COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1064

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

ROBUST NONSTATIONARY REGRESSION

Peter C. B. Phillips

November 1993

Robust Nonstationary Regression

by

Peter C. B. Phillips*

Cowles Foundation for Research in Economics,
Yale University

0. Abstract

This paper provides a robust statistical approach to nonstationary time series regression and inference. Fully modified extensions of traditional robust statistical procedures are developed which allow for endogeneities in the nonstationary regressors and serial dependence in the shocks that drive the regressors and the errors that appear in the equation being estimated. The suggested estimators involve semiparametric corrections to accommodate these possibilities and they belong to the same family as the fully modified least squares (FM-OLS) estimator of Phillips and Hansen (1990). Specific attention is given to fully modified least absolute deviation (FM-LAD) estimation and fully modified M (FM-M)-estimation. The criterion function for LAD and some M-estimators is not always smooth and the paper develops generalized function methods to cope with this difficulty in the asymptotics. The results given here include a strong law of large numbers and some weak convergence theory for partial sums of generalized functions of random variables. The limit distribution theory for FM-LAD and FM-M estimators that is developed includes the case of finite variance errors and the case of heavy-tailed (infinite variance) errors. Some simulations and a brief empirical illustration are reported.

Key words: FM-LAD estimator; FM-M estimator; Generalized functions of random variables; laws of large numbers and weak convergence for generalized functions; non-Gaussian; nonstationarity; regular sequence; robust estimation.

October 1993

*All computations and graphics reported in the paper were performed by the author in programs written in GAUSS 3.0 on a 486-50 PC. My thanks go to Vassilis Hajivassiliou for the use of his optimization routine "gmaxim," to the NSF for research support under Grant No. SES 9122142, and to Glena Ames for her skill and effort in keyboarding the manuscript.

1. Introduction

Many recent empirical applications of nonstationary regression methods have involved financial data sets. Examples include econometric tests of the purchasing power parity theory (Johansen and Juselius, 1993), which use exchange rate data, tests of forward exchange market unbiasedness (Corbae, Lim and Ouliaris, 1993), which use spot and forward exchange rates, and tests of uncovered interest rate parity (Hunter, 1993), which use interest rate and exchange rate data. A well documented characteristic of such financial data is their non Gaussianity. The leptokurtosis and heavy tailed features of exchange rate returns are especially notable, and these features are usually accentuated when the data are sampled more frequently.

For illustration, Figure 1(i) shows daily data for returns (*i.e.* differences in logarithms) of the Australian dollar spot exchange rate measured against the US dollar over the period January 1984-April 1991. Outlier activity is a fairly prominent characteristic of this data set. Figure 1(ii) graphs a nonparametric estimate of the density of this data against that of a normal distribution whose mean and variance are fitted to those of the data. The leptokurtosis and heavy tails of the nonparametric density are evident in comparison with the fitted normal.

- - - - - - - - - - - - - - - - -

Figure 1(i) and 1(ii) about here

- - - - - - - - - - - - - - - - -

Two commonly used regression methods for analyzing such data in levels or log levels form are reduced rank regression (RRR) (Johansen, 1988; Ahn and Reinsel, 1990) and fully modified least squares (FM-OLS) (Phillips and Hansen, 1990). Both these procedures are Gaussian in the sense that they can be deduced as maximum likelihood estimators under certain conditions when the data are Gaussian; and in this event they also deliver optimal estimates in nonstationary cointegrating regression (Phillips, 1991a). These techniques were designed to deal with nonstationarity in the data but, like other least squares and Gaussian methods, they were not designed to deal specifically with data where there is prominent outlier activity. In such cases there would seem to be a need for estimators which are more resistant to the presence of outliers than Gaussian estimators while at the same time being able to cope with data nonstationarity and endogenous regressors.

This need is addressed in the present paper. We develop extensions of robust regression procedures which allow for data nonstationarity and endogeneities in the regressors, and serial dependence in the shocks that drive the regressors and in the errors that appear in the regression equation. Our suggested estimators involve semiparametric corrections to accommodate these possibilities and they belong to the same family as the fully modified least squares (FM-OLS) estimator of Phillips and Hansen (1990). Specifically, we develop a fully modified least absolute deviation (FM-LAD) estimator and a fully modified M (FM-M) estimator from the corresponding LAD- and M-estimators of ordinary regression. These estimators are designed to combine the features of nonstationary regression estimators like FM-OLS with the outlier resistant features of the common robust estimators.

Since the criterion function for the LAD estimator and for some common M-estimators is not smooth, we cannot rely on usual Taylor expansion methods to do the asymptotics. Recently, convex function approximations and stochastic equicontinuity arguments have been used to deal with this type of difficulty -- see Pollard (1990, 1991) for some discussion and illustration of these methods. The approach used here is rather different, although it does retain a convexity argument like that of Knight (1989) to assist in establishing the weak convergence of extremum estimators. Our approach is to treat the objective function in an extremum estimation problem as a generalized function and use generalized Taylor series expansions to extract the asymptotics. To facilitate this process, we introduce the concept of a generalized function of a random variable and give a strong law of large numbers and some weak convergence theory for partial sums of generalized functions of random variables.

The paper is organized as follows. Section 2 gives the model, our main assumptions and the preliminary limit theory. Section 3 introduces the idea of a generalized function of a random variable by means of a class of suitable approximating sequences of ordinary functions of random variables. Some limit theory for generalized functions that is used later in the paper is given in this section. The FM-LAD estimator is constructed and its asymptotic theory is derived in Section 4. Section 5 deals with the FM-M estimator and its asymptotics. Extensions of the asymptotic theory to cover the case of heavy-tailed (infinite variance) errors are given in Section

6. Some simulation results and a brief empirical illustration are reported in Section 7. The paper concludes in Section 8 by mentioning some further extensions of robust nonstationary regression. Proofs are given in Section 9.

2. The Model, Assumptions and Preliminary Limit Theory

We will work with the model

(1a) $y_t = x_t'\beta + u_{0t}$

(1b) $\Delta x_t = u_{xt}$

where x_t is a k-vector of full rank (i.e. not cointegrated) integrated regressors. The error vector $u_t = (u_{0t}, u_{xt}')'$ in (1) is possibly temporally dependent and is required to satisfy Assumption EC below. This assumption is convenient for our purposes here but could be replaced by a variety of similar conditions without materially affecting our subsequent results provided the finite second moment requirement is retained. If that condition is relaxed then the limit theory, and indeed, some rates of convergence, will change. We will discuss this possibility later in the paper. The model (1) can also be extended by the inclusion of deterministic trends and this extension affects our results in the usual way (see Park and Phillips, 1988) provided the finite error variance condition holds.

ASSUMPTION EC (*Error Condition*)

(a) u_t *is a strictly stationary and strong mixing sequence with mixing numbers* α_m *that satisfy*

(2) $\Sigma_1^\infty \alpha_m^{(p-\beta)/p\beta} < \infty$

for some $p > \beta > 2$;

(b) $\|u_t\|_p < \infty$;

(c) *The probability density* $h(\cdot)$ *of* u_{0t} *is symmetric, and is positive and continuous in a neighborhood* $(-b, b)$ *of the origin for some* $b > 0$.

The mixing condition (2) and moment condition (b) in EC are sufficient to ensure the functional weak convergence of partial sum processes of u_{xt}, u_{0t} and bounded functions of u_{0t}, as will be needed later. These conditions will also validate the weak convergence to stochastic integrals of sample covariances between the regressors x_t and the errors u_{0t} and bounded functions of u_{0t}. A requirement like (c) is conventional in the development of an asymptotic theory for the LAD estimator, whose limit theory depends on the value of $h(\cdot)$ at the origin, $h(0)$. However, the symmetry condition on $h(\cdot)$ is stronger than usual and could be relaxed, but it will be convenient in our generalized function proofs.

Under Assumption EC the long-run covariance matrix of u_t exists and we partition this matrix conformably with u_t as

$$\Omega_{uu} = \Sigma_{k=-\infty}^{\infty} E(u_0 u_k') = \begin{bmatrix} \Omega_{00} & \Omega_{0x} \\ \Omega_{x0} & \Omega_{xx} \end{bmatrix}.$$

We also use the transformed error process $v_t = \text{sgn}(u_{0t}) = 1, -1$ for $u_{0t} \geq 0$, $u_{0t} < 0$ respectively, and define $w_t = (v_t, u_{xt}')'$. Since v_t is a bounded function of u_{0t}, the long-run covariance matrix of w_t also exists under EC and we partition this matrix conformably with w_t as follows

$$\Omega_{ww} = \Sigma_{k=-\infty}^{\infty} E(w_0 w_k') = \begin{bmatrix} \Omega_{vv} & \Omega_{vx} \\ \Omega_{xv} & \Omega_{xx} \end{bmatrix}.$$

In a similar way we define and partition the one-sided long-run covariance matrices of u_t and w_t respectively as

$$\Delta_{uu} = \Sigma_{k=0}^{\infty} E(u_0 u_k') = \begin{bmatrix} \Delta_{00} & \Delta_{0x} \\ \Delta_{x0} & \Delta_{xx} \end{bmatrix},$$

and

$$\Delta_{ww} = \Sigma_{k=0}^{\infty} E(w_0 w_k') = \begin{bmatrix} \Delta_{vv} & \Delta_{vx} \\ \Delta_{xv} & \Delta_{xx} \end{bmatrix}.$$

Under Assumption EC a multivariate invariance principle for w_t holds, *viz.*

(3) $$T^{-1/2}\Sigma_1^{[Tr]} w_t \longrightarrow_d B_w(r) \equiv BM(\Omega_{ww}) \;, \quad 0 < r \le 1$$

as shown in Phillips and Durlauf (1986). We partition the limit Brownian motion B in (3) conformably with w_t and Ω using the notation $B_w(r)' = (B_v(r), B_x(r)')$. A similar invariance principle holds for partial sums of u_t, *viz.*

(4) $$T^{-1}\Sigma_1^{[Tr]} u_t \longrightarrow_d B_u(r) \equiv BM(\Omega_{uu}) \;, \quad 0 < r \le 1$$

where the limit process is partitioned as $B_u(r)' = (B_0(r), B_x(r)')$ conformably with $u_t' = (u_{0t}, u_{xt}')$. In addition, EC ensures that sample covariances between the regressors x_t and the error vectors have limits that can be expressed as stochastic integrals with drift. In particular,

(5) $$T^{-1}\Sigma_1^{[Tr]} x_t v_t \longrightarrow_d \int_0^r B_x dB_v + r\Delta_{xv} \;, \quad 0 < r \le 1$$

and

(6) $$T^{-1}\Sigma_1^{[Tr]} x_t u_t' \longrightarrow_d \int_0^r B_x dB_u + r\Delta_{xu} \;, \quad 0 < r \le 1$$

where $\Delta_{xu} = [\Delta_{x0} \;\; \Delta_{xx}]$ -- see Phillips (1988) and Hansen (1992).

3. Generalized Functions of Random Variables and Generalized Limit Theory

Our approach is to treat nonsmooth objective criteria like those that appear in LAD estimation as generalized functions and use generalized Taylor series expansions to represent their local behavior. The basic ideas behind this approach and an application to LAD estimation in a stationary regression were laid out by the author in (1991). We will follow those ideas here and develop some additional concepts to make the approach rigorous.

Our main concerns will involve generalized functions of random variables and stochastic limit operations with partial sums of these generalized functions of random variables. The concept of a generalized function of a random variable is different from the idea of a generalized random process, as it appears in the existing literature on generalized functions [see, for instance, Gel'fand and Vilenkin (1964), Ch. III], wherein such a process is defined as a mapping from a given space of test functions into a random variable. An example of the latter is the continuous linear functional $B(\varphi) = \int_0^1 \varphi(t)dW(t)$, which is here expressed as a stochastic integral of the Wiener process $W(t)$ on $C[0, 1]$.

Instead, our need is to give a meaning to objects such as $\delta(u_t)$, where u_t is a real valued random variable (indexed by discrete time t) and $\delta(\cdot)$ is the Dirac delta generalized function, which has the property that $\int_{-\infty}^{\infty} \delta(x)F(x)dx = F(0)$ for any continuous function $F(x)$. There are, in fact, several ways in which this can be done. In defining generalized functions like $\delta(\cdot)$, *i.e.*, before we deal with such "functions" of random variables -- we will use the "regular sequence" approach given in Lighthill (1958). Associated with (and, in fact, defining) any generalized function $f(x)$ is a sequence $f_m(x)$ of *good functions* (i.e., functions that are continuously differentiable any number of times with derivatives of $O(|x|^{-N})$ as $|x| \to \infty$ for any N; hereafter, simply GF) with the property that

$$(7) \qquad \lim_{m\to\infty} \int_{-\infty}^{\infty} f_m(x)F(x)dx$$

exists for any $F \in GF$. The integral of $f(x)$ is then defined by the equation

$$\int_{-\infty}^{\infty} f(x)F(x)dx := \lim_{m\to\infty} \int_{-\infty}^{\infty} f_m(x)F(x)dx \ .$$

A sequence such as $f_m(x)$ with this property is called a *regular sequence* for $f(x)$.

Since the sequence $f_m(\cdot)$ is measurable, $f_m(u_t)$ has a meaning as an ordinary random variable on the probability space where u_t is itself defined. The generalized function $f(u_t)$ of the random variable u_t is then defined by the associated regular sequence $f_m(u_t)$, or more precisely the class of all regular sequences that are equivalent to $f_m(\cdot)$ in the sense that (7) is the same for each sequence. It follows that if $\text{pdf}(u) \in GF$ is the density of u_t then we can define the expectation of the generalized function $f(\ )$ of u_t by

$$(8) \qquad E(f(u_t)) := \lim_{m\to\infty} E(f_m(u_t)) = \lim_{m\to\infty} \int_{-\infty}^{\infty} f_m(u)\text{pdf}(u)du \ .$$

Provided the limit on the right side of (8) exists, we can relax the requirement that $\text{pdf}(u) \in GF$.

Now suppose we wish to establish a weak law of large numbers (WLLN) or strong law of large numbers (SLLN) for partial sums of the generalized function of random variables $f(u_t)$.

Since $f(\cdot)$ is defined in terms of the regular sequence $f_m(\cdot)$ we can define a WLLN and SLLN for $f(u_t)$, *i.e.*

(9) $\quad T^{-1}\Sigma_1^T f(u_t) \rightarrow_{p,\text{a.s.}} E\{f(u_t)\}$,

by the corresponding weak and strong laws for partial sums of the regular sequence $f_m(u_t)$ of ordinary random variables, *i.e.* by

(10) $\quad T^{-1}\Sigma_1^T f_m(u_t) \rightarrow_{p,\text{a.s.}} E\{f_m(u_t)\}$, $\forall m$

and the limit that appears on the right side of (10) is given by (8). This definition is, in fact, compatible with that of a WLLN or SLLN for ordinary functions of u_t.

3.1. LEMMA (SLLN *for ordinary random variables as generalized functions of random variables*)

Suppose u_t is strictly stationary and ergodic and $f(u_t)$ is an ordinary (measurable) function of u_t. Then (9) *holds in the sense of ordinary random sequences iff it holds in the sense of generalized functions of random sequences, i.e. iff* (10) *holds.*

PROOF. To prove necessity, suppose $f(u_t)$ is an ordinary function of u_t satisfying (9) and $E(f(u_t))$ is finite. We need to demonstrate (10). We construct the following regular sequence of good functions to approximate $f(\cdot)$ [cf. Lighthill (1958), p. 22]

(11) $\quad f_m(u) = \int_{-\infty}^{\infty} f(v)S\{m(v-u)\}me^{-v^2/m^2}dv$.

In (11) the function $S(\cdot)$ is a "smudge function" whose role in $f_m(u)$ is to smudge out $f(v)$ when v is outside the interval $(u - m^{-1}, u + m^{-1})$. $S(\cdot)$ is defined as

(12) $\quad S(y) = s(y)/\int_{-1}^{1} s(y)dy$,

where

$$s(y) = \begin{cases} e^{-1/(1-y^2)} & |y| < 1 \\ 0 & |y| \geq 1 \end{cases}$$

and

$$\int_{-1}^{1} s(y)dy = 2\int_0^1 e^{-1/(1-y^2)}dy = (2/e)\int_0^1 e^{-y^2/(1-y^2)}dy$$

$$= (1/e)\int_0^\infty e^{-z} z^{-1/2}(1+z)^{-3/2}dz \ , \quad \text{with} \quad z = y^2/(1-y^2)$$

$$= (\pi^{1/2}/e)\Psi(1/2,\ 0;\ 1) \ ,$$

where Ψ is the confluent hypergeometric function of the second kind (Erdelyi, 1953, p. 255). Note that $S(y)$ and all of its derivatives are zero at $y = \pm 1$.

Now $S\{m(v - u_t)\}$ is a measurable and integrable function of u_t and therefore constitutes an ergodic sequence, so that

$$T^{-1}\Sigma_1^T S\{m(v - u_t)\} \longrightarrow_{\text{a.s.}} E[S\{m(v - u_t)\}] = \int_{-\infty}^{\infty} S\{m(v-u)\}\text{pdf}(u)du \ .$$

Hence,

$$T^{-1}\Sigma_1^T f_m(u_t) = \int_{-\infty}^{\infty} f(v)T^{-1}\Sigma_1^T S\{m(v - u_t)\}me^{-v^2/m^2}dv$$

$$\longrightarrow_{\text{a.s.}} \int_{-\infty}^{\infty} f(v)E[S\{m(v - u_t)\}]me^{-v^2/m^2}dv$$

$$= E\{f_m(u_t)\} \ , \quad \forall m$$

giving (10) as a necessary condition for (9) in the case of ordinary functions $f(u_t)$ of the random sequence u_t.

To show sufficiency of (10) in this case (i.e. when $f(u_t)$ is an ordinary function of u_t) note first that since $f_m(\cdot)$ is a regular sequence for $f(\cdot)$, $E\{f(u_t)\}$ is finite and is given by the limit shown in (8). (9) then follows directly by the ergodic theorem since $f(u_t)$ is an ordinary function, is measurable (as the limit of a sequence of ordinary measurable functions) and $E\{f(u_t)\}$ exists.□

3.2. EXAMPLE. Let $\delta(u_t)$ be the Dirac delta generalized function of the strictly stationary and ergodic time series u_t with continuous marginal density pdf(u). A corresponding regular sequence for $\delta(u_t)$ is

$$(13) \qquad \delta_m(u_t) = (m/\pi)^{1/2}e^{-mu_t^2} \ .$$

We have from (9)

(14) $$T^{-1}\Sigma_1^T\delta(u_t) \longrightarrow_{p,\text{a.s}} E\{\delta(u_t)\} = \int_{-\infty}^{\infty}\delta(u)\text{pdf}(u)du = \text{pdf}(0) .$$

The corresponding result for the sequence $\delta_m(u_t)$ is

$$T^{-1}\Sigma_1^T\delta_m(u_t) \longrightarrow_{p,\text{a.s.}} E\{\delta_m(u_t)\} = (m/\pi)^{1/2}\int_{-\infty}^{\infty}e^{-mu^2}\text{pdf}(u)du = \text{pdf}(0)\{1 + O(m^{-1})\} ,$$

where the last equality follows by virtue of the Laplace approximation.

3.3. EXAMPLE. Let x_t be the integrated process given in (1b) and suppose u_{0t} satisfies Assumption EC. We wish to show that

(15) $$T^{-2}\Sigma_1^T\delta(u_{0t})x_tx_t' \longrightarrow_d \text{pdf}(0)\int_0^1 B_xB_x' .$$

Note that by changing the probability space this can be written as an almost sure convergence result, in which case we can invoke the earlier definition of a.s. convergence of generalized functions of random variables in terms of regular sequences. The corresponding condition in the original probability space is then

(16) $$T^{-2}\Sigma_1^T\delta_m(u_{0t})x_tx_t' \longrightarrow_d E\{\delta_m(u_{0t})\}\int_0^1 B_xB_x' , \quad \forall m$$

where $\delta_m(\cdot)$ is the regular sequence for the delta function given in (13).

To establish (16) we will show that

(17) $$T^{-2}\Sigma_1^T[\delta_m(u_{0t}) - E\{\delta_m(u_{0t})\}]x_tx_t' \longrightarrow_p 0 .$$

First, in view of (3) we have $T^{-1/2}x_{[T\cdot]} \longrightarrow_d B_x(\cdot)$. Next, since u_{0t} is strong mixing, $z_{mt} = \delta_m(u_{0t}) - E\{\delta_m(u_{0t})\}$ is also (with the same mixing nmbers) and

$$\omega_m = \text{lrvar}(z_{mt}) = \Sigma_{j=-\infty}^{\infty}E(z_{mt}z_{mt+j}) ,$$

which is finite for all m. Note, however, that $\omega_m = O(m^{1/2})$ as $m \to \infty$, as is apparent from the fact that

$$\begin{aligned}
\operatorname{var}\{\delta_m(u_t)\} &= E\{\delta_m(u_t)^2\} - E\{\delta_m(u_t)\}^2 \\
&= (m/\pi)\int_{-\infty}^{\infty} e^{-2mu^2}\operatorname{pdf}(u)du - \{(m/\pi)^{1/2}\int_{-\infty}^{\infty} e^{-mu^2}\operatorname{pdf}(u)du\}^2 \\
&= (m/\pi)^{1/2}\operatorname{pdf}(0)\{1+O(m^{-1})\} - \operatorname{pdf}(0)^2\{1+O(m^{-1})\} \qquad (18)\\
&= 0(m^{1/2}) .
\end{aligned}$$

(Note that higher order covariances, *i.e.* $E\{\delta_m(u_t)\delta_m(u_{t+j})\}$ for $j \geq 1$, are of $O(1)$ as $m \to \infty$.) Thus, ω_m is unbounded as $m \to \infty$. But for all finite m, ω_m exists and we have the functional law

$$(19) \qquad T^{-1/2}\Sigma_1^{[Tr]}z_{mt} \longrightarrow_d B_{z_m}(r) \equiv BM(\omega_m) .$$

To prove (17) we simply note that

$$T^{-2}\Sigma_1^T[\delta_m(u_{0t}) - E\{\delta_m(u_{0t})\}]x_tx_t' = T^{-1/2}\Sigma_1^T(T^{-1/2}z_{mt})(T^{-1/2}x_t)(T^{-1/2}x_t') = O_p(T^{-1/2}) , \quad \forall m .$$

In fact, it is not difficult to establish the explicit limit

$$(20) \qquad T^{-3/2}\Sigma_1^T z_{mt}x_tx_t' \longrightarrow_d \int_0^1 dB_{z_m}B_xB_x' + \Delta_{zx}\int_0^1 B' + \int_0^1 B\Delta_{zx}' , \quad \forall m$$

where $\Delta_{zx} = \Sigma_{j=0}^{\infty}E(z_{mj}x_0)$, which is a limit result that is related to one given in Hansen (1992, Theorem 4.2). Thus, (17) holds and this gives (16) and thereby the required limit (15).

3.4. EXAMPLE. Under Assumption EC we have the functional CLT

$$(21) \qquad T^{-1/2}\Sigma_1^{[Tr]}\operatorname{sgn}(u_{0t}) \longrightarrow_d B_v(r) = BM(\Omega_{vv})$$

(see (3) above) with

$$(22) \qquad \Omega_{vv} = \operatorname{lrvar}(\operatorname{sgn}(u_{0t}) = \Sigma_{j=-\infty}^{\infty}E\{\operatorname{sgn}(u_{0t})\operatorname{sgn}(u_{0t+j})\} .$$

If we treat the ordinary function $\operatorname{sgn}(u_{0t})$ of u_{0t} as a generalized function of u_{0t}, result (21) can be viewed as a functional law for partial sums of generalized functions of random variables. The limit process $B_v(r)$ can then be interpreted as a generalized process although of course it also has meaning as an ordinary random process, *viz*. a Brownian motion with variance Ω_{vv}.

A regular sequence for $\operatorname{sgn}(u_{0t})$ can be constructed as in (11). We get

$$\text{(23)} \qquad \mathrm{sgn}_m(u_{0t}) = \int_{-\infty}^{\infty} \mathrm{sgn}(v) S(m(v-u)) m e^{-v^2/m^2} dv \ .$$

Note that with this construction we have

$$\begin{aligned} \mathrm{sgn}_m(-u_{0t}) &= \int_{-\infty}^{\infty} \mathrm{sgn}(v) S(m(v + u_{0t})) m e^{-v^2/m^2} dv \\ &= -\int_{+\infty}^{-\infty} \mathrm{sgn}(-w) S(m(u_{0t} - w)) m e^{-w^2/m^2} dw \\ &= -\int_{-\infty}^{\infty} \mathrm{sgn}(w) S(m(u_{0t} - w)) m e^{w^2 m^2} dw = -\mathrm{sgn}_m(u_{0t}) \ , \end{aligned}$$

so that $\mathrm{sgn}_m(u)$ is an odd function of u, just like $\mathrm{sgn}(u)$. In consequence,

$$E\{\mathrm{sgn}_m(u_{0t})\} = 0 \ ,$$

since the density of u_{0t} is symmetric.

Being a regular sequence, $\mathrm{sgn}_m(u)$ tends to zero faster than any negative power of $|u|$ as $|u| \to \infty$ (see Lighthill, 1958, p. 22). Indeed, recognizing that for large m the dominant part of the integral (23) comes from integrating in the neighborhood of $v = u$, we have from the Laplace approximation

$$\text{(24)} \qquad \mathrm{sgn}_m(u_{0t}) = \int_{-1}^{1} \mathrm{sgn}(u_{0t} + y/m) e^{-(u_{0t}+y/m)^2/m^2} S(y) dy = \mathrm{sgn}(u_{0t}) e^{-u_{0t}^2/m^2}\{1 + O(m^{-1})\} \ .$$

In view of this behavior for large $|u_{0t}|$, all moments of $\mathrm{sgn}_m(u_{0t})$ exist. Also $\mathrm{sgn}_m(u_{0t})$ is a measurable function of u_{0t} and is therefore mixing (with the same mixing numbers as u_{0t}). It follows that

$$\Omega_m = \mathrm{lrvar}\{\mathrm{sgn}_m(u_{0t})\} = \Sigma_{j=-\infty}^{\infty} E\{\mathrm{sgn}_m(u_{0t})\mathrm{sgn}_m(u_{0t+j})\} < \infty \ ,$$

and we have the functional law

$$\text{(25)} \qquad T^{-1/2}\Sigma_1^{[Tr]}\mathrm{sgn}_m(u_{0t}) \to_d B_m(r) \equiv BM(\Omega_m) \ , \quad \forall m \ .$$

Moreover, in view of (24) $E\{\mathrm{sgn}_m(u_{0t})\mathrm{sgn}_m(u_{0t+j})\} \to E\{\mathrm{sgn}(u_{0t})\mathrm{sgn}(u_{0t+j})\}$ and $\Omega_m \to \Omega_{vv}$ as $m \to \infty$, so that

(26) $$\lim_{m \to \infty} B_m(r) \equiv BM(\Omega_{vv}) .$$

Thus, (25) describes a regular sequence of functional laws whose limit (26) is equivalent to the limit of (21). In this sense, (25) & (26) give an alternative representation of the functional law (21), with the difference that the ordinary random variable $\text{sgn}(u_{0t})$ is treated as a generalized function of u_{0t} (by virtue of the regular sequence $\text{sgn}_m(u_{0t})$). Since $\text{sgn}(u_{0t})$ is an ordinary random variable and the limit process in (26) is an ordinary random process the weak convergence results are equivalent.

3.5. EXAMPLE. Assumption EC also validates weak convergence to stochastic integrals, as in (5) and (6) above. Repeating (5) for $r = 1$ we have

(27) $$T^{-1}\Sigma_1^T x_t \text{ sgn}(u_{0t}) \to_d \int_0^1 B_x dB_v + \Delta_{xv} .$$

As in the last example, we can again treat $\text{sgn}(u_{0t})$ as a generalized function of u_{0t}, using the regular sequence $\text{sgn}_m(u_{0t})$ given in (23). In the same way as we derived the functional law (25) for $\text{sgn}_m(u_{0t})$, we obtain

(28) $$T^{-1}\Sigma_1^T x_t \text{ sgn}_m(u_{0t}) \to_d \int_0^1 B_x dB_m + \Delta_{xm} , \quad \forall m$$

where $B_m \equiv BM(\Omega_m)$ and $\Delta_{xm} = \Sigma_{j=0}^{\infty} E\{u_{xt} \text{sgn}_m(u_{0t+j})\}$. Now $\Delta_{xm} \to \Delta_{xv} = \Sigma_{j=0}^{\infty} E\{u_{xt} \text{ sgn}(u_{0t+j})\}$ as $m \to \infty$ and thus in view of (26) we have

(29) $$\lim_{m \to \infty}\left(\int_0^1 B_x dB_m + \Delta_{xm}\right) \equiv \int_0^1 B_x dB_v + \Delta_{xv} .$$

It follows that (28) describes a regular sequence of weak convergence results whose limit, from (29), is distributionally equivalent to the limit of (27). Thus, (28) and (29) give a generalized function characterization of the limit law (27).

4. LAD AND FM-LAD ESTIMATION

The LAD estimator of β in model (1) is defined as the solution of the extremum problem

$$\beta_{\text{LAD}} = \text{argmin}[\Sigma_1^T |y_t - x_t'\beta|] \ . \tag{30}$$

We examine the asymptotic behavior of the estimator β_{LAD} and use this theory to suggest suitable modifications to the estimator that lead to improved asymptotic performance in nonstationary regression situations. Our approach to the development of the asymptotic theory uses generalized functions of random variables and the limit theory for such functions developed in Section 3 to deal with the fact that the objective criterion in (30) is not differentiable as an ordinary function of β.

We start with β_{LAD} and give its asymptotic distribution in the following result.

4.1. THEOREM. *Under Assumption* EC

$$T(\beta_{\text{LAD}} - \beta) \rightarrow_d [2h(0)\int_0^1 B_x B_x']^{-1}[\int_0^1 B_x dB_v + \Delta_{xv}] \ . \tag{31}$$

4.2. REMARKS

(i) Theorem 4.1 shows that β_{LAD} is consistent at the usual $O(T)$ rate for a nonstationary regression estimator. But like OLS, β_{LAD} suffers from second order asymptotic bias arising from the presence of Δ_{xv} in the second factor of (31) and the fact that the limit Brownian motions B_v and B_x are in general correlated (i.e. $\Omega_{vx} \neq 0$ in Ω_{ww}). In fact, formula (31) is very similar to the limit result for the OLS estimator $\hat{\beta}$, *viz.*

$$T(\hat{\beta} - \beta) \rightarrow_d \left(\int_0^1 B_x B_x'\right)^{-1}(\int_0^1 B_x dB_0 + \Delta_{x0})$$

(from Phillips and Durlauf, 1986).

(ii) The limit distribution (31) depends on the value at the origin of the probability density of u_{0t}, *i.e.* $h(0)$. In this respect (31) is similar to the usual limit theory for the LAD estimator that applies in the stationary or linear regression case. However, since (31) is not mixed normal in general the scale effects of $h(0)$ affect more than just the dispersion of the estimator.

(iii) When $v_t = \text{sgn}(u_{0t})$ is a martingale difference sequence with respect to $\mathcal{F}_{t-1} = \sigma(v_{s-1}, u_{xp} : s = t, t-1, ...; p = ... t+1, t, t-1, ...)$, then $\Delta_{xv} = 0$, $\Omega_{xv} = 0$ and (31) specializes to

$$(32) \qquad T(\beta_{\text{LAD}} - \beta) \to_d \left[2h(0)\int_0^1 B_x B_x'\right]^{-1}\left[\int_0^1 B_x dB_v\right] \equiv MN\left(0,\ (1/2h(0))^2\left(\int_0^1 B_x B_x'\right)^{-1}\right)$$

(since B_v and B_x are independent), which is a mixed normal limit that is comparable in form to the normal limit theory for LAD in stationary models. In this special case x_t is exogenous and the system has no feedback between v_t and u_{xt}.

4.3. THE FM-LAD Estimator

Our purpose is to modify the LAD estimator so that we obtain a mixed normal limit theory like (32) even when x_t is not exogenous. To do so we need to adjust for serial dependence to eliminate the one-sided long-run covariance Δ_{xv} and adjust for the endogeneity of x_t that is manifested in the long-run covariance Ω_{xv}. Our construction is based on the idea of the fully modified OLS estimator developed by Phillips and Hansen (1990). However, in the present case we need to take into account: (i) the extremum estimator properties of LAD (*i.e.* unlike OLS, there is no explicit formula for LAD); and (ii) the fact that the limit theory for LAD, as given in Theorem 4.1, relies on the robust function $v_t = \text{sgn}(u_{0t})$ of the equation errors rather than the errors themselves.

We define the fully modified least absolute deviation (FM-LAD) estimator of β in (1) as the following corrected version of β_{LAD}:

$$(33) \qquad \beta^+_{\text{LAD}} = \beta_{\text{LAD}} - [2\hat{h}(0)X'X]^{-1}[X'\Delta X\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{xv} + T\hat{\Delta}^+_{xv}] \ .$$

In (33) $X'X = \Sigma_1^T x_t x_t'$, $X'\Delta X = \Sigma_1^T x_t \Delta x_t'$, $\hat{h}(0)$ is a (nonparametric) consistent estimator of $h(0)$, the probability density of u_{0t} at the origin, $\hat{\Omega}_{xx}$ and $\hat{\Omega}_{xv}$ are consistent estimates of the long-run variance submatrices Ω_{xx} and Ω_{xv}, and $\hat{\Delta}^+_{xv}$ is a consistent estimate of the one-sided long-run covariance matrix

(34) $$\Delta_{xv}^{+} = \Sigma_{j=0}^{\infty}E(u_{x0}v_j^{+}) = \Delta_{xv} - \Delta_{xx}\Omega_{xx}^{-1}\Omega_{xv} ,$$

where

$$v_t^{+} = v_t - \Omega_{vx}\Omega_{xx}^{-1}\Delta x_t .$$

In order to estimate $\hat{\Delta}_{xv}^{+}$ we need first to estimate error v_t^{+}, which in turn involves the estimation of v_t. This is achieved by a first stage LAD regression which produces the error estimate $\hat{u}_{0t} = y_t - \beta_{\text{LAD}}'x_t$ and consequently $\hat{v}_t = \text{sgn}(\hat{u}_{0t})$. We then construct

(35) $$\hat{v}_t^{+} = \hat{v}_t - \hat{\Omega}_{\hat{v}x}\hat{\Omega}_{xx}^{-1}\Delta x_t ,$$

using conventional kernel estimates of the long-run covariance matrices Ω_{vx} and Ω_{xx}, whereupon we can estimate $\hat{\Delta}_{xv}^{+}$ as given by (34) directly by using a kernel estimate of the one-sided long-run covariance of u_{xt} and $\hat{v}_t^{+}$ [see Park and Phillips (1989), Andrews (1991) and Phillips (1993) for more details on kernel estimation of long-run covariance matrices]. Note from (34) that the estimation of $\hat{\Delta}_{xv}^{+}$ effectively involves the estimation of the four submatrices Δ_{xv}, Δ_{xx}, Ω_{xx} and Ω_{xv}. We use the notation $\hat{\Omega}_{\hat{v}x}$ in (35) to make it clear that our estimate of Ω_{vx} (and Δ_{vx} for that matter) relies on $\hat{v}_t$ rather than v_t, which is unobserved.

We can also write (33) in the form

$$\beta_{\text{LAD}}^{+} = \beta_{\text{LAD}} - [2\hat{h}(0)X'X]^{-1}T\hat{\Delta}_{xv}^{++} ,$$

where

$$\hat{\Delta}_{xv}^{++} = (T^{-1}X'\Delta X - \hat{\Delta}_{xx})\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{xv} + \hat{\Delta}_{xv} .$$

In this formula for $\hat{\Delta}_{xv}^{++}$ the first expression on the right side is an endogeneity correction. This term adjusts the regression estimate for potential endogeneity in the regressor x_t. In LAD estimation what is important is the correlation between Δx_t (the shocks in x_t) and the signed equation error function $v_t = \text{sgn}(u_{0t})$. Since there is persistence in the shocks to x_t we measure this correlation by means of Ω_{xv}. The variable $\Delta x_t'\Omega_{xx}^{-1}\Omega_{xv}$ then adjusts the regression coefficient for the conditional mean of the signed error v_t given Δx_t. The term involving $\hat{\Delta}_{xx}$ adjusts for the effects of serial dependence in Δx_t on the covariance $T^{-1}X'\Delta X$ in the limit. Finally, the second term in $\hat{\Delta}_{xv}^{++}$ above is $\hat{\Delta}_{xv}$ and this adjusts for serial covariance between the past history of shocks Δx_t and the signed error v_t. In all these cases we make the corrections by nonparametric (kernel) density estimation. Thus, β_{LAD}^{+} is a semiparametric LAD estimator with nonparametric corrections for endogeneity in the regressor x_t and serial dependence in the equation errors and shocks to x_t.

4.4 THEOREM. *Under Assumption* EC

$$(36) \qquad T(\beta_{\mathrm{LAD}}^{+} - \beta) \to_d \left[2h(0)\int_0^1 B_x B_x'\right]^{-1}\left[\int_0^1 B_x B_{v\cdot x}\right] \equiv MN\left(0,\ (1/2h(0))^2 \omega_{vv\cdot x}\left[\int_0^1 B_x B_x'\right]^{-1}\right)$$

where $B_{v\cdot x} = B_v - \Omega_{vx}\Omega_{xx}^{-1}B_x \equiv BM(\omega_{vv\cdot x})$ *and* $\omega_{vv\cdot x} = \Omega_{vv} - \Omega_{vx}\Omega_{xx}^{-1}\Omega_{xv} = \mathrm{lrvar}(v_t^{+})$.

4.5. REMARKS

(i) The limit theory of FM-LAD is similar to that of the FM-OLS estimator $\beta^{+} = (X'X)^{-1}(X'y^{+} - T\hat{\Delta}_{x0}^{+})$ where $y^{+} = y - \Delta X'\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{x0}$. This is given by

$$(37) \qquad T(\beta^{+} - \beta) \to_d \left(\int_0^1 B_x B_x'\right)^{-1}\left(\int_0^1 B_x dB_{0x}\right) \equiv MN\left(0,\ \omega_{00\cdot x}\left(\int_0^1 B_x B_x'\right)^{-1}\right),$$

where $B_{0\cdot x} = B_0 - \Omega_{0x}\Omega_{xx}^{-1}B_x \equiv BM(\omega_{00\cdot x})$ and $\omega_{00\cdot x} = \Omega_{00} - \Omega_{0x}\Omega_{xx}^{-1}\Omega_{x0}$. The relative asymptotic efficiency of the two estimators depends on the ratio $\omega_{vv\cdot x}/(2h(0))^2\omega_{00\cdot x}$, so that FM-LAD is more efficient than FM-OLS when

$$(38) \qquad h(0)^2 > \omega_{vv\cdot x}/4\omega_{00\cdot x}\ .$$

In the case where x_t is exogenous and u_{0t} is iid$(0, \sigma_{00}^2)$, we have $\omega_{vv\cdot x} = \omega_{vv} = 1$, $\omega_{00\cdot x} = \omega_{00} = \sigma_{00}^2$ and (38) reduces to

$$h(0)^2 > 1/4\sigma_{00}^2\ ,$$

which corresponds to the criterion for the asymptotic superiority of LAD over OLS in linear regression.

(ii) Wald statistics for testing restrictions on β can be constructed in the usual way from the limit theory in Theorem 4.4. For instance, consider the restrictions

$$H_0 : \varphi(\beta) = 0 \,,$$

where φ is a $q \times 1$ vector function with $\Phi(\beta) = \partial\varphi/\partial\beta'$ of full row rank q. The Wald statistic for testing H_0 based on FM-LAD is

$$(39) \qquad W^{+} = \varphi(\beta^{+}_{\mathrm{LAD}})'\left\{\Phi^{+}\left[4\hat{h}(0)^2X'X\right]^{-1}\Phi^{+\prime}\right\}^{-1}\varphi(\beta^{+}_{\mathrm{LAD}})/\hat{\omega}_{vv\cdot x} \,,$$

where $\hat{\omega}_{vv\cdot x} = \hat{\Omega}_{vv} - \hat{\Omega}_{vx}\hat{\Omega}^{-1}_{xx}\hat{\Omega}_{xv}$ is a consistent estimate of the conditional long-run variance $\omega_{vv\cdot x}$, $\hat{h}(0)$ is a (nonparametric) consistent estimate of $h(0)$ and $\Phi^{+} = \Phi(\beta^{+}_{\mathrm{LAD}})$. In view of (36) we have the limit $W^{+} \to \chi^2_q$ under H_0 by a simple deduction. Thus the statistic W^{+} can be used for testing H_0 in the usual way.

(iii) Fully modified standard errors for the β^{+}_{LAD} estimator can be constructed from (the square roots of)

$$(40) \qquad s_i^2 = (\hat{\omega}_{vv\cdot x}/2\hat{h}(0))[(X'X)^{-1}]_{ii} \,, \quad (i = 1, \ldots, k)$$

where $\hat{\omega}_{vv\cdot x} = \hat{\Omega}_{\hat{v}\hat{v}} - \hat{\Omega}_{\hat{v}v}\hat{\Omega}^{-1}_{xx}\hat{\Omega}_{x\hat{v}}$. The variance estimate (40) is based directly on the (conditional) asymptotic variance matrix that appears in (36). Correspondingly, we have the fully modified LAD t-ratios $t_i = (\beta^{+}_{i\mathrm{LAD}} - \beta_i)/s_i$, which are asymptotically $N(0, 1)$. These statistics simplify the statistical reporting of FM-LAD regressions -- in effect we report the estimated coefficients, standard errors and t-ratios in the usual way. The modifications that are built into these statistics mean that they can be interpreted as in conventional stationary linear regression.

5. FM-M ESTIMATION

A more general class of robust procedures is that of M-estimators. In the present case, these estimators can be defined by the extremum problem

$$(41) \qquad \beta_M = \mathrm{argmin}[\Sigma_1^T \rho(y_t - x_t'\beta)] \,,$$

for some function ρ. When $\rho(u) = |u|$ this includes the LAD estimator. Other common choices are $\rho(u) = |u|^{\delta}$ for $\delta \in [1, 2]$, thereby including OLS when $\delta = 2$, and the Huber (1964) loss function

$$(42) \qquad \rho_c(u) = \begin{cases} (1/2)u^2 & \text{for } |u| \le c \\ c|u| - (1/2)c^2 & \text{for } |u| > c , \end{cases}$$

which combines the OLS criterion for deviations bounded by the parameter c with the LAD criterion for bigger deviations.

The estimator β_M can also be defined as a solution to the equation

$$(43) \qquad \Sigma_1^T x_t \psi(y_t - x_t{}'\beta_M) = 0 ,$$

and when ρ is differentiable and $\psi = \rho'$ (43) are the first order conditions. The definitions (41) and (43) are equivalent when ρ is convex and differentiable because in that case there is only one solution to (43). A scale estimate can also be employed in the criteria (41) and (43) and this can be obtained using the residuals of a preliminary consisent regression (possibly by OLS), as discussed by Huber (1981).

Like the LAD and OLS estimators, β_M needs some modification before it has good asymptotic properties in nonstationary regressions. We will construct a fully modified M-estimator β_M^+ to improve the asymptotic behavior of β_M and the construction is similar to that of β_{LAD}^+. As in the LAD case, we first need the limit theory for the unmodified estimator β_M. This calls for some additional conditions that relate to the properties of the functions that appear in (41) and (43).

ASSUMPTION ML (*M-estimator loss function conditions*)

(a) $\psi(u_t)$ *has mean zero and* $\|\psi(u_t)\|_p < \infty$,

(b) ψ' *is Lipschitz continuous and* $\|\psi'(u_t)\|_p < \infty$, *for some* $p > \beta > 2$, *as in* (2) *above.*

Conditions of this type are fairly standard in the development of M-estimator asymptotics. The p'th moment conditions (which relate to the strong mixing condition (2) in EC) on ψ and ψ' in (a) and (b) are helpful because of the allowance for serial dependence in u_t (cf. Knight, 1991) and because of the need to establish results on weak convergence for sample covariances such as $T^{-1}\Sigma_1^T x_t \psi(u_t)$ to stochastic integrals with drift. However, for many ψ functions these conditions will be implied by the corresponding conditions on u_t, and often ψ and ψ' are bounded, in which case they hold automatically. The centering condition $E\{\psi(u_t)\} = 0$ in ML(a) is the analogue for M-estimation of the zero mean and zero median conditions for OLS and LAD estimation.

Some M-estimators are excluded by the differentiability condition ML(b). When ψ' fails to exist at a finite number of points, we can proceed by treating ψ and ψ' as generalized functions. The asymptotic results given below will then continue to hold under some additional conditions on the probability density $h(u)$ of u_t, so that for instance we could write

$$E\{\psi'(u_t)\} = \int_{-\infty}^{\infty}\psi'(u)h(u)du = -\int_{-\infty}^{\infty}\psi(u)h'(u)du$$

i.e. this linear functional of the generalized function $\psi'(u_t)$ of the random variable u_t is equivalent to $-\int_{-\infty}^{\infty}\psi(u)h'(u)du$, which exists as an ordinary function. In the Addendum to the proof of Theorem 5.1 below (see Section 8.5) we will outline how this particular extension of the theory proceeds. The development follows our analysis of LAD asymptotics using generalized functions of random variables and generalized Taylor series.

Here we will focus attention on the nonstationary regression M-estimator asymptotics and the construction of the FM-M estimator

5.1. THEOREM. *Let Assumptions* EC *and* ML *hold. Suppose also that either of the following two conditions apply:*

(a) ρ *is convex,* $\psi = \rho'$ *and* β_M *satisfies* (41);

(b) β_M *is a solution of* (43) *and* $T^{1/2}(\beta_M - \beta) = o_p(1)$. *Then*

$$(44) \qquad T(\beta_M - \beta) \to_d \left[E\{\psi'(u_{0t})\}\int_0^1 B_x B_x'\right]^{-1}\left[\int_0^1 B_x dB_\psi + \Delta_{x\psi}\right]$$

where

$$B_\psi = BM(\Omega_{\psi\psi})\ , \quad \Omega_{\psi\psi} = \Sigma_{j=-\infty}^{\infty}E\{\psi(u_{0t})\psi(u_{0t+j})\}$$

and

$$\Delta_{x\psi} = \Sigma_{j=0}^{\infty}E\{u_{xt}\psi(u_{0t+j})\}\ .$$

5.2. FM-M ESTIMATION

As with the construction of the FM-LAD estimator, our purpose is to modify the M-estimator β_M so that the second order bias effects in the limit theory (44) are removed and the limit distribution is mixed normal. The required corrections are similar to those used in the LAD case and we define the FM-M estimator as

$$\beta_M^+ = \beta_M - \left[\{T^{-1}\Sigma_1^T\psi'(\hat{u}_{0t})\}X'X\right]^{-1}[X'\Delta X\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{x\psi} + T\hat{\Delta}_{x\psi}^+] , \tag{45}$$

where $\hat{\Omega}_{x\psi}$ is a consistent estimator of

$$\Omega_{x\psi} = \Sigma_{j=-\infty}^{\infty}E\{u_{xt}\psi(u_{0t+j})\}$$

and $\hat{\Delta}_{x\psi}^+$ is a consistent estimator of

$$\Delta_{x\psi}^+ = \Delta_{x\psi} - \Delta_{xx}\Omega_{xx}^{-1}\Omega_{x\psi} .$$

Again, all of these component matrices can be estimated using kernel techniques. But we do need a preliminary consistent estimate of β, say β_M, to construct the residuals $\hat{u}_{0t}$ from which we can form the function $\psi(\hat{u}_{0t})$, which is required for the estimation of $\Omega_{x\psi}$ and $\Delta_{x\psi}$.

5.3. THEOREM. *Under the conditions of Theorem* 5.1

$$\begin{aligned} T(\beta_M^+ - \beta) &\rightarrow_d \left[E\{\psi'(u_{0t})\}\int_0^1 B_xB_x'\right]^{-1}[\int_0^1 B_x dB_{\psi\cdot x}] \\ &\equiv MN\left(0,\ \omega_{\psi\psi\cdot x}[E\{\psi'(u_{0t})\}]^{-2}\left[\int_0^1 B_xB_x'\right]^{-1}\right) \end{aligned} \tag{46}$$

where

$$B_{\psi\cdot x} \equiv BM(\omega_{\psi\psi\cdot x}) , \ \omega_{\psi\psi\cdot x} = \Omega_{\psi\psi} - \Omega_{\psi x}\Omega_{xx}^{-1}\Omega_{x\psi} .$$

5.4. REMARKS

(i) In the case where x_t is exogenous and u_{0t} is iid(0, σ_{00}^2) the limit theory given in (46) reduces to

(47) $$T(\beta_M^+ - \beta) \rightarrow_d \mathrm{var}(\psi(u_{0t}))^{1/2}\Big[E\{\psi'(u_{0t})\}\int_0^1 B_x B_x'\Big]^{-1}[\int_0^1 B_x dW]$$

where W is standard Brownian motion independent of B_x. Observe that the limit (47) depends on ψ only through the factor

(48) $$\mathrm{var}(\psi(u_{0t}))^{1/2}/E(\psi'(u_{0t})) \ .$$

Consequently the efficiency of the estimator β_M^+ depends on this factor also, just as it does in the case of linear regression -- see, for example, Huber (1981, p. 173). If the density $h(u)$ of u_{0t} is continuously differentiable then the M estimators $\hat{\beta}_M$, β_M^+ will be asymptotically efficient in this case (note that these two estimators are asymptotically equivalent under the conditions of this remark) if $\psi(\cdot)$ is chosen to satisfy

(49) $$\psi(u) = -ch'(u)/h(u) \ , \quad \text{for} \quad c \neq 0$$

(cf. Huber, 1981, pp. 70, 176). When the density $h(\cdot)$ is unknown, there is the possibility of adaptive estimation as recently discussed by Jeganathan (1988).

(ii) In the general case we can write the limit (46) as

(50) $$T(\beta_M^+ - \beta) \rightarrow_d \frac{\omega_{\psi\psi\cdot x}^{1/2}}{[E\{\psi'(u_{0t})\}]}\Big(\int_0^1 B_x B_x'\Big)^{-1}(\int_0^1 B_x dW) \ ,$$

where $W \equiv BM(1)$ is independent of B_x. So the limit distribution of the class of all FM-M estimators depends on the $\psi(\cdot)$ function only through the factor

$$\omega_{\psi\psi\cdot x}^{1/2}/E\{\psi'(u_{0t})\} = \mathrm{lrvar}\{\psi(u_{0t})\,|u_{xt}\}^{1/2}/E\{\psi'(u_{0t})\} \ .$$

It will be interesting to consider the issue of an optimal estimator in this class. Note that FM-M estimation is semiparametric and $\psi(\cdot)$ depends on the equation error u_{0t}. Maximum likelihood estimation on the other hand involves the complete specification of the system including the transient dynamics of the vector error process $u_t = (u_{0t}, u_{xt}')'$. If the latter is parametric, like the linear process $u_t = C(L;\,\theta)\varepsilon_t = \Sigma_0^\infty C_j(\theta)\varepsilon_{t-j}$ with $\Sigma_0^\infty j^{1/2}\|C_j\| < \infty$, then the likelihood can be constructed using a form of innovations algorithm (as when u_t is ARMA). For the Gaussian case the results in Phillips (1991, Theorem 1′) confirm that the limit distribution of the maximum likelihood estimator (MLE) of β is of the form given in (50). Indeed, FM-M estimation is optimal with $\psi(u) = u$ in this case, *i.e.* the optimal FM-M estimator is just FM-OLS. It will be interesting to try to extend this theory to the non-Gaussian case and to develop a theory of optimal semiparametric M-estimation. This task will be left for later work.

(iii) Theorem 5.3 can be used as a basis for inference using the FM-M estimator β_M in the same way as FM-LAD (refer to Remark 4.5(ii) & (iii)). Thus, to test H_0 as in 4.5(ii) we can use the Wald statistic

$$W_M^+ = \varphi(\beta_M^+)'\left\{\Phi_M^+\left[T^{-1}\Sigma_1^T\psi'(\hat{u}_{0t})\right]^{-2}(X'X)^{-1}\Phi_M^{+\prime}\right\}^{-1}\varphi(\beta_M^+)/\hat{\omega}_{\psi\psi\cdot x} ,$$

where $\Phi_M^+ = \Phi(\beta_M^+)$, $\hat{\omega}_{\psi\psi\cdot x} = \hat{\Omega}_{\psi\psi} - \hat{\Omega}_{\psi x}\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{x\psi}$ is an estimate of the conditional long-run variance of $\psi(u_{0t})$ given Δx_t, and $\hat{u}_{0t} = y_t - \beta_M^{+\prime}x_t$ is the residual from the FM-M regression. The latter quantity is used in the sample estimate $T^{-1}\Sigma_1^T\psi'(\hat{u}_{0t})$ of $E\{\psi'(u_{0t})\}$ and in the construction of the $\hat{\omega}_{\psi\psi\cdot x}$ which relies on the sample values $\hat{\psi}_t = \psi(\hat{u}_{0t})$. In the light of Theorem 5.3, we have $W_M^+ \sim \chi_q^2$ under the null H_0. FM-M coefficient standard errors and t-ratios are constructed in the same way as in Remark 4.5(iii) for FM-LAD estimation.

6. Extensions to Models with Infinite Variance Errors

This section outlines some extensions of the theory to the case where the errors in model (1) have infinite variance. Our purpose is to sketch the development and indicate some interesting points of departure from the earlier theory.

It is simplest to suppose that the tail behavior of the errors in (1a) and (1b) is of the same form. It is especially helpful to require that the components of u_{xt} have distributions with the same tail shape, for then the normalizing constant in central limit theory for partial sums of u_{xt} is a scalar. The general case of an operator stable law when the components of u_{xt} have different tail shapes (e.g. follow asymptotic Pareto laws with different slope coefficients) does not, to the author's knowledge anyway, seem to have been worked out. However, since one of the main applications of a regression theory in the infinite variance case is to series like spot and forward exchange rates the restriction of comparable tail behavior does not seem to be too limiting. At each point in time, spot and forward rates reflect the same information set and economic fundamentals. As a consequence, it seems reasonable to model such series with distributions that have related tail shape.

Accordingly we will confine our attention to limit laws that are of the symmetric α-stable (SαS) form. Thus, a k-vector ξ has an SαS distribution in $\mathbb{R}^k$ if its characteristic function is of the form

$$(51) \qquad E(e^{i\xi' p}) = \exp\{-\int_{S_k} |p'h|^{\alpha}\Gamma(dh)\} ,$$

where $S_k = \{h \in \mathbb{R}^k : h'h = 1\}$ is the unit sphere in $\mathbb{R}^k$ and $\Gamma(\cdot)$ is a probability measure (possibly discrete) on S_k. Paulauskas (1976) provides a discussion of multivariate stable distributions in this class. The most common examples (arising from discrete measures on S_k) are exponentials of powers of quadratic forms such as $\exp\{-(p'\Sigma p)^{\alpha/2}\}$, which include the multivariate normal when $\alpha = 2$.

We will assume the following condition applies to $u_t = (u_{0t}, u_{xt}')'$ in place of Assumption EC.

ASSUMPTION EC^2 *(Error Condition 2)*

(a) u_t *is generated by the linear process*

$$(52) \qquad u_t = D(L)\varepsilon_t = \Sigma_{j=0}^{\infty} D_j \varepsilon_{t-j} , \quad D_0 = I , \quad |D(1)| \neq 0 ,$$

where ε_t *is an* iid *sequence of random vectors whose components have infinite variance and are each in the domain of normal attraction of a stable law of order* $\alpha \in (0, 2)$. *The coefficient matrices in* (52) *satisfy the summability condition*

$$(53) \qquad \Sigma_0^{\infty} j \|D_j\|^{\delta} < \infty , \quad \text{with} \quad 0 < \delta < \alpha \wedge 1 .$$

(b) *Partial sums of the* ε_t *in* (52) *satisfy the following functional limit law in the product space* $D[0, 1]^{k+1}$ of k+1 *copies of* $D[0, 1]$ *with the product Skorohod topology:*

$$(54) \qquad a_T^{-1} \Sigma_1^{[Tr]} \varepsilon_t \rightarrow_d U_\alpha(r) .$$

The limit process $U_\alpha(r)$ in (54) is an α-stable process in $D[0,1]^{k+1}$ whose increments are SαS, *i.e. have a characteristic function of the form* (51), *and $a_T = T^{1/\alpha}$ is a normalizing constant.*

(c) *The sequence u_t is strong mixing with mixing numbers α_m that satisfy the summability condition*

$$\Sigma_0^\infty \alpha_m < \infty .$$

(d) *Condition* EC(c) *holds.*

Condition EC^2(b) is a "high level" condition. Since each of the components of ε_t is in the domain of normal attraction of a stable law with exponent α simple sufficient conditions for a component-wise version of (54) are available (see, for instance, Resnick, 1986; Chan and Tran, 1989; and Knight, 1991 for earlier applications). Condition (b) requires joint convergence and specifies the limit process to be in the SαS class. Condition (a) specifies that u_t has a linear process form and this facilitates the use of arguments like those in Phillips (1991) and Phillips and Solo (1992) for obtaining the limit distributions of certain functions of partial sums of u_t. The mixing condition (c) is useful because we need to work with and characterize the dependence properties of functions of the error process u_{0t}.

Our main result is the following:

6.1. THEOREM. *Under Assumption* EC^2:

(a) *the estimators β_{LAD} and β^+_{LAD} have the common limit distribution*

$$T^a(\beta_{\text{LAD}} - \beta),\ T^a(\beta^+_{\text{LAD}} - \beta) \rightarrow_d \left(\int_0^1 U_{x\alpha} U_{x\alpha}'\right)^{-1} \left(\int_0^1 U^-_{x\alpha} dB_v\right)$$

$$\equiv MN\left(0,\ (1/2h(0))^2 \Omega_{vv} \left(\int_0^1 U_{x\alpha} U_{x\alpha}'\right)^{-1}\right)$$

where $a = 1/2 + 1/\alpha$, $U_{x\alpha}(r) = D_x' U_\alpha(r)$ and $U^-_{x\alpha}(r) = U_{x\alpha}(r-)$ is the left limit of the process $U_{x\alpha}$. Here D_x is the second submatrix of $D(1)' = [D_0, D_x]$ in a partition of $D(1)'$ that is conformable with $u_t = (u_{0t}, u_{xt}')'$.

(b) *The estimators* β_M *and* β_M^+ *have the common limit distribution*

$$T^a(\beta_M - \beta),\ T^a(\beta_M^+ - \beta) \to_d \left[E\{\psi'(u_{0t})\}\int_0^1 U_{x\alpha}U_{x\alpha}'\right]^{-1}\left[\int_0^1 U_{x\alpha}^- dB_\psi\right]$$

$$= MN\left[0,\ \Omega_{\psi\psi}\left(E\{\psi'(u_{0t})\}\int_0^1 U_{x\alpha}U_{x\alpha}'\right)^{-1}\right].$$

6.2. REMARKS

(i) Theorem 6.1 shows that the robust estimators β_{LAD} and β_M are $O(T^a)$ consistent. Since $a = 1/2 + 1/\alpha > 1$ for $\alpha \in (0, 2)$, these estimators converge faster than the OLS and FM-OLS estimators, whose convergence rate is still $O(T)$ in the infinite variance case. The situation is analogous to the case of coefficient estimation in an AR(1) with a unit root. In that case Knight (1989, 1991) showed that LAD and M-estimators of the unit root have a rate of convergence equal to $O(T^a)$; and Chan and Tran (1989) and Phillips (1991) showed that OLS and semi-parametrically corrected OLS have convergence rates of $O(T)$. Thus, just as in the unit root case, the robust estimators β_{LAD} and β_M are infinitely more efficient than OLS-based estimation procedures when there are infinite variance errors.

(ii) Interestingly, β_{LAD} and β_{LAD}^+ are asymptotically equivalent in the infinite variance case. Thus, there is no need to make corrections for endogeneity or serial correlation when the errors have infinite variance. Intuitively this is because the robust estimators control the effects of outliers in the errors but retain the additional strength in the signal from x_t that arises from the presence of heavy tailed and persistent shocks. In doing so, these estimators not only achieve a higher rate of convergence than OLS and FM-OLS but they also remove the endogeneity effects of the regressors and the effects of dependence between the past history of the shocks that drive x_t and the equation error u_{0t}. In effect, whereas $T^{-2}\Sigma_1^T u_{xt}u_{0t} = O_p(1)$ (it converges weakly to the double stochastic integral or quadratic covariation process $\int_0^1 dU_{x\alpha}dU_{0\alpha}$, where $U_\alpha(r) = (U_{0\alpha}(r),\ U_{x\alpha}'(r))'$) we have $T^{-2/\alpha}\Sigma_1^T u_{xt}\ \mathrm{sgn}(u_{0t}) = o_p(1)$, so that the endogeneity and serial dependence effects wash out in robust estimation with heavy-tailed errors.

(iii) Since no modifications to β_{LAD} are required in the infinite variance case, we may as well use β_{LAD} rather than β_{LAD}^+ if it were known that $\alpha < 2$. On the other hand, if we do use

$\beta^+_{\rm LAD}$ then it follows from the theorem that nothing is lost asymptotically because the modifications in $\beta^+_{\rm LAD}$ wash out in large samples. As we will see, however, in the simulations reported in the next section there is clear evidence that $\beta^+_{\rm LAD}$ does pay a price for the modifications over $\beta_{\rm LAD}$ in terms of additional sampling dispersion.

(iv) The mixed normality of the robust estimators in the limit means that standard errors, t-ratios and Wald tests can be constructed in the usual way as shown in Sections 4 and 5.

7. Some Simulation Results and an Empirical Illustration

7.1. Simulations

A small simulation study was conducted to study the sampling performance of the new robust regression estimators. The model we used for data generation was the following:

$$\begin{aligned} y_t &= \beta x_t + u_{0t}, \quad \beta = 1 \\ \Delta x_t &= u_{xt} \end{aligned} \tag{55}$$

where

$$\begin{aligned} u_{0t} &= \{1/(1+c^2)^{1/2}\}\varepsilon_{1t} + \{c/(1+c^2)^{1/2}\}\varepsilon_{2t}, \\ u_{xt} &= \varepsilon_{2t}, \end{aligned} \tag{56}$$

and ε_{1t} and ε_{2t} are each serially independent, independent of each other and are drawn from the following four distributions:

D(a) : $N(0, 1)$;

D(b) : t distribution with 4 degrees of freedom (t_4);

D(c) : t distribution with 2 degrees of freedom (t_2);

D(d) : standard Cauchy.

According to the construction (56), the equation error u_{0t} is an orthonormal combination of the independent shocks (ε_{1t}, ε_{2t}). The parameter c controls the degree of association between u_{0t} and u_{xt} and therefore measures the amount of dependence in the regressor x_t in (55). When $c = 0$ x_t is exogenous, when $|c| = 1$ the squared correlation between u_{xt} and u_{0t} is 1/2 and when $c \to \infty$ u_{xt} and u_{0t} become linearly dependent.

The parameter values chosen for our small simulation study were $c = -1, 0, 1$ and $T = 100$. We computed FM-OLS, FM-LAD, LAD and RRR estimates of the regression coefficient in (55). From 5,000 replications in each case, kernel density estimates were calculated of the sampling distributions of these estimates. The results are shown in Figures 2–5, where each figure in this sequence displays the outcome for an error distribution in the aforementioned groups D(a)–D(d). The figures show the estimated densities of the LAD, FM-LAD and FM-OLS estimates, as well as the estimates from a reduced rank regression with two lags in the regression (i.e. one lagged difference) which is denoted RRR-2 in the figure legends. The estimates are centered on the true coefficient and are scaled by the sample size, so that the given densities are those of $T(\hat{\beta}-\beta)$ for each estimator $\hat{\beta}$. In each case we show the results for the association parameter value $c = 1$. Very similar results were obtained for $c = -1$ and $c = 0$ with the exception that LAD shows no bias in the latter case, as would be anticipated from the asymptotic theory given in Theorem 4.1 (noting that $\Delta_{xv} = 0$ and B_v and B_x are independent when $c = 0$).

Figure 2 gives the densities for normal errors. LAD is biased ($c = 1$); FM-OLS, FM-LAD and RRR-2 are all well centered; FM-OLS shows the best concentration and, interestingly, FM-LAD has better concentration than RRR-2. Thus, although FM-OLS and RRR-2 are asymptotically optimal in this case, FM-LAD appears to do well and to be superior to RRR-2 in this finite sample case.

Figure 3 gives the results for the case of t_4 error distributions. The outcome is very similar to the case of Gaussian errors. However, FM-LAD is now closer to FM-OLS, although FM-OLS still dominates. FM-LAD dominates RRR-2 by a wider margin than in the Gaussian error case. LAD is still biased (again $c = 1$).

- - - - - - - - - - - - - - - -

Figures 2 and 3 about here

- - - - - - - - - - - - - - - -

Figure 2 gives the outcome for t_2-errors. Under these heavy-tailed error distributions the rankings have changed. FM-LAD dominates both FM-OLS and RRR-2 in terms of concentration. FM-OLS continues to outperform RRR-2. LAD is much less biased in this case.

- - - - - - - - - - - - - - - -
Figures 4 and 5 about here
- - - - - - - - - - - - - - - -

Figure 5 shows the same densities under Cauchy errors. The results are dramatic. FM-OLS and RRR-2 are widely dispersed. FM-LAD dominates FM-OLS and RRR-2 by a wide margin; and LAD is by far the most concentrated. Note that in this case both LAD and FM-LAD have rates of convergence (here order $T^{3/2}$) that exceed that of FM-OLS and RRR-2 (here order T), so we expect FM-OLS and RRR-2 both to be poor in relation to the robust estimates. Although FM-LAD and LAD have the same limit distribution in this case (see Theorem 6.1) the sampling distributions are very different, with the LAD estimator showing much more concentration. Thus, FM-LAD does pay a price in finite samples for the additional correction terms in this case of very heavy-tailed errors.

7.2. An Empirical Illustration

The robust and nonrobust regression procedures were used to estimate the foreign exchange market regression equation

(57) $$s_{t+k} = \alpha + \beta f_{t,k} + u_{t+k}$$

that relates the natural logarithm of the forward exchange rate for a k-period ahead contract delivery $f_{t,k}$ to the logarithm of the future spot rate of the same currency s_{t+k}. Daily exchange rate data for the Australian dollar over the period January 1984–April 1991 were used and the forward contract period was 3 months. There were 1830 observations in total.

- - - - - - - - - - - -
Figure 6 about here
- - - - - - - - - - - -

Figure 6 shows the sample data and the fitted regression lines obtained by FM-LAD, FM-OLS and RRR-6 (reduced rank regression with six lags).

In spite of the large number of observations there are big differences in the regression coefficients. Both FM-OLS and RRR-6 seem to be substantially affected by outlying observations (particularly the small spot rate and moderate forward rate pairs). The FM-LAD regression line seems much less affected by these outliers and seems to follow the general cluster of data more closely. The estimated coefficients are given in Table 1 and these show that the numerical differences between the estimates are indeed substantial. Note that the FM-OLS and RRR-6 estimates of the slope coefficient are both much closer to unity than the FM-LAD estimate. Thus, inference about the forward rate unbiasedness hypothesis (under which $\beta = 1$ in (57)) is affected by the regression procedure: the nonrobust estimates are biased in favor of this hypothesis, while the robust estimates do not support it. The reader is referred to the author's paper (1993) for a detailed empirical analysis of these data.

Table 1: Estimates of Equation (57)
(standard errors in parentheses)

	α	β
FM-LAD	-0.071 (0.012)	0.700 (0.040)
FM-OLS	-0.025 (0.029)	0.883 (0.092)
RRR-6	-0.003	0.935

8. Further Useful Extensions

The robust regression methods developed here are designed for use in single equation nonstationary regression. They can be extended to multivariate regressions or subsystem cointegrating regression where there is more than one cointegrating relation. There is also the possibility of adaptive estimation wherein the error distribution is estimated and used in the estimation of the regression coefficients. Jeganathan (1988) discussed this possibility in the context of regression models like (1) with serially independent errors and exogenous regressors. Given the extensive use of vector autoregressive models in empirical econometric research and the growing use of RRR methods in VAR models it would seem useful to develop adaptive estimation methods for these models also.

9. Proofs

9.1. PROOF OF THEOREM 4.1. We start by defining the process

(P1) $$Z_T(g) = \Sigma_1^T\{|u_{0t} - T^{-1}x_t'g| - |u_{0t}|\} .$$

The vector $\hat{g}$ which minimizes $Z_T(g)$ is just $\hat{g}_T = T(\beta_{\mathrm{LAD}} - \beta)$. Since $Z_T(g)$ is convex, we can make use of the approach given by Knight (1989). In particular, by Knight's Lemma A it follows that if the finite dimensional distributions of $Z_T(g)$ converge to those of a process $Z(g)$ and $Z(g)$ has a unique minimum at $\hat{g}$ then the convexity of Z_T implies that $\hat{g}_T \longrightarrow_d \hat{g}$. This also means that $\beta_{\mathrm{LAD}} \longrightarrow_p \beta$ and a separate argument for consistency of β_{LAD} is not required. (Pollard (1991) used a similar approach to LAD asymptotics but his examples 1 and 2 give normal distribution limits and do not involve random quadratic elements in the limiting process.)

We will establish convergence of the unidimensional distributions of $Z_T(g)$ and then the higher dimensional distributions converge in a corresponding way by applying the Cramer-Wold device. Note that the process $Z_T(g)$ involves the ordinary random functions $|u_{0t} - T^{-1}x_t'g|$ and $|u_{0t}|$ and is itself an ordinary random process. But, it can also be treated as a generalized process (here a stochastic process defined in terms of generalized functions of random variables) by treating the function $f(\xi_t) = |\xi_t|$ of the random variable ξ_t as a generalized function of the random variable ξ_t, *i.e.* by using the regular sequence of random variables

$$f_m(\xi_t) = \int_{-\infty}^{\infty}|v|S(m(v-\xi))me^{-v^2/m^2}dv$$

to represent $f(\xi_t)$ as in (11). Thus, as a generalized process $Z_T(g)$ is defined by the following regular sequence of processes

(P2) $$Z_{Tm}(g) = \Sigma_1^T\{f_m(u_{0t} - T^{-1}x_t'g) - f_m(u_{0t})\} .$$

We now proceed to develop a Taylor expansion of $Z_{Tm}(g)$ and to characterize its limit behavior. Expanding $Z_{Tm}(g)$ in a Taylor series about $g = 0$ we have

(P3) $$Z_{Tm}(g) = -T^{-1}\Sigma_1^T f_m^{(1)}(u_{0t})x_t'g + (1/2)T^{-2}\Sigma_1^T f_m^{(2)}(u_{0t}^*)g'x_tx_t'g ,$$

where $f_m^{(1)}(\cdot)$ and $f_m^{(2)}(\cdot)$ denote the first and second derivatives of $f_m(\cdot)$, and u_{0t}^* lies between

u_{0t} and $u_{0t} - T^{-1}x_t'g$. Since $f(\xi)$ has first derivative everywhere except $\xi = 0$ and $f'(\xi) = \mathrm{sgn}(\xi)$ exists as an ordinary function, it follows that the regular sequence $f_m^{(1)}(\cdot)$ is a regular sequence for sgn(·) treated as a generalized function (Lighthill, 1958, Theorem 10, p. 24). Thus, $f_m^{(1)}(\cdot)$ is equivalent to the regular sequence $\mathrm{sgn}_m(\cdot)$ given in (23). Similarly, $f_m^{(2)}(\cdot)$ is a regular sequence for the generalized function

$$d/d\xi(\mathrm{sgn}(\xi)) = 2\delta(\xi) ,$$

(cf. Lighthill, 1958, p. 23) and is therefore equivalent to the regular sequence $2\delta_m(\cdot)$ given in (13).

Next, we consider the limit behavior of the two components of $Z_{Tm}(g)$ in (P3). First, by Example 3.5 we have

$$\text{(P4)} \qquad T^{-1}\Sigma_1^T \mathrm{sgn}_m(u_{0t})x_t'g \longrightarrow_d (\textstyle\int_0^1 dB_m B_x' + \Delta_{xm}')g$$

and the limit process as $m \to \infty$ is equivalent to $(\int_0^1 dB_v B_x' + \Delta)g$, *i.e.*

$$\text{(P5)} \qquad \lim_{m\to\infty}(\textstyle\int_0^1 dB_m B_x' + \Delta_{xm}')g \equiv (\int_0^1 dB_v B_x' + \Delta_{xv}')g ,$$

which is of course an ordinary random variable.

For the second term of (P3), observe that the regular sequence $\delta_m(\cdot)$ is differentiable and has bounded derivative (with a bound dependent on m) for all m. Thus,

$$|\delta_m(u_{0t}^*) - \delta_m(u_{0t})| \le K_m |T^{-1}x_t'g| , \quad \forall m .$$

and therefore

$$|T^{-2}\Sigma_1^T\{\delta_m(u_{0t}) - \delta_m(u_{0t}^*)\}g'x_tx_t'g|$$

$$\le K_m T^{-3}\Sigma_1^T |(x_t'g)|(g'x_tx_t'g) \longrightarrow_p 0 , \quad \forall m$$

uniformly over g in compact sets. Now using Example 3.3 we have

$$\text{(P6)} \qquad T^{-2}\Sigma_1^T\delta_m(u_{0t})g'x_tx_t'g \longrightarrow_d E\{\delta_m(u_{0t})\}\textstyle\int_0^1(g'B_x)^2 ,$$

whose limit as $m \to \infty$ is $g'\int_0^1 B_x B_x' g$.

Combining (P4) and (P6) we deduce that

$$\text{(P7)} \qquad Z_{Tm}(g) \to_d -\left(\int_0^1 dB_m B_x' + \Delta_{xm}'\right)g + E\{\delta_m(u_{0t})\}g'\left(\int_0^1 B_x B_x'\right)g = Z_m(g)\ , \quad \text{say,} \quad \forall m$$

uniformly over g in compact sets. In view of (P5) and since $\lim_{m\to\infty} E\{\delta_m(u_{0t})\} = E\{\delta(u_{0t})\} = \text{pdf}(0)$, the limit process $Z_m(g)$ has the following equivalent representation as $m \to \infty$

$$\text{(P8)} \qquad Z(g) = -\left(\int_0^1 dB_v B_x' + \Delta_{xv}'\right)g + \text{pdf}(0)g'\left(\int_0^1 B_x B_x'\right)g\ ,$$

which is an ordinary random variable.

Since $Z_{Tm}(g) \to_d Z_m(g)$, $\forall m$, and $\lim_{m\to\infty} Z_m(g) \equiv Z(g)$, we have established the weak convergence of $Z_T(g) \to_d Z(g)$ as generalized processes uniformly over g in compact sets. But both $Z_T(g)$ and $Z(g)$ exist as ordinary random processes so that the weak convergence applies in this sense also. The argument that we can neglect the region outside a suitable compact set for g relies on the convexity of $Z_T(g)$ and is the same as that given in Knight (1989, p. 277). Finally,

$$\text{(P9)} \qquad \hat{g} = \operatorname{argmin} Z(g) = \left[2\text{pdf}(0)\int_0^1 B_x B_x'\right]^{-1}\left[\int_0^1 B_x dB_v + \Delta_{xv}\right]$$

and we deduce that $\hat{g}_T = T(\beta_{\text{LAD}} - \beta) \to_d \hat{g}$ as required.

9.2. PROOF OF THEOREM 4.4. Start by writing the estimation error as

$$T(\beta_{\text{LAD}}^{+} - \beta) = T(\beta_{\text{LAD}} - \beta) - (1/2\hat{h}(0))\left(T^{-2}X'X\right)^{-1}\left[T^{-1}X'\Delta X\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{xv} + \hat{\Delta}_{xv}^{+}\right] .$$

Then, using Theorem 4.1 and (6) we obtain

$$\begin{aligned}
T(\beta_{\text{LAD}}^{+} - \beta) &\to_d \left[2h(0)\int_0^1 \beta_x B_x'\right]^{-1}\left[\int_0^1 B_x dB_v + \Delta_{xv}\right] \\
&\quad - (1/2h(0))\left(\int_0^1 B_x B_x'\right)^{-1}\left[\left(\int_0^1 B_x dB_x + \Delta_{xx}\right)\Omega_{xx}^{-1}\Omega_{xv} + \Delta_{xv}^{+}\right] \\
&= \left[2h(0)\int_0^1 B_x B_x'\right]^{-1}\left[\int_0^1 B_x(dB_v - \Omega_{vx}\Omega_{xx}^{-1}dB_x)\right] \\
&= \left[2h(0)\int_0^1 B_x B_x'\right]^{-1}\int_0^1 B_x dB_{v\cdot x} \\
&\equiv MN\left(0,\ (1/2h(0))^2\omega_{vv\cdot x}\left[\int_0^1 B_x B_x'\right]^{-1}\right)
\end{aligned}$$

as required.

9.3. PROOF OF THEOREM 5.1. The argument follows the general lines of Knight (1989, Theorem 2). Take the case (a) of ρ convex and define

$$Z_T(g) = \Sigma_1^T\{\rho(u_{0t} - T^{-1}x_t'g) - \rho(u_{0t})\} ,$$

so that if $\hat{g}_T$ minimizes $Z_T(g)$ we have $\hat{g}_T = T(\beta_M - \beta)$. Then, by virtue of the convexity of $Z_T(g)$ we have $\hat{g}_T \rightarrow_d g = \text{argmin } Z(g)$ where $Z(g)$ is the weak limit of $Z_T(g)$. As in the proof of Theorem 4.1 above, we need only establish finite dimensional convergence of $Z_T(g)$ to $Z(g)$.

Taylor expansion of $Z_T(g)$ around $g = 0$ gives

$$\text{(P10)} \qquad Z_T(g) = -T^{-1}\Sigma_1^T\psi(u_{0t})x_t'g + (1/2)T^{-2}\Sigma_1^T\psi'(u_{0t}^*)g'x_tx_t'g ,$$

where u_{0t}^* lies between u_{0t} and $u_{0t} - T^{-1}x_t'g$. Now $|\psi'(u_{0t}) - \psi'(u_{0t}^*)| < K|T^{-1}x_t'g|$ for some $K > 0$ and therefore

$$\text{(P11)} \qquad T^{-2}\Sigma_1^T|\psi'(u_{0t} - \psi'(u_{0t}^*)|g'x_tx_t'g \le KT^{-3/2}\Sigma_1^T(T^{-1/2}x_t'g)(T^{-1}g'x_tx_t'g) \rightarrow_p 0 ,$$

uniformly over g in compact sets. Next

$$\text{(P12)} \qquad T^{-2}\Sigma_1^T[\psi'(u_{0t}) - E\{\psi'(u_{0t})\}]g'x_tx_t'g$$

$$= T^{-1/2}g'[\Sigma_1^T T^{-1/2}[\psi'(u_{0t}) - E\{\psi'(u_{0t})\}](T^{-1/2}x_t)(T^{-1/2}x_t')]g$$

$$= O_p(T^{-1/2}) ,$$

uniformly in g because the expression in large square brackets converges to a stochastic integral with random drift, just as in (20). Finally,

$$\text{(P13)} \qquad T^{-2}\Sigma_1^T E\{\psi'(u_{0t})\}g'x_tx_t'g \rightarrow_d E\{\psi'(u_{0t})\}g'\int_0^1 B_xB_x'g ,$$

and

$$\text{(P14)} \qquad T^{-1}\Sigma_1^T\psi(u_{0t})x_t'g \rightarrow_d g'(\int_0^1 B_x dB_\psi + \Delta_{x\psi}) ,$$

since $\psi(u_{0t})$ satisfies the functional law

$$T^{-1/2}\Sigma_1^T\psi(u_{0t}) \rightarrow B_\psi \equiv BM(\Omega_{\psi\psi}) ,$$

and the conditions for the convergence to the stochastic integral with drift in (P12) in view of Assumption ML. Combining (P13) and (P14) with (P11) and (P12) we obtain the following limit for $Z_T(g)$

$$Z_T(g) \longrightarrow_d -g'\left(\int_0^1 B_x dB_\psi + \Delta_{x\psi}\right) + (1/2)E\{\psi'(u_{0t})\}g'\int_0^1 B_x B_x' g = Z(g) \ .$$

We deduce that

$$\hat{g}_T \longrightarrow_d \hat{g} = \left[E\{\psi'(u_{0t})\}\int_0^1 B_x B_x'\right]^{-1}\left[\int_0^1 B_x dB_\psi + \Delta_{x\psi}\right] ,$$

giving the required result.

In case (b) where $T^{1/2}(\beta_M - \beta) = o_p(1)$ and β_M satisfies (41) we expand the first order conditions, giving

$$0 = T^{-1}\Sigma_1^T x_t\psi(u_{0t}) - T^{-2}\Sigma_1^T\psi'(u_{0t})x_t x_t' T(\beta_M - \beta) + T^{-1}R_T ,$$

where

$$T^{-1}R_T = T^{-1}\Sigma_1^T\{\psi'(u_{0t}) - \psi'(u_{0t}^*)\}x_t x_t'(\beta_M - \beta) ,$$

and u_{0t}^* lies between u_{0t} and $u_{0t} + x_t'(\beta - \beta_M)$. Now

$$|T^{-1}R_T| \le KT^{-1}\Sigma_1^T\|x_t\|^3|\beta_M - \beta|^2 = KT^{-1}\Sigma_1^T\|T^{-1/2}x_t\|^3|T(\beta_M - \beta)||T^{1/2}(\beta_M - \beta)|$$
$$= |T(\beta_M - \beta)|o_p(1) \ .$$

Hence

$$T(\beta_M - \beta) = \left[T^{-2}\Sigma_1^T\psi'(u_{0t})x_t x_t' + o_p(1)\right]^{-1}[T^{-1}\Sigma_1^T x_t\psi(u_{0t})$$
$$\longrightarrow_d \left[E\{\psi'(u_{0t})\}\int_0^1 B_x B_x'\right]^{-1}\left[\int_0^1 B_x dB_\psi + \Delta_{x\psi}\right] ,$$

just as in the case of convex ρ.

9.4. ADDENDUM TO THEOREM 5.1: $\psi(\cdot)$ NON SMOOTH

We will consider here the case where $\psi(\ )$ is differentiable except for a countable number of points of $\mathbb{R}$. We will retain the other conditions of Assumption ML. The arguments follow the same general lines as those given in the proof of Theorem 4.1 for the LAD estimator.

Take the case (a) where ρ is convex. As in the LAD proof we need to show that $Z_T(g)$

$= \Sigma_1^T\{\rho(u_{0t} - T^{-1}x_t'g) - \rho(u_{0t})\}$ has a suitable quadratic approximation as $T \to \infty$. Since $\psi(\cdot)$ is not everywhere differentiable we cannot use (P10). Instead, we proceed by treating the ordinary function $\rho(\cdot)$ in $Z_T(g)$ as a generalized function by means of the corresponding regular sequence $\rho_m(\cdot)$ given by

$$\rho_m(u) = \int_{-\infty}^{\infty}\rho(v)S(m(v-u))me^{-v^2/m^2}dv \ .$$

(The existence of this integral poses no practical constraints on $\rho(v)$ which will, for robust estimation purposes, generally be bounded by a function that is at most $O(v^2)$ as $|v| \to \infty$). Then $Z_T(g)$ is defined by the regular sequence of processes

$$Z_{Tm}(g) = \Sigma_1^T\{\rho_m(u_{0t} - T^{-1}x_t'g) - \rho_m(u_{0t})\} \ .$$

Expanding Z_{Tm} in a Taylor series about $g = 0$ we have

(P15) $$Z_{Tm}(g) = T^{-1}\Sigma_1^T\psi_m(u_{0t})x_t'g + (1/2)T^{-2}\Sigma_1^T\psi_m'(u_{0t}^*)g'x_tx_t'g \ ,$$

where $\psi_m(\cdot) = \rho_m'(\cdot)$ is a regular sequence for $\psi(\cdot) = \rho'(\cdot)$ and $\psi_m'(\cdot)$ is a regular sequence for $\psi'(\cdot)$ where both ψ and ψ' are treated as generalized functions.

We examine the limit behavior of the two components of $Z_{Tm}(g)$ separately. First, as in Example 3.5, we get

(P16) $$T^{-1}\Sigma_1^T\psi_m(u_{0t})x_t'g \to_d (\int_0^1 dB_{\psi_m}B_x' + \Delta_{x\psi_m}')g \ ,$$

where $B_{\psi_m} = BM(\Omega_{\psi_m\psi_m})$, $\Omega_{\psi_m\psi_m} = \Sigma_{j=-\infty}^{\infty}E\{\psi_m(u_{0t})\psi_m(u_{0t+j})\}$ and $\Delta_{x\psi_m} = \Sigma_{j=0}^{\infty}E\{u_{xt}\psi_m(u_{0t+j})\}$. The limit process in (P16) as $m \to \infty$ is

(P17) $$\lim_{m\to\infty}(\int_0^1 dB_{\psi_m}B_x' + \Delta_{x\psi_m}')g \equiv (\int_0^1 dB_\psi B_x' + \Delta_{x\psi}')g \ .$$

In the second term of (P15) $\psi_m'(\cdot)$ is a regular sequence and therefore is differentiable with bounded derivative for each m. Thus

$$|\psi_m'(u_{0t}^*) - \psi_m'(u_{0t})| \le K_mT^{-1}x_t'g \ , \quad \forall m$$

for some $K_m > 0$ and

$$T^{-2}\Sigma_1^T\{\psi_m'(u_{0t}) - \psi_m'(u_{0t}^*)\}g'x_tx_t'g \le K_mT^{-3}\Sigma_1^T\|x_t\|^3\|g\|^3 \longrightarrow_p 0\ , \quad \forall m$$

uniformly over g in compact sets. Just as in Example 3.3, we now obtain the limit

(P18) $$T^{-2}\Sigma_1^T\psi_m'(u_{0t})g'x_tx_t'g \longrightarrow_d E\{\psi_m'(u_{0t})\}\int_0^1(g'B_x)^2\ .$$

By definition of the regular sequence $\psi_m'(\cdot)$ we have the limit

$$\lim_{m\to\infty}\int_{-\infty}^{\infty}\psi_m'(u)h(u)du = \int_{-\infty}^{\infty}\psi'(u)h(u)du = -\int_{-\infty}^{\infty}\psi(u)h'(u)du\ ,$$

which exist as ordinary Riemann integrals, *i.e.* we have

(P19) $$E\{\psi_m'(u_{0t})\} \to E\{\psi'(u)\}\ .$$

Combining (P16) and (P18) we have

$$Z_{Tm}(g) \longrightarrow_d -\left(\int_0^1 dB_{\psi_m}B_x' + \Delta_{x\psi_m}'\right)g + (1/2)E\{\psi_m'(u_{0t})\}g'\int_0^1 B_xB_x'g = Z_m(g)\ , \quad \text{say}$$

and in view of (P17) and (P19)

$$\lim_{m\to\infty} Z_m(g) = -\left(\int_0^1 dB_\psi B_x' + \Delta_{x\psi}'\right)g + (1/2)E\{\psi'(u_{0t})\}g'\int_0^1 B_xB_x'g = Z(g)\ .$$

This establishes the weak convergence of $Z_T(g) \longrightarrow_d Z(g)$ as generalized processes uniformly in g over compact sets. The argument then follows as in the proof of Theorem 4.1 and we get

$$\hat{g}_T = \text{argmin}\ Z_T(g) \longrightarrow_d \hat{g} = \text{argmin}\ Z(g)\ ,$$

and thus the conclusion of Theorem 5.1 continues to apply in this case where $\psi(\ )$ is not everywhere differentiable.

An analogous proof for the ρ not necessarily convex (i.e. case (b) of Theorem 5.1) is constructed following the lines of the second part of the proof of Theorem 5.1 and using generalized functions of random variables in the same way as the earlier part of this Addendum.

9.5. PROOF OF THEOREM 5.3. The error of estimation is

$$T(\beta_M^+ - \beta) = T(\beta_M - \beta) - \{1/T^{-1}\Sigma_1^T\psi'(\hat{u}_{0t})\}\left(T^{-2}X'X\right)^{-1}[T^{-1}X'\Delta X\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{x\psi} + \hat{\Delta}_{x\psi}^+] .$$

Note that

$$|T^{-1}\Sigma_1^T\{\psi'(\hat{u}_{0t}) - \psi'(u_{0t})\}| \le KT^{-1}\Sigma_1^T\|x_t\|\|\beta_M - \beta\| \to_p 0$$

and

$$T^{-1}\Sigma_1^T\psi'(u_{0t}) \to_{\text{a.s.}} E\{\psi'(u_{0t})\}$$

so that

$$T^{-1}\Sigma_1^T\psi'(\hat{u}_{0t}) \to_p E\{\psi'(u_{0t})\} .$$

In a similar way we can replace $\psi(u_{0t})$ by $\hat{\psi} = \psi(\hat{u}_{0t})$ in the sample covariances that enter into the formulae for $\hat{\Omega}_{x\psi}$ and $\hat{\Delta}_{x\psi}$ and retain the consistency of these estimators. Then, using Theorem 5.1 we get

$$\begin{aligned} T(\beta_M^+ - \beta) &\to_d \left[E\{\psi'(u_{0t})\}\int_0^1 B_xB_x'\right]^{-1}[\int_0^1 B_x dB_\psi + \Delta_{x\psi}] \\ &\quad - (1/E\{\psi'(u_{0t})\})\left(\int_0^1 B_xB_x'\right)^{-1}[(\int_0^1 B_x dB_x + \Delta_{xx})\Omega_{xx}^{-1}\Omega_{x\psi} + \Delta_{x\psi}^+] \\ &= \left[E\{\psi'(u_{0t})\}\int_0^1 B_xB_x'\right]^{-1}[\int_0^1 B_x dB_\psi - \int_0^1 B_x dB_x'\Omega_{xx}^{-1}\Omega_{x\psi}] \\ &= \left[E\{\psi'(u_{0t})\}\int_0^1 B_xB_x'\right]^{-1}[\int_0^1 B_x dB_{\psi\cdot x}] \\ &\equiv MN\left(0,\ \omega_{\psi\psi\cdot x}[E\{\psi'(u_{0t})\}]^{-2}\left[\int_0^1 B_xB_x'\right]^{-1}\right) \end{aligned}$$

as given in (46).

9.6. PROOF OF THEOREM 6.1. We first consider β_{LAD} and our line of approach is the same as in the proof of Theorem 4.1. However, instead of (P1) we take

$$Z_T(g) = \Sigma_1^T\{|u_{0t} - T^{-a}x_t'g| - |u_{0t}|\} , \quad \text{and} \quad \hat{g}_T = \text{argmin } Z_T(g)$$

with $a = 1/2 + 1/\alpha$. As before we treat $Z_T(g)$ as a generalized process, defined in terms of the regular sequence

$$Z_{Tm}(g) = \Sigma_1^T\{f_m(u_{0t} - T^{-a}x_t'g) - f_m(u_{0t})\} ,$$

and use the Taylor expansion

$$\text{(P20)} \qquad Z_{Tm}(g) = -T^{-a}\Sigma_1^T f_m^{(1)}(u_{0t})x_t'g + (1/2)T^{-2a}\Sigma_1^T f_m^{(2)}(u_{0t}^*)g'x_tx_t'g ,$$

where u_{0t}^* lies between u_{0t} and $u_{0t} - T^{-a}x_t'g$. Here the sequence $f_m^{(1)}(\cdot)$ is equivalent to $\mathrm{sgn}_m(\cdot)$ and $f_m^{(2)}(\cdot)$ to $\delta_m(\cdot)$, as defined earlier.

First consider the second term of (P20). We use the "BN" decomposition

$$\text{(P21)} \qquad u_t = D(L)\varepsilon_t = D(1)\varepsilon_t + \tilde{\varepsilon}_{t-1} - \tilde{\varepsilon}_t$$

(see Phillips and Solo, 1992) where $\tilde{\varepsilon}_t = \tilde{D}(L)\varepsilon_t$ and $\tilde{D}(L) = \Sigma_0^\infty \tilde{D}_j L^j$ with $\tilde{D}_j = \Sigma_{j+1}^\infty D_k$. Now, in view of (53)

$$\Sigma_0^\infty \|\tilde{D}_j\|^\delta \le \Sigma_0^\infty k\|D_k\|^\delta < \infty ,$$

and thus $\tilde{\varepsilon}_t = \Sigma_0^\infty \tilde{D}_j\varepsilon_{t-j}$ converges almost surely and $\tilde{\varepsilon}_t \in \mathcal{D}(\alpha)$. Now set $P_t = \Sigma_1^t \varepsilon_j$ and $W_t = \Sigma_1^t u_j$. We have $W_t = D(1)P_t + \tilde{\varepsilon}_0 - \tilde{\varepsilon}_t$ and then

$$\text{(P22)} \qquad \begin{aligned} T^{-2a}\Sigma_1^T W_tW_t' &= D(1)T^{-2a}\Sigma_1^T P_tP_t'D(1)' + D(1)T^{-2a}\Sigma_1^T P_t(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_t) \\ &+ T^{-2a}\Sigma_1^T(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_t)P_t'D(1)' + T^{-2a}\Sigma_1^T(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_t)(\tilde{\varepsilon}_0 - \tilde{\varepsilon}_t)' . \end{aligned}$$

Note that $\tilde{\varepsilon}_t \in \mathcal{D}(\alpha/2)$, so that $T^{-2/\alpha}\Sigma_1^T\tilde{\varepsilon}_t\tilde{\varepsilon}_t' = O_p(1)$, $T^{-1/\alpha}\Sigma_1^T\tilde{\varepsilon}_t = O_p(1)$ and therefore the final term of (P22) is $o_p(1)$. Also $T^{-1-1/\alpha}\Sigma_1^T P_t = O_p(1)$ and $T^{-2a}\Sigma_1^T P_t\tilde{\varepsilon}_t = o_p(1)$ (the latter can be shown by using a further "BN" decomposition for $\tilde{\varepsilon}_t$). Hence, (P22) is dominated by the first term. However, $T^{-1/\alpha}P_{[Tr]} \to_d U_\alpha(r)$ by (54) and by virtue of the continuous mapping theorem we obtain

$$\text{(P23)} \qquad T^{-2a}\Sigma_1^T W_tW_t' = D(1)T^{-2a}\Sigma_1^T P_tP_t'D(1)' + o_p(1) \to_d D(1)\left(\int_0^1 U_\alpha(r)U_\alpha(r)'dr\right)D(1)' .$$

We deduce that

$$T^{-2a}\Sigma_1^T x_tx_t' \to_d \int_0^1 U_{x\alpha}(r)U_{x\alpha}(r)'dr ,$$

where $U_{x\alpha}(r) = D_x'U_\alpha(r)$ and $D' = [D_0 \; D_x]$ is partitioned conformably with $u_t = (u_{0t}, u_{xt}')'$.

In the same way as in the proof of Theorem 4.1 we can now show that

$$T^{-2a}\Sigma_1^T\{\delta_m(u_{0t}) - \delta_m(u_{0t}^*)\}g'x_tx_t'g \rightarrow_p 0$$

uniformly over g in compact sets, and

$$T^{-2a}\Sigma_1^T\delta_m(u_{0t})g'x_tx_t'g \rightarrow_d E\{\delta_m(u_{0t})\}\int_0^1(g'U_{x\alpha})^2 .$$

Next consider the first term of (P20). Noting that $f_m^{(1)}(u_{0t}) = \mathrm{sgn}_m(u_{0t})$, which is (a sequence of) strictly stationary bounded functions of u_{0t} we have the martingale difference decomposition (see Hall and Heyde, 1980)

(P24) $\qquad \mathrm{sgn}_m(u_{0t}) = Y_{mt} + Q_{mt} - Q_{mt-1}\ , \quad \forall m$

where the Y_{mt} are stationary square integrable ergodic martingale differences (with respect to the filtration generated by $\{u_{0j} : j \le t\}$) and the Q_{mt} are square integrable stationary processes $\forall m$. As in the proof of Lemma 2 of Knight (1991) we can use (P24) and the BN decomposition for u_{xt} that follows from (P21) to establish the weak convergence

(P25) $\qquad T^{-a}\Sigma_1^T\mathrm{sgn}_m(u_{0t})x_t' = \Sigma_1^T(T^{-1/2}\mathrm{sgn}_m(u_{0t}))(T^{-1/\alpha}x_t') \rightarrow_d \int_0^1 dB_mU_{x\alpha}^{-\prime}$

where $U_{x\alpha}^-$ signifies the left limit of $U_{x\alpha}(\cdot)$. In the limiting stochastic integral (P25) the Brownian motion $B_m \equiv BM(\Omega_m)$ is stochastically independent of the stable process $U_{x\alpha}$. There is also no drift or bias term in the limit (P25), unlike the finite variance case. The independence is a consequence of the different rates of convergence to B_m and $U_{x\alpha}$ and follows from a result originally shown by Resnick and Greenwood (1979).

Combining these results we obtain

$$Z_{Tm}(g) \rightarrow_d -\left(\int_0^1 dB_mU_{x\alpha}^{-\prime}\right)g + E\{\delta_m(u_{0t})\}g'\left(\int_0^1 U_{x\alpha}U_{x\alpha}'\right)g = Z_m(g)\ , \quad say\ \forall m$$

and, as in the proof of Theorem 4.1, the convergence holds uniformly over g in compact sets. Again, since $\lim_{m\to\infty} B_m(r) = B_v \equiv BM(\Omega_{vv})$ and $\lim_{m\to\infty} E\{\delta_m(u_{0t})\} = E\{\delta(u_{0t})\} = \mathrm{pdf}(0)$ we have

$$\lim_{m\to\infty} Z_m(g) = -(\int_0^1 dB_v U_{x\alpha}^{-}{}')g + \text{pdf}(0)g'(\int_0^1 U_{x\alpha}U_{x\alpha}')g = Z(g)\ , \quad \text{say}$$

which is an ordinary random variable. The remainder of the argument now follows exactly as in Theorem 4.1 and the result for $T^a(\beta_{\text{LAD}} - \beta)$ is established.

Next consider the estimator β_{LAD}^+. We have

$$\text{(P26)}\quad T^a(\beta_{\text{LAD}}^+ - \beta) = T^a(\beta_{\text{LAD}} - \beta) - \left[2\hat{h}(0)T^{-2a}\Sigma_1^T x_t x_t'\right]^{-1}[T^{-2a}(\Sigma_1^T x_t u_{xt}')\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{xv} + T^{1-2a}\hat{\Delta}_{xv}^{+}]\,.$$

We need to show that the second term on the right of (P26) is $o_p(1)$. Since β_{LAD} is consistent (from the first part of the proof) and LAD residuals are used in the construction of $\hat{h}(0)$ and the long-run variance matrix estimates that appear in (P26), we may proceed as if these estimates were constructed using the true errors u_{0t}. Then $\hat{h}(0) \to_p h(0)$, and following the same line of argument as that given in Section 2.3 of Phillips (1991) we find that $T^{1-2a}\hat{\Omega}_{xx} = O_p(1)$, $T^{1-2a}\hat{\Delta}_{xx} = O_p(1)$, $T^{1-2a}\hat{\Omega}_{xv} = o_p(1)$ and $T^{1-2a}\Delta_{xv} = o_p(1)$. Then

$$\begin{aligned}
&T^{-2a}(\Sigma_1^T x_t u_{xt}')\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{xv} + T^{1-2a}\Delta_{xv}^{+} \\
&= \left(T^{1-2a}\Sigma_1^T x_t u_{xt}'\right)\left(T^{1-2a}\hat{\Omega}_{xx}\right)^{-1}(T^{1-2a}\hat{\Omega}_{xv}) \\
&\quad + T^{-2/\alpha}\Delta_{xv} - (T^{1-2a}\hat{\Delta}_{xx})\left(T^{1-2a}\hat{\Omega}_{xx}\right)^{-1}(T^{-2/\alpha}\hat{\Omega}_{xv}) \\
&= (O_p(1))(O_p(1))^{-1}o_p(1) + o_p(1) - O_p(1)(O_p(1))^{-1}o_p(1) = o_p(1)\ .
\end{aligned}$$

We deduce from (P26) that $T^a(\beta_{\text{LAD}}^+ - \beta) = T^a(\beta_{\text{LAD}} - \beta) + o_p(1)$ and the stated result follows.

A similar argument gives the limit distribution of $T^a(\beta_M - \beta)$ and shows the asymptotic equivalence of β_M and β_M^+.

10. References

Ahn, S. K. and G. C. Reinsel (1990). "Estimation for partially nonstationary multivariate autoregressive models," *Journal of the American Statistical Association*, 85, 813-823.

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

Erdelyi, A. (1953). *Higher Transcendental Functions*, Vol. 1. New York: McGraw Hill.

Gel'fand, I. M. and N. Ya. Vilenkin (1964). *Generalized Functions*, Vol. 4. New York: Academic Press.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Application*. New York: Academic Press.

Hansen, B. E. (1992). "Convergence to stochastic integrals for dependent heterogeneous processes," *Econometric Theory*, 8, 489-500.

Huber, P. J. (1964). "Robust estimation of a location parameter," *Annals of Mathematical Statistics*, 35, 73-101.

Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley.

Hunter, J. (1992). "Tests of cointegrating exogeneity for PPP and uncovered interest rate parity in the United Kingdom," *Journal of Policy Modeling*, 14, 453-464.

Jeganathan, P. (1989). "Some aspects of asymptotic theory with applications to time series models," mimeographed, University of Michigan.

Johansen, S. (1988). "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. and K. Juselius (1992). "Some structural hypotheses in a multivariate cointegration analysis of the purchasing power parity and the uncovered interest parity for the UK," *Journal of Econometrics*.

Knight, K. (1989). "Limit theory for autoregressive parameters in an infinite variance random walk," *Canadian Journal of Statistics*, 261-278.

Knight, K. (1991). "Limit theory for *M*-estimates in an integrated infinite variance process," *Econometric Theory*, 7, 200-212.

Lighthill, M. J. (1958). *Introduction to Fourier Analysis and Generalised Functions*. Cambridge: Cambridge University Press.

Park, J. Y. and P. C. B. Phillips (1988). "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4, 468-497.

Park, J. Y. and P. C. B. Phillips (1989). "Statistical inference in regressions with integrated processes: Part 2," *Econometric Theory*, 5, 95-131.

Paulauskas, V. J. (1976). "Some remarks on multivariate stable distributions," *Journal of Multivariate Analysis*, 6, 356-368.

Phillips, P. C. B. (1991b). "Optimal inference in cointegrated systems," *Econometrica*, 59, 283-306.

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

Phillips, P. C. B. (1993). "Fully modified least squares and vector autoregression," Cowles Foundation Discussion Paper No. 1055.

Phillips, P. C. B. (1988). "Weak convergence of sample covariance matrices to stochastic integrals via martingale approximations," *Econometric Theory*, 4, 528-533.

Phillips, P. C. B. and S. N. Durlauf (1986). "Multiple time series with integrated variables," *Review of Economic Studies*, 53, 473-496.

Phillips, P. C. B. and B. E. Hansen (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies*, 57, 99-125.

Phillips, P. C. B. and V. Solo (1992). "Asymptotics for linear processes," *Annals of Statistics*, 20, 971-1001.

Pollard, D. (1990). "Empirical processes: Theory and applications," *NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 2*. Hayward, CA: Institute of Mathematical Statistics.

Pollard, D. (1991). "Asymptotics for least absolute deviation regression estiamtors," *Econometric Theory*, 7, 186-199.

Resnick, S. I. (1986). "Point processes, regular variation and weak convergence," *Advances in Applied Probability*, 18, 66-138.

Resnick, S. and P. Greenwood (1979). "A bivariate stable characterization and domains of attraction," *Journal of Multivariate Analysis*, 9, 206-221.

Fig 1(i): Australian dollar
spot exchange rate returns

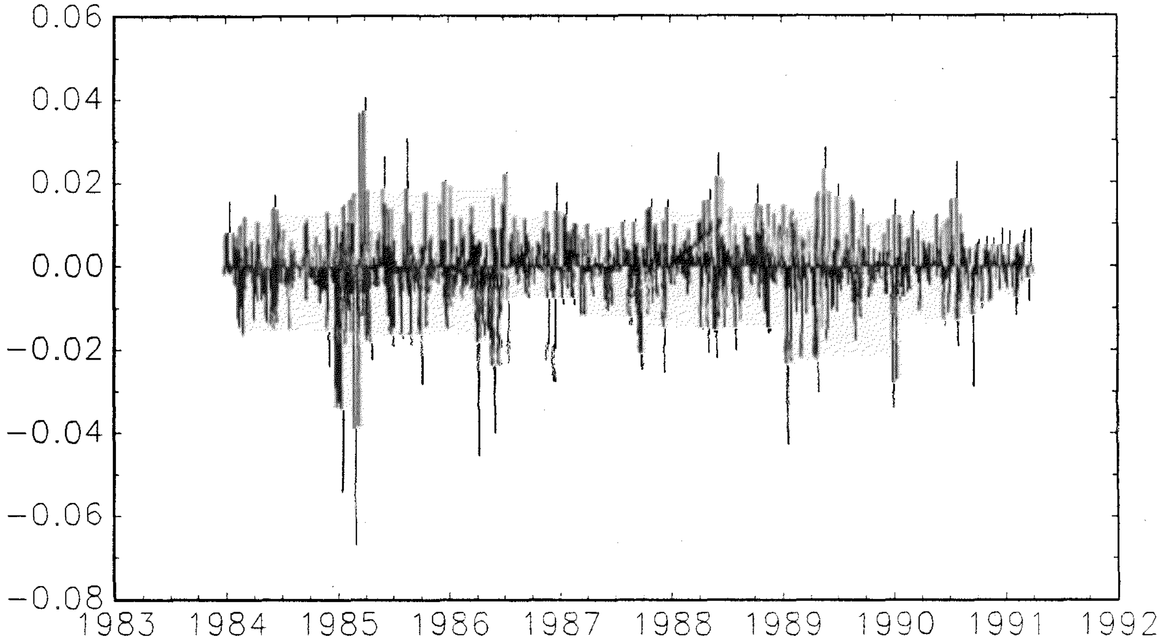

Fig 1(ii): Density Estimates
spot rate returns

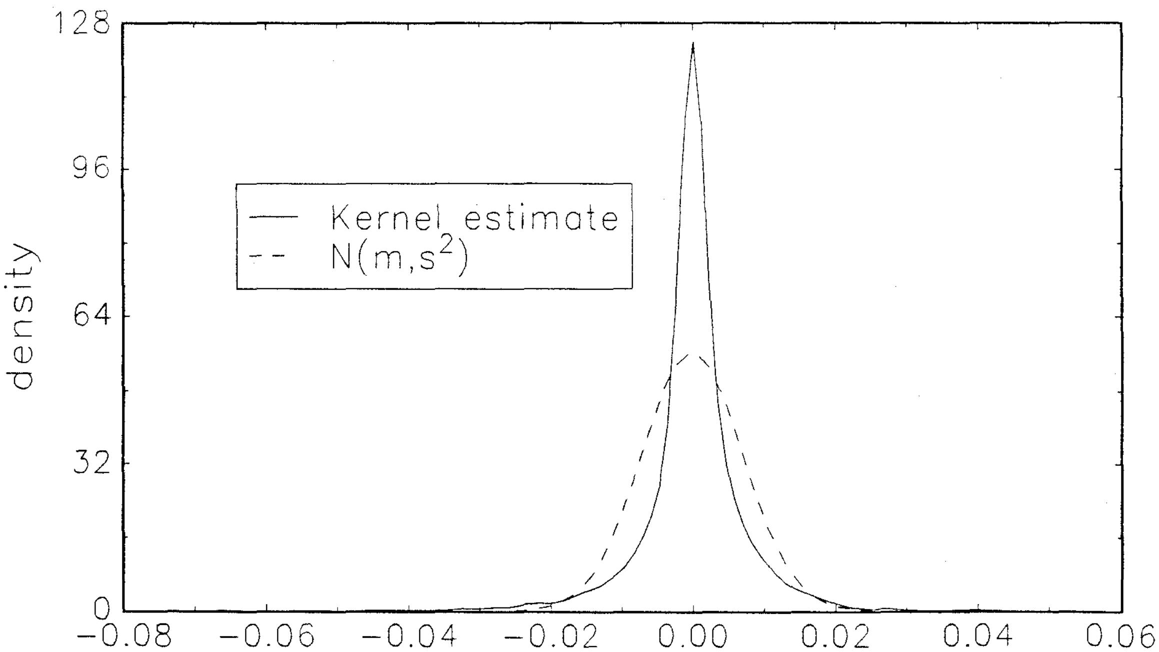

Figure 2: Normal errors

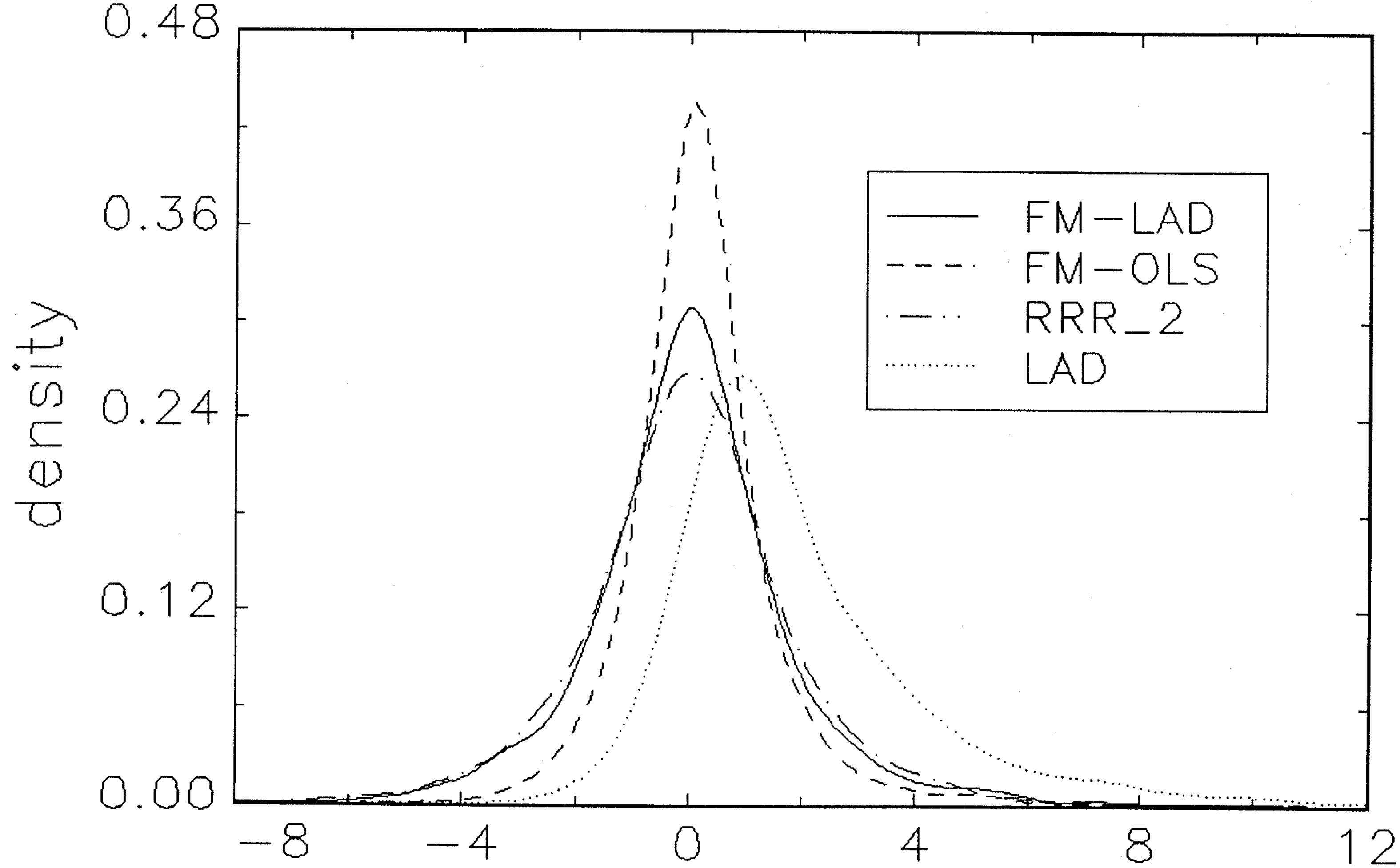

Figure 3: t_4 errors

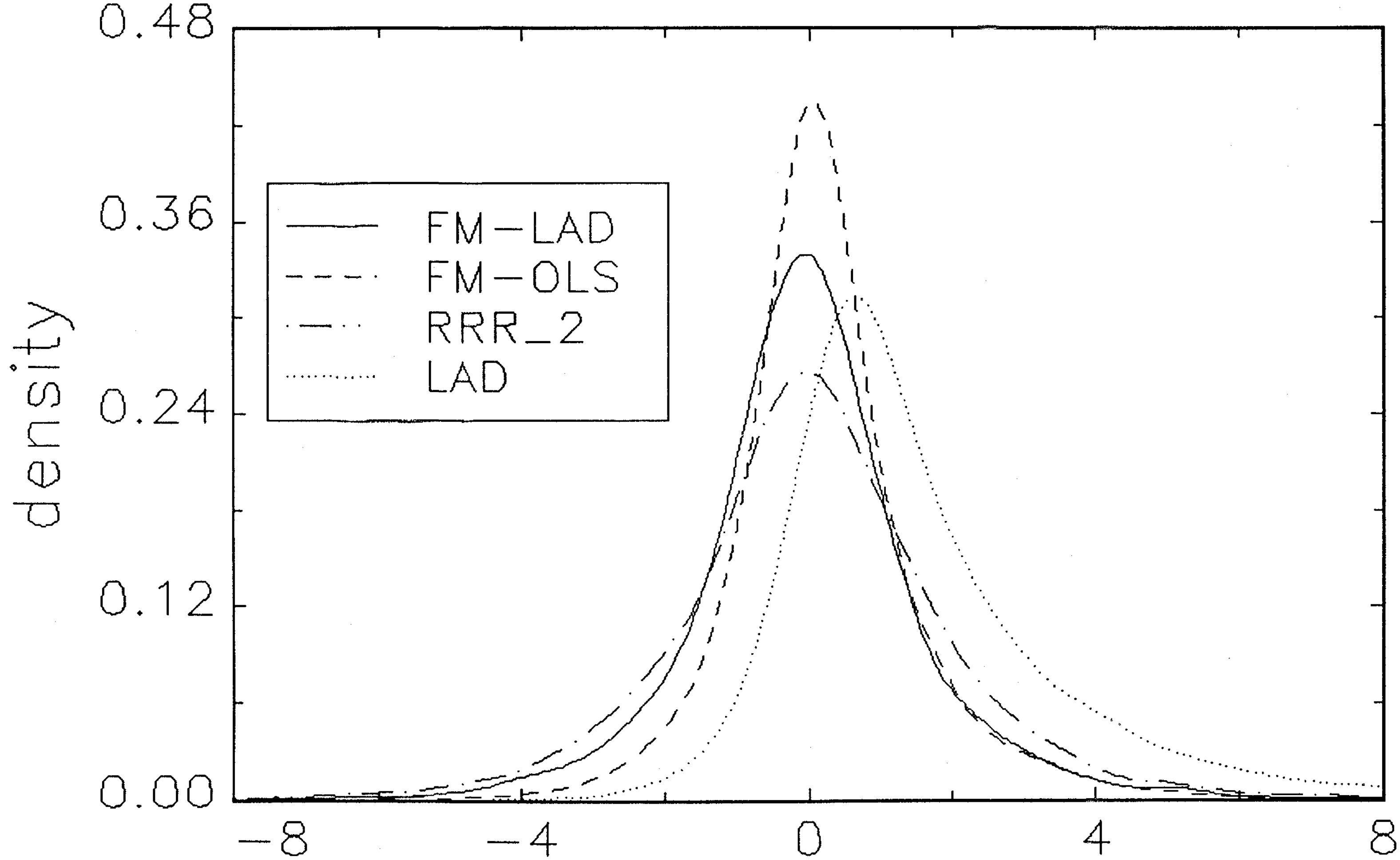

Figure 4: t_2 errors

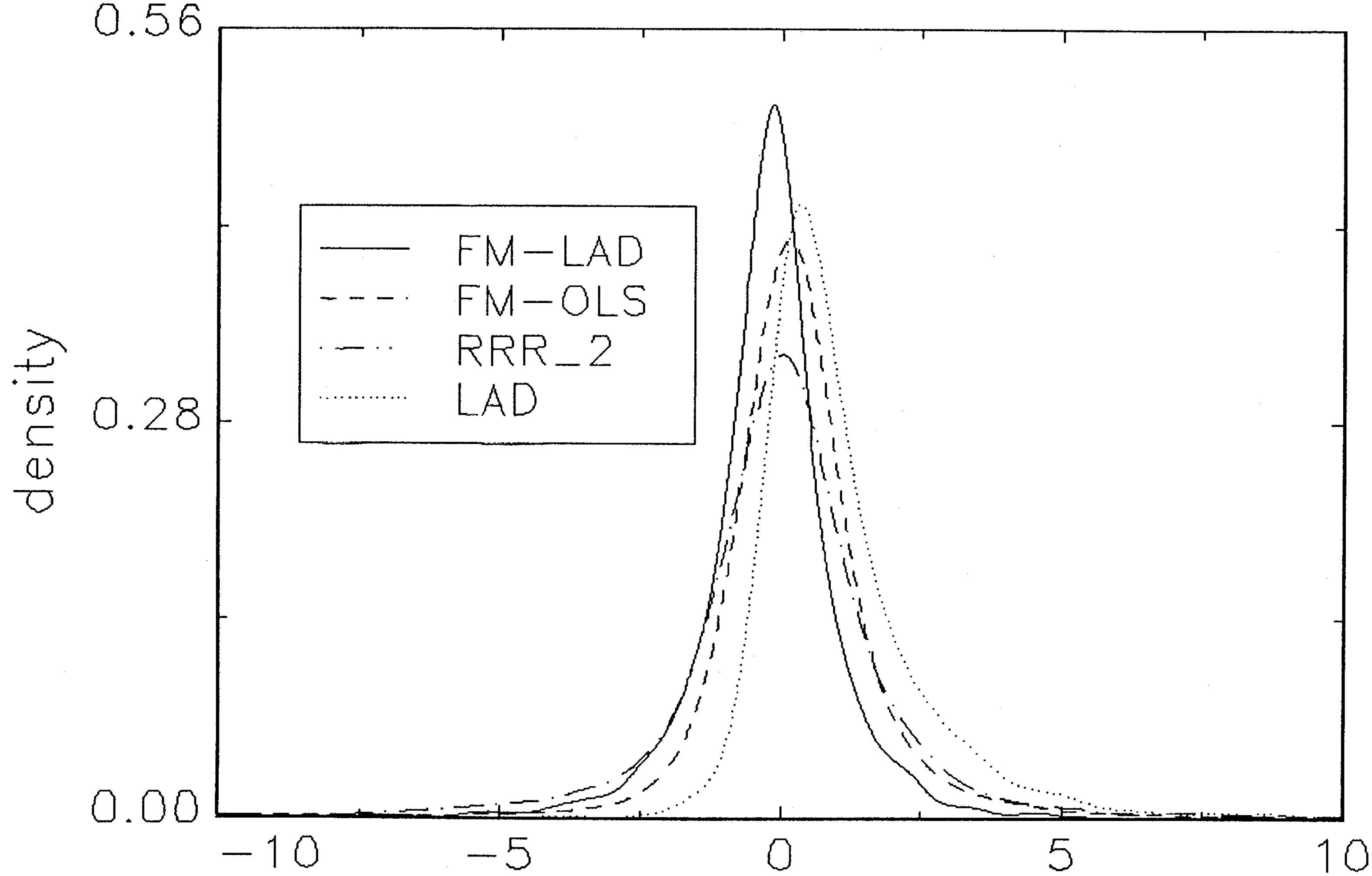

Figure 5: Cauchy errors

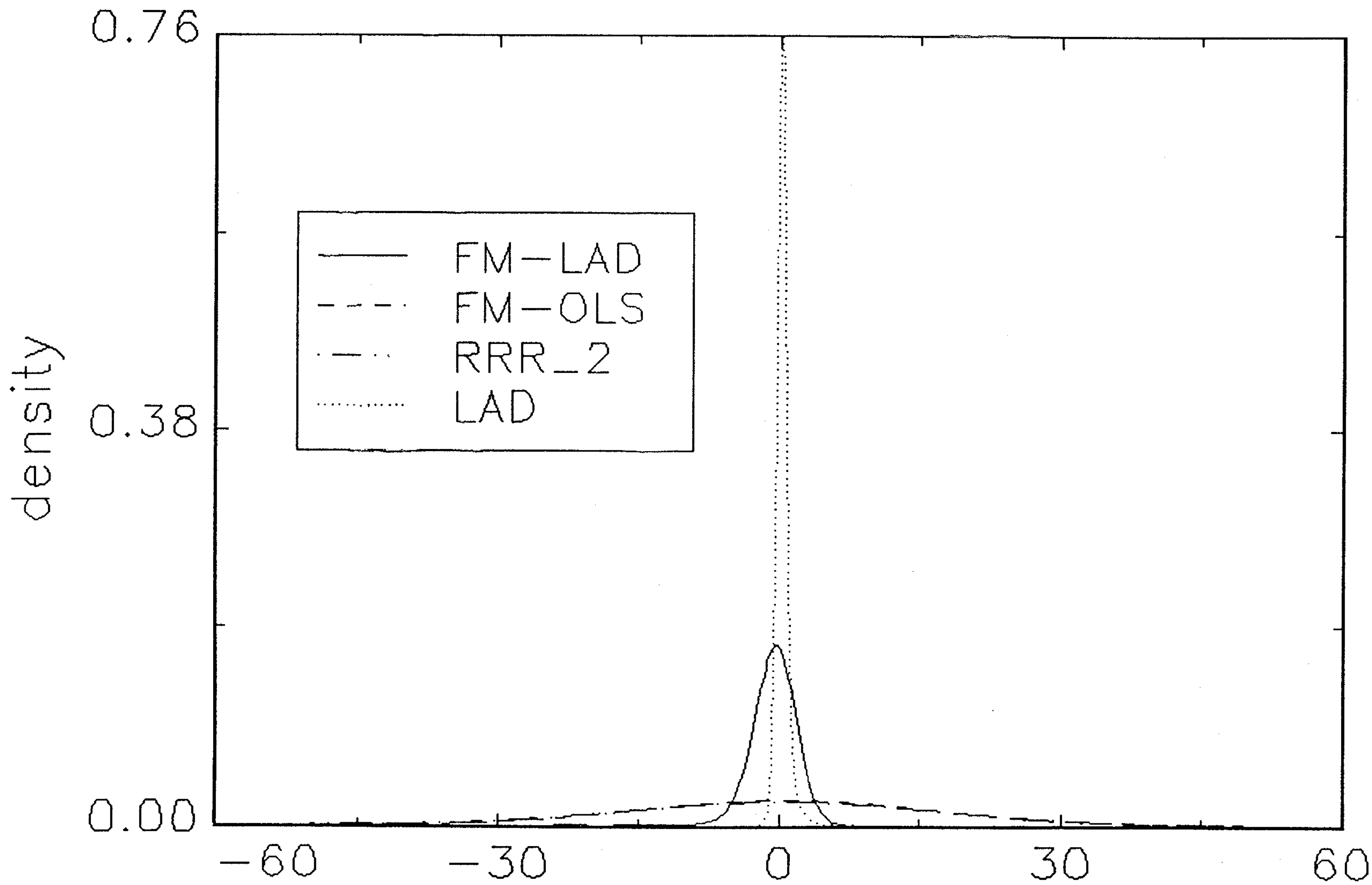

Fig 6: Scatter plot & regressions for equation (55)