



Inference for linear models with dependent errors

Zhou Zhou

University of Toronto, Canada

and Xiaofeng Shao

University of Illinois at Urbana–Champaign, Champaign, USA

[Received August 2010. Final revision June 2012]

Summary. The paper is concerned with inference for linear models with fixed regressors and weakly dependent stationary time series errors. Theoretically, we obtain asymptotic normality for the M -estimator of the regression parameter under mild conditions and establish a uniform Bahadur representation for recursive M -estimators. Methodologically, we extend the recently proposed self-normalized approach of Shao from stationary time series to the regression set-up, where the sequence of response variables is typically non-stationary in mean. Since the limiting distribution of the self-normalized statistic depends on the design matrix and its corresponding critical values are case dependent, we develop a simulation-based approach to approximate the critical values consistently. Through a simulation study, we demonstrate favourable finite sample performance of our method in comparison with a block-bootstrap-based approach. Empirical illustrations using two real data sets are also provided.

Keywords: M -estimation; Non-linear time series; Quantile regression; Self-normalization

1. Introduction

Consider the following p -variate linear model:

$$Y_t = \mathbf{x}_{t,n}^T \beta + e_t, \quad t = 1, 2, \dots, n, \quad (1)$$

where ‘T’ denotes matrix transpose, and $\mathbf{x}_{t,n} = (x_{t1,n}, x_{t2,n}, \dots, x_{tp,n})^T$, $1 \leq t \leq n$, are $p \times 1$ known deterministic design vectors. In what follows we omit the subscript n in the notation if no confusion arises. As a typical estimation procedure, let ρ be a convex function and the unknown parameter vector β is estimated by the minimizer

$$\hat{\beta}_n = \arg \min_{\beta \in \mathbb{R}^p} \sum_{t=1}^n \rho(Y_t - \mathbf{x}_t^T \beta). \quad (2)$$

Note that $\rho(x) = \alpha \max(x, 0) + (1 - \alpha) \max(-x, 0)$, $0 < \alpha < 1$, leads to quantile regression (Koenker, 2005). Other popular choices of ρ are \mathcal{L}^q -regression with $\rho(x) = |x|^q$ and Huber’s estimation with $\rho(x) = x^2 \mathbf{1}(|x| \leq c) / 2 + (cx - c^2/2) \mathbf{1}(|x| > c)$, $c > 0$. Considerable efforts have been made in the study of the limiting behaviour of $\hat{\beta}_n$, when the error e_t is dependent; see Koul (1977), Portnoy (1977, 1979, 1991), Prakasa Rao (1981), Babu (1989), Phillips (1991) and Wu (2007a), among others. Related work includes Cui *et al.* (2004), where e_t is assumed to be a spatially correlated process.

Address for correspondence: Zhou Zhou, Department of Statistics, University of Toronto, 100 St George Street, Toronto, Ontario, M5S 3G3, Canada.
E-mail: zhou@utstat.toronto.edu

In statistical applications, an important problem is to construct a confidence region or to perform hypothesis testing for β . In the case of temporally dependent errors, the asymptotic variance of $\hat{\beta}_n$ admits a complicated form. For example, suppose that $\mathbf{x}_t = 1$, and $\rho(x) = |x|$. Then, under certain regularity conditions, it can be shown that $(\hat{\beta}_n - \beta)\sqrt{n} \Rightarrow N\{0, \sigma_\infty^2 / \varphi'(0)^2\}$, where ‘ \Rightarrow ’ stands for convergence in distribution, and $\varphi(\cdot)$ and σ_∞^2 are defined in conditions 2 and 10 in Section 2.1 respectively. Consistent estimation of σ_∞^2 and $\varphi'(0)$ is possible but quite involved. Basically, σ_∞^2 is the long-run variance of the transformed sequence $\psi(e_t)$ (see condition 1), which can be estimated by the lag window type of estimate with a choice of kernel function and truncation lag. As for $\varphi'(0)$, it corresponds to the so-called sparsity (Koenker, 2005) in the special case of quantile regression, and a consistent estimate can be formed with the choice of another bandwidth parameter. In practice, the practitioner may find the selection of bandwidths a difficult task although there are bandwidth selection algorithms developed for problems of both types. The choice of bandwidth usually depends on another user-chosen number or it admits an analytical expression under the assumption of a parametric model, the violation of which could lead to poor finite sample performance. Therefore it seems desirable to develop alternative methods that do not involve any bandwidth selection and still lead to asymptotically valid inference.

Recently, Shao (2010) proposed the so-called self-normalized (SN) approach to constructing confidence intervals (regions) for quantities that are associated with a stationary time series. A distinctive feature of the SN approach is that no smoothing parameters or bandwidths are involved and the SN statistic is asymptotically pivotal. The major goal of this paper is to extend the SN approach to regression model (1) with time series errors. The extension is very non-trivial and it differs from Shao’s work in three important aspects. First, though their covariance structure is stationary, the response variables $\{Y_t\}_{t=1}^n$ in model (1) are typically non-stationary in mean over time, so the argument on the basis of influence functions in Shao (2010) is not directly applicable. To circumvent the difficulty, we establish a uniform Bahadur representation for $\hat{\beta}_{[sn]}$, $s \in (0, 1]$, where $[a]$ stands for the integer part of $a \in \mathbb{R}$ and $\hat{\beta}_{[sn]}$ stands for the estimate of β based on the subsample $\{\mathbf{x}_t, Y_t\}_{t=1}^{[sn]}$, and thus we bypass the influence-function-based argument. Second, the key technical assumption in Shao (2010) for the validity of the SN approach is not yet verified for certain important statistics, such as sample quantiles. Our regression set-up includes Shao’s as a special case because, if we let $\mathbf{x}_t = 1$ for $t = 1, \dots, n$, then $Y_t = \beta_0 + e_t$ are stationary time series. So our results provide a direct verification of the assumption that is needed in the use of the SN approach for a large class of statistics with sample quantiles as a prominent example. Third, the limiting distribution of our SN statistic is not pivotal and it depends on the design matrix, so we cannot directly use the critical values of the particular limiting distribution that were described in Lobato (2001) and Shao (2010). We propose a simulated-based method to approximate the critical values consistently.

The rest of the paper is organized as follows. Section 2 presents theoretical assumptions and results, and introduces the SN approach in the regression set-up. Section 3 compares the SN approach with the block-bootstrap-based approach for both least squares regression and median regression via simulations. Empirical illustrations are given in Section 4 and some discussions are provided in Section 5. Proofs are gathered in Appendix A.

2. Asymptotic theory and methodology

Throughout the paper, we assume that the error (e_t) is a stationary process which admits the causal representation

$$e_t = G(\dots, \varepsilon_{t-1}, \varepsilon_t), \quad (3)$$

where ε_t , $t \in \mathbb{Z}$, are independent and identically distributed (IID) random variables. The representation (3) can be viewed as a physical system with (ε_t) being the inputs, (e_t) being the outputs and G being the transform that represents the underlying data-generating mechanism. It covers a wide range of linear and non-linear time series models that are encountered in practice; see, for example, Priestley (1980) and Wu (2005, 2007a) for more discussions and examples.

2.1. Dependence measures and regularity conditions

Define the shift process $\mathcal{F}_t = (\dots, \varepsilon_{t-1}, \varepsilon_t)$. Let $(\varepsilon'_t)_{t \in \mathbb{Z}}$ be an IID copy of $(\varepsilon_t)_{t \in \mathbb{Z}}$. For $t \geq 0$, let the coupled process $\mathcal{F}_t^* = (\mathcal{F}_{-1}, \varepsilon'_0, \varepsilon_1, \dots, \varepsilon_t)$. Further let $\|\cdot\|_q := \mathbb{E}[|\cdot|^q]^{1/q}$ and $\|\cdot\| := \|\cdot\|_2$ and denote $X \in \mathcal{L}^q$ if $\|X\|_q < \infty$. Define physical dependence measures for the stationary time series $H(\mathcal{F}_t) \in \mathcal{L}^q$ as

$$\delta_H(k, q) = \|H(\mathcal{F}_k) - H(\mathcal{F}_k^*)\|_q. \quad (4)$$

Adopting the idea of coupling, $\delta_H(k, q)$ measures the dependence of $H(\mathcal{F}_k)$ on the input ε_0 . The above dependence measures are closely related to the data-generating mechanism and therefore are easy to work with. Wu (2005) contains detailed calculations of $\delta_H(k, p)$ for stationary linear processes and a very general class of stationary non-linear processes used in practice. In what follows, we provide a list of technical conditions that are needed in establishing our main results.

Condition 1. Assume that ρ has derivative ψ . $\mathbb{E}[\psi(e_1)] = 0$ and $\|\psi(e_1)\| > 0$.

Condition 2. $\varphi(s) := E[\psi(e_1 + s)] = s\varphi'(0) + O(s^2)$ as $s \rightarrow 0$. Assume that $\varphi'(0) > 0$.

Condition 3. $m(s) := \|\psi(e_1 + s) - \psi(e_1)\| = O(|s|^\lambda)$ as $s \rightarrow 0$ for some $\lambda > 0$.

Condition 4. Uniformly on $[r, 1]$ for some $0 < r < 1$, $\Sigma_{\lfloor sn \rfloor} / (sn) = \Sigma(s) + O(1/\sqrt{n})$, where $\Sigma_t = \Sigma_{j=1}^t \mathbf{x}_j \mathbf{x}_j^\top$. Assume that, for sufficiently large n , there exists $L < \infty$, such that

$$|\Sigma_{\lfloor s_1 n \rfloor} - \Sigma_{\lfloor s_2 n \rfloor}| \leq L(|\lfloor s_2 n \rfloor - \lfloor s_1 n \rfloor|)$$

holds for all $r \leq s_1 < s_2 \leq 1$, and the smallest eigenvalue of $\Sigma(\cdot)$ is bounded away from 0 on $[r, 1]$.

Condition 5. $\max_{\lfloor rn \rfloor \leq t \leq n} |\Sigma_{j=1}^t \psi(e_j - \mathbf{x}_j^\top \hat{\beta}_t) \mathbf{x}_j| = O_{\mathbb{P}}(\max_{1 \leq t \leq n} |\mathbf{x}_t|)$, where r is the same as the r in condition 4.

Condition 6. $\sum_{j=1}^{n-1} |\mathbf{x}_j - \mathbf{x}_{j+1}| + |\mathbf{x}_n| = O(n^{1/4-\delta})$ for some $\delta > 0$.

Condition 7. Define $\psi_k(s; \mathcal{F}_0) = \mathbb{E}[\psi(e_k + s) | \mathcal{F}_0]$. Assume that there is an $\varepsilon_0 > 0$, such that

$$L_t := \sup_{|s_1|, |s_2| \leq \varepsilon_0, s_1 \neq s_2} \frac{|\psi_1(s_1; \mathcal{F}_t) - \psi_1(s_2; \mathcal{F}_t)|}{|s_1 - s_2|} \in \mathcal{L}^1.$$

Condition 8. Suppose that $\psi_1(\cdot, \mathcal{F}_t) \in \mathcal{C}^p$. For some $\varepsilon_0 > 0$ and $l = 0, 1, \dots, p$, we have $\sup_{|\varepsilon| \leq \varepsilon_0} \|\psi_1^{(l)}(\varepsilon; \mathcal{F}_t)\|_4 < \infty$ and

$$\sum_{t=0}^{\infty} \sup_{|\varepsilon| \leq \varepsilon_0} \|\mathbb{E}[\psi_1^{(l)}(\varepsilon; \mathcal{F}_t) | \mathcal{F}_0] - \mathbb{E}[\psi_1^{(l)}(\varepsilon; \mathcal{F}_t^*) | \mathcal{F}_0^*]\|_4 < \infty.$$

Condition 9. $\sum_{k=1}^{\infty} k \delta_{\psi \circ G}(k, 4) < \infty$.

Condition 10. $\sigma_\infty^2 := \Sigma_{t=-\infty}^\infty \text{cov}\{\psi(e_0), \psi(e_t)\} > 0$.

A few remarks on the regularity conditions are in order. Conditions 1–3, 5, 7 and 8 are adopted from Wu (2007a), and they are required to guarantee the asymptotic normality and Bahadur representations of the M -estimates. We refer to Wu (2007a) for more details. In particular, the term on the left-hand side of condition 5 is always 0 if $\psi(\cdot)$ is a continuous function. Additionally, condition 5 is always satisfied in the case of quantile regression by the arguments in lemma 7 of Zhou and Wu (2009).

The Lipschitz-type condition on $\Sigma_{[sn]}$ in condition 4 is to guarantee the tightness of $\hat{\beta}_{[sn]}$. See lemma 6 in Section 6 for more details. Additionally, condition 4 implies that $\Sigma(\cdot)$ is a continuous function on $[r, 1]$. Condition 6 controls the total variation of the sequence (\mathbf{x}_t) and is needed to establish the uniform Gaussian approximation of the M -processes $(\Sigma_{i=1}^t \psi(e_i) \mathbf{x}_i)_{t=1}^n$. See lemma 5 in Appendix A. Proposition 1 below shows that conditions 4 and 6 are satisfied by a very general class of fixed designs. However, note that condition 6 excludes some of the orthogonal designs (e.g. $\mathbf{x}_t = (1, (-1)^t)'$) and the random design when (\mathbf{x}_t) is a realization of a p -dimensional stationary process. In the latter case the results of Lee (2006), which dealt with the random-regressor case, can be used. As pointed out by a referee, the fixed regressor assumption is a limitation in many applications, which we acknowledge. However, it should be noted that our set-up allows complicated deterministic trend functions that are not included in Lee's framework; see example 1 below. From this point of view, our work is a complement to Lee's.

Note that $\psi(e_t) = \psi\{G(\mathcal{F}_t)\}$. Therefore condition 9 asserts that the stationary process $\{\psi(e_t)\}$ is short range dependent. Condition 9 guarantees that σ_∞ in condition 10 is well defined and is finite. A sufficient condition for condition 9 is $\delta_{\psi \circ G}(k, 4) = O\{k^{-2} \log(-^{-1-\varepsilon} k)\}$ for some $\varepsilon > 0$. Condition 10 is mild and it asserts that the long-run variance of the process $\{\psi(e_t)\}$ is positive.

2.1.1. Example 1 (piecewise smooth design)

Consider the design

$$\mathbf{x}_{t,n} = \mathbf{f}_j(t/n) \mathbf{1}(s_j < t/n \leq s_{j+1}), \quad (5)$$

where $0 = s_0 < s_1 < \dots < s_m < s_{m+1} = 1$ are fixed points and $m = O(n^{1/4-\delta})$ for some $\delta > 0$. The functions $\mathbf{f}_j : [s_j, s_{j+1}] \rightarrow \mathbb{R}^p$ are assumed to be uniformly Lipschitz continuous in the sense that their Lipschitz constants are bounded. Clearly \mathbf{x}_t is a piecewise smooth function with m change points, where m can grow with n . In the case that $m = 0$, then $\mathbf{f}_0(x)$ is a Lipschitz continuous function on $[0, 1]$. The latter type of design covers many useful cases in practice such as polynomial trend, periodic trend and exponential trend, among others.

Proposition 1. Assume design (5). If $\Sigma(r)$ is non-singular, then design (5) satisfies conditions 4 and 6.

Theorem 1. Under conditions 1–10, we have

$$(\hat{\beta}_n - \beta) \sqrt{n} \Rightarrow N\{0, \sigma_\infty^2 \Sigma^{-1}(1) / \varphi'(0)^2\}. \quad (6)$$

By condition 4, $\Sigma(1)$ can be consistently estimated by Σ_n/n . Hence theorem 1 implies that (asymptotically) valid inference of β on the basis of the above central limit theorem boils down to consistent estimation of the quantities σ_∞^2 and $\varphi'(0)$. It is worth mentioning that theorem 1 in Wu (2007a) established a very general central limit theorem for $\hat{\beta}_n$. However, the latter central

limit theorem contains an infinite number of nuisance parameters, i.e. $\text{cov}\{\psi(e_0), \psi(e_t)\}$ for all t , and its use is not straightforward for inference. In contrast, our theorem 1 implies that, in many interesting cases, theorem 1 in Wu (2007a) can be simplified into result (6) and thus the number of nuisance parameters can be reduced to 2.

Theorem 2. Under conditions 1–10, we have for all $r \in (0, 1)$ the following uniform Bahadur representation:

$$\sup_{r \leq s \leq 1} \left| \varphi'(0) s \Sigma(s) \hat{\beta}_{\lfloor sn \rfloor} \sqrt{n} - \sum_{j=1}^{\lfloor sn \rfloor} \psi(e_j) \mathbf{x}_j / \sqrt{n} \right| = o_{\mathbb{P}}(1). \quad (7)$$

Furthermore, we have

$$n(\hat{\beta}_n - \beta)^T W_n^{-1} (\hat{\beta}_n - \beta) \Rightarrow U^T(1) \left\{ \int_{s=r}^1 (U(s) - sU(1))(U(s) - sU(1))^T ds \right\}^{-1} U(1),$$

where $W_n = \Sigma_{t=\lfloor rn \rfloor}^n t^2 (\hat{\beta}_t - \hat{\beta}_n)(\hat{\beta}_t - \hat{\beta}_n)^T / n^2$ and $U(\cdot)$ is a zero-mean Gaussian process with covariance $\text{cov}\{U(s_1), U(s_2)\} = \Sigma\{\max(s_1, s_2)\}^{-1}$.

In practice, people are often interested in making inference on $\beta_A := A\beta$, where A is a full rank $q \times p$ matrix with $q \leq p$. For instance, one may want to construct a confidence interval for a single regression coefficient β_i , which corresponds to $A = \mathbf{1}_{i,p}$, where $\mathbf{1}_{i,p}$ represents the row vector of length p with the i th entry being 1 and the other entries being 0.

Proposition 2. Define $\hat{\beta}_{t,A} = A\hat{\beta}_t$ for $t = 1, 2, \dots, n$. Then, under conditions 1–10, we have

$$\begin{aligned} n(\hat{\beta}_{n,A} - \beta_A)^T W_{n,A}^{-1} (\hat{\beta}_{n,A} - \beta_A) \\ \Rightarrow U^T(1) A^T \left\{ A \int_{s=r}^1 (U(s) - sU(1))(U(s) - sU(1))^T ds A^T \right\}^{-1} A U(1), \end{aligned} \quad (8)$$

where $W_{n,A} = \Sigma_{t=\lfloor rn \rfloor}^n t^2 (\hat{\beta}_{t,A} - \hat{\beta}_{n,A})(\hat{\beta}_{t,A} - \hat{\beta}_{n,A})^T / n^2$.

Proposition 2 follows trivially from the proof of theorem 2. The limiting distributions that are stated in theorem 2 and proposition 2 are unknown and they depend on the design matrix in a non-trivial manner. Below we present a simulation-based method to approximate the limiting distribution.

Step 1: generate standard normal random variables $V_{m1}, V_{m2}, \dots, V_{mn}$. Let $\hat{U}_{m,s,A} = A \hat{\Sigma}^{-1}(s) \Sigma_{j=1}^{\lfloor sn \rfloor} V_{mj} \mathbf{x}_j / \sqrt{n}$. Calculate

$$D_m = \hat{U}_{m,1,A}^T \left\{ \int_{s=r}^1 (\hat{U}_{m,s,A} - s\hat{U}_{m,1,A})(\hat{U}_{m,s,A} - s\hat{U}_{m,1,A})^T ds \right\}^{-1} \hat{U}_{m,1,A}.$$

Step 2: repeat step 1 for $m = 1, 2, \dots, B = 1000$ (say) and obtain D_1, \dots, D_B .

Step 3: denote by $Q_{m,\alpha}$ the α th quantile of D_1, \dots, D_B . Then a $100\alpha\%$ confidence region for β_A can be constructed as

$$\{\beta_A : n(\hat{\beta}_{n,A} - \beta_A)^T W_{n,A}^{-1} (\hat{\beta}_{n,A} - \beta_A) \leq Q_{m,\alpha}\}.$$

Under our condition 4, $\Sigma(s)$ can be well estimated by $\hat{\Sigma}(s) := \Sigma_{\lfloor sn \rfloor} / (\lfloor sn \rfloor)$ uniformly in $s \in [r, 1]$. Thus it is straightforward to show that each D_m converges in distribution to

$$U^T(1)A^T \left\{ A \int_{s=r}^1 (U(s) - sU(1))(U(s) - sU(1))^T ds A^T \right\}^{-1} A U(1)$$

in result (8). Hence by proposition (2) the above approximation is consistent. In our inference procedure, the only user-chosen parameter is r , which stands for the trimming proportion in the normalization matrix W_n (or $W_{n,A}$). It is different from the smoothing parameter, such as the truncation lag in the long-run variance estimator, the block size in the block bootstrap methods or the subsampling width in the subsampling approach, as its effect is accounted for in the limiting distribution and its approximation. The effect of r on the coverage and length of the interval will be examined in our simulation studies.

Remark 1. As described in Shao (2010), the SN method is a special case of the so-called fixed b approach (Kiefer and Vogelsang, 2005) in the mean case. In the fixed b asymptotics, the ratio of bandwidth (or truncation lag) and sample size, denoted as b , is held fixed as the sample size $n \rightarrow \infty$. The resulting limiting distribution turns out to be a better approximation to the sampling distribution of the Studentized statistic than the limiting distribution that is obtained under conventional asymptotics (or so-called small b asymptotics), where $b \rightarrow \infty$ as $n \rightarrow \infty$. See Jansson (2004) and Sun *et al.* (2008) for theoretical justifications. Recently, in the context of quantile regression inference, Goh and Knight (2009) proposed to approximate the sampling distribution of the Wald-type test statistic with a fixed m limiting distribution, where m is a smoothing parameter involved in the conditional density estimate for the Wald-type test statistic. The fixed m limiting distribution depends on the distribution of the conditioning variables and is non-standard, but it can be consistently approximated by resampling. The idea that was presented in Goh and Knight (2009) is similar in spirit to the fixed b asymptotics, where the effect from the smoothing parameter appears in the limiting distribution. It is, however, developed for independent samples and seems not applicable to dependent situations, as the simple IID bootstrap procedure may not be able to approximate the fixed m limiting distribution in the presence of dependence consistently.

Remark 2. There is a literature on inference for regression models with dependent errors. For time series linear regression models with random regressors, bootstrap-based inference has been investigated by Fitzenberger (1997), Hidalgo (2003), Goncalves and White (2005) and Romano and Wolf (2006), among others. All the methods except Hidalgo (2003) involve a smoothing parameter, such as the block size in the moving block bootstrap and the subsampling width in the subsampling method. Hidalgo introduced a wild bootstrap method in the frequency domain that seems to work only for the least squares method in the time series regression setting. It is worth noting that Kiefer *et al.* (2000) developed an SN-like approach (fixed b , with $b = 1$ and Bartlett kernel) for inference in time series regression models, but their method does not seem applicable when the regressors are fixed. Also their framework takes advantage of the analytical form of the ordinary least squares (OLS) estimate and does not allow for quantile regression. Additionally, Lee (2006) introduced an SN approach for time series linear regression in the context of M -estimation but his work seems to focus on the random-regressor case. For linear regression with fixed regressors, Sherman (1997) generalized the subsampling method to the regression set-up and showed that the subsampling estimator can be calibrated to yield second-order-correct inference by using the extrapolation method that was introduced by Hall and Jing (1996).

Remark 3. As pointed out by a referee, inference for regression slopes when the error is dependent is widely studied in the econometrics literature and empirical researchers in economics

routinely use inference methods that allow dependence. In particular, work has been done to make the test asymptotically valid regardless of whether the error e_t is weakly dependent (e.g. an auto-regressive AR(1) model with AR(1) coefficient less than 1) or a unit root process. See Vogelsang (1998), Bunzel and Vogelsang (2005), Harvey *et al.* (2007) and Perron and Yabu (2009) among others for this line of research. By contrast, our SN-based test allows for weak dependence in the error but is not robust to strong persistence or dependence in the error, such as the existence of a unit root. However, all the aforementioned work seems focused on least squares estimation, whereas we allow quantile regression in our framework.

3. Simulation studies

In this section we shall conduct two simulation studies to investigate the accuracy and sensitivity of our proposed method in finite samples and compare those with the performance of the block-bootstrap-based methods. For this, we consider the simple regression model

$$Y_t = \beta_0 + \beta_1 \frac{t}{n} + e_t, \quad t = 1, 2, \dots, n, \quad (9)$$

and e_t are assumed to follow the AR(1) model

$$e_t = ae_{t-1} + \varepsilon_t, \quad (10)$$

where ε_t are assumed to be IID standard normal and to have $t(2)$ error with $|a| < 1$.

3.1. Sensitivity with respect to r

In this subsection we are interested in investigating the finite sample performance of our proposed method with respect to the choice of r . Model (9) was used with errors satisfying model (10). We let $a = 0.1, 0.5, 0.8$ represent weak, medium and strong dependence respectively. Additionally, OLS, where $\rho(\cdot) = (\cdot)^2$ and median regression (MR), where $\rho(\cdot) = |\cdot|$, were performed to estimate the unknown coefficients. 2000 replications were conducted with sample size $n = 100$ and $n = 200$. At nominal level 95%, we report the empirical coverage probability and average length of the confidence interval of β_1 with various choices of r . Tables 1 and 2 summarize the results for standard normal and $t(2)$ innovations respectively.

As seen from Tables 1 and 2, the coverage probabilities are not very sensitive with respect to r as long as r is not too small (in most cases) or too large. In fact, if r is too small, then the first several recursive estimates are based on very few samples and those estimates are not stable. In contrast, when r is too large, such as $r = 0.5$, the coverage is apparently inferior to that for small r , although the interval is usually shorter. Intuitively, we lose some efficiency if we trim too much. For all the combinations of a and n , the coverage is fairly stable over a broad range of r , and that range depends on a (i.e. the magnitude of dependence in the errors) and the innovation type ($N(0, 1)$ or $t(2)$). When the dependence becomes stronger, the range slightly drifts towards large values. In the case of standard normal innovations, the coverages appear quite satisfactory for $a = 0.1$ when $r \in [0.01, 0.1]$, for $a = 0.5$ when $r \in [0.05, 0.1]$, and for $a = 0.8$ when $r \in [0.08, 0.3]$. When the innovations are $t(2)$ distributed, the optimal coverages for the OLS method occur when $r \in [0.08, 0.3]$ for all a s, whereas, for the MR method, the optimal ranges all include $r = 0.1, 0.2$ and they drift a little towards large values as a becomes larger. A comparison between OLS and MR shows that the interval that is delivered by the OLS method is shorter or longer with similar coverages when the innovations are respectively $N(0, 1)$ or $t(2)$ distributed, thus confirming the advantage of MR, or quantile regression in general, for heavy-tailed error processes. Not surprisingly, the simulation results show that the performance of our proposed

Table 1. Simulated coverage probabilities and average lengths of the self-normalized confidence interval of β_1 for both OLS and MR at 95% nominal level with standard normal innovations[†]

r	OLS		MR		OLS		MR	
	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
	$a=0.1, n=100$				$a=0.1, n=200$			
0.01	0.945	1.67	0.934	1.89	0.939	1.18	0.927	1.39
0.02	0.952	1.74	0.945	2.01	0.935	1.16	0.927	1.39
0.04	0.944	1.68	0.941	1.99	0.934	1.19	0.932	1.44
0.05	0.942	1.67	0.941	2.01	0.931	1.20	0.933	1.45
0.08	0.939	1.70	0.945	2.06	0.923	1.17	0.927	1.43
0.1	0.937	1.72	0.943	2.09	0.926	1.18	0.923	1.44
0.2	0.907	1.58	0.915	1.95	0.933	1.24	0.924	1.51
0.3	0.898	1.55	0.902	1.92	0.899	1.11	0.902	1.37
0.5	0.864	1.47	0.864	1.83	0.864	1.04	0.858	1.28
	$a=0.5, n=100$				$a=0.5, n=200$			
0.01	0.890	2.43	0.883	2.62	0.908	1.83	0.905	1.99
0.02	0.911	2.65	0.911	2.87	0.913	1.88	0.910	2.06
0.04	0.912	2.69	0.912	2.95	0.927	1.99	0.923	2.20
0.05	0.913	2.73	0.917	2.99	0.929	2.02	0.921	2.23
0.08	0.924	2.87	0.924	3.15	0.924	2.01	0.916	2.22
0.1	0.930	2.92	0.927	3.22	0.920	2.03	0.916	2.24
0.2	0.903	2.76	0.904	3.06	0.920	2.16	0.925	2.39
0.3	0.889	2.71	0.884	3.01	0.892	1.96	0.898	2.18
0.5	0.852	2.52	0.848	2.82	0.858	1.82	0.861	2.04
	$a=0.8, n=100$				$a=0.8, n=200$			
0.01	0.776	4.29	0.769	4.49	0.826	3.57	0.819	3.75
0.02	0.815	4.77	0.807	5.00	0.847	3.76	0.840	3.96
0.04	0.830	5.01	0.823	5.26	0.875	4.16	0.868	4.39
0.05	0.840	5.13	0.833	5.40	0.885	4.29	0.879	4.52
0.08	0.859	5.59	0.856	5.89	0.891	4.40	0.890	4.65
0.1	0.872	5.82	0.869	6.12	0.893	4.51	0.892	4.77
0.2	0.865	5.86	0.861	6.19	0.902	5.00	0.907	5.30
0.3	0.855	5.93	0.854	6.29	0.869	4.61	0.868	4.90
0.5	0.816	5.63	0.823	6.03	0.841	4.35	0.843	4.66

[†]The AR(1) coefficient a is 0.1, 0.5 and 0.8 and r ranges from 0.01 to 0.5.

method deteriorates uniformly as the dependence becomes stronger and improves uniformly as n grows larger.

3.2. Comparison with block bootstrap and normal approximation

In this subsection we shall compare the finite sample coverages of the SN-based approach with the standard normal approximation approach and the bootstrap approach. We consider the same simulation designs as those in Section 3.1, i.e. we generate the data from model (9), where the errors are linear processes with IID $N(0, 1)$ or $t(2)$ innovations. We apply the OLS and MR methods to the $N(0, 1)$ and $t(2)$ cases respectively. We then construct a confidence interval for β_1 by using the following three methods.

- (a) Residual moving block bootstrap without Studentizing: we approximate the sampling distribution of $(\hat{\beta}_{1n} - \beta_1)\sqrt{n}$ by $(\hat{\beta}_{1n}^* - \hat{\beta}_{1n})\sqrt{n}$, where $\hat{\beta}_{1n}^*$ is the estimate of β_1 based on the bootstrap sample $(\mathbf{x}_t, y_t^*)_{t=1}^n$ obtained in the following fashion. Let $\hat{\epsilon}_t = Y_t - \mathbf{x}_t^T \hat{\beta}_n$,

Table 2. Simulated coverage probabilities and average lengths of the self-normalized confidence interval of β_1 for both OLS and MR at 95% nominal level with $t(2)$ innovations†

r	OLS		MR		OLS		MR	
	Coverage	Length	Coverage	Length	Coverage	Length	Coverage	Length
	$a = 0.1, n = 100$				$a = 0.1, n = 200$			
0.01	0.891	3.93	0.939	2.79	0.911	3.16	0.943	1.96
0.02	0.908	4.11	0.943	2.76	0.913	3.21	0.944	1.85
0.04	0.906	4.12	0.936	2.59	0.923	3.37	0.948	1.84
0.05	0.910	4.13	0.936	2.57	0.931	3.50	0.944	1.88
0.08	0.918	4.29	0.939	2.56	0.930	3.44	0.934	1.80
0.1	0.930	4.51	0.944	2.64	0.928	3.44	0.929	1.78
0.2	0.927	4.54	0.932	2.58	0.929	3.46	0.920	1.74
0.3	0.914	4.54	0.917	2.54	0.931	3.74	0.925	1.86
0.5	0.863	3.93	0.862	2.17	0.882	3.33	0.885	1.63
	$a = 0.5, n = 100$				$a = 0.5, n = 200$			
0.01	0.852	5.90	0.885	4.42	0.890	4.89	0.915	3.50
0.02	0.874	6.39	0.907	4.73	0.904	5.13	0.920	3.59
0.04	0.885	6.67	0.913	4.85	0.916	5.50	0.926	3.71
0.05	0.889	6.78	0.913	4.88	0.923	5.73	0.935	3.80
0.08	0.897	7.12	0.917	4.97	0.921	5.72	0.932	3.67
0.1	0.912	7.52	0.922	5.16	0.920	5.74	0.932	3.65
0.2	0.914	7.76	0.919	5.10	0.914	5.84	0.920	3.59
0.3	0.904	7.90	0.906	5.06	0.924	6.29	0.928	3.84
0.5	0.845	6.83	0.859	4.31	0.873	5.58	0.886	3.40
	$a = 0.8, n = 100$				$a = 0.8, n = 200$			
0.01	0.757	11.39	0.778	9.68	0.817	10.22	0.852	8.24
0.02	0.792	12.52	0.812	10.62	0.833	10.79	0.866	8.70
0.04	0.810	13.39	0.830	11.34	0.871	11.97	0.894	9.58
0.05	0.818	13.74	0.836	11.63	0.881	12.33	0.901	9.80
0.08	0.835	14.88	0.861	12.52	0.888	12.70	0.907	9.87
0.1	0.859	16.00	0.878	13.39	0.892	13.06	0.908	10.02
0.2	0.873	17.42	0.887	14.15	0.910	14.62	0.915	10.72
0.3	0.871	17.80	0.882	14.17	0.880	13.57	0.879	9.76
0.5	0.823	15.38	0.830	12.02	0.829	12.50	0.846	8.93

†The AR(1) coefficient a is 0.1, 0.5 and 0.8 and r ranges from 0.01 to 0.5.

$t = 1, \dots, n$, be the residuals. Then $y_t^* = \mathbf{x}_t^T \hat{\beta}_n + e_t^*$, $t = 1, \dots, n$, where $(e_t^*)_{t=1}^n$ is a moving block bootstrap sample from the residuals $(e_t)_{t=1}^n$.

- (b) Normal approximation: we obtain a consistent estimator for $\text{var}(\hat{\beta}_{1n})$ by using the moving block bootstrap estimator $\text{var}^*(\hat{\beta}_{1n}^*)$. In practice, we calculate the bootstrap estimator $\hat{\beta}_{1n}^*$ 500 times and use its sample variance as an estimate. We then approximate $(\hat{\beta}_{1n} - \beta_1)\sqrt{n}/\sqrt{\text{var}^*(\hat{\beta}_{1n}^*)}$ by the standard normal distribution.
- (c) The third method is the SN-based approximation with $r = 0.05$.

In Figs 1 and 2, we plot the empirical coverage probabilities and median lengths of the intervals based on 1000 replications for block sizes $b = 1, 3, 5, 7, 9, 11, 13, 15, 20$. The methods ‘MBB–Nostud’, ‘ $N(0, 1)$ ’ and ‘SN’ in Figs 1 and 2 correspond to schemes (a)–(c) that were described above. From Fig. 1, it can be seen that all methods lead to undercoverage. The coverages for the residual moving block bootstrap without Studentizing are comparable with the normal approximation but are noticeably inferior to the SN-based method. As the correlation grows stronger, the coverages for all methods deteriorate. In contrast, the intervals for the

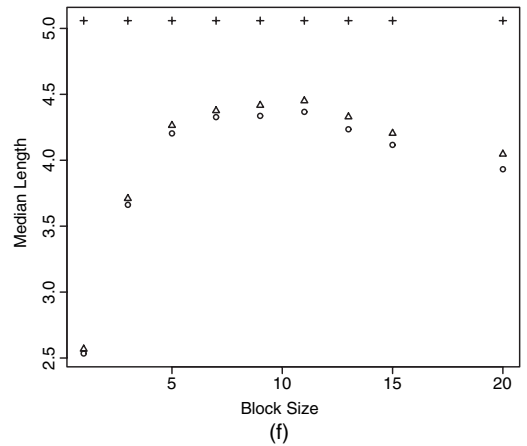
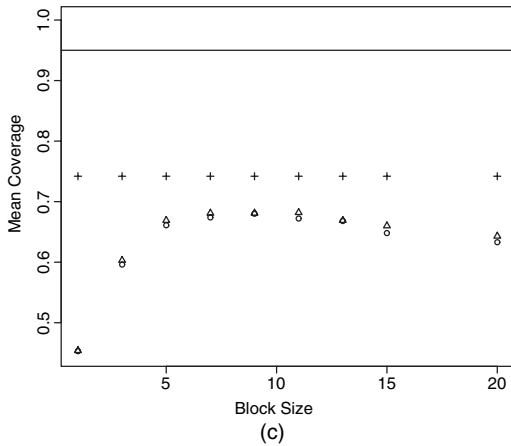
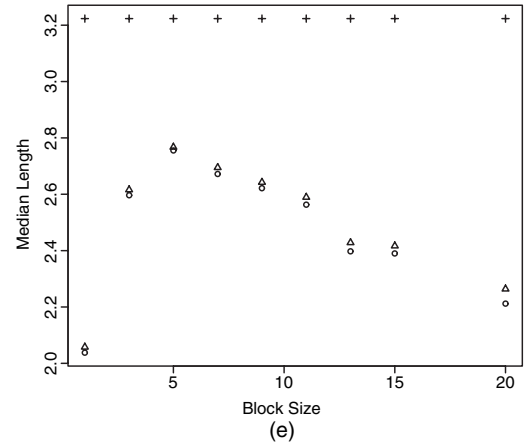
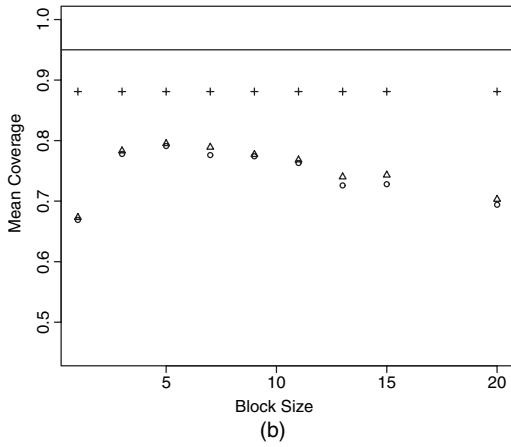
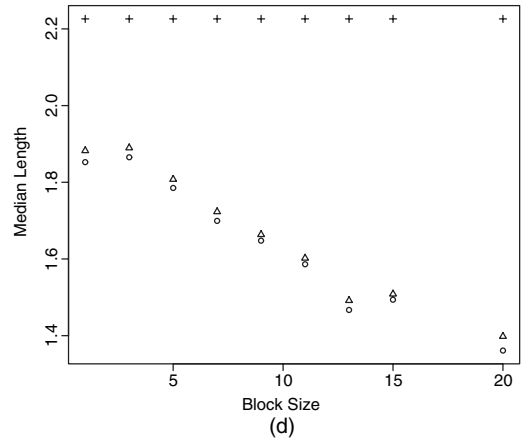
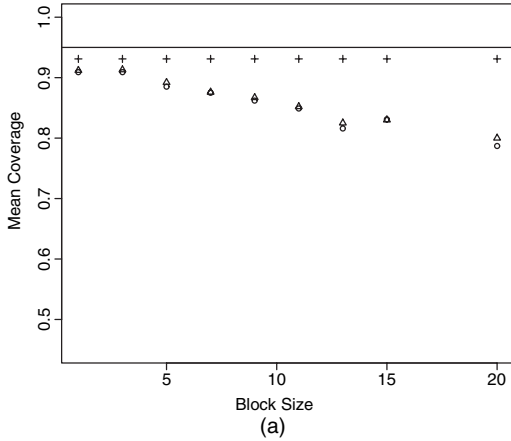


Fig. 1. (a)–(c) Empirical coverage probabilities and (d)–(f) median of interval widths based on the OLS estimator (the sample size $n = 50$ and the number of replications is 1000; \circ , MBB–Nostud method; Δ , normal $N(0, 1)$; $+$, SN method): (a), (d) $\rho = 0.1$; (b), (e) $\rho = 0.5$; (c), (f) $\rho = 0.8$

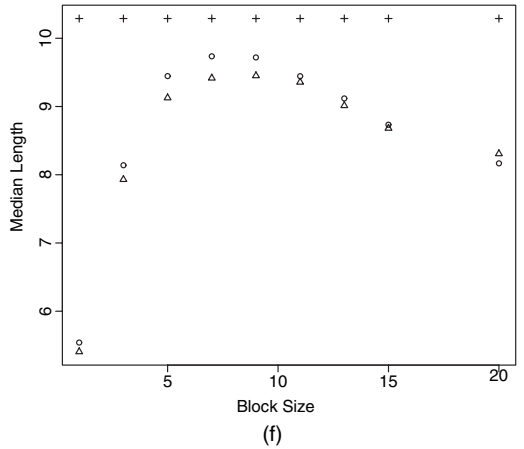
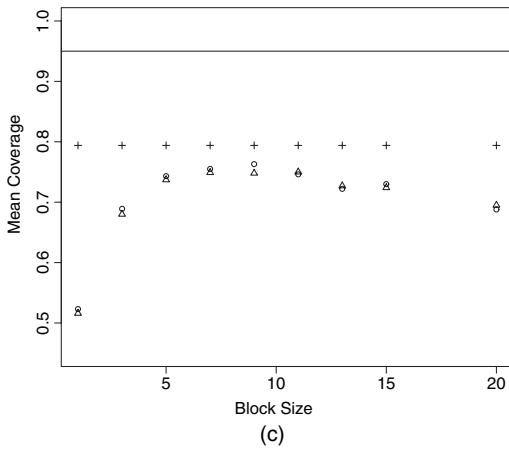
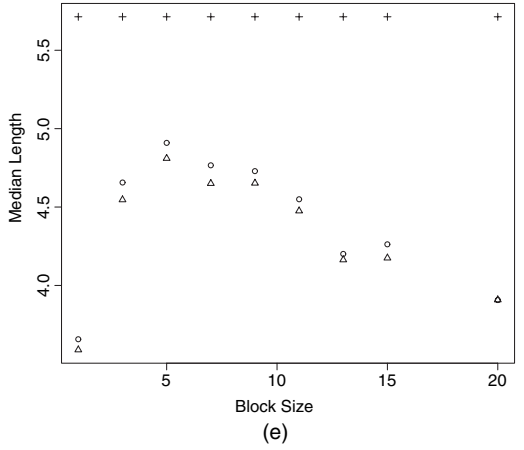
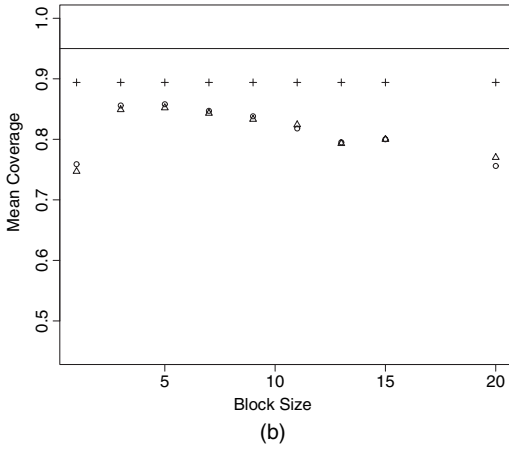
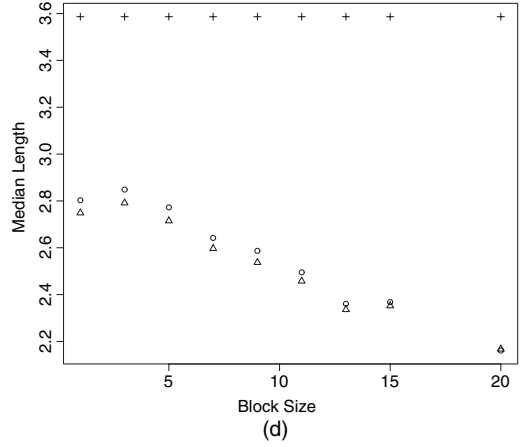
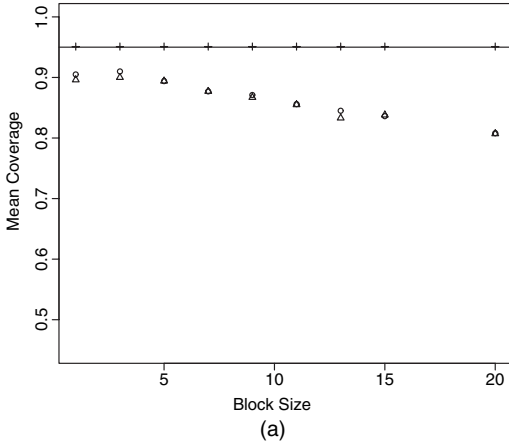


Fig. 2. (a)–(c) Empirical coverage probabilities and (d)–(f) median of interval widths based on the median regression estimator (the sample size $n = 50$ and the number of replications is 1000; \circ , MBB–Nostud; Δ , normal $N(0,1)$; $+$, SN method): (a), (d) $\rho = 0.1$; (b), (e) $\rho = 0.5$; (c), (f) $\rho = 0.8$

normal approximation and residual moving block bootstrap without Studentizing are of similar widths and are shorter than the SN-based intervals. Fig. 2 shows a similar pattern. Overall, the SN-based method delivers more accurate coverages than the normal approximation and the residual block bootstrap without Studentizing at the expense of a wider interval. This finding is consistent with that in Shao (2010) in the stationary time series set-up.

4. Empirical illustrations

In this section, we illustrate the usefulness of our SN-based method by applying it to two real data sets. The first is the luteningizing hormone data provided in Diggle (1990) and the time series

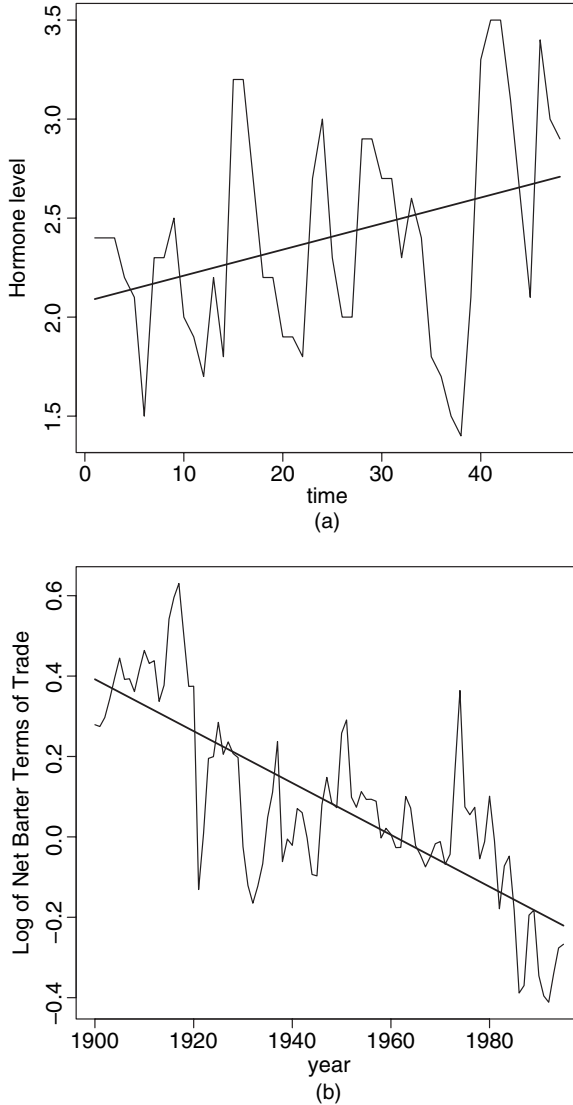


Fig. 3. (a) Plot of hormone level over time and the fitted linear trend based on the least squares method and (b) plot of the logarithm of net barter terms of trade and the fitted linear trend based on the least squares method

Table 3. Confidence intervals for β_1 based on the least squares and MR estimates of model (11) and the SN method for the hormone level data[†]

Method	$\alpha\%/r$	Intervals for the following values of β_1 :				
		0.05	0.10	0.15	0.2	0.25
Least squares	90	[0.21, 1.05]	[0.18, 1.08]	[0.20, 1.06]	[0.20, 1.06]	[0.19, 1.08]
	95	[0.14, 1.14]	[0.06, 1.21]	[0.06, 1.20]	[0.07, 1.19]	[0.09, 1.17]
QR(0.5)	90	[-0.04, 1.00]	[-0.09, 1.05]	[-0.14, 1.10]	[-0.19, 1.15]	[-0.22, 1.18]
	95	[-0.16, 1.12]	[-0.24, 1.20]	[-0.30, 1.26]	[-0.36, 1.32]	[-0.39, 1.35]
QR(0.75)	90	[0.60, 1.24]	[0.57, 1.28]	[0.57, 1.28]	[0.54, 1.30]	[0.55, 1.30]
	95	[0.53, 1.32]	[0.48, 1.37]	[0.48, 1.37]	[0.45, 1.39]	[0.45, 1.39]
QR(0.25)	90	[-0.33, 0.86]	[-0.40, 0.93]	[-0.44, 0.97]	[-0.47, 1.01]	[-0.50, 1.03]
	95	[-0.47, 1.00]	[-0.56, 1.09]	[-0.62, 1.16]	[-0.65, 1.19]	[-0.69, 1.23]

[†]QR(α) stands for quantile regression at the α th quantile.

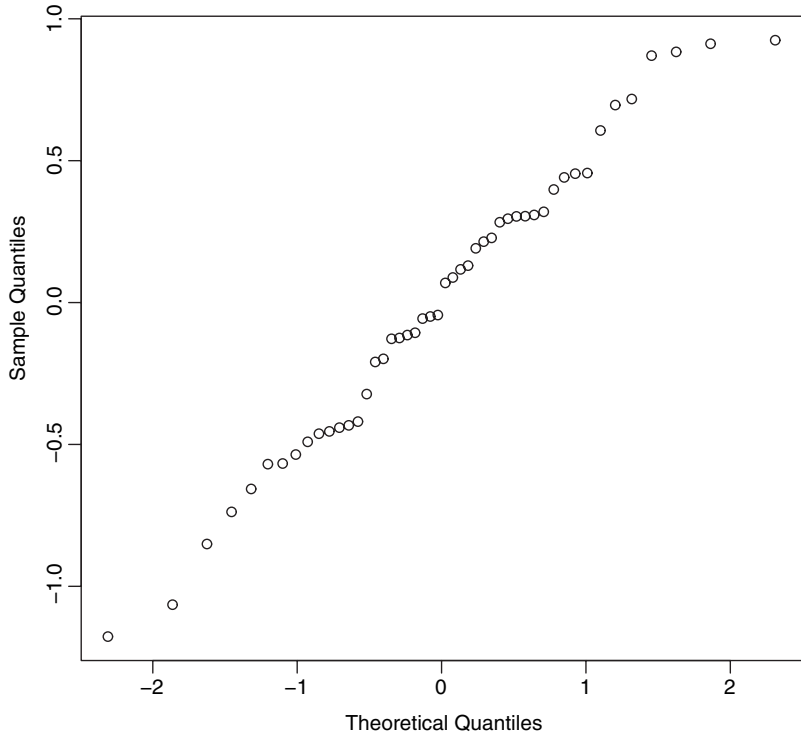
are hormone levels measured on a single woman every 10 min for 8 h; Fig. 3(a). As mentioned in Sherman (1997), a problem of interest is to assess whether the hormone level changes over time for this individual. Sherman (1997) fitted least square regression to the model $Y_t = \alpha_0 + \alpha_1 t + e_t$, where Y_t stands for the observed hormone level at time t , and e_t is a stationary error sequence. He applied the subsampling-based method to construct a confidence interval for α_1 . Again, there is a practical issue with the choice of the subsampling width, which seems not well addressed in Sherman (1997).

To remedy the problem, we consider the linear model

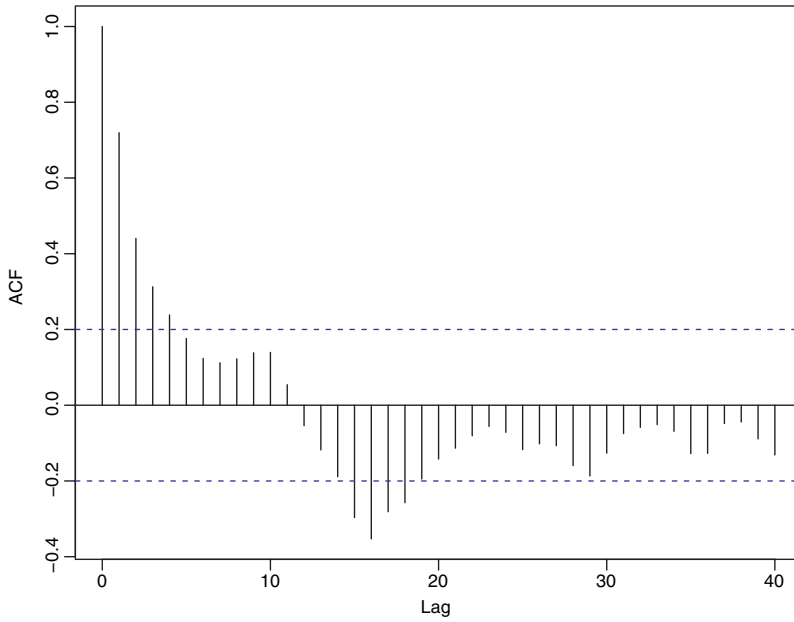
$$Y_t = \beta_0 + \beta_1 \frac{t}{n} + e_t, \quad t = 1, \dots, n = 48, \tag{11}$$

with the fixed time covariates t/n . It is not difficult to see that our assumption on the fixed regressors is satisfied. The least squares estimate for β_1 is 0.63. An inspection of the auto-correlation plot (which is not shown) of the residuals from the least squares fit suggests that there is a significant auto-correlation at lag 1. So the standard error that is provided in the output of the `lm` command in R (<http://stat.ethz.ch/R-manual/R-patched/library/stats/html/lm.html>) is misleading, as independent error is assumed. Table 3 shows the 90% and 95% SN-based confidence intervals for β_1 when $r = 0.05, \dots, 0.25$ and OLS and quantile regression with $\alpha = 0.25, 0.5, 0.75$ are used. It appears that all the OLS-based intervals exclude zero, suggesting that $\beta_1 \neq 0$ at the 5% level of significance. Also the intervals seem quite stable for $r = 0.1, 0.15, 0.2, 0.25$. Therefore, our results based on least squares regression indicate that there is a positive linear trend over time. In contrast, all the MR-based intervals include zero for all r s; thus the trend is not significant at the 10% level. This difference is presumably due to the inefficiency of the MR estimate when the distribution of e_t is close to normal, which seems to be so as seen from the normal $Q-Q$ -plot in Fig. 4(a). Furthermore, the intervals for quantile regression with $\alpha = 0.25$ and $\alpha = 0.75$ indicate that the upward trend is significant for upper quantiles but not for lower quantiles. This new finding nicely illustrates the usefulness of quantile regression and the associated SN-based inference. Note that most of the existing inference methods for trend with dependent SN errors are based on least squares estimation. The proposed SN-based approach can be readily used if the trend of the conditional quantile is of interest and it provides a convenient tool for practitioners.

We further apply our SN-based inference method to the logarithm of the annual net barter



(a)



(b)

Fig. 4. (a) Normal quantile-quantile plot of the residuals from the least squares fit to the model (11) for hormone level and (b) auto-correlation plot of the residuals from the least squares fit to the logarithm of net barter terms of trade

Table 4. Confidence intervals for β_1 based on the least squares and quantile regression estimates of model (11) and the SN method for the net barter terms of trade data[†]

Method	$\alpha\%/r$	Intervals for the following values of β_1 :				
		0.05	0.10	0.15	0.2	0.25
Least squares	90	[-0.81, -0.43]	[-0.83, -0.41]	[-0.84, -0.40]	[-0.84, -0.40]	[-0.83, -0.41]
	95	[-0.86, -0.38]	[-0.89, -0.35]	[-0.90, -0.34]	[-0.89, -0.35]	[-0.89, -0.35]
	99	[-0.95, -0.29]	[-0.97, -0.27]	[-1.00, -0.24]	[-0.98, -0.26]	[-0.97, -0.26]
QR(0.5)	90	[-0.85, -0.48]	[-0.87, -0.46]	[-0.88, -0.45]	[-0.86, -0.47]	[-0.83, -0.50]
	95	[-0.91, -0.43]	[-0.93, -0.40]	[-0.94, -0.40]	[-0.91, -0.42]	[-0.88, -0.46]
	99	[-0.99, -0.34]	[-1.01, -0.32]	[-1.03, -0.30]	[-0.99, -0.34]	[-0.94, -0.39]
QR(0.75)	90	[-0.88, -0.40]	[-0.90, -0.37]	[-0.93, -0.35]	[-0.92, -0.35]	[-0.87, -0.40]
	95	[-0.94, -0.33]	[-0.98, -0.29]	[-1.00, -0.27]	[-0.99, -0.28]	[-0.93, -0.34]
	99	[-1.05, -0.22]	[-1.09, -0.19]	[-1.13, -0.15]	[-1.11, -0.16]	[-1.03, -0.25]
QR(0.25)	90	[-0.78, -0.38]	[-0.80, -0.36]	[-0.82, -0.34]	[-0.83, -0.34]	[-0.82, -0.34]
	95	[-0.84, -0.33]	[-0.87, -0.30]	[-0.88, -0.28]	[-0.88, -0.28]	[-0.89, -0.28]
	99	[-0.93, -0.23]	[-0.96, -0.21]	[-0.99, -0.18]	[-0.98, -0.18]	[-0.98, -0.18]

[†]QR(α) stands for quantile regression at the α th quantile.

terms of trade series (see Fig. 3(b)), constructed by Grilli and Yang (1988) and Lutz (1999), and analysed in Bunzel and Vogelsang (2005) and Harvey *et al.* (2007). The goal of this analysis is to check the so-called Prebisch–Singer hypothesis, which asserts that the net barter terms of trade should be falling over time. As seen from Table 4, which has the same format as Table 3, all the intervals fall on the left-hand side of zero, thus convincingly showing that the negative trend is significant at the 1% level. The auto-correlation plot of the least squares residual (see Fig. 4(b)) suggests that the error is stationary in the sense that no unit root exists, since the correlation decays exponentially for the first few lags and becomes non-significant from zero at large lags. Our conclusion is in accordance with those reached by Bunzel and Vogelsang (2005) and Harvey *et al.* (2007).

5. Discussion

In this paper, we propose an extension of Shao’s SN approach to the regression setting. Our model allows for fixed regressors with various types of design but excludes the random-design case, which was studied by Lee (2006). A uniform Bahadur representation for recursive M -estimators is obtained under some verifiable technical conditions. Simulation results suggest that the SN-based approach delivers better coverage at the expense of a wider interval, compared with the block bootstrap and normal approximation approaches. It is worth noting that, in our extension of the SN method to the regression problem, we need to introduce a user-chosen number $r \in (0, 1)$; see theorem 2 and proposition 2. In practice, the choice of r is similar to the choice of bandwidth(s) as required in the above-mentioned traditional approaches, and different choices of r (bandwidths) may lead to different inference results. However, in theory the choice of r is reflected in the limiting distribution of the SN statistic and associated simulation-based approximation, whereas the choice of bandwidths is typically not captured by the usual standard normal limiting distribution (i.e. two different bandwidths are used by two users with $N(0, 1)$ as the same reference distribution). In our limited Monte Carlo experiments, we find that the choice of $r = 0.1$ performs reasonably well for both MR and OLS methods and for any combination of error type (with varying degree of dependence) in the linear model and sample

size. The optimal data-driven choice of r , which presumably depends on the design matrix and the error dependence, seems to be a very challenging topic and is left for future research.

Acknowledgements

Shao acknowledges partial financial support from National Science Foundation grants DMS-0804937 and DMS-1104545. Zhou's research was supported in part by the National Science and Engineering Research Council of Canada. We thank Professor Helle Bunzel and Professor Matthias Lutz for providing us with the trade data set that is used in the paper. We are also grateful to the Associate Editor and two referees for constructive comments that led to a substantial improvement of the paper.

Appendix A

Without loss of generality, throughout this section we assume that the true parameter $\beta = 0$.

A.1. Proof of proposition 1

Note that $\Sigma_{\lfloor sn \rfloor} = \sum_{j=1}^{\lfloor sn \rfloor} \mathbf{x}_j^\top \mathbf{x}_j$. Let j_s be the largest j , such that $s_j < s$. Then

$$\Sigma_{\lfloor sn \rfloor} = \sum_{k=1}^{s_j} n(s_k - s_{k-1}) \int_{s_{k-1}}^{s_k} \mathbf{f}_{k-1}^\top(x) \mathbf{f}_{k-1}(x) dx + n(s - s_j) \int_{s_j}^s \mathbf{f}_j^\top(x) \mathbf{f}_j(x) dx + O(m).$$

Therefore $\Sigma_{\lfloor sn \rfloor}/sn = \Sigma(s) + O(1/\sqrt{n})$, where

$$\Sigma(s) = \sum_{k=1}^{s_j} (s_k - s_{k-1}) \int_{s_{k-1}}^{s_k} \mathbf{f}_{k-1}^\top(x) \mathbf{f}_{k-1}(x) dx/s + (s - s_j) \int_{s_j}^s \mathbf{f}_j^\top(x) \mathbf{f}_j(x) dx/s.$$

Note that $\max_{0 \leq i \leq m} \sup_{s_i \leq t \leq s_{i+1}} |\mathbf{f}_i^\top(x) \mathbf{f}_i(x)| \leq L$. Therefore

$$|\Sigma_{\lfloor tm \rfloor} - \Sigma_{\lfloor sn \rfloor}| \leq L(\lfloor tm \rfloor - \lfloor sn \rfloor).$$

Observe that, for any $t \geq r$, $(\Sigma_{\lfloor tm \rfloor} - \Sigma_{\lfloor sn \rfloor})/n$ is positive semidefinite. Hence it is easy to see that, if $\Sigma(r)$ is non-singular, then $\Sigma(\cdot)$ is bounded away from 0 on $[r, 1]$.

We now prove condition 6. By the Lipschitz continuity of $\mathbf{f}_j(\cdot)$, if $s_{j-1} < t/n < (t+1)/n \leq s_j$ for some j , then $|\mathbf{x}_t - \mathbf{x}_{t+1}| = O(1/n)$. Otherwise $|\mathbf{x}_t - \mathbf{x}_{t+1}| = O(1)$. However, there are at most m jumps in the sequence $\{x_t\}$. Hence

$$\sum_{t=1}^{n-1} |\mathbf{x}_t - \mathbf{x}_{t+1}| + |\mathbf{x}_n| \leq C \left\{ (n-m) \frac{1}{n} + m + 1 \right\} = O(n^{1/4-\delta}). \quad \square$$

Consider the transformed model

$$Y_t = \mathbf{z}_{t,n}^\top \theta_n + e_t, \quad \mathbf{z}_{t,n} = \Sigma_n^{-1/2} \mathbf{x}_t, \quad \theta_n = \Sigma_n^{1/2} \beta. \quad (12)$$

Observe that $\hat{\theta}_t := \Sigma_n^{1/2} \hat{\beta}_t$ is a minimizer of $\sum_{j=1}^t \rho(e_j - \mathbf{z}_{j,n}^\top \theta)$ and that, by definition, $\mathbf{z}_{t,n} = (z_{t,1,n}, z_{t,2,n}, \dots, z_{t,p,n})^\top$ satisfies $\sum_{t=1}^n \mathbf{z}_{t,n} \mathbf{z}_{t,n}^\top = \text{Id}_p$, the $p \times p$ identity matrix. For $q > 0$ let $\zeta_n(q) = \sum_{t=1}^n |\mathbf{z}_{t,n}|^q$. Then $\zeta_n(2) = p$. Define

$$\Omega_t(\theta) = \sum_{j=1}^t \psi(e_j - \mathbf{z}_{j,n}^\top \theta) \mathbf{z}_{j,n},$$

and $K_t(\theta) = \Omega_t(\theta) - \mathbb{E}[\Omega_t(\theta)]$.

Write $K_t(\theta) = M_t(\theta) + N_t(\theta)$, where

$$M_t(\theta) = \sum_{j=1}^t \{\psi(e_j - \mathbf{z}_{j,n}^\top \theta) - \mathbb{E}[\psi(e_j - \mathbf{z}_{j,n}^\top \theta) | \mathcal{F}_{j-1}]\} \mathbf{z}_{j,n},$$

$$N_t(\theta) = \sum_{j=1}^t \{\mathbb{E}[\psi(e_j - \mathbf{z}_{j,n}^\top \theta) | \mathcal{F}_{j-1}] - \mathbb{E}[\psi(e_j - \mathbf{z}_{j,n}^\top \theta)]\} \mathbf{z}_{j,n}.$$

Lemma 1. Under conditions 4 and 6, we have for sufficiently large n

$$r_n = O(n^{-1/4-\delta}),$$

$$n^{-1/2} = O(r_n),$$

where $r_n = \max_{1 \leq t \leq n} |\mathbf{z}_{t,n}|$ and $\delta > 0$. Additionally, $\zeta_n(2+q) = O(r_n^q)$ for all $q \geq 0$.

Proof. Note that $\sum_{i=1}^n |\mathbf{z}_{i,n}|^2 = p$. Therefore $|r_n| \geq n^{-1/2} \sqrt{p}$. In contrast, note that $|\mathbf{x}_n| + \sum_{i=k}^{n-1} |\mathbf{x}_i - \mathbf{x}_{i+1}| \geq |\mathbf{x}_k|$. Therefore condition 6 implies that $\max_{1 \leq t \leq n} |\mathbf{x}_t| = O(n^{1/4-\delta})$ for some $\delta > 0$. Additionally, condition 4 implies that $\Sigma_n^{-1} = O(n)$. Therefore

$$r_n = \max_{1 \leq t \leq n} |\mathbf{z}_{t,n}| = \max_{1 \leq t \leq n} |\Sigma_n^{-1/2} \mathbf{x}_t| = O(n^{-1/4-\delta}).$$

By the fact that

$$\zeta_n(2+q) = \sum_{i=1}^n |\mathbf{z}_{i,n}|^q \leq r_n^q \sum_{i=1}^n |\mathbf{z}_{i,n}|^2 = p r_n^q,$$

we have $\zeta_n(2+q) = O(r_n^q)$.

Lemma 2. Assume condition 7. Let (δ_n) be a sequence of deterministic positive numbers such that

$$\delta_n \rightarrow \infty,$$

$$\delta_n r_n = \delta_n \max_{1 \leq t \leq n} |\mathbf{z}_{t,n}| \rightarrow 0. \quad (13)$$

Then we have

$$\max_{1 \leq t \leq n} \sup_{|\theta| \leq \delta_n} |M_t(\theta) - M_t(0)| = O_{\mathbb{P}} \log(n) \{\sqrt{\tau_n(\delta_n)} + n^{-3}\}, \quad (14)$$

where $\tau_n(\delta) = \sum_{i=1}^n |\mathbf{z}_{i,n}|^2 \{m^2(|\mathbf{z}_{i,n}| \delta) + m^2(-|\mathbf{z}_{i,n}| \delta)\}$.

By the same martingale exponential inequality and chaining techniques as were used in lemma 4 of Wu (2007a), lemma 2 can be proved. The details are omitted.

Lemma 3. Assume conditions 8 and 13. Then

$$\| \max_{1 \leq t \leq n} \sup_{|\theta| \leq \delta_n} |N_t(\theta) - N_t(0)| \|_4 = O\{n^{1/4} \delta_n \sqrt{\zeta_n(4)}\}.$$

Proof. Let $I = \{\alpha_1, \alpha_2, \dots, \alpha_q\} \subset \{1, 2, \dots, p\}$ be a non-empty set and $1 \leq \alpha_1 < \dots < \alpha_q$. For $\mathbf{u} = (u_1, \dots, u_p)$, let $\mathbf{u}_I = (u_1 \mathbf{1}_{1 \in I}, \dots, u_p \mathbf{1}_{p \in I})$. Define

$$\int_0^{\theta_I} \frac{\partial^q N_t(\mathbf{u}_I)}{\partial \mathbf{u}_I} d\mathbf{u}_I = \int_0^{\theta_{\alpha_1}} \dots \int_0^{\theta_{\alpha_q}} \frac{\partial^q N_t(\mathbf{u}_I)}{\partial u_{\alpha_1} \dots \partial u_{\alpha_q}} du_{\alpha_1} \dots du_{\alpha_q}.$$

Note that

$$\left| \frac{\partial^q N_t(\mathbf{u}_I)}{\partial \mathbf{u}_I} d\mathbf{u}_I \right| = \left| \sum_{j=1}^t \{\psi_1^{(q)}(-\mathbf{z}_{j,n} \mathbf{u}_I; \mathcal{F}_{j-1}) - \varphi^{(q)}(-\mathbf{z}_{j,n}^\top \mathbf{u}_I)\} \mathbf{w}_j \right|, \quad (15)$$

where $\mathbf{w}_j = \mathbf{z}_{j,n} z_{j,\alpha_1} \dots z_{j,\alpha_q}$. Let $|\mathbf{u}| \leq p \delta_n$ and $k \in \mathbb{N}$. Since $\max_{j \leq n} |\mathbf{z}_{j,n} \mathbf{u}| \leq p r_n \delta_n$, we have by Burkholder's inequality

$$\left\| \left\{ \sum_{j=1}^t \mathcal{P}_{j-k} \psi_1^{(q)}(-\mathbf{z}_{j,n} \mathbf{u}_I; \mathcal{F}_{j-1}) \mathbf{w}_j \right\} \right\|_4^2 \leq C \left\| \sum_{j=1}^t |\mathcal{P}_{j-k} \psi_1^{(q)}(-\mathbf{z}_{j,n} \mathbf{u}_I; \mathcal{F}_{j-1}) \mathbf{w}_j|^2 \right\|$$

$$\begin{aligned} &\leq C \sum_{j=1}^t \|\mathcal{P}_{j-k} \psi_1^{(q)}(-\mathbf{z}_{j,n} \mathbf{u}_l; \mathcal{F}_{j-1})\|_4^2 |\mathbf{w}_j|^2 \\ &= \mathcal{P}_0 \psi_1^{(q)}(-\mathbf{z}_{j,n} \mathbf{u}_l; \mathcal{F}_{k-1})\|_4^2 \sum_{j=1}^t |\mathbf{w}_j|^2. \end{aligned}$$

Note that $\sum_{j=1}^t |\mathbf{w}_j|^2 \leq \sum_{j=1}^n |\mathbf{w}_j|^2 \leq \zeta_n (2+2q)$ and

$$\sum_{j=1}^t \{\psi_1^{(q)}(-\mathbf{z}_{j,n} \mathbf{u}_l; \mathcal{F}_{j-1}) - \varphi^{(q)}(-\mathbf{z}_{j,n}^T \mathbf{u}_l)\} \mathbf{w}_j = \sum_{k=1}^{\infty} \sum_{j=1}^t \mathcal{P}_{j-k} \psi_1^{(q)}(-\mathbf{z}_{j,n} \mathbf{u}_l; \mathcal{F}_{j-1}) \mathbf{w}_j.$$

By lemma 1 of Wu (2007a) and condition 8, we have

$$\left\| \frac{\partial^q N_t(\mathbf{u}_l)}{\partial \mathbf{u}_l} \right\|_4 \leq C \sqrt{\zeta_n (2+2q)} \quad \text{uniformly over } |\mathbf{u}| \leq p \delta_n. \quad (16)$$

Since

$$\left\| \max_{1 \leq t \leq n} \left\| \frac{\partial^q N_t(\mathbf{u}_l)}{\partial \mathbf{u}_l} \right\|_4 \right\|_4 \leq \sum_{t=1}^n \left\| \left\| \frac{\partial^q N_t(\mathbf{u}_l)}{\partial \mathbf{u}_l} \right\|_4 \right\|_4,$$

we have by expression (16) that

$$\left\| \max_{1 \leq t \leq n} \left\| \frac{\partial^q N_t(\mathbf{u}_l)}{\partial \mathbf{u}_l} \right\|_4 \right\|_4 \leq C n^{1/4} \sqrt{\zeta_n (2+2q)} \quad \text{uniformly over } |\mathbf{u}| \leq p \delta_n. \quad (17)$$

Consequently,

$$\begin{aligned} \left\| \max_{1 \leq t \leq n} \sup_{|\theta| \leq \delta_n} \int_0^{\theta_l} \left\| \frac{\partial^q N_t(\mathbf{u}_l)}{\partial \mathbf{u}_l} \right\|_4 d\mathbf{u}_l \right\|_4 &\leq \int_{-\delta_n}^{\delta_n} \dots \int_{-\delta_n}^{\delta_n} \left\| \max_{1 \leq t \leq n} \left\| \frac{\partial^q N_t(\mathbf{u}_l)}{\partial \mathbf{u}_l} \right\|_4 \right\|_4 d\mathbf{u}_l \\ &\leq C n^{1/4} \delta_n^q \sqrt{\zeta_n (2+2q)}. \end{aligned}$$

By expression (13), $\delta_n^q \sqrt{\zeta_n (2+2q)} = O\{\delta_n \sqrt{\zeta_n (4)}\}$. So the lemma follows from the identity

$$N_t(\theta) - N_t(0) = \sum_{l \in \{1, 2, \dots, p\}} \int_0^{\theta_l} \frac{\partial^{l|} N_t(\mathbf{u}_l)}{\partial \mathbf{u}_l} d\mathbf{u}_l.$$

Corollary 1. Under expression (13) and conditions 7 and 8, we have

$$\max_{1 \leq t \leq n} \sup_{|\theta| \leq \delta_n} |K_t(\theta) - K_t(0)| = O_{\mathbb{P}}\{\delta_n^\lambda \log(n) \sqrt{\zeta_n (2+2\lambda)} + n^{1/4} \delta_n \sqrt{\zeta_n (4)}\},$$

since $K_t(\theta) = M_t(\theta) + N_t(\theta)$. Corollary 1 follows from lemmas 2 and 3 and condition 3.

Lemma 4. Assume conditions 4, 6 and 9. Then, on a richer probability space, there exist IID Gaussian random variables V_1, V_2, \dots, V_n such that

$$\max_{1 \leq t \leq n} \left| \sum_{j=1}^t \mathbf{z}_{j,n} \psi(e_j) - \sigma_\infty \sum_{j=1}^t \mathbf{z}_{j,n} V_j \right| = o(1), \quad \text{almost surely,}$$

where $\sigma_\infty^2 = \sum_{i \in \mathbb{Z}} \text{cov}\{\psi(e_0), \psi(e_i)\}$.

Proof. Let $S_e(t) = \sum_{j=1}^t \psi(e_j)$. By theorem 3 of Wu (2007b), on a richer probability space, there exist IID Gaussian random variables V_1, V_2, \dots, V_n , such that

$$\max_{1 \leq t \leq n} |S_e(t) - \sigma_\infty S_v(t)| = O\{n^{1/4} \log(n)\} \quad \text{almost surely} \quad (18)$$

where $S_v(t) = \sum_{j=1}^t V_j$. Let $\tilde{V}_t = \psi(e_t) - \sigma_\infty V_t$ and $S_{\tilde{V}}(t) = \sum_{j=1}^t \tilde{V}_j$. Then

$$\begin{aligned}
\max_{1 \leq t \leq n} \left| \sum_{j=1}^t \mathbf{z}_{j,n} \psi(e_j) - \sigma_\infty \sum_{j=1}^t \mathbf{z}_{j,n} V_j \right| &= \max_{1 \leq t \leq n} \left| \sum_{j=1}^t \mathbf{z}_{j,n} \tilde{V}_j \right| \\
&= \max_{1 \leq t \leq n} \left| \sum_{j=1}^{t-1} S_{\tilde{V}}(j)(\mathbf{z}_{j,n} - \mathbf{z}_{j+1,n}) + S_{\tilde{V}}(t) \mathbf{z}_{t,n} \right| \\
&\leq \max_{1 \leq t \leq n} |S_{\tilde{V}}(t)| \max_{1 \leq t \leq n} \left(\sum_{j=1}^{t-1} |\mathbf{z}_{j,n} - \mathbf{z}_{j+1,n}| + |\mathbf{z}_{t,n}| \right) \\
&= O\{n^{1/4} \log(n) n^{1/4-\delta} / n^{1/2}\} \quad \text{almost surely} \\
&= o(1) \quad \text{almost surely.}
\end{aligned}$$

The lemma follows.

Lemma 5. Under the conditions of lemmas 2–4, we have

$$\max_{\lfloor rn \rfloor \leq t \leq n} |\hat{\theta}_t| = O_{\mathbb{P}}\{\log(n)\}.$$

Recall that $\hat{\theta}_t = \Sigma_n^{1/2} \hat{\beta}_t$.

Proof. Define

$$\begin{aligned}
\Lambda_t(\theta) &= \sum_{j=1}^t \{\rho(e_j - \theta^T \mathbf{z}_{j,n}) - \rho(e_j)\}, \\
\Xi_t(\theta) &= \Lambda_t(\theta) + \sum_{j=1}^t \theta^T \mathbf{z}_{j,n} \psi(e_j) := \Lambda_t(\theta) + \theta^T \Psi_t.
\end{aligned} \tag{19}$$

Note that $\Xi_t(\theta) = -\theta^T \int_0^1 \{\Omega_t(s\theta) - \Omega_t(0)\} ds$. Let δ_n satisfy condition (13). Corollary 1 then implies that

$$\begin{aligned}
\max_{1 \leq t \leq n} \sup_{|\theta| \leq \delta_n} |\Xi_t(\theta) - \mathbb{E}[\Xi_t(\theta)]| &= O_p\{\delta_n^{\lambda+1} \log(n) \sqrt{\zeta_n} (2+2\lambda) + n^{1/4} \delta_n^2 \sqrt{\zeta_n} (4)\} \\
&= O_p\{r_n^\lambda \delta_n^{\lambda+1} \log(n) + n^{1/4} r_n \delta_n^2\}.
\end{aligned} \tag{20}$$

By condition 4 and arguments similar to those used in proposition 7 in Wu (2007a), we have

$$\max_{\lfloor rn \rfloor \leq t \leq n} \sup_{|\theta| \leq \delta_n} \left| \mathbb{E}[\Xi_t(\theta)] - \frac{\varphi'(0)}{2} \theta^T \Sigma^{-1/2}(1) \Sigma\left(\frac{t}{n}\right) \Sigma^{-1/2}(1) \theta \right| = o(\delta_n^2). \tag{21}$$

In contrast, lemma 4 and condition 4 imply that

$$\max_{1 \leq t \leq n} |\Psi_t| = O_{\mathbb{P}}\{\log(n)\}. \tag{22}$$

For a sequence $c_n \rightarrow \infty$, let $\tilde{c}_n = \min(c_n, r_n^{-1/3})$ and $\tilde{\delta}_n = \tilde{c}_n \log(n)$. Then it is straightforward to check that this $\tilde{\delta}_n$ satisfies expression (13). Therefore we have by expressions (20)–(22) and lemma 1 that

$$\mathbb{P}\left\{ \max_{\lfloor rn \rfloor \leq t \leq n} \sup_{|\theta| = \tilde{\delta}_n} \Lambda_t(\theta) \leq 0 \right\} \rightarrow 0. \tag{23}$$

Note that $\Lambda_t(\theta)$ is convex in θ and $\Lambda_t(0) = 0$. Hence, for any θ such that $|\theta| > \tilde{\delta}_n$, we have $\Lambda_t(\theta) \geq (|\theta|/\tilde{\delta}_n) \Lambda_t(\tilde{\delta}_n \theta/|\theta|)$. Therefore expression (23) implies that

$$\mathbb{P}\left\{ \max_{\lfloor rn \rfloor \leq t \leq n} \sup_{|\theta| \geq \tilde{\delta}_n} \Lambda_t(\theta) \leq 0 \right\} \rightarrow 0.$$

Therefore the lemma follows from the definition of $\hat{\theta}_t$ and the fact that (c_n) can approach ∞ arbitrarily slowly.

Proposition 3. Under the conditions of lemmas 2–5, we have

$$\sup_{r \leq s \leq 1} \left| \varphi'(0)sn \Sigma(s) \hat{\beta}_{\lfloor sn \rfloor} - \sum_{j=1}^{\lfloor sn \rfloor} \psi(e_j) \mathbf{x}_j \right| = o_{\mathbb{P}}(\sqrt{n}).$$

Proof. Note that $K_t(0) = \sum_{j=1}^t \psi(e_j) \mathbf{z}_{j,n}$. By condition 5,

$$\max_{\lfloor rn \rfloor \leq t \leq n} \left| K_t(\hat{\theta}_t) + \sum_{j=1}^t \varphi(-\mathbf{z}_{j,n}^T \hat{\theta}_t) \mathbf{z}_{j,n} \right| = O_{\mathbb{P}}(r_n).$$

Since $\mathbf{z}_{j,n}^T \hat{\theta}_t = o_{\mathbb{P}}(1)$ uniformly on $[rn, n]$, we have by condition 2 that

$$\varphi(-\mathbf{z}_{j,n}^T \hat{\theta}_t) = -\varphi'(0) \mathbf{z}_{j,n}^T \hat{\theta}_t + O\{(\mathbf{z}_{j,n}^T \hat{\theta}_t)^2\},$$

uniformly on $t \in [rn, n]$. By expression (12), condition 4, lemma 5 and corollary 1, we have

$$\sup_{r \leq s \leq 1} \left| \varphi'(0)sn \Sigma(s) \hat{\beta}_{\lfloor sn \rfloor} - \sum_{j=1}^{\lfloor sn \rfloor} \psi(e_j) \mathbf{x}_j \right| = O_{\mathbb{P}}[\{r_n^\lambda \log(n)^{\lambda+1} + n^{1/4} r_n \log(n)\} \sqrt{n}].$$

By lemma 1, $r_n^\lambda \log(n)^{\lambda+1} + n^{1/4} r_n \log(n) = o(1)$. Therefore lemma 5 follows.

Lemma 6. Recall that V_1, V_2, \dots, V_n was defined in lemma 4. Then, under condition 4, we have

$$\{\tilde{\Psi}_{\lfloor sn \rfloor}\}_{r \leq s \leq 1} \Rightarrow \{\Upsilon_s\}_{r \leq s \leq 1},$$

where $\tilde{\Psi}_t = \sum_{j=1}^t V_j \mathbf{x}_j / \sqrt{n}$ and $\{\Upsilon_s\}$ is a mean 0 Gaussian process with covariance $\text{cov}(\Upsilon_{s_1}, \Upsilon_{s_2}) = \Sigma\{\min(s_1, s_2)\}$.

Proof. The finite dimensional convergence is straightforward. We now prove tightness. For $r \leq s_1 < s_2 < s_3 \leq 1$, if $\lfloor ns_3 \rfloor - \lfloor ns_1 \rfloor \geq 2$, we have

$$\begin{aligned} \mathbb{E}[|\tilde{\Psi}_{\lfloor s_1 n \rfloor} - \tilde{\Psi}_{\lfloor s_2 n \rfloor}|^2 |\tilde{\Psi}_{\lfloor s_2 n \rfloor} - \tilde{\Psi}_{\lfloor s_3 n \rfloor}|^2] &= \frac{|\text{tr}(\Sigma_{\lfloor s_1 n \rfloor}) - \text{tr}(\Sigma_{\lfloor s_2 n \rfloor})| |\text{tr}(\Sigma_{\lfloor s_2 n \rfloor}) - \text{tr}(\Sigma_{\lfloor s_3 n \rfloor})|}{n^2} \\ &\leq \frac{|\text{tr}(\Sigma_{\lfloor s_3 n \rfloor}) - \text{tr}(\Sigma_{\lfloor s_1 n \rfloor})|^2}{n^2} \\ &\leq L^2 (\lfloor ns_3 \rfloor - \lfloor ns_1 \rfloor)^2 / n^2 \\ &\leq (2L)^2 (s_3 - s_1)^2. \end{aligned}$$

However, if $\lfloor ns_3 \rfloor - \lfloor ns_1 \rfloor$ equals 1 or 0, then $|\tilde{\Psi}_{\lfloor s_1 n \rfloor} - \tilde{\Psi}_{\lfloor s_2 n \rfloor}|^2 |\tilde{\Psi}_{\lfloor s_2 n \rfloor} - \tilde{\Psi}_{\lfloor s_3 n \rfloor}|^2 = 0$. Hence, by theorem 13.5 in Billingsley (1999), lemma 6 follows.

A.2. Proof of theorem 2

Equation (7) follows from proposition 3. Furthermore, proposition 3 and lemma 4 imply that

$$\sup_{r \leq s \leq 1} \left| \varphi'(0)s \Sigma(s) \hat{\beta}_{\lfloor sn \rfloor} \sqrt{n} - \sigma_\infty \sum_{j=1}^{\lfloor sn \rfloor} V_j \mathbf{x}_j / \sqrt{n} \right| = o_{\mathbb{P}}(1). \quad (24)$$

Theorem 2 follows easily from lemma 6 and the continuous mapping theorem.

A.3. Proof of theorem 1

Letting $s = 1$ in equation (24), we have

$$\varphi'(0)\Sigma(1)(\hat{\beta}_n - \beta) \sqrt{n} = \sigma_\infty \sum_{j=1}^n V_j \mathbf{x}_j / \sqrt{n} + o_{\mathbb{P}}(1).$$

By the fact that the V_j s are IID standard Gaussian, theorem 1 follows.

References

- Babu, G. J. (1989) Strong representations for LAD estimators in linear models. *Probab. Theor. Reltd Flds*, **83**, 547–558.
- Billingsley, P. (1999) *Convergence of Probability Measures*, 2nd edn. New York: Wiley.
- Bunzel, H. and Vogelsang, T. J. (2005) Powerful trend function tests that are robust to strong serial correlation, with an application to the Prebisch-Singer hypothesis. *J. Bus. Econ. Statist.*, **23**, 381–394.
- Cui, H., He, X. and Ng, K. W. (2004) M-estimation for linear models with spatially-correlated errors. *Statist. Probab. Lett.*, **66**, 383–393.
- Diggle, P. (1990) *Time Series, a Biostatistical Introduction*. Oxford: Clarendon.
- Fitzenberger, B. (1997) The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *J. Econometr.*, **82**, 235–287.
- Goh, S. C. and Knight, K. (2009) Nonstandard quantile-regression inference. *Econometr. Theor.*, **25**, 1415–1432.
- Goncalves, S. and White, H. (2005) Bootstrap standard error estimates for linear regression. *J. Am. Statist. Ass.*, **100**, 970–979.
- Grilli, E. R. and Yang, C. (1988) Primary commodity prices, manufactured goods prices, and the terms of trade of developing countries. *World Bnk Econ Rev.*, **2**, 1–47.
- Hall, P. and Jing, B. (1996) On sample reuse methods for dependent data. *J. R. Statist. Soc. B*, **58**, 727–737.
- Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2007) A simple, robust and powerful test of the trend hypothesis. *J. Econometr.*, **141**, 1302–1330.
- Hidalgo, J. (2003) An alternative bootstrap to moving blocks for time series regression models. *J. Econometr.*, **117**, 367–399.
- Jansson, M. (2004) The error rejection probability of simple autocorrelation robust tests. *Econometrica*, **72**, 937–946.
- Kiefer, N. M. and Vogelsang, T. J. (2005) A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometr. Theor.*, **21**, 1130–1164.
- Kiefer, N. M., Vogelsang, T. J. and Bunzel, H. (2000) Simple robust testing of regression hypotheses. *Econometrica*, **68**, 695–714.
- Koenker, R. (2005) *Quantile Regression*. Cambridge: Cambridge University Press.
- Koul, H. L. (1977) Behavior of robust estimators in the regression model with dependent errors. *Ann. Statist.*, **5**, 681–699.
- Lee, W. (2006) Robust tests of hypotheses in models with m-estimation. *Working Paper*. National Chung Cheng University, Min-Hsiung.
- Lobato, I. N. (2001) Testing that a dependent process is uncorrelated. *J. Am. Statist. Ass.*, **96**, 1066–1076.
- Lutz, M. G. (1999) A general test of the Prebisch-Singer hypothesis. *Rev. Devlpmnt Econ.*, **3**, 44–57.
- Perron, P. and Yabu, T. (2009) Estimating deterministic trends with an integrated or stationary noise component. *J. Econometr.*, **151**, 56–69.
- Phillips, P. C. B. (1991) A shortcut to LAD estimator asymptotics. *Econometr. Theor.*, **7**, 450–463.
- Portnoy, S. (1977) Robust estimation in dependent situations. *Ann. Statist.*, **5**, 22–43.
- Portnoy, S. (1979) Further remarks on robust estimation in dependent situations. *Ann. Statist.*, **7**, 224–231.
- Portnoy, S. (1991) Asymptotic behavior of regression quantiles in non-stationary, dependence cases. *J. Multiv. Anal.*, **38**, 100–113.
- Prakasa Rao, B. L. S. (1981) Asymptotic behavior of M-estimators for the linear model with dependent errors. *Bull. Inst. Math. Acad. Sin.*, **9**, 367–375.
- Priestley, M. B. (1980) State-dependent models: a general approach to non-linear time series analysis. *J. Time Ser. Anal.*, **1**, 47–71.
- Romano, J. P. and Wolf, M. (2006) Improved nonparametric confidence intervals in time series regression. *J. Nonparam. Statist.*, **18**, 199–214.
- Shao, X. (2010) A self-normalized approach to confidence interval construction in time series. *J. R. Statist. Soc. B*, **72**, 343–366.
- Sherman, M. (1997) Subseries methods in regression. *J. Am. Statist. Ass.*, **92**, 1041–1048.
- Sun, Y., Phillips, P. C. B. and Jin, S. (2008) Optimal bandwidth selection in heteroscedasticity-autocorrelation robust testing. *Econometrica*, **76**, 175–194.
- Vogelsang, T. J. (1998) Trend function hypothesis testing in the presence of serial correlation. *Econometrica*, **66**, 123–148.
- Wu, W. B. (2005) Nonlinear system theory: another look at dependence. *Proc. Natn. Acad. Sci. USA*, **102**, 14150–14154.
- Wu, W. B. (2007a) M-estimation of linear models with dependent errors. *Ann. Statist.*, **35**, 495–521.
- Wu, W. B. (2007b) Strong invariance principles for dependent random variables. *Ann. Probab.*, **35**, 2294–2320.
- Zhou, Z. and Wu, W. B. (2009) Local linear quantile estimation for non-stationary time series. *Ann. Statist.*, **37**, 2696–2729.