*Econometric Theory*, **32**, 2016, 686–713.
doi:10.1017/S0266466615000031

# ASYMPTOTIC THEORY FOR NONLINEAR QUANTILE REGRESSION UNDER WEAK DEPENDENCE

WALTER OBERHOFER
*University of Regensburg*

HARRY HAUPT
*University of Passau*

This paper studies the asymptotic properties of the nonlinear quantile regression model under general assumptions on the error process, which is allowed to be heterogeneous and mixing. We derive the consistency and asymptotic normality of regression quantiles under mild assumptions. First-order asymptotic theory is completed by a discussion of consistent covariance estimation.

## 1. INTRODUCTION

The concept of quantile regression introduced in the seminal paper of Koenker and Bassett (1978), has become a widely used and accepted technique in many areas of theoretical and applied econometrics. The first monograph on this topic has been published by Koenker (2005), covering a wide scope of well established foundations and (even a 'twilight zone' of) actual research frontiers. In addition, many of the numerous new concepts in this fast evolving field have been reviewed and summarized in survey articles (see inter alia Buchinsky, 1998; Yu, Lu, and Stander, 2003) and econometric textbooks (e.g., Peracchi, 2001; Wooldridge, 2010). In contrast to the more methodological literature, there are also important, nontechnical attempts to bring the key concepts and especially the applicability of quantile estimation to a wider audience outside the statistical profession (e.g., Koenker and Hallock, 2001).

This paper deals with quantile regressions where the dependent variable $y$ and covariates $x_1, \ldots, x_K$ satisfy a nonlinear model with additive errors. Let $(\Omega, \mathcal{F}, P)$

Our heartfelt thanks to Matei Demetrescu, Bernd Fitzenberger, Keith Knight, Rolf Tschernig, and especially Roger Koenker for encouraging discussions on earlier drafts of the paper. The second author gratefully acknowledges financial support of the International Centre for Mathematical Sciences (ICMS). Finally we thank three anonymous referees and the associate editor Oliver Linton for helpful comments and suggestions. Of course all remaining errors are ours. Address correspondence to Harry Haupt, Department of Statistics, University of Passau, 94030 Passau, Germany; e-mail: harry.haupt@uni-passau.de.

 © Cambridge University Press 2015

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

be a complete probability space and let $\{y_t\}_{t\in\mathbb{N}}$ be an $\mathcal{F}$-measurable scalar random sequence. We consider the regression model

$$y_t - g(x_t, \beta_0) = u_t, \quad 1 \le t \le T, \tag{1}$$

where $\beta_0 \in D_\beta \subset \mathbb{R}^K$ is a vector of unknown parameters, the $1 \times L$ vectors $x_t$ are assumed to be deterministic and given, though in Section 2 we comment on the stochastic regressors case. The dependent variables $y_t$ are observable, $g(x, \beta)$ is, in general, a nonlinear function defined for $x \in D_x$ and $\beta \in D_\beta$ from $D_x \times D_\beta \to \mathbb{R}$, where $x_t \in D_x$ for all $t$, and $\{u_t\}$ is an error process. Quantile regression asymptotics for this model have been studied in Oberhofer (1982), Jureckova and Prochazka (1994), and Koenker (2005).

Oberhofer (1982) considered the consistency and Wang (1995) the asymptotic normality of the least absolute deviations (LAD) estimator under the assumption of independent and identically distributed (i.i.d.) errors, respectively. Liese and Vajda (2003, 2004), and He and Shao (1996) provide very general treatments for other classes of M-estimators in this context.

The i.i.d. assumption has been challenged in different ways in the quantile regression literature. Koenker and Bassett (1982) first investigated the case of heteroscedasticity based on regression quantiles, other authors discussed this case for the most prominent quantile, the median (see for example Knight, 1999; Zhao, 2001, and the literature cited therein). Quantile regression models for (weakly) dependent data have been studied for LAD estimation by Phillips (1991) and Weiss (1991), for unconditional quantiles in a parametric context by Oberhofer and Haupt (2005), for marginal sample quantiles by Dominicy, Hörmann, Ogata, and Veredas (2012), as an alternative for classical periodogram estimators by Li (2012), for linear regression models by Portnoy (1991) and Fitzenberger (1997). The latter also provides an extensive discussion of bootstrap-based consistent covariance estimation. In a nonparametric context, De Gooijer and Zerom (2003) discuss additive models, Ioannides (2004) and Cai (2002) consider nonparametric time series (forecasting) models, where the latter surveys the preceding literature in this context. Recently, El Ghouch and Genton (2009) propose a mixture of parametric and nonparametric approaches in a non-iid framework. Under quite general conditions, Komunjer (2005) introduces the class of 'tick-exponential' quasi-maximum likelihood estimators (QMLE) of possibly misspecified dynamic nonlinear quantile regression models. Under specific distributional assumptions[1], the proposed QMLE embeds traditional quantile regression estimators à la Koenker and Bassett (1978).

Fixed regressors in a time series or dependent data framework have been employed among others in Roussas, Tran, and Ioannides (1992), Tran, Roussas, Yakowitz, and Van (1996), and Robinson (1997), and in the quantile regression context by Phillips (1991), Pollard (1991), Fitzenberger (1997), Ioannides (2004), and Zhou and Shao (2013). Roussas et al. (1992, p. 263) provide a motivating example for such a setup. Pötscher and Prucha (1997, Chap. 6.1) provide a rationale for using mixing conditions in a static model.

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

Other relevant works in this context include Richardson and Bhattacharyya (1987), Koenker and Park (1994), Jureckova and Prochazka (1994), Powell (1991, 1994), White (1994), and, more recently, Zheng (1998), Mukherjee (1999, 2000), Chernozhukov and Umantsev (2001), and Engle and Manganelli (2004) with applications to finance, Kim, Kim, and Hur (2002), Chen, Tran, and Lin (2004), Karlsson (2007, 2009), and Chen and Zhou (2010). The monograph of Koenker (2005) reviews and reflects additional literature on (nonlinear) quantile regression asymptotics. Recently Chen, Koenker, and Xiao (2009) studied copula based nonlinear parametric quantile autoregressions (NLQAR) using similar models as Weiss (1991) and Engle and Manganelli (2004), though improving on the conditions used previously. Interestingly the class of copula-based NLQAR models generically constitutes specific forms of nonlinear regression functions. Other examples using a priori known regression functions in NLQR are Box-Cox transformations (e.g., Powell, 1991; Chamberlain, 1994; Buchinsky, 1995; Fitzenberger, Wilke, and Zhang, 2010).

The major goal of this paper is to extend the work on nonlinear quantile regression (NLQR) of Oberhofer (1982), Jureckova and Prochazka (1994), and Koenker (2005) to a general non-iid framework, where we allow for heterogeneous mixing processes. We provide proofs of the consistency and asymptotic normality of coefficient estimators as well as a consistent estimator of the asymptotic covariance matrix. While improving on several assumptions in the literature this paper is the first to provide detailed proofs of first-order asymptotic theory in such a general model.

The paper is organized as follows. In Section 2, we provide a proof of the weak consistency of nonlinear regression quantiles and a thorough discussion of the underlying assumptions. In Section 3, we thoroughly discuss the assumptions and provide an in-depth treatment and proof of asymptotic normality of nonlinear regression quantiles under weak dependence. In Section 4, we discuss the consistent estimation of the covariance matrix under dependence and heterogeneity. The Appendix contains proofs of our four main theorems and twelve technical lemmas, which may have their own merits.

## 2. CONSISTENCY

Our aim is to analyze the asymptotic behavior of the $\vartheta$-quantile regression estimator $\hat{\beta}_T$, i.e., $\beta = \hat{\beta}_T$ minimizing the asymmetrically weighted absolute deviations objective function

$$\sum_{t=1}^{T} \rho_\vartheta \left( y_t - g(x_t, \beta) \right), \tag{2}$$

where $0 < \vartheta < 1$ and $\rho_\vartheta(z) = z(\vartheta - I[z \leq 0])$ is the check function introduced by Koenker and Bassett (1978) and $I[.]$ is the usual indicator function. From (1) follows that the deviations in (2) can be written as

$$y_t - g(x_t, \beta) = u_t + g(x_t, \beta_0) - g(x_t, \beta) = u_t - h_t(\alpha), \tag{3}$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

where, for the sake of convenience for the derivation and discussion of asymptotic results, we define $\alpha \stackrel{\text{def}}{=} \beta - \beta_0$ with $D_\alpha = \{\alpha | \alpha = \beta - \beta_0, \beta \in D_\beta\}$, and

$$h_t(\alpha) \stackrel{\text{def}}{=} g(x_t, \beta_0 + \alpha) - g(x_t, \beta_0).$$

In order to avoid unnecessary moment requirements, we follow the suggestion of Huber (1967) to replace (2) with the equivalent objective function

$$\begin{aligned} Q_T(\alpha) = \sum_{t=1}^{T} q_t(\alpha) &\stackrel{\text{def}}{=} \sum_{t=1}^{T} [\rho_\vartheta(u_t - h_t(\alpha)) - \rho_\vartheta(u_t)] \\ &= \sum_{t=1}^{T} (h_t(\alpha) - u_t)\,(I[u_t \le h_t(\alpha)] - \vartheta) \\ &\quad + u_t\,(I[u_t \le 0] - \vartheta)\,. \end{aligned} \tag{4}$$

As $\alpha = 0$ corresponds to $\beta_0$, we can study the behavior of the former instead of $\beta = \beta_0$. If an estimator $\hat{\alpha}_T$ results from minimizing (4), we get $\hat{\beta}_T = \beta_0 + \hat{\alpha}_T$. For asymptotic analysis, we are interested in the suitably scaled difference $\hat{\beta}_T - \beta_0$.

Noteworthy, the summands $q_t(\alpha)$ reveal the important inequality $|q_t(\alpha)| \le |h_t(\alpha)|$ (see e.g., Jureckova and Prochazka, 1994, and Lemma 1C, Appendix). As a consequence, every moment of $q_t(\alpha)$ exists for finite $h_t(\alpha)$. Hence, the expected value of $q_t(\alpha)$ exists even if the expected value of $u_t$ does not exist.

Following the approach of Knight (1998), we decompose $Q_T(\alpha)$ according to

$$q_t(\alpha) = b_t(\alpha) + c_t(\alpha), \tag{5}$$

or $Q_T(\alpha) = B_T(\alpha) + C_T(\alpha)$, where $B_T(\alpha) = \sum_{t=1}^{T} b_t(\alpha)$ and $C_T(\alpha) = \sum_{t=1}^{T} c_t(\alpha)$. The summands in (5) are defined as

$$b_t(\alpha) \stackrel{\text{def}}{=} |h_t(\alpha) - u_t|\,(I[0 < u_t \le h_t(\alpha)] + I[h_t(\alpha) < u_t \le 0]), \tag{6}$$

$$c_t(\alpha) \stackrel{\text{def}}{=} -h_t(\alpha)\,\psi_\vartheta(u_t), \tag{7}$$

with $\psi_\vartheta(z) \stackrel{\text{def}}{=} \vartheta - I[z \le 0]$ being the right-hand derivative of the check-function $\rho_\vartheta(z)$. By virtue of (5) we can study the asymptotic behavior of the objective function by studying separately that of $b_t(\alpha)$ and $c_t(\alpha)$. The summand $c_t(\alpha)$ has an interesting interpretation, as its first factor arises from the deviation between the regression function and its true value, and its second factor is a Bernoulli random variable capturing the dependence structure of the present regression problem.

We assume the typical quantile regression normalization under the implicit assumption that the regression function $g$ contains an intercept[2].

**Assumption A.1.** For the distribution $F_t(z)$ of $u_t$ let $F_t(0) = P(u_t \le 0) = \vartheta$, $0 < \vartheta < 1$ for all $t$.

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

As a consequence $E[\psi_\vartheta(u_t)] = 0$ and hence $E[c_t(\alpha)] = 0$ (given Assumption A.5 below).

**Assumption A.2.** $u_t$ is $\alpha$-mixing (see e.g., Doukhan, 1994, p. 3)

**Assumption A.3.** There exist a positive $f_0$ and a positive $\delta_0$, such that for all $|x| \leq \delta_0$ and all $t$,

$$\min[F_t(|x|) - F_t(0), F_t(0) - F_t(-|x|)] \geq f_0|x|.$$

While Assumptions A.1–A.3 refer to the error process, the following assumptions refer to properties of the covariates and the regression function. Together these assumptions allow us to establish a generic ULLN and weak consistency of $\hat{\beta}_T$ minimizing (2).

**Assumption A.4.** $D_\beta$ is compact, $\beta_0$ is an inner point of $D_\beta$, and $g(z, \beta)$ is continuous in $\beta$ for $z \in D_x$.

**Assumption A.5.** The $1 \times L$ vectors $x_t$ are deterministic and known, $t = 1, 2, \ldots$.

**Assumption A.6.** For every $\epsilon > 0$ there exists a positive $\delta_\epsilon$ such that for all $\beta \in D_\beta$

$$\liminf_T \inf_{||\tilde{\beta}-\beta|| \geq \epsilon} T^{-1} \sum_{t=1}^{T} \left| g\left(x_t, \tilde{\beta}\right) - g(x_t, \beta) \right| > \delta_\epsilon.$$

**Assumption A.7.** For some $\epsilon > 0$, and all $\beta \in D_\beta$

$$\limsup_T T^{-1} \sum_{t=1}^{T} |g(x_t, \beta)|^{1+\epsilon} < \infty.$$

**Assumption A.8.** For every $\beta \in D_\beta$ and every $\epsilon > 0$, there exists a $\delta > 0$, such that

$$\limsup_T \sup_{||\tilde{\beta}-\beta|| \leq \delta} T^{-1} \sum_{t=1}^{T} \left| g\left(x_t, \tilde{\beta}\right) - g(x_t, \beta) \right| < \epsilon.$$

THEOREM 1. *In model (1), under Assumptions A.1–A.8,* $\operatorname{plim}_T \hat{\beta}_T = \beta_0$.

**Proof.** Appendix. ■

Discussion of assumptions:

In this framework, the existence of a measurable estimator $\hat{\beta}_T$ is ensured by Pfanzagl (1969, Thm. 3.10), which holds under the assumptions stated above. For a further discussion of compactness and measurability, the reader is referred to the discussion in Pötscher and Prucha (1997, Lemma 3.4, Chap. 4.3).

Violations of Assumption A.3 are treated in Knight (1998) and Rogers (2001).

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

It is not necessary for the regressors to be deterministic as postulated in Assumption A.5, as similar behavior can be expected for random regressors $\{x_t\}$ independent of the disturbances $\{u_t\}$ as in Zhao and Xiao (2014). Consider the example of a linear regression function and let $\{x_t\}$ be a stationary sequence with $E(x_t'x_t)$ finite and nonsingular. Then, almost all realizations would have the necessary limiting properties (see the discussion in Pollard, 1991). In the present NLQR context, stating Assumptions A.6–A.8 for stochastic regressors requires further assumptions differing from the ones stated in Pollard (1991) for the linear case. An example for such an assumption can be seen in the moment condition stated in Pötscher and Prucha (1997, Thm. 6.3), playing an analogous role to our Assumption A.7 (discussed in detail below). Proceeding further in this vein while avoiding the restrictive assumptions (see also the discussion in Fitzenberger, 1997) in Weiss (1991) or in Jureckova and Prochazka (1994) requires setup and detail beyond the scope of the present discussion.

The identifiable uniqueness condition in Assumption A.6 corresponds to the analogous condition $\liminf_T \inf_{||\tilde{\beta}-\beta||\geq\epsilon} T^{-1}\sum_{t=1}^{T}\left[x_t\left(\tilde{\beta}-\beta\right)\right]^2 = \liminf_T \inf_{||\alpha||>\epsilon} \alpha' T^{-1}X'X\alpha > 0$ in the linear regression model $y_t = x_t\beta + u_t$ using least squares estimation, implied by the nonsingularity of the limit of the matrix $T^{-1}X'X$. The dominance condition A.7 rules out a too strong growth of the covariates, while the identification Assumption A.6 guarantees enough variation. The trade-off problem between Assumptions A.6 and A.7, however, is rather involved for the general nonlinear case and lies beyond the focus of this paper. For the linear case, Assumption A.7 is given by $\limsup_T T^{-1}\sum_{t=1}^{T}|x_t\beta|^{1+\epsilon} < \infty$, respectively, implied by the assumption $\limsup_T T^{-1}\sum||x_t||^{1+\epsilon} < \infty$. As pointed out by Haupt and Oberhofer (2009), previous work of Wang (1995, 1996) on $L_1$- and $L_2$-norm estimation asymptotics in nonlinear regression has not addressed this problem (see also Prakasa Rao, 1984, 1987).

As Assumption A.8 guarantees sufficient continuity of the Cesàro sum, it allows to establish a generic ULLN in the spirit of Andrews (1987), Pötscher and Prucha (1989, 1994, 1997), and Gallant and White (1988). Alternatively, a more restrictive Lipschitz condition can be used. E.g., for each $\beta \in D_\beta$, there exists a constant $\eta > 0$ such that $||\tilde{\beta}-\beta|| \leq \eta$ implying $\left|g\left(x_t,\tilde{\beta}\right)-g(x_t,\beta)\right| \leq G_t\left|\left|\tilde{\beta}-\beta\right|\right|$, where $\limsup_T T^{-1}\sum_{t=1}^{T}G_t < \infty$, and $G_t$ and $\eta$ may depend on $\beta$. Insightful discussions of assumptions commonly used to verify (uniform) stochastic equicontinuity conditions can be found among others in Davidson (1994, Chap. 21.4) and Newey and McFadden (1994, Chap. 7.2), or Andrews (1994a, 1994b, Chap. 4, 5), who considers and contrasts Lipschitz and $L_p$ continuity conditions.

Beyond the discussed guidelines to prove such a result, a detailed proof of Theorem 1 in the NLQR context is given only in Oberhofer (1982), and Jureckova and Prochazka (1994)[3]. However, both results consider the iid case, while Theorem 1 allows for heterogeneous and dependent errors. Furthermore, we do not require the monotonicity assumption of Jureckova and Prochazka (1994, A.4).

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

Our assumptions guaranteeing uniqueness of the estimator are less restrictive in comparison to Oberhofer (1982), Jureckova and Prochazka (1994), and Koenker (2005).

## 3. ASYMPTOTIC NORMALITY

The starting point for our derivation of the limiting law is the first order conditions for a local minimum $\hat{\alpha}_T$ of the loss function $Q_T(\alpha)$ defined in (4). In Lemma 2N in the Appendix it is shown that the corresponding first order conditions can be written as inequality

$$A_{lT}(\alpha, w) \leq S_T(\alpha, w) \leq A_{uT}(\alpha, w), \tag{8}$$

where the entities in (8) are defined in (28) and $w$ is the direction of the derivative of the loss function.

For the derivation of asymptotic normality we require the weak consistency of $\hat{\beta}_T$, that is $\text{plim}_T \hat{\beta}_T = \beta_0$ or $\text{plim}_T \hat{\alpha}_T = 0$. From consistency[4] follows that we can employ the restriction $||\alpha|| \leq c$, where $c$ is positive and arbitrarily small, but independent from $T$, in several of the assumptions below. We assume:

**Assumption A.9.** For all $t$ and $||\alpha|| \leq c$, $h_t(\alpha)$ has continuous second derivatives with respect to all $\alpha_i$. Let $\nabla h_t(\alpha)$ denote the $(K \times 1)$-vector with $i$th component $\partial h_t(\alpha)/\partial \alpha_i$, $i = 1, \ldots, K$, and let $\nabla^2 h_t(\alpha)$ denote the $(K \times K)$-matrix with element $\partial^2 h_t(\alpha)/\partial \alpha_i \partial \alpha_j$ in row $i$ and column $j$.

**Assumption A.10.** $\limsup_T T^{-1} \sum_{t=1}^{T} \sup_{||\alpha|| \leq c} ||\nabla h_t(\alpha)||^2 < \infty$.

**Assumption A.11.** Let $w$ and $v$ be arbitrary vectors in $\mathbb{R}^K$ with $||w||=||v||=1$. Then

$$\limsup_T T^{-1} \sum_{t=1}^{T} \sup_{||\alpha|| \leq c} \left| w' \nabla^2 h_t(\alpha) v \right| < \infty.$$

**Assumption A.12.** $\lim_T \sup_{t \leq T} \sup_{||\alpha|| \leq c} T^{-1/2} ||\nabla h_t(\alpha)|| = 0$.

**Assumption A.13.** The density $f_t(z)$ of $F_t(z)$ exists for every $t$ and $z$ and is uniformly continuous in $t$ for $z = 0$.

**Assumption A.14.** $\limsup_t \sup_{||\alpha|| \leq c} f_t(h_t(\alpha)) < \infty$.

**Assumption A.15.** The $u_t$ are $\alpha$-mixing of size $-1$ with mixing coefficients $\mu_k$, $k = 1, 2, \ldots$ Hence, there exists an $\eta > 0$ such that $\mu_k = O(k^{-1-\eta})$.

**Assumption A.16.** The $(K \times K)$-matrix

$$T^{-1} \sum_{t=1}^{T} \nabla h_t(0) f_t(h_t(\lambda(t))) \nabla h_t(0)'$$

is nonsingular for sufficiently large $T$ and all $\lambda(t) \in \mathbb{R}^K$ with $||\lambda(t)|| \leq c$.

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

**Assumption A.17.** The $(K \times K)$-matrix $V_T = T^{-1} \sum_{t=1}^{T} \nabla h_t(0) f_t(0) \nabla h_t(0)'$ converges to the nonsingular matrix $V_0$ for $T \to \infty$.

**Assumption A.18.** The $(K \times 1)$-vector $\xi_T = T^{-1/2} \sum_{t=1}^{T} \nabla h_t(0) \left( I[u_t \leq 0] - \vartheta \right)$ converges in distribution to a normal distribution with mean 0 and covariance matrix $\Sigma_0$ for $T \to \infty$.

THEOREM 2. *In model (1), under Assumptions A.1–A.16, for* $\varphi(T) > 0$ *and* $\lim_T \varphi(T)^2 / T = 0$, *follows* $\operatorname{plim}_T \varphi(T) \left( \hat{\beta}_T - \beta_0 \right) = 0$.

**Proof.** Appendix. ■

THEOREM 3. *In model (1), under Assumptions A.1–A.18,* $\sqrt{T} \left( \hat{\beta}_T - \beta_0 \right)$ *converges in distribution to a normal distribution with mean zero and covariance matrix* $V_0^{-1} \Sigma_0 V_0^{-1}$.

**Proof.** Appendix. ■

The lines of reasoning behind Theorems 2 and 3 can be summarized as follows. Let $\hat{\alpha}_T$ be a solution of the first order conditions for a minimum of $Q_T(\alpha)$. We have to find a sequence of positive numbers $\varphi(T)$, $t = 1, 2, \ldots$, where $\lim_T \varphi(T) = \infty$ such that for every $\epsilon > 0$ there exists an $m' > 0$ and an $m > m'$ with the property

$$\lim_T P \left( m' < \left|\left| \varphi(T) \hat{\alpha}_T \right|\right| < m \right) \geq 1 - \epsilon. \tag{9}$$

Condition (9) only ensures that asymptotically the distribution of $\varphi(T)\hat{\alpha}_T$ is nondegenerate (that is, does neither vanish nor grow without bound), though the limiting distribution need not be Gaussian. From (9) follows

$$\lim_T P \left( m'/\varphi(T) < \left|\left| \hat{\alpha}_T \right|\right| < m/\varphi(T) \right) \geq 1 - \epsilon, \tag{10}$$

implying that from minimization of $Q_T(\alpha)$ for $||\alpha|| \leq m/\varphi(T)$, where $m$ can be chosen arbitrarily large but independent from $T$, we have to find a solution $\hat{\alpha}_T$ satisfying condition (10). Hence an obvious choice is to use the transformation $\gamma = \varphi(T)\alpha$ and to calculate the first order conditions for $Q_T(\alpha) = Q_T(\gamma/\varphi(T))$ as a function of $\gamma$. When we estimate $\gamma$ by $\hat{\gamma}_T$, then $\hat{\alpha}_T = \hat{\gamma}_T/\varphi(T)$ is a solution of the first order conditions corresponding to a minimum of $Q_T(\alpha)$. Choosing a suitable scaling of $Q_T(\alpha)$ ensures that it is nondegenerate without changing its minimand. From the proof of Lemma 4N it is obvious that the scaling factor $\varphi(T)^2/T$ is a suitable choice.

In Theorem 2, it is shown that for choosing $\varphi(T)$ such that $\lim_T \varphi(T)^2/T = 0$ and for $m'$ arbitrarily small follows

$$\lim_T P \left( \left|\left| \varphi(T) \hat{\alpha}_T \right|\right| < m' \right) = 1, \tag{11}$$

implying that the selected $\varphi(T)$ does not satisfy condition (9). Hence, for such a choice of $\varphi(T)$, the estimator $\varphi(T)\hat{\alpha}_T$ cannot follow a nondegenerate

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

limiting distribution. In Theorem 3, it is shown that for choosing $\varphi(T)$ such that $\varphi(T)^2/T = 1$, that is $\varphi(T) = \sqrt{T}$, not only condition (9) holds but the limiting distribution of $\varphi(T)\hat{\alpha}_T$ is Gaussian.

As Lemma 1N is essential for the arguments employed in our asymptotic normality proof we require the compactness of the parameter space. Hence, we restrict our analysis to $\{\gamma \mid ||\gamma|| \leq m\}$, where $m$ is a positive real number which can be chosen arbitrarily large, but independent from $T$. Such a restriction of the parameter space may imply that we can not find a solution for every $\omega \in \Omega$ but only for a possibly empty subset $\Omega_{T,m}$ depending on $T$ and $m$. We have to show that for every $\epsilon > 0$ there exists an $m$ such that $\lim_T P(\Omega_{T,m}) \geq 1 - \epsilon$. In this case the restriction is not critical.

A thorough discussion of the preliminary lemmas necessary to establish Theorems 2 and 3 is provided in the Appendix. The remainder of this section is devoted to a brief discussion of some assumptions: Chen et al. (2009, Assumption 3.6) employ assumptions similar to Assumptions A.9, A.10, A.11, A.16, and A.17. Assumptions A.10, A.11, A.14, and A.16 are local dominance and identification conditions. Note that from Assumptions A.8 and A.10 follows, Assumption A.8. Analogously to our considerations in the previous section, but in contrast to the work of Chen et al. (2009), we employ a mixing assumption in A.15. In the reasoning behind such an assumption we agree to Pötscher and Prucha (1997, Chap. 6) who argue that mixing conditions are "problem adequate" for static models but not for dynamic models. Hence, Assumptions A.15 and (the assumptions implicit in) A.18 restrict the dependence structure imposed on the quantile regression model, while Assumption A.17 controls the form of heteroskedasticity. Assumption A.16 holds if $v'T^{-1}\sum_{t=1}^{T}\nabla h_t(0) f_t(h_t(\lambda(t)))\nabla h_t(0)'v > 0$ for all $v \in \mathbb{R}^K$ with $||v|| = 1$. Thus due to $v'T^{-1}\sum_{t=1}^{T}\nabla h_t(0) f_t(h_t(\lambda(t)))\nabla h_t(0)'v = T^{-1}\sum_{t=1}^{T}\left(v'\nabla h_t(0)\sqrt{f_t(h_t(\lambda(t)))}\right)^2$, Assumption A.16 follows from $T^{-1}\sum_{t=1}^{T}\left(v'\nabla h_t(0)\inf_{||\alpha||\leq c}\sqrt{f_t(h_t(\alpha))}\right)^2 > 0$.

Note that the nondifferentiability of the objective function makes a standard Taylor series argument impossible. However, also a result in Pollard (1984, Thm. 5) cannot be used, as it requires first and second derivatives of the loss function (see Pollard, p. 140/141). However, in the quantile regression context even first derivatives are not defined in all points of the parameter space (see Lemma 2N). The argument can not be saved by assuming that those points have measure zero, as, unfortunately, we have a preferred occurrence of the minimum at such points. In the limit those points occur in arbitrarily small neighborhoods: Hence, the infinitesimal probability given by the density at zero $f_t(0)$ is relevant for our arguments (but not in Pollard, 1984, Thm. 5).

## 4. CONSISTENT COVARIANCE MATRIX ESTIMATION

In order to complete the first-order asymptotic theory for the nonlinear quantile regression model under dependence and heterogeneity, we have to provide an estimator of the asymptotic covariance matrix and prove its consistency. The latter

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

is needed for statistical inference procedures such as for the construction of Wald tests or large-sample confidence regions for the regression parameters. Relative to the literature, the case considered here is more general as Powell's (1991) work, which is based on the independence assumption, whereas Weiss' (1991) analysis rests on martingale difference properties of the influence function, which cannot accommodate serial dependence of the influence function. In the present context, a consistent asymptotic covariance estimator will require both nonparametric estimation of the error densities (at zero) and estimation of the spectral density matrix (at zero) of the subgradient terms in the quantile minimization problem (e.g., Andrews, 1991). Fortunately, these two problems separate, as will be evident from our proof of Theorem 4 below: The outer matrix $V_0$ contains the densities $f_t(0)$, while the covariance structure reflected by the $\omega_{s,t}$ is manifested in the middle matrix $\Sigma_0$, which should be estimable using the heteroskedasticity and autocorrelation consistent (HAC) asymptotic covariance for the normalized subgradient (e.g., Powell, 1991; Fitzenberger, 1997). The outer matrix can be estimated using the procedures for heteroskedasticity-consistent covariance matrix given in Hendricks and Koenker (1992) or Powell (1991), which are contrasted in Koenker (2005, Chap. 3.4.2). The consistency results below thus refers to the case of given $\omega_{s,t}$ and $f_t(0)$.

In the proof of the following Theorem 4 we show that asymptotically $V_0^{-1}\Sigma_0 V_0^{-1}$ results from the probability limit of entities depending on observations. We assume

**Assumption A.19.** $\operatorname{plim}_T \xi_T \xi_T' = \Sigma_0$, where $\xi_T = T^{-1/2}\sum_{t=1}^T \nabla h_t(0)\left(I[u_t \le 0] - \vartheta\right)$.

THEOREM 4. *In model (1), under Assumptions A.1–A.19, a consistent estimator of the covariance matrix $V_0^{-1}\Sigma_0 V_0^{-1}$ of the limiting distribution of $\sqrt{T}\left(\hat\beta_T - \beta_0\right)$ is given by $\hat V_T^{-1}\hat\Sigma_T\hat V_T^{-1}$, where the outer matrix is given by $\hat V_T = T^{-1}\sum_{t=1}^T \nabla g(x_t, \hat\beta_T) f_t(0) \nabla g(x_t, \hat\beta_T)'$ and the estimated covariance structure by $\hat\Sigma_T = T^{-1}\sum_{s,t=1}^T \nabla g(x_s, \hat\beta_T)\omega_{s,t}\nabla g(x_t, \hat\beta_T)'$.*

**Proof.** Appendix. ■

## 5. CONCLUSIONS AND EXTENSIONS

The paper extends the body of asymptotic results on the nonlinear quantile regression model studied by Oberhofer (1982), Jureckova and Prochazka (1994), and Koenker (2005). A complete first-order asymptotic theory is established by consistency and asymptotic normality of the coefficient estimators, and a consistent estimator of the asymptotic covariance matrix under quite general conditions on the heterogeneity and dependence structure of the stochastic part of the regression model. The structure of the consistency proof in Section 2 is classic, though the assumptions are simpler and less restrictive than some conditions invoked in the previous literature (e.g., Oberhofer, 1982; Jureckova and

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

Prochazka, 1994; Koenker, 2005). Further consistency proofs given (in different QR model contexts) in Chen et al. (2009) and Weiss (1991) do not cover the fixed regressor case. The main result of the paper stated as Theorem 3 in Section 3 is, to the best of our knowledge, the first to provide an explicit and detailed proof of asymptotic normality under the given non-iid conditions in the NLQR context, where classic results such as Pollard (1984, Thm. 5) cannot be invoked. The discussion of a consistent estimator of the asymptotic covariance matrix in Section 4 builds on previous work of Powell (1991), Weiss (1991), Hendricks and Koenker (1992), and Fitzenberger (1997), and adapts and extends those arguments to the non-iid NLQR context.

The local nature of quantile specific conditions for modeling dependence and heterogeneity hints at potential for alternative, less restrictive dependence conditions and hence for more specific applications of limit theorems. Clearly, as only local weak dependence matters in the quantile regression context, it is neither necessary to invoke Assumption A.2 that the error process $\{u_t\}$ is strongly mixing, nor to assume that the Bernoulli process $\{\psi_\vartheta(u_t)\}$ defined in Section 2 is strongly mixing. Hence, further lines of research seem promising in the direction of replacing the mixing-assumptions in this paper by the recent seminal developments of weak dependence concepts proposed in Doukhan and Louhichi (1999), Dedecker, Doukhan, Lang, Leon, Louhichi, and Prieur (2007) or in the various works of Wu (see e.g., Wu and Mielniczuk, 2010), building on earlier work of Withers (1981). The concept of Doukhan and Louhichi (1999) is implied by near epoch dependence[5] as has been shown recently by Nze and Doukhan (2004).

## NOTES

1. The approach provides an interesting alternative to existing ones, especially with respect to consistent (HAC) covariance matrix estimation under general conditions (e.g., Newey and McFadden, 1994; Buchinsky, 1995; Fitzenberger, 1997). The required assumption of an a priori specification of the likelihood in the QMLE approach may be seen as one price of this generality and stands in contrast to other work cited in this paper.

2. Note that the inclusion of an intercept is required if $F_t(0)$ is constant and $F_t(0) \neq \vartheta$, where $F_t(z)$ is the distribution of $u_t$.

3. The consistency proof in Chen et al. (2009) does not cover the fixed regressor case.

4. Note that without the requirement of consistency we have to assume that the parameter space is compact and as a consequence $c$ no longer is arbitrarily small.

5. The use of a CLT based on a near epoch dependent (NED) process seems to be overly restrictive in the non-iid NLQR context, as the NED and the mixing process are identical in this case.

## REFERENCES

Andrews, D.W.K. (1987) Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica* 55, 1465–1471.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. (1994a) Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43–72.

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

Andrews, D.W.K. (1994b) Empirical process methods in econometrics. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2247–2294. Elsevier Science.

Buchinsky, M. (1995) Estimating the asymptotic covariance matrix for quantile regression models. *Journal of Econometrics* 68, 303–338.

Buchinsky, M. (1998) Recent advances in quantile regression. *Journal of Human Resources* 33, 88–126.

Cai, Z. (2002) Regression quantiles for time series. *Econometric Theory* 18, 169–192.

Chamberlain, G. (1994) Quantile regression, censoring, and the structure of wages. In C. Sims (ed.), *Advances in Econometrics: Sixth World Congress*, vol. 1, pp. 171–208. Cambridge University Press.

Chen, X., R. Koenker, & Z. Xiao (2009) Copula-based nonlinear quantile autoregression. *Econometrics Journal* 12, S50–S67.

Chen, L.-A., L.T. Tran, & L.-C. Lin (2004) Symmetric regression quantile and its application to robust estimation for the nonlinear regression model. *Journal of Statistical Planning and Inference* 126, 423–440.

Chen, N. & S. Zhou (2010) Simulation-based estimation of cycle time using quantile regression. *IIE Transactions* 43, 176–191.

Chernozhukov, V. & L. Umantsev (2001) Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics* 26, 271–292.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.

Dedecker, J., P. Doukhan, G. Lang, J.R. Leon, S. Louhichi, & C. Prieur (2007) *Weak Dependence: With Examples and Applications*. Springer-Verlag.

De Gooijer, J.G. & D. Zerom (2003) On additive conditional quantiles with high-dimensional covariates. *Journal of the American Statistical Association* 98, 135–146.

Dominicy, Y., S. Hörmann, H. Ogata, & D. Veredas (2012) On sample marginal quantiles for stationary processes. *Statistics and Probability Letters* 83, 28–36.

Doukhan, P. (1994) *Mixing*. Springer-Verlag.

Doukhan, P. & S. Louhichi (1999) A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications* 84, 313–342.

El Ghouch, A. & M.G. Genton (2009) Local polynomial quantile regression with parametric features. *Journal of the American Statistical Association* 104, 1416–1429.

Engle, R.F. & S. Manganelli (2004) CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economics Statistics* 22, 367–381.

Fitzenberger, B. (1997) The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics* 82, 235–287.

Fitzenberger, B., R. Wilke, & X. Zhang (2010) Implementing Box-Cox quantile regression. *Econometric Reviews* 29, 158–181.

Gallant, A.R. & H. White (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell.

Haupt, H. & W. Oberhofer (2009) On asymptotic normality in nonlinear regression. *Statistics and Probability Letters* 79, 848–849.

He, X. & Q.-M. Shao (1996) A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Annals of Statistics* 24, 2608–2630.

Hendricks, W. & R. Koenker (1992) Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association* 87, 58–68.

Huber, P.J. (1967) Behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 129–156, Berkeley. UC Press.

Ioannides, D.A. (2004) Fixed design regression quantiles for time series. *Statistics and Probability Letters* 68, 235–245.

Jureckova, J. & B. Prochazka (1994) Regression quantiles and trimmed least squares estimators in nonlinear regression models. *Journal of Nonparametric Statistics* 3, 201–222.

Karlsson, A. (2007) Nonlinear quantile regression estimation of longitudinal data. *Communications in Statistics – Simulation and Computation* 37, 114–131.

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

Karlsson, A. (2009) Bootstrap methods for bias correction and confidence interval estimation for nonlinear quantile regression of longitudinal data. *Journal of Statistical Computation and Simulation* 79, 1205–1218.

Kim, T.S., H.K. Kim, & S. Hur (2002) Asymptotic properties of a particular nonlinear regression quantile estimation. *Statistics and Probability Letters* 60, 387–394.

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

Knight, K. (1999) Asympotics for $L_1$-estimators of regression parameters under heteroscedasticity. *Canadian Journal of Statistics* 27, 497–507.

Koenker, R. (2005) *Quantile Regression.* Econometric Society Monographs No. 38. Cambridge University Press.

Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 46, 33–50.

Koenker, R. & G. Bassett (1982) Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50, 43–61.

Koenker, R. & K.F. Hallock (2001) Quantile regression. *Journal of Economic Perspectives* 15, 143–156.

Koenker, R. & B. Park (1994) An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics* 71, 265–283.

Komunjer, I. (2005) Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics* 128, 137–164.

Li, T.-H. (2012) Quantile periodograms. *Journal of the American Statistical Association* 107, 765–776.

Liese, F. & I. Vajda (2003) A general asymptotic theory of M-estimators. I. *Mathematical Methods of Statistics* 12, 454–477.

Liese, F. & I. Vajda (2004) A general asymptotic theory of M-estimators. II. *Mathematical Methods of Statistics* 13, 82–95.

Mukherjee, K. (1999) Asymptotics of quantiles and rank scores in nonlinear time series. *Journal of Time Series Analysis* 20, 173–192.

Mukherjee, K. (2000) Linearization of randomly weighted empiricals under long range dependence with applications to nonlinear regression quantiles. *Econometric Theory* 16, 301–323.

Newey, W.K. & D.L. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2111–2245. Elsevier Science.

Nze, P.A. & P. Doukhan (2004) Weak dependence: Models and applications to econometrics. *Econometric Theory* 20, 169–192.

Oberhofer, W. (1982) The consistency of nonlinear regression minimizing the $L_1$ norm. *Annals of Statistics* 10, 316–319.

Oberhofer, W. & H. Haupt (2005) The asymptotic distribution of the unconditional quantile estimator under dependence. *Statistics and Probability Letters* 73, 243–250.

Peracchi, F. (2001) *Econometrics*. Wiley.

Pfanzagl, J. (1969) On the measurability and consistency of minimum contrast estimates. *Metrika* 14, 249–272.

Phillips, P.C.B. (1991) A shortcut to LAD estimator asymptotics. *Econometric Theory* 7, 450–463.

Pötscher, B.M. & I.R. Prucha (1989) A uniform law of large numbers for dependent and heterogeneous data processes. *Econometrica* 57, 675–683.

Pötscher, B.M. & I.R. Prucha (1994) Generic uniform convergence and equicontinuity concepts for random functions. *Journal of Econometrics* 60, 23–63.

Pötscher, B.M. & I.R. Prucha (1997) *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer-Verlag.

Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer-Verlag.

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

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

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

Powell, J.L. (1991) Estimation of monotonic regression models under quantile restrictions. In W. Barnett, J.L. Powell, & G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 357–384. Cambridge University Press.

Powell, J.L. (1994) Estimation of semiparametric models. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2443–2521. Elsevier Science.

Prakasa Rao, B.L.S. (1984) The rate of convergence of the least squares estimator in a nonlinear regression model with dependent errors. *Journal of Multivariate Analysis* 14, 315–322.

Prakasa Rao, B.L.S. (1987) *Asymptotic Theory of Statistical Inference*. Wiley.

Richardson, G.D. & B.B. Bhattacharyya (1987) Consistent $L_1$-estimators in nonlinear regression for a noncompact parameter space. *Sankhya Series A* 49, 377–387.

Robinson, P.M. (1997) Large-sample inference for nonparametric regression with dependent errors. *Annals of Statistics* 25, 2054–2083.

Rogers, A.J. (2001) Least absolute deviations regression under nonstandard conditions. *Econometric Theory* 17, 820–852.

Roussas, G.G., L.T. Tran, & D.A. Ioannides (1992) Fixed design regression for time series: Asymptotic normality. *Journal of Multivariate Analysis* 40, 262–291.

Tran, L.T., G.G. Roussas, S. Yakowitz, & B.T. Van (1996) Fixed-design regression for linear time series. *Annals of Statistics* 24, 975–991.

Wang, J. (1995) Asymptotic normality of $L_1$-estimators in nonlinear regression. *Journal of Multivariate Analysis* 54, 227–238.

Wang, J. (1996) Asymptotics of least-squares estimators for constrained nonlinear regression. *Annals of Statistics* 4, 1316–1326.

Weiss, A.A. (1991) Estimating nonlinear dynamic models using least absolute error estimation. *Econometric Theory* 7, 46–68.

White, H. (1994) *Estimation, Inference and Specification Analysis*. Econometric Society Monographs No. 22. Cambridge University Press.

Withers, C.S. (1981) Central limit theorems for dependent variables. I. *Zeitschrift fr Wahrscheinlichkeitstheorie und verwandte Gebiete* 57, 509–534.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press.

Wu, W.B. & J. Mielniczuk (2010) A new look at measuring dependence. In P. Doukhan et al. (eds.), *Dependence in Probability and Statistics*. Lecture Notes in Statistics, vol. 200, pp. 123–142. Springer.

Yu, K., Z. Lu, & J. Stander (2003) Quantile regression: Applications and current research areas. *The Statistician* 52, 331–350.

Zhao, Q. (2001) Asymptotic efficient median regression in the presence of heteroscedasticity of unknown form. *Econometric Theory* 17, 765–784.

Zhao, Z. & Z. Xiao (2014) Efficient regressions via optimally combining quantile information. *Econometric Theory* 30, 1272–1314.

Zheng, J.X. (1998) A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory* 14, 123–138.

Zhou, Z. & X. Shao (2013) Inference for linear models with dependent errors. *Journal of the Royal Statistical Society B* 75, 323–343.

# APPENDIX: Consistency

The proof of Theorem 1 rests upon the following Lemmas 1C–3C.

LEMMA 1C. *For each pair* $\alpha, \tilde{\alpha} \in D_\alpha$, $D_\alpha = \{\alpha | \alpha = \beta - \beta_0, \beta \in D_\beta\}$, *follows*

$$|q_t(\tilde{\alpha}) - q_t(\alpha)| \leq \max(\vartheta, 1-\vartheta)|h_t(\tilde{\alpha}) - h_t(\alpha)|.$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

**Proof of Lemma 1C.** For $u_t - h_t(\tilde{\alpha}) > 0$ and $u_t - h_t(\alpha) > 0$ the assertion follows directly from (4). The same is valid for $u_t - h_t(\tilde{\alpha}) \le 0$ and $u_t - h_t(\alpha) \le 0$. Furthermore, $u_t - h_t(\tilde{\alpha}) > 0$ and $u_t - h_t(\alpha) \le 0$ imply

$$0 < u_t - h_t(\tilde{\alpha}) \le h_t(\alpha) - h_t(\tilde{\alpha}). \tag{12}$$

Thus, according to (4) we have

$$q_t(\tilde{\alpha}) - q_t(\alpha) = (1-\vartheta)\big[h_t(\tilde{\alpha}) - h_t(\alpha)\big] + u_t - h_t(\tilde{\alpha}). \tag{13}$$

From (12) and (13) follows $(1-\vartheta)[h_t(\tilde{\alpha}) - h_t(\alpha)] < q_t(\tilde{\alpha}) - q_t(\alpha) \le -\vartheta[h_t(\tilde{\alpha}) - h_t(\alpha)]$. Analogous considerations for $u_t - h_t(\tilde{\alpha}) \le 0$ and $u_t - h_t(\alpha) > 0$ are left to the reader. ■

LEMMA 2C. *Under Assumptions A.2, A.4, A.5, A.7, and A.8,*

$$\operatorname*{plim}_T \sup_{\alpha \in D_\alpha} \left| \frac{1}{T} Q_T(\alpha) - E \frac{1}{T} Q_T(\alpha) \right| = 0.$$

**Proof of Lemma 2C.** From Lemma 1C and Assumption A.8 follows for every $\alpha \in D_\alpha$ and for every $\epsilon > 0$ that there exists a $\delta > 0$ such that

$$\limsup_T \sup_{||\tilde{\alpha} - \alpha|| \le \delta} \frac{1}{T} \sum_{t=1}^{T} |q_t(\tilde{\alpha}) - q_t(\alpha)| < \epsilon, \text{ and}$$

$$\limsup_T \sup_{||\tilde{\alpha} - \alpha|| \le \delta} \frac{1}{T} \sum_{t=1}^{T} E\,|q_t(\tilde{\alpha}) - q_t(\alpha)| < \epsilon.$$

These equations imply

$$\limsup_T \sup_{||\tilde{\alpha} - \alpha|| \le \delta} \left| \frac{1}{T} \sum_{t=1}^{T} (q_t(\tilde{\alpha}) - E q_t(\tilde{\alpha})) - \frac{1}{T} \sum_{t=1}^{T} (q_t(\alpha) - E q_t(\alpha)) \right| < 2\epsilon. \tag{14}$$

For a fixed $\alpha \in D_\alpha$, the sequence $q_t(\alpha)$ obeys a weak LLN (law of large numbers) (i) if $q_t(\alpha)$ is strongly mixing (which is the case if $u_t$ is strongly mixing), and (ii) if for an $\epsilon > 0$, we have $\limsup_T T^{-1} \sum_{t=1}^{T} E|q_t(\alpha)|^{1+\epsilon} < \infty$ (e.g., Pötscher and Prucha, 1997, Thm. 6.3). From the mixing property A.2 follows (i). From $q_t(0) = 0$ in Lemma 1C follows $|q_t(\alpha)| \le \max(\vartheta, 1-\vartheta)|h_t(\alpha)|$ and the $c_r$-inequality (e.g., Davidson, 1994, 9.28 Loève's $c_r$-inequality) implies $|h_t(\alpha)|^{1+\epsilon} = |g(x_t, \beta_0 + \alpha) - g(x_t, \beta_0)|^{1+\epsilon} \le 2^\epsilon |g(x_t, \beta_0 + \alpha)|^{1+\epsilon} + 2^\epsilon |g(x_t, \beta_0)|^{1+\epsilon}$. Together with the dominance condition A.7 this establishes (ii).

The assertion of the lemma follows from (14) and that $q_t(\alpha)$ obeys a LLN using the usual arguments, since $D_\alpha$ is compact and admits a finite covering. ■

LEMMA 3C. *Under Assumptions A.3, A.4, and A.5, for every $\alpha \in D_\alpha$,*

$$E\left[\frac{1}{T}\sum_{t=1}^{T} b_t(\alpha)\right] \ge f_0 \min\left(\left[\frac{1}{4T}\sum_{t=1}^{T} |h_t(\alpha)|\right]^2, \delta_0^2\right).$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

**Proof of Lemma 3C.** Due to Assumption A.3 and taking into account the monotonicity of $F_t(x)$, for all $t$ and all positive $\delta \le \delta_0$,

$$\min\left[F_t(|x|) - F_t(0), F_t(0) - F_t(-|x|)\right] \ge \begin{cases} f_0|x| & \text{for} \quad |x| \le \delta, \\ f_0\delta & \text{for} \quad |x| > \delta. \end{cases}$$

From the definition of $b_t(\alpha)$ follows

$$E\left[b_t(\alpha)\right] = \begin{cases} \int_0^{h_t}(h_t - z)dF_t(z) & \text{for} \quad h_t > 0, \\ \int_{h_t}^0 (z - h_t)dF_t(z) & \text{for} \quad h_t \le 0. \end{cases} \tag{15}$$

By limiting the integration domain in (15) to $[0, h_t/2]$ and $[h_t/2, 0]$, respectively, we obtain

$$\begin{aligned} E\left[\frac{1}{T}\sum_{t=1}^{T} b_t(\alpha)\right] &\ge \frac{f_0}{T}\sum_{t,\,|h_t|\le 2\delta}^{T}\left(\frac{h_t}{2}\right)^2 + \frac{f_0}{T}\sum_{t,\,|h_t|>2\delta}^{T}\left|\frac{h_t}{2}\right|\delta \\ &\ge \frac{f_0\delta}{2T}\sum_{t,\,|h_t|>2\delta}^{T}|h_t| \ge \frac{f_0\delta}{2}\left(\frac{1}{T}\sum_{t=1}^{T}|h_t| - 2\delta\right). \end{aligned}$$

For $T$ large enough, the assertion follows from setting $\delta = \min\left(\delta_0, (4T)^{-1}\sum_{t=1}^{T}|h_t|\right)$. ■

**Proof of Theorem 1.** The assertion is equivalent to $\operatorname{plim}_T \hat{\alpha}_T = 0$, where $\alpha = \beta - \beta_0$. For every fixed event $\omega$ the normalized loss function $Q_T(\alpha)/T$ can be written as

$$\frac{1}{T}Q_T(\alpha) = E\left(\frac{1}{T}Q_T(\alpha)\right) + \left(\frac{1}{T}Q_T(\alpha) - E\frac{1}{T}Q_T(\alpha)\right). \tag{16}$$

Due to $E(C_T(\alpha)) = 0$, following from Assumption A.1 and Lemma 3C,

$$E\left(\frac{1}{T}Q_T(\alpha)\right) \ge f_0 \min\left(\left[\frac{1}{4T}\sum_{t=1}^{T}|h_t(\alpha)|\right]^2, \delta_0^2\right). \tag{17}$$

The latter and Assumption A.6 imply for $||\alpha|| \ge \epsilon$, $T$ large enough, and an arbitrary $\epsilon > 0$

$$E\left(\frac{1}{T}Q_T(\alpha)\right) \ge f_0 \min\left(\frac{1}{16}\delta_\epsilon^2, \delta_0^2\right) \stackrel{\text{def}}{=} \eta > 0. \tag{18}$$

According to Lemma 2C, for $T \to \infty$,

$$P\left(\sup_{\alpha\in D_\alpha}\left|\frac{1}{T}Q_T(\alpha) - E\frac{1}{T}Q_T(\alpha)\right| \le \frac{\eta}{2}\right) \to 1. \tag{19}$$

Then, from (16) and under consideration of (18) and (19), for $T \to \infty$ and $||\alpha|| \ge \epsilon$,

$$P\left(\frac{1}{T}Q_T(\alpha) \ge \frac{\eta}{2}\right) \to 1. \tag{20}$$

However, due to $Q_T(\hat{\alpha}_T) \le Q_T(0) = 0$, from (20) follows

$$\lim_{T\to\infty} P\left(||\hat{\alpha}_T|| < \epsilon\right) = 1,$$

for an arbitrary $\epsilon > 0$. ■

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

# APPENDIX: Asymptotic Normality

As a first prerequisite for our asymptotic normality considerations in Lemma 1N we will prove a basic result on uniform convergence in probability, inspired by the various works of Andrews and Pötscher and Prucha on this subject. Its implications will be used repeatedly.

For every $T = 1, 2, \ldots$ and every $\gamma \in \mathcal{C} \subset \mathbb{R}^K$, where $\mathcal{C}$ is compact, let $R_T(\omega, \gamma)$ be random variables with existing expectation $ER_T(\omega, \gamma)$. For each $\gamma \in \mathcal{C}$, we define the open balls $B(\gamma, \rho) = \{\alpha \,|\, \alpha \in \mathcal{C}, ||\gamma - \alpha|| < \rho\}$ with $\rho > 0$. Then, for every $\rho$ we can choose finitely many $\gamma(i, \rho) \in \mathcal{C}$, $i = 1, 2, \ldots, n(\rho)$, admitting the finite covering

$$\mathcal{C} \subset \bigcup_{i=1}^{n(\rho)} B(\gamma(i, \rho), \rho). \tag{21}$$

Further we assume for all $T$, every sufficiently small $\rho$, and the corresponding covering (21), the existence of random variables $\underline{R}_T(i, \rho)$ and $\overline{R}_T(i, \rho)$ for $1 \le i \le n(\rho)$ such that

$$\underline{R}_T(i, \rho) \le R_T(\omega, \gamma) \le \overline{R}_T(i, \rho) \qquad \text{for all } \gamma \in B(\gamma(i, \rho), \rho), \tag{22}$$

$$\operatorname*{plim}_T \left(\underline{R}_T(i, \rho) - E\underline{R}_T(i, \rho)\right) = 0, \quad \operatorname*{plim}_T \left(\overline{R}_T(i, \rho) - E\overline{R}_T(i, \rho)\right) = 0, \tag{23}$$

and

$$\lim_{\rho \to 0} \limsup_T \max_{1 \le i \le n(\rho)} \left|E\overline{R}_T(i, \rho) - E\underline{R}_T(i, \rho)\right| = 0. \tag{24}$$

LEMMA 1N. *Then,*

$$\operatorname*{plim}_T \sup_{\gamma \in \mathcal{C}} \left(R_T(\omega, \gamma) - ER_T(\omega, \gamma)\right) = 0. \tag{25}$$

**Proof of Lemma 1N.** From (22) follows for all $T$, every sufficiently small $\rho$, and $1 \le i \le n(\rho)$,

$$\begin{aligned}
&\underline{R}_T(i, \rho) - E\underline{R}_T(i, \rho) + \left(E\underline{R}_T(i, \rho) - ER_T(\omega, \gamma)\right) \le R_T(\omega, \gamma) - ER_T(\omega, \gamma), \\
&\overline{R}_T(i, \rho) - E\overline{R}_T(i, \rho) + \left(E\overline{R}_T(i, \rho) - ER_T(\omega, \gamma)\right) \ge R_T(\omega, \gamma) - ER_T(\omega, \gamma),
\end{aligned} \tag{26}$$

for all $\gamma \in B(\gamma(i, \rho))$. Thus, due to (24), for every $\epsilon > 0$ and $T$ sufficiently large, there exists a $\rho > 0$ such that

$$\begin{aligned}
&\min_i \left(\underline{R}_T(i, \rho) - E\underline{R}_T(i, \rho)\right) - \epsilon \le R_T(\omega, \gamma) - ER_T(\omega, \gamma), \\
&\max_i \left(\overline{R}_T(i, \rho) - E\overline{R}_T(i, \rho)\right) + \epsilon \ge R_T(\omega, \gamma) - ER_T(\omega, \gamma),
\end{aligned} \tag{27}$$

for all $\gamma \in \mathcal{C}$. Then the assertion follows from (27) and (23). ■

As a next step in Lemma 2N we study the first order conditions resulting from the directional derivatives of the loss function (4).

LEMMA 2N. *In the model (1), the following assertion holds under Assumption A.9: If there exists an $\alpha = \hat{\alpha}_T$ such that*

$$A_{lT}(\alpha, w) \le S_T(\alpha, w) \le A_{uT}(\alpha, w), \tag{28}$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

*where*

$$A_{lT}(\alpha, w) \stackrel{def}{=} -\sum_{t=1}^{T} I[u_t = h_t(\alpha)] \left| w' \nabla h_t(\alpha) \right| I\left[w' \nabla h_t(\alpha) < 0\right],$$

$$S_T(\alpha, w) \stackrel{def}{=} \sum_{t=1}^{T} w' \nabla h_t(\alpha) \left(I[u_t \leq h_t(\alpha)] - \vartheta\right),$$

$$A_{uT}(\alpha, w) \stackrel{def}{=} \sum_{t=1}^{T} I[u_t = h_t(\alpha)] \left| w' \nabla h_t(\alpha) \right| I\left[w' \nabla h_t(\alpha) \geq 0\right],$$

*holds for all* $w \in \mathbb{R}^K$ *with* $||w|| = 1$, *then* $\hat{\alpha}_T$ *is a local minimum of* $Q_T(\alpha)$.

**Proof of Lemma 2N.** In the following we calculate the derivative of the loss function in direction $w$,

$$\lim_{s \to 0} \frac{Q_T(\alpha + sw) - Q_T(\alpha)}{|s|}. \tag{29}$$

By choosing $w$ as usual as the $i$th unit vector $e_i$, and $s > 0$, we get the partial derivative with respect to $\beta_i$. The more general argumentation employed here has the advantage to avoid the use of the index $i$.

We calculate (29) by analyzing the summands $q_t(\alpha)$ defined in equation (4). For convenience of notation we define $m_t \stackrel{\text{def}}{=} \text{sign}(s) w' \nabla h_t(\alpha)$. From Assumption A.9 follows for $h_t(\alpha) \neq u_t$

$$\lim_{s \to 0} \frac{q_t(\alpha + sw) - q_t(\alpha)}{|s|} = m_t \left(I[u_t \leq h_t(\alpha)] - \vartheta\right), \tag{30}$$

and, for $h_t(\alpha) = u_t$,

$$\lim_{s \to 0} \frac{q_t(\alpha + sw) - q_t(\alpha)}{|s|} = m_t \left((1 - \vartheta) I[m_t > 0] - \vartheta\, I[m_t \leq 0]\right). \tag{31}$$

The right hand side of (31) can be written as $m_t \left(I[m_t > 0] - \vartheta\right)$. Thus, from (30) and (31) follows that $Q_T(\alpha)$ has a local minimum if, for all $w$,

$$\sum_{t=1}^{T} I[u_t \neq h_t(\alpha)] m_t \left(I[u_t \leq h_t(\alpha)] - \vartheta\right) + \sum_{t=1}^{T} I[u_t = h_t(\alpha)] m_t \left(I[m_t > 0] - \vartheta\right) \geq 0. \tag{32}$$

By selecting $s > 0$ and then $s < 0$, from (32) follows

$$\begin{aligned} & -\sum_{t=1}^{T} I[u_t = h_t(\alpha)] w' \nabla h_t(\alpha) \left(I\left[w' \nabla h_t(\alpha) > 0\right] - \vartheta\right) \\ \leq & \sum_{t=1}^{T} I[u_t \neq h_t(\alpha)] w' \nabla h_t(\alpha) \left(I[u_t \leq h_t(\alpha)] - \vartheta\right) \\ \leq & -\sum_{t=1}^{T} I[u_t = h_t(\alpha)] w' \nabla h_t(\alpha) \left(I\left[w' \nabla h_t(\alpha) < 0\right] - \vartheta\right). \end{aligned} \tag{33}$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

From adding $\sum_{t=1}^{T} I[u_t = h_t(\alpha)]w'\nabla h_t(\alpha)(1-\vartheta)$ to all three sides of (33) follows the assertion. ■

Whenever the bounds $A_{lT}(\alpha, w)$ and $A_{uT}(\alpha, w)$ vanish, the first order conditions have the usual form. However, due to $A_{lT}(\alpha, w) \leq 0$ and $A_{uT}(\alpha, w) \geq 0$, both limits vanish if and only if

$$A_T(\alpha, w) \stackrel{\text{def}}{=} A_{uT}(\alpha, w) - A_{lT}(\alpha, w) = \sum_{t=1}^{T} \left| w'\nabla h_t(\alpha) \right| I[u_t = h_t(\alpha)] \tag{34}$$

vanishes. Note that it suffices to prove that $\operatorname{plim}_T A_T(\alpha, w) = 0$.

**Proof of Theorem 2.** The proof of Theorem 2 can be split up in three lemmas. Next, we employ Lemma 1N to study the lower and upper limits of the rendered first order conditions (28).

As mentioned above the first order conditions (28) are multiplied by $\varphi(T)/T$ and in Lemma 3N we want to show for $\lim_T \varphi(T)^2/T = 0$ that

$$\operatorname*{plim}_T \sup_{||\gamma|| \leq m} \frac{\varphi(T)}{T} A_T\left(\frac{\gamma}{\varphi(T)}, w\right) = 0. \tag{35}$$

LEMMA 3N. *If* $\lim_T \varphi(T)^2/T = 0$, *then (35) holds for all* $w$ *with* $||w|| = 1$.

**Proof of Lemma 3N.** For a given $\rho > 0$ we consider a finite covering with balls $B(\gamma(i,\rho), \rho)$, $1 \leq i \leq n(\rho)$, for $\{\gamma \,|\, ||\gamma|| \leq m\}$. Then Lemma 1N implies that for this covering we have to find lower and upper bounds $\underline{R}_T(i,\rho)$ and $\overline{R}_T(i,\rho)$ for $[\varphi(T)/T]A_T(\gamma/\varphi(T), w)$ and to verify conditions (23) and (24), respectively.

According to the definition of $A_T(\gamma/\varphi(T), w)$ let

$$R_T(\omega, \gamma) = \frac{\varphi(T)}{T} \sum_{t=1}^{T} \left| w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right) \right| I\left[u_t = h_t\left(\frac{\gamma}{\varphi(T)}\right)\right], \tag{36}$$

and note that Assumption A.13 implies $E R_T(\omega, \gamma) = 0$.

Due to Assumption A.9 for all $T$, $1 \leq t \leq T$, all $w$ with $||w|| = 1$, a sufficiently small $\rho$ and every $\gamma(i,\rho)$, where $1 \leq i \leq n(\rho)$, there exists a $\overline{\gamma} \in \overline{B}(\gamma(i,\rho), \rho)$, where $\overline{B}$ denotes the closure of $B$, such that

$$\left| w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right) \right| \leq \left| w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) \right|,$$

for all $\gamma \in B(\gamma(i,\rho), \rho)$. Analogously there exist $\gamma_* \in \overline{B}(\gamma(i,\rho), \rho)$ and $\gamma^* \in \overline{B}(\gamma(i,\rho), \rho)$ such that

$$h_t\left(\frac{\gamma_*}{\varphi(T)}\right) \leq h_t\left(\frac{\gamma}{\varphi(T)}\right) \leq h_t\left(\frac{\gamma^*}{\varphi(T)}\right), \tag{37}$$

for all $\gamma \in B(\gamma(i,\rho), \rho)$. Note that $\overline{\gamma}$ depends on $w$, $T$, $\rho$, $i$, and $t$, while $\gamma_*$ and $\gamma^*$ depend on $T$, $\rho$, $i$, and $t$, respectively. Now define

$$\overline{R}_T(i,\rho) = \frac{\varphi(T)}{T} \sum_{t=1}^{T} \left| w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) \right| I\left[h_t\left(\frac{\gamma_*}{\varphi(T)}\right) \leq u_t \leq h_t\left(\frac{\gamma^*}{\varphi(T)}\right)\right], \tag{38}$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

and $\underline{R}_T(i,\rho) = 0$. Then, for all $w$, $T$, $i$, and $\rho$,

$$\underline{R}_T(i,\rho) \le R_T(\omega,\gamma) \le \overline{R}_T(i,\rho), \tag{39}$$

for all $\gamma \in B(\gamma(i,\rho),\rho)$. From the mixing assumption A.15 follows (see Doukhan, 1994, Lemma 3, p. 10)

$$Var\left(\overline{R}_T(i,\rho)\right) \le \frac{\varphi(T)^2}{T^2}\sum_{t=1}^{T}\left(w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right)\right)^2 8\sum_{k=0}^{\infty}\mu_k. \tag{40}$$

From Assumptions A.10 and A.15 follows that the right hand side of (40) converges to 0 for sufficiently small $\rho$, and $1 \le i \le n(\rho)$, for $\lim_T \varphi(T)^2/T = 0$, establishing condition (23) of Lemma 1N.

It remains to verify condition (24). For all $T$, $i$, $\rho$, and $w$ with $||w|| = 1$,

$$E\overline{R}_T(i,\rho) = \frac{\varphi(T)}{T}\sum_{t=1}^{T}\left|w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right)\right| d_t(\gamma^*,\gamma_*), \tag{41}$$

where we use the abbreviation

$$d_t(\gamma^*,\gamma_*) \stackrel{\text{def}}{=} F_t\left(h_t\left(\frac{\gamma^*}{\varphi(T)}\right)\right) - F_t\left(h_t\left(\frac{\gamma_*}{\varphi(T)}\right)\right). \tag{42}$$

From Assumptions A.13 and A.9 and a Taylor expansion with remainder of (42) follows an upper bound for (41) given by

$$\frac{1}{T}\sum_{t=1}^{T}\left|w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right)\right| f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\left|\nabla h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)'(\gamma^* - \gamma_*)\right|, \tag{43}$$

where $\tilde{\gamma}$ lies in between $\gamma_*$ and $\gamma^*$. Then, due to Assumptions A.10 and A.14 using the Cauchy-Schwartz-inequality and $||\gamma^* - \gamma_*|| \le \rho$, condition (24) follows from (43). Hence the assertion is proved. ■

As a next step we split up the middle part of the first order conditions (28) according to

$$S_T\left(\frac{\gamma}{\varphi(T)}, w\right) = \left(S_T\left(\frac{\gamma}{\varphi(T)}, w\right) - ES_T\left(\frac{\gamma}{\varphi(T)}, w\right)\right) + ES_T\left(\frac{\gamma}{\varphi(T)}, w\right). \tag{44}$$

In Lemma 4N we analyze the second term on the right hand side of (44).

LEMMA 4N. *For* $|\gamma|| \le m$,

$$E\frac{\varphi(T)}{T}S_T\left(\frac{\gamma}{\varphi(T)}, w\right) = \frac{1}{T}\sum_{t=1}^{T} w'\nabla h_t(0)\, f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\nabla h_t(0)'\gamma,$$

*where* $\tilde{\gamma}$ *is in between 0 and* $\gamma$.

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

**Proof of Lemma 4N.** From the definition of $S_T(\alpha, w)$ in (28) then follows for $||\gamma|| \leq m$

$$E\frac{\varphi(T)}{T}S_T\left(\frac{\gamma}{\varphi(T)}, w\right) = \frac{\varphi(T)}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right)\left(F_t\left(h_t\left(\frac{\gamma}{\varphi(T)}\right)\right) - F_t(0)\right). \quad \textbf{(45)}$$

From Assumptions A.13 and A.9 and a Taylor expansion with remainder, follows for the right hand side of (45), in analogy to (43),

$$\frac{1}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right) f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\nabla h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)'\gamma, \quad \textbf{(46)}$$

where $\tilde{\gamma}$ is in between 0 and $\gamma$. As a first step we show that (46) is asymptotically equivalent to

$$\frac{1}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right) f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\nabla h_t(0)'\gamma, \quad \textbf{(47)}$$

which follows from proving that

$$\lim_T \frac{1}{T}\sum_{t=1}^{T}\left\|\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right)\right\| f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\left\|\nabla h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right) - \nabla h_t(0)\right\| = 0.$$

The latter follows from application of the Cauchy-Schwartz inequality and under consideration of Assumptions A.9, A.10, A.11, and A.14, and $\lim_T \varphi(T) = \infty$. The second step consists of analogously showing that (47) is asymptotically equivalent to the expression in the assertion. ■

In Lemma 5N we analyze the first term on the right hand side of (44).

LEMMA 5N.

$$\operatorname*{plim}_T \sup_{||\gamma||\leq m} \frac{\varphi(T)}{T}\left(S_T\left(\frac{\gamma}{\varphi(T)}, w\right) - ES_T\left(\frac{\gamma}{\varphi(T)}, w\right)\right) = 0. \quad \textbf{(48)}$$

**Proof of Lemma 5N.** Again we employ Lemma 1N. In order to match $S_T(\gamma/\varphi(T), w)$ we define $R_T(\omega, \gamma)$ from Lemma 1N according to

$$R_T(\omega, \gamma) = \frac{\varphi(T)}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right) I\left[u_t \leq h_t\left(\frac{\gamma}{\varphi(T)}\right)\right]. \quad \textbf{(49)}$$

Again, for a given $\rho > 0$ we consider a finite covering with balls $B(\gamma(i,\rho), \rho)$, $1 \leq i \leq n(\rho)$, for $\{\gamma \,|\, ||\gamma|| \leq m\}$. Due to Assumption A.9 for all $T$, $1 \leq t \leq T$, all $w$ with $||w|| = 1$, for a sufficiently small $\rho$ and every $\gamma(i,\rho)$, where $1 \leq i \leq n(\rho)$, there exist $\underline{\gamma}$ and $\overline{\gamma}$, both from $\overline{B}(\gamma(i,\rho), \rho)$, such that

$$w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) \leq w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right) \leq w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right),$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

for all $\gamma \in B(\underline{\gamma}(i,\rho),\rho)$. Analogously (37) holds for all $\gamma \in B(\overline{\gamma}(i,\rho),\rho)$. Now define $\underline{R}_T(i,\rho) = [\varphi(T)/T]\sum_{t=1}^{T} \underline{r}_t(i,\rho)$ and $\overline{R}_T(i,\rho) = [\varphi(T)/T]\sum_{t=1}^{T} \overline{r}_t(i,\rho)$, where

$$\underline{r}_t(i,\rho) \stackrel{\text{def}}{=} w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) I\left[u_t \le h_t\left(\frac{\gamma_*}{\varphi(T)}\right)\right] \quad \text{for } w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) > 0, \tag{50}$$

$$\underline{r}_t(i,\rho) \stackrel{\text{def}}{=} w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) I\left[u_t \le h_t\left(\frac{\gamma^*}{\varphi(T)}\right)\right] \quad \text{for } w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) \le 0, \tag{51}$$

$$\overline{r}_t(i,\rho) \stackrel{\text{def}}{=} w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) I\left[u_t \le h_t\left(\frac{\gamma^*}{\varphi(T)}\right)\right] \quad \text{for } w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) > 0, \tag{52}$$

$$\overline{r}_t(i,\rho) \stackrel{\text{def}}{=} w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) I\left[u_t \le h_t\left(\frac{\gamma_*}{\varphi(T)}\right)\right] \quad \text{for } w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) \le 0. \tag{53}$$

Then for all $i$, where $1 \le i \le n(\rho)$, and all $T$, inequality (39) holds for all $\gamma \in B(\underline{\gamma}(i,\rho),\rho)$. Again we verify conditions (23) and (24) of Lemma 1N. Both $\{\omega \mid u_t \le h_t(\gamma_*/\varphi(T))\}$ and $\{\omega \mid u_t \le h_t(\gamma^*/\varphi(T))\}$ belong to the $\sigma$-algebra generated by $u_t$. Hence, in analogy to the considerations concerning (40), for $\lim_T \varphi(T)^2/T = 0$ and all $i$, the variances of $\underline{R}_T(i,\rho)$ and $\overline{R}_T(i,\rho)$ vanish asymptotically and (23) is established. Next we consider $E\overline{R}_T(i,\rho) - E\underline{R}_T(i,\rho)$. According to the definitions (50)-(53) this difference can be written as

$$\frac{\varphi(T)}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) F_t\left(h_t\left(\frac{\lambda(t)}{\varphi(T)}\right)\right) - \frac{\varphi(T)}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) F_t\left(h_t\left(\frac{\kappa(t)}{\varphi(T)}\right)\right), \tag{54}$$

where $\lambda(t) \in \{\gamma_*, \gamma^*\}$ and $\kappa(t) \in \{\gamma_*, \gamma^*\}$. Expression (54) can be split up according to

$$\frac{\varphi(T)}{T}\sum_{t=1}^{T} w'\left(\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) - \nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right)\right) F_t\left(h_t\left(\frac{\lambda(t)}{\varphi(T)}\right)\right) + \frac{\varphi(T)}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right)\left(F_t\left(h_t\left(\frac{\lambda(t)}{\varphi(T)}\right)\right) - F_t\left(h_t\left(\frac{\kappa(t)}{\varphi(T)}\right)\right)\right). \tag{55}$$

Again, from Assumptions A.13, A.9, and due to (42), we get an upper bound for the absolute value of (55) by

$$\frac{1}{T}\sum_{t=1}^{T}\left|w'\nabla^2 h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)(\overline{\gamma}-\underline{\gamma})\right| + \frac{\varphi(T)}{T}\sum_{t=1}^{T}\left|w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) d_t(\lambda(t),\kappa(t))\right|, \tag{56}$$

where $\tilde{\gamma}$ lies in between $\underline{\gamma}$ and $\overline{\gamma}$. Due to Assumption A.11 and analogously to (43), expression (56) vanishes asymptotically for $\rho \to 0$, completing the assertion. ■

So far we have established the following:

If $\varphi(T)$ grows slower than $\sqrt{T}$, that is $\lim_T \varphi(T)^2/T = 0$, then follows for $|\gamma\,|| \le m$ from using the first order conditions in (28) in Lemma 2N, the decomposition in (44), and Lemmas 3N–5N,

$$o_p(1) \le \frac{1}{T}\sum_{t=1}^{T} w'\nabla h_t(0)\, f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\nabla h_t(0)'\gamma \le o_p(1),$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

where $\tilde{\gamma}$ lies in between 0 and $\gamma$. Thus, for choosing the $K$ unit vectors for $\omega$ and for $||\gamma|| \leq m$, the first order conditions can be written as

$$\frac{1}{T}\sum_{t=1}^{T} \nabla h_t(0)\, f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right) \nabla h_t(0)'\gamma = o_p(1). \tag{57}$$

By construction the estimator $\hat{\gamma}_T$ is a solution of (57) for all $\omega \in \Omega_{T,m} = \{\omega | \, ||\hat{\gamma}_T(\omega)|| \leq m\}$. Since due to Assumption A.16 for all $\gamma(t)$ with $||\gamma(t)|| \leq m$ the matrix

$$\frac{1}{T}\sum_{t=1}^{T} \nabla h_t(0)\, f_t\left(h_t\left(\frac{\gamma(t)}{\varphi(T)}\right)\right) \nabla h_t(0)' \tag{58}$$

is asymptotically nonsingular, from (57) follows that a solution of the first order conditions of $Q_T(\alpha)$ exists for $\omega \in \Omega_{T,m}$ and for every $\epsilon > 0$ there exists an $m$ such that $\lim_T P(\Omega_{T,m}) \geq 1 - \epsilon$. Then the proof of Theorem 2 is complete. ■

**Proof of Theorem 3.** The proof of Theorem 3 can be split up in four lemmas.

The starting point of our considerations is the first order conditions (28) multiplied by $1/\sqrt{T}$ and we have to show that the assertion of Lemma 3N holds when $\varphi(T)$ is replaced by $\sqrt{T}$.

LEMMA 6N.

$$\operatorname*{plim}_T \sup_{||\gamma|| \leq m} \frac{1}{\sqrt{T}} A_T\left(\frac{\gamma}{\sqrt{T}}, w\right) = 0. \tag{59}$$

**Proof of Lemma 6N.** In the proof of Lemma 3N, the factor $\varphi(T)$ played an important role for calculating the variance in (40). Again following Doukhan (1994, Thm. 3, p. 9, setting $p = q$) and due to Assumption A.15, instead of (40) we get for any $p > 1$

$$Var\left(\overline{R}_T(i,\rho)\right) \leq \frac{1}{T}\sum_{t=1}^{T}\left(w'\nabla h_t\left(\frac{\overline{\gamma}}{\sqrt{T}}\right)\right)^2 d_t\left(\gamma^*, \gamma_*\right)^{2/p} 16 \sum_{k=0}^{\infty} \mu_k^{(p-2)/p}, \tag{60}$$

where $d_t(\gamma^*, \gamma_*)$ is defined in analogy to (42) and $\varphi(T) = \sqrt{T}$. For every $\eta > 0$ the sum $\sum_{k=1}^{\infty} k^{-1-\eta}$ converges. Assumption A.15 implies the existence of an $\eta > 0$ such that $\mu_k = O(k^{-1-\eta})$. Hence $\sum_{k=0}^{\infty} \mu_k^{(p-2)/p}$ converges if $(1+\eta)(p-2)/p > 1$. We can always find such a $p$ for a fixed $\eta > 0$. Hence, the sum over $k$ in (60) converges for such a $p$.

Further, from a Taylor expansion with remainder we get (see (42) and (43))

$$d_t\left(\gamma^*, \gamma_*\right) = \frac{1}{\sqrt{T}} f_t\left(h_t\left(\frac{\tilde{\gamma}}{\sqrt{T}}\right)\right) \nabla h_t\left(\frac{\tilde{\gamma}}{\sqrt{T}}\right)'\left(\gamma^* - \gamma_*\right), \tag{61}$$

where $\tilde{\gamma}$ lies in between $\gamma_*$ and $\gamma^*$. Then, due to Assumptions A.14 and A.12, expression (61) is bounded from above by

$$\sup_{1 \leq t \leq T} d_t\left(\gamma^*, \gamma_*\right) = o(1). \tag{62}$$

Together with Assumption A.10, from (60) and (62) follows that $\lim_T Var\left(\overline{R}_T(i,\rho)\right) = 0$ for sufficiently small $\rho$, all $T$, and $1 \leq i \leq n(\rho)$. It remains to establish the condition

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

analogous to (41) using $\sqrt{T}$ instead of $\varphi(T)$. As (43) also holds using $\sqrt{T}$ instead of $\varphi(T)$, the assertion is proved. ■

Now decompose $S_T\left(\gamma/\sqrt{T}, w\right)$ defined in the first order conditions (28) into the following four sums:

$$ES_T\left(\frac{\gamma}{\sqrt{T}}, w\right) = \sum_{t=1}^{T} w' \nabla h_t\left(\frac{\gamma}{\sqrt{T}}\right)\left(F_t\left(h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right) - \vartheta\right),$$

$$S_{T11}\left(\frac{\gamma}{\sqrt{T}}, w\right) = \sum_{t=1}^{T} w'\left(\nabla h_t\left(\frac{\gamma}{\sqrt{T}}\right) - \nabla h_t(0)\right) \times \left(I\left[u_t \le h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right] - F_t\left(h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right)\right),$$

$$S_{T12}\left(\frac{\gamma}{\sqrt{T}}, w\right) = \sum_{t=1}^{T} w' \nabla h_t(0) \times \left(I\left[u_t \le h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right] - I[u_t \le 0] - F_t\left(h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right) + \vartheta\right),$$

$$S_{T2}(w) = \sum_{t=1}^{T} w' \nabla h_t(0)\left(I[u_t \le 0] - \vartheta\right).$$

In Lemma 4N, $ES_T(\gamma/\varphi(T), w)$ has been analyzed. Now, we study the first sum in the decomposition above, $ES_T\left(\gamma/\sqrt{T}, w\right)$.

LEMMA 7N. *For* $||\gamma|| \le m$, $(1/\sqrt{T})ES_T(\gamma/\sqrt{T}, w)$ *is asymptotically equivalent to*

$$\frac{1}{T}\sum_{t=1}^{T} w' \nabla h_t(0) f_t(0) \nabla h_t(0)' \gamma. \tag{63}$$

**Proof of Lemma 7N.** Starting point of our considerations is Lemma 4N valid also for choosing $\varphi(T) = \sqrt{T}$. To begin with we establish that $h_t\left(\gamma/\sqrt{T}\right)$ converges to $h_t(0) = 0$ uniformly in $t$. This follows from a Taylor expansion with remainder, $h_t\left(\gamma/\sqrt{T}\right) = h_t(0) + \left(1/\sqrt{T}\right)\nabla h_t\left(\tilde{\gamma}/\sqrt{T}\right)'\gamma$, where $\tilde{\gamma}$ lies in between 0 and $\gamma$, and Assumption A.12. Then the assertion follows from Assumptions A.10 and A.13. ■

In the following two Lemmas 8N and 9N we analyze the second and third sum in the decomposition above. First we study $S_{T11}\left(\gamma/\sqrt{T}, w\right)$.

LEMMA 8N.

$$\operatorname*{plim}_{T} \sup_{||\gamma|| \le m} \frac{1}{\sqrt{T}} S_{T11}\left(\frac{\gamma}{\sqrt{T}}, w\right) = 0. \tag{64}$$

**Proof of Lemma 8N.** In contrast to (49) in the proof of Lemma 5N we now analyze

$$R_T(\omega, \gamma) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} w'\left(\nabla h_t\left(\frac{\gamma}{\sqrt{T}}\right) - \nabla h_t(0)\right) I\left[u_t \le h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right] \tag{65}$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

and note that $w'\nabla h_t(0)$ does not depend on $\gamma$. The remainder of the proof is similar to that of Lemma 5N. From Assumption A.10 and the triangle inequality follows

$$\limsup_T \sup_{||\gamma||\le m} \frac{1}{T}\sum_{t=1}^{T}\left|\left|\nabla h_t\left(\frac{\gamma}{\sqrt{T}}\right)-\nabla h_t(0)\right|\right|^2 < \infty.$$

Hence, the variances of $\underline{R}_T(i,\rho)$ and $\overline{R}_T(i,\rho)$ converge to 0 as $T\to\infty$, for sufficiently small $\rho$, all $T$, and $1\le i\le n(\rho)$, respectively. For the analysis of $E\overline{R}_T(i,\rho)-E\underline{R}_T(i,\rho)$ we are careful to note that

$$w'\left(\nabla h_t\left(\frac{\overline{\gamma}}{\sqrt{T}}\right)-\nabla h_t(0)\right)-w'\left(\nabla h_t\left(\frac{\underline{\gamma}}{\sqrt{T}}\right)-\nabla h_t(0)\right)$$
$$= w'\left(\nabla h_t\left(\frac{\overline{\gamma}}{\sqrt{T}}\right)-\nabla h_t\left(\frac{\underline{\gamma}}{\sqrt{T}}\right)\right).$$

The proof of the assertion follows in analogy to the line of reasoning employed in the proof of Lemma 5N and is left to the reader. ■

LEMMA 9N.

$$\operatorname*{plim}_T \sup_{||\gamma||\le m}\frac{1}{\sqrt{T}}S_{T12}\left(\frac{\gamma}{\sqrt{T}},w\right)=0. \tag{66}$$

**Proof of Lemma 9N.** In contrast to (49) in the proof of Lemma 5N we now analyze

$$R_T(\omega,\gamma)=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w'\nabla h_t(0)\left(I\left[u_t\le h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right]-I[u_t\le 0]\right).$$

In analogy to the preceding lemmas we suitably define $\underline{R}_T(i,\rho)$ and $\overline{R}_T(i,\rho)$. Further, in analogy to Lemma 6N we show that for all $\gamma\in\overline{B}(\gamma(i,\rho),\rho)$, $\lim_T Var\left(\underline{R}_T(i,\rho)\right)=0$ and $\lim_T Var\left(\overline{R}_T(i,\rho)\right)=0$. Next, we consider $E\overline{R}_T(i,\rho)-E\underline{R}_T(i,\rho)$ and in analogy to (56), due to $\underline{\gamma}=\overline{\gamma}=0$, we now have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w'\nabla h_t(0)d_t(\lambda(t),\kappa(t)), \tag{67}$$

where $\lambda(t)\in\{\gamma_*,\gamma^*\}$ and $\kappa(t)\in\{\gamma_*,\gamma^*\}$ and $d_t(\lambda(t),\kappa(t))$ is defined in analogy to (42). Note that $w'\nabla h_t(0)I[u_t\le 0]$ does not depend on $\gamma$. Again, the proof of the assertion follows in analogy to the line of reasoning employed in the proof of Lemma 6N. ■

The proof of Theorem 3 so far can be summarized as follows: Due to the decomposition of $S_T\left(\gamma/\sqrt{T},w\right)$ (see the considerations after the proof of Lemma 6N) follows from the assertions in Lemmas 6N, 8N, 9N, and the first order conditions (28),

$$o_p(1)\le\frac{1}{\sqrt{T}}ES_T\left(\gamma/\sqrt{T},w\right)+\frac{1}{\sqrt{T}}S_{T2}(w)\le o_p(1).$$

According to Lemma 7N and the definition of $(1/\sqrt{T})S_{T2}(w)=w'\xi_T$ (see Assumption A.18) we have for $||\gamma||\le m$,

$$o_p(1)\le\frac{1}{T}\sum_{t=1}^{T}w'\nabla h_t(0)f_t(0)\nabla h_t(0)'\gamma+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}w'\nabla h_t(0)(I[u_t\le 0]-\vartheta)\le o_p(1).$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

Choosing the $K$ unit vectors for $w$ implies for $\omega \in \Omega_{T,m} = \{\omega | ||\hat{\gamma}_T(\omega)|| \leq m\}$ (see Assumption A.17)

$$o_p(1) \leq V_T \hat{\gamma}_T + \xi_T \leq o_p(1),$$

or

$$\hat{\gamma}_T = -V_T^{-1} \xi_T + o_p(1), \tag{68}$$

where $V_T = T^{-1} \sum_{t=1}^{T} \nabla h_t(0) f_t(0) \nabla h_t(0)'$ and $\lim_T V_T = V_0$ as defined in Assumption A.17.

Since $\xi_T$ is independent from $\gamma$, from Assumption A.18 follows that for every $\epsilon > 0$ there exists an $m$ such that for sufficiently large $T$, $P(\Omega_{T,m}) \geq 1 - \epsilon$. Hence, from the CLT in Assumption A.18 follows that $\hat{\gamma}_T$ converges in distribution to a normal distribution with mean zero and covariance matrix $V_0^{-1} \Sigma_0 V_0^{-1}$, which is the assertion of Theorem 3. ■

# APPENDIX: Consistent Estimator of Asymptotic Covariance Matrix

**Proof of Theorem 4.** From the definition of $\xi_T$ in Assumption A.18 follows

$$\lim_T E \xi_T \xi_T' = \Sigma_0$$

and

$$E \xi_T \xi_T' = T^{-1} \sum_{s,t=1}^{T} \nabla h_s(0) \omega_{s,t} \nabla h_t(0)', \tag{69}$$

with $\omega_{s,t} = F_{s,t}(0,0) - \vartheta^2$ for $s \neq t$, $\omega_{t,t} = \vartheta(1-\vartheta)$, where $F_{s,t}(z,w)$ is the common distribution of $(u_s, u_t)$ for $s \neq t$. Further due to Assumption A.17 follows $\lim_T V_T = V_0$, where

$$V_T = T^{-1} \sum_{t=1}^{T} \nabla h_t(0) f_t(0) \nabla h_t(0)'. \tag{70}$$

We have to estimate the asymptotic covariance matrix $V_0^{-1} \Sigma_0 V_0^{-1}$. For the estimation of $\Sigma_0$ we employ the right hand side of (69). For the estimation of $V_0$ we employ the right hand side of (70).

The problem is that the vector $\nabla h_t(0) = \nabla g(x_t, \beta_0)$, contained in both $V_0$ and $\Sigma_0$, depends on the unknown parameter vector $\beta_0$. Since $\hat{\beta}_T$ is given, we replace $\beta_0$ according to $\hat{\beta}_T = \beta_0 + \hat{\gamma}_T / \sqrt{T}$ and employ $\nabla h_t(\hat{\gamma}_T / \sqrt{T}) = \nabla g(x_t, \beta_0 + \hat{\gamma}_T / \sqrt{T}) = \nabla g(x_t, \hat{\beta}_T)$ instead of $\nabla h_t(0)$.

Thus the estimate of $\Sigma_0$ is

$$\hat{\Sigma}_T = T^{-1} \sum_{s,t=1}^{T} \nabla h_s \left( \frac{\hat{\gamma}_T}{\sqrt{T}} \right) \omega_{s,t} \nabla h_t \left( \frac{\hat{\gamma}_T}{\sqrt{T}} \right)' \tag{71}$$

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

and the estimate of $V_0$ is

$$\hat{V}_T = T^{-1} \sum_{t=1}^{T} \nabla h_t \left( \frac{\hat{\gamma}_T}{\sqrt{T}} \right) f_t(0) \nabla h_t \left( \frac{\hat{\gamma}_T}{\sqrt{T}} \right)'. \tag{72}$$

We will show that $\operatorname{plim}_T \hat{\Sigma}_T = \Sigma_0$ and $\operatorname{plim}_T \hat{V}_T = V_0$, establishing that $\hat{V}_T^{-1} \hat{\Sigma}_T \hat{V}_T^{-1}$ is a consistent estimate of the covariance matrix $V_0^{-1} \Sigma_0 V_0^{-1}$, where $\hat{V}_T$ is asymptotically nonsingular due to Assumption A.17.

As a first step we have to show that

$$\operatorname*{plim}_T \hat{\Sigma}_T = \operatorname*{plim}_T T^{-1} \sum_{s,t=1}^{T} \nabla h_s \left( \frac{\hat{\gamma}_T}{\sqrt{T}} \right) \omega_{s,t} \nabla h_t(0)'. \tag{73}$$

In order to show that (73) holds, we have to establish it for every element of the $(K \times K)$-matrix on its right hand side. Hence, we use the unit vectors $e_i$ and equation (73) follows if for $1 \le i, j \le K$

$$\operatorname*{plim}_T T^{-1} \sum_{s,t=1}^{T} e_i' \nabla h_s \left( \frac{\hat{\gamma}_T}{\sqrt{T}} \right) \omega_{s,t} \left( \nabla h_t \left( \frac{\hat{\gamma}_T}{\sqrt{T}} \right) - \nabla h_t(0) \right)' e_j = 0. \tag{74}$$

For the present, we consider (73) only for all $\omega \in \Omega_{T,m} = \{\omega \mid ||\hat{\gamma}_T|| \le m\}$. Due to the fact that the sum in (74) is a continuous function of $\hat{\gamma}_T$, there exists a $\gamma^*$ with $||\gamma^*|| \le m$ such that the absolute value of the sum on the left hand side of (74) is maximal for all $\omega \in \Omega_{T,m}$, choosing $\gamma^*$ instead of $\hat{\gamma}_T$. Note that $\gamma^*$ depends on $i$, $j$, and $T$. Hence, instead of (74) it suffices to show that

$$\lim_T T^{-1} \sum_{s,t=1}^{T} \left| e_i' \nabla h_s \left( \frac{\gamma^*}{\sqrt{T}} \right) \omega_{s,t} \left( \nabla h_t \left( \frac{\gamma^*}{\sqrt{T}} \right) - \nabla h_t(0) \right)' e_j \right| = 0. \tag{75}$$

From Assumption A.9 follows

$$\nabla h_t \left( \frac{\gamma^*}{\sqrt{T}} \right) - \nabla h_t(0) = \nabla^2 h_t \left( \frac{\tilde{\gamma}}{\sqrt{T}} \right) \frac{\gamma^*}{\sqrt{T}}$$

where $\tilde{\gamma}$ lies in between 0 and $\gamma^*$. Due to $\omega_{s,t} = P(u_s \le 0, u_t \le 0) - P(u_s \le 0) P(u_t \le 0)$ for $s \ne t$, we define $\mu_0 = \omega_{t,t} = \vartheta(1-\vartheta)$, and Assumption A.15 implies $|\omega_{s,t}| \le \mu_{|s-t|}$. As a consequence and by using the Cauchy-Schwartz inequality it can be shown that the absolute value of the left hand side of (75) is bounded by (see Doukhan, 1994, Lemma 3, p. 10)

$$8 \sum_{k=0}^{\infty} \mu_k \sqrt{T^{-1} \sum_{t=1}^{T} \left( e_i' \nabla h_t \left( \frac{\gamma^*}{\sqrt{T}} \right) \right)^2} \sqrt{T^{-1} \sum_{t=1}^{T} \left( \frac{\gamma^{*\prime}}{\sqrt{T}} \nabla^2 h_t \left( \frac{\tilde{\gamma}}{\sqrt{T}} \right)' e_j \right)^2}.$$

Due to $||\gamma^*|| \le m$, the dominance conditions in Assumptions A.10 and A.11, and the mixing condition in Assumption A.15, this bound is $O\left(1/\sqrt{T}\right)$, establishing (75). Analogously we can show that

$$\lim_T T^{-1} \sum_{s,t=1}^{T} \nabla h_s \left( \frac{\gamma^*}{\sqrt{T}} \right) \omega_{s,t} \nabla h_t(0)' = \lim_T T^{-1} \sum_{s,t=1}^{T} \nabla h_s(0) \omega_{s,t} \nabla h_t(0)',$$

establishing the consistency of $\hat{\Sigma}_T$ for $\Omega_{T,m}$.

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press

As a second step we have to show the consistency of $\hat{V}_T$ for $\Omega_{T,m}$. Our line of reasoning is analogous to the previous considerations (until (75)), for $1 \leq i, j \leq K$ leading to

$$\lim_T T^{-1} \sum_{t=1}^{T} \left| e_i' \nabla h_t \left( \frac{\gamma^*}{\sqrt{T}} \right) f_t(0) \left( \nabla h_t \left( \frac{\gamma^*}{\sqrt{T}} \right) - \nabla h_t(0) \right)' e_j \right| = 0. \tag{76}$$

Again invoking Assumption A.9 we can derive an upper bound of (76)

$$T^{-1} \sum_{t=1}^{T} \left| e_i' \nabla h_t \left( \frac{\gamma^*}{\sqrt{T}} \right) \right| \limsup_s f_s(0) \left| \left( \frac{\tilde{\gamma}}{\sqrt{T}} \right)' \nabla^2 h_t \left( \frac{\gamma}{\sqrt{T}} \right)' e_j \right| \frac{m}{\sqrt{T}},$$

where $\tilde{\gamma}$ lies in between 0 and $\gamma^*$.

From applying the Cauchy-Schwartz inequality follows

$$T^{-1} \sum_{t=1}^{T} \left| e_i' \nabla h_t \left( \frac{\gamma^*}{\sqrt{T}} \right) \right| \left| \left( \frac{\tilde{\gamma}}{\sqrt{T}} \right) \nabla^2 h_t \left( \frac{\gamma}{\sqrt{T}} \right)' e_j \right|$$
$$\leq \sqrt{T^{-1} \sum_{t=1}^{T} \left( e_i' \nabla h_t \left( \frac{\gamma^*}{\sqrt{T}} \right) \right)^2} \sqrt{T^{-1} \sum_{t=1}^{T} \left| \left( \frac{\tilde{\gamma}}{\sqrt{T}} \right)' \nabla^2 h_t \left( \frac{\gamma}{\sqrt{T}} \right)' e_j \right|}.$$

Then equation (76) follows due to Assumptions A.10, A.11, and A.14.

Analogously we can show that

$$\lim_T T^{-1} \sum_{t=1}^{T} \nabla h_t \left( \frac{\gamma^*}{\sqrt{T}} \right) f_t(0) \nabla h_t(0)' = \lim_T T^{-1} \sum_{t=1}^{T} \nabla h_t(0) f_t(0) \nabla h_t(0)'.$$

It remains to verify that the restriction $||\hat{\gamma}_T|| \leq m$ is not substantial for the consistency proof. We keep in mind that $m$ can be chosen arbitrarily large and that $\hat{\gamma}_T$ is asymptotically normal with finite covariance. Thus, for every $\epsilon > 0$ we can find an $m$ and a $T_0$ such that $P(\{\omega \in \Omega_{T,m}\}) \geq 1 - \epsilon$ for all $T \geq T_0$. Hence we have established the consistency of $\hat{V}_T$, $\hat{\Sigma}_T$, and thus the consistency of $\hat{V}_T^{-1} \hat{\Sigma}_T \hat{V}_T^{-1}$ and the assertion of Theorem 4 is shown. ■

https://doi.org/10.1017/S0266466615000031 Published online by Cambridge University Press