J. R. Statist. Soc. B (2005)
67, *Part* 3, *pp.* 381–393

Self-weighted least absolute deviation estimation for infinite variance autoregressive models

Shiqing Ling

Hong Kong University of Science and Technology, People's Republic of China

[Received January 2004. Revised December 2004]

Summary. How to undertake statistical inference for infinite variance autoregressive models has been a long-standing open problem. To solve this problem, we propose a self-weighted least absolute deviation estimator and show that this estimator is asymptotically normal if the density of errors and its derivative are uniformly bounded. Furthermore, a Wald test statistic is developed for the linear restriction on the parameters, and it is shown to have non-trivial local power. Simulation experiments are carried out to assess the performance of the theory and method in finite samples and a real data example is given. The results are entirely different from other published results and should provide new insights for future research on heavy-tailed time series.

Keywords: Autoregressive model; Heavy-tailed time series; Infinite variance; Least absolute deviation estimation; Self-weighted least absolute deviation

1. Introduction

Consider stationary autoregressive (AR(p)) time series $\{y_t\}$ which are generated by

$$y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t, \tag{1.1}$$

where $\{\varepsilon_t\}$ is a sequence of independent and identically distributed errors and $\boldsymbol{\phi} = (\phi_0, \phi_1, \ldots, \phi_p)'$ is an unknown parameter vector with its true value $\boldsymbol{\phi}_0$. When $E(\varepsilon_t^2)$ is finite, it is well known that various estimators of $\boldsymbol{\phi}_0$ are asymptotically normal and several methods are available for statistical inference. When $E(\varepsilon_t^2)$ is infinite, model (1.1) is called the infinite variance autoregressive (IVAR) model. Such heavy-tailed models are encountered in several fields, such as teletraffic engineering (see Duffy *et al.* (1994)), hydrology (see Castillo (1988)) and economics and finance (see Koedijk *et al.* (1990) and Jansen and de Vries (1991)). A comprehensive review and more references can be found in Resnick (1997). The statistical theory of the IVAR model is fundamentally different from that of AR models with finite variances.

Kanter and Steiger (1974) showed weak consistency of the least squares estimator of $\boldsymbol{\phi}_0$. Furthermore, Hannan and Kanter (1977) proved its strong consistency with a convergent rate $n^{1/\delta}$, where n is the sample size, $\delta > \alpha$ and $\alpha \in (0, 2)$ is the tail index of ε_t (see also Knight (1987)). The limiting distribution of the least squares estimator was not available until Davis and Resnick (1985, 1986). On the basis of a point process technique, they showed that the least squares estimator converges weakly to a ratio of two stable random variables with the rate $n^{1/\alpha} L(n)$, where $L(n)$ is a slowly varying function. Gross and Steiger (1979) considered

Address for correspondence: Shiqing Ling, Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, People's Republic of China.
E-mail: maling@ust.hk

 1369–7412/05/67381

the least absolute deviation (LAD) estimator and proved its strong consistency. An and Chen (1982) showed that the convergence rate of the LAD estimator is $n^{1/\delta}$. The asymptotic theory of the LAD and M-estimators of ϕ_0 was completely established by Davis *et al.* (1992). They showed that these estimators converge weakly to the minimum of a stochastic process with rate $a_n = \inf\{x : P(|\varepsilon_t| > x) \leqslant n^{-1}\}$. Mikosch *et al.* (1995) studied the Whittle estimator for IVAR moving average models and showed that this estimator converges to a function of a sequence of stable random variables. This result was extended by Kokoszka and Taqqu (1996) for long memory autoregressive fractionally integrated moving average models.

However, none of the limiting distributions in the above has a closed form. We therefore cannot use them for statistical inference in practice. This has been a long-standing open problem for IVAR models. To solve this problem, we propose a self-weighted LAD estimator and show that this estimator is asymptotically normal. Furthermore, a Wald test statistic is developed for testing linear restrictions on the parameters and it is shown to have non-trivial local power. Simulation experiments are undertaken to assess the performance of the theory and method in finite samples and a real data example is given to illustrate its practicality.

We organize the paper as follows. Section 2 presents the main results. Section 3 reports the simulation results, and Section 4 presents the real data example. Concluding remarks are given in Section 5 and all proofs are given in Appendix A.

2. The self-weighted least absolute deviance estimator and the main results

In the regression set-up, one advantage of LAD estimation is that it does not require any moment condition on the errors to obtain asymptotic normality (see Koenker and Bassett (1978)). However, when we use this method for model (1.1), such as in Koul and Saleh (1995), Koenker and Zhao (1996), Davis and Dunsmuir (1997), Mukherjee (1999), Koul (2002) and Ling and McAleer (2004), this advantage disappears. As discussed by Davis *et al.* (1992), an intuitive reason is that the large positive or negative values of ε_t in IVAR models produce y_t which appear to be outliers, and the same ε_t produces many *leverage* points (i.e. the large positive or negative values of y_s as $s > t$) such that $\{y_t\}$ itself has heavy tails. Compared with the least squares estimator, the usual LAD estimator gives less weight to the outliers but it gives essentially the same weight to the leverage points. Thus, the LAD estimator cannot reduce the effect of these points on the covariance matrix and so asymptotic normality no longer holds. From the view of robust estimators, this property may be recovered if we can reduce the effect of these leverage points.

This motivates us to define the objective function

$$L_n(\phi) = \sum_{t=1}^{n} w_t |y_t - X_t'\phi|,$$

where $X_t = (1, y_{t-1}, \ldots, y_{t-p})'$ and w_t is a given function of $\{y_{t-1}, \ldots, y_{t-p}\}$. Let

$$\hat{\phi}_n = \arg\min_{\phi\in\Theta}\{L_n(\phi)\},$$

where Θ is the parameter space. $\hat{\phi}_n$ is called the self-weighted LAD (SLAD) estimator of the true value ϕ_0 in Θ. We give the strictly stationary and ergodic condition of model (1.1) as follows (see proposition 13.3.2 in Brockwell and Davis (1996)).

Assumption 1. The characteristic polynomial $1 - \phi_1 z - \ldots - \phi_p z^p$ has all roots outside the unit circle and $\{\varepsilon_t\}$ are independent and identically distributed with $E|\varepsilon_t|^\delta < \infty$ for some $\delta > 0$.

The purpose of the weight w_t is first to downweight the leverage points in X_t such that the covariance matrices Ω and Σ in theorem 1 below are finite. Second, the w_t allow us to approximate $L_n(\hat{\phi}_n)$ by a quadratic form. From the discussion in Davis *et al.* (1992), we can see that it cannot be approximated by a quadratic form when $w_t = 1$ since the remainder term cannot be ignored. A key technical part in our method is to make this term disappear asymptotically by using the condition $E\{(w_t + w_t^2)\|X_t\|^3\} < \infty$ and $\sup_{x\in R}|f'(x)| < \infty$ (see equations (A.4) and (A.5) in Appendix A). Formally, we state these conditions as follows.

Assumption 2. $w_t = g(y_{t-1}, \ldots, y_{t-p})$ and $g(x_1, \ldots, x_p)$ is a real and positive function on R^p such that $E\{(w_t + w_t^2)(\|X_t\|^2 + \|X_t\|^3)\} < \infty$.

Assumption 3. The errors $\{\varepsilon_t\}$ have zero median and a differentiable density $f(x)$ everywhere in R with $f(0) > 0$ and $\sup_{x\in R}|f'(x)| < \infty$.

Here and in what follows, $\to_{\mathcal{L}}$ denotes convergence in distribution as $n \to \infty$. The following is our main result.

Theorem 1. If assumptions 1–3 hold, then it follows that

$$n^{1/2}(\hat{\phi}_n - \phi_0) \underset{\mathcal{L}}{\to} N\left\{0, \frac{1}{4\,f^2(0)}\Sigma^{-1}\Omega\Sigma^{-1}\right\},$$

where $\Sigma = E(w_t X_t X_t')$ and $\Omega = E(w_t^2 X_t X_t')$.

This result is novel and entirely different from the results that are available in the literature for the IVAR model. To use it for statistical inference, we need to estimate $f(0)$, Σ and Ω. Using the logistic kernel $K(x) = \exp(-x)/\{1 + \exp(-x)\}^2$ and the bandwidth $b_n = c/n^{\nu}$ with $\nu \in (0, \frac{1}{2})$ and a constant $c > 0$, we can estimate $f(0)$ by

$$\hat{f}_n(0) = \frac{1}{\hat{\sigma}_w b_n n}\sum_{t=1}^{n} w_t K\left(\frac{y_t - \hat{\phi}_n' X_t}{b_n}\right), \tag{2.1}$$

where $\hat{\sigma}_w = n^{-1}\Sigma_{t=1}^{n} w_t$. Σ and Ω can be estimated respectively by

$$\begin{aligned}\hat{\Sigma}_n &= \frac{1}{n}\sum_{t=1}^{n}(w_t X_t X_t'),\\ \hat{\Omega}_n &= \frac{1}{n}\sum_{t=1}^{n}(w_t^2 X_t X_t').\end{aligned} \tag{2.2}$$

Theorem 1 and equations (2.1) and (2.2) can now be used to undertake statistical inference for the IVAR model. Here, we consider p_1 linear hypotheses of the form

$$H_0 : \Gamma\phi_0 = \gamma,$$

in the usual notation. The Wald test statistic for the hypothesis H_0 is defined as

$$W_n(p_1) = 4n\,\hat{f}_n^2(0)(\Gamma\hat{\phi}_n - \gamma)'(\Gamma\hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1}\Gamma')^{-1}(\Gamma\hat{\phi}_n - \gamma).$$

The following theorem gives the limiting distribution of $W_n(p_1)$.

Theorem 2. If assumptions 1–3 hold with $E(w_t^2) < \infty$, and $b_n = O(1/n^{\nu})$ with $\nu \in (0, \frac{1}{2})$, then under hypothesis H_0 it follows that $W_n(p_1) \to_{\mathcal{L}} \chi^2_{p_1}$.

When testing an order p_0 against a higher order p, we can take a Γ such that $\Gamma\phi_0 = (\phi_{0,p_0+1}, \ldots, \phi_{0p})'$ and let $\gamma = (0, \ldots, 0)_{p_1\times 1}$ with $p_1 = p - p_0$. As pointed out by a referee, this is very useful in model building. The Wald statistic can be used to fashion a test for independent white noise against an AR alternative, i.e. test whether

$$\phi_0 = \phi_1 = \ldots = \phi_p = 0,$$

by using

$$W_{0n}(p+1) = 4n\, \hat{f}_n^2(0)\hat{\phi}_n'(\hat{\Sigma}_n^{-1}\hat{\Omega}_n\hat{\Sigma}_n^{-1})^{-1}\hat{\phi}_n.$$

We reject the independence hypothesis when $W_{0n}(p+1)$ is large. This also provides a model selection tool. Since the residuals are approximately independent and identically distributed for a well-fitted model, we may use $W_{0n}(p+1)$ to test the independence of the residuals $\{\hat{\varepsilon}_t\}$, where $\hat{\varepsilon}_t = y_t - X_t'\hat{\phi}_n$. If we cannot reject the independence hypothesis, then the fitted model is considered to be adequate for the data.

A natural question is whether or not $W_n(p_1)$ has a local power. For this, we consider the local alternative hypothesis

$$H_{1n} : \Gamma\phi_n = \gamma,$$

where $\phi_n = \phi_0 + \nu/n^{1/2}$ and $\nu \in R^{p+1}$ is a constant vector. The standard method for this is to show that the probability measures of $(y_1, \ldots, y_n)$ under hypotheses H_0 and H_{1n} are contiguous and then to use Le Cam's third lemma. It is not clear whether contiguity holds for the IVAR model. Even if contiguity holds, it would be difficult to prove that in the usual method as in Ling and McAleer (2003). In Appendix A, we prove the following result by a direct method. This result implies that $W_n(p_1)$ has a non-trivial local power.

Theorem 3. If the assumptions of theorem 2 hold and $g(x_1, \ldots, x_p)$ is almost everywhere continuous on R^p, then under hypotheses H_{1n}

(a) $n^{1/2}(\hat{\phi}_n - \phi_0) \underset{\mathcal{L}}{\rightarrow} N\left\{\nu, \dfrac{1}{4\,f^2(0)}\Sigma^{-1}\Omega\Sigma^{-1}\right\}$,

(b) $W_n(p_1) \underset{\mathcal{L}}{\rightarrow} \chi^2_{p_1}(\mu)$,

where $\mu = 4\,f^2(0)\nu'\Gamma'(\Gamma\Sigma^{-1}\Omega\Sigma^{-1}\Gamma')^{-1}\Gamma\nu$ is a non-central parameter.

To use the results in this section, we need to select a weight w_t. It seems reasonable to use the following weight analogue to Huber's (1977) influence function:

$$w_t = \begin{cases} 1 & \text{if } a_t = 0, \\ C^3/a_t^3 & \text{if } a_t \neq 0, \end{cases} \tag{2.3}$$

where $a_t = \Sigma_{i=1}^p |y_{t-i}|(|y_{t-i}| \geqslant C)$ and $C > 0$ is a constant. This weight satisfies assumption 2. It downweights the covariance matrices with leverage points but takes full advantage of all matrices without leverage points. Similar to Huber's estimator for regression models, we need to choose C. When $P(\varepsilon_t > x) = P(\varepsilon_t < -x) = x^{-\alpha}$ with $0 < \alpha < 2$ for $x \geqslant 1$, by expression (2.7) in Davis and Resnick (1985), we can show that

$$\Sigma^{-1}\Omega\Sigma^{-1} = O(1)[E\{y_{t-1}^2 I(|y_{t-1}| \leqslant C)\} + C^{2-\alpha}]^{-1},$$

as $p = 1$ and $C \rightarrow \infty$. Thus, the larger C is, the smaller the asymptotic variance is. However, for a large C, the distribution of $\hat{\phi}_n$ may not be well approximated by its limiting distribution as n is small. We still have no theory to support the choice of C. But our simulation results in Section 3 show that it works well when C is the 90% or 95% quantile of data $\{y_1, \ldots, y_n\}$. Also, we note that a small-tailed index α results in a small asymptotic variance. Obviously, there are many other weights, such as $w_t = (1 + C\|X_t\|^2)^{-3/2}$ and $I(\max_{1\leqslant i\leqslant p} |y_{t-1}| \leqslant C)$, that satisfy assumption 2. However, our simulation results which are not reported in this paper show that the SLAD estimator based on w_t in expression (2.3) is much more efficient than that based on these weights.

3. Simulation studies

This section examines the performance of the asymptotic results in finite samples through Monte Carlo experiments. In all experiments, we used the weight w_t in expression (2.3) with C being the 95% quantile of data $\{y_1, \ldots, y_n\}$. To estimate $f(0)$, the optimal bandwidth b_n that is given in Silverman (1986), page 40, was used, which is automatically selected from the data.

We first study the means and standard deviations of the SLAD estimator. Data are generated through the AR(1) model $y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$. The true parameters are taken to be $(\phi_0, \phi_1) = (0, -0.5), (0, 0.5), (0, 0.8)$. Three error distributions, Cauchy, t_2 and $N(0, 1)$, are considered. Here, $N(0, 1)$ has a finite variance and is given for reference. The sample sizes are $n = 200$ and $n = 400$. The number of replications is 1000. Table 1 summarizes the empirical means, empirical standard deviations SD and asymptotic standard deviations AD of the SLAD estimators. The ADs are calculated by using the estimated covariances in equations (2.1) and (2.2). Table 1 shows that all the biases are very small and all the SDs and ADs are very close, particularly when $n = 400$. As n increases from 200 to 400, all the SDs and ADs become smaller. When the tail index becomes smaller, note that the $N(0, 1)$, t_2- and Cauchy distributions are in order and all the SDs and ADs become smaller. This confirms our discussion in Section 2.

Table 1. Means and standard deviations of the SLAD estimators for AR(1) models†

ϕ_0	ϕ_1		*Results for* $\varepsilon_t \sim$ *Cauchy*		*Results for* $\varepsilon_t \sim t_2$		*Results for* $\varepsilon_t \sim$ *N(0, 1)*	
			$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$	$\hat{\phi}_0$	$\hat{\phi}_1$
n = 200								
0.0	−0.5	Mean	0.009	−0.500	0.001	−0.497	0.000	−0.500
		SD	0.119	0.024	0.099	0.054	0.086	0.082
		AD	0.122	0.023	0.107	0.055	0.091	0.083
0.0	0.5	Mean	0.009	0.499	−0.001	0.494	0.001	0.491
		SD	0.121	0.023	0.101	0.054	0.090	0.082
		AD	0.123	0.021	0.108	0.055	0.092	0.084
0.0	0.8	Mean	0.007	0.799	0.000	0.795	0.000	0.788
		SD	0.123	0.009	0.105	0.034	0.092	0.060
		AD	0.125	0.008	0.110	0.033	0.093	0.059
n = 400								
0.0	−0.5	Mean	0.004	−0.500	0.002	−0.498	0.001	−0.500
		SD	0.081	0.016	0.072	0.038	0.063	0.058
		AD	0.086	0.016	0.075	0.039	0.064	0.059
0.0	0.5	Mean	0.005	0.500	−0.005	0.495	0.001	0.493
		SD	0.082	0.015	0.074	0.038	0.064	0.058
		AD	0.086	0.015	0.075	0.038	0.064	0.059
0.0	0.8	Mean	0.003	0.800	−0.001	0.798	0.000	0.792
		SD	0.083	0.006	0.073	0.023	0.065	0.041
		AD	0.087	0.006	0.075	0.023	0.065	0.041

†1000 replications.

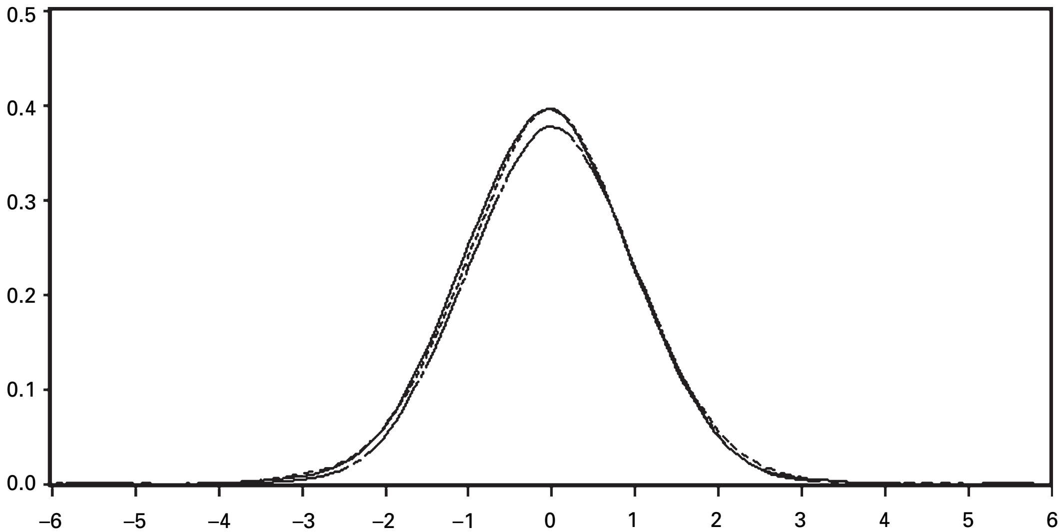

Fig. 1. Density curves of $N(0,1)$ (- — -), N_c (———) and N_t (– – –)

To give an overall view on the distribution of $\hat{\phi}_1$ and its limiting distribution, we simulate 20000 replications for the case with $\phi_1 = 0.5$ and $n = 400$ when the error distributions are Cauchy and t_2. Denote $n^{1/2}(\hat{\phi}_1 - 0.5)/\hat{\sigma}_{\text{SLAD}}$ by N_c and N_t when the distribution is Cauchy and t_2 respectively, where $\hat{\sigma}_{\text{SLAD}}$ are the ADs of $\hat{\phi}_1$. Fig. 1 shows the density curves of N_c, N_t and $N(0,1)$. The density curve of N_c, denoted by $f_c(x)$, is approximated by

$$f_c(x_i) \approx \sum_{r=1}^{20000} I(x_{i-1} \leqslant N_{rc} \leqslant x_i)/20000$$

with $x_0 = -6$ and $x_i = x_{i-1} + 1$, where $i = 1, \ldots, 12$ and N_{rc} is the value of N_c at its rth replication, and similarly for N_t. Fig. 1 shows that the density of $N(0,1)$ approximates reasonably well those of N_c and N_t for all x. In particular, they are very close to each other when $|x| > 1$, which is important in hypothesis testing and in constructing confidence intervals.

We now investigate the size and power of $W_n(1)$ for testing the AR(1) against the AR(2) model when $(\phi_0, \phi_1) = (0, 0.5)$. $W_n(1)$ gives its sizes when $\phi_2 = 0$ under hypothesis H_0 and its powers when $\phi_2 = \pm 0.1$ and $\phi_2 = \pm 2$. Again, the sample sizes are $n = 200$ and $n = 400$ and the number of replications is 1000. Cauchy, t_2- and $N(0,1)$ distributions are used. Table 2 summarizes the results at significance levels 0.1, 0.05 and 0.01. From Table 2, we see that all the sizes are close to the nominal significance levels when $n = 400$ and reasonable when $n = 200$. The powers are increased when n becomes large or when the distance between the alternative and the null H_0 becomes large as expected. These simulation results indicate that the Wald test works well in finite samples.

We also carried out experiments for all the cases in Tables 1 and 2 using the weight w_t in equation (2.3) with C being the 90% quantile of data $\{y_1, \ldots, y_n\}$. The estimators are very similar to those in Table 1, but a little less efficient. The sizes of the Wald test are similar to those in Table 2 when the error distribution is t_2 and $N(0,1)$. When the error distribution is Cauchy, the sizes when $n = 200$ are a little better than those in Table 2. However, they are similar to those in Table 2 when $n = 400$. All the powers are a little lower than those in Table 2. These results are available from the author on request.

Table 2. Sizes and powers of the Wald test for null hypothesis H_0: $\phi_2 = 0.0$ at significance levels α in autoregressive models†

ϕ_0	ϕ_1	ϕ_2	*Results for the Cauchy distribution and the following values of* α*:*			*Results for the* t_2*-distribution and the following values of* α*:*			*Results for the N(0, 1) distribution and the following values of* α*:*		
			0.10	*0.05*	*0.01*	*0.10*	*0.05*	*0.01*	*0.10*	*0.05*	*0.01*
n = 200											
0.0	0.5	−0.2	0.999	0.999	0.992	0.874	0.815	0.618	0.625	0.519	0.296
0.0	0.5	−0.1	0.878	0.822	0.671	0.436	0.317	0.148	0.265	0.191	0.072
0.0	0.5	0.0	*0.123*	*0.071*	*0.024*	*0.104*	*0.052*	*0.015*	*0.093*	*0.048*	*0.015*
0.0	0.5	0.1	0.929	0.892	0.779	0.332	0.230	0.104	0.230	0.136	0.040
0.0	0.5	0.2	0.996	0.992	0.989	0.866	0.804	0.625	0.589	0.487	0.258
n = 400											
0.0	0.5	−0.2	1.000	1.000	1.000	0.990	0.983	0.928	0.869	0.781	0.571
0.0	0.5	−0.1	0.987	0.969	0.912	0.629	0.507	0.288	0.405	0.303	0.143
0.0	0.5	0.0	*0.106*	*0.056*	*0.014*	*0.101*	*0.051*	*0.010*	*0.098*	*0.052*	*0.010*
0.0	0.5	0.1	0.995	0.989	0.969	0.587	0.450	0.241	0.375	0.265	0.107
0.0	0.5	0.2	1.000	1.000	1.000	0.979	0.968	0.906	0.857	0.775	0.559

†1000 replications; numbers in italics denote size.

4. Example

This section examines the Hang Seng Index (HSI) in the Hong Kong stock-market. It is one of the most important indices in the Asian financial markets and has been extensively investigated in the literature. We consider the HSI daily closing index from June 3rd, 1996, to May 31st, 1998, which has a total of 497 observations. Let x_t denote the original data and y_t denote the log-return data, i.e. $y_t = \log(x_t/x_{t-1})$. Figs 2 and 3 are time plots of x_t and $100y_t$ respectively. They display some drastic shocks, which were caused by the South-east Asian economic and financial crisis in 1997.

We first study whether or not $\{y_t\}$ has an infinite variance. For this, we use the Hill estimator for the tail index of y_t. Let $y_{(1)} > y_{(2)} > \ldots > y_{(n)}$ be the order statistic of $\{y_t : 1 \leqslant t \leqslant n\}$. The estimators of the right- and left-tailed indices are defined as

Fig. 2. Time plot of the original HSI (x_t)

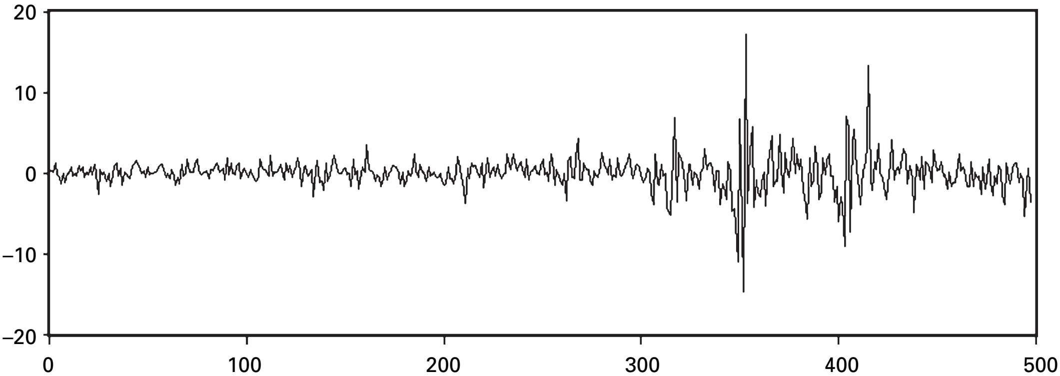

Fig. 3. Time plot of the log-return of the HSI ($100y_t$)

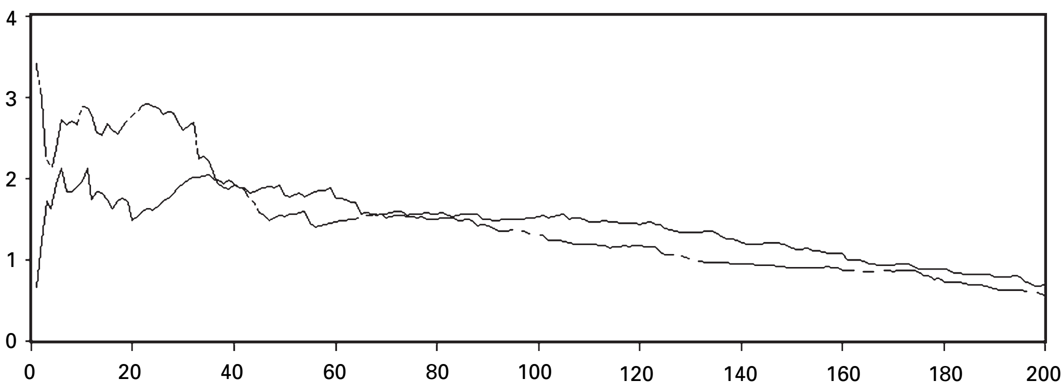

Fig. 4. Hill estimators of the left-hand tail index H_{1k} (———) and the right-hand tail index H_{2k} (- — -)

$$H_{1k}=\left\{\frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{y_{(i)}}{y_{(k+1)}}\right)\right\}^{-1},$$

$$H_{2k}=\left\{\frac{1}{k}\sum_{i=1}^{k}\log\left(\frac{y_{(n-i+1)}}{y_{(n-k)}}\right)\right\}^{-1}$$

respectively. The Hill estimator is consistent when $n,k\to\infty$ and $k/n\to 0$; see for example Resnick (1997). They may not be stable in terms of k selected. Fig. 4 graphs $\{(k,H_{1k}):1\leqslant k\leqslant 200\}$ and $\{(k,H_{2k}):1\leqslant k\leqslant 200\}$. From Fig. 4, it seems difficult to obtain an accurate estimator for the tail index. But it is clear that the right-hand tail index is less than 2 and the left-hand tail index is most likely less than 2. This means that y_t should have an infinite variance and hence it seems reasonable to study $\{y_t\}$ by using the results in Section 2.

We now use $W_{0n}(8)$ and $W_{0n}(12)$ to test whether or not $\{y_t\}$ is independent white noise. Their values are 17.67 and 28.68 respectively. Both reject the independence hypothesis at the 0.05-level. Then, we try to fit an AR(7) model to the data. The estimators are

$$(\hat{\phi}_0,\hat{\phi}_1,\hat{\phi}_2,\hat{\phi}_3,\hat{\phi}_4,\hat{\phi}_5,\hat{\phi}_6,\hat{\phi}_7)=(0.066,0.065,0.000,0.105,0.029,-0.084,0.024,-0.090),$$

and their standard deviations are 0.058, 0.043, 0.044, 0.044, 0.044, 0.044, 0.043 and 0.043

respectively. Again, using independence tests $W_{0n}(8)$ and $W_{0n}(12)$ for the residuals, we obtain that $W_{0n}(8)=1.689$ and $W_{0n}(12)=13.036$. The independence hypothesis cannot now be rejected and hence an AR(7) process should be adequate for the data. For each i, we use $W_n(1)$ to test $H_0:\phi_i=0$. Their values are as follows:

i	0	1	2	3	4	5	6	7
$W_n(1)$	1.266	2.281	0.000	5.710	0.431	3.637	0.316	4.330

$W_n(1)$ rejects only $\phi_3=0$ and $\phi_7=0$ at the 0.05-level. Furthermore, we consider the following model for the data:

$$y_t=\phi_3 y_{t-3}+\phi_7 y_{t-7}+\varepsilon_t. \tag{4.1}$$

The estimators are $\hat{\phi}_3=0.135$ and $\hat{\phi}_7=-0.051$ with standard deviations 0.044 and 0.043 respectively. The values of independence tests $W_{0n}(8)$ and $W_{0n}(12)$ for the residuals are 7.109 and 14.339 respectively, which cannot reject the independence hypothesis and hence model (4.1) should also be adequate for the data. The values of $W_n(1)$ are 9.594 and 1.372 respectively, for testing $\phi_3=0$ and $\phi_7=0$. It only rejects $\phi_7=0$ at the 0.05-level. Finally, we use the following model to fit the data:

$$y_t=\phi_3 y_{t-3}+\varepsilon_t. \tag{4.2}$$

We obtain that $\hat{\phi}_3=0.123$ and its standard deviation is 0.043, and $W_n(1)=8.053$, which rejects $\phi_3=0$ at the 0.05-level. Using independence tests for the residuals, we obtain that $W_{0n}(8)=9.442$ and $W_{0n}(12)=17.152$. They cannot reject the independence hypothesis and hence model (4.2) should be suitable as a final model for the data. A possible interpretation for model (4.2) is that the Hong Kong stock-market has a delayed 3-day reaction to dynamic movements. This may arise because investors in the Hong Kong stock-market prefer to observe other news before making trading decisions.

5. Concluding remarks

This paper has proposed a self-weighted LAD estimator for IVAR models. This estimator was shown to be asymptotically normal and it was used to construct a Wald test for testing the linear restriction on the parameters. The simulation study and a real data example showed that our theory and method should be useful in practice. The self-weight principle can be used for other infinite variance time series models, such as autoregressive moving average and threshold autoregressive models (see Ling (2004) for a self-weighted maximum likelihood estimator of autoregressive moving average–generalized autoregressive conditional heteroscedastic models). Our results can be extended to other estimators, such as the M-, quantile, R- and L-estimators in Koul (2002). This paper provides a new way to handle heavy-tailed time series data and should have a large applicable area in the future.

Acknowledgements

The author greatly appreciates two referees, the Joint Editor, H. Koul, W. K. Li, M. McAleer and V. A. Unkefer for their very helpful comments and suggestions, and thanks the Hong Kong Research Grants Council for financial support through Competitive Earmarked Research Grant HKUST4765/03H.

Appendix A: Proofs

The proof of theorem 1 is similar to those in Davis and Dunsmuir (1997), Knight (1998) and Peng and Yao (2003). We denote the Euclidean norm by $\|\cdot\|$ and a random sequence converging to 0 in probability by $o_p(1)$, and let $\mathcal{F}_t = \sigma\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$.

A.1. Proof of theorem 1

Denote $\hat{\boldsymbol{\theta}}_n = n^{1/2}(\hat{\phi}_n - \phi_0)$ and

$$\tilde{L}_n(\mathbf{u}) = \sum_{t=1}^{n} w_t \left(|\varepsilon_t - \frac{1}{n^{1/2}} \mathbf{u}' X_t| - |\varepsilon_t| \right),$$

where $\mathbf{u} \in R^{p+1}$. Then, $\hat{\boldsymbol{\theta}}_n$ is the minimizer of $\tilde{L}_n(\mathbf{u})$ on R^{p+1}. Using the identity

$$|x-y| - |x| = -y\{I(x>0) - I(x<0)\} + 2\int_0^y \{I(x \leqslant s) - I(x \leqslant 0)\}\,\mathrm{d}s,$$

which holds when $x \neq 0$ (see Knight (1998)), it follows that

$$\tilde{L}_n(\mathbf{u}) = -\mathbf{u}' T_n + 2\sum_{t=1}^{n} \xi_t(\mathbf{u}), \tag{A.1}$$

where $T_n = \Sigma_{t=1}^n w_t X_t \{I(\varepsilon_t > 0) - I(\varepsilon_t < 0)\}/n^{1/2}$ and

$$\xi_t(\mathbf{u}) = w_t \int_0^{\mathbf{u}' X_t / n^{1/2}} \{I(\varepsilon_t \leqslant s) - I(\varepsilon_t \leqslant 0)\}\,\mathrm{d}s.$$

Denote the distribution of ε_t by $F(x)$. By Taylor's expansion, we have

$$\begin{aligned}
\sum_{t=1}^{n} E\{\xi_t(\mathbf{u}) | \mathcal{F}_{t-1}\} &= \sum_{t=1}^{n} w_t \int_0^{\mathbf{u}' X_t / n^{1/2}} \{F(s) - F(0)\}\,\mathrm{d}s \\
&= \sum_{t=1}^{n} w_t \int_0^{\mathbf{u}' X_t / n^{1/2}} \left\{ s f(0) + \frac{1}{2} s^2 f'(s^*) \right\} \mathrm{d}s \\
&= \mathbf{u}' \left\{ \frac{f(0)}{2n} \sum_{t=1}^{n} w_t X_t X_t' \right\} \mathbf{u} + \frac{1}{2} \sum_{t=1}^{n} w_t \int_0^{\mathbf{u}' X_t / n^{1/2}} s^2 f'(s^*)\,\mathrm{d}s,
\end{aligned} \tag{A.2}$$

where $s^* \in (0, s)$. Thus, by equations (A.1) and (A.2), it follows that

$$\tilde{L}_n(\mathbf{u}) = -\mathbf{u}' T_n + f(0) \mathbf{u}' \left(\frac{1}{n} \sum_{t=1}^{n} w_t X_t X_t' \right) \mathbf{u} + R_n(\mathbf{u}), \tag{A.3}$$

where

$$R_n(\mathbf{u}) = \sum_{t=1}^{n} w_t \int_0^{\mathbf{u}' X_t / n^{1/2}} s^2 f'(s^*)\,\mathrm{d}s + 2\sum_{t=1}^{n} [\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u}) | \mathcal{F}_{t-1}\}].$$

By assumptions 2 and 3, for each $\mathbf{u}$, it follows that

$$\left| \sum_{t=1}^{n} w_t \int_0^{\mathbf{u}' X_t / n^{1/2}} s^2 f'(s^*)\,\mathrm{d}s \right| \leqslant \sup_x |f'(x)| \frac{\|\mathbf{u}\|^3}{n n^{1/2}} \sum_{t=1}^{n} w_t \|X_t\|^3 \to 0, \tag{A.4}$$

almost surely as $n \to \infty$. As for equation (A.2), we can show that

$$E\{\xi_t^2(\mathbf{u})\} \leqslant \frac{1}{n^{1/2}} E\left\{ w_t^2 |\mathbf{u}' X_t| \int_0^{|\mathbf{u}' X_t|/n^{1/2}} |I(\varepsilon_t \leqslant s) - I(\varepsilon_t \leqslant 0)| \,\mathrm{d}s \right\}$$
$$\leqslant \frac{1}{n^{1/2}} E\left[w_t^2 |\mathbf{u}' X_t| \int_0^{|\mathbf{u}' X_t|/n^{1/2}} \{F(s) - F(0)\}^{1/2} \,\mathrm{d}s \right]$$
$$\leqslant \frac{1}{n^{1/2}} \max_x \{f(x)^{1/2}\}\, E\left(w_t^2 |\mathbf{u}' X_t| \int_0^{|\mathbf{u}' X_t|/n^{1/2}} s^{1/2} \,\mathrm{d}s \right)$$
$$\leqslant \frac{\|\mathbf{u}\|^{2.5}}{n^{1+1/4}} \max_x \{f(x)^{1/2}\}\, E(w_t^2 \|X_t\|^{2+1/2})$$
$$\leqslant \frac{\|\mathbf{u}\|^{2.5}}{n^{1+1/4}} \max_x \{f(x)^{1/2}\}\, E\{w_t^2(\|X_t\|^2 + \|X_t\|^3)\}.$$

By assumptions 2 and 3, for each $\mathbf{u}$, it follows that

$$E\left(\sum_{t=1}^n [\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\}] \right)^2 = \sum_{t=1}^n E[\xi_t(\mathbf{u}) - E\{\xi_t(\mathbf{u})|\mathcal{F}_{t-1}\}]^2$$
$$\leqslant 2 \sum_{t=1}^n E\{\xi_t^2(\mathbf{u})\} \to 0, \qquad \text{(A.5)}$$

as $n \to \infty$. By inequalities (A.4) and (A.5), we know that $R_n(\mathbf{u}) = o_p(1)$ for each $\mathbf{u}$.

By assumptions 1 and 2 and the ergodic theorem, we have

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n (w_t X_t X_t') &\to \Sigma \qquad \text{almost surely,} \\ \frac{1}{n} \sum_{t=1}^n (w_t^2 X_t X_t') &\to \Omega \qquad \text{almost surely} \end{aligned} \qquad \text{(A.6)}$$

as $n \to \infty$. By the second part of expression (A.6) and the central limit theorem, we readily show that $T_n \to_{\mathcal{L}} \Phi$, where $\Phi \sim N(0, \Omega)$. Thus, by equation (A.3), for each $\mathbf{u}$, we have

$$\tilde{L}_n(\mathbf{u}) \underset{\mathcal{L}}{\to} -\mathbf{u}'\Phi + f(0)\mathbf{u}'\Sigma\mathbf{u}.$$

The limit $-\mathbf{u}'\Phi + f(0)\mathbf{u}'\Sigma\mathbf{u}$ has a unique minimum at $\mathbf{u} = \{2\, f(0)\}^{-1}\Sigma^{-1}\Phi$ almost surely. As for corollary 2 in Knight (1998), by the convexity of $\tilde{L}_n(\mathbf{u})$ for each n, we can show that

$$\hat{\boldsymbol{\theta}}_n = n^{1/2}(\hat{\phi}_n - \phi_0) \underset{\mathcal{L}}{\to} N\left\{ 0, \frac{1}{4 f^2(0)} \Sigma^{-1}\Omega\Sigma^{-1} \right\}.$$

This completes the proof.

A.2. Proof of theorem 2

Denote $A_t(\mathbf{u}) = \int K(y)\, f(b_n y + \mathbf{u}' X_t / n^{1/2})\,\mathrm{d}y$, where $\mathbf{u} \in R^{p+1}$. Then, we have

$$E\left\{ \frac{1}{b_n} K\left(\frac{\varepsilon_t - \mathbf{u}' X_t / n^{1/2}}{b_n} \right) \middle| \mathcal{F}_{t-1} \right\} = \frac{1}{b_n} \int K\left(\frac{x - \mathbf{u}' X_t / n^{1/2}}{b_n} \right) f(x)\,\mathrm{d}x = A_t(\mathbf{u}).$$

Using this with $\int K(x)\,\mathrm{d}x = 1$ and $\int K^2(x)\,\mathrm{d}x < \infty$, for each $\mathbf{u}$, it follows that

$$E\left[\frac{1}{n} \sum_{t=1}^n w_t \left\{ \frac{1}{b_n} K\left(\frac{\varepsilon_t - \mathbf{u}' X_t / n^{1/2}}{b_n} \right) - A_t(\mathbf{u}) \right\} \right]^2$$
$$= \frac{1}{n^2} \sum_{t=1}^n E\left[w_t^2 \left\{ \frac{1}{b_n} K\left(\frac{\varepsilon_t - \mathbf{u}' X_t / n^{1/2}}{b_n} \right) - A_t(\mathbf{u}) \right\} \right]^2$$
$$\leqslant \frac{2}{n^2 b_n} \sum_{t=1}^n E\left\{ \frac{w_t^2}{b_n} K^2\left(\frac{\varepsilon_t - \mathbf{u}' X_t / n^{1/2}}{b_n} \right) \right\} + \frac{2}{n^2} \sum_{t=1}^n E\{w_t^2 A_t^2(\mathbf{u})\}$$

$$\leqslant \frac{2}{n^2 b_n} \sum_{t=1}^{n} E\left\{ w_t^2 \int K^2(y) f\left(b_n y + \frac{\mathbf{u}' X_t}{n^{1/2}} \right) \mathrm{d}y \right\} + \frac{2}{n^2} \sum_{t=1}^{n} E\{ w_t^2 A_t^2(\mathbf{u}) \}$$
$$\leqslant \frac{2 \max_x \{ f(x) \} E(w_t^2)}{n b_n} \int K^2(y) \,\mathrm{d}y + \frac{2 \max_x \{ f(x) \} E(w_t^2)}{n} = o(1), \tag{A.7}$$

by assumption 3. By inequality (A.7) and the continuity of f, we can further show that

$$\sup_{\|\mathbf{u}\| \leqslant M} \left| \frac{1}{n} \sum_{t=1}^{n} w_t \left\{ \frac{1}{b_n} K\left(\frac{\varepsilon_t - \mathbf{u}' X_t / n^{1/2}}{b_n} \right) - A_t(\mathbf{u}) \right\} \right| = o_p(1),$$

for any fixed constant $M > 0$. Let $\hat{\boldsymbol{\theta}}_n = n^{1/2}(\hat{\phi}_n - \phi_0)$, which is bounded in probability. By the preceding equation, it follows that

$$\frac{1}{n} \sum_{t=1}^{n} w_t \left\{ \frac{1}{b_n} K\left(\frac{\varepsilon_t - \hat{\boldsymbol{\theta}}_n' X_t / n^{1/2}}{b_n} \right) - A_t(\hat{\boldsymbol{\theta}}_n) \right\} = o_p(1).$$

Since $\int |y| \, K(y) \,\mathrm{d}y < \infty$, by Taylor's expansion and assumptions 2 and 3, we have

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} w_t |A_t(\hat{\boldsymbol{\theta}}_n) - f(0)| &= \frac{1}{n} \sum_{t=1}^{n} w_t \left| \int K(y) \{ f(b_n y + \hat{\boldsymbol{\theta}}_n' X_t / n^{1/2}) - f(0) \} \,\mathrm{d}y \right| \\
&= \frac{1}{n} \sum_{t=1}^{n} w_t \left| \int K(y) (b_n y + \hat{\boldsymbol{\theta}}_n' X_t / n^{1/2}) f'(\xi_t^*) \,\mathrm{d}y \right| \\
&\leqslant \max_x |f'(x)| \left\{ \frac{b_n}{n} \sum_{t=1}^{n} w_t \int |y| K(y) \,\mathrm{d}y + \frac{\|\hat{\boldsymbol{\theta}}_n\|}{n^{3/2}} \sum_{t=1}^{n} w_t \|X_t\| \right\} \\
&= O_p(b_n) + O_p\left(\frac{1}{n^{1/2}} \right) = o_p(1),
\end{aligned}$$

where ξ_t^* lies between 0 and $b_n y + \hat{\boldsymbol{\theta}}_n' X_t$. By the preceding two inequalities, we can readily show that $\hat{f}_n(0) = f(0) + o_p(1)$. Finally, by theorem 1 and expression (A.6), the conclusion holds. This completes the proof.

A.3. Proof of theorem 3

Note that y_t under hypothesis H_{1n} depends on n. To emphasize this, we denote y_t by y_{nt} under H_{1n}. Here, y_{nt} is a function of n, ϕ, ν and $\{\varepsilon_t\}$ and $y_{nt} = y_t$ when $\nu = 0$, where y_t is defined by model (1.1) under H_0. Similarly define w_{nt} and X_{nt}. It is easy to see that $y_{nt} \to y_t$ almost surely. Since g is continuous almost everywhere, $w_{nt} \to w_t$ almost surely when $n \to \infty$. Now, under H_{1n},

$$\hat{\phi}_n = \arg \min_{\phi \in \Theta} \left(\sum_{t=1}^{n} w_{nt} |y_{nt} - X_{nt-1}' \phi| \right).$$

Let $a_{nt} = w_{nt} X_{nt} \{ I(\varepsilon_t > 0) - I(\varepsilon_t < 0) \}$. By Markov's inequality, the dominated convergence theorem and the ergodic theorem, we can show that

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^{n} a_{nt} a_{nt}' &= \Omega + o_p(1), \\
\frac{1}{n} \sum_{t=1}^{n} E(a_{nt} a_{nt}' | \mathcal{F}_{t-1}) &= \Omega + o_p(1).
\end{aligned} \tag{A.8}$$

Since a_{nt} is a martingale difference in terms of $\mathcal{F}_t$, by the central limit theorem for martingale differences and expression (A.8), it follows that

$$\frac{1}{n^{1/2}} \sum_{t=1}^{n} a_{nt} \underset{\mathcal{L}}{\to} N(0, \Omega). \tag{A.9}$$

Using expressions (A.8) and (A.9) and a similar method as for theorem 1, we can show that part (a) holds. Furthermore, we can show that $\hat{\Sigma}_n = \Sigma + o_p(1)$, $\hat{\Omega}_n = \Omega + o_p(1)$ and $\hat{f}_n(0) = f(0) + o_p(1)$ under hypothesis

H_{1n}. By part (a) of this theorem and noting that $n^{1/2}(\hat{\phi}_n - \phi_n) = n^{1/2}(\hat{\phi}_n - \phi_0) + \nu$, it is straightforward to show that part (b) holds (see also the proof of theorem 6 in Weiss (1991)). This completes the proof.

References

An, H. Z. and Chen, Z. G. (1982) On convergence of LAD estimates in autoregression with infinite variance. *J. Multiv. Anal.*, **12**, 335–345.

Brockwell, P. J. and Davis, R. A. (1986) *Time Series: Theory and Methods*, 2nd edn. New York: Springer.

Castillo, E. (1988) *Extreme Value Theory in Engineering*. Cambridge: Cambridge University Press.

Davis, R. A. and Dunsmuir, W. T. M. (1997) Least absolute deviation estimation for regression with ARMA errors. *J. Theoret. Probab.*, **10**, 481–497.

Davis, R. A., Knight, K. and Liu, J. (1992) M-estimation for autoregressions with infinite variance. *Stoch. Processes Appl.*, **40**, 145–180.

Davis, R. A. and Resnick, S. I. (1985) Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.*, **13**, 179–195.

Davis, R. A. and Resnick, S. I. (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.*, **14**, 533–558.

Duffy, D., Mcintosh, A., Rosenstein, M. and Willinger, W. (1994) Statistical analysis of CCSN/SS7 traffic data from working CCS subnetworks. *IEEE J. Selctd Areas Communs*, **12**, 544–551.

Gross, S. and Steiger, W. L. (1979) Least absolute deviation estimates in autoregression with infinite variance. *J. Appl. Probab.*, **16**, 104–116.

Hannan, E. J. and Kanter, M. (1977) Autoregressive processes with infinite variance. *J. Appl. Probab.*, **14**, 411–415.

Huber, P. J. (1977) *Robust Statistical Procedures*. Philadelphia: Society for Industrial and Applied Mathematics.

Jansen, D. and De Vries, C. (1991) On the frequency of large stock returns: putting booms and busts into perspective. *Rev. Econ. Statist.*, **73**, 18–24.

Kanter, M. and Steiger, W. L. (1974) Regression and autoregression with infinite variance. *Adv. Appl. Probab.*, **6**, 768–783.

Knight, K. (1987) Rate of convergence of centred estimates of autoregressive parameters for infinite variance autoregressions. *J. Time Ser. Anal.*, **8**, 51–60.

Knight, K. (1998) Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.*, **26**, 755–770.

Koedlj, K., Schafgans, M. and De Vries, C. (1990) The tail index of exchange rate returns. *J. Int. Econ.*, **29**, 93–108.

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

Koenker, R. and Zhao, Q. S. (1996) Conditional quantile estimation and inference for ARCH models. *Econometr. Theory*, **12**, 793–813.

Kokoszka, P. S. and Taqqu, M. S. (1996) Parameter estimation for infinite variance fractional ARIMA. *Ann. Statist.*, **24**, 1880–1913.

Koul, H. L. (2002) Weighted empirical processes in dynamic nonlinear models. *Lect. Notes Statist.*, **166**.

Koul, H. L. and Saleh, A. K. Md. E. (1995) Autoregression quantiles and related rank scores processes. *Ann. Statist.*, **23**, 670–689.

Ling, S. (2004) Self-weighted LSE and MLE for ARMA-GARCH models. *J. Econometr.*, to be published.

Ling, S. and McAleer, M. (2003) Adaptive estimation in nonstationary ARMA models with GARCH noises. *Ann. Statist.*, **31**, 642–674.

Ling, S. and McAleer, M. (2004) Regression quantiles for unstable autoregression model. *J. Multiv. Anal.*, **89**, 304–328.

Mikosch, T., Gadrich, T., Klüppelberg, C. and Adler, R. J. (1995) Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.*, **23**, 305–326.

Mukherjee, K. (1999) Asymptotics of quantiles and rank scores in nonlinear time series. *J. Time Ser. Anal.*, **20**, 173–192.

Peng, L. and Yao, Q. W. (2003) Least absolute deviations estimation for ARCH and GARCH models. *Biometrika*, **90**, 967–975.

Resnick, S. I. (1997) Heavy tail modeling and teletraffic data (with discussion). *Ann. Statist.*, **25**, 1805–1869.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

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