UC San Diego
Recent Work

Title
A New Asymptotic Theory for Vector Autoregressive Long-run Variance Estimation and Autocorrelation Robust Testing

Permalink
https://escholarship.org/uc/item/8cx0t4gc

Authors
Sun, Yixiao
Kaplan, David M.

Publication Date
2011-05-01

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/3.0/

eScholarship.org Powered by the California Digital Library
University of California

Fixed-smoothing Asymptotics and Accurate F Approximation Using Vector Autoregressive Covariance Matrix Estimator

Yixiao Sun and David M. Kaplan*

Department of Economics,
University of California, San Diego

Abstract

We develop a new asymptotic theory for autocorrelation robust tests using a vector autoregressive (VAR) covariance matrix estimator. In contrast to the conventional asymptotics where the VAR order goes to infinity but at a slower rate than the sample size, we have the VAR order grow at the same rate, as a fixed fraction of the sample size. Under this fixed-smoothing asymptotic specification, the associated Wald statistic remains asymptotically pivotal. On the basis of this asymptotics, we introduce a new and easy-to-use F^* test that employs a finite sample corrected Wald statistic and uses critical values from an F distribution. We also propose an empirical VAR order selection rule that exploits the connection between VAR variance estimation and kernel variance estimation. Simulations show that the new VAR F^* test with the empirical order selection is much more accurate in size than the conventional chi-square test.

JEL Classification: C13; C14; C32; C51

Keywords: F distribution, Fixed-smoothing Asymptotics, Flat-top Kernel, Heteroscedasticity and Autocorrelation Robust Test, Long Run Variance, Rectangular Kernel, t distribution, Vector Autoregression

---

*Email: yisun@ucsd.edu and dkaplan@ucsd.edu. Sun gratefully acknowledges partial research support from NSF under Grant No. SES-0752443. Correspondence to: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.

1 Introduction

The paper considers statistical inference with time series data when model parameters are estimated by the generalized method of moments (GMM) or the method of estimating equations. To avoid possible misspecification and to be completely general, we often do not parametrize the dependence structure of the moment conditions or the estimating equations. The problem is how to nonparametrically estimate the covariance matrix of the GMM estimator. The recent literature has mainly focused on kernel-based methods, although quite different approaches like the vector autoregressive (VAR) approach (see, for example, Berk 1974, Parzen 1983, den Haan and Levin 1998) have been explored. Under fairly general conditions, den Haan and Levin (1997, 1998) show that the VAR covariance estimator converges at a faster rate than commonly used positive semidefinite kernel covariance estimators. This faster rate of convergence may lead to a chi-square test with good size and power properties. However, Monte Carlo simulations in den Haan and Levin (1998) show that the finite sample performance of the chi-square test based on the VAR covariance estimator is unsatisfactory, especially when there is strong autocorrelation in the data.

The key asymptotic result underlying the chi-square test is the consistency of the VAR covariance estimator. It requires that the VAR order p increase with the sample size T but at a slower rate. While theoretically convenient, the consistency result does not capture the sampling variation of the covariance estimator in finite samples. In addition, the consistency result completely ignores the estimation uncertainty of the model parameters. In this paper, we develop a new asymptotic theory that avoids these drawbacks. The main idea is to view the VAR order p as proportional to the sample size T. That is, $p = bT$ for some fixed constant $b \in (0, 1)$. Under this new statistical thought experiment, the VAR covariance estimator converges in distribution to a random matrix that depends on the VAR order and the estimation error of model parameters. Furthermore, the random matrix is proportional to the true covariance. As a result, the associated test statistic is still asymptotically pivotal under this new asymptotics. More importantly, the new asymptotic distribution captures the sampling variation of the covariance estimator and provides a more accurate approximation than the conventional chi-square approximation.

To develop the new asymptotic theory, we observe that the VAR(p) model estimated by the Yule-Walker method is stationary almost surely and has conditional population autocovariances (conditional on the estimated model parameters) that are identical to the empirical autocovariances up to order p. These crucial observations drive our asymptotic development. Given the stationarity and 'reproducing' properties of the Yule-Walker estimator, we know that the VAR covariance estimator is asymptotically equivalent to the kernel covariance estimator based on the rectangular kernel with bandwidth equal to p. The specification of $p = bT$ is then the same as the so-called fixed-b specification in Kiefer and Vogelsang (2005), Kuan and Lee (2006), Sun, Phillips and Jin (2008), and Sun (2010). The rectangular kernel is not continuous and has not been considered in the literature on fixed-b asymptotics. Existing results does not directly apply. One of the contributions of this paper is to fill in this important gap and develop a new asymptotic theory for the VAR covariance estimator and associated tests.

The new asymptotics obtained under the specification that $p = bT$ for a fixed b may be referred to as the fixed-smoothing asymptotics, as the asymptotically equivalent kernel estimator has a finite and thus fixed effective degree of freedom. On the other hand, when $b \to 0$, the effective degree of freedom increases with the sample size. The conventional asymptotics obtained under the specification that $p \to \infty$ but $b \to 0$ may be referred to as the increasing-smoothing asymptotics. The two specifications can be viewed as different asymptotic devices to obtain approximations to the finite sample distribution. The fixed-smoothing asymptotics does not require that we fix the value of b in finite samples. In fact, in empirical applications, the sample size T is usually given beforehand, and the VAR order needs to be determined using *a priori* information and/or information obtained from the data. While the selected VAR order may be relatively small for large T, it is also true that $b = p/T > 0$, so intuition alone is ambivalent. Additionally, we can show that the two types of asymptotics coincide as $b \to 0$. In other words, the fixed-smoothing approximation is asymptotically valid under the conventional thought experiment.

Following Sun (2010), we show that, after some modification, the nonstandard fixed-smoothing limiting distribution can be approximated by a standard F distribution. The F approximation can also be derived from a high-order Edgeworth expansion under the conventional asymptotics where $p \to \infty$ but at a slower rate than the sample size. On the basis of the F approximation, we propose a new F^* test. The F^* test statistic is equal to the Wald statistic multiplied by a finite sample correction factor, which can be motivated as a Bartlett-type correction. See Bartlett (1937, 1954). In addition to the multiplicative correction, the F^* test employs an F distribution instead of a χ^2 distribution as the reference distribution. The F^* test is as easy to use as the standard Wald test as both the correction factor and the critical values are easy to obtain.

The connection between the VAR spectrum estimator and the kernel spectrum estimator with the rectangular kernel does not seem to be fully explored in the literature. First, the asymptotic equivalence of these two estimators can be used to prove the consistency and asymptotic normality of the VAR estimator as the asymptotic properties of the kernel estimator have been well researched in the literature. Second, the connection sheds some light on the faster rate of convergence of the VAR spectrum estimator and the kernel spectrum estimator based on flat-top kernels. The general class of flat-top kernels, proposed by Politis (2001), includes the rectangular kernel as a special case. Under the conventional asymptotics, Politis (2011, Theorem 2.1) establishes the rate of convergence of flat-top kernel estimators, while den Haan and Levin (1998, Theorem 1) give the rate for the VAR estimator. Careful inspection shows that the rates in Politis (2011) are the same as those in den Haan and Levin (1998), although the routes to them are completely different. In view of the asymptotic equivalence, the identical rates of convergence are not surprising at all. Finally, the present paper gives another example that takes advantage of this connection. Among others, the connection inspires us to specify p to be proportional to the sample size in order to derive the new asymptotics. This rate specification is not obvious from the rate conditions given under the conventional asymptotic theory, say for example, Berk (1974).

Compared with a finite-order kernel estimator, the VAR covariance estimator enjoys the same bias reducing property as any infinite-order flat-top kernel estimator does. Like a flat-top kernel estimator, there is no bias induced by kernel downweighing. This is one of the main reasons for the renewed interest in the rectangular kernel and other flat-top kernels; See for example Politis (2011) and Lin and Sakata (2009). Compared with the rectangular kernel estimator, which may not be positive semidefinite in finite samples, the VAR covariance estimator is guaranteed to be positive semidefinite, a very desirable property in empirical applications. Although our theoretical development exploits the connection between the VAR estimator and the rectangular kernel estimator, we do not recommend the use of the rectangular kernel in finite samples. The rectangular kernel estimator involves hard thresholding in that the autocovariances "cut off" after the truncation lag. In contrast, the VAR estimator extends the autocovariances beyond the truncation lag. This is very much like a general flat-top kernel estimator with a smooth kernel function. The VAR estimator also has the information-theoretic property that it is the maximum entropy estimator, subject to some autocovariance constraints. See Burg (1967) for the original contribution and Cover and Thomas (2006, pp. 415) for a textbook treatment.

The small bias, coupled with the new asymptotic theory that captures the randomness of the VAR covariance estimator, gives the proposed F^* test a great size advantage. This is confirmed in the Monte Carlo experiments. Simulation results indicate that the size of the VAR F^* test with a new empirically determined VAR order is as accurate as, and sometimes more accurate than, the kernel-based F^* tests recently proposed by Sun (2010). The VAR F^* test is as accurate in size as the nonstandard test using simulated critical values. The VAR F^* test is uniformly more accurate in size than the conventional chi-square test. The power of the VAR F^* test is also very competitive relative to the kernel-based F^* test and χ^2 test.

The paper contributes to the recent and growing literature on fixed-smoothing asymptotics for semiparametric testing. The fixed-smoothing asymptotics for kernel covariance estimators has been well researched. It is an open and long standing problem of how to develop the fixed-smoothing asymptotics for the VAR covariance estimator. The paper provides a simple and illuminating solution to this problem. The main innovation lies in the insight that asymptotically unifies the VAR covariance estimator with the class of kernel covariance estimators.

The rest of the paper is organized as follows. Section 2 presents the GMM model and the testing problem. It also provides an overview of the VAR covariance estimator. The next two sections are devoted to the fixed-smoothing asymptotics of the VAR covariance estimator and the associated test statistic. Section 5 details a new method for lag order determination, and Section 6 reports simulation evidence. The last section provides some concluding discussion. Proofs are given in the appendix.

A word on notation: we use $F_{q,K-q+1}$ to denote a random variable that follows the F distribution with degrees of freedom $(q, K-q+1)$. When there is no possibility of confusion, we also use $F_{q,K-q+1}$ to denote the F distribution itself. We use the same convention for other distributions. For notational convenience, we assume that bT is an integer. $W_m(r)$ is the m-dimensional standard Brownian motion.

2 GMM Estimation and Autocorrelation Robust Testing

We are interested in a $d \times 1$ vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^d$. Let v_t denote a vector of observations. Let θ_0 be the true value and assume that θ_0 is an interior point of the compact parameter space Θ. The moment conditions or estimating equations

$$Ef(v_t, \theta) = 0, \ t = 1, 2, \ldots, T$$

hold if and only if $\theta = \theta_0$ where $f(\cdot)$ is an $m \times 1$ vector of continuously differentiable functions with $m \geq d$ and rank $E[\partial f(v_t, \theta_0)/\partial \theta'] = d$. Defining

$$g_t(\theta) = T^{-1} \sum_{j=1}^{t} f(v_j, \theta),$$

the GMM estimator (Hansen, 1982) of θ_0 is then given by

$$\hat{\theta}_T = \arg\min_{\theta \in \Theta} g_T(\theta)' \mathcal{W}_T g_T(\theta),$$

where $\mathcal{W}_T$ is an $m \times m$ positive definite and symmetric weighting matrix. The GMM estimator includes many popular estimators as special cases, for example, the linear and nonlinear least squares estimators, the linear and nonlinear instrumental variable estimators, and general M estimators.

Let

$$G_t(\theta) = \frac{\partial g_t(\theta)}{\partial \theta'} = \frac{1}{T} \sum_{j=1}^{t} \frac{\partial f(v_j, \theta)}{\partial \theta'}.$$

Under some regularity conditions, $\hat{\theta}_T$ satisfies

$$\hat{\theta}_T - \theta_0 = -\left[G_T\left(\tilde{\theta}_T\right)' \mathcal{W}_T G_T\left(\tilde{\theta}_T\right) \right]^{-1} G_T(\theta_0)' \mathcal{W}_T g_T(\theta_0) + o_p\left(\frac{1}{\sqrt{T}}\right),$$

where $\tilde{\theta}_T$ is a value between $\hat{\theta}_T$ and θ_0. If $\text{plim}_{T\to\infty} G_T(\tilde{\theta}_T) = G$, $\text{plim}_{T\to\infty} \mathcal{W}_T = \mathcal{W}$ and $\sqrt{T} g_T(\theta_0) \Rightarrow N(0, \Omega)$, where Ω is the so-called long run variance (LRV) of $f(v_t, \theta)$, then

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \Rightarrow N(0, \mathcal{V}), \tag{1}$$

for $\mathcal{V} = (G'\mathcal{W}G)^{-1} (G'\mathcal{W}\Omega\mathcal{W}G) (G'\mathcal{W}G)^{-1}$. The above asymptotic result provides the basis for inference on θ_0.

Consider the null hypothesis $H_0 : r(\theta_0) = 0$ and the alternative hypothesis $H_1 : r(\theta_0) \neq 0$ where $r(\theta)$ is a $q \times 1$ vector of continuously differentiable functions with first-order derivative matrix $R(\theta) = \partial r(\theta)/\partial\theta'$. Denote $R = R(\theta_0)$. The F-test version of the Wald statistic for testing H_0 against H_1 is

$$F_T = \left[\sqrt{T} r(\hat{\theta}_T)\right]' \hat{\mathcal{V}}_R^{-1} \left[\sqrt{T} r(\hat{\theta}_T)\right] /q,$$

where $\hat{\mathcal{V}}_R$ is an estimator of the asymptotic covariance $\mathcal{V}_R$ of $R\sqrt{T}(\hat{\theta}_T - \theta_0)$. When $r(\cdot)$ is a scalar function, we can construct the t-statistic as $t_T = \sqrt{T} r(\hat{\theta}_T)/\sqrt{\hat{\mathcal{V}}_R}$.

It follows from (1) that $\mathcal{V}_R = R\mathcal{V}R'$. To make inference on θ_0, we have to estimate the unknown quantities in $\mathcal{V}$. $\mathcal{W}$ and G can be consistently estimated by their finite sample versions $\mathcal{W}_T$ and $\hat{G}_T = G_T(\hat{\theta}_T)$, respectively. It remains to estimate Ω. Let $\hat{\Omega}_T$ be an estimator of Ω. Then $\mathcal{V}_R$ can be estimated by

$$\hat{\mathcal{V}}_R = \hat{R}_T \left(\hat{G}_T' \mathcal{W}_T \hat{G}_T\right)^{-1} (\hat{G}_T' \mathcal{W}_T \hat{\Omega}_T \mathcal{W}_T \hat{G}_T) \left(\hat{G}_T' \mathcal{W}_T \hat{G}_T\right)^{-1} \hat{R}_T',$$

where $\hat{R}_T = R(\hat{\theta}_T)$.

Many nonparametric estimators of Ω are available in the literature. The most popular ones are kernel estimators, which are based on the early statistical literature on spectral density estimation. See Priestley (1981). Andrews (1991) and Newey and West (1987) extend earlier results to econometric models where the LRV estimation is based on estimated processes. In this paper, we follow den Haan and Levin (1997, 1998) and consider estimating the LRV by vector autoregression. The autoregression approach can be traced back to Whittle (1954). Berk (1974) provides the first proof of the consistency of the autoregressive spectrum estimator.

Let

$$h_t = \hat{R}_T (\hat{G}_T' \mathcal{W}_T \hat{G}_T)^{-1} \hat{G}_T' \mathcal{W}_T f(v_t, \hat{\theta}_T) \tag{2}$$

be the transformed moment conditions based on the estimator $\hat{\theta}_T$. Note that h_t is a vector process of dimension q. We outline the steps involved in the VAR covariance estimation below.

1. Fit a VAR(p) model to the estimated process h_t using the Yule-Walker method (see, for example, Lütkepohl (2007)):

   $$h_t = \hat{A}_1 h_{t-1} + \ldots + \hat{A}_p h_{t-p} + \hat{e}_t,$$

   where $\hat{A}_1, \ldots, \hat{A}_p$ are estimated autoregression coefficient matrices and $\hat{e}_t$ is the fitted residual. More specifically,

   $$\hat{A} = \left(\hat{A}_1, \ldots, \hat{A}_p\right) = [\hat{\Gamma}_h(1), \ldots, \hat{\Gamma}_h(p)] \hat{\mathbf{\Gamma}}_H^{-1}(p), \tag{3}$$

where

$$\hat{\Gamma}_h(j) = \begin{cases} T^{-1}\sum_{t=j+1}^{T} h_t h_{t-j}', & j \geq 0 \\ T^{-1}\sum_{t=1}^{T+j} h_t h_{t-j}', & j < 0 \end{cases}$$

is a $q \times q$ matrix and

$$\hat{\mathbf{\Gamma}}_H(p) = \begin{bmatrix} \hat{\Gamma}_h(0) & \dots & \hat{\Gamma}_h(p-1) \\ \vdots & \ddots & \vdots \\ \hat{\Gamma}_h(-p+1) & \dots & \hat{\Gamma}_h(0) \end{bmatrix}.$$

2. Compute

$$\hat{\Sigma}_e = \hat{\Gamma}_h(0) - \hat{A}_1\hat{\Gamma}_h(1) - \ldots - \hat{A}_p\hat{\Gamma}_h(p) \tag{4}$$

and estimate $\mathcal{V}_R$ by

$$\hat{\mathcal{V}}_R = \left(\mathbb{I}_q - \hat{A}_1 - \ldots - \hat{A}_p\right)^{-1}\hat{\Sigma}_e\left(\mathbb{I}_q - \hat{A}_1' - \ldots - \hat{A}_p'\right)^{-1}$$

where $\mathbb{I}_q$ is the $q \times q$ identity matrix.

We note that the Yule-Walker estimator can be computed using OLS based on a simple augmented data set. For more details, see Stoica and Moses (2005, pages 97–98). So it is easy to compute the Yule-Walker estimator using standard statistical packages.

It is important to point out that we fit a VAR(p) model to the transformed moment condition h_t instead of the original moment condition $f(v_t, \hat{\theta}_T)$. There are several advantages of doing this. First, the dimension of h_t can be much smaller than the dimension of $f(v_t, \hat{\theta}_T)$, especially when there are many moment conditions. So the VAR(p) model for h_t may have substantially fewer parameters than the VAR model for $f(v_t, \hat{\theta}_T)$. Second, by construction $\sum_{t=1}^{T} h_t = 0$, so an intercept vector is not needed in the VAR for h_t. On the other hand, when the model is overidentified, that is, $m > d$, $\sum_{t=1}^{T} f(v_t, \hat{\theta}_T) \neq 0$ in general. Hence, a VAR model for $f(v_t, \hat{\theta}_T)$ should contain an intercept. Finally and more importantly, h_t is tailored to the null hypothesis under consideration. The VAR order we select will reflect the null directly. In contrast, autoregressive fitting on the basis of $f(v_t, \hat{\theta}_T)$ completely ignores the null hypothesis, and the resulting covariance estimator $\hat{\mathcal{V}}_R$ may be poor in finite samples.

Let

$$\hat{\mathbf{A}} = \begin{bmatrix} \hat{A}_1 & \dots & \hat{A}_{p-1} & \hat{A}_p \\ \mathbb{I}_q & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & \mathbb{I}_q & 0 \end{bmatrix} \text{ and } \hat{\Sigma}_E = \begin{bmatrix} \hat{\Sigma}_e & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

then the Yule-Walker estimators $\hat{\mathbf{A}}$ and $\hat{\Sigma}_E$ satisfy:

$$\hat{\mathbf{\Gamma}}_H\left(p\right) = \hat{\mathbf{A}}\hat{\mathbf{\Gamma}}_H\left(p\right)\hat{\mathbf{A}}' + \hat{\Sigma}_E. \tag{5}$$

It is well known that for scalar time series the estimated AR model obtained via the Yule-Walker method is stationary almost surely. See Brockwell and Davis (1987, ch 8.1) and Lütkepohl (2007, ch 3.3.4). We expect this to hold for vector time series. To the best of authors' knowledge, a rigorous proof for the multivariate case is currently lacking in the statistical literature. We collect the stationarity result in the proposition below and provide a simple proof in the appendix.

Proposition 1 *If* $\hat{\mathbf{\Gamma}}_H\left(p\right)$ *and* $\hat{\mathbf{\Gamma}}_H\left(p+1\right)$ *are Toeplitz matrices and are positive definite almost surely, then* $\left\|\lambda_{\hat{\mathbf{A}}}\right\| < 1$ *almost surely where* $\lambda_{\hat{\mathbf{A}}}$ *is any eigenvalue of* $\hat{\mathbf{A}}$.

Proposition 1 is a crucial result that drives our asymptotic development. It gives precise conditions under which the fitted VAR(p) process is stationary. The proof requires that the covariance estimators $\hat{\mathbf{\Gamma}}_H(p)$ and $\hat{\mathbf{\Gamma}}_H(p+1)$ be Toeplitz matrices. For the Yule-Walker estimator, these conditions are satisfied. Hence the fitted VAR(p) model is stationary almost surely. For the OLS estimator, the corresponding matrices are not Toeplitz. As a result, the fitted VAR(p) model may not be stationary.

3 Fixed-smoothing Asymptotics for the Covariance Estimator

In this section, we derive the asymptotic distribution of $\hat{\mathcal{V}}_R$. Depending on how the VAR order p and the sample size T go to infinity, there are several different types of asymptotics. When the VAR order is set equal to a fixed proportion of the sample size, i.e. $p = bT$ for a fixed constant $b \in (0,1)$, we obtain the so-called fixed-smoothing asymptotics. On the other hand, if $b \to 0$ at the rate given in den Haan and Levin (1998), we obtain the conventional increasing-smoothing asymptotics. Under this type of asymptotics, $b \to 0$ and $T \to \infty$ jointly. So the increasing-smoothing asymptotics is a type of joint asymptotics. An intermediate case is obtained when we let $T \to \infty$ for a fixed b followed by letting $b \to 0$. Given the sequential nature of the limiting behavior of b and T, we call the intermediate case the sequential asymptotics.

An important property of the Yule-Walker estimator is that conditional on $\hat{A}_1, \ldots, \hat{A}_p$ and $\hat{\Sigma}_e$, the fitted VAR(p) process has theoretical autocovariances that are identical to the sample autocovariances up to lag p. To see this, consider a generic VAR(p) process $\tilde{h}_t$,

$$\tilde{h}_t = A_1\tilde{h}_{t-1} + \ldots + A_p\tilde{h}_{t-p} + \tilde{e}_t,$$

where $\tilde{e}_t \sim iid(0, \Sigma_e)$ and A_i are autoregression matrices. Define

$$\mathbf{\Gamma}_H(p) = \begin{bmatrix} \Gamma(0) & \ldots & \Gamma(p-1) \\ \vdots & \ddots & \vdots \\ \Gamma(-p+1) & \ldots & \Gamma(0) \end{bmatrix}$$

where $\Gamma(j) = E\tilde{h}_t\tilde{h}'_{t-j}$. Then the autocovariance sequence satisfies

$$\mathbf{\Gamma}_H(p) = \mathbf{A}\mathbf{\Gamma}_H(p)\mathbf{A}' + \Sigma_E, \tag{6}$$

where $\mathbf{A}$ and Σ_E are defined similarly as $\hat{\mathbf{A}}$ and $\hat{\Sigma}_E$. It follows that

$$vec\left[\mathbf{\Gamma}_H(p)\right] = \left[\mathbb{I}_{p^2q^2} - (\mathbf{A}\otimes\mathbf{A})\right]^{-1} vec\left(\Sigma_E\right).$$

That is, when $\mathbb{I}_{p^2q^2} - (\mathbf{A}\otimes\mathbf{A})$ is invertible, we can represent the autocovariances of $\{\tilde{h}_t\}$ as a function of $A_1, \ldots, A_p$ and Σ_e :

$$\Gamma(j) \equiv \Gamma_j(A_1, \ldots, A_p, \Sigma_e), \; j = 0, 1, \ldots, p. \tag{7}$$

That is, given $A_1, \ldots, A_p, \Sigma_e$, the set of autocovariances is uniquely determined by the function Γ_j in (7).

By the definition of the Yule-Walker estimator, $\hat{\mathbf{A}}$ and $\hat{\Sigma}_E$ satisfy $\hat{\mathbf{\Gamma}}_H(p) = \hat{\mathbf{A}}\hat{\mathbf{\Gamma}}_H(p)\hat{\mathbf{A}}' + \hat{\Sigma}_E$. Comparing this with the theoretical autocovariance sequence in (6) and in view of (7), we have

$$\hat{\Gamma}_h(j) = \Gamma_j\left(\hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e\right), \; j = 0, 1, \ldots, p,$$

provided that $\mathbb{I}_{p^2q^2} - \hat{\mathbf{A}}\otimes\hat{\mathbf{A}}$ is invertible. The almost sure invertibility of $\mathbb{I}_{p^2q^2} - \hat{\mathbf{A}}\otimes\hat{\mathbf{A}}$ is guaranteed by Proposition 1. In other words, conditional on $\hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e$, the autocovariances of the fitted VAR(p) process match exactly with the empirical autocovariances used in constructing the Yule-Walker estimator.

Using this 'reproducing' property of the Yule-Walker estimator, we can relate the VAR covariance estimator to the kernel estimator of $\mathcal{V}_R$ based on the rectangular kernel. Let

$k_{rect}(r) = \mathbf{1}\{|r| \leq 1\}$ and $k_{rect,b}(r) = \mathbf{1}\{|r| \leq b\}$, where $\mathbf{1}\{\cdot\}$ is the indicator function. Then the rectangular kernel estimator of $\mathcal{V}_R$ is

$$\tilde{\mathcal{V}}_R = \sum_{j=-p}^{p} \hat{\Gamma}_h(j) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} h_t h_s' k_{rect}\left(\frac{t-s}{p}\right),$$

where h_t is defined in (2) and p is the bandwidth or truncation lag. By definition, $\hat{\mathcal{V}}_R = \tilde{\mathcal{V}}_R + M_R$ where $M_R = \sum_{|j|>p} \tilde{\Gamma}_h(j)$,

$$\tilde{\Gamma}_h(j) = \sum_{i=1}^{p} \hat{A}_i \hat{\Gamma}_h(j-i) \text{ for } j > p, \text{ and } \tilde{\Gamma}_h(-j) = \tilde{\Gamma}_h(j)' \text{ for } j < -p. \tag{8}$$

Intuitively, the fitted VAR process necessarily agrees exactly up to lag order p with the estimated autocovariances. The values of the autocovariances after lag p are generated recursively in accordance with the VAR(p) model as in (8). The difference between the VAR covariance estimator and the rectangular kernel covariance estimator is that for the former estimator the autocovariances of order greater than p are based on the VAR(p) extrapolation while for the latter estimator these autocovariances are assumed to be zero.

Using the relationship between the VAR covariance estimator and the rectangular kernel covariance estimator, we can establish the asymptotic distribution of the VAR covariance estimator under the fixed-smoothing asymptotics. We make the following assumptions.

Assumption 1 *$plim_{T\to\infty}\hat{\theta}_T = \theta_0$.*

Assumption 2 *$T^{1/2} g_{[rT]}(\theta_0) \Rightarrow \Lambda W_m(r)$ where $\Lambda\Lambda' = \Omega = \sum_{j=-\infty}^{\infty} E u_t u_{t-j}' > 0$ is the LRV of $u_t := f(v_t, \theta_0)$.*

Assumption 3 *$plim_{T\to\infty} G_{[rT]}(\tilde{\theta}_T) = rG$ uniformly in r for any $\tilde{\theta}_T$ between $\hat{\theta}_T$ and θ_0 where $G = E[\partial f(v_j, \theta_0)/\partial\theta']$.*

Assumption 4 *$\sum_{\tau=1}^{T-bT} f(v_{\tau+bT}, \theta_0) g_\tau'(\theta_0) \Rightarrow \Lambda \int_0^{1-b} dW_m(b+r) W_m'(r)\Lambda'$.*

Assumption 5 *$\mathcal{W}_T$ is positive semidefinite, $plim_{T\to\infty}\mathcal{W}_T = \mathcal{W}$, and $G'\mathcal{W}G$ is positive definite.*

Assumption 1 is made for convenience. It can be proved under more primitive assumptions and using standard arguments. Assumptions 2 and 3 are similar to those in Kiefer and Vogelsang (2005), among others. Assumption 2 regulates $\{f(v_t, \theta_0)\}$ to obey a functional central limit theorem (FCLT) while Assumption 3 requires $\{\partial f(v_j,\theta_0)/\partial\theta'\}$ satisfying a uniform law of large numbers (ULLN). Note that FCLT and ULLN hold for serially correlated and heterogeneously distributed data that satisfy certain regularity conditions on moments and the dependence structure over time. These primitive regularity conditions are quite technical and can be found in White (2001). Assumption 4 is a new high-level condition we maintain for the rectangular kernel. It is not needed to derive the fixed-smoothing asymptotics for other kernels. Using the same argument as in de Jong and Davidson (2000), we can show that under some moment and mixing conditions on the process $\{f(v_t,\theta_0)\}$:

$$\sum_{\tau=1}^{T-bT} f\left(v_{\tau+bT},\theta_0\right) g_\tau'\left(\theta_0\right) - \Lambda_T^+ \Rightarrow \Lambda\int_0^{1-b} dW_m(b+r)W_m'(r)\Lambda',$$

where $\Lambda_T^+ = T^{-1}\sum_{\tau=1}^{T-bT}\sum_{t=1}^{\tau} Eu_{\tau+bT}u_t'$. But for $\Gamma_u(s-t)=Eu_su_t'$, we have

$$\begin{aligned}
\Lambda_T^+ &= \frac{1}{T}\sum_{\tau=1}^{T-bT}\sum_{t=1}^{\tau}\Gamma_u\left(\tau+bT-t\right) = \frac{1}{T}\sum_{\tau=1}^{T-bT}\sum_{j=0}^{\tau-1}\Gamma_u\left(bT+j\right)\\
&= \frac{1}{T}\sum_{j=0}^{T-bT-1}\sum_{\tau=j+1}^{T-bT}\Gamma_u\left(bT+j\right) = \sum_{j=0}^{T-bT-1}\left(1-b-\frac{j}{T}\right)\Gamma_u\left(bT+j\right)\\
&= \sum_{j=0}^{\infty}\Gamma_u\left(bT+j\right)+o(1)=o(1),
\end{aligned}$$

where we have assumed the stationarity of $\{f(v_t,\theta_0)\}$ and the absolute summability of its autocovariances. Hence Assumption 4 holds under some regularity conditions.

Lemma 1 *Let Assumptions 1-5 hold. Then under the fixed-smoothing asymptotics,* $M_R = o_p(1)$ *and* $\hat{\mathcal{V}}_R \Rightarrow \mathcal{V}_{R,\infty}$ *where*

$$\mathcal{V}_{R,\infty} = \left[R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda\right]Q_m(b)\left[R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda\right]'$$

$$Q_m(b) = -\left(\int_0^{1-b} dV_m(b+r)V_m'(r) + \int_0^{1-b} V_m(r)dV_m'(r+b)\right) \qquad (9)$$

and $V_m(r)=W_m(r)-rW_m(1)$ *is the standard Brownian bridge process.*

The fixed-smoothing asymptotics of $\hat{\mathcal{V}}_R$ is new in the literature. Because of the discontinuity of the rectangular kernel, the typical arguments for deriving the fixed-smoothing asymptotics do not go through without modification. Here we take advantage of the explicit form of the kernel function and use the indispensable Assumption 4 to prove the lemma.

The asymptotic distribution $\mathcal{V}_{R,\infty}$ is proportional to the true covariance matrix $\mathcal{V}_R$ through $R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda$. This contrasts with the increasing-smoothing asymptotic approximation where $\hat{\mathcal{V}}_R$ is approximated by a constant matrix, i.e. $\mathcal{V}_R$. The advantage of the fixed-smoothing asymptotic result is that the limit of $\hat{\mathcal{V}}_R$ depends on the order of the autoregression through b but is otherwise nuisance parameter free. Therefore, it is possible to obtain a first-order asymptotic distribution theory that explicitly captures the effect of the VAR order used in constructing the VAR covariance estimator.

The following lemma gives an alternative representation of $Q_m(b)$. Using this representation, we can compute the variance of $\mathcal{V}_{R,\infty}$. The representation uses the centered kernel defined by

$$
\begin{aligned}
& k_b^*(r,s) \\
&= k_{rect}\left(\frac{r-s}{b}\right) - \int_0^1 k_{rect}\left(\frac{r-s}{b}\right)dr - \int_0^1 k_{rect}\left(\frac{r-s}{b}\right)ds + \int_0^1\int_0^1 k_{rect}\left(\frac{r-s}{b}\right)drds \\
&= k_{rect}\left(\frac{r-s}{b}\right) + \max(0,r-b) + \max(0,s-b) - \min(1,b+r) - \min(1,b+s) - b\left(b-2\right).
\end{aligned}
\tag{10}
$$

Lemma 2 *(a) $Q_m(b)$ can be represented as*

$$Q_m(b) = \int_0^1\int_0^1 k_b^*(r,s)dW_m(s)dW_m'(r),$$

(b) $E\Lambda Q_m(b)\Lambda' = \mu_1\Omega$ and

$$\operatorname{var}(\operatorname{vec}(\Lambda Q_m(b)\Lambda')) = \mu_2\left(\Omega\otimes\Omega\right)\left(\mathbb{I}_{m^2}+\mathbb{K}_{m^2}\right),$$

(c) $E\mathcal{V}_{R,\infty} = \mu_1\mathcal{V}_R$ and

$$\operatorname{var}(\operatorname{vec}(\mathcal{V}_{R,\infty})) = \mu_2\left(\mathcal{V}_R\otimes\mathcal{V}_R\right)\left(\mathbb{I}_{q^2}+\mathbb{K}_{q^2}\right),$$

where

$$
\begin{aligned}
\mu_1 &= \mu_1(b) = \int_0^1 k_b^*(r,r)dr = (1-b)^2 \\
\mu_2 &= \mu_2(b) = \int_0^1\int_0^1 \left[k_b^*(r,s)\right]^2 drds = \begin{cases} b\left(3b^3+8b^2-15b+6\right)/3, & b\le 1/2 \\ (b-1)^2\left(3b^2-2b+2\right)/3, & b>1/2 \end{cases}
\end{aligned}
$$

and $\mathbb{K}_{m^2}$ is the $m^2\times m^2$ commutation matrix.

It is often tedious to find the moments of the fixed-smoothing limiting distribution. The calculation has to be done on a case-by-case basis. For this reason, explicit moment formulae are not readily available in the literature. An exception is Hashimzade and Vogelsang (2007) who compute the first two moments of the fixed-smoothing limiting distribution for the Bartlett kernel. Lemma 2 is another attempt in this direction. It allows us to compare the first two moments of the fixed-smoothing limiting distribution with those of the increasing-smoothing limiting distribution.

It follows from Lemma 2 that the mean of $\mathcal{V}_{R,\infty}$ is proportional to the true covariance $\mathcal{V}_R$. When $b \to 0$, we have $\mu_1 \to 1$ and $\mu_2(b) \to 0$. So $\text{plim}_{b\to 0}\mathcal{V}_{R,\infty} = \mathcal{V}_R$. A direct implication is that as b goes to zero, the fixed-smoothing asymptotics coincides with the conventional increasing-smoothing asymptotics. More precisely, the probability limits of $\hat{\mathcal{V}}_R$ are the same under the sequential asymptotics and the joint asymptotics.

As $b \to 0$, we have

$$\lim_{b\to 0} b^{-1}\text{var}(\text{vec}(\mathcal{V}_{R,\infty})) = 2\left(\mathcal{V}_R \otimes \mathcal{V}_R\right)\left(\mathbb{I}_{q^2} + \mathbb{K}_{q^2}\right).$$

Note that $\int_{-1}^{1} k_{rect}^2(r)\, dr = \int_{-1}^{1} \mathbf{1}\{|r| \leq 1\}\, dr = 2$. The right hand side is exactly the asymptotic variance one would obtain under the joint asymptotic theory. That is, the second moment of the fixed-smoothing asymptotic distribution becomes arbitrarily close to that of the increasing-smoothing asymptotic distribution. Therefore, $\hat{\mathcal{V}}_R$ has not only the same probability limit but also the same asymptotic variance under the sequential and joint asymptotics.

When $b > 0$ is fixed, $E\mathcal{V}_{R,\infty} - \mathcal{V}_R = b\,(b-2)\,\mathcal{V}_R$. So $\hat{\mathcal{V}}_R$ is not asymptotically unbiased. The asymptotic bias arises from the estimation uncertainty of model parameter θ. It may be called the demeaning bias as the stochastic integral in (9) depends on the Brownian bridge process rather than the Brownian motion process. One advantage of the fixed-smoothing asymptotics is its ability to capture the demeaning bias. In contrast, under the conventional increasing-smoothing asymptotics, the estimation uncertainty of θ is negligible. As a result, the first-order conventional asymptotics does not reflect the demeaning bias.

4 Fixed-smoothing Asymptotics for Test Statistics

In this section, we first establish the asymptotic distribution of F_T under the fixed-smoothing asymptotics. We then develop an F approximation to the nonstandard limiting distribution.

Our asymptotic results can be extended straightforwardly to the t-statistic.

The following theorem can be proved using Lemmas 1 and 2.

Theorem 2 *Let Assumptions 1-5 hold. Assume that R is of full rank q. Under the fixed-smoothing asymptotics where b is held fixed, we have $F_T \Rightarrow F_\infty(q,b)$ where*

$$F_\infty(q,b) = W_q'(1)\left[Q_q(b)\right]^{-1} W_q(1)/q,$$

and $Q_q(b) = \int_0^1 \int_0^1 k_b^*(r,s)dW_q(r)dW_q'(s)$.

Theorem 2 shows that $F_\infty(q,b)$ depends on b but otherwise is nuisance parameter free. So the new asymptotic theory gives rise to an asymptotically pivotal test that reflects the choice of the VAR order. This is in contrast with the asymptotic results under the standard approach where F_T would have a limiting χ_q^2/q distribution regardless of the choice of b or VAR order.

When $b \to 0$, $Q_q(b) \to^p \mathbb{I}_q$ and as a result $F_\infty(q,b) \Rightarrow \chi_q^2/q$. Hence, when $b \to 0$, the fixed-smoothing asymptotics approaches the standard increasing-smoothing asymptotics. In a sequence of papers on kernel LRV estimation with finite order kernels, Sun (2010) and Sun, Phillips and Jin (2008) show that critical values from the fixed-smoothing asymptotics are high-order correct under the conventional joint asymptotics. It is not difficult to generalize their result to infinite order kernels which include the rectangular kernel as a special case. In fact, in a simple Gaussian location model, Jun (2011) has obtained an Edgeworth expansion for the t-statistic using the autoregressive variance estimator. So the fixed-smoothing approximation is not only asymptotically valid but also second-order correct under the increasing-smoothing asymptotics.

The asymptotic distribution $F_\infty(q,b)$ is nonstandard. Critical values are not readily available from statistical tables or software packages. For this reason, we approximate $F_\infty(q,b)$ by a convenient F distribution. Let

$$\kappa = \exp(2qb) \text{ and } F_\infty^*(q,b) = F_\infty(q,b)/\kappa$$

The theorem below follows from Theorem 1 of Sun (2010).

Theorem 3 *As $b \to 0$, we have*

$$P\left(F_\infty^*(q,b) \le z\right) = P\left(F_{q,K^*} \le z\right) + o(b) \tag{11}$$

where

$$K^* = \max(\lceil (2b)^{-1} \rceil - q + 1, 1). \tag{12}$$

Let F^{α}_{q,K^*} and $F^{\alpha}_{\infty}(q,b)$ be the $1-\alpha$ quantiles of the standard F_{q,K^*} distribution and the nonstandard $F_{\infty}(q,b)$ distribution, respectively. Then

$$P\left(F_{\infty}(q,b) > \kappa F^{\alpha}_{q,K^*}\right) = \alpha + o(b), \tag{13}$$

which implies that $F^{\alpha}_{\infty}(q,b) = \kappa F^{\alpha}_{q,K^*} + o(b)$. Since the approximation error is of smaller order $o(b)$ rather than $O(b)$ as $b \to 0$, $\kappa F^{\alpha}_{q,K^*}$ is second-order correct. So for the original F statistic, we can use

$$\mathcal{F}^{\alpha}_{q,b} = \kappa F^{\alpha}_{q,K^*} \tag{14}$$

as the critical value for the test with nominal size α. In the rest of the paper, we call this test the F^* test. To emphasize its reliance on vector autoregression, we also refer to it as the VAR F^* test. Of course, the VAR F^* test is equivalent to the test that is based the modified Wald statistic F_T/κ and uses F^{α}_{q,K^*} as the critical value. We use these two definitions interchangingly.

The correction factor κ in (14) can be regarded as a Bartlett-type correction; see Bartlett (1937, 1954). The idea is to choose κ to make the mean of $F^{*}_{\infty}(q,b)$ closer to that of χ^2_q/q. In addition to the Bartlett-type correction, Theorem 3 approximates the nonstandard distribution by an F distribution rather than a chi-square distribution.

So far the F approximation has been motivated from the fixed-smoothing asymptotics. It can also be obtained by developing a second-order expansion under the conventional increasing-smoothing asymptotics. See Sun (2010) for details. To save space, we do not present them here. So the F approximation can be justified under both the sequential asymptotics and the second-order joint asymptotics. It does not have to be regarded as an approximation to $F_{\infty}(q,b)$. Under the joint asymptotics, both the F distribution and the nonstandard distribution $F_{\infty}(q,b)$ are second-order correct. Without a third-order joint asymptotic theory, it is not clear which distribution provides a more accurate approximation to the finite sample distribution of the test statistic.

We end this section by discussing the t-statistic. All the results for the Wald statistic can be extended to the t-statistic with $q=1$. For example, we can show that $t_T \Rightarrow W_1(1)/\sqrt{Q_1(b)}$. In addition, t_T/κ can be approximated by the t-distribution with degree of freedom K^* as defined in (12). In Gaussian location models, the second order correctness of our t-approximation under the conventional asymptotic theory follows from Jun (2011).

5 VAR Lag Order Determination

For VAR models, it is standard practice to use model selection criteria such as AIC or BIC to choose the lag order. However, the AIC and BIC are not aimed at the testing problem we consider. In this section, we propose a new lag order selection rule that is based on the bandwidth choice for the rectangular kernel LRV estimator. We set the VAR lag order equal to the bandwidth (i.e. bT) for the rectangular kernel LRV estimator.

The question is how to select the bandwidth for the rectangular kernel LRV estimator that is directed at the testing problem at hand. Before addressing this question, we review the method proposed by Sun (2010) who considers finite-order kernel LRV estimators and associated F^* tests. He proposes to select the bandwidth to minimize an asymptotic measure of the type II error while controlling for the asymptotic type I error. More specifically, the testing-optimal bandwidth is given by

$$b^* = \arg\min e_{II}(b), \text{ s.t. } e_I(b) \leq \alpha\tau \tag{15}$$

where $e_I(b)$ and $e_{II}(b)$ are approximate measures of type I and type II errors and $\tau > 1$ is the so-called tolerance parameter.

Under some regularity conditions, for a kernel function $k(x)$ with Parzen exponent ϱ, the type I error of the kernel F^* test is shown to approximately equal

$$e_I(b) = \alpha - (bT)^{-\varrho} G_q'(\mathcal{X}_q^\alpha)\mathcal{X}_q^\alpha \bar{B} \text{ with } \bar{B} = \frac{tr(B\mathcal{V}_R^{-1})}{q}$$

where α is the nominal type I error, $\mathcal{X}_q^\alpha$ is the α-level critical value from $G_q(\cdot)$, the CDF of the χ_q^2 distribution, and B is the asymptotic bias of the kernel LRV estimator for $\mathcal{V}_R$. The average type II error under the local alternative $H_1(\delta_o^2) : r(\theta_0) = (R\mathcal{V}R')^{1/2}\tilde{c}/\sqrt{T}$ where $\tilde{c}$ is uniformly distributed on $\mathcal{S}_q(\delta_o^2) = \{\tilde{c} \in \mathbb{R}^q : \|\tilde{c}\|^2 = \delta_o^2\}$ is

$$e_{II}(b) = G_{q,\delta_o^2}(\mathcal{X}_q^\alpha) + (bT)^{-\varrho} G_{q,\delta_o^2}'(\mathcal{X}_q^\alpha)\mathcal{X}_q^\alpha \bar{B} + \frac{\delta_o^2}{2} G_{(q+2),\delta_o^2}'(\mathcal{X}_q^\alpha)\mathcal{X}_q^\alpha c_2 b$$

where $G_{\ell,\delta_o^2}(\cdot)$ is the CDF of the noncentral $\chi_\ell^2(\delta_o^2)$ distribution and $c_2 = \int_{-\infty}^{\infty} k^2(x)\,dx$. In the above expression, higher-order terms and a term of order $1/\sqrt{T}$ that does not depend on

b have been omitted. Note that δ_o^2 is not known in practice. We choose δ_o^2 such that the first order power of the chi-square test is 75%. This is a reasonable choice when we do not have any information on local alternatives.

The testing optimal bandwidth depends on the sign of $\bar{B}$. When $\bar{B} < 0$, the constraint $e_I(b) \leq \alpha\tau$ is binding and the optimal b^* satisfies $e_I(b^*) = \alpha\tau$. When $\bar{B} > 0$, the constraint $e_I(b) \leq \alpha\tau$ is not binding and the optimal b^* minimizes $e_{II}(b)$.

The problem is that the above testing optimal bandwidth works only for finite-order kernels while the rectangular kernel is an infinite-order kernel. For infinite order kernels, we cannot obtain an exact order of the asymptotic bias with an explicit data-dependent coefficient, although we may be able to obtain that the asymptotic bias is of smaller order than some rate. For example, with enough smoothness conditions, the asymptotic bias can be made of order $o(1/\sqrt{T})$. A similar problem is also present for optimal bandwidth choice under the MSE criterion, as an explicit asymptotic bias formula is lacking. This is a general problem in nonparametric estimation with infinite order kernels. To address this problem, we employ a second-order kernel as the target kernel and use its testing-optimal bandwidth as a basis for bandwidth selection for the rectangular kernel. Our proposal is ad hoc to some degree, but there is no procedure that is free from being ad hoc in this situation. Further, we show that the rectangular kernel with our proposed bandwidth improves upon the size and power properties of the target kernel.

Let $k_{tar}(\cdot)$ be the target kernel and b_{tar}^* be the associated testing-optimal bandwidth parameter. For example, we may let $k_{tar}(\cdot)$ be the Parzen kernel, the QS kernel, or any other commonly used finite-order kernel. We set the bandwidth for the rectangular kernel to be

$$b_{rect}^* = \begin{cases} b_{tar}^*, & \text{if } \bar{B} < 0 \\ (c_{2,tar})(c_{2,rect})^{-1} b_{tar}^*, & \text{if } \bar{B} > 0 \end{cases} \tag{16}$$

where $c_{2,tar} = \int_{-\infty}^{\infty} k_{tar}^2(x)\,dx$, $c_{2,rect} = \int_{-\infty}^{\infty} k_{rect}^2(x)\,dx = 2$, and

$$b_{tar}^* = \begin{cases} \left[\frac{G_p'(\mathcal{X}_p^\alpha)\mathcal{X}_p^\alpha|\bar{B}|}{(\tau-1)\alpha}\right]^{1/q} \frac{1}{T}, & \bar{B} < 0 \\ \left[\frac{2qG'_{p,\delta_o^2}(\mathcal{X}_p^\alpha)|\bar{B}|}{\delta_o^2 G'_{(p+2),\delta^2}(\mathcal{X}_p^\alpha)c_{2,tar}}\right]^{\frac{1}{q+1}} T^{-\frac{q}{q+1}}, & \bar{B} > 0 \end{cases}$$

For example, when the Parzen kernel is used as the target kernel,

$$b_{rect}^* = b_{Parzen}^* \mathbf{1}\left\{\bar{B} < 0\right\} + \frac{0.539285}{2} b_{Parzen}^* \mathbf{1}\left\{\bar{B} > 0\right\}. \tag{17}$$

When the QS kernel is used as the target kernel,

$$b^*_{rect} = b^*_{QS}\mathbf{1}\left\{\bar{B} < 0\right\} + \frac{1}{2}b^*_{QS}\mathbf{1}\left\{\bar{B} > 0\right\}. \tag{18}$$

Given b^*_{rect}, we set the VAR lag order to be $p = \lceil b^*_{rect}T \rceil$. For convenience, we refer to this bandwidth selection and lag order determination method as the method of target kernels (MTK).

When $\bar{B} < 0$, the bandwidth based on the MTK is the same as the testing-optimal bandwidth for the target kernel. In this case, all F^* tests are expected to be over sized, thanks to the asymptotic bias of the associated LRV estimator. For a given bandwidth parameter and under some regularity conditions, the asymptotic bias of the rectangular kernel LRV estimator is of smaller order than that of any finite-order kernel (see Politis, 2011). As a consequence, the bandwidth selected by the MTK is expected to control the type I error better than the testing-optimal bandwidth selection rule for the target kernel.

When $\bar{B} > 0$, the type I error of the F^* test is expected to be capped by the nominal type I error. This gives us the opportunity to select the bandwidth to minimize the type II error without worrying about over rejection. With the bandwidth selected by the MTK, the third term of the form $\delta_o^2 G'_{(q+2),\delta_o^2}\left(\mathcal{X}_q^\alpha\right)\mathcal{X}_q^\alpha c_2 b/2$ in $e_{II}(b)$ is the same for the rectangular kernel and the target kernel, while the second term is expected to be smaller for the rectangular kernel. Therefore, the F^* test based on the rectangular kernel and the MTK is expected to have smaller type II error than the F^* test based on the target kernel with testing-optimal bandwidth choice.

To sum up, when the F^* tests are expected to over-reject, the rectangular kernel with bandwidth selected by the MTK delivers an F^* test with a smaller type I error than the corresponding target kernel. On the other hand, when the F^* tests are expected to under-reject so that the asymptotic type I error is capped by the nominal type I error, the F^* test based on the rectangular kernel and the MTK is expected to have smaller type II error than the F^* test based on the finite-order target kernel.

Our bandwidth selection rule via the MTK bears some resemblance to a rule suggested by Andrews (1991, footnote on page 834). Andrews (1991) employs the MSE criterion and suggests setting the bandwidth for the rectangular kernel equal to the half of the MSE-optimal bandwidth for the QS kernel. Essentially, Andrews (1991) uses the QS kernel as the target kernel. This is a natural choice as the QS kernel is the optimal kernel under the MSE criterion in the class of positive semidefinite kernels. Lin and Sakata (2009) make the same recommendation and show that the resulting rectangular kernel LRV estimator has smaller asymptotic MSE than the QS kernel LRV estimator. When $\bar{B} > 0$, the MTK is analogous to that suggested by Andrews (1991) and Lin and Sakata (2009). However, when $\bar{B} < 0$ such that the F^* tests tend to over-reject, the MTK is different. It suggests using the same bandwidth, rather than a fraction of it, as the bandwidth for the target kernel in order to control the size distortion.

6 Simulation Study

This section provides some simulation evidence on the finite sample performance of the VAR F^* test. We compare the VAR F^* test with the nonstandard VAR test, the standard chi-square test and kernel-based F^* tests recently proposed by Sun (2010).

6.1 Location model

In our first simulation experiment, we consider a multivariate location model of the form

$$y_t = \theta + u_t,$$

where $y_t = (y_{1t}, y_{2t}, y_{3t})'$, $u_t = (u_{1t}, u_{2t}, u_{3t})'$ and $\theta = (\theta_1, \theta_2, \theta_3)'$. The error processes $\{u_{it}\}$ are independent of each other. We consider two cases. In the first case, all components of u_{it} follow the same AR(2) process:

$$u_{it} = \rho_1 u_{it-1} + \rho_2 u_{it-2} + e_{it}$$

where $e_{it} \sim iid(0, \sigma_e^2)$ and $\sigma_e^2 = (1+\rho_2)\left[(1-\rho_2)^2 - \rho_1^2\right](1-\rho_2)^{-1}$. In the second case, all components of u_{it} follow the same MA(2) process:

$$u_{it} = \rho_1 e_{it-1} + \rho_2 e_{it-2} + e_{it}$$

where $e_{it} \sim iid(0, \sigma_e^2)$ and $\sigma_e^2 = \left(1 + \rho_1^2 + \rho_2^2\right)^{-1}$. In both cases, the value of σ_e^2 is chosen such that the variance of u_{it} is one. The error e_{it} follows either a normal distribution or centered chi-square distributions with different degrees of freedom.

We consider the following null hypotheses:

$$H_{0q} : \theta_1 = \ldots = \theta_q = 0$$

for $q = 1, 2, 3$. The corresponding restriction matrix is $R_{0q} = \mathbb{I}_d(1 : q, :)$, i.e., the first q rows of the identity matrix $\mathbb{I}_3$. The local alternative hypothesis is $H_{1q}\left(\delta^2\right) : R_{0q}\theta = c_q/\sqrt{T}$ where $c_q = (R_{0q}\Omega R_{0q}')^{1/2}\tilde{c}$, Ω is the LRV matrix of u_t, $\tilde{c}$ is uniformly distributed over the sphere $\mathcal{S}_q\left(\delta^2\right)$, that is, $\tilde{c} = \delta\xi/\left\|\xi\right\|, \xi \sim N(0, \mathbb{I}_q)$. It is important to point out that δ^2 is not the same as δ_o^2 used in the testing-oriented criterion and the MTK.

We consider the following (ρ_1, ρ_2) combinations:

$$(\rho_1, \rho_2) = (-.8, 0), (-.4, 0), (0, 0), (.4, 0), (.8, 0), (1.5, -.75), (.25, .25), (.35, .35).$$

The last two combinations come from den Haan and Levin (1998). The combination with negative ρ_2 comes from Kiefer and Vogelsang (2002a,b). The remaining combinations consist of simple AR(1) or MA(1) models with different persistence.

We consider two sets of testing procedures. The first set consists of the tests using the VAR covariance estimator. For each restriction matrix R_{0q}, we fit a VAR(p) model to $R_{0q}\left(u_t - \bar{u}\right)$ by OLS. We select the lag order of each VAR model by AIC or BIC. As standard model selection methods, the details on AIC and BIC can be found in many textbooks and papers, see for example, Lütkepohl (2007, sec 4.3) and den Haan and Levin (1998). We also consider selecting the VAR order by the MTK, that is $p = \lceil b_{rect}^* T\rceil$ where b_{rect}^* is defined in (16). We use Parzen and QS kernels as the target kernels. We call the resulting two VAR order selection rules the VAR-Par rule and VAR-QS rule.

For each of the VAR order determination methods, we construct the VAR covariance estimator and compute the Wald statistic. We perform the nonstandard test, the F^* test, and the traditional χ^2 test. The nonstandard test is based on Theorem 2. It employs the Wald statistic F_T and simulated critical values from the nonstandard limiting distribution. For convenience, we call the nonstandard test the S^2 test, reflecting its Simulation-based nature and the similarity between the nonstandard distribution and Hotelling's T^2 distribution. The F^* test employs the Wald statistic F_T and the critical value $\kappa F_{q,\hat{K}}^{\alpha}$ where $\hat{K} = \max(\lceil T/(2\hat{p})\rceil - q + 1, 1)$ and $\hat{p}$ is the selected VAR order. The traditional χ^2 test employs the Wald statistic F_T and the critical value $\mathcal{X}_q^{\alpha}/q$.

The second set of testing procedures consists of kernel-based tests. We consider two commonly used second-order kernels: the Parzen and QS kernels. For each kernel, the bandwidth is determined via the testing-oriented criterion (Sun 2010). The bandwidth is selected to solve the constrained minimization problem in (15). We set $\tau = 1.2$ in the simulation experiment.

As in the VAR case, there are three tests: the S^2 test which uses simulated critical values from a nonstandard fixed-smoothing asymptotic distribution, the F^* test which uses critical values from an F distribution, and the χ^2 test which uses critical values from a χ^2 distribution. The testing-optimal bandwidth requires a plug-in implementation. We use the VAR model selected by the AIC as the approximating parametric model.

To explore the finite sample size of the tests, we generate data under the null hypothesis. To compare the power of the tests, we generate data under the local alternative. For each test, we consider two significance levels $\alpha = 5\%$ and $\alpha = 10\%$, three different sample sizes $T = 100, 200, 500$. The number of simulation replications is 10000.

Tables 1-3 give the type I errors of the 18 testing methods for the AR error with normal distribution and sample size $T = 100$. The significance level is 5%, which is also the nominal type I error. We can make several observations from these tables. First, the conventional chi-square tests can have a large size distortion. The size distortion increases with both the error dependence and the number of restrictions being jointly tested. The size distortion can be very severe. For example, when $(\rho_1, \rho_2) = (.8, 0)$ and $q = 3$, the empirical type I error of a χ^2 test can be more than 50%, which is far from 0.05, the nominal type I error.

Second, the size distortion of the VAR F^* test is often substantially smaller than the corresponding VAR χ^2 test. Note that the lag order underlying the VAR F^* test is the same as that for the corresponding VAR χ^2 test. The VAR F^* test is more accurate in size because it employs an asymptotic approximation that captures the estimation uncertainty of the covariance estimator. Based on this observation, we can conclude that the proposed finite sample correction, coupled with the use of the F critical values, is very effective in reducing the size distortion of the χ^2 test. For the same reason, the size distortion of the F^*-Parzen and F^*-QS tests is also much smaller than that of the corresponding kernel-based χ^2 tests.

Third, among the F^* tests using the VAR covariance estimator, the test based on the MTK has the smallest size distortion. Unreported results show that in an average sense the VAR order selected by the MTK is the largest. In terms of size accuracy, the AIC and BIC appear to be conservative in that they choose smaller VAR lag orders. It is well known that BIC is consistent in selecting the true model. However, the error in approximating the distribution of F_T by an F distribution does not necessarily increase with the VAR lag order. Our simulation result shows that for size accuracy it is advantageous to use a larger model, even though this may lead to some power loss.

Fourth, when the error process is highly persistent, the VAR F^* test with the VAR order selected by the MTK is more accurate in size than the corresponding kernel F^* test. On the other hand, when the error process is not persistent, all the F^* tests have more or less the same size properties. So the VAR F^* test with the VAR order selected by the MTK reduces the size distortion when it is needed most, and maintains the good size property when it is not needed.

Finally, the VAR F^* test is as accurate in size as the nonstandard VAR S^2 test. In fact, when the error process is highly persistent, the VAR F^* test is more accurate that the VAR S^2 test. This provides some evidence that the standard F approximation is a convenient alternative to the nonstandard fixed-smoothing approximation.

Figures 1 and 2 present the finite sample power in the Gaussian AR case for $q = 2$, which is representative of the cases $q = 1$ and $q = 3$. We compute the power using 5% empirical finite sample critical values under the null. The empirical critical values are simulated separately for each testing method with data-driven smoothing parameter. So the finite sample power is size-adjusted and power comparisons are meaningful. It should be pointed out that the size adjustment is not feasible in practice. The parameter configuration is the same as those for Tables 1-3 except that the DGP is generated under the local alternatives. The power curves are for the F^* tests. We do not include S^2 and χ^2 tests as these three types of tests have the same size-adjusted power. It is clear from the figure that the VAR F^* test based on the AIC or BIC is more powerful than the other F^* tests. Among all F^* tests, the VAR F^* test based on the BIC is most powerful. However, this F^* test also has the largest size distortion. The power differences among the F^* tests are small in general. Compared with the kernel F^* test with testing optimal bandwidth, the VAR F^* test based on the MTK has very competitive power—sometimes it is more powerful than the kernel F^* test. Therefore, the VAR F^* test based on the MTK achieves more accurate size without sacrificing much power.

We omit the tables of simulated type I errors for the MA case, as the qualitative observations on size comparison for the AR case remain valid. In fact, these qualitative observations hold for other parameter configurations such as different sample sizes, significance levels, and error distributions. We do not present the power figures for the MA case but note that the qualitative observations on power comparison for the AR case still hold.

6.2 Regression model

In our second simulation experiment, we consider a regression model of the form:

$$y_t = \gamma + x_t'\beta + \varepsilon_t,$$

where x_t is a 3×1 vector process and x_t and ε_t follow either an AR (1) process

$$x_{t,j} = \rho x_{t-1,j} + \sqrt{1-\rho^2} e_{t,j},\ \varepsilon_t = \rho \varepsilon_{t-1} + \sqrt{1-\rho^2} e_{t,0}$$

or an MA(1) process

$$x_{t,j} = \rho e_{t-1,j} + \sqrt{1-\rho^2} e_{t,j},\ \varepsilon_t = \rho e_{t-1,0} + \sqrt{1-\rho^2} e_{t,0}.$$

The error term $e_{t,j} \sim iidN(0,1)$ across t and j. For this DGP, we have $m = d = 4$. Throughout we are concerned with testing for the regression parameter β and set $\gamma = 0$ without the loss of generality.

Let $\theta = (\gamma', \beta')'$. We estimate θ by the OLS estimator. Since the model is exactly identified, the weighting matrix $\mathcal{W}_T$ becomes irrelevant. Let $\tilde{x}_t' = [1, x_t']$ and $\tilde{X} = [\tilde{x}_1, \ldots, \tilde{x}_T]'$, then the OLS estimator is $\hat{\theta}_T - \theta_0 = -G_T^{-1} g_T(\theta_0)$ where $G_T = -\tilde{X}'\tilde{X}/T$, $G = E(G_T)$, $g_T(\theta_0) = T^{-1}\sum_{t=1}^{T} \tilde{x}_t \varepsilon_t$. The asymptotic variance of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is $\mathcal{V} = G^{-1}\Omega G^{-1}$ where Ω is the LRV matrix of the process $\tilde{x}_t \varepsilon_t$.

We consider the following null hypotheses:

$$H_{0q} : \beta_1 = \ldots = \beta_q = 0$$

for $q = 1, 2, 3$. The local alternative hypothesis is $H_{1q}(\delta^2) : R_{0q}\theta = c_q/\sqrt{T}$ where $c_q = \left(R_{0q} G^{-1} \Omega G^{-1} R_{0q}'\right)^{1/2} \tilde{c}$ and $\tilde{c}$ is uniformly distributed over the sphere $\mathcal{S}_q(\delta^2)$.

Tables 4-6 report the empirical type I error of different tests for the AR(1) case with sample size $T = 200$. Other unspecified parameters such as τ and the number of simulation replications are the same as in Tables 1-3. As before, it is clear that the F^* test is more accurate in size than the corresponding χ^2 test. Among the three VAR F^* tests, the test based on the MTK has less size distortion than that based on AIC and BIC. This is especially true when the error is highly persistent. The VAR F^* test based on the MTK is as accurate in size as kernel F^* tests and nonstandard tests.

To sum up, the VAR F^* test has much smaller size distortion than the conventional χ^2 test, as considered by den Haan and Levin (1998). Compared to the kernel F^* tests and VAR and kernel nonstandard tests, it has either more accurate size or equally accurate size. The size accuracy of the VAR F^* test is achieved with no or small power loss.

7 Conclusions

The paper has established a new asymptotic theory for the covariance matrix estimator that is based on fitting a vector autoregressive model to the estimated moment process. The new asymptotic theory assumes that the VAR order is proportional to the sample size. Compared with the conventional asymptotics, the new asymptotic theory has two attractive properties: the limiting distribution reflects the VAR order used and the estimation uncertainty of model parameters. On the basis of this new asymptotic theory, we propose a new and easy-to-use F^* test. The test statistic is equal to a finite sample corrected Wald statistic and the critical values are from the standard F distribution.

The VAR F^* test is attractive in three respects. First, the underlying VAR covariance estimator is automatically positive semidefinite. There is no need to do any modification. Second, the VAR covariance estimator has smaller asymptotic bias than the kernel LRV estimators commonly used in practice. As shown in Sun, Phillips, and Jin (2008), bias reduction is especially important when we focus on size accuracy rather than the MSE of the covariance estimator. Third, the VAR F^* test is as easy to use as the conventional kernel-based and VAR-based chi-square tests but is much more accurate in size than the latter tests.

The new asymptotic theory can be extended to the autoregressive estimator of spectral densities at other frequencies. It is also straightforward to extend the overidentification test of Sun and Kim (2012) and the M tests of Kuan and Lee (2006) to the present setting. All the tests considered in the paper can be combined with prewhitening procedures such as those in Andrews and Monahan (1992) and Lee and Phillips (1994). The idea of the paper may be used to develop more accurate approximations in other problems that employ vector autoregression to model short run dynamics.

Table 1: Type I error of different tests for Gaussian location models with AR errors and $T = 100$: the number of restrictions $q = 1$

(ρ_1, ρ_2)	(-0.8,0)	(-0.4,0)	(0,0)	(0.4,0)	(0.8,0)	(1.5,-.75)	(.25,.25)	(.35,.35)
S^2-VAR-AIC	0.049	0.052	0.056	0.063	0.104	0.050	0.089	0.102
F*-VAR-AIC	0.051	0.053	0.058	0.065	0.106	0.051	0.090	0.104
χ^2-VAR-AIC	0.061	0.062	0.066	0.075	0.119	0.069	0.107	0.125
S^2-VAR-BIC	0.046	0.048	0.052	0.059	0.102	0.047	0.104	0.114
F*-VAR-BIC	0.048	0.050	0.055	0.061	0.105	0.048	0.107	0.117
χ^2-VAR-BIC	0.056	0.058	0.061	0.071	0.115	0.065	0.120	0.135
S^2-VAR-Par	0.049	0.052	0.055	0.056	0.089	0.050	0.064	0.082
F*-VAR-Par	**0.050**	**0.053**	**0.056**	**0.046**	**0.076**	**0.045**	**0.054**	**0.070**
χ^2-VAR-Par	0.066	0.063	0.072	0.122	0.175	0.090	0.133	0.167
S^2-VAR-QS	0.048	0.051	0.055	0.058	0.088	0.050	0.063	0.082
F*-VAR-QS	**0.050**	**0.053**	**0.056**	**0.055**	**0.075**	**0.046**	**0.057**	**0.070**
χ^2-VAR-QS	0.058	0.061	0.067	0.099	0.175	0.089	0.124	0.164
S^2-Parzen	0.048	0.050	0.057	0.062	0.084	0.031	0.072	0.086
F*-Parzen	0.046	0.049	0.056	0.062	0.087	0.033	0.074	0.087
χ^2-Parzen	0.066	0.066	0.070	0.097	0.173	0.081	0.128	0.169
S^2-QS	0.053	0.053	0.057	0.063	0.067	0.032	0.073	0.070
F*-QS	0.054	0.054	0.060	0.072	0.103	0.046	0.091	0.102
χ^2-QS	0.066	0.066	0.069	0.092	0.186	0.079	0.127	0.179

Note: For each smoothing parameter choice, there are three tests that use the same test statistic but critical values from different reference distributions. The S^2 test uses simulated critical values from the nonstandard fixed-smoothing asymptotics; the F^* test uses critical values from an F distribution; and the χ^2 test uses critical values from a chi-square distribution. The VAR orders are selected by AIC, BIC or MTK based on Parzen and QS kernels. The MTK rules are denoted as VAR-Par and VAR-QS, respectively. For the Parzen and QS kernels, the smoothing parameters are determined by a testing-optimal rule developed by Sun (2010). Results for our recommended F*-VAR-Par and F*-VAR-QS tests are in bold fonts.

Table 2: Type I error of different tests for Gaussian location models with AR errors and $T = 100$: the number of restrictions $q = 2$

(ρ_1, ρ_2)	(-0.8,0)	(-0.4,0)	(0,0)	(0.4,0)	(0.8,0)	(1.5,-.75)	(.25,.25)	(.35,.35)
S^2-VAR-AIC	0.040	0.043	0.051	0.067	0.151	0.053	0.120	0.146
F*-VAR-AIC	0.046	0.051	0.057	0.076	0.161	0.053	0.126	0.149
χ^2-VAR-AIC	0.062	0.069	0.077	0.097	0.184	0.089	0.166	0.200
S^2-VAR-BIC	0.038	0.041	0.049	0.064	0.148	0.050	0.169	0.199
F*-VAR-BIC	0.045	0.050	0.056	0.073	0.160	0.051	0.181	0.207
χ^2-VAR-BIC	0.059	0.065	0.073	0.093	0.181	0.085	0.213	0.249
S^2-VAR-Par	0.042	0.045	0.050	0.064	0.135	0.050	0.083	0.122
F*-VAR-Par	**0.046**	**0.051**	**0.051**	**0.034**	**0.087**	**0.043**	**0.048**	**0.079**
χ^2-VAR-Par	0.061	0.069	0.086	0.232	0.347	0.135	0.257	0.330
S^2-VAR-QS	0.041	0.045	0.050	0.063	0.135	0.050	0.083	0.122
F*-VAR-QS	**0.045**	**0.050**	**0.054**	**0.049**	**0.088**	**0.044**	**0.059**	**0.080**
χ^2-VAR-QS	0.060	0.068	0.078	0.168	0.346	0.128	0.229	0.326
S^2-Parzen	0.037	0.043	0.053	0.064	0.101	0.024	0.082	0.102
F*-Parzen	0.034	0.040	0.052	0.059	0.095	0.022	0.076	0.095
χ^2-Parzen	0.064	0.070	0.073	0.141	0.288	0.101	0.201	0.280
S^2-QS	0.045	0.045	0.054	0.067	0.087	0.031	0.090	0.091
F*-QS	0.049	0.048	0.057	0.080	0.146	0.041	0.115	0.142
χ^2-QS	0.069	0.070	0.071	0.135	0.361	0.097	0.207	0.340

Note: see note to table 1

Table 3: Type I error of different tests for Gaussian location models with AR errors and $T = 100$: the number of restrictions $q = 3$

(ρ_1, ρ_2)	(-0.8,0)	(-0.4,0)	(0,0)	(0.4,0)	(0.8,0)	(1.5,-.75)	(.25,.25)	(.35,.35)
S^2-VAR-AIC	0.037	0.044	0.053	0.079	0.224	0.058	0.166	0.206
F*-VAR-AIC	0.043	0.048	0.058	0.085	0.235	0.061	0.174	0.212
χ^2-VAR-AIC	0.067	0.071	0.085	0.117	0.279	0.118	0.241	0.298
S^2-VAR-BIC	0.037	0.042	0.052	0.078	0.224	0.056	0.239	0.349
F*-VAR-BIC	0.042	0.047	0.057	0.085	0.234	0.060	0.251	0.361
χ^2-VAR-BIC	0.065	0.070	0.084	0.115	0.276	0.116	0.301	0.418
S^2-VAR-Par	0.035	0.041	0.051	0.075	0.199	0.057	0.102	0.176
F*-VAR-Par	**0.043**	**0.048**	**0.054**	**0.022**	**0.090**	**0.048**	**0.037**	**0.079**
χ^2-VAR-Par	0.066	0.071	0.097	0.384	0.570	0.194	0.430	0.536
S^2-VAR-QS	0.035	0.041	0.049	0.074	0.199	0.059	0.110	0.177
F*-VAR-QS	**0.042**	**0.048**	**0.056**	**0.051**	**0.092**	**0.052**	**0.065**	**0.084**
χ^2-VAR-QS	0.065	0.070	0.087	0.272	0.567	0.181	0.356	0.530
S^2-Parzen	0.034	0.044	0.058	0.069	0.129	0.025	0.105	0.130
F*-Parzen	0.031	0.040	0.053	0.060	0.101	0.018	0.084	0.102
χ^2-Parzen	0.065	0.078	0.079	0.195	0.435	0.126	0.282	0.420
S^2-QS	0.045	0.046	0.060	0.072	0.116	0.029	0.119	0.130
F*-QS	0.048	0.050	0.057	0.091	0.211	0.041	0.137	0.202
χ^2-QS	0.079	0.076	0.077	0.181	0.564	0.126	0.294	0.523

Note: see note to table 1

Table 4: Type I error of different tests in a regression model with Gaussian AR(1) regressors and error and $T = 200$: the number of restrictions $q = 1$

ρ_1	-0.75	-0.5	-0.25	0	0.25	0.5	0.75	0.9
S^2-VAR-AIC	0.050	0.049	0.051	0.054	0.058	0.066	0.093	0.159
F*-VAR-AIC	0.050	0.049	0.051	0.054	0.058	0.067	0.093	0.159
χ^2-VAR-AIC	0.054	0.054	0.055	0.058	0.064	0.071	0.101	0.169
S^2-VAR-BIC	0.049	0.048	0.049	0.052	0.056	0.065	0.092	0.160
F*-VAR-BIC	0.049	0.048	0.049	0.052	0.056	0.065	0.092	0.160
χ^2-VAR-BIC	0.052	0.052	0.052	0.055	0.061	0.069	0.098	0.166
S^2-VAR-Par	0.045	0.046	0.048	0.051	0.052	0.058	0.078	0.126
F*-VAR-Par	**0.047**	**0.048**	**0.050**	**0.051**	**0.050**	**0.047**	**0.062**	**0.107**
χ^2-VAR-Par	0.055	0.056	0.057	0.059	0.081	0.109	0.152	0.222
S^2-VAR-QS	0.046	0.047	0.049	0.053	0.055	0.059	0.074	0.127
F*-VAR-QS	**0.048**	**0.049**	**0.051**	**0.052**	**0.056**	**0.056**	**0.063**	**0.108**
χ^2-VAR-QS	0.055	0.056	0.057	0.057	0.069	0.088	0.138	0.222
S^2-Parzen	0.044	0.045	0.047	0.053	0.062	0.066	0.078	0.127
F*-Parzen	0.044	0.046	0.048	0.053	0.063	0.068	0.079	0.128
χ^2-Parzen	0.056	0.058	0.058	0.061	0.074	0.090	0.130	0.221
S^2-QS	0.046	0.047	0.047	0.055	0.065	0.068	0.080	0.110
F*-QS	0.049	0.051	0.049	0.055	0.067	0.074	0.092	0.148
χ^2-QS	0.056	0.058	0.058	0.061	0.074	0.089	0.126	0.228

See note to Table 1

Table 5: Type I error of different tests in a regression model with Gaussian AR(1) regressors and error and $T = 200$: the number of restrictions $q = 2$

ρ_1	-0.75	-0.5	0.25	0	0.25	0.5	0.75	0.9
S^2-VAR-AIC	0.061	0.049	0.047	0.048	0.053	0.065	0.105	0.208
F*-VAR-AIC	0.068	0.056	0.053	0.054	0.061	0.074	0.114	0.221
χ^2-VAR-AIC	0.078	0.065	0.062	0.063	0.070	0.086	0.126	0.237
S^2-VAR-BIC	0.060	0.049	0.046	0.047	0.052	0.064	0.104	0.206
F*-VAR-BIC	0.066	0.055	0.053	0.053	0.060	0.073	0.113	0.219
χ^2-VAR-BIC	0.075	0.063	0.061	0.062	0.068	0.084	0.123	0.233
S^2-VAR-Par	0.059	0.050	0.046	0.048	0.052	0.061	0.087	0.173
F*-VAR-Par	**0.043**	**0.052**	**0.052**	**0.052**	**0.051**	**0.043**	**0.050**	**0.118**
χ^2-VAR-Par	0.171	0.082	0.062	0.066	0.106	0.178	0.291	0.405
S^2-VAR-QS	0.060	0.050	0.046	0.047	0.053	0.064	0.088	0.172
F*-VAR-QS	**0.058**	**0.056**	**0.054**	**0.053**	**0.058**	**0.062**	**0.064**	**0.118**
χ^2-VAR-QS	0.117	0.069	0.062	0.063	0.081	0.127	0.243	0.405
S^2-Parzen	0.069	0.054	0.050	0.052	0.063	0.072	0.084	0.149
F*-Parzen	0.070	0.056	0.052	0.054	0.065	0.073	0.086	0.145
χ^2-Parzen	0.107	0.067	0.062	0.064	0.082	0.115	0.197	0.357
S^2-QS	0.072	0.056	0.050	0.053	0.065	0.074	0.086	0.131
F*-QS	0.081	0.059	0.055	0.057	0.069	0.085	0.117	0.205
χ^2-QS	0.106	0.069	0.061	0.064	0.081	0.111	0.188	0.390

See note to Table 1

Table 6: Type I error of different tests in a regression model with Gaussian AR(1) regressors and error and $T = 200$: the number of restrictions $q = 3$

ρ_1	-0.75	-0.5	-0.25	0	0.25	0.5	0.75	0.9
S^2-VAR-AIC	0.090	0.061	0.055	0.054	0.060	0.078	0.133	0.275
F*-VAR-AIC	0.096	0.067	0.061	0.061	0.067	0.086	0.141	0.287
χ^2-VAR-AIC	0.110	0.082	0.073	0.075	0.080	0.102	0.160	0.311
S^2-VAR-BIC	0.088	0.060	0.054	0.054	0.059	0.078	0.132	0.273
F*-VAR-BIC	0.095	0.066	0.061	0.061	0.066	0.085	0.139	0.285
χ^2-VAR-BIC	0.108	0.080	0.072	0.075	0.079	0.100	0.157	0.308
S^2-VAR-Par	0.077	0.058	0.053	0.053	0.058	0.069	0.106	0.223
F*-VAR-Par	**0.028**	**0.054**	**0.059**	**0.060**	**0.056**	**0.040**	**0.032**	**0.093**
χ^2-VAR-Par	0.340	0.139	0.077	0.079	0.137	0.274	0.461	0.602
S^2-VAR-QS	0.074	0.061	0.054	0.054	0.060	0.073	0.109	0.222
F*-VAR-QS	**0.061**	**0.063**	**0.061**	**0.061**	**0.065**	**0.069**	**0.056**	**0.094**
χ^2-VAR-QS	0.202	0.098	0.073	0.076	0.096	0.169	0.374	0.602
S^2-Parzen	0.081	0.069	0.059	0.060	0.068	0.077	0.101	0.171
F*-Parzen	0.081	0.069	0.059	0.060	0.069	0.078	0.094	0.146
χ^2-Parzen	0.166	0.093	0.071	0.074	0.092	0.144	0.277	0.499
S^2-QS	0.085	0.073	0.060	0.061	0.072	0.081	0.102	0.148
F*-QS	0.102	0.076	0.062	0.064	0.076	0.094	0.140	0.262
χ^2-QS	0.159	0.093	0.071	0.074	0.091	0.137	0.261	0.569

See note to Table 1

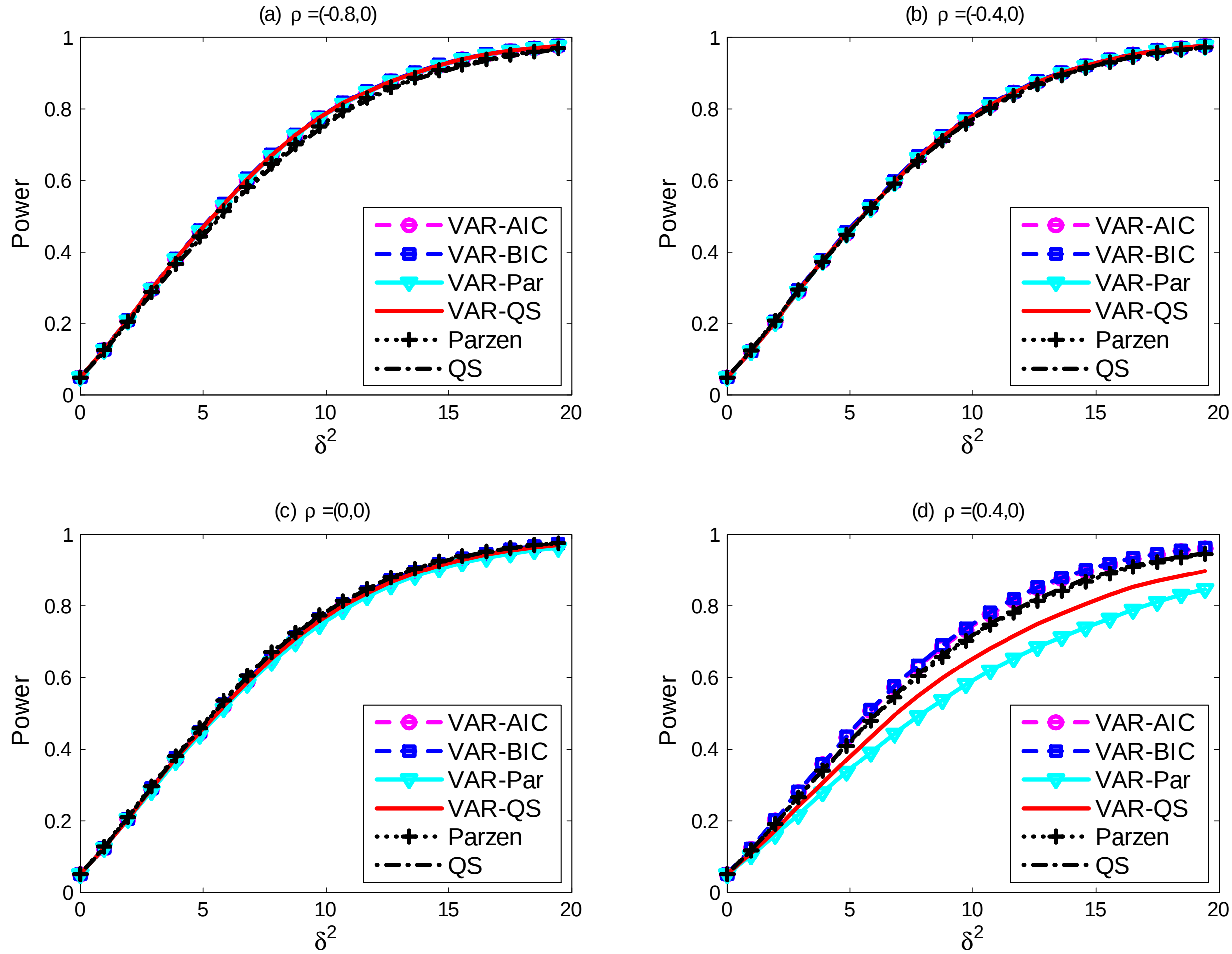

Figure 1: Size-adjusted power of the different F^* tests under the Gaussian location model with AR error, sample size $T = 100$ and number of restrictions $q = 2$.

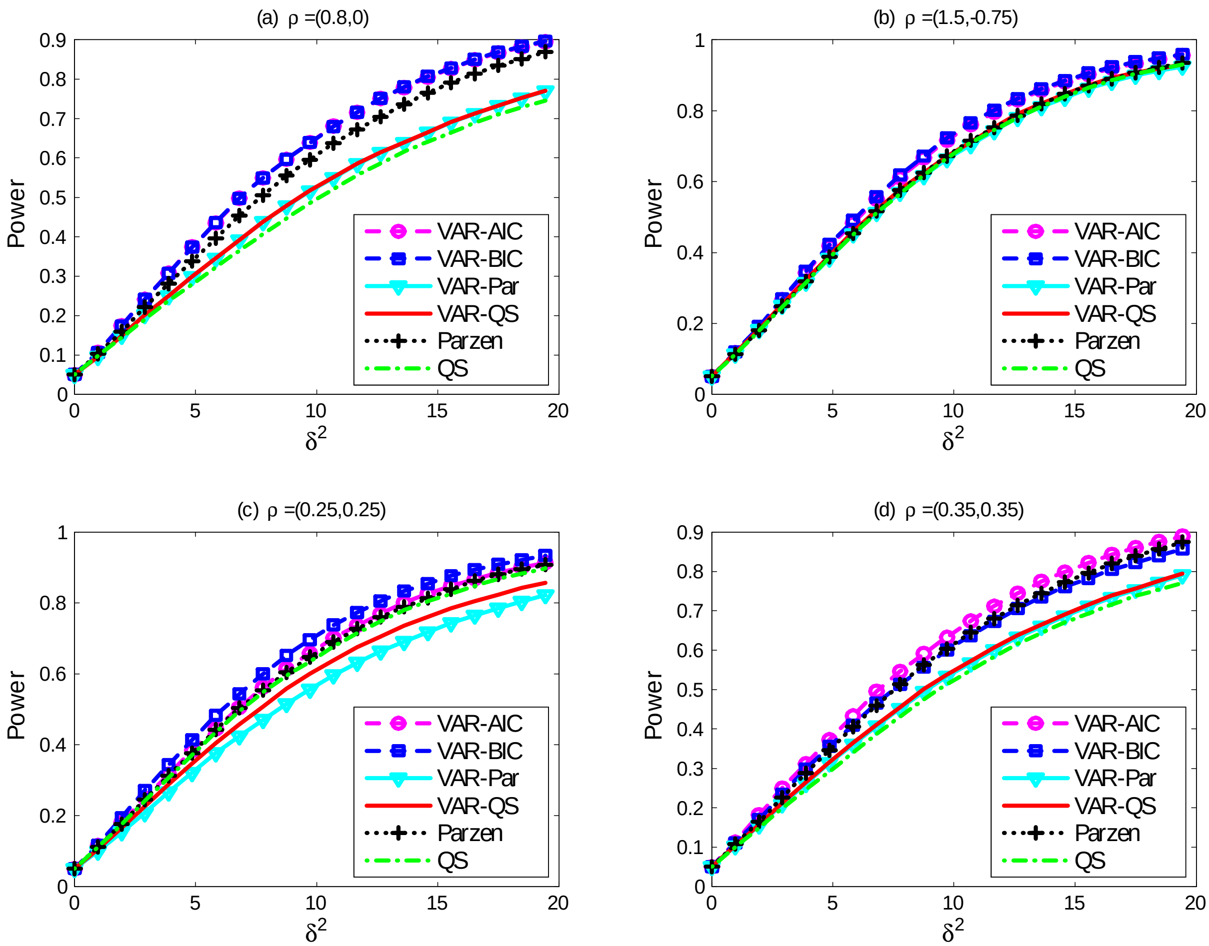

Figure 2: Size-adjusted power of the different F^* tests under the Gaussian location model with AR error, sample size $T = 100$ and number of restrictions $q = 2$.

8 Appendix: Proofs

Proof of Proposition 1. Note that the Yule-Walker estimators $\hat{A}_1, \ldots, \hat{A}_p$ and $\hat{\Sigma}_e$ satisfy (3) and (4). These two equations can be written as $\hat{B}\hat{\mathbf{\Gamma}}_H(p+1) = \hat{C}$, where

$$\hat{B} = \begin{pmatrix} \mathbb{I}_q, & -\hat{A}_1 & \ldots, & -\hat{A}_p \end{pmatrix}, \hat{C} = \begin{pmatrix} \hat{\Sigma}_e, & 0, & \ldots, & 0 \end{pmatrix}.$$

Let λ be an eigenvalue of $\hat{\mathbf{A}}'$ and $x = (x_1', \ldots, x_p')'$ be the corresponding eigenvector. Then

$$\begin{aligned} \hat{A}_1' x_1 + x_2 &= \lambda x_1, \\ \hat{A}_2' x_1 + x_3 &= \lambda x_2, \\ &\ldots \\ \hat{A}_{p-1}' x_1 + x_p &= \lambda x_{p-1}, \\ \hat{A}_p' x_1 &= \lambda x_p. \end{aligned}$$

From these equations, we know that $x \neq 0$ implies $x_1 \neq 0$. Writing these equations more compactly, we have

$$-\hat{B}' x_1 + \begin{pmatrix} x \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} 0 \\ x \end{pmatrix}. \tag{A.1}$$

We consider the case $\lambda \neq 0$. In this case, $\hat{B}' x_1 \neq 0$. It follows from (A.1) and the Toeplitz structure of $\hat{\mathbf{\Gamma}}_H(p+1)$ that

$$\begin{aligned} & x^* \hat{\mathbf{\Gamma}}_H(p)\, x \\ &= \begin{pmatrix} x \\ 0 \end{pmatrix}^* \hat{\mathbf{\Gamma}}_H(p+1) \begin{pmatrix} x \\ 0 \end{pmatrix} = \left[\hat{B}' x_1 + \lambda \begin{pmatrix} 0 \\ x \end{pmatrix}\right]^* \hat{\mathbf{\Gamma}}_H(p+1) \left[\hat{B}' x_1 + \lambda \begin{pmatrix} 0 \\ x \end{pmatrix}\right] \\ &= x_1^* \hat{B} \hat{\mathbf{\Gamma}}_H(p+1)\, \hat{B}' x_1 + \|\lambda\|^2 x^* \hat{\mathbf{\Gamma}}_H(p)\, x + \lambda x_1^* \hat{B} \hat{\mathbf{\Gamma}}_H(p+1) \begin{pmatrix} 0 \\ x \end{pmatrix} + \lambda^* \begin{pmatrix} 0 \\ x \end{pmatrix}^* \hat{\mathbf{\Gamma}}_H(p+1)\, \hat{B}' x_1 \\ &= x_1^* \hat{B} \hat{\mathbf{\Gamma}}_H(p+1)\, \hat{B}' x_1 + \|\lambda\|^2 x^* \hat{\mathbf{\Gamma}}_H(p)\, x + \lambda x_1^* \hat{C} \begin{pmatrix} 0 \\ x \end{pmatrix} + \lambda^* \begin{pmatrix} 0 \\ x \end{pmatrix}^* \hat{C}' x_1 \\ &= x_1^* \hat{B} \hat{\mathbf{\Gamma}}_H(p+1)\, \hat{B}' x_1 + \|\lambda\|^2 x^* \hat{\mathbf{\Gamma}}_H(p)\, x, \end{aligned}$$

where the last line follows because

$$\hat{C} \begin{pmatrix} 0 \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ x \end{pmatrix}^* \hat{C}' = 0.$$

So, we get

$$\|\lambda\|^2 = 1 - \frac{x_1^* \hat{B} \hat{\mathbf{\Gamma}}_H(p+1)\, \hat{B}' x_1}{x^* \hat{\mathbf{\Gamma}}_H(p)\, x}.$$

As a result, $\|\lambda\|^2 < 1$ almost surely if $\hat{\mathbf{\Gamma}}_H(p)$ and $\hat{\mathbf{\Gamma}}_H(p+1)$ are positive definite almost surely. ■

Proof of Lemma 1. Since the fitted VAR process is stationary almost surely, the long run variance

$$\hat{\mathcal{V}}_R = \left(\mathbb{I} - \hat{A}_1 - \ldots - \hat{A}_p\right)^{-1} \hat{\Sigma}_e \left(\mathbb{I} - \hat{A}_1' - \ldots - \hat{A}_p'\right)^{-1}$$

is well-defined almost surely. As a result,

$$\hat{\mathcal{V}}_R = \sum_{j=-p}^{p} \hat{\Gamma}_h(j) + \sum_{|j|>p} \tilde{\Gamma}_h(j) < \infty$$

almost surely. That is, $M_R = \sum_{|j|>p} \tilde{\Gamma}_h(j) = o(1)$ almost surely.

Define $S_t = \sum_{j=1}^{t} h_j$, $S_0 = 0$. It is easy to show that

$$\tilde{\mathcal{V}}_R = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} h_t h_\tau' k_{rect}\left(\frac{t-\tau}{bT}\right) = \frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=1}^{T} D(t,\tau) S_t S_\tau',$$

where

$$D(t,\tau) = k_{rect}(\frac{t-\tau}{bT}) - k_{rect}(\frac{t+1-\tau}{bT}) - k_{rect}(\frac{t-\tau-1}{bT}) + k_{rect}(\frac{t-\tau}{bT}).$$

To simplify the notation, we assume that bT is an integer and write $G_t = G_t(\theta_0)$ and $g_t = g_t(\theta_0)$. Note that $D(t,\tau) \neq 0$ if and only if $|t-\tau| = bT$ or $bT+1$. So

$$\begin{aligned}
\tilde{\mathcal{V}}_R &= T^{-1}\sum_{\tau=1}^{T-bT} S_{\tau+bT}S_\tau' - T^{-1}\sum_{\tau=1}^{T-bT-1} S_{\tau+bT+1}S_\tau', \\
&+ T^{-1}\sum_{t=1}^{T-bT} S_t S_{t+bT}' - T^{-1}\sum_{t=1}^{T-bT-1} S_t S_{t+bT+1}' \\
&= -T^{-1}\sum_{\tau=1}^{T-bT-1} h_{\tau+bT+1}S_\tau' - T^{-1}\sum_{t=1}^{T-bT-1} S_t h_{t+bT+1}'.
\end{aligned}$$

To establish the limiting distribution of $T^{-1}\sum_{\tau=1}^{T-bT-1} h_{\tau+bT+1}S_\tau'$, we write

$$\begin{aligned}
h_t &= \hat{R}_T\left(\hat{G}_T'\mathcal{W}_T\hat{G}_T\right)^{-1}\hat{G}_T'\mathcal{W}_T\left[f(v_t,\theta_0) + \frac{\partial f(v_t,\bar{\theta}_T)}{\partial\theta'}\left(\hat{\theta}_T - \theta_0\right)\right] \\
&= \hat{R}_T\left(\hat{G}_T'\mathcal{W}_T\hat{G}_T\right)^{-1}\hat{G}_T'\mathcal{W}_T\left[f(v_t,\theta_0) - \frac{\partial f(v_t,\bar{\theta}_T)}{\partial\theta'}\left\{\left(\tilde{G}_T'\mathcal{W}_T\tilde{G}_T\right)^{-1}G_T'\mathcal{W}_T g_T + o_p\left(\frac{1}{\sqrt{T}}\right)\right\}\right],
\end{aligned}$$

where $\tilde{G}_T = G_T\left(\tilde{\theta}_T\right)$ and $\bar{\theta}_T$, $\tilde{\theta}_T$ satisfy $\bar{\theta}_T = \theta_0 + O_p(1/\sqrt{T})$ and $\tilde{\theta}_T = \theta_0 + O_p(1/\sqrt{T})$. So

$$S_\tau = \hat{R}_T\left(\hat{G}_T'\mathcal{W}_T\hat{G}_T\right)^{-1}\hat{G}_T'\mathcal{W}_T\left[Tg_\tau - T\bar{G}_\tau\left\{\left(\tilde{G}_T'\mathcal{W}_T\tilde{G}_T\right)^{-1}G_T'\mathcal{W}_T g_T + o_p\left(\frac{1}{\sqrt{T}}\right)\right\}\right],$$

where $\bar{G}_\tau = G_T\left(\bar{\theta}_T\right)$. As a result,

$$T^{-1}\sum_{\tau=1}^{T-bT-1} h_{\tau+bT+1}S_\tau' = \hat{R}_T\left(\hat{G}_T'\mathcal{W}_T\hat{G}_T\right)^{-1}(I_1+I_2+I_3+I_4)\left(\hat{G}_T'\mathcal{W}_T\hat{G}_T\right)^{-1}\hat{R}_T' + o_p(1),$$

where

$$I_1 = \sum_{\tau=1}^{T-bT-1} \hat{G}_T' \mathcal{W}_T f\left(v_{\tau+bT+1}, \theta_0\right) g_\tau' \mathcal{W}_T \hat{G}_T,$$

$$I_2 = -\sum_{\tau=1}^{T-bT-1} \hat{G}_T' \mathcal{W}_T f\left(v_{\tau+bT+1}, \theta_0\right) g_T' \mathcal{W}_T G_T \left(\tilde{G}_T' \mathcal{W}_T \tilde{G}_T\right)^{-1} \bar{G}_\tau' \mathcal{W}_T \hat{G}_T,$$

$$I_3 = -\sum_{\tau=1}^{T-bT-1} \hat{G}_T' \mathcal{W}_T \frac{\partial f(v_{\tau+bT+1}, \bar{\theta}_T)}{\partial \theta'} \left(\tilde{G}_T' \mathcal{W}_T \tilde{G}_T\right)^{-1} (G_T' \mathcal{W}_T g_T)(g_\tau' \mathcal{W}_T \hat{G}_T),$$

$$I_4 = \sum_{\tau=1}^{T-bT-1} \hat{G}_T' \mathcal{W}_T \frac{\partial f(v_{\tau+bT+1}, \bar{\theta}_T)}{\partial \theta'} \left(\tilde{G}_T' \mathcal{W}_T \tilde{G}_T\right)^{-1} G_T' \mathcal{W}_T g_T g_T' \mathcal{W}_T G_T \left(\tilde{G}_T' \mathcal{W}_T \tilde{G}_T\right)^{-1} \bar{G}_\tau' \mathcal{W}_T \hat{G}_T.$$

We consider each of the above terms in turn. For I_1, we use Assumptions 4-5 to obtain

$$I_1 \Rightarrow G' \mathcal{W} \Lambda \int_0^{1-b} dW_m(b+r) W_m'(r) \Lambda' \mathcal{W} G.$$

For I_2, we have, by Assumptions 3-4:

$$\begin{aligned}
I_2 &= -\sum_{\tau=1}^{T-bT-1} G' \mathcal{W} f\left(v_{\tau+bT+1}, \theta_0\right) g_T' \mathcal{W} G \left(G' \mathcal{W} G\right)^{-1} \frac{\tau}{T} G \mathcal{W} G \left(1 + o_p(1)\right) \\
&= -\frac{1}{T} \sum_{\tau=1}^{T-bT-1} G' \mathcal{W} \tau f\left(v_{\tau+bT+1}, \theta_0\right) g_T' \mathcal{W} G \left(1 + o_p(1)\right) \\
&\Rightarrow -G' \mathcal{W} \Lambda \int_0^{1-b} dW_m(b+r) r W_m'(1) \Lambda' \mathcal{W} G.
\end{aligned}$$

For I_3 and I_4, we have

$$\begin{aligned}
I_3 &= -\sum_{\tau=1}^{T-bT-1} G' \mathcal{W} \frac{\partial f(v_{\tau+bT+1}, \bar{\theta}_T)}{\partial \theta'} \left(G' \mathcal{W} G\right)^{-1} (G' \mathcal{W} g_T)(g_\tau' \mathcal{W} G) \left(1 + o_p(1)\right) \\
&= -T \sum_{\tau=1}^{T-bT-1} G' \mathcal{W} \left[\bar{G}_{\tau+bT+1} - \bar{G}_{\tau+bT}\right] \left(G' \mathcal{W} G\right)^{-1} (G' \mathcal{W} g_T)(g_\tau' \mathcal{W} G) \left(1 + o_p(1)\right) \\
&= -\sum_{\tau=1}^{T-bT-1} (G' \mathcal{W} g_T)(g_\tau' \mathcal{W} G) \left(1 + o_p(1)\right) \\
&\Rightarrow -G' \mathcal{W} \Lambda \left[W_m(1) \int_0^{1-b} W_m'(r) dr\right] \Lambda' \mathcal{W} G,
\end{aligned}$$

and

$$
\begin{aligned}
I_4 &= \sum_{\tau=1}^{T-bT-1} G'\mathcal{W}\frac{\partial f(v_{\tau+bT+1},\bar{\theta}_T)}{\partial\theta'}\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}g_T g_T'\mathcal{W}G\left(G'\mathcal{W}G\right)^{-1}\bar{G}_\tau'\mathcal{W}G\left(1+o_p\left(1\right)\right)\\
&= T\sum_{\tau=1}^{T-bT-1} G'\mathcal{W}\left(\bar{G}_{\tau+bT+1}-\bar{G}_{\tau+bT}\right)\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}g_T g_T'\mathcal{W}G\frac{\tau}{T}\left(1+o_p\left(1\right)\right)\\
&= \sum_{\tau=1}^{T-bT-1}\frac{\tau}{T}G'\mathcal{W}g_T g_T'\mathcal{W}_T G\left(1+o_p\left(1\right)\right) \Rightarrow \left(\int_0^{1-b} rdr\right)G'\mathcal{W}\Lambda W_m(1)W_m'(1)\Lambda'\mathcal{W}G\\
&= \frac{1}{2}\left(b-1\right)^2 G'\mathcal{W}\Lambda W_m(1)W_m'(1)\Lambda'\mathcal{W}G.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&I_1+I_2+I_3+I_4\\
&\Rightarrow G'\mathcal{W}\Lambda\left[\int_0^{1-b} dW_m(b+r)W_m'(r)-\int_0^{1-b} dW_m(b+r)rW_m'(1)\right.\\
&\left.-\int_0^{1-b} W_m(1)W_m'(r)dr+\frac{1}{2}\left(b-1\right)^2 W_m(1)W_m'(1)\right]\Lambda'\mathcal{W}G
\end{aligned}
\tag{A.2}
$$

$$
\begin{aligned}
&= G'\mathcal{W}\Lambda\left[\int_0^{1-b} dW_m(b+r)V_m'\left(r\right)\right]\Lambda'\mathcal{W}G\\
&- G'\mathcal{W}\Lambda\left[\int_0^{1-b} W_m(1)W_m'(r)dr-\frac{1}{2}\left(b-1\right)^2 W_m(1)W_m'(1)\right]\Lambda'\mathcal{W}G\\
&= G'\mathcal{W}\Lambda\left\{\left[\int_0^{1-b} dW_m(b+r)V_m'\left(r\right)\right]-\int_0^{1-b} drW_m(1)V_m'\left(r\right)\right\}\Lambda'\mathcal{W}G\\
&= G'\mathcal{W}\Lambda\left(\int_0^{1-b} dV_m(b+r)V_m'\left(r\right)\right)\Lambda'\mathcal{W}G.
\end{aligned}
$$

Combining the above results yields $\tilde{\mathcal{V}}_R \Rightarrow \mathcal{V}_{R,\infty}$. Since $\hat{\mathcal{V}}_R = \tilde{\mathcal{V}}_R + o_p\left(1\right)$, we have $\hat{\mathcal{V}}_R \Rightarrow \mathcal{V}_{R,\infty}$ as stated. ■

Proof of Lemma 2. (a) It follows from equation (A.2) that

$$
\begin{aligned}
&-\int_0^{1-b} dV_m(b+r)V_m'\left(r\right)\\
&= -\int_0^{1-b} dW_m(b+r)W_m'(r)+\int_0^{1-b} dW_m(b+r)rW_m'(1)\\
&+\int_0^{1-b} W_m(1)W_m'(r)dr-\frac{1}{2}\left(b-1\right)^2 W_m(1)W_m'(1)\\
&= -\int_b^1\left[dW_m(s)\int_0^{s-b} dW_m'\left(r\right)\right]+\int_0^1\left[\int_b^1\left(s-b\right)dW_m(s)\right]dW_m'\left(r\right)\\
&+\int_0^{1-b} W_m(1)W_m'(r)dr-\int_0^1\int_0^1\frac{1}{2}\left(b-1\right)^2 dW_m(s)dW_m'\left(r\right).
\end{aligned}
$$

But using integration by parts we have

$$\int_0^{1-b} W_m(1)W_m'(r)dr = \int_0^{1-b} (1-b-r)\left[\int_0^1 dW_m(s)\right] dW_m'(r),$$

so

$$\begin{aligned}
&-\int_0^{1-b} dV_m(b+r)V_m'(r) \\
&= -\int_b^1 \left[dW_m(s)\int_0^{s-b} dW_m'(r)\right] + \int_0^1 \left[\int_b^1 (s-b)\, dW_m(s)\right] dW_m'(r) \\
&- \int_0^{1-b} (b+r-1)\left[\int_0^1 dW_m(s)\right] dW_m'(r) - \int_0^1\int_0^1 \frac{1}{2}(b-1)^2 dW_m(s)dW_m'(r)
\end{aligned}$$

or

$$\begin{aligned}
&-\int_0^{1-b} dV_m(b+r)V_m'(r) \\
&= \int_0^1\int_0^1 -\mathbf{1}\{r\in[0,s-b], s\in[b,1]\} + \mathbf{1}\{r\in[0,1], s\in[b,1]\}(s-b) \\
&+ (1-b-r)\,\mathbf{1}\{r\in[0,1-b], s\in[0,1]\} - \frac{1}{2}(b-1)^2\,\mathbf{1}\{r\in[0,1], s\in[0,1]\}\, dW_m(s)dW_m'(r) \\
&= \int_0^1\int_0^1 \tilde{k}_b(r,s)dW_m(s)dW_m'(r),
\end{aligned}$$

and

$$\tilde{k}_b(r,s) = -\frac{1}{2}(b-1)^2 + \begin{cases} 1-b-r, & \text{if } r\in[0,1-b], s\in[0,b] \\ s-b, & \text{if } r\in[1-b,1], s\in(b,1] \\ s-r-2b, & \text{if } r\in(0,s-b), s\in(b,1] \\ s-r-2b+1, & \text{if } r\in[s-b,1-b), s\in(b,1] \\ 0, & \text{if } r\in(1-b,1], s\in(0,b] \end{cases}$$

For the second term in $Q_m(b)$, we note that

$$\begin{aligned}
&-\int_0^{1-b} V_m(r)dV_m'(r+b) = \left(-\int_0^{1-b} dV_m(b+r)V_m'(r)\right)' \\
&= \int_0^1\int_0^1 \tilde{k}_b(r,s)dW_m(r)dW_m'(s) = \int_0^1\int_0^1 \tilde{k}_b(s,r)dW_m(s)dW_m'(r).
\end{aligned}$$

Therefore $Q_m(b) = \int_0^1\int_0^1 k_b^*(r,s)dW_m(s)dW_m'(r)$, where $k_b^*(r,s) = \tilde{k}_b(r,s) + \tilde{k}_b(s,r)$. Some algebra shows that $k_b^*(r,s)$ can be simplified to the expression given in (10).

(b) Note that $E\Lambda Q_m(b)\Lambda' = \left(\int_0^1 k_b^*(r,r)dr\right)\Omega$. It is easy to show that $\int_0^1 k_b^*(r,r)dr = (1-b)^2$. Hence $E\Lambda Q_m(b)\Lambda' = \mu_1\Omega$.

Let

$$A = \int_0^1\int_0^1 k_b^*(r,s)dW_m(s)dW_m'(r),$$

then

$$vec\left(\Lambda Q_m(b)\Lambda'\right) = vec\left[\Lambda A\Lambda'\right] = (\Lambda\otimes\Lambda)\left[vec(A)\right].$$

To compute var$(vec\,(\Lambda Q_m\,(b)\,\Lambda'))$, it is sufficient to compute var$(vec\,(A))$:

$$var\,(vec\,(A)) = var\left[\int_0^1\int_0^1 k_b^*(r,s)vec\left[dW_m(s)dW_m'\,(r)\right]\right]$$
$$= var\left[\int_0^1\int_0^1 k_b^*\,(r,s)\,[dW_m(r)\otimes dW_m(s)]\right].$$

But

$$var\left[\int_0^1\int_0^1 k_b^*(r,s)\,[dW_m(r)\otimes dW_m(s)]\right]$$
$$= \left(\int_0^1\int_0^1 [k_b^*(r,s)]^2\,drds\right)(\mathbb{I}_{m^2}+\mathbb{K}_{m^2})\,.$$

Consequently,

$$var\left(vec\left(\Lambda Q_m\,(b)\,\Lambda'\right)\right) = \mu_2\,(\Omega\otimes\Omega)\,(\mathbb{I}_{m^2}+\mathbb{K}_{m^2})\,,$$

where $\mu_2 = \int_0^1\int_0^1 [k_b^*(t,\tau)]^2\,dtd\tau$. The expression of μ_2 in the lemma can be obtained by elementary albeit lengthy calculations. To save space, we do not present the details here but they are available upon request. As a result

$$\text{var}(\text{vec}(\mathcal{V}_{R,\infty})) = \mu_2\,(\mathcal{V}_R\otimes\mathcal{V}_R)\left(\mathbb{I}_{q^2}+\mathbb{K}_{q^2}\right).$$

(c) Part (c) follows directly from part (b). Details are omitted here. ■

Proof of Theorem 2. Note that

$$\hat{\mathcal{V}}_R \Rightarrow \mathcal{V}_{R,\infty} := R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda Q_m(b)\Lambda'\mathcal{W}G\left(G'\mathcal{W}G\right)^{-1}R',$$
$$\sqrt{T}r(\hat{\theta}_T) \Rightarrow -R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda W_m\,(1)\,.$$

In addition, it is easy to show that $Q_m(b)$ is positive definite with probability one for any $b\in(0,1)$. So by the continuous mapping theorem, we have

$$F_T \Rightarrow \left[R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda W_m\,(1)\right]'$$
$$\times\left[R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda\int_0^1\int_0^1 k_b^*(r,s)dW_m(s)dW_m'\,(r)\,\Lambda'\mathcal{W}G\left(G'\mathcal{W}G\right)^{-1}R'\right]^{-1}$$
$$\times\left[R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda W_m\,(1)\right]/q.$$

Let

$$R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda W_m\,(r) \overset{d}{=} DW_q(r)$$

for a $q\times q$ matrix D such that

$$DD' = R\left(G'\mathcal{W}G\right)^{-1}G'\mathcal{W}\Lambda\Lambda'\mathcal{W}G\left(G'\mathcal{W}G\right)^{-1}R'.$$

Then

$$F_T \Rightarrow [DW_q(1)]'\left[D\int_0^1\int_0^1 k_b^*(r,s)dW_q(s)dW_q'(r)D'\right]^{-1}DW_q(1)/q$$
$$\overset{d}{=} W_q'\,(1)\,[Q_q\,(b)]^{-1}\,W_q(1)/q$$

as desired. ■

References

[1] Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–854.

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

[3] Brockwell, P.J. and Davis R.A. (1991), *Time Series: Theory and Methods*, Second Edition. Springer, New York.

[4] Bartlett, M.S. (1937), "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society* A, 160, 268–282.

[5] Bartlett, M.S. (1954), "A Note on the Multiplying Factors for Various χ^2 Approximations," *Journal of the Royal Statistical Society* B, 16, 296–298.

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

[7] Burg, J.P. (1967), "Maximum Entropy Spectral Analysis," Proceedings of the 37th Meeting of the Society of Exploration Geophysicists; Reprinted in D.G. Childers, ed. (1978), *Modern Spectrum Analysis*, IEEE Press, pp. 34–41.

[8] Cover, T. and Thomas, J. (1991), *Elements of Information Theory*, John Wiley and Sons, Inc.

[9] den Haan, W.J. and Levin A. (1997), "A Practitioner's Guide to Robust Covariance Matrix Estimation," In *Handbook of Statistics* 15, G.S. Maddala and C.R. Rao, eds., Elsevier, Amsterdam, 299–342.

[10] ——— (1998), "Vector Autoregressive Covariance Matrix Estimation," http://www.wouterdenhaan.com/papers.htm, working paper.

[11] de Jong, R.M. and Davidson J. (2000), "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes," *Econometric Theory*, 16(5), 621–642.

[12] Hansen, L.P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

[13] Hashimzade, N. and T. J. Vogelsang (2007): "Fixed-b Asymptotic Approximation of the Sampling Behavior of Nonparametric Spectral Density Estimators." *Journal of Time Series Analysis* 29, 142–162.

[14] Jun, B. (2011), "Size Distortion of the t-test Using an Autoregressive Spectral Variance Estimator," Working paper, Korea Institute of Public Finance.

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

[16] ——— (2002b), "Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation," *Econometrica*, 70, 2093–2095.

[17] ——— (2005), "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory,* 21, 1130–1164.

[18] Kuan, C.-M. and Lee, W.-M (2006), "Robust M Tests without Consistent Estimation of Asymptotic Covariance Matrix," *Journal of the American Statistical Association*, 101, 1264–1275.

[19] Lee, C. C. and Peter C. B. Phillips (1994), "An ARMA-prewhitened Long Run Variance Estimator," Yale University, mimeographed (korora.econ.yale.edu/phillips/papers/prewhite.pdf).

[20] Lütkepohl, H. (2007), *New Introduction to Multiple Time Series Analysis.* Springer, New York.

[21] Lin, C.-C. and Sakata, S. (2009), "On Long-Run Covariance Matrix Estimation with the Truncated Flat Kernel," Working paper, http://idv.sinica.edu.tw/ever23/Flat-top/lin-sakata.pdf, working paper.

[22] Newey, W.K. and West, K.D. (1987), "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

[23] Parzen, E. (1983), "Autoregressive Spectral Estimation," in *Handbook of Statistics* 3, D.R. Brillinger and P.R. Krishnaiah, eds., Elsevier Press, 221–247.

[24] Politis, D.N. (2001), "On Nonparametric Function Estimation with Infinite-order Flat-top Kernels," In *Probability and Statistical Models with Applications.* Ch. A. Charalambides et al. (eds.), Chapman and Hall/CRC, Boca Raton, 469–483.

[25] ——— (2011), "Higher-order Accurate, Positive Semi-definite Estimation of Large-sample Covariance and Spectral Density Matrices," *Econometric Theory*, 27, 703-744.

[26] Priestley, M.B. (1981), *Spectral Analysis and Time Series*, Academic Press, London and New York.

[27] Stoica, P. and Moses, R. (2005), *Spectral Analysis of Signals*, Pearson Prentice Hall.

[28] Sun, Y., Phillips, P.C.B. and Jin, S. (2008), "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76, 175–194.

[29] Sun, Y. (2010), "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference," http://www.econ.ucsd.edu/~yisun/HAR_paper.pdf, working paper, Department of Economics, UC San Diego.

[30] Sun, Y. and Kim, M.S. (2012), "Simple and Powerful GMM Over-identification Tests with Accurate Size," *Journal of Econometrics* 166, 267–281.

[31] White, H. (2001), *Asymptotic Theory for Econometricians*, revised edition. Academic Press, San Diego.

[32] Whittle, P. (1954), "The Statistical Analysis of a Seiche Record," *Journal of Marine Research,* 13, 76–100.