Econometric Theory, **34**, 2018, 790–814.
doi:10.1017/S026646661700024X

ON STANDARD INFERENCE FOR GMM WITH LOCAL IDENTIFICATION FAILURE OF KNOWN FORMS

JI HYUNG LEE
University of Illinois

ZHIPENG LIAO
UC Los Angeles

This paper studies the GMM estimation and inference problem that occurs when the Jacobian of the moment conditions is rank deficient of known forms at the true parameter values. Dovonon and Renault (2013) recently raised a local identification issue stemming from this type of degenerate Jacobian. The local identification issue leads to a slow rate of convergence of the GMM estimator and a nonstandard asymptotic distribution of the over-identification test statistics. We show that the known form of rank-deficient Jacobian matrix contains nontrivial information about the economic model. By exploiting such information in estimation, we provide GMM estimator and over-identification tests with standard properties. The main theory developed in this paper is applied to the estimation of and inference about the common conditionally heteroskedastic (CH) features in asset returns. The performances of the newly proposed GMM estimators and over-identification tests are investigated under the similar simulation designs used in Dovonon and Renault (2013).

1. INTRODUCTION

The generalized method of moments (GMM) is a popular method for empirical research in economics and finance. Under some regularity and weak dependence conditions, Hansen (1982) showed that the GMM estimator has standard properties, such as $T^{1/2}$-consistency and asymptotic normal distribution. The over-identification test (Sargan's J-test) statistic has an asymptotic Chi-square distribution. On the other hand, when some of the regularity conditions are not satisfied, the GMM estimator may have nonstandard properties. For example, when

We acknowledge useful comments from Don Andrews, Anil Bera, Xu Cheng, Denis Chetverikov, Yanqin Fan, Jonathan Gu, Jinyong Hahn, Guofang Huang, Roger Koenker, Ivana Komunjer, Rosa Matzkin, Peter Phillips, Shuyang Sheng, Ruoyao Shi, Yixiao Sun, and participants in Econometrics seminar at UC Davis, UCLA, UCSD, UIUC, Yale and 2014 Seattle-Vancouver Econometrics Conference. We also thank the Co-Editor, Anna Mikusheva, and two anonymous referees for constructive suggestions which improved this paper through the revision processes. The vast amount of outstanding editorial input by the Editor, Professor Phillips, on our last version of the manuscript is extremely appreciated. Any errors are the responsibility of the authors. Address correspondence to Ji Hyung Lee, Department of Economics, University of Illinois, 1407 W. Gregory Dr., 214 David Kinley Hall, Urbana, IL 61801, USA; e-mail: jihyung@illinois.edu.

the moment conditions only contain weak information, the GMM estimator may be inconsistent and have mixture normal asymptotic distribution (see, e.g., Staiger and Stock, 1997; Stock and Wright, 2000; Andrews and Cheng, 2012).

Dovonon and Renault (2013, hereafter DR) recently pointed out an interesting issue that occurs due to the violation of one regularity condition. When testing for common conditionally heteroskedastic (CH) features in asset returns, DR showed that the Jacobian of the moment conditions is a matrix of zeros at the true parameter value. This causes a slower than $T^{1/2}$ rate of convergence of the GMM estimator. A new limit theory was then developed to investigate the nonstandard asymptotic distribution of the J-test. When H and p are the number of moment conditions and parameters, respectively, the J-test was shown to have an asymptotic distribution that lies between $\chi^2(H-p)$ and $\chi^2(H)$. Their results extend the findings in Sargan (1983) and provide an important empirical caution - the commonly used critical values based on $\chi^2(H-p)$ lead to asymptotically oversized J-tests under the degeneracy of Jacobian moments.

This paper revisits the issue raised in DR and more general nonstandard inferential issues with a rank-deficient expected Jacobian matrix of known form. First, we consider moment functions for which the Jacobian of the moment conditions is known to be a matrix of zeros at the true parameter values due to the functional forms of the moment conditions. We provide alternative GMM-based estimation and inference using the zero Jacobian matrix as additional moment conditions. These additional moment restrictions contain extra information on the economic model. This information is exploited to achieve first-order local identification of the unknown structural parameters. We construct GMM estimators with $T^{1/2}$-consistency and asymptotic normality by adding the zero Jacobian as extra moment conditions. The J-test statistics based on the new set of moments are shown to have asymptotic Chi-square distributions. Other forms of singular expected Jacobian are also addressed and possible ways of recovering inference are discussed in supplementary material to this article, available at Cambridge Journals Online (journals.cambridge.org/ect).

We apply the newly developed theory to the main example—inference on the common feature in the common CH factor model. When using J-tests for the existence of the common feature in this model, DR suggests using the conservative critical values based on $\chi^2(H)$ to avoid the over-rejection issue. We show that, under the same sufficient conditions of DR, the common feature is not only first order locally identified, but also globally identified by the zero Jacobian moment conditions. As a result, our GMM estimators of the common feature have $T^{1/2}$-consistency and asymptotic normality. Our J-test statistic for the existence of the common feature has an asymptotic Chi-square distribution, which enables nonconservative asymptotic inference. Moreover, the Jacobian based GMM estimator of the common feature has a closed form expression, which makes it particularly well suited to empirical applications.

The rest of this paper is organized as follows. Section 2 describes the key idea of our methods in the general GMM framework. Section 3 applies the main results

from Section 2 to the common CH factor models. Section 4 contains simulation studies and Section 5 concludes. Theoretical extensions, proofs and supplementary simulation results are provided in the Online Appendix.

Standard notation is used. Throughout the paper, for any positive integer d, $\mathbb{R}^d$ denotes the real coordinate space of d dimensions; for any positive integer d and any $\alpha \in (0,1)$, $\chi^2_\alpha(d)$ denotes the α-th quantile of $\chi^2(d)$ where $\chi^2(d)$ denotes the chi-square random variable with the degree of freedom d; for any d-dimensional real vector μ_d and $d \times d$ symmetric positive definite real matrix Σ_d, $N(\mu_d, \Sigma_d)$ denotes the normal random vector with mean μ_d and variance-covariance matrix Σ_d; for any real matrix A, A' denotes its transpose and $\|A\|$ denotes its Frobenius norm; $\text{vec}(\cdot)$ denotes vectorization of a matrix; for any $d_1 \times 1$ vector function $f(x)$: $\mathbb{R}^{d_2} \to \mathbb{R}^{d_1}$ we use $\frac{\partial f(x)}{\partial x'}$ to denote the $d_1 \times d_2$ matrix whose i-th row and j-th column element is $\frac{\partial f_i(x)}{\partial x_j}$ where $f_i(\cdot)$ and x_j are the i-th component in $f(\cdot)$ and j-th component in x respectively; for any positive integers k_1 and k_2, I_{k_1} denotes the $k_1 \times k_1$ identity matrix, $\mathbf{0}_{k_1 \times k_2}$ denotes the $k_1 \times k_2$ zero matrix and $\mathbf{1}_{k_1 \times k_2}$ denotes the $k_1 \times k_2$ matrix of ones; $A \equiv B$ means that A is defined as B; $a_n = o_p(b_n)$ means that for any constants $\epsilon_1, \epsilon_2 > 0$, there is $\Pr(|a_n/b_n| \geq \epsilon_1) < \epsilon_2$ eventually; $a_n = O_p(b_n)$ means that for any $\epsilon > 0$, there is a finite constant C_ϵ such that $\Pr(|a_n/b_n| \geq C_\epsilon) < \epsilon$ eventually; "$\to_p$" and "$\to_d$" denote convergence in probability and convergence in distribution, respectively.

2. DEGENERATE JACOBIAN IN GMM MODELS

We are interested in estimating some parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$ which is uniquely identified by H ($H \geq p$) many moment conditions:

$$\psi(\theta_0) \equiv \mathbb{E}[\psi(X_t, \theta_0)] \equiv \mathbb{E}[\psi_t(\theta_0)] = \mathbf{0}_{H \times 1}, \tag{2.1}$$

where X_t is a random vector which is observable in period t. When we have data $\{X_t\}_{t=1}^T$, the GMM estimator $\widehat{\theta}_{\psi,T}$ based on the moment conditions (2.1) is defined as

$$\widehat{\theta}_{\psi,T} = \arg\min_{\theta \in \Theta} J_{\psi,T}(\theta), \tag{2.2}$$

where $J_{\psi,T}(\theta) \equiv T^{-1}\left[\sum_{t=1}^T \psi_t(\theta)\right]' W_{\psi,T}\left[\sum_{t=1}^T \psi_t(\theta)\right]$ and $W_{\psi,T}$ is an $H \times H$ weight matrix. Let $\widehat{\theta}^*_{\psi,T}$ be the GMM estimator in (2.2) with weight matrix $W_{\psi,T} = \hat{\Omega}^{-1}_{\psi,T}$, where $\hat{\Omega}_{\psi,T}$ is a consistent estimator of the asymptotic variance of $T^{-1/2}\sum_{t=1}^T \psi_t(\theta_0)$. The J-test statistic defined as $J_{\psi,T} \equiv J_{\psi,T}(\widehat{\theta}^*_{\psi,T})$ with $W_{\psi,T} = \hat{\Omega}^{-1}_{\psi,T}$ is commonly used to test the validity of the moment conditions (2.1).

As illustrated in Hansen (1982), global identification together with other regularity conditions can be used to show standard properties of the GMM estimator $\widehat{\theta}_{\psi,T}$ and the J-test statistic $J_{\psi,T}$. For any $\theta \in \Theta$, define

$$\Gamma(\theta) = \frac{\partial}{\partial \theta'} \mathbb{E}[\psi_t(\theta)].$$

Then $\Gamma(\theta)$ is an $H \times p$ matrix of functions. The standard properties of the GMM estimator $\widehat{\theta}_{\psi,T}$, such as $T^{1/2}$-consistency and asymptotic normality, rely on the condition that $\Gamma(\theta_0)$ has full rank. When $\Gamma(\theta_0) = \mathbf{0}_{H \times p}$, GMM inference is nonstandard; the convergence rate is slower than $T^{1/2}$ and the J-test statistic $J_{\psi,T}$ has a mixture of asymptotic Chi-square distributions. DR have established these nonstandard properties of GMM inference when $\Gamma(\theta_0) = \mathbf{0}_{H \times p}$, in the context of testing for the common feature in common CH factor models. In this section, we discuss the same issue and propose an alternative solution in a general GMM context.

Define $g(\theta) \equiv vec(\Gamma(\theta)')$, then $\Gamma(\theta_0) = \mathbf{0}_{H \times p}$ implies that $g(\theta_0) = \mathbf{0}_{pH \times 1}$. The zero Jacobian matrix provides pH many extra moment conditions:

$$g(\theta) = \mathbf{0}_{pH \times 1} \text{ when } \theta = \theta_0. \tag{2.3}$$

The new set of moment restrictions ensures the first order local identification of θ_0, when the Jacobian of $g(\theta)$ (or essentially the Hessian of $\psi(\theta)$) evaluated at θ_0 has full column rank. We define the corresponding Jacobian matrix of $g(\theta)$ as

$$\mathbb{H}(\theta) \equiv \frac{\partial}{\partial \theta'} g(\theta),$$

where $\mathbb{H}(\theta)$ is now a $pH \times p$ matrix of functions. When $\mathbb{H} \equiv \mathbb{H}(\theta_0)$ has full column rank the first order local identification of θ_0 could be achieved, which makes it possible to construct GMM estimators and J-tests with standard properties. We next provide a lemma that enables checking the rank condition of the moment conditions based on the Jacobian matrix.

LEMMA 2.1. *Let $\psi_h(\theta)$ be the h-th $(h = 1, \ldots, H)$ component function in $\psi(\theta)$. Suppose that (i) θ_0 belongs to the interior of Θ; and (ii) for any $\theta \in \Theta$,*

$$\left((\theta - \theta_0)' \left(\frac{\partial^2 \psi_h}{\partial \theta \partial \theta'} (\theta_0) \right) (\theta - \theta_0) \right)_{1 \le h \le H} = \mathbf{0}_{H \times 1} \tag{2.4}$$

if and only if $\theta = \theta_0$. Then the matrix $\mathbb{H}$ has full rank.

Lemma 2.1 shows that when the moment conditions in (2.3) are used under the condition (ii), the first order local identification of θ_0 is achieved. The moment conditions in (2.3) alone may not ensure the global/unique identification of θ_0. However, as θ_0 is globally (uniquely) identified by the moment conditions in (2.1), we can use the moment conditions in (2.1) and (2.3) in GMM to ensure both the global identification and the first order local identification of θ_0.

Remark 2.1. Condition (ii) in Lemma 2.1 is the second order local identification condition of θ_0 based on the moment conditions in (2.1). This condition is derived as a general result in DR (see, their Lemma 2.3), and is used as a

high-level sufficient assumption in Dovonon and Gonçalves (2016, hereafter DG) to justify the validity of bootstrap inference procedures. When this condition is violated, the GMM estimator additionally using moment conditions from zero Jacobian may not have the $T^{1/2}$-asymptotic normal distribution. Moreover, when this condition is close to being violated (i.e., *weak local identification*), the finite sample performance of the new GMM estimators and standard inference procedures proposed below may not closely follow the asymptotic analysis in the paper (see, e.g., simulation results under D1.1 and D2.1 in Section 4).[1]

Let $\Gamma_t(\theta) = \partial\psi_t(\theta)/\partial\theta'$, $g_t(\theta) = vec(\Gamma_t(\theta)')$ and $m_t(\theta) = (\psi_t'(\theta), g_t'(\theta))'$. We define the GMM estimator of θ_0 using all moment conditions as

$$\widehat{\theta}_{m,T} = \arg\min_{\theta\in\Theta} J_{m,T}(\theta), \tag{2.5}$$

where $J_{m,T}(\theta) \equiv T^{-1}\left[\sum_{t=1}^T m_t(\theta)\right]' W_{m,T}\left[\sum_{t=1}^T m_t(\theta)\right]$ and $W_{m,T}$ is an $(Hp+H)\times(Hp+H)$ weight matrix. Similarly, we define the GMM estimator of θ_0 using only the moment conditions in (2.3) as

$$\widehat{\theta}_{g,T} = \arg\min_{\theta\in\Theta} J_{g,T}(\theta), \tag{2.6}$$

where $J_{g,T}(\theta) \equiv T^{-1}\left[\sum_{t=1}^T g_t(\theta)\right]' W_{g,T}\left[\sum_{t=1}^T g_t(\theta)\right]$ and $W_{g,T}$ is an $Hp\times Hp$ weight matrix.

Assumption 2.1. (i) The Central Limit Theorem (CLT) holds: $T^{-1/2}\sum_{t=1}^T m_t(\theta_0) \to_d N(0,\Omega_m)$ where Ω_m is a positive definite matrix with

$$\Omega_m = \begin{pmatrix} \Omega_\psi & \Omega_{\psi g} \\ \Omega_{g\psi} & \Omega_g \end{pmatrix},$$

where Ω_ψ is an $H\times H$ matrix and Ω_g is a $pH\times pH$ matrix; (ii) $W_{m,T} \to_p \Omega_m^{-1}$ and $W_{g,T} \to_p W_g$, where W_g is a symmetric, positive definite matrix.

We next state the asymptotic distributions of the GMM estimators $\widehat{\theta}_{m,T}$ and $\widehat{\theta}_{g,T}$.

PROPOSITION 2.1. *Under the conditions of Lemma 2.1, Assumptions 2.1 and B.1 in the Online Appendix, we have (i)* $T^{1/2}(\widehat{\theta}_{m,T}-\theta_0) \to_d N(0,\Sigma_{\theta,m})$, *where* $\Sigma_{\theta,m}^{-1} = \mathbb{H}'(\Omega_g - \Omega_{g\psi}\Omega_\psi^{-1}\Omega_{\psi g})^{-1}\mathbb{H}$; *(ii) moreover if* θ_0 *is uniquely identified by (2.3), then* $T^{1/2}(\widehat{\theta}_{g,T}-\theta_0) \to_d N(0,\Sigma_{\theta,g})$, *where* $\Sigma_{\theta,g} = (\mathbb{H}'W_g\mathbb{H})^{-1}\mathbb{H}'W_g\Omega_g W_g\mathbb{H}(\mathbb{H}'W_g\mathbb{H})^{-1}$.

In the linear IV models, Condition (2.4) does not hold under the zero Jacobian, and hence θ_0 is not identified. Phillips (1989) and Choi and Phillips (1992) showed that if the Jacobian has insufficient rank but is not entirely zero in the linear IV models, some part of θ_0 (after some rotation) is identified and $T^{1/2}$-estimable. Proposition 2.1 shows that in some nonlinear models, the degenerate Jacobian

together with Condition (2.4) and other regularity conditions can be used to derive a $T^{1/2}$-consistent GMM estimator of θ_0. Our results therefore supplement the earlier findings in Phillips (1989) and Choi and Phillips (1992).

Remark 2.2. When $W_g = \Omega_g^{-1}$, we have $\Sigma_{\theta,g}^{-1} = \mathbb{H}'\Omega_g^{-1}\mathbb{H} \leq \Sigma_{\theta,m}^{-1}$, which implies that $\widehat{\theta}_{m,T}$ is preferred to $\widehat{\theta}_{g,T}$ from the asymptotic efficiency perspective. In some examples (e.g., the common CH factor model in the next section), the computation of $\widehat{\theta}_{g,T}$ is much easier than $\widehat{\theta}_{m,T}$ as the former has a closed-form expression. We next propose an estimator which is as efficient as $\widehat{\theta}_{m,T}$, but can be computed similarly to $\widehat{\theta}_{g,T}$. This new procedure shares some similar features with the principle of control variables (see, e.g., Fieller and Hartley, 1954; Kleibergen, 2005; Antoine, Bonnal, and Renault, 2007).[2]

Let $\hat{\Omega}_{g,T}$, $\hat{\Omega}_{\psi g,T}$ and $\hat{\Omega}_{\psi,T}$ be the consistent estimators of Ω_g, $\Omega_{g\psi}$, and Ω_ψ respectively. These variance matrix estimators can be constructed using $\widehat{\theta}_{g,T}$ with an identity weight matrix, for example. The new GMM estimator is defined as

$$\widehat{\theta}_{g^*,T} = \underset{\theta\in\Theta}{\arg\min} T^{-1}\left[\sum_{t=1}^{T}\widehat{g}_{\psi,t}(\theta)\right]' W_{g^*,T}\left[\sum_{t=1}^{T}\widehat{g}_{\psi,t}(\theta)\right], \tag{2.7}$$

where $\widehat{g}_{\psi,t}(\theta) = g_t(\theta) - \hat{\Omega}_{g\psi,T}\hat{\Omega}_{\psi,T}^{-1}\psi_t(\widehat{\theta}_{g,T})$ and $W_{g^*,T}^{-1} = \hat{\Omega}_{g,T} - \hat{\Omega}_{g\psi,T}\hat{\Omega}_{\psi,T}^{-1}\hat{\Omega}_{\psi g,T}$.

THEOREM 2.1. *Under the conditions of Proposition 2.1, we have* $T^{1/2}(\widehat{\theta}_{g^*,T} - \theta_0) \to_d N(0, \Sigma_{\theta,m})$, *where* $\Sigma_{\theta,m}$ *is defined in Proposition 2.1.*

From Theorem 2.1, $\widehat{\theta}_{g^*,T}$ has the same asymptotic variance as $\widehat{\theta}_{m,T}$. Moreover, it is essentially computed based on the moment conditions (2.3), hence it enhances computational simplicity whenever $\widehat{\theta}_{g,T}$ is $T^{1/2}$-consistent and is easy to calculate.

Remark 2.3. Standard inference on θ_0 can be conducted using the asymptotic normal distributions of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ or $\widehat{\theta}_{g^*,T}$. The inference based on $\widehat{\theta}_{m,T}$ or $\widehat{\theta}_{g^*,T}$ may be better than the inference based on $\widehat{\theta}_{g,T}$, as the latter has larger asymptotic variance. By contrast, the asymptotic distribution of $\widehat{\theta}_{\psi,T}$ is nonstandard, which makes the inference based on $\widehat{\theta}_{\psi,T}$ difficult to use in practice. Moreover, the convergence rates of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ are $T^{-1/2}$ which is faster than $T^{-1/4}$, the convergence rate of $\widehat{\theta}_{\psi,T}$. Hence the inference based on $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$, or $\widehat{\theta}_{g^*,T}$ is not only easy to use but also better (in both size and power) from the theoretical perspective.

When GMM estimators have standard asymptotic properties, it is straightforward to construct the over-identification test statistics and show their asymptotic distributions. As the model specification implies both the moment conditions in (2.1) and (2.3), one can jointly test their validity using the following standard result:[3]

$$J_{m,T} \equiv T^{-1}\left[\sum_{t=1}^{T} m_t(\widehat{\theta}_{m,T})\right]' \hat{\Omega}_{m,T}^{-1}\left[\sum_{t=1}^{T} m_t(\widehat{\theta}_{m,T})\right] \to_d \chi^2(Hp+H-p). \quad \textbf{(2.8)}$$

When θ_0 is identified by (2.3), it may be convenient to use the J-test based on (2.6) in practice:

$$J_{g,T} \equiv T^{-1}\left[\sum_{t=1}^{T} g_t(\widehat{\theta}^*_{g,T})\right]' \hat{\Omega}_{g,T}^{-1}\left[\sum_{t=1}^{T} g_t(\widehat{\theta}^*_{g,T})\right] \to_d \chi^2(Hp-p), \quad \textbf{(2.9)}$$

where $\widehat{\theta}^*_{g,T}$ denotes the GMM estimator defined in (2.6) with weight matrix $W_{g,T} = \hat{\Omega}_{g,T}^{-1}$. One interesting aspect of the proposed J-test in (2.8) is that it has the standard degrees of freedom, i.e., the number of moment conditions used in estimation minus the number of parameters we estimate. Among the $H(p+1)$ many moment restrictions, H moments from (2.1) have degenerate Jacobian matrix. By combining these H moments with the extra information provided by the Hp Jacobian moments, we avoid the issue of rank deficiency. Stacking the additional moments from (2.3) provides enough sensitivity of the J-test statistic to parameter variation. As a result, the standard degree of freedom shows up in the asymptotic Chi-square distribution in (2.8).

Without incurring greater computation costs, we prefer testing more valid moment restrictions under the null. For this purpose, one can use the following test statistics:

$$J_{h,T} \equiv T^{-1}\left[\sum_{t=1}^{T} m_t(\widehat{\theta}_{g,T})\right]' W_{h,T}\left[\sum_{t=1}^{T} m_t(\widehat{\theta}_{g,T})\right], \quad \textbf{(2.10)}$$

where $W_{h,T}$ is an $(Hp+H)\times(Hp+H)$ real matrix. Let $m(\theta) = E[m_t(\theta)]$ for any $\theta \in \Theta$. Theorem 2.2 below provides the asymptotic distribution of $J_{h,T}$.

Remark 2.4. The J-test $J_{g,T}$ is expected to have power if a violation of the economic theory implies the invalidity of the moment conditions (2.3) (including the CH factor model studied in the next section). The J-tests $J_{m,T}$ and $J_{h,T}$ test not only the moment conditions in (2.1) but also those in (2.3). Including more moment conditions in the J-test may not necessarily increase or decrease the power of the test.[4] For example, suppose that the null hypothesis of an economic model implies both (2.1) and (2.3). If the moment conditions (2.1) hold under both null and alternative (i.e., only the moment conditions in (2.3) are violated under the alternative hypothesis), then the J-test based on (2.1) will have no power although it has fewer moment conditions than $J_{m,T}$. Ideally, one wants to include only the most misspecified moment conditions in J-test for the power consideration. The degree of misspecification of different moment conditions is, however, unknown in reality. To avoid the potential power loss from testing only mildly misspecified moment conditions, one may want to include all valid moment conditions implied by the null hypothesis in the J-test.

THEOREM 2.2. *Suppose that the conditions of Proposition 2.1 hold and $W_{h,T} \to_p W_h$ where W_h is a nonrandom real matrix. Then we have*

$$J_{h,T} \to_d B'_{H(p+1)} \Omega_m^{\frac{1}{2}} \mathbb{P}' W_h \mathbb{P} \Omega_m^{\frac{1}{2}} B_{H(p+1)},$$

where $\mathbb{P} \equiv I_{H(p+1)} - \frac{\partial m(\theta_0)}{\partial \theta'} (\mathbf{0}_{p\times H}, (\mathbb{H}' W_g \mathbb{H})^{-1} \mathbb{H}' W_g)$ and $B_{H(p+1)}$ denotes an $H(p+1) \times 1$ standard normal random vector.

Remark 2.5. Theorem 2.2 indicates that the asymptotic distribution of $J_{h,T}$ is not pivotal. However critical values of its asymptotic distribution are easy to simulate in practice.[5] Unlike the bootstrap procedures in DG, one does not need to solve the optimization problem of GMM estimation repeatedly to get the simulated critical values. The matrices W_h and W_g in the asymptotic distribution of $J_{h,T}$ depend on the weight matrices $W_{h,T}$ and $W_{g,T}$ respectively. We suggest use of identity matrices for both weight matrices to construct $J_{h,T}$. With such a choice, there are fewer unknown nuisance parameters (only Ω_m, $\mathbb{H}$, and $\partial m(\theta_0)/\partial \theta'$) in the asymptotic distribution of $J_{h,T}$ to be estimated. This may reduce the estimation error in the simulated critical value. The other choices of $W_{h,T}$ and $W_{g,T}$, for example $\hat{\Omega}_{m,T}^{-1}$ and $\hat{\Omega}_{g,T}^{-1}$, will not make the asymptotic distribution of $J_{h,T}$ pivotal and may introduce additional errors to the simulated critical value. Moreover, using the identity matrix for $W_{g,T}$ implies that $\widehat{\theta}_{g,T}$ in (2.10) is the one-step GMM estimator based on the identity weight matrix, which makes the construction of $J_{h,T}$ easy in practice.

Remark 2.6. From the asymptotic analysis in DR, the asymptotic distribution of $J_{\psi,T}$ is bounded by $\chi^2(H-p)$ from below and by $\chi^2(H)$ from above. Let C_1 and C_2 denote the tests based on $J_{\psi,T}$ with critical values $\chi^2_{1-\alpha}(H-p)$ and $\chi^2_{1-\alpha}(H)$, respectively. We would expect: (i) when the sample size is large, the empirical rejection probabilities of the test C_1(or C_2) will be larger(or smaller) than the nominal size α; (ii) when H is large and p is small, the critical values $\chi^2_{1-\alpha}(H-p)$ and $\chi^2_{1-\alpha}(H)$ are close to each other (in relative magnitude) which implies that the empirical rejection probabilities of the tests C_1 and C_2 will be close; (iii) when the sample size is large with large H and small p, the empirical rejection probabilities of the tests C_1 and C_2 should be close to α. Compared with the tests C_1 and C_2, the standard inference based on $J_{m,T}$, $J_{g,T}$ or $J_{h,T}$ is asymptotically exact so that their empirical rejection probabilities converge to the nominal size. Moreover, as the lower bound of the convergence rate of $\widehat{\theta}_{\psi,T}$ is $T^{-1/4}$, the empirical rejection probabilities of the tests C_1 and C_2 converge slowly to their asymptotic values.[6] It is worth mentioning, however, that the conservative test C_2 proposed in DR does not explicitly use the known information of zero Jacobian. Hence the test remains under-sized when the known information is misspecified, for example, when the Jacobian has full rank. Their test is robust from this perspective.

Remark 2.7. Compared with the bootstrap procedures in DG, the standard inference procedures provided in this paper have the following advantages. First,

the leading error terms in the asymptotic approximation of the J-test statistics $J_{m,T}$, $J_{g,T}$, and $J_{h,T}$ converge to zero at the $T^{-1/2}$ rate, while the corresponding term in $J_{\psi,T}$ converges at a rate not faster than $T^{-1/4}$. As the bootstrap critical values in DG approximate the critical values of the asymptotic distribution of $J_{\psi,T}$, the empirical rejection probabilities of the bootstrap inference procedures converge to the nominal size slower than the standard inference based on $J_{m,T}$, $J_{g,T}$, and $J_{h,T}$. Second, the inference based on $J_{m,T}$, $J_{g,T}$, and $J_{h,T}$ does not need to solve the optimization problem in the GMM estimation repeatedly, so are simpler to use in practice. One potential drawback of our inference procedures is that they invoke many moment conditions in GMM estimation and inference. The asymptotic results established in this section assume that both H and p are fixed, hence they may not capture the finite sample distributions of the GMM estimators and J-test statistics well when Hp is large. When there are many moment conditions, estimation methods such as the continuous updated GMM are preferred as they produce GMM estimators with smaller bias (see Newey and Smith, 2004), and the inference procedures which take the (second-order) bias and variance contributed by many moment conditions are preferred (see Newey and Windmeijer, 2009).

Remark 2.8. Under the conditions of Proposition 2.1 and the null hypothesis that the moment conditions in (2.1) and (2.3) hold, one can show that

$$J_{\widehat{g},T}(\widehat{\theta}_{g^*,T}) \to_d \chi^2(Hp-p), \tag{2.11}$$

where $J_{\widehat{g},T}(\widehat{\theta}_{g^*,T}) \equiv T^{-1}\big[\sum_{t=1}^T \widehat{g}_{\psi,t}(\widehat{\theta}_{g^*,T})\big]' W_{g^*,T}\big[\sum_{t=1}^T \widehat{g}_{\psi,t}(\widehat{\theta}_{g^*,T})\big]$. The over-identification test based on $J_{\widehat{g},T}(\widehat{\theta}_{g^*,T})$ and its asymptotic distribution in (2.11) controls the (asymptotic) size. The power of this test requires careful investigation. Consider the local misspecification case for illustration. In such a case, the moment conditions (2.1) and (2.3) become

$$\mathbb{E}[\psi_t(\theta_0)] = d_\psi n^{-1/2} \text{ and } \mathbb{E}[g_t(\theta_0)] = d_g n^{-1/2} \tag{2.12}$$

respectively, where $d_\psi \in \mathbb{R}^H$ and $d_g \in \mathbb{R}^{Hp}$ are real vectors. Suppose that θ_0 is asymptotically uniquely identified by the second set of the moment conditions in (2.12) and Condition (ii) in Lemma 2.1 holds. Then both $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ are still $T^{1/2}$-consistent estimators of θ_0. Let $g_{\psi,t} = g_t(\theta_0) - \Omega_{g\psi}\Omega_\psi^{-1}\psi_t(\theta_0)$. It is easy to show that $J_{\widehat{g},T}(\widehat{\theta}_{g^*,T}) \to_d \left(B_{g_\psi} + d_{g_\psi}\right)' \Xi \left(B_{g_\psi} + d_{g_\psi}\right)$, where $\Xi = I_{HP} - \Sigma_{\theta,m}^{-1/2}\mathbb{H}(\mathbb{H}'\Sigma_{\theta,m}^{-1}\mathbb{H})^{-1}\mathbb{H}'\Sigma_{\theta,m}^{-1/2}$, $d_{g_\psi} = \Sigma_{\theta,m}^{-1/2}(d_g - \Omega_{g\psi}\Omega_\psi^{-1}d_\psi)$ and B_{g_ψ} is an HP dimensional standard normal random vector. The test based on $J_{\widehat{g},T}(\widehat{\theta}_{g^*,T})$ has nontrivial power as long as $d_{g_\psi} \neq 0$ and d_{g_ψ} does not lie in the eigenspace of the zero eigenvalues of Ξ. However, this test may have no (or low) power when $d_{g_\psi} = 0$ (or d_{g_ψ} is close to zero). Note that d_{g_ψ} can be zero or close to zero even if d_g and d_ψ are far away from zero. On the other hand, the J-test based on $J_{m,T}$ or $J_{h,T}$ has nontrivial power for nonzero d_g and d_ψ in general. To avoid the power loss of $J_{\widehat{g},T}(\widehat{\theta}_{g^*,T})$ with small d_{g_ψ} but large d_g and/or d_ψ, we do not recommend this test in practice.

Remark 2.9. In practice, one may want to test whether the Jacobian is zero or not. For such a problem, the null and alternative hypotheses are:

$$H_0 : \mathbb{E}[g_t(\theta_0)] = 0 \text{ v.s. } H_1 : \mathbb{E}[g_t(\theta_0)] \neq 0, \tag{2.13}$$

where $\mathbb{E}[\psi_t(\theta_0)] = 0$ is maintained under both the null and alternative hypotheses. Several tests are available in the literature. First, one can use the J-test statistic $J_{m,T}$ based on the stacked moment conditions $\mathbb{E}[m_t(\theta_0)] = 0$. Second, one can also use the quasi-likelihood ratio (QLR) statistic (see, e.g., Eichenbaum, Hansen, and Singleton, 1988) which is the difference of two J-test statistics: one is $J_{m,T}$ and the other is based on $\mathbb{E}[\psi_t(\theta_0)] = 0$. As explained in Hall (2005), the QLR statistic has the asymptotic $\chi^2(Hp)$ distribution under the standard regularity conditions, potentially providing a more powerful test than the J-test statistic $J_{m,T}$. However, the GMM estimator based on $\mathbb{E}[\psi_t(\theta_0)] = 0$ lacks the local identification under the null hypothesis, so the asymptotic distribution of the QLR statistic becomes nonstandard. On the other hand, the J-test statistic $J_{m,T}$ has the standard chi-square limiting distribution under the null hypothesis and hence is more convenient for testing (2.13).

Remark 2.10. One may be interested in the effect of pre-testing the validity of moment conditions (using $J_{m,T}$, $J_{g,T}$ or $J_{h,T}$) on the subsequent GMM inference of the structural parameter θ_0. As the pseudo true value identified by misspecified moment conditions is in general different from θ_0, the GMM inference based on the asymptotic distributions of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ or $\widehat{\theta}_{g^*,T}$ in Proposition 2.1 and Theorem 2.1 will perform poorly when the J-test commits the type-2 error. On the other hand, the GMM inference of θ_0 using Proposition 2.1 and Theorem 2.1 are valid when the moment conditions in (2.1) and (2.3) are correctly specified. As a result, the J-test which controls size and has better power is preferred because it enables the subsequent GMM inference on θ_0 to suffer less from the pre-test bias. The J-tests based on $J_{m,T}$, $J_{g,T}$ and $J_{h,T}$ are attractive due to their asymptotic exact size control. On the other hand, the asymptotic Chi-square and weighted Chi-square distributions may be poor approximations of the finite sample distributions of $J_{m,T}$, $J_{g,T}$ and $J_{h,T}$ when there are many moment conditions and/or there is a weak local identification problem. In these scenarios, alternative asymptotic theory (for $J_{m,T}$, $J_{g,T}$ and $J_{h,T}$) which captures the effects of the number of moment conditions and/or the weak local identification is needed for better size control in finite samples.

3. APPLICATION TO COMMON CH FACTOR MODEL

Multivariate volatility models commonly assume a smaller number of conditionally heteroskedastic (CH) factors than the number of assets. Thus, a small number of common factors may generate CH behavior of many assets, a phenomenon that may be motivated by economic theories (see, e.g., the early discussion in Engle, Ng, and Rothschild, 1990). Moreover, without imposing a common factor structure, there may be an overwhelming number of parameters to be estimated in

multivariate volatility models. From these theoretical and empirical perspectives, common CH factor models are preferred and widely used. Popular examples include the factor Gaussian generalized autoregressive conditional heteroskedastic (GARCH) models (see, e.g., Section 2.2 of Silvennoinen and Teräsvirta, 2009) and factor stochastic volatility models (see, e.g., Section 2.2 of Broto and Ruiz, 2004; and references therein). The relevant common CH factor model literature includes Diebold and Nerlove (1989), Engle et al. (1990), Fiorentini, Sentana and Shephard (2004), Doz and Renault (2006), and Dovonon and Renault (2013). It is therefore important to test whether a common CH factor structure exists in multivariate asset returns of interest. This test will serve as the main application of our results from Section 2. See DG and Sentana (2015) for other possible applications.

Engle and Kozicki (1993) proposed to detect the existence of common CH factor structure, or equivalently the common CH features, using the GMM overidentification test (Hansen, 1982). Consider an n-dimensional vector of asset returns $Y_{t+1} = (Y_{1,t+1}, \ldots, Y_{n,t+1})'$ which satisfies

$$\mathrm{Var}\,(Y_{t+1}|\mathcal{F}_t) = \Lambda D_t \Lambda' + \Omega, \tag{3.1}$$

where $\mathrm{Var}(\cdot|\mathcal{F}_t)$ denotes the conditional variance given all available information $\mathcal{F}_t$ at period t, $\Lambda = (\lambda_1, \ldots, \lambda_p)$ is an $n \times p$ matrix ($p \leq n$), D_t is a $p \times p$ diagonal matrix whose diagonal entry is $\sigma^2_{i,t}$ for $i = 1, \ldots, p$, Ω is an $n \times n$ positive definite matrix and $\{\mathcal{F}_t\}_{t\geq 0}$ is the increasing filtration to which $\{Y_t\}_{t\geq 0}$ and $\{\sigma^2_{k,t}\}_{1\leq k\leq p, t\geq 0}$ are adapted. Let $Diag\,(D_t) = \left(\sigma^2_{1,t}, \ldots, \sigma^2_{p,t}\right)'$. The Assumptions 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6 below are from DR.

Assumption 3.1. $Rank(\Lambda) = p$ and $Var[Diag\,(D_t)]$ is nonsingular.

Assumption 3.2. $\mathbb{E}\left[Y_{t+1}|\mathcal{F}_t\right] = 0$.

Assumption 3.3. We have H many $\mathcal{F}_t$-measurable random variables z_t such that: (i) $Var\,(z_t)$ is nonsingular, and (ii) $Rank[Cov\,(z_t, Diag\,(D_t))] = p$.

Assumption 3.4. The process (z_t, Y_t) is stationary and ergodic with $\mathbb{E}[\|z_t\|^2] < \infty$ and $\mathbb{E}[\|Y_t\|^4] < \infty$. We also assume a weak dependent structure so that $\left(z_t, vec\left(Y_t Y_t'\right)\right)$ fulfill a central limit theorem.

When $p < n$, e.g., $p = n - 1$, there exists a nonzero $\theta^0_* \in \mathbb{R}^n$ such that $\theta^{0\prime}_* \Lambda = \mathbf{0}_{1\times p}$. The real vector θ^0_* is called the common CH feature in the literature. In the presence of the common CH feature, we have

$$\mathrm{Var}\left(\theta^{0\prime}_* Y_{t+1}|\mathcal{F}_t\right) = \theta^{0\prime}_* \Lambda D_t \Lambda' \theta^0_* + \theta^{0\prime}_* \Omega \theta^0_* = \theta^{0\prime}_* \Omega \theta^0_* = \text{Constant}. \tag{3.2}$$

Note the CH effects are nullified in the linear combination $\theta^{0\prime}_* Y_{t+1}$, while the individual return $Y_{i,t+1}$'s ($i = 1, \ldots, n$) are showing CH volatility. The equations in (3.2) lead to the following moment conditions

$$\mathbb{E}\left[(z_t - \mu_z)\left(\theta'_* Y_{t+1} Y'_{t+1} \theta_*\right)\right] = \mathbf{0}_{H\times 1} \text{ when } \theta_* = \theta^0_* \in \mathbb{R}^n \text{ and } \theta^0_* \neq \mathbf{0}_{n\times 1}, \tag{3.3}$$

where μ_z denotes the population mean of z_t.[7] Given the restrictions in (3.3), one can use GMM to estimate the common feature θ_*^0 and conduct inference about the validity of the moment conditions.

DR have shown that GMM inference using (3.3) is subject to the issue of zero Jacobian moments. The GMM estimator based on (3.3) can therefore be as slow as $T^{-1/4}$ with a nonstandard limiting distribution. As explained earlier, the J-test based on (3.3) has an asymptotic mixture of two different chi-square distributions, $\chi^2(H-p)$ and $\chi^2(H)$. Following DR's empirical suggestion—using critical values based on $\chi^2(H)$ rather than $\chi^2(H-p)$—provides conservative size control. We show that it is possible to construct $T^{1/2}$-consistent and asymptotically normally distributed GMM estimators and nonconservative J-tests by applying theory developed in Section 2 to this common CH factor model.

Following DR, we assume that exclusion restrictions characterize a set $\Theta_* \subset \mathbb{R}^n$ that contains at most one unknown common feature θ_*^0 up to a normalization condition denoted by $\mathcal{N}$.

Assumption 3.5. We have $\theta_* \in \Theta_* \subset \mathbb{R}^n$ such that $\Theta^* = \Theta_* \cap \mathcal{N}$ is a compact set and

$$\left(\theta_* \in \Theta^* \text{ and } \theta_*'\Lambda = \mathbf{0}_{1\times p}\right) \Leftrightarrow \left(\theta_* = \theta_*^0\right). \tag{3.4}$$

The nonzero restriction on θ_*^0 could be imposed in several ways. For example, the unit cost condition $\mathcal{N} = \left\{\theta \in \mathbb{R}^n, \sum_{i=1}^n \theta_i = 1\right\}$ can be maintained without loss of generality.[8] To implement this restriction, we define an $n \times (n-1)$ matrix G_2 as $G_2 = \left(\mathbf{I}_{(n-1)}, -\mathbf{1}_{(n-1)\times 1}\right)'$. Then for any $\theta_* \in \Theta^*$,

$$\theta_* = \left(\theta_1, \ldots, \theta_{n-1}, 1 - \sum_{i=1}^{n-1}\theta_i\right)' = \left(\theta', 1 - \sum_{i=1}^{n-1}\theta_i\right)' = G_2\theta + l_n \tag{3.5}$$

where $\theta = (\theta_1, \ldots, \theta_{n-1})'$ is an $(n-1)$-dimensional real vector and $l_n = (0, \ldots, 0, 1)'$ is an $n \times 1$ vector. Hence, we can write

$$\Theta^* = \{\theta_* : \theta_* = G_2\theta + l_n, \forall \theta \in \Theta\}, \tag{3.6}$$

where Θ is a nondegenerate subspace of $\mathbb{R}^{n-1}$.

Remark 3.1. Under the unit cost condition (3.6), the identification (3.4) requires that $\Lambda' G_2 \theta_0 = -\Lambda' l_n$. The sufficient and necessary condition for the existence of solution θ_0 in equation $\Lambda' G_2 \theta_0 = -\Lambda' l_n$ is that the matrices $\Lambda' G_2$ and $\left(\Lambda' G_2, -\Lambda' l_n\right)$ have the same rank. Moreover, when the solution θ_0 exists, it is unique when the matrix $\Lambda' G_2$ has rank $n-1$. These findings together with the special structure of G_2 and l_n imply the following lemma.

LEMMA 3.1. *Suppose that Assumptions 3.1, 3.2 and 3.3 hold. Then*: (*i*) *the moment conditions* (*3.3*) *hold if and only if there is a* θ_*^0 *such that* $\theta_*^{0\prime}\Lambda = \mathbf{0}_{1\times p}$; (*ii*) *there exists a* θ_*^0 *such that* $\theta_*^{0\prime}\Lambda = \mathbf{0}_{1\times p}$ *if and only if* $rank(\Lambda' G_2) = p$; (*iii*) *the identification* (*3.4*) *holds if and only if* $\Lambda' G_2$ *is invertible and in such a case, we have*

$$\theta_*^0 = \left(\theta_0', 1 - \sum_{i=1}^{n-1} \theta_{0,i}\right)' = G_2\theta_0 + l_n,$$

where $\theta_0 = -(\Lambda' G_2)^{-1}\Lambda' l_n$.

Remark 3.2. Lemma 3.1 is useful for understanding the properties of the over-identification test when it is used for testing the existence of common features. Essentially, the over-identification test based on the moment conditions (3.3) enables one to test the following hypotheses

$$\begin{aligned} &H_0: \text{ there exists a unique common feature in } \Theta^*, \text{ v.s.} \\ &H_1: \text{ no common feature in } \Theta^* \text{ exists among the assets } Y_{t+1}. \end{aligned} \tag{3.7}$$

By Lemma 3.1(i), we know that under the null, the moment conditions in (3.3) are valid. Lemma 3.1(iii) further captures the identified common feature under the null. Under the alternative, by Lemma 3.1(i), we know that the moment conditions in (3.3) are invalid.

Remark 3.3. It is possible that the true data generating process (DGP) is not nested by either the null or the alternative hypothesis in (3.7). There are two cases to consider: (i) the unit cost condition is misspecified; (ii) there are multiple common features among the assets Y_{t+1}. When the unit cost condition is misspecified, there is no $\theta_*^0 \in \Theta^*$ such that $\theta_*^{0\prime}\Lambda = \mathbf{0}_{1\times p}$ which together with Lemma 3.1(i) implies that the moment conditions (3.3) are invalid and the over-identification test may lead to rejection even if a common feature exists. Next, suppose that the unit cost conditions is correctly specified, but there are multiple common features among Y_t, i.e., $p < n-1$. Then there are multiple θ_*^0s in Θ^* such that the moment condition (3.3) holds. In this case, the common feature is not uniquely identified. The asymptotic distribution of the over-identification test in such a case is difficult to study and is left for future research.[9]

Remark 3.4. One possible way of testing the existence of the common features with $p < n-1$ is using only (the first) $p+1$ assets out of the n assets Y_{t+1} and then constructing the J-tests proposed below in this section. As long as the factor structure holds for the selected $p+1$ assets,[10] the J-tests control the size under H_0 in (3.7) and has power in general against H_1 in (3.7). Let $\theta_{0,j}$ $(j = 1, \ldots, n)$ denote the j-th component of θ_0. This procedure essentially imposes $n-p-1$ many extra restrictions (in addition to the unit cost restriction), i.e., $\theta_{0,j} = 0$ for $j = p+2, \ldots, n$, to achieve the unique identification of a common feature in Θ^*.

Assumption 3.6. The vector θ_0 belongs to the interior of Θ.

Following the notations introduced in Section 2, we define

$$\begin{aligned} &\psi_t(\theta) \equiv (z_t - \mu_z)\theta_*'[Y_{t+1}Y_{t+1}' - \mathbb{E}(Y_{t+1}Y_{t+1}')]\theta_* \\ &\text{and } g_t(\theta) \equiv vec\left[\left(\frac{1}{2}\frac{\partial \psi_t(\theta)}{\partial \theta'}\right)'\right] \end{aligned} \tag{3.8}$$

for any $\theta \in \Theta$ and any t, where the relation between θ_* and θ is specified in (3.5).[11] Then by definition, we have $\psi(\theta) = \mathbb{E}[\psi_t(\theta)]$ and $g(\theta) = \mathbb{E}[g_t(\theta)]$ for any $\theta \in \Theta$. Under the integrability condition in Assumption 3.3, the nullity of the moment Jacobian occurs at a true common feature θ_0 in (3.3) because

$$\Gamma(\theta_0) = \mathbb{E}\left[(z_t - \mu_z)\theta_*^{0\prime} Y_{t+1} Y_{t+1}' G_2\right] = \mathbf{0}_{H\times p}. \tag{3.9}$$

We consider using both restrictions in (3.3) and (3.9) by stacking them:

$$m(\theta_0) \equiv \mathbb{E}[m_t(\theta_0)] \equiv \mathbb{E}\begin{bmatrix} \psi_t(\theta_0) \\ g_t(\theta_0) \end{bmatrix} = \mathbf{0}_{(pH+H)\times 1} \tag{3.10}$$

which are the moment conditions implied by H_0 in (3.7). As discussed in the previous section, the first order local identification of θ_0 could be achieved in (3.10), if we could show that the following matrix has full column rank:

$$\frac{\partial m(\theta_0)}{\partial \theta'} = \begin{bmatrix} \frac{\partial \psi(\theta_0)}{\partial \theta'} \\ \frac{\partial g(\theta_0)}{\partial \theta'} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{H\times p} \\ \mathbb{H} \end{bmatrix},$$

where $\mathbb{H} \equiv \partial g(\theta_0)/\partial \theta'$ is a $pH \times p$ matrix.

LEMMA 3.2. *Under Assumptions 3.1, 3.3 and (3.4), the matrix $\mathbb{H}$ has full rank.*

Lemma 3.2 shows that θ_*^0 is (first order) locally identified by the stacked moment conditions. The source of the local identification is from the zero Jacobian matrix, which actually contains more information than that needed for the local identification. We next show that if θ_*^0 is globally identified by (3.3), it is also globally identified by (3.9).

LEMMA 3.3. *Under Assumptions 3.1, 3.3 and (3.4), θ_*^0 is uniquely identified by (3.9).*

By Lemmas 3.2 and 3.3, θ_*^0 is not only (first order) locally identified, but also globally identified by the moment conditions in (3.9). As a result, one may only use these moment conditions to estimate the common feature θ_*^0. It is clear that the moment conditions in (3.9) are linear in θ, which makes the corresponding GMM estimators easy to compute. The GMM estimator based on the stacked moment conditions may be more efficient, as illustrated in Proposition 2.1. However, its computation may be costly, particularly when the dimension of θ_*^0 is high.

Lemmas 3.2 and 3.3 imply that the moment conditions in (3.9) are valid and identify a unique common feature θ_*^0 under the null hypothesis H_0 in (3.7). Hence the J-test based on (3.9) is expected to control the (asymptotic) size. We next show that in general, this test also has power against the hypothesis that there does not exist any common feature among Y_{t+1}.

LEMMA 3.4. *Suppose Assumptions 3.2 and 3.3 hold. If $\lambda_j' G_2 \neq \mathbf{0}_{1\times(n-1)}$ for any $j = 1, \ldots, p$, then there is no $\theta_*^0 \in \Theta^*$ such that (3.9) holds.*

By the definition of G_2, the left eigenvector of the zero eigenvalue of G_2 takes the form $a\mathbf{1}_{1\times n}$ for any $a \neq 0$. Hence the restriction $\lambda_j' G_2 \neq \mathbf{0}_{1\times(n-1)}$ for any $j = 1, \ldots, p$ together with the full rank of Λ is equivalent to the condition that there is no λ_j ($j \in \{1, \ldots, p\}$) such that $\lambda_j = a_j \mathbf{1}_{n\times 1}$ for any $a_j \in \mathbb{R}$. Note that the condition $\lambda_j' G_2 \neq \mathbf{0}_{1\times(n-1)}$ for any $j = 1, \ldots, p$ is a sufficient condition to ensure the moment conditions in (3.9) are invalid when there is no common feature in Θ^*. When this condition does not hold, the J-test based on (3.9) may still have nontrivial power.

For any two real matrices A and B, let $A \otimes B$ denote the Kronecker product of A and B. Using the sample average $\bar{z}$ of z_t as the estimator of μ_z, we construct the feasible moment functions as

$$\widehat{m}_t(\theta) \equiv \begin{bmatrix} \widehat{\psi}_t(\theta) \\ \widehat{g}_t(\theta) \end{bmatrix} \equiv \begin{bmatrix} (z_t - \bar{z})\left(\theta_*' Y_{t+1} Y_{t+1}' \theta_*\right) \\ \left((z_t - \bar{z}) \otimes \mathbf{I}_p\right) G_2' Y_{t+1} Y_{t+1}' \theta_* \end{bmatrix}, \tag{3.11}$$

where θ_* is defined in (3.5). The GMM estimator $\widehat{\theta}_{m,T}$ is defined as

$$\widehat{\theta}_{m,T} = \arg\min_{\theta\in\Theta} T^{-1} \left[\sum_{t=1}^{T} \widehat{m}_t(\theta)\right]' (\hat{\Omega}_{m,T})^{-1} \left[\sum_{t=1}^{T} \widehat{m}_t(\theta)\right] \tag{3.12}$$

where $\hat{\Omega}_{m,T}$ is defined in (3.15) below. In Section D of the Online Appendix, we show that under Assumptions 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6, Proposition 2.1(i) holds for the GMM estimator $\widehat{\theta}_{m,T}$.

From $\theta_* = G_2\theta + l_n$, we can write

$$T^{-1} \sum_{t=1}^{T} \widehat{g}_t(\theta) = \mathbb{H}_T \theta + \mathbb{S}_T,$$

where

$$\mathbb{H}_T \equiv T^{-1} \sum_{t=1}^{T} \left((z_t - \bar{z}) \otimes \mathbf{I}_p\right) G_2' Y_{t+1} Y_{t+1}' G_2,$$

$$\mathbb{S}_T \equiv T^{-1} \sum_{t=1}^{T} \left((z_t - \bar{z}) \otimes \mathbf{I}_p\right) G_2' Y_{t+1} Y_{t+1}' l_n. \tag{3.13}$$

Given the weight matrix $W_{g,T}$, we can compute the GMM estimator $\widehat{\theta}_{g,T}$ as

$$\widehat{\theta}_{g,T} = -\left(\mathbb{H}_T' W_{g,T} \mathbb{H}_T\right)^{-1} \mathbb{H}_T' W_{g,T} \mathbb{S}_T. \tag{3.14}$$

Let $\hat{\Omega}_{\psi,T}$, $\hat{\Omega}_{g\psi,T}$ and $\hat{\Omega}_{g,T}$ be the estimators of Ω_g, $\Omega_{g\psi}$ and Ω_ψ respectively. For example, we can construct

$$\hat{\Omega}_{m,T} = T^{-1} \sum_{t=1}^{T} (\widetilde{m}_t(\widetilde{\theta}_{g,T}) - \overline{m}_T(\widetilde{\theta}_{g,T}))(\widetilde{m}_t(\widetilde{\theta}_{g,T}) - \overline{m}_T(\widetilde{\theta}_{g,T}))' \tag{3.15}$$

where $\widetilde{\theta}_{g,T} = -\left(\mathbb{H}_T'\mathbb{H}_T\right)^{-1}\mathbb{H}_T'\mathbb{S}_T$, $\overline{m}_T(\theta) = T^{-1}\sum_{t=1}^{T}\widehat{m}_t(\theta)$ and

$$\widetilde{m}_t(\theta) = \begin{bmatrix} (z_t - \bar{z})\left(\theta_*'(Y_{t+1}Y_{t+1}' - T^{-1}\sum_{t=1}^{T} Y_{t+1}Y_{t+1}')\theta_*\right) \\ \left((z_t - \bar{z}) \otimes \mathbf{I}_p\right) G_2'(Y_{t+1}Y_{t+1}' - T^{-1}\sum_{t=1}^{T} Y_{t+1}Y_{t+1}')\theta_* \end{bmatrix},$$

and $\hat{\mathbf{\Omega}}_{g,T}$, $\hat{\mathbf{\Omega}}_{g\psi,T}$ and $\hat{\mathbf{\Omega}}_{\psi,T}$ are the leading $H \times H$, upper-right $H \times Hp$ and last $Hp \times Hp$ submatrices of $\hat{\mathbf{\Omega}}_{m,T}$, respectively.

The modified GMM estimator $\widehat{\theta}_{g^*,T}$ defined in (2.7) can be also obtained as

$$\widehat{\theta}_{g^*,T} = -\left(\mathbb{H}_T' W_{g^*,T}\mathbb{H}_T\right)^{-1}\mathbb{H}_T' W_{g^*,T}\left(\mathbb{S}_T - \mathbb{F}_T\right). \tag{3.16}$$

where $W_{g^*,T} = \left(\hat{\mathbf{\Omega}}_{g,T} - \hat{\mathbf{\Omega}}_{g\psi,T}\hat{\mathbf{\Omega}}_{\psi,T}^{-1}\hat{\mathbf{\Omega}}_{g\psi,T}'\right)^{-1}$, $\hat{\mathbf{\Omega}}_{g,T}$, $\hat{\mathbf{\Omega}}_{g\psi,T}$ and $\hat{\mathbf{\Omega}}_{\psi,T}$ are defined above, $\mathbb{F}_T = \hat{\mathbf{\Omega}}_{g\psi,T}\hat{\mathbf{\Omega}}_{\psi,T}^{-1}\mathbb{A}_T$, and

$$\mathbb{A}_T = T^{-1}\sum_{t=1}^{T}(z_t - \bar{z})\left(\left(G_2\widehat{\theta}_{g,T} + l_n\right)' Y_{t+1}Y_{t+1}'\left(G_2\widehat{\theta}_{g,T} + l_n\right)\right).$$

In Section D of the Online Appendix, we show that under Assumptions 3.1, 3.2, 3.3, 3.4 and (3.4), Proposition 2.1(ii) and Theorem 2.1 hold for $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ respectively. The closed form expressions of $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ enable us to show their $T^{1/2}$-normality without the compactness assumption on Θ^*.

After the GMM estimators $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ are obtained, we can use the J-test statistics defined in (2.8), (2.9) and (2.10) to conduct inference about the existence of common feature θ_*^0. It is clear that the test based on $J_{g,T}$ is the easiest one to use in practice because $\widehat{\theta}_{g,T}$ has a closed form solution and $J_{g,T}$ has an asymptotically pivotal distribution. The test using $J_{h,T}$ is also convenient, although one has to simulate the critical value. The test using $J_{m,T}$ is not easy to apply, especially when the dimension of parameter θ_*^0 is high.

4. SIMULATION STUDIES

In this Section, we study the finite sample performances of the proposed GMM estimators and J-tests using the CH factor model. Specifically, Y_{t+1} is an $n \times 1$ vector of asset returns generated from the following model

$$Y_{t+1} = \Lambda F_{t+1} + u_{t+1} \tag{4.1}$$

where Λ is an $n \times p$ matrix of the factor loadings, u_{t+1} is an $n \times 1$ vector of error term, and F_{t+1} is a $p \times 1$ vector of factors $(f_{l,t+1}, l = 1, \ldots, p)$ generated from a GARCH model:

$$f_{l,t+1} = \sigma_{l,t}\varepsilon_{l,t+1} \text{ and } \sigma_{l,t}^2 = \omega_l + \alpha_l f_{l,t}^2 + \beta_l\sigma_{l,t-1}^2 \tag{4.2}$$

where $(\omega_l, \alpha_l, \beta_l)'$ is a real vector with $\alpha_l + \beta_l < 1$ for any $l = 1, \ldots, p$, $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{p,t})'$ is independent with respect to u_s for any t and s, and u_{t+1} and ε_{t+1} are i.i.d. from $N(0, I_n)$ and $N(0, 0.5I_n)$ respectively. The parametrizations of n, p, Λ and $(\omega_l, \alpha_l, \beta_l)'$ for $l = 1, \ldots, p$ considered in our simulation study

are summarized in Table 5.1. The finite sample properties of the GMM estimators and the inference procedures are investigated with 50,000 simulated samples.[12]

We consider the portfolio $(\theta_0', 1 - \sum_{i=1}^{n-1} \theta_{0,i})'$ where θ_0 is an $(n-1) \times 1$ real vector. There are two sets of moment conditions for estimating θ. The moment conditions proposed in DR are:

$$\mathbb{E}\left[(z_{t+1} - \mu_z) \left| Y_{n,t+1} + \sum_{i=1}^{n-1} \theta_{0,i} (Y_{i,t+1} - Y_{n,t+1}) \right|^2 \right] = \mathbf{0}_{n \times 1}, \tag{4.3}$$

where $z_{t+1} = (Y_{1,t}^2, \ldots, Y_{n,t}^2)'$, and the moment conditions from the Jacobian of the moment functions in (4.3) are:

$$\mathbb{E}\left[(z_{t+1} - \mu_z)(Y_{i,t+1} - Y_{n,t+1}) \left(Y_{n,t+1} + \sum_{i=1}^{n-1} \theta_{0,i} (Y_{i,t+1} - Y_{n,t+1}) \right) \right] = \mathbf{0}_{n \times 1}, \tag{4.4}$$

TABLE 5.1. Parametrizations of the DGPs

	for D1.1–D1.4		
	Number of factors p	Factor loading Λ	$((\omega_l, \alpha_l, \beta_l))_{l=1}^p$
D1.1	1	$\begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$	$(\omega_1, \alpha_1, \beta_1) = (0.2, 0.2, 0.6)$
D1.2	1	$\begin{pmatrix} 1 \\ 1/2 \end{pmatrix}$	$(\omega_1, \alpha_1, \beta_1) = (0.2, 0.4, 0.4)$
D1.3	2	I_2	$\begin{pmatrix} \omega_1, \alpha_1, \beta_1 \\ \omega_2, \alpha_2, \beta_2 \end{pmatrix} = \begin{pmatrix} 0.2, 0.2, 0.6 \\ 0.2, 0.4, 0.4 \end{pmatrix}$
D1.4	2	$\begin{pmatrix} 1 & 1/2 \\ 1/2 & 0 \end{pmatrix}$	$\begin{pmatrix} \omega_1, \alpha_1, \beta_1 \\ \omega_2, \alpha_2, \beta_2 \end{pmatrix} = \begin{pmatrix} 0.2, 0.4, 0.4 \\ 0.2, 0.6, 0.2 \end{pmatrix}$
	for D2.1–D2.4		
	Number of factors p	Factor loading Λ	$((\omega_l, \alpha_l, \beta_l))_{l=1}^p$
D2.1	2	$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1/2 & 1/2 \end{pmatrix}$	$\begin{pmatrix} \omega_1, \alpha_1, \beta_1 \\ \omega_2, \alpha_2, \beta_2 \end{pmatrix} = \begin{pmatrix} 0.2, 0.2, 0.6 \\ 0.2, 0.4, 0.4 \end{pmatrix}$
D2.2	2	$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1/2 & 1/2 \end{pmatrix}$	$\begin{pmatrix} \omega_1, \alpha_1, \beta_1 \\ \omega_2, \alpha_2, \beta_2 \end{pmatrix} = \begin{pmatrix} 0.2, 0.4, 0.4 \\ 0.2, 0.5, 0.3 \end{pmatrix}$
D2.3	3	I_3	$\begin{pmatrix} \omega_1, \alpha_1, \beta_1 \\ \omega_2, \alpha_2, \beta_2 \\ \omega_3, \alpha_3, \beta_3 \end{pmatrix} = \begin{pmatrix} 0.2, 0.2, 0.6 \\ 0.2, 0.4, 0.4 \\ 0.1, 0.1, 0.8 \end{pmatrix}$
D2.4	3	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}$	$\begin{pmatrix} \omega_1, \alpha_1, \beta_1 \\ \omega_2, \alpha_2, \beta_2 \\ \omega_3, \alpha_3, \beta_3 \end{pmatrix} = \begin{pmatrix} 0.2, 0.4, 0.4 \\ 0.2, 0.5, 0.3 \\ 0.1, 0.6, 0.2 \end{pmatrix}$

for $i = 1, \ldots, n-1$. In addition to the square of the lagged asset returns $Y_{i,t}^2$ $(i = 1, \ldots, n)$, we include their cross-product terms $Y_{i,t}Y_{j,t}$ $(i, j = 1, \ldots, n$ and $i \neq j)$ as extra IVs and hence augment z_{t+1} to $\widetilde{z}_{t+1} = (z'_{t+1}, Y_{1,t}Y_{2,t}, \ldots, Y_{1,t}Y_{n,t}, \ldots, Y_{n-1,t}Y_{n,t})'$.[13] We consider the GMM estimation and inference with IVs z_{t+1} and $\widetilde{z}_{t+1}$ respectively in the simulation studies.

We consider $n = 2$ and 3 separately in the simulation studies. The proposed J-tests are compared with the bootstrap procedures suggested in DG when $n = 2$. In such a case, the portfolio $(\theta_0, 1-\theta_0)$ has only one unknown parameter to be estimated, which makes the bootstrap procedures less time consuming. When $n = 3$ the number of moment conditions in (4.3) is 3 if the set of IVs z_{t+1} are used, and the number of the moment conditions is increased to 18 if both the moment conditions in (4.3) and (4.4) are used with the augmented set of IVs $\widetilde{z}_{t+1}$. This is useful to investigate the finite sample properties of the proposed GMM estimators and J-tests with many moment conditions.

Four GMM estimators are studied: (i) the optimal weighted GMM estimator $\widehat{\theta}_{\psi,T}$ in (2.2);[14] (ii) the GMM estimator $\widehat{\theta}_{m,T}$ in (3.12); (iii) the GMM estimator $\widehat{\theta}_{g,T}$ in (3.14); and (iv) the GMM estimator $\widehat{\theta}_{g^*,T}$ in (3.16). Five inference procedures are investigated for $n = 2$ and 3 with nominal size of 0.05. In addition to the newly proposed J-tests $J_{m,T}$, $J_{g,T}$ and $J_{h,T}$,[15] we also consider the J-tests C_1 and C_2 based on $J_{\psi,T}$ with critical values $\chi^2_{1-\alpha}(H-p)$ and $\chi^2_{1-\alpha}(H)$ respectively. The bootstrap procedures in DG are investigated only under $n = 2$: their corrected bootstrap test and the continuously-corrected bootstrap test are denoted as B_1 and B_2 respectively.[16] The number of moment conditions and degrees of freedom of the asymptotic distribution of the J-tests, $J_{m,T}$, $J_{g,T}$, $J_{h,T}$, C_1 and C_2 are summarized in Table E.1 in Section E of the Online Appendix.

4.1. Simulation Results with $n = 2$

Four DGPs (labeled as D1.1, D1.2, D1.3, and D1.4 respectively) are considered when $n = 2$. The parametrizations of these DGPs are summarized in Table 5.1. The null hypothesis H_0 in (3.7) holds under D1.1 and D1.2, while the alternative hypothesis H_1 holds under D1.3 and D1.4.

The finite sample bias, standard deviation and root of mean square error (RMSE) of the four GMM estimators in D1.1 and D1.2 are reported in Figures E.1 and E.2 in Section E of the Online Appendix, respectively. From the three left panels (i.e., the top-left, middle-left and bottom-left panels) of Figure E.1, we see that: (i) with the growth of the sample size, the bias of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ goes to zero much faster than $\widehat{\theta}_{\psi,T}$. Moreover, the bias of $\widehat{\theta}_{\psi,T}$ is of the same magnitude as and sometimes bigger than the standard errors, leading to extremely poor coverage of confidence sets and misleading inferences; (ii) the standard deviation of $\widehat{\theta}_{\psi,T}$ is overall smaller than the other three GMM estimators in this DGP; (iii) the GMM estimators which takes the zero Jacobian into estimation does not seem to dominate $\widehat{\theta}_{\psi,T}$ in terms of RMSE when the sample size is not large enough; (iv) the finite sample properties of $\widehat{\theta}_{m,T}$ and $\widehat{\theta}_{g^*,T}$ are very similar

when the sample size becomes large (e.g., $T = 5,000$). From the three right panels (i.e., the top-right, middle-right and bottom-right panels) of Figure E.1, we see that including more moment conditions in GMM estimation slightly increases the bias of the GMM estimators, but reduces their variances.

The finite sample properties of the GMM estimators which take the zero Jacobian into estimation in D1.1 indicate that the additional local identification strength provided by the Jacobian moment conditions is rather weak,[17] see Antoine and Renault (2016) for some related discussion. The identification strength of the Jacobian moment conditions is slightly increased in D1.2 which leads to significant improvements of the GMM estimators $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ as we can see in Figure E.2. The finite sample properties of $\widehat{\theta}_{\psi,T}$ are also improved in D1.2, which may be due to the strengthening of the second order local identification. Specifically, in Figure E.2 we see that: (i) the finite sample bias and variance of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ converge to zero very fast with the growth of the sample size; (ii) the finite sample bias of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ are overall smaller than $\widehat{\theta}_{\psi,T}$ and their finite sample variances are also smaller when the sample size is reasonably large (e.g., $T = 3,000$); (iii) including more valid moment conditions in GMM estimation clearly improves the finite sample properties of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ in terms of RMSE, but has almost no effect on the RMSE of $\widehat{\theta}_{\psi,T}$; (iv) the RMSEs of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ become smaller than $\widehat{\theta}_{\psi,T}$ when the sample size is slightly large.

We next investigate the finite sample properties of the inference procedures under D1.1–D1.4. The results are presented in Figure 5.1. We have the following findings. First, the empirical rejection probabilities of the J-tests $J_{m,T}$, $J_{g,T}$ and $J_{h,T}$ converge to the nominal size slowly under D1.1, while they converge much faster under D1.2. The test $J_{m,T}$ shows over-rejection in D1.1 even when the sample size is relatively large. The tests $J_{g,T}$ and $J_{h,T}$ control size well in both D1.1 and D1.2. The slow convergence rate of the empirical rejection probabilities of the J-tests $J_{m,T}$, $J_{g,T}$ and $J_{h,T}$ in D1.1 is due to the weak local identification of the Jacobian moment conditions. Second, the tests C_1 and C_2 show over-rejection and under-rejection respectively in both D1.1 and D1.2 when the set of IVs z_{t+1} is used. This is expected by the asymptotic analysis in DR. However, when more IVs, i.e., $\widetilde{z}_{t+1}$, are used in the GMM estimation, the tests C_1 and C_2 are both under-sized in D1.2 even with the large sample size. This is rather surprising, because increasing the number of moment conditions (from 2 with z_{t+1} to 3 with $\widetilde{z}_{t+1}$) in D1.1 does not make the test C_1 severely under-sized and one would expect similar results in D1.2 as the local identification is improved in this DGP. This shows that there may be a potential further aspect of limit theory in DR regarding the finite sample distribution of $J_{\psi,T}$. Third, the bootstrap inference B_2 is overall under-sized in both D1.1 and D1.2. In D1.2, B_2 is even more conservative than the conservative test C_2 proposed in DR. The empirical rejection probability of the bootstrap inference B_1 is closer to the nominal size, although it is also under-sized in D1.2 (particularly when more IVs are used in GMM). Fourth, all the inference procedures have nontrivial power in D1.3 and D1.4. The power

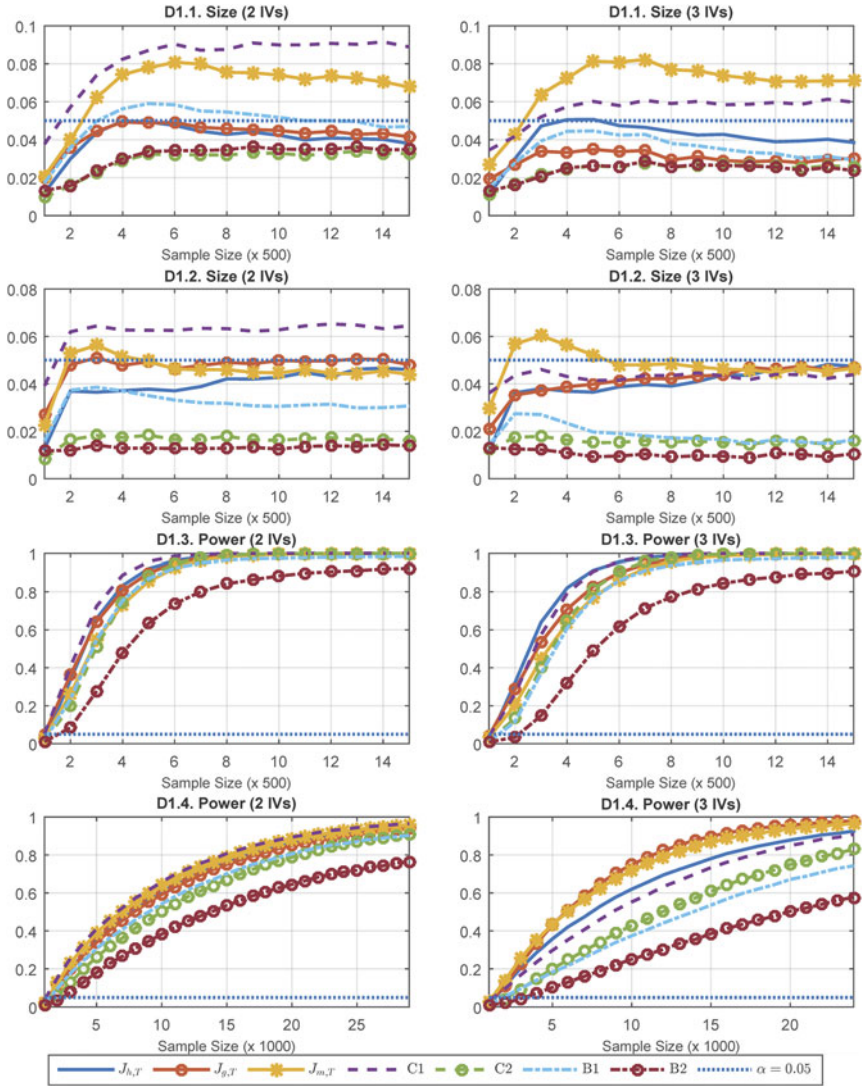

FIGURE 5.1. Empirical rejection probabilities of the GMM inference $J_{h,T}$, $J_{m,T}$, $J_{g,T}$, C_1, C_2, B_1 and B_2 under D1.1–D1.4.

Notes: 1. In the left three panels, the 2 IVs are $(Y_{1,t}^2, Y_{2,t}^2)$; 2. in the right three panels, the 3 IVs are $(Y_{1,t}^2, Y_{2,t}^2, Y_{1,t}Y_{2,t})$; 3. in the first six panels on the top, sample size divided by 500 is reported on the horizontal axis; 4. in the last two panels on the bottom, sample size divided by 1,000 is reported on the horizontal axis; 5. the simulation results are based on 50,000 simulated samples.

of the tests $J_{m,T}$, $J_{g,T}$, and $J_{h,T}$ are better than the bootstrap tests B_1 and B_2. From their parametrizations, we see that D1.4 is a local perturbation of D1.2. The under-rejection of the tests based on C_1, C_2, B_1, and B_2 in D1.2 leads to

power loss in D1.4. Fifth, the condition $\lambda_j' G_2 \neq \mathbf{0}_{1\times(n-1)}$ for any $j \in \{1, \ldots, p\}$ in Lemma 3.4 holds under D1.3 and D.1.4. Hence the Jacobian moment conditions are misspecified in these DGPs, which explains the good power properties of the J-test based on $J_{g,T}$.

4.2. Simulation Results with $n = 3$

Four DGPs (labeled as D2.1, D2.2, D2.3, and D2.4 respectively) are considered when $n = 3$, whose parametrizations are also summarized in Table 5.1. The null hypothesis H_0 in (3.7) holds under D2.1 and D2.2, while the alternative hypothesis H_1 holds under D2.3 and D2.4.

The finite sample properties of the GMM estimators of $(\theta_{0,1}, \theta_{0,2})$ in D2.1 and D2.2 are summarized in Figures E.3–E.6 in Section E of the Online Appendix. The local identification provided by the Jacobian moment conditions is slightly improved in D2.2 when compared with D2.1. From Figures E.3–E.6, we see that: (i) with the growth of the sample size, the RMSEs of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ converge to zero much faster than $\widehat{\theta}_{\psi,T}$; (ii) increasing the number of the moment conditions leads to increase of finite sample bias, decrease of variance and decrease of RMSEs of the GMM estimators in D2.1 and D2.2; (iii) the finite sample properties of $\widehat{\theta}_{m,T}$ and $\widehat{\theta}_{g^*,T}$ are very similar when the sample size becomes slightly large; (iv) $\widehat{\theta}_{m,T}$ has smaller finite sample variance than $\widehat{\theta}_{g,T}$, which shows the efficiency gain of using modified moment conditions; (v) the RMSEs of $\widehat{\theta}_{m,T}$, $\widehat{\theta}_{g,T}$ and $\widehat{\theta}_{g^*,T}$ converge to zero much faster in D2.2 than in D2.1, which shows the improvement of their finite sample properties due to the improvement of the local identification strength in D2.2.

We next investigate the properties of the inference procedures $J_{m,T}$, $J_{g,T}$, $J_{h,T}$, C_1, and C_2 under D2.1–D2.4. The results are presented in Figure 5.2. which illustrates the following findings. First, the empirical rejection probabilities of the J-tests $J_{m,T}$, $J_{g,T}$, and $J_{h,T}$ converge to the nominal size slowly under D2.1, while they converge much faster under D2.2. As we have discussed above, the Jacobian moment conditions provide weak local identification under D2.1, leading to the slow convergence of the empirical size of the J-tests $J_{m,T}$, $J_{g,T}$ and $J_{h,T}$. The strengthened local identification from the Jacobian moment conditions in D2.2 improves not only the finite sample properties of the GMM estimators $\widehat{\theta}_{m,T}$ and $\widehat{\theta}_{g,T}$, but also the performances of the J-tests based on these estimators. Second, the J-test $J_{h,T}$ controls the size well in both D2.1 and D2.2, although it is slightly under-sized in finite sample. It has good power in D2.3 with 3 or 6 IVs and in D2.4 with 3 IVs. Its power suffers slightly in D2.4 when more IVs are included in GMM estimation. Third, in D2.1 and D2.2, we see that including more moment conditions in GMM decreases the empirical rejection probabilities of the J-tests. This effect is almost negligible on $J_{h,T}$, mild on $J_{m,T}$, substantial on $J_{g,T}$ and severe on the test C_1. The under-rejection of the test C_1 with more moment conditions is much more severe in D2.2 than in D1.2. As we have discussed in Remark 2.2, one may expect that the empirical rejection probabilities of the tests

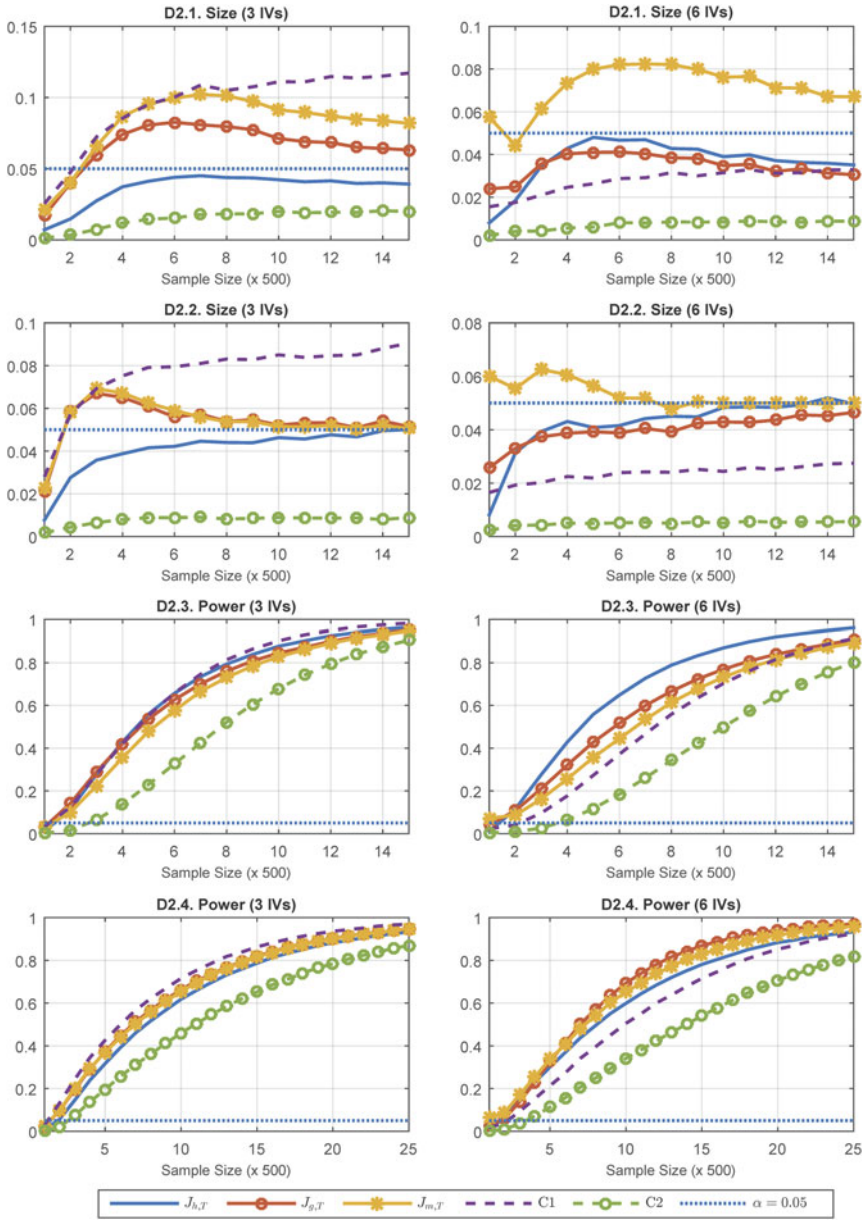

FIGURE 5.2. Empirical rejection probabilities of the GMM inference $J_{h,T}$, $J_{m,T}$, $J_{g,T}$, C_1 and C_2 under D2.1–D2.4.

Notes: 1. In the left three panels, the 3 IVs are $(Y_{1,t}^2, Y_{2,t}^2, Y_{3,t}^2)$; 2. in the right three panels, the 6 IVs are $(Y_{1,t}^2, Y_{2,t}^2, Y_{3,t}^2, Y_{1,t}Y_{2,t}, Y_{1,t}Y_{3,t}, Y_{2,t}Y_{3,t})$; 3. in each panel, sample size divided by 500 is reported on the horizontal axis; 4. the simulation results are based on 50,000 simulated samples.

C_1 and C_2 are close to the nominal size when the sample size is large in D2.1 or D2.2 with 6 IVs, as the critical values $\chi^2_{0.95}(4)$ and $\chi^2_{0.95}(6)$ are close in their relative magnitudes. However, as we can see from Figure 5.2, the empirical rejection probabilities of the tests C_1 and C_2 are close to each other but some away from the nominal size. The empirical rejection probability of C_1 in D2.1 or D2.2 with 6 IVs increases with the sample size but the rate is slow. Thus the finite sample distributions of the asymptotic lower and upper bounds of $J_{\psi,T}$ (i.e., $\chi^2(H-p)$ and $\chi^2(H)$) may deserve further investigation. Fourth, the under-rejection of the tests based on C_1 and C_2 leads to significant loss of power. The poor power of the test C_2 is expected as it is conservative. However, in D2.3 and D2.4, the power of the test C_1 is worse than the tests based on $J_{m,T}$, $J_{g,T}$, and $J_{h,T}$ when 6 IVs are used in GMM estimation. As D2.4 is a (local) perturbation of D2.2, the poor power of C_1 in this DGP is not surprising. Finally, the condition $\lambda_j' G_2 \neq \mathbf{0}_{1\times(n-1)}$ for any $j \in \{1,\ldots,p\}$ in Lemma 3.4 holds under D2.3 and D.2.4. Hence the Jacobian moment conditions are misspecified in these DGPs, explaining the good power properties of the J-test based on $J_{g,T}$.

5. CONCLUSION

This paper investigates the GMM estimation and inference when the Jacobian of the moment conditions has known deficient rank. We show that the degenerated Jacobian contains nontrivial information about the unknown parameters. When such information is employed in estimation, one can possibly construct GMM estimators and over-identification tests with standard properties. Our simulation results in the common CH factor models support the proposed theory. The GMM estimators additionally using the Jacobian-based moment conditions show good finite sample properties. Moreover, the J-tests based on the proposed GMM estimators have good size control and better power than the commonly used GMM inference which ignores the information contained in the Jacobian moments.

NOTES

1. We want to emphasize that weak second-order local identification issue is not allowed in Condition (2.4), because the asymptotic theory of this paper assumes that the latent data generating process (DGP) is fixed with the growth of the sample size. Our paper and other closely related papers (DR and DG) do not theoretically study this issue, but rather focus on the case of local identification failure of known forms with deficient rank Jacobian. The developed theory and efficiency gain from adding additional Jacobian moments are therefore more clearly confirmed through a simulation with first-order local identification failure *without weak second-order local identification* (see, e.g., simulation results under D1.2 and D2.2 in Section 4). Introducing the weak local identification issue further into our framework is interesting but left for future research.

2. We appreciate a referee's comment pointing this out.

3. In reality, one may only be interested in testing the validity of the moment conditions in (2.1). For any $T^{1/2}$-consistent estimator $\widehat{\theta}_T$, one can show that

$$J_{\psi g,T} \equiv T^{-1}\left[\sum_{t=1}^{T}\psi_t(\widehat{\theta}_T)\right]'\hat{\Omega}_{\psi,T}^{-1}\left[\sum_{t=1}^{T}\psi_t(\widehat{\theta}_T)\right] \to_d \chi^2(H),$$

under the null hypothesis that (2.1) and (2.3) hold. In practice, one can let $\widehat{\theta}_T = \widehat{\theta}_{m,T}$, or $\widehat{\theta}_T = \widehat{\theta}_{g,T}$ when θ_0 is identified by (2.3). It is interesting that the estimation error in $\widehat{\theta}_T$ does not affect the asymptotic distribution of the test statistic $J_{\psi g,T}$ under the null, because $J_{\psi g,T}$ has the same asymptotic $\chi^2(H)$ if $\widehat{\theta}_T$ is replaced by θ_0. The reason for this phenomenon is that the Jacobian of the moment conditions in (2.1) is degenerate.

4. See the simulation results under D1.3, D1.4, D2.3, and D2.4 for example. In these DGPs, the tests using the moment conditions (2.3) are more powerful than the tests which only consider the moment conditions (2.1).

5. Using the Hadamard product of matrices in many standard softwares (e.g., Matlab), one can conveniently get the simulated critical values. Let $\mathbb{P}_T$ and $\hat{\Omega}_{m,T}$ be the consistent estimators of $\mathbb{P}$ and Ω_m respectively. Define $\mathbb{M}_T = \hat{\Omega}_m^{\frac{1}{2}} \mathbb{P}_T' W_{T,h} \mathbb{P}_T \hat{\Omega}_m^{\frac{1}{2}}$. Let $\mathbb{B}_R$ be a $R \times (Hp + H)$ matrix whose rows are independently generated from the $Hp + H$ dimensional standard normal distribution. The matrix $\mathbb{B}_R$ can be generated simultaneously in many softwares including Matlab. Let $A \circ B$ denote the Hadamard product of matrices A and B. Define $J_R = ((\mathbb{B}_R \mathbb{M}_T) \circ \mathbb{B}_R) \times 1_{H(p+1)}$, where $1_{H(p+1)}$ is an $(Hp + H) \times 1$ vector of ones. Then the simulated $(1-\alpha)$-th quantile of the asymptotic distribution of $J_{h,T}$ is the $(1-\alpha)R$-th largest value in J_R. In the simulation studies of this paper, we choose $R = 250,000$ and set $W_{h,T}$ to be the identity matrix, and use $\widehat{\theta}_{g,T}$ with identity weight matrix to construct $\mathbb{M}_T$.

6. From the simulation studies in Section 4, when H is increased (e.g., $H = 3$ in D2.1 and D2.2 with 3 IVs but increased to $H = 6$ in these DGPs with 6 IVs) and p is fixed (e.g., $p = 2$ in D2.1 and D2.2), the empirical rejection probabilities of the tests C_1 and C_2 are close to each other but both tests are under-sized even when the sample size is large. The empirical rejection probability of the tests C_1 converges slowly to its asymptotic value. See Figures 5.1 and 5.2 and the related discussion in Section 4 for the details.

7. We assume that μ_z is known when we study the identification of the common feature θ_*^0. This is valid as μ_z is uniquely identified by the moment conditions $\mathbb{E}[z_t - \mu_z] = \mathbf{0}_{H\times 1}$.

8. We use this normalization in the rest of the paper because the main results in DR are also derived under this restriction. However, the issue in DR and our proposed solutions apply irrespective of a specific normalization condition.

9. In their simulation studies, DR found that the J-test based on $J_{\psi,T}$ is undersized in one of the DGPs which has multiple common features (see their simulation results under the DGP labeled as D3). We find similar results for the J-tests based on $J_{m,T}$, $J_{g,T}$ and $J_{h,T}$ in the same DGP. These simulation results are available upon request.

10. Let $(Y_{t+1})_{p+1}$ denote the leading $(p+1) \times 1$ subvector of Y_{t+1}. Then (3.1) implies that

$$\mathrm{Var}\left((Y_{t+1})_{p+1} | \mathcal{F}_t\right) = \Lambda_{p+1} D_t \Lambda_{p+1}' + \Omega,$$

where Λ_{p+1} denotes the leading $(p+1) \times p$ submatrix of Λ. The factor structure of Y_{t+1} is kept in $(Y_{t+1})_{p+1}$ if $rank(\Lambda_{p+1}) = p$.

11. As $\psi_t(\theta)$ is quadractic in θ, we define the Jacobian moment functions as the partial derivative of $\psi_t(\theta)$ divided by 2 to simpify notations.

12. For each simulated sample with sample size T, we generate $T + 2,500$ observations and drop the first 2,500 observations to reduce the effect of the initial conditions of the data generating mechanism on the estimation and inference.

13. The Jacobian of the moment conditions in (4.4) is a matrix of quadratic functions of Y_{t+1}, which motivates us to include the cross-product terms $Y_{i,t+1} Y_{j,t+1}$ $(i \neq j)$ as extra IVs.

14. The optimal weight matrix is constructed using a preliminary GMM estimator $\widetilde{\theta}_{\psi,T}$ based on moment conditions (4.3) and the identity weight matrix.

15. We let $W_{g,T}$ and $W_{h,T}$ be the identity matrix when constructing $J_{h,T}$.

16. We use the same bootstrap estimating equation and bootstrap weight matrix employed in Section 5 of DG. Following DG, we let the bootstrap replication be 399.

17. Indeed, we find that the smallest eigenvalue of the inverse of the asymptotic variance-covariance matrix $\mathbb{H}'(\Omega_g - \Omega_{g\psi}\Omega_\psi^{-1}\Omega_{\psi g})^{-1}\mathbb{H}$ of $\widehat{\theta}_{m,T}$ in D1.1 is very close to zero (around 0.0034).

REFERENCES

Andrews, D.W.K. & X. Cheng (2012) Estimation and inference with weak, semi-strong and strong identification. *Econometrica* 80, 2153–2211.

Antoine, B., H. Bonnal, & E. Renault (2007) On the efficient use of the informational content of estimating equations: Implied probabilities and euclidean empirical likelihood. *Journal of Econometrics* 138(2), 461–487.

Antoine, B. & E. Renault (2017) On the relevance of weaker instruments. *Econometric Reviews*, 1–18.

Broto, C. & E. Ruiz (2004) Estimation methods for stochastic volatility models: A survey. *Journal of Economic Surveys* 18(5), 613–649.

Choi, I. & P.C.B. Phillips (1992) Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. *Journal of Econometrics* 51(1), 113–150.

Diebold, F.X. & M. Nerlove (1989) The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. *Journal of Applied Econometrics* 4(1), 1–21.

Dovonon, P. & S. Gonçalves (2016) Bootstrapping the GMM Overidentification Test Under First-order Underidentification. Working paper, Concordia University.

Dovonon, P. & E. Renault (2013) Testing for common conditionally heteroskedastic factors. *Econometrica* 81(6), 2561–2586.

Doz, C. & E. Renault (2006) Factor stochastic volatility in mean models: A GMM approach. *Econometric Reviews* 25(2–3), 275–309.

Eichenbaum, M., L. Hansen, & K. Singleton (1988) A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics* 103(1), 51–78.

Engle, R.F. & S. Kozicki (1993) Testing for common features. *Journal of Business and Economic Statistics* 11(4), 369–380.

Engle, R.F., V.K. Ng, & M. Rothschild (1990) Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics* 45(1), 213–237.

Fieller, E.C. & H.O. Hartley (1954) Sampling with control variables. *Biometrika* 41(3/4), 494–501.

Fiorentini, G., E. Sentana, & N. Shephard (2004) Likelihood-based estimation of latent generalized ARCH structures. *Econometrica* 72(5), 1481–1517.

Hall, A.R. (2005) *Generalized Methods of Moments*. Oxford University Press.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Kleibergen, F. (2005) Testing parameters in GMM without assuming that they are identified. *Econometrica* 73(4), 1103–1123.

Newey, W.K. & R.J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72(1), 219–255.

Newey, W.K. & F. Windmeijer (2009) Generalized method of moments with many weak moment conditions. *Econometrica* 77(3), 687–719.

Phillips, P.C.B. (1989) Partially identified econometric models. *Econometric Theory* 5, 181–240.

Sargan, J.D. (1983) Identification and lack of identification. *Econometrica* 51, 1605–1633.

Sentana, E. (2015) Finite Underidentification. CEMFI Working paper, 1508.

Silvennoinen, A. & T. Teräsvirta (2009) Multivariate GARCH models. In T.G. Andersen, R.A. Davis, J.-P. Kreiss, & T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 201–229. Springer.

Staiger, D. & J.H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.

Stock, J.H. & J.H. Wright (2000) GMM with weak identification. *Econometrica* 68, 1055–1096.