Should We Go One Step Further?
An Accurate Comparison of One-Step and Two-Step Procedures
in a Generalized Method of Moments Framework

Jungbin Hwang and Yixiao Sun
Department of Economics,
University of California, San Diego

Preliminary. Please do not circulate.

Abstract

According to the conventional asymptotic theory, the two-step Generalized Method of Moments (GMM) estimator and test perform as least as well as the one-step estimator and test in large samples. The conventional asymptotics theory, as elegant and convenient as it is, completely ignores the estimation uncertainty in the weighting matrix, and as a result it may not reflect finite sample situations well. In this paper, we employ the fixed-smoothing asymptotic theory that accounts for the estimation uncertainty, and compare the performance of the one-step and two-step procedures in this more accurate asymptotic framework. We show the two-step procedure outperforms the one-step procedure only when the benefit of using the optimal weighting matrix outweighs the cost of estimating it. This qualitative message applies to both the asymptotic variance comparison and power comparison of the associated tests. A Monte Carlo study lends support to our asymptotic results.

JEL Classification: C12, C32

Keywords: Asymptotic Efficiency, Fixed-smoothing Asymptotics, Heteroskedasticity and Autocorrelation Robust, Increasing-smoothing Asymptotics, Asymptotic Mixed Normality, Non-standard Asymptotics, Two-step GMM Estimation

1 Introduction

Efficiency is one of the most important problems in statistics and econometrics. In the widely-used GMM framework, it is standard practice to employ a two-step procedure to improve the efficiency of the GMM estimator and the power of the associated tests. The two-step procedure requires the estimation of a weighting matrix. According to the Hansen (1982), the optimal weighting matrix is the asymptotic variance of the (scaled) sample moment conditions. For time series data, which is our focus here, the optimal weighting matrix is usually referred to as the

*Email: j6hwang@ucsd.edu, yisun@ucsd.edu. Correspondence to: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508.
long run variance (LRV) of the moment conditions. To be completely general, we often estimate the LRV using the nonparametric kernel or series method.

Under the conventional asymptotics, both the one-step and two-step GMM estimators are asymptotically normal. In general, the two-step GMM estimator has a smaller asymptotic variance. Statistical tests based on the two-step estimator are also asymptotically more powerful than those based on the one-step estimator. A driving force behind these results is that the two-step estimator and the associated tests have the same asymptotic properties as the corresponding ones when the optimal weighting matrix is known. However, given that the optimal weighting matrix is estimated nonparametrically in the time series setting, there is large estimation uncertainty. A good approximation to the distributions of the two-step estimator and the associated tests should reflect this relatively high estimation uncertainty.

One of the goals of this paper is to compare the asymptotic properties of the one-step and two-step procedures when the estimation uncertainty in the weighing matrix is accounted for. There are two ways to capture the estimation uncertainty. One is to use the high order conventional asymptotic theory under which the amount of nonparametric smoothing in the LRV estimator increases with the sample size but at a slower rate. While the estimation uncertainty vanishes in the first order asymptotics, we expect it to remain in high order asymptotics. The second way is to use an alternative asymptotic approximation that can capture the estimation uncertainty even with just a first-order asymptotics. To this end, we consider a limiting thought experiment in which the amount of nonparametric smoothing is held fixed as the sample size increases. This leads to the so-called fixed-smoothing asymptotics in the recent literature.

In this paper, we employ the fixed-smoothing asymptotics to compare the one-step and two-step procedures. For the one-step procedure, the LRV estimator is used in computing the standard errors, leading to the popular heteroskedasticity and autocorrelation robust (HAR) standard errors. See, for example, Newey and West (1987) and Andrews (1991). For the two-step procedure, the LRV estimator not only appears in the standard error estimation but also plays the role of the optimal weighting matrix in the second-step GMM criterion function. Under the fixed-smoothing asymptotics, the weighting matrix converges to a random matrix. As a result, the second-step GMM estimator is not asymptotically normal but rather asymptotically mixed normal. The asymptotic mixed normality reflects the estimation uncertainty of the GMM weighting matrix and is expected to be closer to the finite sample distribution of the second-step GMM estimator. In a recent paper, Sun (2014b) shows that both the one-step and two-step test statistics are asymptotically pivotal under this new asymptotic theory. So a nuisance-parameter-free comparison of the one-step and two-step tests is possible.

Comparing the one-step and two-step procedures under the new asymptotics is fundamentally different from that under the conventional asymptotics. Under the new asymptotics, the two-step procedure outperforms the one-step procedure only when the benefit of using the optimal weighting matrix outweighs the cost of estimating it. This qualitative message applies to both the asymptotic variance comparison and the power comparison of the associated tests. This is in sharp contrast with the conventional asymptotics where the cost of estimating the optimal weighting matrix is completely ignored. Since the new asymptotic approximation is more accurate than the conventional asymptotic approximation, comparing the two procedures under this new asymptotics will give an honest assessment of their relative merits. This is confirmed by a Monte

---

1In this paper, the one-step estimator refers to the first-step estimator in the typical two-step GMM framework. It does not refer to the continuous updating GMM estimator that involves only one step. We use the terms “one-step” and “first-step” interchangingly.
There is a large and growing literature on the fixed-smoothing asymptotics. For kernel LRV variance estimators, the fixed-smoothing asymptotics is the so-called the fixed-b asymptotics first studied by Kiefer, Vogelsang and Bunzel (2002) and Kiefer and Vogelsang (2002a, 2002b, 2005) in the econometrics literature. For other studies, see, for example, Jansson (2004), Sun, Phillips and Jin (2008), Sun and Phillips (2009), Gonçalves and Vogelsang (2011), and Zhang and Shao (2013) in the time series setting; Bester, Conley, Hansen and Vogelsang (2014) and Sun and Kim (2014) in the spatial setting; and Gonçalves (2011), Kim and Sun (2013), and Vogelsang (2012) in the panel data setting. For OS LRV variance estimators, the fixed-smoothing asymptotics is the so-called fixed-K asymptotics. For its theoretical development and related simulation evidence, see, for example, Phillips (2005), Müller (2007), and Sun (2011, 2013). The approximation approaches in some other papers can also be regarded as special cases of the fixed-smoothing asymptotics. This includes, among others, Ibragimov and Müller (2010), Shao (2010) and Bester, Conley, and Hansen (2011). The fixed-smoothing asymptotics can be regarded as a convenient device to obtain some high order terms under the conventional increasing-smoothing asymptotics.

The rest of the paper is organized as follows. The next section presents a simple overidentified GMM framework. Section 3 compares the two procedures from the perspective of point estimation. Section 4 compares them from the testing perspective. Section 5 extends the ideas to a general GMM framework. Section 6 reports simulation evidence on the finite sample performances of the two procedures. Proofs of the main theorems are provided in the Appendix.

A word on notation: for a symmetric matrix $A$, $A^{1/2}$ (or $A_{1/2}$) is a matrix square root of $A$ such that $A^{1/2} (A^{1/2})' = A$. Note that $A^{1/2}$ does not have to be symmetric. We will specify $A^{1/2}$ explicitly when it is not symmetric. If not specified, $A^{1/2}$ is a symmetric matrix square root of $A$ based on its eigendecomposition. By definition, $A^{-1/2} = (A^{1/2})^{-1}$ and so $A^{-1} = (A^{-1/2})' (A^{-1/2})$. For matrices $A$ and $B$, we use “$A \geq B$” to signify that $A - B$ is positive (semi)definite. We use “0” and “$O$” interchangeably to denote a matrix of zeros whose dimension may be different at different occurrences. For two random variables $X$ and $Y$, we use $X \perp Y$ to indicate that $X$ and $Y$ are independent.

# 2 A Simple Overidentified GMM Framework

To illustrate the basic ideas of this paper, we consider a simple overidentified time series model of the form:

\[
\begin{align*}
    y_{1t} &= \theta_0 + u_{1t}, \quad y_{1t} \in \mathbb{R}^d, \\
    y_{2t} &= u_{2t}, \quad y_{2t} \in \mathbb{R}^q
\end{align*}
\]

for $t = 1, \ldots, T$ where $\theta_0 \in \mathbb{R}^d$ is the parameter of interest and the vector process $u_t := (u'_{1t}, u'_{2t})'$ is stationary with mean zero. We allow $u_t$ to have autocorrelation of unknown forms so that the long run variance $\Omega$ of $u_t$:

\[
\Omega = \text{lrvar}(u_t) = \sum_{j=-\infty}^{\infty} Eu_t u'_{t-j}
\]
takes a general form. However, for simplicity, we assume that $\text{var}(u_t) = \sigma^2 I_d$ for the moment.

Our model is just a location model. We initially consider a general GMM framework but later find out that our points can be made more clearly in the simple location model. In fact, the simple location model may be regarded as a limiting experiment in a general GMM framework.

Embedding the location model in a GMM framework, the moment conditions are

$$E(y_t) - \begin{pmatrix} \theta_0 \\ 0_{q \times 1} \end{pmatrix} = 0$$

where $y_t = (y_{1t}', y_{2t}')'$. Let

$$g_T(\theta) = \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T y_{1t} - \theta \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T y_{2t} \end{pmatrix}.$$  

Then a GMM estimator of $\theta_0$ can be defined as

$$\hat{\theta}_{GMM} = \arg \min_{\theta} g_T(\theta)'W_T^{-1}g_T(\theta)$$

for some positive definite weighting matrix $W_T$. Writing

$$W_T = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix},$$

where $W_{11}$ is a $d \times d$ matrix and $W_{22}$ is a $q \times q$ matrix, then it is easy to show that

$$\hat{\theta}_{GMM} = \frac{1}{T} \sum_{t=1}^T (y_{1t} - \beta_W y_{2t}) \text{ for } \beta_W = W_{12}W_{22}^{-1}.$$  

There are at least two different choices of $W_T$. First, we can take $W_T$ to be the identity matrix $W_T = I_m$ for $m = d + q$. In this case, $\beta_W = 0$ and the GMM estimator $\hat{\theta}_{1T}$ is simply

$$\hat{\theta}_{1T} = \frac{1}{T} \sum_{t=1}^T y_{1t}.$$  

Second, we can take $W_T$ to be the ‘optimal’ weighting matrix $W_T = \Omega$. With this choice, we obtain the GMM estimator:

$$\tilde{\theta}_{2T} = \frac{1}{T} \sum_{t=1}^T (y_{1t} - \beta y_{2t}),$$  

where $\beta = \Omega_{12}\Omega_{22}^{-1}$ is the long run regression coefficient matrix. While $\hat{\theta}_{1T}$ completely ignores the information in $\{y_{2t}\}$, $\tilde{\theta}_{2T}$ takes advantage of this information.

---

If $\text{var}(u_t) = \mathbb{V} := \begin{pmatrix} \mathbb{V}_{11} & \mathbb{V}_{12} \\ \mathbb{V}_{21} & \mathbb{V}_{22} \end{pmatrix} \neq \sigma^2 I_d$

for any $\sigma^2 > 0$, we can let

$$\mathbb{V}_{1/2} = \begin{pmatrix} (\mathbb{V}_{12})^{1/2} & \mathbb{V}_{12}(\mathbb{V}_{22})^{-1/2} \\ 0 & (\mathbb{V}_{22})^{1/2} \end{pmatrix}.$$  

Then $\mathbb{V}_{1/2}^{-1}(y_{1t}', y_{2t}')'$ can be written as a location model whose error variance is the identity matrix $I_d$. The estimation uncertainty in estimating $\mathbb{V}$ will not affect our asymptotic results.
Under some moment and mixing conditions, we have
\[ \sqrt{T} \left( \hat{\theta}_{1T} - \theta_0 \right) \overset{d}{\rightarrow} N(0, \Omega_{11}) \text{ and } \sqrt{T} \left( \hat{\theta}_{2T} - \theta_0 \right) \overset{d}{\rightarrow} N(0, \Omega_{1,2}), \]
where
\[ \Omega_{1,2} = \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}. \]
So \( \text{Asym Var}(\hat{\theta}_{2T}) < \text{Asym Var}(\hat{\theta}_{1T}) \) unless \( \Omega_{12} = 0 \). This is a well known result in the literature. Since we do not know \( \Omega \) in practice, \( \hat{\theta}_{2T} \) is infeasible. However, given the feasible estimator \( \hat{\theta}_{1T} \), we can estimate \( \Omega \) and construct a feasible version of \( \hat{\theta}_{2T} \). The common two-step estimation strategy is as follows.

i) Estimate the long run covariance matrix by
\[
\hat{\Omega} := \hat{\Omega}(\hat{u}) = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} Q_h(s, t) \left( \hat{u}_{ts} - \frac{1}{T} \sum_{r=1}^{T} \hat{u}_{rt} \right) \left( \hat{u}_{st} - \frac{1}{T} \sum_{r=1}^{T} \hat{u}_{sr} \right)',
\]
where \( \hat{u}_{ts} = (y_{1t} - \hat{\theta}_{1T}, y_{2t})' \).

ii) Obtain the feasible two-step estimator \( \hat{\theta}_{2T} = T^{-1} \sum_{t=1}^{T} \left( y_{1t} - \hat{\beta} y_{2t} \right) \) where \( \hat{\beta} = \hat{\Omega}_{22}^{-1} \).

In the above definition of \( \hat{\Omega} \), \( Q_h(r, s) \) is a symmetric weighting function that depends on the smoothing parameter \( h \). For conventional kernel estimators, \( Q_h(r, s) = k((r-s)/b) \) and we take \( h = 1/b \). For the orthonormal series (OS) estimators, \( Q_h(r, s) = K^{-1} \sum_{j=1}^{K} \phi_j(r) \phi_j(s) \) and we take \( h = K \), where \( \{\phi_j(r)\} \) are orthonormal basis functions on \( L^2[0,1] \) satisfying \( \int_0^1 \phi_j(r) \, dr = 0 \).

We parametrize \( h \) in such a way so that \( h \) indicates the level or amount of smoothing for both types of LRV estimators.

Note that we use the demeaned process \( \{\hat{u}_{ts} - T^{-1} \sum_{r=1}^{T} \hat{u}_{rt}\} \) in constructing \( \hat{\Omega}(\hat{u}) \). For the location model, \( \hat{\Omega}(\hat{u}) \) is numerically identical to \( \hat{\Omega}(u) \) where the unknown error process \( \{u_t\} \) is used. The moment estimation uncertainty is reflected in the demeaning operation. Had we known the true value of \( \theta_0 \) and hence the true moment process \( \{u_t\} \), we would not need to demean \( \{u_t\} \).

While \( \hat{\theta}_{2T} \) is asymptotically more efficient than \( \hat{\theta}_{1T} \), is \( \hat{\theta}_{2T} \) necessarily more efficient than \( \hat{\theta}_{1T} \) and in what sense? Is the Wald test based on \( \hat{\theta}_{2T} \) necessarily more powerful than that based on \( \hat{\theta}_{1T} \)? One of the objectives of this paper is to address these questions.

### 3 A Tale of Two Asymptotics: Point Estimation

We first consider the conventional asymptotics where \( h \rightarrow \infty \) as \( T \rightarrow \infty \) but at a slower rate, i.e., \( h/T \rightarrow 0 \). The asymptotic direction is represented by the curved arrow in Figure 1. Sun (2014a, 2014b) calls this type of asymptotics the “Increasing-smoothing Asymptotics,” as \( h \) increases with the sample size. Under this type of asymptotics and some regularity conditions, we have \( \hat{\Omega} \rightarrow^p \Omega \). It can then be shown that \( \hat{\theta}_{2T} \) is asymptotically equivalent to \( \theta_{2T} \), i.e., \( \sqrt{T}(\theta_{2T} - \hat{\theta}_{2T}) = o_p(1) \).

As a direct consequence, we have
\[ \sqrt{T} \left( \hat{\theta}_{1T} - \theta_0 \right) \overset{d}{\rightarrow} N(0, \Omega_{11}), \sqrt{T} \left( \hat{\theta}_{2T} - \theta_0 \right) \overset{d}{\rightarrow} N \left[ 0, \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \right]. \]

So \( \hat{\theta}_{2T} \) is still asymptotically more efficient than \( \hat{\theta}_{1T} \).
The conventional asymptotics, as elegant and convenient as it is, does not reflect the finite sample situations well. Under this type of asymptotics, we essentially approximate the distribution of $\hat{\theta}$ by the degenerate distribution concentrating on $\hat{\theta}$. That is, we completely ignore the estimation uncertainty in $\hat{\theta}$. The degenerate approximation is too optimistic, as $\hat{\theta}$ is a nonparametric estimator, which by definition can have high variation in finite samples.

To obtain a more accurate distributional approximation of $\sqrt{T}(\hat{\theta}_T - \theta_0)$, we could develop a high order increasing-smoothing asymptotics that reflects the estimation uncertainty in $\hat{\theta}$. This is possible but requires strong assumptions that cannot be easily verified. In addition, it is also technically challenging and tedious to rigorously justify the high order asymptotic theory. Instead of high order asymptotic theory under the conventional asymptotics, we adopt the type of asymptotics that holds $h$ fixed as $T \rightarrow \infty$. This asymptotic behavior of $h$ and $T$ is illustrated by the arrow pointing to the right in Figure 1. Given that $h$ is fixed, we follow Sun (2014a, 2014b) and call this type of asymptotics the “Fixed-smoothing Asymptotics.” This type of asymptotics takes the sampling variability of $\hat{\theta}$ into consideration.

Sun (2013, 2014a) has shown that the fixed-smoothing asymptotic distribution is high order correct under the conventional increasing-smoothing asymptotics. So the fixed-smoothing asymptotics can be regarded as a convenient device to obtain some high order terms under the conventional increasing-smoothing asymptotics.

To establish the fixed-smoothing asymptotics, we maintain Assumption 1 on the kernel function and basis functions.
Assumption 1 (i) For kernel HAR variance estimators, the kernel function $k(\cdot)$ satisfies the following condition: for any $b \in (0,1]$ and $\rho \geq 1$, $k_b(x)$ and $k^\rho(x)$ are symmetric, continuous, piecewise monotonic, and piecewise continuously differentiable on $[-1,1]$. (ii) For the OS HAR variance estimator, the basis functions $\phi_j(\cdot)$ are piecewise monotonic, continuously differentiable and orthonormal in $L^2[0,1]$ and $\int_0^1 \phi_j(x) \, dx = 0$.

Assumption 1 on the kernel function is very mild. It includes many commonly used kernel functions such as the Bartlett kernel, Parzen kernel, and Quadratic Spectral (QS) kernel.

Define

$$Q^*_h(r,s) = Q_h(r,s) - \sum_{j=1}^{T} \sum_{t=1}^{T} Q_h^*(\frac{s}{T}, \frac{t}{T}) \hat{u}_t \hat{u}'_s.$$ 

which is a centered version of $Q_h(r,s)$, and

$$\hat{\Omega} = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} Q_h^*(\frac{s}{T}, \frac{t}{T}) \hat{u}_t \hat{u}'_s.$$ 

Assumption 1 ensures that $\hat{\Omega}$ and $\hat{\Omega}$ are asymptotically equivalent. Furthermore, under this assumption, Sun (2014a) shows that, for both kernel LRV and OS LRV estimation, the centered weighting function $Q^*_h(r,s)$ satisfies:

$$Q^*_h(r,s) = \sum_{j=1}^{\infty} \lambda_j \Phi_j(r) \Phi_j(s)$$

where $\{\Phi_j(r)\}$ is a sequence of continuously differentiable functions satisfying $\int_0^1 \Phi_j(r) \, dr = 0$ and the series on the right hand side converges to $Q^*_h(r,s)$ absolutely and uniformly over $(r,s) \in [0,1] \times [0,1]$. The representation can be regarded as a spectral decomposition of the compact Fredholm operator with kernel $Q^*_h(r,s)$. See Sun (2014a) for more discussion.

Now, letting $\Phi_0(\cdot) := 1$ and using the basis functions $\{\Phi_j(\cdot)\}_{j=1}^{\infty}$ in the series representation of the weighting function, we make the following assumptions.

Assumption 2 The vector process $\{u_t\}_{t=1}^{T}$ satisfies:

a) $T^{-1/2} \sum_{t=1}^{T} \Phi_j(t/T) u_t$ converges weakly to a continuous distribution, jointly over $j = 0, 1, ..., J$ for every fixed $J$;

b) For every fixed $J$ and $x \in \mathbb{R}^m$,

$$P \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j(t/T) u_t \leq x \right. \) =$$

$$P \left( \Omega_{1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Phi_j(t/T) e_t \leq x \right. \) + o(1) \text{ as } T \to \infty$$

where

$$\Omega_{1/2} = \begin{pmatrix} \Omega_{1/2}^{1/2} & \Omega_{12} \Omega_{22}^{-1/2} \\ 0 & \Omega_{22}^{1/2} \end{pmatrix} > 0$$

is a matrix square root of $\Omega = \sum_{j=-\infty}^{\infty} E u_t u'_t - j$ and $e_t \sim iid N(0, I_m)$.
Assumption 3 \[ \sum_{j=-\infty}^{\infty} \| Eu_t u'_{t-j} \| < \infty. \]

Proposition 1 Let Assumptions 1–3 hold. As \( T \to \infty \) for a fixed \( h \), we have:

(a) \( \hat{\Omega} \overset{d}{\to} \Omega_\infty \) where

\[
\Omega_\infty = \Omega_{1/2} \hat{\Omega}_\infty \Omega_{1/2} := \begin{pmatrix} \Omega_{\infty,11} & \Omega_{\infty,12} \\ \Omega_{\infty,21} & \Omega_{\infty,22} \end{pmatrix}
\]

\( \hat{\Omega}_\infty = \int_0^1 \int_0^1 Q_h(r,s) dB_m(r) dB_m(s)' := \begin{pmatrix} \hat{\Omega}_{\infty,11} & \hat{\Omega}_{\infty,12} \\ \hat{\Omega}_{\infty,21} & \hat{\Omega}_{\infty,22} \end{pmatrix} \)

and \( B_m(\cdot) \) is a standard Brownian motion of dimension \( m = d + q \);

(b) \( \sqrt{T} (\hat{\theta}_{2T} - \theta_0) \overset{d}{\to} \Omega_{1/2} B_m(1) \) where \( \beta_\infty = \beta_\infty(h,d,q) := \Omega_{\infty,12} \Omega_{\infty,22}^{-1} \) is independent of \( B_m(1) \).

Conditional on \( \beta_\infty \), the asymptotic distribution of \( \sqrt{T} (\hat{\theta}_{2T} - \theta_0) \) is a normal distribution with variance

\[
V_2 = \begin{pmatrix} I_d & -\beta_\infty \end{pmatrix} \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \begin{pmatrix} I_d \\ -\beta_\infty' \end{pmatrix} = \Omega_{11} - \Omega_{12} \beta_\infty' - \beta_\infty \Omega_{21} + \beta_\infty \Omega_{22} \beta_\infty'.
\]

That is, \( \sqrt{T}(\hat{\theta}_{2T} - \theta_0) \) is asymptotically mixed-normal rather than normal. Since

\[
V_2 - (\Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}) = \Omega_{12} \Omega_{22}^{-1} \Omega_{21} - \Omega_{12} \beta_\infty' - \beta_\infty \Omega_{21} + \beta_\infty \Omega_{22} \beta_\infty' = (\Omega_{12} \Omega_{22}^{-1} - \beta_\infty) \Omega_{22} (\Omega_{12} \Omega_{22}^{-1} - \beta_\infty)' \geq 0
\]

almost surely, the feasible estimator \( \hat{\theta}_{2T} \) has a large variation than the infeasible estimator \( \hat{\theta}_{2T} \).

This is consistent with our intuition.

Under the fixed-smoothing asymptotics, we still have \( \sqrt{T}(\hat{\theta}_{\mathbf{1}T} - \theta_0) \overset{d}{\to} N(0,\Omega_{11}) \) as \( \hat{\theta}_{\mathbf{1}T} \) does not depend on the smoothing parameter \( h \). So the asymptotic (conditional) variances of \( \sqrt{T}(\hat{\theta}_{\mathbf{1}T} - \theta_0) \) and \( \sqrt{T}(\hat{\theta}_{2T} - \theta_0) \) are both larger than \( \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \) in terms of matrix definiteness. However, it is not straightforward to directly compare them. To evaluate the relative magnitude of the two (conditional) asymptotic variances, we define

\[
\tilde{\beta}_\infty := \beta_\infty(h,d,q) := \hat{\Omega}_{\infty,12} \hat{\Omega}_{\infty,22}^{-1}, \tag{2}
\]

which does not depend on any nuisance parameter but depends on \( h,d,q \). For notational economy, we sometimes suppress this dependence. Direct calculations show that

\[
\beta_\infty = \Omega_{1/2}^\frac{1}{2} \tilde{\beta}_\infty \Omega_{22}^{-1/2} + \Omega_{12} \Omega_{22}^{-1}. \tag{3}
\]

Using this, we have, for any conforming vector \( a \):

\[
Ea' (V_2 - \Omega_{11}) a = Ea' \beta_\infty \Omega_{22} \beta_\infty' a - a' (\Omega_{12} \beta_\infty' + \beta_\infty \Omega_{21}) a
= a' (\Omega_{12}^\frac{1}{2} E \tilde{\beta}_\infty \beta_\infty \Omega_{12}^{-1/2} a - \Omega_{12} \Omega_{22}^{-1} \Omega_{21} a
= a' (\Omega_{12}^\frac{1}{2} \left[ E \tilde{\beta}_\infty \beta_\infty' - \Omega_{12}^{-1/2} \Omega_{12} \Omega_{22}^{-1} \Omega_{21} (\Omega_{12}^{-1/2})' \right] (\Omega_{12}^{-1/2})') a. \tag{4}
\]

The following lemma gives a characterization of \( E \tilde{\beta}_\infty (h,d,q) \tilde{\beta}_\infty (h,d,q)' \).
Lemma 2  For any $d \geq 1$, we have $E[\tilde{\beta}_\infty (h, d, q) \tilde{\beta}_\infty (h, d, q)'] = \left( E[||\tilde{\beta}_\infty (h, 1, q)||^2] \right) \times I_d$.

On the basis of the above lemma, we define

$$g(h, q) := \frac{E[||\tilde{\beta}_\infty (h, 1, q)||^2]}{1 + E[||\tilde{\beta}_\infty (h, 1, q)||^2]} \in (0, 1).$$

Let

$$\rho = \Omega_{11}^{-1/2}\Omega_{12}(\Omega_{22}^{-1/2})' \in \mathbb{R}^{d \times q},$$

which is the long run correlation matrix between $u_{1t}$ and $u_{2t}$. Using \(4\) and rewriting it in terms of $\rho$, we obtain the following proposition.

Proposition 3  Let Assumptions $[7]$ hold. If $g(h, q)I_d - \rho\rho'$ is positive (negative) semidefinite, then $\theta_{2T}$ has a larger (smaller) asymptotic variance than $\theta_{1T}$ under the fixed-smoothing asymptotics.

In the proof of the proposition, we show that the choices of the matrix square roots $\Omega_{11}^{1/2}$ and $\Omega_{22}^{1/2}$ are innocuous. To use the proposition, we need only to check whether the condition is satisfied for a given choice of the matrix square roots.

If $\rho = 0$, then $g(h, q)I_d \geq \rho\rho'$ holds trivially. In this case, the asymptotic variance of $\hat{\theta}_{2T}$ is larger than the asymptotic variance of $\hat{\theta}_{1T}$. Intuitively, when the long run correlation is zero, there is no information that can be explored to improve efficiency. If we insist on using the long run correlation matrix in attempt to improve the efficiency, we may end up with a less efficient estimator, due to the noise in estimating the zero long run correlation matrix. On the other hand, if $\rho\rho' = I_d$ after some possible rotation, which holds when the long run variation of $u_{1t}$ is almost perfectly predicted by $u_{2t}$, then $g(h, q)I_d < \rho\rho'$. In this case, it is worthwhile estimating the long run variance and using it to improve the efficiency $\theta_{2T}$. There are many scenarios in between where the matrix $g(h, q)I_d - \rho\rho'$ is indefinite. In this case we may still be able to access the definiteness of the matrix along some directions; see Theorem $[10]$. The case with $d = 1$ is simplest. In this case, if $\rho\rho' > g(h, q)$, then $\theta_{2T}$ is asymptotically more efficient than $\theta_{1T}$. Otherwise, it is not asymptotically less efficient.

In the case of kernel LRV estimation, it is hard to obtain an analytical expression for $E[||\tilde{\beta}_\infty (h, 1, q)||^2]$ and hence $g(h, q)$, although we can always simulate it numerically. The threshold $g(h, q)$ depends on the smoothing parameter $h = 1/b$ and the degree of overidentification $q$. Tables $[1][3]$ report the simulated values of $g(h, q)$ for $b = 0.00 : 0.01 : 0.20$ and $q = 1 \sim 5$. It is clear that $g(h, q)$ increases with $q$ and decreases with the smoothing parameter $h = 1/b$.

When the OS LRV estimation is used, we do not need to simulate $g(h, q)$ as we can obtain a closed form expression for it. Using this expression, we can prove the following corollary.

Corollary 4  Let Assumptions $[1][3]$ hold. Consider the case of OS LRV estimation. If $\frac{q}{K-1}I_d - \rho\rho'$ is positive (negative) semidefinite, then $\theta_{2T}$ has a larger (smaller) asymptotic variance than $\theta_{1T}$ under the fixed-smoothing asymptotics.

A necessary and sufficient condition for $q/(K-1)I_d - \rho\rho'$ to be positive semidefinite is that the largest eigenvalue of $\rho\rho'$ is smaller than $q/(K-1)$. The eigenvalues of $\rho\rho'$ are the squared long run correlation coefficients between $c'_1u_{1t}$ and $c'_2u_{2t}$ for some $c_1$ and $c_2$, i.e., the square long run canonical correlation coefficients between $u_{1t}$ and $u_{2t}$. So if the largest squared long run
We are interested in testing the null hypothesis

4 A Tale of Two Asymptotics: Hypothesis Testing

asymptotic variance of

\[ \hat{\theta}_{2T} \] has a larger asymptotic variance than \( \hat{\theta}_{1T} \). Similarly, if the smallest squared long run canonical correlation is greater than \( q/(K - 1) \), then \( \hat{\theta}_{2T} \) has a smaller asymptotic variance than \( \hat{\theta}_{1T} \).

Since \( \hat{\theta}_{2T} \) is not asymptotically normal, asymptotic variance comparison does not paint the whole picture. To compare the asymptotic distributions of \( \hat{\theta}_{1T} \) and \( \hat{\theta}_{2T} \), we consider the case that \( d = q = 1 \) and \( K = 4 \) as an example. Figure 2 represents the shapes of probability density functions in the case of OS LRV estimation when \( (\Omega_{11}, \Omega_{12}^2, \Omega_{22}) = (1, 0.10, 1) \). In this case, \( \Omega_{11} - \Omega_{12}^2 \Omega_{22}^{-1} \Omega_{21} = 0.9 \). The first graph shows \( \sqrt{T}(\hat{\theta}_{1T} - \theta_0) \) (red solid line) and \( \sqrt{T}(\theta_{2T} - \theta_0) \) (blue dotted line) under the conventional asymptotics. The conventional limiting distributions for \( \sqrt{T}(\hat{\theta}_{1T} - \theta_0) \) and \( \sqrt{T}(\hat{\theta}_{2T} - \theta_0) \) are both normal but the latter has a smaller variance, so the asymptotic efficiency of \( \hat{\theta}_{2T} \) is always guaranteed. However, this is not true in the second graph of Figure 2 which represents the limiting distributions under the fixed-smoothing asymptotics. Here, as before the red solid line represents \( N(0, 1) \) but the blue dotted line represents the mixed normal distribution \( MN[0, 0.9(1 + \tilde{\beta}_{\infty}^2)] \) with \( \tilde{\beta}_{\infty} = \sum_{i=1}^{K} \xi_{1i} \xi_{2i} / \sum_{i=1}^{K} \xi_{2i} \sim MN(0, 1/\chi^2_K) \). This can be obtained by using (4) with \( a = 1 \). More specifically,

\[ V_2 = \Omega_{12}^{1/2} \beta_{\infty} \beta_{\infty}' (\Omega_{12}^{1/2})' - \Omega_{12} \Omega_{22}^{-1} \Omega_{21} + \Omega_{11} = 0.9 \tilde{\beta}_{\infty}^2 + 0.9 = 0.9(1 + \tilde{\beta}_{\infty}^2). \]  

Comparing these two different families of distributions, we find that the asymptotic distribution of \( \hat{\theta}_{2T} \) has fatter tail than that of \( \hat{\theta}_{1T} \). The asymptotic variance of \( \hat{\theta}_{2T} \) is \( 0.9(1 + 1/2) = 1.35 \) which is larger than the asymptotic variance of \( \hat{\theta}_{1T} \).

More generally, when \( d = q = 1 \) with \( (\Omega_{11}, \Omega_{12}^2, \Omega_{22}) = (1, \rho^2, 1) \), we can use (4) to obtain the asymptotic variance of \( \hat{\theta}_{2T} \) as follows:

\[ 1 + (1 - \rho^2) \left[ E\tilde{\beta}_{\infty}^2 - \frac{\rho^2}{1 - \rho^2} \right] = (1 - \rho^2) \left( 1 + E\tilde{\beta}_{\infty}^2 \right) \]

\[ = (1 - \rho^2) \left( 1 + \frac{1}{K - 2} \right) = (1 - \rho^2) \left( \frac{K - 1}{K - 2} \right). \]

So, for any choice of \( K \), the asymptotic variance of \( \hat{\theta}_{2T} \) decreases as \( \rho^2 \) increases. Under the quadratic loss function, the asymptotic relative efficiency of \( \hat{\theta}_{2T} \) relative to \( \hat{\theta}_{1T} \) is thus equal to \( (1 - \rho^2) (K - 1) / (K - 2) \). The second factor \( (K - 1) / (K - 2) > 1 \) reflects the cost of having to estimate \( \rho \). When \( \rho^2 = 1/(K - 1) \), \( \hat{\theta}_{1T} \) and \( \hat{\theta}_{2T} \) have the same asymptotic variance. For a given choice of \( K \), if \( \rho^2 \leq 1/(K - 1) \), we prefer to use \( \hat{\theta}_{1T} \). This is not captured by the conventional asymptotics, which always chooses \( \hat{\theta}_{2T} \) over \( \hat{\theta}_{1T} \).

4 A Tale of Two Asymptotics: Hypothesis Testing

We are interested in testing the null hypothesis \( H_0 : R\theta_0 = r \) against the local alternative \( H_1 : R\theta_0 = r + \delta_0 \sqrt{T} \) for some \( p \times d \) matrix \( R \) and \( p \times 1 \) vectors \( r \) and \( \delta_0 \). Nonlinear restrictions can be converted into linear ones using the Delta method. We construct the following two Wald statistics:

\[ \mathbb{W}_{1T} := T(R\hat{\theta}_{1T} - r)' \left( R\hat{\Omega}_{11} R \right)^{-1} (R\hat{\theta}_{1T} - r) \]

\[ \mathbb{W}_{2T} := T(R\hat{\theta}_{2T} - r)' \left[ R\hat{\Omega}_{12} R \right]^{-1} (R\hat{\theta}_{2T} - r) \]
Figure 2: Limiting distributions of $\tilde{\theta}_1T$ and $\tilde{\theta}_2T$ based on the OS LRV estimator with $K = 4$.

where $\hat{\Omega}_{1.2} = \hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}$. When $p = 1$ and the alternative is one sided, we can construct the following two $t$ statistics:

$$T_{1T} : = \frac{\sqrt{T} \left( R\hat{\theta}_{1T} - r \right)}{\sqrt{R\hat{\Omega}_{11}R'}}$$

(6)

$$T_{2T} : = \frac{\sqrt{T} \left( R\hat{\theta}_{2T} - r \right)}{\sqrt{R\hat{\Omega}_{1.2}R'}}.$$  

(7)

No matter whether the test is based on $\hat{\theta}_{1T}$ or $\hat{\theta}_{2T}$, we have to employ the long run covariance estimator $\hat{\Omega}$. Define the $p \times p$ matrices $A_1$ and $A_2$ according to

$$\Lambda_1A_1' = R\hat{\Omega}_{11}R'$$

and

$$\Lambda_2A_2' = R\hat{\Omega}_{1.2}R'.$$

In other words, $\Lambda_1$ and $\Lambda_2$ are matrix square roots of $R\hat{\Omega}_{11}R'$ and $R\hat{\Omega}_{1.2}R'$ respectively.
Under the conventional increasing-smoothing asymptotics, it is straightforward to show that under $H_1 : R\theta_0 = r + \delta_0 / \sqrt{T}$:

\[
\mathbb{W}_{1T} \xrightarrow{d} \chi^2_p(\|\Lambda^{-1}_1\delta_0\|^2), \quad \mathbb{W}_{2T} \xrightarrow{d} \chi^2_p(\|\Lambda^{-1}_2\delta_0\|^2), \quad T_{1T} \xrightarrow{d} N(\Lambda^{-1}_1\delta_0, 1), \quad T_{2T} \xrightarrow{d} N(\Lambda^{-1}_2\delta_0, 1).
\]

When $\delta_0 = 0$, we obtain the null distributions:

\[
\mathbb{W}_{1T}, \mathbb{W}_{2T} \xrightarrow{d} \chi^2_p \quad \text{and} \quad T_{1T}, T_{2T} \xrightarrow{d} N(0, 1).
\]

So under the conventional increasing-smoothing asymptotics, the null limiting distributions of $\mathbb{W}_{1T}$ and $\mathbb{W}_{2T}$ are identical. Since $\|\Lambda^{-1}_1\delta_0\|^2 \leq \|\Lambda^{-1}_2\delta_0\|^2$, under the conventional asymptotics, the local asymptotic power function of the test based on $\mathbb{W}_{2T}$ is higher than that based on $\mathbb{W}_{1T}$.

The key driving force behind the conventional asymptotics is that we approximate the distribution of $\bar{\Omega}$ by the degenerate distribution concentrating on $\Omega$. The degenerate approximation does not reflect the finite sample distribution well. As in the previous section, we employ the fixed-smoothing asymptotics to derive more accurate distributional approximations. Let

\[
C_{pp} = \int_0^1 \int_0^1 Q_h^*(r, s) dB_p(r) dB_p(s)', \quad C_{pq} = \int_0^1 \int_0^1 Q_h^*(r, s) dB_p(r) dB_q(s)',
\]

\[
C_{qq} = \int_0^1 \int_0^1 Q_h^*(r, s) dB_q(r) dB_q(s)', \quad C_{pp} = C_{pq}
\]

and

\[
D_{pp} = C_{pp} - C_{pq}C^{-1}_{qq}C_{pq}'
\]

where $B_p(\cdot) \in \mathbb{R}^p$ and $B_q(\cdot) \in \mathbb{R}^q$ are independent standard Brownian motion processes.

**Proposition 5** Let Assumptions 1–3 hold. As $T \to \infty$ for a fixed $h$, we have, under $H_1 : R\theta_0 = r + \delta_0 / \sqrt{T}$:

(a) $\mathbb{W}_{1T} \xrightarrow{d} \mathbb{W}_{1\infty}(\|\Lambda^{-1}_1\delta_0\|^2)$ where

\[
\mathbb{W}_{1\infty}(\|\xi\|^2) = [B_p(1) + \xi]' C^{-1}_{pp} [B_p(1) + \xi] \quad \text{for} \ \xi \in \mathbb{R}^p.
\]

(b) $\mathbb{W}_{2T} \xrightarrow{d} \mathbb{W}_{2\infty}(\|\Lambda^{-1}_2\delta_0\|^2)$ where

\[
\mathbb{W}_{2\infty}(\|\xi\|^2) = [B_p(1) - C_{pq}C^{-1}_{qq}B_q(1) + \xi]' D_{pp}^{-1} [B_p(1) - C_{pq}C^{-1}_{qq}B_q(1) + \xi].
\]

(c) $T_{1T} \xrightarrow{d} \mathbb{T}_{1\infty}(\Lambda^{-1}_1\delta_0) \equiv [B_p(1) + \Lambda^{-1}_1\delta_0] / \sqrt{\bar{C}_{pp}} \quad \text{for} \ p = 1.$

(d) $T_{2T} \xrightarrow{d} \mathbb{T}_{2\infty}(\Lambda^{-1}_2\delta_0) \equiv [B_p(1) - C_{pq}C^{-1}_{qq}B_q(1) + \Lambda^{-1}_2\delta_0] / \sqrt{\bar{D}_{pp}} \quad \text{for} \ p = 1.$

In Proposition 5, we use the notation $\mathbb{W}_{1\infty}(\|\xi\|^2)$, which implies that the right hand side of (8) depends on $\xi$ only through $\|\xi\|^2$. This is true, because for any orthogonal matrix $H$:

\[
[B_p(1) + \xi]' C^{-1}_{pp} [B_p(1) + \xi] = [H B_p(1) + H \xi]' H C^{-1}_{pp} H' [H B_p(1) + H \xi]
\]

\[
\xrightarrow{d} [B_p(1) + H \xi]' C^{-1}_{pp} [B_p(1) + H \xi].
\]
If we choose $H = (\xi/\|\xi\|, \bar{H})'$ for some $\bar{H}$ such that $H$ is orthogonal, then

$$[B_p (1) + \xi'] C_{pp}^{-1} [B_p (1) + \xi] \overset{d}{=} [B_p (1) + \|\xi\| e_p]' C_{pp}^{-1} [B_p (1) + \|\xi\| e_p].$$

So the distribution of $[B_p (1) + \xi'] C_{pp}^{-1} [B_p (1) + \xi]$ depends on $\xi$ only through $\|\xi\|$. Similarly, the distribution of the right hand side of (9) depends only on $\|\xi\|^2$.

When $\delta_0 = 0$, we obtain the limiting distributions of $\mathcal{W}_{1T}, \mathcal{W}_{2T}, \mathcal{T}_{1T}$ and $\mathcal{T}_{2T}$ under the null hypothesis:

$$\mathcal{W}_{1T} \overset{d}{=} \mathcal{W}_{1\infty} := \mathcal{W}_{1\infty} (0) = B_p (1)' C_{pp}^{-1} B_p (1),$$

$$\mathcal{W}_{2T} \overset{d}{=} \mathcal{W}_{2\infty} := \mathcal{W}_{2\infty} (0) = [B_p (1) - C_{pq} C_{qq}^{-1} B_q (1)]' D_{pp}^{-1} [B_p (1) - C_{pq} C_{qq}^{-1} B_q (1)],$$

$$\mathcal{T}_{1T} \overset{d}{=} \mathcal{T}_{1\infty} := \mathcal{T}_{1\infty} (0) = B_p (1)/\sqrt{C_{pp}},$$

$$\mathcal{T}_{2T} \overset{d}{=} \mathcal{T}_{2\infty} := \mathcal{T}_{2\infty} (0) = [B_p (1) - C_{pq} C_{qq}^{-1} B_q (1)] / \sqrt{D_{pp}}.$$

These distributions are different from those under the conventional asymptotics. For $\mathcal{W}_{1T}$ and $\mathcal{T}_{1T}$, the difference lies in the random scaling factor $C_{pp}$ or $\sqrt{C_{pp}}$. The random scaling factor captures the estimation uncertainty of the LRV estimator. For $\mathcal{W}_{2T}$ and $\mathcal{T}_{2T}$, there is an additional difference embodied by the random location shift $C_{pq} C_{qq}^{-1} B_q (1)$ with a consequent change in the random scaling factor.

The proposition below provides some characterization of the two limiting distributions $\mathcal{W}_{1\infty}$ and $\mathcal{W}_{2\infty}$.

**Proposition 6** For any $x > 0$, the following hold:

(a) $\mathcal{W}_{1\infty} (0)$ first-order stochastically dominates (FSD) $\mathcal{W}_{2\infty} (0)$ in that

$$P [\mathcal{W}_{2\infty} (0) \geq x] > P [\mathcal{W}_{1\infty} (0) \geq x].$$

(b) $P [\mathcal{W}_{1\infty} (\|\xi\|^2) \geq x] \text{ strictly increases with } \|\xi\|^2 \text{ and } \lim_{\|\xi\|^2 \to \infty} P [\mathcal{W}_{1\infty} (\|\xi\|^2) \geq x] = 1.$

(c) $P [\mathcal{W}_{2\infty} (\|\xi\|^2) \geq x] \text{ strictly increases with } \|\xi\|^2 \text{ and } \lim_{\|\xi\|^2 \to \infty} P [\mathcal{W}_{2\infty} (\|\xi\|^2) \geq x] = 1.$

Proposition (a) is intuitive. $\mathcal{W}_{2\infty}$ FSD $\mathcal{W}_{1\infty}$ because $\mathcal{W}_{2\infty}$ FSD $\mathcal{B}_p (1)' D_{pp}^{-1} B_p (1)$, which in turn FSD $\mathcal{B}_p (1)' C_{pp}^{-1} B_p (1)$ which is just $\mathcal{W}_{1\infty}$. According to a property of the first-order stochastic dominance, we have

$$\mathcal{W}_{2\infty} \overset{d}{=} \mathcal{W}_{1\infty} + \mathcal{W}_e$$

for some $\mathcal{W}_e > 0$. Intuitively, $\mathcal{W}_{2\infty}$ shifts some of the probability mass of $\mathcal{W}_{1\infty}$ to the right. A direct implication is that the asymptotic critical values for $\mathcal{W}_{2T}$ are larger than the corresponding ones for $\mathcal{W}_{1T}$. The difference in critical values has implications on the power properties of the two tests.

For $x > 0$, we have

$$P (\mathcal{T}_{1\infty} > x) = \frac{1}{2} P (\mathcal{W}_{1\infty} \geq x^2) \text{ and } P (\mathcal{T}_{2\infty} > x) = \frac{1}{2} P (\mathcal{W}_{2\infty} \geq x^2).$$

It then follows from Proposition (a) that $P (\mathcal{T}_{2\infty} > x) \geq P (\mathcal{T}_{1\infty} > x)$ for $x > 0$. So for a one-sided test with the alternative $H_1 : R \theta_0 > r$, critical values from $\mathcal{T}_{2\infty}$ are larger than those from
\(T_{1,\infty}\). Similarly, we have \(P (T_{2,\infty} < x) \geq P (T_{1,\infty} < x)\) for \(x < 0\). This implies that for a one-sided test with the alternative \(H_1 : R_0^0 < r\), critical values from \(T_{2,\infty}\) are smaller than those from \(T_{1,\infty}\).

Let \(\mathbb{W}_{1,\infty}^0\) and \(\mathbb{W}_{2,\infty}^0\) be the \((1 - \alpha)\) quantile from the distributions \(\mathbb{W}_{1,\infty}\) and \(\mathbb{W}_{2,\infty}\), respectively. The local asymptotic power functions of the two tests are

\[
\pi_1 (\|A_1^{-1} \delta_0\|^2) := \pi_1 (\|A_1^{-1} \delta_0\|^2; K, p, q, \alpha) = P \left[ \mathbb{W}_{1,\infty} (\|A_1^{-1} \delta_0\|^2) > \mathbb{W}_{1,\infty}^0 \right],
\]

\[
\pi_2 (\|A_2^{-1} \delta_0\|^2) := \pi_2 (\|A_2^{-1} \delta_0\|^2; K, p, q, \alpha) = P \left[ \mathbb{W}_{2,\infty} (\|A_2^{-1} \delta_0\|^2) > \mathbb{W}_{2,\infty}^0 \right].
\]

While \(\|A_2^{-1} \delta_0\|^2 \geq \|A_1^{-1} \delta_0\|^2\), we also have \(\mathbb{W}_{2,\infty}^0 > \mathbb{W}_{1,\infty}^0\). The effects of the critical values and the noncentrality parameter move in opposite directions. It is not straightforward to compare the two power functions. However, Proposition suggests that if the difference in the noncentrality parameters \(\|A_2^{-1} \delta_0\|^2 - \|A_1^{-1} \delta_0\|^2\) is large enough to offset the increase in critical values, then the two-step test based on \(\mathbb{W}_{2T}\) will be more powerful.

To evaluate \(\|A_2^{-1} \delta_0\|^2 - \|A_1^{-1} \delta_0\|^2\), we define

\[
\rho_R = (R \Omega_{11} R')^{-1/2} (R \Omega_{12}) \Omega_{22}^{-1/2},
\]

which is the long run correlation matrix \(\rho_R\) between \(R u_{1t}\) and \(u_{2t}\). In terms of \(\rho_R \in \mathbb{R}^{p \times q}\) we have

\[
\|A_2^{-1} \delta_0\|^2 - \|A_1^{-1} \delta_0\|^2 = \delta_0 (R \Omega_{11} R' - R \Omega_{12} \Omega_{22}^{-1} \Omega_{21} R')^{-1} \delta_0 - \delta_0' (R \Omega_{11} R')^{-1} \delta_0
\]

\[
= \delta_0 (A_1')^{-1} \left[ I_p - A_1^{-1} R \Omega_{12} \Omega_{22}^{-1} \Omega_{21} R' (A_1')^{-1} \right]^{-1} (A_1^{-1} \delta_0) - \delta_0' (A_1')^{-1} (A_1^{-1} \delta_0)
\]

\[
= \delta_0 (A_1')^{-1} \left[ I_p - \rho_R \rho_R' [I_p - \rho_R \rho_R']^{-1/2} (A_1^{-1} \delta_0) \right] > 0.
\]

So the difference in the noncentrality parameters depends on the matrix \(\rho_R \rho_R'\).

Let \(\rho_R = \sum_{i=1}^{\min(p,q)} \lambda_i a_i b_i^t\) be the singular value decomposition of \(\rho_R\) where \(\{a_i\}\) and \(\{b_i\}\) are orthonormal vectors in \(\mathbb{R}^p\) and \(\mathbb{R}^q\) respectively. Sorted in the descending order, \(\{\lambda_i^2\}\) are the (squared) long run canonical correlation coefficients between \(R u_{1t}\) and \(u_{2t}\). Then

\[
\|A_2^{-1} \delta_0\|^2 - \|A_1^{-1} \delta_0\|^2 = \sum_{i=1}^{\min(p,q)} \frac{\lambda_i^2}{1 - \lambda_i^2} \left[ a_i' A_1^{-1} \delta_0 \right]^2.
\]

Consider a special case that \(\lambda_i^2 := \max_{i=1}^p \{\lambda_i^2\}\) approaches 1. If \(a_i' A_1^{-1} \delta_0 \neq 0\), then \(\|A_2^{-1} \delta_0\|^2 - \|A_1^{-1} \delta_0\|^2\) and hence \(\|A_2^{-1} \delta_0\|^2\) approaches \(\infty\) as \(\lambda_i^2\) approaches 1 from below. This case happens when the second block of moment conditions has very high long run prediction power for the first block. In this case, we expect the \(\mathbb{W}_{2T}\) test to be more powerful, as \(\lim_{|\lambda_i|^{-1}} \pi_2 (\|A_2^{-1} \delta_0\|^2) = 1\).

Consider another special case that \(\max_{i} \{\lambda_i^2\} = 0\), i.e., \(\rho_R\) is a matrix of zeros. In this case, the second block of moment conditions contains no additional information, and we have \(\|A_2^{-1} \delta_0\|^2 = \|A_1^{-1} \delta_0\|^2\). In this case, we expect the \(\mathbb{W}_{2T}\) test to be less powerful.

It follows from Proposition(b)(c) that for any \(\lambda\), there exists a unique \(\tau (\lambda) := \tau (\lambda; h, p, q, \alpha)\) such that

\[
\pi_1 (\lambda) = \pi_2 (\lambda \tau (\lambda)).
\]
Then $\pi_2(\|\Lambda_2^{-1}\delta_0\|^2) < \pi_1(\|\Lambda_1^{-1}\delta_0\|^2)$ if and only if $\|\Lambda_2^{-1}\delta_0\|^2 < \tau(\|\Lambda_1^{-1}\delta_0\|^2) \cdot \|\Lambda_1^{-1}\delta_0\|^2$. But

$$\|\Lambda_2^{-1}\delta_0\|^2 - \tau(\|\Lambda_1^{-1}\delta_0\|^2) \cdot \|\Lambda_1^{-1}\delta_0\|^2 = \tau(\|\Lambda_1^{-1}\delta_0\|^2) \cdot \delta'_0(\Lambda')^{-1} \left[ I_p - \rho_R\rho'_R \right]^{-1/2} \left[ \rho_R\rho'_R - f(\|\Lambda_1^{-1}\delta_0\|^2) I_p \right] \left[ I_p - \rho_R\rho'_R \right]^{-1/2} (\Lambda_1^{-1}\delta_0),$$

where

$$f(\lambda) := f(\lambda; h, p, q, \alpha) = \frac{\tau(\lambda; h, p, q, \alpha) - 1}{\tau(\lambda; h, p, q, \alpha)}.$$ 

So the power comparison depends on whether $f(\|\Lambda_1^{-1}\delta_0\|^2) I_p - \rho_R\rho'_R$ is positive semidefinite.

**Proposition 7** Let Assumptions 1–3 hold and define $\lambda := \|\Lambda_1^{-1}\delta_0\|^2 = \delta'_0(RO_{11}R')^{-1}\delta_0$. If $f(\lambda; h, p, q, \alpha)I_p - \rho_R\rho'_R$ is positive (negative) semidefinite, then the two-step test based on $W_{2T}$ has a lower (higher) local asymptotic power than the one-step test based on $W_{1T}$ under the fixed-smoothing asymptotics.

Note that

$$\rho_R\rho'_R = (RO_{11}R')^{-1/2} R(\Omega_{12}\Omega_{22}^{-1}\Omega_{21}) R' (RO_{11}R')^{-1/2}.$$ 

To pin down $\rho_R\rho'_R$, we need to choose a matrix square root of $R\Omega_{11}R'$. Following the arguments in the proof of Proposition 3, it is easy to see that we can use any matrix square root. In particular, we can use the principal matrix square root.

When $p = 1$, which is of ultimate importance in empirical studies, $\rho_R\rho'_R$ is equal to the sum of the squared long run canonical correlation coefficients. In this case, $f(\lambda; h, p, q, \alpha)$ is the threshold value of $\rho_R\rho'_R$ for assessing the relative efficiency of the two tests. More specifically, when $\rho_R\rho'_R > f(\lambda; h, p, q, \alpha)$, the two-step $W_{2T}$ test is more powerful than the one-step $W_{1T}$ test. Otherwise, the two-step $W_{2T}$ test is less powerful.

Proposition 7 is in parallel with Proposition 3. The qualitative messages of these two propositions are the same — when the long run correlation is high enough, we should estimate and exploit it to reduce the variation of our point estimator and improve the power of the associated tests. However, the thresholds are different quantitatively. The two propositions fully characterize the threshold for each criterion under consideration.

**Proposition 8** Consider the case of OS LRV estimation. For any $\lambda \in \mathbb{R}^+$, we have $\pi_1(\lambda) > \pi_2(\lambda)$ and hence $\tau(\lambda; h, p, q, \alpha) > 1$ and $f(\lambda; h, p, q, \alpha) > 0$.

Proposition 8 is intuitive. When there is no long run correlation between $Ru_{1t}$ and $u_{2t}$, we have $\|\Lambda_2^{-1}\delta_0\|^2 = \|\Lambda_1^{-1}\delta_0\|^2$. In this case, the two-step $W_{2T}$ test is necessarily less powerful. The proof uses the theory of uniformly most powerful invariant tests and the theory of complete and sufficient statistics. It is an open question whether the same strategy can be adopted to prove Proposition 3 in the case of kernel LRV estimation. Our extensive numerical work supports that $\tau(\lambda; h, p, q, \alpha) > 1$ and $f(\lambda; h, p, q, \alpha) > 0$ continue to hold in the kernel case.

It is not easy to give an analytical expression for $f(\lambda; h, p, q, \alpha)$ but we can compute it numerically without any difficulty. In Tables 4 and 5, we consider the case of OS LRV estimation and compute the values of $f(\lambda; K, p, q, \alpha)$ for $\lambda = 1 \sim 10, K = 8, 10, 12, 14, p = 1 \sim 4$ and $q = 1 \sim 3$. Similar to the asymptotic variance comparison, we find that these threshold values increase as the degree of over-identification increases and decrease as the smoothing parameter $K$ increases. Note that the continuity of two power functions $\pi_1(\lambda; K, p, q, \alpha)$ and $\pi_2(\lambda; K, p, q, \alpha)$ guarantees
that for \( \lambda \in [1, 10] \) the two-step \( \mathbb{W}_{2T} \) test has a lower local asymptotic power than the one-step \( \mathbb{W}_{1T} \) test if the largest eigenvalue of \( p_R^T \) is smaller than \( \min_{\lambda \in [1, 10]} f(\lambda; K, p, q, \alpha) \). On the other hand, for \( \lambda \in [1, 10] \) the two-step \( \mathbb{W}_{2T} \) test has a higher local asymptotic power if the smallest eigenvalue of \( p_R^T \) is greater than \( \max f(\lambda; K, p, q, \alpha) \).

For the case of kernel LRV estimator, results not reported here show that \( f(\lambda; h, p, q, \alpha) \) increases with \( q \) and decreases with \( h \). This is entirely analogous to the case of OS LRV estimation.

5 General Overidentified GMM Framework

In this section, we consider the general GMM framework. The parameter of interest is a \( d \times 1 \) vector \( \theta \in \Theta \subseteq \mathbb{R}^d \). Let \( v_t \in \mathbb{R}^{d_v} \) denote the vector of observations at time \( t \). We assume that \( \theta_0 \) is the true value, an interior point of the parameter space \( \Theta \). The moment conditions

\[
E\tilde{f}(v_t, \theta) = 0, \quad t = 1, 2, ..., T.
\]

hold if and only if \( \theta = \theta_0 \) where \( \tilde{f}(v_t, \cdot) \) is an \( m \times 1 \) vector of continuously differentiable functions. The process \( \tilde{f}(v_t, \theta_0) \) may exhibit autocorrelation of unknown forms. We assume that \( m \geq d \) and that the rank of \( E[\partial \tilde{f}(v_t, \theta_0) / \partial \theta^T] \) is equal to \( d \). That is, we consider a model that is possibly over-identified with the degree of over-identification \( q = m - d \).

5.1 One-Step and Two-Step Estimation and Inference

Define the \( m \times m \) contemporaneous covariance matrix \( \tilde{\Sigma} \) and the LRV matrix \( \tilde{\Omega} \) as:

\[
\tilde{\Sigma} = E\tilde{f}(v_t, \theta)\tilde{f}(v_t, \theta)' \quad \text{and} \quad \tilde{\Omega} = \sum_{j=-\infty}^{\infty} \tilde{\Omega}_j \quad \text{where} \quad \tilde{\Omega}_j = E\tilde{f}(v_t, \theta_0)\tilde{f}(v_{t-j}, \theta_0)'.
\]

Let

\[
\tilde{g}_t(\theta) = \frac{1}{\sqrt{T}} \sum_{j=1}^{t} \tilde{f}(v_j, \theta).
\]

Given a simple positive-definite weighting matrix \( \tilde{W}_{0T} \) that does not depend on any unknown parameter, we can obtain an initial GMM estimator of \( \theta_0 \) as

\[
\hat{\theta}_{0T} = \arg \min_{\theta \in \Theta} \tilde{g}_T(\theta)'\tilde{W}_{0T}^{-1}\tilde{g}_T(\theta).
\]

For example, we may set \( \tilde{W}_{0T} \) equal to \( I_m \). In the case of IV regression, we may set \( \tilde{W}_{0T} \) equal to \( Z_T'Z_T/T \) where \( Z_T \) is the matrix of the instruments.

Using \( \tilde{\Sigma} \) or \( \tilde{\Omega} \) as the weighting matrix, we obtain the following two (infeasible) GMM estimators:

\[
\hat{\theta}_{1T} : = \arg \min_{\theta \in \Theta} \tilde{g}_T(\theta)'\tilde{\Sigma}^{-1}\tilde{g}_T(\theta), \quad (11)
\]

\[
\hat{\theta}_{2T} : = \arg \min_{\theta \in \Theta} \tilde{g}_T(\theta)'\tilde{\Omega}^{-1}\tilde{g}_T(\theta). \quad (12)
\]

For the estimator \( \hat{\theta}_{1T} \), we use the contemporaneous covariance matrix \( \tilde{\Sigma} \) as the weighting matrix and ignore all the serial dependency in the moment vector process \( \{\tilde{f}(v_t, \theta_0)\}_{t=1}^{T} \). In contrast to this procedure, the second estimator \( \hat{\theta}_{2T} \) accounts for the long run dependency. The feasible
versions of these two estimators $\hat{\theta}_{1T}$ and $\hat{\theta}_{2T}$ can be naturally defined by replacing $\Sigma$ and $\Omega$ with their estimates $\hat{\Sigma}_{est}(\hat{\theta}_{0T})$ and $\hat{\Omega}_{est}(\hat{\theta}_{0T})$ where

\[
\hat{\Sigma}_{est}(\theta) = \frac{1}{T} \sum_{t=1}^{T} \tilde{f}(v_t, \theta)\tilde{f}(v_t, \theta)',
\]

\[
\hat{\Omega}_{est}(\theta) = \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} \Omega^*_h(s, t) \tilde{f}(v_t, \theta)\tilde{f}(v_s, \theta)'.
\]

To test the null hypothesis $H_0 : R\theta_0 = r$ against $H_1 : R\theta_0 = r + \delta_0/\sqrt{T}$, we construct two different Wald statistics as follows:

\[
W_{1T} = T(R\hat{\theta}_{1T} - r)' \left\{ R\hat{\Omega}_{1T} R' \right\}^{-1} (R\hat{\theta}_{1T} - r),
\]

\[
W_{2T} = T(R\hat{\theta}_{2T} - r)' \left\{ R\hat{\Omega}_{2T} R' \right\}^{-1} (R\hat{\theta}_{2T} - r),
\]

where

\[
\hat{\Omega}_{1T} = \left[ G'_{1T} \hat{\Sigma}_{est}(\hat{\theta}_{1T}) G_{1T} \right]^{-1} \left[ G'_{1T} \hat{\Sigma}_{est}(\hat{\theta}_{1T}) \hat{\Omega}_{est}(\hat{\theta}_{1T}) \hat{\Sigma}_{est}(\hat{\theta}_{1T}) G_{1T} \right] \left[ G'_{1T} \hat{\Sigma}_{est}(\hat{\theta}_{1T}) G_{1T} \right]^{-1}
\]

\[
\hat{\Omega}_{2T} = \left[ G'_{2T} \hat{\Sigma}_{est}(\hat{\theta}_{2T}) G_{2T} \right]^{-1}
\]

and

\[
\hat{\xi}_{1T} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \tilde{f}(v_t, \theta)}{\partial \theta'} \bigg|_{\theta = \hat{\theta}_{1T}}, \quad \hat{\xi}_{2T} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial \tilde{f}(v_t, \theta)}{\partial \theta'} \bigg|_{\theta = \hat{\theta}_{2T}}.
\]

These are the standard Wald test statistics in the GMM framework.

To compare the two estimators $\hat{\theta}_{1T}$ and $\hat{\theta}_{2T}$ and associated tests, we maintain the standard assumptions below.

**Assumption 4** As $T \to \infty$, $\hat{\theta}_{0T} = \theta_0 + o_p(1)$, $\hat{\theta}_{1T} = \theta_0 + o_p(1)$, $\hat{\theta}_{2T} = \theta_0 + o_p(1)$ for an interior point $\theta_0 \in \Theta$.

**Assumption 5** Define

\[
\hat{\xi}_t(\theta) = \frac{1}{\sqrt{T}} \frac{\partial \tilde{g}_t}{\partial \theta'} = \frac{1}{T} \sum_{j=1}^{t} \frac{\partial \tilde{f}(v_t, \theta)}{\partial \theta'} \text{ for } t \geq 1 \text{ and } \hat{\xi}_0(\theta) = 0.
\]

For any $\theta_T = \theta_0 + o_p(1)$, the following holds: (a) $\lim_{T \to \infty} G_{[rT]}(\theta_T) = rG$ uniformly in $r$ where $G = G(\theta_0)$ and $G(\theta) = E\partial \tilde{f}(v_t, \theta)/\partial \theta'$; (b) $\hat{\Sigma}_{est}(\theta_T) \overset{p}{\to} \Sigma > 0$.

With these assumptions and some mild conditions, the standard GMM theory gives us

\[
\sqrt{T}(\hat{\theta}_{1T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \hat{\xi}' \hat{\Sigma}^{-1} G \right]^{-1} \hat{\xi}' \hat{\Sigma}^{-1} \tilde{f}(v_t, \theta_0) + o_p(1).
\]

Under the fixed-smoothing asymptotics, Sun (2014b) establishes the representation:

\[
\sqrt{T}(\hat{\theta}_{2T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \hat{\xi}' \hat{\Omega}_{\infty}^{-1} G \right]^{-1} \hat{\xi}' \hat{\Omega}_{\infty}^{-1} \tilde{f}(v_t, \theta_0) + o_p(1),
\]
where $\tilde{\Omega}_\infty$ is defined in the similar way as $\Omega_\infty$ in Proposition 1. $\tilde{\Omega}_\infty = \tilde{\Omega}_{1/2}^1 \tilde{\Omega}_\infty \tilde{\Omega}_{1/2}^1$.

Due to the complicated structure of two transformed moment vector processes, it is not straightforward to compare the asymptotic distributions of $\hat{\theta}_{1T}$ and $\hat{\theta}_{2T}$ as in Sections 3 and 4. To confront this challenge, we let

$$\tilde{G} = U_{m \times m} \cdot \Xi \cdot V_{d \times d} \cdot \Xi'$$

be a singular value decomposition (SVD) of $\tilde{G}$, where

$$\Xi' = \begin{pmatrix} A & O \end{pmatrix}_{d \times d}$$

$A$ is a $d \times d$ diagonal matrix and $O$ is a matrix of zeros. Also, we define

$$f^*(v_t, \theta_0) = (f_1^*(v_t, \theta_0), f_2^*(v_t, \theta_0))^T := U' \tilde{f}(v_t, \theta_0) \in \mathbb{R}^m,$$

where $f_1^*(v_t, \theta_0) \in \mathbb{R}^d$ and $f_2^*(v_t, \theta_0) \in \mathbb{R}^q$ are the rotated moment conditions. The variance and long run variance matrices of $\{f^*(v_t, \theta_0)\}$ are

$$\Sigma^* := U' \tilde{\Sigma} U = \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix}.$$

and $\Omega^* := U' \tilde{\Omega} U$. To convert the variance matrix into an identity matrix, we define the normalized moment conditions below:

$$f(v_t, \theta_0) = [f_1(v_t, \theta_0)', f_2(v_t, \theta_0)']' := (\Sigma_{1/2}^*)^{-1} f^*(v_t, \theta_0)$$

where

$$\Sigma_{1/2}^* = \begin{pmatrix} (\Sigma_{12}^*)^{1/2} & \Sigma_{12}^* (\Sigma_{22}^*)^{-1/2} \\ 0 & (\Sigma_{22}^*)^{1/2} \end{pmatrix}. \quad (17)$$

More specifically,

$$f_1(v_t, \theta_0) := (\Sigma_{12}^*)^{-1/2} \left[ f_1^*(v_t, \theta_0) - \Sigma_{12}^* (\Sigma_{22}^*)^{-1} f_2^*(v_t, \theta_0) \right] \in \mathbb{R}^d,$$

$$f_2(v_t, \theta_0) := (\Sigma_{22}^*)^{-1/2} f_2^*(v_t, \theta_0) \in \mathbb{R}^q.$$

Then the contemporaneous variance of the time series $\{f(v_t, \theta_0)\}$ is $I_m$ and the long run variance is $\Omega := (\Sigma_{1/2}^*)^{-1} \Omega^*(\Sigma_{1/2}^*)^{-1}$.

**Lemma 9** Let Assumptions 4 and 5 hold by replacing $u_t$ with $f(v_t, \theta_0)$ in Assumptions 3 and 3. Then as $T \to \infty$ for a fixed $h$,

$$\left(\Sigma_{12}^*\right)^{-1/2} AV' \sqrt{T}(\hat{\theta}_{1T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_1(v_t, \theta_0) + o_p(1) \overset{d}{\to} N(0, \Omega_{11}) \quad (18)$$

$$\left(\Sigma_{12}^*\right)^{-1/2} AV' \sqrt{T}(\hat{\theta}_{2T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ f_1(v_t, \theta_0) - \beta_\infty f_2(v_t, \theta_0) \right] + o_p(1) \overset{d}{\to} MN \left(0, \Omega_{11} - \Omega_{12} \beta_\infty' - \beta_\infty \Omega_{21} + \beta_\infty \Omega_{22} \beta_\infty' \right) \quad (19)$$

where $\beta_\infty := \Omega_{\infty, 12} \Omega_{\infty, 22}^{-1}$ is the same as in Proposition 4.
The asymptotic normality and mixed normality in the lemma are not new. The limiting distribution of $\sqrt{T}(\hat{\theta}_{2T} - \theta_0)$ is available from the standard GMM theory while that of $\sqrt{T}(\hat{\theta}_{2T} - \theta_0)$ has been recently established in Sun (2014b). It is the representation that is new and insightful. The representation casts the two estimators in the same form and enables us to directly compare their asymptotic properties.

It follows from the proof of the lemma that

$$(\Sigma_{1,2}^*)^{-1/2} AV'\sqrt{T}(\hat{\theta}_{2T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [f_1(v_t, \theta_0) - \beta_0 f_2(v_t, \theta_0)] + o_p(1)$$

where $\beta_0 = \Omega_{12}\Omega_{22}^{-1}$ as defined before. So the difference between the feasible and infeasible two-step GMM estimators lies in the uncertainty in estimating $\beta_0$. While the true value of $\beta$ appears in the asymptotic distribution of the infeasible estimator $\hat{\theta}_{2T}$, the fixed-smoothing limit of the implied estimator $\hat{\beta} := \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}$ appears in that of the feasible estimator $\hat{\theta}_{2T}$. It is important to point out that the estimation uncertainty in the whole weighting matrix $\hat{\Omega}_{\text{est}}$ matters only through that in $\hat{\beta}$.

If we let $(u_{1t}, u_{2t}) = (f_1(v_t, \theta_0), f_2(v_t, \theta_0))$, then the right hand sides of (18) and (19) are exactly the same as what we would obtain in the location model. The location model, as simple as the observations.

While

$$\rho = \Omega_{11}^{-1/2}\Omega_{12}\Omega_{22}^{-1/2} \in \mathbb{R}^{d \times q} \text{ and } \rho_R = \left(\tilde{R}\Omega_{11}\tilde{R}^{-1}\right)^{-1/2} \left(\tilde{R}\Omega_{12}\right) \Omega_{22}^{-1/2} \in \mathbb{R}^{p \times q}.$$ 

which is compatible with the definition in (2). We define

$$\rho = \Omega_{11}^{-1/2}\Omega_{12}\Omega_{22}^{-1/2} \in \mathbb{R}^{d \times q} \text{ and } \rho_R = \left(\tilde{R}\Omega_{11}\tilde{R}^{-1}\right)^{-1/2} \left(\tilde{R}\Omega_{12}\right) \Omega_{22}^{-1/2} \in \mathbb{R}^{p \times q}.$$ 

While $\rho$ is the long run correlation matrix between $f_1(v_t, \theta_0)$ and $f_2(v_t, \theta_0)$, $\rho_R$ is the long run correlation matrix between $\tilde{R} f_1(v_t, \theta_0)$ and $f_2(v_t, \theta_0)$. Finally, we redefine the noncentrality parameter $\lambda$:

$$\lambda = \delta_0 \left[ R(\tilde{G}^{\prime}\tilde{\Sigma}^{-1}\tilde{G})^{-1}(\tilde{G}^{\prime}\tilde{\Sigma}^{-1}\tilde{G})(\tilde{G}^{\prime}\tilde{\Sigma}^{-1}\tilde{G})^{-1}R \right]^{-1}\delta_0, \quad (21)$$

which is the noncentrality parameter based on the first-step test.

For the location model considered before, the above definition of $\rho_R$ is identical to that in (10). In that case, we have $\tilde{G} = (I_d, O_{d \times q})'$ and so $U = I_m$, $A = I_d$ and $V = I_d$. Given the assumption that $\tilde{\Sigma} = \Sigma^* = I_m$, which implies that $\Sigma_{1,2}^* = I_d$, we have $\tilde{R} = R$. In addition, the noncentrality parameter $\lambda$ reduces to $\lambda = \delta_0 (R\Omega_{11}R')^{-1}\delta_0$ as defined in Proposition 7.
Theorem 10 Let the assumptions in Lemma 9 hold.

(a) If \( g(h, q) \cdot I_d - \rho \rho' \) is positive (negative) semidefinite, then \( \hat{\theta}_{2T} \) has a larger (smaller) asymptotic variance than \( \theta_{1T} \) as \( T \to \infty \) for a fixed \( h \).

(b) If \( g(h, q) \cdot I_p - R \rho R' \) is positive (negative) semidefinite, then \( R \hat{\theta}_{2T} \) has a larger (smaller) asymptotic variance than \( R \theta_{1T} \) as \( T \to \infty \) for a fixed \( h \).

(c) If \( f(\lambda; h, p, q, \alpha) \cdot I_p - R \rho R' \) is positive (negative) semidefinite, then the test based on \( W_{2T} \) is asymptotically less (more) powerful than that based on \( W_{1T} \), as \( T \to \infty \) for a fixed \( h \).

Theorem 10(a) is a special case of Theorem 10(b) with \( R = I_d \). We single out this case in order to compare it with Proposition 3. It is reassuring to see that the results are the same. The only difference is that in the general GMM case we need to rotate and standardize the original moment conditions before computing the long run correlation matrix. Part (b) can also be applied to a general location model with a nonscalar error variance, in which case \( R = R(\Sigma^*_1)^{1/2} \). Theorem 10(c) reduces to Proposition 7 in the case of the location model.

### 5.2 Two-Step GMM Estimation and Inference with a Working Weighting Matrix

In the previous subsection, we employ two specific weighting matrices — the variance and long run variance estimators. In this subsection, we consider a general weighting matrix \( \hat{W}_T(\theta_{OT}) \), which may depend on the initial estimator \( \hat{\theta}_{OT} \) and the sample size \( T \), leading to yet another GMM estimator:

\[
\hat{\theta}_{aT} = \arg \min_{\theta \in \Theta} \hat{g}_T(\theta)' \left[ \hat{W}_T(\hat{\theta}_{OT}) \right]^{-1} \hat{g}_T(\theta)
\]

where the subscript ‘a’ signifies ‘another’ or ‘alternative’.

An example of \( \hat{W}_T(\theta_{OT}) \) is the implied LRV matrix when we employ a simple approximating parametric model to capture the dynamics in the moment process. We could also use the general LRV estimator but we choose a large \( h \) so that the variation in \( \hat{W}_T(\theta_{OT}) \) is small. In the kernel LRV estimation, this amounts to including only autocovariances of low orders in constructing \( \hat{W}_T(\theta_{OT}) \). We assume that \( \hat{W}_T(\theta_{OT}) \to^p \hat{W} \), a positive definite nonrandom matrix under the fixed-smoothing asymptotics. \( \hat{W} \) may not be equal to the variance or long run variance of the moment process. We call \( \hat{W}_T(\theta_{OT}) \) a working weighting matrix. This is in the same spirit of using a working correlation matrix rather than a true correlation matrix in the generalized estimating equations (GEE) setting. See, for example, Liang and Zeger (1986).

In parallel to (15), we construct the test statistic

\[
W_{aT} := T(\hat{R}\hat{\theta}_{aT} - r)' \left\{ R\hat{V}_{aT}^{-1} R' \right\}^{-1} (\hat{R}\hat{\theta}_{aT} - r),
\]

where, for \( \hat{G}_{aT} = \frac{1}{T} \sum_{t=1}^T \partial f(v_t, \theta)/\partial \theta' \bigg|_{\theta = \hat{\theta}_{aT}} \), \( \hat{V}_{aT} \) is defined according to

\[
\hat{V}_{aT} = \left[ \hat{G}_{aT}' \hat{W}_T^{-1}(\hat{\theta}_{aT}) \hat{G}_{aT} \right]^{-1} \left[ \hat{G}_{aT}' \hat{W}_T^{-1}(\hat{\theta}_{aT}) \hat{\Omega}_{est} \left( \hat{\theta}_{aT} \right) \hat{W}_T^{-1}(\hat{\theta}_{aT}) \hat{G}_{aT} \right] \left[ \hat{G}_{aT}' \hat{W}_T^{-1}(\hat{\theta}_{aT}) \hat{G}_{aT} \right]^{-1},
\]

which is a standard variance estimator for \( \hat{\theta}_{aT} \).

Define

\[
W^* = U'\hat{W}U \quad \text{and} \quad W = \Sigma_{1/2}^{-1} W^* (\Sigma_{1/2}^*)^{-1} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}
\]

20
and $\beta_a = W_{12} W_{22}^{-1}$.

Using the same argument for proving Lemma 9, we can show that

$$(\Sigma_{12}^T)^{-1/2} A^{-1} V' \sqrt{T} (\hat{\theta}_{aT} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{i=1}^T [f_1(v_i, \theta_0) - \beta_a f_2(v_i, \theta_0)] + o_p(1).$$ (22)

The above representation is the same as in (19) except that $\beta_\infty$ is now replaced by $\beta_a$.

Define $\tilde{\beta}_a$ according to $\beta_a = \Omega_{12}^{1/2} \tilde{\beta}_a \Omega_{22}^{-1/2} + \beta_0$. That is,

$$\tilde{\beta}_a = \Omega_{12}^{1/2} (\beta_a - \beta_0) \Omega_{22}^{1/2} = \Omega_{12}^{1/2} \beta_a \Omega_{22}^{1/2} - (I_d - \rho_\rho')^{-1/2} \rho.$$ (23)

The relationship between $\beta_a$ and $\tilde{\beta}_a$ is entirely analogous to that between $\beta_\infty$ and $\tilde{\beta}_\infty$; see (3).

In fact, if we let

$$\tilde{W} = \Omega^{-1/2} W \left( \Omega^{-1/2} \right)' = \begin{pmatrix} \tilde{W}_{11} & \tilde{W}_{12} \\ \tilde{W}_{21} & \tilde{W}_{22} \end{pmatrix},$$

then $\tilde{\beta}_a = \tilde{W}_{12} \tilde{W}_{22}^{-1}$, which is the long run regression matrix implied by the normalized weighting matrix $\tilde{W}$.

Let $V_a$ be the long run variance of $\tilde{R} [f_1(v_t, \theta_0) - \beta_a f_2(v_t, \theta_0)]$ and

$$\rho_{a,R} = V_a^{-1/2} \left[ \tilde{R} (\Omega_{12} - \beta_a \Omega_{22}) \right] \Omega_{22}^{-1/2},$$

be the long run correlation matrix between $\tilde{R} [f_1(v_t, \theta_0) - \beta_a f_2(v_t, \theta_0)]$ and $f_2(v_t, \theta_0)$. When $R = I_d$, $\rho_{a,R}$ reduces to

$$\rho_a = \left[ \Omega_{11} + \beta_a \Omega_{22} \beta_a' - \beta_a \Omega_{21} - \Omega_{12} \beta_a' \right]^{-1/2} \Omega_{12} - \beta_a \Omega_{22}) \Omega_{22}^{-1/2}.$$ (4)

Let

$$\tilde{R}_a = R V A^{-1} (\Sigma_{12}^*)^{1/2} \Omega_{12}^{1/2} = \tilde{R} \Omega_{12}^{1/2} \text{ and } \tilde{D}_R = \left( \tilde{R}_a \tilde{R}_a' \right)^{-1/2} \tilde{R}_a = \left( \tilde{R}_a \tilde{R}_a' \right)^{-1/2} \tilde{R} \Omega_{12}^{1/2}.$$ (5)

$\tilde{D}_R$ has the same dimension as $\tilde{R}_a$ and $R$. By construction, $\tilde{D}_R \tilde{D}_R' = I_p$, that is, each row of $\tilde{D}_R$ is a unit vector in $\mathbb{R}^d$ and the rows of $\tilde{D}_R$ are orthogonal to each other. The matrix $\tilde{D}_R$ signifies the orthogonal directions embodied in $\tilde{R}_a$.

With these additional notations, we are ready to state the theorem below.

**Theorem 11** Let the assumptions in Lemma 9 hold. Assume further that $\tilde{W}_T(\theta_{0T}) \rightarrow^p \tilde{W}$, a positive definite nonrandom matrix.

(a) If $E \tilde{\beta}_\infty (h, d, q) \tilde{\beta}_\infty (h, d, q) - \tilde{\beta}_a \tilde{\beta}_a'$ is positive (negative) semidefinite, then $\hat{\theta}_{2T}$ has a larger (smaller) asymptotic variance than $\theta_{aT}$, as $T \rightarrow \infty$ for a fixed $h$.

(b) If $E \tilde{\beta}_\infty (h, p, q) \tilde{\beta}_\infty (h, p, q) - \tilde{D}_R \tilde{\beta}_a \tilde{D}_R'$ is positive (negative) semidefinite, then $R \theta_{2T}$ has a larger (smaller) asymptotic variance than $R \theta_{aT}$, as $T \rightarrow \infty$ for a fixed $h$.

(c) If $g(h, d) \cdot I_d - \rho_a \rho_a'$ is positive (negative) semidefinite, then $\hat{\theta}_{2T}$ has a larger (smaller) asymptotic variance than $\theta_{aT}$, as $T \rightarrow \infty$ for a fixed $h$.

(d) If $g(h, d) \cdot I_p - \rho_{a,P} \rho_{a,R}$ is positive (negative) semidefinite, then $R \hat{\theta}_{2T}$ has a larger (smaller) asymptotic variance than $R \theta_{aT}$, as $T \rightarrow \infty$ for a fixed $h$.

(e) Let $\lambda_a = \| V_a^{-1/2} \delta_0 \|^2$. If $f(\lambda_a; h, p, q, \alpha) I_p - \rho_{a,P} \rho_{a,R}$ is positive (negative) semidefinite, then the test based on $\mathbb{W}_{2T}$ is asymptotically less (more) powerful than that based on $\mathbb{W}_{aT}$, as $T \rightarrow \infty$ for a fixed $h$.  

21
Theorem 11(a) is intuitive. While \(E\tilde{\beta}_\infty(h,d,q)\tilde{\beta}'_\infty(h,d,q)\) can be regarded as the variance inflation arising from estimating the long run regression matrix, \(\tilde{\beta}_a\tilde{\beta}'_a\) can be regarded as the bias effect from not using a true long run regression matrix. When the variance inflation effect dominates the bias effect, the two-step estimator \(\hat{\theta}_{2T}\) will have a larger asymptotic variance than \(\hat{\theta}_{aT}\).

In the special case when \(\hat{\omega} = \hat{\Omega}\), we obtain the infeasible two-step estimator \(\hat{\theta}_{2T}\) and \(\tilde{\beta}_a = 0\). In this case, \(E\tilde{\beta}_\infty(h,d,q)\tilde{\beta}'_\infty(h,d,q) \geq \tilde{\beta}_a\tilde{\beta}'_a\) holds trivially and the feasible two-step estimator \(\theta_{2T}\) has higher variation than its infeasible version \(\hat{\theta}_{2T}\).

In another special case when \(\hat{\omega} = \hat{\Sigma}\), we obtain the first step estimator \(\hat{\theta}_{1T}\) in which case \(\beta_a = 0\) and by (23), \(\tilde{\beta}_a = -(I_d - \rho\rho')^{-1/2} \rho\). Hence

\[
\tilde{\beta}_a\tilde{\beta}'_a = (I_d - \rho\rho')^{-1/2} (\rho\rho') (I_d - \rho\rho')^{-1/2}.
\]

The condition in Theorem 11(a) reduces to whether \(g(h,q) \cdot I_q - \rho\rho'\) is positive semidefinite or not. Theorem 11(a) thus reduces to Theorem 10(a).

Theorem 11(b) is a rotated version of Theorem 11(a). Given that \(\tilde{D}_R\tilde{\beta}(h,d,q) = \tilde{\beta}(h,p,q)\), the condition in Theorem 11(a) implies that in Theorem 11(b). So part (b) of the theorem is stronger than part (a) when \(\hat{R}\) is not a square matrix.

When \(\hat{\omega} = \hat{\Sigma}\), we have \(\beta_a = 0\) and so

\[
\tilde{D}_R\tilde{\beta}_a\tilde{\beta}'_a\tilde{D}'_R = \left(\tilde{R}\Omega_{1:2}\tilde{R}'\right)^{-1/2} \left[\tilde{R}\Omega_{1:2}^2\right] \left(\beta_a\beta'_a\right) \left(\tilde{R}\Omega_{1:2}^2\right)^{-1/2}
\]

\[
= \left(\tilde{R}\Omega_{1:2}\tilde{R}'\right)^{-1/2} \left[R (\beta_a - \beta_0) \Omega_{22} \{(\beta_a - \beta_0)\}' \tilde{R}' \right] \left(\tilde{R}\Omega_{1:2}\tilde{R}'\right)^{-1/2}
\]

\[
= \left[\tilde{R}\Omega_{1:2}\tilde{R}'\right]^{-1/2} \left[R \Omega_{22}^{-1} \Omega_{21} \tilde{R}' \right] \left(\tilde{R}\Omega_{1:2}\tilde{R}'\right)^{-1/2}
\]

\[
= \left[I_p - \rho_{R\rho'_R}\right]^{-1/2} \rho_{R\rho'_R} \left[I_p - \rho_{R\rho'_R}\right]^{-1/2}.
\]

As a result, the condition \(E\tilde{\beta}_\infty(h,p,q)\tilde{\beta}'_\infty(h,p,q) \geq \tilde{D}_R\tilde{\beta}_a\tilde{\beta}'_a\tilde{D}'_R\) in Theorem 11(b) is equivalent to

\[
E\tilde{\beta}_\infty(h,p,q)\tilde{\beta}'_\infty(h,p,q) \geq \left[I_p - \rho_{R\rho'_R}\right]^{-1/2} \rho_{R\rho'_R} \left[I_p - \rho_{R\rho'_R}\right]^{-1/2},
\]

which can be reduced to the same condition in Theorem 10(b).

Theorem 11(c)-(e) is entirely analogous to Theorem 10(a)-(c). The only difference is that the second block of moment conditions is removed from the first block using the implied matrix coefficient \(\tilde{\beta}_a\) before computing the long run correlation coefficient.

To understand 11(e), we can see that the effective moment conditions behind \(R\hat{\theta}_{aT}\) are:

\[
Ef_{1a}(v_t, \theta_0) = 0 \quad \text{for} \quad f_{1a}(v_t, \theta_0) = RV A^{-1} (\Sigma_{1:2}^*)^{1/2} \left[f_1(v_t, \theta_0) - \beta_a f_2(v_t, \theta_0)\right].
\]

\(R\hat{\theta}_{aT}\) uses the information in \(Ef_2(v_t, \theta_0) = 0\) to some extent, but it ignores the residual information that is still potentially available from \(Ef_2(v_t, \theta_0) = 0\). In contrast, \(R\hat{\theta}_{2T}\) attempts to explore the residual information. If there is no long run correlation between \(f_{1a}(v_t, \theta_0)\) and \(f_2(v_t, \theta_0)\), i.e., \(\rho_{aR} = 0\), then all the information in \(Ef_2(v_t, \theta_0) = 0\) has been fully captured by the effective moment conditions underlying \(R\hat{\theta}_{aT}\). As a result, the test based on \(R\hat{\theta}_{aT}\) necessarily outperforms that based on \(R\hat{\theta}_{2T}\). If the long run correlation \(\rho_{aR}\) is large enough in the sense that is given in 11(e), the test based on \(R\hat{\theta}_{2T}\) could be more powerful than that based on \(R\hat{\theta}_{aT}\) in large samples.
5.3 Practical Recommendation

Theorems 10 and 11 give theoretical comparisons of various procedures. But there are some unknown quantities in the two Theorem 10. In practice, given the set of moment conditions $E f(v_t, \theta) = 0$ and the data $\{v_t\}$, suppose that we want to test $H_0: R\theta_0 = r$ against $R\theta_0 \neq r$ for some $R \in \mathbb{R}^{d \times d}$. We can follow the steps below to evaluate the relative merit of one-step and two-step testing procedures.

1. Compute the initial estimator $\hat{\theta}_{0T} = \arg\min_{\theta} \left\| \sum_{t=1}^{T} f(v_t, \theta) \right\|^2.$

2. One the basis of $\hat{\theta}_{0T}$, use a data-driven method to select the smoothing parameter. Denote the data-driven value by $\hat{h}$.

3. Compute $\Sigma_{est}(\hat{\theta}_{0T})$ and $\bar{\Sigma}_{est}(\hat{\theta}_{0T})$ using the formulae in (14) and smoothing parameter $\hat{h}$.

4. Compute $\bar{G}_T(\hat{\theta}_{0T}) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial f(v_t, \theta)}{\partial \theta} |_{\theta = \hat{\theta}_{0T}}$ and its singular value decomposition $\hat{U}\hat{\Sigma}\hat{V}'$ where $\hat{\Sigma}' = (\hat{A}_{d \times d}, O_{d \times q})$ and $\hat{A}_{d \times d}$ is diagonal.

5. Estimate the variance and the long run variance of the rotated moment processes by $\hat{\Sigma}_* := \hat{U}'\Sigma_{est}(\hat{\theta}_{0T})\hat{U}$ and $\hat{\Sigma}_* := \hat{U}'\bar{\Sigma}_{est}(\hat{\theta}_{0T})\hat{U}$.

6. Compute the normalized LRV estimator:

$$\hat{\Omega} = (\hat{\Sigma}_{1/2})^{-1} \hat{\Sigma}_* (\hat{\Sigma}_{1/2})^{-1} := \begin{pmatrix} \hat{\Omega}_{11} & \hat{\Omega}_{12} \\ \hat{\Omega}_{21} & \hat{\Omega}_{22} \end{pmatrix}$$

where

$$\hat{\Sigma}_{1/2} = \begin{pmatrix} \hat{\Sigma}_{1/2}^{1/2} & \hat{\Sigma}_{12} \left( \hat{\Sigma}_{22} \right)^{-1/2} \\ 0 & \left( \hat{\Sigma}_{22} \right)^{1/2} \end{pmatrix}.$$

(24)

7. Let $\tilde{R}_{est} = R\hat{V}\hat{A}^{-1}(\hat{\Sigma}_{1/2})^{1/2}$ and compute

$$\hat{\rho}_R\hat{\rho}'_R = \left[ (\tilde{R}_{est}\hat{\Omega}_{11}\tilde{R}_{est})^{-1/2} \right] \left[ \tilde{R}_{est}\hat{\Omega}_{12}\tilde{R}_{22}\hat{\Omega}_{12}\tilde{R}_{est} \right] \left[ (\tilde{R}_{est}\hat{\Omega}_{11}\tilde{R}_{est})^{-1/2} \right]'$$

and its largest eigenvalue $v_{max}(\hat{\rho}_R\hat{\rho}'_R)$ and smallest eigenvalue $v_{min}(\hat{\rho}_R\hat{\rho}'_R)$.

8. Choose the value of $\lambda^0$ such that $P \left( \chi^2_p(\lambda) > \chi^1_{p} \right) = 75\%$. We may also choose a value of $\lambda$ to reflect scientific interest or economic significance.

9. (a) If $v_{min}(\hat{\rho}_R\hat{\rho}'_R) > f \left( \lambda^0; \hat{h}, p, q, \alpha \right)$, then we use the second-step testing procedure based on $\mathbb{W}_{2T}$.

(b) If $v_{max}(\hat{\rho}_R\hat{\rho}'_R) < f \left( \lambda^0; \hat{h}, p, q, \alpha \right)$, then we use the first-step testing procedure based on $\mathbb{W}_{1T}$.

(c) If either the condition in (a) or in (b) is satisfied, then we use the first-step testing procedure based on $\mathbb{W}_{aT}$.
6 Simulation Evidence

This section compares the finite sample performances of one-step and two-step estimators and tests using the fixed-smoothing approximation.

6.1 Point Estimation

We consider the location model given in (1) with the true parameter value \( \theta_0 = (0, ..., 0) \in \mathbb{R}^d \) but we allow for a nonscalar error variance. The error \( \{u_t^*\} \) follows a VAR(1) process:

\[
\begin{align*}
    u^*_{1t} &= \psi u^*_{1t-1} + \gamma \sum_{j=1}^{q} u^*_{2t-j} + e^*_{1t} \quad \text{for } i = 1, ..., d \\
    u^*_{2t} &= \psi u^*_{2t-1} + e^*_{2t} \quad \text{for } i = 1, ..., q
\end{align*}
\]

where \( e^*_{1t} \sim iid N(0,1) \) across \( i \) and \( t \), \( e^*_{2t} \sim iid N(0,1) \) across \( i \) and \( t \), and \( \{e^*_t, t = 1, 2, ..., T\} \) are independent of \( \{e^*_t, t = 1, 2, ..., T\} \). Let \( u_t^* := (u^*_{1t}, u^*_{2t})' \in \mathbb{R}^m \), then \( u_t^* = \Gamma u_{t-1}^* + e_t^* \) where

\[
\Gamma_m = \begin{pmatrix} \psi I_d & \gamma \sqrt{q} J_{d,q} \\ 0 & \psi I_q \end{pmatrix}, \quad e_t^* = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix} \sim iid N(0, I_m).
\]

Direct calculations give us the expressions for the long run and contemporaneous variances of \( \{u_t^*\} \) as

\[
\Omega^* = \sum_{j=-\infty}^{\infty} Eu_t^* u_{t-j}^* = (I_m - \Gamma)^{-1} (I_m - \Gamma')^{-1}
\]

\[
= \begin{pmatrix} \frac{1}{(1-\psi)^2} I_d + \frac{\gamma^2}{(1-\psi)^3} J_d & \frac{\gamma}{(1-\psi)^2} \sqrt{\psi} \sqrt{q} J_{d,q} \\ \frac{\gamma}{(1-\psi)^2} \sqrt{q} \sqrt{\psi} J_{q,d} & \frac{1}{(1-\psi)^2} I_q \end{pmatrix}
\]

and

\[
\Sigma^* = var(u_t^*) = \begin{pmatrix} \frac{1}{1-\psi^2} I_d + \frac{\gamma^2(1+\psi^2)}{(1-\psi^2)^3} J_d & \frac{\gamma \psi}{(1-\psi^2)^2} \sqrt{q} \sqrt{\psi} J_{d,q} \\ \frac{\gamma \psi}{q (1-\psi^2)^2} \sqrt{\psi} J_{q,d} & \frac{1}{1-\psi^2} I_q \end{pmatrix}
\]

where \( J_{d,q} \) is the \( d \times q \) matrix of ones. In our simulation, we fix the value of \( \psi \) at 0.75 so that each time series is reasonably persistent. Let \( u_{1t} = (\Sigma_{12}^*)^{-1/2} [u^*_{1t} - \Sigma_{12}^* (\Sigma_{22}^*)^{-1} u^*_{2t}] \) and \( u_{2t} = (\Sigma_{22}^*)^{-1/2} u^*_{2t} \) and \( \rho \) be the long run correlation matrix of \( u_t = (u'_{1t}, u'_{2t})' \). With some algebraic manipulations, we have

\[
\rho \rho' = \left( d + \frac{(1-\psi^2)^2}{\gamma^2} \right)^{-1} J_d.
\]

So the maximum eigenvalue of \( \rho \rho' \) is given by \( v_{\max}(\rho \rho') = \left[ 1 + (1-\psi^2)/(d\gamma^2) \right]^{-1} \), which is also the only nonzero eigenvalue. Obviously, \( v_{\max}(\rho \rho') \) is increasing in \( \gamma^2 \) for any given \( \psi \) and \( d \). Given \( \psi \) and \( d \), we choose the value of \( \gamma \) to get a different value of \( v_{\max}(\rho \rho') \). We consider \( v_{\max}(\rho \rho') = 0, 0.09, 0.18, ..., 0.90, 0.99 \) which are obtained by setting \( \gamma^2 = \left( v_{\max}(\rho \rho')(1-\psi^2)^2 / (d(1-\max(\rho \rho'))) \right) \).
For the basis functions in OS HAR estimation, we choose the following orthonormal basis functions \{\Phi_j\}_{j=1}^{\infty} in \mathbb{L}^2[0,1] space:

\[
\Phi_{2j-1}(x) = \sqrt{2} \cos(2j\pi x) \quad \text{and} \quad \Phi_{2j}(x) = \sqrt{2} \sin(2j\pi x) \quad \text{for} \quad j = 1, ..., K/2
\]

We also consider commonly used kernel based LRV estimators by employing three kernels: Bartlett, Parzen, QS kernels. In all our simulations the sample size \(T\) is 200.

We focus on the case with \(d = 1\), under which \(\rho\rho'\) is a scalar and \(\nu_{\text{max}}(\rho\rho') = \rho\rho'\). According to Proposition 3 if \(\rho\rho'\) greater than a threshold value, then \(\text{Var}(\hat{\theta}_{2T})\) will be smaller than \(\text{Var}(\hat{\theta}_{1T})\). Otherwise, \(\text{Var}(\theta_{2T})\) will be larger. We simulate \(\text{Var}(\hat{\theta}_{1T})\) and \(\text{Var}(\theta_{2T})\) using 10,000 simulation replications.

Tables 6-7 report the simulated variances under the OS and kernel HAR estimation with \(q = 3\) and 4 at some given values of the smoothing parameters. We first discuss the case when OS HAR estimation is used. It is clear that \(\text{Var}(\hat{\theta}_{2T})\) becomes smaller than \(\text{Var}(\hat{\theta}_{1T})\) only when \(\nu_{\text{max}}(\rho\rho')\) is large enough. For example, when \(q = 4\) and there is no long run correlation, i.e., \(\nu_{\text{max}}(\rho\rho') = 0\), we have \(\text{Var}(\hat{\theta}_{1T}) = 0.080 < \text{Var}(\hat{\theta}_{2T}) = 0.112\), and so \(\hat{\theta}_{1T}\) is more efficient than \(\theta_{2T}\) with 27% efficiency gain. These numerical observations are consistent with our theoretical result in Proposition 4. \(\theta_{2T}\) becomes more efficient relative to \(\theta_{1T}\) only when the benefit of using long run correlation matrix \(\rho\) outweighs the cost of estimating it. With the choice of \(K = 14\) and \(q = 4\), Table 6 indicates that \(\text{Var}(\hat{\theta}_{2T})\) starts to become smaller than \(\text{Var}(\hat{\theta}_{1T})\) when \(\nu_{\text{max}}(\rho\rho') = \rho\rho'\) crosses a value in the interval [0.270, 0.360] from below. This agrees with the theoretical threshold value \(\rho\rho' = q/(K-1) \approx 0.307\) given in Corollary 4.

In the case when kernel HAR estimation is used, we get the exactly same qualitative messages. For example, consider the case with the Bartlett kernel, \(b = 0.08\), and \(q = 3\). We observe that \(\text{Var}(\hat{\theta}_{2T})\) starts to become smaller than \(\text{Var}(\hat{\theta}_{1T})\) when \(\nu_{\text{max}}(\rho\rho') = \rho\rho'\) crosses a value in the interval [0.09, 0.18] from below. This is compatible with the threshold value 0.152 given in Table 7.

6.2 Hypothesis Testing

In addition to the VAR(1) error process, we also consider the following VARMA(1,1) process for \(u_t^i\):

\[
\begin{align*}
  u_{1t}^i &= \psi u_{1t-1}^i + e_{1t}^i + \frac{\gamma}{\sqrt{q}} \sum_{j=1}^{q} e_{2,t-1}^j \quad \text{for} \quad i = 1, ..., d \\
  u_{2t}^i &= \psi u_{2t-1}^i + e_{2t}^i \quad \text{for} \quad i = 1, ..., q
\end{align*}
\]

where \(e_{k}^i \sim i.i.d. N(0, I_m)\). The corresponding long run covariance matrix \(\Omega^*\) and contemporaneous covariance matrix \(\Sigma^*\) are

\[
\Omega^* = \begin{pmatrix}
\frac{1}{(1-\psi)^2} I_d + \frac{\gamma^2}{(1-\psi)^2} \cdot J_d & \frac{\gamma}{(1-\psi)^2 \sqrt{q}} \cdot J_{d,q} \\
\frac{\gamma}{(1-\psi)^2 \sqrt{q}} \cdot J_{q,d} & \frac{\gamma^2}{(1-\psi)^2} \cdot I_q
\end{pmatrix}
\]

\(3\)To calculate the values of \(K\) and \(b\), we employ the plug-in values \(\Omega^*\) and \(B\) in Phillips (2005) and Andrews (1991)’s AMSE-optimal formulae assuming \(\nu(\rho\rho') = 0.40\).
and
\[ \Sigma^* = \left( \begin{array}{ccc} 1 - \psi^2 & \psi^2 & \psi^2 \\ \psi^2 & 1 - \psi^2 & \psi^2 \\ \psi^2 & \psi^2 & 1 - \psi^2 \end{array} \right) \]

With some additional algebras, we have,
\[ \rho \rho' = \left( d + \frac{1}{(1 - \psi^2)^2} \right)^{-1} J_d. \]

Considering both the VAR(1) and VARMA(1,1) error processes, we implement three testing procedures on the basis of \( \mathbb{W}_{1T}, \mathbb{W}_{2T} \) and \( \mathbb{W}_{aT} \). The significance level is \( \alpha = 0.05 \) and we set \( \nu_{\text{max}}(\rho \rho') \in \{0.00, 0.35, 0.50, 0.60, 0.80, 0.90\} \) with \( d = 3 \) and \( q = 3 \). The null hypotheses of interest are:

\[ H_{01} : \beta_1 = 0, \]
\[ H_{02} : \beta_1 = \beta_2 = 0 \]

where \( p = 1, 2 \) respectively. Here, \( \mathbb{W}_{aT} \) is based on working weighting matrix \( \tilde{W}(\hat{\theta}_{QT}) \) using VAR(1) as the approximating model for the estimated error process \( \{ \hat{u}_t(\hat{\theta}_{QT}) \} \). Note that \( \tilde{W}(\hat{\theta}_{QT}) \) converges in probability to the true long run variance matrix \( \Omega^* \) if the true data generating process is VAR(1) process. However, the probability limit of \( \tilde{W}(\hat{\theta}_{QT}) \) is different from the true LRV when the true DGP is VARMA(1,1) process. For the smoothing parameters, we employ the data driven AMSE optimal bandwidth through VAR(1) plug-in implementation developed by Andrews (1991) and Phillips (2005). All our results below are based on 10,000 simulation replications.

Tables 8–15 report the empirical size of three testing procedures based on the two types of asymptotic approximations with OS HAR and kernel based estimators. It is clear that all of the three tests based on \( \mathbb{W}_{1T}, \mathbb{W}_{aT} \) and \( \mathbb{W}_{2T} \) suffer from severe size distortion if the conventional normal (or chi-square) critical values are used. For example, when the DGP is VAR(1) and OS HAR estimation is implemented, the empirical sizes of the three tests using the OS variance estimator are reported to be around 16% ~ 24% when \( p = 2 \). The relatively large size distortion of the \( \mathbb{W}_{2T} \) test comes from the additional cost in estimating the weighting matrix. However, if the nonstandard critical values \( \mathbb{W}_{aT}^{\infty} \) and \( \mathbb{W}_{2T}^{\infty} \) are used, we observe that the size distortion of both procedures is substantially reduced. The result agrees with the previous literature such as Sun (2013, 2014a&b) and Kiefer and Vogelsang (2005) which highlight the higher accuracy of the fixed-smoothing approximations. Also, we observe that when the fixed-smoothing approximations are used, the \( \mathbb{W}_{1T} \) test is more size-distorted than the \( \mathbb{W}_{2T} \) test in most cases. Similar results for the kernel cases are reported in Tables 10–15.

Next, we investigate the finite sample power performances of the three procedures. We use the finite sample critical values under the null, the power is size-adjusted and the power comparison is meaningful. The DGPs are the same as before except the parameters are from the local null alternatives \( \theta = \theta_0 + \delta_0 / \sqrt{T} \). The degree of over identification considered here is \( q = 3 \). Also, the domain of each power curve is rescaled to be \( \lambda := \delta_0 \Omega_{11}^{-1} \delta_0 \) as in Section 4.

Figures 3–6 show the size-adjusted finite sample power of both procedures by OS HAR estimation. The dashed lines in the figures indicate power curves of \( \mathbb{W}_{aT} \) procedure and \( \mathbb{W}_{1T}, \mathbb{W}_{2T} \) are represented by dash-dot and dotted lines, respectively. We can see that in all figures, the power curve of the two-step test shifts upward as the degree of the long run correlation \( \nu_{\text{max}}(\rho_{RR} \rho'_{RR}) \) increases and it starts to dominate that of the one-step test from certain point \( \nu_{\text{max}}(\rho_{RR} \rho'_{RR}) \in (0, 1) \). This is consistent with Proposition 7.
For example, with \( K = 14 \) and \( p = 1 \), the power curves in Figure 4 show that the power curve of the two-step test \( W_{2T} \) starts to cross with that of the one-step test \( W_{1T} \) when \( \nu_{\max}(\rho_R\rho_R') \) is around 0.25. This matches our theoretical results in Proposition 7 and Table 5 which indicate that the threshold value \( \max_{\lambda \in [1, 10]} f(\lambda; K, p, q, \alpha) \) is about 0.273 when \( K = 14, p = 1 \) and \( q = 3 \). Also, if \( \nu_{\max}(\rho_R\rho_R') \) is as high as 0.75, we can see that the two step test is more powerful than the one-step test in most of cases.

Lastly, when the DGP is based on VAR(1), the performance of \( W_{aT} \) dominates \( W_{1T} \) and \( W_{2T} \) in all ranges of \( \nu_{\max}(\rho_R\rho_R') \in (0, 1) \). This is because the working weighting matrix \( \bar{W}(\tilde{\theta}_0T) \) with VAR(1) approximation can almost precisely capture the true long run variance matrix \( \Omega^* \) and this improves the power of tests whenever there is some long run correlation between \( v_{1t}^* \) and \( u_{2t}^* \). However, in VARMA(1,1) DGP case, Figures 5-6 show that the advantages of \( W_{aT} \) are reduced, especially when \( \nu_{\max}(\rho_R\rho_R') \) is close to one. This is due to the misspecification bias for \( \bar{W}(\tilde{\theta}_0T) \) which is estimated differently from the true model. Nevertheless, we still observe comparable performances of \( W_{aT} \) in most of \( \nu_{\max}(\rho_R\rho_R') \) values. The other Figures 7-9 by kernel LRV estimators deliver the same quantitative messages.

7 Conclusion

In this paper we have provided more accurate and honest comparisons between the popular one-step and two-step GMM estimators and the associated inference procedures. We have given some clear guidance on when we should go one step further and use a two-step procedure. Qualitatively, we want to go one step further only if the benefit of doing so clearly outweighs the cost. When the benefit and cost comparison is not clear-cut, we recommend using the two-step procedure with a working weighting matrix.

The qualitative message of the paper is applicable more broadly. As long as there is additional nonparametric estimation uncertainty in a two-step procedure relative to the one-step procedure, we have to be very cautious about using the two-step procedure. While some asymptotic theory may indicate that the two step procedure is always more efficient, the efficiency gain may not materialize in finite samples. In fact, it may do more harm than good sometimes if we blindly use the two step procedure.

There are many extensions of the paper. We give some examples here. First, we can use the more accurate approximations to compare the continuous updating GMM and other generalized empirical likelihood estimators with the one-step and two-step estimators. While the fixed-smoothing asymptotics captures the nonparametric estimation uncertainty of the weighting matrix estimator, it does not fully capture the estimation uncertainty embodied in the first-step estimator. The source of the problem is that we do not observe the moment process and have to use the estimated moment process based on the first-step estimator to construct the nonparametric variance estimator. It is interesting to develop a further refinement to the fixed-smoothing approximation to capture the first step estimation uncertainty more adequately. Finally, it will be also very interesting to give an honest assessment of the relative merits of the OLS and GLS estimators which are popular in empirical applications.

8 Table of Selected Notations

Moment processes and their variance, LRV and implied regression matrices.
Moment Process \( \dot{u}_t = \dot{f}(v_t, \theta_0) \)  
\( u_t^* = U^* f (v_t, \theta_0) \)  
\( u_t = (\Sigma^*)^{-1/2} u_t^* \)

Variance \( \dot{\Sigma} = E\dot{u}_t \dot{u}_t' \)  
\( \Sigma^* = \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix} \)  
\( I_{d+q} \)

LRV \( \dot{\Omega} = \sum_{j=-\infty}^{\infty} E\dot{u}_t \dot{u}_t'_{-j} \)  
\( \Omega^* = \begin{pmatrix} \Omega_{11}^* & \Omega_{12}^* \\ \Omega_{21}^* & \Omega_{22}^* \end{pmatrix} \)  
\( \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \)

Implied beta \( \beta_{\Sigma^*} = \Sigma_{12}^* (\Sigma_{22}^*)^{-1} \)  
\( \beta^* = \Omega_{12}^* (\Omega_{22}^*)^{-1} \)  
\( \beta = \Omega_{12} \Omega_{22}^{-1} \)

Some limiting quantities.

\[
\tilde{\Omega}_\infty = \int_0^1 f^1_0 \int_0^1 Q_h^*(r, s) dB_m(r) dB_m(s)' = \begin{pmatrix} \Omega_{\infty,11} & \Omega_{\infty,12} \\ \Omega_{\infty,21} & \Omega_{\infty,22} \end{pmatrix} \quad \tilde{\beta}_\infty = \tilde{\beta}_\infty(h, d, q) = \tilde{\Omega}_{\infty,12} \tilde{\Omega}_{\infty,22}^{-1}
\]

\[
\Omega_\infty = \Omega_{1/2} \tilde{\Omega}_\infty \Omega_{1/2}' := \begin{pmatrix} \Omega_{\infty,11} & \Omega_{\infty,12} \\ \Omega_{\infty,21} & \Omega_{\infty,22} \end{pmatrix} \quad \beta_\infty = \beta_\infty(h, d, q) = \Omega_{\infty,12} \Omega_{\infty,22}^{-1}
\]

\[
\int_0^1 f^1_0 \int_0^1 Q_h^*(r, s) dB_{p+q}(r) dB_{p+q}(s)' = \begin{pmatrix} C_{pp} & C_{pq} \\ C_{qp} & C_{qq} \end{pmatrix} \quad D_{pp} = C_{pp} - C_{pq} C_{qq}^{-1} C_{pq}'
\]
Table 1: Threshold values \( g(h, q) \) for asymptotic variance comparison with Bartlett kernel

<table>
<thead>
<tr>
<th>( b )</th>
<th>( q = 1 )</th>
<th>( q = 2 )</th>
<th>( q = 3 )</th>
<th>( q = 4 )</th>
<th>( q = 5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>0.007</td>
<td>0.014</td>
<td>0.020</td>
<td>0.027</td>
<td>0.033</td>
</tr>
<tr>
<td>0.020</td>
<td>0.014</td>
<td>0.027</td>
<td>0.040</td>
<td>0.053</td>
<td>0.065</td>
</tr>
<tr>
<td>0.030</td>
<td>0.020</td>
<td>0.040</td>
<td>0.059</td>
<td>0.078</td>
<td>0.097</td>
</tr>
<tr>
<td>0.040</td>
<td>0.027</td>
<td>0.053</td>
<td>0.079</td>
<td>0.104</td>
<td>0.128</td>
</tr>
<tr>
<td>0.050</td>
<td>0.034</td>
<td>0.066</td>
<td>0.098</td>
<td>0.128</td>
<td>0.157</td>
</tr>
<tr>
<td>0.060</td>
<td>0.040</td>
<td>0.079</td>
<td>0.116</td>
<td>0.152</td>
<td>0.185</td>
</tr>
<tr>
<td>0.070</td>
<td>0.047</td>
<td>0.092</td>
<td>0.135</td>
<td>0.175</td>
<td>0.211</td>
</tr>
<tr>
<td>0.080</td>
<td>0.054</td>
<td>0.104</td>
<td>0.152</td>
<td>0.197</td>
<td>0.237</td>
</tr>
<tr>
<td>0.090</td>
<td>0.061</td>
<td>0.117</td>
<td>0.170</td>
<td>0.218</td>
<td>0.260</td>
</tr>
<tr>
<td>0.100</td>
<td>0.068</td>
<td>0.129</td>
<td>0.186</td>
<td>0.238</td>
<td>0.282</td>
</tr>
<tr>
<td>0.110</td>
<td>0.074</td>
<td>0.141</td>
<td>0.203</td>
<td>0.257</td>
<td>0.303</td>
</tr>
<tr>
<td>0.120</td>
<td>0.081</td>
<td>0.153</td>
<td>0.218</td>
<td>0.274</td>
<td>0.322</td>
</tr>
<tr>
<td>0.130</td>
<td>0.088</td>
<td>0.164</td>
<td>0.233</td>
<td>0.291</td>
<td>0.340</td>
</tr>
<tr>
<td>0.140</td>
<td>0.094</td>
<td>0.175</td>
<td>0.247</td>
<td>0.306</td>
<td>0.356</td>
</tr>
<tr>
<td>0.150</td>
<td>0.101</td>
<td>0.186</td>
<td>0.260</td>
<td>0.321</td>
<td>0.371</td>
</tr>
<tr>
<td>0.160</td>
<td>0.107</td>
<td>0.196</td>
<td>0.273</td>
<td>0.334</td>
<td>0.384</td>
</tr>
<tr>
<td>0.170</td>
<td>0.113</td>
<td>0.206</td>
<td>0.284</td>
<td>0.347</td>
<td>0.397</td>
</tr>
<tr>
<td>0.180</td>
<td>0.119</td>
<td>0.216</td>
<td>0.295</td>
<td>0.358</td>
<td>0.407</td>
</tr>
<tr>
<td>0.190</td>
<td>0.124</td>
<td>0.226</td>
<td>0.306</td>
<td>0.369</td>
<td>0.417</td>
</tr>
<tr>
<td>0.200</td>
<td>0.130</td>
<td>0.235</td>
<td>0.316</td>
<td>0.380</td>
<td>0.425</td>
</tr>
</tbody>
</table>

Note: \( h = 1/b \) indicates the level of smoothing and \( q \) is the degrees of overidentification. When \( pp' > g(h, q) \), then the two-step estimator \( \hat{\theta}_{2T} \) is more efficient than the one-step estimator \( \hat{\theta}_{1T} \).
Table 2: Threshold values $g(h, q)$ for asymptotic variance comparison with Parzen kernel

<table>
<thead>
<tr>
<th>$b$</th>
<th>$q = 1$</th>
<th>$q = 2$</th>
<th>$q = 3$</th>
<th>$q = 4$</th>
<th>$q = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>0.006</td>
<td>0.011</td>
<td>0.016</td>
<td>0.022</td>
<td>0.027</td>
</tr>
<tr>
<td>0.020</td>
<td>0.011</td>
<td>0.022</td>
<td>0.033</td>
<td>0.043</td>
<td>0.054</td>
</tr>
<tr>
<td>0.030</td>
<td>0.017</td>
<td>0.033</td>
<td>0.049</td>
<td>0.065</td>
<td>0.081</td>
</tr>
<tr>
<td>0.040</td>
<td>0.022</td>
<td>0.044</td>
<td>0.065</td>
<td>0.087</td>
<td>0.107</td>
</tr>
<tr>
<td>0.050</td>
<td>0.028</td>
<td>0.055</td>
<td>0.082</td>
<td>0.108</td>
<td>0.134</td>
</tr>
<tr>
<td>0.060</td>
<td>0.033</td>
<td>0.066</td>
<td>0.099</td>
<td>0.130</td>
<td>0.161</td>
</tr>
<tr>
<td>0.070</td>
<td>0.039</td>
<td>0.077</td>
<td>0.115</td>
<td>0.152</td>
<td>0.187</td>
</tr>
<tr>
<td>0.080</td>
<td>0.045</td>
<td>0.088</td>
<td>0.132</td>
<td>0.173</td>
<td>0.213</td>
</tr>
<tr>
<td>0.090</td>
<td>0.051</td>
<td>0.100</td>
<td>0.148</td>
<td>0.194</td>
<td>0.238</td>
</tr>
<tr>
<td>0.100</td>
<td>0.057</td>
<td>0.111</td>
<td>0.164</td>
<td>0.215</td>
<td>0.263</td>
</tr>
<tr>
<td>0.110</td>
<td>0.063</td>
<td>0.122</td>
<td>0.181</td>
<td>0.236</td>
<td>0.288</td>
</tr>
<tr>
<td>0.120</td>
<td>0.069</td>
<td>0.133</td>
<td>0.197</td>
<td>0.257</td>
<td>0.312</td>
</tr>
<tr>
<td>0.130</td>
<td>0.075</td>
<td>0.145</td>
<td>0.213</td>
<td>0.277</td>
<td>0.336</td>
</tr>
<tr>
<td>0.140</td>
<td>0.081</td>
<td>0.156</td>
<td>0.229</td>
<td>0.297</td>
<td>0.359</td>
</tr>
<tr>
<td>0.150</td>
<td>0.087</td>
<td>0.168</td>
<td>0.245</td>
<td>0.317</td>
<td>0.382</td>
</tr>
<tr>
<td>0.160</td>
<td>0.093</td>
<td>0.179</td>
<td>0.261</td>
<td>0.337</td>
<td>0.404</td>
</tr>
<tr>
<td>0.170</td>
<td>0.100</td>
<td>0.191</td>
<td>0.277</td>
<td>0.356</td>
<td>0.426</td>
</tr>
<tr>
<td>0.180</td>
<td>0.106</td>
<td>0.202</td>
<td>0.293</td>
<td>0.375</td>
<td>0.448</td>
</tr>
<tr>
<td>0.190</td>
<td>0.112</td>
<td>0.214</td>
<td>0.308</td>
<td>0.393</td>
<td>0.469</td>
</tr>
<tr>
<td>0.200</td>
<td>0.118</td>
<td>0.225</td>
<td>0.323</td>
<td>0.411</td>
<td>0.489</td>
</tr>
</tbody>
</table>

See notes to Table 1
Table 3: Threshold values $g(h, q)$ for asymptotic variance comparison with QS kernel

<table>
<thead>
<tr>
<th>$b$</th>
<th>$q = 1$</th>
<th>$q = 2$</th>
<th>$q = 3$</th>
<th>$q = 4$</th>
<th>$q = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.010</td>
<td>0.010</td>
<td>0.020</td>
<td>0.030</td>
<td>0.040</td>
<td>0.050</td>
</tr>
<tr>
<td>0.020</td>
<td>0.021</td>
<td>0.041</td>
<td>0.061</td>
<td>0.082</td>
<td>0.102</td>
</tr>
<tr>
<td>0.030</td>
<td>0.031</td>
<td>0.062</td>
<td>0.093</td>
<td>0.124</td>
<td>0.154</td>
</tr>
<tr>
<td>0.040</td>
<td>0.042</td>
<td>0.084</td>
<td>0.126</td>
<td>0.166</td>
<td>0.206</td>
</tr>
<tr>
<td>0.050</td>
<td>0.053</td>
<td>0.106</td>
<td>0.158</td>
<td>0.209</td>
<td>0.258</td>
</tr>
<tr>
<td>0.060</td>
<td>0.065</td>
<td>0.128</td>
<td>0.191</td>
<td>0.252</td>
<td>0.311</td>
</tr>
<tr>
<td>0.070</td>
<td>0.077</td>
<td>0.151</td>
<td>0.225</td>
<td>0.296</td>
<td>0.362</td>
</tr>
<tr>
<td>0.080</td>
<td>0.089</td>
<td>0.175</td>
<td>0.259</td>
<td>0.340</td>
<td>0.414</td>
</tr>
<tr>
<td>0.090</td>
<td>0.102</td>
<td>0.198</td>
<td>0.293</td>
<td>0.382</td>
<td>0.464</td>
</tr>
<tr>
<td>0.100</td>
<td>0.115</td>
<td>0.222</td>
<td>0.326</td>
<td>0.423</td>
<td>0.516</td>
</tr>
<tr>
<td>0.110</td>
<td>0.127</td>
<td>0.247</td>
<td>0.359</td>
<td>0.463</td>
<td>0.565</td>
</tr>
<tr>
<td>0.120</td>
<td>0.140</td>
<td>0.271</td>
<td>0.392</td>
<td>0.502</td>
<td>0.612</td>
</tr>
<tr>
<td>0.130</td>
<td>0.153</td>
<td>0.296</td>
<td>0.426</td>
<td>0.542</td>
<td>0.655</td>
</tr>
<tr>
<td>0.140</td>
<td>0.166</td>
<td>0.321</td>
<td>0.458</td>
<td>0.581</td>
<td>0.697</td>
</tr>
<tr>
<td>0.150</td>
<td>0.179</td>
<td>0.346</td>
<td>0.489</td>
<td>0.619</td>
<td>0.736</td>
</tr>
<tr>
<td>0.160</td>
<td>0.193</td>
<td>0.371</td>
<td>0.520</td>
<td>0.655</td>
<td>0.773</td>
</tr>
<tr>
<td>0.170</td>
<td>0.206</td>
<td>0.395</td>
<td>0.549</td>
<td>0.690</td>
<td>0.806</td>
</tr>
<tr>
<td>0.180</td>
<td>0.220</td>
<td>0.418</td>
<td>0.578</td>
<td>0.722</td>
<td>0.834</td>
</tr>
<tr>
<td>0.190</td>
<td>0.233</td>
<td>0.441</td>
<td>0.605</td>
<td>0.752</td>
<td>0.859</td>
</tr>
<tr>
<td>0.200</td>
<td>0.246</td>
<td>0.463</td>
<td>0.630</td>
<td>0.779</td>
<td>0.879</td>
</tr>
</tbody>
</table>

See notes to Table 1

Table 4: Values of $f(\lambda; K, p, q, \alpha)$ for $\alpha = 0.05$ and $K = 8, 10$.

<table>
<thead>
<tr>
<th>$K$</th>
<th>$p = 1$</th>
<th>$p = 2$</th>
<th>$p = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$q = 1$</td>
<td>$q = 2$</td>
<td>$q = 3$</td>
</tr>
<tr>
<td>1</td>
<td>0.144</td>
<td>0.350</td>
<td>0.514</td>
</tr>
<tr>
<td>5</td>
<td>0.152</td>
<td>0.357</td>
<td>0.494</td>
</tr>
<tr>
<td>9</td>
<td>0.154</td>
<td>0.343</td>
<td>0.489</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>0.162</td>
<td>0.354</td>
</tr>
<tr>
<td>17</td>
<td>0.168</td>
<td>0.360</td>
<td>0.501</td>
</tr>
<tr>
<td>21</td>
<td>0.171</td>
<td>0.363</td>
<td>0.509</td>
</tr>
<tr>
<td>25</td>
<td>0.165</td>
<td>0.379</td>
<td>0.526</td>
</tr>
</tbody>
</table>

Note: If the largest squared long run canonical correlation between the two blocks of moment conditions is smaller than $f(\lambda; K, p, q, \alpha)$, then the two step test is asymptotically less powerful.
Table 5: Values of $f(\lambda; K, p, q, \alpha)$ for $\alpha = 0.05$ and $K = 12, 14$.

<table>
<thead>
<tr>
<th>K</th>
<th>$\lambda$</th>
<th>$q = 1$</th>
<th>$q = 2$</th>
<th>$q = 3$</th>
<th>$q = 1$</th>
<th>$q = 2$</th>
<th>$q = 3$</th>
<th>$q = 1$</th>
<th>$q = 2$</th>
<th>$q = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.092</td>
<td>0.200</td>
<td>0.298</td>
<td>0.146</td>
<td>0.202</td>
<td>0.322</td>
<td>0.163</td>
<td>0.339</td>
<td>0.346</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.101</td>
<td>0.217</td>
<td>0.304</td>
<td>0.124</td>
<td>0.247</td>
<td>0.350</td>
<td>0.118</td>
<td>0.290</td>
<td>0.360</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.093</td>
<td>0.201</td>
<td>0.308</td>
<td>0.127</td>
<td>0.226</td>
<td>0.357</td>
<td>0.117</td>
<td>0.274</td>
<td>0.357</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.104</td>
<td>0.196</td>
<td>0.322</td>
<td>0.121</td>
<td>0.230</td>
<td>0.361</td>
<td>0.116</td>
<td>0.275</td>
<td>0.369</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.092</td>
<td>0.194</td>
<td>0.327</td>
<td>0.124</td>
<td>0.239</td>
<td>0.362</td>
<td>0.123</td>
<td>0.284</td>
<td>0.379</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.119</td>
<td>0.204</td>
<td>0.327</td>
<td>0.116</td>
<td>0.240</td>
<td>0.362</td>
<td>0.122</td>
<td>0.286</td>
<td>0.385</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.113</td>
<td>0.251</td>
<td>0.350</td>
<td>0.105</td>
<td>0.243</td>
<td>0.364</td>
<td>0.123</td>
<td>0.296</td>
<td>0.388</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.071</td>
<td>0.286</td>
<td>0.273</td>
<td>0.101</td>
<td>0.182</td>
<td>0.331</td>
<td>0.135</td>
<td>0.263</td>
<td>0.366</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.091</td>
<td>0.223</td>
<td>0.267</td>
<td>0.133</td>
<td>0.184</td>
<td>0.279</td>
<td>0.110</td>
<td>0.211</td>
<td>0.337</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.084</td>
<td>0.215</td>
<td>0.265</td>
<td>0.118</td>
<td>0.192</td>
<td>0.283</td>
<td>0.118</td>
<td>0.212</td>
<td>0.338</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0.087</td>
<td>0.199</td>
<td>0.257</td>
<td>0.107</td>
<td>0.200</td>
<td>0.283</td>
<td>0.122</td>
<td>0.216</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.087</td>
<td>0.213</td>
<td>0.265</td>
<td>0.105</td>
<td>0.196</td>
<td>0.284</td>
<td>0.128</td>
<td>0.207</td>
<td>0.334</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.103</td>
<td>0.245</td>
<td>0.262</td>
<td>0.097</td>
<td>0.190</td>
<td>0.289</td>
<td>0.119</td>
<td>0.213</td>
<td>0.337</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>0.086</td>
<td>0.244</td>
<td>0.269</td>
<td>0.095</td>
<td>0.201</td>
<td>0.295</td>
<td>0.130</td>
<td>0.205</td>
<td>0.339</td>
<td></td>
</tr>
</tbody>
</table>

Note: If the largest squared long run canonical correlation between the two blocks of moment conditions is smaller than $f(\lambda; K, p, q, \alpha)$, then the two step test is asymptotically less powerful.
Table 6: Finite sample variance comparison between $\hat{\theta}_1$ and $\hat{\theta}_2$ under VAR(1) error with $T = 200$, and $q = 3$.

<table>
<thead>
<tr>
<th>$\nu_{\max}(\rho')$</th>
<th>Var($\theta_{1T}$)</th>
<th>Var($\theta_{2T}$)</th>
<th>Var($\theta_{aT}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\cdot$</td>
<td>OS HAR</td>
<td>Bartlett</td>
<td>Parzen</td>
</tr>
<tr>
<td>$\cdot$</td>
<td>$K=14$</td>
<td>$b=0.08$</td>
<td>$b=0.15$</td>
</tr>
<tr>
<td>0.000</td>
<td>0.081</td>
<td>0.103</td>
<td>0.100</td>
</tr>
<tr>
<td>0.090</td>
<td>0.093</td>
<td>0.105</td>
<td>0.103</td>
</tr>
<tr>
<td>0.180</td>
<td>0.107</td>
<td>0.108</td>
<td>0.105</td>
</tr>
<tr>
<td>0.270</td>
<td>0.124</td>
<td>0.111</td>
<td>0.108</td>
</tr>
<tr>
<td>0.360</td>
<td>0.146</td>
<td>0.115</td>
<td>0.111</td>
</tr>
<tr>
<td>0.450</td>
<td>0.174</td>
<td>0.120</td>
<td>0.116</td>
</tr>
<tr>
<td>0.540</td>
<td>0.214</td>
<td>0.127</td>
<td>0.122</td>
</tr>
<tr>
<td>0.630</td>
<td>0.272</td>
<td>0.137</td>
<td>0.131</td>
</tr>
<tr>
<td>0.720</td>
<td>0.368</td>
<td>0.154</td>
<td>0.145</td>
</tr>
<tr>
<td>0.810</td>
<td>0.554</td>
<td>0.185</td>
<td>0.174</td>
</tr>
<tr>
<td>0.900</td>
<td>1.073</td>
<td>0.274</td>
<td>0.253</td>
</tr>
<tr>
<td>0.990</td>
<td>10.892</td>
<td>1.937</td>
<td>1.731</td>
</tr>
</tbody>
</table>

Table 7: Finite sample variance comparison between $\hat{\theta}_1$ and $\hat{\theta}_2$ under VAR(1) error with $T = 200$, and $q = 4$.

<table>
<thead>
<tr>
<th>$\nu_{\max}(\rho')$</th>
<th>Var($\theta_{1T}$)</th>
<th>Var($\theta_{2T}$)</th>
<th>Var($\theta_{aT}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\cdot$</td>
<td>OS HAR</td>
<td>Bartlett</td>
<td>Parzen</td>
</tr>
<tr>
<td>$\cdot$</td>
<td>$K=14$</td>
<td>$b=0.70$</td>
<td>$b=0.150$</td>
</tr>
<tr>
<td>0.000</td>
<td>0.081</td>
<td>0.112</td>
<td>0.104</td>
</tr>
<tr>
<td>0.090</td>
<td>0.092</td>
<td>0.114</td>
<td>0.106</td>
</tr>
<tr>
<td>0.180</td>
<td>0.106</td>
<td>0.117</td>
<td>0.108</td>
</tr>
<tr>
<td>0.270</td>
<td>0.122</td>
<td>0.124</td>
<td>0.111</td>
</tr>
<tr>
<td>0.360</td>
<td>0.146</td>
<td>0.125</td>
<td>0.115</td>
</tr>
<tr>
<td>0.450</td>
<td>0.175</td>
<td>0.130</td>
<td>0.121</td>
</tr>
<tr>
<td>0.540</td>
<td>0.217</td>
<td>0.139</td>
<td>0.129</td>
</tr>
<tr>
<td>0.630</td>
<td>0.278</td>
<td>0.151</td>
<td>0.141</td>
</tr>
<tr>
<td>0.720</td>
<td>0.379</td>
<td>0.172</td>
<td>0.160</td>
</tr>
<tr>
<td>0.810</td>
<td>0.576</td>
<td>0.213</td>
<td>0.198</td>
</tr>
<tr>
<td>0.900</td>
<td>1.128</td>
<td>0.328</td>
<td>0.305</td>
</tr>
<tr>
<td>0.990</td>
<td>11.627</td>
<td>2.538</td>
<td>2.364</td>
</tr>
</tbody>
</table>
Table 8: Empirical Size of one-step and two-step tests based on the series LRV estimator under VAR(1) error when $\psi = 0.75, p = 1 \sim 2$, and $T = 200$

<table>
<thead>
<tr>
<th>$\nu_{max}(\rho_R^{\ell})$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{1\infty}$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{1\infty}$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{2\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.128</td>
<td>0.098</td>
<td>0.151</td>
<td>0.119</td>
<td>0.187</td>
<td>0.076</td>
</tr>
<tr>
<td>0.15</td>
<td>0.135</td>
<td>0.102</td>
<td>0.135</td>
<td>0.103</td>
<td>0.177</td>
<td>0.061</td>
</tr>
<tr>
<td>0.25</td>
<td>0.135</td>
<td>0.105</td>
<td>0.127</td>
<td>0.094</td>
<td>0.174</td>
<td>0.059</td>
</tr>
<tr>
<td>0.33</td>
<td>0.139</td>
<td>0.107</td>
<td>0.086</td>
<td>0.061</td>
<td>0.154</td>
<td>0.044</td>
</tr>
<tr>
<td>0.57</td>
<td>0.143</td>
<td>0.116</td>
<td>0.046</td>
<td>0.031</td>
<td>0.118</td>
<td>0.032</td>
</tr>
</tbody>
</table>

$p = 2$ and $q = 3$

<table>
<thead>
<tr>
<th>$\nu_{max}(\rho_R^{\ell})$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{1\infty}$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{1\infty}$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{2\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.181</td>
<td>0.111</td>
<td>0.222</td>
<td>0.138</td>
<td>0.290</td>
<td>0.077</td>
</tr>
<tr>
<td>0.26</td>
<td>0.191</td>
<td>0.118</td>
<td>0.219</td>
<td>0.136</td>
<td>0.296</td>
<td>0.069</td>
</tr>
<tr>
<td>0.40</td>
<td>0.192</td>
<td>0.115</td>
<td>0.201</td>
<td>0.120</td>
<td>0.290</td>
<td>0.065</td>
</tr>
<tr>
<td>0.50</td>
<td>0.195</td>
<td>0.119</td>
<td>0.194</td>
<td>0.112</td>
<td>0.290</td>
<td>0.057</td>
</tr>
<tr>
<td>0.73</td>
<td>0.206</td>
<td>0.120</td>
<td>0.168</td>
<td>0.095</td>
<td>0.272</td>
<td>0.057</td>
</tr>
<tr>
<td>0.86</td>
<td>0.206</td>
<td>0.124</td>
<td>0.143</td>
<td>0.082</td>
<td>0.245</td>
<td>0.051</td>
</tr>
</tbody>
</table>

Table 9: Empirical Size of one-step and two-step tests based on the series LRV estimator under VARMA(1,1) error when $\psi = 0.75, p = 1 \sim 2$, and $T = 200$

<table>
<thead>
<tr>
<th>$\nu_{max}(\rho_R^{\ell})$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{1\infty}$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{1\infty}$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{2\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.117</td>
<td>0.091</td>
<td>0.138</td>
<td>0.108</td>
<td>0.181</td>
<td>0.068</td>
</tr>
<tr>
<td>0.15</td>
<td>0.140</td>
<td>0.113</td>
<td>0.142</td>
<td>0.113</td>
<td>0.173</td>
<td>0.071</td>
</tr>
<tr>
<td>0.25</td>
<td>0.144</td>
<td>0.117</td>
<td>0.140</td>
<td>0.113</td>
<td>0.165</td>
<td>0.065</td>
</tr>
<tr>
<td>0.33</td>
<td>0.155</td>
<td>0.127</td>
<td>0.141</td>
<td>0.111</td>
<td>0.160</td>
<td>0.060</td>
</tr>
<tr>
<td>0.57</td>
<td>0.167</td>
<td>0.138</td>
<td>0.128</td>
<td>0.106</td>
<td>0.121</td>
<td>0.043</td>
</tr>
<tr>
<td>0.75</td>
<td>0.168</td>
<td>0.141</td>
<td>0.118</td>
<td>0.096</td>
<td>0.087</td>
<td>0.025</td>
</tr>
</tbody>
</table>

$p = 2$ and $q = 3$

<table>
<thead>
<tr>
<th>$\nu_{max}(\rho_R^{\ell})$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{1\infty}$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{1\infty}$</th>
<th>$\chi^2$</th>
<th>$\mathbb{W}_{2\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.188</td>
<td>0.119</td>
<td>0.227</td>
<td>0.146</td>
<td>0.290</td>
<td>0.080</td>
</tr>
<tr>
<td>0.26</td>
<td>0.202</td>
<td>0.129</td>
<td>0.209</td>
<td>0.136</td>
<td>0.270</td>
<td>0.073</td>
</tr>
<tr>
<td>0.40</td>
<td>0.206</td>
<td>0.135</td>
<td>0.204</td>
<td>0.134</td>
<td>0.254</td>
<td>0.069</td>
</tr>
<tr>
<td>0.50</td>
<td>0.223</td>
<td>0.148</td>
<td>0.215</td>
<td>0.144</td>
<td>0.251</td>
<td>0.065</td>
</tr>
<tr>
<td>0.73</td>
<td>0.221</td>
<td>0.148</td>
<td>0.205</td>
<td>0.138</td>
<td>0.214</td>
<td>0.053</td>
</tr>
<tr>
<td>0.86</td>
<td>0.222</td>
<td>0.156</td>
<td>0.194</td>
<td>0.132</td>
<td>0.178</td>
<td>0.044</td>
</tr>
</tbody>
</table>
Table 10: Empirical Size of one-step and two-step tests based on the Bartlett kernel variance estimator under VAR(1) error when $\psi = 0.75$, $p = 1 \sim 2$ and $T = 200$

<table>
<thead>
<tr>
<th>$\nu_{\max}(p_{\beta})$</th>
<th>One Step($\Sigma^*$) $\chi^2$</th>
<th>One Step($\hat{W}$) $\hat{W}_{1\infty}$</th>
<th>Two Step $\gamma^2$</th>
<th>Two Step $\hat{W}_{2\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$ and $q = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.156</td>
<td>0.138</td>
<td>0.192</td>
<td>0.172</td>
</tr>
<tr>
<td>0.15</td>
<td>0.163</td>
<td>0.138</td>
<td>0.175</td>
<td>0.154</td>
</tr>
<tr>
<td>0.25</td>
<td>0.161</td>
<td>0.138</td>
<td>0.164</td>
<td>0.141</td>
</tr>
<tr>
<td>0.33</td>
<td>0.154</td>
<td>0.127</td>
<td>0.140</td>
<td>0.115</td>
</tr>
<tr>
<td>0.57</td>
<td>0.147</td>
<td>0.119</td>
<td>0.085</td>
<td>0.066</td>
</tr>
<tr>
<td>0.75</td>
<td>0.152</td>
<td>0.128</td>
<td>0.035</td>
<td>0.023</td>
</tr>
<tr>
<td>$p = 2$ and $q = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.239</td>
<td>0.183</td>
<td>0.287</td>
<td>0.228</td>
</tr>
<tr>
<td>0.26</td>
<td>0.230</td>
<td>0.166</td>
<td>0.263</td>
<td>0.196</td>
</tr>
<tr>
<td>0.40</td>
<td>0.231</td>
<td>0.169</td>
<td>0.243</td>
<td>0.170</td>
</tr>
<tr>
<td>0.50</td>
<td>0.228</td>
<td>0.161</td>
<td>0.234</td>
<td>0.159</td>
</tr>
<tr>
<td>0.73</td>
<td>0.228</td>
<td>0.157</td>
<td>0.179</td>
<td>0.118</td>
</tr>
<tr>
<td>0.86</td>
<td>0.230</td>
<td>0.159</td>
<td>0.161</td>
<td>0.108</td>
</tr>
</tbody>
</table>

Table 11: Empirical Size of one-step and two-step tests based on the Bartlett kernel variance estimator under VARMA(1,1) error when $\psi = 0.75$, $p = 1 \sim 2$ and $T = 200$

<table>
<thead>
<tr>
<th>$\nu_{\max}(p_{\beta})$</th>
<th>One Step($\Sigma^*$) $\chi^2$</th>
<th>One Step($\hat{W}$) $\hat{W}_{1\infty}$</th>
<th>Two Step $\gamma^2$</th>
<th>Two Step $\hat{W}_{2\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$ and $q = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.161</td>
<td>0.142</td>
<td>0.196</td>
<td>0.177</td>
</tr>
<tr>
<td>0.15</td>
<td>0.147</td>
<td>0.127</td>
<td>0.165</td>
<td>0.144</td>
</tr>
<tr>
<td>0.25</td>
<td>0.140</td>
<td>0.117</td>
<td>0.149</td>
<td>0.129</td>
</tr>
<tr>
<td>0.33</td>
<td>0.131</td>
<td>0.115</td>
<td>0.134</td>
<td>0.113</td>
</tr>
<tr>
<td>0.57</td>
<td>0.117</td>
<td>0.099</td>
<td>0.083</td>
<td>0.068</td>
</tr>
<tr>
<td>0.75</td>
<td>0.109</td>
<td>0.092</td>
<td>0.035</td>
<td>0.026</td>
</tr>
<tr>
<td>$p = 2$ and $q = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.235</td>
<td>0.180</td>
<td>0.292</td>
<td>0.230</td>
</tr>
<tr>
<td>0.26</td>
<td>0.213</td>
<td>0.157</td>
<td>0.239</td>
<td>0.181</td>
</tr>
<tr>
<td>0.40</td>
<td>0.203</td>
<td>0.147</td>
<td>0.224</td>
<td>0.165</td>
</tr>
<tr>
<td>0.50</td>
<td>0.205</td>
<td>0.146</td>
<td>0.209</td>
<td>0.151</td>
</tr>
<tr>
<td>0.73</td>
<td>0.191</td>
<td>0.136</td>
<td>0.167</td>
<td>0.114</td>
</tr>
<tr>
<td>0.86</td>
<td>0.190</td>
<td>0.133</td>
<td>0.147</td>
<td>0.105</td>
</tr>
</tbody>
</table>
Table 12: Empirical Size of one-step and two-step tests based on the Parzen kernel variance estimator under VAR(1) error when $\psi = 0.75$, $p = 1 \sim 2$ and $T = 200$

<table>
<thead>
<tr>
<th>$\nu_{\max}(p_R^pR_R)$</th>
<th>$\chi^2$</th>
<th>$W_{1\infty}^\psi$</th>
<th>$\chi^2$</th>
<th>$W_{1\infty}\chi^2$</th>
<th>$\chi^2$</th>
<th>$W_{2\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$ and $q = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Step($\Sigma^*$)</td>
<td>$0.00$</td>
<td>$0.145$</td>
<td>$0.108$</td>
<td>$0.182$</td>
<td>$0.139$</td>
<td>$0.214$</td>
</tr>
<tr>
<td>One Step($W$)</td>
<td>$0.15$</td>
<td>$0.148$</td>
<td>$0.105$</td>
<td>$0.173$</td>
<td>$0.125$</td>
<td>$0.223$</td>
</tr>
<tr>
<td>Two Step</td>
<td>$0.25$</td>
<td>$0.142$</td>
<td>$0.102$</td>
<td>$0.161$</td>
<td>$0.115$</td>
<td>$0.220$</td>
</tr>
<tr>
<td></td>
<td>$0.33$</td>
<td>$0.142$</td>
<td>$0.101$</td>
<td>$0.142$</td>
<td>$0.099$</td>
<td>$0.211$</td>
</tr>
<tr>
<td></td>
<td>$0.57$</td>
<td>$0.150$</td>
<td>$0.105$</td>
<td>$0.107$</td>
<td>$0.068$</td>
<td>$0.186$</td>
</tr>
<tr>
<td></td>
<td>$0.75$</td>
<td>$0.141$</td>
<td>$0.101$</td>
<td>$0.054$</td>
<td>$0.030$</td>
<td>$0.147$</td>
</tr>
<tr>
<td>$p = 2$ and $q = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Step($\Sigma^*$)</td>
<td>$0.00$</td>
<td>$0.216$</td>
<td>$0.123$</td>
<td>$0.278$</td>
<td>$0.169$</td>
<td>$0.340$</td>
</tr>
<tr>
<td>One Step($W$)</td>
<td>$0.26$</td>
<td>$0.225$</td>
<td>$0.117$</td>
<td>$0.267$</td>
<td>$0.149$</td>
<td>$0.348$</td>
</tr>
<tr>
<td>Two Step</td>
<td>$0.40$</td>
<td>$0.221$</td>
<td>$0.117$</td>
<td>$0.260$</td>
<td>$0.140$</td>
<td>$0.346$</td>
</tr>
<tr>
<td></td>
<td>$0.50$</td>
<td>$0.219$</td>
<td>$0.112$</td>
<td>$0.241$</td>
<td>$0.123$</td>
<td>$0.331$</td>
</tr>
<tr>
<td></td>
<td>$0.73$</td>
<td>$0.217$</td>
<td>$0.102$</td>
<td>$0.199$</td>
<td>$0.097$</td>
<td>$0.310$</td>
</tr>
<tr>
<td></td>
<td>$0.86$</td>
<td>$0.226$</td>
<td>$0.116$</td>
<td>$0.175$</td>
<td>$0.080$</td>
<td>$0.292$</td>
</tr>
</tbody>
</table>

Table 13: Empirical Size of one-step and two-step tests based on the Parzen kernel variance estimator under VARMA(1,1) error when $\psi = 0.75$, $p = 1 \sim 2$ and $T = 200$

<table>
<thead>
<tr>
<th>$\nu_{\max}(p_R^pR_R)$</th>
<th>$\chi^2$</th>
<th>$W_{1\infty}^\psi$</th>
<th>$\chi^2$</th>
<th>$W_{1\infty}\chi^2$</th>
<th>$\chi^2$</th>
<th>$W_{2\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$ and $q = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Step($\Sigma^*$)</td>
<td>$0.00$</td>
<td>$0.142$</td>
<td>$0.104$</td>
<td>$0.186$</td>
<td>$0.141$</td>
<td>$0.218$</td>
</tr>
<tr>
<td>One Step($W$)</td>
<td>$0.15$</td>
<td>$0.134$</td>
<td>$0.099$</td>
<td>$0.164$</td>
<td>$0.125$</td>
<td>$0.210$</td>
</tr>
<tr>
<td>Two Step</td>
<td>$0.25$</td>
<td>$0.136$</td>
<td>$0.099$</td>
<td>$0.155$</td>
<td>$0.117$</td>
<td>$0.200$</td>
</tr>
<tr>
<td></td>
<td>$0.33$</td>
<td>$0.127$</td>
<td>$0.096$</td>
<td>$0.150$</td>
<td>$0.113$</td>
<td>$0.191$</td>
</tr>
<tr>
<td></td>
<td>$0.57$</td>
<td>$0.122$</td>
<td>$0.087$</td>
<td>$0.110$</td>
<td>$0.079$</td>
<td>$0.156$</td>
</tr>
<tr>
<td></td>
<td>$0.75$</td>
<td>$0.111$</td>
<td>$0.082$</td>
<td>$0.070$</td>
<td>$0.046$</td>
<td>$0.114$</td>
</tr>
<tr>
<td>$p = 2$ and $q = 3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One Step($\Sigma^*$)</td>
<td>$0.00$</td>
<td>$0.220$</td>
<td>$0.124$</td>
<td>$0.279$</td>
<td>$0.171$</td>
<td>$0.338$</td>
</tr>
<tr>
<td>One Step($W$)</td>
<td>$0.26$</td>
<td>$0.204$</td>
<td>$0.112$</td>
<td>$0.248$</td>
<td>$0.142$</td>
<td>$0.320$</td>
</tr>
<tr>
<td>Two Step</td>
<td>$0.40$</td>
<td>$0.198$</td>
<td>$0.108$</td>
<td>$0.226$</td>
<td>$0.135$</td>
<td>$0.303$</td>
</tr>
<tr>
<td></td>
<td>$0.50$</td>
<td>$0.196$</td>
<td>$0.112$</td>
<td>$0.225$</td>
<td>$0.131$</td>
<td>$0.291$</td>
</tr>
<tr>
<td></td>
<td>$0.73$</td>
<td>$0.186$</td>
<td>$0.106$</td>
<td>$0.188$</td>
<td>$0.102$</td>
<td>$0.255$</td>
</tr>
<tr>
<td></td>
<td>$0.86$</td>
<td>$0.182$</td>
<td>$0.105$</td>
<td>$0.156$</td>
<td>$0.083$</td>
<td>$0.219$</td>
</tr>
</tbody>
</table>
Table 14: Empirical Size of one-step and two-step tests based on the QS kernel variance estimator under VAR(1) error when $\psi = 0.75$, $p = 1 \sim 2$ and $T = 200$

<table>
<thead>
<tr>
<th>$p = 1$ and $q = 3$</th>
<th>One Step($\Sigma^*$)</th>
<th>One Step($\hat{W}$)</th>
<th>Two Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_{max}(p_{R_{1}R_{2}})$</td>
<td>$\chi^2$</td>
<td>$\hat{W}_{1\infty}$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>0.00</td>
<td>0.138</td>
<td>0.107</td>
<td>0.174</td>
</tr>
<tr>
<td>0.15</td>
<td>0.138</td>
<td>0.103</td>
<td>0.164</td>
</tr>
<tr>
<td>0.25</td>
<td>0.141</td>
<td>0.106</td>
<td>0.151</td>
</tr>
<tr>
<td>0.33</td>
<td>0.135</td>
<td>0.099</td>
<td>0.145</td>
</tr>
<tr>
<td>0.57</td>
<td>0.149</td>
<td>0.110</td>
<td>0.101</td>
</tr>
<tr>
<td>0.75</td>
<td>0.132</td>
<td>0.099</td>
<td>0.049</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p = 2$ and $q = 3$</th>
<th>One Step($\Sigma^*$)</th>
<th>One Step($\hat{W}$)</th>
<th>Two Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_{max}(p_{R_{1}R_{2}})$</td>
<td>$\chi^2$</td>
<td>$\hat{W}_{1\infty}$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>0.00</td>
<td>0.210</td>
<td>0.124</td>
<td>0.265</td>
</tr>
<tr>
<td>0.26</td>
<td>0.217</td>
<td>0.122</td>
<td>0.261</td>
</tr>
<tr>
<td>0.40</td>
<td>0.216</td>
<td>0.119</td>
<td>0.244</td>
</tr>
<tr>
<td>0.50</td>
<td>0.214</td>
<td>0.114</td>
<td>0.234</td>
</tr>
<tr>
<td>0.73</td>
<td>0.204</td>
<td>0.113</td>
<td>0.188</td>
</tr>
<tr>
<td>0.86</td>
<td>0.214</td>
<td>0.121</td>
<td>0.158</td>
</tr>
</tbody>
</table>

Table 15: Empirical Size of one-step and two-step tests based on the QS kernel variance estimator under VARMA(1,1) error when $\psi = 0.75$, $p = 1 \sim 2$ and $T = 200$

<table>
<thead>
<tr>
<th>$p = 1$ and $q = 3$</th>
<th>One Step($\Sigma^*$)</th>
<th>One Step($\hat{W}$)</th>
<th>Two Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_{max}(p_{R_{1}R_{2}})$</td>
<td>$\chi^2$</td>
<td>$\hat{W}_{1\infty}$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>0.00</td>
<td>0.141</td>
<td>0.112</td>
<td>0.175</td>
</tr>
<tr>
<td>0.15</td>
<td>0.137</td>
<td>0.110</td>
<td>0.164</td>
</tr>
<tr>
<td>0.25</td>
<td>0.130</td>
<td>0.104</td>
<td>0.149</td>
</tr>
<tr>
<td>0.33</td>
<td>0.123</td>
<td>0.096</td>
<td>0.140</td>
</tr>
<tr>
<td>0.57</td>
<td>0.117</td>
<td>0.094</td>
<td>0.113</td>
</tr>
<tr>
<td>0.75</td>
<td>0.110</td>
<td>0.085</td>
<td>0.060</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p = 2$ and $q = 3$</th>
<th>One Step($\Sigma^*$)</th>
<th>One Step($\hat{W}$)</th>
<th>Two Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_{max}(p_{R_{1}R_{2}})$</td>
<td>$\chi^2$</td>
<td>$\hat{W}_{1\infty}$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>0.00</td>
<td>0.213</td>
<td>0.128</td>
<td>0.271</td>
</tr>
<tr>
<td>0.26</td>
<td>0.199</td>
<td>0.123</td>
<td>0.249</td>
</tr>
<tr>
<td>0.40</td>
<td>0.194</td>
<td>0.122</td>
<td>0.231</td>
</tr>
<tr>
<td>0.50</td>
<td>0.183</td>
<td>0.108</td>
<td>0.212</td>
</tr>
<tr>
<td>0.73</td>
<td>0.188</td>
<td>0.114</td>
<td>0.187</td>
</tr>
<tr>
<td>0.86</td>
<td>0.182</td>
<td>0.113</td>
<td>0.156</td>
</tr>
</tbody>
</table>
Figure 3: Size-adjusted power of the three tests based on the OS HAR estimator under VAR(1) error with \( p = 1, q = 3, \psi = 0.75, T = 200, \) and \( K = 14 \).
Figure 4: Size-adjusted power of the three tests based on the OS HAR estimator under VAR(1) error with $p = 2$, $q = 3$, $\psi = 0.75$, $T = 200$, and $K = 14$. 
Figure 5: Size-adjusted power of the three tests based on the OS HAR estimator under VARMA(1,1) error with $p = 1$, $q = 3$ and $\psi = 0.75$, $T = 200$ and $K = 14$. 
Figure 6: Size-adjusted power of the three tests based on the OS HAR estimator under VARMA(1,1) error with $p = 2$, $q = 3$, $\phi = 0.75$, $T = 200$, and $K = 14$. 

$\nu_{\max}(\rho_R \rho'_R) = 0.00$

$\nu_{\max}(\rho_R \rho'_R) = 0.26$

$\nu_{\max}(\rho_R \rho'_R) = 0.40$

$\nu_{\max}(\rho_R \rho'_R) = 0.50$

$\nu_{\max}(\rho_R \rho'_R) = 0.73$

$\nu_{\max}(\rho_R \rho'_R) = 0.86$
Figure 7: Size-adjusted power of the three tests based on the Bartlett HAR estimator under VAR(1) error with $p = 2$, $q = 3$, $\psi = 0.75$, $T = 200$, and $b = 0.078$. 
Figure 8: Size-adjusted power of the three tests based on the Parzen HAR estimator under VAR(1) error with $p = 2$, $q = 3$, $\psi = 0.75$, $T = 200$, and $b = 0.16$. 
Figure 9: Size-adjusted power of the three tests based on the QS HAR estimator under VAR(1) error with \( p = 2, \ q = 3, \ \psi = 0.75, \ T = 200, \) and \( b = 0.079. \)
9 Appendix of Proofs

Lemma 12 Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n}$ be any two matrix square roots of a positive finite symmetric matrix $C \in \mathbb{R}^{n \times n}$ such that $AA' = BB' = C$, then there exists an orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ such that $A = BQ$.

Proof of Lemma 12 It follows from $AA' = BB'$ that $B^{-1}AA'(B')^{-1} = I_n$. Let $Q = B^{-1}A$, then $QQ' = I_n$. That is, $Q$ is an orthogonal matrix. For such a choice of $Q$, we have $A = BQ$, as desired.

Proof of Proposition 1 Part (a) follows from Lemma 2 of Sun (2013). For part (b), we note that $\hat{\beta} \overset{d}{\rightarrow} \beta_{\infty}$ and so

$$\sqrt{T} \left( \hat{\theta}_{2T} - \theta_0 \right) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( (y_{1t} - E_{y_{1t}}) - \hat{\beta} y_{2t} \right)$$

$$= (I_d, -\hat{\beta}) \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (y_{1t} - E_{y_{1t}}) \right) \overset{d}{\rightarrow} (I_d, -\beta_{\infty}) \Omega_{1/2} B_m(1).$$

Proof of Lemma 2 For any $a \in \mathbb{R}^d$, we have

$$Ea' \hat{\beta}_{\infty} (h, d, q) \hat{\beta}_{\infty} (h, d, q)' a$$

$$= E \left[ tr a' \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_q (r) dB_q' (s) \right) \right.$$  

$$\times \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_q (r) dB_q' (s) \right)^{-2} \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_q (r) dB_d (s) \right) a \left. \right]$$

$$= E \left[ tr \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_q (r) dB_q' (s) \right)^{-2} \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_q (r) dB_d (s) \right) a' \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_d (r) dB_q' (s) \right) \right]$$

$$= E \left[ tr \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_q (r) dB_q' (s) \right)^{-2} \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_q (r) dB_d (s) \right) \right.$$  

$$\times \left[ \int_0^1 \int_0^1 Q_h^* (r, s) dB_q (r) \left[ a' dB_d (s) \right] \right] \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_q' (s) \right) \right]\right]$$

$$= \kappa(h, q) a' a,$$

where

$$\kappa(h, q) = E tr \left( \int_0^1 \int_0^1 Q_h^* (r, s) dB_q (r) dB_q' (s) \right)^{-2} \left[ \int_0^1 \int_0^1 Q_h^* (r, \tau) Q_h^* (\tau, s) d\tau \right] dB_q (r) dB_q' (s).$$
So
\[ E\tilde{\beta}_\infty (h, d, q) \tilde{\beta}_\infty (h, d, q)' = \kappa(h, q) \cdot I_d. \]
Since this holds for any \( d \), we have
\[ E\tilde{\beta}_\infty (h, 1, q) \tilde{\beta}_\infty (h, 1, q)' = \kappa(h, q). \]
It then follows that
\[ E\tilde{\beta}_\infty (h, d, q) \tilde{\beta}_\infty (h, d, q)' = \left( E \left\| \tilde{\beta}_\infty (h, 1, q) \right\|^2 \right) \cdot I_d. \]

**Proof of Proposition 3.** We have
\[ \Omega_{1,2}^{1/2} = \left( \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \right)^{1/2} \]
\[ = \left\{ \Omega_{11}^{1/2} \left[ I_d - \Omega_{11}^{-1/2} \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \left( \Omega_{11}^{-1/2} \right)' \right] \left( \Omega_{11}^{-1/2} \right)' \right\}^{1/2}. \]
Note that
\[ \Omega_{11}^{1/2} \left[ I_d - \Omega_{11}^{-1/2} \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \left( \Omega_{11}^{-1/2} \right)' \right]^{1/2} \times \left[ \Omega_{11}^{1/2} \left[ I_d - \Omega_{11}^{-1/2} \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \left( \Omega_{11}^{-1/2} \right)' \right]^{1/2} \right]' \]
is equal to \( \Omega_{1,2} \). Combining this with Lemma 12, we can represent any matrix square root of \( \Omega_{1,2} \) by
\[ \Omega_{1,2}^{1/2} = \Omega_{11}^{1/2} \left[ I_d - \Omega_{11}^{-1/2} \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \left( \Omega_{11}^{-1/2} \right)' \right]^{1/2} Q \]
for an orthogonal matrix \( Q \). Using this representation, we have
\[ \Omega_{1,2}^{-1/2} \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \left( \Omega_{1,2}^{-1/2} \right)' \]
\[ = Q' \left[ I_d - \Omega_{11}^{-1/2} \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \left( \Omega_{11}^{-1/2} \right)' \right]^{-1/2} \]
\[ \times \Omega_{11}^{-1/2} \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \left( \Omega_{11}^{-1/2} \right)' \]
\[ \times \left\{ \left[ I_d - \Omega_{11}^{-1/2} \Omega_{12} \Omega_{22}^{-1} \Omega_{21} \left( \Omega_{11}^{-1/2} \right)' \right]^{-1/2} \right\}' Q \]
\[ = Q' \left( I_d - \rho \rho' \right)^{-1/2} \left( \rho \rho' \right) \left[ I_d - \rho \rho' \right]^{-1/2} Q' \]
(27)

It then follows that \( \hat{\theta}_{2T} \) has a larger (smaller) asymptotic variance than \( \hat{\theta}_{1T} \) under the fixed-smoothing asymptotics if
\[ E\tilde{\beta}_\infty (h, d, q) \tilde{\beta}_\infty (h, d, q)' \geq Q' \left( I_d - \rho \rho' \right)^{-1/2} \left( \rho \rho' \right) \left[ I_d - \rho \rho' \right]^{-1/2} Q \]
for any orthogonal matrix \( Q \). But by Lemma 2, this is equivalent to
\[ E \left\| \tilde{\beta}_\infty (h, 1, q) \right\|^2 \cdot I_d \geq \left( I_d - \rho \rho' \right)^{-1/2} \left( \rho \rho' \right) \left[ I_d - \rho \rho' \right]^{-1/2}. \]
(29)
Note that
\[ \left( I_d - \rho \rho' \right)^{-1/2} \rho \rho' \left[ I_d - \rho \rho' \right]^{-1/2} = (I_d - \rho \rho')^{-1} - I_d. \]
the condition in (29) is equivalent to

\[
\left( E \left\| \tilde{\beta}_\infty (h, 1, q) \right\|^2 + 1 \right) I_d \geq (I_d - \rho' \rho)^{-1}
\]
or

\[
I_d - \rho' \rho \leq \frac{1}{E \left\| \tilde{\beta}_\infty (h, 1, q) \right\|^2 + 1} I_d
\]

which is the same as

\[
\rho' \rho \geq \frac{E \left\| \tilde{\beta}_\infty (h, 1, q) \right\|^2}{E \left\| \tilde{\beta}_\infty (h, 1, q) \right\|^2 + 1} I_d = g(h, q) I_d.
\]

Note that \( \rho \) is not uniquely defined as \( \Omega_{11}^{1/2} \) and \( \Omega_{22}^{1/2} \) may not be uniquely defined. However, the non-uniqueness is innocuous. According to Lemma 12, if \( \tilde{\rho} \) is another version of the long run correlation matrix, then \( \tilde{\rho} = Q_1 \rho Q_2 \) for some orthogonal matrices \( Q_1 \) and \( Q_2 \). So

\[
\tilde{\rho} \tilde{\rho}' = Q_1 \rho' Q_1'
\]

and \( \tilde{\rho} \tilde{\rho}' \geq g(h, q) I_d \) if and only if \( \rho' \rho \geq g(h, q) I_d \). ■

**Proof of Corollary 4.** For the OS LRV estimation, we have

\[
Q_h^* (r, s) = \frac{1}{K} \sum_{i=1}^{K} \Phi_i (r) \Phi_i (s)
\]

and so

\[
\int_{0}^{1} Q_h^* (r, \tau) Q_h^* (\tau, s) d\tau = \int_{0}^{1} \frac{1}{K} \sum_{i=1}^{K} \Phi_i (r) \Phi_i (\tau) \frac{1}{K} \sum_{j=1}^{K} \Phi_j (\tau) \Phi_j (s) d\tau
\]

\[
= \frac{1}{K^2} \sum_{i=1}^{K} \Phi_i (r) \Phi_i (s) = \frac{1}{K} Q_h^* (r, s).
\]

As a result,

\[
\kappa(h, q) = \frac{1}{K} E tr \left( \int_{0}^{1} \int_{0}^{1} Q_h^* (r, s) dB_q (r) dB_q' (s) \right)^{-1}.
\]

Let

\[
\xi_i = \int_{0}^{1} \Phi_i (r) dB_q (r) \sim iid N(0, I_q)
\]

then

\[
\kappa(h, q) = tr E \left[ \left( \sum_{j=1}^{K} \xi_j \xi_j' \right)^{-1} \right] = \frac{q}{K - q - 1}
\]

where the last equality follows from the mean of an inverse Wishart distribution. Using this, we have

\[
g(h, q) = \frac{\kappa(h, q)}{1 + \kappa(h, q)} = \frac{\frac{q}{K - q - 1}}{1 + \frac{q}{K - q - 1}} = \frac{q}{K - 1}.
\]
The corollary then follows from Proposition 3.

Proof of Proposition 5. It suffices to prove parts (a) and (b) as parts (c) and (d) follow from similar arguments. Part (b) is a special case of Theorem 6(a) of Sun (2014b) with \( G = [I_d, O_{d 	imes q}]' \). It remains to prove part (a). Under \( R\theta_0 = r + \delta_0 / \sqrt{T} \), we have:

\[
\sqrt{T}(R\theta_{1T} - r) = \sqrt{T}R(\theta_{1T} - \theta_0) \xrightarrow{d} R\Omega_{11}^{1/2}B_d(1) + \delta_0.
\]

Using Proposition 1(a), we have

\[
(R\hat{\Omega}_{11}R') \xrightarrow{d} R\Omega_{11}^{1/2}C_{dd}(R\Omega_{11}^{1/2})',
\]

where \( C_{dd} = \int_0^1 \int_0^1 Q_n(r, s)dB_r(r)dB_d(s)' \) and \( C_{dd} \perp B_d(1) \). Invoking the continuous mapping theorem yields

\[
\mathbb{W}_{1T} := \sqrt{T}(R\hat{\theta}_{1T} - r)'(R\hat{\Omega}_{11}R')^{-1}\sqrt{T}(R\hat{\theta}_{1T} - r) \xrightarrow{d} \left[ R\Omega_{11}^{1/2}B_d(1) + \delta_0 \right]' \left[ R\Omega_{11}^{1/2}C_{dd}(R\Omega_{11}^{1/2})' \right]^{-1} \left[ R\Omega_{11}^{1/2}B_d(1) + \delta_0 \right].
\]

Now, \( [R\Omega_{11}^{1/2}B_d(1), R\Omega_{11}^{1/2}C_{dd}(R\Omega_{11}^{1/2})'] \) is distributionally equivalent to \([\Lambda_1B_p(1), \Lambda_1C_{pp}\Lambda_1']\), and so

\[
\mathbb{W}_{1T} \xrightarrow{d} [\Lambda_1B_p(1) + \delta_0]'[\Lambda_1C_{pp}\Lambda_1']^{-1}[\Lambda_1B_p(1) + \delta_0] = [B_p(1) + \Lambda_1^{-1}\delta_0]'C_{pp}^{-1}[B_p(1) + \Lambda_1^{-1}\delta_0] \xrightarrow{d} \mathbb{W}_{1\infty}(\|\Lambda_1^{-1}\delta_0\|^2),
\]

as desired.

Proof of Proposition 6. 

Part (a) We first prove that \( P(\chi_p^2(\delta^2) > x) \) increases with \( \delta^2 \) for any integer \( p \) and \( x > 0 \). Note that

\[
P(\chi_p^2(\delta^2) > x) = \sum_{j=0}^{\infty} \frac{e^{-\delta^2/2}(\delta^2/2)^j}{j!} P(\chi_p^{2j} > x),
\]

we have

\[
\frac{\partial P(\chi_p^2(\delta^2) > x)}{\partial \delta^2} = -\frac{1}{2} \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} P(\chi_p^{2j+2} > x) + \frac{1}{2} \sum_{j=1}^{\infty} \frac{(\delta^2/2)^{j-1}}{(j-1)!} e^{-\delta^2/2} P(\chi_p^{2j} > x)
\]

\[
= -\frac{1}{2} \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} P(\chi_p^{2j+2} > x) + \frac{1}{2} \sum_{j=0}^{\infty} \frac{(\delta^2/2)^j}{j!} e^{-\delta^2/2} P(\chi_p^{2j+2} > x)
\]

\[
= 0
\]

as needed.

Let \( \phi \sim N(0, 1) \) and \( \psi \) be a zero mean random variable that is independent of \( \phi \). Using the monotonocity of \( P(\chi_p^2(\delta^2) > x) \) in \( \delta^2 \), we have

\[
P(||\phi + \psi||^2 > x) = E[P(\chi_1^2(\psi^2) > x)|\psi^2]
\]

\[
> P(\chi_1^2 > x) = P(||\phi||^2 > x)
\]

for any \( x \).
Now we proceed to prove the theorem. Note that \( B_p(1) \) and \( B_q(1) \) are independent of \( C_{pq}, C_{pp}, \) and \( C_{qq} \). Let \( D_{pp}^{-1} = \sum_{i=1}^p \lambda_{Di} d_i d_i' \) be the spectral decomposition of \( D_{pp}^{-1} \) where \( \lambda_{Di} \geq 0 \) almost surely and \( \{d_i\} \) are orthonormal in \( \mathbb{R}^p \). Then

\[
[B_p(1) - C_{pq} C_{qq}^{-1} B_q(1)]' D_{pp}^{-1} [B_p(1) - C_{pq} C_{qq}^{-1} B_q(1)] \\
= \sum_{i=1}^p \lambda_{Di} [d_i' B_p(1) - d_i' C_{pq} C_{qq}^{-1} B_q(1)]^2 = \sum_{i=1}^p \lambda_{Di} (\phi_i + \psi_i)^2
\]

where \( \phi_i = d_i' B_p(1), \psi_i = -d_i' C_{pq} C_{qq}^{-1} B_q(1) \), \( \{\phi_i\} \) is independent of \( \{\psi_i\} \) conditional on \( C_{pq}, C_{pp}, \) and \( C_{qq} \). In addition, \( \phi_i \sim \text{iid} \mathcal{N}(0, 1) \) conditionally on \( C_{pq}, C_{pp}, \) and \( C_{qq} \) and unconditionally. So for any \( x > 0 \),

\[
P(\mathbb{W}_{2\infty}(0) > x) = EP(\mathbb{W}_{2\infty}(0) > x | C_{pq}, C_{pp}, C_{qq}) \\
= EP \left( \sum_{i=1}^p \lambda_{Di} (\phi_i + \psi_i)^2 > x | C_{pq}, C_{pp}, C_{qq} \right) \\
= EP \left( \lambda_{D1} (\phi_1 + \psi_1)^2 > x - \sum_{i=2}^p \lambda_{Di} (\phi_i + \psi_i)^2 | C_{pq}, C_{pp}, C_{qq}, \{\phi_i\}_{i=2}^p, \{\psi_i\}_{i=1}^p \right) \\
\geq EP \left( \lambda_{D1} \phi_1^2 > x - \sum_{i=2}^p \lambda_{Di} (\phi_i + \psi_i)^2 | C_{pq}, C_{pp}, C_{qq}, \{\phi_i, \psi_i\}_{i=2}^p \right) \\
= EP \left( \lambda_{D1} \phi_1^2 > x - \sum_{i=2}^p \lambda_{Di} (\phi_i + \psi_i)^2 | C_{pq}, C_{pp}, C_{qq}, \{\psi_i\}_{i=2}^p \right).
\]

Using the above argument repeatedly, we have

\[
P(\mathbb{W}_{2\infty}(0) > x) \geq EP \left( \sum_{i=1}^p \lambda_{Di} \phi_i^2 > x | C_{pq}, C_{pp}, C_{qq} \right) \\
= P \left( \sum_{i=1}^p \lambda_{Di} \phi_i^2 > x \right) = P \left[ B_p(1)' D_{pp}^{-1} B_p(1) > x \right] \\
> P \left[ B_p(1)' C_{pp}^{-1} B_p(1) > x \right] = P(\mathbb{W}_{\infty}(0) > x)
\]

where the last inequality follows from the fact that \( D_{pp}^{-1} > C_{pp}^{-1} \) almost surely.

**Part (b).** Let \( C_{pp}^{-1} = \sum_{i=1}^p \lambda_{Ci} c_i' c_i \) be the spectral decomposition of \( C_{pp}^{-1} \). Since \( C_{pp} > 0 \) with probability one, \( \lambda_{Ci} > 0 \) with probability one. We have

\[
\mathbb{W}_{\infty} \left( \|\xi\|^2 \right) = D [B_p(1) + \|\xi\| e_p]' C_{pp}^{-1} [B_p(1) + \|\xi\| e_p] \\
= \sum_{i=1}^p \lambda_{Ci} [c_i' B_p(1) + \|\xi\| c_i e_p]^2
\]

where \([c_i' B_p(1) + \|\xi\| c_i e_p]^2\) follows independent noncentral chi-square distributions with noncentrality parameter \(\|\xi\|^2 (c_i e_p)^2\), conditional on \(\{\lambda_{Ci}\}_{i=1}^p\) and \(\{c_i\}_{i=1}^p\). Now consider two vectors \(\xi_1\)
and \( \xi_2 \) such that \( \|\xi_1\| < \|\xi_2\| \). We have

\[
P \left[ W_{1 \infty} \left( \|\xi_1\|^2 \right) > x \right]
= P \left\{ \sum_{i=1}^{p} \lambda_{Ci} [c'_i B_p (1) + \|\xi_1\| c'_i e_p]^2 > x \right\}
= EP \left\{ \lambda_{C1} [c'_1 B_p (1) + \|\xi_1\| c'_1 e_p]^2 > x - \sum_{i=2}^{p} \lambda_{Ci} [c'_i B_p (1) + \|\xi_1\| c'_i e_p]^2 \right\} \{\lambda_{Ci}\}_{i=1}^{p}, \{c_i\}_{i=1}^{p}
< EP \left\{ \lambda_{C1} [c'_1 B_p (1) + \|\xi_2\| c'_1 e_p]^2 > x - \sum_{i=2}^{p} \lambda_{Ci} [c'_i B_p (1) + \|\xi_1\| c'_i e_p]^2 \right\} \{\lambda_{Ci}\}_{i=1}^{p}, \{c_i\}_{i=1}^{p}
= P \left\{ \lambda_{C1} [c'_1 B_p (1) + \|\xi_2\| c'_1 e_p]^2 + \sum_{i=2}^{p} \lambda_{Ci} [c'_i B_p (1) + \|\xi_1\| c'_i e_p]^2 > x \right\}
\]

where we have used the strict monotonicity of \( P (\chi^2 (\delta^2) > x) \) in \( \delta^2 \). Repeating the above argument, we have

\[
P \left[ W_{1 \infty} \left( \|\xi_1\|^2 \right) > x \right]
< P \left\{ \lambda_{C1} [c'_1 B_p (1) + \|\xi_2\| c'_1 e_p]^2 + \lambda_{C2} [c'_2 B_p (1) + \|\xi_2\| c'_2 e_p]^2 + \sum_{i=3}^{p} \lambda_{Ci} [c'_i B_p (1) + \|\xi_1\| c'_i e_p]^2 > x \right\}
< P \left\{ \sum_{i=1}^{p} \lambda_{Ci} [c'_i B_p (1) + \|\xi_2\| c'_i e_p]^2 > x \right\}
= P \{ [B_p (1) + \xi_2]' C_{pq}^{-1} [B_p (1) + \xi_2] > x \} = P \left[ W_{1 \infty} \left( \|\xi_2\|^2 \right) > x \right]
\]
as desired.

**Part (e).** We note that

\[
W_{2 \infty} \left( \|\xi\|^2 \right)
= \left[ B_p (1) - C_{pq} C_{qq}^{-1} B_q (1) + \|\xi\| e_p \right]' D_{pp}^{-1} \left[ B_p (1) - C_{pq} C_{qq}^{-1} B_q (1) + \|\xi\| e_p \right]
= \left\{ \left[ I_p + C_{pq} C_{qq}^{-1} C_{qp} \right]^{-1/2} B_p (1) - C_{pq} C_{qq}^{-1} B_q (1) \right\} + \|\xi\| \tilde{e}_p
\]

\[
x \times \left[ I_p + C_{pq} C_{qq}^{-1} C_{qp} \right]^{1/2} D_{pp}^{-1} \left[ I_p + C_{pq} C_{qq}^{-1} C_{qp} \right]^{1/2}
\]

\[
x \times \left\{ \left[ I_p + C_{pq} C_{qq}^{-1} C_{qp} \right]^{-1/2} B_p (1) - C_{pq} C_{qq}^{-1} B_q (1) \right\} + \|\xi\| \tilde{e}_p
\]

where

\[
\tilde{e}_p = \left[ I_p + C_{pq} C_{qq}^{-1} C_{qp} \right]^{-1/2} e_p.
\]

Let \( \sum_{i=1}^{p} \lambda_i, \tilde{d}_i, \tilde{d}_i^* \) be the spectral decomposition of \( \left[ I_p + C_{pq} C_{qq}^{-1} C_{qp} \right]^{1/2} D_{pp}^{-1} \left[ I_p + C_{pq} C_{qq}^{-1} C_{qp} \right]^{1/2} \).

Define

\[
\tilde{u}_{di} = \tilde{d}_i \left[ I_p + C_{pq} C_{qq}^{-1} C_{qp} \right]^{-1/2} \left[ B_p (1) - C_{pq} C_{qq}^{-1} B_q (1) \right].
\]

50
Then conditional on $C_{pq}, C_{pp}$ and $C_{qq}$, $\tilde{\phi}_{di} \sim iid N(0, 1)$. Since the conditional distribution does not depend on $C_{pq}, C_{pp}$ and $C_{qq}$, $\tilde{\phi}_{di} \sim iid N(0, 1)$ unconditionally. Now

$$\mathbb{W}_{2\infty}(\|\xi_1\|^2)$$

$$= \sum_{i=1}^{p} \lambda_{D_i} \left\{ \tilde{d}_i^T [I_p + C_{pq}C_{qq}^{-1}C_{qp}]^{-1/2} [B_p (1) - C_{pq}C_{qq}^{-1}B_q (1)] + \|\xi_1\| \tilde{d}_i \tilde{e}_p \right\}^2$$

$$= \sum_{i=1}^{p} \lambda_{D_i} \left( \tilde{\phi}_{di} + \|\xi_1\| \tilde{d}_i \tilde{e}_p \right)^2$$

and so for two vectors $\xi_1$ and $\xi_2$ such that $\|\xi_1\| < \|\xi_2\|$ we have

$$P \left\{ \mathbb{W}_{2\infty}(\|\xi_1\|^2) > x \right\}$$

$$= EP \left\{ \sum_{i=1}^{p} \lambda_{D_i} \left( \tilde{\phi}_{di} + \|\xi_1\| \tilde{d}_i \tilde{e}_p \right)^2 > x \left| C_{pq}, C_{pp}, C_{qq} \right\}$$

$$< EP \left\{ \sum_{i=1}^{p} \lambda_{D_i} \left( \tilde{\phi}_{di} + \|\xi_2\| \tilde{d}_i \tilde{e}_p \right)^2 > x \left| C_{pq}, C_{pp}, C_{qq} \right\}$$

$$= P \left\{ \sum_{i=1}^{p} \lambda_{D_i} \left( \tilde{\phi}_{di} + \|\xi_2\| \tilde{d}_i \tilde{e}_p \right)^2 > x \right\} = P \left\{ \mathbb{W}_{2\infty}(\|\xi_2\|^2) > x \right\}.$$

**Proof of Proposition 8.** Instead of directly proving $\pi_1 (\lambda) > \pi_2 (\lambda)$ for any $\lambda > 0$, we consider the following testing problem: we observe $(Y, S) \in \mathbb{R}^{p+q} \times \mathbb{R}^{(p+q) \times (p+q)}$ with $Y \perp S$ from the following distributions:

$$Y \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N_{p+q}(\mu, \Omega) \text{ with } \mu = \begin{pmatrix} \delta_0 \\ \begin{pmatrix} (p \times 1) \end{pmatrix} \end{pmatrix}, \Omega = \begin{pmatrix} \Omega_{11} & 0 \\ \begin{pmatrix} (p \times p) \\ (q \times p) \end{pmatrix} & \Omega_{22} \end{pmatrix}$$

$$S \begin{pmatrix} S_{11} \\ S_{21} \\ S_{12} \\ S_{22} \end{pmatrix} \sim \mathcal{W}_{p+q}(K, \Omega)$$

where $\Omega_{11}$ and $\Omega_{22}$ are non-singular matrices and $\mathcal{W}_{p+q}(K, \Omega)$ is the Wishart distribution with $K$ degrees of freedom. We want to test $H_0 : \delta_0 = 0$ against $H_1 : \delta_0 \neq 0$. The testing problem is partially motivated by Das Gupta and Perlman (1974) and Marden and Perlman (1980).

The joint pdf of $(Y, S)$ can be written as

$$f(Y, S|\delta_0, \Omega_{11}, \Omega_{22})$$

$$= \alpha(\delta_0, \Omega_{11}, \Omega_{22}) h(S) \exp \left\{ -\frac{1}{2} tr \left[ \Omega_{11}^{-1} (Y_1 Y_1' + KS_{11}) + \Omega_{22}^{-1} (Y_2 Y_2' + KS_{22}) \right] + Y_1' \Omega_{11}^{-1} \delta_0 \right\}$$

for some functions $\alpha(\cdot)$ and $h(\cdot)$. It follows from the exponential structure that

$$\Pi := (Y_1, S_{11}, Y_2 Y_2' + KS_{22})$$
is a complete sufficient statistic for
\[ \Gamma := (\delta_0, \Omega_{11}, \Omega_{22}). \]
We note that \( Y_1 \sim N(\delta_0, \Omega_{11}), \) \( KS_{11} \sim \mathcal{W}_p(K, \Omega_{11}) \) and \( Y_2Y'_2 + KS_{22} \sim \mathcal{W}_q(K + 1, \Omega_{22}) \) and these three random variables are mutually independent.

Now, we define the following two test functions for testing \( H_0 : \delta_0 = 0 \) against \( H_1 : \delta_0 \neq 0 : \)
\[
\begin{align*}
\phi_1(\Pi) & := 1(\mathcal{V}_1(\Pi) > \mathcal{W}_{1\infty}^\alpha) \\
\phi_2(\Pi) & := E[1(\mathcal{W}_2(Y, S) > \mathcal{W}_{2\infty}^\alpha)|\Pi]
\end{align*}
\]
where
\[
\mathcal{V}_1(\Pi) := Y'_1S_{11}^{-1}Y_1 and \mathcal{W}_2(Y, S) := (Y_1 - S_{12}S_{22}^{-1}Y_2)'(S_{11} - S_{12}S_{22}^{-1}S_{21})^{-1}(Y_1 - S_{12}S_{22}^{-1}Y_2).
\]
We can show that the distributions of \( \mathcal{V}_1(\Pi) \) and \( \mathcal{W}_2(Y, S) \) depend on the parameter \( \Gamma \) only via \( \delta_0'\Omega_{11}^{-1}\delta_0 \). First, it is easy to show that
\[
\mathcal{W}_2(Y, S) = \left( \begin{array}{c}
Y_1 \\
Y_2
\end{array} \right)' \left( \begin{array}{cc}
S_{11} & S_{12} \\
S_{21} & S_{22}
\end{array} \right)^{-1} \left( \begin{array}{c}
Y_1 \\
Y_2
\end{array} \right) - Y'_2S_{22}^{-1}Y_2.
\]
Let
\[
\tilde{Y} := \left( \begin{array}{c}
\tilde{Y}_1 \\
\tilde{Y}_2
\end{array} \right) = \Omega_{11}^{-1/2} \left( \begin{array}{c}
Y_1 \\
Y_2
\end{array} \right) \sim N(0, I_{p+q}), \tilde{\delta} = \left( \begin{array}{c}
\Omega_{11}^{-1/2}\delta_0 \\
0
\end{array} \right) and
\]
\[
\tilde{S} := \left( \begin{array}{c}
\tilde{S}_{11} \\
\tilde{S}_{21} \\
\tilde{S}_{22}
\end{array} \right) = \Omega_{11}^{-1/2} \left( \begin{array}{cc}
S_{11} & S_{12} \\
S_{21} & S_{22}
\end{array} \right) (\Omega_{11}^{-1/2})' \sim \mathcal{W}_{p+q}(K, I_{p+q}).
\]
Then \( \tilde{Y} \perp \tilde{S} \) and
\[
\mathcal{W}_2(Y, S) = \left( \tilde{Y} + \tilde{\delta} \right)' \tilde{S}^{-1} \left( \tilde{Y} + \tilde{\delta} \right) - \tilde{Y}^'_2\tilde{S}_{22}^{-1}\tilde{Y}_2.
\]
It is now obvious that the distribution of \( \mathcal{W}_2(Y, S) \) depends on \( \Gamma \) only via \( \|\tilde{\delta}\|_2^2 \), which is equal to \( \delta_0'\Omega_{11}^{-1}\delta_0 \). Second, we have
\[
\mathcal{V}_1(\Pi) = \left( \tilde{Y}_1 + \Omega_{11}^{-1/2}\delta_0 \right)' \tilde{S}_{11}^{-1} \left( \tilde{Y}_1 + \Omega_{11}^{-1/2}\delta_0 \right)
\]
and so the distribution of \( \mathcal{V}_1(\Pi) \) depends on \( \Gamma \) only via \( \|\Omega_{11}^{-1/2}\delta_0\|_2^2 \) which is also equal to \( \delta_0'\Omega_{11}^{-1}\delta_0 \).

It is easy to show that the null distributions of \( \mathcal{V}_1(\Pi) \) and \( \mathcal{W}_2(Y, S) \) are the same as \( \mathcal{W}_{1\infty} \) and \( \mathcal{W}_{2\infty} \), respectively. In view of the critical values used, both the tests \( \phi_1(\Pi) \) and \( \phi_2(\Pi) \) have the correct level \( \alpha \). Since
\[
E\phi_1(\Pi) = P(\mathcal{V}_1(\Pi) > \mathcal{W}_{1\infty}^\alpha) and E\phi_2(\Pi) = E\{E[1(\mathcal{W}_2(Y, S) > \mathcal{W}_{2\infty}^\alpha)|\Pi]\} = P(\mathcal{W}_2(Y, S) > \mathcal{W}_{1\infty}^\alpha),
\]
the power functions of the two tests \( \phi_1(\Pi) \) and \( \phi_2(\Pi) \) are \( \pi_1(\delta_0'\Omega_{11}^{-1}\delta_0) \) and \( \pi_2(\delta_0'\Omega_{11}^{-1}\delta_0) \), respectively.

We consider a group of transformations \( G \), which consists of the elements in \( \mathcal{A}^{p \times p} := \{ A \in \mathbb{R}^p \times \mathbb{R}^p : A is a (p \times p) non-singular matrix \} \) and acts on the sample space \( \Pi := \mathbb{R}^p \times \mathcal{A}^{p \times p} \times \mathbb{R}^{q \times q} \) for the sufficient statistic \( \Pi \) through the mapping
\[
G : (Y_1, S_{11}, Y'_2, KS_{22}) \Rightarrow (AY_1, AS_{11}A', Y'_2 + KS_{22}).
\]
The induced group of transformations \( \bar{G} \) acting on the parameter space \( \Gamma := \mathbb{R}^p \times \mathbb{S}^{p \times p} \times \mathbb{S}^{q \times q} \) is given by

\[
\bar{G} : \Gamma = (\delta_0, \Omega_{11}, \Omega_{22}) \Rightarrow (A\delta_0, A\Omega_{11}A', \Omega_{22}).
\]

Our testing problem is obviously invariant to this group of transformations.

We can also show that \( \mathbb{V}(\Pi) \) is maximal invariant under \( G \). To do so, we consider two different samples \( \Pi := (Y_1, S_{11}, Y_2Y_2' + KS_{22}) \) and \( \bar{\Pi} := \left( \bar{Y}_1, \bar{S}_{11}, \bar{Y}_2Y_2' + K\bar{S}_{22} \right) \) such that \( \mathbb{V}(\Pi) = \mathbb{V}(\bar{\Pi}) \). To show that there exists an \( n \times p \) non-singular matrix \( A \) such that \( Y_1 = A\bar{Y}_1 \) and \( S_{11} = A(S_{11}A' \) whenever \( Y_1S_{11}^{-1}Y_1 = \bar{Y}_1\bar{S}_{11}^{-1}\bar{Y}_1 \).

By Theorem A9.5 (Vinograd’s Theorem) in Muirhead (2009), there exists an orthogonal \( p \times p \) matrix \( H \) such that \( S_{11}^{-1/2}Y_1 = HS_{11}^{-1/2}\bar{Y}_1 \) and this gives us the non-singular matrix \( A := S_{11}^{1/2}H\bar{S}_{11}^{-1/2} \) satisfying \( Y_1 = A\bar{Y}_1 \) and \( S_{11} = A\bar{S}_{11}A' \). Similarly, we can show that

\[
\nu(\Gamma) := (\delta'_0\Omega_{11}^{-1}\delta_0, \Omega_{22})
\]

is maximal invariant under the induced group \( G \). Therefore, restricting attention to \( G \)-invariant tests, testing \( H_0 : \delta_0 = 0 \) against \( H_1 : \delta_0 \neq 0 \) reduces to testing

\[
H'_0 : \delta'_0\Omega_{11}^{-1}\delta_0 = 0 \text{ against } H'_1 : \delta'_0\Omega_{11}^{-1}\delta_0 > 0
\]

based on the maximal invariant statistic \( \mathbb{V}(\Pi) \).

Let \( f(\mathbb{V}_1; \delta'_0\Omega_{11}^{-1}\delta_0) \) and \( f(\mathbb{V}_2; \Omega_{22}) \) be the marginal pdf’s of \( \mathbb{V}_1 := \mathbb{V}_1(\Pi) \) and \( \mathbb{V}_2 := \mathbb{V}_2(\Pi) \). By construction, \( \mathbb{V}_1(\Pi)K/(K - p + 1) \) follows the noncentral \( F \) distribution \( F_{p, K - p + 1}(\delta'_0\Omega_{11}^{-1}\delta_0) \).

So \( f(\mathbb{V}_1; \delta'_0\Omega_{11}^{-1}\delta_0) \) is the (scaled) pdf of the noncentral \( F \) distribution. It is well known that the noncentral \( F \) distribution has the Monotone Likelihood Ratio (MLR) property in \( \mathbb{V}_1 \) with respect to the parameter \( \delta'_0\Omega_{11}^{-1}\delta_0 \) (e.g. Chapter 7.9 in Lehmann and Romano (2008)). Also, in view of the independence between \( \mathbb{V}_1 \) and \( \mathbb{V}_2 \), the joint distribution of \( \mathbb{V}(\Pi) \) also has the MLR property in \( \mathbb{V}_1 \). By the virtue of the Neyman-Pearson lemma, the test \( \phi_1(\Pi) := 1(\mathbb{V}_1(\Pi) > \mathbb{W}_{1\alpha}^{\infty}) \) is the unique Uniformly Most Powerful Invariant (UMPI) test among all \( G \)-invariant tests based on the complete sufficient statistic \( \Pi \). So if \( \phi_2(\Pi) \) is equivalent to a \( G \)-invariant test, then \( \pi_2(\delta'_0\Omega_{11}^{-1}\delta_0) > \pi_2(\delta'_0\Omega_{11}^{-1}\delta_0) \) for any \( \delta'_0\Omega_{11}^{-1}\delta_0 > 0 \). To show that \( \phi_2(\Pi) \) has this property, we let \( g \in G \) be any element of \( G \) with the corresponding matrix \( A_g \) and induced transformation \( \bar{g} \in \bar{G} \). Then,

\[
E_{\Gamma}[\phi_2(g\Pi)] = E_{\Gamma}[\phi_2(\Pi)] = \pi_2((A_g\delta_0)'(A_g\Omega_{11}A'_g)^{-1}(A_g\delta_0)) = \pi_2(\delta'_0\Omega_{11}^{-1}\delta_0) = E_{\Gamma}[\phi_2(\Pi)]
\]

for all \( \Gamma \). It follows from the completeness of \( \Pi \) that \( \phi_2(g\Pi) = \phi_2(\Pi) \) almost surely and this drives the desired result.

**Proof of Lemma 9.** We prove a more general result by establishing a representation for

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \bar{G}'\bar{M}^{-1}\bar{G} \right]^{-1} \bar{G}'\bar{M}^{-1}\bar{f}(v_t, \theta_0)
\]

in terms of the rotated and normalized moment conditions for any \( m \times m \) (almost surely) positive definite matrix \( \bar{M} \) which can be random. Let

\[
M^* = U'\bar{M}U, M = \Sigma_{1/2}^* M^* \Sigma_{1/2}^{*\prime} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}
\]

53
and $M_{12} = M_{11} - M_{12}M_{22}^{-1}M_{21}$. Using the SVD $U\Sigma V'$ of $\tilde{G}$, we have
\[
\tilde{G}'\tilde{M}^{-1}\tilde{G} = V\Xi'(U'\tilde{M}U)^{-1}\Xi V' \\
= VA\left( I_d, O \right) (M^*)^{-1} ( I_d, O )' AV' \\
= VA\left( I_d, O \right) (\Sigma_{1/2})' \left( \left[ \Sigma_{1/2}^{-1} M^*(\Sigma_{1/2}^*)' \right] \right)^{-1} \Sigma_{1/2} ( I_d, O )' AV' \\
= VA\left( I_d, O \right) (\Sigma_{1/2})' M^{-1} \Sigma_{1/2} ( I_d, O )' AV' \\
= VA(\Sigma_{12})^{-1/2} \left( I_d, O \right) M^{-1} \left[ VA(\Sigma_{12})^{-1/2} \left( I_d, O \right) \right]' \\
= VA(\Sigma_{12})^{-1/2} M_{12}^{-1} (\Sigma_{12}^*)^{-1/2} AV',
\]
where we have used
\[
( I_d, O ) (\Sigma_{1/2}^*)' = \begin{pmatrix} (\Sigma_{12})^{-1/2} \\ (\Sigma_{12}^*)^{-1/2} \end{pmatrix}
\]
In addition,
\[
\tilde{G}'\tilde{M}^{-1}\tilde{f}(v_t, \theta_0) = V\Xi'(U'\tilde{M}U)^{-1}U'\tilde{f}(v_t, \theta_0) = VA\left( I_d, O \right) (M^*)^{-1} f^*(v_t, \theta_0) \\
= VA\left( I_d, O \right) (\Sigma_{1/2})' \left[ \left( \Sigma_{1/2}^{-1} M^*(\Sigma_{1/2}^*)' \right) \right]^{-1} \Sigma_{1/2} f^*(v_t, \theta_0) \\
= VA\left( I_d, O \right) (\Sigma_{1/2})' M^{-1} f(v_t, \theta_0) = VA(\Sigma_{12})^{-1/2} \left( I_d, O \right) M^{-1} f(v_t, \theta_0) \\
= VA(\Sigma_{12})^{-1/2} M_{12}^{-1} \left[ f_1(v_t, \theta_0) - M_{12}M_{22}^{-1} f_2(v_t, \theta_0) \right].
\]
Hence
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \tilde{G}'\tilde{M}^{-1}\tilde{G} \right]^{-1} \tilde{G}'\tilde{M}^{-1}\tilde{f}(v_t, \theta_0) \\
= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ VA(\Sigma_{12})^{-1/2} M_{12}^{-1} (\Sigma_{12}^*)^{-1/2} AV' \right]^{-1} \left[ VA(\Sigma_{12})^{-1/2} M_{12}^{-1} \right] \\
\times \left[ f_1(v_t, \theta_0) - M_{12}M_{22}^{-1} f_2(v_t, \theta_0) \right] \\
= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} VA^{-1}(\Sigma_{12})^{1/2} \left[ f_1(v_t, \theta_0) - M_{12}M_{22}^{-1} f_2(v_t, \theta_0) \right].
\]

Let $\tilde{M} = \tilde{\Sigma}$, we have $M^* = U'\Sigma U = \Sigma$ and $M = \Sigma_{1/2}^{-1} M^*(\Sigma_{1/2}^*)' = I_m$. So $M_{12}M_{22}^{-1} = 0$. As a result
\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \tilde{G}'\tilde{M}^{-1}\tilde{G} \right]^{-1} \tilde{G}'\tilde{M}^{-1}\tilde{f}(v_t, \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} VA^{-1}(\Sigma_{12})^{1/2} f_1(v_t, \theta_0).
\]
\[ \tilde{G}' \tilde{M}^{-1} \tilde{G} = VA (\Sigma_{1,2})^{-1/2} (\Sigma_{1,2})^{-1/2} AV'. \]

Using this and the stochastic expansion of \( \sqrt{T}(\hat{\theta}_{1T} - \theta_0) \), we have

\[
\sqrt{T}(\hat{\theta}_{1T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} VA^{-1} (\Sigma_{1,2})^{1/2} f_1(v_t, \theta_0) + o_p(1).
\]

It then follows that

\[
(\Sigma_{1,2})^{-1/2} AV' \sqrt{T}(\hat{\theta}_{1T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_1(v_t, \theta_0) + o_p(1) \overset{d}{\rightarrow} N(0, \Omega_{11}).
\]

Let \( \tilde{M} = \tilde{\Omega}_\infty \), we have \( M = \Sigma_{1/2} U' \tilde{\Omega}_\infty U \Sigma_{1/2}^{-1} = \Omega_\infty \), and so \( M_1 M_2^{-1} = \Omega_{12} \Omega_{22}^{-1} = \beta_\infty \). As a result

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \tilde{G}' \tilde{\Omega}_\infty^{-1} \tilde{G}\right]^{-1} \tilde{G}' \tilde{\Omega}_\infty^{-1} \tilde{f}(v_t, \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} VA^{-1} (\Sigma_{1,2})^{1/2} [f_1(v_t, \theta_0) - \beta_\infty f_2(v_t, \theta_0)].
\]

Using this, we have

\[
\sqrt{T}(\hat{\theta}_{2T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \tilde{G}' \tilde{\Omega}_\infty^{-1} \tilde{G}\right]^{-1} \tilde{G}' \tilde{\Omega}_\infty^{-1} \tilde{f}(v_t, \theta_0) + o_p(1)
\]

\[
= \frac{1}{\sqrt{T}} VA^{-1} (\Sigma_{1,2})^{1/2} \sum_{t=1}^{T} (f_1(v_t, \theta_0) - \beta_\infty f_2(v_t, \theta_0)) + o_p(1).
\]

It then follows that

\[
(\Sigma_{1,2})^{-1/2} AV' \sqrt{T}(\hat{\theta}_{2T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [f_1(v_t, \theta_0) - \beta_\infty f_2(v_t, \theta_0)] + o_p(1)
\]

\[
\overset{d}{\rightarrow} MN \left( 0, \Omega_{11} - \Omega_{12} \beta_\infty' - \beta_\infty \Omega_{21} + \beta_\infty \Omega_{22} \beta_\infty' \right).
\]

**Proof of Theorem 10.** Parts (a) and (b). Instead of comparing the asymptotic variances of \( R \sqrt{T}(\hat{\theta}_{1T} - \theta_0) \) and \( R \sqrt{T}(\hat{\theta}_{2T} - \theta_0) \) directly, we equivalently compare the asymptotic variances of \( (RA^{-1} \Sigma_{1,2} A^{-1} V' R')^{-1/2} R \sqrt{T}(\hat{\theta}_{1T} - \theta_0) \) and \( (RA^{-1} \Sigma_{1,2} A^{-1} V' R')^{-1/2} R \sqrt{T}(\hat{\theta}_{2T} - \theta_0) \). We can do so because \( (RA^{-1} \Sigma_{1,2} A^{-1} V' R')^{-1/2} \) is nonsingular. Note that the latter two asymptotic variances are the same as those of the respective one-step estimator \( \hat{\theta}_{1T}^R \) and two-step estimator \( \hat{\theta}_{2T}^R \) of \( \theta_0^R \) in the following simple location model:

\[
\begin{cases}
  y_{1T}^R = \theta_{0}^R + u_{1T}^R & \in \mathbb{R}^p \\
  y_{2T} = u_{2T} & \in \mathbb{R}^q
\end{cases}
\]

where

\[
\begin{align*}
\theta_{0}^R &= (RA^{-1} \Sigma_{1,2} A^{-1} V' R')^{-1/2} \theta_0, \\
u_{1T}^R &= (RA^{-1} \Sigma_{1,2} A^{-1} V' R')^{-1/2} RA^{-1} (\Sigma_{1,2})^{1/2} u_{1T}
\end{align*}
\]

55
and the (contemporaneous) variance and long run variance of \( u_t = (u_{1t}', u_{2t}')' \) are \( I_m \) and \( \Omega \) respectively.

It suffices to compare the asymptotic variances of \( \hat{\theta}^R_{1T} \) and \( \hat{\theta}^R_{2T} \) in the above location model. By construction, the variance of \( u_t^R := (u_{1t}^R, u_{2t}^R)' \) is

\[
\text{var}(u_t^R) = \begin{pmatrix} I_p & O \\ O & I_q \end{pmatrix} = I_{p+q}.
\]

So the above location model has exactly the same form as the model in Section 3. We can invoke Proposition 3 to complete the proof. The long run variance of \( u_t^R \) is

\[
\text{lrvar}(u_t^R) = \begin{pmatrix} \Omega_{11}^R & \Omega_{12}^R \\ \Omega_{21}^R & \Omega_{22}^R \end{pmatrix}
\]

where

\[
\begin{align*}
\Omega_{11}^R &= \left[RVA^{-1} \Sigma_{1,2}^* A^{-1} V' R'\right]^{-1/2} \left[RVA^{-1} (\Sigma_{1,2}^*)^{1/2}\right] \\
\Omega_{12}^R &= \left[RVA^{-1} \Sigma_{1,2}^* A^{-1} V' R'\right]^{-1/2} \left[RVA^{-1} (\Sigma_{1,2}^*)^{1/2}\right]^{1/2} \\
\Omega_{22}^R &= \left[RVA^{-1} \Sigma_{1,2}^* A^{-1} V' R'\right]^{-1/2} \left[RVA^{-1} (\Sigma_{1,2}^*)^{1/2}\right]^{1/2} \Omega_{12} + \Omega_{22}^R = \Omega_{22}.
\end{align*}
\]

The implied long run correlation matrix is then

\[
\begin{align*}
\left(\Omega_{11}^R\right)^{-1/2} \Omega_{12}^R (\Omega_{22}^R)^{-1/2} &= \begin{pmatrix} (RVA^{-1} \Sigma_{1,2}^* A^{-1} V' R')^{-1/2} \left[RVA^{-1} (\Sigma_{1,2}^*)^{1/2}\right] \Omega_{11} \\
\times \Omega_{11} & \times \left[RVA^{-1} (\Sigma_{1,2}^*)^{1/2}\right]' \left[RVA^{-1} \Sigma_{1,2}^* A^{-1} V' R'\right]^{-1/2} \left[RVA^{-1} (\Sigma_{1,2}^*)^{1/2}\right] \Omega_{12} \right) \times \Omega_{22}^{-1/2} \\
&= \left(\tilde{R} \tilde{R}\right)^{-1/2} \tilde{R} \tilde{\Omega}_{11} \tilde{R}' \left(\tilde{R} \tilde{R}\right)^{-1/2} \left(\tilde{R} \tilde{R}\right)^{-1/2} \tilde{R} \tilde{\Omega}_{12} \right) \times \Omega_{22}^{-1/2} \\
&= \left(\left(\tilde{R} \tilde{R}\right)^{-1/2} \left(\tilde{R} \tilde{R}\right)^{1/2}\right) \left(\tilde{R} \tilde{R}\right)^{-1/2} \tilde{R} \tilde{\Omega}_{12} \right) \times \Omega_{22}^{-1/2} \\
&= \left(\tilde{R} \tilde{\Omega}_{11} \tilde{R}'\right)^{-1/2} \tilde{R} \tilde{\Omega}_{12} \times \Omega_{22}^{-1/2} = \rho_R.
\end{align*}
\]

Part (b) then follows from Proposition 3.

In the above calculation we have used a particular form of \( \left(\tilde{R} \tilde{R}\right)^{-1/2} \tilde{R} \tilde{\Omega}_{11} \tilde{R}' \left(\tilde{R} \tilde{R}\right)^{-1/2} \right)^{1/2}. \)

If we employ a different matrix square root, then \( \left(\Omega_{11}^R\right)^{-1/2} \Omega_{12}^R (\Omega_{22}^R)^{-1/2} = Q \rho_R \) for some \( p \times p \) orthogonal matrix \( Q \). This is innocuous as our result is rotation invariant.

In the special case when \( R = I_d \), \( \tilde{R} \) is a square matrix, and we have

\[
\rho_R = \left(\tilde{R} \tilde{\Omega}_{11} \tilde{R}'\right)^{-1/2} \left(\tilde{R} \tilde{\Omega}_{12} \right) \Omega_{22}^{-1/2} \]
\[
= \left(\tilde{R} \tilde{\Omega}_{11}^{-1/2}\right) \left(\tilde{R} \tilde{\Omega}_{12} \right) \Omega_{22}^{-1/2} = \Omega_{11}^{-1/2} \Omega_{12} \Omega_{22}^{-1/2} = \rho.
\]

56
So Part (a) is a special case of Part (b). Again, if \((\hat{R}\Omega_{11}\hat{R})^{1/2}\) takes another form instead of \(R\Omega_{11}^{1/2}\), then \(\rho_R = Q\rho\) for some orthogonal matrix \(Q\). Part (a) is still a special case of part (b) because of the rotational invariance.

**Part (c).** The local asymptotic power of the one-step test and two step test are the same as the local asymptotic power of respective one-step and two-step tests in the location model given in \((33)\). Part (c) follows from Proposition \(7\) with

\[
\lambda = \left[( RVA^{-1}\Sigma_1^* A^{-1} V'^{T} R')^{-1/2} \delta_0 \right] (\Omega_{11})^{-1} \left[( RVA^{-1}\Sigma_1^* A^{-1} V'^{T} R')^{-1/2} \delta_0 \right]^{-1} \delta_0.
\]

It remains to show that the last term is equal to what is defined in \((21)\).

Using \((30)\) with \(M^{-1} = \hat{\Sigma}^{-1}\), we have \(M = I_m\) and

\[
\hat{G}'\hat{\Sigma}^{-1}\hat{G} = VA (\Sigma_1^*)^{-1} AV'.
\]

Using \((30)\) again but with \(M^{-1} = \hat{\Sigma}^{-1}\hat{\Omega}\hat{\Sigma}^{-1}\), we obtain the corresponding \(M\) as

\[
M = \Sigma_1^{-1/2} U'\hat{M}U (\Sigma_1^{-1})' = \Sigma_1^{-1/2} U'\hat{\Sigma}^{1/2} \hat{\Sigma}\hat{U} (\Sigma_1^{-1})' = \Sigma_1^{-1/2} U'\hat{\Sigma}^{1/2} (\Sigma_1^{-1})' = \left((\Sigma)^{-1/2}\right)' \hat{\Sigma}^{1/2} (\Sigma_1^{-1})' = \left((\Sigma)^{-1/2}\right)' \hat{\Sigma}^{-1/2}\hat{\Omega}\hat{\Sigma}^{-1} = \Sigma_1^{-1/2} \hat{\Omega}\hat{\Sigma}^{-1} = \Sigma_1^{-1/2} \Omega_{11}^{-1} (\Sigma_1^{-1/2})^{-1/2} AV'.
\]

which implies that \(M^{-1} = \Omega_{11}\). So

\[
\hat{G}' (\Sigma^{-1}\hat{\Omega}\Sigma^{-1})^{-1} \hat{G} = VA (\Sigma_1^*)^{-1/2} \Omega_{11} (\Sigma_1^*)^{-1/2} AV'.
\]

Therefore,

\[
\delta_0 \left[R \left(\hat{G}'\hat{\Sigma}^{-1}\hat{G} \right)^{-1} \left(\hat{G}'\hat{\Sigma}^{-1}\hat{\Omega}\hat{\Sigma}^{-1}\hat{G} \right)^{-1} R' \right]^{-1} \delta_0 = \delta_0 \left[R \left[VA^{-1}\Sigma_1 A^{-1} V' \right] \left[VA (\Sigma_1^*)^{-1/2} \Omega_{11} (\Sigma_1^*)^{-1/2} AV' \right] \left[VA^{-1}\Sigma_1 A^{-1} V' \right]^{-1} \right]^{-1} \delta_0 = \delta_0 \left[R \left[VA^{-1}\Sigma_1 A^{-1} V' \right] \Omega_{11} \left[VA^{-1}\Sigma_1 A^{-1} V' \right]^{-1} \right]^{-1} \delta_0,
\]

which is exactly the same as \(\lambda\) given in \((34)\). \(\blacksquare\)

**Proof of Theorem 11.** Part (a). Plugging \(\tilde{M} = \hat{W}\) into \((31)\) and using the similar notation, we have

\[
\sqrt{T}(\hat{\theta}_a t - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\hat{G}'\hat{W}^{-1}\hat{G} \right]^{-1} \hat{G}'\hat{W}^{-1} f(v_t, \theta_0) + o_p(1)
\]

\[
= VA^{-1} (\Sigma_1^*)^{1/2} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[f_1(v_t, \theta_0) - \beta_1 f_2(v_t, \theta_0) \right] + o_p(1).
\]
So we have

\[
(S_{1:2}^*)^{-1/2} AV' \sqrt{T} (\hat{\theta}_{aT} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [f_1(v_t, \theta_0) - \beta_0 f_2(v_t, \theta_0)] + (\beta_a - \beta_0) f_2(v_t, \theta_0) + o_p(1)
\]

\[
(S_{1:2}^*)^{-1/2} AV' \sqrt{T} (\hat{\theta}_{2T} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [f_1(v_t, \theta_0) - \beta_0 f_2(v_t, \theta_0)] + (\beta_\infty - \beta_0) f_2(v_t, \theta_0) + o_p(1)
\]

(35)

where the first term in each summation has no long run correlation with the second term. It then follows that \( \sqrt{T} (\hat{\theta}_{2T} - \theta_0) \) has a larger asymptotic variance than \( \sqrt{T} (\hat{\theta}_{aT} - \theta_0) \) if

\[
E (\beta_\infty - \beta_0) \Omega_{22} (\beta_\infty - \beta_0)' > (\beta_a - \beta_0) \Omega_{22} (\beta_a - \beta_0)'.
\]

By definition,

\[
\beta_\infty = \Omega_{1:2}^{1/2} \beta_\infty \Omega_{22}^{-1/2} + \beta_0 \quad \text{and} \quad \beta_a = \Omega_{1:2}^{1/2} \beta_a \Omega_{22}^{-1/2} + \beta_0.
\]

The inequality is then equivalent to

\[
E \Omega_{1:2}^{1/2} \beta_\infty \beta_\infty' \Omega_{1:2}^{1/2} > \Omega_{1:2}^{1/2} \beta_a \beta_a' \Omega_{1:2}^{1/2},
\]

which simplifies to \( E \beta_\infty \beta_\infty' > \beta_a \beta_a \) as desired.

**Part (b).** It follows from part (a) that

\[
\text{asymvar} \left[ R \sqrt{T} (\hat{\theta}_{2T} - \theta_0) \right] - \text{asymvar} \left[ R \sqrt{T} (\hat{\theta}_{aT} - \theta_0) \right] = \text{ERVA}^{-1} (\Sigma_{1:2}^*)^{1/2} \left[ (\beta_a - \beta_0) \Omega_{22} (\beta_a - \beta_0)' - (\beta_\infty - \beta_0) \Omega_{22} (\beta_\infty - \beta_0)' \right] \left[ \text{ERVA}^{-1} (\Sigma_{1:2}^*)^{1/2} \right]'
\]

\[
= \text{ERVA}^{-1} (\Sigma_{1:2}^*)^{1/2} \left( \beta_\infty \beta_\infty' - \beta_a \beta_a' \right) \left[ \text{ERVA}^{-1} (\Sigma_{1:2}^*)^{1/2} \right]'
\]

\[
= \text{ERVA}^{-1} (\Sigma_{1:2}^*)^{1/2} \left( \beta_\infty \beta_\infty' - \beta_a \beta_a' \right) \hat{R}_a
\]

\[
= \left( \hat{R}_a \hat{R}_a' \right)^{1/2} \text{ERVA}^{-1} (\Sigma_{1:2}^*)^{1/2} \left( \beta_\infty \beta_\infty' - \beta_a \beta_a' \right) \left[ \hat{R}_a \hat{R}_a' \right]^{-1/2} \hat{R}_a
\]

\[
\text{asymvar} \left[ R \sqrt{T} (\hat{\theta}_{aT} - \theta_0) \right] > \text{asymvar} \left[ R \sqrt{T} (\hat{\theta}_{aT} - \theta_0) \right] \quad \text{if}
\]

\[
E \beta_\infty (h, p) \beta_\infty' (h, p) > \left( \hat{R}_a \hat{R}_a' \right)^{1/2} \text{ERVA}^{-1} (\Sigma_{1:2}^*)^{1/2} \left( \beta_\infty \beta_\infty' - \beta_a \beta_a' \right) \left[ \hat{R}_a \hat{R}_a' \right]^{-1/2} \hat{R}_a.
\]

**Parts (c)(d)(e).** These three parts are similar to Theorem 10(a),(b), and (c) respectively. We only give the proof for part (e) in some details. It is easy to show that under the local alterative \( H_1 : R \theta_0 = r + \delta_0 / \sqrt{T} \), we have \( \mathbb{W}_{aT} \overset{d}{\rightarrow} \mathbb{W}_{1\infty} (|| \nu_a^{-1/2} \delta_0 ||^2) \) where

\[
\nu_a = \text{ERVA}^{-1} (\Sigma_{1:2}^*)^{1/2} (I_d, -\beta_a) \Omega \left( \begin{bmatrix} I_d \\ -\beta_a' \end{bmatrix} \right) \left[ \text{ERVA}^{-1} (\Sigma_{1:2}^*)^{1/2} \right]'.
\]

(36)
Similarly, we have

\[ \mathcal{W}_{2T} \xrightarrow{d} \mathcal{W}_{2\infty}(\| \mathcal{V}_2^{-1/2} \delta_0 \|^2), \]

where

\[ \mathcal{V}_2 = R \left( \hat{G}^\prime \hat{\Omega}^{-1} \hat{G} \right)^{-1} R' \]

\[ = RVA^{-1} (\Sigma_{1,2}^{*})^{1/2} (I_d, -\beta_0) \Omega \left( \begin{array}{c}
I_d \\
-\beta_0' 
\end{array} \right) \left[ RVA^{-1} (\Sigma_{1,2}^{*})^{1/2} \right]', \]

which is the asymptotic variance of \( R\sqrt{T} (\hat{\theta}_{2T} - \theta_0) \) with \( \hat{\theta}_{2T} \) being the infeasible optimal two-step GMM estimator.

The difference in the two matrices \( \mathcal{V}_2 \) and \( \mathcal{V}_2 \) is

\[ \mathcal{V}_2 - \mathcal{V}_2 = RVA^{-1} (\Sigma_{1,2}^{*})^{1/2} (\beta_a - \beta_0) \Omega_{22} (\beta_a - \beta_0)' \left[ RVA^{-1} (\Sigma_{1,2}^{*})^{1/2} \right]', \]

Now

\[ \| \mathcal{V}_2^{-1/2} \delta_0 \|^2 - \tau \| \mathcal{V}_2^{-1/2} \delta_0 \|^2 \]

\[ = \delta_0' \left[ \mathcal{V}_2^{-1} - \tau \mathcal{V}_2^{-1} \right] \delta_0 = \delta_0' \left[ \mathcal{V}_2^{-1/2} \mathcal{V}_2^{-1/2} - \tau I_p \right] \mathcal{V}_2^{-1/2} \delta_0 \]

\[ = \delta_0' \left[ \mathcal{V}_2^{-1/2} \right]' \left\{ \mathcal{V}_2^{-1/2} \mathcal{V}_2^{-1/2} - \tau I_p \right\} \mathcal{V}_2^{-1/2} \delta_0 \]

\[ = \delta_0' \left[ \mathcal{V}_2^{-1/2} \right]' \left\{ \mathcal{V}_2^{-1/2} \mathcal{V}_2^{-1/2} \mathcal{V}_2^{-1/2} - \tau I_p \right\} \mathcal{V}_2^{-1/2} \delta_0 \]

where

\[ \mathcal{V}_2^{-1/2} \mathcal{V}_2 [\mathcal{V}_2^{-1/2}]' = I_p - \mathcal{V}_2^{-1/2} (\mathcal{V}_2 - \mathcal{V}_2) [\mathcal{V}_2^{-1/2}]', \]

and

\[ \mathcal{V}_2^{-1/2} (\mathcal{V}_2 - \mathcal{V}_2) [\mathcal{V}_2^{-1/2}]' = RVA^{-1} (\Sigma_{1,2}^{*})^{1/2} (\beta_a - \beta_0) \Omega_{22} (\beta_a - \beta_0)' \left[ RVA^{-1} (\Sigma_{1,2}^{*})^{1/2} \right]' [\mathcal{V}_2^{-1/2}]'. \]

Define

\[ \rho_{a,R} = \mathcal{V}_2^{-1} RVA^{-1} (\Sigma_{1,2}^{*})^{1/2} (\beta_a - \beta_0) \Omega_{22}^{-1/2}, \]

which is the long run correlation matrix between

\[ RVA^{-1} (\Sigma_{1,2}^{*})^{1/2} [f_1 (v_t, \theta_0) - \beta_a f_2 (v_t, \theta_0)] \text{ and } f_2 (v_t, \theta_0). \]

Then

\[ \mathcal{V}_2^{-1/2} (\mathcal{V}_2 - \mathcal{V}_2) [\mathcal{V}_2^{-1/2}]' = \rho_{a,R} \rho_{a,R}' \text{ and } \mathcal{V}_2^{-1/2} \mathcal{V}_2 [\mathcal{V}_2^{-1/2}]' = I_p - \rho_{a,R} \rho_{a,R}' \]

and

\[ \| \mathcal{V}_2^{-1/2} \delta_0 \|^2 - \tau \| \mathcal{V}_2^{-1/2} \delta_0 \|^2 \]

\[ = \delta_0' \left[ \mathcal{V}_2^{-1/2} \right]' \left( I_p - \rho_{a,R} \rho_{a,R}' \right)^{-1/2} \left( \rho_{a,R} \rho_{a,R}' - \frac{1}{\tau} I_p \right)^{-1/2} \left( I_p - \rho_{a,R} \rho_{a,R}' \right)^{-1/2} \delta_0 \tau. \]

So if \( \rho_{a,R} \rho_{a,R}' > (\tau - 1) / \tau \cdot I_p = f (\lambda_a; h, p, q, \alpha) \cdot I_p \), then \( \| \mathcal{V}_2^{-1/2} \delta_0 \|^2 > \tau \| \mathcal{V}_2^{-1/2} \delta_0 \|^2 \) and the test based on \( \mathcal{W}_{2T} \) is asymptotically more powerful than that based on \( \mathcal{W}_{aT} \).
References


