

A Theory of Robust Long-Run Variance Estimation

Ulrich K. Müller*

Princeton University
Economics Department
Princeton, NJ, 08544

First Draft: September 2003

This Version: April 2004

Abstract

The paper studies the robustness of long-run variance estimators employed for conducting Wald-type tests in standard time series models. It is shown that all long-run variance estimators that are consistent for the variance of Gaussian White Noise lack robustness in the sense that they yield arbitrary results for some underlying process that satisfies a Functional Central Limit Theorem. An analytical measure of robustness of long-run variance estimators is suggested that captures the degree of this fragility. A family of inconsistent long-run variance estimators is derived that optimally trade off this measure of robustness against efficiency. A minor modification of these optimal estimators lead to asymptotically F-distributed test statistics under the null hypothesis, so that robust large sample inference can be conducted very similarly to well-understood small sample Gaussian inference.

JEL Classification: C13, C12, C32

Keywords: Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, Functional Central Limit Theorem, Contiguity, F-distribution, Hotelling T^2

*The author thanks Graham Elliott, Michael Jansson, Barbara Rossi, Bo Honore, Chris Sims, Mark Watson and participants at research seminars at Cornell University, the University of Montreal, the University of Pennsylvania, the University of Michigan, the University of Houston and at the Triangle Econometrics Workshop for very helpful discussions.

1 Introduction

Sums and averages play a crucial role in most econometric techniques. Most estimators can be written as a simple average of (possibly transformed) data. The application of a suitable Central Limit Theorem ensures that, appropriately scaled, the difference between the sample average and the parameter behaves like a zero mean Gaussian variate, at least for a large enough sample. An estimation of the parameter by the average hence leads to an error term that is close to having a mean zero Gaussian distribution, and inference about the parameter boils down to establishing the variance of this Gaussian distribution. In a time series context, this variance is called the 'long-run variance', abbreviated LRV in the following.

Under broad assumptions on the underlying disturbances, it is possible to consistently estimate the LRV. Consistency of the LRV estimator justifies an approximation of the small sample distribution that ignores differences between the LRV estimator and its population value. This approximation does not only lead to simple Gaussian and χ^2 distributions of standard test statistics; consistent LRV estimators also make inference asymptotically efficient as no additional uncertainty is incurred by the lack of knowledge of the population LRV. Out of these reasons, most research has concentrated on deriving consistent LRV estimators. Important contributions include Berk (1974), White (1984), Gallant (1987), Newey and West (1987), Andrews (1991), Hansen (1992) and Newey and West (1994), among others.

In the standard asymptotic thought experiment, consistent LRV estimators lead to simple and efficient tests, so that there appears to be no compelling reason to consider alternative approaches. Unfortunately, however, the approximation of this thought experiment for small samples with moderate to strong amounts of dependence is poor—see Haan and Levin (1997), for instance. While theoretically appealing and simple to apply, inference based on consistent LRV estimators often leads to tests with bad size control in small samples with realistic amounts of dependence. As demonstrated in the seminal paper of Kiefer, Vogelsang, and Bunzel (2000) and further developed in Kiefer and Vogelsang (2002b) and Kiefer and Vogelsang (2003), the small sample approximation can be improved upon for various data generating processes when tests are based instead on certain inconsistent LRV estimators.

While strongly encouraging, these results are not fully satisfying. It is quite straightfor-

ward to analytically describe the loss in power of tests based on inconsistent LRV estimators in the standard asymptotic thought experiment, see Kiefer, Vogelsang, and Bunzel (2000) and Kiefer and Vogelsang (2002b). It is much harder, however, to analytically describe the gain in robustness of such tests, since under standard asymptotics, also consistent LRV estimators lead to correctly sized tests. An interesting but analytically difficult way to approach this question is to derive second-order properties of the error in rejection probability and discriminate LRV estimators on these grounds. Jansson (2004) finds that in a Gaussian location model, a certain class of inconsistent LRV estimators that contains the suggestion by Kiefer, Vogelsang, and Bunzel (2000) lead to tests with an order of magnitude smaller error in rejection probabilities than those from certain consistent LRV estimators. It might be possible to extend these results to other models and classes of LRV estimators; it suffers from the drawback, though, that it classifies LRV estimators as either nonrobust (when the error in rejection probability is of the usual order of magnitude) or robust (when the order of magnitude is smaller). This makes it impossible to compare LRV estimators within each of the groups, whereas simulations indicate that—unsurprisingly—the size control properties of tests in the same class can be very different.

This paper tries to shed some light on these issues from an alternative perspective. The central idea is that the problem of LRV estimation can be cast as the problem of estimating the scale of the partial sum process of the residuals divided by the square root of the sample size T . Under standard assumptions, this partial sum process converges in distribution to a Brownian Bridge scaled by the (matrix) square root of the LRV. Limited amounts of dependence and heterogeneity of the underlying disturbances only lead to differences in the local properties of the partial sum process, whereas the overall shape continues to resemble a scaled Brownian Bridge for a large enough sample. A natural measure of robustness with respect to the dependence and heterogeneity of the underlying disturbances of LRV estimators then becomes whether small deviations of the partial sum process from a scaled Brownian Bridge can induce large differences in the estimation of its scale.

Relying on this perspective of the problem, the paper first develops a precise analytical sense in which *any* consistent LRV estimator be said to lack robustness.¹ It is shown that for large enough T , any consistent LRV estimator can be induced to estimate an arbitrary

¹In this paper, a LRV estimator is labelled 'consistent' if it is consistent for the covariance matrix of (demeaned) vector Gaussian White Noise.

positive definite matrix for a partial sum process of the residuals that is arbitrarily close to a standard Brownian Bridge. Although the overall shape of this partial sum process is almost identical to that of a Brownian Bridge, the consistent LRV estimator estimates its scale arbitrarily badly. As an implication, for any consistent LRV estimator, there exist underlying disturbances that satisfy a Functional Central Limit Theorem, yet a test statistic constructed with the consistent LRV estimator has arbitrarily bad size control. Consistent LRV estimators do hence not lead to valid asymptotic inference in the whole class of disturbances that satisfy a Functional Central Limit Theorem.

Motivated by this insight, the paper then develops a quantitative measure of robustness for LRV estimators by considering the degree to which they can be perturbed when confronted with a partial sum process that is close to a scaled Brownian Bridge in some appropriate sense. Specifically, we consider the largest bias that can be induced by a process that is in some well-defined neighborhood of a Brownian Bridge. For the large class of LRV estimators that can be written as a quadratic form in the estimated disturbances—which includes kernel estimators with data-independent bandwidths—explicit formulae of this robustness measures are derived. Given this measure of robustness, we determine the most robust LRV estimator as well as the form of LRV estimators that optimally trade off efficiency against robustness in this class. A slight modification of these optimal LRV estimators yields a family of simple inconsistent LRV estimators that leads to test statistics with an asymptotic student-t and F-distribution under the null hypothesis.

A numerical investigation of the properties of tests based on a number of LRV estimators shows the proposed robustness measure to be a highly reliable predictor for of the size control properties of the tests. The most robust LRV estimator controls size very well, but at the cost of low power compared to less robust estimators. Conditional on their robustness, none of the considered inconsistent LRV estimators is markedly superior to the estimators that leads to test statistics with an asymptotic student-t and F-distribution, such that in practice, robust inference might as well be based on those.

The rest of the paper is organized as follows. Section 2 introduces the general set-up and notation and establishes in which sense any consistent LRV can be said to lack robustness. Section 3 defines a quantitative measure of robustness, determines the robustness of some kernel estimators and derives the form of optimal LRV estimators for this measure within

a broad class of estimators. Section 4 contains the numerical investigation into the power and size properties, and Section 5 summarizes the implications for applied work. Proofs are collected in an appendix.

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

The general model this paper is concerned with is

$$Y_t = \beta + u_t, \quad t = 1, \dots, T \quad (1)$$

where the $m \times 1$ vectors $\{Y_t\}_{t=1}^T$ are observed, the $m \times 1$ vector β is the parameter of interest and $\{u_t\}_{t=1}^T$ is an unobserved random sequence. The aim is to conduct inference about β via the estimator $\hat{\beta} = T^{-1} \sum_{t=1}^T Y_t$. While straightforward, this problem of estimating the mean of a multivariate series captures many of the popular models in time series econometrics. Specifically, any standard exactly identified General Method of Moments model of Hansen (1982) concerning the data $\{y_t^*\}$ with $m \times 1$ moment condition $E[g(y_t^*, \beta)] = 0$ and estimator $\hat{\beta}$ (such that $\sum_{t=1}^T \frac{\partial g}{\partial \beta} \Big|_{(y_t^*, \hat{\beta}^*)} g(y_t^*, \hat{\beta}) = 0$) can be cast in this set-up when the distribution of $\{u_t\}$ is allowed to depend on T in a double array fashion. With $\hat{h}_t = \frac{\partial g}{\partial \beta} \Big|_{(y_t^*, \hat{\beta})}$ the mapping to model (1) is

$$\begin{aligned} Y_t &= \hat{\beta} + \left[T^{-1} \sum_{s=1}^T \hat{h}_s \hat{h}_s' \right]^{-1} \hat{h}_t g(y_t^*, \hat{\beta}) \\ u_t &= \left[T^{-1} \sum_{s=1}^T \hat{h}_s \hat{h}_s' \right]^{-1} \left[\hat{h}_t g(y_t^*, \hat{\beta}) + T^{-1} \sum_{s=1}^T \hat{h}_s g(y_s^*, \beta) \right]. \end{aligned} \quad (2)$$

Let $\{\hat{u}_t\} = \{Y_t - \hat{\beta}\}$ and for $s \in [0, 1]$ define $W_T(s) = T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} u_t$ and $B_T(s) = T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} \hat{u}_t = W_T(s) - \frac{\lfloor sT \rfloor}{T} W_T(1)$, where $\lfloor \cdot \rfloor$ is the largest smaller integer function. With these definitions, $T^{1/2}(\hat{\beta} - \beta) = T^{-1/2} \sum_{t=1}^T u_t = W_T(1)$.

Under sufficient regularity conditions, a Functional Central Limit Theorem and related arguments imply that

$$(W_T(1), B_T(\cdot)) \Rightarrow (\Omega^{1/2} W(1), \Omega^{1/2} B(\cdot)) \quad (3)$$

where \Rightarrow denotes weak convergence of the underlying probability measures, $W(\cdot)$ is a standard $m \times 1$ vector Wiener process, $B(s) = W(s) - sW(1)$ and Ω is invertible. As long as (3) holds, short-run dynamics of $\{u_t\}$ only drive a wedge between the 'local' properties of $B_T(\cdot)$ and the scaled Brownian Bridge $\Omega^{1/2}B(\cdot)$: Positive autocorrelation, for instance, leads through the integration to smoother paths of $B_T(\cdot)$ compared to $\Omega^{1/2}B(\cdot)$, but the overall shape of $B_T(\cdot)$ still closely resembles that of $\Omega^{1/2}B(\cdot)$ for T large enough.

If (3) holds, the distribution of $T^{1/2}(\hat{\beta} - \beta)$ is approximately $\mathcal{N}(0, \Omega)$, such that inference about $\beta = (\beta_1, \dots, \beta_m)'$ can be made with an appropriate estimator $\hat{\Omega}_T$ of Ω . Specifically, one might rely on the usual t- and F-statistics

$$\tilde{t} = \frac{T^{1/2}(\hat{\beta}_j - \beta_{0,j})}{\sqrt{\hat{\Omega}_{T,jj}}} \quad \text{and} \quad \tilde{F} = \frac{T(R'\hat{\beta} - r_0)'(R'\hat{\Omega}_T R)^{-1}(R'\hat{\beta} - r_0)}{d} \quad (4)$$

where $\hat{\Omega}_{T,jj}$ is the j^{th} diagonal element of $\hat{\Omega}_T$, R is a full column rank $m \times d$ matrix, r_0 is $d \times 1$ and $1 \leq d \leq m$. \tilde{F} rejects the null hypothesis of $R'\beta = r_0$ for large values and a one-sided test of $\beta_j = \beta_{0,j}$ against the alternative $\beta_j > \beta_{0,j}$ based on \tilde{t} rejects for large values, too.

When (3) holds and $\hat{\Omega}_T$ is consistent for Ω , then under the null hypothesis \tilde{t} and $d\tilde{F}$ are asymptotically distributed standard normal and χ^2 with d degrees of freedom, respectively. As demonstrated in Kiefer, Vogelsang, and Bunzel (2000), one can continue to base asymptotically valid tests on \tilde{t} and \tilde{F} even when $\hat{\Omega}_T$ is not consistent. In fact, \tilde{t} and \tilde{F} will have asymptotic distributions free of nuisance parameters as long as $(T^{1/2}(\hat{\beta} - \beta), \hat{\Omega}_T)$ has an asymptotic distribution $(\Omega^{1/2}W(1), \Omega_a)$ such that Ω_a is independent of $W(1)$ and with probability one $\Omega^{-1/2}\Omega_a\Omega^{-1/2}$ is a matrix of rank greater or equal to d whose distribution does not depend on Ω or β .

A natural requirement on LRV estimators $\hat{\Omega}_T$ is that they are location and scale invariant. Location invariant LRV estimators remain numerically unaltered through transformations of $\{Y_t\}$ of the form $\{Y_t\} \rightarrow \{Y_t + b\}$ for any $m \times 1$ vector b . Location invariance ensures that estimators of the LRV are (at least asymptotically under (3)) independent of the unknown value of β . Similarly, scale equivariant LRV estimators react to a transformation of the form $\{Y_t\} \rightarrow \{HY_t\}$ for some nonsingular $m \times m$ matrix H by an appropriate rescaling $\hat{\Omega}_T \rightarrow H\hat{\Omega}_T H'$. All common LRV estimators are location invariant and scale equivariant, and this paper focusses exclusively on such estimators.

It is well-known (Lehmann and Casella (1998), chapter 3) that any location invariant

estimator can be written as a function of a maximal invariant, and one maximal invariant is $\{\hat{u}_t\}$. Any location invariant LRV estimator $\hat{\Omega}_T$ hence must be a function of $\{\hat{u}_t\}$. The partial sum process $B_T(\cdot)$ is a one-to-one mapping from $\{\hat{u}_t\}$ to a continuous time process on the unit interval, since $\hat{u}_t = T^{1/2}[B_T(t/T) - B_T((t-1)/T)]$. Any location invariant LRV estimator can hence be written as a function of $\{B_T(tT^{-1})\}_{t=1}^T$, or equivalently as $\hat{\Omega}_T(B_T(\cdot))$. Without loss of generality we assume in the following that $\hat{\Omega}_T(\cdot)$ is a function of $\{B_T(tT^{-1})\}_{t=1}^T$ alone, i.e. $\hat{\Omega}_T(\cdot)$ does not alter with changes of the argument at points that are not simple fractions in T .

When $B_T(\cdot) \Rightarrow \Omega^{1/2}B(\cdot)$ and $T^{1/2}(\hat{\beta} - \beta) = W_T(1) \Rightarrow \Omega^{1/2}W(1)$, one way to intuitively describe the task of a location invariant LRV estimator is hence to say that it estimates the variation of the unobserved endpoint of the partial sum process $W_T(\cdot)$ by considering the magnitude of the partial sum process $B_T(\cdot)$. In the scalar case, if Ω is four, then the partial sum process $B_T(\cdot)$ roughly looks like a Brownian Bridge multiplied by two. All information that a location invariant LRV estimator can exploit about the variation of $W_T(1)$ must stem from the shape of $B_T(\cdot)$, and in the standard asymptotic framework, $B_T(\cdot)$ looks more and more (in an appropriate sense) like a Brownian Bridge scaled by $\Omega^{1/2}$.

At the same time, note that the random matrix $\hat{\Omega}_T(B(\cdot))$ describes the properties of the LRV estimator $\hat{\Omega}_T$ when the underlying disturbances $\{u_t\}$ are distributed i.i.d. $\mathcal{N}(0, I_m)$, simply because the implied residuals $\hat{u}_t = T^{1/2}[B(t/T) - B((t-1)/T)]$ are demeaned Gaussian White Noise. In this perspective, the Brownian Bridge process $B(\cdot)$ plays a pivotal role in the LRV estimation problem: it is equivalent to the most regular set-up of estimating the mean in a model with Gaussian White noise disturbances, and any assumptions that imply a Functional Central Limit Theorem (3) lead to this process asymptotically. This makes the Brownian Bridge process naturally suited as the starting point for considering the robustness of LRV estimators: Whenever the argument of $\hat{\Omega}_T$ is exactly a Brownian Bridge (such that the underlying disturbances are Gaussian White Noise), any reasonable estimator should be well-behaved, and one might call a LRV estimator robust if it remains well-behaved for processes $B_T(\cdot)$ that are very close to a Brownian Bridge. The following Theorem establishes that any LRV estimator that is consistent for the variance of Gaussian White Noise is not robust in this sense. Here and below, $\|\cdot\|_\infty$ denotes the maximum absolute row sum matrix norm.

Theorem 1 *Let H be any nonsingular $m \times m$ matrix. If a scale equivariant LRV estimator $\hat{\Omega}_T$ satisfies $\hat{\Omega}_T(B(\cdot)) \xrightarrow{p} I_m$, then there exists a sequence of Gaussian processes $\bar{B}_T(\cdot)$ such that $\sup_s \|B(s) - \bar{B}_T(s)\|_\infty \xrightarrow{a.s.} 0$ and $\hat{\Omega}_T(\bar{B}_T(\cdot)) \xrightarrow{p} HH'$.*

Inadequate behavior of estimators of the spectral density at a given point under certain circumstances has been established before, cf. Sims (1971), Sims (1972), Faust (1999) and 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 parameterized models.

Theorem 1 shows aberrant properties only of a subset of all possible LRV estimators, namely those that are consistent for the variance of Gaussian White Noise. The LRV estimator of Kiefer, Vogelsang, and Bunzel (2000), for instance, by being a continuous mapping from $B_T(\cdot)$ to the $m \times m$ matrices in the sup norm, does not share the fragility with respect to infinitesimally small deviations of $B_T(\cdot)$ from $B(\cdot)$. In contrast to the literature cited, Theorem 1 does not state that a certain set of assumptions is too weak to enable useful inference; rather it shows that a certain way of conducting inference leads to potentially arbitrary results.

The argument of Theorem 1 concerns the partial sum process $B_T(\cdot)$ only, and nothing is being assumed or said about the (double-array) underlying disturbances $\{u_t\}$ or residuals $\{\hat{u}_t\}$. Closeness of $\bar{B}_T(\cdot)$ to $B(\cdot)$ does, of course, not imply that $\hat{u}_t = T^{1/2}[\bar{B}_T(t/T) - \bar{B}_T((t-1)/T)]$ behaves anything like demeaned Gaussian White Noise; in fact, $\{\hat{u}_t\}$ will surely not satisfy the typical assumptions, since if it did, these assumptions would ensure that usual LRV estimators are consistent for the scale of the Brownian Bridge the partial sum process of $\{\hat{u}_t\}$ converges to. The question is, however, whether this weakens the implication of Theorem 1 for the lack of robustness of consistent LRV estimators. Long-run variance estimators are employed precisely because a researcher is ignorant about the short-run dynamics of $\{u_t\}$. These short-run dynamics translate through the integration $B_T(s) = T^{-1/2} \sum_{t=1}^{\lfloor sT \rfloor} \hat{u}_t$ into the local properties of $B_T(\cdot)$. As long as the overall behavior of $\{u_t\}$ is sufficiently regular to lead to a $W_T(\cdot)$ that behaves like a scaled Wiener process

in the limit, the object of interest—the asymptotic variance of $W_T(1)$ —is well defined.² In this perspective, the very aim of heteroskedasticity and autocorrelation robust variance estimation is to estimate the overall scale of $B_T(\cdot)$ while ignoring its local properties. But $\bar{B}_T(\cdot)$ of Theorem 1 has the same global scale as a standard Brownian Bridge and differs only in its local properties. By potentially yielding arbitrary estimation results, all consistent LRV estimators hence lack robustness against the underlying process whose dependence structure induces these local properties.

Note that the sequence of processes $\bar{B}_T(\cdot)$ of Theorem 1 converges to a Brownian Bridge in the sup-norm. While correctly estimating the scale one of an exact Brownian Bridge, a consistent LRV estimator can hence be made to consistently yield any positive definite matrix when put subject to a sequence of processes that converges weakly to a Brownian Bridge. This implies the following Corollary.

Corollary 1 *There does not exist a scale equivariant LRV estimator that consistently estimates Ω for all double-array processes $\{\hat{u}_{T,t}\}_{t=1}^T$ satisfying $T^{-1/2} \sum_{t=1}^{\lfloor T \rfloor} \hat{u}_{T,t} \Rightarrow \Omega^{1/2} B(\cdot)$.*

Knowledge of the convergence of the partial sum process of a series to a scaled Brownian Bridge alone is hence not enough to allow for a consistent estimation of the scale of the limit process. It is straightforward to generalize this result to the case where the convergence is to a scaled Wiener process. Even when the disturbances are directly observed, Functional Central Limit Theorems, that imply weak convergence to a scaled Wiener process, are hence generically too weak to enable consistent estimation of the scale of the limit process. Recent 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 LRV estimation (see, for instance, de Jong and Davidson (2000) for a recent contribution). But Corollary 1 reveals that this wedge is of substance: The set of all 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 limit process. When one insists on employing a LRV estimator that is consistent for the variance of Gaussian White Noise in the construction of the test statistics (4), the implication

²While the weak convergence (3) does not, of course, imply $E[W_T(1)W_T(1)'] \rightarrow \Omega$, this has no bearing on this discussion since the asymptotic validity of inference based on (4) does not depend on the convergence—or existence—of any moments.

is that size control is arbitrarily bad for some (double-array) process $\{u_t\}$ that satisfies a Functional Central Limit Theorem. Consistent LRV estimators do not lead to valid inference in the whole class of processes that satisfy a Functional Central Limit Theorem, and lack robustness in this sense.

One could ask at this point whether the additional assumptions typically employed to obtain the consistency of standard LRV estimators are plausible for most data sets, or question the interpretation of double-array asymptotic arguments, and judge the relevance of Theorem 1 and Corollary 1 on these grounds. While a consideration of the properties of the underlying disturbances $\{u_t\}$ is important to motivate approximations based on the Functional Central Limit Theorem, such a discussion would not address the central point. The objective of asymptotic thought experiments is to generate useful small sample approximations. It is well documented (Haan and Levin (1997), for instance) that the approximation provided by the standard thought experiment in which Ω can be estimated consistently becomes quite poor for amounts of time series dependence that are perfectly plausible for most macroeconomic data in levels. The approximations are poor although the small sample data generating processes considered in such studies technically satisfy the assumptions required for consistency of standard LRV estimators. Rather than to abandon asymptotic arguments, a more constructive reaction to this finding is to alter the asymptotic thought experiment in order to render it more relevant to the small sample problem.

The standard asymptotic thought experiment entails that for large enough T , an infinite amount of local variation in $B_T(\cdot)$ accurately reflects its overall scale. But in any small sample, there is necessarily only a limited amount of local variation, and this variation is determined by the short-run dynamics of $\{u_t\}$. For well behaved, not too small samples, the standard thought experiment can still yield useful approximations, since there is enough exploitable local variation in $B_T(\cdot)$ to pin down Ω with little uncertainty. For samples with a high degree of persistency or heterogeneity, however, there is only a very limited amount of global variation that is comparable to the variation in $W_T(1)$. An asymptotic thought experiment that captures this property is one where the local properties of $B_T(\cdot)$ remain uninformative about the variation of $W_T(1)$. If inference procedures behave badly in such an asymptotic thought experiment, then this suggests that they also behave badly in small samples with strong dependence.

In this perspective, it is highly disconcerting to find that standard inference based on consistent LRV estimators can be made to completely break down by a seemingly marginal extension of the usual asymptotic thought experiment to include all processes $\{u_t\}$ that satisfy a Functional Central Limit Theorem, i.e. for which (3) holds. After all, the convergence $B_T(\cdot) \Rightarrow \Omega^{1/2}B(\cdot)$ in the appropriate metric function space does imply that all local variation of $B_T(\cdot)$ that can be described in this space eventually becomes indistinguishable of the the variation of $\Omega^{1/2}B(\cdot)$. As demonstrated by Kiefer, Vogelsang, and Bunzel (2000), it is perfectly possible to construct methods of inference that remain asymptotically valid whenever (3) holds. The contribution of this section is to show that one must *necessarily* abandon the attempt to consistently estimate the LRV in order to gain such robustness.

3 A Quantitative Measure of Robustness

The aim of LRV estimation in model (1) is to assess the variation of $T^{-1/2} \sum_{t=1}^T u_t = W_T(1)$. To this end, any location invariant LRV estimator solely relies on the properties of the partial sum process $B_T(\cdot)$ of the residuals $\{\hat{u}_t\}$. When $\{u_t\}$ is autocorrelated, the local variation of $B_T(\cdot)$ does not correspond to the variation in $W_T(1)$, so LRV becomes the problem of assessing the overall variation of $B_T(\cdot)$ independent of its local properties. The analysis of section 2 shows that any LRV estimator that is consistent for the variance of Gaussian White Noise necessarily depends on the local properties of $B_T(\cdot)$ in such a strong way that it can produce arbitrarily bad estimates of the variation of $W_T(1)$ even though $B_T(\cdot)$ converges weakly to a Brownian Bridge of the same scale as the limit of $W_T(1)$.

An arguably necessary part of a definition of a 'robust' LRV estimator $\hat{\Omega}_T$ is hence that for all sequences $B_T(\cdot)$ that converge weakly to a Brownian Bridge $B(\cdot)$, $\hat{\Omega}_T(B_T(\cdot))$ must have the same asymptotic distribution as $\hat{\Omega}_T(B(\cdot))$. Clearly, any LRV estimator that can be written as a continuous mapping from $B_T(\cdot)$ to the $m \times m$ matrices (in appropriate metric spaces) will have this property by the Continuous Mapping Theorem. Any of these estimators will yield asymptotically justified inference whenever the disturbances satisfy a Functional Central Limit Theorem—which, as argued below Corollary 1, is a larger set than the set of all processes that in addition allow consistent estimation of the scale of the limit process.

While necessary, such a requirement is digital and does not allow the comparison of the

robustness of two alternative continuous mappings. What is more, one can easily construct a continuous mapping that results in an arbitrarily small mean-squared error of the LRV estimator.³ While formally 'robust', such estimators will behave just like their consistent counterparts in small samples. The reason is that the asymptotic thought experiment of considering the properties of LRV estimators when (3) holds still restricts any disparities of the local variations of $B_T(\cdot)$ from its overall scale to be minute: Under (3), all local properties of $B_T(\cdot)$ that can be described by continuous mappings are eventually identical to those of $\Omega^{1/2}B(\cdot)$. This results in a poor approximation for samples with strong time series dependence, for which only the global variation of $B_T(\cdot)$ is comparable to the variation in $W_T(1)$.

In order to make further progress, the usual asymptotic thought experiment must hence be enlarged even further to include sequences $B_T(\cdot)$ which are close to a Brownian Bridge in some sense, but whose local properties are quite arbitrary. Specifically, this is achieved by contaminating a standard Brownian Bridge $B(\cdot)$ by independent continuous time processes that predominantly perturb the local properties of the resulting process $\tilde{B}_T(\cdot)$. Since the local properties of $B_T(\cdot)$ reflect the unknown and essentially arbitrary short-run dynamics of $\{u_t\}$, a robust LRV estimator should yield similar results for *any* such $\tilde{B}_T(\cdot)$.

Definition 2 *Define Gaussian Noise of size δ as any $m \times 1$ vector continuous time Gaussian process $S_T(s) = (S_{T1}(s), \dots, S_{Tm}(s))'$ on $s \in [0, 1]$ with $S_T(0) = S_T(1) = 0$ such that the spectrum of the covariance matrix of $T^{-1/2}(S_{Ti}(T^{-1}), S_{Ti}(2T^{-1}), \dots, S_{Ti}(1))'$ is bounded by δ for all $i = 1, \dots, m$. For any $\delta > 0$, define the contamination neighborhood $\mathfrak{C}_T(\delta)$ of size δ of the process $B(\cdot)$ as the set of all processes $\tilde{B}_T(\cdot)$ satisfying*

$$\tilde{B}_T(\cdot) + \tilde{S}_T(\cdot) = B(\cdot) + S_T(\cdot),$$

where $S_T(\cdot)$ is independent of $B(\cdot)$, $\tilde{S}_T(\cdot)$ is independent of $\tilde{B}_T(\cdot)$ and $S_T(\cdot)$ and $\tilde{S}_T(\cdot)$ are Gaussian Noise of size δ . When $E[\hat{\Omega}_T(\tilde{B}_T(\cdot))]$ exists for all $\tilde{B}_T(\cdot) \in \mathfrak{C}_T(\delta)$, the robustness function $R_T(\delta)$ of the scale equivariant LRV estimator $\hat{\Omega}_T$ is defined as

$$R_T(\delta) = \sup_{\tilde{B}_T(\cdot) \in \mathfrak{C}_T(\delta)} \frac{\text{tr} \left| E[\hat{\Omega}_T(\tilde{B}_T(\cdot))] - I_m \right|}{\delta m},$$

³The variance of the asymptotically unbiased $\hat{\Omega}_{UA}(p)$ as defined in (7) below, for instance, can be made arbitrarily small by choosing p large.

where the absolute value $|\cdot|$ of a matrix is the matrix of the absolute values of its elements.

The idea of examining the properties of inference procedures in contaminated benchmark models has, of course, a long history in statistics. It is particularly attractive for the problem of LRV estimation, since the very aim of LRV estimation is to be robust against deviations from the benchmark of independent and identically distributed disturbances. Contaminations of this benchmark are hence a generic part of the LRV estimation exercise. Viewing LRV estimation as the problem of estimating the scale of the partial sum process $B_T(\cdot)$ as promoted here allows the formalization of this idea in Definition 2. In the standard terminology of the statistical robustness literature, $R_T(\delta)$ describes something akin to the Gross-Error Sensitivity: It is the largest bias in $\hat{\Omega}_T$ that can be induced by contaminating the benchmark model of Gaussian White Noise disturbances.

The processes $\tilde{B}_T(\cdot)$ in the contamination neighborhood of the Brownian Bridge $B(\cdot)$ can be thought of as noisy versions of a Brownian Bridge, where both 'positive' noise $S_T(\cdot)$ and 'negative' noise $\tilde{S}_T(\cdot)$ are allowed: With $\tilde{S}_T(s) = 0 \forall s$, a nonzero $S_T(\cdot)$ increases the local variability of $\tilde{B}_T(\cdot)$ compared to $B(\cdot)$, giving it a more ragged shaped. With $S_T(s) = 0 \forall s$ and a nonzero $\tilde{S}_T(\cdot)$, it is only after having added noise to $\tilde{B}_T(\cdot)$ that one obtains a Brownian Bridge $B(\cdot)$, so $\tilde{B}_T(\cdot)$ is in general smoother than $B(\cdot)$. The neighborhood encompasses arbitrary combinations of this two forms of contaminations, where their magnitude is controlled through the largest eigenvalue of the covariance matrices of $S_{T_i}(\cdot)$ and $\tilde{S}_{T_i}(\cdot)$ evaluated at $T^{-1}, 2T^{-1}, \dots, 1$ for all $i = 1, \dots, m$. This construction allows a multitude of perturbations of the local properties of the Brownian Bridge, which correspond to many alternative short-run dynamics and forms of heterogeneity of the underlying disturbances.

Interesting examples of $\tilde{B}_T(\cdot)$ for $m = 1$ include continuous record sampling from a continuous time process whose covariance kernel is close to the covariance kernel of a Brownian Bridge. Consider, for instance, the processes

$$\begin{aligned} M(s) &= (1 - \tau\omega^2)^{1/2}W_1(s) + \omega\tau^{1/2}W_2(s \wedge \tau) + \omega(1 - \tau)^{1/2}W_3(s \wedge \tau) \\ \tilde{S}_T(s) &= \omega\tau^{1/2}\mathbf{1}[s > \tau]W_4(s - \tau) - \omega\tau^{1/2}sW_4(1 - \tau) \\ S_T(s) &= \omega(1 - \tau)^{1/2}W_3(s \wedge \tau) - \omega(1 - \tau)^{1/2}sW_3(\tau) \end{aligned}$$

for some $0 < \tau < 1$, $0 \leq \omega < \tau^{1/2}$, and independent Wiener processes $W_i(\cdot)$, $i = 1, \dots, 4$, where $\mathbf{1}[\cdot]$ is the indicator function. With $W(s) = (1 - \tau\omega^2)^{1/2}W_1(s) + \omega\tau^{1/2}W_2(s \wedge \tau) +$

$\omega\tau^{1/2}\mathbf{1}_{[s > \tau]}W_4(s - \tau)$, this generates $\tilde{B}_T(s) = M(s) - sM(1)$, i.e. it corresponds to underlying disturbances $\{u_t\}$ that are independent Gaussian with a variance that decreases from $(1 + (1 - \tau)\omega^2)$ to $(1 - \tau\omega^2)$ at $t = \lceil\tau T\rceil$. Alternatively, consider $\tilde{S}_T(s) = 0 \forall s$ and $S_T(s) = Q \sin(2\pi ns)$ for some positive integer n and $Q \sim \mathcal{N}(0, \sigma_Q^2)$. Such a model captures the behavior of LRV estimators when the data contains a seasonal or cycling component. Finally, consider the case of $S_T(s) = 0 \forall s$ and $\tilde{S}_T(s) = G_\gamma(s) - sG_\gamma(1)$ for some $\gamma > 0$, where

$$G_\gamma(s) = (1 - e^{-\gamma s})Z_G + \int_0^s e^{-\gamma(s-r)}dW_G(r)$$

and $Z_G \sim \mathcal{N}(0, (2\gamma)^{-1})$ independent of the standard Wiener process $W_G(\cdot)$. Interestingly, $\tilde{B}_T(s)$ is then given by a scaled, integrated demeaned stationary Ornstein-Uhlenbeck process $J_\gamma(\cdot)$ with parameter γ , i.e. $\tilde{B}_T(s) = \gamma \int_0^s J_\gamma(r)dr - s\gamma \int_0^1 J_\gamma(r)dr$ and $J_\gamma(s) = e^{-\gamma s}Z_J + \int_0^s e^{-\gamma(s-r)}dW_J(r)$, where $Z_J \sim \mathcal{N}(0, (2\gamma)^{-1})$ independent of the standard Wiener process $W_J(\cdot)$. As $\tilde{B}_T(\cdot)$ and $\tilde{S}_T(\cdot)$ are independent and Gaussian, this can be checked by computing the sum of their covariance kernels. In other words, the set-up of Definition 2 captures the popular modelling strategy of letting $\{u_t\}$ be local-to-unity in the sense of Chan and Wei (1987) and Phillips (1987). It is known (Hochstadt (1973), Chapter 6) that as the sampling interval $1/T$ converges to zero, the eigenvalues of the covariance matrix of a discretely sampled continuous time process (scaled by T^{-1}) converge uniformly to those of the (continuous) covariance kernel of the continuous time process. In the examples given here, for large enough T , the minimal size of the the contamination neighborhood δ hence becomes arbitrarily close to the largest eigenvalue of the (sample size independent) covariance kernels of $S_T(\cdot)$ and $\tilde{S}_T(\cdot)$.

The robustness measure $R_T(\delta)$ focusses on the maximal bias of the LRV estimator when the argument is a contaminated Brownian Bridge $\tilde{B}_T(\cdot)$. This construction renders the following analyses mathematically tractable. At the same time, the expectation is only one of many parameters describing the distribution of $\hat{\Omega}_T$, and robustness of the expectation is of course not sufficient for the whole distribution of $\hat{\Omega}_T$ to be robust to the contaminations. On the one hand, a focus on the bias can be motivated by the finding of Andrews and Monahan (1992) that the bias properties of LRV estimators seem to be mainly responsible for size control of the tests (4). On the other hand, Theorem 5 below shows that, at least for a broad class of LRV estimators, a uniformly bounded robustness function (in T) does imply that the distributions of the LRV estimator for an exact Brownian Bridge and a contaminated

Brownian Bridge are arbitrarily close for a small enough contamination.

Given the discussion in Section 2, consistent LRV estimators should not be classified as robust by the robustness measure defined here. And this is indeed the case.

Theorem 3 *For any nonnegative definite, scale equivariant LRV estimator $\hat{\Omega}_T$ satisfying $\hat{\Omega}_T(B(\cdot)) \xrightarrow{p} I_m$, $\sup_T R_T(\delta) = \infty$ for all $\delta > 0$.*

Much of the following discussions will concern a specific class of location invariant and scale equivariant LRV estimators. Let M_e be the $T \times T$ matrix $M_e = I_T - T^{-1}ee'$, where e is a $T \times 1$ vector of ones, and let \hat{u} be the $T \times m$ matrix with \hat{u}'_t in row t . Any LRV estimator that can be written in the form

$$\hat{\Omega}_A = \frac{\hat{u}'A\hat{u}}{\text{tr}[M_eA]} \quad (5)$$

for A a $T \times T$ symmetric, nonnegative definite matrix will be called a *quadratic LRV estimator*. The assumption of nonnegative definiteness of A ensures that it is impossible to obtain a negative definite estimate of the LRV. The denominator $\text{tr}[M_eA]$ is a normalization that ensures that for $\{u_t\}$ with covariance matrix Ω and no autocorrelation, $E[\hat{\Omega}_A] = \Omega$.

The focus of the literature so far have been kernel estimators, which are usually defined as

$$\hat{\Omega} = \sum_{j=-T+1}^{T-1} k(jb_T^{-1})\hat{\Gamma}(j)$$

with $\hat{\Gamma}(j) = T^{-1}\sum_{t=j+1}^T \hat{u}_t\hat{u}'_{t-j}$ for $j \geq 0$ and $\hat{\Gamma}(j) = \hat{\Gamma}(-j)'$ for $j < 0$, where b_T is the bandwidth and $k : \mathbb{R} \rightarrow \mathbb{R}$ is an even and continuous kernel function satisfying $k(0) = 1$ and $\int_0^\infty |k(s)|ds < \infty$. Note that all kernel LRV estimators with even kernel $k(\cdot)$ and data-independent bandwidth b_T are quadratic LRV estimators, where A is a bandmatrix of the form

$$A = T^{-1} \begin{pmatrix} k(0) & \cdot & \cdot & \cdots & \cdot \\ k(b_T^{-1}) & k(0) & \cdot & \cdots & \cdot \\ k(2b_T^{-1}) & k(b_T^{-1}) & k(0) & \cdots & \cdot \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k((T-1)b_T^{-1}) & k((T-2)b_T^{-1}) & k((T-3)b_T^{-1}) & \cdots & k(0) \end{pmatrix}.$$

Specifically, we will be concerned with two classes of kernel estimators: On the one hand, kernel estimators whose kernel $k(\cdot)$ is twice differentiable at the origin, and on the other

hand the Bartlett LRV estimator, whose kernel is given by $k(x) = 1 - |x|$ for $|x| < 1$ and $k(x) = 0$ elsewhere.

Theorem 4 *Let F be the $T \times T$ matrix with zeros above the diagonal and ones elsewhere and let $D = F^{-1}$.*

(a) *The robustness function $R_T(\delta)$ of quadratic LRV estimators $\hat{\Omega}_A$ (5) are constant with robustness coefficient*

$$R_T = R_T(\delta) = \frac{\text{tr}[\text{diag}(I_{T-1}, 0)D'AD]}{\text{tr}[M_e A]} T^2.$$

(b) *For a bandwidth b_T satisfying $b_T = o(T)$ and $b_T \rightarrow \infty$, twice differentiable kernel LRV estimators have robustness coefficients $R_T = -k''(0)T^2b_T^{-2} + o(T^2b_T^{-2})$, and Bartlett estimators have robustness coefficient $R_T = -2T^2b_T^{-1} + o(T^2b_T^{-1})$.*

(c) *For $b_T = cT$, $c \leq 1$, twice differentiable kernel LRV estimators have robustness coefficient*

$$R_T = \frac{-k''(0)/c^2}{1 - 2 \int_0^1 (1-s)k(s/c)ds} + o(1)$$

and the Bartlett LRV estimator

$$R_T = \frac{6}{3c - 3c^2 + c^3} T + o(T).$$

Part (a) of Theorem 4 establishes the robustness function $R_T(\delta)$ of quadratic LRV estimators $\hat{\Omega}_A$ as a function of A . Interestingly, the robustness function is constant, such that the robustness properties of quadratic LRV estimators in the sense of Definition 2 reduce to a single number, the robustness coefficient R_T . The maximal bias of $\hat{\Omega}_A$ is linear in the magnitude δ of the considered contamination neighborhood of the Brownian Bridge, with the coefficient being R_T .

Parts (b) and (c) establish the robustness coefficients for twice differentiable kernel and Bartlett LRV estimators for bandwidths b_T that are $b_T = o(T)$ and $b_T = cT$, respectively. Apart from the normalization that ensures $E[\hat{\Omega}_T(B(\cdot))] = I_m$, the robustness coefficient of twice differentiable kernel estimators depends on the properties of the kernel at the origin alone: The steeper the kernels at the origin, the less robust the resulting estimator as measured by R_T . At the same time, the Bartlett estimator is an order of b_T less robust than twice differentiable kernel estimators.

When $b_T = cT$, Kiefer and Vogelsang (2003) show that the resulting LRV estimators are continuous mappings from $B_T(\cdot)$ to the $m \times m$ real matrices. Interestingly, in this case R_T is uniformly bounded for twice differentiable kernel estimators only, whereas the Bartlett estimator still has robustness coefficient of order T . It is hence possible to induce a nonvanishing bias in the Bartlett estimator with $b_T = T$ even in a shrinking δ_T -neighborhood of the Brownian Bridge, at least as long $T^{1/2}\delta_T$ does not converge to zero. This suggests that the Bartlett estimator is more susceptible to variations of the short-run dynamics of the underlying disturbances than twice differentiable kernel estimators.

While R_T is uniformly bounded for twice differentiable kernel estimators with $b_T = cT$, this only guarantees that the expectation of these estimators remains close to I_m when confronted with a contaminated Brownian Bridge in a small neighborhood of a standard Brownian Bridge. The following Theorem ensures that uniform boundedness of R_T is enough to guarantee robustness of the whole distribution for small contamination neighborhoods.

Theorem 5 *For any process $\tilde{B}_T(\cdot)$ in the contamination neighborhood of size δ of $B(\cdot)$ and quadratic LRV estimator $\hat{\Omega}_T$ with robustness coefficient R_T ,*

$$E[|\hat{\Omega}_T(\tilde{B}_T(\cdot)) - \hat{\Omega}_T(B(\cdot))|_\infty] \leq 2m^2[3\delta R_T + \sqrt{\delta R_T} + \sqrt{\delta R_T + (\delta R_T)^2}].$$

In the scalar examples described below Definition 2, the covariance kernels of the processes $S_T(\cdot)$ and $\tilde{S}_T(\cdot)$ converge to zero uniformly as $\omega \rightarrow 0$, $\sigma_Q^2 \rightarrow 0$ and $\gamma \rightarrow \infty$. For small enough ω and σ_Q^2 or large enough γ , the resulting processes $\tilde{B}_T(\cdot)$ are thus within an arbitrarily small contamination neighborhood of the Brownian Bridge, at least for large enough T . Theorem 5 hence implies that the asymptotic distribution of quadratic LRV estimators with uniformly bounded R_T remains robust for these processes. This in particular implies that inference based on (4) using such LRV estimators for a local-to-unity process $\{u_t\}$ becomes an accurate asymptotic approximation for a large enough local-to-unity parameter γ , since $\text{Var}[\gamma \int_0^1 J_\gamma(s) ds] = 1 + (1 - e^{-\gamma})/\gamma \rightarrow 1$ as $\gamma \rightarrow \infty$. Note that this is not the case for the usual LRV estimators: As shown by Müller (2002), kernel estimators with an arbitrary bandwidth choice $b_T = o_p(T)$ underestimate the variance of $W_T(1)$ by a factor of b_T^2/T^2 in local-to-unity models for any local-to-unity parameter $\gamma > 0$, resulting in arbitrarily bad size control of the tests (4).

With a quantitative robustness measure at hand, the question arises which LRV estimator is the most robust one. The following Theorem provides the answer in the class of

quadratic LRV estimators, along with the form of the LRV estimators that optimally trade off robustness against efficiency. Comparable to Andrews (1991), the variance of the LRV estimators is used as a proxy for their inefficiency when employed in the tests (4).

Theorem 6 *Let $v(l)$ be the eigenvector associated with the l^{th} largest eigenvalue $r_l = [4T^2 \sin(\pi l/(2T))^2]^{-1} = (\pi l)^{-2} + o(1)$ of the $T \times T$ matrix $T^{-2}FM_eF'$, such that the t^{th} element of $v(l)$ is $v(l)_t = \sqrt{2/T} \sin(\pi l t T^{-1})$.*

(a) *The quadratic LRV estimator $\hat{\Omega}_{MR}$ with the smallest robustness coefficient $R_T = r_1^{-1} = \pi^2 + o(1)$ results when $A_{MR} = F'v(1)v(1)'F/(r_1T^2)$.*

(b) *For any $\lambda \geq r_1^{-1}$, let p_λ such that $r_{p_\lambda}^{-1} \leq \lambda < r_{p_\lambda+1}^{-1}$ and define $w_l = (\lambda - r_l^{-1})/(p_\lambda \lambda - \sum_{j=1}^{p_\lambda} r_j^{-1})$ for $l = 1, \dots, p_\lambda$. Let K be any nonzero, nonnegative definite $m^2 \times m^2$ matrix. The quadratic LRV estimators $\hat{\Omega}_{RE}(\lambda)$ that minimizes $\text{tr Var}[K \text{vec} \hat{\Omega}_T(B(\cdot))]$ for a given robustness coefficient $R_T = \sum_{l=1}^{p_\lambda} w_l r_l^{-1} \geq r_1^{-1}$ results when $A = A_{RE}(\lambda)$ is of the form*

$$A_{RE}(\lambda) = \sum_{l=1}^{p_\lambda} w_l F'v(l)v(l)'F/(r_l T^2).$$

Part (a) of Theorem 6 establishes the form of the most robust quadratic LRV estimator $\hat{\Omega}_{MR}$. Its weighting matrix A_{MR} is not band-diagonal, and $\hat{\Omega}_{MR}$ hence does not correspond to a kernel estimator. A_{MR} has a single non-zero eigenvalue. For $m = 1$ and independent standard normal disturbances $\{u_t\}$, $\hat{u} \sim \mathcal{N}(0, M_e)$, such that $T^{-1/2}F\hat{u} \sim \mathcal{N}(0, T^{-1}FM_eF')$. $T^{-1}FM_eF'$ is the covariance matrix of the small sample analogue of a Brownian Bridge, and $v(1)$ is the eigenvector associated with its largest eigenvalue. Of all vectors of unit length, a weighting with $v(1)$ extracts the largest variance of $T^{-1/2}F\hat{u}$, thereby capturing the characteristic of $T^{-1/2}F\hat{u}$ that is most difficult to perturb. The t^{th} element of $T^{-1}v(1)'FM_e$ is a scalar multiple of $\sqrt{2/T} \cos(\pi l(t-1/2)/T)$. From a spectral perspective, $\hat{\Omega}_{MR}$ can hence be interpreted as extracting the lowest frequency of the Discrete Cosine Transform Type II that \hat{u} contains information about.

As robustness is traded off against efficiency, the estimator $\hat{\Omega}_{RE}(\lambda)$ extracts information also from higher frequencies. The estimator is a weighted average of the sample variance of $T^{-1}r_l^{-1/2}v(l)'F\hat{u}$, $l = 1, \dots, p_\lambda$, where $v(l)$ are the eigenvectors of $T^{-1}FM_eF'$ associated with the p_λ largest eigenvalues $\{Tr_l\}_{l=1}^{p_\lambda}$. The weights are a smooth function of λ , and put less weight on higher frequencies. When $\{u_t\}$ is not autocorrelated, $v(l)'F\hat{u}$ and $v(j)'F\hat{u}$ for $l \neq j$ are uncorrelated. As $T \rightarrow \infty$, $\sqrt{T}v(l)_{[Ts]} \rightarrow \phi_l(s) \equiv \sqrt{2} \sin(\pi ls)$. Note that

$\{\sqrt{2}\sin(\pi ls)\}_{l=1}^{\infty}$ are the eigenfunctions of the covariance kernel of a scalar Brownian Bridge $E[B(s)B(r)] = s \wedge r - sr$ corresponding to the eigenvalues $(l\pi)^{-2}$. The Karhunen-Loeve expansion of the Brownian Bridge is hence given by

$$B(s) = \sum_{l=1}^{\infty} \frac{\sqrt{2}\sin(\pi ls)}{\pi l} \xi_l \quad (6)$$

where $\{\xi_l\}$ are independent standard normal, cf. Phillips (1998). The right-hand side of (6) is known to converge uniformly with probability one—see Itô and Nisio (1968). An alternative way to write $\hat{\Omega}_{RE}(\lambda)$ is $\hat{\Omega}_{RE}(\lambda) = \sum_{l=1}^{p_\lambda} w_l \hat{\xi}_l \hat{\xi}_l'$, where $\hat{\xi}_l = T^{-1} r_l^{-1/2} v(l)' F \hat{u} = \sqrt{2}(\pi l) \int_0^1 \sin(\pi ls) B_T(s) ds + o_p(1)$. When $B_T(\cdot) \Rightarrow \Omega^{1/2} B(\cdot)$ the Continuous Mapping Theorem implies that $\hat{\xi}_j$ and $\hat{\xi}_l$ for $l \neq j$ are asymptotically independent distributed $\mathcal{N}(0, \Omega)$ for any finite l and j . But for the inference based on $\hat{\Omega}_{RE}(\lambda)$ to be asymptotically justified, only $\{\hat{\xi}_l\}_{l=1}^{p_\lambda}$ have to jointly converge to independent $\mathcal{N}(0, \Omega)$ variates. Whenever this is the case, $\hat{\Omega}_{RE}(\lambda)$ converges weakly to a weighted sum of independent Wishart random matrices with one degree of freedom. The relative loss in robustness of $\hat{\Omega}_{RE}(\lambda)$ compared to $\hat{\Omega}_{MR}$ hence consists of the requirement of convergence of $\{\hat{\xi}_l\}_{l=1}^{p_\lambda}$ rather than $\hat{\xi}_1$ only. The larger λ and p_λ , the more frequencies of the partial sum process of $\{\hat{u}_t\}$ have to behave like a Brownian Bridge in order to obtain accurate approximations with inference based on $\hat{\Omega}_{RE}(\lambda)$.

The Bartlett estimator with $b_T = T$, denoted $\hat{\Omega}_{BT}(1)$, can also be written as a weighted function of $\{\hat{\xi}_l \hat{\xi}_l'\}$. Since $\hat{\Omega}_{BT}(1)$ is proportional to $T^{-1} \sum_{t=1}^T B_T(tT^{-1}) B_T(tT^{-1})'$ (cf. Kiefer and Vogelsang (2002a)) and $B_T(tT^{-1}) = \sum_{l=1}^{T-1} r_l^{1/2} v(l)_t \hat{\xi}_l$ for any $t = 1, \dots, T$, it follows from the orthonormality of the eigenvectors that $\hat{\Omega}_{BT}(1)$ is proportional to $\sum_{l=1}^{T-1} r_l \hat{\xi}_l \hat{\xi}_l'$. For inference based on $\hat{\Omega}_{BT}(1)$ to be asymptotically justified, the whole sequence $\{\hat{\xi}_l\}_{l=1}^{\infty}$ has to converge weakly to a series of independent $\mathcal{N}(0, \Omega)$ variates. Arguably, this dependence on the whole series $\{\hat{\xi}_l\}_{l=1}^{\infty}$ makes $\hat{\Omega}_{BT}(1)$ an order of magnitude more susceptible to deviations of $B_T(\cdot)$ from a Brownian Bridge compared to $\hat{\Omega}_{RE}(\lambda)$ for any λ , as reflected in an unbounded robustness coefficient of $\hat{\Omega}_{BT}(1)$.

An attractive slight modification to $\hat{\Omega}_{RE}(\lambda)$ is the unweighted average of $\{\hat{\xi}_l \hat{\xi}_l'\}_{l=1}^p$ for some integer $p \geq 1$, i.e.

$$\hat{\Omega}_{UA}(p) = p^{-1} \sum_{l=1}^p \hat{\xi}_l \hat{\xi}_l'. \quad (7)$$

Compared to the optimal LRV estimators $\hat{\Omega}_{RE}(\lambda)$, $\hat{\Omega}_{UA}(p)$ puts somewhat more weight on the sample variances of higher frequency components in $\{\hat{u}_t\}$. In return, when $B_T(\cdot) \Rightarrow \Omega^{1/2} B(\cdot)$,

the asymptotic distribution of $p\hat{\Omega}_{UA}(p)$ is Wishart with p degrees of freedom.

Asymptotically justified inference based on (4) using $\hat{\Omega}_{UA}(p)$ can hence proceed just as standard small sample Gaussian inference for a sample of size $p + 1$: Under the null hypothesis, the asymptotic distribution of \tilde{t} is student- t with p degrees of freedom, and for $p \geq d$, $d\tilde{F}$ has the Hotelling- T^2 distribution with parameters d and p . Under the null hypothesis, the asymptotic distribution of $(p+1-d)\tilde{F}/p$ is hence F with d degrees of freedom in the numerator and $(p+1-d)$ degrees of freedom in the denominator—cf. Hotelling (1931). Critical values for these distributions are, of course, readily available from standard tables.

4 Quantitative Evaluation

This section numerically explores the properties of inference based on various LRV estimators. The aim is to assess the relevance of the analytical considerations in Sections 2 and 3 for some standard small sample Data Generating Processes. There are two main questions: On the one hand, do the standard t and F -tests (4) based on the theoretically motivated most robust LRV estimator $\hat{\Omega}_{MR}$ actually control size much better than inference based on alternative LRV estimators? On the other hand, is the robustness measure R_T a useful predictor of the small sample size control properties of these tests?

The analysis focusses on the following types of inconsistent LRV estimators: the most robust estimator $\hat{\Omega}_{MR}$, the estimators $\hat{\Omega}_{RE}(\lambda)$ that optimally trade-off robustness and efficiency as derived in Theorem 6, the asymptotically Wishart distributed estimator $\hat{\Omega}_{UA}(p)$ defined in (7), the quadratic spectral kernel LRV estimator $\hat{\Omega}_{QS}(c)$ with a bandwidth equal to c times sample size and the Bartlett estimator $\hat{\Omega}_{BT}(c)$ with bandwidth equal to c times sample size. All of these estimators can be written (up to an $o_p(1)$ term) as continuous functions from $B_T(\cdot)$, such that asymptotic distributions of the tests \tilde{t} and \tilde{F} (4) under (3) under both the null and local alternatives follows from the Continuous Mapping Theorem—cf. Kiefer and Vogelsang (2003) and the discussion above.

Prior to considering size control properties of the tests (4) based on various LRV estimators, we first analyze their power. Specifically, we consider their asymptotic local power of the null hypothesis $\beta_j = \beta_{0,j}$ based on \tilde{t} as defined in (4) when $\beta_j - \beta_{0,j} = \kappa_0 T^{-1/2} \sqrt{\Omega_{jj}}$. Figure 1 depicts the asymptotic local power of a 5% level, two-sided t -test using the various inconsistent LRV estimators, along with the power envelope that arises through substitution

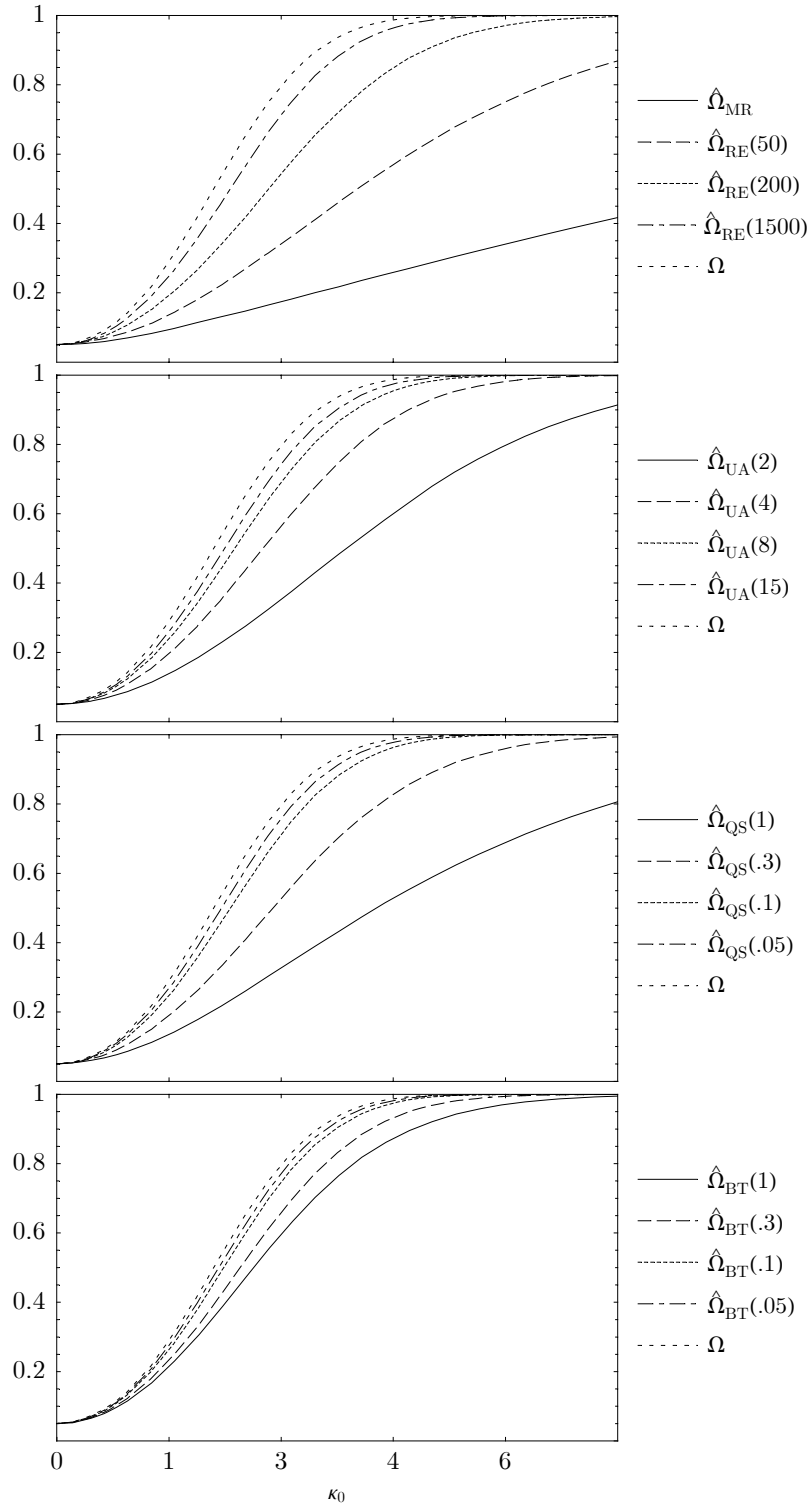


Figure 1: Local Asymptotic Power of 5%-Level two-sided t-tests

Table 1: Robustness and Efficiency of Long-Run Variance Estimators

| | R_T | Pt-25 | Pt-50 | Pt-75 |
|---------------------------|-----------------|-------|-------|-------|
| $\hat{\Omega}_{MR}$ | $9.87 + o(1)$ | 9.91 | 19.2 | 30.7 |
| $\hat{\Omega}_{RE}(50)$ | $16.0 + o(1)$ | 3.04 | 3.67 | 4.47 |
| $\hat{\Omega}_{RE}(200)$ | $49.1 + o(1)$ | 1.72 | 1.82 | 1.91 |
| $\hat{\Omega}_{RE}(1500)$ | $320.2 + o(1)$ | 1.21 | 1.22 | 1.23 |
| $\hat{\Omega}_{UA}(2)$ | $24.7 + o(1)$ | 2.92 | 3.39 | 3.86 |
| $\hat{\Omega}_{UA}(4)$ | $74.0 + o(1)$ | 1.64 | 1.73 | 1.78 |
| $\hat{\Omega}_{UA}(8)$ | $251.7 + o(1)$ | 1.27 | 1.28 | 1.30 |
| $\hat{\Omega}_{UA}(15)$ | $815.8 + o(1)$ | 1.13 | 1.15 | 1.14 |
| $\hat{\Omega}_{QS}(1)$ | $14.6 + o(1)$ | 3.17 | 4.13 | 5.62 |
| $\hat{\Omega}_{QS}(.3)$ | $47.7 + o(1)$ | 1.75 | 1.90 | 2.01 |
| $\hat{\Omega}_{QS}(.1)$ | $323.3 + o(1)$ | 1.21 | 1.22 | 1.23 |
| $\hat{\Omega}_{QS}(.05)$ | $1211.4 + o(1)$ | 1.09 | 1.10 | 1.10 |
| $\hat{\Omega}_{BT}(1)$ | $6T + o(T)$ | 1.46 | 1.57 | 1.70 |
| $\hat{\Omega}_{BT}(.3)$ | $9.13T + o(T)$ | 1.31 | 1.37 | 1.41 |
| $\hat{\Omega}_{BT}(.1)$ | $22.1T + o(T)$ | 1.12 | 1.13 | 1.13 |
| $\hat{\Omega}_{BT}(.05)$ | $42.1T + o(T)$ | 1.06 | 1.07 | 1.06 |

of the LRV estimator with the population value. Similar pictures for the power of a one-sided t -test of level 5% appear in Kiefer and Vogelsang (2003) for $\hat{\Omega}_{QS}(c)$ and $\hat{\Omega}_{BT}(c)$.

There are stark differences in the power of the test statistic constructed around different LRV estimators: The most robust LRV estimator $\hat{\Omega}_{MR}$ has by far the lowest local power, followed by $\hat{\Omega}_{QS}(1)$ and $\hat{\Omega}_{UA}(2)$. The loss of efficiency compared to the benchmark of a known LRV of these tests is quite dramatic. In order to achieve power of 50%, for instance, a test using $\hat{\Omega}_{MR}$ requires a local alternative of around $\kappa_0 = 8.6$, whereas for Ω known, local power of 50% is already achieved when $\kappa_0 = 1.96$. A useful way of characterizing this efficiency loss is the Pitman efficiency: for a given power of the efficient benchmark test, it describes how much more data would be needed to achieve the same power with an inefficient test. For the model here, this is simply the square of the ratio of the local alternatives against which the inefficient test and the efficient test achieve the same local power. Table 1 lists the Pitman efficiencies of the considered tests along with the robustness coefficient as defined in Definition 2. For the LRV estimators with uniformly bounded robustness coefficient, there is

a strong correlation between the robustness measure and the local power of the tests: tests that are constructed around more more robust LRV estimators are less powerful.

We now turn to the small sample size control of tests based on various LRV estimators. For comparison purposes, we include in our experiments two popular LRV estimators that are consistent in the usual asymptotic thought experiment: the quadratic spectral estimator $\hat{\Omega}_{QA}$ with an automatic bandwidth selection using an AR(1) model for the bandwidth determination as suggested by Andrews (1991), and an AR(1) prewhitened LRV estimator $\hat{\Omega}_{PW}$ with a second stage automatic bandwidth quadratic spectral kernel estimator as described in Andrews and Monahan (1992). The (asymptotically justified) critical values of the \tilde{t} and \tilde{F} -tests (4) based on $\hat{\Omega}_{QA}$ and $\hat{\Omega}_{PW}$ are accordingly those from standard normal and χ^2 distributions.

The small sample experiment design closely follows similar exercises in Andrews (1991), Andrews and Monahan (1992) and Kiefer, Vogelsang, and Bunzel (2000). We are interested in the empirical rejection probability of a nominal 5% level, two-sided t -test concerning the second element of the coefficient vector in a OLS regression of $\{y_t\}$ on $\{x_t\}$, where $x_t = (1, \tilde{x}_t)'$ and \tilde{x}_t is a 4×1 vector. The number of observations is $T = 128$ throughout. The regressors $\{x_t\}$ are generated by one of three models: the AR(1) model, the MA(1) model and the AR(1)+MA(1) model. For the AR(1) model, let the 4×1 vector sequence $\{\tilde{x}_t\}_{t=1}^T$ be four independent draws from a mean-zero, Gaussian, stationary AR(1) process of unit variance and common coefficient ρ . Let $\{\bar{x}_t\} = \{\tilde{x}_t - T^{-1} \sum_{s=1}^T \tilde{x}_s\}$. Then $\{\tilde{x}_t\} = \{(T^{-1} \sum_{s=1}^T \bar{x}_s \bar{x}_s')^{-1/2} \bar{x}_t\}$. The transformations ensure that the sample covariance matrix of $\{x_t\}$ is I_5 . In the MA(1) model, the regressors are generated in the same way, except that a Gaussian, stationary MA(1) process of unit variance and common coefficient θ is employed (parametrized such that $\theta > 0$ generates a spectrum with less mass at frequency zero). Finally, for the AR(1)+MA(1), the pretransformed sequence $\{\tilde{x}_t\}$ is given by the sum of the (pretransformed and independent) AR(1) and MA(1) models with common coefficient $\phi = \rho = \theta$ for the AR(1) and MA(1) component.

For the disturbances, we consider three cases: homoskedastic disturbances (HOMO), heteroskedastic disturbances with heteroskedasticity solely with respect to the regressor we are conducting inference about (HET1), and heteroskedastic disturbances with heteroskedasticity with respect to all four nonconstant regressors (HET2). In the HOMO case, the

disturbances $\{\varepsilon_t\}$ are independent draws from the same model as the (pretransformed) regressors. In the heteroskedastic cases, the disturbances are those from the HOMO case ($\{\varepsilon_t\}$) multiplied by the absolute value of a linear combination of the nonconstant regressors, i.e. $\{\varepsilon_t \cdot |\tilde{\omega}'\tilde{x}_t|\}$. In the HET1 case, $\tilde{\omega} = (1, 0, 0, 0)'$, and in the HET2 case $\tilde{\omega} = (1/2, 1/2, 1/2, 1/2)'$.

Tables 2 and 3 contain the empirical rejection probabilities of a subset of the possible combinations of regressor models and type of disturbances, based on 10,000 replications. Simulations for the other combinations were performed but are not reported as the results are qualitatively very similar. As predicted by the analysis of section 3, inference based on $\hat{\Omega}_{MR}$ has by far the smallest size distortions, with an effective rejection probability never exceeding 10%. The—under the usual thought experiment consistent—estimators $\hat{\Omega}_{QA}$ and $\hat{\Omega}_{PW}$ do much worse than $\hat{\Omega}_{MR}$. As the simulations with the AR(1)+MA(1) model show, perceptible gains from prewhitening require the prewhitening regression to be a reasonable approximation for the whole spectrum. For this model, size control using $\hat{\Omega}_{QA}$ and $\hat{\Omega}_{PW}$ is worse than the least robust versions of $\hat{\Omega}_{RE}(\lambda)$, $\hat{\Omega}_{UA}(p)$ and $\hat{\Omega}_{QS}(c)$ considered in this exercise. For the LRV estimators with a bounded R_T , the larger R_T the larger the size distortions become, just as predicted by the theory in section 3. Despite having an order of magnitude larger robustness coefficient, however, inference based on $\hat{\Omega}_{BT}(c)$ is not dramatically worse. As in Andrews (1991), the dire predictions of asymptotic theory for the Bartlett estimator do not seem to fully bear out in samples of size $T = 128$.

In order to gain some overall insight in the correlation between the robustness measure R_T and the behavior of tests based on the various LRV estimators with bounded R_T , Figure 2 depicts scatter plots of the size distortion in the various models against the natural logarithm of R_T for $\rho = \theta = \phi = 0.9$. The size distortion here is the absolute value of the difference between nominal level and effective rejection probability. It is seen that R_T predicts the size control behavior very well: Size distortions are close to being monotonic functions of R_T in all six experiments.

Figure 3 shows that not only size control, but also local asymptotic power at $\kappa_0 = 3$ of tests constructed around the various LRV estimators strongly correlates with R_T . This implies that scatter plots of size distortions against power are also almost monotonic for most experiments. For the range of experiments and tests considered here, more power of a test

Table 2: Rejection probabilities of nominal 5% t-tests based on various LRV estimators for $T = 128$

| ρ | $\hat{\Omega}_{RE}(\lambda)$ | | | | $\hat{\Omega}_{UA}(p)$ | | | | $\hat{\Omega}_{QS}(c)$ | | | | $\hat{\Omega}_{BT}(c)$ | | | | | |
|--------|------------------------------|---------------------|---------------------|------------|------------------------|------|------|------|------------------------|------|------|------|------------------------|------|------|------|------|------|
| | 50 | 200 | 1500 | 2 | 4 | 8 | 15 | 1 | 0.3 | 0.1 | 0.05 | 1 | 0.3 | 0.1 | 0.05 | | | |
| | $\hat{\Omega}_{QA}$ | $\hat{\Omega}_{PW}$ | $\hat{\Omega}_{MR}$ | AR(1)-HOMO | | | | | | | | | | | | | | |
| 0 | 6.6 | 7.0 | 4.8 | 5.3 | 5.8 | 6.2 | 5.3 | 5.9 | 6.0 | 5.8 | 5.4 | 6.1 | 6.4 | 6.1 | 6.2 | 6.3 | 6.5 | 6.1 |
| -0.3 | 7.7 | 7.1 | 4.9 | 5.1 | 5.9 | 6.3 | 5.0 | 5.9 | 6.1 | 6.2 | 5.5 | 6.0 | 6.6 | 6.7 | 5.8 | 6.3 | 6.8 | 7.1 |
| -0.5 | 8.8 | 7.5 | 4.6 | 5.0 | 5.7 | 6.1 | 5.4 | 5.5 | 5.9 | 6.7 | 5.6 | 6.1 | 6.5 | 7.1 | 6.3 | 6.5 | 7.4 | 8.1 |
| 0.3 | 7.3 | 6.7 | 4.9 | 4.9 | 5.4 | 5.4 | 5.0 | 5.5 | 5.5 | 5.7 | 5.5 | 5.8 | 6.2 | 6.2 | 5.6 | 6.1 | 6.5 | 6.7 |
| 0.5 | 8.2 | 7.0 | 5.1 | 5.2 | 5.7 | 6.0 | 5.0 | 5.5 | 6.3 | 6.2 | 5.9 | 6.2 | 6.5 | 6.7 | 6.4 | 6.8 | 7.0 | 7.7 |
| 0.7 | 12.6 | 9.5 | 5.1 | 6.2 | 8.1 | 8.9 | 6.3 | 7.6 | 8.6 | 9.4 | 6.3 | 7.7 | 9.5 | 9.9 | 8.5 | 9.2 | 10.7 | 12.2 |
| 0.9 | 25.7 | 18.5 | 7.7 | 10.1 | 13.3 | 18.8 | 10.2 | 13.4 | 17.8 | 21.2 | 10.4 | 13.3 | 19.3 | 23.4 | 17.0 | 18.6 | 22.8 | 27.7 |
| 0.95 | 39.1 | 28.3 | 9.3 | 15.1 | 22.9 | 30.6 | 15.4 | 23.0 | 28.8 | 33.8 | 15.4 | 22.3 | 31.2 | 37.1 | 26.1 | 29.7 | 35.9 | 41.7 |
| | AR(1)-HET1 | | | | | | | | | | | | | | | | | |
| 0 | 7.2 | 8.3 | 4.6 | 5.1 | 5.2 | 5.7 | 5.1 | 5.2 | 6.0 | 6.2 | 5.9 | 6.0 | 6.2 | 6.4 | 6.2 | 6.4 | 6.2 | 6.9 |
| -0.3 | 8.5 | 8.8 | 5.2 | 5.3 | 5.9 | 6.6 | 5.4 | 6.1 | 6.4 | 6.9 | 5.3 | 6.3 | 7.1 | 7.5 | 6.5 | 6.9 | 7.4 | 8.0 |
| -0.5 | 9.5 | 9.0 | 4.8 | 5.2 | 5.5 | 6.7 | 5.8 | 5.9 | 6.4 | 7.3 | 5.8 | 6.3 | 6.9 | 7.9 | 6.7 | 7.3 | 8.0 | 8.7 |
| 0.3 | 8.7 | 8.6 | 4.9 | 5.2 | 5.8 | 7.0 | 5.2 | 6.3 | 7.2 | 7.2 | 5.4 | 6.5 | 7.4 | 7.7 | 6.3 | 7.1 | 7.7 | 8.0 |
| 0.5 | 9.8 | 9.5 | 4.8 | 5.2 | 5.8 | 6.9 | 5.3 | 6.1 | 7.1 | 7.8 | 5.7 | 6.4 | 7.8 | 8.2 | 6.9 | 7.2 | 8.4 | 9.1 |
| 0.7 | 15.2 | 13.7 | 5.8 | 6.5 | 7.9 | 10.5 | 6.8 | 8.0 | 9.7 | 11.4 | 7.9 | 9.0 | 10.6 | 12.3 | 10.0 | 10.5 | 12.6 | 14.4 |
| 0.9 | 28.1 | 24.9 | 7.5 | 10.0 | 15.0 | 20.7 | 10.9 | 15.7 | 19.6 | 22.7 | 11.2 | 15.4 | 20.8 | 25.1 | 18.6 | 20.5 | 24.6 | 29.5 |
| 0.95 | 38.1 | 32.1 | 8.5 | 14.0 | 21.2 | 29.0 | 14.3 | 21.8 | 27.9 | 32.8 | 14.7 | 21.0 | 30.1 | 35.5 | 25.0 | 28.2 | 34.6 | 40.7 |
| | AR(1)-HET2 | | | | | | | | | | | | | | | | | |
| 0 | 6.2 | 7.4 | 4.7 | 5.1 | 5.3 | 5.4 | 4.7 | 5.3 | 5.9 | 5.7 | 4.8 | 5.5 | 6.1 | 5.8 | 5.4 | 5.9 | 6.1 | 6.2 |
| -0.3 | 7.8 | 8.2 | 5.0 | 5.2 | 5.2 | 6.0 | 4.7 | 5.2 | 5.7 | 6.5 | 5.5 | 5.8 | 6.3 | 6.8 | 6.2 | 6.2 | 6.6 | 7.2 |
| -0.5 | 9.5 | 9.3 | 5.4 | 5.4 | 6.2 | 7.0 | 5.3 | 6.2 | 7.2 | 7.4 | 5.6 | 6.5 | 7.5 | 7.8 | 7.0 | 7.5 | 7.9 | 8.7 |
| 0.3 | 7.6 | 7.8 | 5.2 | 5.6 | 6.1 | 6.1 | 5.2 | 5.6 | 6.3 | 6.6 | 5.6 | 6.2 | 6.6 | 6.6 | 5.6 | 6.2 | 6.8 | 7.1 |
| 0.5 | 9.8 | 9.6 | 5.1 | 5.6 | 5.8 | 6.9 | 5.4 | 6.0 | 7.0 | 7.2 | 6.1 | 6.1 | 7.2 | 7.9 | 7.1 | 7.4 | 7.9 | 8.7 |
| 0.7 | 13.9 | 12.2 | 4.9 | 6.5 | 7.6 | 9.7 | 6.2 | 7.8 | 9.4 | 10.3 | 6.2 | 8.0 | 10.2 | 11.1 | 8.7 | 9.8 | 11.9 | 13.1 |
| 0.9 | 26.6 | 21.7 | 7.4 | 10.3 | 15.3 | 19.4 | 10.5 | 15.4 | 18.9 | 21.3 | 10.4 | 15.2 | 20.3 | 23.5 | 17.5 | 19.6 | 23.5 | 27.1 |
| 0.95 | 36.5 | 28.3 | 8.4 | 13.2 | 19.3 | 27.5 | 13.6 | 20.0 | 25.4 | 30.6 | 14.1 | 20.0 | 28.3 | 33.7 | 23.6 | 26.5 | 32.7 | 38.1 |

Table 3: Rejection probabilities of nominal 5% t-tests based on various LRV estimators for $T = 128$

| θ | $\hat{\Omega}_{RE}(\lambda)$ | | | | | | | | | | $\hat{\Omega}_{QA}$ | | | | | | | | | |
|----------|------------------------------|------|------|-----|------|------|-----|------|------|------|---------------------|------|------|------|------|------|------|------|------|--|
| | 50 | 200 | 1500 | 2 | 4 | 8 | 15 | 1 | 0.05 | 1 | 0.3 | 0.1 | 0.05 | 1 | 0.3 | 0.1 | 0.05 | | | |
| | MA(1)-HOMO | | | | | | | | | | | | | | | | | | | |
| -0.3 | 7.0 | 6.8 | 5.1 | 4.9 | 5.6 | 5.3 | 5.0 | 5.4 | 5.6 | 5.7 | 5.6 | 5.5 | 6.2 | 6.3 | 5.5 | 6.0 | 6.1 | 6.6 | 6.6 | |
| -0.5 | 7.9 | 6.7 | 5.1 | 5.1 | 5.6 | 6.1 | 5.4 | 5.4 | 6.2 | 6.2 | 5.3 | 6.0 | 6.5 | 6.6 | 6.1 | 6.4 | 6.7 | 7.1 | 7.1 | |
| 0.3 | 6.8 | 6.4 | 4.9 | 5.1 | 5.4 | 5.4 | 4.9 | 5.4 | 5.7 | 5.7 | 5.3 | 5.9 | 5.8 | 6.2 | 6.0 | 6.2 | 6.3 | 6.3 | 6.3 | |
| 0.5 | 7.0 | 6.3 | 5.3 | 5.6 | 5.1 | 5.4 | 4.8 | 5.4 | 5.5 | 5.4 | 5.2 | 5.9 | 5.8 | 5.8 | 5.8 | 5.9 | 6.2 | 6.5 | 6.5 | |
| 0.7 | 7.9 | 6.3 | 5.2 | 5.2 | 5.7 | 5.8 | 5.0 | 5.5 | 5.8 | 6.1 | 5.7 | 6.0 | 6.0 | 6.4 | 5.9 | 6.6 | 6.6 | 7.2 | 7.2 | |
| 0.9 | 8.3 | 6.3 | 4.9 | 4.9 | 5.5 | 6.2 | 5.3 | 5.3 | 6.3 | 6.4 | 5.8 | 5.8 | 6.9 | 6.8 | 6.1 | 6.6 | 7.2 | 7.7 | 7.7 | |
| 0.99 | 8.1 | 6.5 | 5.2 | 5.5 | 5.2 | 6.0 | 5.5 | 5.0 | 6.1 | 6.5 | 5.9 | 5.8 | 6.5 | 7.0 | 6.4 | 6.6 | 7.1 | 7.6 | 7.6 | |
| ϕ | AR(1)+MA(1)-HOMO | | | | | | | | | | | | | | | | | | | |
| -0.3 | 6.0 | 6.3 | 4.9 | 4.7 | 4.8 | 5.1 | 5.1 | 5.1 | 5.1 | 5.2 | 5.3 | 5.6 | 5.5 | 5.7 | 5.6 | 5.8 | 5.7 | 5.9 | 5.9 | |
| -0.5 | 5.9 | 6.5 | 4.8 | 4.9 | 4.8 | 5.0 | 4.8 | 4.8 | 5.4 | 5.0 | 4.9 | 5.5 | 5.4 | 5.6 | 5.5 | 5.7 | 5.7 | 5.7 | 5.7 | |
| 0.3 | 6.0 | 6.4 | 4.9 | 5.0 | 5.2 | 5.3 | 4.7 | 5.4 | 5.3 | 5.6 | 4.9 | 5.5 | 5.9 | 6.0 | 5.4 | 5.9 | 6.0 | 6.1 | 6.1 | |
| 0.5 | 6.3 | 6.6 | 5.0 | 4.5 | 5.2 | 5.4 | 4.9 | 5.6 | 5.5 | 5.8 | 4.9 | 6.0 | 6.1 | 6.0 | 5.5 | 6.0 | 6.1 | 6.1 | 6.1 | |
| 0.7 | 8.3 | 8.2 | 4.9 | 5.3 | 5.7 | 6.0 | 5.3 | 5.3 | 6.0 | 5.7 | 5.4 | 5.3 | 5.9 | 6.0 | 5.5 | 5.8 | 6.3 | 6.4 | 6.4 | |
| 0.9 | 19.1 | 19.4 | 5.6 | 6.8 | 8.1 | 10.0 | 6.8 | 8.4 | 9.6 | 11.5 | 7.1 | 8.5 | 10.1 | 13.2 | 10.1 | 10.3 | 12.4 | 15.0 | 15.0 | |
| 0.95 | 26.4 | 26.9 | 6.5 | 8.0 | 10.1 | 15.2 | 7.9 | 10.2 | 13.6 | 17.8 | 7.5 | 10.6 | 15.4 | 19.6 | 12.3 | 14.0 | 18.0 | 21.5 | 21.5 | |
| ϕ | AR(1)+MA(1)-HET1 | | | | | | | | | | | | | | | | | | | |
| -0.3 | 6.5 | 7.4 | 5.3 | 4.8 | 5.0 | 5.6 | 4.5 | 5.2 | 5.6 | 5.9 | 4.9 | 5.8 | 6.3 | 6.2 | 5.8 | 6.1 | 6.5 | 6.4 | 6.4 | |
| -0.5 | 8.3 | 9.3 | 5.7 | 5.4 | 6.1 | 6.6 | 5.7 | 6.4 | 6.9 | 6.6 | 6.6 | 6.9 | 7.5 | 7.2 | 7.2 | 7.5 | 7.4 | 7.7 | 7.7 | |
| 0.3 | 7.3 | 7.8 | 4.6 | 5.1 | 5.2 | 6.6 | 5.1 | 5.0 | 6.3 | 6.2 | 5.2 | 5.7 | 6.8 | 6.7 | 5.6 | 5.9 | 6.8 | 6.9 | 6.9 | |
| 0.5 | 7.9 | 8.6 | 4.7 | 5.7 | 6.4 | 6.8 | 5.6 | 6.1 | 6.7 | 6.8 | 6.0 | 6.6 | 7.2 | 7.3 | 7.0 | 7.1 | 7.5 | 7.4 | 7.4 | |
| 0.7 | 9.9 | 10.7 | 5.7 | 5.2 | 5.8 | 6.9 | 5.7 | 5.7 | 7.4 | 7.6 | 6.0 | 6.1 | 7.5 | 7.8 | 6.9 | 7.0 | 7.9 | 8.4 | 8.4 | |
| 0.9 | 19.7 | 20.7 | 5.9 | 6.8 | 8.5 | 11.2 | 7.0 | 8.4 | 10.5 | 12.8 | 7.0 | 8.8 | 11.7 | 13.9 | 9.9 | 10.9 | 13.4 | 15.3 | 15.3 | |
| 0.95 | 25.3 | 26.8 | 5.7 | 7.3 | 8.8 | 14.1 | 7.2 | 9.2 | 12.6 | 16.6 | 7.2 | 9.4 | 14.5 | 18.5 | 11.7 | 12.9 | 17.1 | 20.6 | 20.6 | |

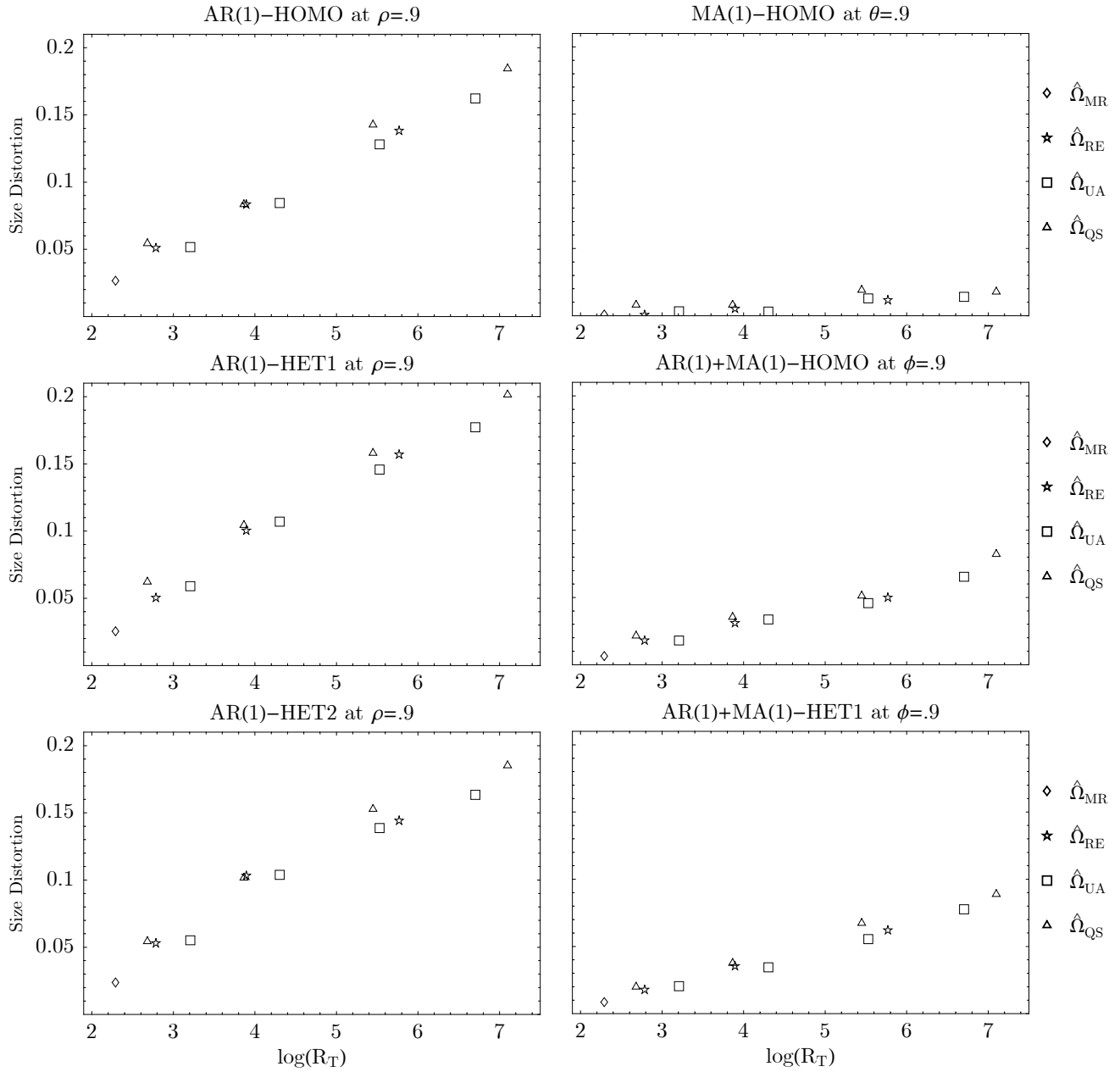


Figure 2: Small Sample Size Distortions and Robustness Coefficients of Various LRV Estimators

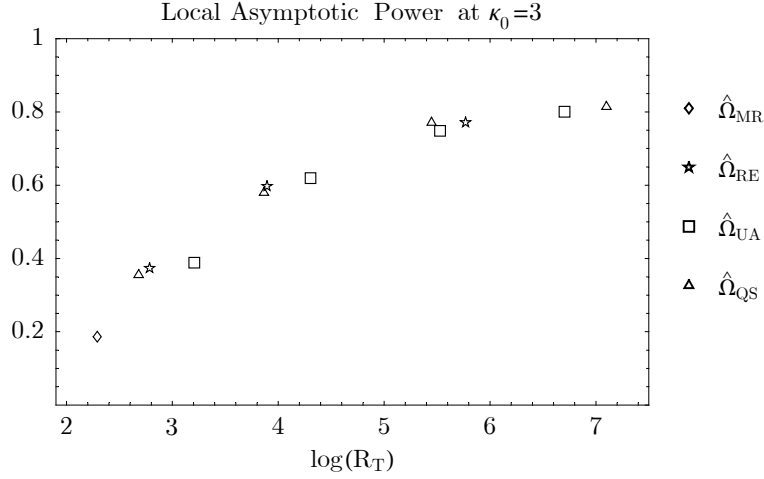


Figure 3: Local Asymptotic Power and Robustness Coefficients of Various LRV Estimators

comes necessarily at the cost of worse size control.⁴ One might argue, then, that the power of tests constructed around a specific LRV estimator is the truly reliable predictor for its size control, and that R_T only correlates with size distortions because it predicts power well. But this is not the case. LRV estimators that extract information only of higher frequencies and discard low-frequency information like $\hat{\xi}_4 \hat{\xi}'_4$ or $\frac{1}{4} \sum_{l=4}^7 \hat{\xi}_l \hat{\xi}'_l$ in the notation developed at the end of section 3 have the same asymptotic distribution as $\hat{\Omega}_{MR}$ and $\hat{\Omega}_{UA}(4)$, respectively, as long as (3) holds. The tests (4) based on the LRV estimators $\hat{\xi}_4 \hat{\xi}'_4$ or $\frac{1}{4} \sum_{l=4}^7 \hat{\xi}_l \hat{\xi}'_l$ hence have the same asymptotic power as tests based on $\hat{\Omega}_{MR}$ and $\hat{\Omega}_{UA}(4)$. At the same time, simulations not reported here show that their small sample size control is much worse, and very much in line with predictions based on their robustness coefficient.

A more natural interpretation of the similarity of the power of tests based on $\hat{\Omega}_{RE}(\lambda)$, $\hat{\Omega}_{UA}(p)$ and $\hat{\Omega}_{QS}(c)$ conditional on the robustness coefficients is therefore that all these LRV estimators are making close to optimal efficiency against robustness trade-offs. One way to describe this analytically is to compare the robustness coefficients of $\hat{\Omega}_{UA}(p)$ and $\hat{\Omega}_{QS}(c)$ with the robustness coefficient of the optimal LRV estimator $\hat{\Omega}_{RE}(\lambda)$ of Theorem 6 where λ is chosen such that the variances of the LRV estimators coincide. In such a comparison

⁴In such a scatterplot of power against size distortions, the Bartlett LRV estimator $\hat{\Omega}_{BT}$ does worse than either of the estimators $\hat{\Omega}_{QS}$, $\hat{\Omega}_{UA}$ or $\hat{\Omega}_{RE}$. At least in this sense, the theoretic prediction of the relative inferiority of inference based on $\hat{\Omega}_{BT}$ as implied by an unbounded robustness coefficient R_T of $\hat{\Omega}_{BT}$ does hold.

one finds R_T to be 5%-20% larger than the smallest possible R_T as generated by $\hat{\Omega}_{RE}(\lambda)$ for the values of p and c considered in the simulation exercise. Given an average change in size distortions induced by a one percent relative increase in R_T of less than 0.05 percentage points in Figure 2, this suggests a very moderate maximum loss of 1 percent higher size distortions of inference based on $\hat{\Omega}_{UA}(p)$ and $\hat{\Omega}_{QS}(c)$ instead of $\hat{\Omega}_{RE}(\lambda)$.

Overall, the results of the small sample experiment conducted here are reliably predicted by the asymptotic analysis of section 2 and 3: Despite being optimal in a certain sense (cf. Andrews (1991)), the consistent automatic bandwidth selection LRV estimators $\hat{\Omega}_{QA}$ and $\hat{\Omega}_{PW}$ perform considerably worse with data exhibiting strong time series dependence than most of the inconsistent LRV estimators, underlining the relevance of Theorem 1. Furthermore, the theoretically motivated robustness measure R_T is highly successful in predicting the small sample size properties of the inconsistent LRV estimators $\hat{\Omega}_{RE}(\lambda)$, $\hat{\Omega}_{UA}(p)$ and $\hat{\Omega}_{QS}(c)$.

5 Implications for Applied Work

The tests (4) based on $\hat{\Omega}_{RE}(\lambda)$, $\hat{\Omega}_{UA}(p)$ and $\hat{\Omega}_{QS}(c)$ have very comparable size and power properties when λ , p and c are such that their robustness coefficient is the same. This suggests that $\hat{\Omega}_{UA}(p)$ is the most convenient choice in practice, since tests based on this LRV estimator have well-known asymptotic distributions under the null hypothesis. In detail, for inference in the OLS regression $y_t = x_t' \beta + \varepsilon_t$ with β being $m \times 1$ one proceeds as follows:

- Compute the ordinary least squares estimator $\hat{\beta}$ of a regression of $\{y_t\}$ on $\{x_t\}$, and construct $\hat{\Sigma}_X = T^{-1} \sum_{t=1}^T x_t x_t'$ and $\{\hat{u}_t\} = \{\hat{\Sigma}_X^{-1} x_t \hat{\varepsilon}_t\}$, where $\{\hat{\varepsilon}_t\}$ are the residuals from the regression.
- Construct the p series $\{\tilde{v}(l)_t\}_{t=1}^T$, $l = 1, \dots, p$, where $\tilde{v}(l)_t = \sqrt{2/T} \cos(l\pi(t-1/2)/T)$.
- Compute the $p \times m$ ordinary least squares estimator $\hat{\xi}$ of a regression of $\{\hat{u}_t\}$ on the p series $\{\tilde{v}(l)_t\}$, $l = 1, \dots, p$.
- Compute $\hat{\Omega}_{UA}(p) = \hat{\xi}' \hat{\xi} / p$.
- For a test concerning the j^{th} element of β , construct the t-statistic $\tilde{t} = T^{1/2}(\hat{\beta}_j - \beta_{0,j}) / \sqrt{\hat{\Omega}_{UA}(p)_{jj}}$, where $\hat{\Omega}_{UA}(p)_{jj}$ is the j^{th} element on the diagonal of $\hat{\Omega}_{UA}(p)$.

- For a test of the restriction $R'\beta = r_0$ for some $m \times d$ matrix R and $d \times 1$ vector r_0 , $1 \leq d \leq p$, compute $\tilde{G} = (p + 1 - d)T(R'\hat{\beta} - r_0)'(R'\hat{\Omega}_{UA}(p)R)^{-1}(R'\hat{\beta} - r_0)/(dp)$.
- Under the null hypothesis, \tilde{t} and \tilde{G} are asymptotically distributed student-t with p degrees of freedom and F with d degrees of freedom in the numerator and $(p + 1 - d)$ degrees of freedom in the denominator, respectively.

For a two stage least squares regression with instruments $\{z_t\}$, $\{x_t\}$ needs to be replaced with the least-squares projection of $\{x_t\}$ on $\{z_t\}$ throughout. The steps above are also applicable to an exactly identified GMM problem, where in the first step, $\hat{\beta}$ is the GMM estimator, $\{x_t\}$ is replaced by the derivative of the moment condition evaluated at the GMM estimator and $\{\hat{\varepsilon}_t\}$ is the moment condition evaluated at the GMM estimator, see (2) above.

The asymptotically justified inference resulting from this procedure is more efficient but less robust for larger p . This raises the important question of how to choose p . While a detailed answer is left to future research, Theorems 1 and 3 above put a concrete limit on any data-dependent procedure: When $\{u_t\}$ is Gaussian White Noise, i.e. as well behaved as it arguably can, a data-dependent method that results in an unbounded p (as a function of the sample size) will necessarily result in a very nonrobust LRV estimator. Intuitively, there always exists deviations from Gaussian White Noise that are large enough to grossly invalidate inference but that are too small to be detectable.

The most conservative way of conducting inference is then to rely solely on the most robust but least efficient estimator with $p = 1$ (for the t-tests and a one-dimensional F-test). And indeed, the numerical results of the last section suggest that extreme deviations from the usual regularity conditions are necessary to grossly distort the nominal level of a test based on this LRV estimator. So if a hypothesis can be rejected on the basis of $\hat{\Omega}_{MR} = \hat{\Omega}_{UA}(1)$, then a researcher can be quite confident that this does not arise from the test not controlling size correctly. But given the low power, many practically important hypotheses will not be rejected, although they are wrong.

From a spectral perspective, p describes the size of the neighborhood of zero of the spectrum of $\{u_t\}$ that can be considered to be constant. For many macroeconomic applications, this view suggests that inference for the LRV should not be made dependent on business cycle frequencies. Taking 10 years as an upper bound for the length of the business cycle, this restricts p to be smaller or equal two times the number of decades of the whole span

of the data. On this ground, for data that spans 40 years, for instance, p may be chosen to equal 8. Such a strategy yields robust inference for all models of the macroeconomic data that imply $\{u_t\}$ to have a flat spectrum below business cycle frequencies. The numerical analysis above suggests for $p = 8$ a loss of efficiency compared to knowing the population LRV of about 20%-30%. At the same time, such a value of p results in reasonable size control over a wide array of dependence structures, including relatively strongly persistent data. For potentially even less well-behaved data, one might want to report test results for smaller values of p , too, allowing readers to assess the strength of the evidence as a function of the amount of regularity imposed on the data.

6 Conclusion

In order for consistent estimators to work, any given data has to satisfy relatively strong regularity conditions. For the problem of long-run variance (LRV) 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 contrast to Jansson (2004), who considers second order asymptotics under the usual asymptotic thought experiment, this paper explores the robustness of LRV estimators by analyzing first order asymptotic properties of LRV estimators in alternative asymptotic thought experiments. It is found that one only needs to consider the set of all processes that satisfy a Functional Central Limit Theorem to obtain arbitrary first order properties of any consistent LRV estimator. When faced with strongly dependent data, this analytical result strongly motivates basing inference on inconsistent LRV estimators instead, for which inference is asymptotically justified whenever a Functional Central Limit Theorem applies.

As a guide for the choice between alternative inconsistent LRV estimators with this property, the paper introduces an analytic measure of robustness of LRV estimators. Specifically, robustness is quantified by considering how much bias can be induced in a LRV estimator when faced with a partial sum process of the residuals that is equal to a contaminated Brownian Bridge. Numerical results show that this measure is highly successful in predicting the size control properties of LRV estimators in a number of small sample experiments. A specific variant of LRV estimators—whose form is very close to LRV estimators that in a certain sense optimally trade off robustness against efficiency—has attractive power and size

properties while leading to test statistics that under the null hypothesis are asymptotically student-t and F distributed, making them a convenient choice for applied work.

The amount of regularity a researcher imposes on the data while employing these LRV estimators is directly related to the number of degrees of freedom of the t- and F-distribution the test statistics converge to under the null hypothesis. For versions of the LRV estimator that result in a small number of degrees of freedom, the robustness of the resulting test is high, but its efficiency is low. The results of this paper imply that it is impossible to develop data dependent methods that are efficient for well behaved data but remain robust for less well behaved data. On a fundamental level, it is impossible to let the data decide how much regularity can be safely assumed. One possibility to impose a certain degree of regularity in a theoretically appealing yet simple way is to employ the LRV estimators suggested here.

7 Appendix

The proof of Theorem 1 relies on a *contiguity* argument. Loosely speaking, the concept of contiguity describes the closeness of two sequences of densities in the limit as $T \rightarrow \infty$. In particular, if a sequence of densities is contiguous to another sequence of densities, then all convergences in probability that apply to the latter sequence also hold for the former sequence. See van der Vaart (1998) or Pollard (2001) for an introduction to the concept.

Lemma 1 *Let $B(\cdot) = \sum_{l=1}^{\infty} \phi_l(\cdot)\xi_l/(\pi l)$, where $\phi_l(s) = \sqrt{2}\sin(l\pi s)$ and ξ_l is distributed independent $\mathcal{N}(0, I_m)$, and let $\check{B}_n(\cdot) = H^{-1} \sum_{l=1}^n \phi_l(\cdot)\xi_l/(\pi l) + \sum_{l=n+1}^{\infty} \phi_l(\cdot)\xi_l/(\pi l)$ for some nonsingular H and fixed n . Then the sequence of densities of $\{\check{B}_n(t/T)\}_{t=1}^T$ are contiguous to the densities of $\{B(t/T)\}_{t=1}^T$.*

Proof. With $r_l = [4T^2 \sin(l\pi/(2T))^2]^{-1} = (l\pi)^{-2} + o(1)$ the l th largest eigenvalue of $T^{-2}FM_eF'$ as defined in Theorem 4, let $\{\check{\xi}_l\}_{l=1}^{T-1} = \{r_l^{-1/2}T^{-1} \sum_{t=1}^T \phi_l(t/T)B(t/T)\}$ and $\{\check{\xi}_l\}_{l=1}^{T-1} = \{r_l^{-1/2}T^{-1} \sum_{t=1}^T \phi_l(t/T)\check{B}_n(t/T)\}$. Clearly, $\{\check{\xi}_l\}_{l=1}^{T-1}$ and $\{\check{\xi}_l\}_{l=1}^{T-1}$ are one-to-one functions of $\{B(t/T)\}_{t=1}^T$ and $\{\check{B}_n(t/T)\}_{t=1}^T$, so it suffices to show contiguity of the sequence of densities of $\{\check{\xi}_l\}_{l=1}^{T-1}$ to the densities of $\{\check{\xi}_l\}_{l=1}^{T-1}$. Since $\{T^{-1/2}\phi_l(t/T)\}_{t=1}^T$ are the elements of the eigenvector of $T^{-2}FM_eF'$ corresponding to the eigenvalue r_l , $\{\check{\xi}_l\}_{l=1}^{T-1}$ is distributed independent $\mathcal{N}(0, I_m)$. Furthermore, for $n < T$, from the orthonormality of the eigenvectors of $T^{-2}FM_eF'$, we find $\{\check{\xi}_l\}_{l=1}^n$ is distributed independent $\mathcal{N}(0, r_l^{-1}(l\pi)^{-2}H^{-1}H^{-1'} + r_l^{-1}(r_l - (l\pi)^{-2})I_m)$, and $\{\check{\xi}_l\}_{l=n+1}^{T-1}$ is distributed independent $\mathcal{N}(0, I_m)$. The only difference between the multivariate normal distributions of $\{\check{\xi}_l\}_{l=1}^{T-1}$ and $\{\check{\xi}_l\}_{l=1}^{T-1}$ is hence the fact that the first n elements of $\{\check{\xi}_l\}_{l=1}^{T-1}$ have covariance matrix $r_l^{-1}(l\pi)^{-2}H^{-1}H^{-1'} + r_l^{-1}(r_l - (l\pi)^{-2})I_m$ rather than I_m . For any fixed l , $r_l^{-1}(l\pi)^{-2}H^{-1}H^{-1'} + r_l^{-1}(r_l - (l\pi)^{-2})I_m \rightarrow H^{-1}H^{-1'}$. The densities of two finite sets of independent Gaussian vectors with positive definite limiting covariance matrices are evidently contiguous, and the result follows. ■

Proof of Theorem 1:

By scale equivariance, $\hat{\Omega}_T(HB(\cdot)) \xrightarrow{p} HH'$. Since the sequence of densities of $\{\check{B}_n(t/T)\}_{t=1}^T$ is contiguous to the densities of $\{B(t/T)\}_{t=1}^T$ for any fixed n by Lemma 1, it follows that $\hat{\Omega}_T(H\check{B}_n(\cdot)) \xrightarrow{p} H'H$, too. Therefore, for any n , there exists a finite number T_n such that for all $T > T_n$

$$P(\|\hat{\Omega}_T(H\check{B}_n(\cdot)) - HH'\|_{\infty} > n^{-1}) < n^{-1}.$$

For any T , let n_T^* be the largest n^* such that $(\max_{n \leq n^*} T_n) < T$. By construction, n_T^* is nondecreasing in T . Furthermore, $n_T^* \rightarrow \infty$ as $T \rightarrow \infty$. Suppose otherwise. Then there exists n_{\max} such that for all T , $T_{n_{\max}+1} > T$. But this is impossible, given that $T_{n_{\max}+1}$ is finite.

Define

$$\bar{B}_T(\cdot) = H\check{B}_{n_T^*}(\cdot) = \sum_{l=1}^{n_T^*} \phi_l(\cdot)\xi_l/(\pi l) + H \sum_{l=n_T^*+1}^{\infty} \phi_l(\cdot)\xi_l/(\pi l).$$

Now for all T , $P(\|\hat{\Omega}_T(\bar{B}_T(\cdot)) - HH'\|_{\infty} > (n_T^*)^{-1}) < (n_T^*)^{-1}$, so that from $n_T^* \rightarrow \infty$ we find $\hat{\Omega}_T(\bar{B}_T(\cdot)) \xrightarrow{p} HH'$.

At the same time, the almost sure uniform convergence (cf. Theorem 5.2 of Itô and Nisio (1968)) of $\lim_{N \rightarrow \infty} \sum_{l=1}^N \phi_l(\cdot)\xi_l/(\pi l)$ to $B(\cdot)$ implies that for any $\eta > 0$, there exists N^* such that for all $N > N^*$, $\sup_s \|\sum_{l=N}^{\infty} \phi_l(s)\xi_l/(\pi l)\|_{\infty} < \eta$ with probability one. For T large enough to make $n_T^* > N^*$, with probability one,

$$\sup_s \|\bar{B}_T(s) - B(s)\|_{\infty} = \sup_s \left\| (H - I_m) \sum_{l=n_T^*}^{\infty} \phi_l(s)\xi_l/(\pi l) \right\|_{\infty} < \|H - I_m\|_{\infty} \eta$$

But η was arbitrary, and the claim of the Theorem follows.

Proof of Theorem 3:

We will show that $\bar{B}_T(\cdot)$ of Theorem 1 for $H = \kappa I_m$ with $\kappa > 1$ is within a neighborhood of $B(\cdot)$ of size δ_T , where $\delta_T = o(1)$. Since $\hat{\Omega}_T(\bar{B}_T(\cdot)) \rightarrow \kappa^2 I_m$ together with $\hat{\Omega}_T$ being nonnegative definite implies $\lim_{T \rightarrow \infty} E[\hat{\Omega}_T(\bar{B}_T(\cdot))] \geq \kappa^2 I_m$, it follows that for any $\delta > 0$, for T chosen sufficiently large to ensure $\delta_T < \delta$

$$R_T(\delta) \geq \frac{\text{tr} \left| E[\hat{\Omega}_T(\bar{B}_T(\cdot))] - I_m \right|}{\delta m} \geq \frac{\kappa^2 - 1}{\delta} + o(1)$$

which can be made arbitrarily large by choosing κ large.

In order to prove that $\bar{B}_T(\cdot)$ is within a $\delta_T = o(1)$ neighborhood of $B(\cdot)$, note that $\bar{B}_T(\cdot)$ might alternative be written as $\bar{B}_T(\cdot) = B(\cdot) + S_T(\cdot)$, where $S_T(\cdot) = (\kappa^2 - 1)^{1/2} \sum_{l=n_T^*+1}^{\infty} \phi_l(\cdot)\psi_l/(\pi l)$, where $\psi_l \sim iid\mathcal{N}(0, I_m)$ independent of $\{\xi_l\}$. Let w_i be the $T \times 1$ vector with elements $w_{it} = S_{T_i}(t/T)$ for any $i \in \{1, \dots, m\}$, $1 \leq t \leq T$, and let a_T be any vector $T \times 1$ vector with elements $a_{T,t}$ satisfying $a_T' a_T = 1$. It is well known that $a_T' E[w_i w_i'] a_T = E[(a_T' w_i)^2]$ is

smaller or equal to the largest eigenvalue of $E[w_i w_i']$. But

$$\begin{aligned} E[(a_T' w_i)^2] &= (\kappa^2 - 1) \sum_{l=n_T^*+1}^{\infty} \left[\sum_{t=1}^T a_{T,t} \frac{\phi_l(t/T)}{l\pi} \right]^2 \\ &\leq (\kappa^2 - 1) \frac{2}{\pi^2} \left[\sum_{l=n_T^*+1}^{\infty} l^{-2} \right] \left[\sum_{t=1}^{T-1} |a_{T,t}| \right]^2 \end{aligned}$$

From Jensen's inequality, we obtain $\left[\sum_{t=1}^{T-1} |a_{T,t}| \right]^2 \leq T$, so that $\sum_{l=n_T^*}^{\infty} l^{-2} \rightarrow 0$ as $T \rightarrow \infty$ implies that the largest eigenvalue of $E[w_i w_i']$ is $o(T)$.

Proof of Theorem 4:

(a) Let w_i, q_i, \tilde{w}_i and \tilde{q}_i be the $T \times 1$ vectors with elements $w_{it} = S_{Ti}(t/T)$, $q_{it} = T^{1/2}[B_i(t/T) - B_i((t-1)/T)]$, $\tilde{w}_{it} = \tilde{S}_{Ti}(t/T)$ and $\tilde{q}_{it} = T^{1/2}[\tilde{B}_{Ti}(t/T) - \tilde{B}_{Ti}((t-1)/T)]$, for $t = 1, \dots, T$ and $i = 1, \dots, m$, respectively, and denote $\text{diag}(\Sigma_i, 0) = E[w_i w_i']$ and $\text{diag}(\tilde{\Sigma}_i, 0) = E[\tilde{w}_i \tilde{w}_i']$. Define $w = (w_1, \dots, w_m)$ and q, \tilde{w} and \tilde{q} analogously. Let A be such that $Ae = 0$ and $\text{tr}[A] = 1$ without loss of generality.

From the independence assumption

$$\begin{aligned} E[(\tilde{q} + T^{1/2} D\tilde{w})' A (\tilde{q} + T^{1/2} D\tilde{w})] &= E[(q + T^{1/2} Dw)' A (q + T^{1/2} Dw)] \\ E[\tilde{q}' A \tilde{q}] + TE[\tilde{w}' D' A D \tilde{w}] &= E[q' A q] + TE[w' D' A D w] \\ E[\tilde{q}' A \tilde{q}] - I_m &= TE[\tilde{w}' D' A D \tilde{w}] - TE[w' D' A D w] \end{aligned}$$

and with $q_i \sim \mathcal{N}(0, M_e)$, we find

$$\text{tr} |E[\tilde{q}' A \tilde{q}] - I_m| = \text{tr} |TE[\tilde{w}' D' A D \tilde{w}] - TE[w' D' A D w]|.$$

Since $TE[\tilde{w}' D' A D \tilde{w}]$ and $TE[w' D' A D w]$ are nonnegative definite, R_T is maximized when either all Σ_i or all $\tilde{\Sigma}_i$ are set to zero. Without loss of generality, let $\tilde{\Sigma}_i = 0 \forall i$. It is well known that for two conformable matrices B and C , $\text{tr}[BC] \leq \bar{\lambda}_B \text{tr} C$, where $\bar{\lambda}_B$ is the largest eigenvalue of B . Hence

$$\text{tr} |E[\tilde{q}' A \tilde{q}] - I_m| \leq \sum_{i=1}^m T \text{tr} [\text{diag}(\Sigma_i, 0) D' A D] \leq m \delta T^2 \text{tr} [\text{diag}(I_{T-1}, 0) D' A D]$$

Furthermore, for $\Sigma_i = \delta T I_{T-1}$ and $\tilde{\Sigma}_i = 0 \forall i$, $\text{tr} |E[\tilde{q}' A \tilde{q}] - I_m| = m \delta T^2 \text{tr} [\text{diag}(I_{T-1}, 0) D' A D]$, so that the result follows.

(b) A direct calculation reveals that the first $T - 1$ diagonal elements of $TD'AD$ are given by $2k(0) - 2k(b_T^{-1})$. For a twice differentiable kernel, a second order Taylor expansion yields $2k(0) - 2k(b_T^{-1}) = k''(0)b_T^{-2} + o(b_T^{-2})$, whereas for the Bartlett kernel $2k(0) - 2k(b_T^{-1}) = 2b_T^{-1} + o(b_T^{-1})$. Now $\text{tr}[M_e A] = \text{tr}[A] - T^{-1}e' Ae$, and $T^{-1}e' Ae \leq 2T^{-1} \sum_{t=0}^T |k(t/b_T)|$. But $b_T^{-1} \sum_{t=0}^{\infty} |k(t/b_T)| \rightarrow \int_0^{\infty} |k(s)| ds < \infty$, so that $T^{-1} \sum_{t=0}^T |k(t/b_T)| \leq T^{-1} \sum_{t=0}^{\infty} |k(t/b_T)| \rightarrow 0$ and the result follows.

(c) From $\text{tr}[MA] = \text{tr}[A] - T^{-1}e' Ae$ and

$$T^{-1}e' Ae = T^{-2} \sum_{j=-T+1}^{T-1} (T-j)k(j/T) \rightarrow 2 \int_0^1 (1-s)k(s)ds$$

the result follows from the same reasoning as in part (b).

Proof of Theorem 5:

Without loss of generality, let A be such that $Ae = 0$ and $\text{tr}[A] = 1$. In the notation of the proof of Theorem 4, by a direct calculation for $i, j \in \{1, \dots, m\}$,

$$(\tilde{q}_i + T^{1/2}D\tilde{w}_i)' A (\tilde{q}_j + T^{1/2}D\tilde{w}_j) = (q_i + T^{1/2}Dw_i)' A (q_j + T^{1/2}Dw_j)$$

so that

$$\begin{aligned} \tilde{q}_i' A \tilde{q}_j + T^{1/2} \tilde{w}_i' D' A \tilde{q}_j + T^{1/2} \tilde{w}_j' D' A \tilde{q}_i + T \tilde{w}_i' D' A D \tilde{w}_j = \\ q_i' A q_j + T^{1/2} w_i' D' A q_j + T^{1/2} w_j' D' A q_i + T w_i' D' A D w_j. \end{aligned}$$

Now

$$2w_i' D' A D w_j = (w_i + w_j)' D' A D (w_i + w_j) - w_i' D' A D w_i - w_j' D' A D w_j,$$

and $E[(w_i - w_j)(w_i - w_j)'] = \Sigma_i + \Sigma_j - E[w_i w_j' + w_j w_i']$ is nonnegative definite, so that the largest eigenvalue of $E[w_i w_j' + w_j w_i']$ is smaller or equal to $2\delta T$. Hence

$$2TE[|w_i' D' A D w_j|] \leq 6\delta T^2 \text{tr}[\text{diag}(I_{T-1}, 0) D' A D] = 6\delta R_T$$

and similarly, $TE[|\tilde{w}_i' D' A D \tilde{w}_j|] \leq 3\delta R_T$.

Furthermore,

$$\begin{aligned} TE[(w_i' D' A q_j)^2] &= TE[q_j' A D w_i w_i' D' A q_j] \\ &= T \text{tr}[M_e A D \text{diag}(\Sigma_i, 0) D' A] \\ &\leq \delta T^2 \text{tr}[\text{diag}(I_{T-1}, 0) D' A^2 D]. \end{aligned}$$

But with $\text{tr}[A] = 1$, the eigenvalues of A necessarily lie between zero and one, so that $TE[(w'_i D' A q_j)^2] \leq \delta R_T$. Also, noting that $E[\tilde{q}_j \tilde{q}'_j] = M_e + TD \text{diag}(\Sigma_j - \tilde{\Sigma}_j, 0) D'$ we find

$$\begin{aligned}
TE[(\tilde{w}'_i D' A \tilde{q}_j)^2] &= TE[\tilde{q}'_j A D \tilde{w}_i \tilde{w}'_i D' A \tilde{q}_j] \\
&= T \text{tr}[(M_e + TD \text{diag}(\Sigma_j - \tilde{\Sigma}_j, 0) D') A D \text{diag}(\tilde{\Sigma}_i, 0) D' A] \\
&\leq \delta T^2 \text{tr}[\text{diag}(I_{T-1}, 0) D' A^2 D] + \delta^2 T^4 \text{tr}[\text{diag}(I_{T-1}, 0) D' A D \text{diag}(I_{T-1}, 0) D' A D] \\
&\leq \delta R_T + (\delta T^2 \text{tr}[\text{diag}(I_{T-1}, 0) D' A D])^2 \\
&\leq \delta R_T + (\delta R_T)^2.
\end{aligned}$$

Proceeding analogously for $TE[(w'_j D' A q_i)^2]$ and $TE[(\tilde{w}'_j D' A \tilde{q}_i)^2]$, we find using Jensen's inequality

$$E[|\hat{\Omega}_{Tij}(\tilde{B}_T(\cdot)) - \hat{\Omega}_{Tij}(B(\cdot))|] \leq 6\delta R_T + 2\sqrt{\delta R_T} + 2\sqrt{\delta R_T + (\delta R_T)^2}$$

and the result follows.

Proof of Theorem 6:

(a) Define \tilde{A} implicitly by $A = T^{-1} F' \text{diag}(P \Lambda^{-1/2} \tilde{A} \Lambda^{-1/2} P', 0) F$, where $P \Lambda P'$ is the spectral decomposition of the upper-left $T - 1 \times T - 1$ block of $T^{-1} F M_e F'$ with the eigenvalues Tr_l decreasing along the diagonal of Λ . Apart from the inconsequential restriction $Ae = 0$, this leaves A as a function of \tilde{A} unrestricted. Since A is nonnegative definite, so is \tilde{A} . Note that

$$\begin{aligned}
\text{tr}[M_e A] &= T^{-1} \text{tr}[F M_e F' \text{diag}(P \Lambda^{-1/2} \tilde{A} \Lambda^{-1/2} P', 0)] \\
&= \text{tr}[\text{diag}(\Lambda^{-1/2} P' P \Lambda P' P \Lambda^{-1/2} \tilde{A}, 0)] = \text{tr}[\tilde{A}]
\end{aligned}$$

and

$$\text{tr}[\text{diag}(I_{T-1}, 0) D' A D] = T^{-1} \text{tr}[P \Lambda^{-1/2} \tilde{A} \Lambda^{-1/2} P'] = T^{-1} \text{tr}[\Lambda^{-1} \tilde{A}],$$

such that in terms of \tilde{A} , $R_T = T \text{tr}[\Lambda^{-1} \tilde{A}] / \text{tr}[\tilde{A}]$.

Write $\tilde{A} = B B'$. The minimization problem may hence be written as

$$\min_B \text{tr}[\Lambda^{-1} B B'] - \lambda_0 \text{tr}[B B']$$

for some $\lambda_0 \geq 0$, yielding the first order condition

$$\Lambda^{-1} B - \lambda_0 B = 0.$$

This necessary condition can only be satisfied when the columns of B are either zero or proportional to a single Cartesian Basis vector, such that \tilde{A} is proportional to $\iota_{T-1,l}\iota'_{T-1,l}$, where ι_l is the l th column of I_{T-1} . The minimizing choice corresponds to the smallest λ_0 , and since the diagonal elements of Λ^{-1} are increasing, this yields the solution $\tilde{A} = \iota_{T-1,1}\iota'_{T-1,1}$.

(b) Building on the results of part (a), note that for $q_i \sim \mathcal{N}(0, M_e)$ for $i = 1, \dots, m$ and q_i and q_j independent for $i \neq j$, $\text{Var}[q'_i A q_i] = 2 \text{tr}[(M_e A)^2] = 2 \text{tr}[\tilde{A}^2]$ and $\text{Var}[q'_i A q_j] = \text{tr}[\tilde{A}^2]$. Hence

$$\text{tr Var}[K \text{vec } \hat{\Omega}_T(B(\cdot))] = \text{tr}[\tilde{A}^2] \text{tr}[K(I_{m^2} + \sum_{i=1}^m \sum_{j=1}^m (\iota_{m,i}\iota'_{m,j}) \otimes (\iota_{m,j}\iota'_{m,i}))K']$$

where $\iota_{m,i}$ is the i th column of I_m and \otimes denotes the Kronecker product.

With $\tilde{A} = BB'$, the optimization problem might hence be written as

$$\min_B \text{tr}[BB'BB'] + \lambda_0 \text{tr}[\Lambda^{-1}BB'] - \lambda_1 \text{tr}[BB'] \quad (8)$$

for λ_0 and λ_1 nonnegative constants, with first order condition

$$2BB'B + \lambda_0 \Lambda^{-1}B - \lambda_1 B = 0.$$

Denoting with \tilde{A}^* the solution to the optimization problem, clearly the symmetric matrix $B^* = (\tilde{A}^*)^{1/2}$ must satisfy the necessary condition. We might hence assume B to be symmetric without loss of generality. Let v_B be any eigenvector of B associated with a nonzero eigenvalue r_B . Then

$$2r_B^3 v_B + r_B(\lambda_0 \Lambda^{-1} - \lambda_1 I_{T-1})v_B = 0$$

which is possible only when v_B is proportional to a Cartesian basis vector, at least when $\lambda_0 > 0$. A solution when $\lambda_0 = 0$ is easily ruled out, and therefore, B^* and \tilde{A}^* are diagonal. Recomputing the first order conditions of the optimization problem (8) in terms of the diagonal elements \tilde{a}_{ll} of \tilde{A} under the conditions $\tilde{a}_{ll} \geq 0 \forall l$, we find $2\tilde{a}_{ll} = (\lambda_1 - \lambda_0 r_l^{-1}) \vee 0$. From the relationship between \tilde{A} and A , the result follows.

References

- ANDREWS, D. (1991): “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation,” *Econometrica*, 59, 817–858.
- ANDREWS, D., AND J. MONAHAN (1992): “An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator,” *Econometrica*, 60, 953–966.
- BERK, K. (1974): “Consistent Autoregressive Spectral Estimates,” *Annals of Statistics*, 2, 489–502.
- 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.
- FAUST, J. (1999): “Conventional Confidence Intervals for Points on Spectrum Have Confidence Level Zero,” *Econometrica*, 67, 629–637.
- GALLANT, A. (1987): *Nonlinear Statistical Models*. John Wiley and Sons, New York.
- HAAN, W. D., 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.
- HANSEN, B. (1992): “Consistent Covariance Matrix Estimation for Dependent Heterogenous Processes,” *Econometrica*, 60, 967–972.
- HANSEN, L. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50, 1029–1054.
- HOCHSTADT, H. (1973): *Integral Equations*. John Wiley and Sons, New York.
- HOTELLING, H. (1931): “The Generalization of Student’s Ratio,” *The Annals of Mathematical Statistics*, 2, 360–378.
- ITÔ, K., AND M. NISIO (1968): “On the Convergence of Sums of Independent Banach Space Valued Random Variables,” *Osaka Journal of Mathematics*, 5, 35–48.

- JANSSON, M. (2004): “The Error in Rejection Probability of Simple Autocorrelation Robust Tests,” *Econometrica*, 72, 937–946.
- KIEFER, N., AND T. VOGELSANG (2002a): “Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel Without Truncation,” *Econometrica*, 70, 2093–2095.
- (2002b): “Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size,” *Econometric Theory*, 18, 1350–1366.
- (2003): “A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests,” *mimeo*, Cornell University.
- KIEFER, N., T. VOGELSANG, AND H. BUNZEL (2000): “Simple Robust Testing of Regression Hypotheses,” *Econometrica*, 68, 695–714.
- LEHMANN, E., AND G. CASELLA (1998): *Theory of Point Estimation*. Springer, New York, 2nd edn.
- MÜLLER, U. (2002): “Size and Power of Tests for Stationarity in Highly Autocorrelated Time Series,” *University of St. Gallen Working Paper 2002-26*.
- NEWBY, W., AND K. WEST (1987): “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703–708.
- (1994): “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economic Studies*, 61, 631–653.
- 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. (2001): “Contiguity,” unpublished draft, <http://www.stat.yale.edu/~pollard/Paris2001/Lectures/Contiguity.pdf>.

- 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.
- SIMS, C. (1971): “Distributed Lag Estimation When the Parameter Space is Explicitly Infinite-Dimensional,” *The Annals of Mathematical Statistics*, 42, 1622–1636.
- (1972): “The Role of Approximate Prior Restrictions in Distributed Lag Models,” *Journal of the American Statistical Society*, 67, 169–175.
- VAART, A. V. D. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.
- WHITE, H. (1984): *Asymptotic Theory for Econometricians*. Academic Press, New York.