INFINITE DENSITY AT THE MEDIAN AND THE TYPICAL SHAPE OF STOCK RETURN DISTRIBUTIONS

BY

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

Infinite Density at the Median and the Typical Shape of Stock Return Distributions

Chirok HAN

Department of Economics, Korea University, Anam-dong, Seongbuk-gu, Seoul, Korea, 136-701
(*chirokhan@korea.ac.kr*)

Jin Seo CHO

School of Economics, Yonsei University, 262 Seongsanno, Seodaemun-gu, Seoul, Korea, 120-749
(*jinseocho@yonsei.ac.kr*)

Peter C. B. PHILLIPS

Cowles Foundation for Research in Economics, Yale University, PO Box 208281, New Haven, CT 06520-8281;
Department of Economics, University of Auckland, Owen G Glenn Building, 12 Grafton Road, Auckland, New
Zealand; School of Social Sciences, University of Southampton, Southampton SO17 1BJ, U.K.; and School of
Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore
(*peter.phillips@yale.edu*)

Statistics are developed to test for the presence of an asymptotic discontinuity (or infinite density or peakedness) in a probability density at the median. The approach makes use of work by Knight (1998) on L_1 estimation asymptotics in conjunction with nonparametric kernel density estimation methods. The size and power of the tests are assessed, and conditions under which the tests have good performance are explored in simulations. The new methods are applied to stock returns of leading companies across major U.S. industry groups. The results confirm the presence of infinite density at the median as a new significant empirical evidence for stock return distributions.

KEY WORDS: Asymptotic leptokurtosis; Infinite density at the median; Kernel density estimation; Least absolute deviations; Stylized facts.

1. INTRODUCTION

Identifying the distributional shape characteristics of economic variables is an important aspect of statistical description and the search for stylized facts about economic data. Knowledge of the appropriate distributional class including tail shape and peakedness can be particularly important in designing suitable methods of inference, in forecasting, in risk analysis, and in decision making on financial investments. Early studies in empirical finance, such as the classic articles of Mandelbrot (1963) and Fama (1965), recognized these advantages and accordingly sought to identify some stylized distributional features of asset returns (such as heavy tailedness) to assist in laying a statistical foundation for methods of empirical finance. More recently, the importance of distributional shape, density estimation and forecasting was acknowledged in the management of financial risk, the measurement of value at risk, and in financial market volatility (e.g., Gabaix et al. 2003; Ibragimov 2007; Ibragimov and Walden 2007).

In much statistical work, it is conventional to suppose that the variables of interest have finite density over their entire support. It is also convenient to rely on normal density functions or modified versions based on mixtures of normals in fitting economic data and in diagnostic statistical analysis. However, the condition of a finite density may be restrictive in some situations, particularly for asset return data, which are generally acknowledged to be highly peaked at the median return. Moreover, imposing the condition of a finite density when it is false will have consequences for inference. For example, applying

goodness-of-fit tests, such as the Kolmogorov–Smirnov test, can easily lead to inappropriate conclusions when the relevant density functions are not finite over their domains.

Heavy tailedness in returns is often accompanied by heavy concentrations of observations around the median return, which is commonly zero. This peakedness or leptokurtosis in the distribution is a stylized fact for most financial asset return data. Sometimes the concentration around the median return may be so great as to produce an asymptote in the density at the median. This "asymptotic leptokurtosis" is one focus of interest in the present article.

Existence of an infinite pole in the density combined with possible heavy tails is also important when evaluating various estimation techniques. In particular, the quality of least squares can be severely compromised if the error distribution has heavy tails, in which case least absolute deviation (LAD) estimation is an attractive alternative. As Knight (1998) pointed out, the finite sample and asymptotic properties of the LAD estimator are determined by how the density behaves around the median. When the density is infinite, then the LAD estimator is super consistent and it is, therefore, possible to construct sharper confidence intervals than the usual intervals that are based on least-squares estimation. However, assumptions relating to the existence of the density at the median were regarded as difficult to verify

(see Knight 1998, p. 756) and no procedures for doing so have yet been suggested. The methodology proposed in the present article provides one solution to this issue.

The possible nonexistence of the probability density was recently considered by Zinde-Walsh (2008) from a different perspective. By means of generalized functions and generalized random processes, Zinde-Walsh examined the asymptotic features of the kernel estimator for the conditional mean under general conditions. The present article differs in that we consider the median rather than the mean and also in that we directly propose a method to test the existence of the density at the median.

Accordingly, the main theoretical goal of the article is to provide a statistical test of infinite density at the median. Our approach is to exploit the asymptotic theory of Knight (1998), and in particular, the mild regularity conditions under which the sample median is asymptotically normal when the density is finite and nonzero at the median. When the density has an infinite discontinuity at the median, the sample median converges to the population median at a faster rate than the usual $\sqrt{n}$ rate, where n is the sample size. This simple differential provides a device for constructing test statistics for asymptotic leptokurtosis that can be applied in general linear econometric models where there are other nuisance parameters to estimate. The approach combines a nonparametric kernel density estimate at the median with the sample median to deliver a simple nonparametrically studentized test statistic.

The empirical goal of the article is to evaluate the leptokurtosis of certain financial asset return data and assess the evidence in support of an infinite density at the median return. Much empirical literature already documents the nonnormality of asset return distributions and the leptokurtosis of these distributions (Mandelbrot 1963; Fama 1965). The present article takes the further step of testing for infinite density in stock returns. More specifically, we apply our tests to the return residuals from a simple autoregression. The empirical findings indicate that a significant number of leading companies in U.S. industries have asset returns with infinite density at the median. Accordingly, there appears to be evidence supporting infinite leptokurtosis as a new empirical evidence for some stock return distributions in the U.S.

The plan of the article is as follows. Section 2 develops the test statistic along with Monte Carlo experiments. Section 3 reports the empirical application and concluding remarks are given in Section 4. Proofs and data information are given in the Appendix.

2. MEDIAN INFINITE DENSITY TESTS

We consider the linear regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{Y}$ and $\mathbf{X}$ are $n \times 1$ and $n \times p$ matrices with of Y_t and $\mathbf{X}_t'$, respectively, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$. Our aim is in testing whether or not the density of ε_t is finite. For this purpose, we first motivate the hypotheses and the test in the case of independently and identically distributed (iid) disturbances and exogenous regressors. Then, the work is extended to the time series case allowing the regression errors to be conditionally heteroscedastic and the regressors to be predetermined (i.e., weakly exogenous). These extensions are important for our empirical application to financial data.

2.1 Motivational Remarks

We motivate our tests in a heuristic way by letting $(\mathbf{X}_t', \varepsilon_t)$ be iid, where $\mathbf{X}_t$ and ε_t are mutually independent. Let $F(\cdot)$ and $f(\cdot)$ be the cumulative distribution function (cdf) and the probability density function (pdf) of ε_t, respectively. It is well known that the LAD estimator, say $\hat{\boldsymbol{\beta}}_n$, is $\sqrt{n}$ consistent and its asymptotic distribution is $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \Rightarrow (2f_0)^{-1}N(0, \mathbf{C}^{-1})$ whenever $f_0 := f(0)$ is positive and finite under suitable regularity conditions, where $\mathbf{C} := \text{plim} \, n^{-1}\mathbf{X}'\mathbf{X}$, and $\boldsymbol{\beta}_0$ is a p-vector of parameters defined as $\boldsymbol{\beta}_0 := \text{argmin}_{\boldsymbol{\beta}} E|Y_t - \mathbf{X}_t'\boldsymbol{\beta}|$.

The meaning of the parameter $\boldsymbol{\beta}_0$ is given in the literature in numerous ways. In regression contexts, $\boldsymbol{\beta}_0$ is identified by the zero conditional median assumption: $\text{median}(\varepsilon_t|\mathbf{X}_t) = 0$. If $\mathbf{X}_t = 1$, then $\boldsymbol{\beta}_0$ is itself the median of Y_t, corresponding to the 0.5th regression quantile of Bassett and Koenker (1978). Bloomfield and Steiger (1983), Pollard (1991), and Phillips (1991) also focused on quantile and/or LAD estimation and confirmed the result in various environments. Phillips (1991) worked under dynamic misspecification, Koenker and Zhao (1996) worked with the quantile regression model using time series data and conditionally heteroscedastic disturbances, and Kim and White (2003) studied a misspecified quantile regression model with conditional heteroscedastic disturbances using iid data.

The situation is very different if $f(x)$ asymptotes to infinity as x tends to zero. In that event, the convergence rate of the LAD estimator is determined by the divergence speed of $f(x)$ as $x \to 0$ and the shape of $F(\cdot)$ near zero. In such conditions, Knight (1998) developed LAD asymptotic theory for iid data under the condition that the sequence of functions $\psi_n(s) := \sqrt{n}[F(a_n^{-1}s) - F(0)]$ converges to a nondegenerate limit function. In this setting the scale component a_n in $\psi_n(s)$ is the convergence rate of the LAD estimator. For example, if $f(x) \simeq \lambda\alpha|x|^{\alpha-1}$ near $x = 0$ for some $\alpha \leq 1$ and $\lambda \in (0, \infty)$, then we have $F(x) - F(0) \simeq \lambda\,\text{sgn}(x)|x|^{\alpha}$ near zero, so that $\psi_n(s) \to \lambda\,\text{sgn}(s)|s|^{\alpha}$ with $a_n = n^{1/2\alpha}$, as demonstrated by Knight (1998). Thus, if $\alpha = 1$ (so $f_0 < \infty$), $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ has a nondegenerate limit, whereas $\sqrt{n}/a_n \to 0$ and $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = (\sqrt{n}/a_n)a_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \to_p 0$ if $\alpha < 1$ (so $f_0 = \infty$).

The present article exploits these differences in the limit behavior of $n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ under the different forms of $f(x)$ in the vicinity of the origin to provide information about distributional shape. A particular focus of attention relates to various leptokurtotic forms including extreme forms in which the density asymptotes at the origin. The relevant hypotheses in this case can be formulated specifically in null and alternative forms as follows:

$$\mathcal{H}_0 : f(0) \in (0, \infty) \quad \text{versus} \quad \mathcal{H}_1 : f(0) = \infty. \tag{1}$$

The main motivation for considering these particular hypotheses stems from empirical observations of financial data. As explained in the Introduction, many financial asset returns exhibit distributions that appear so heavily peaked at the median as to throw into doubt whether the density is finite at the origin. We seek to provide a mechanism for investigating this possibility in a formal manner with a statistical test procedure that enables a formal test of Equation (1).

If $\boldsymbol{\beta}_0$ were known, the goal of the present article could be relatively easily achieved by exploiting

$$\hat{B}_n := 4n\hat{f}_0^2(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)'\mathbf{C}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0), \qquad (2)$$

as a suitable test statistic, where $\hat{f}_0$ is a density estimator for f_0. This quantity converges to χ_p^2 under $\mathcal{H}_0$, whereas under $\mathcal{H}_1$ we have $\sqrt{n}/a_n \to 0$ and $\hat{f}_0 \to_p \infty$, so the limit behavior of

$$\hat{B}_n = 4\underbrace{(\sqrt{n}/a_n)^2}_{=o(1)} \cdot \underbrace{\hat{f}_0^2}_{\neq O_p(1)} \cdot \underbrace{a_n^2(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)'\mathbf{C}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)}_{=O_p(1)},$$

depends on how fast $\sqrt{n}/a_n$ converges to zero and how fast $\hat{f}_0$ diverges. In particular, if the divergence speed of $\hat{f}_0$ is slower than the convergence rate of $\sqrt{n}/a_n$ to zero, then $(\sqrt{n}/a_n)\hat{f}_0 \to_p 0$ and accordingly $\hat{B}_n$ converges to zero in probability under $\mathcal{H}_1$. As we discuss later more specifically, the Nadaraya–Watson estimator based on the LAD residuals works well for this purpose as a density estimator if the bandwidth parameter δ_n converges to zero while $\sqrt{n}\delta_n \to \infty$, i.e.,

$$\delta_n + \frac{1}{\sqrt{n}\delta_n} \to 0, \quad \text{as } n \to \infty. \qquad (3)$$

The test statistic $\hat{B}_n$, therefore, enables a consistent test of $\mathcal{H}_0$ against $\mathcal{H}_1$ by exploiting the different asymptotic behavior of the statistic under the null and alternative. Specifically, we reject the null hypothesis at a given significance level if the test statistic is less than the corresponding left-tailed critical value of the χ_p^2 distribution. (The convergence of $\hat{B}_n$ to zero under the alternative happens because the bandwidth is not too small. So one may suspect that a usual right-tailed test may be available if the bandwidth were chosen to converge to zero faster. But this strategy is not promising because then the accuracy of $\hat{f}_0$ is so poor that the test is considerably oversized.)

In practice, $\boldsymbol{\beta}_0$ is usually unknown, in which case, we can proceed by splitting the sample into two equal sized subsets. If n is even, equal sized subsets can be obtained by taking the first and second half of the sample. If n is odd, we may simply discard the first, the last, or the middle observation to obtain equal-sized subsets. If unequal subsets have to be used, the procedures given in the following may be modified by rewriting in an obvious way analogous to the weighting used in the jackknife (Quenouille 1956).

Let $\tilde{\boldsymbol{\beta}}_{1n}$ and $\tilde{\boldsymbol{\beta}}_{2n}$ be the LAD estimators from the first subset (i.e., for $t = 1, \ldots, n/2$) and the remainder of the sample (i.e., $t = n/2 + 1, \ldots, n$), respectively. When $\mathbf{X}_1$ and $\mathbf{X}_2$ are the equal-sized submatrices of $\mathbf{X}$ such that $\mathbf{X} = (\mathbf{X}_1' \vdots \mathbf{X}_2')'$, if $n^{-1}\mathbf{X}'\mathbf{X} \to_p \mathbf{C}$, then both $(2/n)\mathbf{X}_1'\mathbf{X}_1$ and $(2/n)\mathbf{X}_2'\mathbf{X}_2$ also converge to the same limit $\mathbf{C}$, implying that for $j = 1, 2$, $\sqrt{n/2}(\tilde{\boldsymbol{\beta}}_{jn} - \boldsymbol{\beta}_0) \Rightarrow (2f_0)^{-1}\mathbf{C}^{-1/2}\mathbf{Z}_j$ under $\mathcal{H}_0$, where $(\mathbf{Z}_1, \mathbf{Z}_2)' \sim N(0, I)$. We may consider the differential $\sqrt{2n}f_0(\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n})$ as our test device, which weakly converges to $\mathbf{C}^{-1/2}\mathbf{Z}_1 - \mathbf{C}^{-1/2}\mathbf{Z}_2 \sim N(0, 2\mathbf{C}^{-1})$ under $\mathcal{H}_0$. Thus, it follows that $nf_0^2(\Delta\tilde{\boldsymbol{\beta}}_n)'\mathbf{C}(\Delta\tilde{\boldsymbol{\beta}}_n) \Rightarrow \chi_p^2$ under $\mathcal{H}_0$, where $\Delta\tilde{\boldsymbol{\beta}}_n := \tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n}$. Now a useful test statistic can be constructed from this quantity by replacing $\mathbf{C}$ and f_0 with $n^{-1}\mathbf{X}'\mathbf{X}$ and $\hat{f}_0$, respectively. Again, the null distribution is χ_p^2, whereas the statistic converges to zero in probability under $\mathcal{H}_1$. This heuristic explanation underpinning the test can be formally stated as follows.

Theorem 1. Suppose that $(\mathbf{X}_t, \varepsilon_t)$ is iid over t, and that $\mathbf{X}_t$ and ε_t are independent. Under Assumptions A and B stated in Section 2.2 with $\sigma_t \equiv 1$, the following holds:

(i) If $f(0) \in (0, \infty)$ and $f(s)$ is continuous in a neighborhood of zero, then $\hat{B}_n \Rightarrow \chi_p^2$.

(ii) If $f(0) = \infty$, and Equation (10) in Section 2.2 holds, then $\hat{B}_n \to_p 0$.

This theorem may be obtained as a special case of Theorem 3 in the next section for conditionally heteroscedastic time series data (except that the covariance matrix of the LAD estimator is estimated under the iid and exogeneity assumption). We, therefore, omit the proof and provide further discussion later.

2.2 Extensions to Time Series Contexts

The heuristic arguments that apply for iid data with strictly exogenous regressors can be extended to times series models with lagged dependent variables as regressors on the right-hand side and conditionally heteroscedastic errors. For this purpose, let $\mathcal{F}_t$ be the sigma field generated by $(\mathbf{X}_t, \varepsilon_{t-1}, \mathbf{X}_{t-1}, \varepsilon_{t-2}, \ldots)$ and let $\varepsilon_t := \sigma_t e_t$, where σ_t is adapted to $\mathcal{F}_t$ and e_t is iid with pdf $f^e(\cdot)$. Our interest focuses on testing

$$\mathcal{H}_0' : f^e(0) \in (0, \infty) \quad \text{versus} \quad \mathcal{H}_1' : f^e(0) = \infty.$$

We also let $F_t(s)$ and $f_t(s)$ denote the conditional cdf and pdf of ε_t, respectively, so that $F_t(s) = P(\varepsilon_t \le s | \mathcal{F}_t) = P(e_t \le \sigma_t^{-1}s | \mathcal{F}_t) = F^e(\sigma_t^{-1}s)$, and $f_t(s) := F_t'(s) = \sigma_t^{-1}f^e(\sigma_t^{-1}s)$. Further, the previous definition of the quantity $\psi_n(s)$ is here modified to $\psi_{nt}(s) = \sqrt{n}[F_t(a_n^{-1}s) - F_t(0)]$, where a_n is selected so that $\psi_{nt}(s)$ has a nondegenerate (i.e., nonzero and finite) limit on an open set. If $f_t(0)$ is finite, then $a_n = \sqrt{n}$, and if $f_t(0) = \infty$, then $\sqrt{n}/a_n \to 0$ by the same illustration of the power density given above. Finally, we let $\Psi_{nt}(s) := \int_0^s \psi_{nt}(r)\, dr$ and also denote $f_t(0)$ as f_{0t} for notational simplicity. Here, the functional $\Psi_{nt}(\cdot)$ embodies the nonstochastic component of the centered and rescaled objective function $Z_n(\cdot)$ of Equation (4) below—see also Equation (5). In the classical environment where there is a continuous error density, this term is further expanded to form the "denominator" of the asymptotic distribution of the LAD estimator as shown in Equation (6).

We allow for (σ_t) to be a stochastic process adapted to $\mathcal{F}_t$. Thus, the model can be interpreted within the framework of generalized autoregressive conditional heteroscedasticity (G)ARCH models (Bollerslev, Engle, and Nelson 1994). The motivation for this setup follows from the fact that much economic data, particularly in finance, exhibit heteroscedastic behavior that is well characterized and frequently modeled in practice by the (G)ARCH effect. It is useful to employ the heteroscedasticity process (σ_t) in analyzing heavy-tailed densities although this formulation is not identical to conventional (G)ARCH model effects unless conditional median and mean equations are the same.

We first establish LAD asymptotics by following Knight (1998, 1999). The argument is sketched here to exposit the main ideas and a formal statement is given in Theorem 2, which is proved in Cho, Han, and Phillips (2010, hereafter CHP). The

asymptotic behavior of the LAD estimator is obtained by analyzing the following rescaled and centered objective function:

$$Z_n(\mathbf{u}) := \frac{a_n}{\sqrt{n}} \sum_{t=1}^{n} (|\varepsilon_t - a_n^{-1}\mathbf{X}_t'\mathbf{u}| - |\varepsilon_t|)$$

$$= -\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{X}_t'\mathbf{u} \, \mathrm{sgn}(\varepsilon_t)$$

$$+ \frac{2a_n}{\sqrt{n}} \sum_{t=1}^{n} \int_0^{a_n^{-1}\mathbf{X}_t'\mathbf{u}} [I(\varepsilon_t \le s) - I(\varepsilon_t \le 0)] \, ds$$

$$= Z_n^{(1)}(\mathbf{u}) + Z_n^{(2)}(\mathbf{u}), \quad \text{say,} \tag{4}$$

where the second line follows from the representation $|x - y| - |x| = -y \, \mathrm{sgn}(x) + 2 \int_0^y [I(x \le s) - I(x \le 0)] \, ds$ for all $x \ne 0$ (see Knight 1998). Because $Z_n(u)$ is minimized by the centred and scaled estimator $a_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$, the behavior of $a_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ is determined by that of $Z_n(\cdot)$. In particular, the minimizer $a_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ of $Z_n(\cdot)$ weakly converges to the minimizer of the limit of $Z_n(\cdot)$ under quite mild regularity conditions (e.g., Assumption A). Conveniently, it is sufficient to establish the pointwise probability limit of $Z_n(\cdot)$ because it is a convex function (see Knight 1998 and Geyer 1996).

The limit of $Z_n(\mathbf{u})$ can be derived after obtaining the limits of $Z_n^{(1)}(\mathbf{u})$ and $Z_n^{(2)}(\mathbf{u})$ separately. First, for the limit of $Z_n^{(1)}(\mathbf{u})$, we simply apply a central limit theorem (CLT) for a martingale difference array (MDA) under the usual regularity conditions, so that $Z_n^{(1)}(\mathbf{u}) \Rightarrow -\mathbf{u}'\mathbf{G}$, where $\mathbf{G} \sim N(0, \mathbf{C})$ with $\mathbf{C} := \mathrm{plim}\, n^{-1}\mathbf{X}'\mathbf{X}$ as before. Second, for the limit of $Z_n^{(2)}(\mathbf{u})$, we map s to $a_n s$ and apply a change of variables to get

$$Z_n^{(2)}(\mathbf{u}) = \frac{2}{n} \sum_{t=1}^{n} \int_0^{\mathbf{X}_t'\mathbf{u}} \sqrt{n}[I(\varepsilon_t \le a_n^{-1}s) - I(\varepsilon_t \le 0)] \, ds$$

$$= \frac{2}{n} \sum_{t=1}^{n} \tilde{z}_{nt}, \quad \text{say.}$$

Given this expression, we may decompose $Z_n^{(2)}(\mathbf{u})$ into two sums by introducing the quantity $\xi_{nt} := \tilde{z}_{nt} - E[\tilde{z}_{nt}|\mathcal{F}_t]$, which is a martingale difference. Because $E[\tilde{z}_{nt}|\mathcal{F}_t] = \Psi_{nt}(\mathbf{X}_t'\mathbf{u})$, we have $\tilde{z}_{nt} = \xi_{nt} + \Psi_{nt}(\mathbf{X}_t'\mathbf{u})$, and

$$Z_n^{(2)}(\mathbf{u}) = \frac{2}{n} \sum_{t=1}^{n} \xi_{nt} + \frac{2}{n} \sum_{t=1}^{n} \Psi_{nt}(\mathbf{X}_t'\mathbf{u}). \tag{5}$$

The first term on the right-hand side is an average of an MDA, so that it should be negligible in probability, whereas the second term should follow a law of large numbers (LLN) under suitable regularity conditions. In particular, under $\mathcal{H}_0'$ we have $a_n = \sqrt{n}$ and

$$\frac{2}{n} \sum_{t=1}^{n} \Psi_{nt}(\mathbf{X}_t'\mathbf{u}) = \mathbf{u}'\left(\frac{1}{n} \sum_{t=1}^{n} f_{0t}\mathbf{X}_t\mathbf{X}_t'\right)\mathbf{u} + o_p(1)$$

$$\to_p \mathbf{u}'\mathbf{A}\mathbf{u}, \quad \text{say,} \tag{6}$$

by the ergodic theorem because

$$\Psi_{nt}(\mathbf{X}_t'\mathbf{u}) = E[\tilde{z}_{nt}|\mathcal{F}_{t-1}]$$

$$= \int_0^{\mathbf{X}_t'\mathbf{u}} \sqrt{n}[F_t(a_n^{-1}s) - F_t(0)] \, ds$$

$$= \int_0^{\mathbf{X}_t'\mathbf{u}} f_{0t}s \, ds + o_p(1)$$

$$= \frac{1}{2}f_{0t}\mathbf{u}'\mathbf{X}_t\mathbf{X}_t'\mathbf{u} + o_p(1).$$

Therefore, under $\mathcal{H}_0'$, we find that $Z_n^{(2)}(\mathbf{u}) \to_p \mathbf{u}'\mathbf{A}\mathbf{u}$. Combining the limit behavior of $Z_n^{(1)}(\mathbf{u})$ and $Z_n^{(2)}(\mathbf{u})$ now yields $Z_n(\mathbf{u}) \Rightarrow -\mathbf{u}'\mathbf{G} + \mathbf{u}'\mathbf{A}\mathbf{u}$, which is minimized when $\mathbf{u} = \frac{1}{2}\mathbf{A}^{-1}\mathbf{G}$, and this quantity is distributed as $N(0, \frac{1}{4}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1})$ under $\mathcal{H}_0'$. This is the limit distribution of $a_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ under $\mathcal{H}_0'$. Theorem 2 details this argument more rigorously.

We again construct a more realistic test statistic for the unknown parameter $\boldsymbol{\beta}_0$ by splitting the sample into two subsets. Let $\tilde{\boldsymbol{\beta}}_{1n}$ and $\tilde{\boldsymbol{\beta}}_{2n}$ denote the two LAD estimates from the first and second halves of the sample as before. Because $\sqrt{n/2}(\tilde{\boldsymbol{\beta}}_{jn} - \boldsymbol{\beta}_0) \Rightarrow \frac{1}{2}\mathbf{A}^{-1}\mathbf{G}_j$, where $\mathbf{G}_1$ and $\mathbf{G}_2$ are independent and distributed as $N(0, \mathbf{C})$, we have $\sqrt{n/2}(\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n}) \Rightarrow \frac{1}{2}\mathbf{A}^{-1}(\mathbf{G}_1 - \mathbf{G}_2)$, which is also normally distributed, i.e., $N(0, \frac{1}{2}\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1})$. It follows immediately that $n(\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n})'\mathbf{A}\mathbf{C}^{-1}\mathbf{A}(\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n}) \Rightarrow \chi_p^2$ under $\mathcal{H}_0'$. Finally, the unknown elements $\mathbf{C}$ and $\mathbf{A}$ can be replaced by consistent estimators. As $\mathbf{C}$ is the limit variance of $n^{-1/2}\sum_{t=1}^{n}\mathbf{X}_t\,\mathrm{sgn}(\varepsilon_t)$, it can be estimated consistently by $\hat{\mathbf{C}}_n := n^{-1}\mathbf{X}'\mathbf{X}$. $\mathbf{A}$ can be consistently estimated by

$$\hat{\mathbf{A}}_n := \frac{1}{n} \sum_{t=1}^{n} \hat{f}_{0t}\mathbf{X}_t\mathbf{X}_t', \quad \text{where } \hat{f}_{0t} = \delta_n^{-1}K(\delta_n^{-1}\hat{\varepsilon}_t) \tag{7}$$

for some kernel $K(\cdot)$ and bandwidth δ_n, and $\hat{\varepsilon}_t = y_t - \mathbf{X}_t'\hat{\boldsymbol{\beta}}_n$. The kernel function $K(\cdot)$ and the bandwidth δ_n are required to satisfy some regularity conditions, which are provided in Assumption B and in Theorem 3. The feasible test statistic therefore has the following form

$$B_n := (\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n})'\hat{\mathbf{P}}_n(\mathbf{X}'\mathbf{X})^{-1}\hat{\mathbf{P}}_n(\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n}), \quad \text{where}$$

$$\hat{\mathbf{P}}_n := \sum_{t=1}^{n} \hat{f}_{0t}\mathbf{X}_t\mathbf{X}_t', \tag{8}$$

and as before the null hypothesis is rejected at a given significance level if B_n is smaller than the corresponding left-tailed critical value of the χ_p^2 distribution (e.g., 0.00393214 if $p = 1$ and the significance level is 5%).

We now examine the large sample behavior of B_n. We first provide assumptions necessary for deriving the asymptotics of the LAD estimator. Given that $\varepsilon_t = \sigma_t e_t$, it is convenient and common practice, though not strictly necessary here, to assume that e_t is iid. Further, the local behavior of the density around zero is important for the asymptotics of the LAD estimator. Thus, we may assume that $P(e_t \le s|\mathcal{F}_t) = F^e(s)$ for all s in a neighborhood of zero, which we call *local homogeneity*. Given this, if we let $\psi_n^e(s) := \sqrt{n}[F^e(a_n^{-1}s) - F^e(0)]$ and $\Psi_n^e(x) := \int_0^x \psi_n^e(s) \, ds$, then it trivially follows that $\psi_{nt}(s) = \psi_n^e(\sigma_t^{-1}s)$ and $\Psi_{nt}(x) = \sigma_t\Psi_n^e(\sigma_t^{-1}x)$.

The following assumptions are employed to establish the LAD asymptotics.

Assumption A. The following conditions hold:

(i) $(\mathbf{X}'_t, \sigma_t)$ is stationary and ergodic with $\sigma_t^2 \geq \sigma_*^2 > 0$, such that $n^{-1}\sum_{t=1}^n \sigma_t^2 = O_p(1)$ and $E\|\mathbf{X}_t\|^4 < \infty$.

(ii) (e_t) is iid over t; e_t is independent of $(\mathbf{X}'_t, \sigma_t)$ for each t; and 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 $n > n_0$ for some $n_0 < \infty$.

(iii) For some $\psi^e(\cdot)$, there is a symmetric and nonnegative $\delta_n^*(\cdot)$ such that $\delta_n^*(s)$ is increasing as $|s|$ increases; $|\psi_n^e(\cdot) - \psi^e(\cdot)| \leq \delta_n^*(\cdot)$, $\limsup_{n\to\infty} E[\|\mathbf{X}_t\| \times \delta_n^*(\|\mathbf{X}_t\|)] < \infty$, $\delta_n^*(\cdot)$ converges uniformly to zero on every compact neighborhood of zero.

(iv) $E[\text{sgn}(e_t)|\mathcal{F}_t] = 0$ and $n^{-1/2}\max_{1 \leq t \leq n} \|\mathbf{X}_t\| \to_p 0$.

Assumption A is almost identical to the conditions used in CHP to establish LAD asymptotics in an time series environment that allows for conditional heterogeneity and weak exogeneity. Some remarks on the conditions in Assumption A are in order. First, Assumption A(i) allows for the squared terms of $\mathbf{X}_t$ and ε_t to be correlated, so that $\mathbf{C}$ may not be proportional to $\mathbf{A}$, unlike Knight (1998). Second, the assumption that $\sigma_t^2 \geq \sigma_*^2 > 0$ implies that any heavy mass at zero is attributed to the density of e_t, not to the volatility process σ_t^2. Thus a median infinite density of ε_t is sourced in and equivalent to that of e_t. Third, Assumption A(ii) is satisfied by many densities. For example, it is satisfied if $f^e(0)$ is finite or if $f^e(x) = \lambda\alpha|x|^{\alpha-1}$ (i.e., the power density) for $\alpha < 1$ in a neighborhood of zero, so that $f(x)$ asymptotes to infinity as x tends to zero. Fourth, Assumption A(iii) is provided to establish a limit property of $\psi_n^e(\cdot)$ in a way that its limit is a convex function. Finally, Assumption A(iv) is useful for establishing a CLT for $n^{-1}\sum_{t=1}^n \mathbf{X}_t \text{sgn}(\varepsilon_t)$. More detailed explanations on these conditions can be found in CHP.

The following theorem presents the desired LAD asymptotics under these conditions.

Theorem 2 (Cho, Han, and Phillips 2010). Given Assumption A,

$$a_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \quad \Rightarrow \quad \underset{\mathbf{u}\in\mathbb{R}^p}{\text{argmin}}(-\mathbf{u}'\mathbf{G} + \tau(\mathbf{u})), \qquad (9)$$

where $\tau(\mathbf{u}) := 2\,\text{plim}\, n^{-1}\sum_{t=1}^n \Psi_t(\mathbf{X}'_t\mathbf{u})$, $\mathbf{G} \sim N(\mathbf{0}, \mathbf{C})$ with $\mathbf{C} := \text{plim}\, n^{-1}\mathbf{X}'\mathbf{X}$.

This result from CHP (2010) develops the argument of Knight (1998, 1999) into a time series framework that suits the need of the current article. One difference between Theorem 2 and the CHP result is that the CLT is directly assumed in CHP as a high level condition, whereas here it is derived by exploiting Assumption A(iii). Theorem 2 differs from Knight (1998, 1999) mainly because of the presence of conditional heteroscedasticity. In our time series context, σ_t is not necessarily constant, so that it leads to an information matrix inequality if $f^e(0)$ is finite.

Assumption A holds for many datasets and the power density illustrated previously is only one of many examples covered by

Theorem 2. As in the iid data case, Equation (9) also implies that if $f^e(0) < \infty$, then $a_n = \sqrt{n}$ and $\tau(\mathbf{u}) = \mathbf{u}'\mathbf{A}\mathbf{u}$, where $\mathbf{A}$ was defined while obtaining the probability limit of $Z_n^{(2)}(\mathbf{u})$, thus yielding the conventional result that $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \Rightarrow \frac{1}{2}\mathbf{A}^{-1}\mathbf{G}$, whereas if $f^e(0) = \infty$, then $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = O_p(a_n^{-1}\sqrt{n}) = o_p(1)$ as a_n is proportional to n^γ with $\gamma > 1/2$.

Next, we provide regularity conditions under which the test statistic B_n defined previously has the desired asymptotic behavior under the null and alternative hypothesis on the error density. The conditions required mainly relate to the asymptotic behavior of $\hat{\mathbf{A}}_n$ defined in Equation (7).

Assumption B. The following conditions hold:

(i) On a neighborhood of zero, $f^e(\cdot) > 0$ and $\bar{f}^e(y)/\bar{f}^e(x) \leq \tilde{M}$ for some finite $\tilde{M}$ for all x and y in the same neighborhood such that $|x| \leq |y|$, where $\bar{f}^e(x) = [F^e(x) - F^e(0)]/x$ for $x \neq 0$.

(ii) The kernel function $K(\cdot)$ satisfies:

   (a) $K(\cdot)$ is a uniformly bounded nonnegative function which is symmetric around zero and nonincreasing on the positive domain;

   (b) $\int K(x)\,dx = 1$, $\int K(x)^2\,dx < \infty$;

   (c) for each y in a neighborhood of zero and for each x, $|K(x + y) - K(x)| \leq \dot{K}(x)|y|$, where $\dot{K}(\cdot)$ is uniformly bounded and $\int[\sup_{|y|\geq|x|}\dot{K}(y)]^2\,dx < \infty$.

(iii) The bandwidth sequence δ_n satisfies $\delta_n \to 0$ and $n^{1/2} \times \delta_n \to \infty$.

The Lipschitz condition in Assumption B(ii)(c) is satisfied by many popular kernel functions. For example, if $K(x) = \max(1 - |x|, 0)$, then the condition holds by letting $\dot{K}(x) = I(-2 \leq x \leq 2)$ for $|y| \leq 1$, where $I(\cdot)$ is the indicator function. As another example, if $K(\cdot) = \phi(\cdot)$ is the standard normal kernel, then for $|y| \leq 1$, we can let $\dot{K}(x) = \sup_{|z|\leq 1}\{(x+z)\phi(x+z)\}$ because $\dot{K}(x) \leq \phi(0)(|x| + 1)I(|x| \leq 1) + (|x| + 1)\phi(|x| - 1)I(|x| > 1)$, where the last bound is clearly square integrable. For the popular Epanechinikov kernel $K(x) = \{3/(4\sqrt{5})\}(1 - 0.2x^2)I(x^2 \leq 5)$, we can let $\dot{K}(x) = 0.3I(|x| \leq 3)$. In general, if $K(\cdot)$ is differentiable, then Assumption B(ii)(c) holds when $|K'(\cdot)|$ is uniformly bounded by a symmetric and square-integrable function which is nonincreasing on the positive domain.

The limit behavior of the test statistic B_n is given as follows.

Theorem 3. Under Assumptions A and B, the following results hold:

(i) If $f^e(0) \in (0, \infty)$ and $f^e(s)$ is continuous in a neighborhood of zero, then $B_n \Rightarrow \chi_p^2$.

(ii) If $\bar{f}^e(s) = [F^e(s) - F^e(0)]/s$ and

$$\bar{f}^e(x)/\bar{f}^e(y) \to \infty \quad \text{as } x, y \to 0 \text{ and } x/y \to 0, \quad (10)$$

then $B_n \to_p 0$.

Theorem 3 gives the limit behavior of B_n under the null and alternative hypotheses. A useful aspect of our testing procedure is that direct evaluation of $e_t = \sigma_t^{-1}\varepsilon_t$ is not required, thus modeling the conditional variance process is not a necessary component of our test. Note that the condition in Equation (10) is stronger than simply assuming that $f^e(0) = \infty$ and

characterizes local behavior of $f^e(\cdot)$ at the origin. This condition enables the test to discriminate null pdf's from alternatives. More specifically, this condition controls the divergence speed of $f^e(x)$ as x tends to zero. Even under the alternative of an infinite density, if the divergence speed is too slow then the discriminating information in finite samples of data may be too weak to identify the alternative. So, the condition in Equation (10) serves as a restriction in the class of alternative distributions that ensures test power against these alternatives. Many relevant density functions satisfy the condition in Equation (10) in spite of this restriction. As an example, for some $c > 0$ and $\alpha < 1$ if $F^e(x) \propto 1/2 + c\,\text{sgn}(x)|x|^\alpha$ around zero, then $\bar{f}^e(x) \propto |x|^{\alpha-1}$, and therefore, $\bar{f}^e(x)/\bar{f}^e(y) \propto |x/y|^{\alpha-1}$, which diverges as $x/y \to 0$, so that Equation (10) follows. A symmetrized gamma distribution with a shape parameter smaller than 1 also satisfies Equation (10). In general, the ratio $[F^e(x) - F^e(0)]/[F^e(y) - F^e(0)] \to 0$ under the alternative if x approaches zero faster than y. What Equation (10) further requires is that $[F^e(x) - F^e(0)]/[F^e(y) - F^e(0)] \to 0$ more slowly than $x/y \to 0$, so that the ratio $\bar{f}^e(x)/\bar{f}^e(y)$ diverges. An obvious counterexample to Equation (10) is a density with a logarithmic or other slowly varying discontinuities at the origin. For example, if $\bar{f}^e(x) \sim \log(1/|x|)$ and $y = x^{1-\eta}$ for some $\eta \in (0, 1)$ as $x \to 0+$, we have $\bar{f}^e(x)/\bar{f}^e(y) \to (1 - \eta)^{-1}$ as $x \to 0+$, thereby violating Equation (10). These density functions may not be discriminated from null densities by our test. So the test will not, in general, be powerful against densities with logarithmic type discontinuities at the median.

In simpler cases where ε_t is independent of $\mathbf{X}_t$, the test B_n can be further simplified. In such cases, we may use the statistic $\tilde{B}_n := \tilde{\lambda}_n^2 (\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n})'(\mathbf{X}'\mathbf{X})(\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n})$ to test the same hypothesis, where $\tilde{\lambda}_n$ is defined by

$$\tilde{\lambda}_n := \frac{1}{n} \sum_{t=1}^n \frac{1}{\delta_n} K\left(\frac{Y_t - \mathbf{X}_t'\hat{\boldsymbol{\beta}}_n}{\delta_n}\right).$$

As before, $\tilde{B}_n$ weakly converges to χ_p^2 under $\mathcal{H}_0'$ but converges to zero under $\mathcal{H}_1'$. The intuition behind this test is identical to that underlying the generic test in Equation (2).

Before conducting Monte Carlo experiments for these tests, we remark that the limit distribution of the sample median depends on the local behavior of the probability density in the vicinity of the median and does not depend upon its behavior elsewhere, as shown by Knight (1998) and Rogers (2001). This property ensures that the limit distribution of statistic $\hat{B}_n$ also depends only on the shape characteristics of the probability density near the median. Thus, asymmetry of the density and possible discontinuities at points other than the median (e.g., at the mean if the mean and the median are different) do not affect the validity of the test.

2.3 Monte Carlo Experiments

We conduct a brief Monte Carlo experiment to examine the finite sample performance of the test. We use two data generating processes (DGP's) with autoregressive conditional heteroscedasticity. Specifically, we suppose that $Y_t = 0.4Y_{t-1} + \varepsilon_t$, $\varepsilon_t = \sigma_t e_t$, and $\sigma_t^2 := 1 + 0.3\varepsilon_{t-1}^2$, where e_t is iid, and its density function is (i) a two-sided gamma (double gamma) distribution

whose functional form is $f^e(x) = \frac{1}{2}\Gamma(\alpha)^{-1}|x|^{\alpha-1}\exp(-|x|)$ and (ii) the mixture $\alpha N(0, 1) + (1 - \alpha)0$. Accordingly, $\{\varepsilon_t, \mathcal{F}_t\}$ is an MDA, and the conditional median equation is identical to the conditional mean equation, mainly due to the symmetry of the distribution of e_t. Further, $f^e(0)$ is finite when $\alpha = 1$ and infinite when $\alpha < 1$ for both DGP's. Note that this DGP differs substantially from the usual case considered in the literature where e_t is assumed to follow a standard normal distribution. In practice, we generate e_t by letting $e_t = \text{sgn}(z_t)v_t$ for DGP (i) and $e_t = z_t\{u_t \geq \alpha\}$ for DGP (ii), where $z_t \sim N(0, 1)$, v_t is independently drawn from a gamma distribution with the "shape" parameters α, and $u_t \sim U(0, 1)$. In our simulations, the associated random variables are generated by the `rnorm`, `rgamma`, and `runif` functions in R (R Development Core Team 2008) and the LAD estimators are obtained by the `quantreg` package in R (Koenker 2008). The standard normal kernel is chosen for $K(\cdot)$, and the bandwidth parameter δ_n is set by the "rule of thumb" parameter suggested by Scott (1992) as a variation of Silverman's (1986) parameter, i.e., 1.06 times the minimum of the standard deviation and the interquartile range divided by 1.34 times $n^{-1/5}$. This bandwidth is popularly selected for empirical data analysis in the literature and is easy to compute. It is, therefore, of interest to see how this bandwidth performs in our experiments.

The simulation results are reported in Table 1. The findings indicate that size ($\alpha = 1$) is approximately accurate, and power ($\alpha < 1$) approaches one as the sample size increases or the α parameter gets smaller, which corresponds to sharper asymptotes in the density. (For both DGP's the test seems slightly oversized, but this disappears as the sample size further increases.) For the first DGP, power increases rather slowly when α is close to unity as the sample size increases. This behavior is indicative of the inconsistency in the test that arises when Equation (10) fails. The second DGP is not regular if $\alpha > 0$ because then the disturbance term has a discontinuous cdf at zero. Power behaves normally in this case as well.

Additional experiments were conducted. We first considered an asymmetric conditional variance process similar to

Table 1. Simulation results for an AR(1) and ARCH(1) model from 10,000 iterations (5% level). DGP: $y_t = 0.0 + 0.4y_{t-1} + \varepsilon_t$, $\varepsilon_t = \sigma_t e_t$, $\sigma_t^2 = 1 + 0.3\varepsilon_{t-1}^2$. Bandwidth = $1.06 \min\{\text{SD}, (\text{interquartile range})/1.34\}n^{-1/5}$, $K(\cdot) = \phi(\cdot)$

$\alpha \setminus n$	100	200	400	800
		$e_t \sim \pm \text{Gamma}(\alpha)$		
1.0	5.34	6.18	5.95	6.34
0.9	6.64	6.87	7.72	8.58
0.7	9.88	13.19	15.73	18.40
0.5	19.55	28.29	38.19	50.46
0.3	42.90	62.65	79.35	91.79
0.1	68.26	85.43	95.37	99.53
		$e_t \sim \alpha N(0, 1) + (1 - \alpha)0$		
1.0	4.71	5.03	5.18	5.47
0.9	14.22	26.16	46.55	72.91
0.8	39.85	69.82	92.01	99.44
0.7	70.28	93.30	99.67	100.00
0.6	89.51	99.12	100.00	100.00

NOTE: $\alpha = 1.0$: size, $\alpha < 1$: power.

threshold-GARCH (TARCH), and obtained similar results which are omitted for brevity. Bandwidths selected by cross-validation were also examined, but the finite sample performance of the test in this case was found to be inferior to that of the test based on Scott's rule of thumb. Issues of kernel and bandwidth choice obviously deserve further investigation in the present context. Based on the limit theory and the reported simulations, we used the Gaussian kernel and Scott rule-of-thumb methods in our empirical applications.

3. EMPIRICAL APPLICATIONS

Asset return distributions are well known to exhibit non-normality. As overviewed in Bollerslev, Engle, and Nelson (1994), the early articles of Mandelbrot (1963) and Fama (1965) pointed out the leptokurtic feature of many asset return distributions. Other stylized facts concerning asset returns are the typical heavy tails of their distributions and the volatility clustering manifested in squared returns, various realized volatility measures, and fitted (G)ARCH models.

The focus of the present study is to examine the leptokurtosis of asset return distributions more carefully and test whether there is empirical support for "infinite leptokurtosis" arising from infinite density at the median. This section reports the results of applying our tests to stock returns of leading companies in U.S. industries. More precisely, we apply our test in Equation (8) to the autoregression

$$r_{i,t} = \alpha_{i,0} + \beta_{i,0} r_{i,t-1} + \varepsilon_{i,t}, \tag{11}$$

where $r_{i,t}$ is the return of the ith company stock in period t. The companies used for our empirical applications are the so-called America top 400 large companies as announced by *Forbes.com* on December 22, 2005. These companies are selected according to the *Forbes.com* criteria of helping investors to identify potential star stocks across 26 industries. In collecting this stock price data for the last 15 years (from May 24, 1991 to May 23, 1996) using *Datastream Advance 3.5*, 242 companies are found to provide a full dataset with no missing observations. The companies are listed in Appendix B and the total number of observations is 3799 after eliminating holiday observations.

Time series features of daily returns are analyzed via Equation (11), which attempts to capture any potential serial dependence in daily returns that may be induced from a variety of sources, including microstructure effects. Indeed, Equation (11) is often motivated as a reduced-form equation in the finance literature. For example, Lo and MacKinlay (1988, 1990), Scholes and Willams (1977), Dimson (1979), and Cohen et al. (1983a, 1983b) recognized that the betas in the standard capital asset pricing model (CAPM) cannot be consistently estimated by ordinary least squares (OLS) regression because of serially correlated residuals induced by nonsynchronous trading. Also, from a time series perspective, Nelson (1991) suggested that an autoregression be used to eliminate serial correlation. Accordingly, we specify Equation (11) as a suitable reduced-form time series model for returns, without being specific about the underlying source of the weak dependence.

The disturbance term $\varepsilon_{i,t}$ in Equation (11) is expected to possess time varying volatility features and to satisfy the MDA

condition. Note that the leptokurtosis feature of daily stock returns cannot be separated from the time varying volatility effects, as pointed out by Bollerslev, Engle, and Nelson (1994). We explicitly allow for the presence of time varying volatility in writing $\varepsilon_{i,t} = \sigma_{i,t} e_{i,t}$, where $e_{i,t}$ is iid, and $\sigma_{i,t}$ is adapted to $\mathcal{F}_{i,t}$, which we define as the smallest σ-field generated by $(r_{i,t-1}, r_{i,t-2}, \ldots)$ for each $i = 1, 2, \ldots, 242$. Based on this modeling framework, we test whether or not the density of $\varepsilon_{i,t}$ is finite at zero. As detailed earlier, B_n is consistent even when time varying volatility is part of the DGP, thereby enabling us to examine the leptokurtosis of financial asset returns in a context that accommodates this volatility.

The test is implemented using the following procedure. First, Equation (11) is estimated by both LAD and OLS regression methods, and we compare the prediction errors obtained from these. Note that OLS provides consistent estimates of the Equation (11) when $\{\varepsilon_{i,t}, \mathcal{F}_{i,t}\}$ is an MDA having finite variance. However, as remarked earlier, the limit of the LAD estimator may be different from OLS when the conditional mean and median equations are different. Hence, we first check whether OLS estimation yields symmetric prediction error distributions. For this, we apply the runs test developed by McWilliams (1990) to our OLS residuals and test the following hypotheses:

$$\mathcal{H}_0'' : f_i^v(\cdot) \text{ is symmetric} \quad \text{versus} \quad \mathcal{H}_1'' : f_i^v(\cdot) \text{ is asymmetric},$$

where $f_i^v(\cdot)$ is the pdf of $v_{i,t}$, which is the OLS residual obtained by estimating AR(1) model for each i. According to McWilliams (1990), the runs test is more powerful against a certain family of alternatives than other tests such as the Cramér–von Mises test constructed from the empirical distribution. Also, the runs test does not assume a continuous distribution for $v_{i,t}$, which is violated under the infinite density hypothesis, so that symmetry of the pdf may not be properly tested by tests that rely on the empirical distribution. These properties give the runs test some potential advantages in the present context.

Next we apply the test in Equation (8) to the prediction errors obtained by LAD estimation of all companies and the companies with symmetric densities according to the runs test. In particular, we examine how the test statistics behave over subsamples with different sample sizes. Specifically, we start the data analysis by testing the hypotheses using the dataset with 1850 observations (February 2, 1999 to May 23, 2006) and computing p-values. Then we perform the same testing procedure using enlarged datasets with 2050 observations (April 17, 1998 to May 23, 2006). In this way we continue to increase the sample size and apply the test to multiple datasets growing in size by 200 observations each time until the sample size reaches 3799 (May 24, 1991 to May 23, 2006). The information from this sequence of tests is collected for each company, the number of companies rejecting the null is counted, and some collective conclusions are then deduced concerning the evidence in support of infinite density.

The stated procedure is partly motivated by the fact that we reject the finite density hypothesis for B_n close to zero. Even under the finite density hypothesis, B_n will still realize some values close to zero with low probability. The above sequential testing procedure serves to raise the rejection probability and

increase test power for those companies that do exhibit infinite density at the median.

Due to space constraints, we do not attempt to report the analysis in full for all the companies considered in the study. Instead, we mainly focus on a specific industry—Health Care Equipment and Services (HCES)—for the presentation of detailed findings, as the results for this industry are fairly typical. Later in the discussion we provide some key summary results for all 242 companies and for those companies with symmetric densities according to the runs test.

Table 2 compares the parameter estimates obtained by OLS and LAD estimation methods. For the nine companies in the HCES industry there are close similarities and some differences in the parameter estimates. The estimated intercepts are all very close to zero for both estimation methods. The fitted autoregressive (AR) coefficients are also small and the two estimates have the same signs in each case, but there are some small systematic differences, most notably that the LAD estimates are all closer to zero than the corresponding OLS estimates. Nevertheless, we cannot at the moment test whether or not the estimated LAD parameters converge to zero, as the asymptotic distribution of the t-statistic for the LAD estimator critically depends on the assumption of finite density, which we want to test in the present article.

Table 2 also reports test statistic values and associated p-values (left-tailed) for the infinite density test. The outcomes differ according to the significance level. At the 5% level, for example, two companies (Universal Health and Omnicare) have infinite density at the median, and at the 1% level one company (Omnicare) has infinite density. This aspect is further accompanied by the runs test. At the 5% level, three companies (Becton Dickinson, Varian Medical System, and Coventry Health Care) have asymmetric densities, so that every company with infinite densities according to the statistic B_n turn out to have symmetric error densities. Overall, a significant number of companies seem to display strong empirical evidence in support of infinite density at the median in the HCES industry.

Extending this analysis to other industries, we collect the results together in Tables 3 and 4. First, we report the proportion of rejections of the null hypotheses of finite and symmetric densities in Table 3. The columns of Table 3 contain the results of testing symmetry using the runs test based upon OLS residuals, and the rows indicate the results of testing finite density using B_n when the level of the test is 5%. Thus, 40 companies turn

Table 3. Testing finite and symmetric densities at 5% level. Sample period: May 24, 1991 to May 23, 2006

Densities	Symmetric	Asymmetric	Sum
Infinite	40 (21.39%)	17 (30.91%)	57 (23.55%)
Finite	147 (78.61%)	38 (69.09%)	185 (76.45%)
Sum	187 (100.0%)	55 (100.0%)	242 (100.0%)

out to have infinite and symmetric densities, and this approximately amounts to 21.39% of the companies with symmetric densities (16.53% of all companies). Next, we examine these findings in relation to the overall tendency to reject the null of finite density. Table 4 summarizes results for the full set of 242 companies (denoted by "All") and its subset of 187 companies with symmetric error densities according to Table 3 (denoted by "Sym."). The table provides the number of companies rejecting the null over different subsamples at various significance levels (1% to 11%). For example, if the level of the test is 5%, then 6.61% (resp. 6.95%) out of the 242 (resp. 187) companies reject the null when the sample period is April 17, 1998 to May 23, 2006. What we observe from Table 4 is that the rejection rate for the finite density hypothesis gets larger as the number of observations gets bigger. This is observed not only for the 5% level of significance, but also for the other levels. This aspect is affirmed in Figure 1, which shows the histograms of p-values obtained for some of the sample periods in Table 4. Evidently, more p-values cluster around zero as sample period gets larger for both groups of the companies.

This evidence taken together amounts to strong support of infinite density at the median as a remarkable new stylized distributional feature for U.S. industry stock returns.

4. CONCLUDING REMARKS

This article develops and applies a new testing procedure to evaluate kurtosis and explicitly test whether a probability density has an asymptote or infinite discontinuity at the median. The approach makes use of the limit theory for L_1 estimation pioneered by Knight (1998) and extended in recent work by the authors (CHP 2010), which allows for such discontinuities in the density in time series settings that include conditional heterogeneity and serial dependence. The power of the test stems simply from the fact that the sample median converges to the

Table 2. OLS and LAD estimators and test statistic values. Sample period: May 24, 1991 to May 23, 2006

Companies	$\hat{\alpha}_n^{OLS}$	$\hat{\beta}_n^{OLS}$	$\hat{\alpha}_n^{LAD}$	$\hat{\beta}_n^{LAD}$	B_n	p-value*	Runs test	p-value
Universal Health	6.87e−4	0.014	0.000	0.000	0.000	**0.000**	0.632	0.426
Biomet	6.57e−4	−0.089	0.000	−0.047	7.139	0.971	2.086	0.148
Stryker	7.57e−4	−0.005	0.000	0.000	0.495	0.219	0.221	0.637
Becton Dickinson	4.44e−4	−0.052	0.000	0.000	1.374	0.496	3.856	**0.049**
Omnicare	9.23e−4	0.031	0.000	0.000	0.000	**0.000**	1.403	0.236
Varian Medical Sys.	7.67e−4	−0.031	0.000	0.000	0.201	0.095	6.828	**0.008**
Humana	5.79e−4	−0.030	0.000	0.000	3.195	0.797	0.095	0.757
UnitedHealth Group	9.28e−4	0.037	0.001	0.020	3.161	0.794	0.685	0.407
Coventry Health Care	1.12e−3	−0.012	0.000	0.000	3.535	0.829	5.386	**0.020**

NOTE: *Left-tailed p-values.

Table 4. Proportion of rejected companies out of 242 companies (in percent). Model: $r_{i,t} = \alpha_{i,0} + \beta_{i,0} r_{i,t-1} + \varepsilon_{i,t}$

	Sample Period \ Level	1.00	3.00	5.00	7.00	9.00	11.0
All	02-02-99 ~ 05-23-06	2.89	3.30	4.95	7.43	9.50	10.33
	04-17-98 ~ 05-23-06	2.06	5.38	6.61	7.85	8.67	10.74
	07-01-97 ~ 05-23-06	3.30	4.95	7.85	8.26	9.91	13.63
	09-17-96 ~ 05-23-06	4.13	5.78	8.26	12.80	13.63	15.70
	12-04-95 ~ 05-23-06	5.37	7.85	9.09	12.39	13.63	16.11
	02-21-95 ~ 05-23-06	5.37	8.26	9.91	11.57	13.63	14.87
	05-06-94 ~ 05-23-06	7.85	11.57	14.04	15.28	16.94	19.00
	07-23-93 ~ 05-23-06	10.33	13.63	16.94	19.42	21.48	24.79
	10-08-92 ~ 05-23-06	14.87	18.59	19.83	21.48	25.20	26.03
	12-24-91 ~ 05-23-06	18.59	21.07	21.90	23.55	25.20	26.44
	05-24-91 ~ 05-23-06	21.48	22.31	23.55	25.20	26.85	29.33
Sym	02-02-99 ~ 05-23-06	2.13	2.67	4.81	7.48	9.09	9.62
	04-17-98 ~ 05-23-06	1.60	5.34	6.95	7.48	8.02	9.62
	07-01-97 ~ 05-23-06	2.13	4.27	6.95	6.95	9.09	12.83
	09-17-96 ~ 05-23-06	2.67	4.27	6.41	11.22	11.76	13.90
	12-04-95 ~ 05-23-06	4.81	7.48	8.55	11.22	12.29	14.43
.	02-21-95 ~ 05-23-06	4.27	6.95	8.02	9.62	11.76	13.36
	05-06-94 ~ 05-23-06	6.41	9.62	11.76	13.36	14.97	16.57
	07-23-93 ~ 05-23-06	9.62	11.76	14.43	17.11	19.78	22.45
	10-08-92 ~ 05-23-06	14.43	17.64	19.25	20.32	23.52	24.59
	12-24-91 ~ 05-23-06	16.57	18.71	19.78	21.39	23.52	24.06
	05-24-91 ~ 05-23-06	20.32	20.85	21.39	22.99	24.59	27.27

true median at a rate faster than $\sqrt{n}$ rate when the density is infinite at the median.

The test has some useful features for empirical applications. In particular, it is free from other nuisance parameters, does not rely on particular technical conditions such as differentiability or continuity of the underlying density function, is applicable to a wide class of densities, and can be used in a time series regression context.

Empirical application of the test to stock returns of leading companies across U.S. industry is conclusive and provides strong evidence in support of infinite density at the median as a new significant empirical characteristic for stock return distributions. A significant number of the companies considered in the empirical analysis conducted here reject the null hypothesis of a finite density in favor of infinite density at the median. One implication of this finding is that data analysis in financial econometrics that relies on distributions with finite density at the median, such as t distributions and mixtures of normals, will inevitably involve some distributional misspecification in the presence of infinite density.

APPENDIX A: PROOFS

Lemma 4. Let $\tilde{\mathbf{A}}_n = n^{-1} \sum_{t=1}^{n} \delta_n^{-1} K(\delta_n^{-1} \varepsilon_t) \mathbf{X}_t \mathbf{X}_t'$. If $n^{-1} \times \sum_{t=1}^{n} E\|\mathbf{X}_t\|^3$ is uniformly bounded and $\sqrt{n} \delta_n \to \infty$, then under the conditions for Theorem 2, we have $\hat{\mathbf{A}}_n - \tilde{\mathbf{A}}_n \to_p 0$.

Proof. The proof is straightforward because

$$\|\hat{\mathbf{A}}_n - \tilde{\mathbf{A}}_n\| \leq \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\delta_n} \left| K\left(\frac{\hat{\varepsilon}_t}{\delta_n}\right) - K\left(\frac{\varepsilon_t}{\delta_n}\right) \right| \|\mathbf{X}_t\|^2$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \delta_n^{-1} \dot{K}(\delta_n^{-1} \varepsilon_t) \|\mathbf{X}_t\|^3 \cdot \|\boldsymbol{\Delta}_n\|,$$

where $\boldsymbol{\Delta}_n := \delta_n^{-1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$. The right-hand side is $o_p(1)$ because the first term is $O_p(1)$, which can be shown by taking expectation (first conditional on $\mathcal{F}_t$ for each t and then averaging unconditionally), and noting that $\|\boldsymbol{\Delta}_n\|$ is $o_p(1)$ because $\boldsymbol{\Delta}_n = (n^{1/2} \delta_n)^{-1} \sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$.

Lemma 5. Suppose that the assumptions for Lemma 4 hold. Assume further that $f^e(0) < \infty$, $f^e(s)$ is continuous in a neighborhood of zero, and $n^{-1} \sum_{t=1}^{n} E\|\mathbf{X}_t\|^4$ is uniformly bounded. Let $\tilde{\mathbf{A}}_n^* = n^{-1} \sum_{t=1}^{n} f_t(0) \mathbf{X}_t \mathbf{X}_t'$. Then $\tilde{\mathbf{A}}_n - \tilde{\mathbf{A}}_n^* \to_p 0$.

Proof. Note that $\tilde{\mathbf{A}}_n - \tilde{\mathbf{A}}_n^* = \frac{1}{n} \sum_{t=1}^{n} [\delta_n^{-1} K(\delta_n^{-1} \varepsilon_t) - f_t(0)] \times \mathbf{X}_t \mathbf{X}_t' = \frac{1}{n} \sum_{t=1}^{n} \mathbf{W}_{nt}$. We have

$$\frac{1}{n} \sum_{t=1}^{n} [\mathbf{W}_{nt} - E(\mathbf{W}_{nt}|\mathcal{F}_t)]$$

$$= \frac{1}{n} \sum_{t=1}^{n} \delta_n^{-1} \left[K(\delta_n^{-1} \varepsilon_t) - E[K(\delta_n^{-1} \varepsilon_t)|\mathcal{F}_t] \right] \mathbf{X}_t \mathbf{X}_t',$$

which is the average of an MDA. Its variance is bounded by

$$\frac{1}{n^2} \sum_{t=1}^{n} \delta_n^{-2} E K(\delta_n^{-1} \varepsilon_t)^2 \|\mathbf{X}_t\|^4 \leq \frac{1}{n \delta_n^2} (\max K^2) \frac{1}{n} \sum_{t=1}^{n} E\|\mathbf{X}_t\|^4$$

$$\to 0,$$

if $n \delta_n^2 \to 0$. So far we showed that $n^{-1} \sum_{t=1}^{n} \mathbf{W}_{nt} = n^{-1} \times \sum_{t=1}^{n} E(\mathbf{W}_{nt}|\mathcal{F}_t) + o_p(1)$. Next, when $f_t(0)$ is finite,

$$E(\mathbf{W}_{nt}|\mathcal{F}_t) = \int [\delta_n^{-1} K(\delta_n^{-1} s) f_t(s) - f_t(0)] \, ds \, \mathbf{X}_t \mathbf{X}_t'$$

$$= \int [f_t(s) - f_t(0)] \delta_n^{-1} K(\delta_n^{-1} s) \, ds \, \mathbf{X}_t \mathbf{X}_t'$$

$$= \int [f_t(\delta_n s) - f_t(0)] K(s) \, ds \, \mathbf{X}_t \mathbf{X}_t',$$
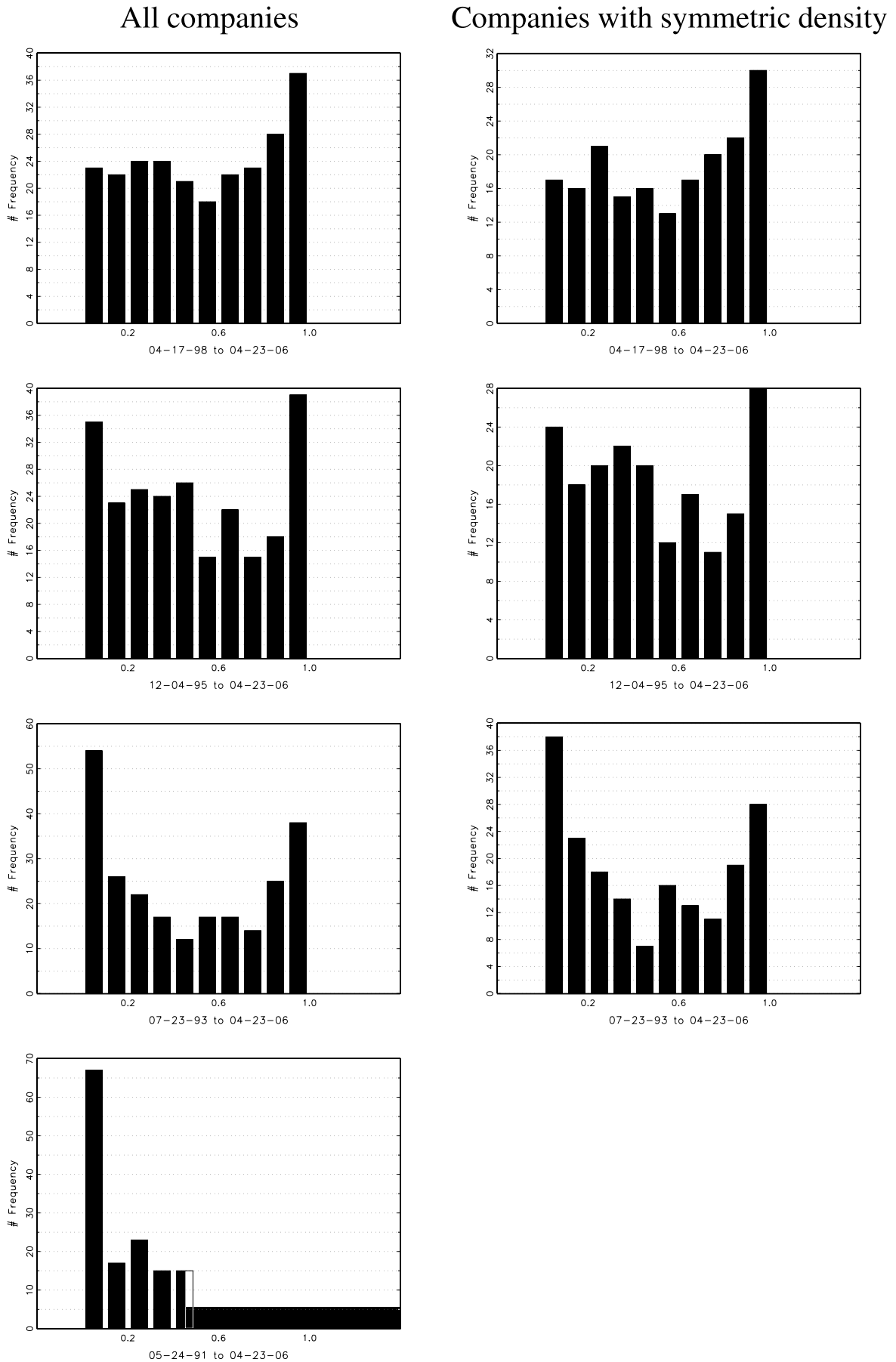
All companies

Companies with symmetric density

Figure 1. Histograms of p-values (left-tailed). Model: $r_{i,t} = \alpha_{i,0} + \beta_{i,0} r_{i,t-1} + \varepsilon_{i,t}$.

so $n^{-1} \sum_{t=1}^{n} E(\mathbf{W}_{nt}|\mathcal{F}_t)$ is integrable, and it converges to zero because

$$E \left\| \frac{1}{n} \sum_{t=1}^{n} E(\mathbf{W}_{nt}|\mathcal{F}_t) \right\|$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} E\|E(\mathbf{W}_{nt}|\mathcal{F}_t)\|$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} E \int |f_t(\delta_n s) - f_t(0)| K(s) \, ds \, \|\mathbf{X}_t\|^2$$

$$\to 0.$$

The result follows immediately.

Proof of Theorem 3(i). Under the null, Lemmas 4 and 5 imply that $\hat{\mathbf{A}}_n - \tilde{\mathbf{A}}_n^* \to_p 0$. The result follows from Theorem 2 because $\tilde{\mathbf{A}}_n^* \to_p \mathbf{A}$.

To handle the case under the alternative hypothesis, we need some technical lemmas. We start with the following.

Lemma 6. Under Assumption B(i) and (ii), if $L(\cdot)$ is continuous, nonnegative and integrates to a strictly positive number over $(-\infty, \infty)$, then for n sufficiently large, $0 < M_0 \leq \frac{\sqrt{n}}{a_n} \int_{-\infty}^{\infty} L(x) f^e(x/a_n) \, dx \leq M_1 < \infty$.

Note that the bounds M_0 and M_1 depend on $L(\cdot)$.

Proof of Lemma 6. We prove the result by considering integration over the positive domain only as the negative domain can be treated similarly. Define $0 = b_0 \leq b_1 \leq b_2 \leq \cdots$ such that $b_k = \inf\{x \geq 0 : L(x) \leq 2^{-k}\}$. Then for $x \in [b_k, b_{k+1}]$, $L(x) \in [2^{-(k+1)}, 2^{-k}]$, so that $\frac{1}{2} A_n^* \leq \frac{\sqrt{n}}{a_n} \int_0^{\infty} L(x) f^e(x/a_n) \, dx \leq A_n^*$, where $A_n^* := \frac{\sqrt{n}}{a_n} \sum_{k=0}^{\infty} \frac{1}{2^k} \int_{b_k}^{b_{k+1}} f^e(x/a_n) \, dx$, which is

$$A_n^* = \sqrt{n} \sum_{k=0}^{\infty} \frac{1}{2^k} [F^e(b_{k+1}/a_n) - F^e(b_k/a_n)]$$

$$= \sum_{k=0}^{\infty} 2^{-k} [\psi_n(b_{k+1}) - \psi_n(b_k)] = \sum_{k=1}^{\infty} \frac{\psi_n(b_k)}{2^k},$$

by change of variables. Here $\psi_n(t) = \sqrt{n}[F^e(t/a_n) - F^e(0)] = (\sqrt{n}/a_n)\bar{f}^e(t/a_n)t$, and therefore, by Assumption B(i), eventually as $n \to \infty$, $\psi_n(t)/\psi_n(1) = \bar{f}^e(t/a_n)t/\bar{f}^e(1/a_n) \leq \tilde{M}t$, if $t \geq 1$. Thus, for n large enough,

$$\sum_{k=1}^{\infty} \frac{\psi_n(b_k)}{2^k} = \sum_{k:b_k \leq 1} \frac{\psi_n(b_k)}{2^k} + \sum_{k:b_k > 1} \frac{\psi_n(b_k)}{2^k}$$

$$\leq \sum_{k:b_k \leq 1} \frac{\psi_n(1)}{2^k} + \tilde{M} \psi_n(1) \sum_{k=1}^{\infty} \frac{b_k}{2^k},$$

which is eventually (as $n \to \infty$) bounded from above by $M_1 = \psi(1)(2 + \tilde{M}L^*) < \infty$ because $\psi_n(1) \to \psi(1)$ and $\sum_{k=1}^{\infty} b_k/2^k \leq \int_0^{\infty} L(x) \, dx = L^*/2$, where $L^* = \int L(x) \, dx$. For the lower bound, $\psi(b_k)$ is strictly positive for some finite k by Assumption B(ii), implying that $\frac{1}{2} A_n^* > 0$ for all n large enough.

Lemma 7. Given the assumptions for Theorems 2 and 3, if Equation (10) holds in addition, then for any uniformly bounded nonnegative function $L(\cdot)$ which is symmetric around zero and nonincreasing over the positive domain, we have $\frac{\sqrt{n}}{a_n} \int L(x) f^e(\delta_n x) \, dx \to 0$.

Proof. We show only that

$$A_n^{**} = \frac{\sqrt{n}}{a_n} \int_0^{\infty} L(x) f^e(\delta_n x) \, dx = o(1), \quad (\text{A.1})$$

because the proof over $(-\infty, 0)$ follows in the same way by virtue of the symmetry of $L(\cdot)$. Without loss of generality, we let $L(0) = 1$ [otherwise, divide $L(x)$ by $L(0)$]. Let m satisfy $a_m \delta_n = 1$. (Again note that the domain of a_n is extended to $\mathbb{R}_+$ and that $m \to \infty$ as $n \to \infty$.) Then clearly $\frac{\sqrt{m}}{a_m} \int_0^{\infty} L(x) f^e(\delta_n x) \, dx = \frac{\sqrt{m}}{a_m} \int_0^{\infty} L(x) f^e(a_m^{-1} x) \, dx = O(1)$ by Lemma 6, and therefore,

$$A_n^{**} = \left[\frac{\sqrt{n}/a_n}{\sqrt{m}/a_m} \right] \frac{\sqrt{m}}{a_m} \int_0^{\infty} L(x) f^e(a_m^{-1} x) \, dx$$

$$= \left[\frac{\sqrt{n}/a_n}{\sqrt{m}/a_m} \right] O(1). \quad (\text{A.2})$$

Since both n and m tend to infinity, both $\psi_n(1)$ and $\psi_m(1)$ converge to $\psi(1)$, we have

$$\frac{\psi_n(1)}{\psi_m(1)} = \frac{\sqrt{n}/a_n}{\sqrt{m}/a_m} \cdot \frac{\bar{f}^e(a_n^{-1})}{\bar{f}^e(a_m^{-1})} \to 1. \quad (\text{A.3})$$

But because $a_m \delta_n = 1$ and $a_n \delta_n \to \infty$, we have $a_n^{-1}/a_m^{-1} = a_m/a_n \to 0$, and thus by Equation (10) we have $\bar{f}^e(a_n^{-1})/\bar{f}^e(a_m^{-1}) \to \infty$, which, by Equation (A.3), implies that $(\sqrt{n}/a_n)/(\sqrt{m}/a_m) \to 0$. This last result and Equation (A.2) imply Equation (A.1), and thereby complete the proof.

Lemma 8. Under Equation (10), $(\sqrt{n}/a_n)\hat{\mathbf{A}}_n \to_p 0$.

Proof. Due to Lemma 4, it suffices to show that $(\sqrt{n}/a_n)\tilde{\mathbf{A}}_n \to_p 0$. [Note that $\sqrt{n}/a_n = O(1)$.] We shall show that $(\sqrt{n}/a_n)E\|\tilde{\mathbf{A}}_n\| \to 0$. We note that $(\sqrt{n}/a_n)E\|\tilde{\mathbf{A}}_n\|$ is bounded by

$$\frac{1}{n} \sum_{t=1}^{n} E\left(\frac{\sqrt{n}}{a_n} \int K(s) f_t(\delta_n s) \, ds \, \|\mathbf{X}_t\|^2 \right)$$

$$\leq \frac{\sqrt{n}}{a_n} \int K(s) f^e(\delta_n s/\sigma_*) \, ds \cdot \frac{1}{n} \sum_{t=1}^{n} E\sigma_t^{-1} \|\mathbf{X}_t\|^2.$$

The second term is obviously $O_p(1)$, and the first term converges to zero by Lemma 7.

Proof of Theorem 3(ii). We have $B_n = [a_n(\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n})]' \times [(\sqrt{n}/a_n)\hat{\mathbf{A}}_n]\hat{\mathbf{C}}^{-1}[(\sqrt{n}/a_n)\hat{\mathbf{A}}_n][a_n(\tilde{\boldsymbol{\beta}}_{1n} - \tilde{\boldsymbol{\beta}}_{2n})] \Rightarrow 0$ by Theorem 2 and Lemma 8.

APPENDIX B: LIST OF COMPANIES INCLUDED IN THE EMPIRICAL APPLICATION

The following lists the companies of each industry included for our empirical analysis. The number of companies is provided in parentheses for each industry.

Aerospace & Defense (5): Goodrich, General Dynamics, Alliant Techsystems, Moog, DRS Technologies.

Banking (13): Citigroup, Synovus Finl, Zions Bancorp., Wells Fargo, Popular, M&T Bank, AmSouth Bancorp., Marshall & Ilsley, Golden West Finl., Wachovia, Commerce Bancorp., Bank of America, Compass Bancshares.

Business Services & Supplies (9): Automatic Data, Paychex, Avery Dennison, Robert Half Intl, Waste Management, ServiceMaster, Manpower, Equifax, World Fuel Services.

Capital Goods (12): Danaher, Valmont Inds., Ingersoll-Rand, Timken, Donaldson, Cummins, JLG Indst., Caterpillar, Ametek, Rockwell Automation, Genlyte Group, Oshkosh Truck.

Chemicals (10): Ecolab, Engelhard, Rohm and Haas, Dow Chemical, Airgas, Sigma-Aldrich, Air Prods & Chems, Valspar, Lubrizol, Georgia Gulf.

Conglomerates (6): General Electric, Dover, Emerson Electric, Fortune Brands, United Technologies, ITT Inds.

Construction (9): Jacobs Engineering, Standard Pacific, Toll Brothers, Lennar, Pulte Homes, MDC Holdings, KB Home, Ryland Group, Meritage Homes.

Consumer Durables (13): Harley-Davidson, Toyota Motor, Honda Motor, Nissan Motor, Volvo, Brunswick, Johnson Controls, Black & Decker, Genuine Parts, Applied Inds., Paccar, Toro, Thor Inds.

Diversified Financials (4): Charles Schwab, Berkshire Hathaway, Franklin Resources, Legg Mason.

Drugs & Biotech (5): Abbott Laboratories, Allergan, Amgen, Johnson & Johnson, Barr Pharmaceuticals.

Food Drink & Tobacco (12): Coca-Cola, General Mills, PepsiCo, Wm Wrigley Jr., Seaboard, PepsiAmericas, McCormick & Co, Hormel Foods, Kellogg, Dean Foods, Constellation Brands, Pilgrim's Pride.

Food Markets (4): Sysco, Weis Markets, Ruddick, Casey's General Store.

Health Care Equipment & Services (9): Universal Health, Biomet, Stryker, Coventry Health Care, Becton Dickinson, Omnicare, Varian Medical Systems, Humana, UnitedHealth Group.

Hotels, Restaurants & Leisure (5): Brinker Intl., Hilton Hotels, Applebee's Intl., MGM Mirage, Carnival Corp.

Household & Personal Products (8): Timberland, Procter & Gamble, Liz Claiborne, Oxford Indst., Alberto-Culver, NIKE, Church & Dwight, Phillips-Van Heusen.

Insurance (10): Chubb, Aflac, Cincinnati Finl., Old Republic Intl., Mercury General, White Mountains Ins., First American, Commerce Group Inc., Selective Ins., Zenith National Ins.

Materials (12): Barrick Gold, Bemis, Worthington Inds., Phelps Dodge, Inco, Harsco, Massey Energy, Nucor, Commercial Metals, Steel Technologies, Quanex, Cleveland-Cliffs.

Media (9): Comcast, Walt Disney, WPP Group, Omnicom Group, EW Scripps, Meredith, RR Donnelley & Sons, McGraw-Hill Cos., Banta.

Oil & Gas Operations (16): Nabors Industries, Baker Hughes, Noble Corp., Marathon Oil, Smith International, Ashland, Apache, EOG Resources, Holly, BJ Services, Murphy Oil, Tesoro, Valero Energy, Sunoco, Western Gas Resources, Occidental Petroleum.

Retailing (17): CVS, Walgreen, Home Depot, Tiffany & Co., Dollar General, Genesco, Sherwin-Williams, Claire's Stores, Lowe's Cos., Fastenal, Staples, AutoNation, Best Buy, Williams-Sonoma, Ross Stores, Nordstrom, Michaels Stores.

Semiconductors (7): Intel, Maxim Integrated Prods, Altera, Linear Technology, Texas Instruments, KLA-Tencor, Lam Research.

Software & Services (6): Microsoft, Adobe Systems, Fiserv, Electronic Arts, CACI International, Autodesk.

Technology Hardware & Equipment (10): EMC, Cisco Systems, Dell, Motorola, Benchmark Electronics, Canon, Harris, Western Digital, Harman Intl., Apple Computer.

Telecommunications Services (5): Verizon Commun., BellSouth, CenturyTel, Sprint Nextel, Alltel.

Transportation (8): Southwest Airlines, SkyWest, FedEx, CSX, Werner Enterprises, Expeditors Intl., Burlington Santa Fe, JB Hunt Transport.

Utilities (18): National Fuel Gas, Nicor, Constellation Energy, Laclede Group, OGE Energy, Scana, MDU Resources, New Jersey Resources, Exelon, AGL Resources, FirstEnergy, Edison Intl., Sempra Energy, Wisconsin Energy, WPS Resources, Questar, Equitable Resources, UGI.

ACKNOWLEDGMENTS

REFERENCES

Bassett, G., and Koenker, R. (1978), "Asymptotic Theory of Least Absolute Error Regression," *Journal of American Statistical Association*, 73, 618–622. [283]

Bloomfield, P., and Steiger, W. (1983), *Least Absolute Deviations: Theory, Applications, and Algorithms*, Boston: Birkhäuser. [283]

Bollerslev, T., Engle, R. F., and Nelson, D. B. (1994), "ARCH Models" in *The Handbook of Econometrics*, Vol. 4, eds. R. F. Engle and D. McFadden, Amsterdam: North-Holland, pp. 2959–3038. [284,288]

Cho, J., Han, C., and Phillips, P. C. B. (2010), "LAD Asymptotics for Conditionally Heteroskedastic Time-Series Data With Possibly Infinite Error Densities," *Econometric Theory*, to appear. [284,286,289]

Cohen, K., Hawawini, G., Maier, S., Schwartz, R., and Whitcomb, D. (1983a), "Friction in the Trading Process and the Estimation of Systematic Risk," *Journal of Financial Economics*, 12, 263–278. [288]

—— (1983b), "Estimating and Adjusting for the Intervalling-Effect Bias in Beta," *Management Science*, 29, 135–148. [288]

Dimson, E. (1979), "Risk Measurement When Shares Are Subject to Infrequent Trading," *Journal of Financial Ecoinomics*, 7, 197–226. [288]

Fama, E. F. (1965), "The Behavior of Stock Market Prices," *Journal of Business*, 38, 34–105. [282,283,288]

Forbes.com Inc. (2005), "America's Best Big Companies," available at *http://www.forbes.com/lists*. [288]

Gabaix, X., Gopikrishnan, P., Plerou, V., and Stanley, H. E. (2003), "A Theory of Power Law Distributions in Financial Market Fluctuations," *Nature*, 423, 267–270. [282]

Geyer, C. (1996), "On the Asymptotics of Convex Stochastic Optimization," unpublished manuscript, School of Statistics, University of Minnesota. [285]

Ibragimov, R. (2007), "Efficiency of Linear Estimators Under Heavy Tailedness: Convolutions of α-Symmetric Distributions," *Econometric Theory*, 23, 501–517. [282]

Ibragimov, R., and Walden, J. (2007), "The Limits of Diversification When Losses May Be Large," *Journal of Banking and Finance*, 31, 2551–2569. [282]

Kim, T., and White, H. (2003), "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regression," in *Maximum Likelihood Esitmation of Misspecified Models: Twenty Years Later*, eds. T. Fomby and R. Hill, New York: Elsevier, pp. 107–132. [283]

Knight, K. (1998), "Limiting Distributions for L_1 Regression Estimators Under General Conditions," *The Annals of Statistics*, 26, 755–770. [282-287,289]

———— (1999), "Asymptotics for L_1-Estimators of Regression Parameters Under Heteroscedasticity," *Canadian Journal of Statistics*, 27, 497–507. [284,286]

Koenker, R. (2008), "quantreg: Quantile Regression," R package version 4.23. Available at *http://www.r-project.org*. [287]

Koenker, R., and Zhao, Q. (1996), "Conditional Quantile Estimation and Inference for ARCH Models," *Econometric Theory*, 12, 793–813. [283]

Lo, A., and MacKinlay, C. (1988), "Stock Market Prices Do Not Follow Random Walks: Evidence From a Simple Specification Test," *Review of Financial Studies*, 1, 41–66. [288]

———— (1990), "An Econometric Analysis of Nonsynchronous Trading," *Journal of Econometrics*, 45, 181–212. [288]

Mandelbrot, B. (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419. [282,283,288]

McWilliams, T. (1990), "A Distribution-Free Test for Symmetry Based on a Statistic," *Journal of the American Statistical Association*, 85, 1130–1133. [288]

Nelson, D. (1991), "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370. [288]

Phillips, P. C. B. (1991), "A Shortcut to LAD Estimator Asymptotics," *Econometric Theory*, 7, 450–463. [283]

Pollard, D. (1991), "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186–199. [283]

Quenouille, M. H. (1956), "Notes on Bias in Estimation," *Biometrika*, 43, 353–360. [284]

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. Available at *http://www.r-project.org*. [287]

Rogers, A. (2001), "Least Absolute Deviations Regression Under Nonstandard Conditions," *Econometric Theory*, 17, 820–852. [287]

Scholes, M., and Williams, J. (1977), "Estimating Betas From Nonsynchronous Data," *Journal of Financial Economics*, 5, 309–327. [288]

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, New York: Wiley. [287]

Silverman, B. W. (1986), *Density Estimation*, London: Chapman & Hall. [287]

Zinde-Walsh, V. (2008), "Kernel Estimation When Density Does Not Exist," *Econometric Theory*, 24, 696–725. [283]