RESEARCH REPORT SERIES
(Statistics #2007-2)

Tail Index Estimation for Parametric Families Using Log Moments

Tucker McElroy

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: January 24, 2007

Tail Index Estimation for Parametric Families Using Log Moments

Tucker McElroy
U.S. Census Bureau

Abstract

For heavy-tailed econometric data it is of interest to estimate the tail index, a parameter that measures the thickness of the tails of the marginal distribution. Common models for such distributions include Pareto and t distributions, and in other applications (such as hydrology) stable distributions are popular as well. This paper constructs square root n consistent estimators of the tail index that are independent of the scale of the data, which are based on an assumed knowledge of the parametric family for the marginal distribution. Given the popularity of parametric modeling for economic time series, this method gives an appealing alternative to nonparametric tail index estimators – such as the Hill and Pickands estimators – that are appropriate when the modeler believes that the data belongs to a certain known parametric family of distributions. The method works fairly well for stationary time series with intermediate memory and infinite variance, and since it is parametric does not depend upon blocking or tuning parameters. Small sample results and full asymptotics are provided in this paper, and simulation studies on various heavy-tailed time series models are given as well.

Keywords. Extreme Value Theory, Heavy Tails, Stable Distributions.

1 Introduction

The last several decades have seen a multiplication of papers that address the presence of heavy tails in economic time series – see Embrechts, Klüppelberg, and Mikosch

(1997) for an overview. In econometics the Student t and Pareto distributions are especially popular, as they are felt to give a reasonable fit (in terms of shape) to a variety of economic data; the wide range of the tail index for these distributions grants a flexibility in modeling that is advantageous. From a more theoretical standpoint, the stable distributions are preferable due to their elegant mathematical properties; see Davis and Resnick (1986). Each of these distributions depends on one key parameter – the tail index – which describes each family of distributions. Clearly, any such heavy-tailed time series model must have a consistent and reliable estimate of the tail index.

Of course, use of the Pareto, Student t, and stable distributions for modeling reflects a parametric approach to time series analysis. Alternatively, one may consider a nonparametric approach, where the marginal distribution is never specified. Starting with the Hill estimator (Hill, 1975) – but including the DEdH, Pickands, and others, see Embrechts et al. (1997) – the literature is saturated with efforts to improve bandwidth selection and optimize these nonparametric estimators for each particular data set. It is natural that improved results can be obtained by using a parametric approach; of course these results are only improved if one truly believes that the data follows the specified parametric distribution. This is the common trade-off between nonparametric and parametric methods – robustness versus accuracy. For this paper, we will focus on deriving tail index estimators in the parametric situation.

In particular, we consider that the data $X_1, X_2, \cdots, X_n$ is an observed stretch of a strictly stationary time series, which may or may not be square integrable. Since we are interested in a scale-free tail index estimate, we develop a target parameter that is a simple invertible function of the unknown tail index α, which is defined in (1) below. The target parameter is the variance of $\log |X|$, where X denotes a common version. Taking the logarithm serves a two-fold purpose: firstly, this separates out the scale parameter; secondly, taking the log ensures that all moments (in particular the second) will exist. It follows from the first point that the variance of $\log |X|$ will not depend on the scale parameter. The method depends upon $Var \log |X|$ being an easily computable function that can be inverted as a function of α.

Section 2 of the paper develops the general method discussed above for *iid* data, and we apply this to the particular case of the stable, Student t, and log-Gamma distributions. In Section 3 we extend the method to long-range dependent time series data, and introduce the class of Noah-Joseph models for simultaneously modeling heavy tails and serial dependence. The method is extended to this general class of processes, and we provide additional illustration through a Pareto-like distribution. In Section 4 we test the method in simulation studies, and we make our conclusions in Section 5. Proofs are contained in a separate Appendix.

2 Basic Theory

In this section we lay out the basic estimator of the tail index, and indicate conditions under which the large sample asymptotics can be known. Section 2.1 discusses the general theory, and the estimator's properties for *iid* data are given in Theorem 1. In Section 2.2 several applications are further developed that are applicable for *iid* data; in particular, we discuss the stable, student t, and log-gamma distributions. For the Pareto distribution, the requisite moment function is difficult to compute; in section 3 we discuss the case of a "Pareto-like" distribution.

2.1 Asymptotic Theory

We suppose that the observed data $X_1, X_2, \cdots, X_n$ is strictly stationary, and is heavy-tailed with tail index α. The notion of tail index is defined in many different ways in the literature, but perhaps the most common (Embrechts et al., 1997) is

$$\mathbb{P}[|X| > x] \sim CL(x)x^{-\alpha} \tag{1}$$

for some constant $C > 0$, a slowly-varying function L, and x large (the notation $f(x) \sim g(x)$ denotes that $f(x)/g(x) \to 1$ as $x \to \infty$). The above definition can be generalized to allow for separate left and right tail behavior, but we will assume that the tail index α is the same for each. Below we describe a few examples.

Pareto. Since $\mathbb{P}[|X| > x] = (\frac{\kappa}{\kappa+x})^{\alpha}$ by definition (Embrechts et al., 1997), the tail index is obviously α. Here κ plays the role of a scale parameter.

Stable. The stable distribution depends in general on four parameters: the location μ, the scale σ, the skewness β, and the characteristic exponent $\alpha \in (0, 2]$ (see Samorodnitsky and Taqqu, 1994). By Proposition 1.2.15 of that work,

$$\mathbb{P}[|X| > x] \sim \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}\sigma^\alpha x^{-\alpha}$$

where Γ denotes the gamma function. Hence the characteristic exponent equals the tail index α.

Student t. We consider the Student t on α degrees of freedom, where α is allowed to be any positive real number. The probability density function (pdf) for X is given by

$$\frac{\Gamma((\alpha+1)/2)}{\sqrt{\pi\alpha}\Gamma(\alpha/2)}\left(1 + x^2/\alpha\right)^{-(\alpha+1)/2}. \tag{2}$$

From this form it can be deduced that the degrees of freedom equals the tail index.

Log-Gamma. From Embrechts et al. (1997) we have the pdf for $x \geq 1$ given by

$$\frac{\alpha^\beta}{\Gamma(\beta)}(\log x)^\beta x^{-(\alpha+1)}.$$

Here β is the shape parameter, and α is the rate. Essentially the log term functions as a slowly-varying function, and so the tail index equals the rate α.

Now we will assume that the data is mean zero (or if the mean does not exist because $\alpha \leq 1$, then we assume that the location parameter is zero). Hence we can write $X_t = \sigma Z_t$, where σ is a scale parameter, and Z_t is a member of the same parametric family (in the Student t and log-gamma examples above, the density was stated for $\sigma = 1$). Then $\log|X_t| = \log\sigma + \log|Z_t|$ and

$$Var\log|X_t| = Var\log|Z_t|.$$

The above equation allows us to remove scale considerations from our tail index estimate, which is very convenient. Now a convenient way of computing $Var\log|Z_t|$ is needed, and the following method is helpful in the case of the stable and Student t distributions. Let $\phi(r) = \mathbb{E}[|Z|^r]$, so that

$$\dot\phi(0) = \mathbb{E}[\log|Z|] \qquad \ddot\phi(0) = \mathbb{E}[\log^2|Z|].$$

We use the $\dot{f}$ notation to denote the derivative of the function f. From this it follows that

$$Var \log |X| = \ddot{\phi}(0) - \left(\dot{\phi}(0)\right)^2 = \frac{d^2}{dr^2} \log \phi(r)|_{r=0},$$

where the last equality follows from $\phi(0) = 1$. Note that the existence of the derivatives of ϕ at the origin depend on a suitably small probability content there; it is sufficient that the pdf is bounded in a neighborhood of zero.

Now the method relies on $Var \log |X|$ being a fairly simple function of α – say $g(\alpha)$ – where g is invertible. The following heuristic justifies this procedure: supposing that Z is supported on $[1, \infty)$ as in the log-gamma case, we have

$$\mathbb{E}[\log |Z|] = \int_0^\infty \mathbb{P}[\log |Z| > x]\, dx = \int_0^\infty \mathbb{P}[|Z| > e^x]\, dx$$

and this tail probability behaves asymptotically like

$$\mathbb{P}[|Z| > e^x] \sim CL(e^x)e^{-\alpha x}$$

by (1). Hence the expected log moment looks roughly like the Laplace transform – evaluated at α – of $L(e^x)$, which may very well be a simple function of α. Section 2.2 illustrates that this is indeed the case for the stable, Student t, and log-gamma distributions. Assuming that g^{-1} exists (or at least that it has only one fiber on the acceptable range of tail indexes $(0, \infty)$), we use the statistic

$$g^{-1}\left(\widehat{Var} \log |X|\right) = \hat{\alpha} \tag{3}$$

as our estimator. Here $\widehat{Var}$ denotes the sample variance statistic; let $W_t = \log |X_t| - \mathbb{E} \log |X_t|$. The following theorem summarizes the statistical properties of $\hat{\alpha}$:

Theorem 1 *Suppose that X_t is an iid sequence of random variables with location zero, such that the first two derivatives of $\phi(r)$ exist at the origin. Then*

$$\sqrt{n}\left(\widehat{Var} \log |X| - Var \log |X|\right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, V)$$

as $n \to \infty$. The limiting variance $V = Var W_t^2$. If the derivative of g^{-1} exists at $g(\alpha)$, then

$$\sqrt{n}\left(\hat{\alpha} - \alpha\right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, V/\dot{g}^2(\alpha)\right).$$

The proof of the theorem is in the appendix. It is possible to obtain some finite sample properties as well, such as the mean and variance of the estimate; these results will be discussed case by case in the examples below. The general expression for V can be worked out:

$$V = \mathbb{E}W_t^4 - (\mathbb{E}W_t^2)^2 = \ddddot{\phi}(0) - 4\dddot{\phi}(0)\dot{\phi}(0) + 8\ddot{\phi}(0)\dot{\phi}^2(0) - \ddot{\phi}^2(0) - 4\dot{\phi}^4(0). \quad (4)$$

It can be shown that this is equal to

$$V = \frac{d^4}{dr^4}\log\phi(r)|_{r=0} + 2\left(\frac{d^2}{dr^2}\log\phi(r)|_{r=0}\right)^2, \quad (5)$$

which may be easier to compute. In general this variance quantity will depend on α, and thus could be estimated by substituting $\hat{\alpha}$.

2.2 Applications to *iid* Models

We first consider the case that the sample consists of *iid* observations from either a stable, Student t, or log-Gamma distribution. In each case, we compute $g(\alpha)$ and determine its inverse so that $\hat{\alpha}$ can be constructed. We also compute the asymptotic variance of $\hat{\alpha}$ explicitly. In the case of the log-Gamma, a nuisance parameter β is dealt with by computing the kurtosis of the logged data.

Stable. The first step is to use the representation

$$Z = \sqrt{\epsilon} \cdot G$$

for independent random variables ϵ and G, where ϵ is distributed as a positively skewed $\alpha/2$ stable with unit scale (denoted $S_{\alpha/2}(1,1,0)$) and G is a standard normal. This representation is found in Samorodnistky and Taqqu (1994). Then $Z^2 = \epsilon \cdot Y$, where Y is a χ^2 random variable with one degree of freedom. Thus

$$\phi(r) = \mathbb{E}[(Z^2)^{r/2}] = \mathbb{E}[\epsilon^{r/2}]\mathbb{E}[Y^{r/2}]$$

follows from independence. Now $\mathbb{E}[Y^{r/2}] = 2^{r/2}\Gamma((r+1)/2)\pi^{-1/2}$, and from a result due to Hardin (1984) we have

$$\mathbb{E}[\epsilon^{r/2}] = \frac{2^{r/2-1}\Gamma(1-r/\alpha)}{\int_0^\infty \frac{r}{2}x^{-(r/2+1)}\sin^2 x\,dx}(\sec(\alpha\pi/4))^{r/\alpha}\cos(r\pi/4).$$

Some elementary real analysis of the term

$$\int_0^\infty t x^{-(t+1)} \sin^2 x \, dx$$

shows that it is integrable; using integration by parts and a change of variable, it can be expressed as

$$\frac{2^t}{2} \int_0^\infty x^{-t} \sin x \, dx = 2^{t-1} \frac{\Gamma(2-t)\cos(t\pi/2)}{1-t}$$

Substituting this and taking logs, we obtain

$$\begin{aligned} \log \phi(r) = \log \Gamma(1 - r/\alpha) - \log \Gamma(2 - r/2) + \log(1 - r/2) + \frac{r}{\alpha} \log \sec(\alpha\pi/4) \\ + \frac{r}{2} \log 2 - \frac{1}{2} \log \pi + \log \Gamma((r+1)/2). \end{aligned} \tag{6}$$

The second derivative at zero yields a simple formula for $g(\alpha)$:

$$g(\alpha) = \Psi_2(1)(1/\alpha^2) - \Psi_2(2)(1/4) - 1/4 + \Psi_2(1/2)(1/4),$$

where Ψ_k denotes the kth derivative of the log-gamma function. Now since $\Gamma(x+1) = x\Gamma(x)$, we have

$$\begin{aligned} \Psi_0(x+1) &= \log x + \Psi_0(x) \\ \Psi_1(x+1) &= 1/x + \Psi_1(x) \\ \Psi_2(x+1) &= -1/x^2 + \Psi_2(x) \end{aligned}$$

so that $\Psi_2(2) = \Psi_2(1) - 1$. This yields

$$\begin{aligned} g(\alpha) &= \frac{1}{4} \big(\Psi_2(1)(4/\alpha^2 - 1) + \Psi_2(1/2)\big) \\ g^{-1}(x) &= \frac{2}{\sqrt{1 + \frac{4x - \Psi_2(1/2)}{\Psi_2(1)}}}, \end{aligned}$$

where the inverse is guaranteed to exist because $\dot{g}(\alpha) = -2\alpha^{-3}\Psi_2(1)$ is always nonzero. Now using (5) to compute V, we have

$$\begin{aligned} V = &\frac{1}{16} \big(\Psi_4(1)(16/\alpha^4) - \Psi_4(2) - 6 + \Psi_4(1/2)\big) \\ &+ \frac{1}{8} \big(\Psi_2(1)(4/\alpha^2 - 1) + \Psi_2(1/2)\big)^2. \end{aligned}$$

Since $\Psi_k(x+1) = (-1)^{k-1}(k-1)! x^{-k} + \Psi_k(x)$, we have $\Psi_4(2) = -6 + \Psi_4(1)$, and hence the asymptotic variance is

$$\frac{V}{\dot{g}^2(\alpha)} = \frac{\alpha^6}{4\Psi_2^2(1)} \left\{ \frac{1}{16} \big(\Psi_4(1)(16\alpha^{-4} - 1) + \Psi_4(.5)\big) + \frac{1}{8} \big(\Psi_2(1)(4\alpha^{-2} - 1) + \Psi_2(.5)\big)^2 \right\}.$$

Thus the variance clearly decreases as α gets closer to zero.

Student t. From the pdf in (2), we can obtain the moment function ϕ. The square of a Student t on α degrees of freedom is an F distribution on $(1, \alpha)$ degrees of freedom (Bickel and Doksum, 1977). This provides a formula for $\phi(2r)$, from which we can deduce the following formulas for $\log \phi(r)$:

$$\log \phi(r) = \frac{r}{2} \log \alpha + \log \Gamma((r+1)/2) + \log \Gamma((\alpha - r)/2) - \log \Gamma(1/2) - \log \Gamma(\alpha/2).$$

The second derivative at zero gives the formula for $g(\alpha)$:

$$g(\alpha) = \frac{1}{4} \left(\frac{\ddot{\Gamma}(\alpha/2)}{\Gamma(\alpha/2)} - \frac{\dot{\Gamma}^2(\alpha/2)}{\Gamma^2(\alpha/2)} + \frac{\ddot{\Gamma}(1/2)}{\Gamma(1/2)} - \frac{\dot{\Gamma}^2(1/2)}{\Gamma^2(1/2)} \right) = \frac{1}{4}(\Psi_2(\alpha/2) + \Psi_2(1/2)).$$

Now the derivative of the inverse of g evaluated at $g(\alpha)$ is equal to $1/\dot{g}(\alpha)$ – as discussed in the proof of Theorem 1 – which is equal to $-8/\Psi_3(\alpha/2)$. Since $\dot{g}(\alpha)$ is nonzero for any α, the inverse of g exists at α by the Inverse Function Theorem. Analytical expressions for Ψ_k are not available, but g^{-1} can be determined by using a look-up table. So $\hat{\alpha}$ can be calculated in practice, and is asymptotically normal. In order to compute its asymptotic variance, note that in general for $k > 1$,

$$\frac{d^k}{dr^k} \log \phi(r)|_{r=0} = 2^{-k} \left(\Psi_k(\alpha/2) + \Psi_k(1/2) \right).$$

Thus using (5) the asymptotic variance is

$$\frac{V}{\dot{g}^2(\alpha)} = 4 \frac{\left(\Psi_4(\alpha/2) + \Psi_4(1/2) + 2(\Psi_2(\alpha/2) + \Psi_2(1/2))^2 \right)}{\Psi_3^2(\alpha/2)}.$$

Log-gamma. In the case of the Log-Gamma distribution, the log moments are just the moments of a Gamma distribution, and thus are computed quite easily. In general,

$$\mathbb{E}[\log^k |Z|] = \frac{1}{\alpha^k} \frac{\Gamma(\beta + k)}{\Gamma(\beta)}$$

so that $g(\alpha) = \beta \alpha^{-2}$. Since this depends on the unknown shape parameter β, we propose calculating a separate quantity that just depends on β. The kurtosis of $\log |X|$ is scale-independent and is defined by

$$Kur \log |X| = \frac{\mathbb{E}[(\log |X| - \mathbb{E} \log |X|)^4]}{Var^2 \log |X|} - 3.$$

Now the centered fourth log moment is

$$
\begin{aligned}
&\mathbb{E}[(\log|X| - \mathbb{E}\log|X|)^4] \\
&= \mathbb{E}\log^4|Z| - 4\mathbb{E}\log^3|Z|\,\mathbb{E}\log|Z| + 6\mathbb{E}\log^2|Z|\,\mathbb{E}^2\log|Z| - 3\mathbb{E}^4\log|Z| \\
&= \frac{1}{\alpha^4}\left((\beta+3)(\beta+2)(\beta+1)\beta - 4(\beta+2)(\beta+1)\beta^2 + 6(\beta+1)\beta^3 - 3\beta^4\right) \\
&= \frac{3(\beta+2)\beta}{\alpha^4}.
\end{aligned}
$$

Thus the kurtosis is equal to $6/\beta$. So one way to proceed is to first estimate the sample kurtosis and obtain an estimate of β:

$$\hat{\beta} = 6/\widehat{Kur}\log|X|$$

This will be consistent for β, since all the log moments exist. Now $g^{-1}(x) = \sqrt{\beta/x}$, and we can substitute $\hat{\beta}$ in for β. So our modified estimate for α is

$$\hat{\alpha} = \sqrt{\hat{\beta}/\widehat{Var}\log|X|} = \frac{\sqrt{6}}{\sqrt{\widehat{Kur}\ \log|X|\widehat{Var}\log|X|}}.$$

For the asymptotic normality result, we have

$$\hat{\alpha} - \alpha = \sqrt{\frac{Kur\log|X|}{\widehat{Kur}\log|X|}}\left(g^{-1}(\widehat{Var}\log|X|) - \alpha\right),$$

and the stochastic term $Kur\log|X|/\widehat{Kur}\log|X|$ tends to one in probability. Therefore $\sqrt{n}(\hat{\alpha} - \alpha)$ is asymptotically normal with variance $V/\dot{g}^2(\alpha)$ as in Theorem 1. Now $\dot{g}(\alpha) = -2\beta\alpha^{-3}$, which is always nonzero, guaranteeing the existence of g^{-1}. As for V, it is easier to work with (4):

$$
\begin{aligned}
V &= \frac{1}{\alpha^4}\left((\beta+3)(\beta+2)(\beta+1)\beta - 4(\beta+2)(\beta+1)\beta^2 + 8(\beta+1)\beta^3 - (\beta+1)^2\beta^2 - 4\beta^4\right) \\
&= \alpha^{-4}2\beta(\beta+3)
\end{aligned}
$$

Thus the asymptotic variance of $\hat{\alpha}$ is $\frac{V}{\dot{g}^2(\alpha)} = \frac{\alpha^2}{2}(1+3/\beta)$. This could be easily estimated using $\hat{\beta}$ and $\hat{\alpha}$.

3 Extended Theory

Since it is of interest to examine the method on serially correlated time series data, Section 3.1 introduces a general **Noah-Joseph** Model that allows for both heavy tails and long range dependence. In Section 3.2 Theorem 2 gives the tail index estimator's properties for data from these types of models. We discuss particular applications to Noah-Joseph models for stable, Student t, and Pareto-like distributed time series.

3.1 Noah-Joseph Models

We now turn to models that are heavy-tailed and serially correlated. Perhaps one of the first models of this kind was the $MA(\infty)$ stable process of Davis and Resnick (1986). Also see Samorodnitsky and Taqqu (1994) for the stable integral moving average model, and Kokoszka and Taqqu (1999) for linear heavy-tail, long memory models. In McElroy and Politis (2006), a new non-linear model that incorporates both heavy-tails and long memory was introduced. This model, once appropriately generalized, is particularly convenient for defining serially dependent stable and Student t models; other heavy-tailed distributions can also be handled. We let each X_t be given as the product of a "volatility" series and a serially correlated Gaussian time series:

$$X_t = \sigma_t \cdot G_t.$$

The volatility time series $\{\sigma_t\}$ is independent of $\{G_t\}$, which can be a short, intermediate, or long memory Gaussian process. The volatility process is typically taken to be *iid*, since the Gaussian process is designed to incorporate all of the serial dependence structure. In general, the volatility series should be some positive time series that makes the data heavy-tailed. We discuss three cases:

- The volatility series is deterministic but time-varying. This results in a heteroscedastic, serially correlated Gaussian process.
- We have $\sigma_t = \sqrt{\epsilon_t}$, where $\epsilon \sim iid S_{\alpha/2}(1,1,0)$ for $\alpha \in (0,2)$, i.e., it is a positively skewed $\alpha/2$-stable random variable with unit scale and location zero. Then (see McElroy and Politis, 2006) X_t is symmetric α-stable with infinite variance; the autocorrelations at nonzero lags exist and are the same as the autocorrelations of $\{G_t\}$.
- We have $\sigma_t = \sqrt{\alpha/Y_t}$ for $\alpha > 0$, where Y_t is *iid* χ^2 on α degrees of freedom. Then X_t has a Student t distribution on α degrees of freedom, and has infinite variance if $\alpha \leq 2$. At nonzero lags, the autocorrelations exist and are the same as the Gaussian autocorrelations.

The second and third cases give general serially dependent stable and Student t models. The latter case is particularly appealing, since a simple moving average model with Student t inputs will not yield a Student t distribution (though this is true of stable random variables). In general, letting $\epsilon_t = \sigma_t^2$ be *iid*, we have the characteristic function of X_t given by – for any real θ –

$$\mathbb{E}\exp\{i\theta\sigma_t G_t\} = \mathbb{E}[\exp\{-\theta^2\epsilon_t/2\}]$$

using conditional expectations. This latter expression is the Laplace transform of the distribution of ϵ_t evaluated at $\theta^2/2$. By setting this expression equal to the characteristic function of a desired heavy-tailed distribution – such as a symmetric Pareto, Burr, log-Normal, or Weibull – one can in theory work out the requisite distribution of ϵ. The general technique is as follows:

$$\int_{-\infty}^{\infty} e^{i\theta x} f_X(x)\,dx = \int_0^{\infty} e^{-\theta^2 x/2} f_\epsilon(x)\,dx$$
$$f_X(x) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-x^2/(2z)} f_\epsilon(z)\, z^{-1/2}\,dz$$

The second equation follows from the first by applying the Fourier Inversion Theorem (Billingsley, 1995) and taking the inverse Fourier Transform of the function $e^{-\cdot\theta^2/2}$ after interchanging the order of integration. It can be interpreted as integrating the Gaussian pdf with respect to its variance z, weighted by $f_\epsilon(z)$. Now this second equation cannot in the general case be solved for f_ϵ in terms of f_X, but some special cases may be worked out. For example, Pareto-like tail behavior can be generated by appropriate selection of f_ϵ, which we describe below. If we know the volatility pdf f_σ, we can write

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-y^2/2} y^{-1} f_\sigma(x/y)\,dy.$$

Now consider the following pdf, which is compactly supported on $z \geq 1$:

$$f_\sigma(z) = \alpha z^{-(\alpha+1)}.$$

Then it follows that

$$f_X(x) = \alpha|x|^{-(\alpha+1)} \frac{1}{\sqrt{2\pi}} \int_0^{|x|} e^{-y^2/2} y^\alpha\,dy,$$

which can be rewritten as

$$\frac{\alpha}{1+|x|^{\alpha+1}} L(|x|)$$

for a slowly-varying function L that tends to

$$C = (2\pi)^{-1/2} \int_0^\infty e^{-y^2/2} y^\alpha \, dy = 2^{\alpha/2-1}\Gamma((\alpha+1)/2)\pi^{-1/2}$$

as $|x| \to \infty$ and tends to $1/(\alpha+1)$ as $x \to 0$. Clearly X has Pareto-like tails. In this way, a general time series model with serially dependent variables and general tail index α can be constructed. Such models will be referred to as **Noah-Joseph** Models, after the terminology of Mandelbrot and Wallis (1968) for heavy tails and long range dependence. This type of model seems useful because it is flexible, encompassing a wide-class of heavy-tailed dependent time series. Moreover one can specify an exact marginal distribution for any desired auto-correlation structure; this is in contrast to $MA(\infty)$-type models, where there is in general no relationship between the distribution of input and output variables (the exception being the stable family). In addition, the tail index estimator of this paper works very nicely on these types of models (when α is small and there is only intermediate memory), assuming that one has a knowledge of the parametric family of the volatility series.

3.2 Tail Index Estimation for Noah-Joseph Models

We now assume that the time series $\{X_t\}$ follows a Noah-Joseph model, such that the Gaussian process $\{G_t\}$ has absolutely summable autocovariance function (ACF), and the tail index is any $\alpha > 0$. We define the sequence $W_t = \log|X_t| - \mathbb{E}\log|X_t|$, which is used in the statement of Theorem 2 below.

Theorem 2 *Suppose that the process $\{X_t\}$ follows a Noah-Joseph model with absolutely summable Gaussian ACF. Suppose that the marginal distribution has location zero, such that the first two derivatives of $\phi(r)$ exist at the origin. Then*

$$\sqrt{n}\left(\widehat{Var}\log|X| - Var\log|Z|\right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0,V)$$

as $n \to \infty$. The limiting variance $V = \sum_{h=-\infty}^{\infty} \gamma_{W^2}(h)$, the sum of the ACF of W_t^2. If the derivative of g^{-1} exists at $g(\alpha)$, then

$$\sqrt{n}\,(\hat{\alpha} - \alpha) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, V/\dot{g}^2(\alpha)\right).$$

This theorem is proved in the appendix; an expression for V can be found in the proof. The method is easily applied to Noah-Joseph models for the stable, Student t, and Pareto-like distribution. For the stable class $\alpha < 2$ implies that the variance is infinite, but since the volatility series is *iid*, the autocovariances at nonzero lags exist so long as $\alpha > 1$. ($\alpha = 2$ corresponds to the Gaussian case, and the volatility series is just a deterministic constant.) So we assume $\alpha \in (0, 2]$; for the Student t and Pareto-like, we will just assume that $\alpha > 0$. Note that if $\alpha \leq 1$, the autocovariances are not defined at any lags, but this will not be a problem for the tail index estimator; we only need assume that the ACF for the Gaussian series $\{G_t\}$ is absolutely summable. Then the asymptotic variance V will be finite; however, it is impracticable to compute. We suggest that either a subsampling estimate of the variance (Politis, Romano, and Wolf, 1999) or a block bootstrap estimate (Kunsch, 1989) be used in practical applications.

Pareto-like Distribution. We work out some additional details for the case of the Pareto-like distribution. Now $Var \log |Z| = \frac{1}{4} Var \log \epsilon + Var \log |G|$, and the second term equals $\Psi_2(1/2)/4$ by the material on the case of the stable distribution in Section 2.2. We can easily compute (for r small)

$$\phi_\epsilon(r) = \alpha \int_1^\infty z^{2r-\alpha-1}\, dz = \frac{\alpha}{\alpha - 2r}.$$

Thus it follows that

$$\log \phi(r) = \log \alpha - \log(\alpha - r) + \frac{r}{2} \log 2 + \log \Gamma((r+1)/2) - \frac{\log \pi}{2}.$$

Hence we have $g(\alpha) = \alpha^{-2} + \Psi_2(1/2)/4$, which can be inverted for α. In particular, $g^{-1}(x) = \sqrt{1/(x - \Psi_2(1/2)/4)}$. In the *iid* case, the variance V is given by

$$V = 8\alpha^{-4} + \alpha^{-2}\Psi_2(1/2) + \frac{1}{8}\Psi_2^2(1/2) + \frac{1}{16}\Psi_4(1/2).$$

Also $\dot{g}(\alpha) = -2\alpha^{-3}$, so the asymptotic variance of $\hat{\alpha}$ is $\alpha^6 V/4$.

4 Simulation Studies

The theoretical results obviate the need for an extensive simulation study, but we show some results for data of a small sample size. We do not compare our method with nonparametric tail index estimators like the Hill estimate, since the comparison would be unfair; the Hill is designed as a general-purpose tool, whereas our methods are only valid when the *specific* marginal distribution is known. The purpose of our simulation studies is two-fold: to examine small-sample performance of the estimator

– and how it is affected by serial correlation – and to determine its robustness with respect to model mis-specification.

Below we apply the method to several heavy-tailed models, in each case assuming that the correct function g is known. There is a small probability that the estimator in each case is undefined; this can happen (e.g., with the Stable class) when $g^{-1}(x)$ is applied to a negative number. For larger values of α this problem becomes increasingly common due to the large variance of $\widehat{Var}\log|X|$. For this reason, we restrict $\alpha \leq 1$ when considering the small sample $n = 100$. For the case that $\alpha \geq 1$, we must increase the sample size to $n = 1000$ to avoid this problem. While this is a noted weakness of the method, the remarkable accuracy of the estimate in the acceptable range of α compensates this. Most of the error is due to variance, the bias being negligible in all cases.

We simulate data from the stable, Student t, and Pareto-like distribution discussed in Sections 2 and 3 above, and consider iid, $MA(1)$, and $AR(1)$ time series models for the Gaussian components of the corresponding Noah-Joseph models; thus we have a total of nine models (with several choices of α for each model). The time series models were selected to represent serial dependence at three levels: none, low, and moderate. The MA parameter is chosen to be 1/2, and the AR parameter is .9, indicating a fairly high degree of persistence (nevertheless the ACF will decay at geometric rate). Hence we are not considering long memory, though see the discussion in Section 5 below. The method generally breaks down when α is large, so we restrict α to the set $.1, .2, .3, \cdots, 1.9$. We consider $10,000$ replications for samples of size 100 (a suitable burn-in was used for the $AR(1)$ model) and 1000, and report the root MSE. Table 1 gives results for sample size $n = 100$ and $\alpha \leq 1$, while Table 2 gives results for $n = 1000$ and $1 \leq \alpha \leq 1.9$.

Table 1. Comparison of Root MSE for various models. Here $.1 \leq \alpha \leq 1$ and sample size is $n = 100$.

	Stable			Student t		
α	WN	MA(1)	AR(1)	WN	MA(1)	AR(1)
.1	.0107	.0106	.0108	.0147	.0145	.0147
.2	.0214	.0210	.0213	.0297	.0296	.0298
.3	.0325	.0326	.0330	.0456	.0456	.0465
.4	.0451	.0448	.0453	.0654	.0653	.0659
.5	.0576	.0586	.0587	.0888	.0877	.0910
.6	.0738	.0739	.0744	.1150	.1175	.1246
.7	.0896	.0903	.0965	.1478	.1571	.1711
.8	.1129	.1122	.1226	.2188	.2041	.2398
.9	.1387	.1377	.1507	2.0393	2.0171	4.0312
1.0	.1698	.1694	.1983	5.6571	4.5008	7.2250

Table 2. Comparison of Root MSE for various models. Here $1 \leq \alpha \leq 1.9$ and sample size is $n = 1000$.

	Stable			Student t		
α	WN	MA(1)	AR(1)	WN	MA(1)	AR(1)
1.0	.0478	.0481	.0507	.0776	.0771	.0820
1.1	.0579	.0577	.0611	.0952	.1003	.1204
1.2	.0690	.0685	.0741	.1133	.1136	.1244
1.3	.0834	.0835	.0897	.1382	.1406	.1541
1.4	.0972	.0964	.1058	.1702	.1660	.1898
1.5	.1167	.1154	.1299	.2026	.1981	.2227
1.6	.1375	.1373	.1521	.2433	.2427	.2724
1.7	.1626	.1597	.1796	.2915	.2907	.3304
1.8	.1881	.1902	.2238	.3502	.3453	.4157
1.9	.2210	.2211	.2670	.4220	.4627	.5221

Using the *ad hoc* rule of thumb that an RMSE of .1 is tolerable, we see that $\alpha \leq .8$ yields acceptable results for the Stable models, when $n = 100$. For the Student t the variance is higher, and $\alpha \leq .5$ gives the acceptable range (with some egregiously bad estimates for $\alpha = .9, 1.0$). If we increase sample size to $n = 1000$, then we can take α as high as 1.4 and 1.0 respectively. All estimates had minute bias – the majority of the error came from the variance. It is interesting that in all cases the bias was positive. The results were affected little by serial dependence in general, though in a few cases the RMSE for the $AR(1)$ was double that of the WN model. The basic pattern is that the method has a harder time with higher values of α, and this can be compensated by taking a larger sample. For small values of α the method is remarkably accurate.

A Note on Simulation It is clear how to simulate Noah-Joseph models for the stable and Student t cases; one needs to generate stable subordinators $S_{\alpha/2}(1, 1, 0)$ and χ^2 variables respectively, along with the serially correlated Gaussian series. To simulate the subordinators, see Chambers, Mallows, and Stuck (1976). For the Pareto-like case, observe that the cumulative distribution function for σ is just

$$F_\sigma(x) = 1 - x^{-\alpha},$$

which is easily inverted yielding $F^{-1}(u) = (1-u)^{-1/\alpha}$. Now we can simulation standard uniforms and plug into $F^{-1}(u)$ to get simulations of σ.

As a second exercise, we now suppose that the true distribution is Pareto-like, but we use a stable to model the process instead. Of course we must assume that $\alpha < 2$ for the exercise to be meaningful. We chose the stable-Pareto-like pair because the g functions are extremely similar for these distributions. The asymptotic bias can be determined exactly; our simulations indicate the small sample bias. We simulate the three Pareto-like models mentioned above, for various values of α; but we construct the estimate by applying g_{stab}^{-1} to $\widehat{Var} \log |X|$, where $g_{stab}(\alpha)$ is the function suitable for the stable class. In particular,

$$g_{stab}^{-1}\left(\widehat{Var}\log|X|\right) - \alpha = g_{stab}^{-1}\left(\widehat{Var}\log|X|\right) - g_{stab}^{-1}\left(g_{par}(\alpha)\right) + g_{stab}^{-1}\left(g_{par}(\alpha)\right) - \alpha$$

where g_{par} is the function suitable for the Pareto-like class. Now the first term above is asymptotically normal with variance $V/(n\dot{g}_{stab}^2(\alpha))$, where V is the variance associated with the Pareto-like class. The second term represents a deterministic asymptotic bias, which is easy to calculate:

$$Bias = \frac{\alpha}{\sqrt{\frac{\alpha^2}{4} + \frac{1}{\Psi_2(1)}}} - \alpha. \tag{7}$$

It is interesting that there is no asymptotic bias when $\alpha = \sqrt{4 - 24/\pi^2} \doteq 1.25$. The bias and variance results for the simulations are shown in Table 3 below.

Table 3. Bias and Root MSE for mis-specification study. Here $.1 \leq \alpha \leq 1.9$ and sample size is $n = 1000$.

	WN		MA(1)		AR(1)	
α	Bias	RMSE	Bias	RMSE	Bias	RMSE
.1	.0283	.0289	.0284	.0290	.0284	.0289
.2	.0552	.0564	.0553	.0564	.0555	.0566
.3	.0788	.0805	.0787	.0805	.0790	.0807
.4	.0986	.1010	.0985	.1009	.0988	.1012
.5	.1126	.1159	.1126	.1159	.1129	.1162
.6	.1202	.1247	.1203	.1249	.1210	.1255
.7	.1213	.1273	.1220	.1281	.1230	.1292
.8	.1159	.1241	.1163	.1244	.1169	.1255
.9	.1040	.1154	.1044	.1162	.1057	.1182
1.0	.0844	.1022	.0837	.1014	.0868	.1057
1.1	.0581	.0872	.0582	.0867	.0612	.0915
1.2	.0257	.0758	.0261	.0767	.0307	.0822
1.3	-.0118	.0798	-.0118	.0808	-.0065	.0854
1.4	-.0558	.1026	-.0560	.1034	-.0507	.1072
1.5	-.1041	.1405	-.1041	.1403	-.0978	.1417
1.6	-.1569	.1869	-.1581	.1885	-.1519	.1884
1.7	-.2159	.2412	-.2132	.2394	-.2050	.2376
1.8	-.2755	.2992	-.2751	.2982	-.2670	.2965
1.9	-.3494	.3602	-.3396	.3608	-.3328	.3593

The bias properties are pretty much as expected, given (7); bias is close to zero for $\alpha = 1.2$ and 1.3, and drops off for low α as well. The RMSE seems tolerable for $.9 \leq \alpha \leq 1.4$ and again for $\alpha \leq .4$. In any event, the RMSE in this mis-specification case is roughly of the same magnitude (in many cases) as the RMSE in the correctly specified situation (Table 2 above). Again, there is little sensitivity to serial dependence.

5 Conclusion

This paper discusses a fairly general parametric procedure for tail index estimation. The method assumes that the marginal distribution of the data is known, but the tail index (and other parameters, such as the scale) are unknown. The approach of using log moments allows for the elimination of the scale parameter, while at the same time computing well-defined moments. The moment approach generates a $\sqrt{n}$-consistent estimator, with asymptotically normal distribution. Because the method is parametric, there is no issue of bandwidth selection, which commonly occurs with nonparametric tail index estimators.

Although the asymptotic results are pleasing, there can be serious difficulties with this method in finite sample when the tail index is too large. Simulations indicate that for sample size $n = 100$ (a moderate length time series in economics), the tail index must be less than unity in order for the method to work; for sample size $n = 1000$, the tail index should be less than two. This is admittedly a drawback; not only are the estimates poor, but in some cases can be meaningless (since one may be required to take the square root of a negative number). Nevertheless, the excellent precision of the estimator for an "acceptable range" of α values is remarkable. The empirical studies indicate that this moment estimator may be more appropriate for extreme value data found in insurance and finance, where the values of α may often be less than unity (see Embrechts et al. (1997)). For finite variance data, this method should be avoided.

The method also seems to work well in the presence of serial correlation. In order to investigate its properties in the context of serial dependence, we introduce the Noah-Joseph models for heavy tails and long memory. These models allow one to utilize a given marginal distribution, but introduce any desired auto-correlation structure. In the asymptotic normality results for the estimator, we focus on the case of intermediate memory, i.e., a summable Gaussian autocovariance function. It is plausible that similar asymptotics hold for the long memory case, though the normalization in the Central Limit Theorem must then be altered to account for this.

Future research should focus on how to extend this type of estimator to be effective with higher values of α. One direction could be to look at higher centered log moments, e.g., a centered fourth log moment. Another direction of research is to derive maximum likelihood estimates of the tail index for Noah-Joseph models, since the likelihood will have an approximately Gaussian form that can be written down explicitly. Again, these types of models should be useful in applied econometric work, where a practitioner may "know" from empirical data analysis that the marginal distributions are Student t and have an auto-correlation structure of an $AR(p)$ (say). This cannot be written as a linear filter of t innovations; but the Noah-Joseph model is very natural and is identifiable.

6 Appendix

Proof of Theorem 1. Since the data are *iid*, the first result is a standard CLT for the sample variance statistic. The second result uses the mapping $g^{-1}(x)$ applied to the first result. By standard Taylor series arguments, the variance gets modified by

$$\left[\frac{d}{dx} g^{-1}(x)|_{x=g(\alpha)}\right]^2 = \dot{g}(\alpha)^{-2},$$

assuming that this quantity exists. $\square$

Proof of Theorem 2. The sample variance of $\log|X_t|$ is equal to

$$\frac{1}{n-1}\sum_{t=1}^{n}(\log|Z_t| - \mathbb{E}\log|Z_t|)^2 - \frac{n}{n-1}\Big(\mathbb{E}\log|Z| - \overline{\log|Z|}\Big)^2,$$

and the second term is $O_P(1/n)$ (assuming that the sample mean of W_t is $O_P(1/\sqrt{n})$, which will follow from arguments made below). Hence asymptotically we may just consider the centered sample mean of W_t^2. Below we compute the covariance function of this process. Let $K(\sigma_t) = \log\sigma_t - \mathbb{E}\log\sigma_t$, and $H(G_t) = \log|G_t| - \mathbb{E}\log|G_t|$. Then

$$\begin{aligned}
&W_t^2 W_{t+h}^2 = (K(\sigma_t)K(\sigma_{t+h}) + H(G_t)K(\sigma_{t+h}) + H(G_{t+h})K(\sigma_t) + H(G_t)H(G_{t+h}))^2 \\
&= K^2(\sigma_t)K^2(\sigma_{t+h}) + 2H(G_t)K^2(\sigma_{t+h})K(\sigma_t) + 2H(G_{t+h})K^2(\sigma_t)K(\sigma_{t+h}) \\
&+ 4H(G_t)H(G_{t+h})K(\sigma_t)K(\sigma_{t+h}) + H^2(G_t)K^2(\sigma_{t+h}) + 2H^2(G_t)H(G_{t+h})K(\sigma_{t+h}) \\
&+ H^2(G_{t+h})K^2(\sigma_t) + 2H(G_t)H^2(G_{t+h})K(\sigma_t) + H^2(G_t)H^2(G_{t+h}).
\end{aligned}$$

Now using the independence of the volatility series from the Gaussian series, along with the fact that $K(\sigma_t)$ and $H(G_t)$ are mean zero, the expectation for nonzero h is

$$\mathbb{E}[W_t^2 W_{t+h}^2] = (\mathbb{E}[K^2(\sigma)])^2 + 2\mathbb{E}[K^2(\sigma)]\mathbb{E}[H^2(G)] + \mathbb{E}[H^2(G_t)H^2(G_{t+h})].$$

Of course $\mathbb{E}[K^2(\sigma)] = Var \log \sigma$ and $\mathbb{E}[H^2(G)] = Var \log |G|$. For $h = 0$ we have

$$\mathbb{E}[W_t^4] = \mathbb{E}[K^4(\sigma)] + 6\mathbb{E}[K^2(\sigma)]\mathbb{E}[H^2(G)] + \mathbb{E}[H^4(G)].$$

Hence it follows that

$$\gamma_{W^2}(0) = Var K^2(\sigma) + 4 Var \log \sigma \cdot Var \log |G| + Var H^2(G)$$
$$\gamma_{W^2}(h) = \gamma_{H^2(G)}(h) \qquad h \neq 0.$$

Here $\gamma_{H^2(G)}$ denotes the ACF of the process $\{H^2(G_t)\}$. Thus V, the limiting variance, is $V = \sum_{h=-\infty}^{\infty} \gamma_{W^2(G)}(h)$. Now for any real θ, the characteristic function of the normalized sample mean of W_t is

$$\mathbb{E} \exp \left\{ i\theta \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(W_t^2 - \mathbb{E}W^2 \right) \right\} = \mathbb{E} \left[\mathbb{E} \exp \left\{ i\theta \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(W_t^2 - \mathbb{E}W^2 \right) \right\} |\mathcal{G} \right],$$

where $\mathcal{G}$ represents the total information in the $\{G_t\}$ process. Since the volatility series is independent of $\mathcal{G}$, we can focus on the interior conditional expectation. Now the W_t^2 variables are independent conditional on $\mathcal{G}$, so

$$\begin{aligned}
&\mathbb{E} \exp\{ i\theta \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left(W_t^2 - \mathbb{E}W^2 \right) \} |\mathcal{G} \\
&= \Pi_{t=1}^{n} \mathbb{E} \exp\{ i\theta \frac{1}{\sqrt{n}} \left(W_t^2 - \mathbb{E}W^2 \right) \} |\mathcal{G} \\
&= \Pi_{t=1}^{n} \left(1 + \frac{i\theta}{\sqrt{n}} \left(\mathbb{E}[W_t^2|\mathcal{G}] - \mathbb{E}W^2 \right) - \frac{\theta^2}{2n} \mathbb{E} \left[\left(W_t^2 - \mathbb{E}W^2 \right)^2 |\mathcal{G} \right] + o(1/n) \right).
\end{aligned}$$

By applying the exponential of the logarithm to this product, we can see that it equals the exponential of

$$\sum_{t=1}^{n} \frac{i\theta}{\sqrt{n}} \left(\mathbb{E}[W_t^2|\mathcal{G}] - \mathbb{E}W^2 \right) - \sum_{t=1}^{n} \frac{\theta^2}{2n} \mathbb{E} \left[\left(W_t^2 - \mathbb{E}W^2 \right)^2 |\mathcal{G} \right] + \sum_{t=1}^{n} \frac{\theta^2}{2n} \left(\mathbb{E}[W_t^2|\mathcal{G}] - \mathbb{E}W^2 \right)^2 \tag{8}$$

plus terms that are $o(1/n)$. This comes from the expansion $\log(1+x) = x - x^2/2 + o(x^2)$; the first two terms in (8) come from the first term in the Taylor series expansion of $\log(1+x)$, while the third term in (8) comes from the second term of $\log(1+x)$. Since these expressions are conditional on random quantities, the $o(1/n)$ can be interpreted as $o_P(1/n)$; the Taylor series approximation is valid because all the relevant moments are bounded. Next,

$$\begin{aligned}
\mathbb{E}[W_t^2|\mathcal{G}] - \mathbb{E}W^2 &= \mathbb{E}\left[K^2(\sigma_t) + 2K(\sigma_t)H(G_t) + H^2(G_t)|\mathcal{G}\right] - \mathbb{E}[K^2(\sigma)] - \mathbb{E}[H^2(G)] \\
&= H^2(G_t) - \mathbb{E}[H^2(G)] \\
\mathbb{E}\left[\left(W_t^2 - \mathbb{E}W^2\right)^2|\mathcal{G}\right] &= \mathbb{E}[K^4(\sigma)] + 4\mathbb{E}[K^3(\sigma)]H(G_t) + 6\mathbb{E}[K^2(\sigma)]H^2(G_t) + H^4(G_t) \\
&\quad - 2\left(\mathbb{E}[K^2(\sigma)] + \mathbb{E}[H^2(G)]\right)\left(\mathbb{E}[K^2(\sigma)] + H^2(G_t)\right) + \left(\mathbb{E}[K^2(\sigma)] + \mathbb{E}[H^2(G)]\right)^2.
\end{aligned}$$

Now the sum of the second and third terms of (8) converges in probability to the limit

$$-\frac{\theta^2}{2}(VarK^2(\sigma) + 4\mathbb{E}[K^2(\sigma)]\mathbb{E}[H^2(G)])$$

after some simplification. This calculation assumes that $H(G_t)$ and its powers satisfy a Strong Law of Large Numbers, which follows from Sun, 1965 (this is where the assumption on the summability of the ACF of $\{G_t\}$ is used). Since the limit of these terms is a constant, they don't affect the convergence of the first term in (8). Hence

$$\begin{aligned}
&\mathbb{E}\exp\left\{i\theta\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(W_t^2 - \mathbb{E}W^2\right)\right\} \\
&\sim \mathbb{E}\exp\left\{\frac{i\theta}{\sqrt{n}}\sum_{t=1}^{n}\left(H^2(G_t) - \mathbb{E}[H^2(G)]\right)\right\}\exp\{-\frac{\theta^2}{2}(VarK^2(\sigma) + 4\mathbb{E}[K^2(\sigma)]\mathbb{E}[H^2(G)])\},
\end{aligned}$$

and the first term is just the characteristic function of $n^{-1/2}\sum_{t=1}^{n} H^2(G_t) - \mathbb{E}[H^2(G)]$. This uses the almost sure convergence of the second two terms of (8) to a negative number together with the boundedness of the exponential function on the negative real line, and an application of the Dominated Convergence Theorem. Again by Sun (1965),

$$n^{-1/2}\sum_{t=1}^{n} H^2(G_t) - \mathbb{E}[H^2(G)] \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, U)$$

with $U = \sum_{h=-\infty}^{\infty} \gamma_{H^2(G)}(h)$. Hence

$$\begin{aligned}
&\mathbb{E}\exp\left\{i\theta\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(W_t^2 - \mathbb{E}W^2\right)\right\} \\
&\to \exp\{-\frac{\theta^2}{2}(U + VarK^2(\sigma) + 4\mathbb{E}[K^2(\sigma)]\mathbb{E}[H^2(G)])\} = e^{-\theta^2 V/2},
\end{aligned}$$

which completes the proof of the theorem. $\square$

References

[1] Bickel, P. and Doksum, K. (1977) *Mathematical Statistics.* Prentice Hall: New Jersey.

[2] Billingsley, P. (1995) *Probability and Measure.* John Wiley and Sons: New York.

[3] Chambers, J., Mallows, C., and Stuck, B. (1976) A method for simulating stable random variables. *Journal of the American Statistical Association* **71**, 340–344.

[4] Davis, R., and Resnick, S. (1986) Limit theory for the sample covariance and correlation functions of moving averages. *Annals of Statistics* **14**, No. 2, 533–558.

[5] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modeling Extremal Events for Insurance and Finance.* Springer-Verlag, Berlin.

[6] Hardin, C.D. (1984) Skewed stable variables and processes. Technical Report 79, Center for Stochastic Processes at the University of North Carolina, Chapel Hill.

[7] Hill, B. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**, 1163–1174.

[8] Kokoszka, P. and Taqqu, M. (1999) Discrete time parametric models with long memory and infinite variance. *Mathematical and Computer Modelling* **29**, 203–215.

[9] Kunsch, H. (1989) The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17**, 1217–1241.

[10] Mandelbrot, B., and Wallis, J.R. (1968) Noah, Joseph, and Operational Hydrology. *Water Resources Res.* **4**, 909-917.

[11] McElroy, T. and Politis, D. (2006) Self-Normalization for Heavy-Tailed Time Series with Long Memory. *Statistica Sinica*, forthcoming.

[12] Politis, D., Romano, J., and Wolf, M. (1999) *Subsampling.* Springer, New York.

[13] Samorodnitsky, G. and Taqqu, M. (1994) *Stable Non-Gaussian Random Processes.* Chapman and Hall.

[14] Sun, T.C. (1965). Some further results on central limit theorems for non-linear functions of a normal stationary process. *Journal of Mathematics and Mechanics* **14**, 71–85.