TESTS OF SHORT MEMORY WITH THICK TAILED ERRORS

Christine Amsler
Michigan State University

Peter Schmidt
Michigan State University

October, 1999

The second author gratefully acknowledges the financial support of the National Science Foundation.

1. INTRODUCTION

This paper considers two tests of the null hypothesis of short memory, the KPSS test of Kwiatkowski *et al.* (1992) and the modified rescaled range (MR/S) test of Lo (1991). The KPSS test was designed as a test of stationarity against the alternative of a unit root. However, the asymptotic distribution theory in KPSS actually assumes short memory, and Lee and Schmidt (1996) have shown that the test has power against stationary and nonstationary long memory alternatives. Conversely, Lo's MR/S test was designed as a test of short memory against stationary long memory alternatives, but it also has power against unit roots.

In this paper we are primarily concerned with the magnitude of possible finite sample size distortions caused by errors with thick tails. Our intuition suggested that these size distortions should be less for the KPSS test than for the MR/S test. Both tests depend on the cumulations of the series, after demeaning and/or detrending as required. However, the KPSS test depends on the sum of squares of the cumulations, while the MR/S test depends on the difference between the maximal and minimal value of the cumulations. Intuitively, maximal and minimal values may be very sensitive to tail thickness, so that the finite sample size distortions caused by thick tails may be worse for the MR/S test than for KPSS. Simulations reported in Section 3 of the paper indicate that this is so.

In order to move beyond a pure simulation study, some further analytical framework is needed. When the data distribution is in the domain of attraction of a stable law, different limit theory applies, based on the Levy process. See, for example, Chan and Tran (1989), Phillips (1990), Knight (1991) and Caner (1997). In this paper we are interested in fat tails *without* infinite variance, for which standard Wiener-process asymptotics apply but may fail to be adequate in finite samples when tails are thick. We therefore consider *local to finite variance* *asymptotics*, based on a representation of the series X_t as follows: $X_t = X_{1t} + (c/g(T))X_{2t}$. Here X_{1t} is a short memory process without thick tails (e.g. normal) and X_{2t} is a symmetric stable distribution. The term "c" is a constant, while the term g(T) is chosen so that the cumulation of X_t converges weakly to a weighted sum of the Wiener and Levy processes, with the weight depending on c. For example, for the case that X_{2t} is Cauchy, $g(T) = T^{1/2}$. In Section 4 we derive the asymptotic distributions of the KPSS and MR/S statistics under this representation of the process; they each depend on the parameter c. We then evaluate the asymptotic distributions to see which distribution is more sensitive to the value of c. The asymptotic distribution of the MR/S statistic is more sensitive to c than the asymptotic distribution of the KPSS statistic, suggesting the prospect of more substantial size distortions with thick tailed errors, as was found in our simulations.

Future research will address the prospect of functionals other than sums of squares or maximum minus minimum, where the aim is to find a test that is less sensitive than KPSS to tail thickness and that still has reasonable power properties. An example might be the point-optimal test for a particular value of c in a data generating mechanism like the one above. Or, less formally, the robustness literature is full of alternatives to sums of squares, and these could be considered.

2. NOTATION AND FURTHER DISCUSSION

Let X_t, t = 1,...,T, be an observed series. We wish to test the null hypothesis that X_t has short memory. Let e_t be the series after demeaning or demeaning and detrending, as necessary.

Thus, if X_t is asserted to have zero mean, $e_t = X_t$; if X_t has non-zero mean but no deterministic trend, $e_t = X_t - \overline{X} \equiv$ "demeaned" X_t; and if X_t has deterministic trend, e_t = residuals from a regression of X_t on [1,t], t = 1,...,T, $\equiv$"demeaned and detrended" X_t. Let the cumulation of the e_t be S_t: $S_t = \Sigma_{j=1}^{t} e_j$. For j = 0, 1, ..., we define the estimated autocovariances $\hat{\gamma} = T^{-1} \Sigma_{t=j+1}^{T} e_t e_{t-j}$. Finally, we define the long run variance estimate $s^2(\ell)$ in the usual way, using "ℓ" lags: for ℓ = 0, we have $s^2(0) = \hat{\gamma}_0$, while for $\ell \geq 1$, we have $s^2(\ell) = \hat{\gamma} + 2\Sigma_{s=1}^{\ell}\ w(s,\ell)\hat{\gamma}_s$, where $w(s,\ell)$ is an optional weighting function, such as the Bartlett window $w(s,\ell) = 1 - s/(\ell+1)$.

We will now define the KPSS and MR/S tests. For the zero mean case, so that $X_t = e_t$, we define the KPSS statistic $\hat{\eta}(\ell)$ and the MR/S statistic $Q(\ell)$:

(1) $$\hat{\eta}(\ell) = T^{-2}\, \Sigma_{t=1}^{T} S_t^2 \,/\, s^2(\ell)\ ,\ \ Q(\ell) = [\max_t S_t - \min_t S_t]/s(\ell).$$

In the case that e_t = demeaned X_t, we will use the notation $\hat{\eta}_\mu(\ell)$ and $Q_\mu(\ell)$, while we will use the notation $\hat{\eta}_\tau(\ell)$ and $Q_\tau(\ell)$ for the case that e_t = demeaned and detrended X_t. This notation is slightly different than in the original articles. Neither KPSS nor Lo explicitly indicated dependence on ℓ. KPSS considered $\hat{\eta}_\mu$ and $\hat{\eta}_\tau$ but not the zero mean case. Lo considered only Q_μ, which he called Q_n for $\ell \geq 1$, and $\tilde{Q}_n$ for ℓ = 0.

Some modified versions of the KPSS test have been proposed. These statistics avoid the long run variance estimate $s^2(\ell)$ by making other types of corrections for short run dynamics. Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994) suggested a parametric correction, based on fitting an ARIMA(p,1,1) model to the data, while Leybourne and McCabe (1999) consider an alternative long run variance estimate. The MR/S test could presumably be modified in the same ways. We will not consider these tests in this paper, but we believe that a similar analysis would apply, since our comparison of the KPSS and MR/S tests depends on features of the numerators of the statistics, not the long run variance estimate.

The KPSS and MR/S tests each have some function of the cumulations S_t in the numerator, and either a long run variance estimate or its square root in the denominator. The potential difference in the robustness properties of the statistics arises from the nature of the functions of the cumulations, $\Sigma_t S_t^2$ versus $[\max_t S_t - \min_t S_t]$. Maxima and minima may be very sensitive to tail thickness, and so the MR/S test might be expected to have poor robustness properties. We can elaborate on this argument as follows. In both cases the fundamental result driving the asymptotic validity of the test is the weak convergence of the normalized cumulation to a variant of the Wiener process W(r). More specifically, under certain standard assumptions of short memory (e.g., Phillips (1987, p. 280)), we have $T^{-½} S_{[rT]} \Rightarrow \sigma W(r)$, where $r \varepsilon [0,1]$, [rT] is the integer part of rT, and σ^2 is the long run variance of the process. (In the demeaned case, we have the same result if we replace W(r) by the Brownian bridge $W_\mu(r) = W(r) - rW(1)$, and in the demeaned and detrended case we replace W(r) by the second-level Brownian bridge $W_\tau(r) = W(r) - (2r - 3r^2)W(1) + (-6r + 6r^2) \int_0^1 W(s)ds$.) The convergence of $T^{-½} S_{[rT]}$ to $\sigma W(r)$ implies the convergence of $T^{-2} \Sigma_t S_t^2$ to $\sigma^2 \int_0^1 W(s)ds$ and the convergence of $T^{-½} [\max_t S_t - \min_t S_t]$ to $\sigma [\max_s W(s) - \min_s W(s)]$. However, the convergence of these functions of the S_t to the corresponding functionals of W(s) need not be equally quick. In particular, the function $\Sigma_t S_t^2$ involves a summation, and thus a second step of averaging (the first step being in the construction of the S_t), that is absent in the maximum and minimum. As a result it is intuitively plausible that the convergence may be faster, and correspondingly the asymptotics may be accurate for smaller samples sizes, for the KPSS test than for the MR/S test.

3. SIMULATIONS

In this section we report the results of some simulations designed to compare the finite sample robustness of the KPSS and MR/S tests to nonnormality. The data generating process is very simple: the X_t, t = 1,...,T, are iid. The data are centered on zero and there is no deterministic trend. We consider sample sizes T = 50, 100, 200, 500, 1000 and 2000. The following distributions for the X_t are considered: standard normal; student *t* with 10, 5 and 2 degrees of freedom; and standard Cauchy. The KPSS and MR/S tests are valid asymptotically for the normal case and for the *t* distributions with 10 or 5 degrees of freedom. The asymptotics are not valid for the other two cases, since the t_2 distribution has an infinite variance, while for the Cauchy the mean does not exist and the variance is infinite.

The data are generated by drawing pseudo-random normal deviates and transforming them as appropriate. (For example, the Cauchy is the ratio of two standard normals, while the *t* is the ratio of a standard normal to the square root of a chi-squared divided by its degrees of freedom, and the chi-squared is generated by the sum of the squares of standard normals.) We used the random number generator GASDEV/RAN3 of Press *et al.* (1986). The number of replications was 5000.

We considered the mean-corrected versions of the KPSS and MR/S tests, $\hat{\eta}_\mu(\ell)$ and $Q_\mu(\ell)$, because they are empirically more relevant than the zero-mean versions. Our results are invariant to the addition of a constant to the data, so that our choice to generate data centered on zero is not restrictive. We consider only the upper tail 5% tests. We chose three values of

$\ell : \ell = 0; \ell = \ell 4 \equiv$ integer[4(T/100)$^{1/4}$], and $\ell = \ell 12 \equiv$ integer[12(T/100)$^{1/4}$], as in Schwert (1989) and many recent simulations. The use of $\ell > 0$ is unnecessary with iid data, but was considered due to its empirical relevance. We could have generated data with short run dynamics, but to do so would have much enlarged the scope of the experiment, and we hoped to focus on the effect of the distribution of the X_t in essentially the simplest possible setting.

Table 1 gives the size of the 5% upper tail KPSS and MR/S tests.[1] We will discuss first the results for the tests with $\ell = 0$, given in the first block of numbers in the table. These results are rather striking. For essentially every sample size and error distribution, the KPSS test has size more nearly equal to nominal size (5%) than does the MR/S test. For the distributions for which the test is asymptotically valid (normal, t_{10} and t_5), the KPSS test displays only very minimal size distortions, even for T as small as 50. The MR/S test suffers from size distortions that were (to us) surprisingly large. For example, for the normal case, size is only 0.010 at T = 50, and it improves rather slowly to 0.028 for T = 200, 0.038 for T = 500, and 0.041 for T = 1000, before finally reaching 0.048 for T = 2000. For the t_{10} and t_5 cases the convergence is even slower, though not strikingly so. The t_5 is perhaps closest to what we might think of as an empirically relevant thick-tailed case, and it is clear that the KPSS test is more reliable for this case than the MR/S test is. However, most of this difference is also present in the normal case; the unreliability of the MR/S test in moderately large samples is mostly just due to slow convergence to its asymptotic distribution in general, not really to the effects of thick tails.

---

[1] For the KPSS and MR/S tests we used the critical values given by KPSS (1992, p.166) and Lo (1991, pp. 1288), respectively. The KPSS critical value of 0.463 was calculated by a simulation with T = 2000. Simulations reported in Section 5 of this paper for T = 10000 yield a slightly larger critical value, 0.476. The use of this critical value instead of the one from KPSS would have reduced the difference in performance of the KPSS and MR/S tests reported below, but only very slightly.

The case of the t_2 distribution is intriguing, even without any asymptotic theory to guide us. From the simulations it would appear that the KPSS test is valid for large T, while this is not apparent for the MR/S test. For the Cauchy distribution, the tests are *not* valid asymptotically (the appropriate asymptotic theory will be discussed in the next section), but the size distortions are very much larger for the MR/S test than for KPSS. The MR/S test essentially never rejects.

We will discuss only very briefly the results for $\ell = \ell 4$ and $\ell 12$, as given in the last two blocks of Table 1. The asymptotics for both tests take hold more slowly here than when $\ell = 0$, so that the finite sample size distortions are larger. However, the comparisons between the KPSS and MR/S tests are still essentially as described above.

The power of the tests is also of importance, of course. Basically we would like to know whether the greater robustness of the KPSS test is purchased with a sacrifice in power. Since the MR/S test underrejects under the null, its apparent power (using standard critical values) will be very low. Correspondingly we will consider size-adjusted power; that is, power using the critical values that would have yielded correct size under the null, for the given sample size and distribution. As usual, this is an attempt to quantify the intrinsic ability of the statistic to distinguish the null and the alternative, rather than to investigate the performance of any feasible procedure. We consider power against unit root alternatives only, although fractional alternatives would presumably also be of interest. Our parameterization of the alternative is a slight generalization of the data generating process in KPSS (1992). They considered the representation of the series: $X_t = \xi t + r_t + \varepsilon_t$, $r_t = r_{t-1} + u_t$. Here the ε_t were taken to be iid $(0, \sigma_\varepsilon^2)$, the u_t were iid $(0, \sigma_u^2)$, and the ε and u processes were independent. The null is $\sigma_u^2 = 0$, and the alternative was indexed by the parameter $\lambda = \sigma_u^2/\sigma_\varepsilon^2$. We consider the no-trend case, so that $\xi =$

0. Also, in order to be able to accommodate cases of infinite variance (t_2 and Cauchy), we make the following slight modification to this parameterization. We generate the series as follows:

(2) $$X_t = \lambda^{½} r_t + \varepsilon_t \ , \ \ r_t = r_{t-1} + u_t \ ,$$

where the ε_t and u_t are both iid and drawn from the same distribution (Normal, *t*, or Cauchy). This agrees with the KPSS parameterization for the finite variance cases and generalizes it in a natural way for the infinite variance cases.

We consider T = 50, 100, 200, 500 and 1000, and λ = 0.001, 0.01, 0.1 and 1.0. These values of T and λ combine to give a broad range of size-adjusted power. We will report our results only for the tests with $\ell = 0$, but the results for $\ell 4$ and $\ell 12$ are similar.

Table 2 gives the results (size adjusted power) for those combinations of T and λ where power is not essentially equal to one. The basic result is that (for a given T, λ and distribution) the size adjusted powers of the KPSS and MR/S tests are quite similar. It appears that the KPSS test is somewhat more powerful than the MR/S test when power is low, while the opposite is true when power is high, but these differences between the tests are outweighed by the similarities between them. For the finite variance distributions, the size adjusted power of each test does not depend much on the distribution. There are some intriguing results for the t_2 and Cauchy cases. For example, for the Cauchy case the power function is less steep (in the λ direction) than in the finite variance cases. However, these results are unrelated to our main interest and we will not pursue them. Our basic conclusion from Table 2 is that the advantage in robustness of the KPSS test relative to the MR/S test does not come at the cost of reduced power.

4. ASYMPTOTIC THEORY: INFINITE VARIANCE CASE

In this section we assume that the X_t are iid and follow a standard symmetric stable law with index α, with $0 < \alpha \leq 2$. We give a brief summary of the asymptotic distribution theory provided by Chan and Tran (1989) and Phillips (1990), and we use it to derive the asymptotic distributions of the KPSS and MR/S tests. This is a prelude to the local to finite variance asymptotics of the next section. We note that, following Chan and Tran (1989) and Phillips (1990), we really need to assume only that the X_t are in the normal domain of attraction of a stable law, but for our purposes this direction of generalization is not very important.

Under the assumptions above, the cumulations of X_t and X_t^2 follow a Levy process. More specifically, define:

(3) $$[U_T(r), V_T(r)] = [T^{-1/\alpha} \Sigma_{j=1}^{[rT]} X_j \, , \, T^{-2/\alpha} \Sigma_{j=1}^{[rT]} X_j^2] \, .$$

Then we have the convergence result: $[U_T(r), V_T(r)] \Rightarrow [U^{(\alpha)}(r), V^{(\alpha)}(r)]$, where $V^{(\alpha)}(r) = \int_0^r (dU^{(\alpha)}(s))^2 \, ds$, and where $U^{(\alpha)}(r)$ is a standard stable process (Levy process) with index α. See, e.g., Phillips (1990, p. 46). Note that the first component of the convergence result (3) asserts, in our previous notation, that $T^{-1/\alpha} S_{[rT]} \Rightarrow U^{(\alpha)}(r)$. Also, for the usual variance estimate, we have $T^{1-2/\alpha} s^2 = T^{-2/\alpha} \Sigma_t X_t^2 \Rightarrow V^{(\alpha)}(1)$. This yields the following limiting results for the KPSS and MR/S statistics, with $\ell = 0$:

(4A) $$\hat{\eta}(0) = T^{-1-2/\alpha} \Sigma_t S_t^2 \, / \, T^{-2/\alpha} \Sigma_t X_t^2 \Rightarrow \int_0^1 U^{(\alpha)}(r)^2 \, dr \, / \, V^{(\alpha)}(1)$$

(4B) $$T^{-1/2} Q(0) = [\max_t T^{-1/\alpha} S_t - \min_t T^{-1/\alpha} S_t] \, / \, [T^{-2/\alpha} \Sigma_t X_t^2]^{1/2}$$

$$\Rightarrow [\max_r U^{(\alpha)}(r) - \min_r U^{(\alpha)}(r)] \, / \, [V^{(\alpha)}(1)]^{1/2}$$

These results can be generalized easily to the case that the data are demeaned or demeaned and detrended. In the case that the data are demeaned, we replace $U^{(\alpha)}(r)$ by $U_{\mu}^{(\alpha)}(r) = U^{(\alpha)}(r) - rU^{(r)}(1)$, the analog of a Brownian bridge, and similarly for $V^{(\alpha)}(r)$. If the data are demeaned and detrended, we replace $U^{(\alpha)}(r)$ by $U_{\tau}^{(\alpha)}(r) = U^{(\alpha)}(r) - (2r - 3r^2)U^{(\alpha)}(1) + (-6r + 6r^2)\int_0^1 U^{(\alpha)}(s)ds$, the analog of a second-level Brownian bridge, and similarly for $V^{(\alpha)}(r)$. See Phillips (1990, p. 55) for further discussion.

The generalization of this analysis to the case of data with short run dynamics, and hence to the $\hat{\eta}(\ell)$ and $Q(\ell)$ statistics where $\ell > 0$ and ℓ grows with T, is more challenging. Phillips (1990, section 2.2) treats the case that X_t is a linear process: $X_t = d(L)_{\varepsilon t} = \Sigma_{j=0}^{\ell}\, \varepsilon_j$. Similarly to before, we suppose that the ε_t are iid symmetric stable with index α, which is stronger than Phillips' assumption. Define $\omega = d(1)$ and $\sigma^2 = \Sigma_{j=1}^{\infty} d_j^2$. Then the convergence result for the cumulations of X_t and X_t^2 is modified slightly: $[U_T(r), V_T(r)] \Rightarrow [\omega U^{(\alpha)}(r), \sigma^2 V^{(\alpha)}(r)]$. Furthermore, the usual long run variance estimate $s^2(\ell)$, which Phillips calls $\hat{\omega}^2$, converges to a random limit: $T^{1-2/\alpha}\, s^2(\ell) \Rightarrow \omega^2 V(1)$. Phillips establishes this result under the assumption that $\ell \to \infty$ as $T \to \infty$, but $\ell = o(T^{½})$. Then it is easy to see that the scale factor involving ω cancels from the limiting distribution for the $\hat{\eta}(\ell)$ and $Q(\ell)$ statistics, so that $\hat{\eta}(\ell)$ has the same asymptotic distribution as was given above in equation (4A) for $\hat{\eta}(0)$, and similarly $T^{-½}Q(\ell)$ has the same asymptotic distribution as was given in equation (4B) for $T^{½}Q(0)$.

In this paper we are interested in thick tails, not short run dependence. However, the results of the previous paragraph are important because we do not want our results to be specific to the case of $\ell = 0$.

5. ASYMPTOTIC THEORY: LOCAL TO FINITE VARIANCE CASE

In this section we consider *local to finite variance* asymptotics. These are based on the following representation of the series X_t:

(5) $$X_t = X_{1t} + (c/T^{(1/\alpha - ½)})\, X_{2t} \,.$$

Here X_{1t} is iid with zero mean and variance one, say, while X_{2t} is iid symmetric stable with index $\alpha < 2$. The quantity "c" is a constant, at this point arbitrary. In the calculations reported below we will consider the case that X_{1t} is standard normal and X_{2t} is Cauchy ($\alpha = 1$), so that the representation of the series is $X_t = X_{1t} + (c/T^{½})X_{2t}$. We call this a local to finite variance representation because X_t has infinite variance for all t, but loosely speaking it is close to having finite variance, and it converges to a finite variance process as $T \rightarrow \infty$. The order in probability of the cumulations of the stable process X_{2t} is larger than that of the cumulations of the finite variance process X_{1t}, and the power of T in the representation (5) is chosen so that the cumulation of the series X_t, suitably normalized, will converge weakly to a stochastic process that is the weighted sum of Wiener and Levy process components.

The point of this construction is to be able to use asymptotic theory to analyze cases that are near the borderline between finite and infinite variance. For example, the t_5 distribution has a finite variance, but thick tails. However, first order asymptotics do not lead to any results that are useful in analyzing the robustness of statistics to errors that are t_5. So, instead of assuming t_5, we postulate the process (5) with a small infinite variance component. Our asymptotics will depend on the quantity c, which measures the size of the infinite variance component, and we can then use asymptotics to calculate how robust various statistics are to such as component. This kind of local asymptotics has been used successfully elsewhere. For example, a number of authors, including Elliott, Rothenberg and Stock (1996), have used local to unit root asymptotics to analyze statistical procedures when a time series is stationary but has a root near to unity, by modelling the root as local to unity. As another example, Staiger and Stock (1997) analyze the weak instrument problem in instrumental variables regression by postulating instruments that are informative, but local to useless. In these two examples, as in our case, the "local to" construction can lead to interesting and potentially useful asymptotics because the rate of convergence of a statistic of interest is of different orders under the two circumstances being considered.

We now proceed with our asymptotics, under the strong assumptions given above; namely, both processes X_{1t} and X_{2t} are iid, and the two processes are independent. As before we define S_t as the cumulation of the X_t: $S_t = \Sigma_{j=1}^{t} X_j$. Then the following weak convergence result holds:

(6) $$T^{-½} S_{[rT]} = T^{-½} \Sigma_{j=1}^{[rT]} X_{1j} + cT^{-1/\alpha} \Sigma_{j=1}^{[rT]} X_{2j} \Rightarrow W(r) + cU^{(\alpha)}(r) \equiv G_c^{(\alpha)}(r).$$

(The independence of the two terms guarantees that the convergence of each term separately leads to a well defined limit for the sum.) This result will ensure weak convergence of the numerators of the KPSS and MR/S statistics. Specifically,

$$(7) \qquad T^{-2} \Sigma_t S_t^2 \Rightarrow \int_0^1 G_c^{(\alpha)}(r)^2 dr \; , \; T^{-½} [\max_t S_t - \min_t S_t] \Rightarrow \max_r G_c^{(\alpha)}(r) - \min_r G_c^{(\alpha)}(r)$$

We also need to consider the denominator of the statistics. We have

$$(8) \qquad T^{-1} \Sigma_t X_t^2 = T^{-1} \Sigma_t X_{1t}^2 + c^2 T^{-2/\alpha} \Sigma_t X_{2t}^2 + 2T^{-(½+1/\alpha)} \Sigma_t X_{1t} X_{2t} \; .$$

The first term on the right hand side of (8) converges in probability to one (the assumed variance of X_{1t}). The second term converges weakly to $c^2 V^{(\alpha)}(1)$, using results in the previous section. We show in the Appendix that the third term converges in probability to zero. Thus we have the result:

$$(9) \qquad T^{-1} \Sigma_t X_t^2 \Rightarrow 1 + c^2 V(1) \; .$$

Combining the results in equations (7) and (9), we obtain the asymptotic distributions of the KPSS and MR/S statistics, with $\ell = 0$.

$$(10A) \qquad \hat{\eta}(0) = T^{½} \Sigma_t S_t^2 / T^{-1} \Sigma_t X_t^2 \Rightarrow \int_0^1 G_c^{(\alpha)}(r)^2 dr / [1 + c^2 V(1)]$$

$$(10B) \qquad T^{-½} Q(0) = T^{-½} [\max_t S_t - \min_t S_t] / [T^{-1} \Sigma_t X_t^2]^{½}$$

$$\Rightarrow [\max_r G_c^{(\alpha)}(r) - \min_r G_c^{(\alpha)}(r)] / [1 + c^2 V(1)]^{½}$$

If the series are demeaned or demeaned and detrended, these results are easily modified. In the demeaned case, we replace $G_c^{(\alpha)}(r)$ by $G_{c,\mu}^{(\alpha)}(r) = W_\mu(r) - cU_\mu^{(\alpha)}(r) = G_c^{(\alpha)}(r) - rG_c^{(\alpha)}(1)$.

Similarly, in the demeaned and detrended case, we replace $G_c^{(\alpha)}(r)$ by $G_{c,\tau}^{(\alpha)}(r) = W_\tau(r) - cU_\tau^{(\alpha)}(r)$ $= G_c^{(\alpha)}(r) - (2r - 3r^2)G_c^{(\alpha)}(1) + (-6r + 6r^2)\int_0^1 G_c^{(\alpha)}(s)ds$.

The generalization of this analysis to the case of data with short run dynamics, and hence to the $\hat{\eta}(\ell)$ and $Q(\ell)$ statistics where $\ell > 0$ and ℓ grows with T, is more challenging. Some incomplete but perhaps still informative results on this case are given in the Appendix.

6. ROBUSTNESS: SENSITIVITY OF THE ASYMPTOTIC DISTRIBUTIONS TO "c"

We now return to the question of the relative robustness of the KPSS and MR/S tests to thick tails. We use the local to finite variance asymptotics in a conceptually simple way. The asymptotic distribution of each of the statistics depends on the quantity "c" that determines the relative weight of the infinite variance component in the representation (5). We will say that one statistic is more robust than another if its asymptotic distribution is less sensitive to the value of c.

Before we make such comparisons, we first present the results of some simulations that investigate the finite sample accuracy of the local to finite variance asymptotics. In these simulations we consider only the statistics with mean correction and with $\ell = 0$ ($\hat{\eta}_\mu(0)$ and $Q_\mu(0)$), and we consider the case that X_{1t} is standard normal while X_{2t} is Cauchy ($\alpha = 1$). We consider T = 100, 1000 and 10000, and c = 0, .1, .316, 1, 3.16, 10 and 31.6. (Here 3.16 is shorthand for $\sqrt{10}$ etc.) Values of c larger than 31.6 had only very minimal effects on the results; we have essentially reached the pure Cauchy case. The number of replications is 20,000.

The results of these simulations are given in Table 3, where we report the size (frequency of rejection) of the 5% upper tail tests. For the KPSS test, the asymptotics are quite reliable, in the sense that, holding c constant, changing T from 100 to 1000 to 10000 did not have much effect on the results. For the MR/S test, the asymptotics are not very reliable for T = 100 (holding c constant, changing T from 100 to 1000 did affect the results, at times considerably) but they seemed fairly reliable for T ℓ 1000.

We note in passing that other types of asymptotics can be found in Table 3. As noted above, as $c \to \infty$ we approach the pure Cauchy case. Another possibility is the following. Our data generating process is: $X_t = X_{1t} + bX_{2t}$, with $b = c/T^{½}$. The local to finite variance asymptotics hold "c" constant, but we could also consider asymptotics that hold "b" constant as T increases. In this case the Cauchy limit theory applies. In fact, the parameter values in Table 3 have been chosen so that "constant b" asymptotics can also be found there. For example, we have b = 3.16 as follows: T = 100, c = .316; T = 1000, c = 1; T = 10000, c = 3.16. That is, constant values of b are found by moving diagonally from northeast to southwest. But, as we would expect, the convergence is much slower with b constant than for the local asymptotics (c constant).

We now return to the main task of assessing the sensitivity of the asymptotic distributions of the KPSS and MR/S statistics to the value of c. This is already apparent in Table 3 for one feature of these distributions, the probability of being above the standard upper tail critical value; the rejection frequencies in Table 3 change much less, as c changes, for KPSS than for MR/S. However, it is worth looking at more features of the distributions than just this one.

It would be nice to be able to make analytical comparisons of the local to finite variance asymptotic distributions, but we have not been able to do so. As usual, the alternative is to resort to simulation. We have evaluated the distributions of the KPSS and MR/S statistics as described above, for T = 10000, which we believe is large enough to be considered asymptotic, based on the evidence in Table 3. We use 20000 replications and various values of c, ranging from 0 to 31.6 (large enough to be essentially infinite). These are the same simulations reported for T = 10000 in Table 3, but now we provide additional detail on the distributions. In particular we provide all of the deciles of the distributions, as well as the 0.001, 0.025, 0.050, 0.950, 0.975 and 0.999 quantiles.

For the KPSS statistic, the shifts in the asymptotic distribution as c changes are relatively minor. Most of the distribution shifts slightly right as c increases. For example, the median is 0.120 for c = 0 and rises to 0.133 for c = 31.6. Most of the other quantiles show a similar pattern. However, the upper tail quantiles are the exception; they *decrease* as c increases. For example, the 0.950 quantile decreases from 0.476 to 0.395 as c moves from zero to 31.6. This explains the smaller number of rejections of the null when c increases, since we are using an upper tail test.

For the MR/S statistic, the asymptotic distribution shifts left as c increases, and this is true across the whole distribution. The shift is largest in the upper tail. For example, as we move from c = 0 to c = 31.6, the median changes from 1.216 to 0.999, while the 0.950 quantile changes from 1.739 to 1.361. These changes are larger for the MR/S statistic than they were for the KPSS statistic, and this at least partially explains the lower degree of robustness of the MR/S test to local deviations from finite variance. However, the thickness of the tails of the distributions of the test statistics under thick tailed errors are also relevant. For example, the change in the 0.95 quantile as we go from normal to Cauchy errors is larger but not terribly larger in percentage terms for the MR/S statistic compared to the KPSS statistic (1.739 becomes 1.361 for MR/S, while 0.476 becomes 0.395 for KPSS). But the implication of this change for rejections under the null is very different, because the longer tail of KPSS (compared to MR/S) in the Cauchy case puts a much higher fraction of observations above the standard critical value.

7. CONCLUDING REMARKS

In this paper we have compared the sensitivity to thick tailed errors of two tests of the null of short memory, namely the KPSS test of Kwiatkowski *et al.* (1992) and the MR/S test of Lo (1991). Our intuition suggested that the MR/S test might not be very robust, because it is based on the maximum and minimum of the cumulated series, and maxima and minima are potentially very sensitive to tail thickness.

We conclude that our intuition was correct. The KPSS test appears to be more robust to thick tails than the MR/S test. This conclusion is based on two types of evidence. First, we did simulations using moderately thick tailed distributions, like student *t* with five degrees of freedom, and found that the actual size was closer to nominal size in finite samples for KPSS than for MR/S. (The size adjusted powers of the two tests were comparable.) Second, we considered local to finite variance asymptotics, in which the series was assumed to be the sum of a finite variance component and a symmetric stable component, with the weight on the stable component going to zero in the limit, but at a rate that ensures a limiting distribution that depends on a weighting parameter (c). We then provided evidence that the asymptotic distribution of the KPSS test is less sensitive to this weighting parameter than the asymptotic distribution of the MR/S test is.

We would like to extend this analysis to other tests, and in particular to be able to find tests that are more robust than either KPSS or MR/S. Here we can mention three possible ways that we might proceed. First, we can simply write down other statistics, formed by taking functions of the cumulations (S_t) other than sum of squares or maximum minus minimum. For example, the sum of the absolute cumulations might be considered, based on a simple analogy to robustness results for linear regression. The question here would appear to be not so much whether we can find more robust tests, but whether this can be done without substantial loss of power. Second, we could proceed more formally by assuming a data generating process like that in KPSS, but where the distribution of the series is as in (5) above, say a weighted sum of normal and Cauchy. Then in principle we could use standard principles for constructing tests. For example, we might consider the LM test, or the Neyman-Pearson likelihood ratio test, etc. These tests would depend on a value of c, specified to reflect an empirically relevant tail thickness. This line of attack requires the derivation of the likelihood, which we have so far not been able to write down in a useful closed form. A third possibility, perhaps more speculative than the first two, is to try to find tests that are completely robust to local deviations from finite variance, that is, whose asymptotic distribution under the local to finite variance assumption in (5) does not depend on the value of c. Tests with this robustness property would have to depend on some feature of the data other than cumulations, and it remains to be seen whether they can be found, but their appeal should be evident.

REFERENCES

Caner, M. (1997), "Weak Convergence to a Matrix Stochastic Integral with Stable Processes," *Econometric Theory*, 13, 506-528.

Chan, N.H. and L.T. Tran (1989), "On the First Order Autoregressive Process with Infinite Variance," *Econometric Theory*, 5, 354-362.

Elliott, G., T.J. Rothenberg and J.H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813-836.

Knight, K. (1991), "Limit Theory for M-Estimates in an Integrated Infinite Variance Process," *Econometric Theory*, 7, 200-212.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Contain a Unit Root?," *Journal of Econometrics*, 54, 159-178.

Lee, D. and P. Schmidt (1996), "On the Power of the KPSS Test of Stationarity against Fractionally-Integrated Alternatives," *Journal of Econometrics*, 73, 285-302.

Leybourne, S.J. and B.P.M. McCabe (1994), "A Consistent Test for a Unit Root," *Journal of Business and Economic Statistics*, 12, 157-166.

Leybourne, S.J. and B.P.M. McCabe (1999), "Modified Stationarity Tests with Data-Dependent Model-Selection Rules," *Journal of Business and Economic Statistics*, 17, 264-270.

Lo, A. (1991), "Long-Term Memory in Stock Market Prices," *Econometrica*, 59, 1279-1313.

Phillips, P.C.B. (1987), "Time Series Regression with a Unit Root," *Econometrica*, 55, 277-301

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

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1986), *Numerical Recipes: The Art of Scientific Computing*, Cambridge: Cambridge University Press.

Saikkonen, P. and R. Luukkonen (1993), "Testing for a Moving Average Unit Root in Autoregressive Integrated Moving Average Models," *Journal of the American Statistical Association*, 88, 596-601.

Schwert, G.W. (1989), "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147-159.

Staiger, D. and J.H. Stock (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.

APPENDIX

A.1 Proof that $T^{-(½+1/\alpha)} \Sigma_t X_{1t}X_{2t} \xrightarrow{p}$

As a matter of notation, let

(A1) $Z_T = T^{-(½+1/\alpha)} \Sigma_t X_{1t}X_{2t}$, $P_{T,\delta} = P(|Z_T| > \delta)$,

for any $\delta > 0$. Then we prove that $Z_T \xrightarrow{p} 0$ by showing that $P_{T,\delta} \to 0$ for any $\delta > 0$.

Define $X_{2\bullet} = [X_{21},...,X_{2T}]$ and $P^c_{T,\delta} = P(|Z_T| > \delta \mid X_{2\bullet})$, so that $P_{T,\delta} = E\, P^c_{T,\delta}$. Now define $Q_T = E(Z_T^2 \mid X_{2\bullet})$. Then we have

(A2) $0 \le P^c_{T,\delta} \le \min(1 , \delta^{-2}Q_T)$,

where $0 \le P^c_{T,\delta} \le 1$ is evident and $P^c_{T,\delta} \le \delta^{-2}Q_T$ is the Chebyshev inequality. But now we calculate Q_T as follows:

(A3) $Q_T = \delta^{-2}\, T^{-1-2/\alpha}\, E[(\Sigma_t X_{1t}X_{2t})^2 \mid X_{2\bullet}]$

$= \delta^{-2}\, T^{-1-2/\alpha}\, \Sigma_t X_{2t}^2$.

Here we have used the fact that the X_{1t} are iid with $E\, X_{1t}^2 = 1$, and are independent of $X_{2\bullet}$.

Furthermore, since $T^{-2/\alpha} \Sigma_t X_{2t}^2 \Rightarrow V^{(\alpha)}(1)$, $Q_T \xrightarrow{p} 0$.

From (A2), the fact that $Q_T \xrightarrow{p} 0$ implies that $P^c_{T,\delta} \xrightarrow{p} 0$, for any δ. Thus, for any δ, from (A2) we see that $P^c_{T,\delta}$ is a uniformly bounded sequence that converges in probability to zero. By the dominated convergence theorem, its expectation converges to zero; that is, $P_{T,\delta} = E\, P^c_{T,\delta} \to 0$.

Therefore $Z_T \xrightarrow{p} 0$.

A.2 Long Run Variance Estimation

We will treat the zero mean case; the other cases are essentially the same. The long run variance estimate is $s^2(\ell) = T^{-1} \Sigma_t X_t^2 + 2\ \Sigma_{j=1}^{\ell} w(j,\ell)\, T^{-1} \Sigma_t X_t X_{t-j}$. Since $X_t = X_{1t} + cT^{-1/\alpha-½}X_{2t}$, simple algebra yields:

(A4) $$s^2(\ell) = s_1^2(\ell) + c^2 s_2^2(\ell) + c\bullet cpt,$$

(*cpt* stands for cross product terms), where

(A5) $$s_1^2(\ell) = T^{-1} \Sigma_t X_{1t}^2 + 2\ \Sigma_{j=1}^{\ell}\ w(j,\ell)\, T^{-1} \Sigma_t X_{1t}X_{1,t-j}$$

(A6) $$s_2^2(\ell) = T^{-2/\alpha} \Sigma_t X_{2t}^2 + 2\ \Sigma_{j=1}^{\ell} w(j,\ell)\, T^{-2/\alpha} \Sigma_t X_{2t}X_{2,t-j}$$

(A7) $$cpt = T^{-1/\alpha-½} \Sigma_t X_{1t}X_{2t} + 2\ \Sigma_{j=1}^{\ell} w(j,\ell)\, T^{-1/\alpha-½} \Sigma_t X_{1t}X_{2,t-j}$$

$$+ T^{-1/\alpha-½} \Sigma_t X_{1t}X_{2t} + 2\ \Sigma_{j=1}^{\ell} w(j,\ell)\, T^{-1/\alpha-½} \Sigma_t X_{2t}X_{1,t-j}\ .$$

In order to say more about these terms, we have to be more specific about the data generating process. Specifically, to have any hope that we will obtain a limiting distribution free of long run variance parameters, we need to assume that X_{1t} and X_{2t} have the same long run variance parameter. To do so, we assume:

(A8) $$X_t = d(L)u_t\ ,\ u_t = u_{1t} + cT^{-1/\alpha-½}u_{2t}\ ,$$

where u_{1t} is iid with mean zero and variance one, and u_{2t} is iid as symmetric stable with parameter α. That is, u_{1t} and u_{2t} satisfy the assumptions that X_{1t} and X_{2t} did previously. Now $X_{1t} = d(L)u_{1t}$ and $X_{2t} = d(L)u_{2t}$. We assume that d(L) satisfies the regularity conditions given by Phillips (1990, p. 50), so that $d(1) \neq 0$ and a summability restriction is obeyed.

As in section 4, we define $\omega = d(1)$. Then, if $\ell \to \infty$ but $\ell = o(T^{½})$, it follows from standard results that $s_1^2(\ell) \xrightarrow{p} \omega^2$, and it follows from Phillips (1990, p. 53, equation (41)) that

$s_2^2(\ell) \Rightarrow \omega^2 V^{(\alpha)}(1)$. Thus we will obtain $s^2(\ell) \Rightarrow 1 + c^2 V^{(\alpha)}(1)$, the same result as in the case of iid X_t and $\ell = 0$, provided that *cpt* $\overset{p}{\rightarrow}$ 0. We now ask under what circumstances this will be so.

We can distinguish two cases. In the first case, $d(L) \equiv 1$. That is, the X_t are iid, and long run variance estimation is being performed needlessly. The expression for *cpt* in (A7) above contains $2\ell+1$ terms, corresponding to different leads or lags of X_2 relative to X_1. We can show that each term individually has probability limit equal to zero, using the essentially the same proof as in section A.1 above. However, since ℓ grows with T, this does not imply that the sum of the $2\ell+1$ terms has probability limit zero. In the second case, we actually have short run dynamics, and d(L) simply has to satisfy some regularity conditions, as above. We do not yet have a satisfactory proof that *cpt* $\overset{p}{\rightarrow}$ 0, in either of these two cases, so for now we will simply leave that as a conjecture.

TABLE 1

SIZE OF 5% UPPER TAIL KPSS AND MR/S TESTS

	KPSS $\hat{\eta}_\mu(0)$				MR/S $Q_\mu(0)$				
T	**N(0,1)**	t_{10}	t_5	t_2	**Cauchy**	**N(0,1)**	t_{10}	t_5	t_2	**Cauchy**
50	.049	.048	.052	.037	.024	.010	.011	.010	.003	.001
100	.047	.052	.043	.043	.027	.021	.022	.019	.008	.000
200	.050	.045	.050	.042	.028	.028	.027	.030	.011	.000
500	.051	.047	.048	.039	.027	.038	.038	.032	.016	.001
1000	.046	.049	.047	.048	.028	.041	.037	.032	.018	.001
2000	.050	.052	.052	.049	.025	.048	.040	.038	.023	.001

	KPSS $\hat{\eta}_\mu(\ell 4)$					MR/S $Q_\mu(\ell 4)$				
T	**N(0,1)**	t_{10}	t_5	t_2	**Cauchy**	**N(0,1)**	t_{10}	t_5	t_2	**Cauchy**
50	.035	.039	.040	.030	.020	.000	.000	.001	.000	.000
100	.040	.046	.041	.037	.025	.004	.005	.003	.002	.000
200	.048	.042	.044	.042	.027	.017	.018	.018	.006	.000
500	.048	.046	.045	.038	.026	.029	.030	.030	.012	.001
1000	.047	.047	.048	.046	.026	.036	.033	.030	.018	.001
2000	.052	.050	.051	.049	.024	.044	.036	.038	.019	.001

	KPSS $\hat{\eta}_\mu(\ell 12)$				MR/S $Q_\mu(\ell 12)$				
T	**N(0,1)**	**t_{10}**	**t_5**	**t_2**	**Cauchy**	**N(0,1)**	**t_{10}**	**t_5**	**t_2**	**Cauchy**
50	.012	.014	.012	.009	.005	.007	.006	.007	.001	.001
100	.032	.031	.025	.026	.017	.002	.001	.002	.002	.002
200	.042	.037	.038	.034	.024	.005	.003	.003	.002	.001
500	.042	.043	.043	.036	.024	.017	.018	.016	.006	.001
1000	.046	.047	.045	.043	.025	.028	.027	.025	.013	.001
2000	.050	.052	.052	.046	.025	.038	.033	.030	.015	.001

TABLE 2

SIZE-ADJUSTED POWER OF 5% UPPER TAIL $\hat{\eta}_\mu(0)$ AND $Q_\mu(0)$ TESTS

$\lambda = 0.001$

	KPSS $\hat{\eta}_\mu(0)$					MR/S $Q_\mu(0)$				
T	**N(0,1)**	**t_{10}**	**t_5**	**t_2**	**Cauchy**	**N(0,1)**	**t_{10}**	**t_5**	**t_2**	**Cauchy**
50	.083	.078	.075	.119	.245	.068	.058	.061	.105	.242
100	.173	.169	.176	.216	.349	.137	.128	.130	.209	.353
200	.410	.408	.417	.442	.494	.371	.356	.380	.433	.522
500	.791	.796	.782	.778	.722	.827	.835	.829	.811	.735
1000	.959	.957	.956	.935	.833	.980	.982	.979	.961	.841

$\lambda = 0.01$

	KPSS $\hat{\eta}_\mu(0)$					MR/S $Q_\mu(0)$				
T	**N(0,1)**	**t_{10}**	**t_5**	**t_2**	**Cauchy**	**N(0,1)**	**t_{10}**	**t_5**	**t_2**	**Cauchy**
50	.307	.297	.277	.365	.447	.238	.238	.244	.342	.457
100	.603	.590	.590	.595	.592	.614	.580	.601	.630	.611
200	.850	.859	.845	.823	.758	.894	.893	.894	.868	.780
500	.989	.989	.988	.973	.985	.998	.998	.996	.984	.987
1000	1.00	1.00	.999	.995	.940	1.00	1.00	1.00	.997	.947

$\lambda = 0.1$

	KPSS $\hat{\eta}_\mu(0)$					MR/S $Q_\mu(0)$				
T	**N(0,1)**	**t_{10}**	**t_5**	**t_2**	**Cauchy**	**N(0,1)**	**t_{10}**	**t_5**	**t_2**	**Cauchy**
50	.732	.722	.712	.727	.684	.766	.764	.764	.765	.713
100	.928	.927	.918	.908	.817	.966	.961	.957	.940	.838
200	.991	.990	.991	.977	.910	.998	.998	.998	.988	.926
500	1.00	1.00	1.00	.998	.961	1.00	1.00	1.00	.999	.967

$\lambda = 1.0$

	KPSS $\hat{\eta}_\mu(0)$					MR/S $Q_\mu(0)$				
T	N(0,1)	t_{10}	t_5	t_2	Cauchy	N(0,1)	t_{10}	t_5	t_2	Cauchy
50	.932	.933	.920	.917	.860	.969	.969	.965	.963	.884
100	.990	.989	.989	.986	.931	.998	.999	.998	.995	.946
200	.999	.999	.999	.998	.970	1.00	1.00	1.00	.999	.975

TABLE 3

SIZE OF 5% UPPER TAIL KPSS AND MR/S TESTS

LOCAL TO FINITE VARIANCE DATA

T = 100

c:	31.6	10	3.16	1	.316	.1	0
KPSS $\hat{\eta}_\mu(0)$	.029	.029	.030	.038	.044	.048	.049
MR/S $Q_\mu(0)$	.001	.001	.002	.009	.016	.019	.020

T = 1000

c:	31.6	10	3.16	1	.316	.1	0
KPSS $\hat{\eta}_\mu(0)$	.026	.026	.030	.036	.043	.047	.050
MR/S $Q_\mu(0)$	.001	.002	.005	.017	.031	.037	.042

T = 10000

c:	31.6	10	3.16	1	.316	.1	0
KPSS $\hat{\eta}_\mu(0)$	.029	.028	.030	.037	.046	.052	.054
MR/S $Q_\mu(0)$	.001	.002	.006	.019	.035	.044	.049

TABLE 4

QUANTILES OF KPSS AND MR/S STATISTICS

LOCAL TO FINITE VARIANCE ASYMPTOTIC DISTRIBUTIONS

KPSS $\hat{\eta}_{\mu}(0)$

Quantile	c: 31.6	10	3.16	1	.316	.1	0
.010	.027	.027	.027	.026	.026	.025	.025
.025	.035	.035	.034	.033	.031	.031	.030
.050	.044	.044	.042	.040	.038	.037	.036
.100	.056	.056	.055	.051	.048	.046	.046
.200	.076	.075	.074	.070	.065	.063	.062
.300	.092	.092	.090	.087	.083	.080	.079
.400	.111	.110	.108	.105	.101	.099	.098
.500	.133	.133	.131	.127	.124	.121	.120
.600	.162	.160	.160	.156	.152	.149	.148
.700	.196	.196	.196	.193	.189	.188	.186
.800	.246	.245	.244	.245	.245	.245	.244
.900	.320	.321	.322	.329	.339	.350	.354
.950	.395	.398	.402	.418	.449	.471	.476
.975	.475	.475	.457	.523	.567	.585	.598
.999	.589	.582	.612	.654	.733	.748	.748

LO $T^{-1/2}Q_{\mu}(0)$

Quantile	c: 31.6	10	3.16	1	.316	.1	0
.010	.715	.720	.744	.758	.754	.751	.747
.025	.761	.768	.790	.809	.804	.803	.799
.050	.803	.810	.829	.856	.861	.854	.855
.100	.853	.860	.882	.911	.923	.919	.919
.200	.917	.923	.945	.976	.994	1.000	1.009
.300	.958	.963	.981	.999	1.038	1.063	1.082
.400	.985	.988	.998	1.036	1.099	1.131	1.149
.500	.999	.999	1.020	1.091	1.163	1.197	1.216
.600	1.028	1.036	1.071	1.154	1.233	1.269	1.288
.700	1.085	1.095	1.135	1.230	1.313	1.349	1.367
.800	1.162	1.171	1.217	1.325	1.411	1.449	1.468
.900	1.267	1.280	1.342	1.465	1.557	1.595	1.616
.950	1.361	1.381	1.450	1.591	1.681	1.724	1.739
.975	1.445	1.463	1.550	1.702	1.802	1.841	1.854
.999	1.537	1.565	1.668	1.844	1.934	1.973	2.000