A Theory of Robust Long-Run Variance Estimation

Ulrich K. Müller*
Princeton University
Economics Department
Princeton, NJ, 08544

Abstract

Long-run variance estimation can typically be viewed as the problem of estimating the scale of a limiting continuous time Gaussian process on the unit interval. A natural benchmark model is given by a sample that consists of equally spaced observations of this limiting process. The paper analyzes the asymptotic robustness of long-run variance estimators to contaminations of this benchmark model. It is shown that any equivariant long-run variance estimator that is consistent in the benchmark model is highly fragile: there always exists a sequence of contaminated models with the same limiting behavior as the benchmark model for which the estimator converges in probability to an arbitrary positive value. A class of robust inconsistent long-run variance estimators is derived that optimally trades off asymptotic variance in the benchmark model against the largest asymptotic bias in a specific set of contaminated models.

JEL classification: C22, C13

Keywords: Heteroskedasticity and Autocorrelation Consistent Variance Estimation, Functional Central Limit Theorem, Bias, Qualitative Robustness

*The author thanks two referees and an anonymous associate editor for constructive and thoughtful comments and suggestions, as well as Mark Watson, Chris Sims and Graham Elliott and seminar participants at the universities of Cornell, Montreal, Pennsylvania, Houston, North Carolina State, Michigan, Oxford, Berlin, Munich, Vienna, Harvard and Boston for helpful discussions. Financial support by the NSF through grant SES-0518036 is gratefully acknowledged.

1 Introduction

The long-run variance ω^2 plays a major role in much of time series inference, such as in regression inference with autocorrelated disturbances, unit root testing or inference with fractional time series. Usually, for second-order stationary processes, the long-run variance is defined as the sum of all autocovariances, or, equivalently, in terms of the spectrum at frequency zero. This paper takes on a different perspective: the starting point is the observable (double-array) scalar sequence $\{u_{T,t}\}_{t=1}^T$, which satisfies

$$u_{T,[\cdot T]} \Rightarrow \omega G(\cdot), \tag{1}$$

where $[\cdot]$ is the greatest lesser integer function, '$\Rightarrow$' denotes weak convergence as $T \to \infty$ and $G(\cdot)$ is a mean zero, almost surely continuous Gaussian process on the unit interval with known continuous and non-degenerate covariance kernel $k(r,s) = E[G(r)G(s)]$. Long-run variance estimation can then be understood as estimation of the (asymptotic) scale of the process $u_{T,[\cdot T]}$. In the context of the OLS regression $y_t = X_t'\beta + \nu_t$ with $\beta = (\beta_1, \cdots, \beta_q)'$, for instance, assume that $T^{-1}\sum_{t=1}^{[sT]} X_t X_t' \xrightarrow{p} s\Sigma_X$ uniformly in $s \in [0,1]$ for some positive definite $q \times q$ matrix Σ_X and that $\{X_t \nu_t\}$ satisfies a Functional Central Limit Theorem (where '$\xrightarrow{p}$' indicates convergence in probability and all limits are takes as $T \to \infty$, if not indicated otherwise). Then $T^{1/2}(\hat{\beta}_1 - \beta_1) \Rightarrow \mathcal{N}(0, \omega^2)$, and the first element of $(T^{-1}\sum_{t=1}^T X_t X_t')^{-1/2} T^{-1/2} \sum_{t=1}^{[\cdot T]} X_t \hat{\nu}_t$ with $\hat{\nu}_t = y_t - X_t'\hat{\beta}$ converges weakly to a Brownian Bridge of scale ω. If instead the regressors contain a time trend or other slowly varying deterministic terms, such as a dummy for a structural break that occurs at a known fixed fraction of the sample, then G is no longer a Brownian Bridge in general, but its covariance kernel is still known.

This paper studies the robustness of long-run variance estimators $\hat{\omega}_T^2$ of ω^2 that are functions of the $T \times 1$ vector $u_T = (u_{T,1}, \cdots, u_{T,T})'$ in an asymptotic framework. We focus on scale equivariant estimators, i.e. on estimators satisfying $\hat{\omega}_T^2(cu_T) = c^2 \hat{\omega}_T^2(u_T)$ for all u_T. Since the units of economic data are typically arbitrary, a restriction to scale equivariant estimators makes sense to ensure coherent results, and all standard long-run variance estimators are scale equivariant. The benchmark model for these estimators is where $u_{T,t} \sim G(t/T)$ for $t = 1, \cdots, T$, i.e. $u_T \sim \mathcal{N}(0, \Sigma_T)$ with $\Sigma_T = [k(s/T, t/T)]_{s,t}$.

We consider the asymptotic robustness of long-run variance estimators $\hat{\omega}_T^2$ in contaminated models $\tilde{u}_T = (\tilde{u}_{T,1}, \cdots, \tilde{u}_{T,T})'$, $\tilde{u}_T \sim \mathcal{N}(0, \tilde{\Sigma}_T)$ with $\tilde{\Sigma}_T$ in some sense close to Σ_T for all (large enough) T. As a motivating example, suppose G is a standard Wiener process W, so that the scaled first differences $T^{1/2}\Delta u_{T,t}$ in (1) satisfy a Functional Central Limit Theorem, and the benchmark model $u_T \sim \mathcal{N}(0, \Sigma_T)$ has $T^{1/2}\Delta u_{T,t}$ distributed as Gaussian White Noise of unit variance. Now consider the contaminated model where $T^{1/2}\Delta\tilde{u}_{T,t}$ follows a Gaussian first order autoregressive process of unit long-run variance, with a root ρ_T that is local-to-unity, i.e. $\rho_T = 1 - \gamma/T$ for fixed $\gamma > 0$. If γ is large (say, $\gamma = 50$), then $T^{1/2}\Delta\tilde{u}_{T,t}$ exhibits strong mean reversion, $\tilde{\Sigma}_T$ is close to Σ_T, and (1) provides a reasonable approximation also for the contaminated model. The asymptotic robustness of a long-run variance estimator would ensure that accordingly, the estimation of the scale of the processes $u_{T,[\cdot T]}$ and $\tilde{u}_{T,[\cdot T]}$ yield similar results, at least for T large. By the usual asymptotic motivation of small sample inference, this in turn suggests that the robust estimator has reasonable properties in a sample of, say, $T = 250$ observations with $\rho_T = 0.8$. In many empirical applications, little is known about the dynamic properties of $T^{1/2}\Delta u_{T,t}$ in (1). It therefore makes sense to require asymptotic robustness over a large set of contaminated models that are close to satisfying $\tilde{u}_{T,[\cdot T]} \Rightarrow \omega G(\cdot)$, in the hope that the set contains one element which provides a good approximation to the actual small sample dynamics.

In Section 2, we establish that *any* scale equivariant long-run variance estimator that is consistent for ω^2 in the benchmark model is highly fragile to contaminations of this kind. In particular, there exists a sequence of contaminated covariance matrices $\tilde{\Sigma}_T$ such that $\tilde{u}_T \sim \mathcal{N}(0, \tilde{\Sigma}_T)$ satisfies $\tilde{u}_{T,[\cdot T]} \Rightarrow \omega G(\cdot)$, yet $\hat{\omega}_T^2(\tilde{u}_T)$ converges in probability to an arbitrary positive value. In Section 3, we derive the form of long-run variance estimators that optimally trade off bias control in a specific set of contaminated models against the variance of the estimator at the benchmark model in the class of all long-run variance estimators that can be written as quadratic forms in u_T. These optimal long-run variance estimators are inconsistent, with a nondegenerate limiting distribution proportional to ω^2 even in the benchmark model. Section 4 presents Monte Carlo evidence on the small sample performance of various long-run variance estimators for two standard data generating processes. Proofs are collected in an appendix.

Most papers that consider robust (in the sense of Huber) inference in time series models are concerned with contaminating outliers, rather than contaminating autocorrelations this paper focusses on. Kleiner, Martin, and Thomson (1979) and Bhansali (1997), for instance, develop spectral density estimators that are robust against contaminating outliers. The general approaches to robust time series inference developed in Künsch (1984) and Martin and Yohai (1986) could in principle be employed to consider robust long-run variance estimation; but they are based on benchmark models with a parametrized dependence structure, and the contaminations these authors consider are substantially different from those analyzed here. The 'nonparametric' starting point (1) of this paper makes it more akin to the work of Hosoya (1978) and Samarov (1987), who consider robust time series forecasting and linear regression inference where contaminated models have spectral density functions that are close to the spectral density function of a known benchmark model.

The large majority of the numerous papers on robust long-run variance estimators use the term 'robustness' in the nonparametric/adaptive sense: They show how to consistently estimate the long-run variance with minimal conditions on moments and dependence properties of the underlying process. See, for instance, Hannan (1957) and Berk (1974) for early contributions, or Newey and West (1987) and Andrews (1991) for popular implementations and Robinson and Velasco (1997) for a survey. Typically, the assumptions in these papers imply a Functional Central Limit Theorem to hold for the underlying disturbances, such that the partial sums of the observed residuals satisfy (1) with G a Brownian Bridge. Robinson (1994, 2005) applies similar methods to fractional time series, such that G based on the residuals becomes a fractional Brownian Bridge. The covariance kernel k of G then depends on the self-similarity index, which is typically unknown. With k unknown, the results of Section 2 concerning the lack of robustness of consistent long-run variance estimators hold a fortiori; the robust long-run variance estimators derived in Section 3, however, crucially depend on knowledge of k.

A large body of work has demonstrated that inference based on consistent long-run variance estimators performs poorly in small samples with strong dependence and heterogeneity, see den Haan and Levin (1997) for a survey. Kiefer, Vogelsang, and Bunzel (2000) have pointed out that it is possible to conduct asymptotically justified inference in a linear time series regression based long-run variance estimators with a nondegenerate limiting distribution, and find that the resulting approximation of the distribution of test statistics leads to better small sample size control in some models. These results were extended to the class of kernel estimators with a bandwidth that is a fixed fraction of the sample size in Kiefer and Vogelsang (2002, 2005). One way to analytically understand these results is to consider higher order expansions of the distribution of test statistics and rejection probabilities; Jansson (2004) and Sun, Phillips, and Jin (2006) find that indeed, in a Gaussian location model, a certain class of quadratic long-run variance estimators leads to an order of magnitude smaller errors in rejection probability than certain consistent long-run variance estimators.

This paper provides a framework to study first order properties of long-run variance estimators in a set of contaminated models. The result of Section 2 shows the fragility of consistent long-run variance estimators in the class of processes that satisfy (1), exposing an inherent limitation of a strategy of adaptive consistent long-run variance estimation. One contribution of this paper is thus an alternative analytical justification for considering time series inference procedures based on inconsistent long-run variance estimators, and the argument presented here is applicable to *all* equivariant consistent long-run variance estimators.

The analysis in Section 3 considers the problem of robust estimation of the long-run variance. This is of immediate interest in contexts where the value of the long-run variance itself is important; the long-run variance of the first differences of an integrated time series describes, for instance, the uncertainty of long-range forecasts. Also the intra-day volatility of the price of financial assets corresponds to the long-run variance of their returns, which at very high frequencies are contaminated by micro-marketstructure noise; see Andersen, Bollerslev, and Diebold (2005) for a survey. In other contexts, the long-run variance estimator is only one element of the inference procedure; think of test statistics concerning the value β_1 in the linear regression example above, unit root tests or parameter stability tests. The validity of these inference procedures depends on the long-run variance estimator to have reasonable properties. The lack of qualitative robustness of consistent long-run variance estimators established in Section 2 hence typically translates into lack of robustness of these procedures—see Müller (2004) for related results on unit root and stationarity tests. Also, it is plausible that such procedures benefit from 'plugging-in' the robust long-run variance estimators determined in Section 3, but the issue is not pursued further. The derivation of robust methods for more general inference problems than the estimation of the long-run variance is an interesting question and is left to future research.

2 The Lack of Qualitative Robustness of Consistent Long-Run Variance Estimators

The deviation of the contaminated model $\tilde{u}_T \sim \mathcal{N}(0, \tilde{\Sigma}_T)$ from the benchmark model $u_T \sim \mathcal{N}(0, \Sigma_T)$ is wholly determined by the difference between the two covariance matrices $\tilde{\Sigma}_T$ and Σ_T. We measure the extent of the difference by two norms on the vector space of $T \times T$ matrices: on the one hand $||A||_\Delta = \max_{i,j} |a_{i,j}|$ and on the other hand $||A||_2$, the square root of the largest eigenvalue of $A'A$. These norms induce two neighborhoods of contaminated models, identified by their covariance matrix

$$\begin{aligned} \mathfrak{C}_T^2(\delta) &= \{\tilde{\Sigma}_T : T^{-1}||\tilde{\Sigma}_T - \Sigma_T||_2 \leq \delta\} \\ \mathfrak{C}_T^\Delta(\delta) &= \{\tilde{\Sigma}_T : ||\tilde{\Sigma}_T - \Sigma_T||_\Delta \leq \delta\}. \end{aligned}$$

Since $||A||_2 \leq T||A||_\Delta$, $\mathfrak{C}_T^\Delta(\delta) \subset \mathfrak{C}_T^2(\delta)$ for any $\delta \geq 0$.

A leading case for long-run variance estimation occurs where G in (1) is a standard Wiener process W, so that $k(r,s) = r \wedge s$. Set-ups that lead to this case include instances where a time series is modelled as being integrated, or instances where a Functional Central Limit Theorem applies to some data $\{a_{T,t}\}_{t=1}^T$, such that $T^{-1/2}\sum_{t=1}^{[\cdot T]} a_{T,t} \Rightarrow \omega W(\cdot)$. For the latter case, the contamination neighborhoods $\{\mathfrak{C}_T^\Delta(\delta)\}_{T\geq 1}$ include the following double-array processes for all large T:

1. A local-to-unity Gaussian AR(1) process in the sense of Chan and Wei (1987) and Phillips (1987), i.e. $a_{T,0} = 0$, $a_{T,t} = \rho_T a_{T,t-1} + (1-\rho_T)\varepsilon_t$ for $\varepsilon_t \sim$i.i.d.$\mathcal{N}(0,1)$

and $T(1-\rho_T)=\gamma$ for a fixed $\gamma>0$. With $\tilde{u}_{T,t}=T^{-1/2}\sum_{s=1}^{t}a_{T,s}$, the $[rT]$,$[sT]$th element of $\tilde{\Sigma}_T$ for $r\leq s$ then converges uniformly to $r-\gamma^{-1}(1+e^{-\gamma(s-r)})(1-e^{-\gamma r})+(2\gamma)^{-1}e^{-\gamma(s-r)}(1-e^{-2\gamma r})$ as $T\to\infty$, which in turn converges to r uniformly as $\gamma\to\infty$.

2. Gaussian White Noise with a relatively low frequency seasonal component, i.e. $a_{T,t}=Z\sigma_Z T^{-1/2}\sin(2\pi\zeta t/T)+\varepsilon_t$ for $Z,\varepsilon_t\sim$i.i.d.$\mathcal{N}(0,1)$ and fixed $\zeta>0$. For $r\leq s$, the $[rT]$,$[sT]$th element of $\tilde{\Sigma}_T$ then converges uniformly to $r+\sigma_Z^2(\zeta\pi)^{-2}(\sin(\zeta\pi r))^2(\sin(\zeta\pi s))^2$ as $T\to\infty$, which in turn converges to r uniformly as $\zeta\to\infty$ for any fixed σ_Z (and also as $\sigma_Z\to 0$ for fixed ζ).

3. Gaussian White Noise with a Gaussian outlier at date $t=[\tau T]$ for fixed $0<\tau<1$, i.e. $a_{T,t}=Z\sigma_Z T^{1/2}\mathbf{1}(t=[\tau T])+\varepsilon_t$, where $\mathbf{1}(\cdot)$ is the indicator function. For $r\leq s$, the $[rT]$,$[sT]$th element of $\tilde{\Sigma}_T$ then converges uniformly to $r+\sigma_Z^2\mathbf{1}(r\geq\tau)(r-\tau)(s-\tau)$ as $T\to\infty$, which converges to r uniformly as $\sigma_Z\to 0$.

For fixed δ, the partial sums of all of these processes are hence elements of $\mathfrak{C}_T^{\Delta}(\delta)$ for sufficiently large γ and ζ and sufficiently small σ_Z, respectively, at least for large enough T. Inference for ω^2 that remains robust to all contaminations $\mathfrak{C}_T^{\Delta}(\delta)$ therefore guards against the impact of strong autocorrelation (example 1), a large peak in spectral density close to the origin (example 2) and outliers (example 3).

Given that $\mathfrak{C}_T^{\Delta}(\delta)\subset\mathfrak{C}_T^{2}(\delta)$, all the examples are also elements of $\mathfrak{C}_T^{2}(\delta)$. An interesting element of $\mathfrak{C}_T^{2}(\delta)$ that is not an element of $\mathfrak{C}_T^{\Delta}(\delta)$ for fixed δ uniformly in T arises from an integrated almost non-invertible MA(1) process, i.e. $a_{T,t}=Th^{-1}(\varepsilon_t-\theta_T\varepsilon_{t-1})$ with $\varepsilon_0=0$ and $T(1-\theta_T)=h$ for fixed $h>0$. Then $\tilde{u}_{T,t}=T^{-1/2}\sum_{s=1}^{t-1}\varepsilon_s+T^{1/2}h^{-1}\varepsilon_t$, such that the $[rT]$,$[sT]$th element of $\tilde{\Sigma}_T$ converges uniformly to $(r\wedge s)+h^{-1}$ for $[rT]\neq[sT]$ and to $r+Th^{-2}$ for $[rT]=[sT]$ as $T\to\infty$. For large enough h and T, the process is hence element of $\mathfrak{C}_T^{2}(\delta)$. The problem of estimating the scale of the permanent component in $\{\tilde{u}_{T,t}\}_{t=1}^{T}$ in such a model arises, for instance, when assessing the extent of instabilities of linear regression models—see Stock and Watson (1998).

Given these examples, it seems desirable that long-run variance estimators are not too fragile in the neighborhood $\mathfrak{C}_T^{2}(\delta)$, or at least $\mathfrak{C}_T^{\Delta}(\delta)$, for small enough δ and large T. But no long-run variance estimator that is consistent in the benchmark model possesses this feature.

Theorem 1 *If a scale equivariant long-run variance estimator* $\hat{\omega}_T^2$ *satisfies* $\hat{\omega}_T^2(u_T) \xrightarrow{p} 1$ *when* $u_T \sim \mathcal{N}(0, \Sigma_T)$*, then for any* $c > 0$ *there exists a sequence* $\tilde{u}_T \sim \mathcal{N}(0, \tilde{\Sigma}_T)$*,* $\tilde{\Sigma}_T \in \mathfrak{C}_T^{\Delta}(\delta_T)$ *with* $\delta_T \to 0$ *satisfying* $\sup_{1 \leq t \leq T} |u_{T,t} - \tilde{u}_{T,t}| \to 0$ *a.s., yet* $\hat{\omega}_T^2(\tilde{u}_T) \xrightarrow{p} c^2$.

For any $\delta > 0$, long-run variance estimators that are consistent in the benchmark model necessarily lack robustness in $\{\mathfrak{C}_T^{\Delta}(\delta)\}_{T \geq 1}$ (and hence $\{\mathfrak{C}_T^{2}(\delta)\}_{T \geq 1}$) for T large. Even highly nonparametric estimators of the long-run variance fail to reasonably estimate the scale of $\tilde{u}_{T,[\cdot T]}$ for some contaminated model $\tilde{\Sigma}_T$, despite the fact that $||\tilde{\Sigma}_T - \Sigma_T||_{\Delta} \to 0$ and $\sup_{1 \leq t \leq T} |u_{T,t} - \tilde{u}_{T,t}| \to 0$ a.s. Given the extreme extent of the fragility, this typically implies a corresponding fragility of more general inference procedures that rely on consistent long-run variance estimators. In particular, stationarity tests, unit root tests or Wald tests of linear regression coefficients that are based on consistent long-run variance estimators have arbitrarily bad asymptotic size control in the contamination neighborhoods $\{\mathfrak{C}_T^{\Delta}(\delta)\}_{T \geq 1}$ for any $\delta > 0$.

Theorem 1 does not imply that consistent long-run variance estimators yield arbitrary results for all specific contaminations in $\mathfrak{C}_T^{\Delta}(\delta)$ and $\mathfrak{C}_T^{2}(\delta)$, such as the four examples discussed above. Rather, it asserts the existence of one arbitrarily small contamination (as measured by $||\Sigma_T - \tilde{\Sigma}_T||_{\Delta}$) that induces arbitrary properties. To gain some insight into the nature of this contamination, consider the special case where G is a Brownian Bridge, so that in the benchmark model, $T^{1/2}\Delta u_{T,t}$ (with $u_{T,0} = 0$) is demeaned Gaussian White Noise of unit variance.

A Brownian Bridge of scale $c > 0$, $B_0(s) \sim cW(s) - csW(1)$, admits the representation (see, for instance, Phillips (1998))

$$B_0(s) = c \sum_{l=1}^{\infty} \frac{\sqrt{2}\sin(\pi l s)}{l\pi} \xi_l \tag{2}$$

where $\xi_l \sim$i.i.d.$\mathcal{N}(0,1)$ and the right hand side converges almost surely and uniformly on $s \in [0,1]$. For $n \geq 1$, define

$$B_n(s) = \sum_{l=1}^{n} \frac{\sqrt{2}\sin(\pi l s)}{l\pi} \xi_l + c \sum_{l=n+1}^{\infty} \frac{\sqrt{2}\sin(\pi l s)}{l\pi} \xi_l.$$

By the consistency of $\hat{\omega}_T^2$ in the benchmark model and scale equivariance, $\hat{\omega}_T^2(u_T^0) \xrightarrow{p} c^2$ when $u_{T,t}^0 \sim B_0(t/T)$, $t \leq T$, $T \geq 1$. The difference between the processes B_n and

B_0 is that the first n components of B_n are of relative scale $1/c$, which may be cast as a difference in the variance of n scalar independent Gaussian variables. The measure of B_n is thus absolutely continuous with respect to the measure of B_0 for any fixed n, which implies that $\hat{\omega}_T^2(u_T^0) \xrightarrow{p} c^2$ entails $\hat{\omega}_T^2(u_T^n) \xrightarrow{p} c^2$, too, where $u_{T,t}^n \sim B_n(t/T)$, $t \leq T$, $T \geq 1$. One can therefore construct a sequence $n_T \to \infty$ satisfying $\hat{\omega}_T^2(\tilde{u}_T) \xrightarrow{p} c^2$, where $\tilde{u}_{T,t} \sim B_{n_T}(t/T)$, $t \leq T$, $T \geq 1$. By the convergence of the right-hand side of (2), $B_{n_T}(s)$ converges to a Brownian Bridge B_∞ of unit scale uniformly on $s \in [0,1]$ almost surely, so that with $u_{T,t} \sim B_\infty(t/T)$ for $t \leq T$, $T \geq 1$, $\sup_{1 \leq t \leq T} |u_{T,t} - \tilde{u}_{T,t}| \to 0$ a.s., and also $\tilde{\Sigma}_T \in \mathfrak{C}_T^\Delta(\delta_T)$ with $\delta_T \to 0$.

Since $T^{1/2}\Delta u_{T,t}$ is distributed as demeaned Gaussian White Noise of unit variance, and $\{\sqrt{2/T}\cos(\pi l(t-1/2)/T)\}_{t=1}^T$, $l = 1, \cdots, T-1$ are the last $T-1$ elements of the orthonormal type II discrete cosine transform, one can deduce[1] that for $c < 1$ and $n \leq T$

$$T^{1/2}\Delta u_{T,t}^n \sim \sqrt{\frac{2}{T}} \sum_{l=1}^{n} \alpha_l \cos(\pi l(t-1/2)/T)\xi_l + c\sqrt{\frac{2}{T}} \sum_{l=n+1}^{T} \cos(\pi l(t-1/2)/T)\xi_l \quad (3)$$

where $\alpha_l^2 = c^2 + (1-c^2)(2T\sin(\pi l/2T)/l\pi)^2 \to 1$ for any fixed l. While $\{T^{1/2}\Delta u_{T,t}^n\}_{t=1}^T$ is not stationary, one might usefully think of $\{T^{1/2}\Delta u_{T,t}^n\}_{t=1}^T$ as a demeaned, approximately stationary Gaussian series with a piece-wise constant spectral density that equals $1/2\pi$ for frequencies of absolute value smaller than $n\pi/T$ (so that the long-run variance is unity) and $c^2/2\pi$ for frequencies of absolute value larger than $n\pi/T$. For $c \ll 1$, such a spectral density is a very rough approximation of the typical spectral shape of economic time series as estimated by Granger (1966), with substantially more spectral mass at low frequencies compared to higher frequencies.

Table 1 describes the behavior of Andrews' (1991) quadratic spectral long-run variance estimator $\hat{\omega}_{QA}^2$ with automatic bandwidth selection based on an AR(1) model for disturbances with distribution (3) for $c = 1/2$ and various n and T, where $n = \infty$ denotes the benchmark model. Along with the 10th and 90th percentile of the empirical

[1] With $\tilde{\xi}_l$ ~i.i.d.$\mathcal{N}(0,1)$ independent of $\{\xi_l\}_{l=1}^\infty$, write $B_n(s) \sim B_0(s) + \sqrt{1-c^2}\sum_{l=1}^n \sqrt{2}\sin(l\pi s)\tilde{\xi}_l/(l\pi)$, and note that $\{\sqrt{2/T}\sum_{l=1}^T \cos(\pi l(t-1/2)/T)\xi_l\}_{t=1}^T$ is distributed as demeaned Gaussian White Noise of unit variance. The result now follows from $\sin(l\pi t/T) - \sin(l\pi(t-1)/T) = 2\cos(\pi l(t-1/2)/T)\sin(\pi l/2T)$ and some rearranging.

Table 1: Behaviour of a Consistent Long-Run Variance under Contamination

	$T = 120$				$T = 240$			
n	5	10	20	∞	5	10	20	∞
10th perc $\hat{\omega}^2_{QA}(u^n_T)$	0.233	0.278	0.425	0.773	0.233	0.257	0.323	0.840
90th perc $\hat{\omega}^2_{QA}(u^n_T)$	0.499	0.761	1.206	1.218	0.369	0.476	0.731	1.156
10th perc sup $\|u_{T,t} - u^n_{T,t}\|$	0.138	0.107	0.080	0	0.146	0.115	0.087	0
90th perc sup $\|u_{T,t} - u^n_{T,t}\|$	0.222	0.163	0.117	0	0.229	0.170	0.123	0
$\|\|\Sigma_T - \Sigma^n_T\|\|_\Delta$	0.019	0.010	0.005	0	0.019	0.010	0.005	0
	$T = 480$				$T = 960$			
n	5	10	20	∞	5	10	20	∞
10th perc $\hat{\omega}^2_{QA}(u^n_T)$	0.235	0.247	0.277	0.886	0.237	0.244	0.258	0.920
90th perc $\hat{\omega}^2_{QA}(u^n_T))$	0.311	0.355	0.461	1.112	0.284	0.302	0.346	1.081
10th perc sup $\|u_{T,t} - u^n_{T,t}\|$	0.151	0.120	0.092	0	0.155	0.124	0.095	0
90th perc sup $\|u_{T,t} - u^n_{T,t}\|$	0.235	0.175	0.128	0	0.239	0.179	0.132	0
$\|\|\Sigma_T - \Sigma^n_T\|\|_\Delta$	0.019	0.010	0.005	0	0.019	0.010	0.005	0

cumulative distribution function of $\hat{\omega}^2_{QA}(u^n_T)$, the table contains the 10th and 90th percentile of the cumulative distribution function of $\sup_{1\leq t\leq T}|u_{T,t} - u^n_{T,t}|$, based on 50,000 replications, and $||\Sigma_T - \Sigma^n_T||_\Delta$, where Σ^n_T denotes the covariance matrix of u^n_T. Theorem 1 and the above discussion implies that for any consistent long-run variance estimator, which of course includes $\hat{\omega}^2_{QA}$, for large enough T there exists n that make $\hat{\omega}^2_T(u^n_T) \xrightarrow{p} c^2$, $\sup_{1\leq t\leq T}|u_{T,t} - u^n_{T,t}| \to 0$ a.s and $||\Sigma_T - \Sigma^n_T||_\Delta \to 0$ accurate approximations. One might say that this is achieved to the greatest extent by $T = 960$ and $n = 10$. More generally, though, $\hat{\omega}^2_{QA}$ has a substantial negative bias for many n and T that are of potential empirical relevance: With 40 years of monthly data (so that $T = 480$), for instance, a value of $n = 10$ approximates a stationary series with twice as much variation below business cycle frequencies (periods larger than 8 years) compared to higher frequency variation. Results not reported here show that for the values of n and T in Table 1, the properties of $\hat{\omega}^2_{QA}$ are essentially the same when the underlying disturbances are exactly stationary and Gaussian with piece-wise constant spectral density $f_{\Delta u}(\lambda) = \mathbf{1}[|\lambda| < n\pi/T]/2\pi + \mathbf{1}[|\lambda| \geq n\pi/T]c^2/2\pi$.

From $\sup_{1 \leq t \leq T} |u_{T,t} - \tilde{u}_{T,t}| \to 0$ a.s. and $\hat{\omega}_T^2(\tilde{u}_T) \xrightarrow{p} c^2$, it follows from Theorem 1 that any consistent long-run variance estimator is necessarily a discontinuous function of $u_{T,[\cdot T]}$, i.e. sample paths $\{u_{T,t}\}_{t=1}^T$ and $\{\tilde{u}_{T,t}\}_{t=1}^T$ that are close in the sup norm do not in general lead to similar long-run variance estimates. Consistent long-run variance estimators might therefore be called 'qualitatively fragile', in analogy to Hampel's (1971) definition that requires qualitatively robust estimators in an i.i.d. setting to be continuous functionals of the empirical cumulative distribution function.

Inadequate behavior of estimators of the spectral density at a given point under certain circumstances has been established before—see, for instance, Sims (1971), Faust (1999) or Pötscher (2002). These papers show the impossibility of obtaining correct confidence intervals for the spectral density at a given point for any sample size when the underlying parametric structure of a time series model is too rich in some sense. Loosely speaking, this literature demonstrates that meaningful inference is impossible in too generously parametrized models, as the relevant convergences do not hold uniformly over the parameter space.

Theorem 1 is different, since it only shows the fragility of long-run variance estimators that are consistent in the benchmark model. Given that $\sup_{1 \leq t \leq T} |u_{T,t} - \tilde{u}_{T,t}| \to 0$ a.s. implies $\tilde{u}_{T,[\cdot T]} \Rightarrow G(\cdot)$, any long-run variance estimator that can be written as a continuous functional of the set of continuous functions on the unit interval is 'qualitatively robust'. What is more, Theorem 2 below demonstrates that it is possible to derive long-run variance estimators that are asymptotically robust in $\mathfrak{C}_T^2(\delta)$ (and hence $\mathfrak{C}_T^\Delta(\delta)$) for small enough δ. Rather than being a statement about the impossibility of valid inference, Theorem 1 shows that a certain class of estimators (those that are consistent in the benchmark model) are necessarily highly fragile.

In the special case where $G \sim W$, the double array of the scaled first differences of $\{u_{T,t}\}_{t=1}^T$ and $\{\tilde{u}_{T,t}\}_{t=1}^T$ of Theorem 1, $\{T^{1/2}\Delta u_{T,t}\}_{t=1}^T$ and $\{T^{1/2}\Delta \tilde{u}_{T,t}\}_{t=1}^T$, satisfy a Functional Central Limit Theorem. Advances in the literature have continuously diminished the wedge between the primitive (on the underlying disturbances) assumptions for Functional Central Limit Theorems and the primitive assumptions for consistent long-run variance estimation (see, for instance, de Jong and Davidson (2000) for a recent contribution). But Theorem 1 reveals that this wedge is of substance: The set of all (double array) processes that satisfy a Functional Central Limit Theorem is strictly larger than the set of all processes that in addition allow consistent estimation of the scale of the limiting Wiener process.

3 Quantitatively Robust Long-Run Variance Estimators

This section derives long-run variance estimators that are asymptotically robust to small contaminations of the form described by $\mathfrak{C}_T^{\Delta}(\delta)$ and $\mathfrak{C}_T^2(\delta)$. As in much of the robustness literature, we focus on the largest asymptotic bias

$$\begin{aligned}\gamma_2(\delta) &= \overline{\lim}_{T\to\infty} \sup_{\tilde{u}_T\sim\mathcal{N}(0,\tilde{\Sigma}_T),\tilde{\Sigma}_T\in\mathfrak{C}_T^2(\delta)} \left|E[\hat{\omega}_T^2(\tilde{u}_T)]-1\right| \\ \gamma_\Delta(\delta) &= \overline{\lim}_{T\to\infty} \sup_{\tilde{u}_T\sim\mathcal{N}(0,\tilde{\Sigma}_T),\tilde{\Sigma}_T\in\mathfrak{C}_T^\Delta(\delta)} \left|E[\hat{\omega}_T^2(\tilde{u}_T)]-1\right|\end{aligned}$$

as the quantitative measures of robustness. Since for any given δ, $\mathfrak{C}_T^{\Delta}(\delta)\subset\mathfrak{C}_T^2(\delta)$, a finite $\gamma_2(\delta)$ also implies bounded $\gamma_\Delta(\delta)$. Note that these measures are relative to the scale of $\tilde{u}_T$: for $\tilde{u}_T\sim\mathcal{N}(0,\omega^2\tilde{\Sigma}_T)$, the largest asymptotic biases of $\hat{\omega}_T^2$ are given by $\omega^2\gamma_2(\delta)$ and $\omega^2\gamma_\Delta(\delta)$.

From Theorem 1, it immediately follows that any non-negative consistent long-run variance estimator has infinite $\gamma_\Delta(\delta)$ and $\gamma_2(\delta)$. The aim of this section is hence to identify robust inconsistent long-run variance estimators.

For this purpose, we consider the class of *quadratic long-run variance estimators*, defined as estimators of the form

$$\hat{\omega}_T^2(u_T)=u_T'A_Tu_T$$

for some positive semi-definite and data independent $T\times T$ matrix A_T with s,t element $a_T(s,t)$ that satisfies $\lim_{T\to\infty}\operatorname{tr}(\Sigma_TA_T)=1$. The normalization ensures asymptotic unbiasedness of this scale equivariant estimator when $u_T\sim\mathcal{N}(0,\omega^2\Sigma_T)$. Note that (possibly after an additional scale normalization) the class of quadratic long-run variance estimators includes the popular kernel estimators

$$\hat{\omega}_\kappa^2 = \sum_{l=-T+1}^{T-1} \kappa(l/b_T)\hat{\gamma}(l) \tag{4}$$

where $\hat{\gamma}(l)$ is the sample autocovariance of $T^{1/2}\Delta u_{T,t}$, i.e. $\hat{\gamma}(l) = \sum_{t=1}^{T-|l|} \Delta u_{T,t+|l|}\Delta u_{T,t}$, κ is a symmetric kernel with $\kappa(0) = 1$ and nonnegative corresponding spectral window generator and b_T is a data independent bandwidth. A popular choice for κ is the Bartlett kernel $\kappa(x) = \mathbf{1}[|x| < 1](1-x)$—see Newey and West (1987). Andrews (1991) has shown that, if $b_T \to \infty$ and $b_T = o(T)$, kernel estimators are consistent for a wide range of underlying disturbances. Theorem 1 implies that all these estimators lack qualitative robustness and have unbounded largest asymptotic bias. In fact, as demonstrated by Müller (2005), these estimators consistently estimate a long-run variance of zero in the local-to-unity example in Section 2 above for any amount of mean reversion $\gamma > 0$.

We thus focus in the following on kernel estimators with a bandwidth that is a fixed fraction of the sample size, $b_T = bT$ for some $b \in (0, 1]$. These 'fixed-b' estimators $\hat{\omega}_\kappa^2(b)$ have been studied by Kiefer and Vogelsang (2002, 2005) for the special case where G is a Brownian Bridge $G(s) \sim W(s) - sW(1)$. In order to satisfy $\lim_{T\to\infty} \operatorname{tr}(\Sigma_T A_T) = 1$, fixed-$b$ estimators require an additional scale normalization: with κ twice continuously differentiable, define

$$\hat{\omega}_{2d}^2(b) = \frac{\hat{\omega}_\kappa^2}{k(1,1) + 2b^{-1}\int \kappa'((1-s)/b)k(1,s)ds - b^{-2}\int\int \kappa''((r-s)/b)k(r,s)drds}$$

where here and in the following, the limits of integration are zero and one, if not indicated otherwise, and for the fixed-b Bartlett estimator $\hat{\omega}_{BT}^2(b)$

$$\hat{\omega}_{BT}^2(b) = \frac{\hat{\omega}_\kappa^2}{k(1,1) + 2b^{-1}(\int_0^1 k(s,s)ds - \int_b^1 k(s,s-b)ds - \int_0^b k(1,1-s)ds)}.$$

See the appendix for details.

For quadratic long-run variance estimators, the bias in a contaminated model with covariance matrix $\tilde{\Sigma}_T$ is given by $\operatorname{tr}((\Sigma_T - \tilde{\Sigma})A_T) + o(1)$, so that it is easy to see that

$$\sup_{\tilde{u}_T \sim \mathcal{N}(0,\tilde{\Sigma}_T), \tilde{\Sigma}_T \in \mathfrak{C}_T^2(\delta)} \left| E[\hat{\omega}_T^2(\tilde{u}_T)] - 1 \right| = \delta T \operatorname{tr} A_T + o(1)$$

where the worst case contamination in $\mathfrak{C}_T^2(\delta)$ is given by $\tilde{\Sigma}_T = \Sigma_T + \delta T I_T$. Quadratic long-run variance estimators with finite $\gamma_2(\delta)$ thus in particular limit the distortionary effect of severe classical measurement error in $\tilde{u}_T$, a feature with potential appeal for, say, the estimation of the volatility of asset returns over short periods of time in the presence of micro-marketstructure noise. For contaminations in $\mathfrak{C}_T^\Delta(\delta)$, we obtain

$$\sup_{\tilde{u}_T \sim \mathcal{N}(0,\tilde{\Sigma}_T), \tilde{\Sigma}_T \in \mathfrak{C}_T^\Delta(\delta)} \left| E[\hat{\omega}_T^2(\tilde{u}_T)] - 1 \right| \leq \delta \sum_{s=1}^T \sum_{t=1}^T |a_T(s,t)| + o(1).$$

Typically, the maximal bias is achieved by the worst case contamination $\tilde{\Sigma}_T = \Sigma_T + \delta S_T \in \mathfrak{C}_T^\Delta(\delta)$, where S_T has elements $S_T(s,t) = \text{sign}(a_T(s,t))$, although S_T might not be positive semi-definite, in which case $\gamma_\Delta(\delta)$ depends on Σ_T and δ. The following results abstract from these complications and focus on the 'generic' asymptotic maximal bias $\bar{\gamma}_\Delta(\delta) = \delta \overline{\lim}_{T\to\infty} \sum_{s=1}^T \sum_{t=1}^T |a_T(s,t)|$.

Theorem 2 *(i) Let φ_l and r_l with $r_1 \geq r_2 \geq \ldots$, $l = 1, 2, \ldots$ be a set of continuous eigenfunctions and eigenvalues of $k(r,s) = E[G(r)G(s)]$, and define $\hat{\xi}_l = r_l^{-1/2} T^{-1} \sum_{t=1}^T \varphi_l(t/T) \tilde{u}_{T,t}$. Among all quadratic long-run variance estimators, the class of estimators indexed by a real number $\lambda > r_1^{-1}$*

$$\hat{\omega}_{RE}^2(\lambda) = \sum_{l=1}^{p(\lambda)} w_l(\lambda) \hat{\xi}_l^2$$

with $p(\lambda)$ the largest l such that $\lambda > r_l^{-1}$ and $w_l(\lambda) = (\lambda - r_l^{-1}) / \sum_{j=1}^{p(\lambda)} (\lambda - r_j^{-1})$ minimizes $\gamma_2(\delta)$ subject to an efficiency constraint $\overline{\lim}_{T\to\infty} E[(\hat{\omega}_T^2(u_T) - 1)^2] \leq \varsigma$ for $u_T \sim \mathcal{N}(0, \Sigma_T)$, and achieves $\gamma_2(\delta) = \delta \sum_{l=1}^{p(\lambda)} w_l(\lambda) r_l^{-1}$ and $\bar{\gamma}_\Delta(\delta) = \delta \int \int | \sum_{l=1}^{p(\lambda)} w_l(\lambda) r_l^{-1} \varphi_l(s) \varphi_l(r) | ds dr$.

(ii) Let $\tau \in \arg\max_{s \in [0,1]} k(s,s)$. Then

$$\hat{\omega}_{R\Delta}^2 = \frac{\tilde{u}_{T,[\tau T]}^2}{k(\tau,\tau)}$$

minimizes $\bar{\gamma}_\Delta(\delta)$ over all quadratic long-run variance estimators, and achieves $\bar{\gamma}_\Delta(\delta) = \delta / k(\tau,\tau)$ and $\gamma_2(\delta) = \infty$.

(iii) Let $\tilde{u}_T$ be any sequence $T = 1, 2, \ldots$ of contaminated models that can be written as $\tilde{u}_T + \tilde{\eta}_T = u_T + \eta_T$ a.s., where $u_T \sim \mathcal{N}(0, \Sigma_T)$, $\tilde{\eta}_T \sim (0, \tilde{V}_T)$, $\eta_T \sim (0, V_T)$, $\tilde{\eta}_T$ is independent of $\tilde{u}_T$, η_T is independent of u_T, and $T^{-1}||\tilde{V}_T||_2 \leq \delta$ and $T^{-1}||V_T||_2 \leq \delta$ uniformly in T. Then for any quadratic long-run variance estimator $\hat{\omega}_T^2$

$$\overline{\lim}_{T\to\infty} E[|\hat{\omega}_T^2(\tilde{u}_T) - \hat{\omega}_T^2(u_T)|] \leq 2(\gamma_2(\delta) + \sqrt{\gamma_2(\delta)} + \sqrt{\gamma_2(\delta) + \gamma_2(\delta)^2}).$$

(iv) For fixed-b kernel estimators with twice differentiable kernel κ,$\gamma_2(\delta) = \infty$ and

$$\bar{\gamma}_\Delta(\delta) = \delta \frac{b^2 + 2b \int |\kappa'((1-s)/b)|ds + \int\int |\kappa''((r-s)/b)|drds}{b^2 k(1,1) + 2b \int \kappa'((1-s)/b)k(1,s)ds - \int\int \kappa''((r-s)/b)k(r,s)drds}$$

and for the Bartlett fixed-b estimator, $\gamma_2(\delta) = \infty$ and

$$\bar{\gamma}_\Delta(\delta) = \frac{b+4}{k(1,1)b + 2\int_0^1 k(s,s)ds - 2\int_b^1 k(s,s-b)ds - 2\int_0^b k(1,1-s)ds}$$

Part (i) of Theorem 2 follows a strategy initially suggested by Hampel (1968) as described in Huber (1996), and used by Künsch (1984) and Martin and Zamar (1993), among others: In a class of estimators and for a given contamination neighborhood, the maximal asymptotic bias is minimized subject to a bound on the asymptotic variance *in the benchmark model*. Just as the bias measure, this (imperfect) measure of asymptotic efficiency is relative to the scale of u_T: $E[(\hat{\omega}_T^2(u_T) - 1)^2] \leq \varsigma$ for $u_T \sim \mathcal{N}(0, \Sigma_T)$ corresponds to $E[(\hat{\omega}_T^2(u_T)/\omega^2 - 1)^2] \leq \varsigma$ for $u_T \sim \mathcal{N}(0, \omega^2\Sigma_T)$. When $r_1^{-1} < \lambda \leq r_2^{-1}$, one obtains the most robust long-run variance estimator $\hat{\omega}_{R2}^2 = \hat{\xi}_1^2$, with asymptotic bias $\gamma_2(\delta) = \delta r_1^{-1}$. As λ increases, more weight is put on the efficiency of the estimator in the benchmark model, leading to estimators that are a weighted average of a finite number of $\hat{\xi}_l^2$, $l = 1, \cdots, p(\lambda)$, with less weight on $\hat{\xi}_l^2$ for l large. These efficient long-run variance estimators cannot be written as kernel estimators (4).

The intuition for the result in part (i) is as follows: Among all square-integrable functions f on the unit interval that satisfy $\int f(s)G(s)ds \sim \mathcal{N}(0,1)$, $f = r_1^{-1/2}\varphi_1$ minimizes $\int f(s)^2 ds$. This property makes $\hat{\xi}_1$ least susceptible to contaminations described by $\mathfrak{C}_T^2(\delta)$ asymptotically, as the differences in the covariance matrices are as little amplified as possible. A requirement of a lower variance in the benchmark model forces exploitation of an additional weighted average of $\{\tilde{u}_{T,t}\}$, and among all square integrable functions f on the unit interval that satisfy $\int f(s)G(s)ds \sim \mathcal{N}(0,1)$ independent of $r_1^{-1/2}\int \varphi_1(s)G(s)ds$, $f = r_2^{-1/2}\varphi_2$ minimizes $\int f(s)^2 ds$, and so forth.

Part (ii) of Theorem 2 identifies the quadratic long-run variance estimator that minimizes the generic maximal bias $\bar{\gamma}_\Delta(\delta)$ under $\mathfrak{C}_T^\Delta(\delta)$ contaminations. The derivation of quadratic long run variance estimators that efficiently trade off $\bar{\gamma}_\Delta(\delta)$ against the asymptotic variance in the benchmark model seems difficult and is not attempted here.

Part (iii) of Theorem 2 shows that controlling the largest asymptotic bias $\gamma_2(\delta)$ of quadratic long-run variance estimators $\hat{\omega}_T^2(u_T)$ in the set of contaminated models with $\tilde{u}_T = u_T + \eta_T - \tilde{\eta}_T$ implies an asymptotic uniform upper bound on the amount of distortion in the *distribution* of $\hat{\omega}_T^2(\tilde{u}_T)$ compared to the benchmark model. Note that this set of contaminated models contains $\tilde{u}_T \sim \mathcal{N}(0,\tilde{\Sigma}_T)$ with $\tilde{\Sigma}_T \in \mathfrak{C}_T^2(\delta)$: let $P\Lambda P'$ be the spectral decomposition of $\tilde{\Sigma}_T - \Sigma_T$, and write $\Lambda = \Lambda^+ + \Lambda^-$, where Λ^+ and Λ^- contain only nonnegative and nonpositive elements, respectively. Letting $\eta_T \sim \mathcal{N}(0, P\Lambda^+P')$ and $\tilde{\eta}_T \sim \mathcal{N}(0, P\Lambda^- P')$ yields $E\tilde{u}_T\tilde{u}_T' = \Sigma_T + P(\Lambda^+ - \Lambda^-)P' = \tilde{\Sigma}_T$, and the claim follows. Also, this distributional robustness allows for some departure from Gaussianity in the contaminated model, as η_T and $\tilde{\eta}_T$ are only assumed to have the specified first and second moments.

Noting that $r_l^{-1/2}\int \varphi_l(s)G(s)ds \sim$i.i.d.$\mathcal{N}(0,1)$, the application of the Continuous Mapping Theorem yields that the asymptotic distribution of $\hat{\omega}_{RE}^2(\lambda)$ is given by a weighted average of independent chi-squared random variables, scaled by ω^2, whenever $\tilde{u}_{T,[\cdot T]} \Rightarrow \omega^2 G(\cdot)$. Applying Theorem 2 (iii) further shows that inference for ω^2 based on $\hat{\omega}_{RE}^2(\lambda)$ using this distributional assumption remains asymptotically accurate for the examples of contaminated models given above Theorem 1 for δ small, that is for large enough γ, ζ and h and small enough σ_Z.

Part (iv) of Theorem 2 show that fixed-b kernel estimators are not robust to $\mathfrak{C}_T^2(\delta)$, just as $\hat{\omega}_{R\Delta}^2$. In terms of the first differences $T^{1/2}\Delta\tilde{u}_{T,t}$, the worst case contamination in $\mathfrak{C}_T^2(\delta)$, $\tilde{\Sigma}_T = \Sigma_T + \delta T I_T$, corresponds to the addition of a non-invertible MA(1) error of variance δT^2 to the first differences $T^{1/2}\Delta u_{T,t}$. Lack of robustness of kernel estimators thus suggests relatively poor performance in underlying models for $T^{1/2}\Delta\tilde{u}_{T,t}$ that are close approximations to a noninvertible MA(1) process.

As one might expect, long-run variance estimators have different robustness properties in different contamination neighborhoods. Ideally, the contamination neighborhood should reflect uncertainty over potential models in a given application. At the same time, Theorem 2 points to $\hat{\omega}^2_{RE}(\lambda)$ as an attractive default class of estimators: The robustness of $\hat{\omega}^2_{RE}(\lambda)$ extends over a very large neighborhood (that includes $\mathfrak{C}^{\Delta}_T(\delta)$), and it is not limited to the first moment of its asymptotic distribution.

4 Monte Carlo Evidence

We now turn to a numerical analysis of the performance of various long-run variance estimators in small samples for two standard data generating processes. Specifically, we consider the estimation of the long-run variance of Gaussian first order autoregressive and moving average processes

$$\begin{aligned} \text{AR}(1) &: \quad a_t = \rho a_{t-1} + (1-\rho)\varepsilon_t \\ \text{MA}(1) &: \quad a_t = (1-\theta)^{-1}(\varepsilon_t - \theta\varepsilon_{t-1}) \end{aligned}$$

with $a_0 = \varepsilon_0 = 0$ and $\varepsilon_t \sim$i.i.d.$\mathcal{N}(0,1)$, such that $\omega^2 = 1$. Under standard asymptotics with ρ and θ fixed, the partial sum process $T^{-1/2}\sum_{t=1}^{[\cdot T]} a_t$ converges weakly to a standard Wiener process W, such that G in (1) corresponds to $G \sim W$. The benchmark model is thus $u_{T,t} = T^{-1/2}\sum_{s=1}^{t}\varepsilon_s$ with $\varepsilon_t \sim$i.i.d.$\mathcal{N}(0,1)$, and $\tilde{u}_{T,t} = T^{-1/2}\sum_{s=1}^{t} a_s$ may be regarded as a contamination of this benchmark model. Note that the eigenvalues and eigenfunctions of $k(r,s) = E[W(r)W(s)] = r \wedge s$ are given by $r_l = \pi^{-2}(l-1/2)^{-2}$ and $\varphi_l(s) = \sqrt{2}\sin(\pi(l-1/2)s)$, $l = 1, 2, \ldots$ (see Phillips (1998)).

For the numerical analysis, we consider the performance of the two most robust long-run variance estimators $\hat{\omega}^2_{R2} = \hat{\xi}^2_1 = \left(r_1^{-1/2}T^{-1}\sum_{t=1}^{T}\varphi_1(t/T)\tilde{u}_{T,t}\right)^2$ and $\hat{\omega}^2_{R\Delta} = \tilde{u}^2_{T,T}$ in $\mathfrak{C}^2_T(\delta)$ and $\mathfrak{C}^{\Delta}_T(\delta)$, respectively, and the class of estimators $\hat{\omega}^2_{RE}(\lambda)$ that efficiently trade off maximal asymptotic bias $\gamma_2(\delta)$ and variance in the benchmark model of Theorem 2. In addition, we consider estimators $\hat{\omega}^2_{UA}(p)$ that are an unweighted average of $\hat{\xi}^2_l$ as defined in Theorem 2 (i), i.e. $\hat{\omega}^2_{UA}(p) = p^{-1}\sum_{l=1}^{p}\hat{\xi}^2_l$. This modification of the efficient estimators have central chi-squared asymptotic distributions with p degrees of freedom, scaled by

ω^2/p, whenever $\tilde{u}_{T,[\cdot T]} \Rightarrow \omega^2 G(\cdot)$. The estimator $\hat{\omega}^2_{UA}(p)$ achieves $\gamma_2(\delta) = \delta p^{-2} \sum_{l=1}^{p} r_l^{-1}$ (and $\bar{\gamma}_\Delta(\delta) = \delta \int \int |\sum_{l=1}^{p} r_l^{-1} \varphi_l(s)\varphi_l(r)| dsdr$), so Theorem 2 (iii) is applicable and the scaled central chi-squared asymptotic distribution is an accurate asymptotic approximation for all contaminated models in $\mathfrak{C}^2_T(\delta)$ for small δ.

We also include two fixed-b kernel estimators. Just as $\hat{\omega}^2_{RE}(\lambda)$ and $\hat{\omega}^2_{UA}(p)$, also these estimators follow a nondegenerate asymptotic distribution in the benchmark model, and the following small sample results are based on this non-degenerate asymptotic approximation. Specifically, we consider the Quadratic Spectral kernel long-run variance estimator $\hat{\omega}^2_{QS}(b)$, and the Bartlett kernel estimator $\hat{\omega}^2_{BT}(b)$ (see Kiefer and Vogelsang (2005) for details). For each class of estimators $\hat{\omega}^2_{RE}(\lambda)$, $\hat{\omega}^2_{UA}(p)$, $\hat{\omega}^2_{QS}(b)$ and $\hat{\omega}^2_{BT}(b)$, we report results for three specific members, where the values of λ, p and b are chosen such that the asymptotic variance in the benchmark model is given by 1, 1/4 and 1/8, respectively.

For comparison, we also consider the quadratic spectral estimator $\hat{\omega}^2_{QA}$ with an automatic bandwidth selection based on an AR(1) model for the bandwidth determination as suggested by Andrews (1991), and Andrews and Monahan's (1992) AR(1) prewhitened long-run variance estimator $\hat{\omega}^2_{PW}$ with a second stage quadratic spectral kernel estimator with automatic bandwidth selection based on an AR(1) model. Both $\hat{\omega}^2_{QA}$ and $\hat{\omega}^2_{PW}$ are, of course, consistent in the benchmark model, and the small sample results are based on the asymptotic approximation of these estimators having point mass at ω^2. This is the approximation typically employed when the long-run variance is a nuisance parameter. Alternatively, one might base inference for ω^2 on the asymptotic Gaussianity of $\hat{\omega}^2_{QA} - \omega^2$ and $\hat{\omega}^2_{PW} - \omega^2$ suitably scaled—see, for instance, Andrews (1991) for some general results. But the mean of this Gaussian approximation depends on unknown quantities that can be estimated in numerous ways, so that for brevity, no such results are presented.

Tables 2 and 3 describe the performance of these long-run variance estimators in the AR(1) and MA(1) for various values of ρ and θ and a sample size of $T = 100$, based on 50,000 replications. For each data generating process and long-run variance estimator, we report the bias, the root mean square error, the largest difference in the cumulative distribution function between the asymptotic distribution F and the small

Table 2: Small Sample Performance of Various LRV Estimators in the AR(1) Model for $T = 100$

					$\hat{\omega}^2_{RE}(\lambda)$			$\hat{\omega}^2_{UA}(p)$			$\hat{\omega}^2_{QS}(b)$			$\hat{\omega}^2_{BT}(b)$		
	$\hat{\omega}^2_{QA}$	$\hat{\omega}^2_{PW}$	$\hat{\omega}^2_{R\Delta}$	$\hat{\omega}^2_{R2}$	63.6	913	3636	2	8	16	0.621	0.130	0.064	1	0.197	0.096
Bias																
$\rho = 0$	0.01	0.01	0.01	0.02	0.02	0.03	0.06	0.02	0.03	0.07	0.00	0.00	0.00	0.00	0.00	0.00
$\rho = -0.7$	0.16	0.00	0.01	0.03	0.03	0.05	0.09	0.03	0.05	0.10	0.02	0.02	0.04	0.04	0.16	0.30
$\rho = 0.7$	-0.16	0.08	-0.04	-0.03	-0.03	-0.11	-0.25	-0.03	-0.13	-0.28	-0.04	-0.13	-0.29	-0.06	-0.17	-0.29
$\rho = 0.9$	-0.25	0.27	-0.14	-0.17	-0.21	-0.48	-0.65	-0.21	-0.52	-0.68	-0.18	-0.49	-0.68	-0.21	-0.47	-0.65
Root Mean Squared Error																
$\rho = 0$	0.20	0.25	1.43	1.45	1.02	0.52	0.40	1.02	0.52	0.41	1.01	0.50	0.35	1.01	0.50	0.36
$\rho = -0.7$	0.29	0.19	1.44	1.45	1.03	0.53	0.41	1.03	0.53	0.42	1.01	0.51	0.36	1.01	0.53	0.48
$\rho = 0.7$	0.49	0.59	1.36	1.36	0.97	0.48	0.40	0.97	0.47	0.40	0.97	0.47	0.40	0.97	0.50	0.42
$\rho = 0.9$	0.79	1.41	1.22	1.18	0.85	0.58	0.68	0.84	0.60	0.71	0.89	0.58	0.70	0.90	0.59	0.68
$\sup_x \lvert F_T(x) - F(x) \rvert$																
$\rho = 0$	0.51	0.53	0.01	0.01	0.01	0.02	0.06	0.01	0.02	0.06	0.00	0.01	0.01	0.00	0.01	0.01
$\rho = -0.7$	0.72	0.52	0.01	0.01	0.01	0.03	0.09	0.01	0.03	0.10	0.01	0.02	0.04	0.04	0.14	0.35
$\rho = 0.7$	0.72	0.55	0.01	0.01	0.02	0.11	0.32	0.02	0.12	0.36	0.02	0.12	0.37	0.04	0.16	0.37
$\rho = 0.9$	0.78	0.58	0.03	0.05	0.10	0.48	0.78	0.09	0.53	0.82	0.09	0.48	0.83	0.13	0.47	0.80
Coverage Rate of 90% CI																
$\rho = 0$	0.00	0.00	0.90	0.90	0.90	0.89	0.88	0.90	0.89	0.87	0.90	0.90	0.90	0.90	0.90	0.90
$\rho = -0.7$	0.00	0.00	0.90	0.90	0.90	0.89	0.87	0.90	0.89	0.87	0.90	0.90	0.90	0.92	0.92	0.84
$\rho = 0.7$	0.00	0.00	0.90	0.90	0.90	0.89	0.76	0.90	0.90	0.76	0.90	0.88	0.75	0.87	0.83	0.69
$\rho = 0.9$	0.00	0.00	0.91	0.91	0.90	0.63	0.18	0.91	0.62	0.14	0.88	0.63	0.12	0.82	0.58	0.16
Asymptotic Properties in Benchmark Model																
Variance	0.00	0.00	2.00	2.00	1.00	0.25	0.13	1.00	0.25	0.13	1.00	0.25	0.13	1.00	0.25	0.13
Lgth $CI_{90\%}$	0.00	0.00	254	254	14.6	2.30	1.37	19.2	2.41	1.40	11.2	2.28	1.37	6.70	2.09	1.32
Analytical Robustness Properties																
$\gamma_2(\delta)/\delta$	∞	∞	∞	2.47	11.4	182	728	12.3	210	841	∞	∞	∞	∞	∞	∞
$\gamma_\Delta(\delta)/\delta$	∞	∞	1	2	7.3	45.6	100	9.1	76.3	195	7.5	45.3	100	5	21.3	42.7

Table 3: Small Sample Performance of Various LRV Estimators in the MA(1) Model for $T = 100$

	$\hat{\omega}^2_{QA}$	$\hat{\omega}^2_{PW}$	$\hat{\omega}^2_{R\Delta}$	$\hat{\omega}^2_{R2}$	$\hat{\omega}^2_{RE}(\lambda)$ 63.6	913	3636	$\hat{\omega}^2_{UA}(p)$ 2	8	16	$\hat{\omega}^2_{QS}(b)$ 0.621	0.130	0.064	$\hat{\omega}^2_{BT}(b)$ 1	0.197	0.096
Bias																
$\theta = -0.7$	-0.02	0.55	-0.01	0.01	0.01	0.02	0.04	0.01	0.02	0.04	-0.01	-0.01	-0.02	-0.01	-0.03	-0.06
$\theta = 0.5$	1.01	1.00	0.03	0.04	0.04	0.09	0.23	0.05	0.10	0.26	0.04	0.06	0.17	0.07	0.23	0.44
$\theta = 0.7$	3.70	3.84	0.10	0.07	0.08	0.22	0.68	0.08	0.24	0.78	0.11	0.23	0.63	0.25	0.88	1.71
$\theta = 0.9$	41.9	43.4	0.98	0.21	0.29	1.85	6.84	0.30	2.10	7.88	1.05	2.47	7.12	2.77	10.03	19.53
Root Mean Squared Error																
$\theta = -0.7$	0.34	0.75	1.39	1.42	1.01	0.51	0.39	1.01	0.51	0.39	0.99	0.50	0.35	0.99	0.50	0.36
$\theta = 0.5$	1.11	1.09	1.47	1.49	1.06	0.55	0.50	1.06	0.56	0.54	1.04	0.53	0.43	1.04	0.57	0.59
$\theta = 0.7$	3.87	4.01	1.57	1.53	1.09	0.64	0.89	1.09	0.68	1.03	1.09	0.63	0.83	1.12	1.06	1.79
$\theta = 0.9$	43.3	45.0	2.96	1.73	1.31	2.32	7.42	1.34	2.75	8.83	2.25	3.11	7.73	3.42	10.3	19.8
$\sup_x \lvert F_T(x) - F(x)\rvert$																
$\theta = -0.7$	0.60	0.90	0.00	0.00	0.01	0.01	0.03	0.00	0.01	0.04	0.01	0.01	0.03	0.01	0.03	0.07
$\theta = 0.5$	1.00	1.00	0.01	0.01	0.02	0.07	0.22	0.02	0.07	0.24	0.01	0.05	0.18	0.05	0.21	0.47
$\theta = 0.7$	1.00	1.00	0.03	0.02	0.03	0.16	0.54	0.03	0.16	0.56	0.05	0.17	0.53	0.18	0.64	0.95
$\theta = 0.9$	1.00	1.00	0.17	0.05	0.10	0.71	0.99	0.10	0.69	0.99	0.28	0.77	1.00	0.82	1.00	1.00
Coverage Rate of 90% CI																
$\theta = -0.7$	0.00	0.00	0.90	0.90	0.90	0.89	0.88	0.90	0.89	0.88	0.90	0.90	0.90	0.90	0.89	0.87
$\theta = 0.5$	0.00	0.00	0.90	0.89	0.89	0.88	0.82	0.89	0.88	0.79	0.90	0.89	0.86	0.92	0.90	0.74
$\theta = 0.7$	0.00	0.00	0.89	0.90	0.89	0.86	0.53	0.89	0.84	0.48	0.90	0.86	0.55	0.93	0.61	0.01
$\theta = 0.9$	0.00	0.00	0.80	0.88	0.87	0.28	0.00	0.86	0.28	0.00	0.76	0.19	0.00	0.45	0.00	0.00
Asymptotic Properties in Benchmark Model																
Variance	0.00	0.00	2.00	2.00	1.00	0.25	0.13	1.00	0.25	0.13	1.00	0.25	0.13	1.00	0.25	0.13
Lgth $CI_{90\%}$	0.00	0.00	254	254	14.6	2.30	1.37	19.2	2.41	1.40	11.2	2.28	1.37	6.70	2.09	1.32
Analytical Robustness Properties																
$\gamma_2(\delta)/\delta$	∞	∞	∞	2.47	11.4	182	728	12.3	210	841	∞	∞	∞	∞	∞	∞
$\bar{\gamma}_\Delta(\delta)/\delta$	∞	∞	1	2	7.3	45.6	100	9.1	76.3	195	7.5	45.3	100	5	21.3	42.7

sample distribution F_T, $\sup_x |F_T(x) - F(x)|$, and the small sample coverage rate of a two-sided asymptotically justified 90% confidence interval, which is equally likely not to include a too small or too large value of ω^2 asymptotically. Given the lack of symmetry of the asymptotic distributions of the inconsistent long-run variance estimators, this is not the shortest 90% confidence interval for ω^2. For comparison, Tables 2 and 3 also report the asymptotic variance and asymptotic average length of this 90% confidence interval for each estimator, as well as the analytical robustness measures $\gamma_2(\delta)$ and $\bar{\gamma}_\Delta(\delta)$.

The numerical results underline the poor quality of approximations of small sample distributions based on consistent long-run variance estimators: in the presence of strong autocorrelations, $\hat{\omega}^2_{QA}$ and $\hat{\omega}^2_{PW}$ exhibit considerable biases and large root mean square errors. This is true even for $\hat{\omega}^2_{PW}$ in the AR(1) model, despite the fact that it is prewhitened based on the correct model of autocorrelation. At the same time, the most robust long-run variance estimator $\hat{\omega}^2_{R2}$ in $\mathfrak{C}^2_T(\delta)$ displays remarkable resilience even in the face of very strong autocorrelations, with relatively little bias and empirical coverage rates of the 90% confidence interval never more than two percentage points off the nominal value. The estimator $\hat{\omega}^2_{R\Delta}$ comes close to this robustness, but it does somewhat worse in the MA(1) model with θ large. These performances, however, come at the cost of $\hat{\omega}^2_{R2}$ and $\hat{\omega}^2_{R\Delta}$ being strikingly inaccurate estimators, with an asymptotic average length of the 90% confidence interval of 254.

The relative small sample performance of the various long-run variance estimators is not particularly well explained by $\gamma_2(\delta)$ or $\bar{\gamma}_\Delta(\delta)$. The measure $\bar{\gamma}_\Delta(\delta)$ successfully ranks the small sample robustness as described by Tables 2 and 3 within the same class of estimators, but not really across: The fixed-b Bartlett estimators $\hat{\omega}^2_{BT}(b)$, for instance, have relatively small $\bar{\gamma}_\Delta(\delta)$ compared to $\hat{\omega}^2_{RE}(\lambda)$, $\hat{\omega}^2_{UA}(p)$ and $\hat{\omega}^2_{QS}(b)$, but perform about equally well in the AR(1) model and worse than these in the MA(1) model. Maybe this should not be too surprising: $\gamma_2(\delta)$ and $\bar{\gamma}_\Delta(\delta)$ are defined with respect to worst case contaminations, and relative performance at these extremes does not necessarily translate into similar relative performance for less extreme contaminations, even asymptotically. As noted in Section 3, the worst case contamination in $\mathfrak{C}^2_T(\delta)$ corresponds to an almost non-invertible MA(1) in the underlying disturbances. Table 3 indeed shows that estimators with small $\gamma_2(\delta)$ do somewhat better in the MA(1) model with $\theta = 0.9$ compared to those with large or infinite $\gamma_2(\delta)$, and unreported simulations for $T = 400$ and $\theta = 0.95$, for instance, reveal much sharper differences.

Overall, these small sample results show competitive performance of $\hat{\omega}^2_{RE}(\lambda)$ and $\hat{\omega}^2_{UA}(p)$ compared to previously studied inconsistent long-run variance estimators, with only minor gains of $\hat{\omega}^2_{RE}(\lambda)$ over $\hat{\omega}^2_{UA}(p)$ conditional on the asymptotic variance in the benchmark model. The convenient standard asymptotic distribution of $\hat{\omega}^2_{UA}(p)$, in combination with its attractive theoretical properties, thus makes $\hat{\omega}^2_{UA}(p)$ a potentially appealing choice for applied work.

5 Conclusion

For consistent estimators to work, any given data has to satisfy relatively strong regularity conditions. For the problem of long-run variance estimation, many real world time series do not seem to exhibit enough regularity such that a substitution of the unknown population value with a consistent estimator yields reliable approximations.

In order to address this issue, this paper develops a framework to analytically investigate the robustness of long-run variance estimators. The starting point is the assumption that the data $\{u_{T,t}\}_{t=1}^T$ satisfies $u_{T,[\cdot T]} \Rightarrow \omega G(\cdot)$ for some mean-zero Gaussian process G with known covariance kernel, where the scalar ω is the square root of the long-run variance. It is found that all equivariant long-run variance estimators that are consistent in the benchmark model $\{u_{T,t}\}_{t=1}^T \sim \{G(t/T)\}_{t=1}^T$ lack qualitative robustness: There always exists a sequence of contaminated disturbances $\{\tilde{u}_{T,t}\}_{t=1}^T$ satisfying $\tilde{u}_{T,[\cdot T]} \Rightarrow G(\cdot)$, yet the long-run variance estimator converges in probability to an arbitrary positive value in this contaminated model. This result may serve as an analytical motivation for considering inconsistent long-run variance estimators that remain robust in the whole class of models satisfying $u_{T,[\cdot T]} \Rightarrow \omega G(\cdot)$, such as those derived in Kiefer, Vogelsang and Bunzel (2000) and Kiefer and Vogelsang (2002, 2005).

Furthermore, we determine the form of optimal inconsistent long-run variance estimators that, among all estimators that can be written as a quadratic form in u_T, efficiently trade off bias in a class of contaminated models against variance in the uncontaminated benchmark model. A minor modification of these efficient estimators yields $\hat{\omega}^2_{UA}(p)$, which conveniently is asymptotically distributed chi-squared with p degrees of freedom, scaled by ω^2/p, whenever $u_{T,[\cdot T]} \Rightarrow \omega G(\cdot)$. Also, this distributional approximation is shown to be uniformly asymptotically accurate in a set of models with small contaminations.

In a Monte Carlo analysis there emerges a stark trade-off between the robustness and efficiency of inconsistent long-run variance estimators, as governed by the parameter p for $\hat{\omega}^2_{UA}(p)$. This raises the important question of how to pick an appropriate value in practice. While a detailed discussion is beyond the scope of this paper, the results obtained here provide an inherent limit to data dependent strategies: Whenever a data dependent choice of p leads to the efficient choice of an unbounded p with probability one in the benchmark model, then the resulting long-run variance estimator is consistent in the benchmark model, and hence qualitatively fragile.

A fruitful approach to the choice of p might result from a spectral perspective. When $u_{T,[\cdot T]} \Rightarrow \omega G(\cdot)$ with G a Brownian Bridge $G(s) \sim W(s) - sW(1)$, an asymptotically equivalent representation of $\hat{\omega}^2_{UA}(p)$ is given by

$$\hat{\omega}^2_{UA}(p) = p^{-1} \sum_{l=1}^{p} \left(\sqrt{2} \sum_{t=1}^{T} \cos(\pi l(t-1/2)/T) \Delta u_{T,t} \right)^2 + o_p(1).$$

The choice of p may hence be interpreted as the size of the neighborhood of zero for which the spectrum of $\{T^{1/2}\Delta u_{T,t}\}$, as described by the low frequencies of the discrete cosine transform type II, is required to be flat. Knowledge about the form of the spectrum of $\{T^{1/2}\Delta u_{T,t}\}$ then suggests appropriate values for p; for macroeconomic time series, for instance, one might want to pick p small enough not to dip into business cycle frequencies. Under the standard convention of the lowest business cycle frequency corresponding to a 8 year period, this would suggest letting $p = [Y/4]$, where Y is the span of the data measured in years. Even if such knowledge about the spectrum of $\{T^{1/2}\Delta u_{T,t}\}$ is elusive, any given choice between robustness and efficiency as embodied by p might be easier to interpret from a spectral perspective.

6 Appendix

Proof of Theorem 1:

Since $k(r,s) = E[G(r)G(s)]$ is continuous, the eigenfunctions $\varphi_1, \varphi_2, \ldots$ of k corresponding to the eigenvalues $r_1 \geq r_2 \geq \ldots$ are continuous and $\sum_{l=1}^{\infty} r_l \varphi_l(r)\varphi_l(s)$ converges uniformly to $k(r,s)$ by Mercer's Theorem—see Hochstadt (1973), p. 90. Let $\xi_l \sim$i.i.d.$\mathcal{N}(0,1)$, $l = 1, 2, \ldots$, and denote with $\mathcal{C}$ the set of continuous functions on the unit interval, equipped with the sup norm. Since k is continuous and the sample paths of G are continuous a.s., G can be constructed as $G(s) = \sum_{l=1}^{\infty} r_l^{1/2}\varphi_l(s)\xi_l$, since the r.h.s. converges a.s. on $\mathcal{C}$, i.e. uniformly in $s \in [0,1]$—see Gilsing and Sottinen (2003).

Let $G_0(s) = cG(s) = c\sum_{l=1}^{\infty} r_l^{1/2}\varphi_l(s)\xi_l$, and define $G_n(s) = \sum_{l=1}^{n} r_l^{1/2}\varphi_l(s)\xi_l + c\sum_{l=n+1}^{\infty} r_l^{1/2}\varphi_l(s)\xi_l$ for $n \geq 1$. We first show that the measures of G_0 and G_n on $\mathcal{C}$ are equivalent: For $x \in \mathcal{C}$, let $\psi_l(x) = \int_0^1 x(s)\varphi_l(s)ds$. Consider the continuous functions $h : \mathcal{C} \mapsto \mathcal{C} \times \mathbb{R}^n$ and $g : \mathcal{C} \times \mathbb{R}^n \mapsto \mathcal{C}$ with $h(x) = (h_1(x), h_2(x)) = (x - \sum_{l=1}^{n} \varphi_l\psi_l(x), (\psi_1(x), \cdots, \psi_n(x))')$ and $g(x, (v_1, \cdots, v_n)') = x + \sum_{l=1}^{n} \varphi_l v_l$ (where the metric on $\mathcal{C} \times \mathbb{R}^n$ is chosen as the sum of the sup norm in $\mathcal{C}$ and the Euclidian norm in $\mathbb{R}^n$). Since $\{\varphi_l\}_{l=1}^{\infty}$ are orthonormal, $h_1(x)$ and $h_2(x)$ are the residual and coefficient vector of a continuous time regression of x on $\{\varphi_l\}_{l=1}^{n}$, respectively. Clearly, $g(h(x)) = x$ for any $x \in \mathcal{C}$. For any measurable $\mathcal{A} \subset \mathcal{C}$, we thus have $P(G_j \in \mathcal{A}) = P(h(G_j) \in g^{-1}(\mathcal{A}))$ for $j \in \{0, n\}$, where $g^{-1}(\mathcal{A}) = (\mathcal{A}_1^-, \mathcal{A}_2^-)$ is the inverse image of $\mathcal{A}$ under g. It thus suffices to show equivalence of the measures of $h(G_0)$ and $h(G_n)$. Since $\{\xi_l\}_{l=1}^{\infty}$ are i.i.d, $P(h(G_j) \in (\mathcal{A}_1^-, \mathcal{A}_2^-)) = P(h_1(G_j) \in \mathcal{A}_1^-)P(h_2(G_j) \in \mathcal{A}_2^-)$, and $P(h_1(G_0) \in \mathcal{A}_1^-) = P(h_1(G_n) \in \mathcal{A}_1^-)$ because $h_1(G_0) = h_1(G_n) = c\sum_{l=n+1}^{\infty} r_l^{1/2}\varphi_l(s)\xi_l$. But $h_2(G_n) = c^{-1}h_2(G_0) = (r_1^{1/2}\xi_1, \cdots, r_n^{1/2}\xi_n)' \sim \mathcal{N}(0, \mathrm{diag}(r_1, \cdots, r_n))$, so that $P(h_2(G_n) \in \mathcal{A}_2^-) = 0$ if and only if $P(h_2(G_0) \in \mathcal{A}_2^-) = 0$, and therefore $P(G_n \in \mathcal{A}) = 0$ if and only if $P(G_0 \in \mathcal{A}) = 0$.

For $n \geq 0$, let the $T \times 1$ vector u_T^n have elements $G_n(1/T), \cdots, G_n(T/T)$, and let $u_T = c^{-1}u_T^0$. For any $\epsilon > 0$, the event $|\hat{\omega}_T^2(u_T^j) - c^2| > \epsilon$ can be equivalently expressed as $G_j \in \mathcal{A}_T(\epsilon) \subset \mathcal{C}$, because u_T^j is a continuous function of G_j for $j \in \{0, n\}$. Since $u_T^0 = cu_T$, scale equivariance of $\hat{\omega}_T^2$ and $\hat{\omega}_T^2(u_T) \xrightarrow{p} 1$ imply $\hat{\omega}_T^2(u_T^0) \xrightarrow{p} c^2$, so that $P(G_0 \in \mathcal{A}_T(\epsilon)) \to 0$. It follows from the equivalence of the measures of G_0 and G_n

that also $P(G_n \in \mathcal{A}_T(\epsilon)) \to 0$ (see Pollard (2002), p. 55). But ϵ was arbitrary, so that $\hat{\omega}_T^2(u_T^n) \xrightarrow{p} c^2$.

There hence exists for any n a finite number T_n such that $P(|\hat{\omega}_T^2(u_T^n) - c^2| > n^{-1}) < n^{-1}$ for all $T \geq T_n$. For any T, let n_T be the largest n^* such that $\max_{n \leq n^*} T_n < T$. Note that $n_T \to \infty$ as $T \to \infty$, as T_n is finite for any n. Let $\tilde{u}_T = u_T^{n_T}$. By construction, $P(|\hat{\omega}_T^2(\tilde{u}_T) - c^2| > n_T^{-1}) < n_T^{-1}$, such that $\hat{\omega}_T^2(\tilde{u}_T) \xrightarrow{p} c^2$. Now

$$\tilde{u}_{T,t} = \sum_{l=1}^{n_T} r_l^{1/2} \varphi_l(t/T)\xi_l + c \sum_{l=n_T+1}^{\infty} r_l^{1/2} \varphi_l(t/T)\xi_l,$$

and hence

$$\sup_{1 \leq t \leq T} |\tilde{u}_{T,t} - u_{T,t}| \leq \sup_{0 \leq s \leq 1} |(c-1) \sum_{l=n_T+1}^{\infty} r_l^{1/2} \varphi_l(s)\xi_l| \to 0 \text{ a.s.}$$

because of $n_T \to \infty$ and the a.s. convergence of $\sum_{l=1}^{\infty} r_l^{1/2} \varphi_l(\cdot)\xi_l$ on $\mathcal{C}$. Also, the covariance kernel of the process G_{n_T} is given by $E[G_{n_T}(r)G_{n_T}(s)] = \sum_{l=1}^{n_T} r_l \varphi_l(r)\varphi_l(s) + c^2 \sum_{l=n_T+1}^{\infty} r_l \varphi_l(r)\varphi_l(s)$, which converges uniformly to $k(s,r)$ as $n_T \to \infty$, so that $||E[\tilde{u}_T \tilde{u}_T'] - \Sigma_T||_\Delta \to 0$, as claimed.

Scale Normalizations of fixed-b estimators:

Let $\bar{A}_T$ be the matrix of the unnormalized kernel estimator, that is the s,t element of the symmetric matrix $\bar{A}_T$ is $\bar{a}_T(s,t) = 2\kappa((s-t)/bT) - \kappa((s-t-1)/bT) - \kappa((s-t+1)/bT)$ for $s,t < T$, $\bar{a}_T(s,t) = \kappa((s-t)/bT) - \kappa((s-t-1)/bT)$ for $s = T$ and $t < T$ and $\bar{a}_T(T,T) = \kappa(0) = 1$. If κ is twice differentiable, by exact first and second order Taylor expansions, $T^2 \bar{a}_T(s,t) = -b^{-2}\kappa''((s-t)/bT) + R_T(s,t)$ for $s,t < T$ and $T\bar{a}_T(T,t) = b^{-1}\kappa'((T-t)/bT) + R_T(T,t)$ for $t < T$, where $\sup_{1 \leq s,t \leq T} |R_T(s,t)| \leq \max(b^{-2} \sup_{r,s \in [0,b^{-1}], |r-s| \leq 2/bT} |\kappa''(r) - \kappa''(s)|, b^{-1} \sup_{r,s \in [0,b^{-1}], |r-s| \leq 2/bT} |\kappa'(r) - \kappa'(s)|) \to 0$ since κ' and κ'' are continuous (and hence uniformly continuous) on $[0, b^{-1}]$. By a direct calculation, $\text{tr}(\bar{A}_T \Sigma_T) = \sum_{s=1}^{T} \sum_{t=1}^{T} k(s/T, t/T)\bar{a}_T(s,t) \to k(1,1) + 2b^{-1} \int \kappa'((1-s)/b)k(1,s)ds - b^{-2} \int\int \kappa''((r-s)/b)k(r,s)drds$, since k, κ' and κ'' are continuous and therefore Riemann integrable. The result for the Bartlett fixed-b estimator follows similarly.

Proof of Theorem 2:

(i) For each T sufficiently large to make the efficiency constraint feasible, we will first derive the quadratic long-run variance estimator that minimizes $\delta T \operatorname{tr} A_T$ subject to $E[(\hat{\omega}_T^2(u_T) - 1)^2] = 2 \operatorname{tr}(\Sigma_T A_T \Sigma_T A_T) \leq \varsigma$ and $\operatorname{tr}(\Sigma_T A_T) = 1$.

Let $Q_T D_T Q_T'$ be the spectral decomposition of Σ_T, and write $A_T = Q_T'(D_T^+)^{1/2}\tilde{A}_T(D_T^+)^{1/2}Q_T$ for some positive definite matrix $\tilde{A}_T$, where $D_T^+ = \operatorname{diag}(d_1^+(1), \cdots, d_T^+(T))$ is the Moore-Penrose inverse of $D_T = \operatorname{diag}(d_T(1), \cdots, d_T(T))$, and $d_T(1) \geq d_T(2) \geq \cdots \geq d_T(T)$. This leaves A_T unrestricted on the space spanned by Σ_T, and it is optimal to restrict A_T to be zero on the null-space of Σ_T: changing A_T on the null-space of Σ_T leaves the variance $\operatorname{tr}(\Sigma_T A_T \Sigma_T A_T)$ and the constraint $\operatorname{tr}(\Sigma_T A_T) = 1$ unaltered while it increases the maximal bias $\delta T \operatorname{tr} A_T$.

Let $\mu_T(1) \geq \mu_T(2) \geq \cdots \geq \mu_T(T) \geq 0$ be the eigenvalues of $\tilde{A}_T$. Then $\operatorname{tr}(A_T \Sigma_T) = \operatorname{tr} \tilde{A}_T = \sum_{l=1}^T \mu_T(l)$, $\operatorname{tr}(\Sigma_T A_T \Sigma_T A_T) = \operatorname{tr}(\tilde{A}_T \tilde{A}_T) = \sum_{l=1}^T \mu_T(l)^2$ and $\operatorname{tr} A_T = \operatorname{tr}(D^+ \tilde{A}_T) \geq \sum_{l=1}^T d_T^+(l)\mu_T(l)$, where the last inequality follows from Theorem H.1.h., page 249, of Marshall and Olkin (1979). So among all matrixes $\tilde{A}_T$ with the same set of eigenvalues, we may always choose $\tilde{A}_T = \operatorname{diag}(\mu_T(1), \cdots, \mu_T(T))$ to minimize $\operatorname{tr}(D^+ \tilde{A}_T)$. It is straightforward to see that the program

$$\begin{aligned} &\min_{\{\mu_T(l)\}_{l=1}^T} T \textstyle\sum_{l=1}^T d_T^+(l)\mu_T(l) \\ &\text{s.t. } \textstyle\sum_{l=1}^T \mu_T(l) = 1, \ \sum_{l=1}^T \mu_T(l)^2 \leq \varsigma/2 \text{ and } \mu_T(l) \geq 0 \text{ for } l = 1, \cdots, T \end{aligned}$$

is solved by

$$\mu_T^*(l) = \frac{(\lambda_T - T d_T^+(l)) \vee 0}{\sum_{j=1}^T ((\lambda_T - T d_T^+(j)) \vee 0)}$$

with λ_T determined by $\sum_{l=1}^T \mu_T^*(l)^2 \leq \varsigma/2$, and the resulting maximal bias is given by $\delta T \sum_{l=1}^T \mu_T^*(l) d_T^+(l)$.

Now it is known that the largest N eigenvalues of $T^{-1}\Sigma_T$ converge to the largest N eigenvalues of k, for any finite N (Hochstadt (1973), chapter 6). Thus λ_T and $\{\mu_T^*(l)\}_{l=1}^N$ converge to limits λ and $\{w_l(\lambda)\}_{l=1}^N$, respectively, and the limit maximal bias and variance of the small sample estimator is given by $\delta \sum_{l=1}^{p(\lambda)} r_l^{-1} w_l(\lambda)$ and $2\sum_{l=1}^{p(\lambda)} w_l(\lambda)^2$, respectively. As an implication of the small sample efficiency of this estimator, no quadratic long-run variance estimator can exist with a better trade-off between $\gamma_2(\delta)$ and the limit superior of the variance in the benchmark model.

It hence suffices to show that $\hat{\omega}^2_{RE}$ is asymptotically unbiased in the benchmark model and achieves the same limiting maximal bias and variance as this sequence of efficient small sample estimators. Denote with A_{RE} the $T \times T$ matrix such that $\hat{\omega}^2_{RE}(\tilde{u}_T) = \tilde{u}'_T A_{RE} \tilde{u}_T$, i.e. $[A_{RE}]_{s,t} = T^{-2} \sum_{l=1}^{p(\lambda)} w_l r_l^{-1} \varphi_l(s/T)\varphi_l(t/T)$. Then

$$\begin{aligned}
\operatorname{tr}(A_{RE}\Sigma_T) &= T^{-2} \sum_{l=1}^{p(\lambda)} w_l(\lambda) r_l^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} \varphi_l(t/T) k(s/T, t/T) \varphi_l(s/T) \\
&\to \sum_{l=1}^{p(\lambda)} w_l(\lambda) r_l^{-1} \int\int \varphi_l(r) k(r,s) \varphi_l(s) ds dr \\
&= \sum_{l=1}^{p(\lambda)} w_l(\lambda) = 1
\end{aligned}$$

$$T \operatorname{tr} A_{RE} = T \sum_{l=1}^{p(\lambda)} w_l(\lambda) r_l^{-1} T^{-2} \sum_{t=1}^{T} \varphi_l(t/T)^2 \to \sum_{l=1}^{p(\lambda)} r_l^{-1} w_l(\lambda)$$

$$\begin{aligned}
\operatorname{tr}(A_{RE}\Sigma_T A_{RE}\Sigma_T) &= \sum_{l=1}^{p(\lambda)} \sum_{m=1}^{p(\lambda)} w_l(\lambda) w_m(\lambda) r_l^{-1} r_m^{-1} (T^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} \varphi_l(t/T) k(s/T, t/T) \varphi_l(s/T))^2 \\
&\to \sum_{l=1}^{p(\lambda)} \sum_{m=1}^{p(\lambda)} w_l(\lambda) w_m r_l^{-1} r_m^{-1} \left(\int\int \varphi_l(r) k(r,s) \varphi_m(s) ds dr \right)^2 \\
&= \sum_{l=1}^{p(\lambda)} w_l(\lambda)^2
\end{aligned}$$

since $\varphi_l(r)k(r,s)\varphi_m(s)$ is continuous in (r,s), and therefore Riemann integrable.

Similarly, the result for $\gamma_2(\delta)$ is an immediate consequence the continuity and thus Riemann integrability of the $[0,1]^2 \mapsto \mathbb{R}$ function defined by $\sum_{l=1}^{p(\lambda)} w_l r_l^{-1} \varphi_l(r)\varphi_l(s)$.

(ii) By the continuity of k, it is obvious that $\hat{\omega}^2_{R\Delta}$ achieves $\bar{\gamma}_\Delta(\delta) = \delta/k(\tau,\tau)$ (and also $\gamma_2(\delta) = \infty$).

Similar to the proof of part (i), consider the optimal quadratic estimator with matrix $A_T^* = [a_T^*(s,t)]_{s,t}$ that minimizes $\sum_{s=1}^T \sum_{t=1}^T |a_T(s,t)|$ subject to $\operatorname{tr}(A_T\Sigma_T) = 1$. Let $\tau_T \in \arg\max_{1 \le t \le T} k(t/T, t/T)$, and $a_T^*(s,t) = 1/k(\tau_T, \tau_T)$ if $s = t = \tau_T$ and $a_T^*(s,t) = 0$ otherwise, so that $\sum_{s=1}^T \sum_{t=1}^T |a_T^*(s,t)| = 1/k(\tau_T,\tau_T)$. This choice is optimal, since $1 = \operatorname{tr}(A_T\Sigma_T) \le \sum_{s=1}^T \sum_{t=1}^T |k(s/T, t/T)||a_T(s,t)| \le$

$\max_{1\leq s,t\leq T}|k(s/T,t/T)|\sum_{s=1}^T\sum_{t=1}^T|a_T(s,t)|$, and $|k(r,s)| \leq \sqrt{k(r,r)k(s,s)} \leq k(r,r)\vee k(s,s)$ for all $r,s\in[0,1]$. Since $k(\tau_T,\tau_T)\leq k(\tau,\tau)$ for all T, $\overline{\lim}_{T\to\infty}\sum_{s=1}^T\sum_{t=1}^T|a_T^*(s,t)|\geq 1/k(\tau,\tau)$, and the result follows.

(iii) We compute

$$\begin{aligned}E|\hat{\omega}_T^2(\tilde{u}_T)-\hat{\omega}_T^2(u_T)| &= E|\eta_T'A_T\eta_T-\tilde{\eta}_T'A_T\tilde{\eta}_T+2\eta_T'A_Tu_T-2\tilde{\eta}_T'A_T\tilde{u}_T| \\ &\leq 2E|\tilde{\eta}_T'A_T\tilde{u}_T|+2E|\eta_T'A_Tu_T|+E\tilde{\eta}_T'A_T\tilde{\eta}_T+E\eta_T'A_T\eta_T.\end{aligned}$$

Now

$$\begin{aligned}E\tilde{\eta}_T'A_T\tilde{\eta}_T &= \operatorname{tr}(A_T\tilde{V}_T)\leq T\delta\operatorname{tr}A_T \\ E\eta_T'A_T\eta_T &= \operatorname{tr}(A_TV_T)\leq T\delta\operatorname{tr}A_T\end{aligned}$$

and

$$\begin{aligned}(E|\eta_T'A_Tu_T|)^2 &\leq E(\eta_T'A_Tu_T)^2 \\ &= \operatorname{tr}(A_TV_TA_T\Sigma_T)\leq T\delta\operatorname{tr}(A_T\Sigma_TA_T)\end{aligned}$$

and with $E\tilde{u}_T\tilde{u}_T'=\Sigma_T+V_T-\tilde{V}_T$, also

$$\begin{aligned}(E|\tilde{\eta}_T'A_T\tilde{u}_T|)^2 &\leq E(\tilde{\eta}_T'A_T\tilde{u}_T)^2 \\ &= \operatorname{tr}(A_T\tilde{V}_TA_T(\Sigma_T+V_T-\tilde{V}_T)) \\ &\leq T\delta\operatorname{tr}(A_T\Sigma_TA_T)+T^2\delta^2\operatorname{tr}(A_TA_T) \\ &\leq T\delta\operatorname{tr}(A_T\Sigma_TA_T)+(T\delta\operatorname{tr}A_T)^2.\end{aligned}$$

Since $\operatorname{tr}(A_T\Sigma_T)\to 1$ and A_T is positive semi-definite, the largest eigenvalue of $A_T\Sigma_T$ has a limit superior bounded by unity, so that $\overline{\lim}_{T\to\infty}(T\delta\operatorname{tr}A_T-T\delta\operatorname{tr}(A_T\Sigma_TA_T))\geq 0$. The result now follows from $\gamma_2(\delta)=\overline{\lim}_{T\to\infty}T\delta\operatorname{tr}A_T$.

(iv) Let $\bar{A}_T$ as in the derivation of the scale normalization of fixed-b estimators above. The results concerning $\bar{\gamma}_\Delta(\delta)$ follow from the same Taylor expansion result derived there for twice continuously differentiable kernels, and from a straightforward computation in the case of the fixed-b Bartlett estimator. The result $\gamma_2(\delta)=\infty$ is an immediate consequence of $\bar{a}_T(T,T)=1$.

References

ANDERSEN, T., T. BOLLERSLEV, AND F. DIEBOLD (2005): "Parametric and Nonparametric Volatility Measurement," in *Handbook of Financial Econometrics*, ed. by L. Hansen, and Y. Ait-Sahalia. North-Holland.

ANDREWS, D. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BERK, K. (1974): "Consistent Autoregressive Spectral Estimates," *The Annals of Statistics*, 2, 489–502.

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

CHAN, N., AND C. WEI (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

DE JONG, R., AND J. DAVIDSON (2000): "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68, 407–423.

DEN HAAN, W., AND A. LEVIN (1997): "A Practitioner's Guide to Robust Covariance Matrix Estimation," in *Handbook of Statistics 15*, ed. by G. Maddala, and C. Rao, pp. 309–327. Elsevier, Amsterdam.

FAUST, J. (1999): "Conventional Confidence Intervals for Points on Spectrum Have Confidence Level Zero," *Econometrica*, 67, 629–637.

GILSING, H., AND T. SOTTINEN (2003): "Power Series Expansions for Fractional Brownian Motions," *Theory of Stochastic Processes*, 9, 38–49.

GRANGER, C. (1966): "The Typical Spectral Shape of an Economic Variable," *Econometrica*, 34, 150–161.

HAMPEL, F. (1968): "Contributions to the Theory of Robust Estimation," *Ph.D. dissertation, University of California, Berkeley.*

——— (1971): "A General Qualitative Definition of Robustness," *The Annals of Mathematical Statistics*, 42, 1887–1896.

Hannan, E. (1957): "The Variance of the Mean of a Stationary Process," *Journal of the Royal Statistical Society B*, 19, 282–285.

Hochstadt, H. (1973): *Integral Equations.* John Wiley and Sons, New York.

Hosoya, Y. (1978): "Robust Linear Extrapolations of Second-Order Stationary Processes," *The Annals of Probability*, 6, 574–584.

Huber, P. (1996): *Robust Statistical Procedures.* SIAM, Philadelphia, 2nd edn.

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

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

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

Kiefer, N., T. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 695–714.

Kleiner, B., R. Martin, and D. Thomson (1979): "Robust Estimation of Power Spectra," *Journal of the Royal Statistical Society B*, 41, 313–351.

Künsch, H. (1984): "Infinitesimal Robustness for Autoregressive Processes," *The Annals of Statistics*, 12, 843–863.

Marshall, A., and I. Olkin (1979): *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

Martin, R., and V. Yohai (1986): "Influence Functionals for Time Series," *The Annals of Statistics*, 14, 781–818.

Martin, R., and R. Zamar (1993): "Efficiency-Constrained Bias-Robust Estimation of Location," *The Annals of Statistics*, 21, 338–354.

Müller, U. (2004): "The Impossibility of Consistent Discrimination Between I(0) and I(1) Processes," *mimeo, Princeton University.*

——— (2005): "Size and Power of Tests of Stationarity in Highly Autocorrelated Time Series," *Journal of Econometrics*, 128, 195–213.

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

Phillips, P. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

——— (1998): "New Tools for Understanding Spurious Regression," *Econometrica*, 66, 1299–1325.

Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge, UK.

Pötscher, B. (2002): "Lower Risk Bounds and Properties of Confidence Sets for Ill-Posed Estimation Problems with Applications to Spectral Density and Persistence Estimation, Unit Roots, and Estimation of Long Memory Parameters," *Econometrica*, 70, 1035–1065.

Robinson, P. (1994): "Semiparametric Analysis of Long-Memory Time Series," *The Annals of Statistics*, 22, 515–539.

——— (2005): "Robust Covariance Matrix Estimation: "HAC" Estimates with Long Memory/Antipersistence Correction," *Econometric Theory*, 21, 171–180.

Robinson, P., and C. Velasco (1997): "Autocorrelation-Robust Inference," in *Handbook of Statistics 15*, ed. by G. Maddala, and C. Rao, pp. 267–298, Amsterdam. Elsevier.

Samarov, A. (1987): "Robust Spectral Regression," *The Annals of Statistics*, 15, 99–111.

Sims, C. (1971): "Distributed Lag Estimation When the Parameter Space is Explicitly Infinite-Dimensional," *The Annals of Mathematical Statistics*, 42, 1622–1636.

Stock, J., and M. Watson (1998): "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model," *Journal of the American Statistical Association*, 93, 349–358.

Sun, Y., P. Phillips, and S. Jin (2006): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Cowles Foundation Discussion Paper 1545*.