Econometric Theory, **26**, 2010, 953–962.
doi:10.1017/S0266466609990703

NOTES AND PROBLEMS

LAD ASYMPTOTICS UNDER CONDITIONAL HETEROSKEDASTICITY WITH POSSIBLY INFINITE ERROR DENSITIES

JIN SEO CHO
Yonsei University

CHIROK HAN
Korea University

PETER C.B. PHILLIPS
Yale University, University of Auckland
University of Southampton
and
Singapore Management University

Least absolute deviations (LAD) estimation of linear time series models is considered under conditional heteroskedasticity and serial correlation. The limit theory of the LAD estimator is obtained without assuming the finite density condition for the errors that is required in standard LAD asymptotics. The results are particularly useful in application of LAD estimation to financial time series data.

1. INTRODUCTION

This note derives asymptotics for the least absolute deviations (LAD) estimator in a linear time series model under conditional heteroskedasticity and serial correlation. Our methods follow the approach pioneered by Knight (1998, 1999), who derived LAD asymptotics under nonstandard conditions that allow for possibly infinite or zero error density function conditions. Nonstandard conditions such as these, particularly an infinite error density at the origin, are likely to arise in many

We thank Paolo Paruolo and three referees for helpful comments on the original version. The paper is motivated by technical considerations that arose in revising Han, Cho, and Phillips (2009, manuscript) for the *Journal of Business & Economics Statistics*, and we are grateful to the *JBES* review for raising them. Han acknowledges research support from a Korea University Special Faculty Research Fund. Phillips acknowledges support from a Kelly Fellowship and the NSF under grant SES 06-47086. Address correspondence to Chirok Han, Department of Economics, Korea University, Seoul, Korea; e-mail: chirokhan@korea.ac.kr.

© Cambridge University Press 2010 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press

empirical applications to financial data, which are well known to be characterized with high kurtosis. In recent related work, the authors (Han, Cho, and Phillips, 2009) developed a test for infinite density in stock returns and found in empirical applications that a significant number of leading companies in U.S. industries have asset returns with infinite density at the median. The present asymptotic results assist in validating such applications for financial data. In particular, the current paper considers time series models with predetermined variables as regressors, conditionally heterogeneous errors, and possibly infinite error densities, thereby expanding the range of potential empirical applications of LAD estimation theory.

Previous work in the literature has developed asymptotic theory for LAD estimation in various time series settings. Phillips (1991) developed LAD limit theory for stationary data under the assumptions of a finite error density and exogenous regressors. Koenker and Zhao (1996) considered the influence of conditional heteroskedasticity on the LAD estimator, also under finite error density conditions. Knight (1999) established LAD asymptotics for a linear model with exogenous regressors and heteroskedastic errors under possibly infinite error density conditions, and Rogers (2001) generalized the framework in Knight (1998) to stationary and integrated time series contexts. These authors do not consider models with conditional heteroskedasticity, weakly exogenous regressors, and possibly discontinuous density at the median.

The present note extends the existing literature by establishing an asymptotic theory for LAD estimation with time series data that exhibits conditional heteroskedasticity and whose errors have possibly infinite density. The model, regularity conditions, and main result are given in Section 2, and Section 3 provides proofs.

2. MAIN RESULTS

We consider the following linear median regression model:

$$y_t = x_t'\beta + \varepsilon_t, \qquad \mathrm{E}[x_t \operatorname{sgn}(\varepsilon_t)] = 0 \qquad \text{iff } \beta = \beta_0,$$

where the regressor vector $x_t \in \mathbb{R}^k$ may contain weakly exogenous variables and $\operatorname{sgn}(x) := 1\{x > 0\} - 1\{x < 0\}$. The error term ε_t is conditionally heteroskedastic in the sense that $\varepsilon_t = \sigma_t e_t$, where σ_t is adapted to $\mathcal{F}_t$, the sigma field generated by $\{x_t, \varepsilon_{t-1}, x_{t-1}, \varepsilon_{t-2}, \ldots\}$, and the primitive shock e_t is assumed to have a distribution function $F^e(z) := P(e_t \le z|\mathcal{F}_t)$ for all t and for all z. We maintain the following assumptions throughout the paper.

Assumption A.

(i) $\sigma_t \ge \sigma_* > 0$;
(ii) (x_t, σ_t) is stationary and ergodic;
(iii) $\mathrm{E}\sigma_t^2 < \infty$; and
(iv) $\mathrm{E}\|x_t\|^4 < \infty$.

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press

The asymptotic behavior of the LAD estimator is well known to depend on the distribution of e_t near zero. Specifically, as Knight (1998, 1999) shows, if we let $\psi_n^e(x) := \sqrt{n}[F^e(a_n^{-1}x) - F^e(0)]$, where a_n is selected in a way that ensures that $\psi_n^e(\cdot)$ converges to a nondegenerate function, then $\psi_n^e(x)$ turns out to be the critical component in determining LAD asymptotics, as detailed subsequently. The sequence a_n is determined differently depending on whether or not $f^e(0)$ is finite. For example, if $f^e(0)$ is finite, so that $F^e(\cdot)$ has a finite slope at zero, then we can let a_n be $\sqrt{n}$, implying that $\psi_n^e(x) = f^e(n^{-1/2}\tilde{x})x$ for some $\tilde{x}$ between x and zero by first-order Taylor expansion. As n tends to ∞, $f^e(n^{-1/2}\tilde{x})x$ converges to $f^e(0)x$, which is nondegenerate. As another example, for some $\alpha \leq 1$, suppose $F^e(x) = F^e(0) + \lambda \operatorname{sgn}(x)|x|^\alpha$ in a neighborhood of zero; then $\psi_n^e(x) = n^{1/2}\lambda \operatorname{sgn}(x) a_n^{-\alpha}|x|^\alpha$, so that $\psi_n^e(x) = \lambda \operatorname{sgn}(x)|x|^\alpha$ when $a_n = n^{1/(2\alpha)}$. In this case, note that for $\alpha < 1$, the density $f^e(0)$ is not finite. Also, $\psi_n^e(x)$ increases with respect to $|x|$, and $\psi_n^e(x) = 0$ if and only if $x = 0$.

It is useful in the development of the asymptotics to make some assumptions regarding the functional shape and properties of $\psi_n^e(\cdot)$ and also some general regularity conditions to facilitate central limit theory. We maintain the following assumptions.

Assumption B. For a function $h(\cdot)$, $|F^e(x) - F^e(0)| \leq h(x)$ for all x in an open interval V containing zero such that $h(x)$ increases with respect to $|x|$, and for some finite C_0 and n_0, $n^{1/2}h(a_n^{-1}x) \leq C_0(1 + |x|)$ for all $x \in \mathbb{R}$ provided that $n > n_0$.

In general, Assumption B is satisfied by a wide class of density functions. If the density of e_t is finite and continuous at zero, for example, then we can let $a_n = \sqrt{n}$, and $F^e(x) - F^e(0) = f^e(\tilde{x})x$ for some $\tilde{x}$ between x and 0. If we further let f^* be the maximum density in the neighborhood, then $|F^e(x) - F^e(0)| = f^e(\tilde{x})|x| \leq f^*|x|$, so that Assumption B follows by letting $h(x) = f^*|x|$, because $n^{1/2}h(a_n^{-1}x) = f^*|x| \leq f^*(1 + |x|)$. Assumption B allows for infinite densities also. For example, if a density around zero can be approximated by a power density, then it also satisfies Assumption B. More specifically, suppose that the density of e_t is bounded by two power functions with the same exponent, that is, for some $\alpha \leq 1$, $\frac{1}{2}\lambda\alpha|x|^{\alpha-1} \leq f^e(x) \leq 2\lambda\alpha|x|^{\alpha-1}$ on an open interval containing zero, then we have $a_n = n^{1/(2\alpha)}$, and $|F^e(x) - F^e(0)| \leq \lambda|x|^\alpha = h(x)$ on the same interval. We note that $h(\cdot)$ satisfies the last part of Assumption B because $n^{1/2}h(a_n^{-1}x) = \lambda|x|^\alpha \leq \lambda(1 + |x|)$ for all $x \in \mathbb{R}$, where the last inequality holds because $\alpha \leq 1$.

Next we assume that $\psi_n^e(\cdot)$ converges to some $\psi^e(\cdot)$ in a proper mode as follows.

Assumption C. For some $\psi^e(\cdot)$, there is a symmetric and nonnegative sequence $\delta_n^*(\cdot)$ such that $\delta_n^*(s)$ is increasing as $|s|$ increases, $|\psi_n^e(\cdot) - \psi^e(\cdot)| \leq \delta_n^*(\cdot)$,

$$\limsup_{n\to\infty} \mathrm{E}\left[\|x_t\|\delta_n^*(\|x_t\|)\right] < \infty, \tag{1}$$

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press

and $\delta_n^*(\cdot)$ converges uniformly to zero on every compact neighborhood of zero.

Assumption C is pivotal in proving Lemma 3 given in the following section, which enables us to relate the LAD asymptotics to the limit function $\psi^e(\cdot)$ instead of the sample function $\psi_n^e(\cdot)$. It also slightly differs from conditions in the previous literature. First, primitive conditions are provided in Assumption C for the convergence of $\psi_n^e(\cdot)$ to $\psi^e(\cdot)$ such that $\Psi^e(x) := \int_0^x \psi^e(s)ds$ is convex with respect to x, as proved in Lemma 4 in Section 3, whereas Knight (1999) assumes this directly. Second, we allow for x_t to be stochastic, differently from Knight (1999).

Many distribution functions satisfy Assumption C. For example, uniformly finite densities on a neighborhood of zero satisfy the conditions. For this kind of density, we have $a_n = n^{1/2}$, and $\psi_n^e(x) = f^e(n^{-1/2}\tilde{x})x = f^e(0)x + [f^e(n^{-1/2}\tilde{x}) - f^e(0)]x = \psi^e(x) + \delta_n(x)$ for some $\tilde{x}$ between x and zero. Thus, if we let $\delta_n^*(x) = \sup_{|\tilde{x}|\le|x|} |f^e(n^{-1/2}\tilde{x}) - f^e(0)| \cdot |x|$ and $\mathrm{E}\|x_t\|^2 < \infty$, then $\delta_n^*(\cdot)$ trivially satisfies all the conditions in Assumption C. Power densities and distributions for which $F^e(x) - F^e(0) = \lambda x \ln(|x|^{-1})$ in a neighborhood of zero (Knight, 1998, p. 761) are other examples satisfying Assumption C.

As the final regularity condition, it is convenient to impose the following high level central limit theorem.

Assumption D. For some positive definite A, $n^{-1/2}\sum_{t=1}^n x_t \operatorname{sgn}(\varepsilon_t) \Rightarrow \zeta \sim N(0, A)$.

If the median of ε_t conditional on $\mathcal{F}_t$ is zero so that $x_t \operatorname{sgn}(\varepsilon_t)$ is a martingale difference, then Assumption D follows by the Lindeberg condition, which is implied by the negligibility condition $\max_{1\le t\le n}\left(n^{-1/2}|x_t|\right) \to_p 0$ under the given moment conditions and $A = \operatorname{plim} n^{-1}\sum_{t=1}^n x_t x_t'$ (e.g., Phillips and Solo, 1992, Thm. 2.5; McLeish, 1974, Thm. 2.3). Assumption D is also implied by stationarity and certain mixing conditions (e.g., Phillips, 1991, p. 458).

To present our main results, we briefly explain some technical details adapted mainly from Knight (1998). First, note that for any a_n, the centered and rescaled LAD estimator $\hat{\theta}_n := a_n(\hat{\beta}_n - \beta_0)$ minimizes

$$Z_n(\theta) := \frac{a_n}{\sqrt{n}} \sum_{t=1}^n \left(|\varepsilon_t - a_n^{-1} x_t'\theta| - |\varepsilon_t| \right).$$

For every $x \neq 0$, we have $|x-y| - |x| = -y\operatorname{sgn}(x) + 2\int_0^y [I(x \le s) - I(x \le 0)]ds$, as shown by Knight (1998, 1999) and Rogers (2001), where $I(\cdot)$ denotes the indicator function. Thus,

$$Z_n(\theta) = -\frac{1}{\sqrt{n}} \sum_{t=1}^n \theta' x_t \operatorname{sgn}(\varepsilon_t) + \frac{2a_n}{\sqrt{n}} \sum_{t=1}^n \int_0^{a_n^{-1}x_t'\theta} \Big[I(\varepsilon_t \le s) - I(\varepsilon_t \le 0) \Big] ds \quad \textbf{(2)}$$

$$= Z_n^{(1)}(\theta) + Z_n^{(2)}(\theta), \quad \text{say.}$$

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press

Here, $Z_n^{(1)}(\theta) \Rightarrow -\theta'\zeta$ by Assumption D, and

$$Z_n^{(2)}(\theta) = \frac{2}{n}\sum_{t=1}^{n}\int_0^{x_t'\theta} n^{1/2}\Big[I(\varepsilon_t \le a_n^{-1}s) - I(\varepsilon_t \le 0)\Big]ds =: \frac{2}{n}\sum_{t=1}^{n} D_{nt}, \tag{3}$$

which converges in probability to a nonrandom quantity as shown by Lemmas 1, 3, and 4 in the following section. Specifically, we can split $n^{-1}\sum_{t=1}^{n} D_{nt}$ into the sum of $n^{-1}\sum_{t=1}^{n}\mathrm{E}(D_{nt}|F_t)$ and $n^{-1}\sum_{t=1}^{n}[D_{nt} - \mathrm{E}(D_{nt}|F_t)]$. Lemma 1 in the next section shows that the deviation term is negligible, and the first term is handled by $\mathrm{E}(D_{nt}|F_t) = \Psi_{nt}(x_t'\theta)$, where

$$\Psi_{nt}(x) := \int_0^x \psi_{nt}(s)\,ds, \qquad \psi_{nt}(s) := \sqrt{n}\Big[F_t(a_n^{-1}s) - F_t(0)\Big]. \tag{4}$$

Here, we note that $F_t(s) = F^e(\sigma_t^{-1}s)$ because $\varepsilon_t = \sigma_t e_t$, so that $\psi_{nt}(s) = \psi_n^e(\sigma_t^{-1}s)$, where $\psi_n^e(s) := n^{1/2}[F^e(a_n^{-1}s) - F^e(0)]$. Thus, $\mathrm{E}(D_{nt}|F_t)$ is indeed a function of x_t and σ_t with σ_t being a function of x_t and all lagged variables. Given this formulation, Lemma 3 further establishes under regularity condition (1) that the limit of $n^{-1}\sum_{t=1}^{n}\Psi_{nt}(x_t'\theta)$ is equal to the limit of $n^{-1}\sum_{t=1}^{n}\Psi_t(x_t'\theta)$, where $\Psi_t(x) := \int_0^x \psi_t(s)ds = \lim_{n\to\infty}\Psi_{nt}(x)$, and $\psi_t(\cdot)$ is the limit of $\psi_{nt}(\cdot)$. Therefore, under (1), the limit of $Z_n^{(2)}(\theta)$ is the same as the limit of $(2/n)\sum_{t=1}^{n}\Psi_t(x_t'\theta)$, which is also identical to $(2/n)\sum_{t=1}^{n}\sigma_t\Psi^e(\sigma_t^{-1}x_t'\theta)$ because $\Psi_t(x) = \sigma_t\Psi^e(\sigma_t^{-1}x)$, where $\Psi^e(x) := \int_0^x \psi^e(s)\,ds$ and $\psi^e(s) := \lim_{n\to\infty}\psi_n^e(s)$.

Denoting the limit of $Z_n^{(2)}(\theta)$ by $\tau(\theta)$, the limit of $Z_n(\theta)$ has the form $Z(\theta) = -\theta'\zeta + \tau(\theta)$. Importantly, both $Z_n(\theta)$ and its limit $Z(\theta)$ are convex functions of θ, as proved in Lemma 4. Hence, if $Z(\cdot)$ is minimized at a unique point under some regularity conditions for the properly chosen a_n, we can invoke an "argmin" continuous mapping theorem to obtain the limit distribution of $a_n(\hat\beta_n - \beta_0)$ without establishing stochastic equicontinuity (e.g., Geyer, 1996).

The following LAD asymptotic theory is established for linear models with weakly exogenous regressors and conditionally heteroskedastic disturbances.

THEOREM 1. *Under Assumptions A–D, the centered and rescaled LAD estimator* $\hat\theta_n := a_n(\hat\beta_n - \beta_0)$, *which minimizes* $Z_n(\cdot)$, *converges in distribution to* $\mathrm{argmin}_{\theta\in\mathbb{R}^k} Z(\theta)$, *provided that the limit function* $Z(\cdot) = -\theta'\zeta + \tau(\theta)$ *is uniquely minimized almost surely.*

Theorem 1 is the main result of the current paper and extends the range of potential applications of the Knight (1998, 1999) limit theory to a time series framework that allows for conditional heterogeneity and serial dependence. A few remarks follow. First, the minimizer of $Z(\cdot)$ is unique if $Z(\cdot)$ is strictly convex, which would be implied if $f^e(\cdot)$ is strictly positive, as discussed after the proof of Lemma 4 in the next section. Second, our limit theory differs from earlier work in the way that $\tau(\theta)$ is deployed. Specifically, the stochastic regressors x_t and the conditional heteroskedasticity process σ_t interact so that σ_t plays the

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press

role of a standardizing factor, which differs from previous limit theory and distinguishes the limit form of $\tau(\theta)$ in the present paper. See the examples discussed subsequently. Third, inferences on β_0 may be drawn by combining Theorem 1 and extensions of the generalized bootstrap in Bose and Chatterjee (2001) to a time series context,[1] a project that is left for future research. Finally, Theorem 1 can be still obtained even when, for z only in a neighborhood of zero, $P(e_t \le z|\mathcal{F}_t)$ is identical for all t. This is so because for any finite x, $a_n^{-1}x$ eventually belongs to the neighborhood as $a_n \to \infty$, so that $\psi_n^e(x)$ is identical for all t at the limit.

We now analyze two special examples using Theorem 1 to show explicitly how the results in Knight (1998, 1999) are modified in the presence of conditional heteroskedasticity and weakly exogenous regressors.

Example (Finite density)
Let $0 < f^e(0) < \infty$. Then for $a_n = \sqrt{n}$, we have $\psi_n^e(s) \to f^e(0)s$ by first-order Taylor expansion, which implies that $\psi_{nt}(s) \to f_t(0)s$ with $f_t(0) = \sigma_t^{-1} f^e(0)$. Thus $\Psi_t(x) = f_t(0)\int_0^x s\,ds = \frac{1}{2} f_t(0)x^2$, and

$$Z_n^{(2)}(\theta) = \frac{2}{n}\sum_{t=1}^{n} \tfrac{1}{2} f_t(0)(\theta' x_t)^2 + o_p(1) \to_p \tau(\theta) = \theta' B\theta,$$

$$\text{where } B := \operatorname{plim} \frac{1}{n}\sum_{t=1}^{n} f_t(0)x_t x_t'.$$

Thus, $Z_n(\theta) \Rightarrow -\theta'\zeta + \theta' B\theta$, and this weak limit is minimized almost surely by $(2B)^{-1}\zeta \sim N(0, \frac{1}{4}B^{-1}AB^{-1})$. Therefore, this expression gives the limit distribution of the LAD estimator according to Theorem 1. Further, we note that $f_t(0)x_t x_t' = f_t^e(0)\sigma_t^{-1}x_t x_t'$, implying that $B = f^e(0)\operatorname{plim} n^{-1}\sum_{t=1}^n \sigma_t^{-1}x_t x_t'$ and thereby showing how σ_t and x_t influence B.

Example (Power density)
If e_t has a power density, so that for some $\alpha \le 1$, $f^e(x) = \lambda\alpha|x|^{\alpha-1}$ in a neighborhood of zero, then we have $\psi_n^e(s) = \lambda\,\mathrm{sgn}(s)|s|^\alpha = \psi^e(s)$ by letting $a_n = n^{1/(2\alpha)}$. Therefore, $\Psi^e(x) = \lambda(\alpha+1)^{-1}|x|^{\alpha+1}$, and this implies that $\Psi_t(x) = \lambda(\alpha+1)^{-1}\sigma_t^{-\alpha}|x|^{\alpha+1}$ by virtue of the fact that $\Psi_t(x) = \sigma_t\Psi^e(\sigma_t^{-1}x)$. From this, for each θ,

$$\frac{2}{n}\sum_{t=1}^{n}\Psi_t(x_t'\theta) = \frac{2\lambda}{\alpha+1}\cdot\frac{1}{n}\sum_{t=1}^{n}\sigma_t^{-\alpha}|x_t'\theta|^{\alpha+1} \to_p \tau(\theta).$$

If we further suppose that x_t is a scalar, then

$$\tau(\theta) = \frac{2\lambda h_\alpha}{\alpha+1}\cdot|\theta|^{\alpha+1}, \quad \text{where } h_\alpha := \operatorname{plim}\frac{1}{n}\sum_{t=1}^{n}\sigma_t^{-\alpha}|x_t|^{\alpha+1},$$

so that $\zeta = 2\lambda h_\alpha\,\mathrm{sgn}(\theta_*)|\theta_*|^\alpha$ when θ_* minimizes $-\theta'\zeta + \tau(\theta)$. Thus, $\mathrm{sgn}(\zeta) = \mathrm{sgn}(\theta_*)$, and we obtain that $\theta_* = (2\lambda h_\alpha)^{-1/\alpha}\,\mathrm{sgn}(\zeta)|\zeta|^{1/\alpha}$. This result is very

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press

close to Example 1 in Knight (1998) except that his λ is replaced with our λh_α because of the influence of σ_t in $\tau(\theta)$.

3. PROOFS

For the proof of Theorem 1, the most difficult part is associated with (2), for which we need to prove a law of large numbers for $Z_n^{(2)}(\theta)$. We start by writing $Z_n^{(2)}(\theta) = 2n^{-1}\sum_{t=1}^n D_{nt}$ using a change of variables, where $D_{nt} := \int_0^{x_t'\theta} \sqrt{n}\big[I(\varepsilon_t \le a_n^{-1}s) - I(\varepsilon_t \le 0)\big]ds$, and decompose this sum as follows:

$$\frac{1}{n}\sum_{t=1}^n D_{nt} = \frac{1}{n}\sum_{t=1}^n \Psi_t(x_t'\theta) + \frac{1}{n}\sum_{t=1}^n \Big[\Psi_{nt}(x_t'\theta) - \Psi_t(x_t'\theta)\Big] + \frac{1}{n}\sum_{t=1}^n \Big[D_{nt} - \Psi_{nt}(x_t'\theta)\Big]. \tag{5}$$

The proof proceeds by showing that the second and third terms on the right side of (5) are negligible in probability, as verified in Lemmas 3 and 1. Then, Lemma 4 shows that the first term on the right side obeys the ergodic theorem. These lemmas are integrated into the overall proof of the theorem to aid readability in what follows.

We begin with the third term on the right side of (5), noting that $\Psi_{nt}(x_t'\theta) = \mathrm{E}(D_{nt}|\mathcal{F}_t)$.

LEMMA 1. *Under Assumptions A(i) and (ii) and B, if* $\mathrm{E}\|x_t\|^3 < \infty$, *then*

$$\frac{1}{n}\sum_{t=1}^n \Big[D_{nt} - \mathrm{E}(D_{nt}|\mathcal{F}_t)\Big] \to_p 0.$$

Proof. If we let $\xi_{nt} = D_{nt} - \mathrm{E}(D_{nt}|\mathcal{F}_t)$, then $\mathrm{E}(\xi_{nt}) = 0$, so that it suffices to show that the variance of $n^{-1}\sum_{t=1}^n \xi_{nt}$ goes to zero. We have $\mathrm{var}(\xi_{nt}|\mathcal{F}_t) = \mathrm{E}(D_{nt}^2|\mathcal{F}_t) - \mathrm{E}(D_{nt}|\mathcal{F}_t)^2 \le \mathrm{E}(D_{nt}^2|\mathcal{F}_t)$, where

$$\begin{aligned}
\mathrm{E}(D_{nt}^2|\mathcal{F}_t) &= \int_0^{x_t'\theta}\int_0^{x_t'\theta} n\mathrm{E}\Big[I(\varepsilon_t \le a_n^{-1}(s\wedge r)) - I(\varepsilon_t \le a_n^{-1}(s\wedge 0)) \\
&\qquad - I(\varepsilon_t \le a_n^{-1}(r\wedge 0)) + I(\varepsilon_t \le 0)\Big|\mathcal{F}_t\Big]drds \\
&= \sqrt{n}\int_0^{x_t'\theta}\int_0^{x_t'\theta}\Big[\psi_{nt}(s\wedge r) - \psi_{nt}(s\wedge 0) - \psi_{nt}(r\wedge 0)\Big]drds \\
&= \sqrt{n}\int_0^{x_t'\theta}\int_0^{x_t'\theta}\psi_{nt}(s\wedge r)\,drds - 2\sqrt{n}\theta' x_t\int_0^{x_t'\theta}\psi_{nt}(s\wedge 0)\,ds.
\end{aligned}$$

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press

Therefore, if $x_t'\theta > 0$, then $0 \le \psi_{nt}(s \wedge r) \le \psi_{nt}(s)$ over the domain of the second integral, so that

$$\mathrm{E}(D_{nt}^2|\mathcal{F}_t) \le \sqrt{n}\int_0^{x_t'\theta}\int_0^{x_t'\theta} \psi_{nt}(s)\,dr ds + 0 = \sqrt{n}\,\theta' x_t \Psi_{nt}(x_t'\theta).$$

On the other hand, if $x_t'\theta < 0$, then $0 \ge \psi_{nt}(s \wedge r) \ge \psi_{nt}(s)$ for all s and r between zero and $x_t'\theta$, so that $\int_0^{x_t'\theta} \psi_{nt}(s \wedge r)\,dr \le \int_0^{x_t'\theta} \psi_{nt}(s)dr = \theta' x_t \psi_{nt}(s)$ for $x_t'\theta \le s \le 0$, implying that

$$\mathrm{E}(D_{nt}^2|\mathcal{F}_t) \le \sqrt{n}\,\theta' x_t \Psi_{nt}(x_t'\theta) - 2\sqrt{n}\,\theta' x_t \Psi_{nt}(x_t'\theta) = -\sqrt{n}\,\theta' x_t \Psi_{nt}(x_t'\theta).$$

Thus, we have $\mathrm{E}(D_{nt}^2|\mathcal{F}_t) \le \sqrt{n}\,|\theta' x_t| \Psi_{nt}(x_t'\theta)$ regardless of the value of $x_t'\theta$, and exploiting this inequality yields that

$$\begin{aligned}\mathrm{var}\left(\frac{1}{n}\sum_{t=1}^n \xi_{nt}\right) &= \frac{1}{n^2}\sum_{t=1}^n \mathrm{E}\Big[\mathrm{E}(\xi_{nt}^2|\mathcal{F}_t)\Big] \le \frac{1}{n^2}\sum_{t=1}^n \mathrm{E}\Big[\mathrm{E}(D_{nt}^2|\mathcal{F}_t)\Big] \qquad (6)\\ &\le \mathrm{E}\left[\frac{1}{n^{3/2}}\sum_{t=1}^n |\theta' x_t| \Psi_{nt}(x_t'\theta)\right] = \frac{1}{n^{3/2}}\sum_{t=1}^n \mathrm{E}\Big[|\theta' x_t| \sigma_t \Psi_n^e(\sigma_t^{-1} x_t'\theta)\Big],\end{aligned}$$

where the second inequality follows from the fact that $\mathrm{E}(D_{nt}^2|\mathcal{F}_t) \le \sqrt{n}\,|\theta' x_t| \Psi_{nt}(x_t'\theta)$. The final equality of (6) holds by the fact that $\Psi_{nt}(x_t'\theta) = \sigma_t \Psi_n^e(\sigma_t^{-1} x_t'\theta)$. Now Lemma 2 shows that $|\psi_n^e(x)| \le C(1 + |x|)$ for all x, and thus

$$\Psi_n^e(x) \le C\int_0^{|x|}(1+s)\,ds = C(|x| + \tfrac{1}{2}|x|^2) \le C(|x| + |x|^2), \qquad (7)$$

implying that $|\theta' x_t| \sigma_t \Psi_n^e(\sigma_t^{-1} x_t'\theta) \le C(|\theta' x_t|^2 + \sigma_*^{-1}|\theta' x_t|^3)$. Hence, the right side of (6) is bounded by $Cn^{-3/2}\sum_{t=1}^n (\mathrm{E}|\theta' x_t|^2 + \sigma_*^{-1}\mathrm{E}|\theta' x_t|^3)$, which is $O(n^{-1/2})$ by Assumptions A(i) and (ii) and the fact that $\mathrm{E}\|x_t\|^3 < \infty$ by assumption. ■

LEMMA 2. *Under Assumption B, for some $C < \infty$, $|\psi_n^e(x)| \le C(1 + |x|)$ for all $x \in \mathbb{R}$.*

Proof. First, we suppose that the function $h(\cdot)$ satisfies Assumption B on $V = (c_1, c_2)$ for some $c_1 < 0$ and $c_2 > 0$. Second, we let $M := \max\{1, \frac{1}{2}h(c_1)^{-1}, \frac{1}{2}h(c_2)^{-1}\} < \infty$. Then, $|F^e(x) - F^e(0)| \le Mh(x)$ for every x in V; and if x is not an element of V, then $Mh(c_1) \ge \frac{1}{2}$ and $Mh(c_2) \ge \frac{1}{2}$, whereas $|F^e(x) - F^e(0)| \le \frac{1}{2}$, so that $|F^e(x) - F^e(0)| \le \frac{1}{2} \le Mh(c_j) \le Mh(x)$, $j = 1, 2$. Thus, for some $M < \infty$, we can conclude that $|F^e(x) - F^e(0)| \le Mh(x)$ for *all* $x \in \mathbb{R}$. Third, therefore we have $\psi_n^e(x) \le Mn^{1/2}h(a_n^{-1}x) \le MC_0(1 + |x|)$ for all $x \in \mathbb{R}$, where the last inequality holds by the final condition in Assumption B. The desired result now follows by letting $C = MC_0 < \infty$. ■

Next, we show that the second term on the right-hand side of (5) converges to zero.

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press

LEMMA 3. *Under Assumptions A and C,*

$$\Delta_n := \frac{1}{n}\sum_{t=1}^{n}\left[\Psi_{nt}(x_t'\theta) - \Psi_t(x_t'\theta)\right] \to_p 0.$$

Proof. We have $\Delta_n = n^{-1}\sum_{t=1}^{n}\sigma_t\xi_{nt}$, where $\xi_{nt} := \int_0^{\sigma_t^{-1}x_t'\theta}[\psi_n^e(s) - \psi^e(s)]\,ds$. By Assumptions A and C, we also have

$$|\xi_{nt}| \le \int_0^{\sigma_t^{-1}|x_t'\theta|}\delta_n^*(s)\,ds \le \sigma_t^{-1}|x_t'\theta|\delta_n^*(\sigma_t^{-1}|x_t'\theta|) \le \sigma_t^{-1}|x_t'\theta|\delta_n^*(\sigma_*^{-1}|x_t'\theta|)$$
$$= \sigma_t^{-1}\eta_{nt}, \quad \text{say},$$

where the second inequality follows from the fact that $\delta_n^*(\cdot) \le \sigma_t^{-1}|x_t'\theta|$ over the domain of the integral by virtue of the monotonicity condition of $\delta_n^*(\cdot)$ in Assumption C, so that

$$|\Delta_n| \le \frac{1}{n}\sum_{t=1}^{n}\eta_{nt} = \frac{1}{n}\sum_{t=1}^{n}\eta_{nt}I(\eta_{nt} \le M) + \frac{1}{n}\sum_{t=1}^{n}\eta_{nt}I(\eta_{nt} > M).$$

By the stationarity of x_t and (1), the second term on the right-hand side can be made as small as desired by increasing M. For the same M, the first term converges to zero because its mean converges to zero by the dominated convergence. ■

Finally, the asymptotic limit of the first component on the right-hand side of (5) is established by ergodicity as follows.

LEMMA 4. *Under Assumptions A–C,* $2n^{-1}\sum_{t=1}^{n}\Psi_t(x_t'\theta) \to_p \tau(\theta)$, *where* $\tau(\cdot)$ *is a nonrandom convex function.*

Proof. When (x_t, σ_t) is stationary and ergodic, $\Psi_t(x_t'\theta) = \sigma_t\Psi^e(\sigma_t^{-1}x_t'\theta)$ is also stationary and ergodic, and the convergence of $n^{-1}\sum_{t=1}^{n}\Psi_t(x_t'\theta)$ follows from the ergodic theorem. The convexity of $\tau(\cdot)$ follows from the convexity of $\Psi_t(\cdot)$, which, in turn, is implied by the convexity of $\Psi_n^e(x) := \int_0^x \psi_n^e(s)ds$ with respect to x. Thus, we focus on the latter. Note that

$$\Psi_n^e(x) = \sqrt{n}a_n\int_0^{a_n^{-1}x}\left[F^e(s) - F^e(0)\right]ds,$$

by change of variables. Because $F^e(s) - F^e(0)$ is increasing in s, $\Psi_n^e(\cdot)$ is convex, and its weighted average $\tau_n(\theta) := n^{-1}\sum_{t=1}^{n}\Psi_{nt}(x_t'\theta) = n^{-1}\sum_{t=1}^{n}\sigma_t\Psi_n^e(\sigma_t^{-1}x_t'\theta)$ and its limit $\tau(\theta)$ must be convex in θ also. ■

Note that strict convexity of $\tau(\theta)$ would obtain if $F^e(\cdot)$ is strictly increasing, that is, if $f^e(\cdot)$ is strictly positive, in a neighborhood of zero. We now continue with the proof of the theorem.

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press

Proof of Theorem 1. By (5) and Lemmas 1, 3, and 4, $Z_n^{(2)}(\theta) \to_p \tau(\theta)$ under Assumptions A–C, where $Z_n^{(2)}(\theta)$ is defined in (2). Assumption D also implies that $Z_n^{(1)}(\theta) \Rightarrow -\theta'\zeta$, where $Z_n^{(1)}(\theta)$ is defined in (2). Thus, $Z_n(\theta) \Rightarrow -\theta'\zeta + \tau(\theta)$ for every θ, where the limit is convex by Lemma 4 and is uniquely minimized almost surely by supposition. The desired result follows from Geyer (1996) in the same manner as Knight (1998, 1999). ■

NOTE

1. We thank the referees for raising this possibility.

REFERENCES

Bose, A. & S. Chatterjee (2001) Generalised bootstrap in non-regular M-estimation problems. *Statistics and Probability Letters* 55, 319–328.

Geyer, C.J. (1996) On the Asymptotics of Convex Stochastic Optimization. Manuscript, University of Minnesota, School of Statistics.

Han, C., J.S. Cho, & P.C.B. Phillips (2009) Infinite Density at the Median and the Typical Shape of Stock Return Distributions. Cowles Foundation Discussion Paper No. 1701, Yale University.

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

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

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

McLeish, D.L. (1974) Dependent central limit theorems and invariance principles. *Annals of Probability* 2, 620–628.

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

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

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

https://doi.org/10.1017/S0266466609990703 Published online by Cambridge University Press