LAD ASYMPTOTICS UNDER CONDITIONAL HETEROSKEDASTICITY WITH POSSIBLY INFINITE ERROR DENSITIES

By

Jin Seo Cho, Chirok Han, and Peter C. B. Phillips

June 2009

COWLES FOUNDATION DISCUSSION PAPER NO. 1703

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

LAD Asymptotics under Conditional Heteroskedasticity with Possibly Infinite Error Densities

Jin Seo Cho
Korea University

Chirok Han
Korea University

Peter C. B. Phillips
Yale University, University of York
University of Auckland & Singapore Management University

April, 2009

Abstract

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.

Keywords: Asymptotic leptokurtosis, Convex function, Infinite density, Least absolute deviations, Median, Weak convergence.

Subject Classifications: Primary C12, Secondary G11.

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 in 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 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) have 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, but conditional heteroskedasticity and weakly exogenous regressors were not considered. Finally, Rogers (2001) generalized the framework in Knight (1998) to stationary and integrated time-series contexts without handling conditional heteroskedasticity.

The present note extends this 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:

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

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 $F^e(z) := P(e_t \leq z|\mathcal{F}_t)$ for all t and for all z in a neighborhood of zero. We maintain the following assumption throughout the paper:

Assumption A. *(i)* $\sigma_t \geq \sigma_* > 0$*; (ii)* (x_t, σ_t) *is stationary and ergodic; (iii)* $n^{-1}\sum_{t=1}^n \sigma_t^2 = O_p(1)$*; and (iv)* $E\|x_t\|^4 < \infty$.

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 $\psi_n^e(\cdot)$ converges to a nondegenerate function, then $\psi_n^e(x)$ turns out to be the critical component in determining LAD asymptotics, as detailed below. 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 0 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 non-degenerate. 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 $\psi_n^e(\cdot)$ as well as 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 $n^{1/2}h(a_n^{-1}x) = f^*|x| \leq f^*(1 + |x|)$. These are not the only densities satisfying Assumption B. 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, i.e., 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, $|\psi_n^e(\cdot) - \psi^e(\cdot)| \leq \delta_n^*(\cdot)$,*

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

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

Assumption C 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 5 below, 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 $\delta_n^*(x) = \sup_{|\tilde{x}|\leq|x|} |f^e(n^{-1/2}\tilde{x}) - f^e(0)| \cdot |x|$ and $E\|x_t\|^2 < \infty$, then $\delta_n^*(\cdot)$ trivially satisfies all the conditions in Assumption C. Power densities are another example satisfying Assumption C.

We further illustrate the use of Assumption C by considering one of the examples given in Knight (1998, p. 761). Suppose $F^e(x) - F^e(0) = \lambda x \ln(|x|^{-1})$ in a neighborhood of zero so that a_n is $n^{1/2}(\ln n)/2$, and for every x and for large enough n such that $a_n^{-1}x$ is in the neighborhood. We then have

$$\psi_n^e(x) = n^{1/2}\lambda a_n^{-1}x[\ln(a_n) + \ln(|x|^{-1})] = \lambda x + \frac{2\lambda x(\ln\ln n - \ln 2 + \ln(|x|^{-1}))}{\ln n}.$$

Letting $\psi^e(x) = \lambda x$ yields

$$|\psi_n^e(x) - \psi^e(x)| \leq \frac{2\lambda|x|(\ln\ln n + 1 + |x| + |x|^{-1})}{\ln n} \leq 2\lambda(1 + |x| + |x|^2)\left(\frac{\ln\ln n + 1}{\ln n}\right),$$

because $0 \leq \ln(|x|^{-1}) \leq |x|^{-1}$ for $|x| \leq 1$ and $0 > \ln(|x|^{-1}) > -|x|$ for $|x| > 1$. We may denote the right hand side (RHS) of the final inequality by $\delta_n^*(x)$, which is an increasing function of $|x|$ and converges to zero uniformly on every compact set as $n \to \infty$. The other conditions stated in Assumption C trivially hold if $E\|x_t\|^3 < \infty$.

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

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 x_t is zero, then Assumption D follows by the Lindeberg condition and $A = \operatorname{plim} n^{-1}\sum_{t=1}^n x_t x_t'$ in this case.

In order 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

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

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

$$(2) \quad Z_n(\theta) = -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \theta' x_t \,\mathrm{sgn}(\varepsilon_t) + \frac{2a_n}{\sqrt{n}}\sum_{t=1}^{n}\int_0^{a_n^{-1}x_t'\theta} \Big[I(\varepsilon_t \leq s) - I(\varepsilon_t \leq 0)\Big]ds$$
$$= Z_n^{(1)}(\theta) + Z_n^{(2)}(\theta), \quad \text{say.}$$

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

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

which converges in probability to a nonrandom quantity as shown by Lemmas 2, 4 and 5 in the following section. Denoting this limit by $\tau(\theta)$, the weak 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 5. Hence, if $Z(\theta)$ 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,* $a_n(\hat{\beta}_n - \beta_0) \Rightarrow \mathrm{argmin}_{\theta\in\mathbb{R}^k}\{-\theta'\zeta + \tau(\theta)\}$ *provided that the limit function* $Z(\theta) = -\theta'\zeta + \tau(\theta)$ *is uniquely minimized almost surely.*

Theorem 1 is the main result of the current paper and extends the range of potential applications of Knight's (1998, 1999) limit theory to our time-series framework that allows for conditional heterogeneity and serial dependence. Also, our limit theory differs from earlier work in the way that $\tau(\theta)$ is deployed. To see this more clearly let

$$(4) \quad D_{nt} := \int_0^{x_t'\theta} \sqrt{n}\Big[I(\varepsilon_t \leq a_n^{-1}s) - I(\varepsilon_t \leq 0)\Big]ds,$$

so $Z_n^{(2)}(\theta) = (2/n)\sum_{t=1}^n D_{nt}$. We can split $n^{-1}\sum_{t=1}^n D_{nt}$ into the sum of $n^{-1}\sum_{t=1}^n E(D_{nt}|\mathcal{F}_t)$ and $n^{-1}\sum_{t=1}^n [D_{nt} - E(D_{nt}|\mathcal{F}_t)]$. Lemma 2 in the next section shows that the second term is asymptotically negligible, whereas for the first term, we have

$$(5) \quad \frac{1}{n}\sum_{t=1}^{n} E(D_{nt}|\mathcal{F}_t) = \frac{1}{n}\sum_{t=1}^{n}\Psi_{nt}(x_t'\theta),$$

where $\Psi_{nt}(x_t'\theta) := \int_0^{x_t'\theta} \psi_{nt}(s)ds$ and $\psi_{nt}(s) := \sqrt{n}\Big[F_t(a_n^{-1}s) - F_t(0)\Big]$. Further, Lemma 4 in the next section shows that when $\psi_{nt}(s) \to \psi_t(s)$ pointwise for each t, the probability limit of (5) equals the probability limit of $n^{-1}\sum_{t=1}^n \int_0^{x_t'\theta} \psi_t(s)ds$. Thus, essentially we have

$$(6) \qquad Z_n^{(2)}(\theta) = \frac{2}{n}\sum_{t=1}^n \Psi_t(\theta) + o_p(1), \quad \text{where } \Psi_t(\theta) := \int_0^{x_t'\theta} \psi_t(s)ds.$$

The first term on the right side of (6) should obey the ergodic theorem, and we have $\psi_t(s) = \sigma_t^{-1}\psi^e(\sigma_t^{-1}s)$, where $\psi^e(s)$ is the limit of $\psi_n^e(s) := n^{-1}[F^e(a_n^{-1}s) - F^e(0)]$. Consequently, $\tau(\theta)$ which is defined as the probability limit of $2n^{-1}\sum_{t=1}^n \Psi_t(\theta)$, is in fact the probability limit of $2n^{-1}\sum_{t=1}^n \sigma_t \Psi^e(\sigma_t^{-1}x_t'\theta)$. In this expression we note that the stochastic regressors x_t and the conditional heteroskedasticity process σ_t interact so that σ_t plays the role of a standardizing factor, which differs from previous limit theory and distinguishes the limit form of $\tau(\theta)$ in the present paper.

We now analyze two special examples using Theorem 1 to show explicity 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 $\Psi_t(x) = f_t(0)\int_0^x sds = \frac{1}{2}f_t(0)x^2$, and

$$Z_n^{(2)}(\theta) = \frac{2}{n}\sum_{t=1}^n \tfrac{1}{2}f_t(0)(\theta'x_t)^2 + o_p(1) \to_p \tau(\theta) = \theta' B\theta, \quad \text{where } B := \operatorname{plim}\frac{1}{n}\sum_{t=1}^n f_t(0)x_t x_t'.$$

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,

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

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 $\zeta = 2\lambda h_\alpha \operatorname{sgn}(\theta_*)|\theta_*|^\alpha$ when θ_* minimizes $-\theta'\zeta + \tau(\theta)$. Thus, $\operatorname{sgn}(\zeta) = \operatorname{sgn}(\theta_*)$, and we obtain that $\theta_* = (2\lambda h_\alpha)^{-1/\alpha} \operatorname{sgn}(\zeta)|\zeta|^{1/\alpha}$. This result is very close to Example 1 in Knight (1998) except that his λ is replaced with our λh_α due to 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 (LLN) 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} is defined in (4), and decompose this sum as follows

$$\frac{1}{n}\sum_{t=1}^n D_{nt} = \frac{1}{n}\sum_{t=1}^n \Psi_t(x_t'\theta) + \frac{1}{n}\sum_{t=1}^n \Big[\Psi_{nt}(x_t'\theta) - \Psi_t(x_t'\theta)\Big] + \frac{1}{n}\sum_{t=1}^n \Big[D_{nt} - \Psi_{nt}(x_t'\theta)\Big]. \tag{7}$$

The proof proceeds by showing that the second and third terms on the right side of (7) are negligible in probability, as verified in Lemmas 4 and 2 below. Then, Lemma 5 below 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 (7), noting that $\Psi_{nt}(x_t'\theta) = E(D_{nt}|\mathcal{F}_t)$.

Lemma 2. *Under Assumptions A(i,ii) and B, if $E\|x_t\|^3 < \infty$, then*

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

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

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

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

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

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

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

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

$$\begin{aligned} (8) \quad \operatorname{var}\left(\frac{1}{n}\sum_{t=1}^n \xi_{nt}\right) &= \frac{1}{n^2}\sum_{t=1}^n E\Big[E(\xi_{nt}^2|\mathcal{F}_t)\Big] \leq \frac{1}{n^2}\sum_{t=1}^n E\Big[E(D_{nt}^2|\mathcal{F}_t)\Big] \\ &\leq E\left[\frac{1}{n^{3/2}}\sum_{t=1}^n |\theta' x_t| \Psi_{nt}(x_t'\theta)\right] = \frac{1}{n^{3/2}}\sum_{t=1}^n E\Big[|\theta' x_t| \sigma_t \Psi_n^e(\sigma_t^{-1} x_t'\theta)\Big], \end{aligned}$$

where the second inequality follows from the fact that $E(D_{nt}^2|\mathcal{F}_t) \leq \sqrt{n}\,|\theta' x_t| \Psi_{nt}(x_t'\theta)$. The final equality of (8) holds by the fact that $\Psi_{nt}(x_t'\theta) = \sigma_t \Psi_n^e(\sigma_t^{-1} x_t'\theta)$. Now Lemma 3 below shows that $|\psi_n^e(x)| \leq C(1 + |x|)$ for all x, and thus

$$(9) \quad \Psi_n^e(x) \leq C\int_0^{|x|} (1+s)ds = C(|x| + \tfrac{1}{2}|x|^2) \leq C(|x| + |x|^2),$$

implying that $|\theta' x_t| \sigma_t \Psi_n^e(\sigma_t^{-1} x_t'\theta) \leq C(|\theta' x_t|^2 + \sigma_*^{-1}|\theta' x_t|^3)$. Hence, the right side of (8) is bounded by $Cn^{-3/2}\sum_{t=1}^n (E|\theta' x_t|^2 + \sigma_*^{-1} E|\theta' x_t|^3)$, which is $O(n^{-1/2})$ by Assumption A(*i,ii*) and the fact that $E\|x_t\|^3 < \infty$ by assumption. □

Lemma 3. *Under Assumption B, for some $C < \infty$, $|\psi_n^e(x)| \leq C(1 + |x|)$ 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 $M := \max\{1, 1/2h(c_1), 1/2h(c_2)\} < \infty$. Then, $|F^e(x) - F^e(0)| \leq Mh(x)$ for every x in V obviously; and if x is not an element of V, then $Mh(c_1) \geq 1/2$ and $Mh(c_2) \geq 1/2$ whereas $|F^e(x) - F^e(0)| \leq 1/2$, so that $|F^e(x) - F^e(0)| \leq 1/2 \leq Mh(c_j) \leq Mh(x)$, $j = 1, 2$. Thus, for some $M < \infty$, we can conclude that $|F^e(x) - F^e(0)| \leq Mh(x)$ for *all* $x \in \mathbb{R}$. Third, therefore, we have $\psi_n^e(x) \leq Mn^{1/2} h(a_n^{-1}x) \leq MC_0(1 + |x|)$ 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 side of (7) converges to zero.

Lemma 4. *Under Assumptions A and C,*

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

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

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

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

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

By the stationarity of x_t and (1), the second term in the RHS 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 side of (7) is established by ergodicity as follows:

Lemma 5. *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.

We now continue with the proof of the 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

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

by change of variables. For notational simplicity, let $H'(s) := F^e(s) - F^e(0)$ and $H(x) := \int_0^x H'(s)ds$ and then $H(\cdot)$ is convex, which follows by showing that for any z_1, z_2 and $\lambda \in [0,1]$, $\lambda H(z_1) + (1-\lambda)H(z_2) \geq H(\lambda z_1 + (1-\lambda)z_2)$. If we let $z_1 < z_2$ without loss of generality and that $H_1 := H(z_1)$, $H_2 := H(z_2)$, and $\tilde{z}_\lambda := \lambda z_1 + (1-\lambda)z_2$, then

$$\begin{aligned}\lambda H_1 + (1-\lambda)H_2 &= H_1 + (1-\lambda)(H_2 - H_1) \\ &= H(\tilde{z}_\lambda) + (1-\lambda)(H_2 - H_1) - [H(\tilde{z}_\lambda) - H_1] = H(\tilde{z}_\lambda) + \delta(\lambda), \quad \text{say.}\end{aligned}$$

The proof is completed if for all $\lambda \leq [0,1]$, $\delta(\lambda) \geq 0$. We note that

$$\delta'(\lambda) = -(H_2 - H_1) + H'(\tilde{z}_\lambda)(z_2 - z_1),$$

which is a decreasing function in λ because $\tilde{z}_\lambda$ decreases in λ and $H'(\cdot)$ is increasing. Further, $\delta'(0) = -(H_2 - H_1) + H'(z_2)(z_2 - z_1) \geq 0$ by the mean-value theorem and the monotonicity of $H'(z)$ in z. Similarly, we obtain that $\delta'(1) = -(H_2 - H_1) + H'(z_1)(z_2 - z_1) \leq 0$. Therefore, $\delta(\lambda) \geq \min\{\delta(0), \delta(1)\} = 0$, yielding that $H(\cdot)$ is convex. This proves the convexity of $\Psi_n^e(\cdot)$, so that its weighted average $\tau_n(\theta) := n^{-1} \sum_{t=1}^n \Psi_{nt}(x_t'\theta)$ and its limit $\tau(\theta)$ must be convex in θ as well. □

Proof of Theorem 1. By (7) and Lemmas 2, 4 and 5, $Z_n^{(2)}(\theta) \to_p \tau(\theta)$ 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 5 and is uniquely minimized almost surely by supposition. The desired result follows from Geyer (1996) in the same manner as Knight (1998, 1999). □

References

Geyer, C. J. (1996) On the Asymptotics of Convex Stochastic Optimization. Unpublished Manuscript.

Han, C., J. S. Cho, & P. C. B. Phillips (2009) Infinite Density at the Median and the Typical Shape of Stock Return Distributions. Unpublished Manuscript.

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

——— (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.

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

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