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

Yixiao Sun and David M. Kaplan*

Department of Economics, University of California, San Diego

First version: July 30, 2010. This version: Sept. 21, 2010

Abstract

In this paper, we develop a new asymptotic theory of the long run variance estimator obtained by fitting a vector autoregressive model to the transformed moment processes in a GMM framework. In contrast to the conventional asymptotics where the VAR lag order $p$ goes to infinity but at a slower rate than the sample size, we assume that $p$ grows at the same rate as the sample size. Under this asymptotic specification, the long run variance estimator is not consistent, but the associated Wald statistic and $t$-statistic remain asymptotically pivotal. On the basis of the new asymptotic theory, 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 long run variance estimation and kernel long run variance estimation. Simulations show that the new VAR $F^*$ test with the empirical order selection is more accurate in size than the conventional chi-square test and kernel-based $F^*$ tests with no or minimal power loss. The paper complements the recent paper by Sun (2010d) who considers kernel-based $F^*$ tests.

JEL Classification: C13; C14; C32; C51

Keywords: F-distribution, Flat-top Kernel, Generalized Method of Moments, Heteroscedasticity and Autocorrelation Robust Test, Hotelling’s T-square Distribution, Long-run Variance, Rectangular Kernel, 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 parameter inference in GMM models with time series data. To avoid possible mis-specification and to be completely general, we often do not parametrize the dependent structure of the moment conditions. The problem is how to nonparametrically estimate the long run variance (LRV) matrix of these moment conditions. The recent literature on heteroscedasticity and autocorrelation robust (HAR) estimators has mainly focused on the kernel-based method, although quite different approaches like the vector autoregressive (VAR) approach (see, for example, Berk, 1974, Parzen, 1983, den Haan and Levin, 1998) and series type approach (Phillips, 2005, Sun, 2010a,b,c) have been explored. Under fairly general conditions, den Haan and Levin (1997, 1998) show that the vector autoregressive HAR variance estimator converges at a faster rate than any positive semi-definite kernel HAR variance estimator. 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 variance 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 variance estimator. The consistency result requires that the VAR order $p$ increases with the sample size $T$ but at a slower rate. While appealing in practical situations, the consistency result does not capture the sampling variation of the variance estimator in finite samples. In addition, the consistency result also completely ignores the estimation uncertainty of the model parameters. In this paper, we develop a new asymptotic theory that avoids these two 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 variance estimator is inconsistent but 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 long run variance. 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 variance estimator and is likely to provide 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 has conditional population autocovariances (conditional on the estimated model parameters) that are identical to the empirical autocovariances up to order $p$. This crucial observation drives all of our asymptotic development. Given this ‘reproducing’ property of the Yule-Walker estimator, we know that the VAR variance estimator is asymptotically equivalent to the kernel LRV 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) and Sun, Phillips and Jin (2008). The rectangular kernel is not continuous and has not been considered in the literature on the fixed-$b$ asymptotics. One of the contributions of this paper is to fill in this gap. As a corollary, we obtain the new asymptotic theory for the VAR variance estimator.

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 necessarily 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. Very often, the selected VAR order is larger for a larger sample size but is still small relative to the sample size. So the empirical situations appear to be more compatible with the conventional asymptotics. Fortunately, we can show that the two types of asymptotics coincide as \( b \to 0 \). In other words, the fixed-smoothing asymptotics is asymptotically valid under the conventional thought experiment.

A further contribution of the paper is to provide an approximation to the nonstandard fixed-smoothing asymptotic distribution of the Wald statistic. We show that the fixed-smoothing asymptotic distribution of the VAR variance estimator is equal to a weighted sum of independent Wishart distributions. In addition, the VAR variance estimator is asymptotically independent of the model parameter estimator. Motivated from the early statistics literature on spectral density estimation, we approximate the weighted sum of Wishart distributions by a single Wishart distribution with equivalent degree of freedom. A direct implication is that the limiting distribution of the Wald statistic is approximately equal to a quadratic form in standard normal variates with an independent and Wishart-distributed weighting matrix. But such a quadratic form is exactly the same as Hotelling’s (1931) \( T^2 \) distribution. Using the well-known relationship between the \( T^2 \) distribution and the \( F \) distribution, we show that, after some modification, the nonstandard fixed-smoothing limiting distribution can be approximated by a standard \( F \) distribution.

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 \( \exp (-2bq) \) or an asymptotically equivalent factor, where \( q \) is the number of restrictions being jointly tested. The finite sample correction can be regarded as an example of the Bartlett or Bartlett-type correction. See Bartlett (1937, 1954). It corrects for the demeaning bias of the VAR variance estimator, which is due to the estimation uncertainty of model parameters, and the randomness of the scaling matrix in the Wald statistic. The critical value used in the \( F^* \) test comes from the \( F \) distribution with degree of freedom \( q \) and \( K = \max(5, [1/ (2b)]) \). Compared with the standard \( \chi^2 \) critical values, the \( F \) critical values capture the randomness of the LRV estimator. 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 autoregressive 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 autoregressive 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 autoregressive 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 (2010, Theorem 2.1) establishes the rate of convergence of flat-top kernel estimators, while den Haan and Levin (1998, Theorem 1) gives the rate for the vector autoregressive estimator. See Lee and Phillips (1994) for related results on the ARMA prewhitened estimator. Careful inspection shows that the rates in
Politis (2010) 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.

Compared with a finite-order kernel estimator, the VAR variance estimator enjoys the bias reducing property same as any infinite-order flat-top kernel estimator. The small bias, coupling with the new asymptotic theory that captures the randomness the VAR variance estimator, may give the proposed \( F^* \) test some 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 often more accurate than, the kernel-based \( F^* \) tests recently proposed by Sun (2010d). 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 the \( \chi^2 \) test.

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 variance estimator. The next two sections are devoted to the fixed-smoothing asymptotics of the VAR variance 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 and some technical results are given in the appendix.

## 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 \( \theta_0 \) be the true value and assume that \( \theta_0 \) is an interior point of the compact parameter space \( \Theta \). The moment conditions

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

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 \( \theta_0 \) is then given by

\[
\hat{\theta}_T = \arg \min_{\theta \in \Theta} g_T^T(\theta) \mathcal{W}_T(\theta) g_T(\theta),
\]

where \( \mathcal{W}_T \) is an \( m \times m \) positive definite weighting matrix.

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

\[
\hat{\theta}_T - \theta_0 = - \left[ G_T(\hat{\theta}_T)^T \mathcal{W}_T G_T(\hat{\theta}_T) \right]^{-1} G_T(\theta_0)^T \mathcal{W}_T g_T(\theta_0),
\]

3
where $\hat{\theta}_T$ is a value between $\hat{\theta}_T$ and $\theta_0$. If \( \text{plim}_{T \to \infty} G_T(\hat{\theta}_T) = G \), \( \text{plim}_{T \to \infty} \mathcal{W}_T = \mathcal{W} \) and \( \sqrt{T} g_T(\theta_0) \Rightarrow N(0, \Omega) \), then

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

(1)

where \( \mathcal{V} = (G' \mathcal{W} G)^{-1} (G' \mathcal{W} \Omega \mathcal{W} G) (G' \mathcal{W} G)^{-1} \). The above asymptotic result provides the basis for inference on $\theta_0$.

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

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

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

\[
t_T = \frac{\sqrt{T} r(\hat{\theta}_T)}{\sqrt{\hat{V}_R}}.
\]

It follows from (1) that $\mathcal{V}_R = R V R'$. To make inference on $\theta_0$, we have to estimate the unknown quantities in $\mathcal{V}$. $\mathcal{W}$ and $G$ can be consistently estimated by their finite sample versions $\hat{\mathcal{W}}_T$ and $\hat{G}_T = G_T(\hat{\theta}_T)$, respectively. It remains to estimate $\Omega$. Let $\hat{\Omega}_T$ be an estimator of $\Omega$. Then $\mathcal{V}_R$ can be estimated by

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

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

Many nonparametric estimators of $\Omega$ 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, 1994) 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 autoregressive spectral estimators.

Let

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

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 variance estimation below.

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

\[
h_t = \hat{A}_1 h_{t-1} + \ldots + \hat{A}_p h_{t-p} + \hat{\epsilon}_t,
\]

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

\[
\hat{A} = (\hat{A}_1, \ldots, \hat{A}_p) = (\hat{\Gamma}_h (1), \ldots, \hat{\Gamma}_h (p)) \hat{\Sigma}_H^{-1} (p),
\]
where
\[
\hat{\Gamma}_h(j) = \begin{cases} 
T^{-1}\sum_{t=j+1}^{T} h_th'_{t-j}, & j \geq 0 \\
T^{-1}\sum_{t=1}^{T+j} h_th'_{t-j}, & j < 0 
\end{cases}
\]
and
\[
\hat{\Gamma}_h(p) = \begin{bmatrix} 
\hat{\Gamma}_h(0) & \ldots & \hat{\Gamma}_h(p-1) \\
\vdots & \ddots & \vdots \\
\hat{\Gamma}_h(-p+1) & \ldots & \hat{\Gamma}_h(0) 
\end{bmatrix}_{pq \times pq}
\]

2. Compute
\[
\hat{\Sigma}_e = \hat{\Gamma}_h(0) - \hat{A}_1\hat{\Gamma}_h(1) - \ldots - \hat{A}_p\hat{\Gamma}_h(p)
\]
and estimate \( \tilde{V}_R \) by
\[
\tilde{V}_R = \left( \mathbb{I}_q - \hat{A}_1 - \ldots - \hat{A}_p \right)^{-1}\hat{\Sigma}_e \left( \mathbb{I}_q - \hat{A}_1' - \ldots - \hat{A}_p' \right)^{-1}.
\]

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 \( f(v_t, \theta_T) \). There are several advantages. First, the dimension of \( h_t \) can be much smaller than the dimension of \( f(v_t, \theta_T) \), especially when there are many moment conditions. So the VAR(\( p \)) model for \( h_t \) may have substantially fewer parameter than the VAR model for \( f(v_t, \theta_T) \). Second, by construction \( \mathbb{P}Tt=1h_t=0 \); so an intercept vector is not needed in the VAR for \( h_t \). On the other hand, when the model is overidentified, \( \sum_{t=1}^{T} f(v_t, \theta_T) \neq 0 \). Hence, a VAR model for \( f(v_t, \theta_T) \) 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, autoregression fitting on the basis of \( f(v_t, \theta_T) \) completely ignores the null hypothesis, and the resulting variance estimator \( \tilde{V}_R \) may be poor in finite samples.

Let
\[
\hat{A} = \begin{bmatrix} 
\hat{A}_1 & \ldots & \hat{A}_{p-1} & \hat{A}_p \\
\mathbb{I}_q & \ldots & 0 & 0 \\
\vdots & \ddots & \vdots & \vdots \\
0 & \ldots & \mathbb{I}_q & 0 
\end{bmatrix}_{pq \times pq}
\]
and
\[
\hat{\Sigma}_E = \begin{bmatrix} 
\hat{\Sigma}_e & \ldots & 0 & 0 \\
0 & \ldots & 0 & \vdots \\
\vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 
\end{bmatrix}_{pq \times pq}
\]

then the Yule-Walker estimators \( \hat{A} \) and \( \hat{\Sigma}_E \) satisfy:
\[
\hat{\Gamma}_h(p) = \hat{A}\hat{\Gamma}_h(p)\hat{A}' + \hat{\Sigma}_E.
\] (3)

It is well known that the estimated VAR model obtained via the Yule-Walker method is stationary almost surely. See Brockwell and Davis (1987, ch 8.1), Lütkepohl (2007, ch 3.3.4), and Reinsel (1993, ch 4.4). These books discuss either the scalar case or the multivariate case but without giving a proof. To the best of author’s knowledge, a rigorous proof for the multivariate case is currently lacking in the literature. We collect the stationarity result in the proposition below and provide a simple proof in the appendix.

**Proposition 1** If \( \hat{\Gamma}_h(p) \) and \( \hat{\Gamma}_h(p+1) \) are positive definite Toeplitz matrices, then
\[
\|\lambda_{\hat{A}}\| < 1,
\]
where \( \lambda_{\hat{A}} \) is any eigenvalue of \( \hat{A} \).
Proposition 1 gives precise conditions under which the fitted VAR($p$) process is stationary. For the OLS estimator, the $X'X$ matrix is not a Toeplitz matrix and the fitted VAR($p$) model may not be stationary.

3 Fixed-Smoothing Asymptotics for the Variance Estimator

In this section, we derive the asymptotic distribution of $\hat{\Sigma}_R$. 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. In this case, $\hat{\Sigma}_R$ is asymptotically equivalent to a kernel smoothing estimator where the smoothing is over a fixed number of quantities, even in large samples. On the other hand, if $b \to 0$ at the rate given in den Haan and Levin (1998), we obtain the conventional asymptotics under which $\hat{\Sigma}_R$ is consistent and $F_T$ is asymptotically $\chi^2_q$ distributed. In this case, the smoothing is taken over increasingly many quantities and the asymptotics may be referred to as the 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 $b \to 0$:

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 a theoretical autocovariance that is identical to the sample autocovariance up to lag $p$:

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

where $\hat{e}_t \sim iid(0, \Sigma_e)$. Define

$$\Gamma_H (p) = \begin{bmatrix} \Gamma_h (0) & \ldots & \Gamma_h (p - 1) \\ \vdots & \ddots & \vdots \\ \Gamma_h (-p + 1) & \ldots & \Gamma_h (0) \end{bmatrix}_{pq \times pq},$$

where $\Gamma_h (j) = E \tilde{h}_t \tilde{h}_{t-j}$. Then the autocovariance sequence satisfies

$$\Gamma_H (p) = A \Gamma_H (p) A' + \Sigma_E,$$

(4)

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

$$vec [\Gamma_H (p)] = [I_{p^2 q^2} - (A \otimes A)]^{-1} vec (\Sigma_E).$$

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

$$\Gamma_h (j) \equiv \Gamma_{h,j} (A_1, \ldots, A_p, \Sigma_e), \quad j = 0, 1, \ldots, p.$$  

(5)

By the definition of the Yule-Walker estimator, $\hat{A}$ and $\hat{\Sigma}_E$ satisfy $\hat{\Gamma}_H (p) = \hat{A} \hat{\Gamma}_H (p) \hat{A}' + \hat{\Sigma}_E$. Comparing this with the theoretical autocovariance sequence in (4) and in view of (5), we have

$$\hat{\Gamma}_h (j) = \Gamma_{h,j} (\hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e), \quad j = 0, 1, \ldots, p.$$
provided that \( I_{p, q}^2 - \hat{A} \otimes \hat{A} \) is invertible. 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 the ‘reproducing’ property of the Yule-Walker estimator, we can relate the VAR variance estimator to the kernel estimator of \( \psi_R \) based on the rectangular kernel. Let \( k_{\text{rect}}(r) = 1 \{ |r| \leq 1 \} \) and \( k_{\text{rect},b}(r) = 1 \{ |r| \leq b \} \), then the rectangular kernel estimator of \( \psi_R \) is

\[
\hat{\psi}_R = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} h_t h_s k_{\text{rect}} \left( \frac{|t-s|}{p} \right),
\]

where \( h_t \) is defined in (2) and \( p \) is the bandwidth or truncation lag. By definition,

\[
\hat{\psi}_R = \sum_{j=-p}^{p} \Gamma_{h,j} \left( \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e \right) + \sum_{|j| > p} \Gamma_{h,j} \left( \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e \right) = \hat{\psi}_R + M_R,
\]

where

\[
M_R = \sum_{|j| > p} \Gamma_{h,j} \left( \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e \right)
\]

and \( \hat{\Gamma}_h(j) = \hat{\Gamma}_h(-j)' \).

\[
\Gamma_{h,j} \left( \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e \right) = \sum_{i=1}^{p} \hat{A}_i \Gamma_{h,j-i} \left( \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e \right) \quad \text{for} \quad j > p. \tag{6}
\]

Intuitively, the fitted VAR process necessarily agrees exactly up to lag \( p \) with the estimated autocovariances. The values of the autocovariances after lag \( p \) are generated recursively in accordance with the VAR\((p)\) model as in (6). The difference between the VAR variance estimator and the rectangular kernel LRV 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 variance estimator and the rectangular kernel LRV estimator, we can establish the asymptotic distribution of the VAR variance estimator under the fixed-smoothing asymptotics. We make the following assumptions.

**Assumption 1** \( \text{plim}_{T \to \infty} \theta_T = \theta_0 \).

**Assumption 2** \( T^{1/2} g_{[\tau,T]}(\theta_0) \Rightarrow \Lambda W_m(r) \) where \( \Lambda \Lambda' = \Omega = \sum_{j=-\infty}^{\infty} E u_t u_{t-j} \) is the LRV of \( u_t := f(v_t, \theta_0) \) and \( W_m(r) \) is the \( m \)-dimensional standard Brownian motion.

**Assumption 3** \( \text{plim}_{T \to \infty} G_{[\tau,T]}(\bar{\theta}_T) = r G \) uniformly in \( r \) for any \( \bar{\theta}_T \) between \( \hat{\theta}_T \) and \( \theta_0 \) where \( G = E \left[ \partial f(v_j, \theta_0) / \partial \theta' \right] \).

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

**Assumption 5** \( W_T \) is positive semidefinite, \( \text{plim}_{T \to \infty} \mathcal{W}_T = \mathcal{W} \), 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), and Sun (2010c,d). 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 weak law of large numbers (WLLN). Note that FCLT and WLLN 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), among others. Assumption 4 is a high-level condition. Using the same argument as in Hansen (1992) and de Jong and Davidson (2000), we can show that under some moment and mixing conditions on the process \( \{f(v_t, \theta_0)\} \), we have

\[
T^{-bT} \sum_{\tau=1}^{T-bT} f(v_{r+bT}; \theta_0) g'_r(\theta_0) - \Lambda^+ T \Rightarrow \Lambda \int_0^{1-b} dW_m(b + r) W'_m(r) \Lambda',
\]

where \( \Lambda^+ = T^{-1} \sum_{\tau=1}^{T-bT} \sum_{t=1}^\tau E u_{r+bT} u'_t \). But

\[
\Lambda^+ = \frac{1}{T} \sum_{\tau=1}^{T-bT} \sum_{t=1}^\tau \Gamma_u(\tau + bT - t) = \frac{1}{T} \sum_{\tau=1}^{T-bT} \sum_{j=0}^{\tau-1} \Gamma_u(bT + j) = \frac{1}{T} \sum_{j=0}^{T-bT-1} \Gamma_u(bT + j) = \sum_{j=0}^{\infty} \Gamma_u(bT + j) + o(1) = o(1),
\]

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

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

\[
Q_m(b) = \left( \int_0^{1-b} dV_m(b + r) V'_m(r) + \int_0^{1-b} V_m(r) dV_m(r + b) \right)'
\]

and

\[
V_m(r) = W_m(r) - r W_m(1)
\]

is the standard Brownian Bridge process.

It follows from the above lemma that \( \hat{\mathcal{V}}_R \Rightarrow \mathcal{V}_{R,\infty} \). That is, under the fixed-smoothing asymptotics, \( \hat{\mathcal{V}}_R \) converges to a random matrix \( \mathcal{V}_{R,\infty} \). Note that \( \mathcal{V}_{R,\infty} \) is proportional to the true variance matrix \( \mathcal{V}_R \) through \( R(G'WG)^{-1} G'W \Lambda \) and its transpose. This contrasts with the increasing-smoothing asymptotic approximation where \( \hat{\mathcal{V}}_R \) is approximated by a constant \( \mathcal{V}_R \). 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 variance estimator.

The following lemma gives an alternative representation of $Q_m(b)$. Using this representation, we can compute the variance of $V_{R, \infty}$. The representation uses the following definition:

$$k_b^*(r, s) = k_{\text{rect}} \left( \frac{r - s}{b} \right) - \int_0^1 k_{\text{rect}} \left( \frac{r - s}{b} \right) \, dr - \int_0^1 k_{\text{rect}} \left( \frac{r - s}{b} \right) \, ds + \int_0^1 \int_0^1 k_{\text{rect}} \left( \frac{r - s}{b} \right) \, dr \, ds$$

$$= k_{\text{rect}} \left( \frac{r - s}{b} \right) + \max(0, r - b) + \max(0, s - b) - \min(1, b + r) - \min(1, b + s) - b(b - 2).$$

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

$$\text{var}(\text{vec}(\Lambda Q_m(b) \Lambda')) = \mu_2 (\Omega \otimes \Omega) (I_m^2 + K_{m^2}),$$

(c) $EV_{R, \infty} = \mu_1 V_R$ and

$$\text{var}(\text{vec}(V_{R, \infty})) = \mu_2 (V_R \otimes V_R) (I_m^2 + K_{m^2}),$$

where

$$\mu_1 = \mu_1(b) = \int_0^1 k_b^*(r, r) dr = (1 - b)^2$$

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

and $K_{m^2}$ is the $m^2 \times m^2$ commutation matrix.

The transformed kernel function $k_b^*(r, s)$ is graphed in Figure 1 for the case $b = 0.2$. In the center of the $(r, s)$ domain, the function is flat and is equal to $-b^2 - 2b + 1$. The edge effects, as clearly seen from the figure, ensure that $\int_0^1 k_b^*(r, s) dr = \int_0^1 k_b^*(r, s) ds = 0$ for any $r, s$. That is, the function $k_b^*(r, s)$ integrates to zero along each coordinate. As $b$ becomes closer to zero, the function becomes more concentrated around the line $r = s$. In the limit as $b \to 0$, $k_b^*(r, s) \to 1 \{r = s\}$. In this case, $Q_m(b) \to^p I_m$ and $V_{R, \infty} \to^p V_R$.

The convergence of $V_{R, \infty}$ to $V_R$ as $b \to 0$ also follows from the mean and variance calculations. The lemma shows that the mean of $V_{R, \infty}$ is proportional to the true variance $V_R$. When $b \to 0$, we have $(1 - b)^2 \to 1$ and $\mu_2(b) \to 0$. In other words, as $b \to 0$, the mean of $V_{R, \infty}$ converges to $V_R$ and the variance converges to zero. A direct implication is that $\text{plim}_{b \to 0} V_{R, \infty} = V_R$. We can conclude that as $b$ goes to zero, the fixed-smoothing asymptotics coincides with the conventional increasing-smoothing asymptotics. More precisely,
the probability limits of $\hat{\nu}_R$ are the same under the sequential asymptotics and the joint asymptotics.

Figure 2 presents the function $\mu_2(b)/b$, the multiplicative factor in $b^{-1}\text{var}(\text{vec}(\nu_{R,\infty}))$. It is clear that this function is decreasing in $b$. As $b \to 0$, we have

$$\lim_{b \to 0} b^{-1}\text{var}(\text{vec}(\nu_{R,\infty})) = 2 (\nu_R \otimes \nu_R) (I_m^2 + K_m^2).$$

Note that $\int_{-1}^{1} k^2(r) \, dr = \int_{-1}^{1} 1 \{|r| \leq 1\} \, dr = 2$. The right hand side is exactly the asymptotic variance one would obtain under the joint asymptotic theory. Therefore, $\hat{\nu}_R$ has not only the same probability limit but also the same asymptotic variance under the sequential and joint asymptotics. Note that $\lim_{b \to 1} b^{-1}\text{var}(\text{vec}(\nu_{R,\infty})) = 0$ and $\lim_{b \to 1} E \nu_{R,\infty} = 0$. As a result, $\text{plim}_{b \to 1} \nu_{R,\infty} = 0$. This is not surprising, as when $b = 1$,

$$\hat{\nu}_R = T^{-1} \left( \sum_{t=1}^{T} h_t \right) \left( \sum_{t=1}^{T} h_t \right)' = 0$$

by the first-order condition for the GMM estimator.

When $b$ is fixed, $E \nu_{R,\infty} = \nu_R = b(b - 2) \nu_R$. So $\hat{\nu}_R$ is not asymptotically unbiased. The asymptotics bias arises from the estimation uncertainty of model parameter $\theta$. It may be called the demeaning bias as the stochastic integral in (7) 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 $\theta$ is negligible. As a result, the first-order conventional asymptotics does not reflect the demeaning bias.

## 4 Fixed-Smoothing Asymptotics for the Wald Test

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

**Theorem 2** Let Assumptions 1-5 hold. Under the fixed-smoothing asymptotics where $b$ is held fixed, we have

$$F_T \Rightarrow F_\infty (q, b), \; t_T \Rightarrow t_\infty (b),$$

where

$$F_\infty (q, b) = W_q' (1) [Q_q (b)]^{-1} W_q (1)/q,$$
$$t_\infty (b) = W_1 (1)/\sqrt{Q_1 (b)},$$

$W_q (s)$ is $q$-dimensional standard Brownian motion, and

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

Theorem 2 is analogous to Theorem 3 in Kiefer and Vogelsang (2005). It shows that $F_T$ depends on $b$ but otherwise is nuisance parameter free. We have thus obtained asymptotically pivotal tests that reflect 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 $\chi_q^2/q$ distribution and $t_T$ a limiting $N(0, 1)$ distribution regardless of the choice of $b$. When $b \to 0$, $Q_q (b) \to^p \mathbb{I}_q$ and as a result $W_q' (1) [Q_q (b)]^{-1} W_q (1)/q \to^p \chi_q^2/q$. Hence, when $b \to 0$, the fixed-smoothing asymptotics approximates the standard increasing-smoothing asymptotics. In other words, the sequential limit of $F_T$ is identical to the joint limit of $F_T$. Compared with the fixed-smoothing asymptotics, the sequential asymptotics invokes an additional approximation, which could lead to additional approximation error. In a sequence of papers, Sun (2010a,b,c,d) and Sun, Phillips and Jin (2008) show that critical values from the fixed-smoothing asymptotics are high-order correct under the conventional joint asymptotics.

Figure 2: Graph of $\mu_2 (b)/b$
The asymptotic distribution $F_\infty (q,b)$ or $t_\infty (b)$ is nonstandard. The critical values are not readily available from statistical tables or software packages. For this reason, we proceed to approximate the critical values in closed form. Since $k_b^*(r,s) \in L^2 ([0,1] \times [0,1])$, it has a Fourier series representation:

$$k_b^*(r,s) = \sum_{\ell=1}^{\infty} \sum_{n=1}^{\infty} \lambda_{\ell n} \Phi_\ell (r) \Phi_n (s)$$

where \{\Phi_\ell (r) \Phi_n (s)\} is an orthonormal basis for $L^2 ([0,1] \times [0,1])$ and the convergence is in the $L^2$ space. For example, we may take $\Phi_\ell (r) = (1/\sqrt{2}) \cos 2\ell \pi r$ or $\sin 2\ell \pi r$. Since $\int_0^1 k_b^*(r,s)dr = \int_0^1 k_b^*(r,s)ds = 0$, we know that

$$\int_0^1 \Phi_\ell (r)dr = 0 \text{ for } \ell = 1,2,\ldots$$

In addition, by the symmetry of $k_b^*(r,s)$, we can deduce that $\lambda_{\ell n} = \lambda_{n\ell}$.

Using the Fourier series representation, we can write

$$Q_q(b) = \sum_{\ell=1}^{\infty} \sum_{n=1}^{\infty} \lambda_{\ell n} \xi_n \xi_n'$$

where $\xi_n = \int_0^1 \Phi_n (s) dW_q(s) \sim iidN(0,\mathbb{I}_q)$. The above equality holds because

$$E \int_0^1 \int_0^1 \left( k_b^*(r,s) - \sum_{\ell=1}^{\infty} \sum_{n=1}^{\infty} \lambda_{\ell n} \Phi_\ell (r) \Phi_n (s) \right) dW_q(s)dW_q(r)' = 0,$$

$$\text{var} \left[ \int_0^1 \int_0^1 \left( k_b^*(r,s) - \sum_{\ell=1}^{\infty} \sum_{n=1}^{\infty} \lambda_{\ell n} \Phi_\ell (r) \Phi_n (s) \right) dW_q(s)dW_q(r)' \right] = 0.$$

Let $\eta = W_q(1)$, then $\eta$ is independent of $\xi_n$ and

$$F_\infty (q,b) = \eta' \left( \sum_{\ell=1}^{\infty} \sum_{n=1}^{\infty} \lambda_{\ell n} \xi_n \xi_n' \right)^{-1} \eta/q.$$ 

To simplify the above representation, we note that

$$\sum_{\ell=1}^{\infty} \sum_{n=1}^{\infty} \lambda_{\ell n} \xi_n \xi_n' = \lim_{N \rightarrow \infty} \sum_{\ell=1}^{N} \sum_{n=1}^{N} \lambda_{\ell n} \xi_n \xi_n' = \lim_{N \rightarrow \infty} \xi' \Delta_N \xi,$$

where $\xi_N = (\xi_1, \ldots, \xi_N)'$ and $\Delta_N = (\lambda_{ij})$ is a $N \times N$ symmetric matrix. Let $\Delta_N = HD_N H'$ be the spectral decomposition of $\Delta_N$, where $D_N = \text{diag}(\lambda_1, \ldots, \lambda_N)$ and $H' H = HH' = \mathbb{I}_N$. Then

$$\lim_{N \rightarrow \infty} \xi' \Delta_N \xi = \lim_{N \rightarrow \infty} (\xi' H) D_N (H' \xi) = \sum_{i=1}^{\infty} \lambda_i \zeta_i \zeta_i',$$

where $\zeta = (\zeta_1, \ldots, \zeta_N)' = H' \xi$. So $Q_q(b)$ can be represented as

$$Q_q(b) = \sum_{i=1}^{\infty} \lambda_i \zeta_i \zeta_i'.$$  

(9)
It is easy to see that \( \zeta_i \sim iidN(0, I_q) \). By definition, \( \zeta_i\zeta_i' \) follows a Wishart distribution \( \mathcal{W}_q(I_q, 1) \), so \( \sum_{i=1}^{\infty} \lambda_i\zeta_i\zeta_i' \) is an infinite weighted sum of independent Wishart distributions.

Using the representation in (9), we have

\[
F_\infty(q, b) = \eta' \left( \sum_{i=1}^{\infty} \lambda_i\zeta_i\zeta_i' \right)^{-1} \eta/q,
\]

where \( \zeta_i \sim iidN(0, I_q) \), \( \eta \sim N(0, I_q) \) and \( \zeta_i \) is independent of \( \eta \) for all \( i \). That is, \( F_\infty(q, b) \) is equal in distribution to a quadratic form in standard normal variates with an independent and random weighting matrix.

Define \( \Phi = \mu_1^{-1} \sum_{i=1}^{\infty} \lambda_i\zeta_i\zeta_i' \). Then

\[
\mu_1 q F_\infty(q, b) = \eta' \Phi^{-1} \eta.
\]

We want to approximate the distribution of \( \Phi \) by a scaled Wishart distribution \( \mathcal{W}_q(I_q, K)/K \) for some integer \( K > 0 \). We select \( K \) to match their first two moments. Let \( \Psi \sim \mathcal{W}_q(I_q, K)/K \). By construction \( E\Phi = E\Psi = I_q \). For any symmetric matrix \( D \), we have

\[
E\Psi D \Phi = \frac{1}{K} tr(D)I_q + \left( 1 + \frac{1}{K} \right) D. \tag{10}
\]

See Example 7.1 in Bilodeau and Brenner (1999). Using this result, we can show that

\[
E\Phi D \Phi = \frac{\mu_2}{\mu_1^2} tr(D)I_q + \left( 1 + \frac{\mu_2}{\mu_1^2} \right) D. \tag{11}
\]

In view of (10) and (11), we can set

\[
K = \left\lceil \frac{\mu_2}{\mu_1^2} \right\rceil = \left\lceil \frac{1}{2b} \times J(b) \right\rceil,
\]

where \( \lceil x \rceil \) denotes the smallest integer greater than \( x \) and

\[
J(b) := \begin{cases} 
\frac{(1-b)^4}{(b^3/2+4b^2/3-5b/2+1)}, & b \leq 1/2 \\
\frac{6b(b-1)^2}{3b^2-2b+2}, & b > 1/2
\end{cases}
\]

That is, \( \Phi \approx^d \mathcal{W}_q(I_q, K)/K \) for the above \( K \) value where \( \approx^d \) denotes “is approximately equal to in distribution.” As a result

\[
\mu_1 q F_\infty(q, b) \approx^d \eta' \Psi^{-1} \eta.
\]

But \( \eta' \Psi^{-1} \eta \) is the Hotelling’s \( T^2(q, K - q + 1) \) distribution (Hotelling, 1931). By the well-known relationship between the \( F \) distribution and the \( T \)-square distribution, we have

\[
\frac{\mu_1 (K - q + 1)}{K} F_\infty(q, b) \approx^d \frac{(K - q + 1)}{Kq} \eta' \Psi^{-1} \eta = F_{q, K - q + 1}. \tag{12}
\]

The following theorem makes the above approximation more precise.

13
Theorem 3  Let \( \tilde{F}_\infty(q, b) \) be the corrected limiting variate defined by

\[
\tilde{F}_\infty(q, b) = \frac{\mu_1 (K - q + 1)}{K} F_\infty(q, b).
\]  

As \( b \to 0 \), we have

\[
P \left( \tilde{F}_\infty(q, b) \leq z \right) = P \left( F_{q,K} \leq z \right) + o(b)
\]  

The parameter \( K \) can be called the “equivalent degree of freedom” (EDF) of the LRV estimator. The idea of approximating a weighted sum of independent Wishart distributions by a simple Wishart distribution with equivalent degree of freedom can be motivated from the early statistical literature on spectral density estimation. In the scalar case, the distribution of the spectral density estimator is often approximated by a chi-square distribution with equivalent degree of freedom; see Priestley (1981, p. 467). In addition, Hall (1983) suggests approximating the distribution of a sum of independent scalar random variables by a chi-square distribution instead of a normal distribution. The advantage of using a chi-square distribution is that it can pick up any positive skewness component in the true distribution. Kollo and von Rosen (1995) extend the idea to multivariate cases and use a Wishart distribution as the approximating distribution. This is exactly the approach we employ here.

When \( b \to 0 \), the EDF is equal to \( 1/(2b) \) up to smaller order terms where the asymptotic variance of the LRV estimator is proportional to \( 2b \). Figure 3 graphs the function \( J(b) \), which provides high-order adjustment to the first-order EDF \( 1/(2b) \). It is clear that as \( b \) decreases, i.e. as the degree of smoothing increases, the EDF increases and the variance decreases. In other words, the higher the EDF is, the larger the degree of smoothing is, and the smaller the variance is.

It is easy to see that Theorem 3 remains valid if the correction factor \( \mu_1 (K - q + 1) / K \) is replaced by its asymptotically equivalent form. More specifically, as \( b \to 0 \),

\[
\frac{\mu_1 (K - q + 1)}{K} = (1 - b)^2 \left( 1 - (q - 1) 2b \times J^{-1}(b) \right)
\]

\[
= 1 - 2qb + o(b) = \frac{1}{1 + 2qb} + o(b)
\]

\[
= \exp(-2qb) + o(b).
\]

In addition

\[
P \left( F_{q,K} \leq z \right) = P \left( F_{q,K} \leq z \right) + o(b).
\]

Combining the above two equations and Theorem 3, we obtain the corollary below.

Corollary 4  Let \( F_\infty^*(q, b) \) be the corrected limiting variate defined by

\[
F_\infty^*(q, b) = \kappa F_\infty(q, b),
\]  

where

\[
\kappa = \frac{1}{1 + 2qb} \text{ or } \exp(-2qb).
\]

As \( b \to 0 \), we have

\[
P \left( F_\infty^*(q, b) \leq z \right) = P \left( F_{q,K} \leq z \right) + o(b)
\]  

where \( K = \max((2b)^{-1}, 5) \).
In the above corollary, we modify the second degree of freedom in the $F$ approximation. The reason is that the variance of an $F$ distribution exists only if its second degree of freedom is larger than 4. Obviously, the modification does not have any impact on the asymptotic result, as $K = [(2b)^{-1}]$ when $b \to 0$.

Let $F_{q,K}^\alpha$ and $F_{q,b}^\alpha$ be the $1 - \alpha$ quantiles of the standard $F_{q,K}$ distribution and the nonstandard $F_{q,b}$ distribution, respectively. Then

$$P \left( F_{\infty}^q(q,b) > F_{q,K}^\alpha \right) = P \left( F_{q,K} > F_{q,K}^\alpha \right) + o(b).$$

That is,

$$P \left( F_{\infty}^q(q,b) > \kappa^{-1}F_{q,K}^\alpha \right) = \alpha + o(b).$$

In other words,

$$F_{\infty}^\alpha(q,b) = \kappa^{-1}F_{q,K}^\alpha + o(b).$$

So for the original $F$ statistic, we can use

$$F_{q,b}^\alpha = \kappa^{-1}F_{q,K}^\alpha$$

as the critical value for the test with nominal size $\alpha$.

As an approximation to the nonstandard critical value, the critical value $F_{q,b}^\alpha$ is second-order correct as the approximation error in (18) is of smaller order $o(b)$ rather than $O(b)$ as $b \to 0$. The second-order critical value is larger than the standard critical value from $\chi^2_q/q$ for two reasons. First, $F_{q,K}^\alpha$ is larger than the corresponding critical value from $\chi^2_q/q$ due to the presence of a random denominator in the $F$ distribution. Second, the correction factor $\kappa^{-1}$ is larger than 1. As $b$ increases, both the correction factor and $F$ critical value $F_{q,K}^\alpha$ increase. As a result, the second-order critical value $F_{q,b}^\alpha$ is an increasing function of $b$. 

Figure 3: Graph of $J(b)$
The second-order critical value is also increasing in \( q \), the number of restrictions being jointly tested. This result is especially interesting when \( q \) is large. In this case, the size distortion of the usual Wald test is large. For example, Burnside and Eichenbaum (1996) show that the small sample size of the Wald test increases sharply with the number of hypotheses being jointly tested. The second-order critical value takes this into account and adjusts the critical value monotonically with \( q \). Our result provides a theoretical explanation of the finite sample results reported by Ray and Savin (2008) and Ray, Savin and Tiwari (2009) who find that the fixed-\( b \) type of asymptotic approximation can substantially reduce size distortion in testing joint hypotheses, especially when the number of hypotheses being tested is large.

The correction in (15) can be regarded as a Bartlett-type correction. Bartlett (1937, 1954) considers likelihood ratio tests, but the basic idea can be applied to Wald tests as well. See Cribari-Neto and Cordeiro (1999) for a more recent survey. The argument goes as follows. Suppose that \( X \sim F_\infty(q,b) \) and \( EX = 1 + bC + o(b) \) for some constant \( C \), then as \( b \to 0 \), \( F_\infty(q,b)/(1+bC) \) is closer to the \( \chi^2_q/q \) distribution than the original distribution \( F_\infty(q,b) \). Using the result in the appendix, we can show that

\[
EX = E \left[ \nu_{11} - \nu_{12} \nu_{21}^{-1} \nu_{21} \right] = 1 + 2bq + o(b).
\]

So the constant \( C \) should be \( C = 2q \). We have thus provided another way to motivate the correction in (13). Essentially, the correction makes the mean of the \( F_\infty(q,b) \) distribution closer to that of the \( \chi^2_q/q \) distribution. In addition to the Bartlett-type correction, Theorem 3 approximates the nonstandard distribution by an \( F \)-distribution rather than a chi-square distribution.

Define \( F^*_T = \kappa F_T \). It then follows from Theorems 2 and 3 that under the sequential asymptotics, we have

\[
\lim_{T \to \infty} P (F^*_T < z) = P (F_{q,K} \leq z) + o(b)
\]

as \( b \to 0 \). In the rest of the paper, we call the test based on the corrected Wald statistic \( F^*_T \) and using the critical values from the \( F_{q,K} \) distribution the \( F^* \) test. To emphasize its reliance on the vector autoregression, we also refer to it as the VAR \( F^* \) test.

### 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 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 (2010d) 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

\[
b^* = \arg \min_{b} e_{II}(b), \text{ s.t. } e_I(b) \leq \alpha \tau
\]
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 $q$, the type I error of the kernel $F^*$ test is shown to approximately equal

$$e_I (b) = \alpha - (bT)^{-q} G_q' (\alpha q) X^\alpha \tilde{B}$$

where $\alpha$ is the nominal type I error, $X^\alpha$ is the $\alpha$-level critical value from $G_q (\cdot)$, the CDF of the $\chi^2$ distribution, and $(bT)^{-q} \tilde{B}$ is the asymptotic bias of the kernel LRV estimator.

A similar problem is also present for optimal bandwidth choice under the MSE criterion, as the conventional squared bias and variance tradeoff does not apply to the rectangular kernel. However, his results do not directly apply as the kernels considered in Sun (2010d) are finite-order kernels while the rectangular kernel is an infinite-order kernel.

The average type II error under the local alternative

$$H_1 (\delta_0^2) : \tau (\theta_0) = (R_0 V_{\infty} R_0')^{1/2} \bar{e} / \sqrt{T}$$

where $\bar{e}$ is uniformly distributed on $S_p (\delta_0) = \{ \bar{e} : \| \bar{e} \|^2 = \delta_0^2 \}$ is

$$e_{II} (b) = G_{q, \delta_0^2} (\alpha q) + (bT)^{-q} G_{q, \delta_0^2}' (\alpha q) X^\alpha \tilde{B} + \frac{\delta_0^2}{2} G_{q+2, \delta_0^2}' (\alpha q) X^\alpha c_2 b$$

where $G_{q, \delta_0^2} (\cdot)$ is the CDF of the noncentral $\chi^2 (\delta_0^2)$ 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.

The testing optimal bandwidth in Sun (2010d) depends on the sign of $\tilde{B}$. When $\tilde{B} < 0$, the constraint $e_I (b) \leq \alpha \tau$ is binding and the optimal $b^*$ satisfies $e_I (b^*) = \alpha \tau$. When $\tilde{B} > 0$, the constraint $e_I (b) \leq \alpha \tau$ is not binding and the optimal $b^*$ minimizes $e_{II} (b)$.

In principle, we can follow Sun (2010d) and select the testing-optimal bandwidth for the rectangular kernel. However, his results do not directly apply as the kernels considered in Sun (2010d) are finite-order kernels while the rectangular kernel is an infinite-order kernel. A similar problem is also present for optimal bandwidth choice under the MSE criterion, as the conventional squared bias and variance tradeoff does not apply to the rectangular kernel. To solve 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.

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

$$b^*_{rect} = \begin{cases} 
  b^*_{tar}, & \text{if } \tilde{B} < 0 \\
  (c_{2, tar})^{-1} b^*_{tar}, & \text{if } \tilde{B} > 0
\end{cases}$$

where $c_{2, tar} = \int_{-\infty}^{\infty} k_{tar}^2 (x) \, dx$ and $c_{2, rect} = \int_{-\infty}^{\infty} k_{rect}^2 (x) \, dx = 2$. For example, when the Parzen kernel is used as the target kernel,

$$b^*_{rect} = b^*_{Parzen} \frac{0.539285}{2} b^*_{Parzen} \begin{cases} 
  1 \{ \tilde{B} < 0 \} & + \\
  1 \{ \tilde{B} > 0 \} & - \end{cases}$$

When the QS kernel is used as the target kernel,

$$b^*_{rect} = b^*_{QS} \begin{cases} 
  1 \{ \tilde{B} < 0 \} & + \\
  1 \{ \tilde{B} > 0 \} & - \end{cases}$$

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

When $\tilde{B} < 0$, 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, 2010). As a consequence, the bandwidth selected by the MTK is expected to control the type I error at least as well as 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 $\delta_2^2 G_{(q+2),q} G_{q}^{\alpha} (X_q^\alpha) X_q^\alpha c_2b/\varepsilon_{\theta_1}(b)$ 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 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 AMSE 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 chi-square tests as well as kernel-based $F^*$ tests recently proposed by Sun (2010d).

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:

$$u_{it} = \rho_1 u_{i,t-1} + \rho_2 u_{i,t-2} + e_{it}$$
where \( e_{it} \sim iidN(0, \sigma_e^2) \) 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 \( e_{it} \sim iidN(0, \sigma_e^2) \) and \( \sigma_e^2 = (1 + \rho_1^2 + \rho_2^2)^{-1} \). In both cases, the value of \( \sigma_e^2 \) is chosen such that the variance of \( u_{it} \) is one.

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 \( H_{1q}(\theta) : R_{0q} \theta = c_q / \sqrt{T} \) where \( c_q = (R_{0q} \Omega R_{0q}')^{1/2} \tilde{c}, \Omega \) is the long run variance matrix of \( u_t, \tilde{c} \) is uniformly distributed over the sphere \( S_q(\tilde{\sigma}^2) \), that is, \( \tilde{c} = \sigma / \| \xi \|, \xi \sim N(0, \mathbb{I}_q) \). It is important to point out that \( \sigma^2 \) is not the same as \( \sigma_0^2 \) used in the testing-oriented criterion and the MTK.

We consider the following \((p_1, p_2)\) combinations:

\[
(p_1, p_2) = (-.8, 0), (-.4, 0), (0, 0), (4, 0), (8, 0), (1.5, -.75), (.25, .25), (.35, .35).
\]

The last two combinations come from den Haan and Levin (1998). The combination with negative \( \rho_2 \) 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 variance estimator. For each restriction matrix \( R_{0q} \), we fit a VAR\((p)\) model to \( R_{0q} (u_t - \tilde{u}_t) \) 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 \( p = [b_{rect}^* T] \) where \( b_{rect}^* \) is defined in (22). 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 variance estimator and compute the Wald statistic. We perform both the \( F^* \) test proposed in this paper and the traditional \( \chi^2 \) test. The \( F^* \) test employs the modified Wald statistic \( \exp(-2qab) F_T \) and critical values from the \( F_{q,K} \) distribution where \( K = \max([T/(2p)], 5) \) and \( \hat{p} \) is the selected VAR order. The traditional \( \chi^2 \) test employs the unmodified Wald statistic and critical values from the \( \chi^2_q / q \) distribution. We refer to these tests as \( F^*-VAR-AIC, \chi^2-VAR-AIC, F^*-VAR-BIC, \chi^2-VAR-BIC, F^*-VAR-Par, \chi^2-VAR-Par, F^*-VAR-QS, \) and \( \chi^2-VAR-QS \), respectively.

The second set of testing procedures consists of kernel-based Wald-type tests. We consider two commonly used second-order kernels: the Parzen and QS kernels. For each kernel, the bandwidth is determined via either a modified MSE criterion (Andrews, 1991) or the testing-oriented criterion (Sun, 2010d). In the former case, we consider the asymptotic MSE of the LRV estimator for the transformed moment process \( h_t \). This is in contrast with the original MSE criterion in Andrews (1991), which is based on the moment process \( f_t \). To some extent, the modified MSE is tailored to the null hypothesis. The modification makes the MSE-based method as competitive as possible. In the latter case, the bandwidth is selected to solve the constrained minimization problem in 21. We set \( \tau = 1.2 \) in the
simulation experiment. The conventional tests using MSE-optimal bandwidth and the \( \chi^2/q \) critical values are referred to as \( \chi^2 \)-Parzen and \( \chi^2 \)-QS, respectively. The tests proposed by Sun (2010d) are referred to as \( F^* \)-Parzen and \( F^* \)-QS as they use critical values from \( F \) distributions. Both the MSE-optimal bandwidth and the testing-optimal bandwidth require a plug-in implementation. We use the VAR model selected by the BIC 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.

Table 1 gives the type I errors of the ten testing methods for the AR error with sample size \( T = 100 \). The significance level is 5\%, which is also the nominal type I error. Several patterns emerge. First, as it is clear from the table, 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 errors of the conventional Wald tests are 0.475 and 0.452 respectively for the Parzen and QS kernels. These empirical type I errors are far from 0.05, the nominal type I error.

Second, the size distortion of the VAR \( F^* \) test is substantially smaller than the corresponding \( \chi^2 \) test. Note that the lag order underlying the VAR \( F^* \) test is the same as that for the corresponding VAR \( \chi^2 \) test. The VAR \( F^* \) test is more accurate in size because it employs an asymptotic approximation that captures the estimation uncertainty of the LRV 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 \( \chi^2 \) test.

Third, the size distortion of the \( F^* \)-Parzen and \( F^* \)-QS tests is also much smaller than that of the corresponding \( \chi^2 \) tests. There are two reasons for this observation. For the kernel \( F^* \) tests, the bandwidth is chosen to control the asymptotic type I error, which captures the empirical type I error to some extent. In addition, the kernel \( F^* \) tests also employ more accurate asymptotic approximations. So it is not surprising that the kernel \( F^* \) tests have more accurate size than the corresponding \( \chi^2 \) tests.

Fourth, among the \( F^* \) tests based on the VAR variance estimator, the test based on the MTK has the smallest size distortion while the test based on the BIC has the largest size distortion. Unreported results show that in an average sense the VAR order selected by the MTK is the largest while that selected by the BIC is the smallest. In terms of the size accuracy, the AIC and BIC appear to be too conservative in choosing the AR lag order. This is especially true for the BIC. den Haan and Levin (1998) obtain the same result. Hall (1994) and Ng and Perron (1995) obtain similar results in Monte Carlo studies on the choice of AR order for the augmented Dickey-Fuller unit root test.

Finally, 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. This observation confirms the advantage of using the VAR variance estimator as compared to the kernel LRV estimator using a finite-order kernel. 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 \( F^* \) 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.

Figures 4-6 present the finite sample power in the AR case for different values of \( q \). We
compute the power using the 5% empirical finite sample critical values obtained from the
null distribution. 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 Table 1 except that the DGP is
generated under the local alternatives. The power curves are for the $F^*$ tests. We do not
include chi-square tests as Sun (2010d) has shown that the kernel-based $F^*$ tests are as
powerful as the conventional chi-square tests. Three observations can be drawn from these
figures. First, the VAR $F^*$ test based on the AIC or BIC are 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. Second, the power differences
among the $F^*$ tests are small in general. An exception is the $F^*$-QS test, which incurs
some power loss when the processes are highly persistent and the number of restrictions
being jointly tested is relatively large. Third, 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 power. This is especially true
for the $F^*$-VAR-QS test.

Table 2 presents the simulated type I errors for the MA error. The qualitative ob-
servations on size comparison for the AR case remain valid. In fact, these qualitative
observations hold for other parameter configurations such as different sample sizes and sig-
nificance levels. Quantitatively, the empirical type I errors for the MA case are smaller
than those for the AR case. 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:

$$ y_t = \gamma + x_0^t \beta + \varepsilon_t, $$

where $x_t$ is a $3 \times 1$ vector process and $x_t$ and $\varepsilon_t$ follow either an AR (1) process

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

or an MA(1) process

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

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

Let $\theta = (\gamma', \beta')'$. We estimate $\theta$ by the OLS estimator. Since the model is exactly
identified, the weighting matrix $W_T$ becomes irrelevant. Let $\tilde{x}_t' = [1, x_t']$ and $\tilde{X} = [\tilde{x}_1, \ldots, \tilde{x}_T]'$, then the OLS estimator is $\tilde{\theta}_T - \theta_0 = -G_T^{-1} g_T (\theta_0)$ where $G_T = -\tilde{X}' \tilde{X} / T, \ G_0 = E(G_T), \ g_T (\theta_0) = T^{-1} \sum_{t=1}^T \tilde{x}_t \varepsilon_t$. The asymptotic variance of $\sqrt{T}(\tilde{\theta}_T - \theta_0)$ is $V_\infty = G_0^{-1} \Omega G_0^{-1}$ where $\Omega$ 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 $H_{1q}(\delta_2^2) : R_{0q}\theta = c_q/\sqrt{T}$ where $c_q = (R_{0q}G_0^{-1}\Omega G_0^{-1}R_{0q}')^{1/2}\tilde{c}$ and $\tilde{c}$ is uniformly distributed over the sphere $S_q(\delta_2^2)$.

Table 3 reports the empirical type I error of different tests for the AR(1) case. As before, it is clear that the $F^*$ test is more accurate in size than the corresponding $\chi^2$ 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. Compared with the kernel $F^*$ test, the VAR $F^*$ test based on the MTK is more accurate in size.

To sum up, the $F^*$-VAR-QS test has much smaller size distortion than the conventional $\chi^2$ test, as considered by den Haan and Levin (1998). It also has more accurate size than the kernel $F^*$ tests proposed by Sun (2010d). The size accuracy of the $F^*$-VAR-QS test is achieved with no or small power loss. In fact, the $F^*$-VAR-QS test is more powerful than the corresponding kernel test in some scenarios.

7 Conclusions

The paper has established a new asymptotic theory for long run variance estimators that are 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 moment conditions. 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. Simulations show that the $F^*$ test has much smaller size distortion than the conventional chi-square test.

The new asymptotic theory can be extended to the autoregressive estimator of spectral densities at other frequencies. The idea of the paper can be used to tackle other econometric problems that employ vector autoregression as an approximating parametric model to capture autoregressive or moving average components, for example, in the augmented Dickey-Fuller unit root test of Said and Dickey (1984) and the fully modified regression of Phillips and Hansen (1990).

In this paper, the VAR order is determined by the AIC, BIC or the MTK that exploits the connection between autoregressive estimators and kernel estimators. Although the MTK works very well, it is interesting to develop alternative model selection criteria for LRV estimation and the associated autocorrelation robust tests.
Table 1: Type I error of different tests for location models with AR errors with $T = 100$

| $(p_1, p_2)$ | (−0.8, 0) | (−0.4, 0) | (0, 0) | (0.4, 0) | (0.8, 0) | (1.5, −0.75) | (0.25, 0.25) | (0.35, 0.35) |
|-------------|-----------|-----------|-------|---------|---------|-------------|-------------|-------------|
| $q = 1$     |           |           |       |         |         |             |             |             |
| F*-VAR-AIC  | 0.051     | 0.053     | 0.058 | 0.065  | 0.106  | 0.051       | 0.090       | 0.104       |
| $\chi^2$-VAR-AIC | 0.061 | 0.062 | 0.066 | 0.075 | 0.119 | 0.069 | 0.107 | 0.125 |
| F*-VAR-BIC  | 0.048     | 0.050     | 0.055 | 0.061  | 0.105  | 0.048       | 0.107       | 0.117       |
| $\chi^2$-VAR-BIC | 0.056 | 0.058 | 0.061 | 0.071 | 0.115 | 0.065 | 0.120 | 0.135 |
| F*-VAR-Par  | 0.048     | 0.051     | 0.054 | 0.045  | 0.076  | 0.039       | 0.054       | 0.070       |
| $\chi^2$-VAR-Par | 0.063 | 0.058 | 0.068 | 0.125 | 0.176 | 0.109 | 0.132 | 0.169 |
| F*-VAR-QS   | 0.048     | 0.050     | 0.054 | 0.054  | 0.076  | 0.040       | 0.062       | 0.072       |
| $\chi^2$-VAR-QS | 0.056 | 0.058 | 0.063 | 0.100 | 0.176 | 0.106 | 0.120 | 0.164 |
| F*-Parzen   | 0.044     | 0.047     | 0.054 | 0.062  | 0.082  | 0.038       | 0.084       | 0.089       |
| $\chi^2$-Parzen | 0.044 | 0.050 | 0.060 | 0.093 | 0.179 | 0.107 | 0.135 | 0.191 |
| F*-QS       | 0.047     | 0.047     | 0.055 | 0.063  | 0.084  | 0.042       | 0.090       | 0.118       |
| $\chi^2$-QS | 0.049     | 0.049     | 0.060 | 0.091  | 0.172  | 0.103       | 0.136       | 0.199       |

| $q = 2$     |           |           |       |         |         |             |             |             |
| F*-VAR-AIC  | 0.047     | 0.051     | 0.057 | 0.076  | 0.161  | 0.054       | 0.127       | 0.151       |
| $\chi^2$-VAR-AIC | 0.062 | 0.069 | 0.077 | 0.097 | 0.184 | 0.089 | 0.166 | 0.200 |
| F*-VAR-BIC  | 0.045     | 0.050     | 0.056 | 0.073  | 0.160  | 0.052       | 0.182       | 0.208       |
| $\chi^2$-VAR-BIC | 0.059 | 0.065 | 0.073 | 0.093 | 0.181 | 0.085 | 0.213 | 0.249 |
| F*-VAR-Par  | 0.045     | 0.050     | 0.051 | 0.048  | 0.113  | 0.040       | 0.064       | 0.099       |
| $\chi^2$-VAR-Par | 0.059 | 0.065 | 0.082 | 0.235 | 0.348 | 0.184 | 0.253 | 0.334 |
| F*-VAR-QS   | 0.045     | 0.050     | 0.055 | 0.056  | 0.113  | 0.042       | 0.074       | 0.101       |
| $\chi^2$-VAR-QS | 0.059 | 0.065 | 0.075 | 0.169 | 0.347 | 0.174 | 0.195 | 0.320 |
| F*-Parzen   | 0.034     | 0.040     | 0.052 | 0.062  | 0.114  | 0.038       | 0.105       | 0.122       |
| $\chi^2$-Parzen | 0.043 | 0.048 | 0.067 | 0.131 | 0.306 | 0.161 | 0.205 | 0.335 |
| F*-QS       | 0.040     | 0.042     | 0.052 | 0.063  | 0.131  | 0.045       | 0.123       | 0.198       |
| $\chi^2$-QS | 0.044     | 0.047     | 0.065 | 0.127  | 0.292  | 0.151       | 0.298       | 0.364       |

| $q = 3$     |           |           |       |         |         |             |             |             |
| F*-VAR-AIC  | 0.043     | 0.048     | 0.058 | 0.086  | 0.236  | 0.064       | 0.176       | 0.216       |
| $\chi^2$-VAR-AIC | 0.067 | 0.071 | 0.085 | 0.117 | 0.279 | 0.118 | 0.241 | 0.298 |
| F*-VAR-BIC  | 0.042     | 0.048     | 0.057 | 0.086  | 0.235  | 0.063       | 0.253       | 0.363       |
| $\chi^2$-VAR-BIC | 0.065 | 0.070 | 0.084 | 0.115 | 0.276 | 0.116 | 0.301 | 0.418 |
| F*-VAR-Par  | 0.042     | 0.048     | 0.055 | 0.063  | 0.183  | 0.049       | 0.086       | 0.157       |
| $\chi^2$-VAR-Par | 0.065 | 0.070 | 0.096 | 0.386 | 0.572 | 0.284 | 0.429 | 0.549 |
| F*-VAR-QS   | 0.042     | 0.048     | 0.057 | 0.067  | 0.182  | 0.051       | 0.099       | 0.157       |
| $\chi^2$-VAR-QS | 0.065 | 0.070 | 0.086 | 0.274 | 0.569 | 0.269 | 0.296 | 0.531 |
| F*-Parzen   | 0.031     | 0.040     | 0.052 | 0.069  | 0.176  | 0.040       | 0.117       | 0.152       |
| $\chi^2$-Parzen | 0.044 | 0.046 | 0.076 | 0.168 | 0.475 | 0.210 | 0.264 | 0.444 |
| F*-QS       | 0.042     | 0.042     | 0.053 | 0.070  | 0.227  | 0.055       | 0.129       | 0.235       |
| $\chi^2$-QS | 0.044     | 0.044     | 0.073 | 0.161  | 0.452  | 0.205       | 0.267       | 0.469       |

Note: F*-VAR-AIC, F*-VAR-BIC, F*-VAR-Par, F*-VAR-QS: the $F^*$ test based on the VAR variance estimator with the VAR order selected by AIC, BIC or MTK based on Parzen and QS kernels. $\chi^2$-VAR-AIC, $\chi^2$-VAR-BIC, $\chi^2$-VAR-Par, $\chi^2$-VAR-QS are the same as the corresponding $F^*$ tests but use critical values from the chi-square distribution. F*-Parzen and $F^*$-QS are the $F^*$ tests proposed by Sun (2010d). $\chi^2$-Parzen and $\chi^2$-QS are conventional chi-square tests.
Table 2: Type I error of different tests for location models with MA errors with $T = 100$

|         | $q = 1$ |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|
|         | $(-0.8, 0)$ | $(-0.4, 0)$ | $(0, 0)$ | $(0.4, 0)$ | $(0.8, 0)$ | $(1.5, -0.75)$ | $(0.25, 0.25)$ | $(0.35, 0.35)$ |
| $F^*$-VAR-AIC | 0.006 | 0.035 | 0.058 | 0.059 | 0.065 | 0.047 | 0.067 | 0.068 |
| $\chi^2$-VAR-AIC | 0.013 | 0.045 | 0.066 | 0.071 | 0.099 | 0.063 | 0.081 | 0.087 |
| $F^*$-VAR-BIC | 0.001 | 0.022 | 0.055 | 0.049 | 0.070 | 0.040 | 0.068 | 0.064 |
| $\chi^2$-VAR-BIC | 0.002 | 0.029 | 0.061 | 0.058 | 0.094 | 0.051 | 0.079 | 0.076 |
| $F^*$-VAR-Par | 0.000 | 0.021 | 0.054 | 0.046 | 0.056 | 0.045 | 0.047 | 0.049 |
| $\chi^2$-VAR-Par | 0.000 | 0.027 | 0.068 | 0.111 | 0.088 | 0.067 | 0.112 | 0.120 |
| $F^*$-VAR-QS | 0.000 | 0.021 | 0.054 | 0.053 | 0.044 | 0.042 | 0.053 | 0.054 |
| $\chi^2$-VAR-QS | 0.000 | 0.027 | 0.063 | 0.087 | 0.068 | 0.056 | 0.090 | 0.103 |
| $F^*$-Parzen | 0.001 | 0.034 | 0.054 | 0.053 | 0.054 | 0.038 | 0.062 | 0.058 |
| $\chi^2$-Parzen | 0.000 | 0.029 | 0.060 | 0.073 | 0.076 | 0.041 | 0.091 | 0.092 |
| $F^*$-QS | 0.001 | 0.034 | 0.055 | 0.054 | 0.053 | 0.037 | 0.063 | 0.060 |
| $\chi^2$-QS | 0.000 | 0.025 | 0.060 | 0.072 | 0.073 | 0.039 | 0.090 | 0.090 |

|         | $q = 2$ |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|
| $F^*$-VAR-AIC | 0.001 | 0.021 | 0.057 | 0.057 | 0.074 | 0.044 | 0.079 | 0.077 |
| $\chi^2$-VAR-AIC | 0.005 | 0.034 | 0.077 | 0.079 | 0.143 | 0.071 | 0.106 | 0.111 |
| $F^*$-VAR-BIC | 0.000 | 0.011 | 0.056 | 0.041 | 0.088 | 0.023 | 0.090 | 0.082 |
| $\chi^2$-VAR-BIC | 0.000 | 0.016 | 0.073 | 0.055 | 0.124 | 0.034 | 0.113 | 0.102 |
| $F^*$-VAR-Par | 0.000 | 0.011 | 0.051 | 0.041 | 0.058 | 0.038 | 0.047 | 0.048 |
| $\chi^2$-VAR-Par | 0.000 | 0.016 | 0.082 | 0.210 | 0.124 | 0.089 | 0.206 | 0.230 |
| $F^*$-VAR-QS | 0.000 | 0.010 | 0.055 | 0.050 | 0.041 | 0.031 | 0.055 | 0.055 |
| $\chi^2$-VAR-QS | 0.000 | 0.016 | 0.075 | 0.138 | 0.088 | 0.052 | 0.137 | 0.168 |
| $F^*$-Parzen | 0.000 | 0.034 | 0.054 | 0.053 | 0.054 | 0.038 | 0.062 | 0.059 |
| $\chi^2$-Parzen | 0.000 | 0.016 | 0.067 | 0.100 | 0.102 | 0.036 | 0.118 | 0.123 |
| $F^*$-QS | 0.000 | 0.023 | 0.052 | 0.052 | 0.056 | 0.025 | 0.067 | 0.061 |
| $\chi^2$-QS | 0.000 | 0.014 | 0.065 | 0.095 | 0.097 | 0.032 | 0.117 | 0.118 |

|         | $q = 3$ |         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|---------|---------|
| $F^*$-VAR-AIC | 0.000 | 0.017 | 0.058 | 0.056 | 0.091 | 0.043 | 0.098 | 0.089 |
| $\chi^2$-VAR-AIC | 0.003 | 0.029 | 0.085 | 0.085 | 0.191 | 0.074 | 0.136 | 0.135 |
| $F^*$-VAR-BIC | 0.000 | 0.008 | 0.057 | 0.043 | 0.076 | 0.016 | 0.106 | 0.098 |
| $\chi^2$-VAR-BIC | 0.000 | 0.014 | 0.084 | 0.062 | 0.114 | 0.026 | 0.142 | 0.131 |
| $F^*$-VAR-Par | 0.000 | 0.008 | 0.055 | 0.053 | 0.069 | 0.037 | 0.059 | 0.062 |
| $\chi^2$-VAR-Par | 0.000 | 0.014 | 0.096 | 0.346 | 0.264 | 0.109 | 0.345 | 0.385 |
| $F^*$-VAR-QS | 0.000 | 0.008 | 0.057 | 0.059 | 0.059 | 0.028 | 0.068 | 0.068 |
| $\chi^2$-VAR-QS | 0.000 | 0.014 | 0.086 | 0.207 | 0.197 | 0.053 | 0.207 | 0.266 |
| $F^*$-Parzen | 0.000 | 0.018 | 0.052 | 0.058 | 0.059 | 0.023 | 0.068 | 0.065 |
| $\chi^2$-Parzen | 0.000 | 0.013 | 0.076 | 0.123 | 0.129 | 0.038 | 0.145 | 0.158 |
| $F^*$-QS | 0.000 | 0.018 | 0.053 | 0.060 | 0.060 | 0.021 | 0.070 | 0.067 |
| $\chi^2$-QS | 0.000 | 0.011 | 0.073 | 0.116 | 0.124 | 0.033 | 0.139 | 0.148 |

See note to Table 1
Table 3: Type I error of different tests in a regression model with AR(1) regressor and error and $T = 100$

| $q$ | (-0.75.0) | (-0.5.0) | (-0.25.0) | (0.0) | (0.25.0) | (0.5.0) | (0.75.0) | (0.9.0) |
|-----|------------|----------|-----------|-------|----------|---------|---------|--------|
| 1   |            |          |           |       |          |         |         |        |
| F*-VAR-AIC | 0.047 | 0.049 | 0.051 | 0.055 | 0.064 | 0.082 | 0.127 | 0.251 |
| $\chi^2$-VAR-AIC | 0.054 | 0.059 | 0.061 | 0.067 | 0.075 | 0.092 | 0.143 | 0.269 |
| F*-VAR-BIC | 0.045 | 0.047 | 0.049 | 0.053 | 0.061 | 0.078 | 0.128 | 0.251 |
| $\chi^2$-VAR-BIC | 0.051 | 0.055 | 0.057 | 0.063 | 0.070 | 0.088 | 0.139 | 0.265 |
| F*-VAR-Par | 0.045 | 0.047 | 0.049 | 0.052 | 0.052 | 0.057 | 0.089 | 0.184 |
| $\chi^2$-VAR-Par | 0.059 | 0.055 | 0.057 | 0.068 | 0.108 | 0.148 | 0.194 | 0.318 |
| F*-VAR-QS | 0.045 | 0.047 | 0.049 | 0.053 | 0.058 | 0.064 | 0.089 | 0.182 |
| $\chi^2$-VAR-QS | 0.051 | 0.055 | 0.057 | 0.064 | 0.084 | 0.121 | 0.191 | 0.317 |
| F*-Parzen | 0.040 | 0.043 | 0.045 | 0.053 | 0.065 | 0.073 | 0.102 | 0.199 |
| $\chi^2$-Parzen | 0.041 | 0.047 | 0.049 | 0.060 | 0.084 | 0.113 | 0.185 | 0.334 |
| F*-QS | 0.042 | 0.043 | 0.045 | 0.054 | 0.066 | 0.076 | 0.106 | 0.209 |
| $\chi^2$-QS | 0.043 | 0.046 | 0.046 | 0.059 | 0.084 | 0.113 | 0.182 | 0.330 |
| 2   |            |          |           |       |          |         |         |        |
| F*-VAR-AIC | 0.092 | 0.068 | 0.065 | 0.065 | 0.075 | 0.099 | 0.177 | 0.352 |
| $\chi^2$-VAR-AIC | 0.114 | 0.088 | 0.082 | 0.087 | 0.097 | 0.125 | 0.204 | 0.388 |
| F*-VAR-BIC | 0.087 | 0.065 | 0.062 | 0.063 | 0.073 | 0.097 | 0.173 | 0.348 |
| $\chi^2$-VAR-BIC | 0.107 | 0.083 | 0.078 | 0.084 | 0.093 | 0.120 | 0.197 | 0.380 |
| F*-VAR-Par | 0.061 | 0.061 | 0.060 | 0.061 | 0.060 | 0.066 | 0.117 | 0.251 |
| $\chi^2$-VAR-Par | 0.242 | 0.121 | 0.082 | 0.094 | 0.167 | 0.270 | 0.368 | 0.533 |
| F*-VAR-QS | 0.073 | 0.064 | 0.062 | 0.063 | 0.068 | 0.079 | 0.119 | 0.252 |
| $\chi^2$-VAR-QS | 0.184 | 0.095 | 0.078 | 0.086 | 0.116 | 0.204 | 0.360 | 0.533 |
| F*-Parzen | 0.089 | 0.064 | 0.058 | 0.063 | 0.073 | 0.089 | 0.125 | 0.260 |
| $\chi^2$-Parzen | 0.156 | 0.093 | 0.070 | 0.077 | 0.110 | 0.161 | 0.291 | 0.519 |
| F*-QS | 0.093 | 0.066 | 0.058 | 0.063 | 0.075 | 0.092 | 0.131 | 0.285 |
| $\chi^2$-QS | 0.154 | 0.090 | 0.067 | 0.075 | 0.107 | 0.158 | 0.282 | 0.509 |
| 3   |            |          |           |       |          |         |         |        |
| F*-VAR-AIC | 0.133 | 0.082 | 0.072 | 0.076 | 0.086 | 0.120 | 0.221 | 0.436 |
| $\chi^2$-VAR-AIC | 0.168 | 0.114 | 0.100 | 0.104 | 0.120 | 0.157 | 0.267 | 0.484 |
| F*-VAR-BIC | 0.130 | 0.081 | 0.072 | 0.075 | 0.085 | 0.118 | 0.217 | 0.430 |
| $\chi^2$-VAR-BIC | 0.164 | 0.111 | 0.099 | 0.102 | 0.118 | 0.154 | 0.261 | 0.477 |
| F*-VAR-Par | 0.092 | 0.074 | 0.070 | 0.072 | 0.072 | 0.085 | 0.160 | 0.328 |
| $\chi^2$-VAR-Par | 0.439 | 0.227 | 0.113 | 0.119 | 0.230 | 0.424 | 0.557 | 0.725 |
| F*-VAR-QS | 0.097 | 0.080 | 0.071 | 0.074 | 0.081 | 0.103 | 0.160 | 0.328 |
| $\chi^2$-VAR-QS | 0.358 | 0.149 | 0.101 | 0.105 | 0.154 | 0.295 | 0.550 | 0.724 |
| F*-Parzen | 0.111 | 0.085 | 0.066 | 0.068 | 0.084 | 0.110 | 0.162 | 0.341 |
| $\chi^2$-Parzen | 0.253 | 0.134 | 0.093 | 0.093 | 0.135 | 0.214 | 0.400 | 0.689 |
| F*-QS | 0.116 | 0.086 | 0.066 | 0.068 | 0.085 | 0.112 | 0.173 | 0.414 |
| $\chi^2$-QS | 0.246 | 0.131 | 0.088 | 0.090 | 0.131 | 0.207 | 0.385 | 0.678 |

See note to Table 1
Figure 4: Size-adjusted power of the different $F^*$ tests under the AR case with sample size $T = 100$ and number of restrictions $q = 1$. 
Figure 5: Size-adjusted power of the different $F^*$ tests under the AR case with sample size $T = 100$ and number of restrictions $q = 2$. 
Figure 6: Size-adjusted power of the different $F^*$ tests under the AR case with sample size $T = 100$ and number of restrictions $q = 3$. 
8 Appendix

Proof of Proposition 1. Note that the Yule-Walker estimators $\hat{A}_1, \ldots, \hat{A}_p$ and $\hat{\Sigma}_e$ satisfy the equation system:

$$\hat{B} \hat{\Gamma}_H (p + 1) = \hat{C},$$

where

$$\hat{B} = \left( \begin{array}{ccc} \mathbb{I}_q & -\hat{A}_1 & \ldots, & -\hat{A}_p \end{array} \right), \hat{C} = \left( \begin{array}{ccc} \hat{\Sigma}_e & 0 & \ldots, & 0 \end{array} \right).$$

Let $\lambda$ be an eigenvalue of $\hat{A}'$ and $x = (x'_1, \ldots, x'_p)'$ be the corresponding eigenvector. Then

$$\hat{A}'_1 x_1 + x_2 = \lambda x_1,$$

$$\hat{A}'_2 x_1 + x_3 = \lambda x_2,$$

$$\ldots$$

$$\hat{A}'_{p-1} x_1 + x_p = \lambda x_{p-1},$$

$$\hat{A}'_p x_1 = \lambda x_p.$$

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

$$-\hat{B}' x_1 + \begin{pmatrix} x \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} 0 \\ x \end{pmatrix}. \quad (25)$$

We consider the case $\lambda \neq 0$. In this case, $\hat{B}' x_1 \neq 0$. It follows from (25) and the Toeplitz structure of $\hat{\Gamma}_H (p + 1)$ that

$$x^* \hat{\Gamma}_H (p) x$$

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

$$= x_1^* \hat{B} \hat{\Gamma}_H (p + 1) \hat{B}' x_1 + ||\lambda||^2 x^* \hat{\Gamma}_H (p) x + \lambda x_1^* \hat{B} \hat{\Gamma}_H (p + 1) \begin{pmatrix} 0 \\ x \end{pmatrix} + \lambda \begin{pmatrix} 0 \\ x \end{pmatrix}^* \hat{\Gamma}_H (p + 1) \hat{B}' x_1$$

$$= x_1^* \hat{B} \hat{\Gamma}_H (p + 1) \hat{B}' x_1 + ||\lambda||^2 x^* \hat{\Gamma}_H (p) x + \lambda x_1 \hat{C}_x \begin{pmatrix} 0 \\ x \end{pmatrix} + \lambda \begin{pmatrix} 0 \\ x \end{pmatrix}^* \hat{C}' x_1$$

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

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

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

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

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

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

29
is well-defined almost surely. As a result,
\[ \hat{\nu}_R = \sum_{j=-p}^{p} \Gamma_{h,j} \left( \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e \right) + \sum_{|j|>p} \Gamma_{h,j} \left( \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e \right) < \infty \]
almost surely. That is, \( \sum_{|j|>p} \Gamma_{h,j} \left( \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e \right) = o(1) \) almost surely. As a direct implication,
\[ M_R = \sum_{|j|>p} \Gamma_{h,j} \left( \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_e \right) = o_p(1). \]

Define \( S_t = \sum_{j=1}^T h_j, S_0 = 0 \). It is easy to show that
\[ \hat{\nu}_R = \frac{1}{T} \sum_{t=1}^{T-1} \sum_{\tau=1}^{T-1} h_t h'_t k_{\text{rect}} \left( \frac{t-\tau}{bT} \right) = \frac{1}{T} \sum_{t=1}^{T-1} \sum_{\tau=1}^{T-1} D(t, \tau) S_t S'_\tau, \]
where
\[ D(t, \tau) = k_{\text{rect}} \left( \frac{t-\tau}{bT} \right) - k_{\text{rect}} \left( \frac{t+1-\tau}{bT} \right) - k_{\text{rect}} \left( \frac{t-\tau-1}{bT} \right) + k_{\text{rect}} \left( \frac{t-\tau}{bT} \right). \]
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
\[ \hat{\nu}_R = T^{-1} \sum_{\tau=1}^{T-bT} S_{\tau+bT} S'_\tau - T^{-1} \sum_{\tau=1}^{T-bT} S_{\tau+bT+1} S'_\tau, \]
\[ + T^{-1} \sum_{t=1}^{T-bT-1} S_t S'_{t+bT} - T^{-1} \sum_{t=1}^{T-bT-1} S_t S'_{t+bT+1} \]
\[ = -T^{-1} \sum_{\tau=1}^{T-bT-1} h_{\tau+bT+1} S'_\tau - T^{-1} \sum_{t=1}^{T-bT-1} S_t h'_{t+bT+1}. \]
To establish the limiting distribution of \( T^{-1} \sum_{\tau=1}^{T-bT-1} h_{\tau+bT+1} S'_\tau \), we write
\[ h_t = \hat{R}_T \left( \hat{G}'_T W_T \hat{G}_T \right)^{-1} \hat{G}'_T W_T \left[ f(v_t, \theta_0) + \frac{\partial f(v_t, \hat{\theta}_T)}{\partial \theta} \left( \hat{\theta}_T - \theta_0 \right) \right] \]
\[ = \hat{R}_T \left( \hat{G}'_T W_T \hat{G}_T \right)^{-1} \hat{G}'_T W_T \left[ f(v_t, \theta_0) - \frac{\partial f(v_t, \hat{\theta}_T)}{\partial \theta} \left( \hat{G}'_T W_T \hat{G}_T \right)^{-1} \hat{G}'_T W_T g_t \right], \]
where \( \hat{G}_T = G_T \left( \hat{\theta}_T \right) \) and \( \hat{\theta}_T \) satisfy \( \hat{\theta}_T = \theta_0 + O_p(1/\sqrt{T}) \) and \( \hat{\theta}_T = \theta_0 + O_p(1/\sqrt{T}) \). So
\[ S_\tau = \hat{R}_T \left( \hat{G}'_T W_T \hat{G}_T \right)^{-1} \hat{G}'_T W_T \left[ T g_\tau - T \hat{G}_T \left( \hat{G}'_T W_T \hat{G}_T \right)^{-1} \hat{G}'_T W_T g_T \right], \]
where \( \hat{G}_T = G_T \left( \hat{\theta}_T \right) \). As a result,
\[ T^{-1} \sum_{\tau=1}^{T-bT-1} h_{\tau+bT+1} S'_\tau = \hat{R}_T \left( \hat{G}'_T W_T \hat{G}_T \right)^{-1} \left( I_1 + I_2 + I_3 + I_4 \right) \left( \hat{G}'_T W_T \hat{G}_T \right)^{-1} \hat{R}_T, \]
where

\[ I_1 = \sum_{\tau=1}^{T-bT-1} \hat{G}_T'W_T f(v_{r+bT+1}, \theta_0) g'_T W_T \hat{G}_T, \]

\[ I_2 = - \sum_{\tau=1}^{T-bT-1} \hat{G}_T'W_T f(v_{r+bT+1}, \theta_0) g'_T W_T G_T \left( \hat{G}_T'W_T \hat{G}_T \right)^{-1} \hat{G}_T' W_T \hat{G}_T, \]

\[ I_3 = - \sum_{\tau=1}^{T-bT-1} \hat{G}_T'W_T \frac{\partial f(v_{r+bT+1}, \theta_T)}{\partial \theta'} \left( \hat{G}_T'W_T \hat{G}_T \right)^{-1} (G_T'W_T g_T)(g'_T W_T \hat{G}_T), \]

\[ I_4 = \sum_{\tau=1}^{T-bT-1} \hat{G}_T'W_T \frac{\partial f(v_{r+bT+1}, \theta_T)}{\partial \theta'} \left( \hat{G}_T'W_T \hat{G}_T \right)^{-1} G_T'W_T g_T g'_T W_T G_T \left( \hat{G}_T'W_T \hat{G}_T \right)^{-1} \hat{G}_T' W_T \hat{G}_T. \]

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

\[ I_1 \Rightarrow G' \mathcal{W} \Lambda \int_0^{1-b} dW_m(b + r) W'_m(r) \Lambda'WG. \]

For \( I_2 \), we have, by Assumptions 4-3:

\[ I_2 \Rightarrow -G' \mathcal{W} \Lambda \int_0^{1-b} dW_m(b + r) W'_m(1) \Lambda'WG. \]

For \( I_3 \) and \( I_4 \), we have

\[ I_3 \Rightarrow -G' \mathcal{W} \Lambda \left[ W_m(1) \int_0^{1-b} W_m(r) dr \right] \Lambda'WG; \]
and

\[ I_4 = \sum_{T^{-bT-1}} G'W \frac{\partial f(v_{T+bT+1, \theta_T})}{\partial \theta_T} (G'WG)^{-1} G'WG_{B \cdot}WG (G'WG)^{-1} \tilde{G}'WG (1 + o_p(1)) \]

\[ = T \sum_{\tau = 1}^{T^{-bT-1}} G'W (\tilde{G}_{\tau+bT} - \tilde{G}_{\tau+bT}) (G'WG)^{-1} G'WG_{B \cdot}WG \frac{T}{T} (1 + o_p(1)) \]

\[ = \sum_{\tau = 1}^{T^{-bT-1}} G'WG_{B \cdot}WG (1 + o_p(1)) \Rightarrow \left( \int_0^{1-b} r dr \right) G'WA \cdot W_m(1) W_m'(1) \Lambda'WG \]

\[ = \frac{1}{2} (b - 1)^2 G'WA \cdot W_m(1) W_m'(1) \Lambda'WG \]

Hence,

\[ I_1 + I_2 + I_3 + I_4 \]

\[ \Rightarrow G'WA \left[ \int_0^{1-b} dW_m(b + r) W_m'(r) - \int_0^{1-b} dW_m(b + r) r W_m'(1) \right] \Lambda'WG \]

\[ = G'WA \left[ \int_0^{1-b} dW_m(b + r) V_m'(r) \right] \Lambda'WG - G'WA \left[ \int_0^{1-b} W_m(1) W_m(r) dr - \frac{1}{2} (b - 1)^2 W_m(1) W_m'(1) \right] \Lambda'WG \]

\[ = G'WA \left[ \int_0^{1-b} dW_m(b + r) V_m'(r) \right] - \int_0^{1-b} dr W_m(1) V_m'(r) \right] \Lambda'WG \]

\[ = G'WA \left( \int_0^{1-b} dV_m(b + r) V_m'(r) \right) \Lambda'WG, \]

where

\[ V_m(r) = W_m(r) - r W_m(1). \]

Now

\[ \bar{V}_R \Rightarrow R (G'WG)^{-1} G'WAQ_m(b) \Lambda'WG (G'WG)^{-1} R', \]

where

\[ Q_m(b) = - \left( \int_0^{1-b} dV_m(b + r) V_m'(r) + \int_0^{1-b} V_m(r) dV_m(r + b) \right). \]
**Proof of Lemma 2.** (a) It follows from equation (26) that

\[- \int_0^{1-b} dV_m(b + r)W_m'(r) \]

\[
= - \int_0^{1-b} dW_m(b + r)W_m'(r) + \int_0^{1-b} dW_m(b + r)rW_m'(1)
\]

\[
+ \int_0^{1-b} W_m(1)W_m(r)'dr - \frac{1}{2} (b - 1)^2 W_m(1)W_m'(1)
\]

\[
= - \int_b^1 \left[ dW_m(s) \int_0^{s-b} dW_m'(r) \right] + \int_b^1 \left[ \int_b^1 (s - b) dW_m(s) \right] dW_m'(r)
\]

\[
+ \int_0^{1-b} W_m(1)W_m(r)'dr - \int_0^1 \int_0^1 \frac{1}{2} (b - 1)^2 dW_m(s)dW_m'(r)
\]

But

\[
\int_0^{1-b} W_m(1)W_m(r)'dr
\]

\[
= W_m(1)W_m(r)'|_0^{1-b} - W_m(1) \int_0^{1-b} r dW_m'(r)
\]

\[
= (1 - b) W_m(1)W_m(1 - b)' - W_m(1) \int_0^{1-b} r dW_m'(r)
\]

\[
= (1 - b) \left[ \int_0^1 dW_m(s) \right] dW_m'(r) - \int_0^1 \left[ \int_0^1 dW_m(s) \right] dW_m'(r)
\]

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

so

\[
- \int_0^{1-b} dV_m(b + r)W_m'(r)
\]

\[
= - \int_b^1 \left[ dW_m(s) \int_0^{s-b} dW_m'(r) \right] + \int_0^1 \left[ \int_b^1 (s - b) dW_m(s) \right] dW_m'(r)
\]

\[
- \int_0^{1-b} (b + r - 1) \left[ \int_0^1 dW_m(s) \right] dW_m'(r) - \int_0^1 \int_0^1 \frac{1}{2} (b - 1)^2 dW_m(s)dW_m'(r)
\]

or

\[
- \int_0^{1-b} dV_m(b + r)W_m'(r)
\]

\[
= \int_0^1 \int_0^1 - \{r \in [0, s - b], s \in [0, b, 1]\} + \{r \in [0, 1], s \in [0, 1]\} (s - b)
\]

\[
+ (1 - b - r) \{r \in [0, 1 - b], s \in [0, 1]\} - \frac{1}{2} (b - 1)^2 \{r \in [0, 1], s \in [0, 1]\} dW_m(s)dW_m'(r)
\]

\[
= \int_0^1 \int_0^1 \tilde{k}_b(r, s) dW_m(s)dW_m'(r)
\]

\[
= 33
\]
where \{\cdot\} is the indicator function 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

\[- \int_0^{1-b} V_m(r) dV_m(r + b) = \left( - \int_0^{1-b} dV_m(b + r)V_m'(r) \right)'
\]

\[= \int_0^1 \int_0^1 \tilde{k}_b(r, s) dW_m(r) dW_m'(s) = \int_0^1 \int_0^1 \tilde{k}_b(s, r) dW_m(s) dW_m'(r).
\]

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 (8).

(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 \(var(vec(A))\):

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

\[ \text{var} \left[ \int_0^1 \int_0^1 k_b^*(r, s) \left[ dW^m(r) \otimes dW^m(s) \right] \right] \\
= E \left[ \int_0^1 \int_0^1 k_b^*(r, s) k_b^*(t, \tau) \left[ dW^m(r) \otimes dW^m(s) \right] \left[ dW^m(t) \otimes dW^m(\tau) \right] \right] \\
- \left( \int_0^1 k_b^*(r, r) dr \right)^2 \text{vec} (\mathbb{I}_m) \text{vec} (\mathbb{I}_m)' \\
= E \left[ \int_0^1 \int_0^1 k_b^*(t, \tau) k_b^*(t, \tau) \left[ dW^m(t) dW^m(t) \otimes dW^m(\tau) dW^m(\tau) \right] \right] \\
+ E \left[ \int_0^1 \int_0^1 k_b^*(\tau, t) k_b^*(\tau, t) \left[ dW^m(\tau) dW^m(\tau) \otimes dW^m(t) dW^m(t) \right] \right] K_{m2} \\
= \left( \int_0^1 \int_0^1 [k_b^*(t, \tau)]^2 dt d\tau \right) (\mathbb{I}_{m2} + \mathbb{K}_{m2}). \]

Consequently,

\[ \text{var} \left( \text{vec} \left( \Lambda Q^m (b) \Lambda' \right) \right) = \mu_2 (\Omega \otimes \Omega) (\mathbb{I}_{m2} + \mathbb{K}_{m2}), \]

where \( \mu_2 = \int_0^1 \int_0^1 [k_b^*(t, \tau)]^2 dt d\tau. \)

Some algebra shows that

\[ \mu_2 = \left( \int_0^1 \int_0^1 k_{\text{rect}, b}(r - s) dr ds \right)^2 + \int_0^1 \int_0^1 k_{\text{rect}, b}^2(r - s) dr ds \\
- 2 \int_0^1 \int_0^1 \int_0^1 k_{\text{rect}, b}(r - p) k_{\text{rect}, b}(r - q) dr dp dq. \]

Now

\[ \int_0^1 \int_0^1 k_{\text{rect}, b}(r - s) dr ds = \int_0^1 \int_0^1 k_{\text{rect}, b}^2(r - s) dr ds = -b(b - 2), \]

and

\[ \int_0^1 \int_0^1 \int_0^1 k_{\text{rect}, b}(r - p) k_{\text{rect}, b}(r - q) dr dp dq \\
= \int_0^1 \int_0^1 \int_0^1 \left\{ \frac{|r - p|}{b} \leq 1 \right\} \left\{ \frac{|r - q|}{b} \leq 1 \right\} dr dp dq \\
= \int_0^1 \int_0^1 \left[ \int_0^1 \{ -b + p \leq r \leq b + p \} \{ -b + q \leq r \leq b + q \} dr \right] dp dq \\
= 2 \int_0^1 \int_0^1 \left[ \{ p < q < 2b + p \} \int_0^1 \{ -b + q \leq r \leq b + q \} dr \right] dp dq \\
= 2 \int_0^1 \int_0^1 \{ p < q < 2b + p \} \min(b + p, 1) dp dq \\
- 2 \int_0^1 \int_0^1 \{ p < q < 2b + p \} \max(0, -b + q) dp dq. \]
We compute the above two integrals in turn:

\[
2 \int_0^1 \int_0^1 \{ p < q < 2b + p \} \min(b + p, 1) dp dq \\
= 2 \int_0^{1-b} \left[ \int_0^1 \{ p < q < 2b + p \} dq \right] (b + p) dp + 2 \int_{1-b}^1 \left( \int_0^1 \{ p < q < 2b + p \} dq \right) dp \\
= 2 \int_0^{1-b} (\min(2b + p, 1) - p) (b + p) dp + 2 \int_{1-b}^1 (\min(2b + p, 1) - p) dp \\
= \left\{ b < \frac{1}{2} \right\} \left\{ 2 \int_0^{1-2b} 2b (b + p) dp + 2 \int_{1-2b}^{1-b} (1 - p) (b + p) dp + 2 \int_{1-b}^1 (1 - p) dp \right\} \\
+ \left\{ b \geq \frac{1}{2} \right\} \left\{ 2 \int_0^{1-b} (1 - p) (b + p) dp + 2 \int_{1-b}^1 (1 - p) dp \right\} \\
= - \left\{ b < \frac{1}{2} \right\} \left\{ \frac{1}{3} b (5b^2 - 6) + \left\{ b \geq \frac{1}{2} \right\} \left\{ \frac{1}{3} b^3 + b + \frac{1}{3} \right\} \right\}.
\]

\[
2 \int_0^1 \int_0^1 \{ p < q < 2b + p \} \max(0, -b + q) dp dq \\
= 2 \int_0^1 \int_0^1 \{ p < q < 2b + p \} \{ b < q \} (q - b) dp dq \\
= 2 \int_0^1 \int_0^1 \{ -b + p < q - b < b + p \} \{ 0 < q - b \} (q - b) dp dq \\
= 2 \int_0^1 \left[ \int_0^{1-b} 1 \{ -b + p < y < b + p \} y dy \right] dp \\
= 2 \int_0^b \left[ \int_0^{1-b} 1 \{ -b + p < y < b + p \} y dy \right] dp + 2 \int_{b}^1 \left[ \int_0^{1-b} 1 \{ -b + p < y < b + p \} y dy \right] dp \\
= 2 \int_0^b \left[ \int_0^{1-b} 1 \{ y < b + p \} y dy \right] dp + 2 \int_{b}^1 \left[ \int_{-b+p}^{1-b} 1 \{ y < b + p \} y dy \right] dp
\]

When \( b > 1/2 \),

\[
2 \int_0^1 \int_0^1 \{ p < q < 2b + p \} \max(0, -b + q) dp dq \\
= \left[ 2 \int_0^b \left( \int_0^{1-b} y dy \right) dp + 2 \int_{b}^1 \left( \int_{-b+p}^{1-b} y dy \right) dp \right] \\
= \frac{1}{3} (b - 1)^2 (b + 2).
\]
When \( b \leq 1/2 \),
\[
2 \int_0^1 \int_0^1 \{ p < q < 2b + p \} \max(0, -b + q) dp dq \\
= 2 \int_0^b \left[ \int_0^{\min(1-b,b+p)} ydy \right] dp + 2 \int_b^1 \left[ \int_{-b+p}^{\min(1-b,b+p)} ydy \right] dp \\
= 2 \int_0^b \{ p < 1 - 2b \} \int_0^{b+p} ydy + 2 \int_b^1 \{ p > 1 - 2b \} \int_0^{1-b} ydy \\
+ 2 \int_b^1 \{ p < 1 - 2b \} \int_{-b+p}^{b+p} ydy + 2 \int_b^1 \{ p > 1 - 2b \} \int_{-b+p}^{1-b} ydy \\
= \left\{ b > \frac{1}{3} \right\} \left\{ \int_0^{1-2b} (\int_0^{b+p} ydy) dp + \{ b \leq \frac{1}{3} \} \left\{ \int_0^{b} (\int_0^{b+p} ydy) dp \\
+ \left\{ b _{\frac{1}{3}} \right\} \left\{ \int_0^{1-2b} (\int_{-b+p}^{1-b} ydy) dp + \left\{ b \leq \frac{1}{3} \right\} \left\{ \int_0^{1} (\int_{-b+p}^{1-b} ydy) dp \\
+ 2 \left\{ b \leq \frac{1}{3} \right\} \left\{ \int_0^{b+p} ydy \right\} + \int_{-b+p}^{1-2b} (\int_{-b+p}^{1-b} ydy) dp + \int_0^{1-b} (\int_{-b+p}^{1-b} ydy) dp \right\} \\
= \frac{1}{3} b (5b^2 - 12b + 6) \\
\right. \\
So when \( b \leq 1/2 \),
\[
\mu_2 = (-b (b - 2))^2 - b (b - 2) + 2 \left[ b \right] (5b^2 - 6) + \frac{2}{3} b (5b^2 - 12b + 6) \\
= \frac{1}{3} b (3b^3 + 8b^2 - 15b + 6) .
\]
When \( b > 1/2 \)
\[
\mu_2 = (-b (b - 2))^2 - b (b - 2) - 2 \left( -\frac{1}{3} b^3 + b + \frac{1}{3} \right) + \frac{2}{3} (b - 1b)^2 (b + 2) \\
= \frac{1}{3} (b - 1)^2 (3b^2 - 2b + 2) .
\]
As a result
\[
\text{vec} \left( \Lambda Q_m (b) \Lambda' \right) = \mu_2 (\Omega \otimes \Omega) (I_{m^2} + K_{m^2}) .
\]
(c) Part (c) follows directly from part (b). Details are omitted here.

Proof of Theorem 2. Since
\[
\hat{V}_R \Rightarrow R \left( G'WG \right)^{-1} G'WQA_m (b) \Lambda' \Lambda WG \left( G'WG \right)^{-1} R' \\
\sqrt{\text{Tr}(\hat{\theta}_T)} \Rightarrow -R \left( G'WG \right)^{-1} G'WAW_m (1) ,
\]
we have

\begin{align*}
F_T & \Rightarrow \left[ R (G'WG)^{-1} G'WAW_m (1) \right]' \\
& \times \left[ R (G'WG)^{-1} G'WA \int_0^1 \int_0^1 k^*_b (r, s) dW_m (s) dW_m (r) A'WG (G'WG)^{-1} R' \right]^{-1} \\
& \times \left[ R (G'WG)^{-1} G'WAW_m (1) \right] / q.
\end{align*}

Let

\[ R (G'WG)^{-1} G'WAW_m (1) \overset{d}{=} DW_q (r) \]

for a \( q \times q \) matrix \( D \) such that

\[ DD' = R (G'WG)^{-1} G'WAA'WG (G'WG)^{-1} R'. \]

Then

\begin{align*}
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
\end{align*}

as desired. \( \blacksquare \)

**Proof of Theorem 3.** Let \( \mathcal{H} \) be an orthonormal matrix such that \( \mathcal{H} = (\eta/ \|\eta\|, \Pi)' \) where \( \Pi \) is a \( q \times (q - 1) \) matrix, then

\begin{align*}
F^{*}_x (q, b) & = \frac{\mu_1 (K - q + 1)}{Kq} \|\eta\|^2 (\mathcal{H} \eta)' \left( \sum_{n=1}^{\infty} \lambda_n \mathcal{H} \zeta_n (\mathcal{H} \zeta_n)' \right)^{-1} (\mathcal{H} \eta) \\
& = \frac{\mu_1 (K - q + 1)}{Kq} \|\eta\|^2 e_1' \left( \sum_{n=1}^{\infty} \lambda_n \mathcal{H} \zeta_n (\mathcal{H} \zeta_n)' \right)^{-1} e_1,
\end{align*}

where \( e_1 = (1, 0, 0, \ldots, 0, 0)' \). Note that \( \|\eta\|^2 \) is independent of \( \mathcal{H} \) and that \( \mathcal{H} W_q (r) \) has the same distribution as \( W_q (r) \), so we can write

\begin{align*}
F^{*}_x (q, b) & \overset{d}{=} \frac{\mu_1 (K - q + 1)}{K} \|\eta\|^2 q e_1' \left( \sum_{n=1}^{\infty} \lambda_n \zeta_n c_n' \right)^{-1} e_1.
\end{align*}

As a result,

\begin{align*}
P (F^{*}_x (q, b) < z) & = P \left( \frac{(K - q + 1)}{K} \|\eta\|^2 q e_1' \left( \frac{1}{\mu_1} \sum_{n=1}^{\infty} \lambda_n \zeta_n c_n' \right)^{-1} e_1 < z \right) \\
& = E q \left( \frac{K}{K - q + 1} \left[ e_1' \left( \frac{1}{\mu_1} \sum_{n=1}^{\infty} \lambda_n \zeta_n c_n' \right)^{-1} e_1 \right]^{-1} \right),
\end{align*}

38
where \( G_q \) is the CDF of the random variable \( \chi_q^2/q \).

Let

\[
\frac{1}{\mu_1} \sum_{n=1}^{\infty} \lambda_n \zeta_n \zeta'_n = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}.
\]

Then

\[
\left[ e_1' \left( \frac{1}{\mu_1} \sum_{n=1}^{\infty} \lambda_n \zeta_n \zeta'_n \right)^{-1} e_1 \right]^{-1} = \phi_{11} - \phi_{12} \phi_{22}^{-1} \phi_{21} := \phi_{11,2},
\]

and so

\[
P(F_{\infty}(q, b) < z) = EG_q \left( \frac{z}{K - q + 1} \phi_{11,2} \right). \tag{27}
\]

Let

\[
\frac{1}{K} \sum_{n=1}^{K} \zeta_n \zeta'_n = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix}.
\]

Using the same argument, we have

\[
P(F_{q,K-q+1} \leq z) = E \left[ \frac{(K - q + 1) \eta'}{K} \left( \frac{1}{K} \sum_{n=1}^{K} \zeta_n \zeta'_n \right)^{-1} \eta/q < z \right] = EG_q \left( \frac{z}{K - q + 1} \psi_{11,2} \right), \tag{28}
\]

where \( \psi_{11,2} = \psi_{11} - \psi_{12} \psi_{22}^{-1} \psi_{21} \).

In view of (27) and (28), we have written \( P(F_{\infty}(q, b) < z) \) and \( P(F_{q,K-q+1} \leq z) \) in the same format. It suffices to show the distribution of \( \phi_{11,2} \) is close to that of \( \psi_{11,2} \). Note that \( \phi_{11,2} \) and \( \psi_{11,2} \) can be written in a unified way if we define

\[
\sum_{n=1}^{\infty} \lambda_n^* \zeta_n \zeta'_n = \begin{pmatrix} \nu_{11} & \nu_{12} \\ \nu_{21} & \nu_{22} \end{pmatrix} \quad \text{and} \quad \nu_{11,2} = \nu_{11} - \nu_{12} \nu_{22}^{-1} \nu_{21}
\]

for some \( \lambda_n^* \) satisfying \( \sum_{n=1}^{\infty} \lambda_n = 1 \), \( \sum_{n=1}^{\infty} (\lambda_n^*)^2 = O(b) \) and \( \sum_{n=1}^{\infty} (\lambda_n^*)^m = O(b^{m-1}) \).

Then \( \phi_{11,2} = \nu_{11,2} \) if \( \lambda_n = \lambda_n^* / \mu_1 \) and \( \psi_{11,2} = \nu_{11,2} \) if \( \lambda_n^* = K^{-1} \{1 \leq n \leq K\} \).

Let \( \zeta'_n = (\zeta'_n, \eta'_n) \) where \( \eta_n \in \mathbb{R}^{q-1} \). Following the same argument in Sun (2010d), we can show that

\[
E(\nu_{11} - \nu_{12} \nu_{22}^{-1} \nu_{21}) = E \sum_{n=1}^{\infty} \lambda_n^* \zeta_{n,1} \zeta'_{n,1} = -E \sum_{n=1}^{\infty} \lambda_n^* \zeta_{n,1} \eta'_n \sum_{n=1}^{\infty} \lambda_n^* \eta_n \eta'_n \sum_{n=1}^{\infty} \lambda_n^* \eta_n \zeta_{n,1} \eta'_n
\]

\[
= 1 - E \text{tr} \left( \sum_{n=1}^{\infty} \lambda_n^* \eta_n \eta'_n \right)^{-1} \left( \sum_{n=1}^{\infty} \lambda_n^* \eta_n \eta'_n \right)^{-1} \left( \sum_{n=1}^{\infty} \lambda_n^* \eta_n \zeta_{n,1} \eta'_n \right)^{-1}
\]

\[
= 1 - E \left[ \text{tr} \left( \sum_{n=1}^{\infty} \lambda_n^* \eta_{-1} \right)^{-1} + O_p(b) \right] \left( \sum_{n=1}^{\infty} \lambda_n^* \right)^2 \eta_n \eta'_n
\]

\[
= 1 - \left( \sum_{n=1}^{\infty} \lambda_n^* \right)^{-1} \sum_{n=1}^{\infty} \lambda_n^* \left( q - 1 \right) + o(b).
\]
In addition,
\[ E \left( \nu_{11} - \nu_{12} \nu_{22}^{-1} \nu_{21} \right)^2 = E \nu_{11}^2 + E \nu_{12} \nu_{22}^{-1} \nu_{21} \nu_{12} \nu_{22}^{-1} \nu_{21} - 2E \nu_{11} \nu_{12} \nu_{22}^{-1} \nu_{21}, \]
where
\[ E \nu_{11}^2 = \left( \sum_{n=1}^{\infty} \lambda_n^* \right)^2 + 2 \sum_{n=1}^{\infty} (\lambda_n^*)^2, \]

\[ E \nu_{11} \nu_{12} \nu_{22}^{-1} \nu_{21} = \left( \sum_{n=1}^{\infty} \lambda_n^* \right) \text{Etr} \left( \sum_{n=1}^{\infty} \lambda_n^* \eta_n \eta_n' \right)^{-1} \left( \sum_{n=1}^{\infty} (\lambda_n^*)^2 \eta_n \eta_n' \right) \]
\[ + 2 \text{Etr} \left( \sum_{n=1}^{\infty} \lambda_n^* \eta_n \eta_n' \right)^{-1} \left( \sum_{n=1}^{\infty} (\lambda_n^*)^3 \eta_n \eta_n' \right) \]
\[ = \sum_{n=1}^{\infty} (\lambda_n^*)^2 (q - 1) + o(b), \]
and
\[ E \nu_{12} \nu_{22}^{-1} \nu_{21} \nu_{12} \nu_{22}^{-1} \nu_{21} \]
\[ = \text{Etr} \left( \sum_{n=1}^{\infty} \lambda_n^* \eta_n \eta_n' \right)^{-1} \left[ \sum_{l=1}^{\infty} (\lambda_l^*)^2 \eta_l \eta_l' \right] \text{tr} \left( \sum_{n=1}^{\infty} \lambda_n^* \eta_n \eta_n' \right)^{-1} \left( \sum_{l=1}^{\infty} (\lambda_l^*)^2 \eta_l \eta_l' \right) \]
\[ + 2 \text{Etr} \left\{ \sum_{n=1}^{\infty} \lambda_n^* \eta_n \eta_n' \right\}^{-1} \left( \sum_{l=1}^{\infty} (\lambda_l^*)^2 \eta_l \eta_l' \right) \left[ \sum_{n=1}^{\infty} \lambda_n^* \eta_n \eta_n' \right]^{-1} \left( \sum_{l=1}^{\infty} (\lambda_l^*)^2 \eta_l \eta_l' \right) \]
\[ = o(b). \]

Hence
\[ E \left( \nu_{11} - \nu_{12} \nu_{22}^{-1} \nu_{21} \right)^2 = \left( \sum_{n=1}^{\infty} \lambda_n^* \right)^2 + 2 \sum_{n=1}^{\infty} (\lambda_n^*)^2 + \sum_{n=1}^{\infty} (\lambda_n^*)^2 (q - 1) + o(b). \]

We therefore have shown that the first and second moments of \( \nu_{11,2} \) depend only on \( \sum_{n=1}^{\infty} \lambda_n^* \) and \( \sum_{n=1}^{\infty} (\lambda_n^*)^2 \) up to smaller order \( o(b) \). Furthermore, we can show that higher-order moments of \( \nu_{11,2} \) are of order \( o(b) \). But \( \sum_{n=1}^{\infty} \lambda_n^* \) and \( \sum_{n=1}^{\infty} (\lambda_n^*)^2 \) are the same for \( \phi_{11,2} \) and \( \psi_{11,2} \). This result, combined with a Taylor expansion, yields
\[ P \left( F_{\infty}(q, b) < z \right) = P \left( F_{q, K+q+1} \leq z \right) + o(b). \]
References

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

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

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

[4] Bartlett, M.S. (1954): “A Note on the Multiplying Factors for Various $\chi^2$ Approximations.” *Journal of the Royal Statistical Society* B, 16, 296–298.

[5] Berk, K.N. (1974): “Consistent Autoregressive Spectral Estimates.” *The Annals of Statistics*, 2, 489–502.

[6] Bilodeau, M. and D. Brenner (1999): *Theory of Multivariate Statistics*. Springer, New York.

[7] Burnside, C. and M. Eichenbaum (1996): “Small-Sample Properties of GMM-Based Wald Tests.” *Journal of Business and Economic Statistics*, 14(3), 294–308.

[8] Cribari-Neto, F. and G.M. Cordeiro (1996): “On Bartlett and Bartlett-type Corrections.” *Econometric Reviews*, 15, 339–367.

[9] den Haan, W.J. and A. Levin (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] den Haan, W. and A. Levin (1998): “Vector Autoregressive Covariance Matrix Estimation.” Working paper. http://www1.feb.uva.nl/mint/wdenhaan/papers.htm

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

[12] Hall, P. (1983): “Chi-squared Approximations to the Distribution of a Sum of Independent Random Variables.” *Annals of Probability*, 11, 1028–1036.

[13] Hall, A.R. (1994): “Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection.” *Journal of Business and Economic Statistics*, 12, 461–70.

[14] Hansen, B. (1992): “Convergence to Stochastic Integrals for Dependent Heterogeneous Processes.” *Econometric Theory*, 8(4), 489–500.

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

[16] Hotelling, H. (1931): “The Generalization of Student’s Ratio.” *Annals of Mathematical Statistics*, 2, 360–378.
[17] Kiefer, N.M. and T.J. Vogelsang (2002a): “Heteroskedasticity-autocorrelation Robust Testing Using Bandwidth Equal to Sample Size.” *Econometric Theory*, 18, 1350–1366.

[18] ——— (2002b): “Heteroskedasticity-autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation.” *Econometrica*, 70, 2093–2095.

[19] ——— (2005): “A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests.” *Econometric Theory*, 21, 1130–1164.

[20] Kollo, T. and D. von Rosen (1995): “Approximating by the Wishart Distribution.” *Annals of the Institute of Statistical Mathematics*, 47(4), 767–783.

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

[22] Lee, C.C. and P.C.B. Phillips (1994): “An ARMA Prewhitened Long-run Variance Estimator.” Manuscript, Yale University.

[23] Lin, C.-C. and S. Sakata (2009): “On Long-Run Covariance Matrix Estimation with the Truncated Flat Kernel.” Working paper, Department of Economics, University of British Columbia.

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

[25] ——— (1994): “Automatic Lag Selection in Covariance Estimation.” *Review of Economic Studies*, 61, 631–654.

[26] Ng, S., and P. Perron (1995): “Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag.” *Journal of the American Statistical Association*, 90, 268–81.

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

[28] Phillips, P.C.B. and B. Hansen (1990): “Statistical Inference in Instrumental Variables Regression with I(1) Processes.” *Review of Economic Studies* 57, 99–125.

[29] Phillips, P. C. B. (2005): “HAC Estimation by Automated Regression.” *Econometric Theory*, 21, 116–142.

[30] 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.

[31] Politis, D.N. (2010): “Higher-order Accurate, Positive Semi-definite Estimation of Large-sample Covariance and Spectral Density Matrices.” *Econometric Theory*, forthcoming.

[32] Priestley, M.B. (1981): *Spectral Analysis and Time Series*, Academic Press, London and New York.
Ray, S. and N.E. Savin (2008): “The Performance of Heteroskedasticity and Autocorrelation Robust Tests: a Monte Carlo Study with an Application to the Three-factor Fama-French Asset-pricing Model.” *Journal of Applied Econometrics*, 23, 91–109.

Ray, S., N.E. Savin, and A. Tiwari (2009): “Testing the CAPM Revisited.” Working paper, Department of Finance, University of Iowa.

Reinsel, G.C. (1997): *Elements of Multivariate Time Series Analysis*, Springer, New York.

Said, S.E. and D. Dickey (1984): “Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order.” *Biometrika*, 71(3): 599–607.

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

Sun, Y. (2010a): “Autocorrelation Robust Inference Using Nonparametric Series Method and Testing-optimal Smoothing Parameter.” Working paper, Department of Economics, UC San Diego.

Sun, Y. (2010b): “Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter.” Working paper, Department of Economics, UC San Diego.

Sun, Y. (2010c): “Calibrating Optimal Smoothing Parameter in Heteroscedasticity and Autocorrelation Robust Tests.” Working paper, Department of Economics, UC San Diego.

Sun, Y. (2010d): “Let’s Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference.” Working paper, Department of Economics, UC San Diego.

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

Whittle, P. (1954): “The Statistical Analysis of a Seiche Record.” *Journal of Marine Research*, 13, 76–100.