Statistica Sinica **17**(2007), 199-220

SELF-NORMALIZATION FOR HEAVY-TAILED TIME SERIES WITH LONG MEMORY

Tucker McElroy[1,2] and Dimitris Politis[2]

[1] *U.S. Census Bureau and* [2] *University of California, San Diego*

Abstract: Many time series data sets have heavy tails and/or long memory, both of which are well-known to greatly influence the rate of convergence of the sample mean. Typically time series analysts consider models with either heavy tails or long memory; we consider both. The paper is essentially a theoretical case study that explores the growth rate of the sample mean for a particular heavy-tailed, long memory time series model. An exact rate of convergence, which displays the competition between memory and tail thickness in fostering sample mean growth, is obtained in our main theorem. An appropriate self-normalization is used to produce a studentized sample mean statistic, computable without prior knowledge of the tail and memory parameters. This paper presents a novel heavy-tailed time series model that also has long memory in the sense of sums of well-defined autocovariances; we explicitly show the role that memory and tail thickness play in determining the sample mean's rate of growth, and we construct an appropriate studentization. Our model is a natural extension of long memory Gaussian models to data with infinite variance, and therefore pertains to a wide range of applications, including finance, insurance, and hydrology.

Key words and phrases: Heavy-tailed data, infinite variance, long-range dependence, studentization.

1. Introduction

The phenomena of heavy tails and long memory have been observed in many branches of science, as well as in insurance and economics; see Samorodnitsky and Taqqu (1994, pp.586-590) for a historical review and many references. Heavy-tailed data frequently exhibit large extremes, and may even have infinite variance, while long memory data exhibit great serial persistence, behaving similar to a random walk in many cases. The literature on time series models that capture these phenomena has often times followed two separate paths − heavy-tailed, intermediate (or short) memory models and finite variance, long memory models. Davis and Resnick (1985, 1986) introduced infinite order moving average time series models with heavy-tailed inputs, and these seminal papers have sparked a wealth of interest in such models and their applications. However, these models do not have well-defined autocovariances, and the common modern conception of long memory cannot be formulated within that umbrella. On the other hand, long memory models have received much attention (see Beran (1994, Chap.1) and Granger and Joyeux (1980) for example), but the literature generally assumes that the data has finite variance.

There is a need to explore time series models with both thick tails and long memory. Indeed, much of the early work (Mandelbrot and Wallis (1968)) in this field noted that long memory time series often were heavy-tailed and self-similar as well. So the joint presence of heavy tails and long memory in many data sets has been noted for several decades; for more recent work in this area, see Heyde and Yang (1997), Hall (1997), Rachev and Samorodnitsky (2001) and Mansfield, Rachev and Samorodnitsky (2001). Also Heath, Resnick, and Samorodnitsky (1998, 1999) show some interesting theoretical work on fluid models that incorporate both heavy tails and long memory. Again in finance, where long memory models have seen some popularity (Greene and Fielitz (1977)), the data are known to be heavy-tailed (Embrechts, Klüppelberg and Mikosch (1997)). Thus, there is a large body of literature documenting the joint presence of long memory and heavy tails in time series data.

A common modern approach to modeling long memory is through describing the rate of decay of the autocovariance function. But this definition requires that the autocovariance sequence is well-defined in the first place; for infinite variance heavy-tailed data, it is not clear how to proceed. Indeed, the finite and infinite variance approaches to time series analysis tend to differ drastically, with correspondingly different mathematical tools and methods. For example, in nonparametric time series analysis one can explore long-range dependence through covariance functions and mixing coefficients, but there is a trade-off: the more dependence that is present in the data, the more moments one is required to assume exist. Perhaps for this reason, infinite variance models tend to be parametric − for example, consider the popular moving average models of Davis and Resnick (1985, 1986). However, the drawback for these models is that the summability conditions on the moving average coefficients preclude parametrizing long memory through the decay rate of pseudo-autocorrelations. In contrast, the model that we present has infinite variance *and* finite autocovariances, thus permitting a simple adaption of the modern definition of long memory. In addition, our model is conditionally Gaussian, so that many of the finite variance methods may still be applied.

Consider a stationary time series Y_t centered at zero; let Y denote a random variable with the same distribution as that of Y_t. If the cumulative distribution function (cdf) of Y is in the normal domain of attraction of an α-stable distribution (written $D(\alpha)$, see Embrechts, Klüppelberg and Mikosch (1997, Chap.2) and the data are independent or an infinite order moving average (with absolutely summable coefficients) of *i.i.d.* errors, then $\sum_{t=1}^{n} Y_t = O_P(n^{1/\alpha})$ so that $1/\alpha$ gives the appropriate rate of growth (see Davis and Resnick (1985)). On the other hand, if Y is square integrable, with autocovariances γ_Y satisfying $\sum_{h=1}^{n} \gamma_Y(h) = O(n^{\beta})$ for some $\beta \in [0,1)$ as $n \to \infty$, then $\sum_{t=1}^{n} Y_t = O_P(n^{(\beta+1)/2})$, which follows from computing the sum's variance. If Y could somehow share both of these properties, then its rate of growth would be the greater of $n^{1/\alpha}$ and $n^{(\beta+1)/2}$. Restricting to $\alpha \in (1,2)$ (so that the mean exists but the variance is infinite) yields the picture given in Figure 1.

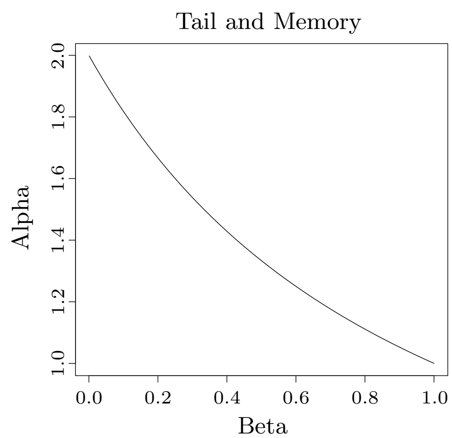

Figure 1. Tail Thickness vs. Memory

We refer to the lower left-hand region as *Tail*, because here $2/\alpha > \beta + 1$ so that tail thickness determines the growth rate of the partial sums. The upper right-hand region is *Mem*, where $2/\alpha < \beta + 1$ so that memory dominates. The curve represents the boundary line $2/\alpha = \beta + 1$, where both tails and memory give an equal contribution. The convexity of this curve shows that *Mem* is more prevalent; its area is $2 - \log 4 = 0.6137$ versus the $\log 4 - 1 = 0.3863$ area of *Tail*.

We present our model in Section 2, along with a few basic properties. In Section 3 we compute the asymptotic distribution of the sample mean, the sample variance, and their joint asymptotic distribution. Finally, we present a self-normalization for the memory parameter, and show how a studentized mean can be computed without prior knowledge of α or β. Section 4 is the conclusion, which discusses some possible extensions of the model and the methods of proof; proofs are contained in the appendix.

2. The Model

Let our observed series be $\{X_t,\, t = 1, \ldots, n\}$ which has location parameter

η. Then define

$$Y_t = X_t - \eta = \sigma_t G_t, \tag{1}$$

so that Y_t is the product of a "volatility" time series σ_t and a long memory time series G_t. We make the following additional assumptions.

(A) The series $\{\sigma_t\}$ and $\{G_t\}$ are independent.

(B) Let $\sigma_t = \sqrt{\epsilon_t}$, where ϵ_t are *i.i.d.* and have the marginal distribution of an $\alpha/2$ stable random variable with skewness parameter 1, scale parameter $\tau = (\cos(\pi\alpha/4))^{2/\alpha}$, and location parameter zero.

(C) The tail index α is a constant in $(1, 2)$.

(D) $\{G_t\}$ is stationary (mean zero) Gaussian.

(E) The Gaussian series is purely non-deterministic, i.e., the one-step ahead prediction errors have nonzero variance − see Brockwell and Davis (1991, p.187).

This construction is based on the sub-Gaussian processes discussed in Samorodnitsky and Taqqu (1994, Chap.2), but here the subordinator ϵ_t is a process instead of a single fixed variable. Since $\alpha \in (1, 2)$, the model has finite mean but infinite variance. Let us denote the autocovariance function of $\{G_t\}$ by γ_G; we also know by assumptions (D) and (E) that we can represent $\{G_t\}$ by a linear process. The following proposition summarizes some of the most salient properties of this model.

Proposition 1. *Given* (A) *through* (E), *the following statements are true of the model at* (1):

1. *The series* $\{X_t\}$ *is strictly stationary, and the marginal distribution is symmetric* α *stable (sαs) with scale parameter* $\sqrt{\gamma_G(0)/2}$ *and mean* η.
2. *The mean of* σ_t *exists and is* $\mu := \Gamma(1 - 1/\alpha)/2\Gamma(3/2)$.
3. *The second moment of* X_t *is infinite, but* $\mathrm{Cov}\,[X_t, X_{t+h}] = \mu^2\gamma_G(h)$ *is finite for* $h \neq 0$ *and depends only on the lag* $|h|$.

Several of the assumptions in the model can be generalized to encompass a wider class of data sets. Assumption (A) is crucial to the perspective here, and cannot be relaxed without altering the analytical methods used. As for (B), there is clearly no loss of generality in specifying scale τ for the volatility, due to the multiplicative structure of the model. Indeed, if we originally assumed that ϵ had some generic unspecified scale C, then we could redefine ϵ by $\epsilon \cdot \tau/C$ and scale the Gaussian process by C/τ.

Remark 1. Note that we can extend the model to encompass $\alpha = 2$, which corresponds to the Gaussian case. Then the volatility σ_t is deterministic (and constant to ensure stationarity), but we say little about this case in our paper, since it has been well-studied.

Remark 2. Suppose that the volatility series is only assumed m-dependent. Then, for any $|h| \leq m$, we have $\mathrm{Cov}\,[X_t, X_{t+h}] = \mathbb{E}[\sigma_t \sigma_{t+h}]\mathbb{E}[G_t G_{t+h}]$ and there are no guarantees that $\mathbb{E}[\sigma_t \sigma_{t+h}]$ is finite. However, if $|h| > m$, the same calculation yields $\mu^2 \gamma_G(h)$, which is well-defined. For example, suppose that ϵ_t is a causal moving average of an *i.i.d.* stochastic volatility series

$$\epsilon_t = \psi_0 \delta_t + \psi_1 \delta_{t-1} + \psi_2 \delta_{t-2} + \cdots + \psi_m \delta_{t-m}.$$

Then the series is clearly m-dependent and, if $|h| \leq m$, we have

$$\begin{aligned}\mathbb{E}[\sigma_t \sigma_{t+h}] = \mathbb{E}[\sqrt{\epsilon_t \epsilon_{t+h}}] &\geq \mathbb{E}\left[\sqrt{\sum_j \psi_j \psi_{j+h} \delta_{t-j}^2}\right] \\ &\geq \mathbb{E}\left[\sqrt{\psi_0 \psi_h \delta_t^2}\right] = \mathbb{E}\left[\sqrt{\psi_0 \psi_{-h}}\delta_t\right] = +\infty\end{aligned}$$

assuming that all the coefficients ψ_j are positive. But if $|h| > m$, we obtain $\mu^2 \gamma_G(h)$. Clearly, if the stochastic volatility series is autoregressive, then the autocovariance of the Y series is never finite at any lag, which is not convenient for defining long memory.

Definition 1. A stationary process is said to have long memory if its autocovariance function γ satisfies

$$\sum_{0<|h|<n} \gamma(h) \sim C n^{\beta} \quad \text{and} \quad \sum_{0<|h|<n} |\gamma(h)| = O(n^{\beta})$$

as $n \to \infty$, and $\beta \in [0, 1)$, where $C > 0$ is a constant. Note that $a_n \sim b_n$ iff $a_n / b_n \to 1$ as $n \to \infty$. The condition on γ is referred to as $LM(\beta)$ for $\beta \in [0, 1)$, but $LM(0)$ is sometimes referred to as denoting intermediate memory $-$ namely that the autocovariance function is absolutely summable.

One easy consequence of $LM(\beta)$ for $\beta > 0$ is that $\gamma(n) \sim (\beta C/2) n^{\beta-1}$, since $\sum_{|h|<n} \gamma(h) = 2 \sum_{h=1}^{n-1} \gamma(h)$ and L'Hopital's Rule gives

$$\frac{C}{2} = \lim_{n\to\infty} \frac{\sum_{h=1}^{n} \gamma(h)}{n^{\beta}} = \lim_{n\to\infty} \frac{\gamma(n)}{\beta n^{\beta-1}}.$$

This property is used whenever $LM(\beta)$ holds with $\beta > 0$.

3. Results

Consider the sample mean $\overline{X}$ as an estimator of η, $\hat{\eta} = \overline{X} = (1/n) \sum_{t=1}^{n} X_t$, so $n(\overline{X} - \eta) = \sum_{t=1}^{n} Y_t$. The following theorem gives the asymptotics of the partial sums; the rate of convergence depends delicately on whether the pair

(α, β) is located in *Tail* or *Mem*. Take $\zeta = \max\{1/\alpha, (\beta+1)/2\}$ to measure the dominant contributor to growth.

Theorem 1. *Suppose* (A) *through* (E) *hold, and assume* $LM(\beta)$ *with* $\beta \in [0, 1)$. *Then the partial sums of the* $\{Y_t\}$ *series, normalized by* n^ζ, *converge to an absolutely continuous random variable. In particular,*

$$n^{-\zeta} \sum_{t=1}^{n} Y_t \stackrel{\mathcal{L}}{\Longrightarrow} \begin{cases} S & \text{if } \frac{1}{\alpha} > \frac{\beta+1}{2} \\ V & \text{if } \frac{1}{\alpha} < \frac{\beta+1}{2} \\ S+V & \text{if } \frac{1}{\alpha} = \frac{\beta+1}{2}, \end{cases} \tag{2}$$

where $\stackrel{\mathcal{L}}{\Longrightarrow}$ *denotes convergence in distribution.* S *is a* $s\alpha s$ *variable with zero location parameter, and scale* $\sqrt{\gamma_G(0)/2}$, *whereas* V *is a mean zero Gaussian variable with variance* $\tilde{C}\mu^2/(\beta+1)$, *where* $\tilde{C} = C - \gamma_G(0) 1_{\{\beta=0\}}$. *In the third case,* S *and* V *are independent.*

Remark 3. The distribution of S is exactly the same as the marginal distribution of the original $\{Y_t\}$ series, by Proposition 1. Hence if it were known that $(\alpha, \beta) \in$ *Tail*, then we would have a method for constructing confidence intervals for η.

Remark 4. We can easily extend this theorem to the $\alpha = 2$ case, which has already been studied in Taqqu (1975). If $\alpha = 2$ then $\mu = 1$, and the variance of V will be either $C - \gamma_G(0)$ or $C/(\beta+1)$ depending on whether $\beta = 0$ or not. Also, when $\alpha = 2$, the random variable S will be a mean zero Gaussian with variance $\gamma_G(0)$. Thus in the $\alpha = 2$, $\beta = 0$ case, the limit is $S + V$ which is Gaussian with mean zero and variance $C = \sum_{h \in \mathbb{Z}} \gamma_G(h)$. Hence in the intermediate memory Gaussian model, we obtain the classical limit.

The comment in Remark 3 does not give a practical method for constructing confidence intervals, since it is difficult in practice to verify whether *Tail* is true. Instead we adopt a different approach that considers the joint asymptotics of the sample mean with some measure of scale. In McElroy and Politis (2002), joint asymptotics for sample mean and sample variance were used to eliminate α from the resulting confidence interval via the trick of self-normalization, or studentization. In that paper, a linear process with heavy-tailed inputs was considered, and similar results for a stable moving average model can be found in McElroy and Politis (2004). Below we establish joint asymptotics for sample mean and sample variance normalized by rates involving ζ. Interestingly, the sample variance is always $O_P(n^{2/\alpha})$, even in the *Mem* case.

Theorem 2. *Suppose* (A) *through* (E) *hold, and assume* $LM(\beta)$ *with* $\beta \in [0, 1)$. *Then the sample first and second moments of the* $\{Y_t\}$ *series, normalized by* n^ζ, *converge jointly to absolutely continuous random variables. In particular,*

$$\left(n^{-\zeta}\sum_{t=1}^{n}Y_t, n^{-2\zeta}\sum_{t=1}^{n}Y_t^2\right) \stackrel{\mathcal{L}}{\Longrightarrow} \begin{cases} (S,U) & \text{if } \frac{1}{\alpha} > \frac{\beta+1}{2} \\ (V,0) & \text{if } \frac{1}{\alpha} < \frac{\beta+1}{2} \\ (S+V,U) & \text{if } \frac{1}{\alpha} = \frac{\beta+1}{2}, \end{cases} \tag{3}$$

S and V as in Theorem 1, *and U is $\alpha/2$ stable with zero location parameter, skewness one, and scale proportional to $\tau\gamma_G(0)$. V is independent of S and U, but S and U are dependent. The joint Fourier/Laplace Transform of $S+V, U$ is (θ real, $\phi > 0$)*

$$\begin{aligned}&\mathbb{E}[\exp\{i\theta(S+V)-\phi U\}]\\ &= \exp\left\{-(\frac{\gamma_G(0)}{2})^{\frac{\alpha}{2}}\mathbb{E}|\theta+\sqrt{2\phi}Z|^{\alpha}1_{\{\frac{2}{\alpha}\geq\beta+1\}} - \frac{\theta^2}{2}\tilde{C}\frac{\mu^2}{\beta+1}1_{\{\frac{2}{\alpha}\leq\beta+1\}}\right\},\end{aligned}$$

where Z has a standard normal distribution. Finally, (3) *remains true if $n^{-2\zeta}\sum_{t=1}^{n}Y_t^2$ is replaced by $n^{-2\zeta}\sum_{t=1}^{n}(X_t-\overline{X})^2$.*

By Theorem 2, the sample variance can be used to studentize $\overline{X}$ in the *Tail* case. We now need to find an appropriate normalization when $2/\alpha < \beta+1$. In the literature on long memory, the log periodogram has been used to estimate β − see Robinson (1995) for example. We are not interested in estimation of β, but in computing a positive statistic that grows at rate $n^{\beta+1}$. To this end, fix $\rho \in (0,1)$ and take

$$\widehat{LM}(\rho) = \left|\sum_{|h|=1}^{\lfloor n^{\rho}\rfloor}\frac{1}{n-|h|}\sum_{t=1}^{n-|h|}\left(X_tX_{t+h}-\overline{X}^2\right)\right|^{\frac{1}{\rho}},$$

which is essentially the sum of the first n^{ρ} sample autocovariances, all raised to the absolute power $1/\rho$. Note that if we replaced the sample autocovariances by the real autocovariances, this quantity would be of order n^{β}. Therefore we propose to use $\widehat{LM}(\rho)$ for some $\rho \in (0,1)$ as a second normalization to our sample mean. The following theorem gives the asymptotic behavior of $\widehat{LM}(\rho)$.

Theorem 3. *Suppose* (A) *through* (E) *hold, as well as $LM(\beta)$ with $\beta \in [0,1)$. Let $\rho \in (0,1)$ be a user-defined rate. Then $\widehat{LM}(\rho)$ converges in probability to a constant at rate n^{β}. In particular, $n^{-\beta}\widehat{LM}(\rho) \stackrel{P}{\longrightarrow} \mu^{2/\rho}C^{1/\rho}$.*

So we may use the centered sample second moments, together with the long memory estimator, to form a rate of growth normalization for the sample sum. This is because the sample second moments grow at rate $n^{2/\alpha}$, regardless of the relationship between α and β, whereas $\widehat{LM}(\rho)$ grows at rate n^{β} − also regardless of *Tail* or *Mem*. The following theorem basically summarizes the work of Theorem 2 and Theorem 3:

Theorem 4. *Suppose* (A) *through* (E) *hold, as well as* $LM(\beta)$ *with* $\beta \in [0,1)$. *Let* $\rho \in (0,1)$ *be a user-defined rate. Then the following joint weak convergence holds:*

$$\left(n^{-\zeta}\sum_{t=1}^{n}(X_t-\eta), n^{-2\zeta}\sum_{t=1}^{n}(X_t-\overline{X})^2, n^{-2\zeta+1}\widehat{LM}(\rho)\right)$$
$$\stackrel{\mathcal{L}}{\Longrightarrow}\begin{cases}(S,U,0) & \text{if } \frac{1}{\alpha}>\frac{\beta+1}{2}\\ (V,0,\mu^{\frac{2}{\rho}}C^{\frac{1}{\rho}}) & \text{if } \frac{1}{\alpha}<\frac{\beta+1}{2}\\ (S+V,U,\mu^{\frac{2}{\rho}}C^{\frac{1}{\rho}}) & \text{if } \frac{1}{\alpha}=\frac{\beta+1}{2}.\end{cases} \tag{4}$$

The normalized statistic also converges weakly. For

$$\hat{\eta}_{SN}=\frac{\sqrt{n}(\overline{X}-\eta)}{\sqrt{\frac{1}{n}\sum_{t=1}^{n}(X_t-\overline{X})^2+\widehat{LM}(\rho)}}$$

and an absolutely continuous random variable

$$Q=\begin{cases}\frac{S}{\sqrt{U}} & \text{if } \frac{1}{\alpha}>\frac{\beta+1}{2}\\ \frac{V}{\mu^{\frac{1}{\rho}}C^{\frac{1}{2\rho}}} & \text{if } \frac{1}{\alpha}<\frac{\beta+1}{2}\\ \frac{S+V}{\sqrt{U+\mu^{\frac{2}{\rho}}C^{\frac{1}{\rho}}}} & \text{if } \frac{1}{\alpha}=\frac{\beta+1}{2},\end{cases}$$

one has

$$\hat{\eta}_{SN}\stackrel{\mathcal{L}}{\Longrightarrow}Q. \tag{5}$$

In (4), it is interesting that the borderline case $2/\alpha=\beta+1$ gives a weak limit that is essentially the sum of the other cases. In (5), the distribution of $S/\sqrt{U}$ is numerically explored in Logan, Mallows, Rice and Shepp (1973). Appealing to the joint characteristic function of S and U in the proof of Theorem 2, we can write $S=\sqrt{\gamma_G(0)/2}S'$ and $U=\gamma_G(0)U'$ where S' and U' no longer depend on the scale parameters $\gamma_G(0)$ of the model; they depend only on α. However, there will not be cancellation of $\mu^{1/\rho}C^{1/2\rho}$ with the standard deviation of V – this could only happen in the case that $\rho=1$, which is prohibited by construction. Therefore the limiting distribution in the case that $\beta+1\geq 2/\alpha$ will depend on ρ through its scale parameter. Thus we have the following corollary:

Corollary 1. *With* Q *as in Theorem* 4 *it follows that an approximate* $1-p$ *confidence interval for* η *is*

$$\left[\overline{X}-\frac{\hat{\sigma}}{\sqrt{n}}\,q_{1-\frac{p}{2}},\overline{X}-\frac{\hat{\sigma}}{\sqrt{n}}\,q_{\frac{p}{2}}\right],$$

where q_p is the quantile of Q such that $p = \mathbb{P}[Q \leq q_p]$ and $\hat{\sigma}$ is the measure of scale

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(X_t - \overline{X}\right)^2 + \widehat{LM}(\rho)}.$$

Note that since $\hat{\sigma} = O_P(n^{\zeta - 1/2})$ and $\zeta < 1$, $\hat{\sigma}/\sqrt{n}$ contracts to zero. Note that the construction of $\hat{\eta}_{SN}$ does not require explicit knowledge of α and β and their relationship; however, the distribution of Q does depend on these parameters as well as ρ, and its quantiles will be difficult to determine. Following the approach of McElroy and Politis (2002), one might be tempted to use subsampling in order to estimates Q's quantiles. Work by Hall, Lahiri and Jing (1998) indicates that subsampling is valid for stationary time series that can be expressed as certain instantaneous functions of a long memory Gaussian process. The strong mixing approach (see Politis, Romano and Wolf (1999, Appendix A), which provides a sufficient (but not necessary) condition for the validity of subsampling, is satisfied in the Gaussian case precisely when the spectral density exists at zero, i.e., $\beta = 0$. But it is unclear whether subsampling is valid for our heavy-tailed, long memory model (1).

4. Conclusion

This paper has introduced a stationary time series with both infinite variance heavy tails and finite autocovariances that exhibit long memory behavior. This model facilitated specific results on the rate of growth of the sample mean − the rate depended on whether tails ($1/\alpha$) or memory ($(1+\beta)/2$) dominate. In the latter case, a central limit theorem was derived, whereas in the former case a non-central limit theorem holds.

We have also produced a new self-normalization, which essentially combines the heavy-tail self-normalization of Logan, Mallows, Rice and Shepp (1973) and the long memory normalization suggested by the log periodogram estimator of Robinson (1995). In order for Corollary 1 to be practical, it is necessary to determine the quantiles of the limit Q. There are many obstacles to this: Q depends on *Tail* versus *Mem*, it depends on μ and C in the *Mem* case, and the distribution of $S/\sqrt{U}$ is unknown in analytical form. We suggest the use of subsampling to empirically determine Q's quantiles, although the validity of this procedure has not been verified. The verification of subsampling methods for infinite variance, long-range dependent data is challenging (since the usual methods require knowledge of the mixing coefficients as well as higher moments), but worthy of further investigation. By Hall, Lahiri and Jing (1998), subsampling would be valid (they provide a slightly different studentization) for the $\{G_t\}$ series alone.

Since $\{\sigma_t\}$ is *i.i.d.* and independent of the Gaussian sequence, it stands to reason that multiplication by the volatility series will not contribute any dependence; hence subsampling should be valid for $\{Y_t = \sigma_t G_t\}$. The authors intend to explore this question further.

Finally, we should say something about extending these results to more general processes. By assuming that $\{\epsilon_t\}$ is *i.i.d.* with cdf in the $\alpha/2$ domain of attraction − instead of assuming them to be exactly $\alpha/2$ stable themselves − we can generalize the tail thickness property in a natural way. Then $\sum_{t=1}^{n} \epsilon_t = O_P(n^{2/\alpha} L(n))$ for some slowly varying function L. Likewise, we may consider any light-tailed long memory stationary time series $\{G_t\}$, generalizing from the Gaussian case. One interesting approach, in the spirit of Taqqu (1975) and Hall, Lahiri and Jing (1998), is to consider $\{h(G_t)\}$ for a function h that is integrable with respect to the Gaussian distribution, and hence equipped with a Hermite polynomial expansion. Then $\sum_{t=1}^{n} h(G_t) = O_P(n^{\xi})$ where ξ depends on β and the Hermite rank of h. One can also generalize the $LM(\beta)$ condition to allow a slowly varying function in the sum of autocovariances. Then the overall rate of $\sqrt{\epsilon_t} \cdot h(G_t)$ would depend on α, ξ and h. The resulting model would encompass a wider-class of stationary, heavy-tailed, long memory time series, while still being amenable to the analytical techniques introduced in this paper.

Acknowledgements

Some of the research was conducted by the first author while at the U.S. Census Bureau. This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the author and not necessarily those of the U.S. Census Bureau. The first author would like to thank David Findley for his encouragement, and Arif Dowla for suggesting the graphic for Figure 1. The authors thank a referee for many useful references on the long memory literature.

Appendix. Proofs

Proof of Proposition 1. Strict stationarity is a simple exercise in probability, using Assumptions (A) and (D) and the fact that the $\{G_t\}$ series is strictly stationary. The marginal distribution is computed from Proposition 1.3.1 of Samorodnitsky and Taqqu (1994, p.20), noting that the scale parameter of the Gaussian variable G_t is $\sqrt{\gamma_G(0)/2}$. The second assertion follows from Property 1.2.16 and 1.2.17 of Samorodnitsky and Taqqu (1994, Chap.1), observing that $\sigma_t = \sqrt{\epsilon_t}$. Finally, the third point is a simple calculation:

$$\mathbb{E}[(X_t - \eta)(X_{t+h} - \eta)] = \mathbb{E}[\sigma_t G_t \sigma_{t+h} G_{t+h}] = \mathbb{E}[\sigma_t]\mathbb{E}[\sigma_{t+h}]\mathbb{E}[G_t G_{t+h}] = \mu^2 \gamma_G(h)$$

for $|h| > 0$.

Proof of Theorem 1. Here a technique of proof is introduced that reappears throughout. Let $\mathcal{E}$ denote the σ-field generated by the entire volatility series ϵ, from past to future. That is, $\mathcal{E} = \sigma(\epsilon) = \sigma(\epsilon_t, t \in \mathbb{Z})$. In the same manner, let the total information of G be $\mathcal{G} = \sigma(G) = \sigma(G_t, t \in \mathbb{Z})$. Of course, that the σ-fields $\mathcal{E}$ and $\mathcal{G}$ are independent with respect to the underlying probability measure $\mathbb{P}$ follows from assumption (A). First assume $LM(0)$ so that $\zeta = 1/\alpha$; then the characteristic function for the normalized sum is

$$\mathbb{E}\exp\{i\nu n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} Y_t\} = \mathbb{E}\Big[\mathbb{E}\Big[\exp\{i\nu n^{-\frac{1}{\alpha}}\sum_{t=1}^{n}\sigma_t G_t\}|\mathcal{E}\Big]\Big],$$

where ν is any real number and $i = \sqrt{-1}$. If we consider the inner conditional characteristic function, we have

$$\mathbb{E}\Big[\exp\{i\nu n^{-\frac{1}{\alpha}}\sum_{t=1}^{n}\sigma_t G_t\}|\mathcal{E}\Big] = \exp\Big\{-\frac{\nu^2}{2}n^{-\frac{2}{\alpha}}\sum_{s,t=1}^{n}\sigma_s\sigma_t\gamma_G(s-t)\Big\}$$

using the stability property of Gaussians. The double sum naturally splits into the diagonal and off-diagonal terms:

$$n^{-\frac{2}{\alpha}}\Big(\sum_{t=1}^{n}\sigma_t^2\gamma_G(0) + \sum_{s\neq t}\sigma_s\sigma_t\gamma_G(s-t)\Big). \tag{6}$$

Now viewing the volatility series as random, it follows from $LM(0)$ that the second term tends to zero in probability. Indeed, the first absolute moment is

$$\begin{aligned}\mathbb{E}\Big|n^{-\frac{2}{\alpha}}\sum_{h\neq 0}^{n-1}\sum_{t=1}^{n-|h|}\sigma_t\sigma_{t+h}\gamma_G(h)\Big| &\leq n^{-\frac{2}{\alpha}}\sum_{h\neq 0}^{n-1}\sum_{t=1}^{n-|h|}\mu^2|\gamma_G(h)| \\ &\leq \mu^2 n^{1-\frac{2}{\alpha}}\sum_{|h|\leq n}(1-\frac{|h|}{n})|\gamma_G(h)|\end{aligned}$$

and, by $LM(0)$ and the Dominated Convergence Theorem, the summation tends to $\sum_{h\in\mathbb{Z}}|\gamma_G(h)|$, which is finite. Finally, since $\alpha < 2$, the whole bound tends to zero as $n \to \infty$. By an $\mathbb{L}_1$-Markov inequality, we conclude that the so-called off-diagonal terms in (6) must tend to zero in probability, hence also in distribution.

The first term of (6) actually has the same distribution as the ϵ series, with scale $\gamma_G(0)$; this follows from stability, assumption (D), and the fact that $\sigma_t^2 = \epsilon_t$. Using the boundedness of $\exp\{-\nu^2/2\cdot\}$ and weak convergence – see Theorem 25.8 of Billingsley (1995, Section 25) – we see that

$$\mathbb{E}\exp\{i\nu n^{-\frac{1}{\alpha}}\sum_{t=1}^{n} Y_t\} \to \mathbb{E}\exp\{-\frac{\nu^2}{2}\gamma_G(0)\epsilon_\infty\},$$

where ϵ_∞ denotes a random variable with the same stable distribution as the volatility series ϵ. Using Proposition 1.2.12 of Samorodnitsky and Taqqu (1994, Chap.1), this limit becomes

$$\exp\left\{-\frac{\tau^{\frac{\alpha}{2}}}{\cos(\frac{\pi\alpha}{4})}\left(\frac{\nu^2}{2}\gamma_G(0)\right)^{\frac{\alpha}{2}}\right\} = \exp\Big\{-|\nu|^\alpha\Big(\frac{\gamma_G(0)}{2}\Big)^{\frac{\alpha}{2}}\Big\},$$

which is recognized as the characteristic function of a sαs variable with scale $\sqrt{\gamma_G(0)/2}$, as desired.

Case $1/\alpha > (\beta+1)/2$. Now assume that $1/\alpha > (\beta+1)/2$, so that $\zeta = 1/\alpha$. Then the $\mathbb{L}^1$ bound of the second term of (6) is $O(n^{1-2/\alpha}n^\beta)$ by $LM(\beta)$, and tends to zero as $n \to \infty$; the rest of the proof is identical.

Case $1/\alpha < (\beta+1)/2$. If we assume $1/\alpha < (\beta+1)/2$, then $\zeta = (\beta+1)/2$, so that (6) becomes

$$n^{-(\beta+1)}\left(\sum_{t=1}^n \sigma_t^2\gamma_G(0) + \sum_{s\neq t}\sigma_s\sigma_t\gamma_G(s-t)\right). \tag{7}$$

The first term is $O_P(n^{2/\alpha-(\beta+1)})$, which tends to zero as n increases. Now (7) can be rewritten as

$$n^{-(\beta+1)}\left(\sum_{|h|>0}^{n-1}\sum_{t=1}^{n-|h|}\sigma_t\sigma_{t+h}\gamma_G(h)\right).$$

Let $A_{h,n} = \sum_{t=1}^{n-|h|}\sigma_t\sigma_{t+h} - \mu^2$, which is a mean zero triangular array defined for $0 \leq |h| \leq n-1$. We require the following lemma.

Lemma 1. *Suppose that $X_1, \ldots, X_n$ have mean zero and are i.i.d. with cdf in $D(\alpha)$ and $1 < \alpha < 2$. Then there exists a rate a_n such that the normalized sums $U_n = a_n^{-1}\sum_{t=1}^n X_t$ converge weakly, and such that the absolute expectation is uniformly bounded, i.e., $U_n \stackrel{\mathcal{L}}{\Longrightarrow} U$ and $\sup_n \mathbb{E}|U_n| < C$ for a constant $C > 0$. The rate $a_n = n^{1/\alpha}L(n)$, where L is a slowly-varying function. The same results also hold if $X_t = Y_tY_{t+h} - \mathbb{E}[Y_t]^2$ for fixed h, where Y_t are i.i.d. with cdf in $D(\alpha)$.*

Proof of Lemma 1. The weak convergence result is well-known, since this is the defining property of $D(\alpha)$, and the formula for a_n is also well-known. Now the absolute expectation can be written as

$$\mathbb{E}|U_n| = \int_0^\infty \mathbb{P}[|U_n| > z]\,dz = \int_0^\infty \mathbb{P}\Big[|\sum_{t=1}^n X_t| > a_n z\Big]\,dz$$
$$\leq 1 + \int_1^\infty \mathbb{P}\Big[|\sum_{t=1}^n X_t| > a_n z\Big]\,dz.$$

By adapting the proof of Lemma 3 in Meerschaert and Scheffler (1998), for any $\delta > 0$ we can bound the above probability by $Cz^{-\alpha+\delta}$ for some constant $C > 0$. Simply choose δ so small such that $\alpha - \delta > 1$, so that the bound on the probabilities is an integrable function; this provides a bound on $\mathbb{E}|U_n|$.

For the case that $X_t = Y_t Y_{t+h} - \mathbb{E}[Y_t]^2$, we obtain weak convergence from Theorem 3.3 of Davis and Resnick (1986). To prove the bound on the corresponding U_n, divide the sum into $h+1$ sums of the form $\{Y_j Y_{j+h} - \mathbb{E}[Y^2], Y_{j+h+1} Y_{j+2h+1} - \mathbb{E}[Y^2], \cdots\}$ for $j = 1, 2, \ldots, h+1$. Now each of the $h+1$ sub-sums consists of independent terms, so the above argument can be applied. Using the triangle inequality, a bound for $\mathbb{E}|U_n|$ can be obtained by producing a bound for each of the $h+1$ sub-sums. This concludes the proof.

Returning to $A_{h,n}$, let $B_t(h) = \sigma_t \sigma_{t+h} - \mu^2$, so that $A_{h,n} = \sum_{t=1}^{n-|h|} B_t(h)$. Now for each fixed h, the variables $B_t(h)$ are mean zero with a cdf in $D(\alpha)$, which follows from Theorem 3.3 of Cline (1983) together with the fact that $\mathbb{E}|\sigma_t|^\alpha = \mathbb{E}\epsilon_t^{\alpha/2} = \infty$, and the regular variation of the tails of σ_t, i.e.,

$$\mathbb{P}[|\sigma_t| > x] = \mathbb{P}[\epsilon_t > x^2] \sim (x^2)^{-\frac{\alpha}{2}} L(x^2) = x^{-\alpha} L(x^2)$$

as $x \to \infty$, where $L(x)$ (and hence $L(x^2)$) is slowly-varying. Therefore by Lemma 1, for each fixed $h \neq 0$, $\mathbb{E}|A_{h,n} a_{n-|h|}^{-1}| \to \mathbb{E}|U_h|$ as $n \to \infty$, where a_n is the rate of growth of the sum of $B_t(h)$ and depends on h. In particular, $a_n = n^{1/\alpha} L_h(n)$ for a slowly-varying function L_h depending on h. It follows that, for some arbitrarily small $\delta > 0$, $\mathbb{E}|A_{h,n}| \leq C(n - |h|)^{\delta+1/\alpha} \leq C n^{\delta+1/\alpha}$ for all $|h| \leq n-1$ and n sufficiently large and a sufficiently large constant $C > 0$ (notice that the slowly-varying functions are eventually dominated by the polynomial growth). Therefore

$$\mathbb{E}\left|\sum_{|h|>0}^{n-1} A_{h,n} \gamma_G(h)\right| \leq \sum_{|h|>0}^{n-1} \mathbb{E}|A_{h,n}||\gamma_G(h)| \leq C\, n^{\delta+\frac{1}{\alpha}} \sum_{|h|=1}^{n} |\gamma_G(h)|,$$

which is of order $n^{1/\alpha+\beta+\delta}$ by $LM(\beta)$. Returning to (7), we have

$$\begin{aligned} & n^{-(\beta+1)} \sum_{|h|>0}^{n-1} \sum_{t=1}^{n-|h|} \sigma_t \sigma_{t+h} \gamma_G(h) \\ &= n^{-(\beta+1)} \sum_{|h|>0}^{n-1} \sum_{t=1}^{n-|h|} A_{h,n} \gamma_G(h) + n^{-(\beta+1)} \mu^2 \sum_{|h|>0}^{n-1} \sum_{t=1}^{n-|h|} \gamma_G(h). \end{aligned}$$

Now the first term is $O_P(n^{\delta+1/\alpha-1})$ by the $\mathbb{L}_1$ Markov inequality, and this tends to zero since $\alpha > 1$ and δ can be chosen less than $1 - 1/\alpha$. The second term is just

$$n^{-(\beta+1)}\mu^2 \sum_{|h|>0}^{n-1} (n-|h|)\gamma_G(h) = n^{-(\beta+1)}\mu^2 \sum_{s\neq t} \gamma_G(s-t).$$

Now it follows from Lemma 3.1 of Taqqu (1975) (using $m=1$) that for $\gamma(h) \sim C' h^{\beta-1}$, $\sum_{s,t}^{n} \gamma(s-t) \sim n^{\beta+1} 2C'/\beta(\beta+1)$. Letting $C' = \beta C/2$ and noting that $\beta > 0$, since $2/\alpha < \beta+1$, yields

$$n^{-(\beta+1)}\mu^2 \sum_{s\neq t}^{n} \gamma_G(s-t) \to \begin{cases} \frac{C\mu^2}{\beta+1} & \text{if} \quad \beta > 0 \\ (C-\gamma_G(0))\mu^2 & \text{if} \quad \beta = 0 \end{cases} = \frac{\tilde{C}\mu^2}{\beta+1}.$$

Hence the limiting characteristic function is (again by boundedness of integrand together with weak convergence)

$$\mathbb{E}\exp\{-\frac{\nu^2}{2}\frac{\tilde{C}\mu^2}{\beta+1}\} = \exp\{-\frac{\nu^2}{2}\frac{\tilde{C}\mu^2}{\beta+1}\},$$

which corresponds to a mean zero Gaussian with variance $\tilde{C}\mu^2/(\beta+1)$. This completes the proof of this case.

Case $1/\alpha = (\beta+1)/2$. Here we combine the two other cases, with both diagonal and off-diagonal terms converging. Since the diagonal converges weakly to an $\alpha/2$ stable but the off-diagonal converges in probability to a constant, we obtain weak convergence of their sum by Slutsky's Theorem. The resulting characteristic function is

$$\mathbb{E}\exp\Big\{-\frac{\nu^2}{2}\Big(\gamma_G(0)\epsilon_\infty + \frac{\tilde{C}\mu^2}{\beta+1}\Big)\Big\} = \exp\Big\{-|\nu|^\alpha\Big(\frac{\gamma_G(0)}{2}\Big)^{\frac{\alpha}{2}}\Big\}\cdot\exp\{-\frac{\nu^2}{2}\frac{\tilde{C}\mu^2}{\beta+1}\},$$

which corresponds to the sum of two independent random variables, a stable S and a normal V. This completes the proof.

Proof of Theorem 2. First we establish that the last assertion holds, given the first weak convergence. Sine

$$\sum_{t=1}^{n}(X_t-\overline{X})^2 = \sum_{t=1}^{n}(Y_t-\overline{Y})^2 = \sum_{t=1}^{n} Y_t^2 - n\overline{Y}^2.$$

The second term is, by Theorem 1, bounded in probability of order $n^{2\zeta-1}$; when multiplied by the rate $n^{2/\alpha}$, this yields

$$n^{2\zeta-1-\frac{2}{\alpha}} = \begin{cases} -1 & \text{if} \quad \frac{2}{\alpha} > \beta+1 \\ \beta-\frac{2}{\alpha} & \text{if} \quad \frac{2}{\alpha} \leq \beta+1, \end{cases}$$

and note that the second quantity is always negative since $\beta < 1$ and $\alpha < 2$. Therefore,

$$n^{-\frac{2}{\alpha}} \sum_{t=1}^{n} (X_t - \overline{X})^2 = o_P(1) + n^{-\frac{2}{\alpha}} \sum_{t=1}^{n} Y_t^2.$$

Here we consider the joint Fourier/ Laplace Transform of the first and second sample moments. It is sufficient to take a Laplace transform in the second component, since the sample second moment is a positive random variable (see Fitzsimmons and McElroy (2006)). So, for any real θ and $\phi > 0$, we have

$$\begin{aligned}
&\mathbb{E}\exp\{i\theta n^{-\zeta}\sum_{t=1}^{n} Y_t - \phi n^{-2\zeta}\sum_{t=1}^{n} Y_t^2\} \\
&= \mathbb{E}\exp\{i\theta n^{-\zeta}\sum_{t=1}^{n} Y_t + i\sqrt{2\phi}n^{-\zeta}\sum_{t=1}^{n} Y_t N_t\} \\
&= \mathbb{E}\exp\{in^{-\zeta}\sum_{t=1}^{n} \sigma_t G_t(\theta + \sqrt{2\phi}N_t)\} \\
&= \mathbb{E}[\mathbb{E}[\exp\{in^{-\zeta}\sum_{t=1}^{n} \sigma_t G_t(\theta + \sqrt{2\phi}N_t)\}|\mathcal{E},\mathcal{N}]] \\
&= \mathbb{E}[\exp\{-\frac{1}{2}n^{-\zeta}\sum_{t,s}^{n} \sigma_t\sigma_s\gamma_G(t-s)(\theta + \sqrt{2\phi}N_t)(\theta + \sqrt{2\phi}N_s)\}].
\end{aligned}$$

The sequence of random variables N_t are *i.i.d.* standard normal, and are all independent of the Y_t series. Their common information is denoted by $\mathcal{N}$. In the first equality, we have conditioned on $\mathcal{E}, \mathcal{G}$, noting that

$$\mathbb{E}[\exp\{i\sqrt{2\phi}n^{-\zeta}\sum_{t=1}^{n} Y_t N_t\}|\mathcal{E},\mathcal{G}] = \exp\{-\phi n^{-2\zeta}\sum_{t=1}^{n} Y_t^2\}$$

by the definition of the multivariate normal characteristic function. Now we break the double sum in our Fourier/Laplace Transform into diagonal and off-diagonal terms, as in the proof of Theorem 1. The off-diagonal term is

$$n^{-2\zeta}\sum_{|h|>0}^{n-1}\sum_{t=1}^{n-|h|} \sigma_t\sigma_{t+h}(\theta + \sqrt{2\phi}N_t)(\theta + \sqrt{2\phi}N_{t+h})\gamma_G(h). \tag{8}$$

First take the case that $2/\alpha > \beta + 1$. Then the absolute expectation of (8) is

$$\mathbb{E}[\mathbb{E}[n^{-\frac{2}{\alpha}} \left| \sum_{|h|>0}^{n-1} \sum_{t=1}^{n-|h|} \sigma_t \sigma_{t+h} (\theta + \sqrt{2\phi} N_t)(\theta + \sqrt{2\phi} N_{t+h}) \gamma_G(h) \right| |\mathcal{N}]]$$

$$\leq \mathbb{E}[n^{-\frac{2}{\alpha}} \sum_{|h|>0}^{n-1} \sum_{t=1}^{n-|h|} \mu^2 |\theta + \sqrt{2\phi} N_t| |\theta + \sqrt{2\phi} N_{t+h}| |\gamma_G(h)|]$$

$$= \mu^2 n^{-\frac{2}{\alpha}} \sum_{|h|>0}^{n-1} (n - |h|) (\mathbb{E}|\theta + \sqrt{2\phi} N|)^2 |\gamma_G(h)|$$

$$= \mu^2 (\mathbb{E}|\theta + \sqrt{2\phi} N|)^2 n^{1-\frac{2}{\alpha}} \sum_{|h|>0}^{n-1} (1 - |h|/n) |\gamma_G(h)|.$$

The summation over h is $O(n^{\beta})$ by $LM(\beta)$, so the overall order is $n^{\beta+1-2/\alpha}$ which tends to zero as $n \to \infty$. By the $\mathbb{L}_1$ Markov inequality, this shows that the off-diagonal term tends to zero in probability.

In the case $2/\alpha \leq \beta + 1$, we write (8) as

$$n^{-(\beta+1)} \sum_{|h|>0}^{n-1} \sum_{t=1}^{n-|h|} (\sigma_t \sigma_{t+h} - \mu^2)(\theta + \sqrt{2\phi} N_t)(\theta + \sqrt{2\phi} N_{t+h}) \gamma_G(h)$$

$$+ \mu^2 n^{-(\beta+1)} \sum_{|h|>0}^{n-1} \sum_{t=1}^{n-|h|} \gamma_G(h) (\theta + \sqrt{2\phi} N_t)(\theta + \sqrt{2\phi} N_{t+h}).$$

Let $V_t = \theta + \sqrt{2\phi} N_t$, so that these form an *i.i.d.* sequence of normals with mean θ and variance 2ϕ. For each $h \neq 0$,

$$\sum_{t=1}^{n-|h|} (\sigma_t \sigma_{t+h} - \mu^2) V_t V_{t+h}$$
$$= \sum_{t=1}^{n-|h|} (\sigma_t \sigma_{t+h} V_t V_{t+h} - \mu^2 \theta^2) - \sum_{t=1}^{n-|h|} \mu^2 (V_t V_{t+h} - \theta^2). \tag{9}$$

Letting $W_t = \sigma_t V_t$, these form an *i.i.d.* sequence so that $W_t W_{t+h}$ has cdf in $D(\alpha)$. By Lemma 1, there exists a $\delta > 0$ such that for all h,

$$\mathbb{E} \left| \sum_{t=1}^{n-|h|} W_t W_{t+h} - \mu^2 \theta^2 \right| = O(n^{\delta + \frac{1}{\alpha}}).$$

Since $\{V_t V_{t+h}\}$ are $h+1$-dependent, the Strong Law of Large Numbers applies, so that

$$\mathbb{E}\left| n^{-\frac{1}{2}} \sum_{t=1}^{n-|h|} V_t V_{t+h} - \theta^2 \right| \leq \mathbb{E}\left(n^{-\frac{1}{2}} \sum_{t=1}^{n-|h|} V_t V_{t+h} - \theta^2 \right)^2 = O(1).$$

Hence the absolute expectation of (9) is $O(n^{\delta+1/\alpha})$, since $1/2 < \delta + 1/\alpha$. Thus the first term of (8) is bounded in probability of order $n^{-(\beta+1)+\beta+\delta+1/\alpha}$, which tends to zero for small δ. For the second term of (8), we break it down further as

$$\mu^2 n^{-(\beta+1)} \sum_{|h|>0}^{n-1} \sum_{t=1}^{n-|h|} \gamma_G(h) V_t V_{t+h}$$
$$= \mu^2 \theta^2 n^{-(\beta+1)} \sum_{|h|>0}^{n-1} \gamma_G(h)(n-|h|) + \mu^2 n^{-(\beta+1)} \sum_{|h|>0}^{n-1} \gamma_G(h) \sum_{t=1}^{n-|h|} (V_t V_{t+h} - \theta^2).$$

For the second term, the absolute expectation is $O(n^{-(\beta+1)+\beta+1/2})$, which tends to zero. The first term becomes

$$n^{-(\beta+1)} \mu^2 \theta^2 \sum_{|h|>0}^{n-1} (n-|h|) \gamma_G(h) \to \tilde{C} \theta^2 \frac{\mu^2}{\beta+1}$$

as $n \to \infty$, as in the proof of Theorem 1. Putting both cases together, we see that the off-diagonal terms converge in probability to $\tilde{C}[\mu^2\theta^2/(\beta+1)] 1_{\{2/\alpha \leq \beta+1\}}$.

Since the off-diagonal terms tend to a constant, by the Dominated Convergence Theorem we can examine the characteristic function of the diagonal terms separately. First,

$$\mathbb{E}[\exp\{-\frac{1}{2}\gamma_G(0) n^{-2\zeta} \sum_{t=1}^{n} \sigma_t^2 V_t^2\}] = \mathbb{E}[\mathbb{E}[\exp\{-\frac{1}{2}\gamma_G(0) n^{-2\zeta} \sum_{t=1}^{n} \epsilon_t V_t^2\} | \mathcal{N}]]$$
$$= \mathbb{E}[\mathbb{E}[\exp\{-\frac{1}{2}\gamma_G(0) n^{-2\zeta} \left(\sum_{t=1}^{n} |V_t|^\alpha \right)^{\frac{2}{\alpha}} \epsilon_\infty\} | \mathcal{N}]]$$
$$= \mathbb{E}[\exp\{-(\frac{\gamma_G(0)}{2})^{\frac{\alpha}{2}} n^{-\alpha\zeta} \sum_{t=1}^{n} |V_t|^\alpha\}].$$

In the case that $2/\alpha \geq \beta + 1$, $\zeta = 1/\alpha$, so that by the Law of Large Numbers,

$$n^{-\alpha\zeta} \sum_{t=1}^{n} |V_t|^\alpha \xrightarrow{P} \mathbb{E}|V|^\alpha.$$

But if $2/\alpha < \beta + 1$, this same sum tends to zero in probability, since $\zeta\alpha > 1$ in that case. By the Dominated Convergence Theorem, the limit as $n \to \infty$ can be taken through the expectation, so that

$$\mathbb{E}[\exp\{-(\frac{\gamma_G(0)}{2})^{\frac{\alpha}{2}} n^{-\alpha\zeta} \sum_{t=1}^{n} |V_t|^\alpha\}] \\ \to \exp\{-(\frac{\gamma_G(0)}{2})^{\frac{\alpha}{2}} \mathbb{E}|\theta + \sqrt{2\phi} N|^\alpha 1_{\{\frac{2}{\alpha} \geq \beta+1\}}\}.$$

Combining this with the off-diagonal terms produces the joint Fourier/Laplace functional stated in Theorem 2. When $2/\alpha > \beta+1$, it becomes $\exp\{-(\gamma_G(0)/2)^{\alpha/2} \mathbb{E}|\theta + \sqrt{2\phi} N|^\alpha\}$. The $\theta = 0$ case provides the Laplace Transform of U, which is $\exp\{-\phi^{\alpha/2}\mathbb{E}|Z|^\alpha(\gamma_G(0)/2)^{\alpha/2}\}$. Letting $\phi = 0$ yields $\exp\{-|\theta|^\alpha(\gamma_G(0)/2)^{\alpha/2}\}$, which is the characteristic function of the S of Theorem 1. On the other hand, letting $2/\alpha < \beta+1$ yields $\exp\{-[(\tilde{C}\mu^2\theta^2)/(2(\beta+1))]\}$, which is the characteristic function of V from Theorem 1. Since there is no ϕ-argument in the limit in this case, the sample second moment converges in probability to zero when normalized by rate $n^{-2\zeta}$. This completes the proof.

Proof of Theorem 3. We suppress the greatest integer notation on n^ρ for ease of presentation. First considering the Y series, we have

$$\sum_{|h|>0}^{n^\rho} \frac{1}{n-|h|} \sum_{t=1}^{n-|h|} Y_t Y_{t+h} = \sum_{|h|>0}^{n^\rho} \frac{1}{n-|h|} \sum_{t=1}^{n-|h|} \sigma_t \sigma_{t+h} G_t G_{t+h} \\ = \sum_{|h|>0}^{n^\rho} \mu^2 \gamma_G(h) + \sum_{|h|>0}^{n^\rho} \frac{1}{n-|h|} \gamma_G(h) \sum_{t=1}^{n-|h|} (\sigma_t\sigma_{t+h} - \mu^2) \\ + \sum_{|h|>0}^{n^\rho} \frac{1}{n-|h|} \sum_{t=1}^{n-|h|} \sigma_t\sigma_{t+h}(G_tG_{t+h} - \gamma_G(h)).$$

The first term is, by $LM(\beta)$, asymptotic to $\mu^2 C n^{\beta\rho}$. We need to show that the other two terms are $o(n^{\beta\rho})$. The absolute expectation of the second term is bounded by

$$\sum_{|h|>0}^{n^\rho} \frac{1}{n-n^\rho} |\gamma_G(h)| \, \mathbb{E}\left| \sum_{t=1}^{n-|h|} (\sigma_t\sigma_{t+h} - \mu^2) \right|.$$

Since $\rho < 1$, $n - n^\rho = n(1+o(1))$. Making use of Lemma 1 once again, the expectation is $O(n^{\delta+1/\alpha})$ for arbitrarily small $\delta > 0$ and all $|h| \leq n^\rho$. Hence the overall bound on the above expectation is $O(n^{\beta\rho-1+\delta+1/\alpha})$, whose quotient by $n^{\beta\rho}$ tends to zero.

For the third term, we also take an absolute expectation, yielding the bound

$$\begin{aligned}
&\sum_{|h|>0}^{n^\rho} \frac{1}{n-|h|}\mathbb{E}|\sum_{t=1}^{n-|h|} \sigma_t\sigma_{t+h}(G_tG_{t+h} - \gamma_G(h))| \\
&\quad \leq \sum_{|h|>0}^{n^\rho} \frac{1}{n-|h|}\mathbb{E}\sum_{t=1}^{n-|h|} \sigma_t\sigma_{t+h}|G_tG_{t+h} - \gamma_G(h)| \\
&\quad = \sum_{|h|>0}^{n^\rho} \frac{1}{n-|h|}\mathbb{E}[\mathbb{E}[\sum_{t=1}^{n-|h|} \sigma_t\sigma_{t+h}|G_tG_{t+h} - \gamma_G(h)||\mathcal{E}]] \\
&\quad \leq \sum_{|h|>0}^{n^\rho} \frac{1}{n-|h|}\mathbb{E}[\sum_{t=1}^{n-|h|} \sigma_t\sigma_{t+h}\mathrm{Var}\,[G_tG_{t+h}]] \\
&\quad \leq \sum_{|h|>0}^{n^\rho} \mathrm{Var}\,[G_0G_h]\mu^2.
\end{aligned}$$

Now by Assumption (E), $\{G_t\}$ has an $MA(\infty)$ representation, and one can easily compute

$$\mathrm{Var}\,[G_0G_h] = \mathbb{E}[G_0^2G_h^2] - \gamma_G^2(h) = 2\gamma_G(h)^2 \sim \frac{\beta^2C^2}{2}h^{2(\beta-1)}$$

as $|h| \to \infty$. This is summable for $\beta < 1/2$. Therefore

$$\sum_{h=1}^{n^\rho} \gamma_G^2(h) = \begin{cases} O(1) & \text{if} \quad \beta < \frac{1}{2} \\ O(\log n) & \text{if} \quad \beta = \frac{1}{2} \\ O(n^{\rho(2\beta-1)}) & \text{if} \quad \beta > \frac{1}{2}. \end{cases}$$

In the last case the sum grows fastest, but $n^{\rho(2\beta-1)}n^{-\beta\rho} = n^{(\beta-1)\rho} \to 0$. Therefore the third term is $o_P(n^{\beta\rho})$, and

$$n^{-\beta\rho}\sum_{|h|>0}^{n^\rho} \frac{1}{n-|h|}\sum_{t=1}^{n-|h|} Y_tY_{t+h} \xrightarrow{P} \mu^2C$$

as $n \to \infty$. Note that there are no restrictions on the choice of ρ, i.e., it does not depend on α or β.

Now we turn to $\widehat{LM}(\rho)$. For each fixed h,

$$\frac{1}{n-|h|}\sum_{t=1}^{n-|h|} X_t X_{t+h} - \overline{X}^2 = \frac{1}{n-|h|}\sum_{t=1}^{n-|h|} Y_t Y_{t+h} - \overline{Y}^2 + \frac{\eta}{n-|h|}\sum_{t=1}^{n-|h|}(Y_t + Y_{t+h} - 2\overline{Y}).$$

The second term, by Theorem 1, is $O_P(n^{2(\zeta-1)})$. If we sum over h and multiply by $n^{-\beta\rho}$, the second term is bounded in probability of order $n^{(1-\beta)\rho+2(\zeta-1)}$, which tends to zero. Consider the first part of the third term:

$$\frac{1}{n-|h|}\sum_{t=1}^{n-|h|} Y_t - \overline{Y} = \frac{-1}{n-|h|}\sum_{t=n-|h|+1}^{n} Y_t + \frac{|h|}{n(n-|h|)}\sum_{t=1}^{n} Y_t.$$

Consider the worst case scenario, for all $|h| \leq n^\rho$. The sum $\sum_{t=n-|h|+1}^{n} Y_t$ has at most n^ρ terms, and so is $O_P(n^{\zeta\rho})$ by Theorem 1. Also, $|h|/(n(n-|h|)) = O(n^{\rho-2})$. So we have the sum of an $O_P(n^{\zeta\rho-1})$ term and an $O_P(n^{\zeta+\rho-2})$ term; using $(1-\rho)(1-\zeta) > 0$, the first sum has a greater rate of growth. Now summing over h and multiplying by $n^{-\beta\rho}$, we have $O_P(n^{(1-\beta)\rho+\zeta\rho-1}) \to 0$. This demonstrates that

$$n^{-\beta\rho}(\widehat{LM}(\rho))^\rho = o_P(1) + \left| n^{-\beta\rho}\sum_{|h|>0}^{n^\rho}\frac{1}{n-|h|}\sum_{t=1}^{n-|h|} Y_t Y_{t+h}\right| \xrightarrow{P} \mu^2 C,$$

which proves the theorem by taking ρth powers.

Proof of Theorem 4. The first convergence (4) follows from Theorems 2 and 3, together with Slutsky's Theorem. We also used the equivalence of the centered sample second moment of the $\{X_t\}$ series with the sample second moment of the $\{Y_t\}$ series, as established in Theorem 2. The second convergence uses the Continuous Mapping Theorem, noting that the denominators are nonzero ($C > 0$ by assumption).

Proof of Corollary 1. It follows from Theorem 4 that

$$1-p = \mathbb{P}[q_{p/2} \leq Q \leq q_{1-p/2}] \approx \mathbb{P}\left[q_{p/2} \leq \sqrt{n}\frac{(\overline{X}-\eta)}{\hat{\sigma}} \leq q_{1-\frac{p}{2}}\right] = \mathbb{P}\left[\overline{X} - \frac{\hat{\sigma}}{\sqrt{n}} q_{1-\frac{p}{2}} \leq \eta \leq \overline{X} - \frac{\hat{\sigma}}{\sqrt{n}} q_{\frac{p}{2}}\right].$$

References

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman and Hall, New York.

Billingsley, P. (1995). *Probability and Measure.* Wiley, New York.

Brockwell, P. and Davis, R. (1991). *Time Series: Theory and Methods.* Springer-Verlag, New York.

Cline, D. (1983). Infinite series of random variables with regularly varying tails. Technical Report, 83-24, Institute of Apllied Mathematics and Statistics, University of British-Columbia.

Davis, R. and Resnick, S. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.* **13**, 179-195.

Davis, R. and Resnick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. *Ann. Statist.* **14**, 533-558.

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

Fitzsimmons, P. and McElroy, T. (2006). On joint Fourier-Laplace transforms. *Mimeo.* http:// www.math.ucsd.edu/ politis/PAPER/FL.pdf.

Granger, C. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Ser. Anal.* **1**, 15-29.

Greene, M. T. and Fielitz, B. D. (1977) Long-Term Dependence in Common Stock Returns. *J. Finan. Econom.* **4**, 339-349.

Hall, P. (1997). Defining and measuring long-range dependence. *Fields Inst. Commun.* **11**, 153-160.

Hall, P., Lahiri, S. and Jing, B. (1998). On the sampling window method for long-range dependent data. *Statist. Sinica* **8**, 1189-1204.

Heath, D., Resnick, S. and Samorodnitsky, G. (1998). Heavy tails and long range dependence in on/off processes and associated fluid models. *Math. Oper. Res.* **23**, 145-165.

Heath, D., Resnick, S. and Samorodnitsky, G. (1999). How system performance is affected by the interplay of averages in a fluid queue with long range dependence induced by heavy tails. *Ann. Appl. Probab.* **9**, 352-375.

Heyde, C. and Yang, Y. (1997). On defining long-range dependence. *J. Appl. Probab.* **34**, 939-944.

Logan, B. F., Mallows, C. L., Rice, S. O. and Shepp, L. A. (1973). limit distributions of self-normalized sums. *Ann. Probab.* **1**, 788-809.

Mandelbrot, B. and Wallis, J. R. (1968). Noah, Joseph and operation hydrology. *Water Resources Res.* **4**, 909-917.

Mansfield, P., Rachev, S. and Samorodnitsky, G. (2001). Long strange segments of a stochastic process and long-range dependence. *Ann. Appl. Probab.* **11**, 878-921.

McElroy, T. and Politis, D. (2002). Robust inference for the mean in the presence of serial correlation and heavy tailed distributions. *Econom. Theory* **18**, 1019-1039.

McElroy, T. and Politis, D. (2004). Large sample theory for statistics of stable moving averages. *J. Nonparametr. Stat.* **16**, 623-657.

Meerschaert, M. and Scheffler, H. (1998). A simple robust estimator for the thickness of heavy tails. *J. Statist. Plann. Inference* **71**, 19-34.

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

Rachev S. and Samorodnitsky, G. (2001). Long strange segments in a long range dependent moving average. *Stochastic Process. Appl.* **93**, 119-148.

Robinson, P. (1995). Gaussian estimation of long range dependence. *Ann Statist.* **23**, 1630-1661.

Samorodnitsky, G. and Taqqu, M. (1994). *Stable Non-Gaussian Random Processes.* Chapman and Hall, New York.

Taqqu, M. (1975). Weak convergence to fractional Brownian and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31**, 287-302.

Mathematical Statistician, Statistical Research Division, U.S. Census Bureau.
University of California, San Diego La Jolla, CA 92093-0112, U.S.A.

E-mail: tucker.s.mcelroy@census.gov

Department of Mathematics, University of California, San Diego La Jolla, CA 92093-0112, U.S.A.

E-mail: dpolitis@ucd.edu

(Received June 2005; accepted June 2006)