the least favorable parametric submodel $\mathcal{P}^*$, θ parametrizes 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). 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 $f_t^0(y)$ which is continuous.

With the semiparametric efficiency bound V_T^0 in hand, we now turn to the problem of constructing a conditional quantile estimator which actually attains the bound.

5. Efficient Conditional Quantile Estimator

As already pointed out in Section 4, the shape ξ_t^* of the optimal objective function $\varphi(\cdot,\cdot,\xi_t^*)$ in Equation (4) is that of the true conditional distribution $F_t^0(\cdot)$, which is unknown. Hence, the M–estimator θ_T^* is in reality infeasible. We construct our (feasible) efficient conditional quantile estimator $\hat{\theta}_T$ by replacing $F_t^0(\cdot)$ in Equation (4) by a nonparametric estimator $\hat{F}_t(\cdot)$. It remains to be shown that the estimator $\hat{\theta}_T$ retains the same asymptotic variance V_T^0. Note that $\hat{\theta}_T$ is constructed without using any knowledge about the true $F_t^0(\cdot)$. It will then follow that the semiparametric efficiency bound V_T^0 can be attained, and that the feasible estimator $\hat{\theta}_T$ is semiparametrically efficient.

We let $g_t^0(\cdot)$ and $\bar{g}_T^0(\cdot)$ be the true density of W_t and the average true density $\bar{g}_T^0(\cdot) \equiv T^{-1}\sum_{t=1}^T g_t^0(\cdot)$ of $\{W_1,\ldots,W_T\}$ respectively, and make the following assumptions:[18]

(A8) *for every* $T \geqslant 1$, $\bar{g}_T^0(\cdot)$ *is continuously differentiable of order* $R \geqslant 1$ *on* $\mathbb{R}^m$ *with* $\sup_{T\geqslant 1}\sup_{w\in\mathbb{R}^m}|D^r\bar{g}_T^0(w)| < \infty$ *for every* $0 \leqslant |r| \leqslant R$.

(A9) *(i) for every* t, $1 \leqslant t \leqslant T, T \geqslant 1$, $F_t^0(\cdot) = F^0(\cdot|W_t)$ *and* $f_t^0(\cdot) = f^0(\cdot|W_t)$*; (ii) the function* $F^0(\cdot|\cdot) : \mathbb{R}^{m+1} \to [0,1]$ *is continuously differentiable of order* $R+2$ *with* $\sup_{(y,w)\in\mathbb{R}^{m+1}}|D^rF^0(y|w)| < \infty$ *for every* $0 \leqslant |r| \leqslant R+2$.

(A10) *for some* $\gamma > 0$ *and any vanishing sequence* $\{c_T\}$*: (i)* $\int_{\{w:\bar{g}_T^0(w)<c_T\}}\bar{g}_T^0(w)dw = o(1)$*, (ii)* $\int_{\{w:\bar{g}_T^0(w)<c_T\}}|\nabla_\theta q_\alpha(w,\theta_0)|\,\bar{g}_T^0(w)dw = O(c_T^\gamma)$*, and (iii)* $\int_{\{w:\bar{g}_T^0(w)<c_T\}} f^0[q_\alpha(w,\theta_0)|w] \times |\nabla_\theta q_\alpha(w,\theta_0)|\bar{g}_T^0(w)dw = O(c_T^{2\gamma})$.

Assumptions (A8) and (A9)(ii) are standard smoothness assumptions on the true densities $g_t^0(\cdot)$ and $f_t^0(\cdot)$; they adapt assumptions NP2 and NP3 used in Andrews (1995) to the case where the regression function is the conditional distribution (and density) of Y_t. On the other hand, assumption (A9)(i) is an additional assumption we need to impose on the true distribution of Y_t conditional upon $\mathcal{W}_t$ in order to construct an estimator that attains the semiparametric efficiency bound. The content of this assumption is twofold. First, it states that no information other than that contained in W_t is useful in constructing the conditional distribution (and density) of Y_t. Note that this is a strengthening of our assumption (A1) which says that W_t contains all the relevant information for the conditional α-quantile of Y_t. Second, assumption (A9)(i) implies that the distribution of Y_t conditional on W_t should be the same as that of Y_s conditional on W_s, for any $s \neq t$.

[18]Recall from Section 2.1 that all the components of W_t are continuous.

Assumption (A10)(i) is weak as it is satisfied if the sequence of probability measures $\{\bar{P}_T^0(\cdot)\}$ associated with the average densities $\{\bar{g}_T^0(\cdot)\}$ is tight, which is itself implied by the tightness of $\{W_t\}$ or equivalently $W_t = O_p(1)$.[19] The latter is obviously satisfied if the W_t's are identically distributed, but it also holds for dependent and heterogenous W_t's if $\{W_t\}$ is uniformly integrable and a fortiori if $\sup_{1\leqslant t\leqslant T, T t\geqslant 1} E[|W_t|^{1+\epsilon}] < \infty$ for some $\epsilon > 0$. Assumptions (A10)(ii) and (A10)(iii) are stronger and used to ensure that the bias of $\hat{\Psi}_T(\theta)$ vanishes at a $\sqrt{T}$-rate. It is similar to conditions that eliminate the asymptotic bias when a stochastic trimming is employed as in Hardle and Stoker (1989) and Lavergne and Vuong (1996). It requires that the tails of $\bar{g}_T^0(\cdot)$ vanish sufficiently fast given the tail behaviors of $|\nabla_\theta q_\alpha(\cdot,\theta_0)|$ and $f^0[q_\alpha(\cdot,\theta_0)|\cdot]$. For instance, if $\sup_{w\in\mathbb{R}^m} |\nabla_\theta q_\alpha(\cdot,\theta_0)| < \infty$ and $\sup_{(y,w)\in\mathbb{R}^{m+1}} f^0(y|w) < \infty$, a sufficient (but not necessary) condition for (A10) is that $\int_{\{w:\bar{g}_T^0(w)<c_T\}} \bar{g}_T^0(w)dw = O(c_T^{2\gamma})$, which is a condition on the vanishing rate of the tails of the average density $\bar{g}_T^0(\cdot)$.

The true conditional distribution $F^0(\cdot|\cdot)$ can be estimated by the kernel estimator $\hat{F}(\cdot|\cdot)$ defined as $\hat{F}(\cdot|w) = 0$ if $\hat{g}(w) = 0$, and $\hat{F}(y|w) \equiv \hat{G}(y,w)/\hat{g}(w)$ if $\hat{g}(w) \neq 0$ with

$$\hat{G}(y,w) \equiv \frac{1}{T h_{wT}^m} \sum_{s=1}^{T} L(\frac{y - Y_s}{h_{yT}}) K(\frac{w - W_s}{h_{wT}}), \tag{7}$$

$$\hat{g}(w) \equiv \frac{1}{T h_{wT}^m} \sum_{s=1}^{T} K(\frac{w - W_s}{h_{wT}}), \tag{8}$$

where $L(y) \equiv \int \mathbb{1}(y-u)K_0(u)du$, $K(\cdot)$ is a multivariate kernel, $K_0(\cdot)$ is a univariate kernel and h_{wT} and h_{yT} are two nonstochastic positive bandwidths. The corresponding kernel estimator of the true conditional density $f^0(\cdot|\cdot)$ is given by $\partial\hat{F}(\cdot|\cdot)/\partial y$, while $\hat{g}(\cdot)$ can be viewed as a kernel estimator of the average true density $\bar{g}_T^0(\cdot)$.

In order to eliminate aberrant behavior of kernel estimators for the conditional distribution (density) of Y_t in regions where the densities of $\{W_t\}$ are small, we define $\hat{F}_t(\cdot) \equiv d_t \hat{F}(\cdot|W_t)$, where $d_t \equiv \mathbb{1}(\hat{g}(W_t) - b_T)$ effectively deletes (trims out) observations for which $\hat{g}(W_t) < b_T$ with $\{b_T\}$ a sequence of positive constants. That is, $\hat{F}_t(\cdot)$ is a trimmed nonparametric estimator of the true conditional distribution $F_t^0(\cdot)$ which we now use to construct our (feasible) estimator $\hat{\theta}_T$. Namely, $\hat{\theta}_T$ is obtained by minimizing the objective function $\hat{\Psi}_T(\theta) \equiv$

[19] By definition (Bilingsley, 1995) the tightness of $\{\bar{P}_T^0(\cdot)\}$ means that for every $\epsilon \in (0,1)$ there exists $M_\epsilon < \infty$ such that $\inf_{1\leqslant t\leqslant T, T\geqslant 1} \bar{P}_T^0([-M_\epsilon, M_\epsilon]) > 1 - \epsilon$. Now, $\int_{\{w:\bar{g}_T^0(w)<c_T\}}^0 \bar{g}_T^0(w)dw = \int_{\{w\in[-M_\epsilon,M_\epsilon]^m:\bar{g}_T^0(w)<c_T\}} \bar{g}_T^0(w)dw + \int_{\{w\notin[-M_\epsilon,M_\epsilon]^m:\bar{g}_T^0(w)<c_T\}} \bar{g}_T^0(w)dw \leqslant c_T(2M_\epsilon)^m + \bar{P}_T^0(\mathbb{R}^m \backslash [-M_\epsilon, M_\epsilon]^m)$ $< c_T(2M_\epsilon)^m + \epsilon$ showing that (A10)(i) holds as $c_T = o(1)$ and ϵ is arbitrary.

$T^{-1}\sum_{t=1}^{T}\varphi(Y_t, q_\alpha(W_t,\theta),\hat{\xi}_t)$, in which

$$\varphi(Y_t, q_\alpha(W_t,\theta),\hat{\xi}_t) \equiv [\alpha - \mathbb{1}(q_\alpha(W_t,\theta) - Y_t)][\hat{F}_t(Y_t) - \hat{F}_t(q_\alpha(W_t,\theta))], \tag{9}$$

for every t, $1 \leqslant t \leqslant T, T \geqslant 1$. In other words, our (feasible) estimator $\hat{\theta}_T$ minimizes a modified version $\hat{\Psi}_T(\cdot)$ of the efficient M–objective function $\Psi_T^*(\cdot)$ in which we have replaced the true conditional distribution of Y_t given W_t with a nonparametric estimator. As a consequence, $\hat{\theta}_T$ is a MINPIN-type estimator (Andrews, 1994a).[20] The shape parameter $\hat{\xi}_t$ of the objective function in Equation (9) is now equal to $\hat{F}_t(\cdot)$.

In order to establish the asymptotic properties of our feasible estimator $\hat{\theta}_T$ we impose the following conditions on the kernels:

(A11) *(i) for any $r = (r_1, \ldots, r_m) \in \mathbb{N}^m$, the kernel $K(\cdot)$ satisfies $\sup_{w\in\mathbb{R}^m}|K(w)| < \infty$, $\int K(w)dw = 1$, $\int w^r K(w)dw = 0$ if $1 \leqslant |r| \leqslant R-1$, and $\int w^r K(w)dw < \infty$ if $|r| = R$; (ii) $K(\cdot)$ has a Fourier transform $\phi(\cdot)$ that is absolutely integrable, i.e. $\int|\phi(w)|dw < \infty$; (iii) $\sup_{y\in\mathbb{R}}|K_0(y)| < \infty$, $\int K_0(y)dy = 1$, $\int y^r K_0(y)dy = 0$ if $1 \leqslant r \leqslant R-1$ and $\int y^R K_0(y)dy < \infty$, (iv) the kernel $K_0(\cdot)$ is continuously differentiable on $\mathbb{R}$ with derivative satisfying $\sup_{y\in\mathbb{R}}|K_0'(y)| < \infty$.*

Assumptions (A11)(i)-(iv) are standard and satisfied, for example, by the multivariate normal-based kernels considered by Bierens (1987): $K(x) = (2\pi)^{-m/2}\sum_{j=1}^{J} a_j|b_j|^{-m}\exp[-ww'/(2b_j^2)]$, where $J \geqslant R/2$ is a positive integer and $\{(a_j, b_j) : j \leqslant J\}$ are constants that satisfy $\sum_{j=1}^{J} a_j = 1$ and $\sum_{j=1}^{J} a_j b_j^{2l} = 0$, for $l = 1, ..., J-1$.

We now turn to the asymptotic properties of our feasible estimator $\hat{\theta}_T$. Note that the shape $\hat{\xi}_t$ of the objective function in Equation (9) depends on all the data up to time T, hence is not $\mathcal{W}_t$-measurable as required by assumption (A2)(i). In consequence, the results of Theorems 1 and 2 do not apply to $\hat{\theta}_T$ and its asymptotic properties need to be derived separately. We first establish the consistency of $\hat{\theta}_T$.

Theorem 5 (Consistency of $\hat{\theta}_T$). *Suppose that (A0)-(A1), (A5)-(A7)(i), (A8)-(A10)(i), (A11) hold. If $b_T = o(1)$ with $b_T\sqrt{T}h_{wT}^m \to \infty$, $b_T/h_{wT}^R \to \infty$ and $b_T/h_{yT}^R \to \infty$ as $T \to \infty$, then $\hat{\theta}_T \xrightarrow{p} \theta_0$.*

The assumptions on the trimming parameter b_T and bandwidths h_{yT} and h_{wT} imply that b_T does not vanish too rapidly and that $h_{yT} \to 0$, $h_{wT} \to 0$ and $\sqrt{T}h_{wT}^m \to \infty$ as

[20] Though $\hat{\theta}_T$ is a member of the MINPIN family, our objective function associated with Equation (9) does not satisfy the assumptions used by Andrews (1994a).

T goes to infinity. Though stronger than necessary, the latter condition is typically used when deriving uniform convergence rates using the Fourier transform $\phi(\cdot)$ of $K(\cdot)$ (Bierens, 1983; Andrews, 1995). In particular, when $R \leqslant m/2$, this condition excludes the optimal bandwidth $h_{wT}^{opt} \sim T^{-1/(2R+m)}$ obtained by Stone (1980, 1982) and Truong and Stone (1992).

In order to derive the asymptotic normality of our efficient estimator $\hat{\theta}_T$, we strengthen our dependence assumption (A6):

(A6') *the sequence* $\{(Y_t, W_t')'\}$ *is (i) strictly stationary and (ii)* β*-mixing with* β *of size* $-r/(r-2)$*, with* $2 < r < 3$*;*

The proof of our result uses Lemma 3 in Arcones (1995) which requires strict stationarity and β-mixing with $r > 2$. Note that β-mixing (or absolute regularity) in (A6')(ii) is a condition intermediate between α-mixing (strong mixing)—which is the weakest form of strong mixing—and ϕ-mixing (uniform mixing)—which is the strongest form of mixing (Bradley, 1986). As such, our weak dependence assumption is stronger than that of α-mixing used by Robinson (1983), for example. Assumption (A6')(ii) also requires the size of the β-mixing process to be comprised between $-\infty$ and -3. In other words, we limit the amount of dependence allowed in $\{(Y_t, W_t')'\}$.[21] In particular, Truong and Stone (1992) use the condition $\beta_t = O(\rho^t)$ as $t \to \infty$ for some ρ with $0 < \rho < 1$ in order to estimate the conditional quantile nonparametrically at the optimal rate. Their condition implies β-mixing of arbitrary size and hence of size $-r/(r-2)$, with some r, $2 < r < 3$.

We can now establish the efficiency of $\hat{\theta}_T$.

Theorem 6 (Efficiency of $\hat{\theta}_T$). *Suppose that Assumptions (A0)-(A1), (A5), (A6'), (A7)(i) and (A8)-(A11) hold. If* $b_T = o(T^{-1/(4\gamma)})$ *with* $b_T T^{1/4} h_{yT} h_{wT}^m \to \infty$*,* $b_T/(T^{1/4} h_{wT}^R) \to \infty$*, and* $b_T/(T^{1/4} h_{yT}^R) \to \infty$*, as* $T \to \infty$*, then* $\hat{\theta}_T$ *is efficient:* $(\bar{V}_T^0)^{-1/2}\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$*, where*

$$\bar{V}_T^0 \equiv \alpha(1-\alpha)\{T^{-1}\sum\nolimits_{t=1}^{T} E[(f^0(q_\alpha(W_t,\theta_0)|W_t))^2 \nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'\}^{-1}$$

is the semiparametric efficiency bound.

[21] Note that the $-\infty$ case, obtained when $r = 2$, corresponds to independence. As the proof of Lemma 10 shows, the assumption (A6')(ii) is stronger than necessary: we can replace it by β-mixing with mixing coefficients β_t that satisfy $\sum_{t=1}^{T-1} t\beta_t^{(r-2)/r} = O(\sqrt{T})$.

The conditions on the trimming parameter and bandwidths are stronger than in Theorem 5. They can be written as:

$$\max\left\{\frac{1}{T^{1/4}h_{yT}h_{wT}^{m}}, T^{1/4}h_{wT}^{R}, T^{1/4}h_{yT}^{R}\right\} \ll b_T \ll \frac{1}{T^{1/(4\gamma)}},$$

where $a_T \ll c_T$ means that $a_T < c_T$ for T sufficiently large. This implies $T^{1/(4\gamma)-1/4} \ll h_{yT}h_{wT}^{m} \ll T^{-[(m+1)/R][1/(4\gamma)+1/4]}$. Hence, necessary conditions are $\gamma > 1$ and $R > (m+1)(\gamma+1)/(\gamma-1)$.[22] For instance, when $m = 1$, $R = 3$ and $\gamma = 6$, a feasible choice is: $h_{yT} \propto T^{-1/10}$, $h_{wT} \propto T^{-1/10}$, and $b_T \propto T^{-1/21}$. Moreover, if $R > (m+1)(3\gamma+1)/[2(\gamma-1)]$, one can choose the L^2-optimal bandwidths $h_{yT}^{*} \propto T^{-2R/[(2R+m)(2R+m+1)]}$ and $h_{wT}^{*} \propto T^{-1/(2R+m)}$ for estimating $f^0(y|w)\bar{g}_T^0(w)$ and $\bar{g}_T^0(w)$.[23] For instance, when $m = 1$, $R = 4$ and $\gamma = 6$, then the L^2-optimal bandwidths $h_{yT}^{*} \propto T^{-8/90}$, $h_{wT}^{*} \propto T^{-1/9}$ with trimming parameter $b_T \propto T^{-1/21}$ can be chosen. In particular, our estimator $\hat{\theta}_T$ differs from many semiparametric ones that are $\sqrt{T}$-asymptotically normal under assumptions that imply undersmoothing and thus exclude the L^2-optimal bandwidth.

Without assumption (A9)(i) we would not be able to construct a conditional quantile estimator which attains $\bar{V}_T^0$. Note however that our general expression for V_T^0 derived in Theorem 3 remains valid whether or not we are able to construct an efficient estimator—this is one of the advantages of using the "supremum" characterization of the semiparametric efficiency bound.

Our efficient M–estimator $\hat{\theta}_T$ is asymptotically equivalent to: the 'one-step' estimator proposed by Newey and Powell (1990), the weighted quantile regression estimator by Zhao (2001), and the CEL estimator by Otsu (2003). Two important features distinguish our efficient estimator from the previous ones. First, similar to Otsu's (2003) CEL estimator,

[22] As indicated in Lavergne and Vuong (1996, p.209), we have $c_T^2 = o\left(\int_{\{w:\bar{g}_T^0(w)<c_T\}} \bar{g}_T^0(w)dw\right)$ when $\bar{g}_T^0(\cdot)$ is continuously differentiable on $\mathbb{R}^m$ and monotonically decreasing in the tails, whether or not the support of $\bar{g}_T^0(\cdot)$ is bounded. Under the same conditions, it can be shown that $c_T = o\left(\int_{\{w:\bar{g}_T^0(w)<c_T\}} |w|\bar{g}_T^0(w)dw\right)$ when the support of $\bar{g}_T^0(\cdot)$ is $\mathbb{R}^m$. Hence, (A10)(ii) implies $\gamma < 1$ when $q_\alpha(w,\theta_0) = w'\theta_0$, which contradicts $\gamma > 1$. On the other hand, when the support of $\bar{g}_T^0(\cdot)$ is bounded (uniformly in T), it can be shown that $\int_{\{w:\bar{g}_T^0(w)<c_T\}} |w|\bar{g}_T^0(w)dw = O(c_T^{2-\delta})$, where $\delta > 0$ can be arbitrarily close to zero depending on $\bar{g}_T^0(\cdot)$. Hence, our assumptions allow the linear quantile specification $q_\alpha(w,\theta_0) = w'\theta_0$, provided the support of W_t is bounded (uniformly in T). Bounded supports, however, are not required as our assumptions allow for unbounded ones. In this case $f^0[q_\alpha(w,\theta^0)|w]$ and $|\nabla_\theta q_\alpha(w,\theta^0)|$ should vanish in the tails of W_t at appropriate rates for (A10) and the trimming/bandwidth conditions to be compatible.

[23] See Stone (1980, 1982) where h_{yT}^{*} solves $(h_{yT}^{*}h_{wT}^{*m})^{1/(m+1)} \propto T^{-1/(2R+m+1)}$.

our M–estimator $\hat{\theta}_T$ does not require a preliminary consistent estimate of θ_0. It is well established that such a preliminary step causes poor small sample performance in GMM estimation (Altonji and Segal, 1996).[24] Second, the objective functions $\varphi(\cdot, \cdot, \hat{\xi}_t)$ used in the construction of $\hat{\theta}_T$ depend on a nonparametric estimator of the distribution function $F^0(\cdot|\cdot)$. Newey and Powell's (1990) and Zhao's (2001) efficient estimators on the other hand depend on nonparametric estimators of the density $f^0(\cdot|\cdot)$.[25] Both features can potentially affect the small sample properties of these efficient estimators.

6. Conclusion

The contributions of this paper are twofold: first, it derives the semiparametric efficiency bound V_T^0 for parameters of conditional quantiles in time series models with weakly dependent and/or heterogeneous data. Our bound V_T^0 generalizes expressions previously derived by the literature on efficient conditional quantile estimation. In particular we allow the data to exhibit dependence and/or conditional heteroskedasticity. The second result of the paper is to show that efficient estimation is possible in models for conditional quantiles in which the true conditional distribution does not depend on any other variables than those entering the quantile. In such models, the semiparametric efficiency bound equals $\bar{V}_T^0$ and we are able to construct an M–estimator $\hat{\theta}_T$ which actually attains the bound. Our efficient estimator is different from previous ones and is of the MINPIN-type as the efficient M–objective function that it minimizes depends on a nonparametric estimator of the conditional distribution.

An interesting by-product of the paper is to show that the class of M–estimators is rich enough to contain estimators that are efficient, at least in models for conditional quantiles. 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 do not belong to the class of QMLEs,

[24] In models with unconditional moment restrictions, Newey and Smith (2004) show how empirical likelihood based methods improve the finite sample properties of GMM.

[25] In particular, when estimating $F^0(\cdot|\cdot)$ and $f^0(\cdot|\cdot)$ by kernel estimators, there is always one smoothing parameter less to choose for conditional distributions (Hansen, 2004a,b). For example, in the iid case, our efficient estimator $\hat{\theta}_T$ can be constructed by using the empirical distribution function.

but 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. However, important efficiency improvements are made by going from QMLEs to M–estimators.

Finally, the "supremum" approach we use to derive the semiparametric efficiency bound V_T^0 does not seem to suffer from strong independence assumptions traditionally imposed by 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. We conjecture that it can thus be generalized fairly easily to accommodate for general moment restrictions. The steps to follow in the construction of semiparametric efficiency bounds in models with time series data seem to be: (1) construct the largest class of M–estimators which are consistent for the true parameter θ_0 of the conditional moment restriction in hand; (2) within this class, find the minimum asymptotic covariance matrix—this is a candidate matrix V for the bound—and the M–estimator which attains this minimum; (3) use its expression to derive the least favorable parametric submodel of the initial semiparametric model; (4) show that the inverse of the Fisher information matrix in this submodel equals V. It then follows that V is the semiparametric efficiency bound. While step (3) is perhaps the crucial one, we have little guidance on how exactly to construct the least favorable parametric submodel under general moment restrictions. This seems to be an important topic which we leave for future research.

7. Proofs

Proof of Theorem 1. First, note that (A2)-(A3) together with the compactness of the parameter space Θ, are sufficient conditions for θ_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 under correct conditional quantile model specification assumption (A1), we have: $\theta_\infty^0 = \theta_0$ for any $T \geqslant 1$ if and only if there exist a real function $A(\cdot, \xi_t) : \mathbb{R} \to \mathbb{R}$, twice continuously differentiable and strictly increasing $a.s. - P$ on $\mathcal{Q}$ with derivative $a(y, \xi_t) \equiv \partial A(y, \xi_t)/\partial y$, and a real function $B(\cdot, \xi_t) : \mathbb{R} \to \mathbb{R}$, such that, for any $T \geqslant 1$ and every t, $1 \leqslant t \leqslant T$,

$$\varphi(Y_t, q_t, \xi_t) = [\alpha - \mathbb{1}(q_t - Y_t)][A(Y_t, \xi_t) - A(q_t, \xi_t)] + B(Y_t, \xi_t), a.s. - P, \tag{10}$$

on $\mathbb{R} \times \mathcal{Q} \times E_t$.

We treat separately the two implications contained in the above equivalence. We start with the sufficiency part of the proof and show that if, for any $T \geqslant 1$ and every t, $1 \leqslant t \leqslant T$,

$\varphi(\cdot,\cdot,\xi_t)$ is as in equation (10) above, then $\theta_\infty^0 = \theta_0$ for any $T \geqslant 1$, i.e. θ_0 is also a minimizer of $E[\Psi_T(\theta)]$ on $\mathring{\Theta}$. Given (A3)(i) we know that $\nabla_\theta E[\Psi_T(\theta)] = T^{-1}\sum_{t=1}^T E[\nabla_\theta \varphi(Y_t, q_\alpha(W_t,\theta),\xi_t)]$. From (10) and the $a.s.-P$ twice continuous differentiability of $A(\cdot,\xi_t)$ on $\mathcal{Q}$, for any t, $1 \leqslant t \leqslant T, T \geqslant 1$, we have:

$$\begin{aligned}
&E[\nabla_\theta \varphi(Y_t, q_\alpha(W_t,\theta),\xi_t)] \\
&= E\{\nabla_\theta q_\alpha(W_t,\theta) a(q_\alpha(W_t,\theta),\xi_t)[\mathbb{1}(q_\alpha(W_t,\theta) - Y_t) - \alpha]\} \\
&= E\{\nabla_\theta q_\alpha(W_t,\theta) a(q_\alpha(W_t,\theta),\xi_t) E[\mathbb{1}(q_\alpha(W_t,\theta) - Y_t) - \alpha|\mathcal{W}_t]\},
\end{aligned}$$

so that by using the correct model specification assumption (A1) we get $E[\mathbb{1}(q_\alpha(W_t,\theta_0) - Y_t) - \alpha|\mathcal{W}_t] = 0, a.s.-P$, for every t, $1 \leqslant t \leqslant T, T \geqslant 1$, and hence $\nabla_\theta E[\Psi_T(\theta_0)] = 0$. Similarly, $\Delta_{\theta\theta} E[\Psi_T(\theta)] = T^{-1}\sum_{t=1}^T E[\Delta_{\theta\theta}\varphi(Y_t, q_\alpha(W_t,\theta),\xi_t)]$ and

$$\begin{aligned}
&E[\Delta_{\theta\theta}\varphi(Y_t, q_\alpha(W_t,\theta),\xi_t)] \\
&= E\left\{[\frac{\partial a(q_\alpha(W_t,\theta),\xi_t)}{\partial y}\nabla_\theta q_\alpha(W_t,\theta)\nabla_\theta q_\alpha(W_t,\theta)'\right. \\
&\left. + a(q_\alpha(W_t,\theta),\xi_t)\Delta_{\theta\theta} q_\alpha(W_t,\theta)][\mathbb{1}(q_\alpha(W_t,\theta) - Y_t) - \alpha]\right\} \\
&+ E\left[\nabla_\theta q_\alpha(W_t,\theta)\nabla_\theta q_\alpha(W_t,\theta)' a(q_\alpha(W_t,\theta),\xi_t)\delta(q_\alpha(W_t,\theta) - Y_t)\right] \\
&= E\left\{[\frac{\partial a(q_\alpha(W_t,\theta),\xi_t)}{\partial y}\nabla_\theta q_\alpha(W_t,\theta)\nabla_\theta q_\alpha(W_t,\theta)'\right. \\
&\left. + a(q_\alpha(W_t,\theta),\xi_t)\Delta_{\theta\theta} q_\alpha(W_t,\theta)] E[\mathbb{1}(q_\alpha(W_t,\theta) - Y_t) - \alpha|\mathcal{W}_t]\right\} \\
&+ E\left\{\nabla_\theta q_\alpha(W_t,\theta)\nabla_\theta q_\alpha(W_t,\theta)' a(q_\alpha(W_t,\theta),\xi_t) E[\delta(q_\alpha(W_t,\theta) - Y_t)|\mathcal{W}_t]\right\}
\end{aligned}$$

so that by using (A1)

$$\begin{aligned}
&\Delta_{\theta\theta} E[\Psi_T(\theta_0)] \\
&= T^{-1}\sum_{t=1}^T E\{\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)' a(q_\alpha(W_t,\theta_0),\xi_t) E[\delta(q_\alpha(W_t,\theta_0) - Y_t)|\mathcal{W}_t]\} \\
&= T^{-1}\sum_{t=1}^T E[\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)' a(q_\alpha(W_t,\theta_0),\xi_t) f_t^0(q_\alpha(W_t,\theta_0))], \qquad (11)
\end{aligned}$$

where for every t, $1 \leqslant t \leqslant T$, $f_t^0(\cdot)$ 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 (11), we know that for any $\chi \in \mathbb{R}^k$, $\chi'\Delta_{\theta\theta} E[\Psi_T(\theta_0)]\chi = 0$ only if $T^{-1}\sum_{t=1}^T E[\chi'\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'\chi a(q_\alpha(W_t,\theta_0),\xi_t)\times$

$f_t^0(q_\alpha(W_t,\theta_0))] = 0$. Now, note that for any t, $1 \leqslant t \leqslant T$ and $T \geqslant 1$,

$$\begin{aligned} & E[\chi'\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'\chi a(q_\alpha(W_t,\theta_0),\xi_t)f_t^0(q_\alpha(W_t,\theta_0))] \\ & = E[(\chi'\nabla_\theta q_\alpha(W_t,\theta_0))^2 a(q_\alpha(W_t,\theta_0),\xi_t)f_t^0(q_\alpha(W_t,\theta_0))] \geqslant 0, \end{aligned} \tag{12}$$

for any $\chi \in \mathbb{R}^k$, since we know that $a(q_\alpha(W_t,\theta_0),\xi_t) > 0, a.s. - P$ and $f_t^0(q_\alpha(W_t,\theta_0)) > 0, a.s. - P$. Taking into account the inequality in (12) we have that for any $\chi \in \mathbb{R}^k$, $\chi'\Delta_{\theta\theta}E[\Psi_T(\theta_0)]\chi = 0$ 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, together with (A0)(iii), implies that $\chi = 0$. From there we conclude that $\Delta_{\theta\theta}E[\Psi_T(\theta_0)] \gg 0$ and therefore θ_0 is a minimizer $E[\Psi_T(\theta)]$ on $\mathring{\Theta}$. Since by (A3)(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.

We now show that the functional form of $\varphi(\cdot,\cdot,\xi_t)$ in (10) is necessary for $\theta_\infty^0 = \theta_0$ to hold for any $T \geqslant 1$. Given the differentiability of $E[\Psi_T(\theta)]$ on Θ by (A3)(i), a necessary requirement for $\theta_\infty^0 = \theta_0$ is that the first order condition $\nabla_\theta E[\Psi_T(\theta_0)] = 0$ be satisfied, which is equivalent to

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

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

$$\begin{aligned} E[\mathbb{1}(q_\alpha(W_t,\theta_0) - Y_t) - \alpha|\mathcal{W}_t] &= 0, a.s. - P \\ &\Rightarrow E[\frac{\partial\varphi}{\partial q_t}(Y_t,q_\alpha(W_t,\theta_0),\xi_t)|\mathcal{W}_t] = 0, a.s. - P, \end{aligned} \tag{13}$$

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

$$\frac{\partial\varphi}{\partial q_t}(Y_t,q_\alpha(W_t,\theta_0),\xi_t) = a(q_\alpha(W_t,\theta_0),\xi_t)[\mathbb{1}(q_\alpha(W_t,\theta_0) - Y_t) - \alpha], a.s. - P, \tag{14}$$

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 (13). Using a generalized Farkas lemma (Lemma 8.1, p 240, vol 1) in Gourieroux and Monfort (1995), (13) implies there exists a $\mathcal{W}_t$-measurable

random variable a_t such that

$$\frac{\partial \varphi}{\partial q_t}(Y_t, q_\alpha(W_t,\theta_0), \xi_t) = a_t[\mathbb{1}(q_\alpha(W_t,\theta_0) - Y_t) - \alpha], a.s. - P.$$

Since the left-hand side only depends on Y_t, $q_\alpha(W_t,\theta_0)$ and ξ_t, the same must hold for the right-hand side. Hence, a_t can only depend on $q_\alpha(W_t,\theta_0)$ and ξ_t and we can write $a_t = a(q_\alpha(W_t,\theta_0),\xi_t)$; so the equality in (14) 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 (A3)(i)).[26] Taking into account (14) and our previous computations leading to (11), we have

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

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 θ_0 implies that $a(q_t,\xi_t) > 0, a.s. - P$ for any $q_t \in \mathcal{Q}$ and for all t, $1 \leqslant t \leqslant T, T \geqslant 1$.

The remainder of the proof is straightforward: we need to integrate the necessary condition (14) with respect to q_t. Note that (14) can be written

$$\frac{\partial \varphi}{\partial q_t}(Y_t, q_\alpha(W_t,\theta_0),\xi_t) = \begin{cases} (1-\alpha)a(q_\alpha(W_t,\theta_0),\xi_t), \text{ if } Y_t \leqslant q_\alpha(W_t,\theta_0), \\ -\alpha a(q_\alpha(W_t,\theta_0),\xi_t), \text{ if } Y_t > q_\alpha(W_t,\theta_0), \end{cases}, a.s. - P,$$

for any $\theta_0 \in \mathring{\Theta}$ and for any t, $1 \leqslant t \leqslant T, T \geqslant 1$. Together with the continuity of $\varphi(Y_t,\cdot,\xi_t)$ $a.s. - P$ on $\mathcal{Q}$ in (A2)(ii), the above integrates into

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

$a.s. - P$, where for every t, $1 \leqslant t \leqslant T, T \geqslant 1$, $A(\cdot,\xi_t)$ is an indefinite integral of $a(\cdot,\xi_t)$, $A(q_t,\xi_t) \equiv \int_a^{q_t} a(r,\xi_t)dr$, $a \in \mathbb{R}$, and $B(\cdot,\xi_t) : \mathbb{R} \to \mathbb{R}$ is a real function. Note that the above

[26] Note that this requirement is weaker than the positive definiteness of $\Delta_{\theta\theta} E[\Psi_T(\theta_0)]$, $\Delta_{\theta\theta} E[\Psi_T(\theta_0)] \gg 0$, which is a sufficient condition for θ_0 to be a minimum.

equality has to hold for any $\theta_0 \in \mathring{\Theta}$ so that

(15)

$$\varphi(Y_t, q_\alpha(W_t,\theta),\xi_t) = B(Y_t,\xi_t) + [\alpha - \mathbb{1}(q_\alpha(W_t,\theta) - Y_t)][A(Y_t,\xi_t) - A(q_\alpha(W_t,\theta),\xi_t)], a.s. - P,$$

for every t, $1 \leqslant t \leqslant T, T \geqslant 1$, and for all $\theta \in \Theta$; this is a necessary condition for the M–estimator θ_T to be consistent for θ_0. Equality (15) implies that for any t, $1 \leqslant t \leqslant T, T \geqslant 1$, $\varphi(Y_t, q_t, \xi_t) = B(Y_t,\xi_t) + [\alpha - \mathbb{1}(q_t - Y_t)][A(Y_t,\xi_t) - A(q_t,\xi_t)]$, $a.s. - P$ on $\mathbb{R} \times \mathcal{Q} \times E_t$. □

Proof of Theorem 2. To show that Theorem 2 holds, we first show that under primitive conditions given in (A0)-(A2) and (A4)-(A7), θ_T is consistent for θ_0, i.e. $\theta_T - \theta_0 \xrightarrow{p} 0$. We proceed by checking that all the assumptions for consistency used by Komunjer (2005b) 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 θ_T defined in (A2). Assumptions A2 and A3 in Komunjer (2005b) are satisfied by imposing our (A5) and (A4), respectively. The α-mixing condition A4 in Komunjer (2005b) and the assumption that W_t is a function of some finite number of lags of X_t stated in A0.iv in Komunjer (2005b) are used to ensure that $\{(Y_t, W_t')'\}$ is α-mixing of 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 (A6), which is sufficient for the proof of Theorem 3 in Komunjer (2005b) to go through. Finally, the moment conditions A5 in Komunjer (2005b) directly follow from our (A7) 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 (2005b)—corresponding to the case where the conditional quantile model is correctly specified (A1)—which proves the consistency of θ_T. Similarly, we derive asymptotic normality by using the results of Corollary 5 in Komunjer (2005b). The boundedness of the second derivative of $A(\cdot,\xi_t)$ contained in assumption A3' in Komunjer (2005b) is directly implied by (A4). The moment condition in assumption A5' in Komunjer (2005b) follows from our (A7). 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 (2005b). Hence, from Corollary 5 in Komunjer (2005b) 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

$$\Delta_T^0 = T^{-1}\textstyle\sum_{t=1}^T E[a(q_\alpha(W_t,\theta_0),\xi_t) f_t^0(q_\alpha(W_t,\theta_0))\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'], \tag{16}$$

and

$$\Sigma_T^0 = T^{-1}\sum_{t=1}^T \alpha(1-\alpha)E[(a(q_\alpha(W_t,\theta_0),\xi_t))^2\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)']. \tag{17}$$

□

Proof of Theorem 3. The proof of this theorem 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:

$$\begin{aligned}
&(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} - V_T^0 \\
&= V_T^0(V_T^0)^{-1}V_T^0 - V_T^0\Delta_T^0(\Delta_T^0)^{-1} - (\Delta_T^0)^{-1}\Delta_T^0V_T^0 + (\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} \\
&= T^{-1}\sum_{t=1}^T E\{V_T^0[\frac{(f_t^0(q_\alpha(W_t,\theta_0)))^2}{\alpha(1-\alpha)}\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)']V_T^0 \\
&\quad - V_T^0[f_t^0(q_\alpha(W_t,\theta_0))a(q_\alpha(W_t,\theta_0),\xi_t)\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'](\Delta_T^0)^{-1} \\
&\quad - (\Delta_T^0)^{-1}[f_t^0(q_\alpha(W_t,\theta_0))a(q_\alpha(W_t,\theta_0),\xi_t)\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)']V_T^0 \\
&\quad + (\Delta_T^0)^{-1}[\alpha(1-\alpha)(a(q_\alpha(W_t,\theta_0),\xi_t))^2\nabla_\theta q_\alpha(W_t,\theta_0)\nabla_\theta q_\alpha(W_t,\theta_0)'](\Delta_T^0)^{-1}\},
\end{aligned}$$

so that

$$(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} - V_T^0 = \frac{1}{\alpha(1-\alpha)}T^{-1}\sum_{t=1}^T E[\chi_t\chi_t'],$$

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. In other words, the matrix 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.

We now show that this lower bound V_T^0 is attained by an M–estimator θ_T^* if and only if its objective function corresponds to $\Psi_T^*(\theta) \equiv T^{-1}\sum_{t=1}^T \varphi(Y_t, q_\alpha(W_t,\theta),\xi_t^*)$ with

$$\varphi(Y_t,q_t,\xi_t^*) = [\alpha - \mathbb{I}(q_t - Y_t)][F_t^0(Y_t) - F_t^0(q_t)], a.s. - P, \tag{18}$$

on $\mathbb{R}\times\mathcal{Q}\times E_t$, for every $1 \leqslant t \leqslant T, T \geqslant 1$. We first show the necessary part of this equivalence: V_T^0 is attained only if for any $T \geqslant 1$, there exist ξ_t^* and $A(\cdot, \xi_t^*)$, $1 \leqslant t \leqslant T, T \geqslant 1$, such that

$$T^{-1}\sum_{t=1}^{T} E(\chi_t \chi_t') = 0. \tag{19}$$

The above needs to hold for any $T \geqslant 1$, hence (19) implies that $E(\chi_t\chi_t') = 0$ for every t, $1 \leqslant t \leqslant T, T \geqslant 1$. Taking into account the positivity $a.s. - P$ of the quadratic form $\chi_t\chi_t'$, the latter equalities holds only if for every t, $1 \leqslant t \leqslant T, T \geqslant 1$, we have $\chi_t\chi_t' = 0$, $a.s. - P$. Hence, $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} = V_T^0$ for any $T \geqslant 1$, if and only if for every t, $1 \leqslant t \leqslant T, T \geqslant 1$, $\chi_t = 0$, $a.s. - P$, which combined with (A0)(iii) is equivalent to

$$\frac{f_t^0(q_\alpha(W_t, \theta_0))}{\alpha(1-\alpha)a(q_\alpha(W_t, \theta_0), \xi_t^*)} V_T^0\Delta_T^0 = \mathrm{Id}, a.s. - P,$$

for all t, $1 \leqslant t \leqslant T, T \geqslant 1$. This in turn implies that for every t, $1 \leqslant t \leqslant T, T \geqslant 1$ and any $q_t \in \mathcal{Q}$,

$$a(q_t, \xi_t^*) = c\frac{f_t^0(q_t)}{\alpha(1-\alpha)} \text{ and } V_T^0\Delta_T^0 = c \cdot \mathrm{Id},$$

where c is some strictly positive real constant, $c > 0$. Note that the above condition is equivalent to $a(q_t, \xi_t^*) = cf_t^0(q_t)/[\alpha(1-\alpha)]$ alone, which by integration with respect to q_t gives that for every t, $1 \leqslant t \leqslant T, T \geqslant 1$, and any $q_t \in \mathcal{Q}$

$$A(q_t, \xi_t^*) = c\frac{F_t^0(q_t)}{\alpha(1-\alpha)} + d, \tag{20}$$

with $d \in \mathbb{R}$. Condition (20) is both a necessary and a sufficient condition for the equality in (19) to hold for any $T \geqslant 1$. It is important to note that changing the value of $A(\cdot, \xi_t^*)$ outside $\mathcal{Q}$ does not affect the minima of $E[\Psi_T]$ so $A(\cdot, \xi_t^*)$ can take arbitrary values on $\mathbb{R}\backslash\mathcal{Q}$. To keep the notation simple and without altering the general validity of our result, we set $A(y, \xi_t^*) = cF_t^0(y)/[\alpha(1-\alpha)] + d$, for all $y \in \mathbb{R}$. Moreover, changing the constants c and d does not affect the value of $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1}$ so that they can be arbitrarily chosen in $\mathbb{R}_+^* \times \mathbb{R}$ for any $T \geqslant 1$. For example, we can let $c = \alpha(1-\alpha)$ and $d = 0$ in which case

$$A(y, \xi_t^*) = F_t^0(y), \tag{21}$$

for all $y \in \mathbb{R}$; this completes the proof of the necessary part.

Now, we show that under (A0)-(A1), (A5)-(A6) and (A7)(i), the M–estimator θ_T^*–obtained by minimizing $\Psi_T^*(\theta)$ associated with (18)–is such that $\sqrt{T}(V_T^0)^{-1/2}(\theta_T^* - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$. Note that the shape ξ_t^* of $A(\cdot, \xi_t^*)$ corresponds to the true conditional distribution $F_t^0(\cdot)$ which is stochastic and $\mathcal{W}_t$-measurable thereby satisfying (A2)(i). Moreover, $F_t^0(\cdot)$ is twice

continuously differentiable with bounded $f_t^0(y)$ and $|df_t^0(y)/dy|$, which satisfies (A2)(ii) and (A4). Moreover, $F_t^0(\cdot)$ being bounded by 1 the moment conditions in (A7)(ii) automatically hold. Hence, we can apply Theorem 2 to show that, under (A0)-(A1), (A5)-(A6) and (A7)(i), θ_T^* with $A(\cdot, \xi_t^*)$ as in (21), is asymptotically normally distributed $\sqrt{T}(\Sigma_T^0)^{-1/2}\Delta_T^0(\theta_T - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathrm{Id})$ with

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

and $\Sigma_T^0 = \alpha(1-\alpha)\Delta_T^0$, so that $(\Delta_T^0)^{-1}\Sigma_T^0(\Delta_T^0)^{-1} = V_T^0$. □

Proof of Theorem 4. The following lemma shows that $(i) - (iii)$ in Theorem 4 hold:

Lemma 7. *The parametric submodel $\mathcal{P}^*$ defined by* (6) *is a submodel of $\mathcal{S}$.*

In order to show that $\mathcal{P}^*$ is the least favorable model, consider estimating the parameter θ in $\mathcal{P}^*$ by using 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 (A0)(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 Θ 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 (2005b). To simplify the notation let

$$x(\theta) \equiv [\mathbb{1}(q_t(\theta) - Y_t) - \alpha][1 - \mathbb{1}(q_t(\theta) - Y_t) - F_t^0(q_t(\theta))] \text{ and } u(z) \equiv \frac{\exp z}{1 - \exp z}, \tag{22}$$

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

$$\begin{aligned}\nabla_\theta \ln f_t^*(Y_t,\theta) = {} & \frac{\nabla_\theta \lambda(\theta)}{\lambda(\theta)} + \nabla_\theta\lambda(\theta)[F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbb{1}(q_t(\theta) - Y_t) - \alpha] \\ & - \lambda(\theta) f_t^0(q_t(\theta))\nabla_\theta q_t(\theta)[\mathbb{1}(q_t(\theta) - Y_t) - \alpha] \\ & + \lambda(\theta)[F_t^0(Y_t) - F_t^0(q_t(\theta))]\delta(q_t(\theta) - Y_t)\nabla_\theta q_t(\theta) \\ & + u(\lambda(\theta)x(\theta))\nabla_\theta(\lambda(\theta)x(\theta)), \end{aligned} \tag{23}$$

where $\nabla_\theta(\lambda(\theta)x(\theta)) = \nabla_\theta\lambda(\theta)x(\theta) + \lambda(\theta)\nabla_\theta x(\theta)$ and

$$\nabla_\theta x(\theta) = \left\{ f_t^0(q_t(\theta))[\alpha - \mathbb{1}(q_t(\theta) - Y_t)] + \delta(q_t(\theta) - Y_t)[\alpha - F_t^0(q_t(\theta))] \right\} \nabla_\theta q_t(\theta). \tag{24}$$

(The equality in (24) follows from (22) and the fact that $[\mathbb{1}(\cdot)]^2 = \mathbb{1}(\cdot)$.) 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 θ_0 so

$$\begin{aligned} u(\lambda(\theta)x(\theta))\nabla_\theta(\lambda(\theta)x(\theta)) &= -\frac{\nabla_\theta(\lambda(\theta)x(\theta))}{\lambda(\theta)x(\theta)} + o_p(1) \\ &= -\frac{\nabla_\theta\lambda(\theta)}{\lambda(\theta)} - \frac{\nabla_\theta x(\theta)}{x(\theta)} + o_p(1), \end{aligned} \tag{25}$$

in the neighborhood of θ_0. In particular, combining (23) (25), (24) and (22) we get

$$\begin{aligned} & \nabla_\theta \ln f_t^*(Y_t,\theta_0) \\ & = -\nabla_\theta q_t(\theta_0) \left\{ \frac{f_t^0(q_t(\theta_0))[\alpha - \mathbb{1}(q_t(\theta_0) - Y_t)] + \delta(q_t(\theta_0) - Y_t)[\alpha - F_t^0(q_t(\theta_0))]}{[\mathbb{1}(q_t(\theta_0) - Y_t) - \alpha][1 - \mathbb{1}(q_t(\theta_0) - Y_t) - F_t^0(q_t(\theta_0))]} \right\} \\ & = -\frac{1}{\alpha(1-\alpha)} \nabla_\theta q_t(\theta_0) f_t^0(q_t(\theta_0))[\mathbb{1}(q_t(\theta_0) - Y_t) - \alpha], \end{aligned} \tag{26}$$

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

$$\begin{aligned} \sup_{\theta\in\Theta} |\nabla_\theta \ln f_t^*(Y_t,\theta)| \quad \leqslant \quad & 2\sup_{\theta\in\Theta} |\nabla_\theta\lambda(\theta)| + \sup_{\theta\in\Theta} |\lambda(\theta)| M_0 |\nabla_\theta q_t(\theta)| + \\ & + C_1 \sup_{\theta\in\Theta} \left| \frac{f_t^0(q_t(\theta))\nabla_\theta q_t(\theta)}{1 - \mathbb{1}(q_t(\theta) - Y_t) - F_t^0(q_t(\theta))} \right|, a.s.-P, \end{aligned} \tag{27}$$

where $C_1 \equiv \sup_{x\in[0,\sup_{\theta\in\Theta}\lambda(\theta)]} |\frac{x}{1-\exp(-x)}| < \infty$. We have $\sup_{t\geqslant 1}\sup_{\theta\in\Theta} F_t^0(q_t(\theta)) \in (a,b)$ with $a>0$ and $b<1$, so $C_2 \equiv \sup_{t\geqslant 1}\sup_{y\in\mathbb{R}}\sup_{\theta\in\Theta}(|1-\mathbb{1}(q_t(\theta)-y)-F_t^0(q_t(\theta))|^{-1}) < \infty$ 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 (A7)(i) we know that $E[\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)|] < \infty$, 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 (2005b) holds; together with (A6) 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 (2005b): 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

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

Hence

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

The facts that $P(\mathbb{1}\{q_t(\tilde{\theta}_T^*) = Y_t\} \neq 0) = 0$ and that $E[\left|\frac{\partial\lambda(\tilde{\theta}_T^*)}{\partial\theta_j}\right| + \left|\lambda(\tilde{\theta}_T^*)f_t^0(q_t(\tilde{\theta}_T^*))\frac{\partial q_t(\tilde{\theta}_T^*)}{\partial\theta_j}\right|]$ 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 (A5), θ_0 is an interior point of Θ 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 (23) and (24) 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. For $u(z)$ in (22) we have

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

$$
\begin{aligned}
&\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta) \\
&= \frac{\Delta_{\theta\theta}\lambda(\theta)}{\lambda(\theta)} - \frac{\nabla_\theta \lambda(\theta) \nabla_\theta \lambda(\theta)'}{[\lambda(\theta)]^2} + \Delta_{\theta\theta}\lambda(\theta)[F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbb{1}(q_t(\theta) - Y_t) - \alpha] \\
&+ 2\nabla_\theta \lambda(\theta) \nabla_\theta q_t(\theta)' \left\{ f_t^0(q_t(\theta))[\alpha - \mathbb{1}(q_t(\theta) - Y_t)] + \delta(q_t(\theta) - Y_t)[F_t^0(Y_t) - F_t^0(q_t(\theta))] \right\} \\
&+ \lambda(\theta) \nabla_\theta q_t(\theta) \nabla_\theta q_t(\theta)' \left\{ \frac{df_t^0(q_t(\theta))}{dq}[\alpha - \mathbb{1}(q_t(\theta) - Y_t)] \right. \\
&\left. -2 f_t^0(q_t(\theta))\delta(q_t(\theta) - Y_t) + [F_t^0(Y_t) - F_t^0(q_t(\theta))] \frac{d\delta(q_t(\theta) - Y_t)}{dq} \right\} \\
&+ \lambda(\theta)\Delta_{\theta\theta} q_t(\theta) \left\{ f_t^0(q_t(\theta))[\alpha - \mathbb{1}(q_t(\theta) - Y_t)] + [F_t^0(Y_t) - F_t^0(q_t(\theta))]\delta(q_t(\theta) - Y_t) \right\} \\
&+ \left[u(\lambda(\theta)x(\theta)) + (u(\lambda(\theta)x(\theta)))^2 \right] (\nabla_\theta(\lambda(\theta)x(\theta))) (\nabla_\theta(\lambda(\theta)x(\theta)))' \\
&+ u(\lambda(\theta)x(\theta)) \Delta_{\theta\theta}(\lambda(\theta)x(\theta)), \qquad (28)
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
&\left[u(\lambda(\theta)x(\theta)) + (u(\lambda(\theta)x(\theta)))^2 \right] (\nabla_\theta(\lambda(\theta)x(\theta))) (\nabla_\theta(\lambda(\theta)x(\theta)))' \\
&= \frac{\nabla_\theta\lambda(\theta)\nabla_\theta\lambda(\theta)'}{[\lambda(\theta)]^2} + 2\frac{\nabla_\theta\lambda(\theta)\nabla_\theta x(\theta)'}{\lambda(\theta)x(\theta)} + \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' \times \\
&\left\{ f_t^0(q_t(\theta)) \frac{[\alpha - \mathbb{1}(q_t(\theta) - Y_t)]}{x(\theta)} + \delta(q_t(\theta) - Y_t) \frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\}^2 + o_p(1), \qquad (29)
\end{aligned}
$$

in the neighborhood of θ_0. Similarly,

$$
\begin{aligned}
& u(\lambda(\theta)x(\theta))\Delta_{\theta\theta}(\lambda(\theta)x(\theta)) \\
& = -\frac{\Delta_{\theta\theta}\lambda(\theta)}{\lambda(\theta)} - \frac{1}{2}\Delta_{\theta\theta}\lambda(\theta)x(\theta) - 2\frac{\nabla_\theta\lambda(\theta)\nabla_\theta x(\theta)'}{\lambda(\theta)x(\theta)} \\
& - \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' \left\{ \frac{df_t^0(q_t(\theta))}{dq}\frac{[\alpha - \mathbb{1}(q_t(\theta) - Y_t)]}{x(\theta)} \right. \\
& \left. -2\frac{f_t^0(q_t(\theta))\delta(q_t(\theta) - Y_t)}{x(\theta)} + \frac{d\delta(q_t(\theta) - Y_t)}{dq}\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} \\
& - \Delta_{\theta\theta}q_t(\theta) \left\{ f_t^0(q_t(\theta))\frac{[\alpha - \mathbb{1}(q_t(\theta) - Y_t)]}{x(\theta)} + \delta(q_t(\theta) - Y_t)\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} + o_p(1),
\end{aligned} \tag{30}
$$

in the neighborhood of θ_0. Combining (28) with (29) and (30), we then get that, for any t, $1 \leqslant t \leqslant T, T \geqslant 1$,

$$
\begin{aligned}
& \Delta_{\theta\theta} \ln f_t^*(Y_t, \theta) \\
& = \Delta_{\theta\theta}\lambda(\theta) \left\{ [F_t^0(Y_t) - F_t^0(q_t(\theta))][\mathbb{1}(q_t(\theta) - Y_t) - \alpha] - \frac{1}{2}x(\theta) \right\} \\
& + \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' \left\{ f_t^0(q_t(\theta))\frac{[\alpha - \mathbb{1}(q_t(\theta) - Y_t)]}{x(\theta)} + \delta(q_t(\theta) - Y_t)\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\}^2 \\
& - \nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' \left\{ \frac{df_t^0(q_t(\theta))}{dq}\frac{[\alpha - \mathbb{1}(q_t(\theta) - Y_t)]}{x(\theta)} \right. \\
& \left. -2\frac{f_t^0(q_t(\theta))\delta(q_t(\theta) - Y_t)}{x(\theta)} + \frac{d\delta(q_t(\theta) - Y_t)}{dq}\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} \\
& - \Delta_{\theta\theta}q_t(\theta) \left\{ f_t^0(q_t(\theta))\frac{[\alpha - \mathbb{1}(q_t(\theta) - Y_t)]}{x(\theta)} + \delta(q_t(\theta) - Y_t)\frac{[\alpha - F_t^0(q_t(\theta))]}{x(\theta)} \right\} + o_p(1),
\end{aligned} \tag{31}
$$

in the neigborhood of θ_0. Using $\alpha = F_t^0(q_t(\theta_0))$ and $x(\theta_0) = -\alpha(1-\alpha)$ we have

$$
\begin{aligned}
|\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta_0)| \qu\leqslant \quad & |\Delta_{\theta\theta}\lambda(\theta_0)|\frac{5}{2} + |\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'| \left(\frac{M_0^2}{[\alpha(1-\alpha)]^2} + \frac{M_1}{\alpha(1-\alpha)} \right) \\
& + |\Delta_{\theta\theta}q_t(\theta_0)|\frac{M_0}{\alpha(1-\alpha)} + o_p(1),
\end{aligned}
$$

with $|\Delta_{\theta\theta}\lambda(\theta_0)| < \infty$. From (A7)(i) we have $E[|\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)'|] < \infty$ 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

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

since

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

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

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

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

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

where the last equality uses $E_t\left(\mathbb{1}(q_t(\theta_0) - Y_t) - \alpha\right) = 0$, $a.s. - P$ and $E_t(\delta(q_t(\theta_0) - Y_t)) = f_t^0(q_t(\theta_0))$, $a.s. - P$. Finally, using the same reasoning gives

$$E\left(\Delta_{\theta\theta} q_t(\theta_0)\left\{f_t^0(q_t(\theta_0))\frac{[\alpha - \mathbb{1}(q_t(\theta_0) - Y_t)]}{x(\theta_0)} + \delta(q_t(\theta_0) - Y_t)\frac{[\alpha - F_t^0(q_t(\theta_0))]}{x(\theta_0)}\right\}\right) = 0.$$

Combining the above results then yields, by (31),

$$E[\Delta_{\theta\theta} \ln f_t^*(Y_t, \theta_0)] = -E\left(\nabla_\theta q_t(\theta_0)\nabla_\theta q_t(\theta_0)' \frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)}\right), \tag{32}$$

for all t, $1 \leqslant t \leqslant T, T \geqslant 1$. Hence, for any $\chi \in \mathbb{R}^k$,

$$\chi' \nabla_\theta E[\nabla_\theta' L_T(\theta_0)]\chi = -T^{-1}\sum_{t=1}^{T} E\left(|\nabla_\theta q_t(\theta_0)'\chi|^2 \frac{[f_t^0(q_t(\theta_0))]^2}{\alpha(1-\alpha)}\right) \leqslant 0,$$

with equality if and only if $\chi = 0$. Hence $\nabla_\theta E[\nabla_\theta' L_T(\theta_0)]$ is negative definite (therefore nonsingular).

We now check condition (iv) of Theorem 7.2 by using a CLT for α-mixing sequences (e.g. Theorem 5.20 in White, 2001, p.130). By (A6), 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., Theorem 3.49 in White, 2001, p.50). Moreover, using (23) and (A1), $E[\nabla_\theta \ln f_t^*(Y_t, \theta_0)] = 0$ and using (A7)(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,

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

where the first equality uses $E_t\left(\nabla_\theta \ln f_t^*(Y_t, \theta_0)\right) = 0$, $a.s. - P$, implied by (A1), 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 $(V_T^0)^{-1/2}\sqrt{T}\nabla_\theta L_T(\theta_0) \xrightarrow{d} \mathcal{N}(0, \text{Id})$ 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 θ_0, we have, from (23), (26) and (28), $r_t(\theta) \leqslant r_t^{(1)}(\theta) + r_t^{(2)}(\theta) + r_t^{(3)}(\theta) + o_p(1)$, where

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

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

$$E\left(\sup_{\theta \in \mathring{\Theta}: |\theta - \theta_0| < \varepsilon_1} r_t^{(1)}(\theta) \right) < \infty. \tag{33}$$

Now, note that $|f_t^0(q_t(\theta))[\mathbb{1}(q_t(\theta) - Y_t) - \alpha]| \leqslant M_0$ for any $\theta \in \mathring{\Theta}$, so

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

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

$$
\begin{aligned}
& E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} r_t^{(2)}(\theta)\right) \\
& \leqslant E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}\left\{\frac{M_0}{2} + E_t\left(\delta(q_t(\theta)-Y_t)\left|\frac{F_t^0(q_t(\theta))}{2} - F_t^0(Y_t) + \frac{\alpha}{2}\right|\right)\right\}\right. \\
& \left.\times\,|\nabla_\theta\lambda(\theta_c)|\cdot|\nabla_\theta q_t(\theta)|\right) \\
& \leqslant \frac{M_0}{2}\cdot E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}|\nabla_\theta\lambda(\theta_c)|\cdot|\nabla_\theta q_t(\theta)|\right) \\
& + \frac{1}{2}E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}|\nabla_\theta\lambda(\theta_c)|\cdot|\nabla_\theta q_t(\theta)|\cdot|\alpha - F_t^0(q_t(\theta))|f_t^0(q_t(\theta))\right) \\
& \leqslant \frac{M_0}{2}\cdot\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}|\nabla_\theta\lambda(\theta_c)|\right) \\
& \times\left[E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}|\nabla_\theta q_t(\theta)|\right) + M_0 E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}|\nabla_\theta q_t(\theta)|\cdot|\nabla_\theta q_t(\theta_d)|\right)\right] \\
& < \infty, \qquad (34)
\end{aligned}
$$

where the last inequality uses the continuity of $\nabla_\theta\lambda(\cdot)$ on $\mathbb{R}^k$, (A7)(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

$$
\begin{aligned}
& E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1} r_t^{(3)}(\theta)\right) \\
& \leqslant E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}[|r_x(\theta)| + |R_x(\theta)|]/|x(\theta)|\right) \\
& \leqslant E\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}[1/|x(\theta)|]\left(\sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}|r_x(\theta)| + \sup_{\theta\in\mathring{\Theta}:|\theta-\theta_0|<\varepsilon_1}|R_x(\theta)|\right)\right) \\
& < \infty, \qquad (35)
\end{aligned}
$$

where the last inequality uses the fact that $\sup_{t\geqslant 1}\sup_{\theta\in\Theta}F_t^0(q_t(\theta)) \in (a,b)$ with $a > 0$ and $b < 1$, so $C_3 \equiv \sup_{t\geqslant 1}\sup_{y\in\mathbb{R}}\sup_{\theta\in\Theta}\left(|[\mathbb{1}(q_t(\theta)-Y_t)-\alpha][1-\mathbb{1}(q_t(\theta)-y)-F_t^0(q_t(\theta))]|^{-1}\right) < \infty$. Combining results $(33)-(35)$ 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 θ in a neighborhood of θ_0 we have $T^{-1}\sum_{t=1}^T \Delta_{\theta\theta}\ln f_t^*(Y_t,\theta) \xrightarrow{p} \nabla_\theta E[\nabla_\theta' L_T(\theta)]$. By (A6), 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. Theorem 3.49 in White, 2001, p.50). Now note that given $\theta \in \mathring{\Theta}$, there exists $\theta_a = a\theta_0 + (1-a)\theta$, $a \in (0,1)$, such that for any $\eta > 0$

$$\begin{aligned} P\left(\delta(q_t(\theta)-Y_t)\left|\alpha - F_t^0(q_t(\theta))\right| > \eta\right) &\leqslant E\left(\left|\alpha - F_t^0(q_t(\theta))\right| f_t^0(q_t(\theta))\right)/\eta \\ &\leqslant |\theta-\theta_0| E\left(|\nabla_\theta q_t(\theta_a)| f_t^0(q_t(\theta_a)) f_t^0(q_t(\theta))\right)/\eta \\ &\leqslant |\theta-\theta_0| M_0^2 E[\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)|]/\eta, \end{aligned} \tag{36}$$

so that in a neighborhood of θ_0, $\delta(q_t(\theta)-Y_t)\left|\alpha - F_t^0(q_t(\theta))\right| = o_p(1)$. Similarly,

$$\begin{aligned} P\left([d\delta(q_t(\theta)-Y_t)/dq]\left|\alpha - F_t^0(q_t(\theta))\right| > \eta\right) &\leqslant E\left(\left|\alpha - F_t^0(q_t(\theta))\right| df_t^0(q_t(\theta))/dq\right)/\eta \\ &\leqslant |\theta-\theta_0| M_0 M_1 E[\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)|]/\eta \end{aligned} \tag{37}$$

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 (31) we have that for any t, $1 \leqslant t \leqslant T, T \geqslant 1$,

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

in a neighborhood of θ_0, which combined with (36) and (37) gives

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

so that for a given $\varepsilon > 0$, there is a positive constant $n_{r,\varepsilon}$ such that

$$
\begin{aligned}
&|\Delta_{\theta\theta} \ln f_t^*(Y_t,\theta)|^{r+\varepsilon} \\
&\leqslant n_{r,\varepsilon} \Big(|\Delta_{\theta\theta}\lambda(\theta)|^{r+\varepsilon} (5/2)^{r+\varepsilon} + |\nabla_\theta q_t(\theta) \nabla_\theta q_t(\theta)'|^{r+\varepsilon} C_3^{2(r+\varepsilon)} \left\{ M_0^2 + M_1 \right\}^{r+\varepsilon} \\
&+ |\Delta_{\theta\theta} q_t(\theta)|^{r+\varepsilon} C_3^{r+\varepsilon} M_0^{r+\varepsilon} \Big) + o_p(1),
\end{aligned}
$$

in a neigborhood of θ_0, and so using (A7)(i) and the fact that $|\Delta_{\theta\theta}\lambda(\theta)| < \infty$ in a neighborhood of θ_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 $\tilde{\theta}_T^*$. □

Proof of Lemma 7. We proceed in two steps.

STEP1: 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 (6) 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 $\mathbb{1}(q_t(\theta) - y) = \mathbb{1}[\lambda(\theta)F_t^0(q_t(\theta)) - u]$, we have

$$
\begin{aligned}
\int_{\mathbb{R}} f_t^*(y,\theta)dy &= \int_0^{\lambda(\theta)F_t^0(q_t(\theta))} \frac{\alpha(1-\alpha)\exp\{(1-\alpha)[u - \lambda(\theta)F_t^0(q_t(\theta))]\}}{1 - \exp[-(1-\alpha)\lambda(\theta)F_t^0(q_t(\theta))]} du \\
&\quad + \int_{\lambda(\theta)F_t^0(q_t(\theta))}^{\lambda(\theta)} \frac{\alpha(1-\alpha)\exp\{-\alpha[u - \lambda(\theta)F_t^0(q_t(\theta))]\}}{1 - \exp\{-\alpha\lambda(\theta)[1 - F_t^0(q_t(\theta))]\}} du \\
&= \frac{\alpha \exp[-(1-\alpha)\lambda(\theta)F_t^0(q_t(\theta))]}{1 - \exp[-(1-\alpha)\lambda(\theta)F_t^0(q_t(\theta))]} \Big[\exp[(1-\alpha)u \Big]_0^{\lambda(\theta)F_t^0(q_t(\theta))} \\
&\quad + \frac{(1-\alpha)\exp[\alpha\lambda(\theta)F_t^0(q_t(\theta))]}{1 - \exp\{-\alpha\lambda(\theta)[1 - F_t^0(q_t(\theta))]\}} \Big[-\exp(-\alpha u) \Big]_{\lambda(\theta)F_t^0(q_t(\theta))}^{\lambda(\theta)} \\
&= \alpha + (1-\alpha) = 1,
\end{aligned}
$$

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 θ_0 and that $f_t^*(\cdot,\theta_0) = f_t^0(\cdot)$. For this, let

$$
P_t(\theta) \equiv \alpha(1-\alpha)\lambda(\theta)\exp\{\lambda(\theta)[F_t^0(y) - F_t^0(q_t(\theta))][\mathbb{1}(q_t(\theta) - y) - \alpha]\}, \tag{38}
$$

$$
Q_t(\theta) \equiv 1 - \exp\{\lambda(\theta)[1 - F_t^0(q_t(\theta)) - \mathbb{1}(q_t(\theta) - y)][\mathbb{1}(q_t(\theta) - y) - \alpha]\}, \tag{39}
$$

so that $f_t^*(y,\theta) = f_t^0(y)P_t(\theta)/Q_t(\theta)$. By (A0)(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 as

$$P_t(\theta) = \sum_{|l|\leqslant 2} \frac{D^l P_t(\theta_0)}{l!}(\theta-\theta_0)^l + o(|\theta-\theta_0|^2), \tag{40}$$

$$Q_t(\theta) = \sum_{|l|\leqslant 2} \frac{D^l Q_t(\theta_0)}{l!}(\theta-\theta_0)^l + o(|\theta-\theta_0|^2). \tag{41}$$

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

$$P_t(\theta_0) = 0, D^1 P_t(\theta_0) = 0, D^2 P_t(\theta_0) = \alpha(1-\alpha)D^2\lambda(\theta_0), \tag{42}$$

and

$$Q_t(\theta_0) = 0, D^1 Q_t(\theta_0) = 0, D^2 Q_t(\theta_0) = \alpha(1-\alpha)D^2\lambda(\theta_0), \tag{43}$$

Hence

$$P_t(\theta) = \tfrac{1}{2}\alpha(1-\alpha)D^2\lambda(\theta_0)(\theta-\theta_0)^2 + o(|\theta-\theta_0|^2), \tag{44}$$

$$Q_t(\theta) = \tfrac{1}{2}\alpha(1-\alpha)D^2\lambda(\theta_0)(\theta-\theta_0)^2 + o(|\theta-\theta_0|^2). \tag{45}$$

Given the nonsingularity of $\Delta_{\theta\theta}\Lambda(0)$, an immediate consequence of l'Hôpital's rule and (44)−(45) 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.

STEP 2: 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 (A1) by assumption. When $\theta \neq \theta_0$, using again the change of variable $u \equiv \lambda(\theta)F_t^0(y)$, we have

$$\begin{aligned} E_\theta[\mathbb{1}(q_t(\theta) - Y_t)|\mathcal{W}_t] &= \int_{-\infty}^{q_t(\theta)} f_t^*(y,\theta)dy \\ &= \int_0^{\lambda(\theta)F_t^0(q_t(\theta))} \frac{\alpha(1-\alpha)\exp\{(1-\alpha)[u-\lambda(\theta)F_t^0(q_t(\theta))]\}}{1-\exp[-(1-\alpha)\lambda(\theta)F_t^0(q_t(\theta))]}du = \alpha. \end{aligned}$$

□

Proof of Theorem 5. From Theorem 3 we know that θ_T^* which minimizes $\Psi_T^*(\theta)$ is consistent for θ_0. Thus, in order to establish the consistency of $\hat{\theta}_T$, it suffices to show that $\hat{\Psi}_T(\theta) - \Psi_T^*(\theta)$ converges uniformly (in θ) to zero, i.e. $\sup_{\theta\in\Theta}|\hat{\Psi}_T(\theta) - \Psi_T^*(\theta)| = o_p(1)$. For this we need a

uniform consistency property of $D^\lambda \hat{G}(\cdot,\cdot)$, where D^λ denotes the λth derivative with respect to y.

Lemma 8. *Suppose that (A6), (A8)-(A9), (A11) hold. Then,* $\sup_{(y,w)\in\mathbb{R}^{m+1}} |D^\lambda \hat{G}(y,w) - H^0_{\lambda T}(y,w)| = O_p[1/(\sqrt{T}h^\lambda_{yT}h^m_{wT})] + O_p(h^R_{wT}) + O_p(h^R_{yT})$*, where* $H^0_{\lambda T}(y,w) \equiv D^\lambda F^0(y|w)\bar{g}^0_T(w)$ *and* $\lambda = 0, 1, 2$.

We will also need the uniform consistency of $\hat{g}(\cdot)$ for $\bar{g}^0_T(\cdot) \equiv (1/T)\sum_{t=1}^T g^0_t(\cdot)$:

$$\sup_{w\in\mathbb{R}^m} |\hat{g}(w) - \bar{g}^0_T(w)| = O_p[1/(\sqrt{T}h^m_{wT})] + O_p(h^R_{wT}), \tag{46}$$

which follows from Theorem 1(a) in Andrews (1995) with $\eta = \infty$ given (A6), (A8) and (A11)(i)-(ii). We let $q_t(\theta) = q_\alpha(W_t,\theta)$ as previously, and $b^\epsilon_T \equiv b_T + \epsilon_T$, $d^\epsilon_t \equiv \mathbb{1}[\bar{g}^0_T(W_t) - b^\epsilon_T]$ and $\Psi^\epsilon_T(\theta)$ be equal to $\hat{\Psi}_T(\theta)$ where $\hat{F}_t(\cdot) \equiv d_t\hat{F}(\cdot|W_t)$ is replaced by $d^\epsilon_t\hat{F}(\cdot|W_t)$, i.e.

$$\Psi^\epsilon_T(\theta) = \frac{1}{T}\sum_{t=1}^T d^\epsilon_t[\alpha - \mathbb{1}(q_t(\theta) - Y_t)][\hat{F}(Y_t|W_t) - \hat{F}(q_t(\theta)|W_t)], \tag{47}$$

where $\{\epsilon_T\} > 0$ is an appropriate vanishing sequence. The remainder of the proof adapts the consistency proof of Theorem 1 in Lavergne and Vuong (1996). Let ϵ_T be such that $\epsilon_T = o(b_T)$, $\epsilon_T\sqrt{T}h^m_{wT} \to \infty$, $\epsilon_T/h^R_{wT} \to \infty$ and $\epsilon_T/h^R_{yT} \to \infty$. As

$$\sup_{\theta\in\Theta} |\hat{\Psi}_T(\theta) - \Psi^*_T(\theta)| \leqslant \sup_{\theta\in\Theta} |\hat{\Psi}_T(\theta) - \Psi^\epsilon_T(\theta)| + \sup_{\theta\in\Theta} |\Psi^\epsilon_T(\theta) - \Psi^*_T(\theta)|,$$

where $\Psi^\epsilon_T(\theta)$ is defined in Equation (47), it suffices to prove that both terms in the right-hand side are $o_p(1)$. Given Lemma 8 and Equation (46) we will use

$$a_T^{-1} \sup_{(y,w)\in\mathbb{R}^{m+1}} |\hat{G}(y,w) - H^0_{0T}(y,w)| = o_p(1), \tag{48}$$

$$a_T^{-1} \sup_{w\in\mathbb{R}^m} |\hat{g}(w) - \bar{g}^0_T(w)| = o_p(1), \tag{49}$$

which hold for any sequence $\{a_T\}$ satisfying $a_T\sqrt{T}h^m_{wT} \to \infty$, $a_T/h^R_{wT} \to \infty$ and $a_T/h^R_{yT} \to \infty$. We will also use the identity

$$\hat{F}(y|w) - F^0(y|w) = \frac{1}{\hat{g}(w)}[\hat{G}(y,w) - H^0_{0T}(y,w)] - \frac{F^0(y|w)}{\hat{g}(w)}[\hat{g}(w) - \bar{g}^0_T(w)]. \tag{50}$$

STEP 1: We first show that $\sup_{\theta\in\Theta} |\hat{\Psi}_T(\theta) - \Psi^\epsilon_T(\theta)| = o_p(1)$. We have

$$\begin{aligned}\hat{\Psi}_T(\theta) - \Psi^\epsilon_T(\theta) &= \frac{1}{T}\sum_{t=1}^T (J_t - H_t)[\alpha - \mathbb{1}(q_t(\theta) - Y_t)][\hat{F}(Y_t|W_t) - \hat{F}(q_t(\theta)|W_t)] \\ &= \Delta\hat{\Psi}_{1T} - \Delta\hat{\Psi}_{2T} + \Delta\hat{\Psi}_{3T},\end{aligned}$$

where $J_t = d_t(1-d_t^\epsilon)$, $H_t = (1-d_t)d_t^\epsilon$ and

$$\begin{aligned}
\Delta\hat{\Psi}_{1T} &= \frac{1}{T}\sum_{t=1}^{T}(J_t - H_t)[\alpha - \mathbb{1}(q_t(\theta) - Y_t)][\hat{F}(Y_t|W_t) - F^0(Y_t|W_t)],\\
\Delta\hat{\Psi}_{2T} &= \frac{1}{T}\sum_{t=1}^{T}(J_t - H_t)[\alpha - \mathbb{1}(q_t(\theta) - Y_t)][\hat{F}(q_t(\theta)|W_t) - F^0(q_t(\theta)|W_t)],\\
\Delta\hat{\Psi}_{3T} &= \frac{1}{T}\sum_{t=1}^{T}(J_t - H_t)[\alpha - \mathbb{1}(q_t(\theta) - Y_t)][F^0(Y_t|W_t) - F^0(q_t(\theta)|W_t)].
\end{aligned}$$

As $H_t \leqslant \mathbb{1}[|\hat{g}(W_t) - \bar{g}_T^0(W_t)| - \epsilon_T]$ and the event $\{\sup_w |\hat{g}(w) - \bar{g}_T^0(w)| > \epsilon_T\}$ has asymptotic probability 0 because Property (49) holds with $a_T = \epsilon_T$ by construction of ϵ_T, we have $\sup_{1\leqslant t\leqslant T, T\geqslant 1} H_t = 0$ with probability approaching one. Hence, we need to consider the J_t terms only. Namely, it suffices to show that $\sup_{\theta\in\Theta}\Delta\hat{\Psi}_{jT}^J = o_p(1)$ for $j = 1,2,3$. Using Identity (50) and the definition of J_t, we obtain

$$|\Delta\hat{\Psi}_{1T}^J| \leqslant b_T^{-1}\left[\sup_{(y,w)\in\mathbb{R}^{m+1}}|\hat{G}(y,w) - H_{0T}^0(y,w)| + \sup_{w\in\mathbb{R}^m}|\hat{g}(w) - g_T^0(w)|\right]\frac{1}{T}\sum_{t=1}^{T}J_t$$

Because $(1/T)\sum_{t=1}^T J_t \leqslant 1$, we get $\sup_{\theta\in\Theta}\Delta\hat{\Psi}_{1T}^J = o_p(1)$ in view of Properties (48) – (49) with $a_T = b_T$ under our assumptions on b_T. Similarly, $\sup_{\theta\in\Theta}|\Delta\hat{\Psi}_{2T}^J| = o_p(1)$. Regarding $\Delta\hat{\Psi}_{3T}^J$, we have $|\Delta\hat{\Psi}_{3T}^J| \leqslant (1/T)\sum_{t=1}^T J_t$. But $(1/T)\sum_{t=1}^T J_t \leqslant (1/T)\sum_{t=1}^T(1-d_t^\epsilon)$ with

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(1-d_t^\epsilon)\right] = \frac{1}{T}\sum_{t=1}^{T}\int_{\{w:\bar{g}_T^0(w)<b_T^\epsilon\}} g_t(w)dw = \int_{\{w:\bar{g}_T^0(w)<b_T^\epsilon\}}\bar{g}_T^0(w)dw = o(1),$$

where the last equality follows by taking $c_T = b_T^\epsilon$ in (A10)(i). Hence, $(1/T)\sum_{t=1}^T(1-d_t^\epsilon) = o_p(1)$ by Markov inequality. Thus,

$$\frac{1}{T}\sum_{t=1}^{T}J_t = o_p(1), \tag{51}$$

and $\sup_{\theta\in\Theta}\Delta\hat{\Psi}_{3T}^J = o_p(1)$.

STEP 2: We next show that $\sup_{\theta\in\Theta}|\Psi_T^\epsilon(\theta)-\Psi_T^*(\theta)|=o_p(1)$. We have

$$\begin{aligned}\Psi_T^\epsilon(\theta)-\Psi_T^*(\theta) &= \frac{1}{T}\sum_{t=1}^T d_t^\epsilon[\alpha-\mathbb{1}(q_t(\theta)-Y_t)][\hat{F}(Y_t|W_t)-F^0(Y_t|W_t)]\\ &\quad -\frac{1}{T}\sum_{t=1}^T d_T^\epsilon[\alpha-\mathbb{1}(q_t(\theta)-Y_t)][\hat{F}(q_t(\theta)|W_t)-F^0(q_t(\theta)|W_t)]\\ &\quad -\frac{1}{T}\sum_{t=1}^T (1-d_T^\epsilon)[\alpha-\mathbb{1}(q_t(\theta)-Y_t)][F^0(Y_t|W_t)-F^0(q_t(\theta)|W_t)]\\ &\equiv \Delta\Psi_{1T}^\epsilon-\Delta\Psi_{2T}^\epsilon-\Delta\Psi_{3T}^\epsilon\end{aligned}$$

Thus, it suffices to show that $\sup_{\theta\in\Theta}\Delta\Psi_{jT}^\epsilon=o_p(1)$ for $j=1,2,3$. Because $\epsilon_T=o(b_T)$, $b_T^\epsilon\equiv b_T+\epsilon_T$ is a sequence satisfying $b_T^\epsilon\sqrt{T}h_{wT}^m\to\infty$, $b_T^\epsilon/h_{wT}^R\to\infty$ and $b_T^\epsilon/h_{yT}^R\to\infty$ so that Properties (48) – (49) hold with $a_T=b_T^\epsilon$. In particular, Property (49) implies $P\left(\inf_{\{w:\bar{g}_T^0(w)\geqslant b_T^\epsilon\}}\hat{g}(w)>b_T^\epsilon(1-\eta)\right)\to 1$ as $T\to\infty$ for any $\eta\in(0,1)$. Thus, using Identity (50), we have

$$|\Delta\Psi_{1T}^\epsilon|\leqslant(b_T^\epsilon)^{-1}(1-\eta)^{-1}\left\{\sup_{(y,w)\in\mathbb{R}^{m+1}}|\hat{G}(y,w)-H_{0T}^0(y,w)|+\sup_{w\in\mathbb{R}^m}|\hat{g}(w)-\bar{g}_T^0(w)|\right\}\frac{1}{T}\sum_{t=1}^T d_t^\epsilon,$$

with probability approaching 1, where $(1/T)\sum_{t=1}^T d_t^\epsilon\leqslant 1$. Hence, $\sup_{\theta\in\Theta}\Delta\Psi_{1T}^\epsilon=o_p(1)$ using Properties (48) – (49) with $a_T=b_T^\epsilon$. Similarly, $\sup_{\theta\in\Theta}\Delta\Psi_{2T}^\epsilon=o_p(1)$. Regarding $\Delta\Psi_{3T}^\epsilon$, we have $\sup_{\theta\in\Theta}|\Delta\Psi_{3T}^\epsilon|\leqslant(1/T)\sum_{t=1}^T(1-d_T^\epsilon)=o_p(1)$ from Step 1. □

Proof of Lemma 8. The proof adapts that of Lemma A-1 in Andrews (1995) to incorporate the supremum over y-values, which leads to the additional term $O_p(h_{yT}^R)$. It is done in three steps. Recall that $L(\cdot)$ was defined as $L(y)\equiv\int\mathbb{1}(y-u)K_0(u)du$. Let $I_t(y)$ be $L[(y-Y_t)/h_{yT}]$ if $\lambda=0$, $K_0[(y-Y_t)/h_{yT}]$ if $\lambda=1$, and $K_0'[(y-Y_t)/h_{yT}]$ if $\lambda=2$. Thus, omitting the subscript T, we have

$$\begin{aligned}&\sup_{(y,w)\in\mathbb{R}^{m+1}}|D^\lambda\hat{G}(y,w)-H_\lambda^0(y,w)|\\ &=\sup_{(y,w)\in\mathbb{R}^{m+1}}\left|\frac{1}{Th_y^\lambda h_w^m}\sum_{t=1}^T I_t(y)K\left(\frac{w-W_t}{h_w}\right)-D^\lambda F^0(y|w)\frac{1}{T}\sum_{t=1}^T g_t^0(w)\right|\end{aligned}$$

for $\lambda = 0, 1, 2$. The desired result then follows from: (i)

$$\sup_{(y,w)\in\mathbb{R}^{m+1}} \left| \frac{1}{Th_y^\lambda h_w^m} \sum_{t=1}^T I_t(y) K\left(\frac{w - W_t}{h_w}\right) - \frac{1}{Th_y^\lambda h_w^m} \sum_{t=1}^T E\left[I_t(y) K\left(\frac{w - W_t}{h_w}\right)\right] \right| = O_p\left(\frac{1}{\sqrt{T} h_y^\lambda h_w^m}\right), \tag{52}$$

which is proved in Step 1 by adapting Andrews' (1995) proof of Lemma A-2, (ii)

$$\sup_{(y,w)\in\mathbb{R}^{m+1}} \left| \frac{1}{Th_y^\lambda h_w^m} \sum_{t=1}^T E\left[I_t(y) K\left(\frac{w - W_t}{h_w}\right)\right] - \frac{1}{Th_w^m} \sum_{t=1}^T E\left[D^\lambda F^0(y|W_t) K\left(\frac{w - W_t}{h_w}\right)\right] \right| = O_p\left(h_y^R\right), \tag{53}$$

which is proved in Step 2, and (iii)

$$\sup_{(y,w)\in\mathbb{R}^{m+1}} \left| \frac{1}{Th_w^m} \sum_{t=1}^T E\left[D^\lambda F^0(y|W_t) K\left(\frac{w - W_t}{h_w}\right)\right] - D^\lambda F^0(y|w) \frac{1}{T} \sum_{t=1}^T g_t^0(w) \right| = O_p\left(h_w^R\right), \tag{54}$$

which is proved in Step 3.

STEP1: When $\lambda = 0$, note that $|I_t(y)| \leqslant \int |K_0(u)| du < \infty$ by (A11)(iii). When $\lambda = 1$, $|I_t(y)| \leqslant \sup_{y\in\mathbb{R}} |K_0(y)| < \infty$ by (A11)(iii). When $\lambda = 2$, $|I_t(y)| \leqslant \sup_{y\in\mathbb{R}} |K_0'(y)| < \infty$ by (A11)(iv). Hence, $I_t(y)$ is bounded by some $C_0 < \infty$. Moreover, (A6) and Theorem 3.49 in White (2001) guarantee that for every y, the sequence $\{(I_t(y), W_t')'\}$ is strong mixing with α of size $-r/(r-2)$, $r > 2$. Hence, for any (t, s), $1 \leqslant t, s \leqslant T$, $T \geqslant 1$, we have $\alpha(|t-s|) = O(|t-s|^{-r/(r-2)-\epsilon})$ for some $\epsilon > 0$ (see Definition 3.45 in White, 2001), and $C_1 \equiv \sum_{s=0}^\infty \alpha(s) < \infty$. Thus, by Billingsley (1995, Lemma 2, p.365), we have

$$\left| \mathrm{Cov}\Big(I_t(y)\cos(v'W_t), I_u(y)\cos(v'W_u)\Big)\right| \leqslant 4C_0^2 \alpha(|t-u|),$$

for any $v \in \mathbb{R}^m$ and any $y, t, u \in \mathbb{R}$. Hence, instead of (A.15) in Andrews (1995), we have

$$\mathrm{Var}\left(\frac{1}{T} \sum_{t=1}^T I_t(y)\cos(v'W_t)\right) \leqslant 8C_0^2 \frac{1}{T} \sum_{t=0}^{T-1} \alpha(s) \leqslant \frac{8C_0^2 C_1}{T}.$$

As this also holds for $\sin(\cdot)$ replacing $\cos(\cdot)$, Lyapunov inequality implies

$$E\left| \frac{1}{T} \sum_{t=1}^T \Big\{ I_t(y)\exp(iv'W_t) - E[I_t(y)\exp(iv'W_t)]\Big\}\right| \leqslant 2C_0 \sqrt{\frac{8C_1}{T}},$$

for any $v \in \mathbb{R}^m$ and any $y \in \mathbb{R}$. Let L_T denote the left-hand side of (52). Using (A.11) in Andrews (1995) with $\lambda = 0$ and the above inequality, we obtain

$$\begin{aligned} E(L_T) &\leqslant E\left(\int \sup_{y\in\mathbb{R}} \left| \frac{1}{Th_y^\lambda} \sum_{t=1}^{T} \left\{ I_t(y)\exp(iv'W_t) - E[I_t(y)\exp(iv'W_t)] \right\} \right| |\phi(h_w v)|dv \right) \\ &\leqslant 2C_0\sqrt{\frac{8C_1}{T}}\frac{1}{h_y^\lambda}\int |\phi(h_w v)|dv = \frac{C_2}{\sqrt{T}h_y^\lambda h_w^m}, \end{aligned}$$

where $C_2 = 2C_0\sqrt{8}C_1 \int |\phi(u)|du < \infty$ by (A11)(ii) using the change of variable $u = h_w v$. By Markov inequality Equation (52) follows.

STEP 2: Consider first $\lambda = 0$. Using Fubini's Theorem, we note that

$$\begin{aligned} E[I_t(y)|W_t = w] &= \int L\left(\frac{y-Y}{h_y}\right) dF^0(Y|w) \\ &= \int\int \mathbb{1}\left(\frac{y-Y}{h_y} - u\right) K_0(u)du dF^0(Y|w) \\ &= \int F^0(y - h_y u|w)K_0(u)du, \end{aligned} \tag{55}$$

which does not depend on t because of (A9)(i). When $\lambda = 1$, using the change of variable $Y = y - h_y u$, we note that

$$E\left[\frac{I_t(y)}{h_y}|W_t = w\right] = \int \frac{1}{h_y}K_0\left(\frac{y-Y}{h_y}\right) f^0(Y|w)dY = \int f^0(y - h_y u|w)K_0(u)du. \tag{56}$$

When $\lambda = 2$, using the change of variable $Y = y - h_y u$ and integration by parts, we have

$$\begin{aligned} E\left[\frac{I_t(y)}{h_y^2}|W_t = w\right] &= \int \frac{1}{h_y^2}K_0'\left(\frac{y-Y}{h_y}\right) f^0(Y|w)dY = \frac{1}{h_y}\int f^0(y - h_y u|w)K_0'(u)du \\ &= \int Df^0(y - h_y u|w)K_0(u)du. \end{aligned} \tag{57}$$

Now, let $L_T(y,w)$ denote the term inside the absolute value on the left-hand side of Equation (53). Combining Equations (55)-(57) with (A11)(iii), we have

$$\begin{aligned}
&L_T(y,w)\\
&= \frac{1}{Th_w^m}\sum_{t=1}^{T} E\left\{\left[\frac{I_t(y)}{h_y^\lambda} - D^\lambda F^0(y|W_t)\right] K\left(\frac{w-W_t}{h_w}\right)\right\}\\
&= \frac{1}{Th_w^m}\sum_{t=1}^{T} E\left\{\left[\int [D^\lambda F^0(y-h_y u|W_t) - D^\lambda F^0(y|W_t)]K_0(u)du\right] K\left(\frac{w-W_t}{h_w}\right)\right\}\\
&= \int\left[\int [D^\lambda F^0(y-h_y u|W) - D^\lambda F^0(y|W)]K_0(u)du\right] K\left(\frac{w-W}{h_w}\right)\frac{1}{Th_w^m}\sum_{t=1}^{T} g_t^0(W)dW.
\end{aligned}$$

Hence, taking an Rth-order Taylor expansion of $D^\lambda F^0(y-h_{yT}u|W)$ at y, and using (A11)(iii) we obtain

$$\begin{aligned}
\sup_{(y,w)\in\mathbb{R}^{m+1}}|L_T(y,w)| \;\leqslant\; & h_y^R \sup_{(y,w)\in\mathbb{R}^{m+1}}|D^{\lambda+R}F^0(y|w)|\int|u^R K_0(u)|du\int|K(\tilde{W})|d\tilde{W}\\
&\times \sup_{T\geqslant 1}\sup_{w\in\mathbb{R}^m}\bar{g}_T^0(w),
\end{aligned}$$

which establishes Equation (53) because of (A8), (A9)(ii), and (A11)(i,iii).

STEP 3: The study of the bias (54) is standard as in the proof of Lemma A-3 in Andrews (1995). Using (A9)(i) we have

$$\begin{aligned}
\frac{1}{Th_w^m}\sum_{t=1}^{T} E\left[D^\lambda F^0(y|W_t)K\left(\frac{w-W_t}{h_w}\right)\right] &= \frac{1}{Th_w^m}\sum_{t=1}^{T}\int D^\lambda F^0(y|W)K\left(\frac{w-W}{h_w}\right)g_t^0(W)dW\\
&= \int H_\lambda^0(y,w-h_w\tilde{W})K(\tilde{W})d\tilde{W},
\end{aligned}$$

where $\tilde{W}=(w-W)/h_w$. Hence, using a Taylor expansion of order R at w together with (A11)(i) we obtain

$$\begin{aligned}
&\frac{1}{Th_w^m}\sum_{t=1}^{T} E\left[D^\lambda F^0(y|W_t)K\left(\frac{w-W_t}{h_w}\right)\right] - D^\lambda F^0(y|w)\frac{1}{T}\sum_{t=1}^{T} g_t^0(w)\\
&= \int\left[H_\lambda^0(y,w-h_w\tilde{W}) - H_\lambda^0(y,w)\right]K(\tilde{W})d\tilde{W}\\
&= \int\left[\sum_{|r|=R}\frac{(-1)^R}{R!}h_w^R\frac{\partial^R H_\lambda^0(y|w-\tilde{h}_w\tilde{W})}{\partial W_1^{r_1}\ldots\partial W_m^{r_m}}\tilde{W}_1^{r_1}\ldots\tilde{W}_m^{r_m}\right]K(\tilde{W})d\tilde{W},
\end{aligned}$$

where $0<\tilde{h}_w<h_w$. This establishes Equation (54) using (A8), (A9)(ii) and (A11)(i). □

Proof of Theorem 6. From Equation (9), the first order conditions associated with $\hat{\theta}_T$ are

$$\sqrt{T}\nabla_\theta \hat{\Psi}_T(\hat{\theta}_T) = 0, a.s. - P, \tag{58}$$

where $\nabla_\theta \hat{\Psi}_T(\theta) = (1/T)\sum_{t=1}^T \{\mathbb{1}[q_t(\theta) - Y_t] - \alpha\} \hat{f}_t[q_t(\theta)] \nabla_\theta q_t(\theta)$ and $\hat{f}_t(y) = d_t \left[\partial \hat{G}(y, W_t)/\partial y\right] / \hat{g}(W_t)$. Given (A11)(iv) $\hat{f}_t(\cdot)$ is continuously differentiable on $\mathbb{R}$. Thus, a first-order Taylor expansion of the condition (58) at θ_0 gives

$$\sqrt{T}\nabla_\theta \hat{\Psi}_T(\theta_0) + \Delta_{\theta\theta}\hat{\Psi}_T(\bar{\theta}_T^c)\sqrt{T}(\hat{\theta}_T - \theta_0) = 0, a.s. - P, \tag{59}$$

where $\bar{\theta}_T^c \equiv c\theta_0 + (1-c)\hat{\theta}_T$ for some $c \in (0,1)$. To establish the theorem, we need two lemmas:

Lemma 9. *Suppose that (A0)-(A1), (A5)-(A7)(i), (A8)-(A10)(i) and (A11) hold. If* $b_T \to 0$ *with* $b_T\sqrt{T}h_{yT}^2 h_{wT}^m \to \infty$, $b_T/h_{wT}^R \to \infty$ *and* $b_T/h_{yT}^R \to \infty$, *then* $\Delta_{\theta\theta}\hat{\Psi}_T(\bar{\theta}_T^c) - \Delta_{\theta\theta}\Psi_T^*(\theta_0) = o_p(1)$.

In particular, the conditions in Theorem 6 imply the conditions in Lemma 9. Thus $\Delta_{\theta\theta}\hat{\Psi}_T(\bar{\theta}_T^c) - \Delta_{\theta\theta}\Psi_T^*(\theta_0) = o_p(1)$.

Lemma 10. *Suppose that all the conditions of Theorem 6 hold. Then,* $\sqrt{T}[\nabla_\theta \hat{\Psi}_T(\theta_0) - \nabla_\theta \Psi_T^*(\theta_0)] = o_p(1)$.

The remainder of the proof is straightforward: Equation (59), Lemmas 9 and 10 together imply: $\sqrt{T}(\hat{\theta}_T - \theta_0) = -\left[\Delta_{\theta\theta}\Psi_T^*(\theta_0) + o_p(1)\right]^{-1}\left(\sqrt{T}\nabla_\theta \Psi_T^*(\theta_0) + o_p(1)\right), a.s. - P$. Thus $\hat{\theta}_T$ is $\sqrt{T}$-asymptotically equivalent to θ_T^*. The desired result follows. □

Proof of Lemma 9. Note that the assumptions of Theorem 5 are satisfied under those of Lemma 9. Hence, $\hat{\theta}_T \xrightarrow{p} \theta_0$. Moreover, because $\bar{\theta}_T^c = c\theta_0 + (1-c)\hat{\theta}_T$ for some $c \in (0,1)$, we have $\bar{\theta}_T^c \xrightarrow{p} \theta_0$. Thus, it suffices to prove that $\sup_{\theta\in\Theta} |\Delta_{\theta\theta}\hat{\Psi}_T(\theta) - \Delta_{\theta\theta}\Psi_T^*(\theta)| = o_p(1)$, where

$$\begin{aligned}
&\Delta_{\theta\theta}\hat{\Psi}_T(\theta) \\
&= \frac{1}{T}\sum_{t=1}^T d_t \left\{\mathbb{1}[q_t(\theta) - Y_t] - \alpha\right\} \left\{D^2\hat{F}[q_t(\theta)|W_t]\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' + D\hat{F}[q_t(\theta)|W_t]\Delta_{\theta\theta}q_t(\theta)\right\}, \\
&\Delta_{\theta\theta}\Psi_T^*(\theta) \\
&= \frac{1}{T}\sum_{t=1}^T \left\{\mathbb{1}[q_t(\theta) - Y_t] - \alpha\right\} \left\{D^2F^0[q_t(\theta)|W_t]\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' + DF^0[q_t(\theta)|W_t]\Delta_{\theta\theta}q_t(\theta)\right\}.
\end{aligned}$$

Let ϵ_T be such that $\epsilon_T = o(b_T)$, $\epsilon_T\sqrt{T}h_{yT}^2 h_{wT}^m \to \infty$, $\epsilon_T/h_{wT}^R \to \infty$ and $\epsilon_T/h_{yT}^R \to \infty$. As

$$\sup_{\theta\in\Theta}|\Delta_{\theta\theta}\hat{\Psi}_T(\theta) - \Delta_{\theta\theta}\Psi_T^*(\theta)| \leqslant \sup_{\theta\in\Theta}|\Delta_{\theta\theta}\hat{\Psi}_T(\theta) - \Delta_{\theta\theta}\Psi_T^\epsilon(\theta)| + \sup_{\theta\in\Theta}|\Delta_{\theta\theta}\Psi_T^\epsilon(\theta) - \Delta_{\theta\theta}\Psi_T^*(\theta)|,$$

where $\Psi_T^\epsilon(\theta)$ is defined in Equation (47), it suffices to prove that both terms in the right-hand side of the above inequality are $o_p(1)$. Given Lemma 8 and Equation (46) we will use

$$\begin{aligned} a_T^{-1}\sup_{(y,w)\in\mathbb{R}^{m+1}}|D^\lambda\hat{G}(y,w) - H_{\lambda T}^0(y,w)| &= o_p(1), \\ a_T^{-1}\sup_{w\in\mathbb{R}^m}|\hat{g}(w) - \bar{g}_T^0(w)| &= o_p(1), \end{aligned}$$

for $\lambda = 1, 2$, which hold for any sequence $\{a_T\}$ satisfying $a_T\sqrt{T}h_{yT}^2 h_{wT}^m \to \infty$, $a_T/h_{wT}^R \to \infty$ and $a_T/h_{yT}^R \to \infty$. For $\lambda = 1, 2$ we will also use the identity

$$D^\lambda\hat{F}(y|w) - D^\lambda F^0(y|w) = \frac{1}{\hat{g}(w)}[D^\lambda\hat{G}(y,w) - H_{\lambda T}^0(y,w)] - \frac{D^\lambda F^0(y|w)}{\hat{g}(w)}[\hat{g}(w) - \bar{g}_T^0(w)]$$

which follows from Equation (50). The proof then draws from that of Theorem 5. Specifically, in Step 1 we deal with $\Delta_{\theta\theta}\hat{\Psi}_T(\theta) - \Delta_{\theta\theta}\hat{\Psi}_T^\epsilon(\theta) = -\Delta\hat{\Psi}_{1T}^{\theta\theta} - \Delta\hat{\Psi}_{2T}^{\theta\theta} - \Delta\hat{\Psi}_{3T}^{\theta\theta}$, where $\Delta\hat{\Psi}_{jT}^{\theta\theta}$ are equal to $\Delta\hat{\Psi}_{jT}$, for $j = 1, 2, 3$, where $\hat{F}(Y_t|W_t) - F^0(Y_t|W_t)$, $\hat{F}(q_t(\theta)|W_t) - F^0(q_t(\theta)|W_t)$, $F^0(Y_t|W_t) - F^0(q_t(\theta)|W_t)$ are replaced by $\{D^2\hat{F}[q_t(\theta)|W_t] - D^2F^0[q_t(\theta)|W_t]\}\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'$, $\{D\hat{F}[q_t(\theta)|W_t] - DF^0[q_t(\theta)|W_t]\}\Delta_{\theta\theta}q_t(\theta)$, and $D^2F^0[q_t(\theta)|W_t]\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)' + DF^0[q_t(\theta)|W_t]\Delta_{\theta\theta}q_t(\theta)$, respectively. We then obtain

$$\begin{aligned} |\Delta\hat{\Psi}_{1T}^{\theta\theta J}| \leqslant b_T^{-1}\Bigg[&\sup_{(y,w)\in\mathbb{R}^{m+1}}|D^2\hat{G}(y,w) - H_{2T}^0(y,w)| \\ &+ \sup_{(y,w)\in\mathbb{R}^{m+1}}|D^2F^0(y|w)|\sup_{w\in\mathbb{R}^m}|\hat{g}(w) - \bar{g}_T^0(w)|\Bigg]\left[\frac{1}{T}\sum_{t=1}^T J_t\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|\right]. \end{aligned}$$

Thus, $\sup_{\theta\in\Theta}\Delta\hat{\Psi}_{1T}^{\theta\theta J} = o_p(1)$ as Cauchy-Schwarz inequality gives

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^T J_t\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'| &\leqslant \left(\frac{1}{T}\sum_{t=1}^T J_t\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^T\left(\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|\right)^2\right)^{1/2} \\ &= o_p(1), \end{aligned} \tag{60}$$

by Equation (51) and $(1/T)\sum_{t=1}^T(\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|)^2 = O_p(1)$, which follows from

$$E\left[\frac{1}{T}\sum_{t=1}^T\left(\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|\right)^2\right] \leqslant \sup_{1\leqslant t\leqslant T, T\geqslant 1}E\left(\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|\right)^2 < \infty,$$

using (A7)(i) and Markov inequality. Similarly, $\sup_{\theta\in\Theta}\Delta\hat{\Psi}_{2T}^{\theta\theta J}=o_p(1)$ using

$$\frac{1}{T}\sum_{t=1}^{T}J_t\sup_{\theta\in\Theta}|\Delta_{\theta\theta}q_t(\theta)|=o_p(1). \tag{61}$$

Regarding $\Delta\hat{\Psi}_{3T}^{\theta\theta J}$, we have

$$\begin{aligned}|\Delta\hat{\Psi}_{3T}^{\theta\theta J}| \;\leqslant\; & \sup_{(y,w)\in\mathbb{R}^{m+1}}|D^2F^0(y|w)|\frac{1}{T}\sum_{t=1}^{T}J_t\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'| \\ & + \sup_{(y,w)\in\mathbb{R}^{m+1}}|DF^0(y|w)|\frac{1}{T}\sum_{t=1}^{T}J_t\sup_{\theta\in\Theta}|\Delta_{\theta\theta}q_t(\theta)|,\end{aligned}$$

showing that $\sup_{\theta\in\Theta}\Delta\hat{\Psi}_{3T}^{\theta\theta J}=o_p(1)$ using Equations (60) − (61) and (A9)(ii).

In Step 2, we deal with $\Delta_{\theta\theta}\hat{\Psi}_T^{\epsilon}(\theta)-\Delta_{\theta\theta}\Psi_T^{*}(\theta)=-\Delta\Psi_{1T}^{\epsilon\theta\theta}-\Delta\Psi_{2T}^{\epsilon\theta\theta}+\Delta\Psi_{3T}^{\epsilon\theta\theta}$, where $\Delta\Psi_{jT}^{\epsilon\theta\theta}$ are equal to $\Delta\Psi_{jT}^{\epsilon}$, for $j=1,2,3$, where $\hat{F}(Y_t|W_t)-F^0(Y_t|W_t)$, $\hat{F}(q_t(\theta)|W_t)-F^0(q_t(\theta)|W_t)$, $F^0(Y_t|W_t)-F^0(q_t(\theta)|W_t)$ are replaced by $\{D^2\hat{F}[q_t(\theta)|W_t]-D^2F^0[q_t(\theta)|W_t]\}\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'$, $\{D\hat{F}[q_t(\theta)|W_t]-DF^0[q_t(\theta)|W_t]\}\Delta_{\theta\theta}q_t(\theta)$, and $D^2F^0[q_t(\theta)|W_t]\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'+DF^0[q_t(\theta)|W_t]$ $\Delta_{\theta\theta}q_t(\theta)$, respectively. We then obtain

$$\begin{aligned}|\Delta\Psi_{1T}^{\epsilon\theta\theta}| \leqslant (b_T^{\epsilon})^{-1}(1-\eta)^{-1}\Bigg[&\sup_{(y,w)\in\mathbb{R}^{m+1}}|D^2\hat{G}(y,w)-H_{2T}^{0}(y,w)| \\ &+\sup_{(y,w)\in\mathbb{R}^{m+1}}|D^2F^0(y|w)|\sup_{w\in\mathbb{R}^m}|\hat{g}(w)-\bar{g}_T^0(w)|\Bigg]\left[\frac{1}{T}\sum_{t=1}^{T}d_t^{\epsilon}\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|\right],\end{aligned}$$

with probability approaching 1, where $(1/T)\sum_{t=1}^{T}d_t^{\epsilon}\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|\leqslant(1/T)\sum_{t=1}^{T}$ $\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|=O_p(1)$ by Markov inequality and (A7)(i). Hence, $\sup_{\theta\in\Theta}\Delta\Psi_{1T}^{\epsilon\theta\theta}=$

$o_p(1)$. Similarly, $\sup_{\theta\in\Theta}\Delta\Psi_{2T}^{\epsilon\theta\theta}=o_p(1)$. Regarding $\Delta\Psi_{3T}^{\epsilon\theta\theta}$, we have

$$
\begin{aligned}
&|\Delta\Psi_{3T}^{\epsilon\theta\theta}| \\
&\leqslant \left[\sup_{(y,w)\in\mathbb{R}^{m+1}}|D^2F^0(y|w)|\right]\left[\frac{1}{T}\sum_{t=1}^{T}(1-d_T^{\epsilon})\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|\right] \\
&+\left[\sup_{(y,w)\in\mathbb{R}^{m+1}}|DF^0(y|w)|\right]\left[\frac{1}{T}\sum_{t=1}^{T}(1-d_T^{\epsilon})\sup_{\theta\in\Theta}|\Delta_{\theta\theta}q_t(\theta)|\right] \\
&\leqslant \left[\sup_{(y,w)\in\mathbb{R}^{m+1}}|D^2F^0(y|w)|\right]\left[\frac{1}{T}\sum_{t=1}^{T}(1-d_T^{\epsilon})^2\right]^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\sup_{\theta\in\Theta}|\nabla_\theta q_t(\theta)\nabla_\theta q_t(\theta)'|\right)^2\right]^{1/2} \\
&+\left[\sup_{(y,w)\in\mathbb{R}^{m+1}}|DF^0(y|w)|\right]\left[\frac{1}{T}\sum_{t=1}^{T}(1-d_T^{\epsilon})^2\right]^{1/2}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\sup_{\theta\in\Theta}|\Delta_{\theta\theta}q_t(\theta)|\right)^2\right]^{1/2}
\end{aligned}
$$

which is an $o_p(1)$ as $(1/T)\sum_{t=1}^{T}(1-d_T^{\epsilon})^2=(1/T)\sum_{t=1}^{T}(1-d_T^{\epsilon})=o_p(1)$ from Step 1 of Theorem 5. □

Proof of Lemma 10. Note that for any density $f_t(\cdot)$, we have $E\Big(f_t[q_t(\theta_0)]\nabla_\theta q_t(\theta_0)\{\mathbb{1}[q_t(\theta_0)-Y_t]-\alpha\}\Big)=0$. Thus, Lemma 10 could be established from (i) the stochastic equicontinuity at $f^0(\cdot|\cdot)$ of the vector process $\nu_T(f)=(1/\sqrt{T})\sum_{t=1}^{T}f[q_t(\theta_0)|W_t]\nabla_\theta q_t(\theta_0)\{\mathbb{1}[q_t(\theta_0)-Y_t]-\alpha\}$ with respect to some pseudo-metric $\rho(f_1,f_2)$, and (ii) the consistency of $\hat{f}(\cdot|\cdot)=\mathbb{1}[\hat{g}(\cdot)-b_T]D\hat{G}(\cdot,\cdot)/\hat{g}(\cdot)$ to $f^0(\cdot|\cdot)$ with respect to $\rho(\cdot,\cdot)$. See Andrews (1994b) for some general results on stochastic equicontinuity. These require, however, a more elaborate trimming than the one used here in view of Andrews (1995, p.571). We thus prove Lemma 10 directly.

Though more complex, our proof draws from the asymptotic normality proof of Theorem 1 in Lavergne and Vuong (1996). For similar asymptotic normality proofs in the iid case see also Robinson (1988) and Hardle and Stoker (1989). As previously, we let $q_t^0=q_\alpha(W_t,\theta_0)$. Moreover, let ϵ_T be such that $\epsilon_T=o(b_T)$, $\epsilon_T T^{1/4}h_{yT}h_{wT}^m\to\infty$, $\epsilon_T/(T^{1/4}h_{wT}^R)\to\infty$ and $\epsilon_T/(T^{1/4}h_{yT}^R)\to\infty$. As

$$
\nabla_\theta\hat{\Psi}_T(\theta_0)-\nabla_\theta\Psi_T^*(\theta_0)=\left[\nabla_\theta\hat{\Psi}_T(\theta_0)-\nabla_\theta\Psi_T^{\epsilon}(\theta_0)\right]+\left[\nabla_\theta\Psi_T^{\epsilon}(\theta_0)-\nabla_\theta\Psi_T^*(\theta_0)\right],
$$

where $\Psi_T^{\epsilon}(\theta)$ is defined in Equation (47), it suffices to prove that both terms on the right-hand side of the above equality are $o_p(T^{-1/2})$. Given Lemma 8 and Equation (46) we shall

use

$$a_T^{-2} \sup_{(y,w)\in\mathbb{R}^{m+1}} |D\hat{G}(y,w) - H_{1T}^0(y,w)|^2 = o_p(T^{-1/2}), \tag{62}$$

$$a_T^{-2} \sup_{(y,w)\in\mathbb{R}^{m+1}} |D\hat{G}(y,w) - H_{1T}^0(y,w)| \sup_{w\in\mathbb{R}^m} |\hat{g}(w) - \bar{g}_T^0(w)| = o_p(T^{-1/2}), \tag{63}$$

$$a_T^{-2} \sup_{w\in\mathbb{R}^m} |\hat{g}(w) - \bar{g}_T^0(w)|^2 = o_p(T^{-1/2}), \tag{64}$$

which hold for any sequence $\{a_T\}$ satisfying $a_T T^{1/4} h_{yT} h_{wT}^m \to \infty$, $a_T/(T^{1/4}h_{wT}^R) \to \infty$ and $a_T/(T^{1/4}h_{yT}^R) \to \infty$. We shall also use the identities

$$\begin{aligned} \hat{f}(y|w) - f^0(y|w) &= \frac{1}{\hat{g}(w)}[D\hat{G}(y,w) - H_{1T}^0(y,w)] - \frac{f^0(y|w)}{\hat{g}(w)}[\hat{g}(w) - \bar{g}_T^0(w)], \quad (65)\\ &= \frac{D\hat{G}(y,w) - f^0(y|w)\hat{g}(w)}{\bar{g}_T^0(w)} + \frac{f^0(y|w)}{\hat{g}(w)\bar{g}_T^0(w)}[\hat{g}(w) - \bar{g}_T^0(w)]^2 \\ &\quad -\frac{1}{\hat{g}(w)\bar{g}_T^0(w)}[D\hat{G}(y,w) - H_{1T}^0(y,w)][\hat{g}(w) - \bar{g}_T^0(w)]. \quad (66) \end{aligned}$$

STEP1: We first show that $\nabla_\theta \hat{\Psi}_T(\theta_0) - \nabla_\theta \Psi_T^\epsilon(\theta_0) = o_p(T^{-1/2})$. We have

$$\begin{aligned} \sqrt{T}\left[\nabla_\theta \hat{\Psi}_T(\theta_0) - \nabla_\theta \Psi_T^\epsilon(\theta_0)\right] &= \frac{1}{\sqrt{T}}\sum_{t=1}^T (J_t - H_t)[\mathbb{1}(q_t^0 - Y_t) - \alpha]\hat{f}(q_t^0|W_t)\nabla_\theta q_t^0 \\ &= \sqrt{T}\Delta\hat{\Psi}_{1T}^\theta(\theta_0) + \sqrt{T}\Delta\hat{\Psi}_{2T}^\theta(\theta_0), \end{aligned}$$

where $J_t = d_t(1-d_t^\epsilon)$, $H_t = (1-d_t)d_t^\epsilon$ and

$$\begin{aligned} \sqrt{T}\Delta\hat{\Psi}_{1T}^\theta(\theta_0) &= \frac{1}{\sqrt{T}}\sum_{t=1}^T (J_t - H_t)[\mathbb{1}(q_t^0 - Y_t) - \alpha][\hat{f}(q_t^0|W_t) - f^0(q_t^0|W_t)]\nabla_\theta q_t^0, \\ \sqrt{T}\Delta\hat{\Psi}_{2T}^\theta(\theta_0) &= \frac{1}{\sqrt{T}}\sum_{t=1}^T (J_t - H_t)[\mathbb{1}(q_t^0 - Y_t) - \alpha]f^0(q_t^0|W_t)\nabla_\theta q_t^0. \end{aligned}$$

As $H_t \leqslant \mathbb{1}[|\hat{g}(W_t) - \bar{g}_T^0(W_t)| - \epsilon_T]$ and the event $\{\sup_w |\hat{g}(w) - \bar{g}_T^0(w)| > \epsilon_T\}$ has asymptotic probability 0 because Property (64) holds with $a_T = \epsilon_T$ by construction of ϵ_T, we have $\sup_{1\leqslant t\leqslant T, T\geqslant 1} H_t = 0$ with probability approaching one. Hence, we need to consider the J_t terms only. Namely, it suffices to show that $\Delta\hat{\Psi}_{jT}^{\theta J}(\theta_0) = o_p(T^{-1/2})$ for $j = 1, 2$. Using

Equality (65) and the definition of J_t, we obtain

$$\begin{aligned}
|\Delta\hat{\Psi}_{1T}^{\theta J}(\theta_0)| &\leqslant b_T^{-1}\left[\sup_{(y,w)\in\mathbb{R}^{m+1}}|D\hat{G}(y,w)-H_{1T}^0(y,w)|\right.\\
&\quad\left.+\sup_{(y,w)\in\mathbb{R}^{m+1}}f^0(y|w)\sup_{w\in\mathbb{R}^m}|\hat{g}(w)-\bar{g}_T^0(w)|\right]\left[\frac{1}{T}\sum_{t=1}^T J_t|\nabla_\theta q_t^0|\right],\\
|\Delta\hat{\Psi}_{2T}^{\theta J}(\theta_0)| &\leqslant \frac{1}{T}\sum_{t=1}^T J_t f^0(q_t^0|W_t)|\nabla_\theta q_t^0|.
\end{aligned}$$

But $(1/T)\sum_{t=1}^T J_t|\nabla_\theta q_t^0| \leqslant (1/T)\sum_{t=1}^T(1-d_t^\epsilon)|\nabla_\theta q_t^0| = O_p(b_T^\gamma)$ and $(1/T)\sum_{t=1}^T J_t f^0(q_t^0|W_t)$ $|\nabla_\theta q_t^0| \leqslant (1/T)\sum_{t=1}^T(1-d_t^\epsilon)f^0(q_t^0|W_t)|\nabla_\theta q_t^0| = O_p(b_T^{2\gamma})$ by Markov inequality combined with (A10)(ii)-(iii) where $c_T = b_T^\epsilon = O(b_T)$. Hence, using $\sup_{(y,w)\in\mathbb{R}^{m+1}} f^0(y|w) < \infty$ by (A9)(ii) combined with Properties (62) and (64) where $a_T = b_T$, we obtain $\Delta\hat{\Psi}_{1T}^{\theta J}(\theta_0) = O_p(T^{-1/4}b_T^\gamma)$ and $\Delta\hat{\Psi}_{2T}^{\theta J}(\theta_0) = O_p(b_T^{2\gamma})$. Since $b_T = o(T^{-1/(4\gamma)})$ we obtain $\Delta\hat{\Psi}_{jT}^{\theta J}(\theta_0) = o_p(T^{-1/2})$ for $j = 1, 2$, as desired.

STEP 2: We next show that $\nabla_\theta\Psi_T^\epsilon(\theta_0) - \nabla_\theta\Psi_T^*(\theta_0) = o_p(T^{-1/2})$. We have $\nabla_\theta\Psi_T^\epsilon(\theta_0) = \mu_0 + [\nabla_\theta\Psi_T^\epsilon(\theta_0) - \mu_0]$, where $\mu_0 \equiv T^{-1}\sum_{t=1}^T d_t^\epsilon[\mathbb{1}(q_t^0 - Y_t) - \alpha]f^0(q_t^0|W_t)\nabla_\theta q_t^0$ and

$$\begin{aligned}
&\nabla_\theta\Psi_T^\epsilon(\theta_0) - \mu_0\\
&= \frac{1}{T}\sum_{t=1}^T d_t^\epsilon[\mathbb{1}(q_t^0 - Y_t) - \alpha][\hat{f}(q_t^0|W_t) - f^0(q_t^0|W_t)]\nabla_\theta q_t^0\\
&= \frac{1}{T}\sum_{t=1}^T d_t^\epsilon[\mathbb{1}(q_t^0 - Y_t) - \alpha]\frac{D\hat{G}(q_t^0, W_t) - f^0(q_t^0|W_t)\hat{g}(W_t)}{\bar{g}_T^0(W_t)}\nabla_\theta q_t^0\\
&+ \frac{1}{T}\sum_{t=1}^T d_t^\epsilon[\mathbb{1}(q_t^0 - Y_t) - \alpha]\frac{f^0(q_t^0|W_t)}{\hat{g}(W_t)\bar{g}_T^0(W_t)}[\hat{g}(W_t) - \bar{g}_T^0(W_t)]^2\nabla_\theta q_t^0\\
&- \frac{1}{T}\sum_{t=1}^T d_t^\epsilon[\mathbb{1}(q_t^0 - Y_t) - \alpha]\frac{D\hat{G}(q_t^0, W_t) - H_{1T}^0(q_t^0, W_t)}{\hat{g}(W_t)\bar{g}_T^0(W_t)}[\hat{g}(W_t) - \bar{g}_T^0(W_t)]\nabla_\theta q_t^0\\
&\equiv \mu_1 + \mu_2 - \mu_3,
\end{aligned}$$

using Equality (66). Hence, $\nabla_\theta\Psi_T^\epsilon(\theta_0) = \mu_0 + \mu_1 + \mu_2 - \mu_3$. Thus, the proof is complete if $\mu_0 = \nabla_\theta\Psi_T^*(\theta_0) + o_p(T^{-1/2})$ and $\mu_j = o_p(T^{-1/2})$ for $j = 1, 2, 3$, as shown next.

STEP 2a: We show that $\mu_0 = \nabla_\theta\Psi_T^*(\theta_0) + o_p(T^{-1/2})$. We have

$$\mu_0 = \nabla_\theta\Psi_T^*(\theta_0) - \frac{1}{T}\sum_{t=1}^T(1-d_t^\epsilon)[\mathbb{1}(q_t^0 - Y_t) - \alpha]f^0(q_t^0|W_t)\nabla_\theta q_t^0.$$

Let μ_{02} denote the second term on the right-hand side of the above equality. Thus, it suffices to show that $\mu_{02} = o_p(T^{-1/2})$. But, from Step 1 we know that $|\mu_{02}| \leqslant T^{-1}\sum_{t=1}^{T}(1 - d_t^\epsilon) f^0(q_t^0|W_t)|\nabla_\theta q_t^0| = O_p(b_T^{2\gamma})$. The desired result follows from $b_T = o(T^{-1/(4\gamma)})$.

STEP 2b: Next, we show that $\mu_2 = \mu_3 = o_p(T^{-1/2})$. We have

$$|\sqrt{T}\mu_2| \leqslant \frac{\sqrt{T}}{b_T^\epsilon \inf_{\{w:\bar{g}_T^0(w)>b_T^\epsilon\}}|\hat{g}(w)|} \sup_{(y,w)\in\mathbb{R}^{m+1}} f^0(y|w) \sup_{w\in\mathbb{R}^m}[\hat{g}(w) - \bar{g}_T^0(w)]^2 \frac{1}{T}\sum_{t=1}^{T}|\nabla_\theta q_t^0|.$$

But Property (64) with $a_T = b_T^\epsilon$ implies $(b_T^\epsilon)^{-1}\sup_w|\hat{g}(w) - \bar{g}_T^0(w)| = o_p(T^{-1/4}) = o_p(1)$. Hence, for any $\eta \in (0,1)$ we have $\inf_{\{w:\bar{g}_T^0(w)/b_T^\epsilon>1\}}|\hat{g}_T^0(w)|/b_T^\epsilon > 1-\eta$ with probability approaching one. Thus, with probability approaching one

$$|\sqrt{T}\mu_2| \leqslant \frac{\sqrt{T}}{(b_T^\epsilon)^2(1-\eta)} \sup_{(y,w)\in\mathbb{R}^{m+1}} f^0(y|w) \sup_{w\in\mathbb{R}^m}[\hat{g}(w) - \bar{g}_T^0(w)]^2 \frac{1}{T}\sum_{t=1}^{T}|\nabla_\theta q_t^0|,$$

which is an $o_p(1)$ by Property (64) with $a_T = b_T^\epsilon$ as $\sup_{(y,w)\in\mathbb{R}^{m+1}} f^0(y|w) < \infty$ and $(1/T)\sum_{t=1}^{T}|\nabla_\theta q_t^0| = O_p(1)$ as noted earlier. That is, $\mu_2 = o_p(T^{-1/2})$. Similarly,

$$\begin{aligned}|\sqrt{T}\mu_3| \leqslant{}& \frac{\sqrt{T}}{b_T^\epsilon \inf_{\{w:\bar{g}_T^0(w)>b_T^\epsilon\}}|\hat{g}(w)|} \\ &\times \sup_{(y,w)\in\mathbb{R}^{m+1}}|D\hat{G}(y,w) - H_{1T}^0(y,w)| \sup_{w\in\mathbb{R}^m}|\hat{g}(w) - \bar{g}_T^0(w)| \frac{1}{T}\sum_{t=1}^{T}|\nabla_\theta q_t^0|,\end{aligned}$$

which shows that $\mu_3 = o_p(T^{-1/2})$ using the same argument with Property (63).

STEP 2c: Lastly, we show that $\mu_1 = o_p(T^{-1/2})$. Let $K_{0T}(\cdot) \equiv (1/h_{yT})K_0(\cdot/h_{yT})$ and $K_T(\cdot) \equiv (1/h_{wT}^m)K_0(\cdot/h_{wT})$. Thus, from the definitions of $D\hat{G}(y,w)$ and $\hat{g}(w)$ we have

$$\begin{aligned}\mu_1 &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} d_t^\epsilon \frac{\mathbb{1}(q_t^0 - Y_t) - \alpha}{\bar{g}_T^0(W_t)}\left[K_{0T}(q_t^0 - Y_s) - f^0(q_t^0|W_t)\right] K_T(W_t - W_s)\nabla_\theta q_t^0 \\ &\equiv L + \frac{T-1}{T}U,\end{aligned}$$

where L and U are the diagonal and U-statistic parts defined as

$$\begin{aligned}
L &\equiv \frac{1}{T^2}\sum_{t=1}^{T} d_t^{\epsilon}\frac{\mathbb{I}(q_t^0 - Y_t) - \alpha}{\bar{g}_T^0(W_t)}\left[K_{0T}(q_t^0 - Y_t) - f^0(q_t^0|W_t)\right] K_T(0)\nabla_\theta q_t^0 \\
U &\equiv \frac{1}{T(T-1)}\sum_{1\leqslant t\neq s\leqslant T} u_{Tts} \\
u_{Tts} &= \frac{1}{2}\left(u_{Tts}^0 + u_{Tst}^0\right) \equiv h_T(Y_t, W_t, Y_s, W_s) \\
u_{Tts}^0 &= \left[K_{0T}(q_t^0 - Y_s) - f^0(q_t^0|W_t)\right] K_T(W_t - W_s) d_t^{\epsilon}\frac{\mathbb{I}(q_t^0 - Y_t) - \alpha}{\bar{g}_T^0(W_t)}\nabla_\theta q_t^0 \\
u_{Tst}^0 &= \left[K_{0T}(q_s^0 - Y_t) - f^0(q_s^0|W_s)\right] K_T(W_s - W_t) d_s^{\epsilon}\frac{\mathbb{I}(q_s^0 - Y_s) - \alpha}{\bar{g}_T^0(W_s)}\nabla_\theta q_s^0,
\end{aligned}$$

for $1 \leqslant t \neq s \leqslant T$. Note that $h_T(Y_t, W_t, Y_s, W_s)$ is symmetric in (Y_t, W_t) and (Y_s, W_s). Hence, it suffices to show that L and U are both $o_p(T^{-1/2})$.

For L we have

$$|\sqrt{T}L| \leqslant \frac{1}{\sqrt{T}b_T^{\epsilon}h_{wT}^m}\left[\frac{1}{h_{yT}}\sup_{y\in\mathbb{R}}|K_0(y)| + \sup_{(y,w)\in\mathbb{R}^{m+1}} f^0(y|w)\right]|K(0)|\frac{1}{T}\sum_{t=1}^{T}|\nabla_\theta q_t^0|$$

where $\sup_{(y,w)\in\mathbb{R}^{m+1}} f^0(y|w) < \infty$, $\sup_{y\in\mathbb{R}}|K_0(y)| < \infty$ and $|K(0)| < \infty$ by (A11)(iii) and (A9)(ii). As $(1/T)\sum_{t=1}^{T}|\nabla_\theta q_t^0| = O_p(1)$ by (A7)(i), (A5) and Markov inequality, we obtain $\sqrt{T}L = o_p(1)$ because $\sqrt{T}b_T^{\epsilon}h_{yT}h_{wT}^m = b_T^{\epsilon}T^{1/4}h_{yT}h_{wT}^m T^{1/4} \to \infty$ using $b_T^{\epsilon} = b_T(1 + o(1))$.

It remains to be shown that $U = o_p(T^{-1/2})$. Because of the stationarity assumption (A6')(i), we have $\bar{g}_T^0(\cdot) = g_t^0(\cdot) \equiv g^0(\cdot)$. Moreover, from the Hoeffding decomposition (see e.g. Arcones (1995, eq. 1.7)), we have $U = U_0 + 2U_1 + U_2$ where

$$U_0 = \int\int\int\int h_T(y_1, w_1, y_2, w_2)\prod_{t=1}^{2}[f^0(y_t|w_t)g^0(w_t)dy_t dw_t] \tag{67}$$

$$U_1 = \frac{1}{T}\sum_{t=1}^{T} h_{T1}(Y_t, W_t) \tag{68}$$

$$U_2 = \frac{1}{T(T-1)}\sum_{1\leqslant t\neq s\leqslant T} h_{T2}(Y_t, W_t, Y_s, W_s) \tag{69}$$

$$h_{T1}(y_1, w_1) = \int\int h_T(y_1, w_1, y_2, w_2) f^0(y_2|w_2)g^0(w_2)dy_2 dw_2 - U_0 \tag{70}$$

$$h_{T2}(y_1, w_1, y_2, w_2) = h_T(y_1, w_1, y_2, w_2) - h_{T1}(y_1, w_1) - h_{T1}(y_2, w_2) - U_0. \tag{71}$$

Note that $U_0 \neq E[U]$ as $\prod_{t=1}^{2}[f^0(y_t|w_t)g^0(w_t)]$ is not the joint density of (Y_1, W_1, Y_2, W_2), while $h_{T1}(\cdot)$ and $h_{T2}(\cdot)$ are canonical kernels, i.e. symmetric kernels satisfying $E[h_{T1}(Y_1, W_1)]$

$= 0$ and $E[h_{T2}(y_1, w_1, Y_2, W_2)] = 0$, respectively, as noted by Arcones (1995). Thus, it suffices to show that $\sqrt{T}U_k = o_p(1)$ for $k = 0, 1, 2$.

STEP 2c(i): We start by showing that$\sqrt{T}U_0 = o_p(1)$. In fact, we have $U_0 = 0$ as Equation (67) gives

$$\begin{aligned} U_0 &= \int\int\int\int \frac{1}{2}(u^0_{T12} + u^0_{T21}) \prod_{t=1}^{2} [f^0(y_t|w_t)g^0(w_t)dy_t dw_t] \\ &= \frac{1}{2}\int\int \left\{ \int \left[K_{0T}(q^0_1 - y_2) - f^0(q^0_1|w_1)\right] f^0(y_2|w_2)dy_2 \right\} \\ &\quad \times \left\{ \int \left[\mathbb{1}(q^0_1 - y_1) - \alpha\right] f^0(y_1|w_1)dy_1 \right\} K_T(w_1 - w_2) \frac{d^\epsilon_1 \nabla_\theta q^0_1}{g^0(w_1)} g^0(w_1)g^0(w_2)dw_1 dw_2 \\ &\quad + \frac{1}{2}\int\int \left\{ \int \left[K_{0T}(q^0_2 - y_1) - f^0(q^0_2|w_2)\right] f^0(y_1|w_1)dy_1 \right\} \\ &\quad \times \left\{ \int \left[\mathbb{1}(q^0_2 - y_2) - \alpha\right] f^0(y_2|w_2)dy_2 \right\} K_T(w_2 - w_1) \frac{d^\epsilon_2 \nabla_\theta q^0_2}{g^0(w_2)} g^0(w_1)g^0(w_2)dw_1 dw_2, \end{aligned}$$

where $\int [\mathbb{1}(q^0_t - y_t) - \alpha] f^0(y_t|w_t)dy_t = 0$ for any t by assumptions (A1) and (A9)(i).

STEP 2c(ii): We now show that $\sqrt{T}U_1 = o_p(1)$. By Markov inequality it suffices to show that $E(TU_1^2) = o(1)$. But assumption (A6') and Lemma 3 in Arcones (1995) with $p = r$ imply

$$E(TU_1^2) = T^{-1}E\left[\Big(\sum_{1\leqslant t\leqslant T} h_{T1}(Y_t, W_t)\Big)^2\right] \leqslant c\Big(T^{-1} + T^{-1}\sum_{t=1}^{T-1} t\beta_t^{(r-2)/r}\Big)M^2_{T1}, \tag{72}$$

where β_t are the mixing coefficients of $\{(Y_t, W_t')'\}$, c is a universal constant and $M_{T1} = \sup_{1\leqslant t<\infty} [E|h_{T1}(Y_t, W_t)|^r]^{1/r}$. Note that Lemma 3 in Arcones (1995) is written for canonical kernels that are independent of T. It is, however, easy to see from his proofs that this lemma and Lemma 8, which is used to prove it, both hold even when canonical kernels depend on T as in $h_{T1}(\cdot)$ and $h_{T2}(\cdot)$. From (A6')(ii) we know that $\sum_{t=1}^{\infty} \beta_t^{(r-2)/r} < \infty$ (see e.g. White 2001 for the definition of the size) hence $T^{-1} + T^{-1}\sum_{t=1}^{T-1} t\beta_t^{(r-2)/r} = O(1)$. We now show that $M_{T1} \to 0$. As $U_0 = 0$ and the integral of u^0_{T21} with respect to $f^0(y_2|w_2)g^0(w_2)dy_2dw_2$ is

zero because $\int [\mathbb{1}(q_2^0 - y_2) - \alpha] f^0(y_2|w_2)dy_2 = 0$, we have from Equation (70)

$$
\begin{aligned}
&|h_{T1}(y_1, w_1)| \\
&= \left| \frac{1}{2} \left[\mathbb{1}(q_1^0 - y_1) - \alpha \right] \frac{d_1^\epsilon \nabla_\theta q_1^0}{g^0(w_1)} \right. \\
&\left. \times \int \left\{ \int \left[K_{0T}(q_1^0 - y_2) - f^0(q_1^0|w_1) \right] f^0(y_2|w_2)dy_2 \right\} K_T(w_1 - w_2) g^0(w_2)dw_2 \right| \\
&\leqslant \frac{|\nabla_\theta q_1^0|}{2b_T^\epsilon} \left| \int \left\{ \int K_0(u) \left[f^0(q_1^0 - uh_{yT}|w_2) - f^0(q_1^0|w_1) \right] du \right\} K_T(w_1 - w_2) g^0(w_2)dw_2 \right| \\
&\leqslant \frac{|\nabla_\theta q_1^0|}{2b_T^\epsilon} \left| \int \left\{ \int K_0(u) \left[f^0(q_1^0 - uh_{yT}|w_2) - f^0(q_1^0|w_2) \right] du \right\} K_T(w_1 - w_2) g^0(w_2)dw_2 \right| \\
&+ \frac{|\nabla_\theta q_1^0|}{2b_T^\epsilon} \left| \int \left[f^0(q_1^0|w_2) - f^0(q_1^0|w_1) \right] K_T(w_1 - w_2) g^0(w_2)dw_2 \right| \\
&\leqslant \frac{|\nabla_\theta q_1^0|}{2b_T^\epsilon} \int \left| \int K_0(u) \left[f^0(q_1^0 - uh_{yT}|w_2) - f^0(q_1^0|w_2) \right] du \right| |K_T(w_1 - w_2)| g^0(w_2)dw_2 \\
&+ \frac{|\nabla_\theta q_1^0|}{2b_T^\epsilon} \left| \int \left[f^0(q_1^0|w_2)g^0(w_2) - f^0(q_1^0|w_1)g^0(w_1) \right] K_T(w_1 - w_2)dw_2 \right| \\
&+ \frac{|\nabla_\theta q_1^0|}{2b_T^\epsilon} f^0(q_1^0|w_1) \left| \int \left[g^0(w_2) - g^0(w_1) \right] K_T(w_1 - w_2)dw_2 \right| \\
&\leqslant \frac{|\nabla_\theta q_1^0|}{b_T^\epsilon} O(h_{yT}^R) \int |K(v)| g^0(w_1 - vh_{wT})dv \\
&+ \frac{|\nabla_\theta q_1^0|}{2b_T^\epsilon} \left| \int \left[f^0(q_1^0|w_1 - vh_{wT})g^0(w_1 - vh_{wT}) - f^0(q_1^0|w_1)g^0(w_1) \right] K(v)dv \right| \\
&+ \frac{|\nabla_\theta q_1^0|}{2b_T^\epsilon} f^0(q_1^0|w_1) \left| \int \left[g^0(w_1 - vh_{wT}) - g^0(w_1) \right] K(v)dv \right| \\
&\leqslant \frac{|\nabla_\theta q_1^0|}{b_T^\epsilon} \left\{ O(h_{yT}^R) + [1 + f^0(q_1^0|w_1)] O(h_{wT}^R) \right\},
\end{aligned}
$$

so

$$
|h_{T1}(y_1, w_1)| \leqslant \frac{|\nabla_\theta q_1^0|}{b_T^\epsilon} \left\{ O(h_{yT}^R) + O(h_{wT}^R) \right\}, \tag{73}
$$

where we have used the change of variables $u = (q_1^0 - y_2)/h_{yT}$ and $v = (w_1 - w_2)/h_{wT}$ combined with (A8), (A9)(ii), (A11)(i,iii) and Taylor expansions of order R of the integrands. As $E|\nabla_\theta q_1^0|^r < \sup_{1 \leqslant t \leqslant T, T \geqslant 1} E[\sup_{\theta \in \Theta} |\nabla_\theta q_\alpha(W_t, \theta)|^r] < \infty$ by (A5) and (A7)(i), it follows that $(E|h_{T1}(Y_t, W_t)|^r)^{1/r} \leqslant (1/b_T^\epsilon) \left\{ O(h_{yT}^R) + O(h_{wT}^R) \right\}$ uniformly in t. Hence, $M_{T1} \leqslant (1/b_T^\epsilon) \left\{ O(h_{yT}^R) + O(h_{wT}^R) \right\}$. Given $b_T^\epsilon = b_T[1 + o(1)]$, $h_{yT}^R = o(b_T)$ and $h_{wT}^R = o(b_T)$,

which follow from $b_T/(T^{1/4}h_{yT}^R) \to \infty$ and $b_T/(T^{1/4}h_{wT}^R) \to \infty$ respectively, we have that $M_{T1} = o(1)$. Combining Property (72) and (A6')(i) then gives $E(TU_1^2) = o(1)$ as desired.

STEP 2c(iii): Finally we show that $\sqrt{T}U_2 = o_p(1)$. Again, by Markov inequality it suffices to show that $E(TU_2^2) = o(1)$. Similar to the previous case, Assumption (A6') and Lemma 3 in Arcones (1995) with $p = 2r$ imply

$$\begin{aligned} E(TU_2^2) &= \left(\frac{T}{T-1}\right)^2 T^{-3} E\left[\Big(\sum_{1\leqslant t\neq s\leqslant T} h_{T2}(Y_t, W_t, Y_s, W_s)\Big)^2\right] \\ &\leqslant \left(\frac{T}{T-1}\right)^2 c\Big(T^{-1} + T^{-1}\sum_{t=1}^{T-1} t\beta_t^{(r-1)/r}\Big) M_{T2}^2, \end{aligned} \tag{74}$$

where c is a universal constant and $M_{T2} = \sup_{1\leqslant t\neq s<\infty} [E|h_{T2}(Y_t, W_t, Y_s, W_s)|^{2r}]^{1/(2r)}$. We now show that $T^{-1} + T^{-1}\sum_{t=1}^{T-1} t\beta_t^{(r-1)/r} = O(1/\sqrt{T})$ and that $M_{T2} = o(T^{1/4})$. The first property is implied by $\sum_{t=1}^{\infty} t\beta_t^{(r-1)/r} < \infty$ for which it suffices to show that $\sum_{t=\tau}^{2\tau} t\beta_t^{(r-1)/r} \to 0$ as $\tau \to \infty$. As previously, from (A6')(ii) we know that $\sum_{t=1}^{\infty} \beta_t^{(r-2)/r} < \infty$ hence $\beta_t^{(r-2)/r} t \to 0$ as $t \to \infty$ and $\beta_t \leqslant t^{r/(2-r)}$ for t large enough. Thus $\sum_{t=\tau}^{2\tau} t\beta_t^{(r-1)/r} \leqslant \sum_{t=\tau}^{2\tau} t^{-1/(r-2)}$ which vanishes when $2 < r < 3$ as assumed. For the second property, we bound M_{T2}. From Equations (71), (73) and $U_0 = 0$ we obtain

$$\begin{aligned} |h_{T2}(y_1, w_1, y_2, w_2)| &\leqslant |h_T(y_1, w_1, y_2, w_2)| + \frac{|\nabla_\theta q_1^0| + |\nabla_\theta q_2^0|}{b_T^\epsilon}\left\{O(h_{yT}^R) + O(h_{wT}^R)\right\} \\ &\leqslant \frac{|\nabla_\theta q_1^0| + |\nabla_\theta q_2^0|}{b_T^\epsilon}\left\{O(h_{yT}h_{wT}^m)^{-1} + O(h_{yT}^R) + O(h_{wT}^R)\right\}, \end{aligned}$$

where the second equality follows from the definitions of u_{T12}^0 and u_{T21}^0, where $\sup_{y\in\mathbb{R}} |K_0(y)| < \infty$, $\sup_{w\in\mathbb{R}^m} |K(w)| < \infty$ and $\sup_{(y,w)\in\mathbb{R}^{m+1}} f^0(y|w) < \infty$ by (A11)(i,iii) and (A9)(ii). Thus, by Minkowski inequality we obtain

$$\begin{aligned} M_{T2} &\leqslant \sup_{1\leqslant t\neq s<\infty}\left\{\left[E|\nabla_\theta q_t^0|^{2r}\right]^{1/(2r)} + \left[E|\nabla_\theta q_s^0|^{2r}\right]^{1/(2r)}\right\} \times \\ &\quad \left\{O\left(\frac{1}{b_T^\epsilon h_{yT}h_{wT}^m}\right) + O\left(\frac{h_{yT}^R}{b_T^\epsilon}\right) + O\left(\frac{h_{wT}^R}{b_T^\epsilon}\right)\right\} \\ &= O\left(\frac{1}{b_T^\epsilon h_{yT}h_{wT}^m}\right) + O\left(\frac{h_{yT}^R}{b_T^\epsilon}\right) + O\left(\frac{h_{wT}^R}{b_T^\epsilon}\right), \end{aligned}$$

by (A7)(i). Given $b_T^\epsilon = b_T[1 + o(1)]$, $h_{yT}^R = o(b_T)$, $h_{wT}^R = o(b_T)$ and $b_T T^{1/4} h_{yT}h_{wT}^m \to \infty$, we have $M_{T2} = o(T^{1/4})$ as desired. Thus, Equation (74) implies $E(TU_2^2) = o(1)$. □

References

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Altonji, J., and L. Segal (1996): "Small Sample Bias in GMM Estimation of Covariance Structures," *Journal of Economic and Business Statistics*, 14, 353–366.

Andrews, D. W. K. (1994a): "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity," *Econometrica*, 62, 43–72.

——— (1994b): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, pp. 2247–2294. Elsevier Science.

——— (1995): "Nonparametric Kernel Estimation for Semiparametric Econometric Models," *Econometric Theory*, 11, 560–596.

Angrist, J., V. Chernozhukov, and I. Fernandez-Val (2006): "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–564.

Arcones, M. A. (1995): "On the Central Limit Theorem for U-Statistics under Absolute Regularity," *Statistics and Probability Letters*, 24, 245–249.

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 (1998): *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag, New York.

Bierens, H. J. (1983): "Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions," *Journal of the American Statistical Association*, 78, 699–707.

——— (1987): "Kernel Estimators of Regression Functions," in *Advances in Econometrics: Fifth World Congress*, ed. by T. Bewley, vol. 1 of *Advances in Econometrics: Fifth World Congress*, pp. 99–144. Cambridge University Press, New York.

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

Bilingsley, P. (1995): *Probability and Measure*. John Wiley and Sons, New York, 3rd edn.

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

BRADLEY, R. C. (1986): "Basic Properties of Strong Mixing Conditions," in *Dependence in Probability and Statistics*, ed. by E. Eberlein, and M. S. Taqqu, pp. 165–192. Birkhauser, Boston.

BROWN, B., AND W. K. NEWEY (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66, 453–464.

BUCHINSKY, M. (1994): "Changes in the US Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, 62, 405–458.

BUCHINSKY, M., AND J. HAHN (1998): "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66, 653–671.

CAI, Z. (2002): "Regression Quantiles for Time Series," *Econometric Theory*, 18, 169–192.

CHAMBERLAIN, G. (1986): "Asymptotic Efficiency in Semi-Parametric Models with Censoring," *Journal of Econometrics*, 32, 189–218.

——— (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

CHERNOZHUKOV, V., AND H. HONG (2002): "Three-Step Censored Quantile Regression and Extramarital Affairs," *Journal of the American Statistical Association*, 97, 872–882.

COSSLETT, S. R. (2004): "Efficient Semiparametric Estimation of Censored and Truncated Regressions Via a Smoothed Self-Consistency Equation," *Econometrica*, 72, 1277–1293.

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.

HAHN, J. (1997): "Efficient Estimation of Panel Data Models with Sequential Moment Restrictions," *Journal of Econometrics*, 79, 1–21.

HANSEN, B. E. (2004a): "Nonparametric Estimation of Smooth Conditional Distributions," University of Winsconsin, Madison.

——— (2004b): "Uniform Convergence Rates for Kernel Estimation," University of Winsconsin, Madison.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moment Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L. P., J. HEATON, AND M. OGAKI (1988): "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association*, 83, 863–871.

HARDLE, W., AND T. STOKER (1989): "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 89, 986–995.

HJORT, N. L., AND D. POLLARD (1993): "Asymptotics for Minimizers of Convex Processes," Yale University.

HOROWITZ, J., AND V. G. SPOKOINY (2002): "An Adaptive Rate-Optimal Test of Linearity for Median Regression Models," *Journal of the American Statistical Association*, 97, 822–835.

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.

KHAN, S. (2001): "Two-Stage Rank Estimation of Quantile Index Models," *Journal of Econometrics*, 100, 319–335.

KIM, T.-H., AND H. WHITE (2003): "Estimation, Inference, and Specification Analysis for Possibly Misspecified Quantile Regression," in *Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later*, ed. by T. Fromby, and R. C. Hill, pp. 107–132. Elsevier, New York.

KITAMURA, Y., G. TRIPATHI, AND H. AHN (2004): "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, 1667–1714.

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.

——— (1982): "Robust Tests for Heteroscedasticity Based on Regression Quantiles," *Econometrica*, 50(1), 43–62.

KOENKER, R., AND K. F. HALLOCK (2001): "Quantile Regression," *Journal of Economic Perspectives*, 15(4), 143–156.

KOENKER, R., AND Z. XIAO (2002): "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612.

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

KOMUNJER, I. (2005a): "Asymmetric Power Distribution: Theory and Applications to Risk Measurement," University of California, San Diego.

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

LAVERGNE, P., AND Q. VUONG (1996): "Nonparametric Selection of Regressors: The Nonnested Case," *Econometrica*, 64, 207–219.

NEWEY, W., AND R. J. SMITH (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–256.

NEWEY, W. K. (1990a): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58, 809–837.

——— (1990b): "Semiparametric Effficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135.

——— (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in *Handbook of Statistics, Volume 11: Econometrics*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, pp. 419–454. North Holland, Amsterdam.

——— (2004): "Efficient Semiparametric Estimation Via Moment Restrictions," *Econometrica*, 72, 1877–1897.

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.

OTSU, T. (2003): "Empirical Likelihood for Quantile Regression," University of Wisconsin Madison.

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

PORTNOY, S. (1991): "Behavior of Regression Quantiles in Non-Stationary, Dependent Cases," *Journal of Multivariate Analysis*, 38, 100–113.

POWELL, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

——— (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

ROBINSON, P. M. (1983): "Nonparametric Estimators for Time Series," *Journal of Time Series Analysis*, 4, 185–207.

——— (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica*, 55, 875–891.

——— (1988): "Root-N Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

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

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.

STONE, C. J. (1980): "Optimal Rates of Convergence for Nonparametric Estimators," *Annals of Statistics*, 8, 1348–1360.

——— (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10, 1040–1053.

TRUONG, Y. K., AND C. J. STONE (1992): "Nonparametric Function Estimation Involving Time Series," *Annals of Statistics*, 20, 77–97.

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.

ZHENG, J. X. (1998): "A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions," *Econometric Theory*, 14, 123–138.