Econometrica, Vol. 71, No. 1 (January, 2003), 285–317

INFERENCE IN ARCH AND GARCH MODELS WITH HEAVY-TAILED ERRORS

BY PETER HALL AND QIWEI YAO[1]

ARCH and GARCH models directly address the dependency of conditional second moments, and have proved particularly valuable in modelling processes where a relatively large degree of fluctuation is present. These include financial time series, which can be particularly heavy tailed. However, little is known about properties of ARCH or GARCH models in the heavy-tailed setting, and no methods are available for approximating the distributions of parameter estimators there. In this paper we show that, for heavy-tailed errors, the asymptotic distributions of quasi-maximum likelihood parameter estimators in ARCH and GARCH models are nonnormal, and are particularly difficult to estimate directly using standard parametric methods. Standard bootstrap methods also fail to produce consistent estimators. To overcome these problems we develop percentile-t, subsample bootstrap approximations to estimator distributions. Studentizing is employed to approximate scale, and the subsample bootstrap is used to estimate shape. The good performance of this approach is demonstrated both theoretically and numerically.

KEYWORDS: Autoregression, bootstrap, dependent data, domain of attraction, financial data, limit theory, percentile-t bootstrap, quasi-maximum likelihood, semiparametric inference, stable law, studentize, subsample bootstrap, time series.

1. INTRODUCTION

IN CONTRAST TO TRADITIONAL time series analysis, which focuses on modelling the conditional first moment, an ARCH or GARCH model takes the dependency of the conditional second moments explicitly into consideration. See, for example, Engle (1982), Bollerslev (1986), and Taylor (1986). The practical motivation for doing so lies in the increasingly important need to explain and model risk and uncertainty in, for example, financial time series. Early successes of ARCH/GARCH modelling of financial time series were confined to the case of Normal errors, for which an explicit conditional likelihood function is readily available to facilitate estimation of parameters in the model. Investigation of non-Normal cases has been partly driven by empirical evidence that financial time series can be very heavy-tailed (e.g., Mittnik, Rachev, and Paolella (1998); Mittnik and Rachev (2000)).

This leads to semiparametric ARCH and GARCH models in which the error distributions are unknown (Engle and Gonzalez-Rivera (1991)). Nevertheless, conditional Gaussian likelihood functions still motivate parameter estimators, which might be called quasi-maximum likelihood estimators. See, for example, Bollerslev and Wooldridge (1992), and Chapter 4 of Gouriéroux (1997). Other

[1] We are grateful to two reviewers for their helpful comments.

methods include adaptive estimation for ARCH models (Linton (1993)), and Whittle estimation for a general ARCH (∞) process (Giraitis and Robinson (2001)).

It is known that, provided the error distribution has finite fourth moment, quasi-maximum likelihood estimators are asymptotically Normally distributed in the case of an ARCH model (Weiss (1986)), and also for a GARCH(1, 1) model (Lee and Hansen (1994); Lumsdaine (1996)). However, little more than consistency is available in other settings, least of all in the case of relatively heavy-tailed error distributions that are of particular interest in applications to finance. In this paper we develop a very general account of theory for estimators in ARCH and GARCH models, paying special attention to the heavy-tailed case. There the limit distributions that arise are multivariate stable laws, and are particularly difficult to estimate directly. While this is arguably the most interesting aspect of our work, even in the case of finite fourth moment (for example, in the setting of GARCH(p, q) models with $(p, q) \neq (1, 1)$) our results are new. Moreover, it is possible to obtain Normal limiting distributions without assuming finite fourth moment.

We suggest bootstrap methods for estimating parameter distributions. Now, it is well known that in settings where the limiting distribution of a statistic is not Normal, standard bootstrap methods are generally not consistent when used to approximate the distribution of the statistic. See, for example, Mammen (1992), Athreya (1987a, 1987b), Knight (1989b), and Hall (1990). To some extent, subsampling methods can be used to overcome the problem of inconsistency. See Bickel, Götze, and van Zwet (1995) and Politis, Romano, and Wolf (1999) for recent accounts of the subsampling method. However, while this approach consistently approximates the distribution of a statistic, it does so only for a value of sample size that is smaller than the size of the actual sample. The surrogate that it uses for sample size is the size of the subsample, which has to be an order of magnitude less than the sample size. As a result, the "scale" of the distribution that is approximated by the subsample bootstrap is generally an order of magnitude larger than that for the true sample size. Therefore, a confidence or prediction procedure based directly on the subsample bootstrap can be very conservative. In the absence of an accurate method for adjusting scale, subsampling can be unattractive.

To overcome this problem we suggest a new approach based on a percentile-t form of the subsample bootstrap. The percentile-t method is usually employed in order to attain a high order of accuracy in approximations where the limiting distribution is Normal. That is not our main goal in the present setting. Instead, we use a form of the percentile-t subsample bootstrap to ensure consistent distribution estimation in a particularly wide range of settings, where the limiting distribution can be either Normal or non-Normal. We studentize primarily to determine the scale of the test statistic. The subsample bootstrap can then be employed to estimate just the shape of the distribution, rather than both shape and scale. In this way we avoid the difficulties noted in the previous paragraph.

In more regular cases, where relatively high-order moments of the error distribution are finite, conventional percentile-t methods can be developed. They differ from the techniques discussed in the present paper in that they use the standard bootstrap rather than the subsample bootstrap, and they studentize using an estimator of the square root of the covariance matrix of the vector of parameter estimators. (In this paper we studentize using a scalar quantity; the covariance matrix is generally not well defined in the heavy-tailed case.) When sufficiently high-order moments can be assumed, the method for implementing conventional bootstrap approximations closely parallels that used for the bootstrap in linear time series; see for example Bose (1988). For the sake of brevity we shall not discuss such methods further here.

Classical work on financial time series with non-Normal errors includes that of Mandelbrot (1963) and Fama (1965), although of course without the benefit of critical recent developments, particularly in extreme value analysis. Even ARCH or GARCH models with Normal errors can have heavy tails; see for example Kesten (1973), Goldie (1991), Embrechts, Klüppelberg, and Mikosch (1997), Davis and Mikosch (1998), and Mikosch and Starica (2000). Statistical aspects of financial modelling of heavy-tailed data are discussed by, for example, Shephard (1996) and Rydberg (2000).

Our results are in the spirit of some of the findings of Mikosch and Straumann (2001), who show that poor rates of convergence can occur if X_t is a GARCH process and $E(X_t^8)$ is infinite. Part (a) of Theorem 2.1 in Section 2.3 below has also been obtained by Berkes, Horváth, and Kokoszka (2001) under different conditions; see also Comte and Lieberman (2000).

2. MAIN THEORETICAL RESULTS

2.1. *Model*

Assume $X_t = \sigma_t \epsilon_t$ for $-\infty < t < \infty$, where the random variables ϵ_t are independent and identically distributed with zero mean and unit variance, ϵ_t is independent of $\{X_{t-i}, i \geq 1\}$,

$$\sigma_t^2 = c + \sum_{i=1}^{p} a_i X_{t-i}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2, \tag{2.1}$$

$c > 0$, $a_i \geq 0$, $b_j \geq 0$, $p \geq 0$, and $q \geq 0$, the latter two quantities of course being integers. If $q = 0$, then the model is of autoregressive conditional heteroscedastic, or ARCH, type. If $q \geq 1$ it is of generalized ARCH, or GARCH, form. To avoid pathological cases we shall assume throughout that $p \geq 1$ and $a_p > 0$, and $b_q > 0$ when $q \geq 1$.

It is known that a necessary and sufficient condition for the process $\{X_t, -\infty < t < \infty\}$ to be strictly stationary with finite mean square is

$$\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j < 1; \tag{2.2}$$

see Nelson (1990), Bougerol and Picard (1992), and Giraitis, Kokoszka, and Leipus (2000), and also Bollerslev (1986). In this case, $E(X_t) = 0$ and

$$E(X_t^2) = c\left(1 - \sum_i a_i - \sum_j b_j\right)^{-1}.$$

We shall assume (2.2) throughout, and that the process is in its stationary distribution. In this case it may be shown that (2.1) implies

$$\sigma_t^2 = \frac{c}{1 - \sum_j b_j} + \sum_{i=1}^{p} a_i X_{t-i}^2 + \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} X_{t-i-j_1-\cdots-j_k}^2, \tag{2.3}$$

where the multiple series vanishes if $q = 0$. Since each a_i and b_j is nonnegative, and since the expected value of the multiple series is finite, then the series converges with probability 1. In this notation we may write $\sigma_t = \sigma_t(a, b, c)$, expressing σ_t as a function of $a = (a_1, \ldots, a_p)$, $b = (b_1, \ldots, b_q)$, c, and the data X_t.

In practice the data are observed only over a finite time interval, say $1 \le t \le n$, and σ_t^2 has to be approximated by a truncated series. For simplicity and clarity, however, we shall assume for the present that insofar as calculation of σ_t is concerned we may use values of X_u for $-\infty < u \le n$, even though in other respects our inference will be confined to X_t for $1 \le t \le n$. The contrary case will be discussed in Section 2.4. There we shall show that our main results do not change when an appropriately truncated approximation is employed.

2.2. *Estimators*

Conditional maximum likelihood estimators in problems of this type were discussed by Engle (1982) and Bollerslev (1986). They can be motivated by temporarily assuming that the errors ϵ_t are Gaussian, which would imply that if (a, b, c) took their true values, then the variables $X_t/\sigma_t(a, b, c)$ would be independent and identically $N(0, 1)$. Therefore, without requiring Gaussian errors from this point on, it is suggested that we minimize

$$L(a, b, c) = \sum_{t=1}^{n} \left\{ \frac{X_t^2}{\sigma_t(a, b, c)^2} + \log \sigma_t(a, b, c)^2 \right\} \tag{2.4}$$

with respect to the $r = p + q + 1$ variables in (a, b, c). There is an extensive literature on using Normal-based methods for non-Normal data, including data with infinite variance. See, for example, Cline (1989) and references cited therein.

The derivatives of $L(a, b, c)$ may be deduced from the formulae

$$\frac{\partial \sigma_t^2}{\partial c} = \frac{1}{1 - \sum_j b_j}, \tag{2.5}$$

$$\frac{\partial \sigma_t^2}{\partial a_i} = X_{t-i}^2 + \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} X_{t-i-j_1-\cdots-j_k}^2, \tag{2.6}$$

$$(2.7)\qquad \frac{\partial\sigma_t^2}{\partial b_j} = \frac{c}{(1-\sum_i b_i)^2} + \sum_{i=1}^{p} a_i \sum_{k=0}^{\infty}(k+1) \times \sum_{j_1=1}^{q}\cdots\sum_{j_k=1}^{q} b_{j_1}\cdots b_{j_k} X^2_{t-i-j-j_1-\cdots-j_k}.$$

To interpret (2.7) relative to (2.3), note that in the latter result we could have dropped the second term on the right-hand side if we had summed from $k=0$ to ∞, instead of from $k=1$ to ∞, in the multiple series.

2.3. *General Central Limit Theorem*

Let $U = U(a,b,c)$ denote the r-vector of first derivatives $\sigma_1^2 = \sigma_1(a,b,c)^2$ with respect to the components of a, b, and c. Put $M = E_0(\sigma_1^{-4}UU^T)$, an $r \times r$ matrix, where E_0 denotes expectation when a, b, and c take their true values a^0, b^0, and c^0. It may be deduced from (2.3) and (2.5)–(2.7) that if none of $a_1, \dots, a_p$ vanishes, and if (for $q \ge 1$) none of $b_1, \dots, b_q$ vanishes, then each component of $\sigma_1^{-2}U$ has all its moments finite, and in particular there exists a constant $C > 0$ such that $E(\sigma_1^{-2}\|U\|)^\nu \le \nu!\, C^\nu$ for all integers $\nu \ge 1$, where $\|\cdot\|$ denotes the Euclidean norm. See Section 2.5 for discussion. Hence, the existence and finiteness of the components of M are guaranteed. We shall assume M is nonsingular. Let $(\hat a, \hat b, \hat c)$ be any local minimum of $L(a,b,c)$ that occurs within radius η of (a^0, b^0, c^0), for sufficiently small but fixed $\eta > 0$, and write $\hat\theta$ for the column vector of length r whose components are those of $(\hat a, \hat b, \hat c)$. Likewise let θ denote the column vector of components of (a,b,c), and let θ^0 be the version of θ for the true parameter values.

Recall that we assume throughout that the distribution of ϵ has zero mean and unit variance, and in particular that $E(\epsilon^2) < \infty$. When $E(\epsilon^4) = \infty$, but the distribution of ϵ^2 is in the domain of attraction of the Normal law, put

$$(2.8)\qquad H(\lambda) = E\{\epsilon^4 I(\epsilon^2 \le \lambda)\} \qquad \text{and} \qquad \lambda_n = \inf\{\lambda > 0 : nH(\lambda) \le \lambda^2\}.$$

The function H is slowly varying at infinity; see Section IX.8 of Feller (1966).

Next consider the case where the distribution of ϵ^2 is in the domain of attraction of a stable law with exponent $\alpha \in [1,2)$. Redefine λ_n by

$$(2.9)\qquad \lambda_n = \inf\{\lambda > 0 : nP(\epsilon^2 > \lambda) \le 1\}.$$

The properties of a stable law imply that λ_n is regularly varying at infinity with exponent $1/\alpha$. That is, $\lambda_n = n^{1/\alpha}\ell(n)$ where ℓ is a slowly varying function, meaning that an appropriate extension of ℓ to the real line satisfies $\ell(cn)/\ell(n) \to 1$ as $n \to \infty$, for each $c > 0$. Examples of slowly varying functions include polynomials in the logarithm function, or in iterates of that function.

Let $Y_1, Y_2, \dots$ represent the infinite extension of the multicomponent joint extreme-value distribution of the first type, with exponent α. That is, for each k the distribution of $(Y_1, \dots, Y_k)$ is the limiting joint distribution of the k largest values of a sample of size n drawn from a distribution in the domain of attraction of the first type of extreme-value distribution, after appropriate normalization for scale. As is standard, we assume the normalization is chosen so that Y_1 has distribution function $\exp(-y^{-\alpha})$ for $y > 0$. Then it may be shown that for each $k \geq 1$ the marginal distribution function of Y_k is given by

$$F_k(y) = \exp(-y^{-\alpha}) \sum_{j=0}^{k-1} y^{-j\alpha}/j!, \quad y > 0. \tag{2.10}$$

Hall (1978a) formulated a representation of the distribution of the full process $Y_1, Y_2, \ldots$.

Let $V_1, V_2, \ldots$ be independent and identically distributed as $\sigma_1^{-2} M^{-1} U$, where we take $\theta = \theta^0$, and let them be independent also of $Y_1, Y_2, \ldots$.

When $1 < \alpha < 2$, put

$$W_0 = \sum_{k=1}^{\infty} \{Y_k V_k - E(Y_k) E(V_1)\}, \tag{2.11}$$

and for $\alpha = 1$ let

$$W_1 = Y_1 V_1 + \sum_{k=2}^{\infty} \{Y_k V_k - E(Y_k) E(V_1)\} \tag{2.12}$$

and $\mu_n = n\lambda_n^{-1} E\{\epsilon^2 I(\epsilon^2 > \lambda_n)\}$. (Note that $E(Y_1) = \infty$ when $\alpha = 1$; hence the need to work with both W_0 and W_1.) Convergence of the infinite series at (2.11) and (2.12) is guaranteed; see part (e) of the theorem below. When $\alpha = 1, \mu_n$ is an unbounded, slowly varying function of n. Let γ denote Euler's constant.

By Theorem 1.4.5 of Samorodnitsky and Taqqu (1994, pp. 28), the marginal distributions of W_0 and W_1 are stable with exponent α. It follows from that property, and characterizations of multivariate stable laws, that the multivariate distributions of W_0 and W_1 are multivariate stable, although the characteristic functions in this setting are awkward to write down and the distributions are more difficult still to estimate directly. The characteristic functions of the jth components of W_0 and W_1 are respectively

$$\begin{aligned} E\{\exp(itW_0^{(j)})\} &= \exp[-s_j^{\alpha}|t|^{\alpha}\{1 - i\beta_j \operatorname{sgn} t \tan(\alpha\pi/2)\}], \\ E\{\exp(itW_1^{(j)})\} &= \exp[i\phi_j t - s_j|t|\{1 + i\beta_j(2/\pi) \operatorname{sgn} t \log|t|\}], \end{aligned} \tag{2.13}$$

where $\operatorname{sgn} t$ denotes the sign of t (taken equal to 1 if $t = 0$), $s_j^{\alpha} = E|V^{(j)}|^{\alpha}/C_{\alpha}$, $\beta_j E|V^{(j)}|^{\alpha} = E\{|V^{(j)}|^{\alpha} \operatorname{sgn} V^{(j)}\}$, $V^{(j)}$ denotes a random variable with the distribution of the jth component of $V_k, C_1 = 2/\pi, C_{\alpha} = (1-\alpha)/\{\Gamma(2-\alpha)\cos(\alpha\pi/2)\}$ for $0 < \alpha < 1$, and ϕ_j may be deduced from Samorodnitsky and Taqqu (1994, pp. 28–29). In interpreting s_j and β_j in the case of (2.13), note that $\alpha = 1$ there.

THEOREM 2.1: *Assume M is nonsingular, that $p \geq 1$, that all of $a_1, \ldots, a_p$ are nonzero, that if $q \geq 1$ then all of $b_1, \ldots, b_q$ are nonzero, that $c > 0$, and that η*

(employed in the definition of $\hat{\theta}$) is strictly positive and sufficiently small. (a) If $E(\epsilon^4) < \infty$, then $n^{1/2}(\hat{\theta} - \theta^0)$ is asymptotically Normally distributed with zero mean and variance matrix $\tau^2 M^{-1}$, where $\tau^2 = E(\epsilon^4) - 1$. (b) If $E(\epsilon^4) = \infty$ but the distribution of ϵ^2 is in the domain of attraction of the Normal law, then $n\lambda_n^{-1}(\hat{\theta} - \theta^0)$ is asymptotically Normally distributed with zero mean and variance matrix M^{-1}. (c) If the distribution of ϵ^2 is in the domain of attraction of a stable law with exponent $\alpha \in (1, 2)$, then $n\lambda_n^{-1}(\hat{\theta} - \theta^0)$ converges in distribution to W_0. (d) If the distribution of ϵ^2 is in the domain of attraction of a stable law with exponent $\alpha = 1$, and if $n^{-1}\lambda_n \mu_n^2 \to 0$, then

$$n\lambda_n^{-1}(\hat{\theta} - \theta^0) + \mu_n E(V_1) - \gamma E(V_1) \tag{2.14}$$

converges in distribution to W_1. (e) For $1 < \alpha < 2$ the infinite series at (2.11) converges with probability 1, and for $\alpha = 1$ the infinite series at (2.12) converges with probability 1.

A proof of the theorem is given in Section 5.1.

The assumption $n^{-1}\lambda_n \mu_n^2 \to 0$, imposed in part (d) of the theorem, is satisfied by some but not all distributions of ϵ^2 that lie in the domain of attraction of stable laws with exponent 1. When the assumption fails, relatively complex results alternative to that in part (d) of the theorem may be proved using arguments similar to those we shall give in Section 5. They require additional location-correction terms, analogues of $\mu_n E(V_1)$ at (2.14). The terms are of respective sizes $(\lambda_n/n)^{k-1}\mu_n^k$, where $1 \le k \le k_0$ and k_0 is the least integer such that $(\lambda_n/n)^{k_0-1}\mu_n^{k_0} \to 0$. (Under the constraint $n^{-1}\lambda_n\mu_n^2 \to 0$ we may take $k_0 = 2$.) If $P(\epsilon^2 > x) \sim \text{const.}\, x^{-1}(\log x)^{-1-\delta}$ as $x \to \infty$, for some $\delta > 0$, then k_0 is the least integer that strictly exceeds $1 + \delta^{-1}$.

Results related to parts (c) and (d) of the theorem, in the context of convergence of sums of independent random variables to stable laws, have been obtained by Darling (1952), Arov and Bobrov (1960), Dwass (1966), Hall (1978b), Samorodnitsky and Taqqu (1994), and Resnick (1986, 1987).

Limit theory for autoregressions, moving averages, and related process in the setting of heavy-tailed error distributions, includes that of Kanter and Steiger (1974), Hannan and Kanter (1977), Yohai and Maronna (1977), Gross and Steiger (1979), Cline and Brockwell (1985), Davis and Resnick (1985a,b, 1989, 1996), Bhansali (1988, 1995, 1997), Knight (1987, 1989a, 1993), Chan and Tran (1989), Chan (1990, 1993), Phillips (1990), Kokoszka and Taqqu (1994, 1996a, b, 1999), Mikosch and Klüppelberg (1995), Mikosch et al. (1995), Kokoszka and Mikosch (1997), Hsing (1999), Leipus and Viano (2000), and Koul and Surgailis (2001). However, results in this setting do not display the same very wide range of possible limit behavior found in the context of ARCH and GARCH models.

2.4. *Truncated Likelihood*

In practice, $\sigma_t(a, b, c)^2$ cannot be computed using the multiple series at (2.3), since X_t^2 is observed only for $1 \le t \le n$. Therefore, the likelihood at (2.4) cannot be calculated exactly. However, $\sigma_t(a, b, c)^2$ can be computed in an approximate, truncated form, as follows. For $2 \le t \le n$, define

$$\text{(2.15)} \qquad \tilde{\sigma}_t(a, b, c)^2 = \frac{c}{1 - \sum_j b_j} + \sum_{i=1}^{\min(p, t-1)} a_i X_{t-i}^2 + \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} \times X_{t-i-j_1-\cdots-j_k}^2 I(t - i - j_1 - \cdots - j_k \ge 1).$$

The indicator function here (denoted by I), and the truncation of the first series over i, ensure that the definition of $\tilde{\sigma}_t(a, b, c)^2$ uses only the data $X_1, \dots, X_t$. However, for small t the accuracy of this approximation to σ_t^2 will be severely curtailed, suggesting that when conducting inference we should avoid early terms in the series. Thus, a practicable version of L might be defined by

$$\text{(2.16)} \qquad \widetilde{L}_\nu(a, b, c) = \sum_{t=\nu}^{n} \left\{ \frac{X_t^2}{\tilde{\sigma}_t(a, b, c)^2} + \log \tilde{\sigma}_t(a, b, c)^2 \right\},$$

where the integer $\nu = \nu(n)$ diverges with n but at a rate sufficiently slow to ensure $\nu/n \to 0$ as $n \to \infty$. Our next result shows that for appropriate choice of ν, the results summarized by Theorem 2.1 continue to hold if estimators are computed using the truncated likelihood $\widetilde{L}_\nu$.

THEOREM 2.2: *Assume the initial conditions of Theorem* 2.1, *as well as the additional conditions for any one of parts* (*a*)–(*c*) *of that theorem. Suppose too that* $\nu = \nu(n)$ *satisfies* $\nu/\log n \to \infty$ *and* $\nu/n \to 0$ *as* $n \to \infty$. *Then, if the estimator* $\hat{\theta}$ *is defined by minimizing* $\widetilde{L}_\nu$ *defined at* (2.16), *instead of L at* (2.4), *the respective conclusions of parts* (*a*)–(*c*) *of Theorem* 2.1 *hold*.

There is also a version of part (d) of Theorem 2.1 in the context of truncated likelihood. However, it involves a new centering constant that depends on ν. For brevity we do not give it here. Our development of bootstrap methods in Section 3 will be founded on estimators calculated using truncated likelihood.

2.5. *Moments of* $\sigma_1^{-2}U$

It is readily seen that the components of $\sigma_1^{-2}U$ that correspond to derivatives of $\sigma_1(a, b, c)^2$ with respect to the components of a, or with respect to c, are bounded. Therefore, to show that all moments of $\sigma_1^{-2}\|U\|$ are finite it suffices to show that the components of $\sigma_1^{-2}U$ that correspond to derivatives with respect to components of b have all moments finite. To this end, given a weight function w, define

$$W(w) = \sum_{i=1}^{p} a_i \sum_{k=1}^{\infty} w(k) \sum_{j_1=1}^{q} \cdots \sum_{j_k=1}^{q} b_{j_1} \cdots b_{j_k} X_{t-i-j_1-\cdots-j_k}^2,$$

and put $V = W(w)$ for $w(k) \equiv 1$ and $V_K = W(w)$ for $w(k) \equiv kI(k \geq K)$. Then it is sufficient to prove that $E\{V_1/(V+1)\}^\nu \leq \nu!\, c_1^\nu$ for a constant $C_1 > 0$ and all integers $\nu \geq 1$. This will follow if we show that

$$P\{V_1 > K(V+1)\} \leq C_2 C_3^K \tag{2.17}$$

for constants $C_2 > 0$ and $C_3 \in (0,1)$, and all integers $K \geq 1$. Now, $V_1 > K(V+1)$ implies $V_K > K$, and

$$KP(V_K > K) \leq E(V_K) \leq C_4 \sum_{k=K}^{\infty} k C_5^k \leq C_6 K C_5^K,$$

where $C_4, C_6 > 0$ and $C_5 = \sum_j b_j < 1$. Therefore (2.17) holds with $(C_2, C_3) = (C_6, C_5)$.

Similarly it may be proved that if U_r denotes the vector of rth derivatives of $\sigma_1(a,b,c)^2$, then $E(\sigma_1^{-2}\|U_r\|)^\nu < \infty$ for all $\nu \geq 1$. Likewise, all moments of the supremum of $\sigma_1^{-2}\|U_r\|$ over values of (a,b,c) in a sufficiently small neighborhood of the true values of these parameters are finite.

3. BOOTSTRAP METHODS

3.1. *Determining Scale by Studentizing*

First we discuss the scales associated with the limit results described by different parts of Theorem 2.1. It may be deduced from the theorem that if the distribution of ϵ^2 lies in the domain of attraction of a stable law (including the Normal law) having exponent strictly greater than 1, and if we define ℓ_n to be the infimum of values $\lambda > 0$ such that $nH(\lambda) \leq \lambda^2$, where $H(\lambda) = E\{\epsilon^4 I(\epsilon^2 \leq \lambda)\}$ (the same definition as at (2.8)), then

(3.1) $\quad n\ell_n^{-1}(\hat\theta - \theta^0)$ has a proper, nondegenerate limiting distribution.

Indeed, if $\tau^2 \equiv \mathrm{var}(\epsilon^2) < \infty$, then $\ell_n \sim n^{1/2}(\tau^2+1)^{1/2}$, and so by part (a) of Theorem 2.1 the limit distribution claimed at (3.1) is Normal with zero mean and variance matrix $(1+\tau^{-2})^{-1}M^{-1}$. If $\tau^2 = \infty$ but ϵ^2 is in the domain of attraction of the Normal law, then ℓ_n is identical to λ_n defined at (2.8), and so by part (b) of the theorem the limit is Normal $N(0, M^{-1})$. And if the stable law for the domain of attraction has exponent $\alpha \in (1,2)$, then ℓ_n is asymptotic to a constant multiple of the quantity λ_n defined at (2.9), and so the limiting distribution claimed at (3.1) is a rescaled form of that given in part (c) of Theorem 2.1. In each of these cases we shall let W denote a random variable with the limiting distribution of $n\ell_n^{-1}(\hat\theta - \theta^0)$.

Result (3.1) implies that in very general circumstances the scale of $(\hat\theta - \theta^0)$ is accurately described by $n^{-1}\ell_n$. In particular, the assertion $\hat\theta - \theta^0 = O_p(n^{-1}\ell_n)$ gives an accurate account of the order or magnitude of $\hat\theta - \theta^0$. However, the size of ℓ_n depends intimately on the particular law in whose domain of attraction the distribution of ϵ^2 lies. The law is unknown, and so it is quite awkward to determine the scale empirically; this is the root of the difficulty of accurately approximating the distribution of $\hat{\theta}-\theta^0$.

This difficulty can be overcome by observing that, if we define

$$\hat{\tau}^2 = \frac{1}{n}\sum_{t=1}^{n}\epsilon_t^4 - \left(\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2\right)^2, \tag{3.2}$$

then $n^{1/2}\hat{\tau}$ has scale ℓ_n. Indeed, we claim that in very general circumstances $n^{1/2}\ell_n^{-1}\hat{\tau}$ converges in distribution to a proper, nonzero limit as $n\to\infty$. If $\tau<\infty$, then the limiting distribution is clearly degenerate at a positive constant. The same holds true if $\tau=\infty$ but the distribution of ϵ^2 lies in the domain of attraction of the Normal law. The limiting constant here is in fact 1; this follows from the so-called "weak law of large numbers with infinite mean" (e.g., Theorem 3, page 223 of Feller (1966)). If $\tau=\infty$ and the distribution of ϵ^2 lies in the domain of attraction of a stable distribution with exponent $\alpha\in(1,2)$, then $(n^{1/2}\ell_n^{-1}\hat{\tau})^2$ converges in distribution to a strictly positive stable law with exponent $\frac{1}{2}\alpha$; this may be deduced from Section IX.8 of Feller (1966).

Not only do $n\ell_n^{-1}(\hat{\theta}-\theta^0)$ and $n^{1/2}\ell_n^{-1}\hat{\tau}$ both have proper limiting distributions, but their weak convergence is joint, as our next result shows. There, and in the other results in this section, it is assumed that all parameter estimators (including bootstrap estimators) are constructed by minimizing the negative log-likelihood within a fixed but sufficiently small distance, $\eta>0$, of the vector of true parameter values.

THEOREM 3.1: *Assume the initial conditions of Theorem* 2.1, *as well as the additional conditions for any one of parts* (*a*)–(*c*) *of that result*. *Then*

$$\ell_n^{-1}(n(\hat{\theta}-\theta^0), n^{1/2}\hat{\tau}) \to (W,S) \tag{3.3}$$

in distribution, *where the random variable S satisfies $P(0<S<\infty)=1$.*

In the contexts of cases (a) or (b) of Theorem 2.1, (3.3) follows trivially from the theorem, since (as noted two paragraphs above) S is then degenerate at a positive constant. In case (c) of the theorem it can be shown that we may write $(W,S)=(c_1W_0, c_2Y)$, where c_1, c_2 are strictly positive constants, W_0 is given by (2.11), and $Y^2=\sum_{k\ge 1}Y_k^2$, with Y_k being the same as at (2.11). The method of proof is similar to that of part (c) of Theorem 2.1. The distribution of Y^2 is identical to the limiting distribution of $\lambda_n^{-2}\sum_t\epsilon_t^4$, where λ_n is defined at (2.9), and has a stable law with exponent $\frac{1}{2}\alpha$; see the corollary of Hall (1978b). The constants c_1 and c_2 may differ from 1 only because the divisor ℓ_n at (3.3) may differ from the norming constants used in Theorem 2.1. In particular, if in the context of part (c) of the theorem we replace ℓ_n by λ_n at (3.3), then that result holds with $(W,S)=(W_0,Y)$ and Y as defined just above.

Of course, (3.1) implies that

$$n^{1/2}\frac{\hat{\theta}-\theta^0}{\hat{\tau}} \to \frac{W}{S} \tag{3.4}$$

in distribution. Comparing (3.1) and (3.4) we see that in normalizing by $\hat{\tau}$ we have eliminated the unknown scale factor ℓ_n from the distribution of $\hat{\theta}-\theta^0$. The limiting distribution in the case of (3.4) is unknown only in terms of shape, not scale. In view of this result it would be straightforward to approximate the distribution of the left-hand side of (3.4) using the subsample bootstrap, except that the errors ϵ_t used to compute $\hat{\tau}$ are unknown. However, they may be replaced by residuals, which we introduce in the next section.

3.2. *Resampling from a GARCH Process*

Suppose we are given a sample $\mathscr{X} = \{X_1, \ldots, X_n\}$, generated by the model described in Section 2.1. Define the truncated version $\tilde{\sigma}_t^2$ of σ_t^2, and the truncated version $\widetilde{L}_\nu$ of L, by (2.15) and (2.16) respectively. Both are functions only of the data in $\mathscr{X}$. Choose $(\tilde{a}, \tilde{b}, \tilde{c})$ to minimize $\widetilde{L}_\nu(a, b, c)$ over nonnegative parameter values. (Our theory permits the minimization to take place only over (a, b, c) within some radius η of (a^0, b^0, c^0), although the same sufficiently small η may be used throughout the bootstrap algorithm. In numerical practice, however, it appears that the minimization may be done globally.) Put $\hat{\sigma}_t = \tilde{\sigma}_t(\tilde{a}, \tilde{b}, \tilde{c})$. In this notation the "raw" residuals are $\tilde{\epsilon}_t = X_t/\hat{\sigma}_t$ for $\nu \le t \le n$. They can be standardized for location and scale, by defining

$$\hat{\epsilon}_t = \frac{\tilde{\epsilon}_t - n_1^{-1}\sum_u \tilde{\epsilon}_u}{\{n_1^{-1}\sum_u \tilde{\epsilon}_u^2 - (n_1^{-1}\sum_u \tilde{\epsilon}_u)^2\}^{1/2}}, \quad \nu \le t \le n, \tag{3.5}$$

where $n_1 = n - \nu + 1$ and each sum over u is taken over $\nu \le u \le n$.

Draw ϵ_t^*, for $-\infty < t < \infty$, by sampling randomly, with replacement, from the centered residuals $\hat{\epsilon}_\nu, \ldots, \hat{\epsilon}_n$. (The series of $\hat{\epsilon}_t$'s has already been truncated, but it could be further reduced if necessary, to remove suspected edge effects. Our theory requires only that the number of $\hat{\epsilon}_t$'s exceed a positive constant multiple on n, for all sufficiently large n. In practical implementation we draw ϵ_t^* for $-K \le t \le n$, where $K > 0$ is a sufficiently large integer.) Consider the stationary process (conditional on $\mathscr{X}$) defined by $X_t^* = \sigma_t^* \epsilon_t^*$ for $-\infty < t < \infty$, where, by analogy with (2.1),

$$(\sigma_t^*)^2 = \tilde{c} + \sum_{i=1}^{p} \tilde{a}_i (X_{t-i}^*)^2 + \sum_{j=1}^{q} \tilde{b}_j (\sigma_{t-j}^*)^2, \quad -\infty < t < \infty.$$

We know from Theorem 2.2 that $(\tilde{a}, \tilde{b}, \tilde{c})$ is consistent for (the true value of) (a, b, c). Therefore, if $c > 0$ and each component of a and b is nonzero (provided

$p \geq 1$ and $q \geq 1$ respectively), the same properties carry over to $\tilde{a}, \tilde{b}$, and $\tilde{c}$, with probability converging to 1. Likewise, for each $\eta > 0$ the probability that

$$\left| \sum_{i=1}^{p} \tilde{a}_i + \sum_{j=1}^{q} \tilde{b}_j - \left(\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j \right) \right| \leq \eta$$

converges to 1 as $n \to \infty$. Therefore, provided the original process X_t was stationary, i.e. (2.2) holds, it will also be true that the probability, conditional on $\mathscr{X}$, of X_t^* being stationary converges to 1 as $n \to \infty$. Here we have used a conditional form of a result of Giraitis, Kokoszka, and Leipus (2000).

Next we introduce a version of the m-out-of-n bootstrap. Let $m < n$, and compute estimators $(\tilde{a}^*, \tilde{b}^*, \tilde{c}^*)$ of (a, b, c) using the dataset $\mathscr{X}^* = \{X_1^*, \ldots, X_m^*\}$ and the truncated likelihood approach described in Section 2.4. In particular, $(\tilde{a}^*, \tilde{b}^*, \tilde{c}^*)$ are defined in the same way as functions of $\mathscr{X}^*$, as were $(\tilde{a}, \tilde{b}, \tilde{c})$ as functions of $\mathscr{X}$. We can use the same value of ν as before, provided $\nu/m \to 0$. Let $\tilde{\theta}$ and $\tilde{\theta}^*$ denote the vectors formed by concatenating the components of $(\tilde{a}, \tilde{b}, \tilde{c})$ and $(\tilde{a}^*, \tilde{b}^*, \tilde{c}^*)$, respectively, and put $m_1 = m - \nu + 1$,

$$\tilde{\tau}^2 = \frac{1}{n_1} \sum_{t=\nu}^{n} \tilde{\epsilon}_t^4 - \left(\frac{1}{n_1} \sum_{t=\nu}^{n} \tilde{\epsilon}_t^2 \right)^2, \qquad (\tilde{\tau}^*)^2 = \frac{1}{m_1} \sum_{t=\nu}^{m} (\tilde{\epsilon}_t^*)^4 - \left\{ \frac{1}{m_1} \sum_{t=\nu}^{m} (\tilde{\epsilon}_t^*)^2 \right\}^2,$$

these being the empirical and bootstrap versions, respectively, of $\hat{\tau}^2$ defined at (3.2).

3.3. *Bootstrap Approximation*

The distribution of $m^{1/2}(\tilde{\tau}^*)^{-1}(\tilde{\theta}^* - \tilde{\theta})$, conditional on $\mathscr{X}$, is our bootstrap approximation to the unconditional distribution of $n^1/2\tilde{\tau}^{-1}(\tilde{\theta} - \theta^0)$. Both distributions enjoy the limit property at (3.4); this follows from Theorem 3.2 below. Here it is necessary to assume that the size m of the bootstrap subsample $\mathscr{X}^*$ is strictly smaller than the sample size n, although still diverging to infinity with n.

THEOREM 3.2: *Assume the conditions of Theorem* 3.1. *Suppose too that* $m = m(n)$ *satisfies* $m \to \infty$ *and* $m/n \to 0$ *as* $n \to \infty$, *and that the truncation point* ν *used to construct the likelihood* $\widetilde{L}_\nu$ *at* (2.16) *satisfies* $\nu/\log n \to \infty$ *and* $\nu/m \to 0$. *Then*

$$\ell_n^{-1}(n(\tilde{\theta} - \theta^0), n^{1/2}\tilde{\tau}) \to (W, S)$$

in distribution, *and*

$$P\{\ell_m^{-1}(m(\tilde{\theta}^* - \tilde{\theta}), m^{1/2}\tilde{\tau}^*) \in [w_1, w_2] \times [s_1, s_2] \big| \mathscr{X}\} \to P\{(W, S) \in [w_1, w_2] \times [s_1, s_2]\}$$

in probability for each $-\infty < w_1 < w_2 < \infty$ *and all continuity points* $0 < s_1 < s_2 < \infty$ *of the distribution of* S, *where the random vector* (W, S) *is as at* (3.3).

COROLLARY: *Assume the conditions of Theorem* 3.2. *Then for all convex subsets* $\mathscr{C}$,

$$\left|P\{m^{1/2}(\tilde{\tau}^*)^{-1}(\tilde{\theta}^* - \tilde{\theta}) \in \mathscr{C}|\mathscr{X}\} - P\{n^{1/2}\tilde{\tau}^{-1}(\tilde{\theta} - \theta^0) \in \mathscr{C}\}\right| \to 0$$

in probability as $n \to \infty$.

An outline proof of Theorem 3.2 will be given in Section 5.3. The corollary follows from Theorem 3.2 on noting that the distribution of W/S is continuous.

When the distribution of ϵ^2 is in the domain of attraction of the Normal distribution (i.e. in cases (a) and (b) of Theorem 2.1), Theorem 3.2 holds without requiring $m/n \to 0$. In particular, in that setting it holds for $m = n$. However, the condition $m/n \to 0$, along with $m \to \infty$, is essential in the heavy-tailed case, where the error distribution is in the domain of attraction of a non-Normal stable law. Hall (1990) has given necessary and sufficient conditions for the standard, n-out-of-n bootstrap to produce consistent distribution estimators in the simpler case of estimating the distribution of a sample mean, and analogous results may be derived in the present setting.

Of course, these remarks address only asymptotic results. The way in which the bootstrap approximation depends on m for finite sample sizes n is not explicitly clear from Theorem 3.2 and its Corollary. Section 4 will take up this issue directly, and show that performance of the bootstrap approximation is robust against variation in m.

In conclusion we explain intuitively why, in the heavy-tailed case, the n-out-of-n bootstrap gives inconsistent results. In effect, it fails to accurately model relationships among extreme order statistics in the sample. For example, for each fixed $k \geq 2$ the probability that the k largest values in a resample are equal does not converge to 0 in the case of the n-out-of-n bootstrap. The probability does converge to 0 for the m-out-of-n bootstrap, provided $m/n \to 0$. And of course, it converges to 0 for the sample itself. In the case of heavy-tailed error distributions the limit properties of parameter estimators are dictated by the behavior of extreme order statistics. In particular this is why, in the heavy-tailed case, the distributions of the limit variables W and S are expressed in terms of extreme-value distributions.

3.4. *Confidence Regions*

In principle, simultaneous multivariate confidence regions for the components of θ^0 can be developed using the asymptotic approximation suggested by the Corollary. However, such regions can be difficult to interpret, and moreover their construction requires a determination of region shape. In the present general setting it is unclear how to do this. Therefore we shall consider only one-sided confidence intervals for individual parameter components. Two-sided intervals may be obtained in the usual way, on taking the intersection of two one-sided intervals.

The vectors θ^0, $\tilde{\theta}$, and $\tilde{\theta}^*$ are each of length $r = p+q+1$. Use the superscript notation $^{(k)}$ to denote the kth component, where $1 \le k \le r$. Given $\pi \in [0, 1]$, for example $\pi = 0.90$ or 0.95, put

$$\hat{u}_\pi = \inf\{u : P[m^{1/2}(\tilde{\tau}^*)^{-1}(\tilde{\theta}^* - \tilde{\theta})^{(k)} \le u | \mathscr{X}] \ge \pi\}.$$

We may of course compute $\hat{u}_\pi$ to arbitrary numerical accuracy by Monte Carlo simulation of the bootstrap distribution. Let $\mathscr{I}_\pi = [\tilde{\theta}^{(k)} - n^{-1/2}\hat{\tau}\hat{u}_\pi, \infty)$ be a potential confidence region for $(\theta^0)^{(k)}$. It follows from the Corollary that $\mathscr{I}_\pi$ has nominal coverage π, and that this coverage is asymptotically correct in the sense that, under the conditions assumed in the Corollary, $P\{(\theta^0)^{(k)} \in \mathscr{I}_\pi\} \to \pi$ as $n \to \infty$.

Our approach can be employed to construct consistent confidence regions even when the error distribution does not lie in any domain of attraction. Compare, for example, Hall and LePage (1996). However, on the present occasion such a degree of generality would be a significant distraction, and we do not pursue it.

4. NUMERICAL PROPERTIES

We report results of a simulation study of ARCH(2) and GARCH(1, 1) models. The latter are the most commonly found GARCH models in the literature, and enjoy significant application in the finance setting. In both cases we took the errors ϵ_t to have Student's t distribution with d degrees of freedom, for $d = 3, 4$, or 5. Note that $E|\epsilon_t|^d = \infty$. For ARCH(2) models we employed $c = 1$, $a_1 = 0.5$, and $a_2 = 0.4$. We used the same c and a_1 for GARCH(1, 1) models, and took $b_1 = 0.4$. It follows that our ARCH and GARCH processes both have the same variance.

We draw 1000 samples of size $n = 500$ and 1000, respectively, in each setting. We truncated likelihood functions at $\nu = p+1 = 2$ in the case of the ARCH model, and $\nu = 20$ for the GARCH model. Parameters were estimated by maximizing the likelihood $\widetilde{L}$ at (2.4). Boxplots of the average absolute errors (AAEs) are presented in Figure 1. The AAE is defined as $\frac{1}{3}(|\hat{c}-c| + |\hat{a}_1 - a_1| + |\hat{a}_2 - a_2|)$ for ARCH(2), and $\frac{1}{3}(|\hat{c}-c| + |\hat{a}_1 - a_1| + |\hat{b}_1 - b_1|)$ for GARCH(1, 1). The AAE is larger when the tail of the error distribution is heavier (i.e., d is smaller), although the deterioration is only slight. Moreover, the deterioration of estimator performance as we pass from the ARCH model to the relatively complex GARCH case is also only slight.

Bootstrap confidence intervals were constructed for each parameter in each model. For the sake of simplicity we give results only for the one-sided intervals $[\hat{\theta}^{(k)} - n^{-1/2}\hat{\tau}\hat{u}_\pi, \infty)$ introduced in Section 3.4, in the case $\pi = 0.9$. Here, $\hat{u}_\pi$ is the 10% quantile of a bootstrap sample drawn from the conditional distribution of $m^{1/2}(\hat{\theta}^* - \theta)/\hat{\tau}^*$ for sample size m. We took $m = 250, 300, 350, 400$, and 500 when $n = 500$, and $m = 500, 600, 700, 800$, and 1000 when $n = 1000$. Thus, we included the case $m = n$ in our simulations. Each bootstrap sampling step was repeated $B = 1000$ times, and 1000 samples were drawn for each configuration of parameters. The relative frequency of the event that a bootstrap interval covers

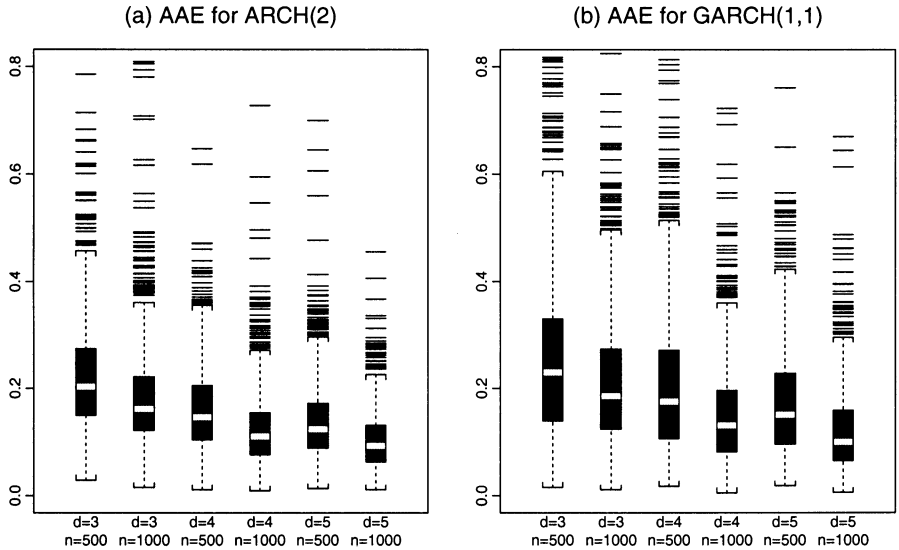

FIGURE 1.— *Simulated quasi-maximum likelihood estimates.* Panels (a) and (b) show boxplots of average absolute errors of estimators in the cases of (a) ARCH(2) and (b) GARCH(1, 1) models, respectively, with errors distributed as Student's t with $d = 3, 4, 5$ degrees of freedom, and sample sizes $n = 500$ and 1000.

the true value of the parameter was taken as our approximation to the true confidence level of the bootstrap interval.

Figure 2 displays approximate coverage levels for parameters c, a_1, and a_2 in the ARCH(2) model. In each case the level is close to its nominal value, 0.90, although accuracy is noticeably greater for the larger sample size. It can be seen that in the ARCH(2) case, distributions with lighter tails tend to produce relatively conservative confidence intervals. Nevertheless, only when $n = 500$, and for the parameter a_1, would the anticonservatism of the extreme heavy-tailed case (i.e. $d = 3$) be a potential problem. Note particularly that coverage error is quite robust against changes in m. The finite sample properties in the case $m = n$, as demonstrated in the figure, are broadly similar to those for $m < n$.

Figure 3 shows coverage levels in the case of the GARCH(1, 1) model. The method is having somewhat greater difficulty in this more complex setting, although serious problems occur only when constructing confidence intervals for b_1. As in the ARCH case, lighter-tailed distributions tend to produce relatively conservative confidence regions, and coverage error is again robust against varying m.

The method is less robust against choice of m when the error distribution is asymmetric. To illustrate this point we simulated samples with $n = 500$ and 1000 from ARCH(2) and GARCH(1, 1) models when the error distribution was Pareto, with density $3/(1+x)^4$ for $x \geq 0$ except that it was recentered and rescaled so as to have zero mean and unit variance. The results are depicted in Figure 4.

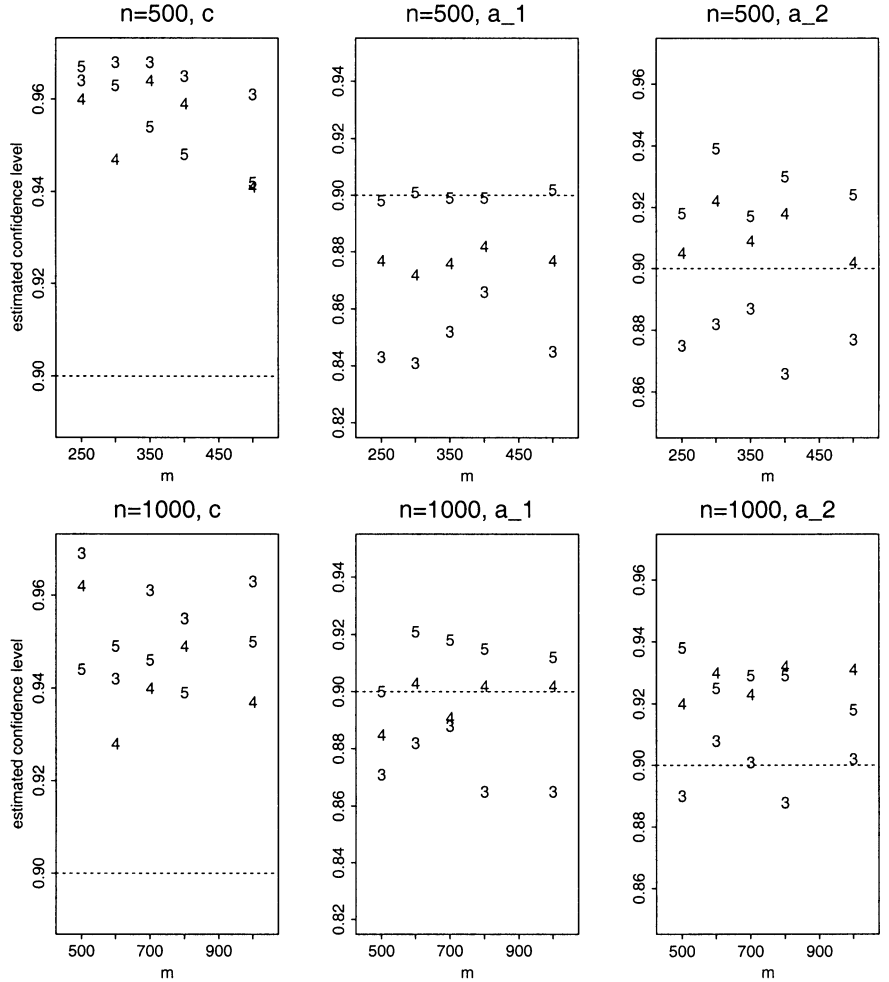

FIGURE 2.— *Confidence levels of bootstrap intervals for ARCH*(2) *models with symmetric errors.* The estimated confidence levels are plotted against m. The labels "3," "4," and "5" correspond to the number of degrees of freedom, d, of the error distribution. The dotted line indicates the nominal confidence level, 0.9.

For either model, minimal coverage error was obtained with m approximately 350 when $n = 500$, and with $m \approx 700$ or 800 when $n = 1000$.

Results for one-sided confidence intervals of the opposite parity, i.e. of the form $[0, \hat{\theta}^{(k)} - n^{-1/2}\hat{\tau}\hat{u}_{1-\pi}]$, are broadly similar. For the ARCH(2) model they show a tendency towards relatively less conservatism. By way of contrast, in the case of the GARCH(1, 1) model, lower-tailed confidence intervals for b_1 have greater coverage, and substantially greater coverage accuracy, than those in the upper tail (the latter are addressed in Figure 3). However, lower-tailed intervals for c and a_1 have reduced coverage relative to their upper-tailed counterparts.

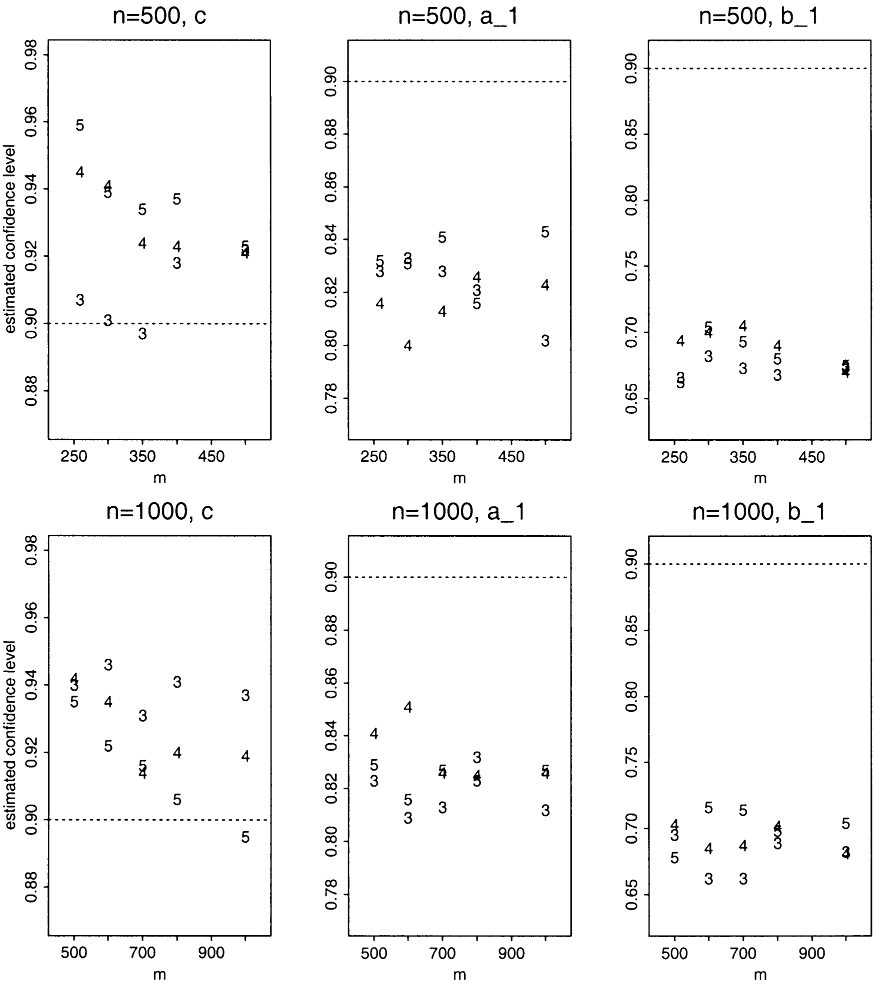

FIGURE 3.— *Confidence levels of bootstrap intervals for GARCH*(1, 1) *models with symmetric errors.* Legend is as for Figure 2.

5. TECHNICAL ARGUMENTS

5.1. *Proof of Theorem 2.1*

Step (*i*): *Preliminary Expansion.* Recall that (a, b, c) has been concatenated into a vector θ of length $r = p+q+1$. Let $\rho_t(\theta)$ denote the r-vector whose ith component is $\sigma_t(\theta)^{-4}\partial\sigma_t(\theta)^2/\partial\theta_i$. In this notation the likelihood equations, defining the extremum of the negative log-likelihood at (2.4), are

$$\sum_{t=1}^{n}\{X_t^2 - \sigma_t(\theta)^2\}\rho_t(\theta) = 0. \tag{5.1}$$

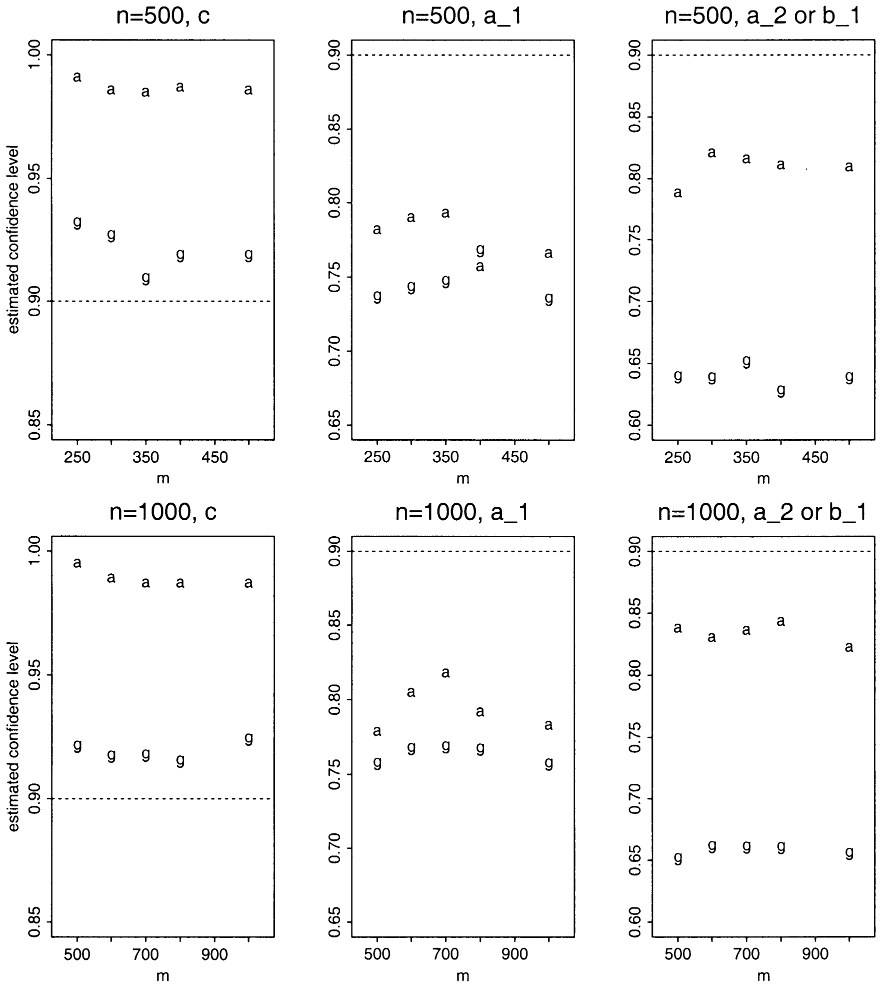

FIGURE 4.— *Confidence levels of bootstrap intervals for ARCH*(2) *and GARCH*(1, 1) *models with asymmetric errors.* Symbols "a" and "g" indicate results in ARCH and GARCH cases, respectively. The dotted line indicates the nominal confidence level, 0.9.

Let $A_t(\theta)$ and $B_t(\theta)$ be the r-vector and $r \times r$ matrix, respectively, of derivatives of $\sigma_t(\theta)^2$ and $\rho_t(\theta)$, respectively, with respect to θ. Using the fact that $p \geq 1$, that none of $a_1, \ldots, a_p$ vanish, and that if $q \geq 1$, then none of $b_1, \ldots, b_q$ vanish; and noting (2.5)–(2.7) and the results in Section 2.5; it may be shown by Taylor expansion that

$$\begin{aligned} (5.2) \qquad & \sigma_t(\theta)^2 = \sigma_t(\theta^0)^2 + A_t(\theta^0)^T(\theta - \theta^0) + \|\theta - \theta^0\|^2 R_{1t}(\theta)\sigma_t(\theta^0)^2, \\ & \rho_t(\theta) = \rho_t(\theta^0) + B_t(\theta^0)(\theta - \theta^0) + \|\theta - \theta^0\|^2 R_{2t}(\theta)\sigma_t(\theta^0)^{-2}, \end{aligned}$$

where $R_{1t}(\theta)$ and $R_{2t}(\theta)$ are an r-vector and an $r \times r$ matrix, respectively, and for a constant $C > 0$ not depending on η provided the latter is sufficiently small, and with $R_t = R_{1t}$ and $R_t = R_{2t}$,

$$(5.3) \qquad P\left\{n^{-1}\sum_{t=1}^{n} \sup_{\|\theta-\theta^0\|\le\eta} |R_t(\theta)| \le C\right\} \to 1,$$

with (5.3) interpreted component-wise in each case. The maximum value of η depends on the difference between the right-hand side and left-hand side at (2.2).

The uniformity claimed at (5.3) may be derived by expressing each of the formulae at (5.2) as a Taylor expansion with exact remainder, the latter in the form of a quadratic in $\theta - \theta^0$ having its coefficients expressed as functions evaluated at a point ω that satisfies $\omega = p\theta + (1-p)\theta^0$ for some $0 \le p \le 1$. Using the fact that ω can be made arbitrarily close to θ^0 by choosing η sufficiently small, an explicit upper bound to each of the coefficients can be derived, valid uniformly in $\|\theta - \theta^0\| \le \eta$ and indexed by t. Then, arguing as in Section 5.2 it can be shown that each moment of each coefficient is bounded, uniformly in t. In particular, defining

$$S_t = \sup_{\|\theta-\theta^0\|\le\eta} |R_t(\theta)|,$$

we have $\sup_t E(|S_t|) < \infty$. This implies (5.3).

Hence, since $X_t^2 - \sigma_t(\theta^0)^2 = (\epsilon_t^2 - 1)\sigma_t(\theta^0)^2$, equation (5.1) may be written as

$$(5.4) \qquad \sum_{t=1}^{n}(\epsilon_t^2 - 1)\sigma_t(\theta^0)^2\rho_t(\theta^0) + \sum_{t=1}^{n}\{(\epsilon_t^2-1)\sigma_t(\theta^0)^2 B_t(\theta^0) - \rho_t(\theta^0)A_t(\theta^0)^T\}(\theta-\theta^0) + \|\theta-\theta^0\|^2 nR(\theta) = 0,$$

where, by (5.3) and for a constant $C > 0$ not depending on η provided the latter is sufficiently small,

$$(5.5) \qquad P\left\{\sup_{\|\theta-\theta^0\|\le n} |R(\theta)| \le C\right\} \to 1.$$

It may be proved from the ergodic theorem, using the property $E(\epsilon^2) < \infty$, that

$$(5.6) \qquad n^{-1}\sum_{t=1}^{n}(\epsilon_t^2-1)\sigma_t(\theta^0)^2 B_t(\theta^0) \to 0,$$

$$(5.7) \qquad n^{-1}\sum_{t=1}^{n}\rho_t(\theta^0)A_t(\theta^0)^T \to M = E\{\rho_1(\theta^0)A_1(\theta^0)^T\}$$

as $n \to \infty$, where both convergences are in probability.

From (5.4)–(5.7) we may deduce that

$$(5.8) \qquad \{M + o_p(1)\}(\theta-\theta^0) + \|\theta-\theta^0\|^2 R(\theta) = n^{-1}\sum_{t=1}^{n}(\epsilon_t^2-1)w_t,$$

where the "$o_p(1)$" term does not depend on θ, $w_t = \sigma_t(\theta^0)^2\rho_t(\theta^0)$, and R satisfies (5.5). Since $\eta > 0$ is arbitrarily small, although fixed, then this implies that

$$(5.9)\qquad \{M + o_p(1)\}(\hat{\theta} - \theta^0) = n^{-1}\sum_{t=1}^{n}(\epsilon_t^2 - 1)w_t.$$

To derive (5.9) from (5.8), let $M_1 = M + o_p(1)$ be identical to the term within braces on the left-hand side of (5.8). Multiply both sides of (5.8) on the left by M_1^{-1}; denote the resulting right-hand side by $\zeta = O_p(n^{-1/2})$; and put $S(\theta) = M_1^{-1}R(\theta)$. Let C_1 equal the constant C at (5.5) multiplied by twice the inverse of the absolute value of the smallest eigenvalue of M; observe that by (5.5),

$$P\left\{\sup_{\|\theta-\theta^0\|\le\eta}|S(\theta)| \le C_1\right\} \to 1$$

as $n \to \infty$; and note, from the discussion immediately preceding (5.5), that C_1 does not depend on η, provided the latter is sufficiently small. Put $\phi = \theta - \theta^0$ and $T(\phi) = S(\theta)$. If $|T(\phi)| \le C_1$ then the norm of any solution of $\phi + \|\phi\|^2T(\phi) = \zeta$ either is not less than $C_1^{-1} + o_p(1)$, or equals $O_p(n^{-1/2})$. If we insist that $0 < \eta < \frac{1}{2}C_1$, and restrict ϕ so that $\|\phi\| \le \eta$, then the probability that the former type of solution arises converges to 0 as $n \to \infty$. Therefore, $\hat{\theta} - \theta^0 = O_p(n^{-1/2})$. In this case, returning to (5.8), we deduce (5.9).

Result (5.9) implies that, in order to establish the limit results claimed for $\hat{\theta} - \theta^0$ in parts (a)–(c) of the theorem, it suffices to show that they apply to $n^{-1}M^{-1}\sum_t(\epsilon_t^2 - 1)w_t$.

Part (d) of the theorem is more complex, however, since the centering constant μ_n in (2.14) diverges to infinity. We now outline a method that can be used in this case. Put $M_1 = E\{B_1(\theta^0)\}$. Arguing as in Step (iii) of the present proof we may show that in the context of part (d) of the theorem, (5.6) may be refined by showing that the random variable

$$\lambda_n^{-1}\sum_{t=1}^{n}(\epsilon_t^2 - 1)\sigma_t(\theta^0)^2B_t(\theta^0) + \mu_nM_1$$

has a proper limiting distribution. Therefore,

$$n^{-1}\sum_{t=1}^{n}(\epsilon_t^2 - 1)\sigma_t(\theta^0)^2B_t(\theta^0) = -n^{-1}\lambda_n\mu_nM_1 + O_p(\lambda_n/n).$$

It can be shown that (5.7) may be sharpened to

$$(5.10)\qquad n^{-1}\sum_{t=1}^{n}\rho_t(\theta^0)A_t(\theta^0)^T = M + O_p(n^{-\eta})$$

for some $\eta > 0$. (The method of proof involves approximating the summands by their counterparts in which we set to zero all values of ϵ_u for which $u < t - \nu$, where ν denotes the integer part of n^{ξ} and $0 < \xi < 1$. Then the summands with indices t and u are independent if $|t-u| > \nu$, whence it follows that the variance of the series on the left-hand side of (5.10), after the summands have been modified, equals $O(n\nu)$. Moreover, the differences between the original series and its modified counterpart, and between the expected values of those two series, both equal $O(n^{-C})$ for all $C > 0$. Result (5.10) follows on combining these properties.) Therefore, by (5.4) and since λ_n/n is a slowly varying function of n,

$$(5.11)\qquad n\lambda_n^{-1}\{I + n^{-1}\lambda_n\mu_n M^{-1}M_1 + O_p(\lambda_n/n)\}(\hat{\theta}-\theta^0) = M^{-1}\lambda_n^{-1}\sum_{t=1}^{n}(\epsilon_t^2-1)w_t + O_p(n\lambda_n^{-1}\|\hat{\theta}-\theta^0\|^2) + o_p(1).$$

Result (5.11), in company with the assumption that $n^{-1}\lambda_n\mu_n^2 \to 0$, and the property

$$(5.12)\qquad M^{-1}\lambda_n^{-1}\sum_{t=1}^{n}(\epsilon_t^2-1)w_t + \mu_n E(V_1) - \gamma E(V_1) \text{ converges in distribution to } W_1 \text{ as } n\to\infty,$$

which we shall establish in step (iii) of the present proof, implies that $\hat{\theta}-\theta^0 = O_p(\lambda_n\mu_n/n)$. Again using the fact that $n^{-1}\lambda_n\mu_n^2 \to 0$ we may now deduce first that $n\lambda_n^{-1}\|\hat{\theta}-\theta^0\|^2 \to 0$ in probability, and thence, from (5.11), that

$$n\lambda_n^{-1}(\hat{\theta}-\theta^0) = M^{-1}\lambda_n^{-1}\sum_{t=1}^{n}(\epsilon_t^2-1)w_t + o_p(1).$$

The result claimed in part (d) of the theorem follows from the latter expansion and (5.12).

Step (*ii*)*: Parts* (*a*) *and* (*b*) *of Theorem.* Case (a) in the theorem follows directly from (5.9); the series at (5.9) may be expressed as a multivariate square-integrable martingale, and a martingale central limit theorem such as the multivariate form of that of Brown (1971) may be used to obtain the result. Alternatively, case (a) is effectively covered by our treatment of case (b), which we give next.

Define D_t to equal any linear combination of the components of the vector w_t at (5.9). We shall prove that under the conditions for case (b), or indeed if $E(\epsilon_t^4) < \infty$,

$$(5.13)\qquad \lambda_n^{-1}\sum_{t}(\epsilon_t^2-1)D_t \text{ is asymptotically Normally distributed with variance } \operatorname{var}(\epsilon_t^2)E(D_1^2).$$

Part (b) of the theorem follows from this property via the Cramér-Wold device.

Let λ_n be as at (2.8) and define $I_{tn} = I(|\epsilon_t^2 - 1| \le \lambda_n)$,

$$(5.14)\qquad \begin{aligned} &J_{tn} = 1 - I_{tn}, \quad \delta_n = E\{(\epsilon_t^2 - 1)I_{tn}\},\\ &S_1 = \sum_t (\epsilon_t^2 - 1)D_t, \quad S_2 = \sum_{t=1}^n \{(\epsilon_t^2 - 1)I_{tn} - \delta_n\}D_t,\\ &S_3 = \sum_{t=1}^n (\epsilon_t^2 - 1)J_{tn}D_t, \quad S_4 = \sum_{t=1}^n D_t. \end{aligned}$$

Note that $S_1 = S_2 + S_3 + \delta_n S_4$. Since H, defined at (2.8), is slowly varying at infinity, then

$$(5.15)\qquad \lambda^2 P(|\epsilon^2 - 1| > \lambda)/H(\lambda) \to 0$$

as $\lambda \to \infty$. See (8.5), of Feller (1966, p. 303). It therefore follows from the definition of λ_n at (2.8) that

$$(5.16)\qquad nP(|\epsilon^2 - 1| > \lambda_n) \sim \lambda_n^2 H(\lambda_n)^{-1} P(|\epsilon^2 - 1| > \lambda_n) \to 0$$

as $n \to \infty$, and so $P(S_3 = 0) \to 1$. Analogously to (5.15) it may be proved that $\lambda E\{|\epsilon^2 - 1| I(|\epsilon^2 - 1| > \lambda)\}/H(\lambda) \to 0$ as $\lambda \to \infty$, and so analogously to (5.16),

$$n|\delta_n| \le nE\{|\epsilon^2 - 1| I(|\epsilon^2 - 1| > \lambda_n)\} = o\{n\lambda_n^{-1} H(\lambda_n)\} = o(\lambda_n).$$

Therefore, $|\delta_n S_4| = O_p(n|\delta_n|) = o_p(\lambda_n)$. Combining the results in the paragraph we deduce that

$$(5.17)\qquad S_1 = S_2 + o_p(\lambda_n).$$

Let $D_t(k)$ be the random variable obtained by setting to 0 each ϵ_u for $u < t - k$, in the formula for D_t. Now, $\{D_t(k), -\infty < t < \infty\}$ is a stationary time series, and for a constant $C_1 > 0$, $|D_1|$ and $|D_1(k)|$ are both less than C_1 with probability 1 for all $k \ge 1$. Define $Q_t = (\epsilon_t^2 - 1)I_{tn} - \delta_n$, $s_{tn}(k) = Q_t D_t(k)$, $v_{tn}(k) = Q_t\{D_t - D_t(k)\}$, $S_6 = \sum_t s_{tn}(k)$, and $S_7 = \sum_t v_{tn}(k)$. Then

$$(5.18)\qquad S_2 = S_6 + S_7.$$

Given $\xi > 0$, choose k_0 so large that $E\{D_1 - D_1(k)\}^2 \le \xi$ for all $k \ge k_0$. Note that for a constant $C_2 > 0$ not depending on k or n, $E(Q_1^2) \le C_2 H(\lambda_n)$. Observe too that $E(v_{tn}v_{un}) = 0$ if $t \ne u$. Therefore if $k \ge k_0$,

$$\begin{aligned} E(S_7^2) &= nE\{v_{1n}(k)^2\}\\ &= nE(Q_1^2)E\{D_1 - D_1(k)\}^2 \le C_2 nH(\lambda_n)\xi \sim C_2\lambda_n^2\xi. \end{aligned}$$

Since this holds for each $\xi > 0$, then

$$(5.19)\qquad \lim_{k\to\infty} \limsup_{n\to\infty} E(S_7/\lambda_n)^2 = 0.$$

Next we prove a central limit theorem for S_6. Distribute the summands $s_{tn}(k), 1 \le t \le n$, among blocks of alternating lengths k and ℓ. Let the sums of $s_{tn}(k)$ within the respective blocks be T_{1u} and T_{2u}, in the cases of blocks of lengths k and ℓ respectively. Thus, excepting a possible residual block at the very end, each T_{1u} or T_{2u} is a sum of k or ℓ, respectively, adjacently indexed values of $s_{tn}(k)$. Denote by n_1 and n_2 the numbers of indices u for T_{1u} and T_{2u}, respectively. Then $|n_1 - n_2| \le 1$, both n_1 and n_2 are asymptotic to $n/(k+\ell)$, and

$$S_6 = \sum_{u=1}^{n_1} T_{1u} + \sum_{u=1}^{n_2} T_{2u}. \tag{5.20}$$

Note too that for a constant $C_3(k) > 0$,

$$E(T_{11}^2) \le k^2 E\{s_{11}(t)^2\} \le 2k^2 C_1^2 E\{(\epsilon_1^2 - 1)^2 I_{11}\} \le C_3(k) H(\lambda_n).$$

If $\ell > k$, then the variables T_{1u}, for $u \ge 1$, are independent, and so

$$\begin{aligned} E\left(\lambda_n^{-1} \sum_u T_{1u}\right)^2 = \lambda_n^{-2} \sum_u E(T_{1u}^2) &\sim n\lambda_n^{-2}(k+\ell)^{-1} E(T_{11}^2) \\ &\sim H(\lambda_n)^{-1}(k+\ell)^{-1} E(T_{11}^2). \end{aligned}$$

Therefore, for each fixed $k \ge 1$,

$$\lim_{\ell \to \infty} \limsup_{n \to \infty} E\left(\lambda_n^{-1} \sum_u T_{1u}\right)^2 = 0. \tag{5.21}$$

The summands $s_{tn}(k)$ are k-dependent, and so the variables T_{2u}, for $1 \le u \le n_2$, are independent. Excepting a possible residual block at the end, they are also identically distributed with finite variance, although the distribution and variance depend on n. Lindeberg's central limit theorem may be applied to the series $\lambda_n^{-1} \sum_u T_{2u}$ to show that it is asymptotically Normally distributed with zero mean and variance $\beta(k, \ell)^2$, say. (When showing that Lindeberg's condition is satisfied, note that the function H is slowly varying.) It may be proved by elementary calculus that for each fixed $k \ge 1$, $\beta(k, \ell)^2 \to \beta^2 \equiv E(D_1^2)$ as $\ell \to \infty$. Therefore, writing Z for a variable with the standard Normal random distribution, we have for each fixed k,

$$\lim_{\ell \to \infty} \lim_{n \to \infty} \sup_{-\infty < x < \infty} \left| P\left(\lambda_n^{-1} \sum_u T_{2u} \le x\right) - P(Z\beta \le x) \right| = 0.$$

Combining this result with (5.20) and (5.21) we deduce that for each fixed $k \ge 1$, $\lambda_n^{-1} S_6$ is asymptotically Normal $N(0, \beta^2)$. Result (5.13) follows from this property and (5.17)–(5.19).

Step (iii): Parts (c)–(e) of Theorem. We shall show only that each component of $n\lambda_n^{-1}(\hat{\theta}-\theta^0)$ converges weakly to the corresponding component of the distribution of W_0 or W_1 (after the appropriate location change if $\alpha=1$). Our argument has a straightforward multivariate version, in which a multivariate metric between distributions is used in place of the Lévy distance that we employ. The longer argument differs only in notational complexity. Part (e) of the theorem may be proved as in Samorodnitsky and Taqqu (1994, Theorem 1.4.5). In particular,

$$S(2)=\sum_{k=1}^{\infty}(Y_k v_k - EY_k ED_1) \tag{5.22}$$

converges almost surely when $1<\alpha<2$.

Let D_t equal any one of the components of the vector $M^{-1}w_t$, where w_t is as at (5.9). Take $I_{tn}=I(|\epsilon_t^2-1|\le C_4\lambda_n)$, where λ_n has the meaning it assumes in parts (c) and (d) of the theorem and $C_4>0$ will be taken small and fixed. In this new notation, define J_{tn} and δ_n by (5.14), and let

$$S_1=\sum_{t=1}^{n}(\epsilon_t^2-1)D_t,\quad S_2=\sum_{t=1}^{n}(\epsilon_t^2-1)J_{tn}D_t,$$
$$S_3=\sum_{t=1}^{n}\{(\epsilon_t^2-1)I_{tn}-\delta_n\}D_t,\quad S_4=\sum_{t=1}^{n}D_t.$$

Then,

$$S_1=S_2+S_3+\delta_n S_4. \tag{5.23}$$

We shall first prove that

$$\lim_{C_4\to 0}\limsup_{n\to\infty}E(S_3/\lambda_n)^2=0, \tag{5.24}$$

$$\delta_n S_4/\lambda_n=-\Delta+o_p(1), \tag{5.25}$$

where, defining

$$\beta=\beta(n,C_4)=\begin{cases}\alpha(\alpha-1)^{-1}C_4^{1-\alpha} & \text{if } 1<\alpha<2,\\ \mu_n-\log C_4 & \text{if } \alpha=1,\end{cases}$$

we put $\Delta=\beta E(D_1)$. Combining (5.23)–(5.25) we see that if $L(Q_1,Q_2)$ represents the Lévy distance between the distributions of random variables Q_1 and Q_2, then

$$\lim_{C_4\to 0}\limsup_{n\to\infty}|L(S_1/\lambda_n, S_2\lambda_n^{-1}-\Delta)|=0. \tag{5.26}$$

To prove (5.24), note that with $Q_t=(\epsilon_t^2-1)I_{tn}-\delta_n$ we have by Karamata's theorem (denoted by KT, say; see Bingham, Goldie, and Teugels (1987, Section 1.6)) that

$$E(Q_t^2)\sim E\{\epsilon^4 I(\epsilon^2\le C_4\lambda_n)\}\sim\alpha(2-\alpha)^{-1}C_4^{2-\alpha}\lambda_n^2/n,$$

and so $E(S_3/\lambda_n)^2 \sim \alpha(2-\alpha)^{-1}C_4^{2-\alpha}E(D_1^2)$ as $n \to \infty$. Since $\alpha < 2$ then this implies (5.24).

Repeated use of KT and the uniform convergence theorem for slow variation (denoted by UCT, say; see Bingham, Goldie, and Teugels (1987, Section 1.2)) allows us to prove that

$$(5.27) \qquad \begin{aligned} &E\{\epsilon^2 I(C_4\lambda_n < \epsilon^2 \le C_4\lambda_n + 1)\} = o(\lambda_n/n), \\ &-\delta_n = E\{\epsilon^2 I(\epsilon^2 > C_4\lambda_n)\} + o(\lambda_n/n). \end{aligned}$$

When $1 < \alpha < 2$, $nE\{\epsilon^2 I(\epsilon^2 > C_4\lambda_n)\} \sim \beta\lambda_n$ by KT and UCT, and so in view of (5.27), $-n\delta_n/\lambda_n \to \beta$; call this property (P). The variables D_1 and D_t are asymptotically uncorrelated as $|t| \to \infty$, and the process $\{D_t\}$ is stationary and essentially bounded, so

$$(5.28) \qquad n^{-1}S_4 \to E(D_1) \quad \text{in probability.}$$

Result (5.25), for $1 < \alpha < 2$, follows from this property and (P). (Result (5.28) can be derived using the argument leading to (5.10); see the parenthetical remarks below that formula.)

To complete the proof of (5.25) when $\alpha = 1$, note that by KT and UCT,

$$E\{\epsilon^2 I(C_4\lambda_n < \epsilon^2 < \lambda_n)\} = -(\lambda_n/n)\log C_4 + o(\lambda_n/n).$$

Hence by (5.27),

$$-\delta_n/\lambda_n = E\{\epsilon^2 I(\epsilon^2 > \lambda_n)\} - n^{-1}\log C_4 + o(n^{-1}).$$

Result (5.25) follows from this formula, the fact that $E\{\epsilon^2 I(\epsilon^2 > \lambda_n)\}$ is slowly varying in n, and (5.27).

For large n, $J_{tn} = I(\epsilon_t^2 > C_4\lambda_n + 1)$ for all $1 \le t \le n$. Furthermore, for each $C_4 > 0$,

$$(5.29) \qquad \lim_{k\to\infty} \liminf_{n\to\infty} P(\text{no more than } k \text{ out of } \epsilon_1^2, \dots, \epsilon_n^2 \text{ exceed } C_4\lambda_n) = 1.$$

Therefore the number of nonzero terms in S_2 equals $O_p(1)$ as $n \to \infty$. Hence, if we take $\epsilon_{(n)}^2 \ge \cdots \ge \epsilon_{(1)}^2$ to be the ordered values of $\epsilon_1^2, \dots, \epsilon_n^2$, if we let $D_{(n)}, \dots D_{(1)}$ be the concomitant values of $D_1, \dots, D_n$, and if we put $Z_{(t)} = \epsilon_{(t)}^2/\lambda_n$, then

$$(5.30) \qquad S_2/\lambda_n = \sum_{t=1}^{n} Z_{(t)}D_{(t)}I(Z_{(t)} > C_4) + o_p(1).$$

Recall that D_t denotes a particular component of $M^{-1}w_t$, say the sth; let v_k be the sth component of V_k. Then v_k and D_t have the same distribution, and in particular, $E(v_k) = E(D_t)$. The joint limiting distribution of $(Z_{(n)}, D_{(n)}), \dots, (Z_{(n-k+1)}, D_{(n-k+1)})$ is the joint distribution of $(Y_1, v_1), \dots, (Y_k, v_k)$, for each fixed k. (Call this result (R); we shall outline a proof in the next paragraph.) Note too that the variables $Z_{(t)}$ are nonincreasing with t. Combining these properties and (5.29), and applying Lemma 5.1 below to the series on the right-hand side of (5.30), we deduce that

$$(5.31)\qquad S_2/\lambda_n \to S(1) \equiv \sum_{k=1}^{\infty} Y_k v_k I(Y_k > C_4),$$

where the convergence is in distribution as $n \to \infty$. The infinite series here converges because, reflecting (5.29), with probability 1 it contains only a finite number of nonzero terms. By (5.26) and (5.31),

$$(5.32)\qquad \lim_{C_4 \to 0} \limsup_{n\to\infty} L\{S_1/\lambda_n, S(1) - \Delta\} = 0.$$

Next we outline a derivation of result (R). Let r_j, a random integer, denote the index r such that $e_{(n-j+1)} = e_r^2$. If (R) did not hold, then there would exist a subsequence of values of n along which the joint distribution of $(Z_{(n)}, D_{(n)}), \dots, (Z_{(n-k+1)}, D_{(n-k+1)})$ converged, as $n \to \infty$, but to a sub-distribution limit that did not have the form claimed under (R). Then, noting that the separations of the integers r_j diverge with sample size, we could choose a sub-subsequence n_m, say, for which there existed a sequence of positive integers ν_n diverging to infinity and with the property that, as n increased along the sub-subsequence, $|r_{j_1} - r_{j_2}|/\nu_n$ diverged to infinity with probability 1 whenever $1 \le j_1 \ne j_2 \le k$. Call this property (P). Consider the version of the problem in which each D_t is replaced by its approximant D'_t, the latter defined by replacing by 0 each ϵ_s for which $s < t - \nu_n$. Let $D'_{(n-j)}$ denote the corresponding concomitant of $Z_{(n-j)}$. Note that D'_{t_1} and D'_{t_2} are independent if $|t_1 - t_2| > c_n$. It may be shown using this result and property (P) that $(Z_{(n)}, D'_{(n)}), \dots, (Z_{(n-k+1)}, D'_{(n-k+1)})$ converges to the limit distribution claimed for $(Z_{(n)}, D_{(n)}), \dots, (Z_{(n-k+1)}, D_{(n-k+1)})$ under (R). However, $D'_{(n-j)} - D_{(n-j)}$ converges in probability to zero for each $0 \le j \le k-1$, and so (R) must be true at least along the sub-subsequence $\{n_m\}$. This contradiction completes the proof of (R).

The proof of Lemma 5.1 is given in the Appendix.

LEMMA 5.1: *Let* (U_{ni}, V_{ni}), *for* $1 \le i \le n < \infty$, *denote a triangular array of random* 2-*vectors, and let* $(U_1, V_1), (U_2, V_2), \dots$ *be an infinite sequence of* 2-*vectors. Assume that for each* $k \ge 1$ *the joint distribution of the ordered sequence* $(U_{ni}, V_{ni}), 1 \le i \le k$, *converges weakly to the joint distribution of* $(U_i, V_i), 1 \le i \le k$, *as* $n \to \infty$. *Suppose too that for each* k, $(V_1, \dots, V_k)$ *has a continuous distribution, and that for each* $C > 0$,

$$(5.33)\qquad \lim_{k\to\infty} \liminf_{n\to\infty} P(\textit{no more than the first } k \textit{ of } V_{n1}, \dots, V_{nn} \textit{ exceed } C) = 1,$$

$$\lim_{k\to\infty} P(\textit{no more than the first } k \textit{ of } V_1, V_2, \dots \textit{ exceed } C) = 1.$$

Then for each $C > 0$,

$$\sum_{i=1}^{n} U_{ni} I(V_{ni} > C) \to \sum_{i=1}^{\infty} U_i I(V_i > C)$$

in distribution as $n \to \infty$.

Reverting to the case $1 < \alpha < 2$, and with $S(1)$ and $S(2)$ given by (5.31) and (5.22), we may write

$$S(1) = S(2) - S(3) + S(4), \tag{5.34}$$

where

$$S(3) = \sum_{k=1}^{\infty} [Y_k v_k I(Y_k \le C_4) - E\{Y_k(Y_k \le C_4)\}E(D_1)], \tag{5.35}$$

$$S(4) = E(D_1) \sum_{k=1}^{\infty} E\{Y_k(Y_k > C_4)\}. \tag{5.36}$$

The argument used to establish convergence of the infinite series $S(2)$ may be employed to prove that $S(3)$ also converges (as an infinite series) with probability 1. Note too that $E\{S(3)\} = 0$ and $\operatorname{var}\{S(3)\} \to 0$ as $C_4 \to 0$. Therefore $S(3) \to 0$ in probability as $C_4 \to 0$. Combining these results with (5.32) we deduce that

$$\lim_{C_4 \to 0} \limsup_{n \to \infty} L\{S_1/\lambda_n, S(2) + S(4) - \Delta\} = 0. \tag{5.37}$$

Continuing to assume $1 < \alpha < 2$, and defining F_k as at (2.10), we have

$$\begin{aligned} S(4)/E(D_1) &= \sum_{k=1}^{\infty} \int_{C_4}^{\infty} y\, dF_k(y) \\ &= \int_{C_4}^{\infty} \left[\sum_{k=1}^{\infty} \{1 - F_k(y)\} \right] dy + C_4 \sum_{k=1}^{\infty} \{1 - F_k(C_4)\}. \end{aligned} \tag{5.38}$$

Now, $\sum_{k \ge 1} \{1 - F_k(y)\} = y^{-\alpha}$. It may therefore be shown that if $1 < \alpha < 2$, then $S(4)/E(D_1) = \beta$. Hence, (5.37) is equivalent to: $\lim_{n \to \infty} L\{S_1/\lambda_n, S(1)\} = 0$, which implies that S_1/λ_n converges in distribution to $S(1)$. This proves part (c) of the theorem, in a component-wise sense.

Next we treat $\alpha = 1$. Continue to define $S(2)$, $S(3)$, and $S(4)$ by (5.22), (5.35), and (5.36), except that now the series should be taken only over $k \ge 2$. (We also define $S(1)$ by (5.31), in this case continuing to take the sum over $k \ge 1$.) Then instead of (5.34),

$$S(1) = S(2) - S(3) + S(4) + Y_1 v_1 I(Y_1 > C_4).$$

We have as before that $S(3) \to 0$ in probability as $C_4 \to 0$. Therefore, from (5.32) we deduce in place of (5.37) that

$$\lim_{C_4 \to 0} \limsup_{n \to \infty} L\{S_1/\lambda_n, S(2) + S(4) - \Delta + Y_1 v_1\} = 0, \tag{5.39}$$

where γ denotes Euler's constant.

Result (5.38) continues to hold when $\alpha = 1$, provided now that both series on the right-hand side are taken over $k \geq 2$. Since (in the case $\alpha = 1$)

$$\sum_{k=2}^{\infty} \{1 - F_k(y)\} = F_1(y) + y^{-1} - 1 = \exp(-y^{-1}) + y^{-1} - 1,$$

and

$$\gamma = \int_0^{\infty} \{\exp(-y^{-1}) + \min(y^{-1} - 1, 0)\}\, dy,$$

then

$$\int_{C_4}^{\infty} \left[\sum_{k=2}^{\infty} \{1 - F_k(y)\} \right] dy + C_4 \sum_{k=2}^{\infty} \{1 - F_k(C_4)\} = \gamma - \log C_4 + o(1)$$

as $C_4 \to 0$. Therefore,

$$S(4) = (\gamma - \log C_4) E(D_1) + o(1) \tag{5.40}$$

as $C_4 \to 0$. Noting (5.39), (5.40), and the fact that $\Delta = (\mu_n - \log C_4) E(D_1)$ in the case $\alpha = 1$, we deduce that

$$\lim_{n \to \infty} L\{S_1/\lambda_n, S(2) + Y_1 v_1 + \gamma E(D_1) - \mu_n E(D_1)\} = 0,$$

which implies that $(S_1/\lambda_n) + \mu_n E(D_1) - \gamma E(D_1)$ converges in distribution to $S(2) + Y_1 v_1$. This proves the component-wise form of (5.12), and as argued at the beginning of the current step, a proof of the vector form is virtually identical. We showed, at the end of step (i), how part (d) of the theorem follows from (5.12).

5.2. *Outline Proof of Theorem 2.2*

Recall the definitions (2.4) and (2.15) of σ_t^2 and $\tilde{\sigma}_t^2$. The property $\nu/\log n \to \infty$, and the fact that $E(X_t^2) < \infty$, ensure that for all $C > 0$,

$$\sup_{(a,b,c) \in \mathcal{N}} \sup_{\nu \leq t \leq n} |\tilde{\sigma}_t(a,b,c)^2 - \sigma_t(a,b,c)^2| = O_p(n^{-C}), \tag{5.41}$$

where $\mathcal{N}$ denotes a sufficiently small, but fixed, open neighborhood of the true parameter values (a^0, b^0, c^0). Let L_ν denote the version of L, defined at (2.4), that is obtained when the sum over $1 \leq t \leq n$ on the right-hand side of (2.4) is replaced by the sum over $\nu \leq t \leq n$. It may be proved from (5.41) that for all $C > 0$, $\widetilde{L}_\nu(a,b,c) - L_\nu(a,b,c) = O_p(n^{-C})$, uniformly in $(a,b,c) \in \mathcal{N}$. The difference between the vectors of derivatives of the likelihoods is likewise of the same order, uniformly in $(a,b,c) \in \mathcal{N}$. The proof of Theorem 2.2 may now be completed by incorporating terms of order $O_p(n^{-C})$, for all $C > 0$, in all expansions, and noting that they are too small to have any first-order influence on the limiting distributions. In the reworked proof, sums that previously were over $1 \leq t \leq n$ are now over $\nu \leq t \leq n$. However, since $\nu/n \to 0$, then this alteration, too, does not affect the limiting distributions.

5.3. *Outline Proof of Theorem 3.2*

Recall that $\hat{\sigma}_t = \tilde{\sigma}_t(\tilde{a}, \tilde{b}, \tilde{c})$. Starting from (2.15), in which we take (a, b, c) equal to (a^0, b^0, c^0) in one instance and to $(\tilde{a}, \tilde{b}, \tilde{c})$ in another, and noting that each component of the estimators $\tilde{a}, \tilde{b}, \tilde{c}$ differs from its true value by $O_p(n^{-\xi})$, for some $\xi > 0$ (see Theorem 2.2), it may be proved that

$$\hat{\sigma}_t / \tilde{\sigma}_t(a^0, b^0, c^0) = 1 + O_p(n^{-\xi}) \tag{5.42}$$

uniformly in $\nu \le t \le n$, for some $\xi > 0$. The argument uses the assumption that (if $p \ge 1$) at least one of $a_1, \dots, a_p$ is nonzero, and (if $q \ge 1$) at least one of $b_1, \dots, b_q$ is nonzero. It also requires $c > 0$.

Note too that, in view of the results obtained in Section 5.2, and writing σ_t to denote $\sigma_t(a^0, b^0, c^0)$, we have

$$\sigma_t / \tilde{\sigma}_t(a^0, b^0, c^0) = 1 + O_p(n^{-C}) \tag{5.43}$$

uniformly in $\nu \le t \le n$, for all $C > 0$. Combining (5.42) and (5.43) we deduce that $\hat{\sigma}_t / \sigma_t = 1 + O_p(n^{-\xi})$ uniformly in $\nu \le t \le n$. Equivalently, $|(\tilde{\epsilon}_t / \epsilon_t) - 1| = O_p(n^{-\xi})$ uniformly in $\nu \le t \le n$. From this property, and the definition of $\hat{\epsilon}_t$ at (3.5), it may be deduced that for some $\xi > 0$,

$$\hat{\epsilon}_t = \epsilon_t \{1 + O_p(n^{-\xi})\} + O_p(n^{-\xi}), \tag{5.44}$$

where both "O_p" terms are of the stated orders uniformly in $\nu \le t \le n$.

Using (5.44) we may rework the proof of Theorem 2.1 in the bootstrap case, to obtain Theorem 3.2. When error variance is finite the argument hardly alters; when error variance is infinite but the error distribution is in the domain of attraction of the Normal law, we borrow an argument from Hall (1990); and when the error distribution is in the domain of attraction of a stable law with exponent $\alpha \in (1, 2)$ we note that, in view of (5.44), large values of ϵ_t^2 are identified with large values of $\hat{\epsilon}_t^2$. For example, for each fixed k we may show that with probability converging to 1 as $n \to \infty$, the indices $t_1, \dots, t_k$ at which the k largest values of ϵ_t^2 (for $1 \le t \le n$) occur are identical to the indices at which the k largest values of $\hat{\epsilon}_t^2$ (for $\nu \le t \le n$) occur.

Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia,

and

Department of Statistics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom.

Manuscript received May, 2001; final revision received February, 2002.

APPENDIX: PROOF OF LEMMA 5.1

Let $L(\cdot, \cdot)$ denote the Lévy metric, interpreted as in Step (iii) of Section 5.1. Suppose $\delta > 0$ is given. Using properties (5.33), choose $k(\delta) \ge 1$ so large that for all sufficiently large n,

$$\begin{aligned} &P\{\text{no more than the first } k(\delta) \text{ of } V_{n1}, \dots, V_{nn} \text{ exceed } C\} \ge 1 - \delta, \\ &P\{\text{no more than the first } k(\delta) \text{ of } V_1, V_2, \dots \text{ exceed } C\} \ge 1 - \delta. \end{aligned} \tag{A.1}$$

The assumptions in the lemma imply that, if we define

$$Q_k(n) = \sum_{i=1}^{k} U_{ni} I(V_{ni} > C) \quad \text{and} \quad Q_k = \sum_{i=1}^{k} U_i I(V_i > C),$$

then for each fixed k,

$$Q_k(n) \to Q_k \quad \text{in distribution} \tag{A.2}$$

as $n \to \infty$. Result (A.1) implies that for sufficiently large n,

$$P\{Q_{k(\delta)}(n) \neq Q_n(n)\} \leq \delta \text{ and } P\{Q_{k(\delta)} \neq Q_\infty\} \leq \delta,$$

and hence that for the same values of n,

$$L\{Q_{k(\delta)}(n), Q_n(n)\} \leq \eta(\delta), \quad L\{Q_{k(\delta)}, Q_\infty\} \leq \eta(\delta), \tag{A.3}$$

where $\eta(\delta)$ does not depend on n and converges to zero as $\delta \to 0$. In view of (A.2),

$$L\{Q_k(n), Q_k\} \to 0 \tag{A.4}$$

as $n \to \infty$. Combining (A.3) and (A.4) we deduce that

$$\begin{aligned} L\{Q_n(n), Q_\infty\} &\leq L\{Q_n(n), Q_{k(\delta)}(n)\} + L\{Q_{k(\delta)}(n), Q_{k(\delta)}\} + L\{Q_{k(\delta)}, Q_\infty\} \\ &\leq 2\eta(\delta) + o(1), \end{aligned}$$

where the first inequality holds for all sufficiently large n and the second as $n \to \infty$. Letting first $n \to \infty$ and then $\delta \to 0$, we deduce that $L\{Q_n(n), Q_\infty\} \to 0$, or equivalently, $Q_n(n) \to Q_\infty$ in distribution, as claimed in the lemma.

REFERENCES

Arov, D. Z., and A. A. Bobrov (1960): "The Extreme Members of a Sample and their Role in the Sum of Independent Variables," *Teorija Verojatnosteĭi ee Primenenija*, 5, 415–435.

Athreya, K. B. (1987a): "Bootstrap of the Mean in the Infinite Variance Case," *Annals of Statistics*, 15, 724–731.

——— (1987b): "Bootstrap of the Mean in the Infinite Variance Case," in *Proceedings of the 1st World Congress of the Bernoulli Society*, 2. Utrecht: VNU Scientific Press, 95–98.

Berkes, I., L. Horváth, and P. Kokoszka (2001): "GARCH Processes: Structure and Estimation," Preprint.

Bhansali, R. (1988): "Consistent Order Determination for Processes with Infinite Variance," *Journal of the Royal Statistical Society*, Series B, 50, 6–60.

——— (1995): "Recent Developments in Analysis of Time Series with Infinite Variance: A Review, in *Proceedings of the Athens Conference on Applied Probability and Time Series Analysis*, Vol II, Springer Lecture Notes in Statistics 115. New York: Springer, pp. 42–54.

——— (1997): "Robustness of the Autoregressive Spectral Estimate for Linear Processes with Infinite Variance, *Journal of Time Series Analysis*, 18, 213–229.

Bickel, P., F. Götze, and W. R. van Zwet (1995): "Resampling Fewer than n Observations: Gains, Losses, and Remedies for Losses," *Statistica Sinica*, 7, 1–31.

Bingham, N. H., C. M. Goldie, and J. Teugels (1987): *Regular Variation.* Cambridge: Cambridge University Press.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

BOLLERSLEV, T., AND J. M. WOOLDRIDGE (1992): "Quasi Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," *Econometric Reviews*, 11, 143–172.

BOSE, A. (1988): "Edgeworth Correction by Bootstrap in Autoregressions," *Annals of Statistics*, 16, 1709–1722.

BOUGEROL, P., AND N. PICARD (1992): "Stationarity of GARCH Processes and of some Nonnegative Time Series," *Journal of Econometrics,* 52, 115–127.

BROWN, B. M. (1971): "Martingale Central Limit Theorems," *Annals of Mathematical Statistics*, 42, 59–66.

CHAN, N. H. (1990): "Inference for Near-Integrated Time Series with Infinite Variance," *Journal of the American Statistical Association*, 85, 1069–1074.

——— (1993): "On the Noninvertible Moving Average Time Series with Infinite Variance," *Econometric Theory*, 9, 680–685.

CHAN, N. H., AND L. T. TRAN (1989): "On the First-Order Autoregressive Process with Infinite Variance," *Econometric Theory*, 5, 354–362.

CLINE, D. B. H. (1989): "Consistency for Least-Squares Regression Estimators with Infinite Variance Data," *Journal of Statistical Planning and Inference*, 23, 163–179.

CLINE, D. B. H., AND P. J. BROCKWELL (1985): "Linear Prediction of ARMA Processes with Infinite Variance," *Stochastic Processes and their Applications*, 19, 281–296.

COMTE, F., AND O. LIEBERMAN (2000): "Asymptotic Theory for Multivariate GARCH Processes," Preprint.

DARLING, D. A. (1952): "The Influence of the Maximum Term in the Addition of Independent Random Variables," *Transactions of the American Mathematical Society*, 73, 95–107.

DAVIS, R. A., AND T. MIKOSCH (1998): "The Sample Autocorrelations of Heavy-Tailed Processes with Applications to ARCH," *Annals of Statistics*, 26, 2049–2080.

DAVIS, R. A., AND S. RESNICK (1985a): "Limit Theory for Moving Averages of Random Variables with Regularly Varying Tail Probabilities," *Annals of Probability*, 13, 179–195.

——— (1985b): "More Limit Theory for the Sample Correlation Function of Moving Averages," *Stochastic Processes and their Applications*, 20, 257–279.

——— (1989): "Basic Properties and Prediction of Max-ARMA Processes," *Advances in Applied Probability*, 21, 781–803.

——— (1996): "Limit Theory for Bilinear Processes with Heavy-Tailed Noise," *Annals of Applied Probability*, 6, 1191–1210.

DWASS, M. (1966): "Extremal Processes. II," *Illinois Journal of Mathematics*, 10, 381–391.

EMBRECHTS, P., C. KLÜPPELBERG, AND T. MIKOSCH (1997): *Modelling Extremal Events.* Berlin: Springer-Verlag.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

ENGLE, R. F., AND G. GONZALEZ-RIVERA (1991): "Semiparametric ARCH Models," *Journal of Economics and Business Statistics*, 9, 345–359.

FAMA, E. (1965): "The Behaviour of Stock Market Prices," *Journal of Business*, 38, 34–105.

FELLER, W. (1966): *An Introduction to Probability Theory and its Applications.* New York: Wiley.

GOLDIE, C. M. (1991): "Implicit Renewal Theory and Tails of Solutions of Random Equations," *Annals of Applied Probability*, 1, 126–166.

GIRAITIS, L., P. KOKOSZKA, AND R. LEIPUS (2000): "Stationary ARCH Models: Dependence Structure and Central Limit Theorem," *Econometric Theory*, 16, 3–22.

GIRAITIS, L., AND P. M. ROBINSON (2001): "Whittle Estimation of ARCH Models," *Econometric Theory*, 17, 608–623.

GOURIÉROUX, C. (1997): *ARCH Models and Financial Applications.* New York: Springer-Verlag.

GROSS, S., AND W. L. STEIGER (1979): "Least Absolute Deviation Estimates in Autoregression with Infinite Variance," *Journal of Applied Probability*, 16, 104–116.

HALL, P. (1978a): "Representations and Limit Theorems for Extreme Value Distributions," *Journal of Applied Probability*, 15, 639–644.

——— (1978b): "On the Extreme Terms of a Sample from the Domain of Attraction of a Stable Law," *Journal of the London Mathematical Society*, 18, 181–191.

——— (1990): "Asymptotic Properties of the Bootstrap for Heavy-Tailed Distributions," *Annals of Probability*, 18, 1342–1360.

Hall, P., and R. Lepage (1996): "On Bootstrap Estimation of the Distribution of the Studentized Mean," *Annals of the Institute of Statistical Mathematics*, 48, 403–421.

Hannan, E. J., and M. Kanter (1977): "Autoregressive Processes with Infinite Variance," *Journal of Applied Probability*, 14, 411–415.

Hsing, T. (1999): "On the Asymptotic Distributions of Partial Sums of Functionals of Infinite-Variance Moving Averages," *Annals of Probability*, 27, 1579–1599.

Kanter, M., and W. Steiger (1974): "Regression and Autoregression with Infinite Variance," *Advances in Applied Probability*, 6, 768–783.

Kesten, H. (1973): "Random Difference Equations and Renewal Theory for Products of Random Matrices," *Acta Mathematica*, 131, 207–248.

Knight, K. (1987): "Rate of Convergence of Centred Estimates of Autoregressive Parameters for Infinite Variance Autoregressions," *Journal of Time Series Analysis*, 8, 51–60.

——— (1989a): "Consistency of Akaike's Information Criterion for Infinite Variance Autoregressive Processes," *Annals of Statistics*, 17, 824–840.

——— (1989b): On the Bootstrap of the Sample Mean in the Infinite Variance Case," *Annals of Statistics*, 17, 1168–1175.

——— (1993): "Estimation in Dynamic Linear Regression Models with Infinite Variance Errors," *Econometric Theory*, 9, 570–588.

Kokoszka, P. S., and T. Mikosch (1997): "The Integrated Periodogram for Long-Memory Processes with Finite or Infinite Variance," *Stochastic Processes and their Applications*, 66, 55–78.

Kokoszka, P. S., and M. S. Taqqu (1994): "Infinite Variance Stable ARMA Processes," *Journal of Time Series Analysis*, 15, 203–220.

——— (1996a): "Parameter Estimation for Infinite Variance Fractional ARIMA," *Annals of Statistics*, 24, 1880–1913.

——— (1996b): "Infinite Variance Stable Moving Averages with Long Memory," *Journal of Econometrics*, 73, 79–99.

——— (1999): "Discrete Time Parametric Models with Long Memory and Infinite Variance," *Mathematical and Computer Modelling*, 29, 203–215.

Koul, H. L., and D. Surgailis (2001): "Asymptotics of Empirical Processes of Long Memory Moving Averages with Infinite Variance," *Stochastic Processes and their Applications*, 91, 309–336.

Lee, S.-W., and B. E. Hansen (1994): "Asymptotic Theory for the GARCH(1, 1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

Leipus, R., and M.-C. Viano (2000): "Modelling Long-Memory Time Series with Finite or Infinite Variance: A General Approach," *Journal of Time Series Analysis*, 21, 61–74.

Linton, O. (1993): "Adaptive Estimation in ARCH Models," *Econometric Theory*, 9, 539–569.

Lumsdaine, R. L. (1996): "Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1, 1) and Covariance Stationary GARCH(1, 1) Models," *Econometrica*, 64, 575–596.

Mammen, E. (1992): *When Does Bootstrap Work? Asymptotic Results and Simulations*, Lecture Notes in Statistics, 77. New York: Springer.

Mandelbrot, B. (1963): "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.

Mikosch, T., T. Gadrich, C. Klüppelberg, and R. Adler (1995): "Parameter Estimation for ARMA Models with Infinite Variance Innovations," *Annals of Statistics*, 23, 305–326.

Mikosch, T., and C. Klüppelberg (1995): "On Strong Consistency of Estimators for Infinite Variance Time Series," *Theory of Probability and Mathematical Statistics*, 53, 127–136.

Mikosch, T., and C. Starica (2000): "Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1, 1) Process, *Annals of Statistics*, 28, 1427–1451.

Mikosch, T., and D. Straumann (2001): "Whittle Estimation in a Heavy-Tailed GARCH(1, 1) Model," Manuscript.

Mittnik, S., and S. T. Rachev (2000): *Stable Paretian Models in Finance.* New York: Wiley.

MITTNIK, S., S. T. RACHEV, AND M. S. PAOLELLA (1998): "Stable Paretian Modeling In Finance: Some Empirical and Theoretical Aspects, in *A Practical Guide to Heavy Tails*, ed. by R. J. Adler, R. E. Feldman, and M. S. Taqqu. Boston: Birkhäuser, pp. 79–110.

NELSON, D. B. (1990): "Stationarity and Persistence in the GARCH(1, 1) Model," *Econometric Theory*, 6, 318–334.

PHILLIPS, P. C. B. (1990): "Time Series Regression with a Unit Root and Infinite-Variance Errors, *Econometric Theory*, 6, 44–62.

POLITIS, D., J. P. ROMANO, AND M. WOLF (1999): *Subsampling.* New York: Springer.

RESNICK, S. I. (1986): "Point Processes, Regular Variation and Weak Convergence," *Advances in Applied Probability*, 18, 66–138.

——— (1987): *Extreme Values, Regular Variation, and Point Processes.* New York: Springer-Verlag.

RYDBERG, T. (2000): "Realistic Statistical Modelling of Financial Data," *International Statistical Review*, 68, 233–258.

SAMORODNITSKY, G., AND M. S. TAQQU (1994): *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance. Stochastic Modeling.* New York: Chapman and Hall.

SHEPHARD, N. (1996): "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen. London: Chapman and Hall, pp. 1–67.

TAYLOR, S. J. (1986): *Modelling Financial Time Series.* Chichester: John Wiley.

WEISS, A. (1986): "Asymptotic Theory for ARCH Models: Estimation and Testing," *Econometric Theory*, 2, 107–131.

YOHAI, V. J., AND R. A. MARONNA (1977): "Asymptotic Behavior of Least-Squares Estimates for Autoregressive Processes with Infinite Variances," *Annals of Statistics*, 5, 554–560.