Title
Asymptotic F Tests under Possibly Weak Identification

Authors
Martínez-Iriarte, Julián
Sun, Yixiao
Wang, Xuexin

Publication Date
2019-03-12

Asymptotic F Tests under Possibly Weak Identification*

Julián Martínez-Iriarte and Yixiao Sun
Department of Economics
UC San Diego, USA

Xuexin Wang
School of Economics and WISE
Xiamen University, China

This version: March 9, 2019

Abstract

This paper develops asymptotic F tests robust to weak identification and temporal dependence. The test statistics are modified versions of the S statistic of Stock and Wright (2000) and the K statistic of Kleibergen (2005), both of which are based on the continuous updating generalized method of moments. In the former case, the modification involves only a multiplicative degree-of-freedom adjustment. In the latter case, the modification involves an additional multiplicative adjustment that uses a J statistic for testing overidentification. By adopting fixed-smoothing asymptotics, we show that both the modified S statistic and the modified K statistic are asymptotically F-distributed. The asymptotic F theory accounts for the estimation errors in the underlying heteroskedasticity and autocorrelation robust variance estimators, which the asymptotic chi-squared theory ignores. Monte Carlo simulations show that the F approximations are much more accurate than the corresponding chi-squared approximations in finite samples.

Keywords: Heteroskedasticity and autocorrelation robust variance, continuous updating GMM, F distribution, fixed-smoothing asymptotics, weak identification.

JEL Classification Numbers: C12, C14, C32, C36

*Email: jum024@ucsd.edu; yisun@ucsd.edu; xuexinwang@xmu.edu.cn. Sun gratefully acknowledges the partial research support from the NSF under Grant No. SES-1530592. Address correspondence to Yixiao Sun, Department of Economics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA.

1 Introduction

For models defined by moment restrictions, we often allow the moment process to have a nonparametric autocorrelation structure, especially in macroeconomic and financial applications. The generalized method of moments (GMM), initially proposed by Hansen (1982), is a workhorse for these models. Inference in the GMM framework typically employs a nonparametric estimator of the long-run variance (LRV) of the moment process. In the regression setting, which is a special case of the GMM, this nonparametric LRV estimator is more commonly referred to as the heteroskedasticity and autocorrelation robust (HAR) variance estimator. Conventional asymptotic theory dictates that the HAR variance estimator is consistent. However, the accuracy of the resulting asymptotic approximation is often unsatisfactory. In particular, HAR tests tend to over-reject, sometimes substantially, under the null hypothesis in finite samples. See, for example, Andrews (1991), Hansen, Heaton, and Yaron (1996), and the other papers in the July 1996 special issue of the *Journal of Business & Economic Statistics* for evidence.

The recent literature has introduced alternative asymptotics to combat this problem. Unlike the conventional asymptotics where the amount of nonparametric smoothing increases with the sample size, the alternative asymptotics holds the amount of nonparametric smoothing fixed. Hence, the alternative asymptotics is also called the fixed-smoothing asymptotics. There are ample numerical evidence and theoretical results on the higher accuracy of fixed-smoothing asymptotic approximations relative to conventional asymptotic approximations. See, for example, Jansson (2004) and Sun, Phillips, and Jin (2008) for location models, and Sun (2014a, 2014b) for the GMM framework.

Another source of the poor finite sample properties of HAR tests is related to the weak identification of the model parameters, as Stock and Wright (2000) point out. Nonetheless, it is possible to perform hypothesis testing and construct confidence intervals without assuming that the parameters are identified. Stock and Wright (2000) propose using the S statistic constructed directly from the GMM objective function, while Kleibergen (2005) proposes using the K statistic, a Lagrangian-multiplier-type statistic constructed from the first-order derivative of the continuous updating GMM (henceforth, CU-GMM) objective function.

Under some regularity conditions, both the K statistic and the S statistic are shown to follow chi-squared distributions asymptotically under the null hypothesis, regardless of the strength of the identification. However, these chi-squared approximations rely on the assumption that the underlying LRV estimators are consistent in the presence of the heteroskedasticity and autocorrelation structure. In other words, the chi-squared approximations completely ignore the estimation errors in the LRV estimators and effectively assume that these LRVs are known. For this reason, the approximating chi-squared distributions can be far from the finite sample distributions.

The problem is especially severe for the K statistic, as all the underlying LRV estimators employ the same smoothing parameter. To illustrate the point, consider the kernel approach to LRV estimation with compactly supported kernels. In this case, the kernel estimators of the LRVs of the moment process and Jacobian process take a weighted sum of the respective sample autocovariances up to the same order (see Assumption 2 in Kleibergen (2005)). For these LRV estimators to be consistent, the maximum order of autocovariances (i.e., the truncated lag) must be sufficiently large such that none of the LRV estimators suffers from a large bias. However, when the maximum order of autocovariances is large, all the LRV estimators have high variation. Hence, to improve the accuracy of approximations, the potentially high estimation uncertainty should not be ignored.

This paper develops fixed-smoothing asymptotics for the K and S statistics to account for the estimation uncertainty in the underlying LRV estimators. It is built upon Sun (2014b) and Hwang and Sun (2017), which establish fixed-smoothing asymptotics for the usual statistics including the Lagrangian-multiplier (LM), Wald, and likelihood-ratio types of statistics in a strongly identified GMM framework. The main departure is that we consider the CU-GMM framework with possibly weak identification. As in Hwang and Sun (2017), we employ the orthonormal series approach to LRV estimation. The simplest and most familiar example of this estimator is the average periodogram estimator, which involves taking a simple average of the first few periodograms. More generally, this approach involves first projecting the time series onto a sequence of orthonormal basis functions such as the sine and cosine functions and then taking the simple average of the squared projection coefficients as the LRV estimator. The number of basis functions, which characterizes the amount of smoothing, is the smoothing parameter underlying this orthonormal series LRV estimator.

By holding the number of basis functions fixed as the sample size increases, we establish a novel result that a modified K statistic and a modified J statistic follow independent F distributions in large samples. There are substantial challenges to establishing this intriguing result. First, Kleibergen's (2005) asymptotic chi-squared theory relies on the key result that the Jacobian for the CU-GMM problem is asymptotically independent of the empirical moment vector. Asymptotic independence may not hold when the estimation uncertainties in the LRV estimators are accounted for. Second, the weighting matrix in the CU-GMM objective function converges to a random matrix under fixed-smoothing asymptotics. This is in sharp contrast to Kleibergen's (2005) setting in which the weighting matrix converges to a deterministic matrix. As a result, the weighting matrix is, by definition, asymptotically independent of all the involved random variables including the Jacobian for the CU-GMM under conventional asymptotics. Such independence does not hold under fixed-smoothing asymptotics. The absence of these two independence results thus makes it challenging to show that the K statistic and J statistic are asymptotically pivotal.

To overcome these challenges, we employ singular value decompositions (SVDs) and conditioning arguments. We also use the rotational invariance of the standard normal distribution and Wishart distribution repeatedly. Along the way, we introduce some simple modifications to the original K, J, and S statistics. For the J and S statistics, such modifications amount to a multiplicative degree-of-freedom correction that is free of nuisance parameters. For the K statistic, the modification involves both a degree-of-freedom correction and a correction that employs the original J statistic. We show that the seemingly complicated asymptotic distributions are not only pivotal but also equal to the standard F distributions.

The asymptotic F tests based on the modified K and S statistics are just as easy to use as the chi-squared tests proposed by Kleibergen (2005) and Stock and Wright (2000). Like chi-squared critical values, F critical values are readily available in standard programming environments and software packages. Moreover, the asymptotic independence of the modified K statistic and modified J statistic allows us to design a new and simple test to overcome the well-known power deficiency of the usual K test. Monte Carlo simulations show that the F distributions provide more accurate approximations of the distributions of the modified K, J, and S statistics than the chi-squared distributions.

We use the following notation throughout the paper: $\mathbb{E}(a)$ is the expected value of the random variable a and $\text{vec}(A)$ stands for the column vectorization of the $m \times d$ dimensional matrix A : $\text{vec}(A) = (A_1', \ldots, A_d')'$ when $A = (A_1, \ldots, A_d)$. $\mathbb{P}_A = A\,(A'A)^{-1}\,A'$ and $\mathbb{M}_A = I_m - \mathbb{P}_A$. "$\to^p$" indicates convergence in probability and "$\Rightarrow$" indicates convergence in distribution. We use "**0**" or "O" to represent a matrix of zeros.

The rest of the paper is organized as follows. Section 2 presents the K and S statistics and the modified K, J, and S statistics, and reviews the asymptotic chi-squared theory for the K and S statistics. Section 3 develops the fixed-smoothing asymptotics for the K, J, and S statistics. Section 4 is devoted to establishing the asymptotic F theory for the modified K and J statistics. While Sections 2–4 focus on testing the whole vector of the model parameters, Section 5 considers testing a subvector. Section 6 develops the asymptotic F theory for the modified S statistic. Section 7 uses the asymptotic properties of the K and J statistics to design a more powerful test. Section 8 reports simulation evidence and presents an empirical application. The last section concludes. All proofs are given in Appendix A.

2 Basic Setting and Asymptotic Chi-squared Theory

2.1 The K, J, and S statistics and their modifications

Let $Y_t \in \mathbb{R}^{d_Y}$ be a vector of observations at time t, and $f(Y_t, \cdot)$ an $m \times 1$ vector of continuously differentiable functions such that the following moment restriction

$$\mathbb{E}f(Y_t, \theta_0) = 0, \text{ for } t = 1, \ldots, T, \tag{1}$$

holds for $\theta_0 \in \Theta \in \mathbb{R}^d$. The degree of potential overidentification is denoted by $q := m - d$. We allow the process $f(Y_t, \theta_0)$ to exhibit heteroskedasticity and autocorrelation of unknown forms. We are interested in testing

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta \neq \theta_0.$$

Since the null hypothesis pins down the full parameter vector, testing without assuming identification is feasible[1]. For example, we can use the K test of Kleibergen (2005) and the S test of Stock and Wright (2000), both of which are based on the CU-GMM objective function.

To introduce the CU-GMM objective function, we let $V_{ff}(\theta)$ be the long-run variance of the moment process, that is,

$$V_{ff}(\theta) = \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t, \theta)\right).$$

A nonparametric estimator of the LRV takes the quadratic form

$$\hat{V}_{ff}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\omega_h\left(\frac{t}{T}, \frac{s}{T}\right)\left[f(Y_t, \theta) - \bar{f}(Y, \theta)\right]\left[f(Y_s, \theta) - \bar{f}(Y, \theta)\right]',$$

where $\bar{f}(Y, \theta) = T^{-1}\sum_{t=1}^{T} f(Y_t, \theta)$, $\omega_h(\cdot, \cdot)$ is a weighting function and h is the smoothing parameter indicating the amount of nonparametric smoothing. For example, we can take $\omega_h(t/T, s/T) = k((t-s)/(hT))$ for a kernel function $k(\cdot)$, leading to the usual kernel LRV estimator. With the LRV estimator $\hat{V}_{ff}(\theta)$, the CU-GMM objective function is[2]

$$Q_T(\theta) = \frac{1}{2}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t, \theta)\right]'\hat{V}_{ff}^{-1}(\theta)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t, \theta)\right].$$

[1] We consider subvector testing in Section 5.

[2] The multiplicative factor of 1/2 is introduced for notational convenience in our later development.

The K statistic is based on the first-order derivative of $Q_T(\theta)$. Define

$$\begin{aligned} g_j(Y_t,\theta) &= \frac{\partial f(Y_t,\theta)}{\partial \theta_j} \in \mathbb{R}^{m\times 1}, j=1,\dots,d, \\ g(Y_t,\theta) &= \frac{\partial f(Y_t,\theta)}{\partial \theta'} = (g_1(Y_t,\theta),\dots,g_d(Y_t,\theta)) \in \mathbb{R}^{m\times d}, \\ \bar{g}(Y_t,\theta) &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial f(Y_t,\theta)}{\partial \theta'} \in \mathbb{R}^{m\times d}. \end{aligned}$$

Taking the first-order and second-order derivatives of $\hat{V}_{ff}(\theta)$ with respect to θ_j, we obtain the following LRV estimators[3]:

$$\begin{aligned} \hat{V}_{g_j f}(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\omega_h\left(\frac{t}{T},\frac{s}{T}\right)\left[g_j(Y_t,\theta)-\bar{g}(Y,\theta)\right]\left[f(Y_s,\theta)-\bar{f}(Y,\theta)\right]', \\ \hat{V}_{g_j g_j}(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\omega_h\left(\frac{t}{T},\frac{s}{T}\right)\left[g_j(Y_t,\theta)-\bar{g}_j(Y,\theta)\right]\left[g_j(Y_s,\theta)-\bar{g}_j(Y,\theta)\right]'. \end{aligned}$$

Then,

$$\frac{\partial Q_T(\theta)}{\partial \theta} = D_T(\theta)'\hat{V}_{ff}^{-1}(\theta)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t,\theta)\right],$$

where

$$\begin{aligned} D_T(\theta) &= \left[D_{T,1}(\theta),\dots,D_{T,d}(\theta)\right] \in \mathbb{R}^{m\times d} \text{ with} \\ D_{T,j}(\theta) &= \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_j(Y_t,\theta)\right] - \hat{V}_{g_j f}(\theta)\hat{V}_{ff}^{-1}(\theta)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t,\theta)\right] \in \mathbb{R}^{m\times 1}. \end{aligned} \tag{2}$$

The K statistic for testing $H_0:\theta=\theta_0$ against $H_1:\theta\neq\theta_0$ is

$$\mathcal{K}_T(\theta_0) = \left(\frac{\partial Q_T(\theta_0)}{\partial\theta}\right)'\left[D_T(\theta_0)'\hat{V}_{ff}^{-1}(\theta_0)D_T(\theta_0)\right]^{-1}\left(\frac{\partial Q_T(\theta_0)}{\partial\theta}\right),$$

where, for a function $\phi(\theta)$, $\partial\phi(\theta_0)/\partial\theta$ is defined to be

$$\frac{\partial\phi(\theta_0)}{\partial\theta} = \left.\frac{\partial\phi(\theta)}{\partial\theta}\right|_{\theta=\theta_0} \in \mathbb{R}^{d\times 1}.$$

The S statistic is proportional to the CU-GMM objective function and is given by

$$\mathcal{S}_T(\theta_0) = 2Q_T(\theta_0).$$

To develop our fixed-smoothing asymptotics, we employ the orthonormal series LRV estimator so that

$$\omega_h\left(\frac{t}{T},\frac{s}{T}\right) = \frac{1}{G}\sum_{\ell=1}^{G}\Phi_\ell\left(\frac{t}{T}\right)\Phi_\ell\left(\frac{s}{T}\right),$$

[3]Strictly speaking, $\hat{V}_{g_j f}(\theta)$ is a long-run covariance estimator, but we refer to it as LRV for convenience.

where G is the smoothing parameter for this estimator and $\Phi_\ell(\cdot)$ is a set of basis functions on $L^2[0,1]$. With this choice of the weighting function $\omega_h(\cdot,\cdot)$, the LRV estimator $\hat{V}_{ff}(\theta)$ becomes[4]

$$\hat{V}_{ff}(\theta) = \frac{1}{G}\sum_{\ell=1}^{G}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[f(Y_t,\theta)-\bar{f}(Y,\theta)\right]\right\}\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[f(Y_t,\theta)-\bar{f}(Y,\theta)\right]\right\}'.$$

In the econometrics literature, the orthonormal series LRV estimator has recently been used by, for example, Phillips (2005), Müller (2007), Sun (2011, 2013, 2014a,b,c), Liu and Sun (2019), and Lazarus, Lewis, Stock, and Watson (2016, 2018).

Our asymptotic F theory will be developed for modified versions of the usual statistics. The modification involves the following J statistic:

$$\begin{aligned}\mathcal{J}_T(\theta_0) &= \mathcal{S}_T(\theta_0)-\mathcal{K}_T(\theta_0)\\ &= \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(Y_t,\theta_0)\right]'\hat{V}_{ff}^{-1}(\theta_0)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(Y_t,\theta_0)\right]\\ &\quad-\left(\frac{\partial Q_T(\theta_0)}{\partial\theta}\right)'\left[D_T(\theta_0)'\hat{V}_{ff}^{-1}(\theta_0)D_T(\theta_0)\right]^{-1}\left(\frac{\partial Q_T(\theta_0)}{\partial\theta}\right). \end{aligned} \tag{3}$$

Define

$$\tilde{f}(Y_t,\theta_0) = \left\{I_m - D_T(\theta_0)\left[D_T(\theta_0)'\hat{V}_{ff}^{-1}(\theta_0)D_T(\theta_0)\right]^{-1}D_T(\theta_0)'\hat{V}_{ff}^{-1}(\theta_0)\right\}f(Y_t,\theta_0).$$

Then, some algebra shows that

$$\mathcal{J}_T(\theta_0) = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{f}(Y_t,\theta_0)\right]'\hat{V}_{ff}^{-1}(\theta_0)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{f}(Y_t,\theta_0)\right].$$

Note that $\mathcal{J}_T(\theta_0)$ is not the same as the usual J statistic, which is given by

$$\tilde{\mathcal{J}}_T(\hat{\theta}) = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f\left(Y_t,\hat{\theta}\right)\right]'\hat{V}_{ff}^{-1}(\hat{\theta})\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f\left(Y_t,\hat{\theta}\right)\right],$$

for an estimator $\hat{\theta}$ of θ_0.

The modified K, J, and S statistics are defined as

$$\begin{aligned}\mathcal{K}_T^* &:= \mathcal{K}_T^*(\theta_0) = \frac{G-m+1}{Gd}\frac{\mathcal{K}_T(\theta_0)}{1+\mathcal{J}_T(\theta_0)/G},\\ \mathcal{J}_T^* &:= \mathcal{J}_T^*(\theta_0) = \frac{G-q+1}{Gq}\mathcal{J}_T(\theta_0),\\ \mathcal{S}_T^* &:= \mathcal{S}_T^*(\theta_0) := \frac{G-m+1}{Gm}\mathcal{S}_T(\theta_0).\end{aligned}$$

We show that all three statistics are asymptotically F-distributed under the fixed-smoothing asymptotics.

[4] The literature typically uses K in place of G. To avoid the possible confusion with the K statistic, we use G to denote the number of basis functions in this paper.

2.2 Conventional asymptotic chi-squared theory

Before deriving the fixed-smoothing asymptotics, it is useful to review the conventional asymptotic chi-squared theory for the K and S statistics. To find the conventional asymptotic distributions of the original K statistic $\mathcal{K}_T(\theta_0)$ and the S statistic $\mathcal{S}_T(\theta_0)$, we maintain the following two assumptions.

Assumption 2.1 *The following CLT holds:*

$$\left(\begin{array}{c} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [f(Y_t, \theta_0) - \mathbb{E} f(Y_t, \theta_0)] \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \operatorname{vec}[g(Y_t, \theta_0) - \mathbb{E} g(Y_t, \theta_0)] \end{array} \right) \Rightarrow \left(\begin{array}{c} \psi_f \\ \psi_g \end{array} \right),$$

where $\psi_f \in \mathbb{R}^{m\times 1}$, $\psi_g \in \mathbb{R}^{md\times 1}$, *and*

$$\left(\begin{array}{c} \psi_f \\ \psi_g \end{array} \right) \sim N[0, V(\theta_0)]$$

for

$$V(\theta_0) = \left(\begin{array}{cc} V_{ff}(\theta_0) & V_{fg}(\theta_0) \\ V_{gf}(\theta_0) & V_{gg}(\theta_0) \end{array} \right).$$

Assumption 2.2 $\hat{V}_{ff}(\theta_0) \to^p V_{ff}(\theta_0)$, $V_{ff}(\theta_0)$ *is positive definite, and for* $j = 1, \ldots, d$

$$\hat{V}_{g_j f}(\theta_0) \to^p V_{g_j f}(\theta_0).$$

Assumption 2.1 is the same as Assumption 1 in Kleibergen (2005). Assumption 2.2 is the working assumption maintained throughout Kleibergen (2005).

We introduce a notational convention. When a (random) variable depends on the true parameter value, we often suppress this dependence. For example, we write

$$V(\theta_0) := V = \left(\begin{array}{cc} V_{ff} & V_{fg} \\ V_{gf} & V_{gg} \end{array} \right) \text{ and } D_T(\theta_0) := D_T.$$

For easy reference, we call $D_T(\theta_0)$ the CU Jacobian.

Now, under Assumptions 2.1 and 2.2, we have

$$\begin{aligned} & D_{T,j}(\theta_0) - \sqrt{T}\mathbb{E} g_j(Y_t, \theta_0) \\ & = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [g_j(Y_t, \theta_0) - \mathbb{E} g_j(Y_t, \theta_0)] - \hat{V}_{g_j f}(\theta_0) \hat{V}_{ff}^{-1}(\theta_0) \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(Y_t, \theta_0) \right] \\ & \Rightarrow \psi_{g_j} - V_{g_j f} V_{ff}^{-1} \psi_f := \psi_{g_j \cdot f}. \end{aligned}$$

Note that $\psi_{g_j \cdot f}$ is independent of ψ_f because they are normally distributed with zero covariance:

$$\operatorname{cov}(\psi_{g_j \cdot f}, \psi_f) = \operatorname{cov}(\psi_{g_j} - V_{g_j f} V_{ff}^{-1} \psi_f, \psi_f) = V_{g_j f} - V_{g_j f} V_{ff}^{-1} V_{ff} = 0.$$

This means that $D_T(\theta_0)$ is asymptotically independent of $\sum_{t=1}^{T} f(Y_t, \theta_0)/\sqrt{T}$. It then follows that

$$\mathcal{K}_T(\theta_0) \Rightarrow \chi_d^2 \text{ under the null } H_0.$$

For the S statistic, Assumptions 2.1 and 2.2 imply that[5]

$$\mathcal{S}_T(\theta_0) \Rightarrow \chi^2_m \text{ under the null } H_0.$$

A component of $\partial \hat{V}_{ff}(\theta_0)/\partial\theta_j$ is $\hat{V}_{g_j f}(\theta_0)$. Assumption 2.2 requires not only that $\hat{V}_{ff}(\theta_0)$ is a consistent estimator of $V_{ff}(\theta_0)$ but also that each derivative $\partial \hat{V}_{ff}(\theta_0)/\partial\theta_j$ is a consistent estimator of the corresponding derivative $\partial V_{ff}(\theta_0)/\partial\theta_j$. Using the consistency results amounts to approximating the distributions of the two sets of LRV estimators, namely $\hat{V}_{ff}(\theta_0)$ and $\{\hat{V}_{g_j f}(\theta_0)\}_{j=1}^d$, by their respective degenerate distributions. This may not be much of a problem if such an approximation is applied to only one LRV estimator and the underlying process has very weak dependence. However, here we apply the degenerate approximations to potentially many LRV estimators, each of which is governed by the same smoothing parameter used in the construction of $\hat{V}_{ff}(\theta_0)$. The single smoothing parameter must be tailored toward both the moment processes and the Jacobian processes, some of which may be strongly autocorrelated. For this reason, the smoothing parameter used in $\hat{V}_{ff}(\theta_0)$ must account for the autocovariances of high orders, even though this is not necessary for estimating $V_{ff}(\theta_0)$ itself. The estimation uncertainties in $\hat{V}_{ff}(\theta_0)$ and hence in $\{\hat{V}_{g_j f}(\theta_0)\}_{j=1}^d$ can therefore be very high. Ignoring them may lead to a poor approximation in finite samples.

3 Fixed-smoothing Asymptotics for the K, J, and S Statistics

The analysis in the previous section motivates us to study the fixed-smoothing asymptotics under which G is held fixed as $T \to \infty$. Under this type of asymptotics, the estimation uncertainties in $\hat{V}_{ff}(\theta_0)$ and $\{\hat{V}_{g_j f}(\theta_0)\}_{j=1}^d$ are all retained in the limit, and as a result, the fixed-smoothing asymptotic approximations are more accurate than the chi-squared approximations. Since G is fixed, this type of asymptotics may be referred to as the fixed-G asymptotics. We consider the fixed-G asymptotics hereafter.

We first present an assumption on the model identification.

Assumption 3.1 *For $\kappa \in [0, 1/2]$,*

$$T^\kappa \mathbb{E}g(Y_t, \theta_0) \to \Pi = (\Pi_1, \ldots, \Pi_d) \in \mathbb{R}^{m\times d}$$

for a constant matrix Π, and $\Pi \neq 0$ for $\kappa \in [0, 1/2)$.

When $\kappa = 0$ and Π has a full column rank, θ_0 can be consistently estimated at the usual $\sqrt{T}$ rate. When $\kappa \in (0, 1/2)$ and Π has a full column rank, θ_0 can still be consistently estimated, albeit at a rate slower than $\sqrt{T}$. In later sections, we refer to the case with $\kappa \in [0, 1/2)$ as the case with possibly weak identification. When $\kappa = 1/2$ and $\Pi = 0$, θ_0 cannot be consistently estimated, and the model contains no information about θ_0 to the first order. We refer to this case as the case with complete identification failure. The case with $\kappa = 1/2$ and $\Pi \neq 0$ is the intermediate case in which θ_0 cannot be consistently estimated, but the model contains some information about θ_0.

The next assumption generalizes Assumption 2.1.

[5] Weaker assumptions can be used to obtain this result. For example, it suffices that $\hat{V}_{ff}(\theta_0) \to^p V_{ff}(\theta_0)$ for a positive definite matrix $V_{ff}(\theta_0)$ and that $T^{-1/2}\sum_{t=1}^T f(Y_t, \theta_0) \Rightarrow N(0, V_{ff}(\theta_0))$.

Assumption 3.2 *The following functional CLT (FCLT) holds:*

$$
\begin{aligned}
&\begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\left[f\left(Y_t,\theta_0\right)-\mathbb{E}f\left(Y_t,\theta_0\right)\right] \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}\operatorname{vec}\left[g\left(Y_t,\theta_0\right)-\mathbb{E}g\left(Y_t,\theta_0\right)\right] \end{pmatrix} \\
&\Rightarrow \begin{pmatrix} B_f\left(r\right) \\ B_g\left(r\right) \end{pmatrix} = \begin{pmatrix} V_{ff} & V_{fg} \\ V_{gf} & V_{gg} \end{pmatrix}^{1/2} \begin{pmatrix} W_f\left(r\right) \\ W_g\left(r\right) \end{pmatrix} \\
&= \begin{pmatrix} V_{ff}^{1/2} & 0 \\ V_{gf}V_{ff}^{-1/2} & V_{g\cdot f}^{1/2} \end{pmatrix} \begin{pmatrix} W_f\left(r\right) \\ W_g\left(r\right) \end{pmatrix},
\end{aligned}
$$

where

$$B_g\left(r\right)=\left(B_{g_1}'\left(r\right),\ldots,B_{g_d}'\left(r\right)\right)'\in\mathbb{R}^{dm\times 1}$$

is a vector Brownian motion process and

$$W_g\left(r\right)=\left(W_{g_1}'\left(r\right),\ldots,W_{g_d}'\left(r\right)\right)'\in\mathbb{R}^{dm\times 1}$$

is a standard vector Brownian motion process.

Assumption 3.2 is stronger than necessary. It can be replaced by Assumption 2.1 and a multivariate CLT that ensures (5) and (6) below. See Sun (2014a,b) for details. Sufficient moment and mixing conditions for the CLT or FCLT can be found, for example, in Theorem 7.18 of White (2001).

For each $B_{g_j}\left(r\right)\in\mathbb{R}^{m\times 1}$, $j=1,\ldots,d$, we write

$$B_{g_j}\left(r\right)=V_{g_jf}V_{ff}^{-1}B_f\left(r\right)+B_{g_j\cdot f}\left(r\right), \tag{4}$$

where $B_f\left(r\right)$ and $B_{g_j\cdot f}\left(r\right)$ are independent Brownian motions. Furthermore, let $V_{g_j\cdot f}=V_{g_jg_j}-V_{g_jf}V_{ff}^{-1}V_{fg_j}$, and define

$$W_{g_j}\left(r\right)=V_{g_jg_j}^{-1/2}B_{g_j}\left(r\right),\ W_f\left(r\right)=V_{ff}^{-1/2}B_f\left(r\right),\text{ and }W_{g_j\cdot f}\left(r\right)=V_{g_j\cdot f}^{-1/2}B_{g_j\cdot f}\left(r\right).$$

Then, $W_{g_j}\left(r\right)$, $W_f\left(r\right)$, and $W_{g_j\cdot f}\left(r\right)$ are standard Brownian motions. Plugging them into (4) leads to

$$W_{g_j}\left(r\right)=V_{g_jg_j}^{-1/2}V_{g_jf}V_{ff}^{-1/2}\cdot W_f\left(r\right)+V_{g_jg_j}^{-1/2}V_{g_j\cdot f}^{1/2}\cdot W_{g_j\cdot f}\left(r\right).$$

For the basis functions used in the LRV estimation, we maintain the following assumption, which is the same as Assumption 1(b) in Sun (2014b).

Assumption 3.3 *For $\ell=1,2,\ldots,G$, the basis functions $\Phi_\ell\left(\cdot\right)$ are piecewise monotonic, continuously differentiable, and orthonormal in $L^2[0,1]$ and satisfy $\int_0^1\Phi_\ell\left(x\right)dx=0$.*

Under Assumptions 3.2 and 3.3, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[f\left(Y_t,\theta_0\right)-\bar{f}(Y,\theta_0)\right]\Rightarrow\int_0^1\Phi_\ell\left(r\right)dB_f\left(r\right):=\xi_{f,\ell} \tag{5}$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[g_j\left(Y_t,\theta_0\right)-\bar{g}_j(Y,\theta_0)\right]\Rightarrow\int_0^1\Phi_\ell\left(r\right)dB_{g_j}\left(r\right):=\xi_{g_j,\ell} \tag{6}$$

jointly over $j=1,2,\ldots,d$ and $\ell=1,\ldots,G$. Let $\xi_{g_j\cdot f,\ell}=\int_0^1\Phi_\ell(r)\,dB_{g_j\cdot f}(r)$. Then, we have

$$\xi_{g_j,\ell}=V_{g_jf}V_{ff}^{-1}\xi_{f,\ell}+\xi_{g_j\cdot f,\ell}.$$

Let $\eta_{f,\ell},\eta_{g_j,\ell}$, and $\eta_{g_j\cdot f,\ell}$ be the scale-free versions of $\xi_{f,\ell},\xi_{g_j,\ell}$, and $\xi_{g_j\cdot f,\ell}$. We have

$$\begin{aligned}
\eta_{f,\ell}&:=\int_0^1\Phi_\ell(r)\,dW_f(r)\sim iidN(0,I_m)\text{ over }\ell=1,\ldots,G;\\
\eta_{g_j,\ell}&:=\int_0^1\Phi_\ell(r)\,dW_{g_j}(r)\sim iidN(0,I_m)\text{ over }\ell=1,\ldots,G;\\
\eta_{g_j\cdot f,\ell}&:=\int_0^1\Phi_\ell(r)\,dW_{g_j\cdot f}(r)\sim iidN(0,I_m)\text{ over }\ell=1,\ldots,G.
\end{aligned}$$

The iid properties hold because $\{\Phi_\ell(\cdot),\ell=1,\ldots,G\}$ are orthonormal on $L_2[0,1]$ and $W_{g_j}(r)$, $W_f(r)$, and $W_{g_j\cdot f}(r)$ are standard Brownian motions. We collect some useful results in the following lemma.

Lemma 3.1 *Let Assumptions 3.2 and 3.3 hold.*

(a) $\left\{\eta_{f,\ell},\ell=1,\ldots,G\right\}$ *and* $\left\{\eta_{g_j\cdot f,\ell}:j=1,\ldots,d,\ell=1,\ldots,G\right\}$ *are independent.*

(b) $W_f(1)$ *is independent of* $\left\{\eta_{f,\ell},\ell=1,\ldots,G\right\}$ *and* $W_{g_j\cdot f}(1)$ *is independent of* $\{\eta_{g_j\cdot f,\ell},\ell=1,\ldots,G\}$.

(c) $\{W_{g_j\cdot f}(1),j=1,\ldots,d\}$ *are independent of* $\left\{\eta_{f,\ell},\ell=1,\ldots,G\right\}$ *and* $W_f(1)$ *is independent of* $\{\eta_{g_j\cdot f,\ell}:j=1,\ldots,d,\ell=1,\ldots,G\}$.

Lemma 3.1(a) holds because $W_f(\cdot)$ and $W_{g_j\cdot f}(\cdot)$ are independent. Lemmas 3.1(b) and 3.1(c) follow from the condition that $\int_0^1\Phi_\ell(r)\,dr=0$ and the independence between $W_f(\cdot)$ and $W_{g_j\cdot f}(\cdot)$.

The following lemma establishes the asymptotic behavior of the LRV estimators and CU Jacobian for different values of κ.

Lemma 3.2 *Let Assumptions 3.1, 3.2, and 3.3 hold. Then,*

$$\hat V_{ff}(\theta_0)\Rightarrow\frac{1}{G}\sum_{\ell=1}^G\xi_{f,\ell}\xi_{f,\ell}',\ \hat V_{g_jf}(\theta_0)\Rightarrow\frac{1}{G}\sum_{\ell=1}^G\xi_{g_j,\ell}\xi_{f,\ell}',$$

and

$$\frac{D_T(\theta_0)}{T^{1/2-\kappa}}\Rightarrow D_\infty,$$

where

$$D_\infty=(D_{\infty,1},\ldots,D_{\infty,d})\in\mathbb{R}^{m\times d}\text{ for }D_{\infty,j}=\Pi_j+1\left\{\kappa=\frac{1}{2}\right\}\tilde B_{g_j\cdot f}(1)$$

and

$$\tilde B_{g_j\cdot f}(1)=B_{g_j}(1)-\left[\frac{1}{G}\sum_{\ell=1}^G\xi_{g_j,\ell}\xi_{f,\ell}'\right]\left[\frac{1}{G}\sum_{\ell=1}^G\xi_{f,\ell}\xi_{f,\ell}'\right]^{-1}B_f(1). \qquad (7)$$

Note that $\tilde{B}_{g_j \cdot f}(1)$ is not the same as $B_{g_j \cdot f}(1)$, which is equal to $B_{g_j}(1) - V_{g_j f} V_{ff}^{-1} B_f(1)$. The difference lies in the implicit projection coefficient. While $B_{g_j \cdot f}(1)$ is based on the true projection coefficient $V_{g_j f} V_{ff}^{-1}$, $\tilde{B}_{g_j \cdot f}(1)$ is based on the limit of the estimated version $\hat{V}_{g_j f} \hat{V}_{ff}^{-1}$. Under the fixed-smoothing asymptotics, the limit of $\hat{V}_{g_j f} \hat{V}_{ff}^{-1}$ is not equal to $V_{g_j f} V_{ff}^{-1}$. The difference has profound implications on our asymptotic development for the case that $\kappa = 1/2$. While $B_{g_j \cdot f}(1)$ is independent of $B_f(1)$, $\tilde{B}_{g_j \cdot f}(1)$ is not. The Jacobian for the CUE is therefore not asymptotically independent of the moment process. This is in sharp contrast to Kleibergen (2005), where asymptotic independence holds even for the case with $\kappa = 1/2$. Section 4.2 provides additional details.

Using Lemma 3.2, we can establish the fixed-smoothing asymptotics of the K, J, and S statistics.

First, for the K statistic, we have

$$
\begin{aligned}
&\mathcal{K}_T(\theta_0) \\
&= \left\{ \frac{D_T'}{T^{1/2-\kappa}} \hat{V}_{ff}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(Y_t, \theta_0) \right\}' \left[\frac{D_T'}{T^{1/2-\kappa}} \hat{V}_{ff}^{-1} \frac{D_T}{T^{1/2-\kappa}} \right]^{-1} \frac{D_T'}{T^{1/2-\kappa}} \hat{V}_{ff}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(Y_t, \theta_0) \\
&\Rightarrow B_f(1)' \left[\frac{1}{G} \sum_{\ell=1}^{G} \xi_{f,\ell} \xi_{f,\ell}' \right]^{-1} D_\infty \left\{ D_\infty' \left[\frac{1}{G} \sum_{\ell=1}^{G} \xi_{f,\ell} \xi_{f,\ell}' \right]^{-1} D_\infty \right\}^{-1} D_\infty' \left[\frac{1}{G} \sum_{\ell=1}^{G} \xi_{f,\ell} \xi_{f,\ell}' \right]^{-1} B_f(1) \\
&:= \mathcal{K}_{\infty,\theta}.
\end{aligned}
$$

Representing $\mathcal{K}_{\infty,\theta}$ using scale-free random variables and processes, we have

$$
\begin{aligned}
\mathcal{K}_{\infty,\theta} &= \left[V_{ff}^{-1/2} B_f(1) \right]' \left[\frac{1}{G} \sum_{\ell=1}^{G} V_{ff}^{-1/2} \xi_{f,\ell} \left(V_{ff}^{-1/2} \xi_{f,\ell} \right)' \right]^{-1} \\
&\quad \times V_{ff}^{-1/2} D_\infty \left\{ \left(V_{ff}^{-1/2} D_\infty \right)' \left[\frac{1}{G} \sum_{\ell=1}^{G} V_{ff}^{-1/2} \xi_{f,\ell} \left(V_{ff}^{-1/2} \xi_{f,\ell} \right)' \right]^{-1} V_{ff}^{-1/2} D_\infty \right\}^{-1} \left(V_{ff}^{-1/2} D_\infty \right)' \\
&\quad \times \left[\frac{1}{G} \sum_{\ell=1}^{G} V_{ff}^{-1/2} \xi_{f,\ell} \left(V_{ff}^{-1/2} \xi_{f,\ell} \right)' \right]^{-1} V_{ff}^{-1/2} B_f(1) \\
&= W_f(1)' \tilde{C}_\infty^{-1} \tilde{D}_\infty \left\{ \tilde{D}_\infty' \tilde{C}_\infty^{-1} \tilde{D}_\infty \right\}^{-1} \tilde{D}_\infty' \tilde{C}_\infty^{-1} W_f(1) \\
&= \left\| \mathbb{P}_{\tilde{C}_\infty^{-1/2} \tilde{D}_\infty} \tilde{C}_\infty^{-1/2} W_f(1) \right\|^2,
\end{aligned}
\tag{8}
$$

where $\tilde{D}_\infty = V_{ff}^{-1/2} D_\infty$ and

$$
\tilde{C}_\infty = \frac{1}{G} \sum_{\ell=1}^{G} \eta_{f,\ell} \eta_{f,\ell}'.
$$

Second, for the J statistic, we note that

$$
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{f}(Y_t, \theta_0) \Rightarrow V_{ff}^{1/2} \left[W_f(1) - \tilde{D}_\infty \left\{ \tilde{D}_\infty' \tilde{C}_\infty^{-1} \tilde{D}_\infty \right\}^{-1} \left\{ \tilde{D}_\infty' \tilde{C}_\infty^{-1} W_f(1) \right\} \right],
$$

and so

$$\begin{aligned}
\mathcal{J}_T(\theta_0) &\Rightarrow \left[W_f(1) - \tilde{D}_\infty \left\{\tilde{D}_\infty' \tilde{C}_\infty^{-1} \tilde{D}_\infty\right\}^{-1} \left\{\tilde{D}_\infty' \tilde{C}_\infty^{-1} W_f(1)\right\}\right]' \tilde{C}_\infty^{-1} \\
&\quad \times \left[W_f(1) - \tilde{D}_\infty \left\{\tilde{D}_\infty' \tilde{C}_\infty^{-1} \tilde{D}_\infty\right\}^{-1} \left\{\tilde{D}_\infty' \tilde{C}_\infty^{-1} W_f(1)\right\}\right] \\
&= \left\| \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{D}_\infty} \tilde{C}_\infty^{-1/2} W_f(1) \right\|^2 := \mathcal{J}_{\infty,\theta},
\end{aligned}$$

which holds jointly with $\mathcal{K}_T(\theta_0) \Rightarrow \mathcal{K}_{\infty,\theta}$. Finally, given that $\mathcal{S}_T(\theta_0) = \mathcal{J}_T(\theta_0) + \mathcal{K}_T(\theta_0)$, we have

$$\mathcal{S}_T(\theta_0) \Rightarrow \left\| \tilde{C}_\infty^{-1/2} W_f(1) \right\|^2 = W_f(1)' \tilde{C}_\infty^{-1} W_f(1) := \mathcal{S}_{\infty,\theta}.$$

Although the fixed-smoothing asymptotic distributions of $\mathcal{K}_T(\theta_0)$, $\mathcal{J}_T(\theta_0)$, and $\mathcal{S}_T(\theta_0)$ look complicated, they provide the basis for developing the asymptotic F theory for $\mathcal{K}_T^*(\theta_0)$, $\mathcal{J}_T^*(\theta_0)$, and $\mathcal{S}_T^*(\theta_0)$. In the next two sections, we focus on the modified K and J statistics and obtain their asymptotic distributions under different assumptions on the strength of the identification. Since the distribution of the modified S statistic does not depend on the identification assumptions, its analysis is carried out separately in Section 6.

4 Asymptotic F Theory for the Modified K and J Statistics

4.1 The case with possibly weak identification

In this subsection, we consider the case $\kappa \in [0, 1/2)$. When $\kappa = 0$, we have the usual case of strong identification, and θ_0 is $\sqrt{T}$ estimable. When $\kappa \in (0, 1/2)$, θ_0 can still be consistently estimated, but the rate of convergence is slower than the usual parametric $\sqrt{T}$ rate. See, for example, Caner (2010).

By Lemma 3.2, when $\kappa \in [0, 1/2)$, the limit D_∞ of the normalized CU Jacobian, namely $D_T / T^{1/2-\kappa}$, is a deterministic matrix. By definition, D_∞ is independent of any random variable including $B_f(1)$, the limit of the empirical moment vector. Therefore, the normalized CU Jacobian is asymptotically independent of the empirical moment vector. This brings our case with $\kappa \in [0, 1/2)$ close to what is considered in Kleibergen (2005). Nevertheless, there is a major difference. Under the usual asymptotics that Kleibergen (2005) employs, the LRV estimator $\hat{V}_{ff}$ converges in probability to the true LRV V_{ff}, and so the randomness of the weighting matrix in the CU-GMM objective function $Q_T(\theta_0)$ vanishes. By contrast, under the fixed-smoothing asymptotics, the randomness is retained in the limit with the consequential effect on the asymptotic distribution of the K statistic. The GMM with a random weighting matrix in the limit has been studied by Sun (2014b) and Hwang and Sun (2017), who consider a two-step GMM framework with strong identification. Despite this difference, we can adopt the idea in Sun (2014b) and Hwang and Sun (2017) to prove the following theorem.

Theorem 4.1 *Let Assumptions 3.1–3.3 hold. If $\kappa \in [0, 1/2)$ and Π has a full column rank, then*

$$\left[\mathcal{K}_T^*(\theta_0), \mathcal{J}_T^*(\theta_0)\right] \Rightarrow \left[\mathcal{K}_{\infty,\theta}^*, \mathcal{J}_{\infty,\theta}^*\right] =^d \left[F_{d,G-m+1}, F_{q,G-q+1}\right],$$

where $F_{d,G-m+1}$ and $F_{q,G-q+1}$ are independent F variates.

Partitioning the matrix $\tilde{C}_\infty$ and the vector $W_f(1)$, we write

$$\tilde{C}_\infty = \frac{1}{G}\sum_{\ell=1}^{G} \eta_{f,\ell}\eta_{f,\ell}' = \begin{pmatrix} \tilde{C}_{dd}, & \tilde{C}_{dq} \\ \tilde{C}_{qd}, & \tilde{C}_{qq} \end{pmatrix} \text{ and } W_f(1) = \begin{pmatrix} W_{f,d}(1) \\ W_{f,q}(1) \end{pmatrix}, \tag{9}$$

where $\tilde{C}_{dd} \in \mathbb{R}^{d\times d}$, $\tilde{C}_{qq} \in \mathbb{R}^{q\times q}$, $\tilde{C}_{dq} \in \mathbb{R}^{d\times q}$, $\tilde{C}_{qd} \in \mathbb{R}^{q\times d}$, $W_{f,d}(1) \in \mathbb{R}^{d\times 1}$, and $W_{f,q}(1) \in \mathbb{R}^{q\times 1}$. To obtain the asymptotic F distribution for $\mathcal{K}_T^*(\theta_0)$ in Theorem 4.1, we first use the SVD of $\tilde{D}_\infty$ and the partitioned matrix inversion formula to show that

$$\mathcal{K}_{\infty,\theta} =^d \left[W_{f,d}(1) - \tilde{C}_{dq}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]' \tilde{C}_{dd\cdot q}^{-1}\left[W_{f,d}(1) - \tilde{C}_{dq}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right],$$

where $\tilde{C}_{dd\cdot q} = \tilde{C}_{dd} - \tilde{C}_{dq}\tilde{C}_{qq}^{-1}\tilde{C}_{qd}$. That is, $\mathcal{K}_{\infty,\theta}$ is equal in distribution to a quadratic form in the vector $W_{f,d}(1) - \tilde{C}_{dq}\tilde{C}_{qq}^{-1}W_{f,q}(1)$ with $\tilde{C}_{dd\cdot q}$ as the (inverse) weighting matrix. While Hotelling's T^2 distribution also takes a quadratic form, the underlying vector is standard normal and the weighting matrix follows an independent Wishart distribution. This inspires us to transform the original K statistic $\mathcal{K}_T(\theta_0)$ so that the resulting limiting distribution takes the form of a T^2 distribution. This transformation involves the J statistic $\mathcal{J}_T(\theta_0)$ and amounts to dividing the original K statistic by $1 + G^{-1}\mathcal{J}_T(\theta_0)$. With additional multiplicative rescaling, we then turn the T^2 distribution into the standard F distribution. This explains how the modified K statistic $\mathcal{K}_T^*(\theta_0)$ is constructed.

Theorem 4.1 shows that the modified K and J statistics converge weakly to independent F distributions. This intriguing result is used later to design a new test that overcomes a drawback of the LM-type or score-type tests such as the K test.

4.2 The case with complete identification failure

We consider $\kappa = 1/2$, in which case, by Lemma 3.2,

$$D_{T,j}(\theta_0) \Rightarrow D_{\infty,j} = \Pi_j + \tilde{B}_{g_j\cdot f}(1) = D_{\infty,j} \in \mathbb{R}^{m\times 1}.$$

We further assume that $\Pi = 0$ so that $\sqrt{T}\mathbb{E}g(Y_t,\theta_0) \to 0$ as $T \to \infty$. That is, $\mathbb{E}g(Y_t,\theta_0)$ converges to a zero matrix at a rate faster than $1/\sqrt{T}$. When $\kappa = 1/2$ and $\Pi = 0$, the model parameter θ_0 is not identified, and we have complete identification failure. An extreme case is $\mathbb{E}g(Y_t,\theta_0) = 0$, in which case the moment conditions provide no information about θ_0 and none of the linear combinations of θ_0 is identified. An example of this case is an instrumental variable (IV) regression where the instruments are completely irrelevant.

In the case of complete identification failure, $D_T(\theta_0)$ converges to a random matrix D_∞. The columns of the random matrix D_∞ are $\{\tilde{B}_{g_j\cdot f}(1)\}$, which are not independent of $\{\tilde{C}_\infty^{-1}, W_f(1)\}$. This scenario is completely different from that of Kleibergen (2005), where the CU Jacobian is asymptotically independent of the empirical moment vector. We also cannot use the same arguments as in the proofs of Theorem 4.1 directly. Nevertheless, we can establish the same result as in Theorem 4.1.

Theorem 4.2 *Let Assumptions 3.1–3.3 hold. If $\kappa = 1/2$ and $\Pi = 0$, then*

$$[\mathcal{K}_T^*(\theta_0), \mathcal{J}_T^*(\theta_0)] \Rightarrow \left[\mathcal{K}_{\infty,\theta}^*, \mathcal{J}_{\infty,\theta}^*\right] =^d [F_{d,G-m+1}, F_{q,G-q+1}],$$

where $F_{d,G-m+1}$ and $F_{q,G-q+1}$ are independent F variates.

To prove the first part of the theorem, we first represent $D_{\infty,j} = \tilde{B}_{g_j \cdot f}(1)$ in terms of the scale-free random variables and processes:

$$\tilde{B}_{g_j \cdot f}(1) = V_{g_j \cdot f}^{1/2} \tilde{W}_{g_j \cdot f}(1),$$

where

$$\tilde{W}_{g_j \cdot f}(1) = W_{g_j \cdot f}(1) - \left[\frac{1}{G} \sum_{\ell=1}^{G} \eta_{g_j \cdot f,\ell} \eta_{f,\ell}' \right] \left[\frac{1}{G} \sum_{\ell=1}^{G} \eta_{f,\ell} \eta_{f,\ell}' \right]^{-1} W_f(1). \tag{10}$$

Owing to the presence of the second term in (10), which captures the projection uncertainty, $\tilde{W}_{g_j \cdot f}(1)$ is not independent of $\{\tilde{C}_\infty^{-1},\ W_f(1)\}$. The question is whether a transformation of $\tilde{W}_{g_j \cdot f}(1)$ is independent of $\{\tilde{C}_\infty^{-1}, W_f(1)\}$. In view of the conditional normality of $\tilde{W}_{g_j \cdot f}(1)$ conditional on $\{\eta_{f,\ell} : \ell = 1, \ldots, G\}$ and $W_f(1)$, it is natural to employ rescaling by its conditional standard deviation as the transformation. We show in the proof of Theorem 4.2 that the conditional variance of $\tilde{W}_{g_j \cdot f}(1)$ is $[1 + G^{-1} W_f(1)' \tilde{C}_\infty^{-1} W_f(1)] I_m$, a scalar matrix. This motivates us to define

$$\tilde{D}_{\infty,j}^* := \frac{\tilde{D}_{\infty,j}}{\sqrt{1 + G^{-1} W_f(1)' \tilde{C}_\infty^{-1} W_f(1)}}. \tag{11}$$

The key step in proving Theorem 4.2 is to show that $\{\tilde{D}_{\infty,j}^*\}$ are independent of $\tilde{C}_\infty^{-1}$ and $W_f(1)$. Then, by conditioning on $\{\tilde{D}_{\infty,j}^*\}$, we effectively reduce the problem to the previous case in which the limit of the CU Jacobian is a deterministic matrix.

Note that $\mathcal{S}_T(\theta_0) \Rightarrow W_f(1)' \tilde{C}_\infty^{-1} W_f(1)$. In terms of the K and J statistics, the rescaling defined in (11) amounts to replacing $D_T(\theta_0)$ by $D_T(\theta_0) / \sqrt{1 + G^{-1} \mathcal{S}_T(\theta_0)}$ in the definitions of $\mathcal{K}_T(\theta_0)$ and $\mathcal{J}_T(\theta_0)$. Such a replacement has no effect on these two statistics, as the relevant terms take a self-normalized form. We employ the rescaling in (11) only in our theoretical development. There is no need to change the original or modified statistic.

It is reassuring that $[\mathcal{K}_T^*(\theta_0), \mathcal{J}_T^*(\theta_0)] \Rightarrow [F_{d,G-m+1}, F_{q,G-q+1}]$, regardless of whether θ_0 is weakly identified or unidentified. We can use the F approximations without knowing the strength of the identification.

4.3 The intermediate case

We consider the case with $\kappa = 1/2$ but now

$$D_{T,j}(\theta_0) \Rightarrow D_{\infty,j} = \Pi_j + \tilde{B}_{g_j \cdot f}(1),$$

for some $\Pi \neq 0$. That is, $\sqrt{T} \mathbb{E} g_j(Y_t, \theta_0) \neq 0$ for some j. In this case, θ_0 is not identified, as the noise and signal are of the same order. However, there is still some information about θ_0. The case is analogous to the simple IV regression where the covariance between the instrument and endogenous variable is small in that it goes to zero at the exact rate of $1/\sqrt{T}$.

We proceed as in Section 4.2 to define

$$\begin{aligned}\tilde{D}^*_{\infty,j} &= \frac{\tilde{D}_{\infty,j}}{\sqrt{1+G^{-1}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} \\ &= V_{ff}^{-1/2}\frac{\Pi_j}{\sqrt{1+G^{-1}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} + V_{ff}^{-1/2}V_{g_j\cdot f}^{1/2}\frac{\tilde{W}_{g_j\cdot f}(1)}{\sqrt{1+G^{-1}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} \\ &= V_{ff}^{-1/2}\frac{\Pi_j}{\sqrt{1+G^{-1}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} + V_{ff}^{-1/2}V_{g_j\cdot f}^{1/2}\tilde{W}^*_{g_j\cdot f}(1).\end{aligned}$$

Compared with $\tilde{D}^*_{\infty,j}$ in Section 4.2, $\tilde{D}^*_{\infty,j}$ now contains an additional term, namely the first term in the above expression. The presence of this first term invalidates our arguments in Section 4.2 because $W_f(1)'\tilde{C}_\infty^{-1}W_f(1)$ and hence the first term are not independent of $\{\tilde{C}_\infty^{-1}, W_f(1)\}$. More fundamentally, it is the mixture of the two distributions in $D_{\infty,j}$ that creates a dilemma. If only Π_j is present for all j, we can use the arguments in Section 4.1 to establish the asymptotic F theory. If only $\tilde{B}_{g_j\cdot f}(1)$ is present for all j, then we can use the arguments in Section 4.2 to establish the same theory. However, the arguments in Sections 4.1 and 4.2 are different and cannot be included in a unified framework.

In general, when $V_{g\cdot f} \neq 0$, the limiting distribution $\mathcal{K}_{\infty,\theta}$ is not free of nuisance parameters. Appendix B provides detailed arguments. However, under the increasing-smoothing asymptotics wherein $G \to \infty$ but $G/T \to 0$ as $T \to \infty$, we have

$$\mathcal{K}_T^* = \frac{\mathcal{K}_T(\theta_0)}{d}(1+o_p(1)) \Rightarrow \chi^2_d/d,$$

using the same arguments in proving the chi-squared approximation in Section 2.2. In the meanwhile, we have

$$F_{d,G-m+1} \Rightarrow \chi^2_d/d,$$

as $G \to \infty$. Therefore, the F approximation $F_{d,G-m+1}$ is asymptotically valid for the modified K statistic $\mathcal{K}_T^*$ under the increasing-smoothing asymptotics. Similarly, the F approximation $F_{q,G-q+1}$ is asymptotically valid for the modified J statistic $\mathcal{J}_T^*$.

Although we cannot establish the asymptotic F theory for the intermediate case, we recommend using it in practice. Our simulation results show that the F approximations are still more accurate than the chi-squared approximations in finite samples. This is expected, as the F approximations still capture much of the uncertainty in the underlying LRV estimators, while the chi-squared approximations completely ignore it.

5 Testing Hypotheses on Subsets of the Parameters

In the previous sections, the null hypothesis pins down all the elements of θ_0. However, in many empirical applications, we may be interested in testing only a subset of the parameters. Suppose now that θ can be partitioned as $\theta = (\alpha', \beta')'$, where $\alpha \in \mathbb{R}^{d_\alpha}$, $\beta \in \mathbb{R}^{d_\beta}$, and $d_\alpha + d_\beta = d$. Assume further that α is locally strongly identified (see Assumption 5.1), whereas β may not be. We would like to test the null of

$$H_0^*: \beta = \beta_0$$

while leaving α unspecified. Under the null hypothesis above, we estimate α by

$$\hat{\alpha}\left(\beta_0\right) = \arg\min_{\alpha\in\mathcal{A}} Q_T\left(\alpha, \beta_0\right),$$

where $\mathcal{A}$ is a compact parameter space. For notational simplicity, we write $\hat{\alpha} = \hat{\alpha}\left(\beta_0\right)$ with the understanding that $\hat{\alpha}$ is the restricted CU-GMM estimator under the null H_0^*. Let α_0 be the probability limit of $\hat{\alpha}_0$. Denote $\theta_0 = (\alpha_0', \beta_0')'$ and $\hat{\theta}_0 = (\hat{\alpha}', \beta_0')'$.

The K statistic for testing H_0^* is

$$\mathcal{K}_T(\hat{\theta}_0) = \left(\frac{\partial Q_T(\hat{\theta}_0)}{\partial\theta}\right)' \left[D_T(\hat{\theta}_0)'\hat{V}_{ff}^{-1}(\hat{\theta}_0)D_T(\hat{\theta}_0)\right]^{-1} \left(\frac{\partial Q_T(\hat{\theta}_0)}{\partial\theta}\right),$$

and the S statistic is $\mathcal{S}_T(\hat{\theta}_0) = 2Q_T(\hat{\theta}_0)$. The construction of the statistics $\mathcal{K}_T(\hat{\theta}_0)$ and $\mathcal{S}_T(\hat{\theta}_0)$ is the same as in the case for testing the full parameter vector. The only difference is that all the statistics are now evaluated at $\hat{\theta}_0$.

The modified K statistic involves the J statistic:

$$\mathcal{J}_T(\hat{\theta}_0) = \mathcal{S}_T(\hat{\theta}_0) - \mathcal{K}_T(\hat{\theta}_0) = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{f}(Y_t,\hat{\theta}_0)\right]' \hat{V}_{ff}^{-1}(\hat{\theta}_0)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{f}(Y_t,\hat{\theta}_0)\right],$$

which takes the same form as before but is now evaluated at $\hat{\theta}_0$. With $\mathcal{J}_T(\hat{\theta}_0)$, we define the modified K, J, and S statistics as

$$\begin{aligned}
\mathcal{K}_T^*(\hat{\theta}_0) &= \frac{G-d_\beta-q+1}{Gd_\beta}\frac{\mathcal{K}_T(\hat{\theta}_0)}{1+G^{-1}\mathcal{J}_T(\hat{\theta}_0)}, \\
\mathcal{J}_T^*(\hat{\theta}_0) &= \frac{G-q+1}{Gq}\mathcal{J}_T(\hat{\theta}_0), \\
\mathcal{S}_T^*(\hat{\theta}_0) &= \frac{G-d_\beta-q+1}{G\left(d_\beta+q\right)}\mathcal{S}_T(\hat{\theta}_0).
\end{aligned}$$

To obtain an explicit expression for the modified K statistic, we partition each of $D_T\left(\theta\right)$ and $g(Y_t,\theta)$ into two blocks:

$$\begin{aligned}
D_T\left(\theta\right) &= \left[D_{T,\alpha}\left(\theta\right), D_{T,\beta}\left(\theta\right)\right] \text{ with } D_{T,\alpha}\left(\theta\right)\in\mathbb{R}^{m\times d_\alpha} \text{ and } D_{T,\beta}\left(\theta\right)\in\mathbb{R}^{m\times d_\beta} \\
g(Y_t,\theta) &= \left[g_\alpha(Y_t,\theta), g_\beta(Y_t,\theta)\right] \text{ with } g_\alpha(Y_t,\theta)\in\mathbb{R}^{m\times d_\alpha} \text{ and } g_\beta(Y_t,\theta)\in\mathbb{R}^{m\times d_\beta}.
\end{aligned}$$

Then,

$$\begin{aligned}
\frac{\partial Q_T(\hat{\theta}_0)}{\partial\theta} &= D_T(\hat{\theta}_0)'V_{ff}^{-1}(\hat{\theta}_0)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(Y_t,\hat{\theta}_0)\right] \\
&= \begin{pmatrix} D_{T,\alpha}(\hat{\theta}_0)'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(Y_t,\hat{\theta}_0)\right] \\ D_{T,\beta}(\hat{\theta}_0)'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(Y_t,\hat{\theta}_0)\right] \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{0}_{d_\alpha\times d} \\ D_{T,\beta}(\hat{\theta}_0)'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(Y_t,\hat{\theta}_0)\right] \end{pmatrix}
\end{aligned}$$

by the first-order conditions for estimating $\hat{\alpha}$.

Since the first block of $\frac{\partial Q_T(\hat{\theta}_0)}{\partial \theta}$ is zero, to compute the K statistic, we need only to find the lower-right block of the matrix $[D_T(\hat{\theta}_0)'\hat{V}_{ff}^{-1}(\hat{\theta}_0)D_T(\hat{\theta}_0)]^{-1}$. By the partitioned matrix inverse formula, this lower-right block is

$$\left[\hat{D}_{T,\beta}'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\hat{D}_{T,\beta} - \hat{D}_{T,\beta}'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\hat{D}_{T,\alpha}\left(\hat{D}_{T,\alpha}'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\hat{D}_{T,\alpha}\right)^{-1}\hat{D}_{T,\alpha}'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\hat{D}_{T,\beta}\right]^{-1},$$

where $\hat{D}_{T,\beta} = D_{T,\beta}(\hat{\theta}_0)$ and $\hat{D}_{T,\alpha} = D_{T,\alpha}(\hat{\theta}_0)$. Hence, the K statistic becomes

$$\begin{aligned}
\mathcal{K}_T(\hat{\theta}_0) &= \left[D_{T,\beta}(\hat{\theta}_0)'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t,\hat{\theta}_0)\right]' \\
&\quad \times \left[\hat{D}_{T,\beta}'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\hat{D}_{T,\beta} - \hat{D}_{T,\beta}'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\hat{D}_{T,\alpha}\left(\hat{D}_{T,\alpha}'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\hat{D}_{T,\alpha}\right)^{-1}\hat{D}_{T,\alpha}'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\hat{D}_{T,\beta}\right]^{-1} \\
&\quad \times \left[D_{T,\beta}(\hat{\theta}_0)'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t,\hat{\theta}_0)\right].
\end{aligned}$$

To establish the fixed-smoothing asymptotic distributions of $\mathcal{K}_T(\hat{\theta}_0)$ and $\mathcal{J}_T(\hat{\theta}_0)$ and hence those of $\mathcal{K}_T^*(\hat{\theta}_0)$ and $\mathcal{J}_T^*(\hat{\theta}_0)$, we make the following assumptions.

Assumption 5.1 *(a) The $m \times d_\alpha$ Jacobian matrix*

$$\mathbb{E}g_\alpha(Y_t,\theta_0) = \Pi_\alpha := (\Pi_1,\ldots,\Pi_{d_\alpha})$$

has full column rank d_α, and $\mathbb{E}g_\alpha(Y_t,\theta)$ is a continuous function of θ at θ_0. (b) The $m \times d_\beta$ Jacobian matrix $\mathbb{E}g_\beta(Y_t,\theta_0)$ satisfies: for $\kappa \in [0,1/2]$,

$$T^\kappa \mathbb{E}g_\beta(Y_t,\theta_0) \to \Pi_\beta := (\Pi_{d_\alpha+1},\ldots,\Pi_d),$$

and Π_β has full column rank for $\kappa \in [0,1/2)$.

Assumption 5.2 *$f(Y_t,\cdot)$ is twice continuously differentiable and*
(a) (i) uniformly over $s \in [0,1]$,

$$\frac{1}{T}\sum_{t=1}^{[Ts]} g_\alpha(Y_t,\hat{\theta}_0) = \Pi_\alpha \cdot s + o_p(1),$$

(ii) uniformly over $s \in [0,1]$,

$$\frac{1}{T}\sum_{t=1}^{[Ts]} \frac{\partial g_j(Y_t,\hat{\theta}_0)}{\partial \alpha'} = \Pi_{(j),\alpha} \cdot s + o_p(1),$$

for $j = 1,\ldots,d_\alpha$ where $\Pi_{(j),\alpha}$ is an $m \times d_\alpha$ constant matrix for each $j = 1,\ldots,d_\alpha$.
(b) $\sqrt{T}(\hat{\alpha}-\alpha_0) = O_p(1)$ and

$$\sqrt{T}(\hat{\alpha}-\alpha_0) = -\left[\frac{D_{T,\alpha}'}{\sqrt{T}}\hat{V}_{ff}^{-1}\frac{D_{T,\alpha}}{\sqrt{T}}\right]^{-1}\left\{\frac{D_{T,\alpha}'}{\sqrt{T}}\hat{V}_{ff}^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t,\theta_0)\right]\right\} + o_p(1),$$

where $D_{T,\alpha} = D_{T,\alpha}(\theta_0)$ *and* $\hat{V}_{ff}^{-1} = \hat{V}_{ff}^{-1}(\theta_0)$.

(*c*) (*i*) $T^{\kappa}\mathbb{E}\left[g_\beta(Y_t,\theta_0) - g_\beta(Y_t,\hat{\theta}_0)\right] = o_p(1)$ *and* (*ii*)

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[g_\beta(Y_t,\hat{\theta}_0) - Eg_\beta(Y_t,\hat{\theta}_0)\right] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[g_\beta(Y_t,\theta_0) - Eg_\beta(Y_t,\theta_0)\right] + o_p(1).$$

Assumption 5.2 (a) is used to prove Lemma 5.1(a) below. This is a standard assumption in the literature on fixed-smoothing asymptotics.

Under fixed-smoothing asymptotics, the weighting matrix in the GMM criterion does not converge to a deterministic matrix. For this reason, Assumption 5.2 (b) does not follow directly from standard textbook arguments. Nevertheless, sufficient primitive conditions can be found in Zhang (2016), who considers the fixed-smoothing asymptotics under strong identification for generalized empirical likelihood estimators, which include the CU-GMM as a special case.

Assumption 5.2 (c.i) holds by the dominated convergence theorem if $\mathbb{E}\sup_{\alpha\in\mathcal{A}}\|T^{\kappa}g_\beta(Y_t,\alpha,\beta_0)\| < \infty$. Assumption 5.2 (c.ii) is a stochastic equicontinuity condition, which holds under some moment and mixing conditions.

Lemma 5.1 *Let Assumptions 3.2, 3.3, 5.1, and 5.2 hold.*

(*a*) *For any* $\tilde{\theta}_0$ *between* $\hat{\theta}_0$ *and* θ_0, *we have*

$$T^{-1}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)g_\alpha\left(Y_t,\tilde{\theta}_0\right) = o_p(1)$$
$$T^{-1}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\frac{\partial g_j(Y_t,\tilde{\theta}_0)}{\partial\alpha'} = o_p(1) \text{ for } j = 1,\ldots,d_\alpha.$$

(*b*) *The following convergence results hold jointly*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[f\left(Y_t,\hat{\theta}_0\right) - \bar{f}(Y,\hat{\theta}_0)\right] \Rightarrow \int_0^1\Phi_\ell(r)\,dB_f(r) := \xi_{f,\ell} \tag{12}$$
$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[g_j\left(Y_t,\hat{\theta}_0\right) - \bar{g}_j(Y,\hat{\theta}_0)\right] \Rightarrow \int_0^1\Phi_\ell(r)\,dB_{g_j}(r) := \xi_{g_j,\ell} \tag{13}$$

for $j = 1,\ldots,d$ *and* $\ell = 1,\ldots,G$.

(*c*) *For* $C_\infty = G^{-1}\sum_{\ell=1}^{G}\xi_{f,\ell}\xi_{f,\ell}'$, *we have*

$$\begin{aligned}&V_{ff}^{-1/2}(\hat{\theta}_0)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(Y_t,\hat{\theta}_0)\right]\\ &= \mathbb{M}_{\hat{V}_{ff}^{-1/2}D_{T,\alpha}}\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{V}_{ff}^{-1/2}f(Y_t,\theta_0)\right)(1+o_p(1)) \Rightarrow \mathbb{M}_{C_\infty^{-1/2}\Pi_\alpha}\cdot C_\infty^{-1/2}B_f(1).\end{aligned}$$

(*d*) *Let*

$$\tilde{B}^*_{g_j\cdot f}(1) = B_{g,j}(1) - \left[\frac{1}{G}\sum_{\ell=1}^{G}\xi_{g_j}\xi_{f,\ell}'\right]C_\infty^{-1/2}\mathbb{M}_{C_\infty^{-1/2}\Pi_\alpha}\cdot C_\infty^{-1/2}B_f(1),$$

$$D_{\infty,\beta,j} = \Pi_{d_\alpha+j} + 1\left\{\kappa = \frac{1}{2}\right\}\tilde{B}^*_{g_{d_\alpha+j}\cdot f}(1), j = 1,\ldots,d_\beta,$$

and

$$D_{\infty,\beta} = \left(D_{\infty,\beta,1},\ldots,D_{\infty,\beta,d_\beta}\right) \in \mathbb{R}^{m\times d_\beta}.$$

Then,

$$D_{T,\beta}(\hat{\theta}_0)/T^{1/2-\kappa} \Rightarrow D_{\infty,\beta}.$$

Lemma 5.1(a) follows from Assumptions 3.3 and 5.2(a). It is used to prove Lemma 5.1(b). Lemma 5.1(b) shows that the asymptotic distributions are invariant to the estimation error in $\hat{\theta}_0$. As a result, the limiting distributions of $\hat{V}_{ff}(\hat{\theta}_0)$, $\hat{V}_{g_jg_j}(\theta_0)$, and $\hat{V}_{fg_j}(\hat{\theta}_0)$ are the same as those of $\hat{V}_{ff}(\theta_0)$, $\hat{V}_{g_jg_j}(\theta_0)$, and $\hat{V}_{fg_j}(\theta_0)$. The key assumption behind this invariance to the estimation error in $\hat{\theta}_0$ is $\int_0^1\Phi_\ell(r)\,dr = 0$. Lemma 5.1(c) shows that the estimation error in $\hat{\theta}_0$ has an asymptotic effect on the normalized moment conditions. The effect is captured via the projection matrix $\mathbb{M}_{\hat{V}_{ff}^{-1/2}D_{T,\alpha}}$. The limit of $D_{T,\beta}(\hat{\theta}_0)/T^{1/2-\kappa}$ in Lemma 5.1(d) is similar to that in Lemma 3.2. The difference is that compared with $B^*_{g_j\cdot f}(1)$ defined in (7), $\tilde{B}^*_{g_j\cdot f}(1)$ contains the additional projection matrix $\mathbb{M}_{C_\infty^{-1/2}\Pi_\alpha}$, which captures the effect from estimating α_0.

In the next two subsections, we establish the fixed-smoothing asymptotic distributions for the modified K and J statistics. We defer the treatment of the modified S statistic to Section 6.

5.1 The case with possibly weak identification

In this subsection, we consider $\kappa \in [0,1/2)$. In this case, $D_{T,\beta}(\hat{\theta}_0)/T^{1/2-\kappa}$ converges to a deterministic matrix $D_{\infty,\beta} = \Pi_\beta$. Hence,

$$\begin{aligned}
\mathcal{K}_T(\hat{\theta}_0) &= \left[\frac{D_{T,\beta}(\hat{\theta}_0)'}{T^{1/2-\kappa}}\hat{V}_{ff}^{-1}(\hat{\theta}_0)\frac{1}{\sqrt{T}}\sum_{t=1}^T f(Y_t,\hat{\theta}_0)\right]' \\
&\times\left[\frac{\hat{D}'_{T,\beta}}{T^{1/2-\kappa}}\hat{V}_{ff}^{-1}\frac{\hat{D}_{T,\beta}}{T^{1/2-\kappa}} - \frac{D'_{T,\beta}}{T^{1/2-\kappa}}\hat{V}_{ff}^{-1}\frac{D_{T,\alpha}}{\sqrt{T}}\left(D'_{T,\alpha}\hat{V}_{ff}^{-1}D_{T,\alpha}\right)^{-1}\frac{D'_{T,\alpha}}{\sqrt{T}}\hat{V}_{ff}^{-1}\frac{D_{T,\beta}}{T^{1/2-\kappa}}\right]^{-1} \\
&\times\left[\frac{D_{T,\beta}\left(\hat{\theta}_0\right)'}{T^{1/2-\kappa}}\hat{V}_{ff}^{-1}(\hat{\theta}_0)\frac{1}{\sqrt{T}}\sum_{t=1}^T f(Y_t,\hat{\theta}_0)\right]' \\
&\Rightarrow\left[\Pi'_\beta C_\infty^{-1/2}\mathbb{M}_{C_\infty^{-1/2}\Pi_\alpha}C_\infty^{-1/2}B_f(1)\right]' \\
&\times\left[\Pi'_\beta C_\infty^{-1}\Pi_\beta - \Pi'_\beta C_\infty^{-1}\Pi_\alpha\left[\Pi'_\alpha C_\infty^{-1}\Pi_\alpha\right]^{-1}\Pi'_\alpha C_\infty^{-1}\Pi_\beta\right]^{-1} \\
&\times\left[\Pi'_\beta C_\infty^{-1/2}\mathbb{M}_{C_\infty^{-1/2}\Pi_\alpha}C_\infty^{-1/2}B_f(1)\right] := \mathcal{K}_{\infty,\beta}.
\end{aligned}$$

In terms of the scale-free variables and processes, we have

$$\begin{aligned}
\mathcal{K}_{\infty,\beta} &= \left[\tilde{\Pi}'_\beta\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}W_f(1)\right]' \\
&\times\left[\tilde{\Pi}'_\beta\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\beta\right]^{-1}\times\left[\tilde{\Pi}'_\beta\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}W_f(1)\right] \qquad (14) \\
&= \left\|\mathbb{P}_{\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\left[\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\beta\right]}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}W_f(1)\right\|^2,
\end{aligned}$$

where

$$\tilde{\Pi}_\beta = V_{ff}^{-1/2}\Pi_\beta \text{ and } \tilde{\Pi}_\alpha = V_{ff}^{-1/2}\Pi_\alpha.$$

Owing to the presence of the projection matrix $\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}$, the strategy for simplifying the limiting distribution $\mathcal{K}_{\infty,\theta}$ in the whole vector case does not work here straightforwardly. Instead, we show in the proof of the next theorem that $\mathcal{K}_{\infty,\beta}$ can be represented by

$$\begin{aligned}&\mathcal{K}_{\infty,\beta}\\ &=^d \left\{R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}\tilde{\Pi}'\tilde{C}_\infty^{-1}W_f(1)\right\}'\left[R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}R_\beta'\right]^{-1}\left\{R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}\tilde{\Pi}'\tilde{C}_\infty^{-1}W_f(1)\right\},\end{aligned}$$

where $\tilde{\Pi} = (\tilde{\Pi}_\alpha, \tilde{\Pi}_\beta)$ and

$$R_\beta = \begin{pmatrix} O_{d_\beta\times d_\alpha}, & I_{d_\beta\times d_\beta}\end{pmatrix}.$$

This representation is identical to the fixed-smoothing limit of the Wald statistic in Hwang and Sun (2017), who consider testing $R_\beta\theta = \beta_0$ in a two-step efficient GMM framework. Using the distributional representation in Hwang and Sun (2017), we find that

$$\mathcal{K}_{\infty,\beta} =^d \left[W_{f,d_\beta}(1) - \tilde{C}_{d_\beta,q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]' C_{d_\beta d_\beta\cdot q}^{-1}\left[W_{f,d_\beta}(1) - \tilde{C}_{d_\beta,q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]. \tag{15}$$

Therefore, the distribution of $\mathcal{K}_{\infty,\beta}$ takes the same quadratic form as the distributional representation of $\mathcal{K}_{\infty,\theta}$ given in (8). The only difference lies in the dimensionality: while the vector in the quadratic form for $\mathcal{K}_{\infty,\theta}$ is of dimension d, that for $\mathcal{K}_{\infty,\beta}$ is of dimension d_β.

In view of the similarity between $\mathcal{K}_{\infty,\beta}$ and $\mathcal{K}_{\infty,\theta}$, it is not surprising that the asymptotic distribution of $\mathcal{K}_T^*(\hat{\theta}_0)$ is also an F distribution. The theorem below presents this result together with the asymptotic distribution of $\mathcal{J}_T^*(\hat{\theta}_0)$.

Theorem 5.1 *Let Assumptions 3.2, 3.3, 5.1, and 5.2 hold. If $\kappa \in [0, 1/2)$ and Π has a full column rank, then*

$$\left[\mathcal{K}_T^*(\hat{\theta}_0), \mathcal{J}_T^*(\hat{\theta}_0)\right] \Rightarrow \left[F_{d_\beta,G-d_\beta-q+1}, F_{q,G-q+1}\right],$$

where $F_{d_\beta,G-d_\beta-q+1}$ and $F_{q,G-q+1}$ are independent F variates.

Theorem 5.1 is similar to Theorem 4.1. When α is the null or empty vector, we have $d_\beta = d_\theta = d$, $F_{d_\beta,G-d_\beta-q+1} = F_{d,G-m+1}$. Hence, in this special case, Theorem 5.1 reduces to Theorem 4.1.

5.2 The case with complete identification failure

We consider the case with $\kappa = 1/2$ and assume that $\Pi_\beta = 0$. Replacing $\tilde{\Pi}_\beta$ by $\tilde{D}_{\infty,\beta}$ in (14), we now have

$$\begin{aligned}\mathcal{K}_{\infty,\beta} &= \left\{\tilde{D}_{\infty,\beta}'\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}W_f(1)\right\}'\\ &\quad\times\left\{\tilde{D}_{\infty,\beta}'\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}\tilde{D}_{\infty,\beta}\right\}^{-1}\times\left\{\tilde{D}_{\infty,\beta}'\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}W_f(1)\right\},\end{aligned}$$

where

$$\tilde{D}_{\infty,\beta} = \left(\tilde{D}_{\infty,\beta,1}, \ldots, \tilde{D}_{\infty,\beta,d_\beta}\right) \in \mathbb{R}^{m\times d_\beta}$$

for $\tilde{D}_{\infty,\beta,j} = V_{ff}^{-1/2} \tilde{B}^{*}_{g_{d_\alpha+j} \cdot f}(1)$. Noting that

$$
\begin{aligned}
& \tilde{B}^{*}_{g_j \cdot f}(1) \\
&= B_{g_j}(1) - \left[\frac{1}{G} \sum_{\ell=1}^{G} \xi_{g_j,\ell} \eta'_{f,\ell} \right] \tilde{C}_\infty^{-1/2} \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha} \cdot \tilde{C}_\infty^{-1/2} W_f(1) \\
&= V_{g_j f} V_{ff}^{-1/2} W_f(1) + V_{g_j \cdot f}^{1/2} W_{g_j \cdot f}(1) \\
&- \frac{1}{G} \sum_{\ell=1}^{G} \left[V_{g_j f} V_{ff}^{-1/2} \eta_{f,\ell} \eta'_{f,\ell} + V_{g_j \cdot f}^{1/2} \eta_{g_j \cdot f,\ell} \eta'_{f,\ell} \right] \tilde{C}_\infty^{-1/2} \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha} \cdot \tilde{C}_\infty^{-1/2} W_f(1) \\
&= V_{g_j f} V_{ff}^{-1/2} \left[I - \tilde{C}_\infty \tilde{C}_\infty^{-1/2} \mathbb{M}_{C_\infty^{-1/2} \tilde{\Pi}_\alpha} \cdot C_\infty^{-1/2} \right] W_f(1) \\
&+ V_{g_j \cdot f}^{1/2} \left[W_{g_j \cdot f}(1) - \left(\frac{1}{G} \sum_{\ell=1}^{G} \eta_{g_j \cdot f,\ell} \eta'_{f,\ell} \right) \tilde{C}_\infty^{-1/2} \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha} \cdot \tilde{C}_\infty^{-1/2} W_f(1) \right] \\
&:= V_{g_j f} V_{ff}^{-1/2} W_f^{(1)}(1) + V_{g_j \cdot f}^{1/2} W_{g_j \cdot f}^{(2)}(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{W}_f^{(1)}(1) &= \left[I - \tilde{C}_\infty \tilde{C}_\infty^{-1/2} \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha} \cdot \tilde{C}_\infty^{-1/2} \right] W_f(1) \\
&= \tilde{\Pi}_\alpha \left(\tilde{\Pi}'_\alpha \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha \right)^{-1} \tilde{\Pi}'_\alpha \tilde{C}_\infty^{-1} W_f(1)
\end{aligned}
$$

and

$$
\tilde{W}_{g_j \cdot f}^{(2)}(1) = W_{g_j \cdot f}(1) - \left(\frac{1}{G} \sum_{\ell=1}^{G} \eta_{g_j \cdot f,\ell} \eta'_{f,\ell} \right) \tilde{C}_\infty^{-1/2} \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha} \cdot \tilde{C}_\infty^{-1/2} W_f(1),
$$

we have

$$
\tilde{D}_{\infty,\beta,j} = V_{ff}^{-1/2} V_{g_{d_\alpha+j} f} V_{ff}^{-1/2} W_f^{(1)}(1) + V_{ff}^{-1/2} V_{g_{d_\alpha+j} \cdot f}^{1/2} W_{g_{d_\alpha+j} \cdot f}^{(2)}(1).
$$

Since $\tilde{D}_{\infty,\beta,j}$ has two components, this case is more complicated than what is considered in Section 4.2 where the counterpart $\tilde{D}_{\infty,j}$ has only one component. It is more difficult to show that a rescaled version of $\tilde{D}_{\infty,\beta,j}$ is independent of $W_f(1)$ and $\tilde{C}_\infty^{-1}$.

The presence of the term containing $W_f^{(1)}(1)$ warrants the use of the SVD of $\tilde{\Pi}_\alpha$. Denote the SVD by $U_\alpha \Lambda_\alpha S_\alpha$, where

$$
\Lambda_\alpha = \begin{pmatrix} A_\alpha \\ O_\alpha \end{pmatrix}, \; U_\alpha = (U_{\alpha 1}, U_{\alpha 2})
$$

with $A_\alpha \in \mathbb{R}^{d_\alpha \times d_\alpha}, O_\alpha \in \mathbb{R}^{(m-d_\alpha) \times d_\alpha}, U_{\alpha 1} \in \mathbb{R}^{m \times d_\alpha}$, $U_{\alpha 2} \in \mathbb{R}^{m \times (m-d_\alpha)}$ and $S_\alpha \in \mathbb{R}^{d_\alpha \times d_\alpha}$. As in any SVD, A_α is a diagonal matrix, O_α is a matrix of zeros, and U_α and S_α are orthogonal matrices. We also need to partition $\tilde{C}_\infty$, $W_f(1)$, and $\eta_{f,\ell}$ differently to deal with the effect from estimating α_0. We now write

$$
\tilde{C}_\infty = \begin{pmatrix} \tilde{C}_{\alpha\alpha}, & \tilde{C}_{\alpha\tilde{q}} \\ \tilde{C}_{\tilde{q}\alpha}, & \tilde{C}_{\tilde{q}\tilde{q}} \end{pmatrix}, \; \tilde{C}_\infty^{-1} = \begin{pmatrix} \tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1}, & -\tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1} \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \\ -\tilde{C}_{\tilde{q}\tilde{q}\cdot\alpha}^{-1} \tilde{C}_{\tilde{q}\alpha} \tilde{C}_{\alpha\alpha}^{-1}, & \tilde{C}_{\tilde{q}\tilde{q}\cdot\alpha}^{-1} \end{pmatrix}
$$

and

$$
W_f(1) = \begin{pmatrix} W_{f,\alpha}(1) \\ W_{f,\tilde{q}}(1) \end{pmatrix}, \; \eta_{f,\ell} = \begin{pmatrix} \eta_{f,\alpha,\ell} \\ \eta_{f,\tilde{q},\ell} \end{pmatrix},
$$

where $\tilde{C}_{\alpha\alpha} \in \mathbb{R}^{d_\alpha \times d_\alpha}, \tilde{C}_{\tilde{q}\tilde{q}} \in \mathbb{R}^{\tilde{q}\times\tilde{q}}, W_{f,\alpha}(1) \in \mathbb{R}^{d_\alpha \times 1}, W_{f,\tilde{q}}(1) \in \mathbb{R}^{\tilde{q}\times 1}$, $\eta_{f,\alpha,\ell} \in \mathbb{R}^{d_\alpha \times 1}$, $\eta_{f,\tilde{q},\ell} \in \mathbb{R}^{\tilde{q}\times 1}$ for $\tilde{q} = q + d_\beta = m - d_\alpha$.

The lemma below provides the distributional representations of the quantities appearing in $\tilde{D}_{\infty,\beta}$ and $\mathcal{K}_{\infty,\beta}$.

Lemma 5.2 *The following distributional representations hold jointly:*
(a) $\tilde{W}_f^{(1)}(1) =^d U_{\alpha 1}\left[W_{f,\alpha}(1) - \tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right]$;
(b) $\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}W_f(1) =^d U_{\alpha 2}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)$;
(c) $\tilde{W}_{g_j\cdot f}^{(2)}(1) =^d U_\alpha\left[W_{g_j\cdot f}(1) - \left(\frac{1}{G}\sum_{\ell=1}^G \eta_{g_j\cdot f,\ell}\eta_{f,\tilde{q},\ell}'\right)\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right]$;
(d) $\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2} =^d U_{\alpha 2}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}U_{\alpha 2}'$.

With some abuse of the notation, we identify the random variables in Lemma 5.2 with their distributional representations in the rest of the paper.

Let $\tilde{\mathcal{J}}_\infty = W_{f,\tilde{q}}(1)'\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)$. Using Lemma 5.2, we can show that conditioning on $W_{f,\tilde{q}}(1)$ and $\{\eta_{f,\tilde{q},\ell}, \ell = 1, 2, \ldots, G\}$, $U_{\alpha 1}'W_f^{(1)}(1)$ and $U_\alpha'\tilde{W}_{g_j\cdot f}^{(2)}(1)$ are normal with the conditional variances $(1+G^{-1}\tilde{\mathcal{J}}_\infty)I_{d_\alpha}$ and $(1+G^{-1}\tilde{\mathcal{J}}_\infty)I_m$, respectively. It is unsurprising that both $U_{\alpha 1}'W_f^{(1)}(1)$ and $U_\alpha'\tilde{W}_{g_j\cdot f}^{(2)}(1)$ have scalar conditional variance matrices. What is surprising is that the underlying scalars are the same. This inspires us to rescale $\tilde{D}_{\infty,\beta,j}$ by the square root of $1 + G^{-1}\tilde{\mathcal{J}}_\infty$, leading to the following definition:

$$
\begin{aligned}
\tilde{D}_{\infty,\beta,j}^* &= \frac{\tilde{D}_{\infty,\beta,j}}{\sqrt{1+G^{-1}\tilde{\mathcal{J}}_\infty}} \\
&= V_{ff}^{-1/2}V_{g_{d_\alpha+j}\cdot f}V_{ff}^{-1/2}\frac{W_f^{(1)}(1)}{\sqrt{1+G^{-1}\tilde{\mathcal{J}}_\infty}} + V_{ff}^{-1/2}V_{g_{d_\alpha+j}\cdot f}^{1/2}\frac{W_{g_{d_\alpha+j}\cdot f}^{(2)}(1)}{\sqrt{1+G^{-1}\tilde{\mathcal{J}}_\infty}}.
\end{aligned}
$$

Let

$$
\tilde{D}_{\infty,\beta}^* = \left(\tilde{D}_{\infty,\beta,1}^*, \ldots, \tilde{D}_{\infty,\beta,d_\beta}^*\right).
$$

Using Lemma 5.2, we have

$$
\begin{aligned}
\mathcal{K}_{\infty,\beta} =^d & \left\{\left(\tilde{D}_{\infty,\beta}^*\right)' U_{\alpha 2}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right\}' \\
& \times\left\{\left(\tilde{D}_{\infty,\beta}^*\right)' U_{\alpha 2}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}U_{\alpha 2}'\tilde{D}_{\infty,\beta}^*\right\}^{-1} \times \left\{\left(\tilde{D}_{\infty,\beta}^*\right)' U_{\alpha 2}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right\}.
\end{aligned}
$$

By establishing that $\{\tilde{D}_{\infty,\beta,j}^*\}$ are independent of $\tilde{C}_{\tilde{q}\tilde{q}}$ and $W_{f,\tilde{q}}(1)$, we can prove the theorem below.

Theorem 5.2 *Let Assumptions 3.2, 3.3, 5.1, and 5.2 hold. If $\kappa = 1/2$, $\Pi_\beta = 0$, then*

$$
\left[\mathcal{K}_T^*(\hat{\theta}_0), \mathcal{J}_T^*(\hat{\theta}_0)\right] \Rightarrow \left[F_{d_\beta, G-d_\beta-q+1}, F_{q,G-q+1}\right],
$$

where $F_{d,G-d_\beta-q+1}$ and $F_{q,G-q+1}$ are independent F variates.

We have thus far considered cases parallel to those in Sections 4.1 and 4.2. Up to the adjustment of the degrees of freedom, the asymptotic distributions of the modified K and J statistics are the same as before.

For the intermediate case in which $\kappa = 1/2$ and $\Pi_\beta \neq 0$, pivotal inference based on $[\mathcal{K}_T^*(\hat{\theta}_0), \mathcal{J}_T^*(\hat{\theta}_0)]$ is not possible under the fixed-smoothing asymptotics, as asymptotically pivotal inference is not possible even in the simpler case of testing the whole vector θ_0. Nevertheless, we can argue as before that we can still use the F approximations for $\mathcal{K}_T^*(\hat{\theta}_0)$ and $\mathcal{J}_T^*(\hat{\theta}_0)$ given in Theorems 5.1 and 5.2, as they are asymptotically valid under the increasing-smoothing asymptotics wherein $G \to \infty$, $G/T \to 0$, as $T \to \infty$.

6 Asymptotic F Theory for the Modified S Statistic

To establish the asymptotic F theory for the modified S statistic, we maintain the following assumption, which is part of Assumption 3.2.

Assumption 6.1 $T^{-1/2} \sum_{t=1}^{[Tr]} [f(Y_t, \theta_0)] \Rightarrow B_f(r)$.

Theorem 6.1 *Let Assumptions 3.3 and 6.1 hold. We have*

$$\mathcal{S}_T^*(\theta_0) := \frac{G-m+1}{Gm} \mathcal{S}_T(\theta_0) \Rightarrow F_{m,G-m+1}.$$

In addition, let Assumptions 5.1(a) and 5.2(a.i) hold and assume that $\hat{V}_{g_j f}(\theta_0) = o_p(\sqrt{T})$ *for* $j = 1, \ldots, d_\alpha$. *We have*

$$\mathcal{S}_T^*(\hat{\theta}_0) := \frac{G-d_\beta-q+1}{G(d_\beta+q)} \mathcal{S}_T(\hat{\theta}_0) \Rightarrow F_{d_\beta+q, G-d_\beta-q+1}.$$

The assumption that $\hat{V}_{g_j f}(\theta_0) = o_p(\sqrt{T})$ for $j = 1, \ldots, d_\alpha$ holds if the FCLT for the Jacobian process holds as in Assumption 3.2. Here, we do not need the asymptotic distribution of $\hat{V}_{g_j f}(\theta_0)$ and only impose a mild rate condition on it.

The assumptions for the asymptotic F theory for the modified S statistic are much weaker than those for the K and J statistics. The reason is that the S statistic does not involve the Jacobian with respect to the parameter subvector fully specified under the null. The drawback of the F test based on the modified S statistic is that the first degree of freedom in the approximating F distribution is larger than that for the modified K statistic, leading to potential power loss.

The asymptotic F theory for the modified S statistic holds regardless of the strength of the identification of the parameters pinned down by the null hypothesis. This contrasts with the K and J statistics, where the asymptotic F theory cannot be theoretically established in the intermediate identification case.

7 J-K* Test: Improving the Power of the Modified K Test

The K test can suffer severe power loss for some parameter values under the alternatives. In this section, we design a new test to improve the power of the K test when it is low.

We first consider testing the full parameter vector. Recall that the K statistic is based on the first-order derivative of the CU-GMM objective function. Suppose that the first-order derivative is zero at some $\theta^* \neq \theta_0$, where θ_0 is the true parameter value of θ, that is,

$$\frac{\partial Q_T(\theta^*)}{\partial \theta} = D_T(\theta^*)' \hat{V}_{ff}^{-1}(\theta^*) \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(Y_t, \theta^*) \right] = 0.$$

Furthermore, suppose that θ^* is not a minimizer of the CU-GMM objective function $Q_T(\theta)$. Such a point exists because in general $Q_T(\theta)$ is not a strictly convex function for any sample. For such a choice of θ^*, the K statistic $\mathcal{K}_T(\theta^*)$ is zero by construction. As a result, we fail to reject the null hypothesis of $\theta = \theta^*$, and we commit a type II error. No such θ^* works for all samples. However, there may exist a parameter $\theta^* \neq \theta_0$ such that $\mathcal{K}_T(\theta^*)$ is small with a high probability and $Q_T(\theta^*)$ is substantially different from the minimal value of $Q_T(\theta)$, namely $\min_{\theta\in\Theta} Q_T(\theta)$. The K test then has low power against θ^*.

To fix the problem, we can employ a test to screen out such θ^*. A defining feature of θ^* is that $2Q_T(\theta^*) - 2\min_{\theta\in\Theta} Q_T(\theta)$ is large. Note that $\min_{\theta\in\Theta} Q_T(\theta)$ does not depend on θ^* and $2Q_T(\theta^*) = \mathcal{K}_T(\theta^*) + \mathcal{J}_T(\theta^*)$. Hence, for a given value of $\mathcal{K}_T(\theta^*)$, if $\mathcal{J}_T(\theta^*)$ is sufficiently large, then we should reject the null $H_0 : \theta = \theta^*$. If $\mathcal{J}_T(\theta^*)$ is not large, then we proceed to use the K statistic $\mathcal{K}_T(\theta^*)$ to decide on whether H_0 is true. This is the J-K test of Kleibergen (2005).

We can also modify the J-K test to obtain a J-K* test based on the $\mathcal{J}_T^*$ and $\mathcal{K}_T^*$ statistics. The J-K* test rejects the null of $H_0 : \theta = \theta_0$ if

$$\mathcal{J}_T^*(\theta_0) \geq F_{q,G-q+1}^{\alpha_J}$$

or

$$\mathcal{J}_T^*(\theta_0) < F_{q,G-q+1}^{\alpha_J} \text{ and } \mathcal{K}_T^*(\theta_0) \geq F_{d,G-m+1}^{\alpha_K},$$

where $F_{q,G-q+1}^{\alpha_J}$ is the $1-\alpha_J$ quantile of the F distribution $F_{q,G-q+1}$ and $F_{d,G-m+1}^{\alpha_K}$ is $1-\alpha_K$ quantile of the F distribution $F_{d,G-m+1}$.

In the cases with possibly weak identification and complete identification failure, $\mathcal{J}_T^*(\theta_0)$ and $\mathcal{K}_T^*(\theta_0)$ are asymptotically independent. Hence, the asymptotic null rejection probability of the J-K* test is

$$\alpha_J + (1-\alpha_J)\alpha_K = \alpha_J + \alpha_K - \alpha_J\alpha_K.$$

To obtain a level-α J-K* test, we can take

$$\alpha_K = \frac{\alpha - \alpha_J}{1-\alpha_J}$$

for a given α_J. For example, if we set $\alpha = 5\%$ and $\alpha_J = 1\%$, then we take

$$\alpha_K = \frac{0.05 - 0.01}{1 - 0.01} = 4.04\%.$$

We use these choices of α_J and α_K in our simulation study.

For the intermediate case when $\kappa = 1/2$ and $\Pi \neq 0$, the above asymptotic level calculation cannot be rigorously justified under the fixed-smoothing asymptotics, but it is still asymptotically valid under the increasing-smoothing asymptotics.

To test a subvector of the model parameters $H_0^* : \beta = \beta_0$, we can design the J-K* test in the same way but with θ_0 replaced by $\hat{\theta}_0$. The subvector J-K* test rejects if

$$\mathcal{J}_T^*(\hat{\theta}_0) \geq F_{q,G-q+1}^{\alpha_J}$$

or

$$\mathcal{J}_T^*(\hat{\theta}_0) < F_{q,G-q+1}^{\alpha_J} \text{ and } \mathcal{K}_T^*(\hat{\theta}_0) \geq F_{d_\beta,G-d_\beta-q+1}^{\alpha_K}.$$

The asymptotic validity of the test can be justified in the same way.

8 Monte Carlo Simulations and Empirical Application

In this section, we investigate the finite sample performances of the proposed tests in a linear IV regression model and a stochastic discount factor (SDF) model. While the SDF model is calibrated to an empirical dataset, the IV regression model is not. The reason for studying the IV regression model is that we have the flexibility to vary the key parameters, and so we can examine their effects on the test performances.

8.1 Linear IV regression

For the linear IV regression, the data-generating process is

$$Y_t = X_{1,t}\beta_1 + X_{2,t}\beta_2 + \varepsilon_{y,t},$$

where $X_{1,t}$ and $X_{2,t}$ are correlated with $\varepsilon_{y,t}$. The instruments are $\{Z_{1,t}, Z_{2,t}, Z_{3,t}\}$, which are uncorrelated with $\varepsilon_{y,t}$, and

$$X_{1,t} = Z_{1,t}\beta_{xz} + Z_{3,t}\beta_{xz} + \varepsilon_{x_1,t}$$
$$X_{2,t} = Z_{2,t}\beta_{xz} + Z_{3,t}\beta_{xz} + \varepsilon_{x_2,t}.$$

The error terms $[\varepsilon_{y,t}, \varepsilon_{x_1,t}, \varepsilon_{x_2,t}]$ follow an $AR(1)$ process:

$$\varepsilon_{y,t} = \phi_\varepsilon \varepsilon_{y,t-1} + (1-\phi_\varepsilon^2)^{1/2} e_{y,t}$$
$$\varepsilon_{x_1,t} = \phi_\varepsilon \varepsilon_{x_1,t-1} + (1-\phi_\varepsilon^2)^{1/2} e_{x_1,t}$$
$$\varepsilon_{x_2,t} = \phi_\varepsilon \varepsilon_{x_2,t-1} + (1-\phi_\varepsilon^2)^{1/2} e_{x_2,t}$$

with $[e_{y,t}, e_{x_1,t}, e_{x_2,t}] \sim N(0, V_e)$. The variance-covariance matrix V_e is

$$V_e = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}.$$

The instruments also follow an $AR(1)$ process:

$$Z_{j,t} = \phi_z Z_{j,t-1} + (1-\phi_z^2)^{1/2} e_{z_j,t}$$

for $j = 1, 2, 3$ with $[e_{z_1,t}, e_{z_2,t}, e_{z_3,t}] \sim N(0, V_e)$. Finally, $[e_{y,t}, e_{x_1,t}, e_{x_2,t}]$ is independent of $[e_{z_1,t}, e_{z_2,t}, e_{z_3,t}]$.

In the simulations, we set $\phi_z = \phi_\varepsilon = \phi$. We target a specified value of the R^2 of the first-stage regression (a regression of X_t on $Z_{1,t}$ and $Z_{2,t}$). Using some simple calculations, the population R^2 of the first-stage regression is

$$R^2 = \frac{2\beta_{xz}^2(1+\rho)}{2\beta_{xz}^2\,(1+\rho)+1}.$$

From this, we back out β_{xz}^2 as

$$\beta_{xz}^2 = \frac{R^2}{2(1+\rho)(1-R^2)}.$$

With this choice, the correlation coefficient between X_{it} and $\varepsilon_{y,t}$ is

$$\text{corr}(X_{i,t}, \varepsilon_{y,t}) = \rho\left(1 - R^2\right) \text{ for } i = 1 \text{ and } 2.$$

There are three key parameters in the model: ϕ, ρ, and R^2. While ϕ measures the degree of persistence, ρ measures the degree of endogeneity and R^2 measures the strength of the instruments. We consider the following parameter configurations: $\phi = 0, 0.6, 0.9$, $\rho = 0.3$, and 0.9, and R^2 ranges from 0 to 0.2 in increments of 0.025.

The null and alternative hypotheses of interest are $H_0 : \beta_1 = \beta_2 = 0$ and $H_1 : \beta_1 \neq 0$ or $\beta_2 \neq 0$. We consider two groups of tests. The first group consists of the two-step Wald test, the two-step LM test, the CU Wald test, and their modified versions[6]. The two-step tests in the first group employ the standard IV estimator as the initial first-step estimator. The CU Wald test is based on the Wald statistic evaluated at the CU estimator. See Sun (2014b) and Hwang and Sun (2017) for more details on the two-step tests. The first group of tests is not robust to weak identification. The second group of tests consists of the K, J, S, and J-K tests and their modified versions developed in this study[7]. For the OS LRV estimators, we use the Fourier basis functions and select the number of basis functions G using the AMSE criterion in Phillips (2005).

Figure 1 reports the empirical sizes of the nominal 5% tests when $\phi = 0$ and $\rho = 0.9$. The average G is around 18 for every level of R^2. Several patterns emerge. First, the chi-squared tests including the Wald, LM, Wald CUE, K, J, S, and J-K tests all have a large size distortion. For the Wald, LM, and Wald CUE tests, the size distortion can be severe. Second, when the instruments are weak (i.e., when R^2 is small), the Wald, LM, and Wald CUE tests have larger size distortion than the K, J, S, and J-K tests, as shown by the scale difference between the top and bottom panels of Figure 1. In addition, the size of the Wald, LM, and Wald CUE tests improves with the strength of the identification, while the size of the K, J, S, and J-K tests does not. These results are consistent with the theoretical prediction: while the K, J, S, and J-K tests are robust to weak identification, the Wald, LM, and Wald CUE tests are not. Third, a chi-squared test has larger size distortion than the corresponding F test presented in this paper or that in Hwang and Sun (2017). In fact, the K*, J*, S*, and J-K* tests, all of which are F-type tests, have an accurate size. For the F tests based on the modified Wald, LM, and Wald CUE statistics, the size distortion diminishes to zero as the identification becomes stronger. We can thus conclude that the asymptotic F approximations are much more accurate than the asymptotic chi-squared approximations.

Figure 2 reports the empirical size of the nominal 5% tests when $\phi = 0.9$ and $\rho = 0.9$. The configurations are the same as in Figure 1, but the processes are now more persistent. The average G is around 6 for every level of R^2. The product of two independent AR(1) processes with the same AR parameter ϕ is an AR(1) process with the AR parameter ϕ^2. Hence, the moment process under consideration follows AR(1) with the AR parameter $0.9^2 = 0.81$, which is high but not empirically implausible. The qualitative observations we made for Figure 1 continue to apply. The difference is that the chi-squared tests are more size distorted than before. This is expected, as the LRV estimators have a large downward bias when the moment process is

[6] For brevity, we refer to these three tests as Wald, Wald CUE, and LM, respectively. Their modified versions are referred to as Wald*, LM*, and Wald CUE*, respectively.

[7] We refer to the modified versions as K*, J*, S*, and J-K*, respectively.

positively autocorrelated. Regarding the F tests, the S* and J-K* tests now slightly under-reject, while the J* test continues to over-reject and the K* test still has a size lower than 5%.

Figure 3 has the same parameter configuration as Figure 2 except that now $\rho = 0.3$, indicating less endogeneity. Here, the average G is 6 as well, although slightly smaller than in the previous case. The figure provides further evidence that an asymptotic F test is more accurate than the corresponding chi-squared test. Owing to the smaller degree of endogeneity, the Wald, LM, and Wald CUE tests do not suffer from such a large size distortion as in Figure 1 when the instruments are very weak.

Figure 4 reports the empirical sizes for $\phi = 0$ and $\rho = 0$ (i.e., no serial dependence and no endogeneity). Here, the average G is around 27. The size distortion of the Wald, LM, and Wald CUE tests changes with the strength of the identification. When the instruments are weak, the Wald, LM, and Wald CUE tests are under-sized. The F tests based on the modified Wald, LM, and Wald CUE statistics are even more under-sized. This is a scenario when the F tests are less accurate. When the instruments become stronger, the Wald, LM, and Wald CUE tests become over-sized, and the corresponding F tests have an improved size accuracy. Consistent with our theoretical prediction, the empirical size of the weak-identification-robust tests, namely the K, J, S, and J-K tests and their modified versions, is invariant to the strength of the identification.

To sum up, our simulation evidence lends strong support to the higher accuracy of the F approximations in most cases. In an overall sense, the K*, S*, and J-K* tests are the most accurate among all the tests considered.

To simulate the power of the tests under consideration, we generate the data under the local alternative so that

$$Y_t = (X_{1,t}, X_{2,t})\beta + \varepsilon_{y,t},$$

for

$$\beta = (\beta_1, \beta_2)' + \frac{\xi}{\|\xi\|} \cdot \frac{c}{\sqrt{T}}$$

where $\xi \sim N(0, I_2)$. We let ξ be different for different simulation replications. That is, we do not specify the direction of the local departure. Effectively, we simulate the average local power averaged over all directions uniformly.

Figures 5–9 report the power of all the tests under consideration. The parameter configurations are the same as in Figures 1–4. For the strength of the instruments, we consider an intermediate value of $R^2 = 0.5$. The reported power is size-adjusted, and thus the power comparison is meaningful. While such a size adjustment is not possible in practice, it is feasible in Monte Carlo experiments. Moreover, the size-adjusted power of the S test is the same as that of the S* test, as they are based on the same test statistic up to a multiplicative constant. Similarly, the J test and J* test have the same size-adjusted power. We can make a number of observations from these four power figures. First, the first group of tests is, in general, more powerful than the second group. However, Figure 5 shows that in the high endogeneity case with $\rho = 0.9$, the second group of tests is more powerful when the local departure from the null is not large. Second, among the second group of tests, the J test is in general less powerful. It is not a good idea to use the J test alone. Third, for the first group of tests, the F test based on the modified Wald, LM, or Wald CUE statistic is as powerful as the corresponding chi-squared test. For the second group of tests, the K* test and J-K* are less powerful than the corresponding chi-squared test. For the K and J-K tests, there is a cost to achieving size accuracy. Finally, the J-K test appears to be more powerful than the K test, and the J-K* test is more powerful than the K* test.

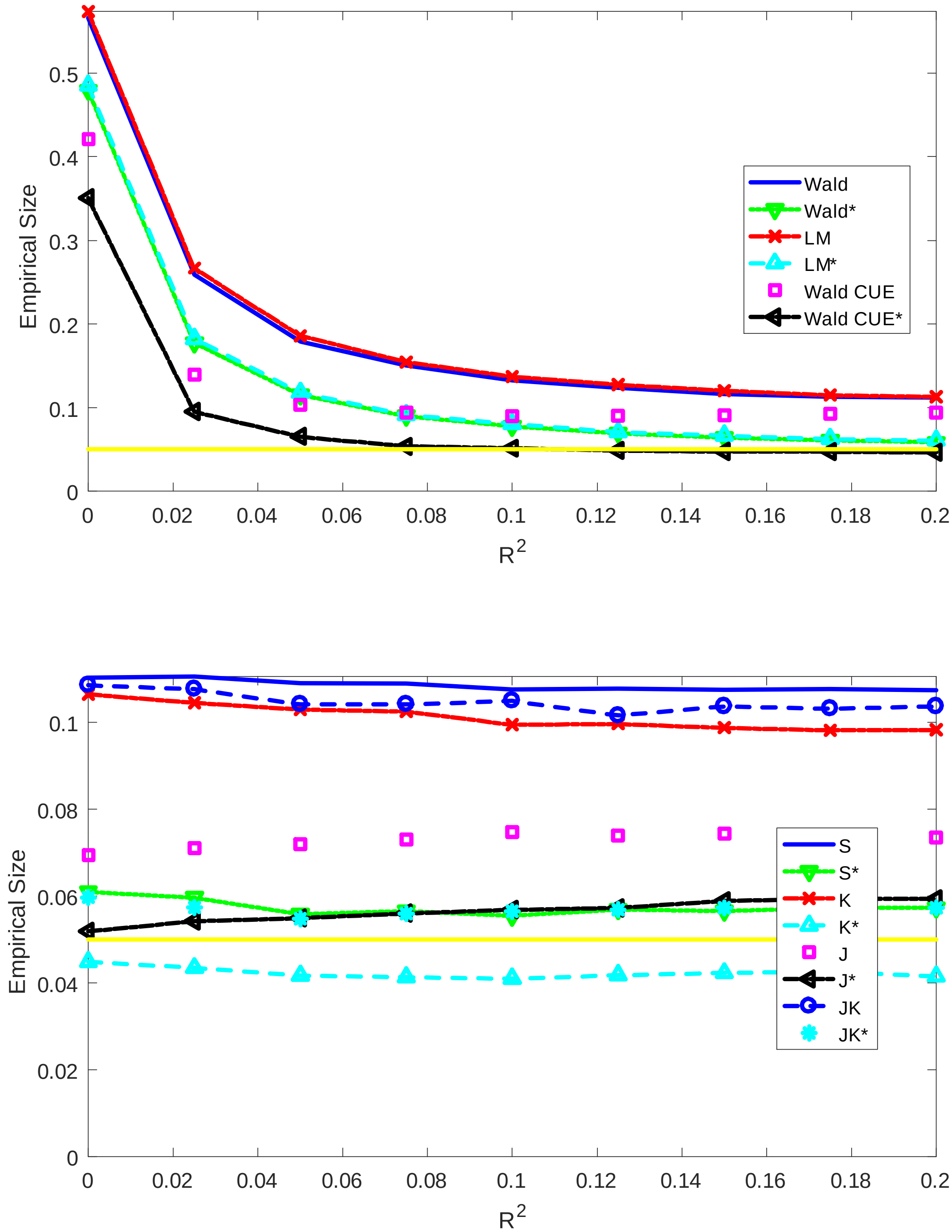


Figure 1: Empirical sizes of the 5% tests when $\phi = 0$ and $\rho = 0.9$.

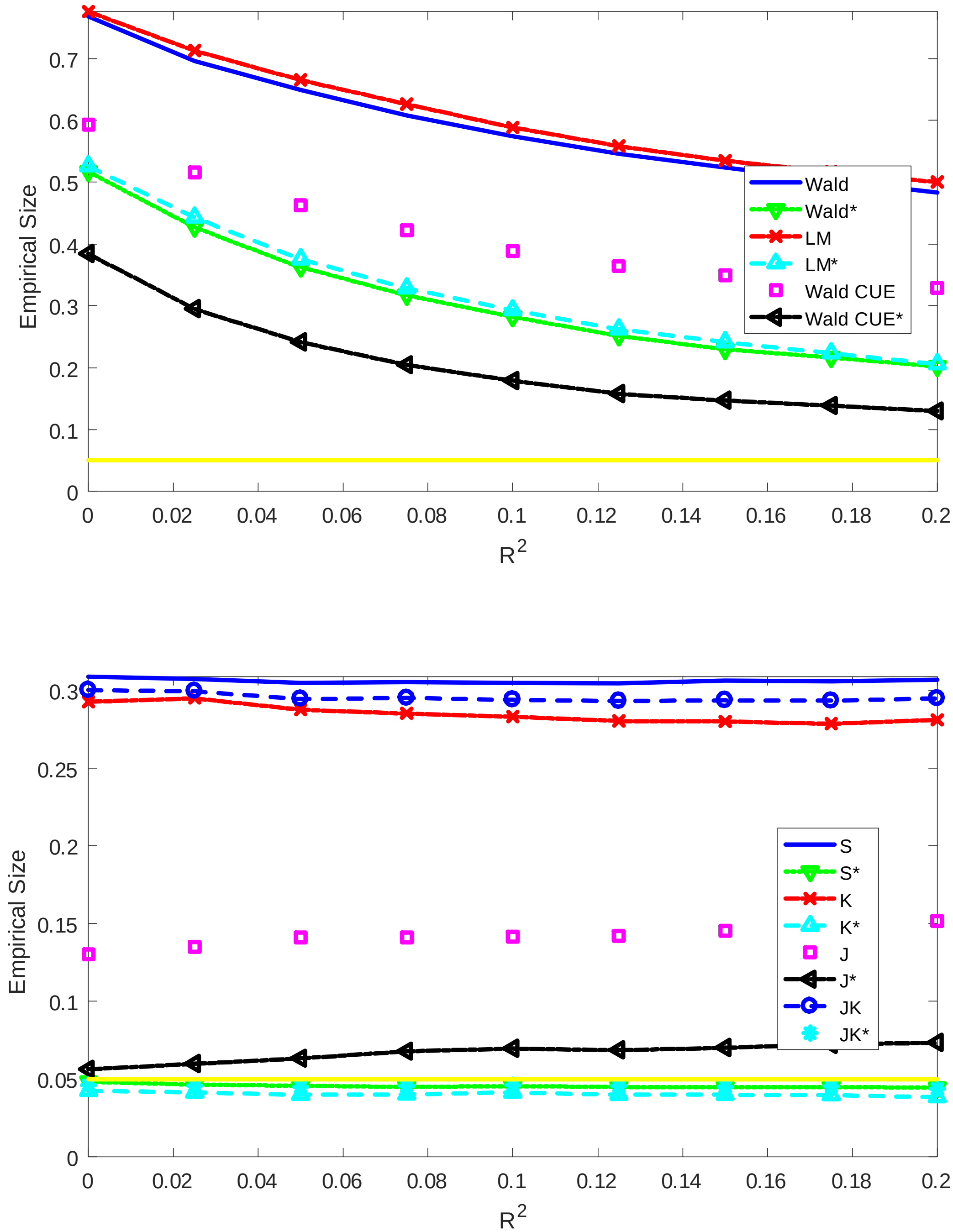


Figure 2: Empirical sizes of the 5% tests when $\phi = 0.9$ and $\rho = 0.9$.

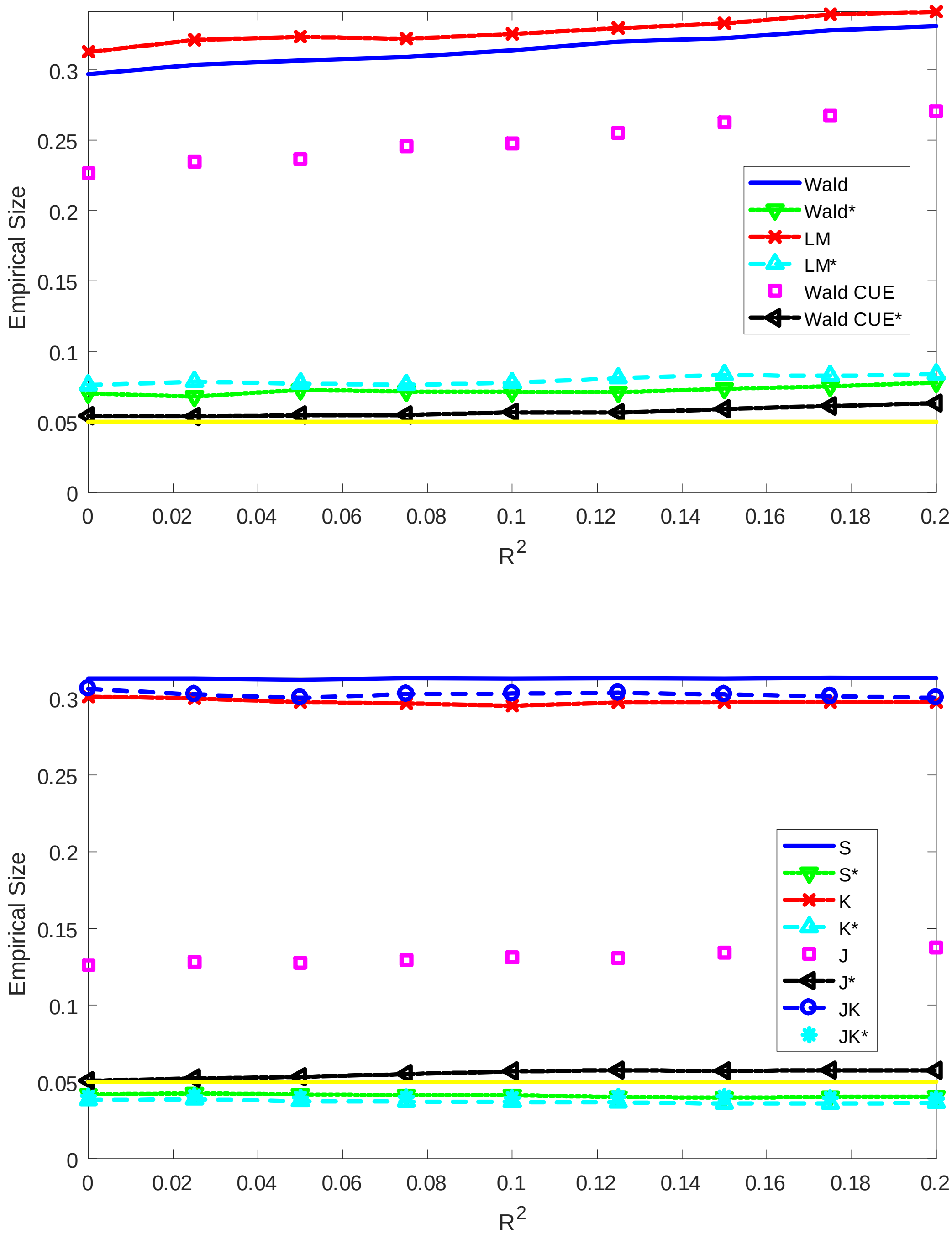


Figure 3: Empirical sizes of the 5% tests when $\phi = 0.9$ and $\rho = 0.3$.

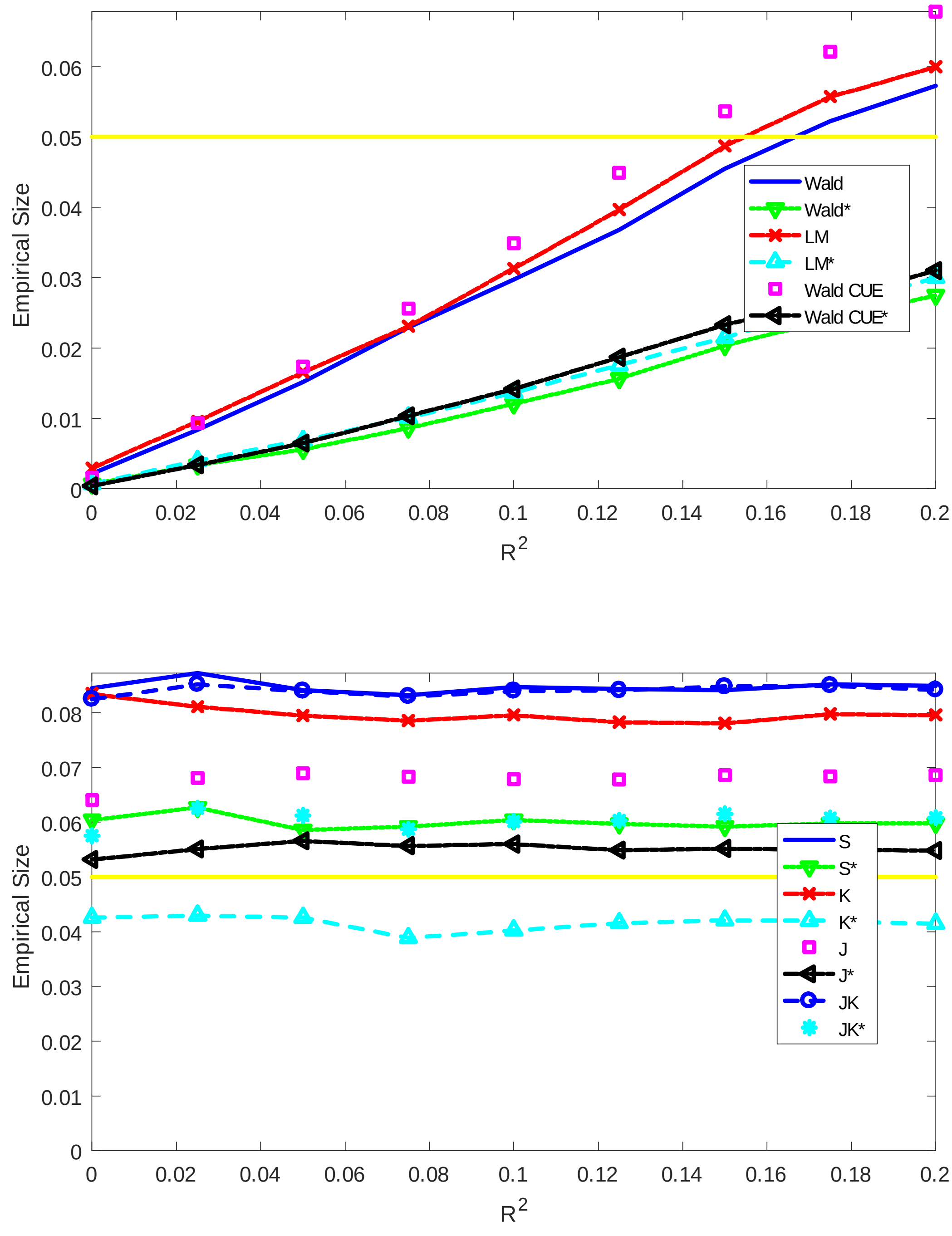


Figure 4: Empirical sizes of the 5% tests when $\phi = 0.0$ and $\rho = 0.0$.

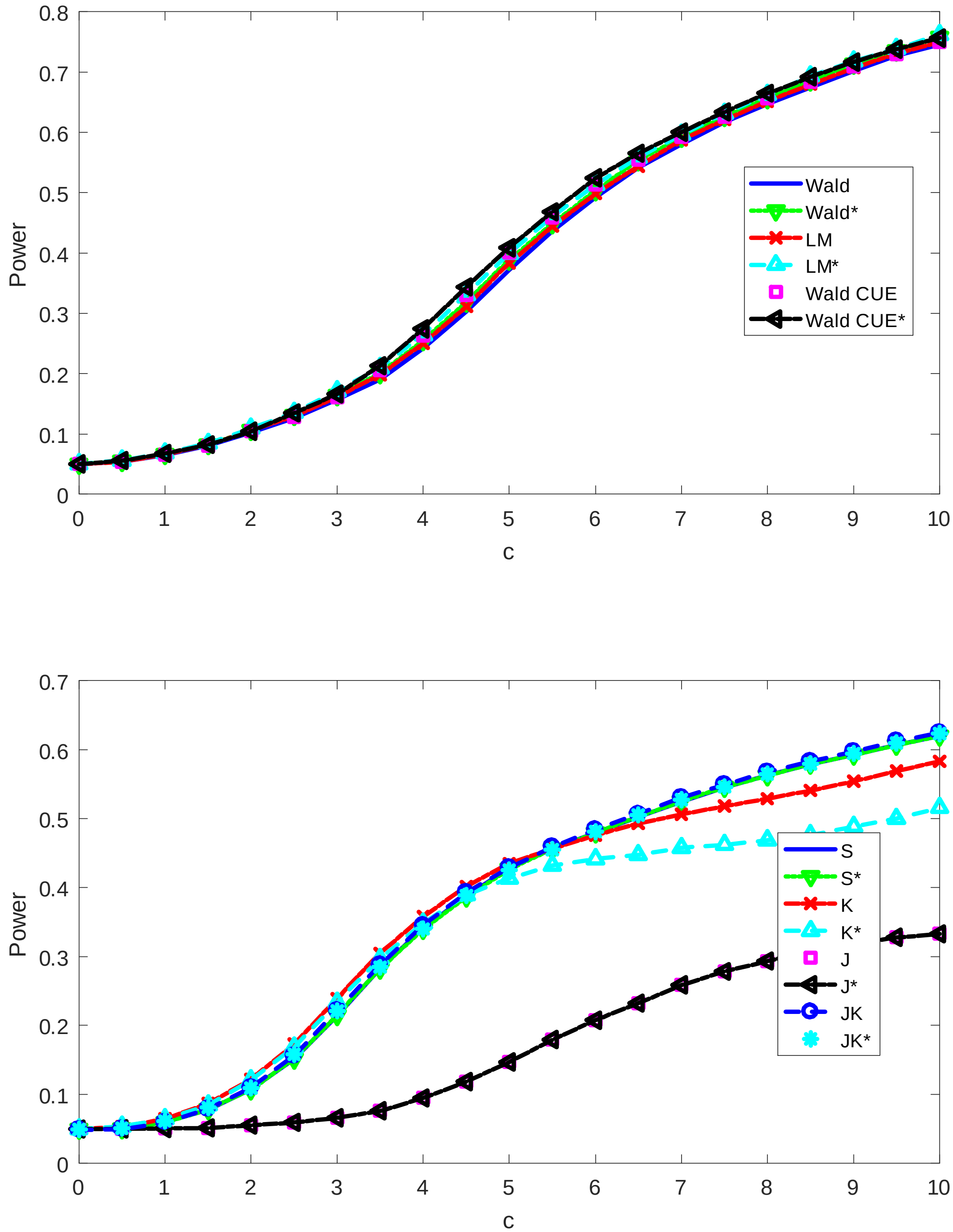


Figure 5: Size-adjusted power when $\phi = 0, \rho = 0.9$ and $R^2 = 0.5$.

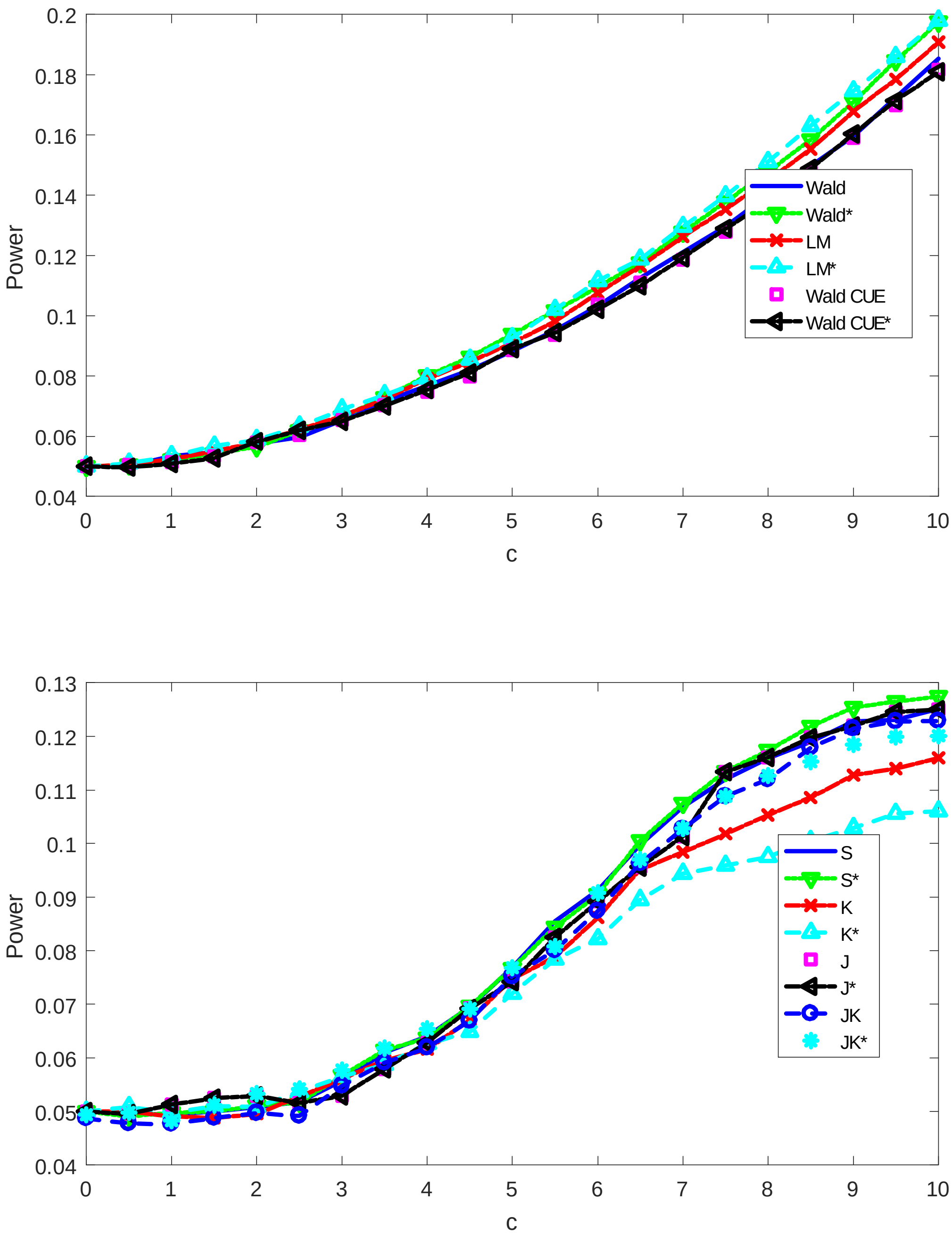


Figure 6: Size-adjusted power when $\phi = 0.9, \rho = 0.9$ and $R^2 = 0.5$.

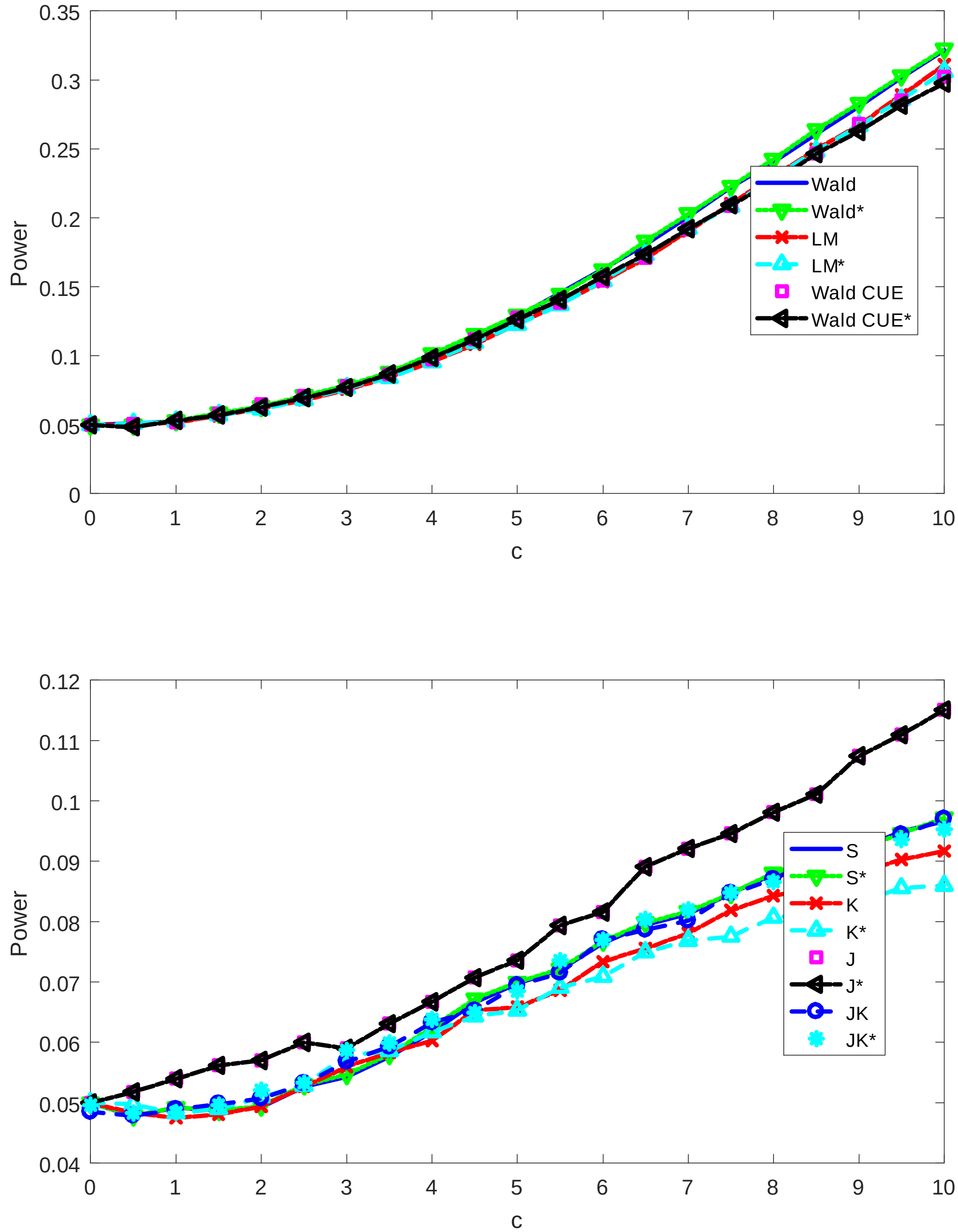
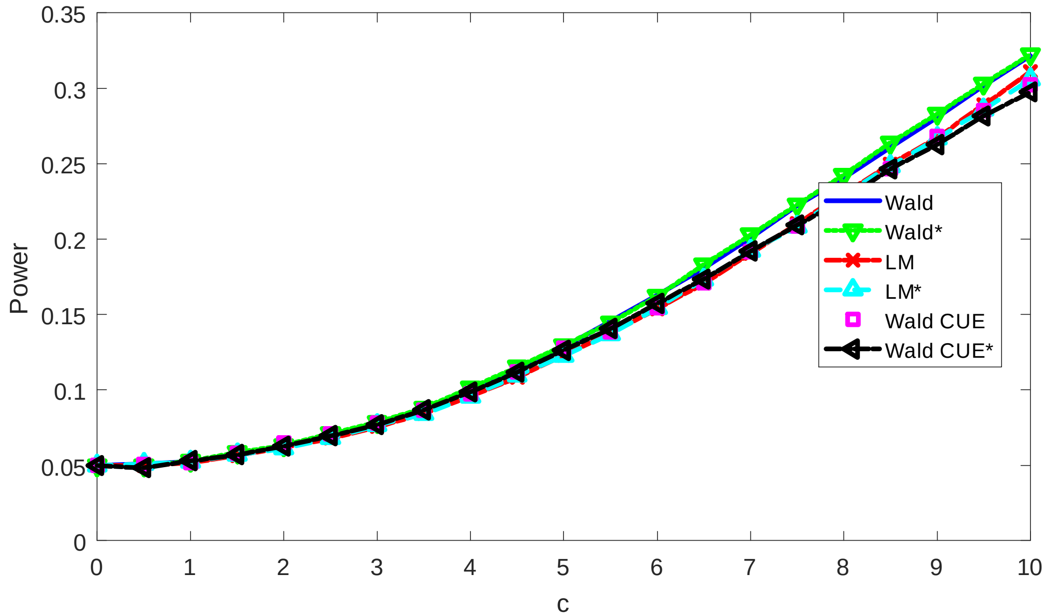


Figure 7: Size-adjusted power when $\phi = 0, \rho = 0.3$ and $R^2 = 0.5$.



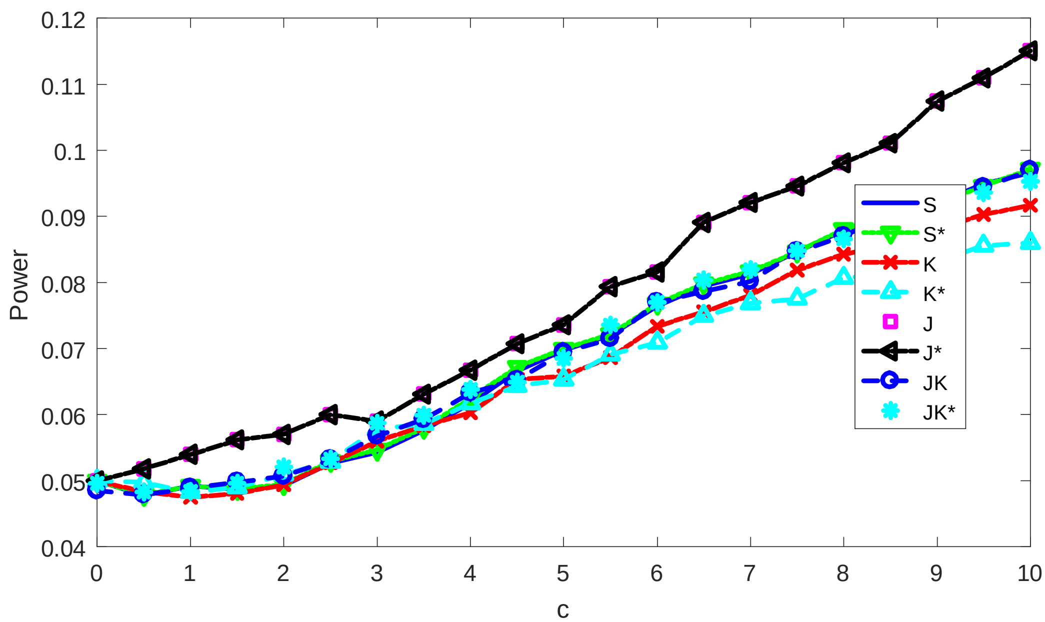


Figure 8: Size-adjusted power when $\phi = 0.9, \rho = 0.3$ and $R^2 = 0.5$.

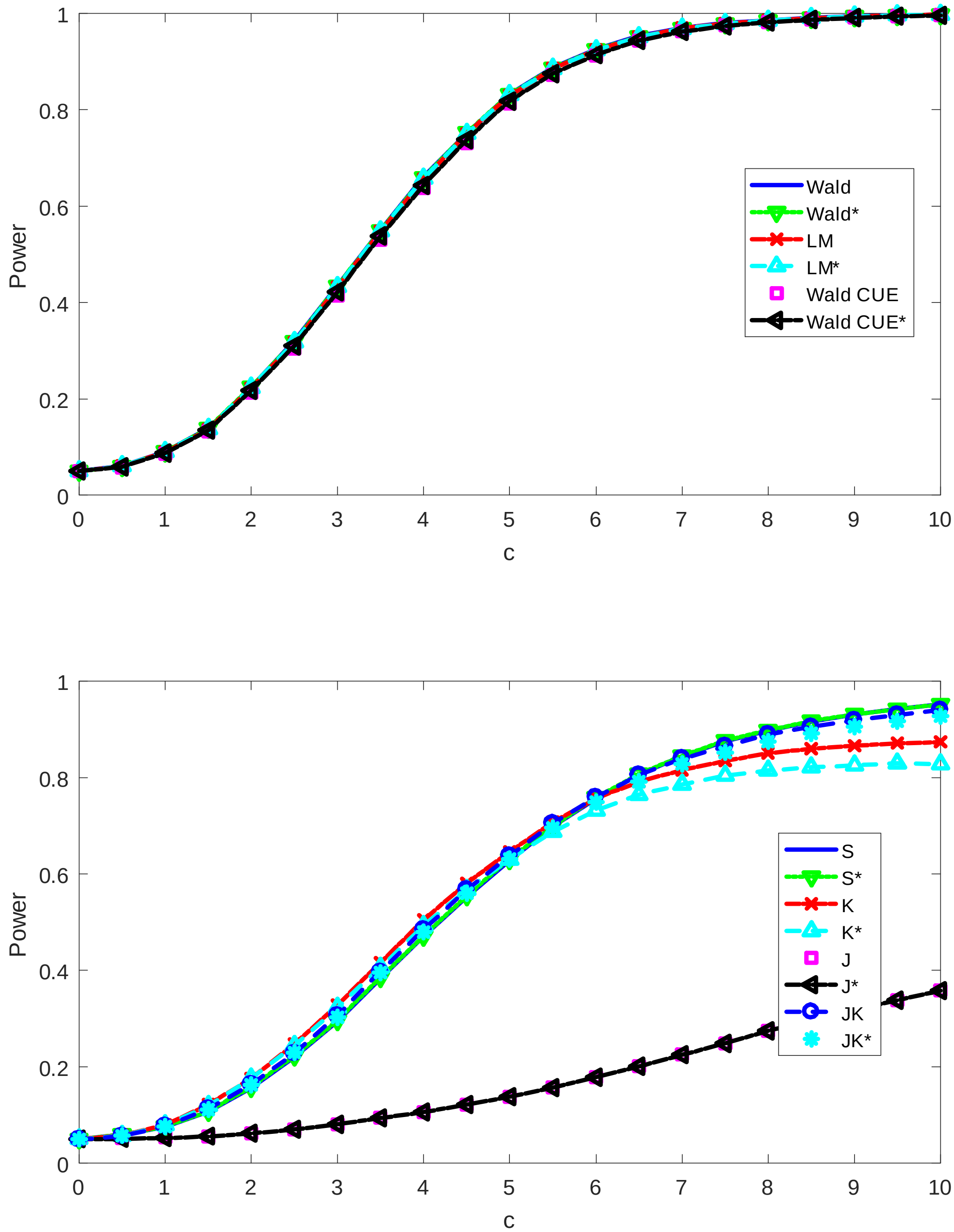


Figure 9: Size-adjusted power when $\phi = 0, \rho = 0$ and $R^2 = 0.5$

8.2 Stochastic discount factor model

To describe the data-generating process in this experiment design, we start with the asset pricing equation derived from the agent's intertemporal optimization problem with CRRA utility:

$$p_t = \frac{1}{\alpha_0}\mathbb{E}_t\left\{\left[\left(\frac{C_{t+1}}{C_t}\right)^{-\beta_0}(p_{t+1}+D_{t+1})\right]\right\},$$

where α_0^{-1} is the discount factor, β_0 is the risk aversion coefficient, p_t is the ex-dividend real price of the asset under consideration, D_t is the real dividend, and C_t is the real consumption. In the above, $\mathbb{E}_t(\cdot)$ is the conditional expectation conditioning on the information set at time t.

Let

$$d_t = \frac{D_t}{D_{t-1}}, c_t = \frac{C_t}{C_{t-1}},\ v_t = \frac{p_t}{D_t}.$$

Then, the pricing equation can be rewritten as

$$v_t = \frac{1}{\alpha_0}\mathbb{E}_t\left[(c_{t+1})^{-\beta_0}(1+v_{t+1})\,d_{t+1}\right]. \tag{16}$$

Following Kleibergen (2005), we assume that (c_t, d_t) evolves according to the following Gaussian VAR(1) process:

$$\begin{pmatrix} \log c_t \\ \log d_t \end{pmatrix} = \begin{pmatrix} 0.021 \\ 0.004 \end{pmatrix} + \begin{pmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{pmatrix}\begin{pmatrix} \log c_{t-1} \\ \log d_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{c,t} \\ \varepsilon_{d,t} \end{pmatrix}, \tag{17}$$

where $(\varepsilon_{c,t}, \varepsilon_{d,t})'$ are iid normally distributed with $\text{var}(\varepsilon_{d,t}) = 0.014$, $\text{var}(\varepsilon_{c,t}) = 0.0012$, and $\text{corr}(\varepsilon_{c,t}, \varepsilon_{d,t}) = 0.43$. The parameters in the Gaussian VAR(1) are calibrated to the log-growth rate of U.S. per capita real annual consumption and the log-growth rate of real annual dividends on the Standard & Poor's 500 (SP500). Kocherlakota (1990), Hansen, Heaton, and Yaron (1996), and Stock and Wright (2000) also use the above VAR(1) in their simulation studies.

To solve for v_t under the above Gaussian VAR(1) dynamics for the state variables c_t and d_t, we follow Tauchen (1986a, 1986b) and approximate the Gaussian VAR(1) using a discrete Markov chain with 10^2 states (10 states for each of the two variables c_t and d_t). Using the discrete Markov chain approximation, we can solve for v as a function of c and d. With some abuse of the notation, we write this function as $v(c, d; \alpha_0, \beta_0)$.

The data for our experiments are generated as follows. Given the model parameters α_0 and β_0, we generate c_t and d_t according to the Gaussian VAR(1) model in (17) and then compute $v_t = v(c_t, d_t; \alpha_0, \beta_0)$, which is the solution to the Euler equation in (16). With $\{c_t, d_t, v_t\}_{t=1}^T$, we compute the real returns of the asset

$$R_{t+1} := \frac{p_{t+1}+D_{t+1}}{p_t} - 1 = \frac{(1+v_{t+1})}{v_t}d_{t+1} - 1.$$

Our simulated data then consist of the vector time series $\{(c_t, d_t, R_t)\}_{t=1}^T$.

In terms of c_{t+1} and R_{t+1}, the Euler equation becomes

$$\mathbb{E}_t\left[(c_{t+1})^{-\beta_0}(1+R_{t+1}) - \alpha_0\right] = 0, t = 1, \ldots, T. \tag{18}$$

The above conditional moment restriction implies the following unconditional moment restrictions:

$$\mathbb{E}f(Y_t, \alpha_0, \beta_0) = 0, t = 1, \ldots, T, \tag{19}$$

where $Y_t = (c_{t+1}, R_{t+1}, Z_t)$,

$$f(Y_t, \alpha, \beta) = \left[(c_{t+1})^{-\beta}(1 + R_{t+1}) - \alpha\right] \otimes Z_t,$$

and the vector of instruments Z_t consists of a constant and the three lagged values of consumption growth and asset returns:

$$Z_t = (1,\ c_t, c_{t-1}, c_{t-2}, R_t, R_{t-1}, R_{t-2}) \in \mathbb{R}^7.$$

We are interested in testing $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$. All the tests under consideration are based on the GMM using (19) as the moment conditions.

In the simulation experiments, we set $\alpha_0 = 1/(0.97)$ and $\beta_0 = 1.3$. For the parameters in the Gaussian VAR(1) process, we also consider a model in which $\text{corr}(\varepsilon_c, \varepsilon_d) = 0.95$ and a model (strong identification) in which the intercept and VAR matrix are equal to 2 times those in (17). As in the linear IV simulation study, we use the Fourier basis functions and select the number of basis functions G using the AMSE criterion in Phillips (2005). The average G (Avg G) and the interquartile range (IQR G) are reported with the results. The sample size is 250 and the number of simulation replications is 10,000.

Table 1 reports the empirical size of the weak identification-robust tests, showing that the F tests have a more accurate size than the corresponding chi-squared tests. In addition, the size properties of all the tests are not affected by the identification strength and the correlation between ε_c and ε_d.

Table 1: Empirical sizes of the 5% tests

	Weaker ID	Weaker ID	Stronger ID	Stronger ID
	$\text{corr}(\varepsilon_c, \varepsilon_d) = .43$	$\text{corr}(\varepsilon_c, \varepsilon_d) = .95$	$\text{corr}(\varepsilon_c, \varepsilon_d) = .43$	$\text{corr}(\varepsilon_c, \varepsilon_d) = .95$
K	0.114	0.106	0.111	0.106
K*	0.051	0.047	0.049	0.049
J	0.121	0.115	0.124	0.112
J*	0.049	0.050	0.053	0.049
S	0.143	0.138	0.146	0.135
S*	0.049	0.047	0.052	0.046
J-K	0.136	0.127	0.134	0.129
J-K*	0.098	0.095	0.100	0.096
Avg G	32.674	33.870	33.004	34.182
IQR G	2	0	2	2

"Weaker ID" refers to the empirically calibrated VAR(1) model for $(\log(c_t), \log(d_t))$. "Stronger ID" refers to the VAR(1) model whose AR matrix is twice the AR matrix in the empirically calibrated VAR(1) model.

Figures 10–13 report the power functions for the tests. The general pattern that emerges is that the F tests have lower power, but the shape of their power functions closely matches that of the power functions of the original chi-squared tests. The S and J-K tests and their modified versions seem to have good power properties. However, as shown in Table 1, the J-K* test has a larger size distortion than the S* test. In summary, the F tests not only have a more accurate size, but also retain the power properties of the chi-squared tests.

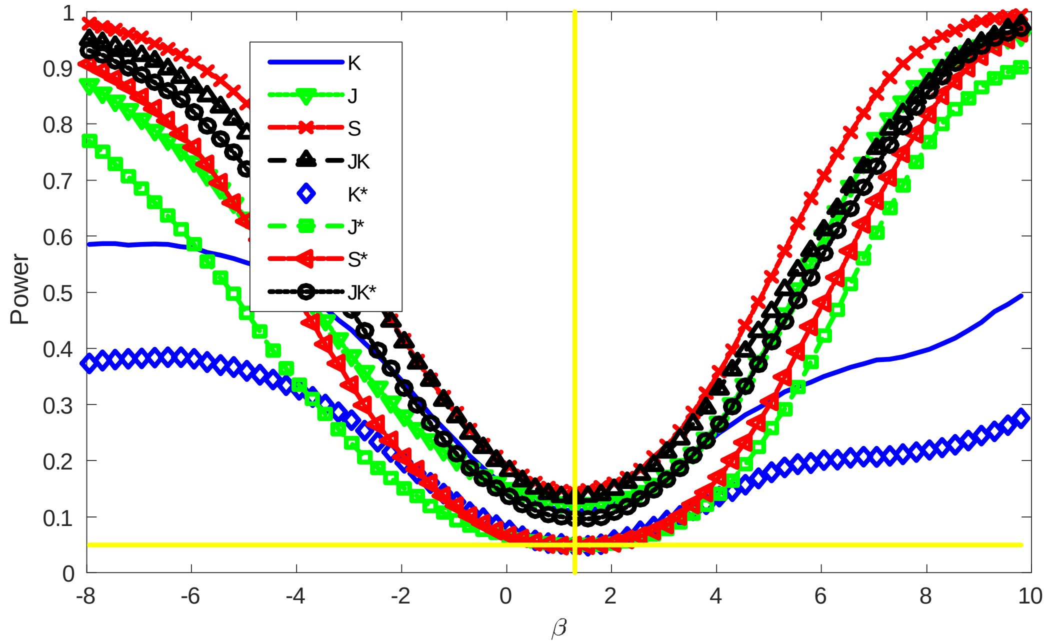


Figure 10: Power for the "Weaker ID" model, and $\text{corr}(\varepsilon_c, \varepsilon_d) = .43$

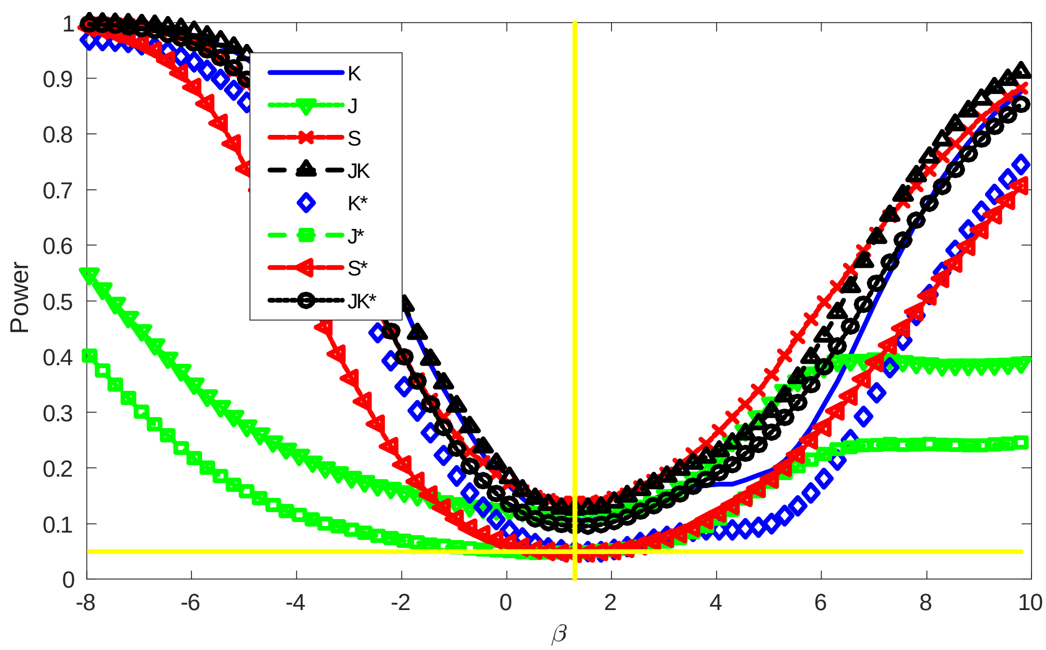


Figure 11: Power for the "Weaker ID" model, and $\text{corr}(\varepsilon_c, \varepsilon_d) = .95$

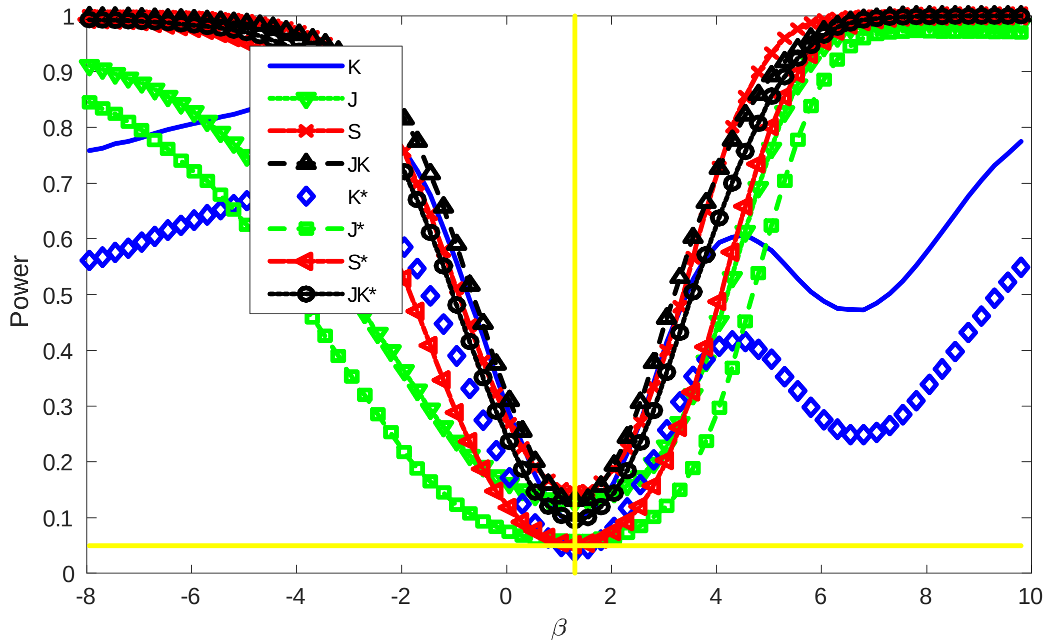


Figure 12: Power for the “Stronger ID” model, and $\mathrm{corr}(\varepsilon_c, \varepsilon_d) = .43$

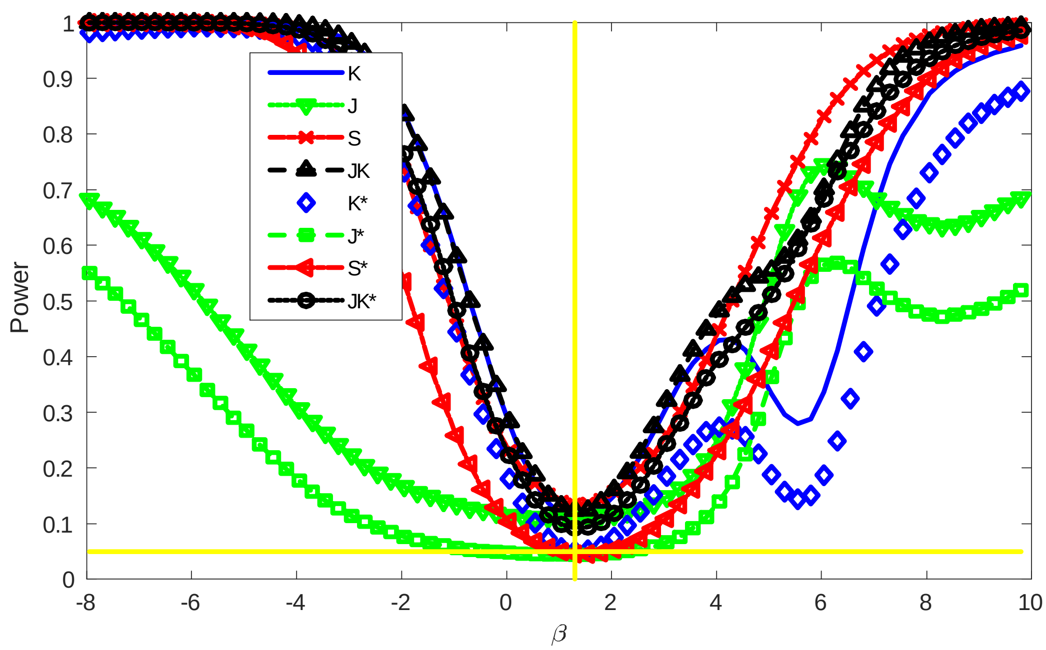


Figure 13: Power for the “Stronger ID” model, and $\mathrm{corr}(\varepsilon_c, \varepsilon_d) = .95$

8.3 SDF model for consumption growth and return on SP500

In this section, we use yearly observations (1871–1993) of U.S. consumption and stock returns to construct asymptotic confidence sets for the risk aversion parameter of the SDF model described in Section 8.2. The dataset we adopt is the same as that used by Stock and Wright (2000) and Kleibergen (2005). As explained therein, the consumption series consists of nondurables and services per capita, while the stock returns are based on the Cowles Commission index, followed by the annual average price of the Standard & Poor's monthly composite index.

To obtain a confidence set for β, we specify a sequence of increasing values of β. We compute the statistics of interest (K, J, and S statistics and their modified versions) for each value of the sequence of β. Then, the $(1-\alpha)100\%$ confidence set of β associated with a test consists of the values of β that are not rejected by the test at the level of $100\alpha\%$.

We use the moment conditions in (19) with two sets of instruments. The first set consists of a constant and the one-period lagged observations of the two series, that is, $Z_t = (1, c_t, R_t)$. The second set consists of a constant and the two-period lagged observations of the two series, that is, $Z_t = (1, c_{t-1}, R_{t-1})$. We follow the argument of Hall (1988) and lag the observations by two periods to avoid a problem with aggregating the consumption data.

For the conventional K, J, and S statistics, the covariance matrices are estimated using Newey–West covariance matrix estimators with one lag as in Kleibergen (2005). For our modified K, J, and S statistics, we apply the data-driven procedure shown in the previous section to select the number of basis functions used in the LRV estimation. Figures 14 and 15 report the difference between 1 and the p-value (i.e., one minus the p-value) for the tests that test the null $H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$ for a sequence of values of β_0. The horizontal line at 95% in the figures helps us construct the 95% confidence set. More specifically, the β values whose "one minus the p-value" is below 95% belong to the 95% confidence set.

In the first case with $Z_t = (1, c_t, R_t)$, the "one minus the p-value" plots of the K test and K* test are similar. The 95% confidence set for β based on the K test is slightly larger than that based on the K* test. In the second case with $Z_t = (1, c_{t-1}, R_{t-1})$, however, the "one minus the p-value" plots of the K test and K* test differ considerably. The 95% confidence set for β resulting from the K* test is much smaller than that from the conventional K test. However, the "one minus the p-value" plot of the J* test indicates that support for the moment equations is limited, as its "one minus the p-values" are above 0.95 in most cases.

9 Conclusion

In this paper, we combine the ideas of weak identification and fixed-smoothing asymptotics in a CU-GMM framework. The S statistic of Stock and Wright (2000) and the K statistic of Kleibergen (2005) allow us to conduct hypothesis testing in weakly identified models. On the other hand, Sun (2014b) and Hwang and Sun (2017) obtain the asymptotic distribution of the trinity of test statistics in a two-step GMM framework using fixed-smoothing asymptotics. The literature on weakly identified models does not pay particular attention to the estimation of the LRV of the moment process, while the literature on fixed-smoothing asymptotics usually imposes strong identification assumptions. We bridge this gap in the literature by applying this latter approach in a potentially weakly identified CU-GMM framework.

One striking result of the fixed-smoothing literature is that, after some simple modifications, the commonly used test statistics are asymptotically F-distributed. A key result of this paper is that this property continues to hold for the weak-identification-robust test statistics. In par-

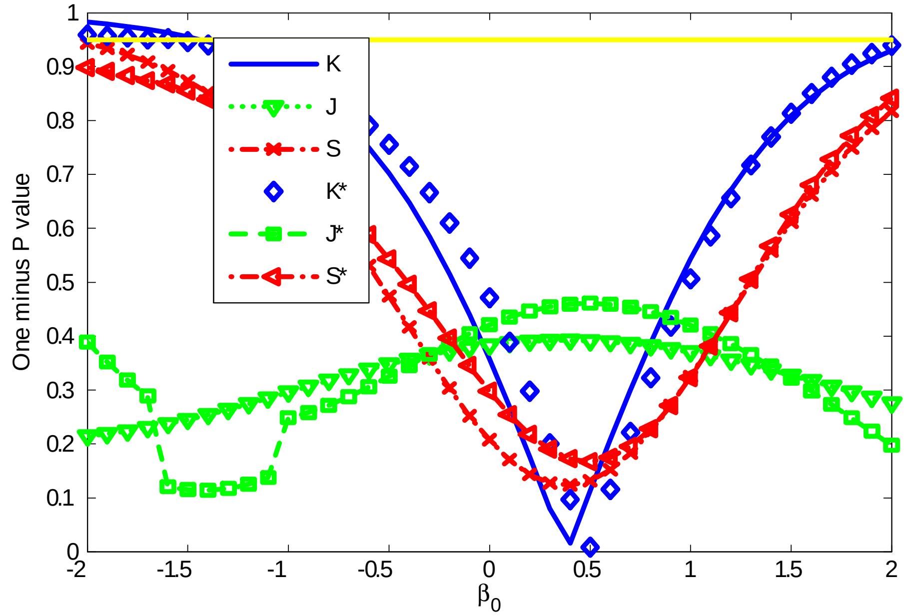


Figure 14: "One minus the p-value" plots for the tests when the instument set is $Z_t = (1, c_t, R_t)$.

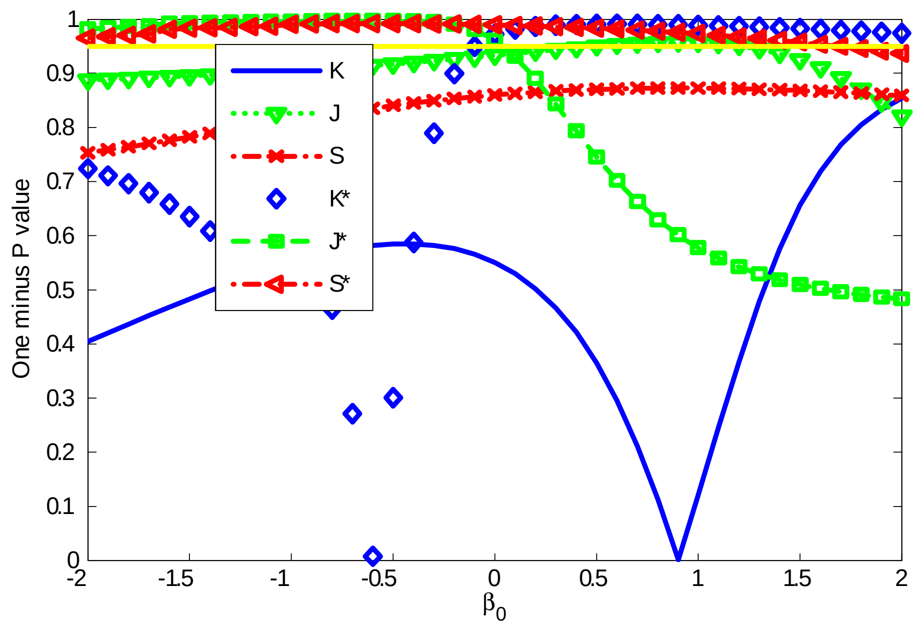


Figure 15: "One minus the p-value" plots for the tests when the instument set is $Z_t = (1, c_{t-1}, R_{t-1})$.

ticular, the modified versions of the K and S statistics are asymptotically F-distributed. This is convenient, since we can avoid using simulations to obtain the critical values for these tests. The simulation results show that in finite samples, the F approximations deliver better results than the conventional chi-squared approximations.

10 Appendix A: Proofs of the Main Results

Proof of Lemma 3.2. Using Assumptions 3.2 and 3.3, and the weak convergence results in (5) and (6), we have

$$
\begin{aligned}
&\hat{V}_{ff}(\theta_0) \\
&= \frac{1}{G}\sum_{\ell=1}^{G}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[f(Y_t,\theta_0)-\bar{f}(Y,\theta_0)\right]\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[f(Y_t,\theta_0)-\bar{f}(Y,\theta_0)\right]\right]' \\
&\Rightarrow \frac{1}{G}\sum_{\ell=1}^{G}\left(\int_0^1\Phi_\ell(r)\,dB_f(r)\right)\left(\int_0^1\Phi_\ell(r)\,dB_f(r)\right)' = \frac{1}{G}\sum_{\ell=1}^{G}\xi_{f,\ell}\xi_{f,\ell}',
\end{aligned}
$$

and

$$
\begin{aligned}
&\hat{V}_{g_jf}(\theta_0) \\
&= \frac{1}{G}\sum_{\ell=1}^{G}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[g_j(Y_t,\theta_0)-\bar{g}_j(Y,\theta)\right]\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[f(Y_t,\theta_0)-\bar{f}(Y,\theta)\right]\right]' \\
&\Rightarrow \frac{1}{G}\sum_{\ell=1}^{G}\left(\int_0^1\Phi_\ell(r)\,dB_{g_j}(r)\right)\left(\int_0^1\Phi_\ell(r)\,dB_f(r)\right)' = \frac{1}{G}\sum_{\ell=1}^{G}\xi_{g_j,\ell}\xi_{f,\ell}'
\end{aligned}
$$

jointly. Therefore,

$$
\begin{aligned}
&D_{T,j}(\theta_0) - \sqrt{T}\mathbb{E}g_j(Y_t,\theta_0) \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[g_j(Y_t,\theta_0)-\mathbb{E}g_j(Y_t,\theta_0)\right] - \frac{\partial\hat{V}_{ff}(\theta_0)}{\partial\theta_j}\hat{V}_{ff}^{-1}(\theta_0)\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f(Y_t,\theta_0)\right] \\
&\Rightarrow B_{g_j}(1) - \left[\frac{1}{G}\sum_{\ell=1}^{G}\xi_{g_j,\ell}\xi_{f,\ell}'\right]\left[\frac{1}{G}\sum_{\ell=1}^{G}\xi_{f,\ell}\xi_{f,\ell}'\right]^{-1}B_f(1) := \tilde{B}_{g_j\cdot f}(1).
\end{aligned}
$$

By Assumption 3.1, we have

$$
\frac{1}{T^{1/2-\kappa}}D_{T,j}(\theta_0) \Rightarrow D_{\infty,j} \in \mathbb{R}^{m\times 1}
$$

for

$$
D_{\infty,j} = \Pi_j + 1\left\{\kappa = \frac{1}{2}\right\}\tilde{B}_{g_j\cdot f}(1).
$$

■

Proof of Theorem 4.1. We first relate the limiting distribution $\mathcal{K}_{\infty,\theta}$ of the original K statistic to the limiting distribution of the rescaled two-step Wald statistic in Hwang and Sun

(2017, hereafter HS (2017)). If we replace R by the $d\times d$ identity matrix and G_Λ by $\tilde{D}_\infty$ in their equation (19) on page 286, then the limiting distribution $d\cdot\mathcal{F}_\infty$ of the two-step Wald statistic in HS (2017) takes the same form of $\mathcal{K}_{\infty,\theta}$ here. In fact, the two distributions are seemingly different only in terms of the notations. Noting that $B_m(1)$ and $\tilde{\mathcal{W}}_\infty$ in HS (2017) are the same as $W_f(1)$ and $\tilde{C}_\infty$ here, the limiting distribution $\mathcal{K}_{\infty,\theta}$ is identical to $d\cdot\mathcal{F}_\infty$ in HS (2017). Careful inspection reveals that the limiting distribution $\mathcal{J}_{\infty,\theta}$ of the J statistic $\mathcal{J}_T(\theta_0)$ is identical to $\mathcal{J}_\infty$ in HS (2017, equation (20), page 286).

Recall that

$$\tilde{C}_\infty=\frac{1}{G}\sum_{\ell=1}^{G}\eta_{f,\ell}\eta_{f,\ell}'=\begin{pmatrix}\tilde{C}_{dd} & \tilde{C}_{dq}\\ \tilde{C}_{qd} & \tilde{C}_{qq}\end{pmatrix}\text{ and }W_f(1)=\begin{pmatrix}W_{f,d}(1)\\ W_{f,q}(1)\end{pmatrix}.$$

Let $\tilde{C}_{dd\cdot q}=\tilde{C}_{dd}-\tilde{C}_{dq}\tilde{C}_{qq}^{-1}\tilde{C}_{qd}$, $\tilde{C}_{qq\cdot d}=\tilde{C}_{qq}-\tilde{C}_{qd}\tilde{C}_{dd}^{-1}\tilde{C}_{dq}$. By definition, $G\tilde{C}_\infty\sim\mathbb{W}_m(G,I_m)$, a Wishart distribution with G degrees of freedom. Using a basic property of Wishart distributions, we know that $G\tilde{C}_{dd\cdot q}\sim\mathbb{W}_d(G-q,I_d)$, and that $\tilde{C}_{dd\cdot q}$ is independent of $(\tilde{C}_{dq},\tilde{C}_{qq})$.

It then follows from HS (2017) that

$$\begin{aligned}\begin{pmatrix}\mathcal{K}_T(\theta_0)\\ \mathcal{J}_T(\theta_0)\end{pmatrix}&\Rightarrow\begin{pmatrix}\mathcal{K}_{\infty,\theta}\\ \mathcal{J}_{\infty,\theta}\end{pmatrix}\\ &\overset{d}{=}\begin{pmatrix}\left[W_{f,d}(1)-\tilde{C}_{dq}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]'\tilde{C}_{dd\cdot q}^{-1}\left[W_{f,d}(1)-\tilde{C}_{dq}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]\\ W_{f,q}(1)'\tilde{C}_{qq}^{-1}W_{f,q}(1)\end{pmatrix}.\end{aligned}$$

By the continuous mapping theorem, we have

$$\begin{pmatrix}\mathcal{K}_T^*(\theta_0)\\ \mathcal{J}_T^*(\theta_0)\end{pmatrix}\Rightarrow\begin{pmatrix}\frac{G-m+1}{Gd}(\psi^*)'\tilde{C}_{dd\cdot q}^{-1}\psi^*\\ \frac{G-q+1}{Gq}W_{f,q}(1)'\tilde{C}_{qq}^{-1}W_{f,q}(1)\end{pmatrix}:=\begin{pmatrix}\mathcal{K}_{\infty,\theta}^*\\ \mathcal{J}_{\infty,\theta}^*\end{pmatrix}\tag{20}$$

where

$$\psi^*:=\frac{\left[W_{f,d}(1)-\tilde{C}_{dq}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]}{\sqrt{1+\frac{1}{G}W_{f,q}'(1)\tilde{C}_{qq}^{-1}W_{f,q}(1)}}.$$

Using the same conditional argument as in HS (2017), we can show that $\psi^*\sim N(0,I_d)$ and ψ^* is independent of $W_{f,q}(1)'\tilde{C}_{qq}^{-1}W_{f,q}(1)$ and $\tilde{C}_{qq}$. Note that $(W_f(1),\tilde{C}_{dq},\tilde{C}_{qq})$ is independent of $\tilde{C}_{dd\cdot q}$. So, as a function of $(W_f(1),\tilde{C}_{dq},\tilde{C}_{qq})$, ψ^* is also independent of $\tilde{C}_{dd\cdot q}$. We have therefore shown that $\mathcal{K}_{\infty,\theta}^*$ is independent of $\mathcal{J}_{\infty,\theta}^*$ and that both $\mathcal{K}_{\infty,\theta}^*$ and $\mathcal{J}_{\infty,\theta}^*$ are equal to a rescaled Hotelling's T^2 distribution, i.e., a rescaled quadratic form in a standard normal vector with an independent Wishart weighting matrix. The rescaling is designed to turn a T^2 distribution into an F distribution. Using the relationship between the T^2 distribution and the F distribution, we have

$$\begin{pmatrix}\mathcal{K}_T^*(\theta_0)\\ \mathcal{J}_T^*(\theta_0)\end{pmatrix}\Rightarrow\begin{pmatrix}F_{d,G-m+1}\\ F_{q,G-q+1}\end{pmatrix}$$

where $F_{d,G-m+1}$ and $F_{q,G-q+1}$ are independent F distributions. ■

Proof of Theorem 4.2. We have

$$\begin{aligned}
\tilde{B}_{g_j \cdot f}(1) &= B_{g_j}(1) - \left[\frac{1}{G}\sum_{\ell=1}^{G}\left(\xi_{g_j,\ell}\eta'_{f,\ell}\right)\right]\tilde{C}_\infty^{-1}W_f(1) \\
&= \left[V_{g_j f}V_{ff}^{-1}B_f(1) + B_{g_j \cdot f}(1)\right] - \frac{1}{G}\sum_{\ell=1}^{G}\left[V_{g_j f}V_{ff}^{-1/2}\eta_{f,\ell}\eta'_{f,\ell} + \xi_{g_j \cdot f,\ell}\eta'_{f,\ell}\right]\tilde{C}_\infty^{-1}W_f(1) \\
&= \left[V_{g_j f}V_{ff}^{-1/2}W_f(1) + B_{g_j \cdot f}(1)\right] - V_{g_j f}V_{ff}^{-1/2}W_f(1) - \left[\frac{1}{G}\sum_{\ell=1}^{G}\xi_{g_j \cdot f,\ell}\eta'_{f,\ell}\right]\tilde{C}_\infty^{-1}W_f(1) \\
&= B_{g_j \cdot f}(1) - \left[\frac{1}{G}\sum_{\ell=1}^{G}\xi_{g_j \cdot f,\ell}\eta'_{f,\ell}\right]\tilde{C}_\infty^{-1}W_f(1) \\
&= V_{g_j \cdot f}^{1/2}\left\{W_{g_j \cdot f}(1) - \left[\frac{1}{G}\sum_{\ell=1}^{G}\eta_{g_j \cdot f,\ell}\eta'_{f,\ell}\right]\tilde{C}_\infty^{-1}W_f(1)\right\} \\
&:= V_{g_j \cdot f}^{1/2}\tilde{W}_{g_j \cdot f}(1)
\end{aligned}$$

where

$$\tilde{W}_{g_j \cdot f}(1) = W_{g_j \cdot f}(1) - \left[\frac{1}{G}\sum_{\ell=1}^{G}\eta_{g_j \cdot f,\ell}\eta'_{f,\ell}\right]\left[\frac{1}{G}\sum_{\ell=1}^{G}\eta_{f,\ell}\eta'_{f,\ell}\right]^{-1}W_f(1).$$

Conditional on $\{\eta_{f,\ell}\}_{\ell=1}^{G}$ and $W_f(1)$, $\tilde{W}_{g_j \cdot f}(1)$ is normal with mean zero and variance

$$\begin{aligned}
& I_m + \mathbb{E}\left[\left[\frac{1}{G}\sum_{\ell=1}^{G}\eta_{g_j \cdot f,\ell} \times \eta'_{f,\ell}\tilde{C}_\infty^{-1}W_f(1)\right]\left[\frac{1}{G}\sum_{\tilde{\ell}=1}^{G}W_f(1)'\tilde{C}_\infty^{-1}\eta_{f,\tilde{\ell}} \times \eta'_{g_j \cdot f,\tilde{\ell}}\right]\Bigg|\{\eta_{f,\ell}\}_{\ell=1}^{G}, W_f(1)\right] \\
&= I_m + \mathbb{E}\left[\left[\frac{1}{G}\sum_{\ell=1}^{G}\eta'_{f,\ell}\tilde{C}_\infty^{-1}W_f(1) \times \eta_{g_j \cdot f,\ell}\right]\left[\frac{1}{G}\sum_{\tilde{\ell}=1}^{G}\eta'_{g_j \cdot f,\tilde{\ell}} \times W_f(1)'\tilde{C}_\infty^{-1}\eta_{f,\tilde{\ell}}\right]\Bigg|\{\eta_{f,\ell}\}_{\ell=1}^{G}, W_f(1)\right] \\
&= I_m + \mathbb{E}\left[\frac{1}{G^2}\sum_{\ell=1}^{G}\sum_{\tilde{\ell}=1}^{G}\eta'_{f,\ell}\tilde{C}_\infty^{-1}W_f(1) \times \left(\eta_{g_j \cdot f,\ell}\eta'_{g_j \cdot f,\tilde{\ell}}\right) \times W_f(1)'\tilde{C}_\infty^{-1}\eta_{f,\tilde{\ell}}\Bigg| \mid \{\eta_{f,\ell}\}_{\ell=1}^{G}, W_f(1)\right] \\
&= I_m + \left[\frac{1}{G^2}\sum_{\ell=1}^{G}\eta'_{f,\ell}\tilde{C}_\infty^{-1}W_f(1) \times W_f(1)'\tilde{C}_\infty^{-1}\eta_{f,\ell}\Bigg|\right] I_m \\
&= I_m\left[1 + \frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)\right].
\end{aligned}$$

So, conditional on $\{\eta_{f,\ell}\}_{\ell=1}^{G}$ and $W_f(1)$,

$$\tilde{W}^*_{g_j \cdot f}(1) := \frac{\tilde{W}_{g_j \cdot f}(1)}{\sqrt{1 + \frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} \sim N(0, I_m).$$

Given that the conditional distribution does not depend on the conditioning variables, we have

$$\tilde{W}^*_{g_j \cdot f}(1) \sim N(0, I_m)$$

unconditionally. The covariance between $\tilde{W}^*_{g_j \cdot f}(1)$ and $W_f(1)$ is

$$
\begin{aligned}
& cov\left(\tilde{W}^*_{g_j \cdot f}(1), W_f(1)\right) \\
&= cov\left(\frac{W_{g_j \cdot f}(1) - \left[\frac{1}{G}\sum_{\ell=1}^G \eta_{g_j \cdot f,\ell}\eta'_{f,\ell}\right]\tilde{C}_\infty^{-1}W_f(1)}{\sqrt{1+\frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}}, W_f(1)\right) \\
&= \mathbb{E}\frac{W_{g_j \cdot f}(1)W_f(1)'}{\sqrt{1+\frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} - \mathbb{E}\frac{\left[\frac{1}{G}\sum_{\ell=1}^G \eta_{g_j \cdot f,\ell}\eta'_{f,\ell}\right]\tilde{C}_\infty^{-1}W_f(1)W_f(1)'}{\sqrt{1+\frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} \\
&= 0 - 0 = 0,
\end{aligned}
$$

using the law of iterated expectations. Therefore, $\tilde{W}^*_{g_j \cdot f}(1)$ is independent of $W_f(1)$. In addition, $\tilde{W}^*_{g_j \cdot f}(1)$ is independent of $\eta_{f,\ell}$ for any $\ell = 1, \ldots, G$, as

$$
\begin{aligned}
& cov\left(\tilde{W}^*_{g_j \cdot f}(1), \eta_{f,\ell}\right) \\
&= cov\left(\frac{W_{g_j \cdot f}(1) - \left[\frac{1}{G}\sum_{k=1}^G \eta_{g_j \cdot f,k}\eta'_{f,k}\right]\tilde{C}_\infty^{-1}W_f(1)}{\sqrt{1+\frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}}, \eta_{f,\ell}\right) \\
&= \mathbb{E}\frac{W_{g_j \cdot f}(1)\eta'_{f,\ell} - \left[\frac{1}{G}\sum_{k=1}^G \eta_{g_j \cdot f,k}\eta'_{f,k}\right]\tilde{C}_\infty^{-1}W_f(1)\eta'_{f,\ell}}{\sqrt{1+\frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} \\
&= \mathbb{E}\left\{\left[\frac{\mathbb{E}\left[W_{g_j \cdot f}(1) \mid \{\eta_{f,k}\}, W_f(1), \tilde{C}_\infty\right]\eta'_{f,\ell}}{\sqrt{1+\frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}}\right]\right\} \\
&\quad - \mathbb{E}\frac{\frac{1}{G}\sum_{k=1}^G \mathbb{E}\left[\eta_{g_j \cdot f,\ell} \mid \{\eta_{f,k}\}, W_f(1), \tilde{C}_\infty\right]\eta'_{f,k}\tilde{C}_\infty^{-1}W_f(1)\eta'_{f,\ell}}{\sqrt{1+\frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} = 0.
\end{aligned}
$$

As a consequence, $\tilde{W}^*_{g_j \cdot f}(1)$ is independent of $\tilde{C}_\infty$.

Let

$$
\begin{aligned}
\tilde{D}^*_{\infty,j} &= \frac{\tilde{D}_{\infty,j}}{\sqrt{1+\frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} \\
&= V_{ff}^{-1/2}V_{g_j \cdot f}^{1/2}\frac{\tilde{W}_{g_j \cdot f}(1)}{\sqrt{1+\frac{1}{G}W_f(1)'\tilde{C}_\infty^{-1}W_f(1)}} = V_{ff}^{-1/2}V_{g_j \cdot f}^{1/2}\tilde{W}^*_{g_j \cdot f}(1)
\end{aligned}
$$

and $\tilde{D}^*_\infty = (\tilde{D}^*_{\infty,1}, \ldots, \tilde{D}^*_{\infty,d})$. Then

$$
\mathcal{K}_{\infty,\theta} = W_f(1)'\tilde{C}_\infty^{-1}\tilde{D}^*_\infty\left\{\tilde{D}^{*\prime}_\infty\tilde{C}_\infty^{-1}\tilde{D}^*_\infty\right\}^{-1}\tilde{D}^{*\prime}_\infty\tilde{C}_\infty^{-1}W_f(1),
$$

where $\tilde{D}^*_\infty$ is independent of $\tilde{C}_\infty$ and $W_f(1)$.

Next,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{f}\left(Y_t,\theta_0\right)\Rightarrow V_{ff}^{1/2}\left[W_f\left(1\right)-\tilde{D}_\infty^*\left\{\tilde{D}_\infty^{*\prime}\tilde{C}_\infty^{-1}\tilde{D}_\infty^*\right\}^{-1}\left\{\tilde{D}_\infty^{*\prime}\tilde{C}_\infty^{-1}W_f\left(1\right)\right\}\right]$$

and so

$$\begin{aligned}\mathcal{J}_T\left(\theta_0\right)\Rightarrow\mathcal{J}_{\infty,\theta}:=&\left[W_f\left(1\right)-\tilde{D}_\infty^*\left\{\tilde{D}_\infty^{*\prime}\tilde{C}_\infty^{-1}\tilde{D}_\infty^*\right\}^{-1}\left\{\tilde{D}_\infty^{*\prime}\tilde{C}_\infty^{-1}W_f\left(1\right)\right\}\right]'\tilde{C}_\infty^{-1}\\&\times\left[W_f\left(1\right)-\tilde{D}_\infty^*\left\{\tilde{D}_\infty^{*\prime}\tilde{C}_\infty^{-1}\tilde{D}_\infty^*\right\}^{-1}\left\{\tilde{D}_\infty^{*\prime}\tilde{C}_\infty^{-1}W_f\left(1\right)\right\}\right]\end{aligned}$$

which holds jointly with $\mathcal{K}_T\left(\theta_0\right)\Rightarrow\mathcal{K}_{\infty,\theta}$.

Given that $\tilde{D}_\infty^*$ is independent of $\tilde{C}_\infty^{-1}$ and $W_f\left(1\right)$, we can use the conditioning argument (conditional on $\tilde{D}_\infty^*$) and the same proof for Theorem 4.1 to obtain

$$\begin{pmatrix}\mathcal{K}_T^*\left(\theta_0\right)\\ \mathcal{J}_T^*\left(\theta_0\right)\end{pmatrix}\Rightarrow\begin{pmatrix}F_{d,G-m+1}\\ F_{q,G-q+1}\end{pmatrix}$$

conditionally on $\tilde{D}_\infty^*$. But the conditional distribution does not depend on $\tilde{D}_\infty^*$. So the above also holds unconditionally. ■

Proof of Lemma 5.1. Part (a). This part can be proved using Sun (2014a). Details are omitted.

Part (b). We prove the marginal convergence in (12) only. The marginal convergence for the other part can be proved similarly. The joint convergence can be proved using the Cramer-Wold device. Using Assumptions 3.2 and 5.2(a.i), we have

$$\begin{aligned}\bar{f}(Y,\hat{\theta}_0)&=\frac{1}{T}\sum_{t=1}^{T}f\left(Y_t,\hat{\theta}_0\right)\\&=\frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f\left(Y_t,\theta_0\right)+\frac{1}{\sqrt{T}}\left[\frac{1}{T}\sum_{t=1}^{T}g_\alpha\left(Y_t,\tilde{\theta}_0\right)\right]\sqrt{T}\left(\hat{\alpha}-\alpha_0\right)\left(1+o_p\left(1\right)\right)\\&=\frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f\left(Y_t,\theta_0\right)+\frac{1}{\sqrt{T}}\Pi_\alpha\cdot\sqrt{T}\left(\hat{\alpha}-\alpha_0\right)\left(1+o_p\left(1\right)\right)=O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

So, by Assumptions 3.3, 5.2(b) and using Part (a), we have

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[f\left(Y_t,\hat{\theta}_0\right)-\bar{f}(Y,\hat{\theta}_0)\right]\\&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)\left[f\left(Y_t,\hat{\theta}_0\right)\right]+o_p\left(1\right)\\&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)f\left(Y_t,\theta_0\right)+\left[\frac{1}{T}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)g_\alpha\left(Y_t,\tilde{\theta}_0\right)\right]\sqrt{T}\left(\hat{\alpha}-\alpha_0\right)+o_p\left(1\right)\\&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_\ell\left(\frac{t}{T}\right)f\left(Y_t,\theta_0\right)+o_p\left(1\right)\Rightarrow\int_0^1\Phi_\ell\left(r\right)dB_f\left(r\right):=\xi_{f,\ell}.\end{aligned}$$

Part (c). Note that for $j = 1, \ldots, d_\alpha$,

$$\frac{1}{\sqrt{T}} D_{T,j}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} g_j(Y_t, \theta_0) - \frac{1}{\sqrt{T}} \hat{V}_{g_j f}(\theta_0) \hat{V}_{ff}^{-1}(\theta_0) \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t, \theta_0)\right]$$
$$= \frac{1}{T}\sum_{t=1}^{T} g_j(Y_t, \theta_0) + O_p\left(\frac{1}{\sqrt{T}}\right).$$

This holds because by Part (b) we have

$$\hat{V}_{g_j f}(\theta_0) \Rightarrow \frac{1}{G}\sum_{\ell=1}^{G} \xi_{g_j,\ell}\xi_{f,\ell}', \hat{V}_{ff}(\theta_0) \Rightarrow \frac{1}{G}\sum_{\ell=1}^{G} \xi_{f,\ell}\xi_{f,\ell}',$$

and $T^{-1/2}\sum_{t=1}^{T} f(Y_t, \theta_0) = O_p(1)$. As a result,

$$\frac{1}{T}\sum_{t=1}^{T} g_\alpha(Y_t, \theta_0) = \frac{D_{T,\alpha}}{\sqrt{T}} + o_p(1).$$

Using this result and Assumption 5.2 (b), we have

$$\hat{V}_{ff}^{-1/2}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f\left(Y_t, \hat{\theta}_0\right)\right]$$
$$= \hat{V}_{ff}^{-1/2}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(Y_t, \theta_0) - \left(\frac{1}{T}\sum_{t=1}^{T} g_\alpha\left(Y_t, \tilde{\theta}_0\right)\right)\sqrt{T}(\hat{\alpha} - \alpha_0)\right]$$
$$= \hat{V}_{ff}^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{I_m - \left(\frac{1}{T}\sum_{t=1}^{T} g_\alpha(Y_t, \theta_0)\right)\left[\frac{D_{T,\alpha}'}{\sqrt{T}}\hat{V}_{ff}^{-1}\frac{D_{T,\alpha}}{\sqrt{T}}\right]^{-1}\frac{D_{T,\alpha}'}{\sqrt{T}}\hat{V}_{ff}^{-1}\right\} f(Y_t, \theta_0)(1 + o_p(1))$$
$$= \left\{I_m - \left(\hat{V}_{ff}^{-1/2}\frac{D_{T,\alpha}}{\sqrt{T}}\right)\left[\frac{D_{T,\alpha}'}{\sqrt{T}}\hat{V}_{ff}^{-1}\frac{D_{T,\alpha}}{\sqrt{T}}\right]^{-1}\frac{D_{T,\alpha}'}{\sqrt{T}}\hat{V}_{ff}^{-1/2}\right\}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{V}_{ff}^{-1/2} f(Y_t, \theta_0)(1 + o_p(1))$$
$$= M_{\hat{V}_{ff}^{-1/2} D_{T,\alpha}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{V}_{ff}^{-1/2} f(Y_t, \theta_0)\right)(1 + o_p(1))$$
$$\Rightarrow \mathbb{M}_{C_\infty^{-1/2}\Pi_\alpha} \cdot C_\infty^{-1/2} B_f(1).$$

Combining this with $\hat{V}_{ff}^{-1/2}(\hat{\theta}_0) = \hat{V}_{ff}^{-1/2}(\theta_0) + o_p(1) = \hat{V}_{ff}^{-1/2} + o_p(1)$ leads to the desired result.

Part (d). The columns of $D_{T,\beta}(\hat{\theta}_0)$ are

$$D_{T,\beta,j-d_\alpha}(\hat{\theta}_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_j(Y_t, \hat{\theta}_0) - \hat{V}_{g_j f}(\hat{\theta}_0)\hat{V}_{ff}^{-1}(\hat{\theta}_0)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f\left(Y_t, \hat{\theta}_0\right)$$
$$= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_j(Y_t, \hat{\theta}_0) - \hat{V}_{g_j f}\hat{V}_{ff}^{-1/2} M_{\hat{V}_{ff}^{-1/2} D_{T,\alpha}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\hat{V}_{ff}^{-1/2} f(Y_t, \theta_0)\right)(1 + o_p(1))$$

for $j = d_\alpha + 1, \ldots, d_\alpha + d_\beta$. It follows from Assumption 5.2 (c.ii) and Part (c) that for $j = d_\alpha + 1, \ldots, d_\alpha + d_\beta$,

$$
\begin{aligned}
& D_{T,\beta,j-d_\alpha}(\hat{\theta}_0) - \sqrt{T}\mathbb{E}g_j(Y_t, \hat{\theta}_0) \\
& = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[g_j(Y_t, \theta_0) - E g_j(Y_t, \theta_0) \right] + o_p(1) \\
& - \hat{V}_{g_j f} \hat{V}_{ff}^{-1/2} M_{\hat{V}_{ff}^{-1/2} D_{T,\alpha}} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\hat{V}_{ff}^{-1/2} f(Y_t, \theta_0) \right) (1 + o_p(1)) \\
& \Rightarrow B_{g_j}(1) - \left[\frac{1}{G} \sum_{\ell=1}^{G} \xi_{g_j,\ell} \xi_{f,\ell}' \right] C_\infty^{-1/2} \mathbb{M}_{C_\infty^{-1/2} \Pi_\alpha} \cdot C_\infty^{-1/2} B_f(1) := \tilde{B}^*_{g_j \cdot f}(1).
\end{aligned}
$$

When $\kappa \in [0, 1/2)$, we have

$$
\begin{aligned}
\frac{D_{T,\beta}(\hat{\theta}_0)}{T^{1/2-\kappa}} &= T^\kappa \mathbb{E} g_\beta(Y_t, \hat{\theta}_0) + o_p(1) \\
&= T^\kappa \mathbb{E} g_\beta(Y_t, \theta_0) - T^\kappa \mathbb{E} \left[g_\beta(Y_t, \theta_0) - g_\beta(Y_t, \hat{\theta}_0) \right] + o_p(1) \Rightarrow \Pi_\beta,
\end{aligned}
$$

where we have used Assumption 5.2 (c.i). Similarly, when $\kappa = 1/2$, we have

$$
\frac{D_{T,\beta,j-d_\alpha}(\hat{\theta}_0)}{T^{1/2-\kappa}} \Rightarrow \Pi_j + \tilde{B}^*_{g_j \cdot f}(1)
$$

for $j = d_\alpha + 1, \ldots, d_\alpha + d_\beta$. Combining the two cases $\kappa \in [0, 1/2)$ and $\kappa = 1/2$ yields

$$
D_{T,\beta}(\hat{\theta}_0)/T^{1/2-\kappa} \Rightarrow D_{\infty,\beta}.
$$

■

Proof of Theorem 5.1. By simple calculations, we have

$$
\begin{aligned}
& \left[\tilde{\Pi}_\beta' \tilde{C}_\infty^{-1/2} \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha} \tilde{C}_\infty^{-1/2} \tilde{\Pi}_\beta \right]^{-1} \\
& = \left\{ \tilde{\Pi}_\beta' \tilde{C}_\infty^{-1} \tilde{\Pi}_\beta - \tilde{\Pi}_\beta' \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha \left[\tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha \right]^{-1} \tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\beta \right\}^{-1} \\
& = R_\beta \left(\tilde{\Pi}' \tilde{C}_\infty^{-1} \tilde{\Pi} \right)^{-1} R_\beta'.
\end{aligned}
$$

Note that

$$
\begin{aligned}
& \tilde{\Pi}_\beta' \tilde{C}_\infty^{-1/2} \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha} \tilde{C}_\infty^{-1/2} W_f(1) \\
& = \left\{ \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha} \left[\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\beta \right] \right\}' \tilde{C}_\infty^{-1/2} W_f(1) \\
& = \left\{ \left[I - \tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha \left(\tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha \right)^{-1} \tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1/2} \right] \tilde{C}_\infty^{-1/2} \tilde{\Pi}_\beta \right\}' \tilde{C}_\infty^{-1/2} W_f(1) \\
& = \left\{ C_\infty^{-1/2} \tilde{\Pi}_\beta - \tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha \left(\tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha \right)^{-1} \tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\beta \right\}' \tilde{C}_\infty^{-1/2} W_f(1) \\
& = \left[\tilde{\Pi}_\beta - \tilde{\Pi}_\alpha \left(\tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha \right)^{-1} \tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\beta \right]' \tilde{C}_\infty^{-1} W_f(1),
\end{aligned}
$$

and

$$
\begin{aligned}
& R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}\tilde{\Pi}'\tilde{C}_\infty^{-1}W_f(1) \\
&= R_\beta\begin{pmatrix} \tilde{\Pi}_\alpha'\tilde{C}_\infty^{-1}\tilde{\Pi}_\alpha & \tilde{\Pi}_\alpha'\tilde{C}_\infty^{-1}\tilde{\Pi}_\beta \\ \tilde{\Pi}_\beta'\tilde{C}_\infty^{-1}\tilde{\Pi}_\alpha, & \tilde{\Pi}_\beta'\tilde{C}_\infty^{-1}\tilde{\Pi}_\beta \end{pmatrix}^{-1}\begin{pmatrix} \tilde{\Pi}_\alpha'\tilde{C}_\infty^{-1} \\ \tilde{\Pi}_\beta'\tilde{C}_\infty^{-1} \end{pmatrix}W_f(1) \\
&= \left\{\tilde{\Pi}_\beta'\tilde{C}_\infty^{-1}\tilde{\Pi}_\beta - \tilde{\Pi}_\beta'\tilde{C}_\infty^{-1}\tilde{\Pi}_\alpha\left[\tilde{\Pi}_\alpha'\tilde{C}_\infty^{-1}\tilde{\Pi}_\alpha\right]^{-1}\tilde{\Pi}_\alpha'\tilde{C}_\infty^{-1}\tilde{\Pi}_\beta\right\}^{-1} \\
&\times\left[\tilde{\Pi}_\beta - \tilde{\Pi}_\alpha\left(\tilde{\Pi}_\alpha'\tilde{C}_\infty^{-1}\tilde{\Pi}_\alpha\right)^{-1}\tilde{\Pi}_\alpha'\tilde{C}_\infty^{-1}\tilde{\Pi}_\beta\right]'\tilde{C}_\infty^{-1}W_f(1) \\
&= \left[R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}R_\beta'\right]\left[\tilde{\Pi}_\beta - \tilde{\Pi}_\alpha\left(\tilde{\Pi}_\alpha'\tilde{C}_\infty^{-1}\tilde{\Pi}_\alpha\right)^{-1}\tilde{\Pi}_\alpha'\tilde{C}_\infty^{-1}\tilde{\Pi}_\beta\right]'\tilde{C}_\infty^{-1}W_f(1) \\
&= \left[R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}R_\beta'\right]\left[\tilde{\Pi}_\beta'\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}W_f(1)\right].
\end{aligned}
$$

We have

$$
\begin{aligned}
& \tilde{\Pi}_\beta'\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2}W_f(1) \\
&= \left[R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}R_\beta'\right]^{-1}R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}\tilde{\Pi}'\tilde{C}_\infty^{-1}W_f(1).
\end{aligned}
$$

Therefore,

$$
\mathcal{K}_{\infty,\beta} = \left\{R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}\tilde{\Pi}'\tilde{C}_\infty^{-1}W_f(1)\right\}'\left[R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}R_\beta'\right]^{-1}\left\{R_\beta\left(\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right)^{-1}\tilde{\Pi}'\tilde{C}_\infty^{-1}W_f(1)\right\}.
$$

Comparing this with the distribution $\mathcal{F}_\infty$ in HS (2017), we can see that $\mathcal{K}_{\infty,\beta}$ is the same as $p\mathcal{F}_\infty$ if we replace G_Λ and R in HS (2017, equation (19), page 286) by $\tilde{\Pi}$ and R_β, respectively.

To derive the limiting distribution of $\mathcal{J}_T(\hat{\theta}_0)$, we let

$$
F_T = \begin{pmatrix} \frac{1}{T^{1/2-\kappa}}I_{d_\beta} & O \\ O & \frac{1}{T^{1/2}}I_{d_\alpha} \end{pmatrix}.
$$

Then $\hat{D}_{FT} := D_T(\hat{\theta}_0)F_T \to^p \Pi$, and

$$
\begin{aligned}
& \frac{1}{\sqrt{T}}\sum_{t=1}^T \tilde{f}\left(Y_t,\hat{\theta}_0\right) \\
&= \hat{V}_{ff}^{1/2}(\hat{\theta}_0)\left\{I_m - \hat{V}_{ff}^{-1/2}(\hat{\theta}_0)\hat{D}_{FT}\left[\hat{D}_{FT}'\hat{V}_{ff}^{-1}(\hat{\theta}_0)\hat{D}_{FT}\right]^{-1}\hat{D}_{FT}\hat{V}_{ff}^{-1/2}(\hat{\theta}_0)\right\}\hat{V}_{ff}^{-1/2}(\hat{\theta}_0)\frac{1}{\sqrt{T}}\sum_{t=1}^T f(Y_t,\hat{\theta}_0) \\
&\Rightarrow C_\infty^{1/2}\left\{I_m - C_\infty^{-1/2}\Pi\left[\Pi'C_\infty^{-1}\Pi\right]^{-1}\Pi'C_\infty^{-1/2}\right\}\mathbb{M}_{C_\infty^{-1/2}\Pi_\alpha}\cdot C_\infty^{-1/2}B_f(1) \\
&= V_{ff}^{1/2}\tilde{C}_\infty^{1/2}\left\{I_m - \tilde{C}_\infty^{-1/2}\tilde{\Pi}\left[\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right]^{-1}\tilde{\Pi}'\tilde{C}_\infty^{-1/2}\right\}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\cdot\tilde{C}_\infty^{-1/2}W_f(1) \\
&= V_{ff}^{1/2}\tilde{C}_\infty^{1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}}\cdot\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\cdot\tilde{C}_\infty^{-1/2}W_f(1) \\
&= V_{ff}^{1/2}\tilde{C}_\infty^{1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}}\cdot\tilde{C}_\infty^{-1/2}W_f(1) \\
&= V_{ff}^{1/2}\left[W_f(1) - \tilde{\Pi}\left\{\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right\}^{-1}\left\{\tilde{\Pi}'\tilde{C}_\infty^{-1}W_f(1)\right\}\right].
\end{aligned}
$$

Therefore,

$$\mathcal{J}_T(\hat{\theta}_0) \Rightarrow \mathcal{J}_{\infty,\beta} := \left[W_f(1) - \tilde{\Pi}\left\{\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right\}^{-1}\left\{\tilde{\Pi}'\tilde{C}_\infty^{-1}W_f(1)\right\}\right]'\tilde{C}_\infty^{-1}$$
$$\times\left[W_f(1) - \tilde{\Pi}\left\{\tilde{\Pi}'\tilde{C}_\infty^{-1}\tilde{\Pi}\right\}^{-1}\left\{\tilde{\Pi}'\tilde{C}_\infty^{-1}W_f(1)\right\}\right]$$

The weak convergence holds jointly with the weak convergence: $\mathcal{K}_T(\hat{\theta}_0) \Rightarrow \mathcal{K}_{\infty,\beta}$. With some notational changes, we can see that $\mathcal{J}_{\infty,\beta}$ is the same as $\mathcal{J}_\infty$ in HS (2017, equation (20), page 286).

It now follows from HS (2017) that

$$\begin{pmatrix} \mathcal{K}_T(\hat{\theta}_0) \\ \mathcal{J}_T(\hat{\theta}_0) \end{pmatrix} \Rightarrow \begin{pmatrix} \mathcal{K}_{\infty,\beta} \\ \mathcal{J}_{\infty,\beta} \end{pmatrix}$$
$$\stackrel{d}{=} \begin{pmatrix} \left[W_{f,d_\beta}(1) - \tilde{C}_{d_\beta,q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]' C_{d_\beta d_\beta\cdot q}^{-1}\left[W_{f,d_\beta}(1) - \tilde{C}_{d_\beta,q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right] \\ W_{f,q}(1)'\tilde{C}_{qq}^{-1}W_{f,q}(1) \end{pmatrix}.$$

Hence

$$\begin{pmatrix} \mathcal{K}_T^*(\hat{\theta}_0) \\ \mathcal{J}_T^*(\hat{\theta}_0) \end{pmatrix} \Rightarrow \begin{pmatrix} \frac{G-d_\beta-q+1}{Gd_\beta}\left(\psi_\beta^*\right)' C_{d_\beta d_\beta\cdot q}^{-1}\psi_\beta^* \\ \frac{G-q+1}{Gq}W_{f,q}(1)'\tilde{C}_{qq}^{-1}W_{f,q}(1) \end{pmatrix},$$

where

$$\psi_\beta^* = \frac{W_{f,d_\beta}(1) - \tilde{C}_{d_\beta,q}\tilde{C}_{qq}^{-1}W_{f,q}(1)}{\sqrt{1 + W_{f,q}(1)'\tilde{C}_{qq}^{-1}W_{f,q}(1)/G}}.$$

Using the same proof as that for Theorem 4.1, we then have

$$\begin{pmatrix} \mathcal{K}_T^*(\hat{\theta}_0) \\ \mathcal{J}_T^*(\hat{\theta}_0) \end{pmatrix} \Rightarrow \begin{pmatrix} F_{d_\beta,G-d_\beta-q+1} \\ F_{q,G-q+1} \end{pmatrix},$$

where the two F variates are independent. ■

Proof of Lemma 5.2. Since U_α is an $m \times m$ orthogonal matrix, $\left[U_\alpha'\tilde{C}_\infty^{-1}U_\alpha, U_\alpha'W_f(1)\right]$ has the same distribution as $\left[\tilde{C}_\infty^{-1}, W_f(1)\right]$. In term of the distributional equivalence, we can replace $\left[U_\alpha'\tilde{C}_\infty^{-1}U_\alpha, U_\alpha'W_f(1)\right]$ by $\left[\tilde{C}_\infty^{-1}, W_f(1)\right]$. We will do so throughout the proof so that the distributional representations in the lemma hold jointly.

Part (a).

$$
\begin{aligned}
\tilde{W}_f^{(1)}(1) &= \tilde{\Pi}_\alpha \left(\tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha\right)^{-1} \tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} W_f(1) \\
&= U_\alpha \Lambda_\alpha S_\alpha \left(S_\alpha' \Lambda_\alpha' U_\alpha' \tilde{C}_\infty^{-1} U_\alpha \Lambda_\alpha S_\alpha\right)^{-1} S_\alpha' \Lambda_\alpha' U_\alpha' \tilde{C}_\infty^{-1} U_\alpha U_\alpha' W_f(1) \\
&=^d U_\alpha \Lambda_\alpha S_\alpha \left(S_\alpha' \Lambda_\alpha' \tilde{C}_\infty^{-1} \Lambda_\alpha S_\alpha\right)^{-1} S_\alpha' \Lambda_\alpha' \tilde{C}_\infty^{-1} W_f(1) \\
&= U_\alpha \Lambda_\alpha \left(\Lambda_\alpha' \tilde{C}_\infty^{-1} \Lambda_\alpha\right)^{-1} \Lambda_\alpha' \tilde{C}_\infty^{-1} W_f(1) \\
&= U_\alpha \Lambda_\alpha \left(\Lambda_\alpha' \tilde{C}_\infty^{-1} \Lambda_\alpha\right)^{-1} \left(A_\alpha', O_\alpha'\right) \tilde{C}_\infty^{-1} W_f(1) \\
&= U_\alpha \Lambda_\alpha \left(A_\alpha' \tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1} A_\alpha\right)^{-1} A_\alpha' \tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1} \left[W_{f,\alpha}(1) - \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1)\right] \\
&= U_\alpha \begin{pmatrix} I_\alpha \\ O_\alpha \end{pmatrix} \left[W_{f,\alpha}(1) - \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1)\right] \\
&= (U_{\alpha 1}, U_{\alpha 2}) \begin{pmatrix} \left[W_{f,\alpha}(1) - \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1)\right] \\ O_{\tilde{q}\times 1} \end{pmatrix} \\
&= U_{\alpha 1} \left[W_{f,\alpha}(1) - \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1)\right].
\end{aligned}
$$

Part (b). In the proof of Part (a), we have shown that

$$
\tilde{\Pi}_\alpha \left(\tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha\right)^{-1} \tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} W_f(1) = U_\alpha \begin{pmatrix} W_{f,\alpha}(1) - \tilde{C}_{\alpha,\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1) \\ O_{\tilde{q}\times 1} \end{pmatrix}.
$$

Using this result, we have

$$
\begin{aligned}
&\tilde{C}_\infty^{-1/2} \mathbb{M}_{\tilde{C}_\infty^{-1/2} \tilde{\Pi}_\alpha} \tilde{C}_\infty^{-1/2} W_f(1) \\
&= \tilde{C}_\infty^{-1} \left[I_m - \tilde{\Pi}_\alpha \left(\tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha\right)^{-1} \tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1}\right] W_f(1) \\
&= \tilde{C}_\infty^{-1} W_f(1) - \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha \left(\tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} \tilde{\Pi}_\alpha\right)^{-1} \tilde{\Pi}_\alpha' \tilde{C}_\infty^{-1} W_f(1) \\
&=^d U_\alpha \tilde{C}_\infty^{-1} U_\alpha' U_\alpha W_f(1) - U_\alpha \tilde{C}_\infty^{-1} U_\alpha' U_\alpha \begin{pmatrix} W_{f,\alpha}(1) - \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1) \\ O_{\tilde{q}\times 1} \end{pmatrix} \\
&= U_\alpha \tilde{C}_\infty^{-1} \left[\begin{pmatrix} W_{f,\alpha}(1) \\ W_{f,\tilde{q}}(1) \end{pmatrix} - \begin{pmatrix} W_{f,\alpha}(1) - \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1) \\ O_{\tilde{q}\times 1} \end{pmatrix}\right] \\
&= U_\alpha \tilde{C}_\infty^{-1} \begin{pmatrix} \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1) \\ W_{f,\tilde{q}}(1) \end{pmatrix} = U_\alpha \tilde{C}_\infty^{-1} \begin{pmatrix} \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \\ I_{\tilde{q}\tilde{q}} \end{pmatrix} W_{f,\tilde{q}}(1) \\
&= U_\alpha \begin{pmatrix} \tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1}, & -\tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1} \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \\ -\tilde{C}_{\tilde{q}\tilde{q}\cdot\alpha}^{-1} \tilde{C}_{\tilde{q}\alpha} \tilde{C}_{\tilde{q}\tilde{q}}^{-1}, & \tilde{C}_{\tilde{q}\tilde{q}\cdot\alpha}^{-1} \end{pmatrix} \begin{pmatrix} \tilde{C}_{\alpha\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \\ I_{\tilde{q}\tilde{q}} \end{pmatrix} W_{f,\tilde{q}}(1) \\
&= U_\alpha \begin{pmatrix} O_{d_\alpha \times \tilde{q}} \\ \tilde{C}_{\tilde{q}\tilde{q}\cdot\alpha}^{-1} \left(I - \tilde{C}_{\tilde{q}\alpha} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \tilde{C}_{\alpha,\tilde{q}} \tilde{C}_{\tilde{q}\tilde{q}}^{-1}\right) \end{pmatrix} W_{f,\tilde{q}}(1) \\
&= (U_{\alpha 1}, U_{\alpha 2}) \begin{pmatrix} O_{d_\alpha \times \tilde{q}} \\ \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \end{pmatrix} W_{f,\tilde{q}}(1) = U_{\alpha 2} \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1).
\end{aligned}
$$

Part (c) Using the presentation

$$\tilde{C}_{\infty}^{-1/2}\mathbb{M}_{\tilde{C}_{\infty}^{-1/2}\tilde{\Pi}_{\alpha}}\tilde{C}_{\infty}^{-1/2}W_f(1) =^d U_{\alpha}\begin{pmatrix} O_{d_{\alpha}\times\tilde{q}} \\ \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \end{pmatrix} W_{f,\tilde{q}}(1)$$

in part (b), we have

$$\begin{aligned}
&\tilde{W}_{g_j\cdot f}^{(2)}(1) \\
&= W_{g_j\cdot f}(1) - \left(\frac{1}{G}\sum_{\ell=1}^{G}\eta_{g_j\cdot f,\ell}\eta_{f,\ell}'\right)\tilde{C}_{\infty}^{-1/2}\mathbb{M}_{\tilde{C}_{\infty}^{-1/2}\tilde{\Pi}_{\alpha}}\cdot\tilde{C}_{\infty}^{-1/2}W_f(1) \\
&=^d U_{\alpha}W_{g_j\cdot f}(1) - U_{\alpha}\left(\frac{1}{G}\sum_{\ell=1}^{G}\eta_{g_j\cdot f,\ell}\eta_{f,\ell}'\right)U_{\alpha}'U_{\alpha}\begin{pmatrix} O_{d_{\alpha}\times\tilde{q}} \\ \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \end{pmatrix} W_{f,\tilde{q}}(1) \\
&= U_{\alpha}\left[W_{g_j\cdot f}(1) - \left(\frac{1}{G}\sum_{\ell=1}^{G}\eta_{g_j\cdot f,\ell}\eta_{f,\ell}'\right)\begin{pmatrix} O_{d_{\alpha}\times\tilde{q}} \\ \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \end{pmatrix} W_{f,\tilde{q}}(1)\right] \\
&= U_{\alpha}\left[W_{g_j\cdot f}(1) - \left(\frac{1}{G}\sum_{\ell=1}^{G}\eta_{g_j\cdot f,\ell}\eta_{f,\tilde{q},\ell}'\right)\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right].
\end{aligned}$$

Part (d). We have

$$\begin{aligned}
&\tilde{C}_{\infty}^{-1/2}\mathbb{M}_{\tilde{C}_{\infty}^{-1/2}\tilde{\Pi}_{\alpha}}\tilde{C}_{\infty}^{-1/2} \\
&= \tilde{C}_{\infty}^{-1}\left[I_m - \tilde{\Pi}_{\alpha}\left(\tilde{\Pi}_{\alpha}'\tilde{C}_{\infty}^{-1}\tilde{\Pi}_{\alpha}\right)^{-1}\tilde{\Pi}_{\alpha}'\tilde{C}_{\infty}^{-1}\right] \\
&=^d U_{\alpha}\tilde{C}_{\infty}^{-1}U_{\alpha}'\left[I_m - U_{\alpha}\Lambda_{\alpha}\left(\Lambda_{\alpha}'\tilde{C}_{\infty}^{-1}\Lambda_{\alpha}\right)^{-1}\left(A_{\alpha}', O_{\alpha}'\right)\tilde{C}_{\infty}^{-1}U_{\alpha}'\right] \\
&= U_{\alpha}\tilde{C}_{\infty}^{-1}U_{\alpha}'\left[I_m - U_{\alpha}\Lambda_{\alpha}\left(A_{\alpha}'\tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1}A_{\alpha}\right)^{-1}A_{\alpha}'\tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1}\left(I_{\tilde{q}\tilde{q}}, -\tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}\right)U_{\alpha}'\right] \\
&= U_{\alpha}\tilde{C}_{\infty}^{-1}U_{\alpha}'\left[I_m - U_{\alpha}\Lambda_{\alpha}A_{\alpha}^{-1}\tilde{C}_{\alpha\alpha\cdot\tilde{q}}\left(A_{\alpha}'\right)^{-1}A_{\alpha}'\tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1}\left(I_{\tilde{q}\tilde{q}}, -\tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}\right)U_{\alpha}'\right] \\
&= U_{\alpha}\tilde{C}_{\infty}^{-1}U_{\alpha}'\left[I_m - U_{\alpha}\begin{pmatrix} I_{\alpha} \\ O_{\alpha} \end{pmatrix}\left(I_{\tilde{q}\tilde{q}}, -\tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}\right)U_{\alpha}'\right] \\
&= U_{\alpha}\tilde{C}_{\infty}^{-1}U_{\alpha}'\left[I_m - \left(U_{\alpha 1}, U_{\alpha 2}\right)\begin{pmatrix} I_{\tilde{q}\tilde{q}}, & -\tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1} \\ O, & O \end{pmatrix}U_{\alpha}'\right] \\
&= U_{\alpha}\tilde{C}_{\infty}^{-1}U_{\alpha}'\left[I_m - \begin{pmatrix} U_{\alpha 1}, & -U_{\alpha 1}\tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1} \end{pmatrix}\begin{pmatrix} U_{\alpha 1}' \\ U_{\alpha 2}' \end{pmatrix}\right].
\end{aligned}$$

Using $U_{\alpha 1}U_{\alpha 1}' + U_{\alpha 2}U_{\alpha 2}' = I_m$, $U_{\alpha 1}'U_{\alpha 2} = 0$, $U_{\alpha 1}'U_{\alpha 1} = I_{d_{\alpha}}$, $U_{\alpha 2}'U_{\alpha 2} = I_{m-d_{\alpha}}$, we can simplify the above expression as

$$
\begin{aligned}
&\tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\tilde{C}_\infty^{-1/2} \\
&=^d U_\alpha \tilde{C}_\infty^{-1} U_\alpha' \left[I_m - U_{\alpha 1}U_{\alpha 1}' + U_{\alpha 1}\tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}U_{\alpha 2}' \right] \\
&= U_\alpha \tilde{C}_\infty^{-1} \begin{pmatrix} U_{\alpha 1}' \\ U_{\alpha 2}' \end{pmatrix} \left[U_{\alpha 2}U_{\alpha 2}' + U_{\alpha 1}\tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}U_{\alpha 2}' \right] \\
&= U_\alpha \tilde{C}_\infty^{-1} \begin{pmatrix} \tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}U_{\alpha 2}' \\ U_{\alpha 2}' \end{pmatrix} \\
&= U_\alpha \begin{pmatrix} \tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1}, & -\tilde{C}_{\alpha\alpha\cdot\tilde{q}}^{-1}\tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1} \\ -\tilde{C}_{\tilde{q}\tilde{q}\cdot\alpha}^{-1}\tilde{C}_{\tilde{q}\alpha}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}, & \tilde{C}_{\tilde{q}\tilde{q}\cdot\alpha}^{-1} \end{pmatrix} \begin{pmatrix} \tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1} \\ I_{\tilde{q}\tilde{q}} \end{pmatrix} U_{\alpha 2}' \\
&= U_\alpha \begin{pmatrix} O \\ \tilde{C}_{\tilde{q}\tilde{q}\cdot\alpha}^{-1} - \tilde{C}_{\tilde{q}\tilde{q}\cdot\alpha}^{-1}\tilde{C}_{\tilde{q}\alpha}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}\tilde{C}_{\alpha\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1} \end{pmatrix} U_{\alpha 2}' \\
&= (U_{\alpha 1}, U_{\alpha 2}) \begin{pmatrix} O \\ \tilde{C}_{\tilde{q}\tilde{q}}^{-1} \end{pmatrix} U_{\alpha 2}' = U_{\alpha 2}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}U_{\alpha 2}'.
\end{aligned}
$$

■

Proof of Theorem 5.2. We start by proving two key results. First, conditional on $\{\eta_{f,\tilde{q},\ell}\} = \{\eta_{f,\tilde{q},\ell}, \ell = 1, \ldots, G\}$ and $W_{f,\tilde{q}}(1)$, $\tilde{W}_{g_j\cdot f}^{(2)}(1)$ is normal with mean zero and variance

$$\left(1 + \frac{1}{G}W_{f,\tilde{q}}(1)' \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1)\right) I_m = \left(1 + \frac{1}{G}\tilde{\mathcal{J}}_\infty\right) I_m.$$

So, conditional on $\{\eta_{f,\tilde{q},\ell}\}$ and $W_{f,\tilde{q}}(1)$,

$$\frac{\tilde{W}_{g_j\cdot f}^{(2)}(1)}{\sqrt{1 + G^{-1}\tilde{\mathcal{J}}_\infty}} \sim N(0, I_m).$$

Given that the conditional distribution does not depend on the conditioning variables $\{\eta_{f,\tilde{q},\ell}\}$ and $W_{f,\tilde{q}}(1)$, we have

$$\frac{\tilde{W}_{g_j\cdot f}^{(2)}(1)}{\sqrt{1 + G^{-1}\tilde{\mathcal{J}}_\infty}} \sim N(0, I_m),$$

unconditionally and $W_{g_j\cdot f}^{(2)}(1)/\sqrt{1 + G^{-1}\tilde{\mathcal{J}}_\infty}$ is independent of $\{\eta_{f,\tilde{q},\ell}\}$ and $W_{f,\tilde{q}}(1)$.

Second, conditional on $\{\eta_{f,\tilde{q},\ell}\}$ and $W_{f,\tilde{q}}(1)$, $U_{\alpha 1}'W_f^{(1)}(1)$ is normal with mean zero and variance

$$
\begin{aligned}
&\mathbb{E}\left\{ \left[W_{f,\alpha}(1) - \tilde{C}_{\alpha,\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1) \right] \left[W_{f,\alpha}(1) - \tilde{C}_{\alpha,\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1) \right]' \right\} \\
&= \left(1 + \frac{1}{G}W_{f,\tilde{q}}(1)' \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1)\right) I_{d_\alpha}
\end{aligned}
$$

where I_{d_α} is the $d_\alpha \times d_\alpha$ identity matrix. So, conditional on $\{\eta_{f,\tilde{q},\ell}\}$ and $W_{f,\tilde{q}}(1)$

$$\frac{W_f^{(1)}(1)}{\sqrt{1 + G^{-1}\tilde{\mathcal{J}}_\infty}} \sim N(0, U_{\alpha 1}U_{\alpha 1}').$$

The conditional distribution does not depend on the conditioning variables. So $W_f^{(1)}(1)/\sqrt{1+G^{-1}\tilde{\mathcal{J}}_\infty}$ is unconditionally normal and is independent of $\{\eta_{f,\tilde{q},\ell}\}$ and $W_{f,\tilde{q}}(1)$.

Note that $\tilde{C}_{\tilde{q}\tilde{q}}$ is a function of $\{\eta_{f,\tilde{q},\ell}\}$. It follows from the above results that $\tilde{D}^*_{\infty,\beta,j}$ is independent of $\tilde{C}_{\tilde{q}\tilde{q}}^{-1}$ and $W_{f,\tilde{q}}(1)$. Therefore, conditional on $U'_{\alpha 2}\tilde{D}^*_{\infty,\beta}\in\mathbb{R}^{\tilde{q}\times d_\beta}$, the distribution of $\mathcal{K}_{\infty,\beta}$, which is given by

$$\mathcal{K}_{\infty,\beta}=\left\|\mathbb{P}_{\tilde{C}_{\tilde{q}\tilde{q}}^{-1/2}U'_{\alpha 2}\tilde{D}^*_{\infty,\beta}}\left[\tilde{C}_{\tilde{q}\tilde{q}}^{-1/2}W_{f,\tilde{q}}(1)\right]\right\|^2,$$

takes the same form as that of $\mathcal{K}_{\infty,\theta}=\left\|\mathbb{P}_{\tilde{C}_\infty^{-1/2}\tilde{D}_\infty}\left[\tilde{C}_\infty^{-1/2}W_f(1)\right]\right\|^2$ given in (8). The only difference lies in the dimensionality. We can use the same argument to simplify $\mathcal{K}_{\infty,\beta}$. More specifically, we write

$$\tilde{C}_{\tilde{q}\tilde{q}}=\begin{pmatrix}\tilde{C}_{\beta\beta}, & \tilde{C}_{\beta q}\\ \tilde{C}_{q\beta}, & \tilde{C}_{qq}\end{pmatrix}\text{ and } W_{f,\tilde{q}}(1)=\begin{pmatrix}W_{f,\beta}(1)\\ W_{f,q}(1)\end{pmatrix},$$

and let $U_D\Lambda_D S_D$ be the SVD of $(\tilde{D}^*_{\infty,\beta})'U_{\alpha 2}$. Then

$$\begin{aligned}\mathcal{K}_{\infty,\beta}&=^d\left\{U_D\Lambda_D S_D\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right\}'\left\{U_D\Lambda_D S_D\tilde{C}_{\tilde{q}\tilde{q}}^{-1}S'_D\Lambda'_D U'_D\right\}^{-1}\\
&\times\left\{U_D\Lambda_D S_D\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right\}.\\
&=^d\left\{\Lambda_D\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right\}'\left\{\Lambda_D\tilde{C}_{\tilde{q}\tilde{q}}^{-1}\Lambda'_D\right\}^{-1}\left\{\Lambda_D\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right\}\\
&=^d\left[W_{f,\beta}(1)-\tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]'\tilde{C}_{\beta\beta\cdot q}^{-1}\left[W_{f,\beta}(1)-\tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right].\end{aligned}$$

Next, we find the limit of $\mathcal{J}_T(\hat{\theta}_0)$. In the proof of Theorem 6.1, we will show that

$$\mathcal{S}_T(\hat{\theta}_0)\Rightarrow W_{f,\tilde{q}}(1)'\tilde{C}_{\tilde{q}\tilde{q}}^{-1}\tilde{C}_{\tilde{q}\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)=W_{f,\tilde{q}}(1)'\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1).$$

The above convergence holds jointly with the weak convergence:

$$\mathcal{K}_T(\hat{\theta}_0)\Rightarrow\left[W_{f,\beta}(1)-\tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]'\tilde{C}_{\beta\beta\cdot q}^{-1}\left[W_{f,\beta}(1)-\tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right].$$

So

$$\begin{aligned}&\mathcal{J}_T(\hat{\theta}_0)=\mathcal{S}_T(\hat{\theta}_0)-\mathcal{K}_T(\hat{\theta}_0)\\
&\Rightarrow W_{f,\tilde{q}}(1)'\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)-\left[W_{f,\beta}(1)-\tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]'\tilde{C}_{\beta\beta\cdot q}^{-1}\left[W_{f,\beta}(1)-\tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right].\end{aligned}$$

But

$$\begin{aligned}&W_{f,\tilde{q}}(1)'\tilde{C}_{qq}^{-1}W_{f,\tilde{q}}(1)\\
&=\begin{pmatrix}W_{f,\beta}(1)\\ W_{f,q}(1)\end{pmatrix}'\begin{pmatrix}\tilde{C}_{\beta\beta}, & \tilde{C}_{\beta q}\\ \tilde{C}_{q\beta}, & \tilde{C}_{qq}\end{pmatrix}^{-1}\begin{pmatrix}W_{f,\beta}(1)\\ W_{f,q}(1)\end{pmatrix}\\
&=\left\|\begin{pmatrix}\tilde{C}_{\beta\beta\cdot q}^{-1/2}, & -\tilde{C}_{\beta\beta\cdot q}^{-1/2}\tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}\\ O, & \tilde{C}_{qq}^{-1/2}\end{pmatrix}\begin{pmatrix}W_{f,\beta}(1)\\ W_{f,q}(1)\end{pmatrix}\right\|\\
&=W_{f,q}(1)'\tilde{C}_{qq}^{-1}W_{f,q}(1)\\
&+\left[W_{f,\beta}(1)-\tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]'\tilde{C}_{\beta\beta\cdot q}^{-1}\left[W_{f,\beta}(1)-\tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]'.\end{aligned}$$

We have

$$\mathcal{J}_T(\hat{\theta}_0) \Rightarrow W_{f,q}(1)' \tilde{C}_{qq}^{-1} W_{f,q}(1).$$

Combining the limits of $\mathcal{K}_T(\hat{\theta}_0)$ and $\mathcal{J}_T(\hat{\theta}_0)$, we have:

$$\begin{pmatrix} \mathcal{K}_T(\hat{\theta}_0) \\ \mathcal{J}_T(\hat{\theta}_0) \end{pmatrix} \Rightarrow \begin{pmatrix} \left[W_{f,\beta}(1) - \tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right]' \tilde{C}_{\beta\beta\cdot q}^{-1}\left[W_{f,\beta}(1) - \tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)\right] \\ W_{f,q}(1)' \tilde{C}_{qq}^{-1} W_{f,q}(1) \end{pmatrix}.$$

As a result,

$$\begin{pmatrix} \mathcal{K}_T^*(\hat{\theta}_0) \\ \mathcal{J}_T^*(\hat{\theta}_0) \end{pmatrix} \Rightarrow \begin{pmatrix} \frac{G-d_\beta-q+1}{Gd_\beta}\left(\psi_\beta^*\right)' \tilde{C}_{\beta\beta\cdot q}^{-1}\psi_\beta^* \\ \frac{G-q+1}{Gq} W_{f,q}(1)' \tilde{C}_{qq}^{-1} W_{f,q}(1) \end{pmatrix} := \begin{pmatrix} \mathcal{K}_{\infty,\beta}^* \\ \mathcal{J}_{\infty,\beta}^* \end{pmatrix},$$

where

$$\psi_\beta^* = \frac{W_{f,\beta}(1) - \tilde{C}_{\beta q}\tilde{C}_{qq}^{-1}W_{f,q}(1)}{\sqrt{1 + G^{-1}W_{f,q}(1)\tilde{C}_{qq}^{-1}W_{f,q}(1)}}.$$

The form of $(\mathcal{K}_{\infty,\beta}^*, \mathcal{J}_{\infty,\beta}^*)'$ is the same as that of $(\mathcal{K}_{\infty,\theta}^*, \mathcal{J}_{\infty,\theta}^*)'$ in (20). Using the same argument at the end of the proof for Theorem 4.1, we obtain

$$\begin{pmatrix} \mathcal{K}_T^*(\hat{\theta}_0) \\ \mathcal{J}_T^*(\hat{\theta}_0) \end{pmatrix} \Rightarrow \begin{pmatrix} F_{d_\beta, G-d_\beta-q+1} \\ F_{q,G-q+1} \end{pmatrix}.$$

■

Proof of Theorem 6.1. The weak convergence result for $\mathcal{S}_T^*(\theta_0)$ holds because

$$\begin{aligned}
\mathcal{S}_T^*(\theta_0) &\Rightarrow \frac{G-m+1}{Gm} B_f(1)' \left[\frac{1}{G}\sum_{\ell=1}^G \xi_{f,\ell}\xi_{f,\ell}'\right]^{-1} B_f(1) \\
&= \frac{G-m+1}{Gm} W_f(1)' \left[\frac{1}{G}\sum_{\ell=1}^G \eta_{f,\ell}\eta_{f,\ell}'\right]^{-1} W_f(1) \\
&=^d F_{m,G-m+1}.
\end{aligned}$$

It remains to prove the result for $\mathcal{S}_T(\hat{\theta}_0)$. Note that Lemma 5.1(c) holds under the assumptions given in the theorem. Using 5.1(c) and 5.2(b), we have

$$\begin{aligned}
\mathcal{S}_T(\hat{\theta}_0) &\Rightarrow \left[\mathbb{M}_{C_\infty^{-1/2}\Pi_\alpha} \cdot C_\infty^{-1/2}B_f(1)\right]' \mathbb{M}_{C_\infty^{-1/2}\Pi_\alpha} \cdot C_\infty^{-1/2}B_f(1) \\
&= \left[W_f(1)' \tilde{C}_\infty^{-1/2}\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}\right]\left[\mathbb{M}_{\tilde{C}_\infty^{-1/2}\tilde{\Pi}_\alpha}' \tilde{C}_\infty^{-1/2} W_f(1)\right] \\
&=^d \left[U_{\alpha 2}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1)\right]' \left(U_\alpha \tilde{C}_\infty^{1/2}U_\alpha'\right)\left(U_\alpha \tilde{C}_\infty^{1/2}U_\alpha'\right) U_{\alpha 2}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1) \\
&= W_{f,\tilde{q}}(1)' \tilde{C}_{\tilde{q}\tilde{q}}^{-1}\left(U_{\alpha 2}'U_\alpha \tilde{C}_\infty U_\alpha' U_{\alpha 2}\right)\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1) \\
&= W_{f,\tilde{q}}(1)' \tilde{C}_{\tilde{q}\tilde{q}}^{-1}\left[\begin{pmatrix} O \\ I_{\tilde{q}} \end{pmatrix}' \tilde{C}_\infty \begin{pmatrix} O \\ I_{\tilde{q}} \end{pmatrix}\right]\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1) \\
&= W_{f,\tilde{q}}(1)' \tilde{C}_{\tilde{q}\tilde{q}}^{-1}\tilde{C}_{\tilde{q}\tilde{q}}\tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1) = W_{f,\tilde{q}}(1)' \tilde{C}_{\tilde{q}\tilde{q}}^{-1}W_{f,\tilde{q}}(1).
\end{aligned}$$

Therefore,

$$\mathcal{S}_T^*(\hat{\theta}_0) = \frac{G - \tilde{q} + 1}{G\tilde{q}} \mathcal{S}_T(\hat{\theta}_0) \Rightarrow \frac{G - \tilde{q} + 1}{G\tilde{q}} W_{f,\tilde{q}}(1)' \tilde{C}_{\tilde{q}\tilde{q}}^{-1} W_{f,\tilde{q}}(1) =^d F_{\tilde{q},G-\tilde{q}+1}.$$

■

11 Appendix B: the nonpivotality of the fixed-smoothing asymptotic distribution in the intermediate case

To prove the negative result, we can consider the special case with $m = 2$ and $d = 1$. In this case, $\tilde{D}_\infty^*$ is a 2×1 vector:

$$\begin{aligned}
&\tilde{D}_\infty^* \\
&= \frac{V_{ff}^{-1/2}\Pi}{\sqrt{1 + \frac{1}{G} W_f(1)' C_\infty^{-1} W_f(1)}} + V_{ff}^{-1/2} V_{g\cdot f}^{1/2} \tilde{W}_{g\cdot f}^*(1) \\
&= \begin{pmatrix} \Pi_1^* + \delta_1 \\ \Pi_2^* + \delta_2 \end{pmatrix} \text{ for } \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} = V_{ff}^{-1/2} V_{g\cdot f}^{1/2} \tilde{W}_{g\cdot f}^*(1),
\end{aligned}$$

where $(\delta_1, \delta_2)' \in \mathbb{R}^2$ is $N(0, V_{ff}^{-1/2} V_{g\cdot f} V_{ff}^{-1/2})$. In the proof of Theorem 4.2, we have shown that $(\delta_1, \delta_2)'$ is independent of $W_f(1)$ and $\tilde{C}_\infty^{-1}$. Note that

$$\tilde{D}_\infty^o := \frac{\tilde{D}_\infty^*}{||\tilde{D}_\infty^*||}$$

lies on the unit circle and its distribution is completely characterized by its polar angle. The distribution of the angle clearly depends on the distribution of the ratio

$$\frac{\Pi_1^* + \delta_1}{\Pi_2^* + \delta_2},$$

which in turn depends on the nuisance parameters $V_{ff}^{-1/2}\Pi$ and $V_{ff}^{-1/2} V_{g\cdot f}^{1/2}$.

Now,

$$\begin{aligned}
\mathcal{K}_{\infty,\theta} &= W_f(1)' \tilde{C}_\infty^{-1} \tilde{D}_\infty^o \left[\tilde{D}_\infty^{o\prime} \tilde{C}_\infty^{-1} \tilde{D}_\infty^o \right]^{-1} \tilde{D}_\infty^{o\prime} \tilde{C}_\infty^{-1} W_f(1) \\
&= \frac{\left[W_f(1)' \tilde{C}_\infty^{-1} \tilde{D}_\infty^o \right]^2}{\left[\tilde{D}_\infty^{o\prime} \tilde{C}_\infty^{-1} \tilde{D}_\infty^o \right]} = \frac{\left\langle \tilde{C}_\infty^{-1/2} W_f(1), \tilde{C}_\infty^{-1/2} \tilde{D}_\infty^o \right\rangle^2}{\left\| \tilde{C}_\infty^{-1/2} W_f(1) \right\|^2 \left\| \tilde{C}_\infty^{-1/2} \tilde{D}_\infty^o \right\|^2} \left\| \tilde{C}_\infty^{-1/2} W_f(1) \right\|^2 \\
&= \cos^2(\Upsilon) \left\| \tilde{C}_\infty^{-1/2} W_f(1) \right\|^2,
\end{aligned}$$

where Υ is an angle spanned by the vectors $\tilde{C}_\infty^{-1/2} W_f(1) / ||\tilde{C}_\infty^{-1/2} W_f(1)||$ and $\tilde{C}_\infty^{-1/2} \tilde{D}_\infty^o / ||\tilde{C}_\infty^{-1/2} \tilde{D}_\infty^o||$. Let Υ_1 and Υ_2 be the polar angles (in $[0, 2\pi]$) of the two vectors $\tilde{C}_\infty^{-1/2} W_f(1) / ||\tilde{C}_\infty^{-1/2} W_f(1)||$ and $\tilde{C}_\infty^{-1/2} \tilde{D}_\infty^o / ||\tilde{C}_\infty^{-1/2} \tilde{D}_\infty^o||$, respectively. Then we can take $\Upsilon = \Upsilon_1 - \Upsilon_2$.

We can make three observations. First, we can show that $\tilde{C}_{\infty}^{-1/2}W_f(1)/||\tilde{C}_{\infty}^{-1/2}W_f(1)||$ is uniform on the unit circle. For any orthogonal matrix U, we have

$$U\frac{\tilde{C}_{\infty}^{-1/2}W_f(1)}{\left\|\tilde{C}_{\infty}^{-1/2}W_f(1)\right\|}=\frac{(U\tilde{C}_{\infty}^{-1/2}U')UW_f(1)}{\left\|(U\tilde{C}_{\infty}^{-1/2}U')UW_f(1)\right\|}:=^d\frac{\tilde{C}_{\infty}^{-1/2}W_f(1)}{\left\|\tilde{C}_{\infty}^{-1/2}W_f(1)\right\|}.$$

That is, the distribution of $\tilde{C}_{\infty}^{-1/2}W_f(1)/||\tilde{C}_{\infty}^{-1/2}W_f(1)||$ is invariant to any rotation. This implies that $\tilde{C}_{\infty}^{-1/2}W_f(1)/||\tilde{C}_{\infty}^{-1/2}W_f(1)||$ is indeed uniform on the unit circle. That is, Υ_1 is uniform over $[0,2\pi]$. Second, while $\tilde{C}_{\infty}^{-1/2}\tilde{D}_{\infty}^{o}/||\tilde{C}_{\infty}^{-1/2}\tilde{D}_{\infty}^{o}||$ lies on the unit circle too, it is neither symmetric or concentrated on a single point on the unit circle, as $\tilde{D}_{\infty}^{o}$ does not satisfy these properties either and there is dependence between $\tilde{C}_{\infty}^{-1/2}$ and $\tilde{D}_{\infty}^{o}$ when $\Pi\neq 0$. The distribution of $\tilde{C}_{\infty}^{-1/2}\tilde{D}_{\infty}^{o}/||\tilde{C}_{\infty}^{-1/2}\tilde{D}_{\infty}^{o}||$ depends on the nuisance parameters in a complicated way. So, Υ_2 is not free from nuisance parameters. The first two observations imply that the distribution of Υ is not nuisance-parameters free. Third, $||\tilde{C}_{\infty}^{-1/2}W_f(1)||^2$ is free from nuisance parameters. It follows from these three observations that the distribution of $\mathcal{K}_{\infty,\theta}$ is not nuisance-parameters free under the fixed-smoothing asymptotics.

References

[1] Andrews, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59, 817–858.

[2] Antoine, B. and Renault, E. (2009): "Efficient GMM with Nearly-weak Instruments." *Econometrics Journal* 12, 135–171.

[3] Caner, M. (2009): "Testing, Estimation in GMM and CUE with Nearly-Weak Identification." *Econometric Reviews* 29(3), 330–363.

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

[5] Hansen, L. P., Heaton, J., and Yaron, A. (1996): "Finite Sample Properties of Some Alternative GMM Estimators." *Journal of Business & Economic Statistics* 14, 262–280.

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

[7] Hwang, J. and Sun, Y. (2017): "Asymptotic F and t Tests in an Efficient GMM Setting." *Journal of Econometrics* 198(2), 277–295.

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

[9] Kiefer, N. and Vogelsang, T. (2002b): "Heteroskedasticity-Autocorrelation Robust Standard Errors Using the Bartlett Kernel without Truncation." *Econometrica* 70, 2093–2095.

[10] Kiefer, N. and Vogelsang, T. (2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests." *Econometric Theory* 21, 1130–1164.

[11] Kim, M. S. and Sun, Y. (2013): "Heteroskedasticity and Spatiotemporal Dependence Robust Inference for Linear Panel Models with Fixed Effects." *Journal of Econometrics* 177(1), 85–108.

[12] Kleibergen, F. (2005): "Testing Parameters in GMM without Assuming that They Are Identified." *Econometrica* 73, 1103–1123.

[13] Kocherlakota, N. R. (1990): "On Tests of Representative Consumer Asset Pricing Models." *Journal of Monetary Economics* 26, 285–304.

[14] Lazarus, E., Lewis, D. J., Stock, J. H. and Watson, M. W. (2016): "HAR Inference: Kernel Choice, Size Distortions, and Power Losses." Working paper, Department of Economics, Harvard University.

[15] Lazarus, E., Lewis, D. J., Stock, J. H. and Watson, M. W. (2018): "HAR Inference: Recommendations for Practice." *Journal of Business & Economic Statistics* 36(4), 541–559.

[16] Liu, C. and Sun, Y. (2019): "Simple and Trustworthy Asymptotic t Tests in Difference-in-Differences Regressions." *Journal of Econometrics*, forthcoming.

[17] Müller, U. K. (2007): "A Theory of Robust Long-run Variance Estimation." *Journal of Econometrics* 141(2), 1331–1352.

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

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

[20] Stock, J. H., and Wright, J. H. (2000): "GMM with Weak Identification." *Econometrica* 68, 1055–1096.

[21] Sun, Y. (2011): "Robust Trend Inference with Series Variance Estimator and Testing-optimal Smoothing Parameter." *Journal of Econometrics* 164(2), 345–366.

[22] Sun, Y. (2013): "A Heteroskedasticity and Autocorrelation Robust F Test Using Orthonormal Series Variance Estimator." *Econometrics Journal* 16, 1–26.

[23] Sun, Y. (2014a): "Let's Fix It: Fixed-b Asymptotics versus Small-b Asymptotics in Heteroscedasticity and Autocorrelation Robust Inference." *Journal of Econometrics* 178(3), 659–677.

[24] Sun, Y. (2014b): "Fixed-smoothing Asymptotics in a Two-step GMM Framework." *Econometrica* 82(6), 2327–2370.

[25] Sun, Y. (2014c): "Fixed-smoothing Asymptotics and Asymptotic F and t Tests in the Presence of Strong Autocorrelation." *Advances in Econometrics* 33, 23–63.

[26] Sun, Y. and Phillips, P. C. B. (2009): "Bandwidth Choice for Interval Estimation in GMM Regression." Working paper, Department of Economics, UC San Diego.

[27] Sun, Y., Phillips, P. C. B., and Jin, S. (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing." *Econometrica* 76(1), 175–194.

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

[29] Sun, Y. and Kim, M. S. (2015): "Asymptotic F Test in a GMM Framework with Cross Sectional Dependence." *Review of Economics and Statistics* 97(1), 210–223.

[30] Tauchen, G. (1986a): "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data." *Journal of Business & Economic Statistics* 4, 397–425.

[31] Tauchen, G. (1986b): "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions." *Economic Letters* 20, 177–181.

[32] White, H. (2001): *Asymptotic Theory for Econometricians*, Academic Press, San Diego, CA.

[33] Zhang, X. (2016): "Fixed-smoothing Asymptotics in the Generalized Empirical Likelihood Estimation Framework." *Journal of Econometrics* 193, 123–146.