ASYMPTOTIC THEORY FOR NONLINEAR QUANTILE REGRESSION UNDER WEAK DEPENDENCE

Walter Oberhofer and Harry Haupt

University of Regensburg and University of Passau

This version: April 16, 2014

Running title: NONLINEAR QUANTILE REGRESSION

Address correspondence to:

Harry Haupt (harry.haupt@uni-passau.de)

Department of Statistics, University of Passau

94030 Passau, Germany.

Abstract: 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 recent survey articles (see inter alia Buchinsky, 1998, and Yu et al., 2003) and econometric textbooks (e.g., Peracchi, 2001, and Wooldridge, 2010). In contrast to the more methodological literature, there are also important, non-technical 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 $(\Omega, \mathcal{F}, P)$ 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

$$y_t - g(x_t, \beta_0) = u_t, \quad 1 \leq t \leq T, \tag{1.1}$$

where $\beta_0 \in D_\beta \subset \mathbb{R}^K$ is a vector of unknown parameters, the $1 \times L$ vectors x_t are deterministic and given, 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 there). 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 et al. (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 a la Koenker and Bassett (1978).

The fixed regressor framework in connection with time series or dependent data has been employed among others in Roussas et al. (1992), Tran et al. (1996), Robinson (1997), and recently in the quantile regression context by 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, Ch. 6) provide a rationale for using mixing conditions in a static model.

Other relevant works in this context include Richardson and Bhattacharyya (1987), Koenker

[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, and 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.

and Park (1994), Jureckova and Prochazka (1994), Powell (1991, 1994), White (1994), and, more recently, Zheng (1998), Mukherjee (1999, 2000), Chernozhukov and Umantsev (2001), Engle and Manganelli (2004) with applications to finance, Kim et al. (2002), 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 et al. (2009) studied copula based nonlinear parametric quantile autoregressions (NLQAR) using similar models as Weiss (1991), 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, and Fitzenberger et al., 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 consistencyof nonlinear regression quantiles and a thorough discussion of the underlying assumptions. In Section 3 we derive the assumptions for asymptotic normality of regression quantiles under weak dependence for nonlinear regression functions. In Section 4 we discuss the consistent estimation of the covariance matrix under dependence and heterogeneity. The Appendix contains proofs of our main theorems and technical lemmas, which may have their own merits.

2 CONSISTENCY

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

$$\sum_{t=1}^{T} \rho_\vartheta(y_t - g(x_t, \beta)), \tag{2.1}$$

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 (1.1) follows that the deviations in (2.1) can be written as

$$y_t - g(x_t, \beta) = u_t + g(x_t, \beta_0) - g(x_t, \beta) = u_t - h_t(\alpha), \tag{2.2}$$

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

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

In order to avoid unnecessary moment requirements, we follow the suggestion of Huber (1967) to replace (2.1) with the equivalent objective function

$$\begin{aligned} Q_T(\alpha) = \sum_{t=1}^{T} q_t(\alpha) \quad &\overset{\text{def}}{=} \quad \sum_{t=1}^{T} [\rho_\vartheta(u_t - h_t(\alpha)) - \rho_\vartheta(u_t)] \\ &= \quad \sum_{t=1}^{T} (h_t(\alpha) - u_t)\,(I[u_t \leq h_t(\alpha)] - \vartheta) + u_t(I[u_t \leq 0] - \vartheta). \end{aligned} \tag{2.3}$$

As $\alpha = 0$ corresponds to β_0, we can study the behavior of the former instead of $\beta = \beta_0$. If an estimator $\hat{\alpha}_T$ results from minimizing (2.3), 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 $|q_t(\alpha)| \leq |h_t(\alpha)|$ (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

$$q_t(\alpha) = b_t(\alpha) + c_t(\alpha), \tag{2.4}$$

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 (2.4) are defined as

$$b_t(\alpha) \overset{\text{def}}{=} |h_t(\alpha) - u_t| \left(I[0 < u_t \leq h_t(\alpha)] + I[h_t(\alpha) < u_t \leq 0] \right), \tag{2.5}$$

$$c_t(\alpha) \overset{\text{def}}{=} -h_t(\alpha)\psi_\vartheta(u_t), \tag{2.6}$$

with $\psi_\vartheta(z) \overset{\text{def}}{=} \vartheta - I[z \leq 0]$ being the right-hand derivative of the check-function $\rho_\vartheta(z)$. By virtue of (2.4) 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 $F_t(0) = P(u_t \leq 0) = \vartheta$, $0 < \vartheta < 1$ for all t.

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 α-mixing (see e.g., Doukhan, 1994, p. 3)

(Assumption A.3)

There exist a positive f_0 and a positive δ_0, such that for all $|x| \leq \delta_0$ and all t,

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

[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.

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 (2.1).

(Assumption A.4)

D_β is compact, β_0 is an inner point of D_β, and $g(z,\beta)$ is continuous in β 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 δ_ϵ such that for all $\beta \in D_\beta$

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

(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

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

THEOREM 1. *In the model (1.1), 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 Theorem 3.10 of Pfanzagl (1969), 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 and Ch. 4.3).

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

It is not necessary for the regressors to be deterministic as postulated in Assumption A.5, as similar behavior can be expected of random regressors $\{x_t\}$ independent of the disturbances $\{u_t\}$. Consider the example of a linear regression function and let $\{x_t\}$ be a stationary sequence with $E(x_t'x_t)$ finite and non-singular. Then, almost all realizations would have the necessary limiting properties (see the discussion in Pollard, 1991).

The identifiable uniqueness condition in Assumption A.6 corresponds to the analogous condition $\liminf_T \inf_{||\tilde{\beta}-\beta|| \geq \epsilon} T^{-1} \sum_{t=1}^T [x_t(\tilde{\beta} - \beta)]^2 = \liminf_T \inf_{||\alpha|| > \epsilon} \alpha' T^{-1} X'X\alpha > 0$ in the linear regression model $y_t = x_t\beta + u_t$ using least squares estimation, implied by the non-singularity 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 has been 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.

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 $|g(x_t, \tilde{\beta}) - g(x_t, \beta)| \leq G_t ||\tilde{\beta} - \beta||$, where $\limsup_T T^{-1} \sum_{t=1}^T G_t < \infty$, and G_t and η may depend on β. Insightful discussions of assumptions commonly used to verify (uniform) stochastic equicontinuity conditions can be found among others in Davidson (1994, Ch. 21.4) and Newey and McFadden (1994, Ch. 7.2), or Andrews (1994b, Ch. 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. Further, we do not require the monotonicity assumption of Jureckova and Prochazka (1994, A.4) and 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 are the first order conditions for a local minimum $\hat{\alpha}_T$ of the loss function $Q_T(\alpha)$ defined in (2.3). In Lemma 2N in the Appendix it is shown that the corresponding first order conditions can be written as inequality

$$A_{lT}(\alpha, w) \leq S_T(\alpha, w) \leq A_{uT}(\alpha, w), \tag{3.1}$$

where the entities in (3.1) are defined in (5.17) in Lemma 2N 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 α_i. Let $\nabla h_t(\alpha)$ denote the $(K \times 1)$-vector with ith 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.

[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.

(Assumption A.10)

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

(Assumption A.11)

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

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

(Assumption A.12)

$\lim_T \sup_{t \le T} \sup_{||\alpha|| \le c} T^{-1/2} ||\nabla h_t(\alpha)|| = 0$.

(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)

$\limsup_t \sup_{||\alpha|| \le c} f_t(h_t(\alpha)) < \infty$.

(Assumption A.15)

The u_t are α-mixing of size -1 with mixing coefficients μ_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 non-singular for sufficiently large T and all $\lambda(t) \in \mathbb{R}^K$ with $||\lambda(t)|| \le c$.

(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 for $T \to \infty$ to the non-singular matrix V_0.

(Assumption A.18)

The $(K \times 1)$-vector $\xi_T = T^{-1/2} \sum_{t=1}^T \nabla h_t(0) \left(I[u_t \le 0] - \vartheta \right)$ converges in distribution to a normal distribution with mean 0 and covariance matrix Σ_0 for $T \to \infty$.

THEOREM 2. *In the model (1.1), 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 the model (1.1), 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

$$\lim_T P(m' < ||\varphi(T)\hat{\alpha}_T|| < m) \geq 1 - \epsilon. \tag{3.2}$$

Condition (3.2) ensures only that asymptotically the distribution of $\varphi(T)\hat{\alpha}_T$ is non-degenerate (that is, does neither vanish nor grow without bound), though the limiting distribution must not be Gaussian. From (3.2) follows

$$\lim_T P(m'/\varphi(T) < ||\hat{\alpha}_T|| < m/\varphi(T)) \geq 1 - \epsilon, \tag{3.3}$$

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 (3.3). 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 γ. When we estimate γ 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 non-degenerate 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

$$\lim_T P(||\varphi(T)\hat{\alpha}_T|| < m') = 1, \tag{3.4}$$

implying that the selected $\varphi(T)$ does not satisfy condition (3.2). Hence for such a choice of $\varphi(T)$ the estimator $\varphi(T)\hat{\alpha}_T$ cannot follow a non-degenerate 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 (3.2) 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 $\{\gamma \,|\, ||\gamma|| \leq m\}$, 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.9 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, Ch. 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 non-differentiability of the objective function makes a standard Taylor series argument impossible. However, also a result such as Theorem 5 in Pollard (1984) 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 cannot 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 probabality given by the density at zero $f_t(0)$ is relevant for our arguments (but not in Theorem 5 of Pollard, 1984).

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 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 Σ_0, which should be estimable using the heteroskedasticity and autocorrelation consistent (HAC) asymptotic covariance for the normalized subgradient (e.g., Powell, 1991 and 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, Ch. 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)

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

THEOREM 4. *In the model (1.1), 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 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*

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

Proof of Lemma 1C. For $u_t - h_t(\tilde{\alpha}) > 0$ and $u_t - h_t(\alpha) > 0$ the assertion follows directly from (2.3). The same is valid for $u_t - h_t(\tilde{\alpha}) \leq 0$ and $u_t - h_t(\alpha) \leq 0$. Furthermore, $u_t - h_t(\tilde{\alpha}) > 0$ and $u_t - h_t(\alpha) \leq 0$ imply

$$0 < u_t - h_t(\tilde{\alpha}) \leq h_t(\alpha) - h_t(\tilde{\alpha}). \tag{5.1}$$

Thus, according to (2.3) we have

$$q_t(\tilde{\alpha}) - q_t(\alpha) = (1-\vartheta)\left[h_t(\tilde{\alpha}) - h_t(\alpha)\right] + u_t - h_t(\tilde{\alpha}). \tag{5.2}$$

From (5.1) and (5.2) follows $(1-\vartheta)[h_t(\tilde{\alpha}) - h_t(\alpha)] < q_t(\tilde{\alpha}) - q_t(\alpha) \leq -\vartheta[h_t(\tilde{\alpha}) - h_t(\alpha)]$. Analogous considerations for $u_t - h_t(\tilde{\alpha}) \leq 0$ 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

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

These equations imply

$$\limsup_T \sup_{||\tilde{\alpha}-\alpha|| \leq \delta} \left| \frac{1}{T} \sum_{t=1}^{T} \left(q_t(\tilde{\alpha}) - Eq_t(\tilde{\alpha})\right) - \frac{1}{T} \sum_{t=1}^{T} \left(q_t(\alpha) - Eq_t(\alpha)\right) \right| < 2\epsilon. \tag{5.3}$$

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, Theorem 6.3). From the mixing property A.2 follows (i). From $q_t(0) = 0$ in Lemma 1C follows $|q_t(\alpha)| \leq \max(\vartheta, 1-\vartheta)|h_t(\alpha)|$ and the c_r-inequality (e.g., Davidson, 1994, 9.28) implies $|h_t(\alpha)|^{1+\epsilon} = |g(x_t, \beta_0+\alpha) - g(x_t, \beta_0)|^{1+\epsilon} \leq 2^\epsilon |g(x_t, \beta_0+\alpha)|^{1+\epsilon} + 2^\epsilon |g(x_t, \beta_0)|^{1+\epsilon}$. Together with the dominance condition A.7 this establishes (ii).

The assertion of the Lemma follows from (5.3) and that $q_t(\alpha)$ obeys a LLN using the usual arguments, since D_α is compact and admits a finite covering. ■

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

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

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 $\delta \leq \delta_0$,

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

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

$$E\big[b_t(\alpha)\big] = \begin{cases} \int_0^{h_t}(h_t - z)dF_t(z) & \text{for} \quad h_t > 0, \\ \int_{h_t}^0 (z - h_t)dF_t(z) & \text{for} \quad h_t \leq 0. \end{cases} \tag{5.4}$$

By limiting the integration domain in (5.4) to $[0, h_t/2]$ and $[h_t/2, 0]$, respectively, we obtain

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

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 $\text{plim}_T\, \hat{\alpha}_T = 0$, where $\alpha = \beta - \beta_0$. For every fixed event ω the normalized loss function $Q_T(\alpha)/T$ can be written as

$$\frac{1}{T}Q_T(\alpha) = E\left(\frac{1}{T}Q_T(\alpha)\right) + \left(\frac{1}{T}Q_T(\alpha) - E\frac{1}{T}Q_T(\alpha)\right). \tag{5.5}$$

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

$$E\left(\frac{1}{T}Q_T(\alpha)\right) \geq f_0 \min\left(\left[\frac{1}{4T}\sum_{t=1}^{T}|h_t(\alpha)|\right]^2, \delta_0^2\right). \tag{5.6}$$

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

$$E\left(\frac{1}{T}Q_T(\alpha)\right) \geq f_0 \min\left(\frac{1}{16}\delta_\epsilon^2, \delta_0^2\right) \stackrel{\text{def}}{=} \eta > 0. \tag{5.7}$$

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

$$P\left(\sup_{\alpha \in D_\alpha}\left|\frac{1}{T}Q_T(\alpha) - E\frac{1}{T}Q_T(\alpha)\right| \leq \frac{\eta}{2}\right) \to 1. \tag{5.8}$$

Then, from (5.5) and under consideration of (5.7) and (5.8), for $T \to \infty$ and $||\alpha|| \geq \epsilon$,

$$P\left(\frac{1}{T}Q_T(\alpha) \geq \frac{\eta}{2}\right) \to 1. \tag{5.9}$$

However, due to $Q_T(\hat{\alpha}_T) \leq Q_T(0) = 0$, from (5.9) follows

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

for an arbitrary $\epsilon > 0$. ■

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

$B(\gamma,\rho) = \{\alpha \mid \alpha \in \mathcal{C}, ||\gamma - \alpha|| < \rho\}$ with $\rho > 0$. Then, for every ρ we can choose finitely many $\gamma(i,\rho) \in \mathcal{C}$, $i = 1, 2, \ldots, n(\rho)$, admitting the finite covering

$$\mathcal{C} \subset \bigcup_{i=1}^{n(\rho)} B(\gamma(i,\rho),\rho). \tag{5.10}$$

Further we assume for all T, every sufficiently small ρ, and the corresponding covering (5.10), 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

$$\underline{R}_T(i,\rho) \le R_T(\omega,\gamma) \le \overline{R}_T(i,\rho) \qquad \text{for all } \gamma \in B(\gamma(i,\rho),\rho), \tag{5.11}$$

$$\operatorname*{plim}_T \left(\underline{R}_T(i,\rho) - E\underline{R}_T(i,\rho)\right) = 0, \quad \operatorname*{plim}_T \left(\overline{R}_T(i,\rho) - E\overline{R}_T(i,\rho)\right) = 0, \tag{5.12}$$

and

$$\lim_{\rho\to 0} \limsup_T \max_{1\le i\le n(\rho)} \left| E\overline{R}_T(i,\rho) - E\underline{R}_T(i,\rho)\right| = 0. \tag{5.13}$$

LEMMA 1N. *Then,*

$$\operatorname*{plim}_T \sup_{\gamma\in\mathcal{C}} \left(R_T(\omega,\gamma) - ER_T(\omega,\gamma)\right) = 0. \tag{5.14}$$

Proof of Lemma 1N. From (5.11) follows for all T, every sufficiently small ρ, and $1 \le i \le n(\rho)$,

$$\begin{aligned} \underline{R}_T(i,\rho) - E\underline{R}_T(i,\rho) + \left(E\underline{R}_T(i,\rho) - ER_T(\omega,\gamma)\right) &\le R_T(\omega,\gamma) - ER_T(\omega,\gamma), \\ \overline{R}_T(i,\rho) - E\overline{R}_T(i,\rho) + \left(E\overline{R}_T(i,\rho) - ER_T(\omega,\gamma)\right) &\ge R_T(\omega,\gamma) - ER_T(\omega,\gamma), \end{aligned} \tag{5.15}$$

for all $\gamma \in B(\gamma(i,\rho))$. Thus, due to (5.13), for every $\epsilon > 0$ and T sufficiently large, there exists a $\rho > 0$ such that

$$\begin{aligned} \min_i \left(\underline{R}_T(i,\rho) - E\underline{R}_T(i,\rho)\right) - \epsilon &\le R_T(\omega,\gamma) - ER_T(\omega,\gamma), \\ \max_i \left(\overline{R}_T(i,\rho) - E\overline{R}_T(i,\rho)\right) + \epsilon &\ge R_T(\omega,\gamma) - ER_T(\omega,\gamma), \end{aligned} \tag{5.16}$$

for all $\gamma \in \mathcal{C}$. Then the assertion follows from (5.16) and (5.12). ■

As a next step in Lemma 2N we study the first order conditions resulting from the directional derivatives of the loss function (2.3).

LEMMA 2N. *In the model (1.1), the following assertion holds under Assumption A.9: If there exists an* $\alpha = \hat{\alpha}_T$ *such that*

$$A_{lT}(\alpha, w) \leq S_T(\alpha, w) \leq A_{uT}(\alpha, w), \tag{5.17}$$

where

$$\begin{aligned} A_{lT}(\alpha, w) &\stackrel{\text{def}}{=} -\sum_{t=1}^{T} I[u_t = h_t(\alpha)] |w' \nabla h_t(\alpha)| I[w' \nabla h_t(\alpha) < 0], \\ S_T(\alpha, w) &\stackrel{\text{def}}{=} \sum_{t=1}^{T} w' \nabla h_t(\alpha) \left(I[u_t \leq h_t(\alpha)] - \vartheta \right), \\ A_{uT}(\alpha, w) &\stackrel{\text{def}}{=} \sum_{t=1}^{T} I[u_t = h_t(\alpha)] |w' \nabla h_t(\alpha)| I[w' \nabla h_t(\alpha) \geq 0], \end{aligned}$$

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,

$$\lim_{s \to 0} \frac{Q_T(\alpha + sw) - Q_T(\alpha)}{|s|}. \tag{5.18}$$

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

We calculate (5.18) by analyzing the summands $q_t(\alpha)$ defined in equation (2.3). 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 $h_t(\alpha) \neq u_t$

$$\lim_{s \to 0} \frac{q_t(\alpha + sw) - q_t(\alpha)}{|s|} = m_t \left(I[u_t \leq h_t(\alpha)] - \vartheta \right), \tag{5.19}$$

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

$$\lim_{s \to 0} \frac{q_t(\alpha + sw) - q_t(\alpha)}{|s|} = m_t \left((1 - \vartheta) I[m_t > 0] - \vartheta I[m_t \leq 0] \right). \tag{5.20}$$

The right hand side of (5.20) can be written as $m_t\left(I[m_t>0]-\vartheta\right)$. Thus, from (5.19) and (5.20) follows that $Q_T(\alpha)$ has a local minimum if, for all w,

$$\sum_{t=1}^{T} I[u_t\neq h_t(\alpha)]m_t\left(I[u_t\leq h_t(\alpha)]-\vartheta\right)+\sum_{t=1}^{T} I[u_t=h_t(\alpha)]m_t\left(I[m_t>0]-\vartheta\right)\geq 0. \tag{5.21}$$

By selecting $s>0$ and then $s<0$, from (5.21) follows

$$\begin{aligned} & -\sum_{t=1}^{T} I[u_t=h_t(\alpha)]w'\nabla h_t(\alpha)\left(I[w'\nabla h_t(\alpha)>0]-\vartheta\right) \\ \leq & \quad \sum_{t=1}^{T} I[u_t\neq h_t(\alpha)]w'\nabla h_t(\alpha)\left(I[u_t\leq h_t(\alpha)]-\vartheta\right) \\ \leq & -\sum_{t=1}^{T} I[u_t=h_t(\alpha)]w'\nabla h_t(\alpha)\left(I[w'\nabla h_t(\alpha)<0]-\vartheta\right). \end{aligned} \tag{5.22}$$

From adding $\sum_{t=1}^{T} I[u_t=h_t(\alpha)]w'\nabla h_t(\alpha)(1-\vartheta)$ to all three sides of (5.22) 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

$$A_T(\alpha,w)\overset{\text{def}}{=}A_{uT}(\alpha,w)-A_{lT}(\alpha,w)=\sum_{t=1}^{T}|w'\nabla h_t(\alpha)|I[u_t=h_t(\alpha)] \tag{5.23}$$

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 (5.17).

As mentioned above the first order conditions (5.17) are multiplied by $\varphi(T)/T$ and in Lemma 3N we want to show for $\lim_T \varphi(T)^2/T=0$ that

$$\operatorname*{plim}_{T}\sup_{||\gamma||\leq m}\frac{\varphi(T)}{T}A_T\left(\frac{\gamma}{\varphi(T)},w\right)=0. \tag{5.24}$$

LEMMA 3N. *If* $\lim_T \varphi(T)^2/T=0$, *then (5.24) 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 $\{\gamma \mid ||\gamma|| \leq m\}$. 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 (5.12) and (5.13), respectively.

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

$$R_T(\omega,\gamma) = \frac{\varphi(T)}{T}\sum_{t=1}^{T}\left|w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right)\right| I\left[u_t = h_t\left(\frac{\gamma}{\varphi(T)}\right)\right], \tag{5.25}$$

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

Due to Assumption A.9 for all T, $1 \leq t \leq T$, all w with $||w|| = 1$, a sufficiently small ρ 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

$$h_t\left(\frac{\gamma_*}{\varphi(T)}\right) \leq h_t\left(\frac{\gamma}{\varphi(T)}\right) \leq h_t\left(\frac{\gamma^*}{\varphi(T)}\right), \tag{5.26}$$

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

$$\overline{R}_T(i,\rho) = \frac{\varphi(T)}{T}\sum_{t=1}^{T}\left|w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right)\right| I\left[h_t\left(\frac{\gamma_*}{\varphi(T)}\right) \leq u_t \leq h_t\left(\frac{\gamma^*}{\varphi(T)}\right)\right], \tag{5.27}$$

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

$$\underline{R}_T(i,\rho) \leq R_T(\omega,\gamma) \leq \overline{R}_T(i,\rho), \tag{5.28}$$

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

$$Var\left(\overline{R}_T(i,\rho)\right) \leq \frac{\varphi(T)^2}{T^2}\sum_{t=1}^{T}\left(w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right)\right)^2 8\sum_{k=0}^{\infty}\mu_k. \tag{5.29}$$

From Assumptions A.10 and A.15 follows that the right hand side of (5.29) converges to 0 for sufficiently small ρ, and $1 \le i \le n(\rho)$, for $\lim_T \varphi(T)^2/T = 0$, establishing condition (5.12) of Lemma 1N.

It remains to verify condition (5.13). For all T, i, ρ, and w with $||w|| = 1$,

$$E\overline{R}_T(i,\rho) = \frac{\varphi(T)}{T}\sum_{t=1}^{T}\left|w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right)\right| d_t(\gamma^*,\gamma_*), \tag{5.30}$$

where we use the abbreviation

$$d_t(\gamma^*,\gamma_*) \overset{\text{def}}{=} F_t\left(h_t\left(\frac{\gamma^*}{\varphi(T)}\right)\right) - F_t\left(h_t\left(\frac{\gamma_*}{\varphi(T)}\right)\right). \tag{5.31}$$

From Assumptions A.13 and A.9 and a Taylor expansion with remainder of (5.31) follows an upper bound for (5.30) given by

$$\frac{1}{T}\sum_{t=1}^{T}\left|w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right)\right| f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\left|\nabla h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)'(\gamma^*-\gamma_*)\right|, \tag{5.32}$$

where $\tilde{\gamma}$ lies in between γ_* and γ^*. Then, due to Assumptions A.10 and A.14 using the Cauchy-Schwartz-inequality and $||\gamma^* - \gamma_*|| \le \rho$, condition (5.13) follows from (5.32). Hence the assertion is proved. ■

As a next step we split up the middle part of the first order conditions (5.17) according to

$$S_T\left(\frac{\gamma}{\varphi(T)},w\right) = \left(S_T\left(\frac{\gamma}{\varphi(T)},w\right) - ES_T\left(\frac{\gamma}{\varphi(T)},w\right)\right) + ES_T\left(\frac{\gamma}{\varphi(T)},w\right). \tag{5.33}$$

In Lemma 4N we analyze the second term on the right hand side of (5.33).

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* γ.

Proof of Lemma 4N. From the definition of $S_T(\alpha, w)$ in (5.17) then follows for $||\gamma|| \le m$

$$E\frac{\varphi(T)}{T}S_T\left(\frac{\gamma}{\varphi(T)},w\right) = \frac{\varphi(T)}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right)\left(F_t\left(h_t\left(\frac{\gamma}{\varphi(T)}\right)\right) - F_t(0)\right). \tag{5.34}$$

From Assumptions A.13 and A.9 and a Taylor expansion with remainder, follows for the right hand side of (5.34), in analogy to (5.32),

$$\frac{1}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right) f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\nabla h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)'\gamma, \tag{5.35}$$

where $\tilde{\gamma}$ is in between 0 and γ. As a first step we show that (5.35) is asymptotically equivalent to

$$\frac{1}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right) f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\nabla h_t\left(0\right)'\gamma, \tag{5.36}$$

which follows from proving that

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

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 (5.36) is asymptotically equivalent to the expression in the assertion.

■

In Lemma 5N we analyze the first term on the right hand side of (5.33).

LEMMA 5N.

$$\operatorname*{plim}_T \sup_{||\gamma||\leq m} \frac{\varphi(T)}{T}\left(S_T\left(\frac{\gamma}{\varphi(T)}, w\right) - ES_T\left(\frac{\gamma}{\varphi(T)}, w\right)\right) = 0. \tag{5.37}$$

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

$$R_T(\omega,\gamma) = \frac{\varphi(T)}{T}\sum_{t=1}^{T} w'\nabla h_t\left(\frac{\gamma}{\varphi(T)}\right) I\left[u_t \leq h_t\left(\frac{\gamma}{\varphi(T)}\right)\right]. \tag{5.38}$$

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 ρ 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

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

for all $\gamma \in B(\gamma(i,\rho),\rho)$. Analogously (5.26) holds for all $\gamma \in B(\gamma(i,\rho),\rho)$. 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

$$\underline{r}_t(i,\rho) \stackrel{\text{def}}{=} w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) I\left[u_t \le h_t\left(\frac{\gamma_*}{\varphi(T)}\right)\right] \quad \text{for } w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) > 0, \tag{5.39}$$

$$\underline{r}_t(i,\rho) \stackrel{\text{def}}{=} w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) I\left[u_t \le h_t\left(\frac{\gamma^*}{\varphi(T)}\right)\right] \quad \text{for } w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) \le 0, \tag{5.40}$$

$$\overline{r}_t(i,\rho) \stackrel{\text{def}}{=} w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) I\left[u_t \le h_t\left(\frac{\gamma^*}{\varphi(T)}\right)\right] \quad \text{for } w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) > 0, \tag{5.41}$$

$$\overline{r}_t(i,\rho) \stackrel{\text{def}}{=} w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) I\left[u_t \le h_t\left(\frac{\gamma_*}{\varphi(T)}\right)\right] \quad \text{for } w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) \le 0. \tag{5.42}$$

Then for all i, where $1 \le i \le n(\rho)$, and all T, inequality (5.28) holds for all $\gamma \in B(\gamma(i,\rho),\rho)$. Again we verify conditions (5.12) and (5.13) of Lemma 1N. Both $\{\omega \,|\, u_t \le h_t(\gamma_*/\varphi(T))\}$ and $\{\omega \,|\, u_t \le h_t(\gamma^*/\varphi(T))\}$ belong to the σ-algebra generated by u_t. Hence, in analogy to the considerations concerning (5.29), 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 (5.12) is established. Next we consider $E\overline{R}_T(i,\rho) - E\underline{R}_T(i,\rho)$. According to the definitions (5.39)-(5.42) this difference can be written as

$$\frac{\varphi(T)}{T}\sum_{t=1}^T w'\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) F_t\left(h_t\left(\frac{\lambda(t)}{\varphi(T)}\right)\right) - \frac{\varphi(T)}{T}\sum_{t=1}^T w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) F_t\left(h_t\left(\frac{\kappa(t)}{\varphi(T)}\right)\right), \tag{5.43}$$

where $\lambda(t) \in \{\gamma_*, \gamma^*\}$ and $\kappa(t) \in \{\gamma_*, \gamma^*\}$. Expression (5.43) can be split up according to

$$\begin{aligned}&\frac{\varphi(T)}{T}\sum_{t=1}^T w'\left(\nabla h_t\left(\frac{\overline{\gamma}}{\varphi(T)}\right) - \nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right)\right) F_t\left(h_t\left(\frac{\lambda(t)}{\varphi(T)}\right)\right)\\ &+\frac{\varphi(T)}{T}\sum_{t=1}^T w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right)\left(F_t\left(h_t\left(\frac{\lambda(t)}{\varphi(T)}\right)\right) - F_t\left(h_t\left(\frac{\kappa(t)}{\varphi(T)}\right)\right)\right).\end{aligned} \tag{5.44}$$

Again, from Assumptions A.13, A.9, and due to (5.31), we get an upper bound for the absolute value of (5.44) by

$$\frac{1}{T}\sum_{t=1}^T \left| w'\nabla^2 h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)(\overline{\gamma} - \underline{\gamma})\right| + \frac{\varphi(T)}{T}\sum_{t=1}^T \left| w'\nabla h_t\left(\frac{\underline{\gamma}}{\varphi(T)}\right) d_t(\lambda(t), \kappa(t))\right|, \tag{5.45}$$

where $\tilde{\gamma}$ lies in between $\underline{\gamma}$ and $\overline{\gamma}$. Due to Assumption A.11 and analogously to (5.32), expression (5.45) 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 $|\gamma|| \leq m$ from using the first order conditions in (5.17) in Lemma 2N, the decomposition in (5.33), and Lemmas 3N-5N,

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

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

$$\frac{1}{T}\sum_{t=1}^{T} \nabla h_t(0)\, f_t\left(h_t\left(\frac{\tilde{\gamma}}{\varphi(T)}\right)\right)\nabla h_t(0)'\,\gamma = o_p(1). \tag{5.46}$$

By construction the estimator $\hat{\gamma}_T$ is a solution of (5.46) 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

$$\frac{1}{T}\sum_{t=1}^{T} \nabla h_t(0)\, f_t\left(h_t\left(\frac{\gamma(t)}{\varphi(T)}\right)\right)\nabla h_t(0)' \tag{5.47}$$

is asymptotically non-singular, from (5.46) 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 are the first order conditions (5.17) 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.

$$\operatorname*{plim}_T \sup_{||\gamma|| \leq m} \frac{1}{\sqrt{T}} A_T\left(\frac{\gamma}{\sqrt{T}}, w\right) = 0. \tag{5.48}$$

Proof of Lemma 6N. In the proof of Lemma 3N, the factor $\varphi(T)$ played an important role for calculating the variance in (5.29). Again following Doukhan (1994, Theorem 3, p. 9, setting $p = q$) and due to Assumption A.15, instead of (5.29) we get for any $p > 1$

$$Var\left(\overline{R}_T(i,\rho)\right) \leq \frac{1}{T}\sum_{t=1}^{T}\left(w'\nabla h_t\left(\frac{\overline{\gamma}}{\sqrt{T}}\right)\right)^2 d_t(\gamma^*,\gamma_*)^{2/p} 16 \sum_{k=0}^{\infty}\mu_k^{(p-2)/p}, \tag{5.49}$$

where $d_t(\gamma^*,\gamma_*)$ is defined in analogy to (5.31) 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 (5.49) converges for such a p.

Further, from a Taylor expansion with remainder we get (see (5.31) and (5.32))

$$d_t(\gamma^*,\gamma_*) = \frac{1}{\sqrt{T}} f_t\left(h_t\left(\frac{\tilde{\gamma}}{\sqrt{T}}\right)\right)\nabla h_t\left(\frac{\tilde{\gamma}}{\sqrt{T}}\right)'(\gamma^*-\gamma_*), \tag{5.50}$$

where $\tilde{\gamma}$ lies in between γ_* and γ^*. Then, due to Assumptions A.14 and A.12, expression (5.50) is bounded from above by

$$\sup_{1\leq t\leq T} d_t(\gamma^*,\gamma_*) = o(1). \tag{5.51}$$

Together with Assumption A.10, from (5.49) and (5.51) follows that $\lim_T Var\left(\overline{R}_T(i,\rho)\right) = 0$ for sufficiently small ρ, all T, and $1 \leq i \leq n(\rho)$. It remains to establish the condition analogous to (5.30) using $\sqrt{T}$ instead of $\varphi(T)$. As (5.32) also holds using $\sqrt{T}$ instead of $\varphi(T)$, the assertion is proved. ■

Now decompose $S_T(\gamma/\sqrt{T}, w)$ defined in the first order conditions (5.17) into the following four sums:

$$\begin{aligned}
ES_T\left(\frac{\gamma}{\sqrt{T}},w\right) &= \sum_{t=1}^{T} w'\nabla h_t\left(\frac{\gamma}{\sqrt{T}}\right)\left(F_t\left(h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right)-\vartheta\right),\\
S_{T11}\left(\frac{\gamma}{\sqrt{T}},w\right) &= \sum_{t=1}^{T} w'\left(\nabla h_t\left(\frac{\gamma}{\sqrt{T}}\right)-\nabla h_t(0)\right)\left(I\left[u_t \leq h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right]-F_t\left(h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right)\right),\\
S_{T12}\left(\frac{\gamma}{\sqrt{T}},w\right) &= \sum_{t=1}^{T} w'\nabla h_t(0)\left(I\left[u_t \leq h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right]-I[u_t\leq 0]-F_t\left(h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right)+\vartheta\right),\\
S_{T2}(w) &= \sum_{t=1}^{T} w'\nabla h_t(0)\left(I[u_t\leq 0]-\vartheta\right).
\end{aligned}$$

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

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

$$\frac{1}{T}\sum_{t=1}^{T} w'\nabla h_t(0) f_t(0) \nabla h_t(0)'\gamma. \tag{5.52}$$

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 $h_t(\gamma/\sqrt{T})$ converges to $h_t(0) = 0$ uniformly in t. This follows from a Taylor expansion with remainder, $h_t(\gamma/\sqrt{T}) = h_t(0) + (1/\sqrt{T})\nabla h_t(\tilde{\gamma}/\sqrt{T})'\gamma$, where $\tilde{\gamma}$ lies in between 0 and γ, 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 $S_{T11}(\gamma/\sqrt{T}, w)$.

LEMMA 8N.

$$\operatorname*{plim}_{T} \sup_{||\gamma|| \leq m} \frac{1}{\sqrt{T}} S_{T11}\left(\frac{\gamma}{\sqrt{T}}, w\right) = 0. \tag{5.53}$$

Proof of Lemma 8N. In contrast to (5.38) in the proof of Lemma 5N we now analyze

$$R_T(\omega, \gamma) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} w'\left(\nabla h_t\left(\frac{\gamma}{\sqrt{T}}\right) - \nabla h_t(0)\right) I\left[u_t \leq h_t\left(\frac{\gamma}{\sqrt{T}}\right)\right] \tag{5.54}$$

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

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

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 ρ, all T, and $1 \leq i \leq n(\rho)$, 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.

$$\operatorname*{plim}_{T} \sup_{||\gamma||\leq m} \frac{1}{\sqrt{T}} S_{T12}\left(\frac{\gamma}{\sqrt{T}}, w\right) = 0. \tag{5.55}$$

Proof of Lemma 9N. In contrast to (5.38) in the proof of Lemma 5N we now analyze

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

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 (5.45), due to $\underline{\gamma} = \overline{\gamma} = 0$, we now have

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} w'\nabla h_t(0) d_t(\lambda(t),\kappa(t)), \tag{5.56}$$

where $\lambda(t) \in \{\gamma_*, \gamma^*\}$ and $\kappa(t) \in \{\gamma_*, \gamma^*\}$ and $d_t(\lambda(t),\kappa(t))$ is defined in analogy to (5.31). Note that $w'\nabla h_t(0) I[u_t \leq 0]$ does not depend on γ. 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 $S_T(\gamma/\sqrt{T}, w)$ (see the considerations after the proof of Lemma 6N) follows from the assertions in Lemmas 6N, 8N, 9N, and the first order conditions (5.17),

$$o_p(1) \leq \frac{1}{\sqrt{T}} E S_T(\gamma/\sqrt{T}, w) + \frac{1}{\sqrt{T}} S_{T2}(w) \leq o_p(1).$$

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

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

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

$$\hat{\gamma}_T = -V_T^{-1}\xi_T + o_p(1), \tag{5.57}$$

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 ξ_T is independent from γ, 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 ξ_T in Assumption A.18 follows

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

and

$$E\xi_T\xi_T' = T^{-1}\sum_{s,t=1}^{T} \nabla h_s(0)\omega_{s,t}\nabla h_t(0)', \tag{5.58}$$

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

$$V_T = T^{-1}\sum_{t=1}^{T} \nabla h_t(0) f_t(0) \nabla h_t(0)'. \tag{5.59}$$

We have to estimate the asymptotic covariance matrix $V_0^{-1}\Sigma_0 V_0^{-1}$. For the estimation of Σ_0 we employ the right hand side of (5.58). For the estimation of V_0 we employ the right hand side of (5.59).

The problem is that the vector $\nabla h_t(0) = \nabla g(x_t, \beta_0)$, contained in both V_0 and Σ_0, depends on the unknown parameter vector β_0. Since $\hat{\beta}_T$ is given, we replace β_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 Σ_0 is

$$\hat{\Sigma}_T = T^{-1} \sum_{s,t=1}^{T} \nabla h_s \left(\frac{\hat{\gamma}_T}{\sqrt{T}} \right) \omega_{s,t} \nabla h_t \left(\frac{\hat{\gamma}_T}{\sqrt{T}} \right)' \tag{5.60}$$

and the estimate of V_0 is

$$\hat{V}_T = T^{-1} \sum_{t=1}^{T} \nabla h_t \left(\frac{\hat{\gamma}_T}{\sqrt{T}} \right) f_t(0) \nabla h_t \left(\frac{\hat{\gamma}_T}{\sqrt{T}} \right)'. \tag{5.61}$$

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 non-singular due to Assumption A.17.

As a first step we have to show that

$$\operatorname*{plim}_T \hat{\Sigma}_T = \operatorname*{plim}_T T^{-1} \sum_{s,t=1}^{T} \nabla h_s \left(\frac{\hat{\gamma}_T}{\sqrt{T}} \right) \omega_{s,t} \nabla h_t (0)'. \tag{5.62}$$

In order to show that (5.62) 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 (5.62) follows if for $1 \leq i, j \leq K$

$$\operatorname*{plim}_T T^{-1} \sum_{s,t=1}^{T} e_i' \nabla h_s \left(\frac{\hat{\gamma}_T}{\sqrt{T}} \right) \omega_{s,t} \left(\nabla h_t \left(\frac{\hat{\gamma}_T}{\sqrt{T}} \right) - \nabla h_t(0) \right)' e_j = 0. \tag{5.63}$$

For the present we consider (5.62) only for all $\omega \in \Omega_{T,m} = \{\omega | \, ||\hat{\gamma}_T|| \leq m\}$. Due to the fact that the sum in (5.63) is a continuous function of $\hat{\gamma}_T$, there exists a γ^* with $||\gamma^*|| \leq m$ such that the absolute value of the sum on the left hand side of (5.63) is maximal for all $\omega \in \Omega_{T,m}$, choosing γ^* instead of $\hat{\gamma}_T$. Note that γ^* depends on i, j, and T. Hence, instead of (5.63) it suffices to show that

$$\lim_T T^{-1} \sum_{s,t=1}^{T} \left| e_i' \nabla h_s \left(\frac{\gamma^*}{\sqrt{T}} \right) \omega_{s,t} \left(\nabla h_t \left(\frac{\gamma^*}{\sqrt{T}} \right) - \nabla h_t(0) \right)' e_j \right| = 0. \tag{5.64}$$

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 γ^*. Due to $\omega_{s,t} = P(u_s \leq 0, u_t \leq 0) - P(u_s \leq 0)P(u_t \leq 0)$ for $s \neq t$, we define $\mu_0 = \omega_{t,t} = \vartheta(1-\vartheta)$, and Assumption A.15 implies $|\omega_{s,t}| \leq \mu_{|s-t|}$. As a consequence and by using the Cauchy-Schwartz inequality it can be shown that the absolute value of the left hand side of (5.64) 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 $||\gamma^*|| \leq m$, the dominance conditions in Assumptions A.10 and A.11, and the mixing condition in Assumption A.15, this bound is $O(1/\sqrt{T})$, establishing (5.64). 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 $\Omega_{T,m}$.

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 (5.64)), for $1 \leq i, j \leq K$ leading to

$$\lim_T T^{-1}\sum_{t=1}^{T}\left|e_i'\nabla h_t\left(\frac{\gamma^*}{\sqrt{T}}\right)f_t(0)\left(\nabla h_t\left(\frac{\gamma^*}{\sqrt{T}}\right) - \nabla h_t(0)\right)' e_j\right| = 0. \tag{5.65}$$

Again invoking Assumption A.9 we can derive an upper bound of (5.65)

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

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

From applying the Cauchy-Schwartz inequality follows

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

Then equation (5.65) 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. ■

References

[1] Andrews, D.W.K. (1987) Consistency in nonlinear econometric models: a generic uniform law of large numbers. *Econometrica*, 55, 1465-1471.

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

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

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

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

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

[7] Cai, Z. (2002) Regression quantiles for time series. *Econometric Theory*, 18, 169-192.

[8] Castellana, J.V., & M.R. Leadbetter (1986) On smoothed probability density estimation for stationary processes. *Stochastic Processes and their Applications*, 21, 179-193.

[9] Chamberlain, G. (1994) Quantile Regression, Censoring, and the Structure of Wages. In Sims, C. (ed.), *Advances in Econometrics: Sixth World Congress*, 1, 171-208.

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

[11] 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.

[12] Chen, X., R. Koenker, & Z. Xiao (2009) Copula-Based Nonlinear Quantile Autoregression. *Econometrics Journal*, 12, S50-S67.

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

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

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

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

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

[18] Doukhan, P. (1994): *Mixing.* Springer Verlag, New York.

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

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

[21] Engle, R.F., & S. Manganelli (2004) CAViaR: Conditional autoregressive Value at Risk by regression quantiles. *Journal of Business and Economics Statistics*, 22, 367-381.

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

[23] Fitzenberger, B., R. Wilke, R., & X. Zhang (2010) Implementing Box-Cox Quantile Regression. *Econometric Reviews*, 29, 158-181.

[24] Gallant, A.R., & H. White (1988) *A unified theory of estimation and inference for nonlinear dynamic models.* Basil Blackwell, Oxford.

[25] Geyer, C.J. (1996) On the asymptotics of convex stochastic optimization. unpublished manuscript.

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

[27] 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.

[28] 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.

[29] Huber, P.J. (1967) Behavior of Maximum Likelihood estimates under nonstandard conditions. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, UC Press.

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

[31] Ivanov, A.V. (1976) An asymptotic expansion for the distribution of the least squares estimator of the non-linear regression parameter. *Theory of Probability and its Applications*, 21, 557-570.

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

[33] Karlsson, A. (2007) Nonlinear Quantile Regression Estimation of Longitudinal Data. *Communications in Statistics – Simulation and Computation*, 37, 114-131.

[34] 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.

[35] 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.

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

[37] Knight, K. (1999) Asympotics for L_1 of regression parameters under heteroscedasticity. *Canadian Journal of Statistics*, 27, 497-507.

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

[39] Koenker, R., & G. Bassett (1978) Regression quantiles. *Econometrica*, 46, 33-50.

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

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

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

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

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

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

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

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

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

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

[50] Nze, P.A., & P. Doukhan (2004) Weak dependence: models and applications to econometrics. *Econometric Theory*, 20, 169-192.

[51] Oberhofer, W. (1982) The consistency of nonlinear regression minimizing the L_1 norm. *Annals of Statistics*, 10, 316-319.

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

[53] Peracchi, F. (2001) *Econometrics.* J. Wiley & Sons

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

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

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

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

[58] Pötscher, B.M., & I.R. Prucha (1997) *Dynamic nonlinear econometric models: asymptotic theory.* Springer-Verlag.

[59] Pollard, D. (1984) *Convergence of stochastic processes.* Springer-Verlag.

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

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

[62] 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*, Cambridge University Press, 357-384.

[63] Powell, J.L. (1994) Estimation of semiparametric models. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, 4, Elsevier Science, Amsterdam, 2443-2521.

[64] 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.

[65] Prakasa Rao, B.L.S. (1987) *Asymptotic theory of statistical inference.* Wiley, New York.

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

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

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

[69] 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.

[70] Tran, L.T., G.G. Roussas, S. Yakowitz, & B.T. Van Fixed-design regression for linear time series. *Annals of Statistics*, 24, 75-991.

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

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

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

[74] White, H. (1994) *Estimation, inference and specification analysis.* Econometric Society Monographs No. 22. Cambridge University Press.

[75] Withers, C.S. (1981) Central Limit Theorems for Dependent Variables. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 509-534.

[76] Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data.* 2e. MIT Press

[77] Yu, K., Z. Lu, & J. Stander (2003) Quantile regression: applications and current research areas. *The Statistician*, 52, 331-350.

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

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

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