HAC Corrections for Strongly Autocorrelated Time Series

Ulrich K. Müller
Department of Economics, Princeton University, Princeton, NJ 08544 (*umueller@princeton.edu*)

Applied work routinely relies on heteroscedasticity and autocorrelation consistent (HAC) standard errors when conducting inference in a time series setting. As is well known, however, these corrections perform poorly in small samples under pronounced autocorrelations. In this article, I first provide a review of popular methods to clarify the reasons for this failure. I then derive inference that remains valid under a specific form of strong dependence. In particular, I assume that the long-run properties can be approximated by a stationary Gaussian AR(1) model, with coefficient arbitrarily close to one. In this setting, I derive tests that come close to maximizing a weighted average power criterion. Small sample simulations show these tests to perform well, also in a regression context.

KEYWORDS: AR(1); Local-to-unity; Long-run variance.

1. INTRODUCTION

A standard problem in time series econometrics is the derivation of appropriate corrections to standard errors when conducting inference with autocorrelated data. Classical references include Berk (1974), Newey and West (1987), and Andrews (1991), among many others. These articles show how one may estimate "heteroscedasticity and autocorrelation consistent" (HAC) standard errors, or "long-run variances" (LRV) in econometric jargon, in a large variety of circumstances.

Unfortunately, small sample simulations show that these corrections do not perform particularly well as soon as the underlying series displays pronounced autocorrelations. One potential reason is that these classical approaches ignore the sampling uncertainty of the LRV estimator (i.e., t- and F-tests based on these corrections employ the usual critical values derived from the normal and chi-squared distributions, as if the true LRV was plugged in). While this is justified asymptotically under the assumptions of these articles, it might not yield accurate approximations in small samples.

A more recent literature, initiated by Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2002a, 2005), seeks to improve the performance of these procedures by explicitly taking the sampling variability of the LRV estimator into account. (Also see Jansson 2004; Müller 2004, 2007; Phillips 2005; Phillips, Sun, and Jin 2006, 2007; Sun, Phillips, and Jin 2008; Gonalves and Vogelsang 2011; Atchade and Cattaneo 2012; Sun and Kaplan 2012.) This is accomplished by considering the limiting behavior of LRV estimators under the assumption that the bandwidth is a fixed fraction of the sample size. Under such "fixed-b" asymptotics, the LRV estimator is no longer consistent, but instead converges to a random matrix. The resulting t- and F-statistics have limiting distributions that are nonstandard, with randomness stemming from both the parameter estimator and the estimator of its variance. In these limiting distributions, the true LRV acts as a common scale factor that cancels, so that appropriate nonstandard critical values can be tabulated.

While this approach leads to relatively better size control in small sample simulations, it still remains the case that strong underlying autocorrelations lead to severely oversized tests. This might be expected, as the derivation of the limiting distributions under fixed-b asymptotics assumes the underlying process to display no more than weak dependence.

This article has two goals. First, I provide a review of consistent and inconsistent approaches to LRV estimation, with an emphasis on the spectral perspective. This clarifies why common approaches to LRV estimation break down under strong autocorrelations.

Second, I derive valid inference methods for a scalar parameter that remain valid even under a specific form of strong dependence. In particular, I assume that the long-run properties are well approximated by a stationary Gaussian AR(1) model. The AR(1) coefficient is allowed to take on values arbitrarily close to one, so that potentially, the process is very persistent. In this manner, the problem of "correcting" for serial correlation remains a first-order problem also in large samples. I then numerically determine tests about the mean of the process that (approximately) maximize weighted average power, using insights of Elliott, Müller, and Watson (2012).

By construction, these tests control size in the AR(1) model in large samples, and this turns out to be very nearly true also in small samples. In contrast, all standard HAC corrections have arbitrarily large size distortions for values of the autoregressive root sufficiently close to unity. In more complicated settings, such as inference about a linear regression coefficient, the AR(1) approach still comes fairly close to controlling size. Interestingly, this includes Granger and Newbold's (1974) classical spurious regression case, where two independent random walks are regressed on each other.

The remainder of the article is organized as follows. The next two sections provide a brief overview of consistent and inconsistent LRV estimation, centered around the problem of inference about a population mean. Section 4 contains the derivation of the new test in the AR(1) model. Section 5 relates these results

© 2014 American Statistical Association
Journal of Business & Economic Statistics
July 2014, Vol. 32, No. 3
DOI: 10.1080/07350015.2014.931238

to more general regression and generalized method of moment (GMM) problems. Section 6 contains some small sample results, and Section 7 concludes.

2. CONSISTENT LRV ESTIMATORS

For a second-order stationary time series y_t with population mean $E[y_t] = \mu$, sample mean $\hat{\mu} = T^{-1}\sum_{t=1}^{T} y_t$, and absolutely summable autocovariances $\gamma(j) = E[(y_t - \mu)(y_{t-j} - \mu)]$, the LRV ω^2 is defined as

$$\omega^2 = \lim_{T\to\infty} \mathrm{var}[T^{1/2}\hat{\mu}] = \sum_{j=-\infty}^{\infty} \gamma(j). \tag{1}$$

With a central limit theorem (CLT) for $\hat{\mu}$, $T^{1/2}(\hat{\mu} - \mu) \Rightarrow \mathcal{N}(0, \omega^2)$, a consistent estimator of ω^2 allows the straightforward construction of tests and confidence sets about μ.

It is useful to take a spectral perspective on the problem of estimating ω^2. The spectral density of y_t is given by the even function $f : [-\pi, \pi] \mapsto [0, \infty)$ defined via $f(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} \cos(j\lambda)\gamma(j)$, so that $\omega^2 = 2\pi f(0)$. Assume T odd for notational convenience. The discrete Fourier transform is a one-to-one mapping from the T values $\{y_t\}_{t=1}^{T}$ into $\hat{\mu}$, and the $T-1$ trigonometrically weighted averages $\{Z_l^{\cos}\}_{l=1}^{(T-1)/2}$ and $\{Z_l^{\sin}\}_{l=1}^{(T-1)/2}$, where

$$Z_l^{\cos} = T^{-1/2}\sqrt{2}\sum_{t=1}^{T} \cos(2\pi l(t-1)/T)y_t,$$
$$Z_l^{\sin} = T^{-1/2}\sqrt{2}\sum_{t=1}^{T} \sin(2\pi l(t-1)/T)y_t. \tag{2}$$

The truly remarkable property of this transformation (see Proposition 4.5.2 in Brockwell and Davis 1991) is that all pairwise correlations between the T random variables $T^{1/2}\hat{\mu}$, $\{Z_l^{\cos}\}_{l=1}^{(T-1)/2}$ and $\{Z_l^{\sin}\}_{l=1}^{(T-1)/2}$ converge to zero as $T \to \infty$, and

$$\sup_{l\le(T-1)/2} |E[(Z_l^{\cos})^2] - 2\pi f(2\pi l/T)| \to 0,$$
$$\sup_{l\le(T-1)/2} |E[(Z_l^{\sin})^2] - 2\pi f(2\pi l/T)| \to 0. \tag{3}$$

Thus, the discrete Fourier transform converts autocorrelation in y_t into heteroscedasticity of $(Z_l^{\sin}, Z_l^{\cos})$, with the shape of the heteroscedasticity governed by the spectral density.

This readily suggests how to estimate the LRV ω^2: collect the information about the frequency $2\pi l/T$ in the lth periodogram ordinate $p_l = \frac{1}{2}((Z_l^{\cos})^2 + (Z_l^{\sin})^2)$, so that p_l becomes an approximately unbiased estimator of $2\pi f(2\pi l/T)$. Under the assumption that f is flat over the frequencies $[0, 2\pi n/T]$ for some integer n, one would naturally estimate $\hat{\omega}_{p,n}^2 = n^{-1}\sum_{l=1}^{n} p_l$ (the subscript p of $\hat{\omega}_{p,n}^2$ stands for "periodogram"). Note that asymptotically, it is permissible to choose $n = n_T \to \infty$ with $n_T/T \to 0$, since any spectral density continuous at 0 becomes effectively flat over $[0, 2\pi n_T/T]$. Thus, a law of large numbers (LLN) applied to $\{p_l\}_{l=1}^{n_T}$ yields $\hat{\omega}_{p,n_T}^2 \xrightarrow{p} \omega^2$.

Popular consistent LRV estimators are often written as weighted averages of sample autocovariances, mimicking the definition (1)

$$\hat{\omega}_{k,S_T}^2 = \sum_{j=-T+1}^{T-1} k(j/S_T)\hat{\gamma}(j), \tag{4}$$

where $\hat{\gamma}(j) = T^{-1}\sum_{t=|j|+1}^{T}(y_t - \hat{\mu})(y_{t-|j|} - \hat{\mu})$ and k is an even weight function with $k(0) = 1$ and $k(x) \to 0$ as $|x| \to \infty$, and the bandwidth parameter S_T satisfies $S_T \to \infty$ and $S_T/T \to 0$ (the subscript k of $\hat{\omega}_{k,S_T}^2$ stands for "kernel"). The Newey and West (1987) estimator, for instance, has this form with k equal to the Bartlett kernel $k(x) = \max(1 - |x|, 0)$. Up to some approximation $\hat{\omega}_{k,S_T}^2$ can be written as a weighted average of periodogram ordinates

$$\hat{\omega}_{k,S_T}^2 \approx \sum_{l=1}^{(T-1)/2} K_{T,l}\, p_l,\ K_{T,l} = \frac{2}{T}\sum_{j=-T+1}^{T-1} \cos(2\pi jl/T)k(j/S_T), \tag{5}$$

where the weights $K_{T,l}$ approximately sum to one, $\sum_{l=1}^{(T-1)/2} K_{T,l} \to 1$. See Appendix for details. Since $(T/S_T)K_{T,l_T} = O(1)$ for $l_T = O(T/S_T)$, these estimators are conceptionally close to $\hat{\omega}_{p,n_T}^2$ with $n_T \approx T/S_T \to \infty$.

As an illustration, consider the problem of constructing a confidence interval for the population mean of the U.S. unemployment rate. The data consist of $T = 777$ monthly observations and is plotted in the left panel of Figure 1. The right panel shows the first 24 log-periodogram ordinates $\log(p_j)$, along with the corresponding part of the log-spectrum (scaled by 2π) of a fitted AR(1) process.

Now consider a Newey–West estimator with bandwidths chosen as $S_T = 0.75T^{1/3} \approx 6.9$ (a default value suggested in Stock and Watson's (2011) textbook for weakly autocorrelated data) and $S_T = 115.9$ (the value derived in Andrews (1991) based on the AR(1) coefficient 0.973). The normalized weights $K_{T,l}/K_{T,1}$ of (5) for the first 24 periodogram ordinates, along with the AR(1) spectrum normalized by $2\pi/\omega^2$, are plotted in Figure 2. Assuming the AR(1) model to be true, it is immediately apparent that both estimators are not usefully thought of as approximately consistent: the estimator with $S_T = 6.9$ is severely downward biased, as it puts most of its weight on periodogram ordinates with expectation much below ω^2. The estimator with $S_T = 115.9$ is less biased, but it has very substantial sampling variability, with 75% of the total weight on the first three periodogram ordinates.

The example demonstrates that there is no way of solving the problem with a more judicious bandwidth choice: to keep the bias reasonable, the bandwidth has to be chosen very large. But such a large bandwidth makes $\hat{\omega}_{k,S_T}^2$ effectively an average of very few periodogram ordinates, so that the LLN provides a very poor approximation, and sampling variability of $\hat{\omega}_{k,S_T}^2$ cannot reasonably be ignored.

An alternative approach (Berk 1974; den Haan and Levin 1997) is to model y_t as an autoregressive moving average (ARMA) process with parameter θ and spectrum $f_{\mathrm{ARMA}}(\lambda; \theta)$, say, and to estimate ω^2 from the implied spectrum $\hat{\omega}_{\mathrm{ARMA}}^2 = 2\pi f_{\mathrm{ARMA}}(0; \hat{\theta})$. The discrete Fourier approximation (3) implies

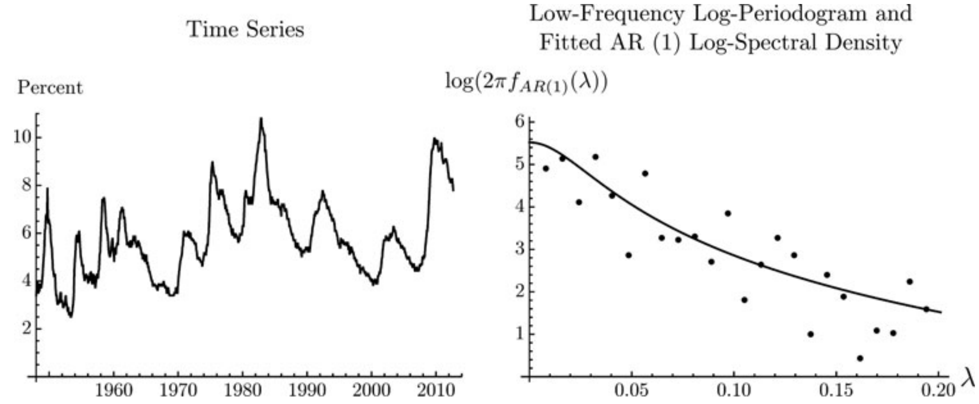

Figure 1. U.S. unemployment. Notes: Series LNS14000000 of the Bureau of Labor Statistics from 1948:1 to 2012:9. The AR(1) log-spectral density has coefficient 0.973 and variance 0.182, the MLEs from a "low-frequency likelihood" (Equation (11) of Section 4 with $q = 48$).

Whittle's (1957, 1962) log-likelihood approximation

$$\frac{T}{2}\log(2\pi) - \sum_{l=1}^{(T-1)/2} \log(2\pi f_{\mathrm{ARMA}}(2\pi l/T;\theta)) - \sum_{l=1}^{(T-1)/2} \frac{p_l}{2\pi f_{\mathrm{ARMA}}(2\pi l/T;\theta)}$$

to the Gaussian ARMA log-likelihood. The (quasi-) maximum likelihood estimator (MLE) $\hat{\theta}$ that defines $\hat{\omega}^2_{\mathrm{ARMA}}$ is thus determined by information about all frequencies, as encoded by the whole periodogram. This is desirable if there are good reasons to assume a particular spectral shape for y_t; otherwise, it leads to potential misspecification, as $\hat{\theta}$ maximizes fit on average, but not necessarily for frequencies close to zero.

Also note that this approach only delivers a useful "consistent" estimator for ω^2 if the estimation uncertainty in $\hat{\theta}$ is relatively small. The LRV of an AR(1) model, for instance, is given by $\omega^2 = \sigma^2/(1-\rho)^2$. This mapping is very sensitive to estimation errors if ρ is close to one. As an illustration, consider the AR(1) model for the unemployment series with $\rho = 0.973$. An estimation error of one standard error in $\hat{\rho}$, $\sqrt{(1-\rho^2)/T} \approx 0.008$, leads to an estimation error in $\hat{\omega}^2_{\mathrm{AR(1)}}$ by a factor of 0.6 and 2.0, respectively. So even if the AR(1) model was known to be correct over all frequencies, one would still expect $\hat{\omega}^2_{\mathrm{AR(1)}}$ to have poor properties for ρ close to one.

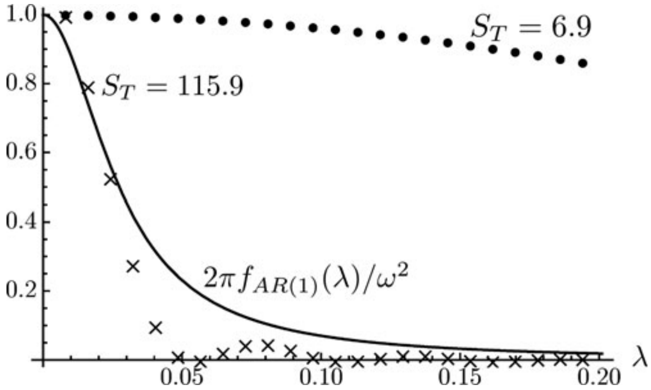

Figure 2. Newey–West weights on periodogram ordinates and normalized AR(1) spectral density. Notes: The dots and crosses correspond to the approximate weights on the first 24 periodogram ordinates of a Newey–West LRV estimator with bandwidth equal to S_T, normalized by the weight on the first periodogram ordinate. Total weight $\sum_{l=1}^{(T-1)/2} K_{T,l}$ is 0.99 and 0.85 for $S_T = 6.9$ and 115.9, respectively. The line is the spectral density of an AR(1) process with coefficient 0.973, scaled by $2\pi/\omega^2$.

A hybrid approach between kernel and parametric estimators is obtained by so-called prewhitening. The idea is to use a parsimonious parametric model to get a rough fit to the spectral density, and to then apply kernel estimator techniques to account for the misspecification of the parametric model near frequency zero. For instance, with AR(1) prewhitening, the overall LRV estimator is given by $\hat{\omega}^2 = \hat{\omega}^2_e/(1-\hat{\rho})^2$, where $\hat{\rho}$ is the estimated AR(1) coefficient, and $\hat{\omega}^2_e$ is a kernel estimator applied to the AR(1) residuals. Just as for the fully parametric estimator, this approach requires the estimation error in the prewhitening stage to be negligible.

3. INCONSISTENT LRV ESTIMATORS

The above discussion suggests that for inference with persistent time series, one cannot safely ignore estimation uncertainty in the LRV estimator. Under the assumption that the spectral density is approximately flat over some thin frequency band around zero, one might still rely on an estimator of the type $\hat{\omega}^2_{p,n} = n^{-1}\sum_{l=1}^{n} p_l$ introduced in the last section. But in contrast to the discussion there, one would want to acknowledge that reasonable n are small, so that no LLN approximation holds. The randomness in the t-statistic $\sqrt{T}(\hat{\mu}-\mu)/\hat{\omega}_{p,n}$ then not only stems from the numerator, as usual, but also from the denominator $\hat{\omega}_{p,n}$.

To make further progress, a distributional model for the periodogram ordinates is needed. Here, a central limit argument may be applied: under a linear process assumption for y_t, say, the line of reasoning in Theorems 10.3.1 and 10.3.2 of Brockwell and Davis (1991) implies that any finite set of the trigonometrically weighted averages $(Z^{\sin}_l, Z^{\cos}_l)$ is asymptotically jointly normal. As a special case, the approximation (3) is thus strengthened to $(Z^{\sin}_l, Z^{\cos}_l)'$ distributed approximately independent and normal with covariance matrix $2\pi f(2\pi l/T)I_2$, $l = 1,\ldots,n$. With an assumption of flatness of the spectral density over the frequencies $[0, 2\pi n/T]$, that is, $2\pi f(2\pi l/T) \approx \omega^2$ for $l = 1,\ldots,n$, one therefore obtains $\hat{\omega}^2_{p,n}$ to be approximately distributed chi-squared with $2n$ degrees of freedom, scaled by ω^2. Since the common scale ω cancels from the numerator and denominator, the t-statistic $\sqrt{T}(\hat{\mu}-\mu)/\hat{\omega}_{p,n}$ is then approximately distributed Student-t with $2n$ degrees of freedom. The difference between the critical value derived from the normal distribution, and that derived from this Student-t distribution, accounts for the estimation uncertainty in $\hat{\omega}_{p,n}$.

Very similar inconsistent LRV estimator have been suggested in the literature, although they are based on trigonometrically weighted averages slightly different from $(Z_l^{\sin}, Z_l^{\cos})$. In particular, the framework analyzed in Müller (2004, 2007) and Müller and Watson (2008) led these authors to consider the averages

$$Y_l = T^{-1/2}\sqrt{2}\sum_{t=1}^{T}\cos(\pi l(t-1/2)/T)y_t,\, l \geq 1, \tag{6}$$

so that for integer l, $Y_{2l} = Z_l^{\cos}$, but $Z_l^{\sin}$ is replaced by Y_{2l+1}, a cosine weighted average of frequency just between $2\pi l/T$ and $2\pi(l+1)/T$. This difference is quite minor, though: just like the discrete Fourier transform, (6) is a one-to-one transformation that maps $\{y_t\}_{t=1}^T$ into the T variables $\hat{\mu}$ and $\{Y_l\}_{l=1}^{T-1}$, and Y_l^2 is an approximately unbiased estimator of the spectral density at frequency $2\pi f(\pi l/T)$. Under the assumption of flatness of the spectrum over the band $[0, \pi q/T]$, and a CLT for $\{Y_l\}_{l=1}^q$, we then obtain Müller's (2004, 2007) estimator

$$\hat{\omega}^2_{Y,q} = \frac{1}{q}\sum_{l=1}^{q} Y_l^2, \tag{7}$$

which implies an approximate Student-t distribution with q degrees of freedom for the t-statistic $\sqrt{T}(\hat{\mu}-\mu)/\hat{\omega}_{Y,q}$. There is an obvious trade-off between robustness and efficiency, as embodied by q: choosing q small makes minimal assumptions about the flatness of the spectrum, but leads to a very variable $\hat{\omega}^2_{Y,q}$ and correspondingly large critical value of the t-statistic $\sqrt{T}(\hat{\mu}-\mu)/\hat{\omega}_{Y,q}$. Similar approaches, for potentially different weighted averages, are pursued in Phillips (2005) and Sun (2013).

In the example of the unemployment series, one might, for instance, be willing to assume that the spectrum is flat below business cycle frequencies. Defining cycles of periodicity larger than 8 years as below business cycle, this leads to a choice of q equal to the largest integer smaller than the span of the data in years divided by 4. In the example with $T = 777$ monthly observations, $(777/12)/4 \approx 16.2$, so that $q = 16$. Given the relationship between Y_l and $(Z_l^{\sin}, Z_l^{\cos})$, this is roughly equivalent to the assumption that the first eight periodogram ordinates in Figure 1 have the same mean. Clearly, if the AR(1) model with $\rho = 0.973$ was true, then this is a fairly poor approximation. At the same time, the data do not overwhelmingly reject the notion that $\{Y_l\}_{l=1}^{16}$ is iid mean-zero normal: Müller and Watson's (2008) low-frequency stationarity test, for instance, fails to reject at the 5% level (although it does reject at the 10% level).

The inconsistent LRV estimators that arise by setting $S_T = bT$ for some fixed b in (4) as studied by Kiefer and Vogelsang (2005) are not easily cast in purely spectral terms, as Equation (5) no longer provides a good approximation when $S_T = bT$ in general. Interestingly, though, the original (Kiefer, Vogelsang, and Bunzel 2000; Kiefer and Vogelsang 2002b) suggestion of a Bartlett kernel estimator with bandwidth equal to the sample size can be written exactly as

$$\hat{\omega}^2_{\text{KVB}} = \sum_{l=1}^{T-1}\kappa_{T,l}Y_l^2, \tag{8}$$

where $1/\kappa_{T,l} = 4T^2\sin(\pi l/(2T))^2 \to \pi^2 l^2$ (see Appendix for details). Except for an inconsequential scale factor, the estimator $\hat{\omega}^2_{\text{KVB}}$ is thus conceptually close to a weighted average of periodogram ordinates (5), but with weights $1/(\pi l)^2$ that do not spread out even in large samples. The limiting distribution under weak dependence is nondegenerate and equal to a weighted average of chi-square-distributed random variables, scaled by ω^2. Sixty percent of the total weight is on Y_1^2, and another 30% on $Y_2^2, \ldots, Y_6^2$. So $\hat{\omega}^2_{\text{KVB}}$ can also usefully be thought of as a close cousin of (7) with q very small.

As noted above, inference based on inconsistent LRV estimators depends on a distributional model for $\hat{\omega}^2$, generated by CLT arguments. In contrast, inference based on consistent LRV estimators only requires the LLN approximation to hold. This distinction becomes important when considering nonstationary time series with pronounced heterogeneity in the second moments. To fix these ideas, suppose $y_t = \mu + (1 + \mathbf{1}[t > T/2])\varepsilon_t$ with $\varepsilon_t \sim \text{iid}(0, \sigma^2)$, that is, the variance of y_t quadruples in the second half of the sample. It is not hard to see that despite this nonstationarity, the estimator (4) consistently estimates $\omega^2 = \lim_{T\to\infty}\text{var}[T^{1/2}\hat{\mu}] = \frac{5}{2}\sigma^2$, so that standard inference is justified. At the same time, the nonstationarity invalidates the discrete Fourier approximation (3), so that $\{(Z_l^{\sin}, Z_l^{\cos})'\}_{l=1}^n$ are no longer uncorrelated, even in large samples, and $\sqrt{T}(\hat{\mu}-\mu)/\hat{\omega}_{p,n}$ is no longer distributed Student-t. The same holds for $\{Y_l\}_{l=1}^q$ and $\sqrt{T}(\hat{\mu}-\mu)/\hat{\omega}_{Y,q}$, and also the asymptotic approximations derived in Kiefer and Vogelsang (2005) for fixed-b estimators are no longer valid. In general, then, inconsistent LRV estimators require a strong degree of homogeneity to justify the distributional approximation of $\hat{\omega}^2$, and are not uniformly more "robust" than consistent ones.

One exception is the inference method suggested by Ibragimov and Müller (2010). In their approach, the parameter of interest is estimated q times on q subsets of the whole sample. The subsets must be chosen in a way that the parameter estimators are approximately independent and Gaussian. In a time series setting, a natural default for the subsets is a simple partition of the T observations into q nonoverlapping consecutive blocks of (approximately) equal length. Under weak dependence, the q resulting estimators of the mean $\hat{\mu}_l$ are asymptotically independent Gaussian with mean μ. The variances of $\hat{\mu}_l$ are approximately the same (and equal to $q\omega^2/T$) when the time series is stationary, but are generally different from each other when the second moment of y_t is heterogenous in time. The usual t-statistic computed from the q observations $\hat{\mu}_l$, $l = 1, \ldots, q$,

$$\frac{\sqrt{q}(\bar{\hat{\mu}}-\mu)}{\sqrt{q^{-1}\sum_{l=1}^{q}(\hat{\mu}_l-\bar{\hat{\mu}})^2}} \tag{9}$$

with $\bar{\hat{\mu}} = q^{-1}\sum_{l=1}^q \hat{\mu}_l \approx \hat{\mu}$, thus has approximately the same distribution as a t-statistic computed from independent and zero-mean Gaussian variates of potentially heterogenous variances. Now Ibragimov and Müller (2010) invoked a remarkable result of Bakirov and Székely (2005), who showed that ignoring potential variance heterogeneity in the small sample t-test about independent Gaussian observations, that is, to simply employ the usual Student-t critical value with $q-1$ degrees of freedom, still leads to valid inference at the 5% two-sided level and below. Thus, as long as y_t is weakly dependent, simply comparing (9) to a Student-t critical value with $q-1$ degrees of freedom leads to approximately correct inference by construction, even under very pronounced forms of second moment nonstationarities in y_t. From a spectral perspective, the choice of q in (9) roughly corresponds to an assumption that the spectral density is flat over $[0, 1.5\pi q/T]$. The factor of 1.5 relative to the assumption justifying $\hat{\omega}^2_{Y,q}$ in (7) reflects the relatively poorer frequency extraction by the simple block averages $\hat{\mu}_l$ relative to the cosine weighted averages Y_l. See Müller and Watson (2013) for related computations.

4. POWERFUL TESTS UNDER AR(1) PERSISTENCE

All approaches reviewed in the last two sections exploit flatness of the spectrum close to the origin: the driving assumption is that there are at least some trigonometrically weighted averages that have approximately the same variance as the simple average $\sqrt{T}\hat{\mu}$. But as Figure 2 demonstrates, there might be only very few, or even no such averages for sufficiently persistent y_t.

An alternative approach is to take a stand on possible shapes of the spectrum close to the origin, and to exploit that restriction to obtain better estimators. The idea is analogous to using a local polynomial estimator, rather than a simple kernel estimator, in a nonparametric setting. Robinson (2005) derived such consistent LRV estimators under the assumption that the underlying persistence is of the "fractional" type. This section derives valid inference when instead the long-run persistence is generated by an autoregressive root close to unity. In contrast to the fractional case, this precludes consistent estimation of the spectral shape, even if this parametric restriction is imposed on a wide frequency band.

4.1. Local-To-Unity Asymptotics

For very persistent series, the asymptotic independence between $\{(Z_l^{\sin}, Z_l^{\cos})'\}_{l=1}^n$, or $\{Y_l\}_{l=1}^q$, no longer holds. So we begin by developing a suitable distributional theory for $\{Y_l\}_{l=0}^q$ for fixed q, where $Y_0 = \sqrt{T}\hat{\mu}$.

Suppose initially that y_t is exactly distributed as a stationary Gaussian AR(1) with mean μ, coefficient ρ, $|\rho| < 1$, and variance σ^2: with $y = (y_1, \ldots, y_T)'$ and $e = (1, \ldots, 1)'$,

$$y \sim \mathcal{N}(\mu e, \sigma^2 \Sigma(\rho)),$$

where $\Sigma(\rho)$ has elements $\Sigma(\rho)_{i,j} = \rho^{|i-j|}/(1-\rho^2)$. Define H as the $T \times (q+1)$ matrix with first column equal to $T^{-1/2}e$, and $(l+1)$th column with elements $T^{-1/2}\sqrt{2}\cos(\pi l(t-1/2)/T)$, $t = 1, \ldots, T$, and ι_1 as the first column of I_{q+1}. Then $Y = (Y_0, \ldots, Y_q)' = H'y \sim \mathcal{N}(T^{1/2}\mu\iota_1, \sigma^2\Omega(\rho))$ with $\Omega(\rho) = H'\Sigma(\rho)H$.

The results reviewed above imply that for any fixed q and $|\rho| < 1$, as $T \to \infty$, $\sigma^2\Omega(\rho)$ becomes proportional to I_{q+1} (with proportionality factor equal to $\omega^2 = \sigma^2/(1-\rho)^2$). But for any fixed T, there exists a ρ sufficiently close to one for which $\Omega(\rho)$ is far from being proportional to I_{q+1}. This suggests that asymptotics along sequences $\rho = \rho_T \to 1$ yield approximations that are relevant for small samples with sufficiently large ρ.

The appropriate rate of convergence of ρ_T turns out to be $\rho_T = 1 - c/T$ for some fixed number $c > 0$, leading to so-called "local-to-unity" asymptotics. Under these asymptotics, a calculation shows that $T^{-2}\Omega(\rho_T) \to \Omega_0(c)$, where the $(l+1), (j+1)$ element of $\Omega_0(c)$ is given by

$$\frac{1}{2c}\int_0^1\int_0^1 \phi_l(s)\phi_j(r)e^{-c|r-s|}dsdr \qquad (10)$$

with $\phi_l(s) = \sqrt{2}\cos(\pi l s)$ for $l \geq 1$ and $\phi_0(s) = 1$. This expression is simply the continuous time analogue of the quadratic form $H'\Sigma(\rho)H$: the weighting functions ϕ_l corresponds to the limit of the columns of H, $T(1-\rho_T^2) \to 2c$ and $e^{-c|r-s|}$ corresponds to the limit of $\rho^{|i-j|}$.

Now the assumption that y_t follows exactly a stationary Gaussian AR(1) is obviously uncomfortably strong. So suppose instead that y_t is stationary and satisfies $y_t = \rho_T y_{t-1} + (1-\rho_T)\mu + u_t$, where u_t is some weakly dependent mean-zero disturbance with LRV σ^2. Under suitable conditions, Chan and Wei (1987) and Phillips (1987) showed that $T^{-1/2}(y_{[\cdot T]} - \mu)$ then converges in distribution to an Ornstein–Uhlenbeck process $\sigma J_c(\cdot)$, with covariance kernel $E[J_c(r)J_c(s)] = e^{-c|r-s|}/(2c)$. Noting that this is exactly the kernel in (10), the approximation

$$T^{-1}Y \sim \mathcal{N}(T^{-1/2}\mu\iota_1, \sigma^2\Omega_0(c)) \qquad (11)$$

is seen to hold more generally in large samples. The dimension of Y, that is, the choice of q, reflects over which frequency band the convergence to the Ornstein–Uhlenbeck process is deemed an accurate approximation (see Müller 2011 for a formal discussion of asymptotic optimality under such a weak convergence assumption). Note that for large c, $e^{-c|r-s|}$ in (10) becomes very small for nonzero $|r-s|$, so that $\int_0^1\int_0^1 \phi_l(s)\phi_j(r)e^{-c|r-s|}dsdr \propto \int_0^1 \phi_l(s)\phi_j(s)ds = \mathbf{1}[i = j]$, recovering the proportionality of $\Omega_0(c)$ to I_{q+1} that arises under a spectral density that is flat in the $1/T$ neighborhood of the origin. For a given q, inference that remains valid under (11) for all $c > 0$ is thus strictly more robust than t-statistic-based inference with $\hat{\omega}^2_{Y,q}$ in (7).

4.2. Weighted Average Power Maximizing Scale Invariant Tests

Without loss of generality, consider the problem of testing $H_0: \mu = 0$ (otherwise, simply subtract the hypothesized mean from y_t) against $H_1: \mu \neq 0$, based on the observation Y with distribution (11). The derivation of powerful tests is complicated by the fact that the alternative is composite (μ is not specified under H_1), and the presence of the two nuisance parameters σ^2 and c.

A useful device for dealing with the composite nature of the alternative hypothesis is to seek tests that maximize weighted average power. For computational convenience, consider a weighting function for μ that is mean-zero Gaussian with variance η^2. As argued by King (1987), it makes sense to choose η^2 in a way that good tests have approximately 50% weighted average power. Now for c and T large, $\text{var}[T^{-1}Y_0] \approx T^{-2}\sigma^2/(1-\rho_T)^2 \approx \sigma^2/c^2$. Furthermore, if σ and c were known, the best 5% level test would simply reject if $|T^{-1}Y_0| > 1.96 \cdot \sigma/c$. This motivates a choice of $\eta^2 = 10T\sigma^2/c^2$, since this would induce this (infeasible) test to have power of approximately $2P(\mathcal{N}(0, 11) > 1.96) \approx 56\%$. Furthermore, by standard arguments, maximizing this weighted average power criterion for given σ and c is equivalent to maximizing power against the alternative

$$T^{-1}Y \sim \mathcal{N}(0, \sigma^2\Omega_1(c)) \tag{12}$$

with $\Omega_1(c) = \Omega_0(c) + (10/c^2)\iota_1\iota_1'$.

The testing problem has thus been transformed into H_0' : $T^{-1}Y \sim \mathcal{N}(0, \sigma^2\Omega_0(c))$ against H_1' : $T^{-1}Y \sim \mathcal{N}(0, \sigma^2\Omega_1(c))$, which is still complicated by the presence of the two nuisance parameters σ^2 and c. For σ^2, note that in most applications, it makes sense to impose that if the null hypothesis is rejected for some observation Y, then it should also be rejected for the observation aY, for any $a > 0$. For instance, in the unemployment example, this ensures that measuring the unemployment rate as a ratio (values between 0 and 1) or in percent (values between 0 and a 100) leads to the same results. Standard testing theory (see chap. 6 in Lehmann and Romano 2005) implies that any test that satisfies this scale invariance can be written as a function of $Y^s = Y/\sqrt{Y'Y}$. One might thus think of Y^s as the effective observation, whose density under H_i', $i = 0, 1$ is equal to (see Kariya 1980; King 1980)

$$f_{i,c}(y^s) = \kappa_q|\Omega_i(c)|^{-1/2}(y^{s\prime}\Omega_i(c)^{-1}y^s)^{-(q+1)/2} \tag{13}$$

for some constant κ_q.

A restriction to scale invariant tests has thus further transformed the testing problem into H_0'' :"Y^s has density $f_{0,c}$" against H_1'' : "Y^s has density $f_{1,c}$." This is a nonstandard problem involving the key nuisance parameter $c > 0$, which governs the degree of persistence of the underlying time series. Elliott, Müller, and Watson (2012) studied nonstandard problems of this kind, and showed that a test that approximately maximizes weighted average power relative to the weighting function F rejects for large values of $\int f_{1,c}(Y^s)dF(c)/\int f_{0,c}(Y^s)d\tilde{\Lambda}(c)$, where $\tilde{\Lambda}$ is a numerically determined approximately least favorable distribution. I implement this approach with F discrete and uniform on the 15 points $c_j = e^{(j-1)/2}$, $j = 1, \ldots, 15$, so that weighted average power is approximately maximized against a uniform distribution on $\log(c)$ on the interval [0, 7]. The endpoints of this interval are chosen such that the whole span of very nearly unit root behavior ($c = 1$) to very nearly stationary behavior ($c = e^7 \approx 1097$) is covered. The substantial mass on relatively large values of c induces good performance also under negligible degrees of autocorrelation.

Figure 3 plots the power of these tests for $q \in \{12, 24, 48\}$ against the alternatives (12) as a function of $\log(c)$. As can be seen from Figure 3, none of these tests have power when $c \to 0$. To understand why, go back to the underlying stationary AR(1) model for y_t. For small c (say, $c < 1$), $y_t - y_1$ is essentially indistinguishable from a random walk process. Decreasing c further thus leaves the distribution of $\{y_t - y_1\}_{t=1}^T$ largely unaffected, while it still increases the variance of y_1, $\text{var}[y_1] = \sigma^2/(1-\rho_T^2) \approx \sigma^2T/(2c)$. But for a random walk, the

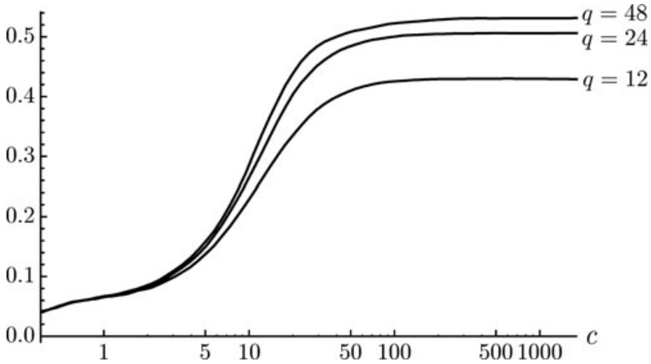

Figure 3. Asymptotic weighted average power of tests under AR(1) persistence. Notes: Asymptotic weighted average power of 5% level hypothesis tests about the mean of an AR(1) model with unit innovation variance and coefficient $\rho = \rho_T = 1 - c/T$ with an $\mathcal{N}(0, 10T/c^2)$ weighting function on the difference between population and hypothesized mean, based on q low-frequency cosine weighted averages of the original data.

initial condition $y_1 - \mu$ and the mean μ are not separately identified, as they both amount to translation shifts. With the mean-zero Gaussian weighting on μ, and under scale invariance, the alternative H_1'' :"Y^s has density $f_{1,c}$" for some small $c = c_1$ is thus almost indistinguishable from the null H_0'' : "Y^s has density f_{0,c_0}" for some $c_0 < c_1$, so that power under such alternatives cannot be much larger than the nominal level. As a consequence, if the tests are applied to an exact unit root process ($c = 0$, which is technically ruled out in the derivations here) with arbitrary fixed initial condition, they only reject with 5% probability. This limiting behavior might be considered desirable, as the population mean of a unit root process does not exist, so a valid test should not systematically rule out any hypothesized value.

The differences in power for different values of q in Figure 3 reflect the value of the additional information contained in Y_l, l large, for inference about μ. This information has two components: on the one hand, additional observations Y_l help to pin down the common scale, analogous to the increase in power of tests based on $\hat{\omega}_{Y,q}^2$ in (7) as a function of q. On the other hand, additional observations Y_l contain information about the shape parameter c. For instance, as noted above, $c \to \infty$ corresponds to flatness of the spectral density in the relevant $1/T$ neighborhood of the origin. The tests based on $\hat{\omega}_{Y,q}^2$ dogmatically impose this flatness and have power of {52.4%, 54.0%, 54.7%} for $q \in \{12, 24, 48\}$ against the alternative (12) with $c \to \infty$. For $q = 12$, this power is 9.4% larger than the power of the test in Figure 3 for c large, reflecting that 12 observations are not sufficient to learn about the flatness of the spectrum. For $q = 24$ and $q = 48$, however, the difference is only 3.3% and 1.6%, respectively.

Even with $q \to \infty$, the shape of the spectral density cannot be consistently estimated under local-to-unity asymptotics. In unreported results, I derive weighted average power maximizing tests based on all observations of a Gaussian AR(1), and find that asymptotic overall weighted average power in this $q = \infty$ case is only 1.9% larger than what is obtained with $q = 48$. As a practical matter, it thus makes little sense to consider tests with q larger than 48.

Table 1. Constants for computation of S_q

q	B	$\text{cv}_{0.01}$	$\text{cv}_{0.05}$	$\text{cv}_{0.10}$		δ_1	δ_2	δ_3	δ_4	δ_5
12	6.2	0.70	1.00	3.25		1.74	−0.44	0.75	2.11	1.80
24	10.0	0.74	1.00	4.23		1.72	−2.16	0.95	1.45	0.96
48	12.0	0.68	1.00	4.27		1.64	−0.81	1.04	1.18	0.49
	δ_6	δ_7	δ_8	δ_9	δ_{10}	δ_{11}	δ_{12}	δ_{13}	δ_{14}	δ_{15}
12	1.75	1.82	1.27	0.32	−0.12	−0.54	−0.80	−1.07	−1.47	−1.82
24	0.01	1.33	1.45	1.48	1.52	0.28	−0.44	−0.90	−1.36	−1.70
48	0.90	0.52	0.89	0.65	1.10	1.29	0.97	−0.01	−0.66	−0.77

4.3. Suggested Implementation and Empirical Illustration

The suggested test statistic is a slightly modified version of $\int f_{1,c}(Y^s)dF(c)/\int f_{0,c}(Y^s)d\tilde{\Lambda}(c)$. A first modification approximates $\Omega_0(c)$ by diagonal matrices with elements $1/(c^2+\pi^2 j^2)$; these diagonal values correspond to the suitably scaled limit of the spectral density of an AR(1) process with $\rho = \rho_T = 1 - c/T$ at frequency $\pi j/T$. This avoids the slightly cumbersome determination of the elements of $\Omega_0(c)$ as a double integral in (10). The second modification bounds the value of $|Y_0|$ relative to $(Y_1, \ldots, Y_q)'$. The bound is large enough to leave weighted average power almost completely unaffected, but it eliminates the need to include very small values of c in the support of $\tilde{\Lambda}$, and helps to maintain the power of the test against distant alternatives when c is large. (Without this modification, it turns out that power functions are not monotone for c large.) Taken together, these modifications lead to a loss in overall weighted average power relative to the 5% level test reported in Figure 3 of approximately one percentage point for all considered values of q.

In detail, the suggested test about the population mean $H_0: \mu = \mu_0$ of an observed scalar time series $\{y_t\}_{t=1}^T$ is computed as follows.

1. Compute the $q+1$ values $Y_0 = T^{-1/2}\sum_{t=1}^T (y_t - \mu_0)$ and $Y_l = T^{-1/2}\sqrt{2}\sum_{t=1}^T \cos(\pi l(t-1/2)/T)y_t, \quad l = 1, \ldots, q$. (So that $Y_0 = \sqrt{T}(\hat{\mu} - \mu_0)$, and Y_l, $l \geq 1$ is as in (6).)
2. Replace Y_0 by $\min(|Y_0|, B\sqrt{q^{-1}\sum_{l=1}^q Y_l^2})$, where B is given in Table 1. (For reasons described above.)
3. Define $d_{i,l}^0 = (c_i^2 + (\pi l)^2)/c_i^2$, where $c_i = e^{(i-1)/2}$, and $d_{i,l}^1 = d_{i,l}^0$ for $l \geq 1$, and $d_{i,0}^1 = 1/11$, $i = 1, \ldots, 15$. (These are the elements of the diagonal approximations to $\Omega_0(c)^{-1}$ and $\Omega_1(c)^{-1}$, scaled by $1/c^2$ for convenience.)
4. Compute

$$S_q = \frac{\sum_{i=1}^{15}(\prod_{l=0}^q d_{i,l}^1)^{1/2}(\sum_{l=0}^q d_{i,l}^1 Y_l^2)^{-(q+1)/2}}{\sum_{i=1}^{15}\exp(\delta_i)(\prod_{l=0}^q d_{i,l}^0)^{1/2}(\sum_{l=0}^q d_{i,l}^0 Y_l^2)^{-(q+1)/2}},$$

where δ_i depends on q and is given in Table 1. (This corresponds to the ratio of $\int f_{1,c}(Y^s)dF(c)/\int f_{0,c}(Y^s)d\tilde{\Lambda}(c)$ with $f_{i,c}$ as in (13) and $\tilde{\Lambda}$ is described by point masses at c_i with relative weights e^{δ_i}.)
5. Reject the null hypothesis at level α if $S_q > \text{cv}_\alpha$, where the critical values cv_α for $\alpha \in \{0.01, 0.05, 0.10\}$ are given in Table 1. (The 5% level critical values are all equal to one, as the appropriate cut-off value for the (approximate) ratio $\int f_{1,c}(Y^s)dF(c)/\int f_{0,c}(Y^s)d\tilde{\Lambda}(c)$ is subsumed in δ_i.)

The values of B, δ_i, and cv_α are numerically determined such that the test S_q is of nominal level under (11), for arbitrary $c > 0$, and that it come close to maximizing weighted average power relative to the weighting function F and (12) at the 5% level.

A confidence interval for μ can be constructed by inverting this test, that is, by determining the set of values of μ_0 for which the test does not reject. This is most easily done by a simple grid search over plausible values of μ_0 (note that different values of μ_0 leave Y_l, $l \geq 1$ unaffected). Numerical calculations suggest that 95% confidence intervals are never empty, as S_q does not seem to take on values larger than the critical value whenever $Y_0 = 0$, that is, when $\mu_0 = \hat{\mu}$. They can be equal to the real line, though, as S_q might not reject for any value of μ_0. When ρ is very close to one (c is very small), this happens necessarily for almost 95% of the draws, as such series contain essentially no information about the population mean. Under asymptotics that correspond to weak dependence ($c \to \infty$), unbounded 95% confidence intervals still arise for 8.6% of the draws by inverting S_{12}, but essentially never ($< 0.05\%$) when inverting S_{24} or S_{48}.

Table 2 reports 95% confidence intervals for the population mean of U.S. unemployment using this test. As a comparison, the table also includes 95% confidence intervals based on three "consistent" estimators (i.e., standard normal critical values are employed) and five "inconsistent" estimators. Specifically, the first group includes Andrews' (1991) estimator $\hat{\omega}_{\text{A91}}^2$ with a quadratic spectral kernel k and bandwidth selection using an AR(1) model; Andrews and Monahan's (1992) $\hat{\omega}_{\text{AM}}^2$ suggestion of the same estimator, but after prewhitening with an AR(1) model; and the fully parametric estimator $\hat{\omega}_{\text{AR}(12)}^2$ based on an

Table 2. 95% confidence intervals for unemployment population mean

	S_{12}	S_{24}	S_{48}	$\hat{\omega}_{\text{A91}}^2$	$\hat{\omega}_{\text{AM}}^2$	$\hat{\omega}_{\text{AR}(12)}^2$	$\hat{\omega}_{\text{KVB}}^2$	$\hat{\omega}_{\text{Y},12}^2$	$\hat{\omega}_{\text{Y},24}^2$	$\hat{\omega}_{\text{SPJ}}^2$	IM_8	IM_{16}
m.e.	∞	1.31	1.34	0.75	2.21	0.88	1.46	0.85	0.65	2.69	1.02	0.77

NOTES: Unemployment data same as in Figure 1. All confidence intervals are symmetric around the sample mean $\hat{\mu} = 5.80$ with endpoints $\hat{\mu}\pm$m.e., where the margin of error m.e. is reported in the table.

AR(12) model. The second group includes Kiefer, Vogelsang, and Bunzel's (2000) Bartlett kernel estimator with lag-length equal to sample size $\hat{\omega}^2_{\mathrm{KVB}}$; Müller's (2007) estimators $\hat{\omega}^2_{Y,12}$ and $\hat{\omega}^2_{Y,24}$; Sun, Phillips, and Jin's (2008) quadratic spectral estimator $\hat{\omega}^2_{\mathrm{SPJ}}$ with a bandwidth that trades off asymptotic Type I and Type II errors in rejection probabilities, with the shape of the spectral density approximated by an AR(1) model and with their weight parameter w equal to 30; and Ibragimov and Müller's (2010) inference with 8 and 16 groups, IM_8 and IM_{16}. The confidence interval based on S_{12} is equal to the whole real line; the 12 lowest cosine transforms (6) of the unemployment rate do not seem to exhibit sufficient evidence of mean reversion for the test to reject any value of the population mean. The full sample AR(1) coefficient estimate is equal to 0.991; this very large value seems to generate the long intervals based on $\hat{\omega}^2_{\mathrm{AM}}$ and $\hat{\omega}^2_{\mathrm{SPJ}}$.

5. GENERALIZATION TO REGRESSION AND GMM PROBLEMS

The discussion of HAC corrections has so far focused on the case of inference about the mean μ of an observable time series $\{y_t\}_{t=1}^T$. But the approaches can be generalized to inference about scalar parameters of interest in regression and GMM contexts.

Consider first inference about a regression parameter. Denote by β the $k \times 1$ regression coefficient, and suppose we are interested in its first element $\beta_1 = \iota_1'\beta$, where in this section, ι_1 denotes the first column of I_k. The observable regressand R_t and $k \times 1$ regressors X_t are assumed to satisfy

$$R_t = X_t'\beta + e_t, \qquad E[e_t | X_{t-1}, X_{t-2}, \ldots] = 0, \qquad t = 1, \ldots, T.$$

Let $\hat{\Sigma}_X = T^{-1}\sum_{t=1}^T X_t X_t'$, and let $\hat{\beta}$ be the ordinary least-square (OLS) estimator $\hat{\beta} = \hat{\Sigma}_X^{-1} T^{-1} \sum_{t=1}^T X_t R_t$. Under suitable regular conditions, $\hat{\Sigma}_X \xrightarrow{p} \Sigma_X$, so that $\sqrt{T}(\hat{\beta}_1 - \beta_1)$ has variance

$$\begin{aligned} \mathrm{var}[\sqrt{T}\hat{\beta}_1] &\approx \iota_1'\Sigma_X^{-1}\mathrm{var}\Big[T^{-1/2}\sum_{t=1}^T X_t e_t\Big]\Sigma_X^{-1}\iota_1 \\ &= \mathrm{var}\Big[T^{-1/2}\sum_{t=1}^T \tilde{y}_t\Big], \end{aligned}$$

where $\tilde{y}_t = \iota_1'\Sigma_X^{-1}X_t e_t$. The problem of estimating the variance of $\hat{\beta}_1$ is thus cast in the form of estimating the LRV of the scalar series $\tilde{y}_t$.

The series $\tilde{y}_t$is not observed, however. So consider instead the observable series $\hat{y}_t = \iota_1'\hat{\Sigma}_X^{-1}X_t\hat{e}_t$, with $\hat{e}_t = R_t - X_t'\hat{\beta}$ the OLS residual, and suppose the HAC corrections discussed in Sections 2 and 3 are computed for $y_t = \hat{y}_t$. The difference between $\tilde{y}_t$ and $\hat{y}_t$ is given by

$$\hat{y}_t = \tilde{y}_t + \iota_1'(\hat{\Sigma}_X^{-1} - \Sigma_X^{-1})X_t e_t - \iota_1'\hat{\Sigma}_X^{-1}X_t X_t'(\hat{\beta} - \beta). \tag{14}$$

With $\hat{\Sigma}_X \xrightarrow{p} \Sigma_X$, the middle term on the right-hand side of (14) is asymptotically negligible. Furthermore, since $\hat{\beta} \xrightarrow{p} \beta$, the last term cannot substantially affect many periodogram ordinates at the same time, so that for *consistent* LRV estimators the underlying LLN still goes through. For *inconsistent* estimators, one obtains the same result as discussed in Section 3 if averages of $X_t X_t'$ are approximately the same in all parts of the sample, $(rT - sT)^{-1}\sum_{t=sT+1}^{rT} X_t X_t' \xrightarrow{p} \Sigma_X$, for all $0 \le s < r \le 1$. Under this homogeneity assumption, $\hat{\Sigma}_X^{-1}X_t X_t'$ averages in large samples to I_k when computing Fourier transforms (2), or cosine transforms (6), for any fixed l. Thus, $\hat{y}_t$ behaves just like the demeaned series $\tilde{y}_t - T^{-1}\sum_{s=1}^T \tilde{y}_s \approx \tilde{y}_t - (\hat{\beta}_1 - \beta_1)$. Consequently, $\hat{\omega}^2_{p,n}$, $\hat{\omega}^2_{Y,q}$, or $\hat{\omega}^2_{\mathrm{KVB}}$ computed from $y_t = \hat{y}_t$ are asymptotically identical to the infeasible estimators computed from $y_t = \tilde{y}_t$, as the underlying weights are all orthogonal to a constant. One may thus rely on the same asymptotically justified critical values; for instance, under weak dependence, the t-statistic $\sqrt{T}(\hat{\beta}_1 - \beta_1)/\hat{\omega}_{Y,q}$ is asymptotically Student-t with q degrees of freedom.

The appropriate generalization of the Ibragimov and Müller (2010) approach to a regression context requires q estimations of the regression on the q blocks of data, followed by the computation of a simple t-statistic from the q estimators of β_1.

For the tests S_q derived in Section 4, suppose the hypothesis to be tested is $H_0 : \beta_1 = \beta_{1,0}$. One would expect that under the null hypothesis, the product of the OLS residuals of a regression of $R_t - \iota_1'X_t\beta_{1,0}$ and $\iota_1'X_t$ on PX_t, respectively, where the $k \times (k-1)$ matrix P collects the last $k-1$ columns of I_k, forms a mean-zero series, at least approximately. Some linear regression algebra shows that this product, scaled by $1/\iota_1'\hat{\Sigma}_X^{-1}\iota_1$, may equivalently be written as

$$y_t = \iota_1'\hat{\Sigma}_X^{-1}X_t\hat{e}_t + \frac{\iota_1'\hat{\Sigma}_X^{-1}X_t X_t'\hat{\Sigma}_X^{-1}\iota_1}{\iota_1'\hat{\Sigma}_X^{-1}\iota_1}(\hat{\beta}_1 - \beta_{1,0}). \tag{15}$$

Thus, the suggestion is to compute (15), followed by the implementation described in Section 4.3. If $X_t = 1$, this reduces to what is suggested there for inference about a population mean.

The construction of the tests S_q assumes y_t to have low-frequency dynamics that resemble those of a Gaussian AR(1) with coefficient possibly close to one, and that alternatives correspond to mean shifts of y_t. This might or might not be a useful approximation under (15), depending on the long-run properties of X_t and e_t. For instance, if X_t and e_t are scalar independent stationary AR(1) processes with the same coefficient close to unity, then y_t in (15) follows a stationary AR(1) with a slightly smaller coefficient, but it is not Gaussian, and incorrect values of $\beta_{1,0}$ do not amount to translation shifts of y_t. Neither the validity nor the optimality of the tests S_q thus goes through as such. One could presumably derive weighted average power maximizing tests that are valid by construction for any particular assumption of this sort. But in practice, it is difficult to specify strong parametric restrictions for the joint long-run behavior X_t and e_t. And for any given realization X_t, y_t in (15) may still follow essentially any mean-zero process via a sufficiently peculiar conditional distribution of $\{e_t\}_{t=1}^T$ given $\{X_t\}_{t=1}^T$. Finally, under the weak dependence and homogeneity assumptions that justify inconsistent estimators in a linear regression context, Y_l for $l \ge 1$ computed from $\hat{y}_t = \iota_1'\hat{\Sigma}_X^{-1}X_t\hat{e}_t$, and Y_l computed from y_t in (15), are asymptotically equivalent, and they converge to mean-zero-independent Gaussian variates of the same variance as $\sqrt{T}(\hat{\beta}_1 - \beta_1)$. Since iid Gaussian Y_l corresponds to the special case of $c \to \infty$ in the analysis of Section 4, the tests S_q are thus also valid under such conditions. So as practical matter, the tests S_q under definition (15) might still be useful to improve the quality of small sample inference, even if the underlying assumptions are unlikely to be met exactly for finite c in a regression context.

Now consider a potentially overidentified GMM problem. Let θ be the $k \times 1$ parameter, and suppose the hypothesis of interest concerns the scalar parameter $\theta_1 = \iota_1'\theta$, $H_0 : \theta_1 = \theta_{1,0}$. Let $\hat\theta$ be the $k \times 1$ GMM estimator, based on the $r \times 1$ moment condition $g_t(\theta)$ with $r \times k$ derivative matrix $\hat G = T^{-1}\sum_{t=1}^T \partial g_t(\theta)/\partial\theta'|_{\theta=\hat\theta}$ and $r \times r$ weight matrix W. In this notation, the appropriate definition for $\hat y_t$ is

$$\hat y_t = -\iota_1'(\hat G' W \hat G)^{-1}\hat G' W g_t(\hat\theta).$$

For the analogue of y_t, define $\hat\theta^0$ as the GMM estimator under the constraint $\theta_1 = \theta_{1,0}$, and $\hat G^0 = T^{-1}\sum_{t=1}^T \partial g_t(\theta)/\partial\theta'|_{\theta=\hat\theta^0}$. Then set

$$y_t = -\iota_1'(\hat G^{0\prime} W \hat G^0)^{-1}\hat G^{0\prime} W g_t(\hat\theta^0).$$

These definitions reduce to what is described above for the special case of a linear regression.

6. SMALL SAMPLE COMPARISON

This section contains some evidence on the small sample size and power performance of the different approaches to HAC corrections.

In all simulations, the sample size is $T = 200$. The first two simulations concern the mean of a scalar time series. In the "AR(1)" design, the data are a stationary Gaussian AR(1) with coefficient ρ and unit innovation variance. In the "AR(1) + Noise" design, the data are the sum of such an AR(1) process and independent Gaussian white noise of variance 4.

Table 3 reports small sample size and size-adjusted power of the same 12 two-sided tests of 5% nominal level that were reported in Table 2 of Section 4.3 in the unemployment illustration. Due to the much smaller sample size in the simulation, the parametric estimator $\hat\omega^2_{\mathrm{AR}}$ is based on 4 lags, however. The size-adjustment is performed on the ratio of test statistic and critical value; this ensures that data-dependent critical values are appropriately subsumed in the effective test.

The AR(1) design is exactly the data-generating process for which the test S_q was derived, and correspondingly, size control is almost exact. All other approaches lead to severely oversized tests as ρ becomes large. (All LRV estimators $\hat\omega^2$ considered here are translation invariant, so that for fixed T, they have a well-defined limiting distribution as $\rho \to 1$. At the same time, for any M, $P(|\hat\mu - \mu| > M) \to 1$ as $\rho \to 1$. Thus, all LRV-based tests have arbitrarily poor size control for ρ sufficiently close to one, and the same also holds for the IM_q tests.) By construction,

Table 3. Small sample performance for inference about population mean

ρ	S_{12}	S_{24}	S_{48}	$\hat\omega^2_{\mathrm{A91}}$	$\hat\omega^2_{\mathrm{AM}}$	$\hat\omega^2_{\mathrm{AR(4)}}$	$\hat\omega^2_{\mathrm{KVB}}$	$\hat\omega^2_{\mathrm{Y,12}}$	$\hat\omega^2_{\mathrm{Y,24}}$	$\hat\omega^2_{\mathrm{SPJ}}$	IM_8	IM_{16}
Size under AR(1)												
0.0	4.7	4.9	5.0	5.5	5.5	6.7	5.0	5.2	5.2	5.1	5.1	5.1
0.7	4.9	4.9	5.0	10.0	7.0	8.1	6.1	6.3	8.8	5.5	6.0	8.1
0.9	5.0	4.8	5.3	17.2	10.7	12.0	8.9	13.8	24.7	6.6	10.5	19.1
0.95	5.0	5.0	5.1	25.5	15.3	16.8	13.0	25.8	41.3	8.8	18.6	32.8
0.98	4.9	4.7	5.0	44.2	26.5	27.7	23.0	48.3	62.0	13.0	37.3	54.3
0.999	4.8	4.6	4.5	87.7	68.1	68.7	71.2	88.1	92.0	44.5	84.3	89.7
Size-adjusted power under AR(1)												
0.0	35.7	42.8	47.1	50.0	49.9	47.6	37.3	44.3	47.4	48.8	40.1	45.8
0.7	34.8	41.7	44.8	46.4	47.0	45.1	36.2	45.1	47.8	39.9	41.5	46.4
0.9	28.8	34.5	35.6	42.7	41.4	41.9	34.7	46.4	48.5	31.6	42.7	46.7
0.95	21.0	23.2	25.3	40.3	38.5	38.6	33.9	46.9	48.5	27.9	43.6	47.0
0.98	11.3	12.5	12.2	40.0	38.1	37.6	35.1	51.4	52.9	26.9	47.2	51.4
0.999	5.6	5.6	5.6	92.6	74.3	73.9	80.0	98.5	99.0	53.6	97.4	98.5
Size under AR(1)+Noise												
0.0	4.6	4.7	4.8	5.3	5.3	6.4	4.8	5.1	4.7	4.8	4.9	5.0
0.7	4.9	5.2	5.9	12.6	13.1	8.1	5.9	5.8	7.4	6.4	5.8	7.1
0.9	5.2	5.7	9.5	27.7	37.5	14.5	8.8	12.8	23.1	10.6	10.0	17.8
0.95	5.3	6.7	12.6	38.0	52.4	21.1	12.8	25.5	40.1	15.9	18.3	31.8
0.98	5.4	7.1	15.1	52.5	67.9	33.6	23.1	47.9	61.5	26.3	37.3	53.8
0.999	5.4	7.4	17.7	86.2	91.9	73.6	71.1	88.1	91.9	68.0	84.2	89.6
Size-adjusted power under AR(1)+Noise												
0.0	35.9	43.9	48.6	51.3	51.0	48.9	37.5	44.8	49.5	50.2	40.9	46.5
0.7	34.4	41.9	46.2	48.9	49.6	46.6	35.8	45.5	48.3	46.0	41.4	46.8
0.9	28.9	35.8	39.0	46.1	48.6	43.0	34.5	46.8	48.9	42.6	42.3	47.6
0.95	21.3	23.5	26.7	43.2	47.2	39.9	34.0	47.6	49.4	39.2	43.6	48.3
0.98	11.8	12.0	12.6	43.0	48.8	39.5	35.5	51.5	53.5	37.7	47.9	51.3
0.999	5.54	5.56	5.53	85.6	91.0	76.8	80.7	98.7	99.1	72.2	97.3	98.6

NOTES: Entries are rejection probability in percent of nominal 5% level tests. Under the alternative, the population mean differs from the hypothesized mean by $2T^{-1/2}(1-\rho)^{-1}$ and $2T^{-1/2}(4+(1-\rho)^{-2})^{1/2}$, respectively. Based on 20,000 replications.

Table 4. Small sample performance for inference about regression coefficient

ρ	S_{12}	S_{24}	S_{48}	$\hat{\omega}^2_{\mathrm{A91}}$	$\hat{\omega}^2_{\mathrm{AM}}$	$\hat{\omega}^2_{\mathrm{AR(4)}}$	$\hat{\omega}^2_{\mathrm{KVB}}$	$\hat{\omega}^2_{\mathrm{Y,12}}$	$\hat{\omega}^2_{\mathrm{Y,24}}$	$\hat{\omega}^2_{\mathrm{SPJ}}$	IM_8	IM_{16}
Size under scalar nonconstant regressor												
0.0	4.7	4.9	5.0	5.9	6.0	7.1	5.1	5.4	5.5	5.5	5.0	5.1
0.7	5.1	5.0	5.1	10.2	8.0	9.6	6.7	7.3	8.2	7.3	5.3	5.5
0.9	5.2	4.6	4.4	17.9	12.8	14.5	10.9	12.2	17.2	10.3	5.7	5.8
0.95	5.0	4.3	4.2	25.4	18.1	20.2	15.4	19.3	28.5	13.8	5.6	5.5
0.98	4.2	3.5	3.6	36.4	26.0	28.2	22.4	31.2	43.2	19.0	5.2	5.2
0.999	2.9	2.3	2.6	51.6	37.0	39.5	33.6	47.1	58.9	26.3	4.8	5.2
Size-adjusted power under scalar nonconstant regressor												
0.0	47.9	59.1	64.6	68.9	68.7	66.3	51.9	62.1	66.0	67.2	53.7	54.6
0.7	36.0	44.8	49.0	49.9	49.4	48.4	38.6	46.4	49.9	45.5	45.8	53.3
0.9	30.6	36.8	39.6	41.0	39.1	40.1	32.8	41.2	43.1	34.9	53.4	73.6
0.95	27.5	31.3	33.1	40.5	36.6	38.0	33.4	41.0	42.9	32.5	70.3	91.2
0.98	25.9	29.7	29.8	44.6	39.3	40.5	38.0	46.4	48.7	34.6	93.2	99.7
0.999	31.2	37.4	36.5	93.6	87.4	87.6	89.1	95.3	95.8	81.5	100	100
Size under four-dimensional nonconstant regressor												
0.0	4.8	5.1	5.2	6.0	6.0	7.1	5.2	5.7	5.6	5.6	4.9	4.8
0.7	5.9	5.8	5.8	11.0	8.8	10.5	7.5	8.1	8.6	8.0	5.3	5.3
0.9	7.2	7.0	6.8	20.3	15.6	17.6	12.7	14.7	18.8	12.9	5.4	5.1
0.95	8.4	8.2	8.1	28.7	22.1	24.6	18.3	22.1	28.5	17.9	5.2	4.9
0.98	10.7	10.3	10.2	38.6	30.3	33.0	25.6	31.4	39.6	24.9	4.9	4.6
0.999	13.4	12.9	12.7	45.5	36.3	39.9	31.7	38.3	46.2	30.7	4.7	4.7
Size-adjusted power under four-dimensional nonconstant regressor												
0.0	47.2	57.9	63.5	67.9	67.6	64.9	51.4	60.8	64.8	66.2	47.9	41.7
0.7	35.5	44.5	47.9	49.3	49.0	47.9	37.4	45.9	48.6	45.2	45.3	48.9
0.9	31.7	38.4	40.2	42.4	40.4	41.2	34.6	41.7	43.8	37.2	64.8	80.2
0.95	30.9	36.9	38.0	44.4	41.6	42.8	36.6	44.0	45.9	38.2	85.2	95.9
0.98	34.0	40.4	41.4	55.6	52.3	53.1	45.6	55.7	57.8	46.5	98.8	99.9
0.999	47.2	56.8	57.5	99.6	98.4	98.4	97.4	99.6	99.8	96.8	100	100

NOTES: Entries are rejection probability in percent of nominal 5% level tests. Under the alternative, the population regression coefficient differs from the hypothesized coefficient by $2.5T^{-1/2}(1-\rho^2)^{-1/2}$. Based on 20,000 replications.

the tests S_q come close to maximizing weighted average power. Yet their size-adjusted power is often substantially below those of other tests. This is no contradiction, as size adjustment is not feasible in practice; a size-adjusted test that rejects if, say, $|\hat{\mu}-\mu|$ is large is, of course, as good as the oracle test that uses a critical value computed from knowledge of ρ.

The addition of an independent white-noise process to an AR(1) process translates the AR(1) spectral density upward. The peak at zero when ρ is large is then "hidden" by the noise, making appropriate corrections harder. This induces size distortions in all tests, including those derived here. The distortions of S_q are quite moderate for $q = 12$, but more substantive for larger q. Assuming the AR(1) approximation to hold over a wider range of frequencies increases power, but, if incorrect, induces more severe size distortions.

The second set of simulations concerns inference about a scalar regression coefficient. The regressions all contain a constant, and the nonconstant regressors and regression disturbances are independent mean-zero Gaussian AR(1) processes with common coefficient ρ and unit innovation variance. The parameter of interest is the coefficient on the first nonconstant regressor. Table 4 reports the small sample performance of the same set of tests, implemented as described in Section 4, for a scalar nonconstant regressor, and a four-dimensional nonconstant regressor. The tests S_q continue to control size well, at least with a single nonconstant regressor. Most other tests overreject substantively for sufficiently large ρ.

The marked exception is the IM_q test, which has outstanding size and power properties in this design. As a partial explanation, note that if the regression errors follow a random walk, then the q estimators of the parameter of interest from the q blocks of data are still exactly independent and conditionally Gaussian, since the block-specific constant terms of the regression soak up any dependence that arises through the *level* of the error term. The result of Bakirov and Székely (2005) reviewed in Section 3 thus guarantees coverage for both $\rho = 0$ and $\rho \to 1$. (I thank Don Andrews for this observation). At the same time, this stellar performance of IM_q is partially due to the specific design; for instance, if the zero-mean AR(1) regression error is multiplied by the regressor of interest (which amounts to a particular form of heteroscedasticity), unreported results show IM_q to be severely oversized, while S_q continues to control size much better.

7. CONCLUSION

Different forms of HAC corrections can lead to substantially different empirical conclusions. For instance, the lengths of standard 95% confidence intervals for the population mean of the U.S. unemployment rate reported in Section 4.3 vary by a factor of three. It is therefore important to understand the rationale of different approaches.

There are good reasons to be skeptical of methods that promise to automatically adapt to any given dataset. All inference requires some a priori knowledge of exploitable regularities. The more explicit and interpretable these driving assumptions, the easier it is to make sense of empirical results.

In my view, the relatively most interpretable way of expressing regularity for deriving HAC corrections is in spectral terms. Under an assumption that the spectral density is flat over a thin frequency band around the origin, an attractive approach is to perform inference with the estimator derived in Müller (2007). For instance, for empirical analyses involving macroeconomic variables, a natural starting point is the assumption that the spectrum is flat below business cycle frequencies. Roughly speaking, this means that business cycles are independent from one another. With a business cycle frequency cut-off of 8 years, this suggests using Müller's (2007) estimator with the number of cosine weighted averages equal to the span of the sample in years divided by four.

This article derives an alternative approach, where the regularity consists of the assumption that an AR(1) model provides a good approximation to the spectral density over a thin frequency band. Flatness of the spectral density over this band is covered as a special case, so that the resulting inference is strictly more robust than that based on Müller's (2007) estimator. It is not obvious over which frequencies one would necessarily want to make this assumption, but the numerical evidence of this article suggests the particular test S_{24} to be a reasonable default.

Especially if second moment instabilities are a major concern, another attractive approach is to follow the suggestion of Ibragimov and Müller (2010). The regularity condition there is that estimating the model on consecutive blocks of data yields approximately independent and Gaussian estimators of the parameter of interest. In contrast to other approaches to inconsistent LRV estimation, no assumption about the homogeneity of second moments is required for this method. Under an assumption of a flat spectrum below 8 year cycles, it makes sense to choose equal-length blocks of approximately 10 years of data.

Econometrics only provides a menu of inference methods derived under various assumptions. Monte Carlo exercises, such as the one in Section 6 above, are typically performed with very smooth spectra in simple parametric families, and it remains unclear to which extent their conclusions about the "empirically" best HAC corrections are relevant for applied work. Ultimately, researchers in the field have to judge which set of regularity conditions makes the most sense for a specific problem.

APPENDIX

Approximation (5)

Since for $|j| < T - 1$, $\frac{2}{T}\sum_{l=1}^{(T-1)/2}\cos(2\pi jl/T)p_l = \hat{\gamma}(j) + \hat{\gamma}(T - |j|)$ (cf. equation (7.6.10) in Priestley (1981)), one can write $\hat{\omega}^2_{k,S_T} = \sum_{l=1}^{(T-1)/2} K_{T,l}\,p_l - \sum_{j=-T+1}^{T-1} k(j/S_T)\hat{\gamma}(T - j)$. With $S_T \ll T$, the remainder term $\sum_{j=-T+1}^{T-1} k(j/S_T)\hat{\gamma}(T - j)$ is typically small, as $\hat{\gamma}(T - j)$ is small for $j \ll T$, and $k(j/S_T)$ is small for $j \gg S_T$.

Furthermore, note that for $0 < |j| < T$, $\sum_{l=1}^{(T-1)/2}\cos(2\pi jl/T) = -1/2$, so that $\sum_{l=1}^{(T-1)/2} K_{T,l} = (T - 1)/T - 2T^{-1}\sum_{j=1}^{T-1} k(j/S_T) \approx 1$.

Equation (8)

Note that with $\phi_l(t)$ equal to either $\sqrt{2}\cos(\pi l(t - 1/2)/T)$ or $\sqrt{2}\sin(\pi lt/T)$, $\sum_{t=1}^{T}\phi_j(t)\phi_l(t) = \mathbf{1}[l = j]T$ for $j, l = 1, \ldots, T - 1$. Thus, $y_t - \hat{\mu} = \sqrt{2}T^{-1/2}\sum_{l=1}^{T}\cos(\pi l(t - 1/2)/T)Y_l$, and by a direct calculation, $\sum_{s=1}^{t}(y_s - \bar{y}) = T^{-1/2}\sum_{l=1}^{T}\frac{\sqrt{2}\sin(\pi lt/T)}{2\sin(\pi l/(2T))}Y_l$, so that $\hat{\omega}^2_{KVB} = T^{-2}\sum_{t=1}^{T}(\sum_{s=1}^{t}(y_s - \bar{y}))^2 = \sum_{l=1}^{T}\kappa_{T,l}Y_l^2$.

ACKNOWLEDGMENTS

The author thanks Graham Elliott and Mark Watson for helpful comments, and Liyu Dou for excellent research assistance. Support by the National Science Foundation through grants SES-0751056 and SES-1226464 is gratefully acknowledged.

[Received March 2014. Revised June 2014.]

REFERENCES

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [311,312]

Andrews, D. W. K., and Monahan, J. C. (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966. [317]

Atchade, Y. F., and Cattaneo, M. D. (2012), "Limit Theorems for Quadratic Forms of Markov Chains," Working Paper, University of Michigan, available at *http://arxiv.org/pdf/1108.2743.pdf*. [311]

Bakirov, N. K., and Székely, G. J. (2005), "Student's T-Test for Gaussian Scale Mixtures," *Zapiski Nauchnyh Seminarov POMI*, 328, 5–19. [315,320]

Berk, K. N. (1974), "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics*, 2, 489–502. [311,312]

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (2nd ed.), New York: Springer. [312,313]

Chan, N. H., and Wei, C. Z. (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063. [315]

den Haan, W. J., and Levin, A. T. (1997), "A Practitioner's Guide to Robust Covariance Matrix Estimation," in *Handbook of Statistics* (Vol. 15), eds. G. S. Maddala, and C. R. Rao, Amsterdam: Elsevier, pp. 299–342. [312]

Elliott, G., Müller, U. K., and Watson, M. W. (2012), "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," Working Paper, Princeton University, available at *www.princeton.edu/∼umueller/nuisance.pdf*. [311,316]

Gonalves, S., and Vogelsang, T. (2011), "Block Bootstrap and HAC Robust Tests: The Sophistication of the Naive Bootstrap," *Econometric Theory*, 27, 745–791. [311]

Granger, C. W. J., and Newbold, P. (1974), "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120. [311]

Ibragimov, R., and Müller, U. K. (2010), "T-Statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28, 453–468. [314,318,321]

Jansson, M. (2004), "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72, 937–946. [311]

Kariya, T. (1980), "Locally Robust Test for Serial Correlation in Least Squares Regression," *The Annals of Statistics*, 8, 1065–1070. [316]

Kiefer, N. M., and Vogelsang, T. J. (2002a), "Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size," *Econometric Theory*, 18, 1350–1366. [311]

——— (2002b), "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation," *Econometrica*, 70, 2093–2095. [314]

——— (2005), "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130–1164. [311,314]

Kiefer, N. M., Vogelsang, T. J., and Bunzel, H. (2000), "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 695–714. [311,314]

King, M. L. (1980), "Robust Tests for Spherical Symmetry and Their Application to Least Squares Regression," *The Annals of Statistics*, 8, 1265–1271. [316]

——— (1987), "Towards a Theory of Point Optimal Testing," *Econometric Reviews*, 6, 169–218. [315]

Lehmann, E. L., and Romano, J. P. (2005), *Testing Statistical Hypotheses*, New York: Springer. [316]

Müller, U. K. (2004), "A Theory of Robust Long-Run Variance Estimation," Working paper, Princeton University. [311,314]

——— (2007), "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141, 1331–1352. [311,314,318,321]

——— (2011), "Efficient Tests Under a Weak Convergence Assumption," *Econometrica*, 79, 395–435. [315]

Müller, U. K., and Watson, M. W. (2008), "Testing Models of Low-Frequency Variability," *Econometrica*, 76, 979–1016. [314]

——— (2013), "Measuring Uncertainty About Long-Run Forecasts," Working Paper, Princeton University. [315]

Newey, W. K., and West, K. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708. [311,312]

Phillips, P. (2005), "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142. [311,314]

Phillips, P., Sun, Y., and Jin, S. (2006), "Spectral Density Estimation and Robust Hypothesis Testing Using Steep Origin Kernels Without Truncation," *International Economic Review*, 47, 837–894. [311]

Phillips, P., Sun, Y., and Jin, S. (2007), "Long Run Variance Estimation and Robust Regression Testing Using Sharp Origin Kernels With No Truncation," *Journal of Statistical Planning and Inference*, 137, 985–1023. [311]

Phillips, P. C. B. (1987), "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547. [315]

Robinson, P. M. (2005), "Robust Covariance Matrix Estimation: HAC Estimates With Long Memory/Antipersistence Correction," *Econometric Theory*, 21, 171–180. [315]

Stock, J., and Watson, M. (2011), *Introduction to Econometrics* (3rd ed.), Boston: Addison Wesley. [312]

Sun, Y. (2013), "Heteroscedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator," *The Econometrics Journal*, 16, 1–26. [314]

Sun, Y., and Kaplan, D. M. (2012), "Fixed-Smoothing Asymptotics and Accurate F Approximation Using Vector Autoregressive Covariance Matrix Estimator," Working Paper, University of California, San Diego. [311]

Sun, Y., Phillips, P. C. B., and Jin, S. (2008), "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76, 175–794. [311]

Whittle, P. (1957), "Curve and Periodogram Smoothing," *Journal of the Royal Statistical Society*, Series B, 19, 38–63. [313]

——— (1962), "Gaussian Estimation in Stationary Time Series," *Bulletin of the International Statistical Institute*, 39, 105–129. [313]

Comment

Nicholas M. Kiefer

Departments of Economics and Statistical Science, Cornell University, Ithaca, NY 14853 and CREATES, University of Aarhus, Aarhus, Denmark (*nicholas.kiefer@cornell.edu*)

1. INTRODUCTION

Müller looks at the problems of interval estimation and hypothesis testing in autocorrelated models from a frequency domain point of view. This leads to good insights as well as proposals for new methods. The new methods may be more robust than existing approaches, though this is more suggested than firmly established. This discussion begins with a speculative overview of the problem and the approach. The theme is that the issue involved is essentially the choice of a conditioning ancillary. Then I turn, perhaps more usefully, to some specific technical comments. Finally, I agree wholeheartedly with the general point that comes through clearly: the more we know or are willing to assume about the underlying process the better we can do.

2. ASYMPTOTICS AND CONDITIONING

The point of asymptotic theory is sometimes lost, especially when new approaches are being considered. The goal is to find a manageable approximation to the sampling distribution of a statistic. The approximation should be as accurate as possible. The assumptions needed to develop the asymptotics are not a model of any actual physical process.

The "trick" is to model the rate of information accumulation leading to the asymptotic approximation, so that the resulting limit distribution can be calculated and as accurately as possible mimics the sampling distribution of interest. There are many ways to do this. These are not "correct" or "incorrect," just different models. What works?

One way to frame the choice of assumptions is as specification of an ancillary statistic. An example will make this specific. Suppose we are estimating μ and the sufficient statistic is S. Suppose S can be partitioned into $(\hat{\mu}, a)$ with a ancillary. With data y sufficiency implies the factorization

$$p(y|\mu) = g(y)p(S|\mu)$$

and ancillarity implies

$$p(S|\mu) = p(\hat{\mu}|a, \mu)p(a).$$

The key is choosing a so its distribution does not depend on μ—or in the local case, does not depend "much." See Christensen and Kiefer (1994). S may have the dimension of the dataset.

It is widely agreed—mostly from examples, not theorems—that inference can (and perhaps should) be based on the conditional distribution. See Barndorff-Nielsen (1984), Berger et al. (1988), and the review by Reid (1995). In the normal mean model, we could set $a = (s^2, a')$ and condition on a, obtaining normal inference, or condition on a' alone obtaining the t. With autocorrelation, $a = (\hat{\rho}, s^2, a') = (\psi, a')$ and conditioning on a

Journal of Business & Economic Statistics
July 2014, Vol. 32, No. 3
DOI: 10.1080/07350015.2014.926816