On Standard Inference for GMM
with Seeming Local Identification Failure

Ji Hyung Lee †  Zhipeng Liao‡

First Version: April 2014; This Version: December, 2014

Abstract

This paper studies the GMM estimation and inference problem that occurs when the Jacobian of the moment conditions is known to be a matrix of zeros 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 non-standard asymptotic distribution of the over-identification test statistics. We show that the zero Jacobian matrix contains non-trivial 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 same simulation designs used in Dovonon and Renault (2013).

Keywords: Degenerate Jacobian; Conditionally Heteroskedastic Factors, GMM, Local Identification Failure, Non-standard Inference, Over-identification Test, Asymptotically Exact Inference

1 Introduction

The generalized method of moments (GMM) is a popular method for empirical research in economics and finance. Under some regularity conditions, Hansen (1982) showed that the GMM estimator has standard properties, such as $\sqrt{T}$-consistency and asymptotic normal distribution. The over-identification test (J-test) statistics has an asymptotic Chi-square distribution. On the other hand, when some of the regularity conditions are not satisfied, the GMM estimator may have non-standard

*We acknowledge useful comments from Don Andrews, Xu Cheng, Denis Chetverikov, Yanyin Fan, Jinyong Hahn, Guofang Huang, Ivana Komunjer, Rosa Matzkin, Peter Phillips, Shuyang Sheng, Ruoyao Shi, Yixiao Sun, and participants in Econometrics seminar at UC Davis, UCLA, UCSD, Yale and the Seattle-Vancouver Econometrics Conference. Any errors are the responsibility of the authors.

†Department of Economics, University of Washington, 336 Savery Hall, Box 353330, Seattle, WA 98195. Email: jihyung2@uw.edu
‡Department of Economics, UC Los Angeles, 8379 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. Email: zhipeng.liao@econ.ucla.edu
properties. For example, when the moment conditions only contain weak information, the GMM estimator may be inconsistent and have mixture normal asymptotic distribution (see, e.g., Andrews and Cheng, 2012, Staiger and Stock, 1997 and Stock and Wright 2000).

Dovonon and Renault (2013, hereafter DR) have recently pointed out an interesting issue that occurs due to the violation of a regularity condition. When testing for common conditionally heteroskedastic (CH) features in asset returns, DR show that the Jacobian of the moment conditions is a matrix of zeros at the true parameter value. This causes a slower than $\sqrt{T}$ rate of convergence of the GMM estimator because the second term in the Taylor expansion of the moment functions now plays the leading role in asymptotics. The role of the second order term results in the $J$-test statistics as being a minimum of even polynomial of degree 4, which can be interpreted as a minimum of polynomial of degree 2 with a positivity constraint on the minimizer. As a result, the constrained minimizer is greater than the unconstrained minimizer (the standard $J$-test statistics, $\chi^2(H - p)$) but it is less than the quadratic function of the sample moment conditions at true parameters ($\chi^2(H)$). The $J$-test statistics therefore lies between $\chi^2(H - p)$ and $\chi^2(H)$, where $H$ and $p$ are the number of moment conditions and parameters, respectively. These results by DR 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 oversized $J$-tests under the degeneracy of Jacobian moments.

This paper revisits the issue raised in DR. We also 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 estimation and inference using the zero Jacobian matrix as additional moment conditions. These additional moment restrictions contain extra information of the economic model. This additional information is exploited to achieve the first-order local identification of the unknown structural parameters. We construct GMM estimators with $\sqrt{T}$-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.

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 $\sqrt{T}$-consistency and asymptotic normality. Our $J$-test statistic for the existence of the common feature have asymptotic Chi-square distribution, which enables non-conservative asymptotic inference. Moreover, the Jacobian based GMM estimator of the common feature has the 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 developed in Section 2 to
the common CH factor models. Section 4 contains simulation studies and Section 5 concludes. Tables, figures, main proofs are given in the Appendix, while selected proofs and further technical arguments are available from the supplementary Appendix.

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:

\[
\rho(\theta_0) \equiv E \left[ \psi(X_t, \theta_0) \right] \equiv E \left[ \psi_t (\theta_0) \right] = 0, \tag{2.1}
\]

where \( X_t \) is a random vector which is observable in period \( t \). As illustrated in Hansen (1982), the global identification together with other regularity conditions can be used to show standard properties of the GMM estimator of \( \theta_0 \). For any \( \theta \in \Theta \), define

\[
\Gamma (\theta) = \frac{\partial}{\partial \theta^T} E [\psi_t (\theta)].
\]

Then \( \Gamma (\theta) \) is an \( H \times p \) matrix of functions. The standard properties of the GMM estimator, such as \( \sqrt{T} \)-consistency and asymptotic normality, rely on the condition that \( \Gamma (\theta_0) \) has full rank. When \( \Gamma (\theta_0) = 0_{H \times p} \), the properties of the GMM estimator are nonstandard. For example its convergence rate is slower than \( \sqrt{T} \) and the associated over-identification J-test statistics has a mixture of asymptotic Chi-square distributions. DR have established these nonstandard properties of GMM inference when \( \Gamma (\theta_0) = 0_{H \times p} \), in the context of testing for the common feature in the common CH factor models. In this Section, we discuss the same issue and a possible solution in a general GMM context.

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

\[
g(\theta) = 0_{pH \times 1} \text{ when } \theta = \theta_0. \tag{2.2}
\]

The new set of moment restrictions ensures the first order local identification of \( \theta_0 \), when the Jacobian of \( g(\theta) \) (or essentially the Hessian of \( \rho(\theta) \)) evaluated at \( \theta_0 \) has full column rank. We define the corresponding Jacobian matrix of \( g(\theta) \) as

\[
\mathbb{H} (\theta) \equiv \frac{\partial}{\partial \theta^T} g(\theta),
\]

where \( \mathbb{H} (\theta) \) is now a \( pH \times p \) matrix of functions. When \( \mathbb{H} (\theta_0) \equiv \mathbb{H} \) has full column rank the first order local identification of \( \theta_0 \) 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 $\rho_h(\theta)$ be the $h$-th ($h = 1, \ldots, H$) component function in $\rho(\theta)$. Suppose that (i) $\theta_0$ belongs to the interior of $\Theta$; and (ii) for any $\theta \in \Theta$,

$$
(\theta - \theta_0)' \left( \frac{\partial^2 \rho_h}{\partial \theta \partial \theta} (\theta_0) \right) (\theta - \theta_0) \right)_{1 \leq h \leq H} = 0
$$

(2.3)

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

Condition (ii) in Lemma 2.1 is the second order local identification condition of $\theta_0$ 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 Gonalves (2014). Lemma 2.1 shows that when the moment conditions in (2.2) are used under the condition (ii), the first order local identification of $\theta_0$ is achieved. The moment conditions in (2.2) alone may not ensure the global/unique identification of $\theta_0$. However, as $\theta_0$ is globally (uniquely) identified by the moment conditions in (2.1), we can use the moment conditions in (2.1) and (2.2) in GMM to ensure both the global identification and the first order local identification of $\theta_0$.

Let $\Gamma_t(\theta) = \frac{\partial}{\partial \theta} \psi_t(\theta)$ and $g_t(\theta) = vec(\Gamma_t(\theta)')$. We can define the GMM estimator of $\theta_0$ using all moment conditions as

$$
\widehat{\theta}_{m,T} = \arg \min_{\theta \in \Theta} \left[ \sum_{t=1}^{T} m_t(\theta) \right]' W_{m,T} \left[ \sum_{t=1}^{T} m_t(\theta) \right]
$$

(2.4)

where $m_t(\theta) = (\psi_t'(\theta), g_t'(\theta))'$ is an $H(p+1)$ dimensional vector of functions, and $W_{m,T}$ is an $(Hp+H) \times (Hp+H)$ weight matrix. Similarly, we define the GMM estimator of $\theta_0$ using only the moment conditions in (2.2) as

$$
\widehat{\theta}_{g,T} = \arg \min_{\theta \in \Theta} \left[ \sum_{t=1}^{T} g_t(\theta) \right]' W_{g,T} \left[ \sum_{t=1}^{T} g_t(\theta) \right]
$$

(2.5)

where $W_{g,T}$ is an $Hp \times Hp$ weight matrix.

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

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

where $\Omega_{\psi}$ is an $H \times H$ matrix and $\Omega_{g}$ is a $pH \times pH$ matrix; (ii) $W_{m,T} \rightarrow_p \Omega_m^{-1}$ and $W_{g,T} \rightarrow_p W_g$, 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] Assumption 2.1 and Assumption A.1 in the Appendix, we have

\[ \sqrt{T}(\hat{\theta}_{m,T} - \theta_0) \rightarrow_d N(0, \Sigma_{\theta,m}), \]

where \( \Sigma_{\theta,m}^{-1} = \mathbb{H}'(\Omega_g - \Omega_{g:\psi}\Omega_{\psi}^{-1}\Omega_{\psi:})^{-1}\mathbb{H}. \) Moreover, if \( \theta_0 \) is uniquely identified by \( (2.2) \), then

\[ \sqrt{T}(\hat{\theta}_g,T - \theta_0) \rightarrow_d N(0, \Sigma_{\theta,g}) \]

where \( \Sigma_{\theta,g} = (\mathbb{H}'W_g\mathbb{H})^{-1}\mathbb{H}'W_g\Omega_gW_g\mathbb{H}(\mathbb{H}'W_g\mathbb{H})^{-1}. \)

In the linear IV models, Condition \( (2.3) \) does not hold under the zero Jacobian, and hence \( \theta_0 \) 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 \( \theta_0 \) (after some rotation) is identified and \( \sqrt{T} \)-estimable. Proposition 2.1 shows that in some non-linear models, the degenerate Jacobian together with Condition (2.3) and other regularity conditions can be used to derive a \( \sqrt{T} \)-consistent GMM estimator of \( \theta_0 \). Our results therefore supplement the earlier findings in Phillips (1989) and Choi and Phillips (1992).

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 \( \hat{\theta}_{m,T} \) is preferred to \( \hat{\theta}_{g,T} \) from the efficiency perspective, even when \( \theta_0 \) is globally identified by (2.2). In some examples (e.g., the common CH factor model in the next Section), the computation of \( \hat{\theta}_{g,T} \) may be easier than \( \hat{\theta}_{m,T} \). We next propose an estimator which is as efficient as \( \hat{\theta}_{m,T} \), and can be computed similarly to \( \hat{\theta}_{g,T} \).

Let \( \tilde{\Omega}_{g,T}, \tilde{\Omega}_{g:\psi,T} \) and \( \tilde{\Omega}_{\psi,T} \) be the consistent estimators of \( \Omega_g, \Omega_{g:\psi} \) and \( \Omega_m \) respectively. These variance matrix estimators can be constructed using \( \tilde{\theta}_{g,T} \), for example. The new GMM estimator is defined as

\[ \tilde{\theta}_{g^*,T} = \arg \min_{\theta \in \Theta} \left[ \sum_{t=1}^{T} \tilde{g}_{\psi,t}(\theta) \right]'W_{g^*,T}\left[ \sum_{t=1}^{T} \tilde{g}_{\psi,t}(\theta) \right] \]

where \( \tilde{g}_{\psi,t}(\theta) = g_t(\theta) - \tilde{\Omega}_{g:\psi,t}\tilde{\Omega}_{g:\psi,T}^{-1}\tilde{\theta}_{g,T} \) and \( W_{g^*,T}^{-1} = \tilde{\Omega}_{g,T} - \tilde{\Omega}_{g:\psi,T}\tilde{\Omega}_{\psi,T}^{-1}\tilde{\Omega}_{g,T} \).

Theorem 2.1  Under the conditions of Proposition 2.1 we have

\[ \sqrt{T}(\tilde{\theta}_{g^*,T} - \theta_0) \rightarrow_d N(0, \Sigma_{\theta,m}) \]

where \( \Sigma_{\theta,m} \) is defined in Proposition 2.1.

From Theorem 2.1 \( \tilde{\theta}_{g^*,T} \) has the same asymptotic variance as \( \hat{\theta}_{m,T} \). It is essentially computed based on the moment conditions \( (2.2) \), hence it benefits computational simplicity whenever \( \tilde{\theta}_{g,T} \) is easy to calculate.

When the GMM estimators have the standard 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 \( (2.2) \), one can jointly test their validity using the
following standard result:

\[ J_{m,T} \equiv T^{-1} \left[ \sum_{t=1}^{T} m_t(\hat{\theta}_{m,T}) \right]' \hat{\Omega}_{m,T}^{-1} \left[ \sum_{t=1}^{T} m_t(\hat{\theta}_{m,T}) \right] \rightarrow_d \chi^2(Hp + H - p). \quad (2.7) \]

When \( \theta_0 \) is identified by (2.2), it may be convenient to use the J-test based on (2.5) in practice:

\[ J_{g,T} \equiv T^{-1} \left[ \sum_{t=1}^{T} g_t(\hat{\theta}_{g,T}) \right]' \hat{\Omega}_{g,T}^{-1} \left[ \sum_{t=1}^{T} g_t(\hat{\theta}_{g,T}) \right] \rightarrow_d \chi^2(Hp - p), \quad (2.8) \]

where \( \hat{\theta}_{g,T} \) denotes the GMM estimator defined in (2.5) with weight matrix \( W_{g,T} = \hat{\Omega}_{g,T}^{-1} \). One interesting aspect of the proposed J-test in (2.7) is that it has standard degrees of freedom, i.e. the number of moment conditions used in estimation minus the number of parameters we estimate. Among the \((Hp + H)\) many moment restrictions, \( H \) moments from (2.1) have degenerate Jacobian moments. By combining this information on \( H \) moments with the extra information provided by the \( Hp \) Jacobian moments, we avoid the issue of rank deficiency. Stacking the additional moments from (2.2) provides enough sensitivity of the J-test statistic to parameter variation. As a result, the standard degrees of freedom show up in the asymptotic Chi-square distribution in (2.7).

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

\[ J_{h,T} \equiv T^{-1} \left[ \sum_{t=1}^{T} m_t(\hat{\theta}_{g,T}) \right]' W_{h,T} \left[ \sum_{t=1}^{T} m_t(\hat{\theta}_{g,T}) \right] \quad (2.9) \]

where \( W_{h,T} \) is an \((Hp + 1) \times (Hp + 1)\) real matrix. For any two square matrices \( A \) and \( B \), we use \( \text{diag}(A,B) \) to denote the diagonal matrix with \( A \) being the leading sub-matrix and \( B \) being the last sub-matrix. The following theorem provides the asymptotic distribution of \( J_{h,T} \).

**Theorem 2.2** Suppose that the conditions of Proposition 2.1 hold and \( W_{h,T} \rightarrow_p W_h \) where \( W_h \) is non-random real matrix. Then we have

\[ J_{h,T} \rightarrow_d B'_{H(p+1)} \Omega_{m}^{1/2} \mathbb{P} W_h \mathbb{P} \Omega_{m}^{1/2} B_{H(p+1)} \]

where \( \mathbb{P} \equiv I_{H(p+1)} - \text{diag}(0_{H \times H}, \mathbb{H}(\mathbb{H}' W_g \mathbb{H})^{-1}\mathbb{H}' W_g) \) and \( B_{m,H(p+1)} \) denotes an \((Hp + 1) \times 1\) standard normal random vector.

Theorem 2.2 indicates that the asymptotic distribution of \( J_{h,T} \) is not pivotal. However, its critical values are easy to simulate in practice. When \( W_{g,T} \) is an identity matrix, \( \mathbb{P} \) becomes an idempotent matrix. Hence if we also let \( W_{h,T} \) be an identity matrix, Theorem 2.2 implies that

\[ J_{h,T} \rightarrow_d B'_{H(p+1)} \Omega_{m}^{1/2} \mathbb{P} \Omega_{m}^{1/2} B_{H(p+1)} \]

which makes the simulation of critical values relatively easy. The performance of the test statistics
$J_{m,T}$, $J_{g,T}$ and $J_{h,T}$ are investigated in the simulation study below in the common CH factor model.

### 3 Application to Common CH Factor Model

Multivariate volatility models commonly assume fewer number of conditionally heteroskedastic (CH) factors than the number of assets. A small number of common factors may generate CH behavior of many assets, which can be motivated by economic theories (see, e.g., the early discussion in Engle et al., 1990). Moreover, without imposing the common factor structure, there may be an overwhelming number of parameters to be estimated in the multivariate volatility models. From these theoretical and empirical perspectives, common CH factor models are preferred and widely used. Popular examples include the Factor-GARCH models (Silvennoien and Terasvirta; 2009, Section 2.2) and Factor Stochastic Volatility models (see, e.g., Section 2.2 of Broto and Ruiz (2004) and references therein). It is therefore important to test whether a common CH factor structure exists in multivariate asset returns of interest.

Engle and Kozicki (1993) propose to detect the existence of common CH factor structure, or equivalently the common CH features, using the GMM over-identification 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

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

where $\text{Var}(\cdot|\mathcal{F}_t)$ denotes the conditional variance given all available information $\mathcal{F}_t$ at period $t$, $\Lambda$ is an $n \times p$ matrix ($p \leq n$), $D_t = \text{diag}(\sigma_{1,t}^2, \ldots, \sigma_{p,t}^2)$ is a $p \times p$ diagonal matrix, $\Omega$ 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 $\{\sigma_{k,t}^2\}_{t \geq 0}$ are adapted. The Assumptions 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6 below are from DR.

**Assumption 3.1** $\text{Rank}(\Lambda) = p$ and $\text{Var}[\text{Diag}(D_t)]$ is non-singular.

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

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

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

When $p < n$, e.g., $p = n - 1$, there exists a nonzero $\theta^0_0 \in \mathbb{R}^n$ such that $\theta^0_0 \Lambda = 0$. The real vector $\theta^0_0$ is called the common CH feature in the literature. In the presence of the common CH feature, we have

$$\text{Var}(\theta^0_0 Y_{t+1}|\mathcal{F}_t) = \theta^0_0 \Lambda D_t \Lambda' \theta^0_0 + \theta^0_0 \Omega \theta^0_0 = \theta^0_0 \Omega \theta^0_0 = \text{Constant.} \quad (3.1)$$

Note the CH effects are nullified in the linear combination $\theta^0_0 Y_{t+1}$, while the individual return $Y_{i,t+1}$'s ($i = 1, \ldots, n$) are showing CH volatility. The equations in (3.1) lead to the following
moment conditions:

\[ E \left[ (z_t - \mu_z) (\theta'_s Y_{t+1} Y'_{t+1} \theta_s) \right] = 0_{H \times 1} \text{ when } \theta_s = \theta^0_s \in \mathbb{R}^n, \ \theta^0_s \neq 0. \quad (3.2) \]

where \( \mu_z \) denotes the population mean of \( z_t \). Given the restrictions in (3.2), one can use GMM to estimate the common feature \( \theta^0_s \) and conduct inference about the validity of the moment conditions.

DR have shown that GMM inference using (3.2) is subject to the issue of zero Jacobian moments. The GMM estimator based on (3.2) can therefore be as slow as \( T^{1/4} \) with a nonstandard limiting distribution. As explained earlier, the J-test based on (3.2) 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 \( \sqrt{T} \)-consistent and asymptotically normally distributed GMM estimators and non-conservative 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 \( \Theta^* \subset \mathbb{R}^n \) of parameters that contains at most one unknown common feature \( \theta^0_s \) up to a normalization condition.

**Assumption 3.5** We have \( \theta^*_s \in \Theta^* \subset \mathbb{R}^n \) such that \( \Theta^* = \Theta^*_s \cap \mathcal{N} \) is a compact set and

\[ (\theta^*_s \in \Theta^* \text{ and } \theta^*_s \Lambda = 0_{1 \times p}) \iff (\theta^*_s = \theta^0_s). \]

The non-zero restriction on \( \theta^*_0 \) could be imposed in several ways. For example, the unit cost condition \( \mathcal{N} = \{ \theta \in \mathbb{R}^n, \sum^n_{i=1} \theta_i = 1 \} \) can be maintained without loss of generality\(^1\). To implement this restriction, we define an \( n \times (n - 1) \) matrix \( G_2 \) as

\[
G_2 = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 \\
-1 & -1 & \cdots & -1
\end{pmatrix}_{n \times (n - 1)}
= \begin{pmatrix}
I_{(n-1)} \\
-1_{1 \times (n-1)}
\end{pmatrix}_{n \times (n-1)}.
\]

where \( I \) is an identity matrix and \( 1 \) is a (column) vector of ones. Then we can write

\[
\theta^*_s = \left( \theta_1, \ldots, \theta_{n-1}, 1 - \sum^{n-1}_{i=1} \theta_i \right)' = \left( \theta', 1 - \sum^{n-1}_{i=1} \theta_i \right)' = G_2 \theta + l_n
\]

where \( \theta = (\theta_1, \ldots, \theta_{n-1})' \) is an \( (n - 1) \)-dimensional real vector and \( l_n = (0, \cdots, 0, 1)' \) is an \( n \times 1 \) vector. The following lemma studies the restriction that the existence and uniqueness of \( \theta^0_s \) impose on the factor loading matrix \( \Lambda \).

\(^1\)By definition, \( E[z_t (\theta' Y_{t+1})^2 - E[(\theta' Y_{t+1})^2]] = Cov(z_t, (\theta' Y_{t+1})^2) = E[(z_t - \mu_z) (\theta' Y_{t+1})^2] \). It is clear that we are using the same moment restrictions as in DR.

\(^2\)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 are irrespective of a specific normalization condition.
Lemma 3.1 Suppose that Assumptions 3.1, 3.2, 3.3, 3.4, and 3.5 hold. Then there exists a unique \( \theta^*_0 \) if and only if (i) \( p = n - 1 \) and (ii) \( \Lambda' G_2 \) is invertible. In such case, \( \theta^*_0 \) has the following expression:

\[
\theta^*_0 = \left( \theta'_0, 1 - \sum_{i=1}^{n-1} \theta_{0,i} \right)' = G_2 \theta_0 + l_n
\]

where \( \theta_0 = -(\Lambda' G_2)^{-1} \Lambda' l_n \) is an \((n-1) \times 1\) vector.

In the rest of this section, we assume that there is a unique \( \theta^*_0 \) satisfying the restrictions in (3.3) and (3.2). This means that the moment conditions in (3.3) together with the unit cost restriction (3.2) identify a unique common feature \( \theta^*_0 \). The moment conditions in (3.2) become

\[
\rho(\theta) = E \left[ (z_t - \mu_z) \left( \theta'_0 Y_{t+1} Y_{t+1}' \theta^*_0 \right) \right] = 0_{H \times 1} \text{ when } \theta = \theta_0 \in \mathbb{R}^p,
\]

(3.4)

where the relation between \( \theta^*_0 \) and \( \theta \) is specified in (3.3). We use \( \arg \min_{\theta^*_0 \in \Theta^*} = \arg \min_{\theta \in \Theta^*} \) interchangeably when defining the GMM estimators below.

Assumption 3.6 The vector \( \theta_0 \) belongs to the interior of \( \Theta \subset \mathbb{R}^p \).

Following the notations introduced in Section 2, we define

\[
\psi_t(\theta) \equiv (z_t - \mu_z) \left( \theta'_0 Y_{t+1} Y_{t+1}' \theta^*_0 \right) \text{ and } g_t(\theta) \equiv \text{vec} \left[ \left( \frac{\partial \psi_t(\theta)}{\partial \theta^0} \right) \right]
\]

(3.5)

for any \( \theta \in \Theta \) and any \( t \). Then by definition, we have

\[
\rho(\theta) = E \left[ \psi_t(\theta) \right] \text{ and } g(\theta) = E \left[ g_t(\theta) \right] \text{ for any } \theta \in \Theta.
\]

Under the integrability condition in Assumption 3.3, the nullity of the moment Jacobian occurs at a true common feature \( \theta_0 \) in (3.4) because

\[
\Gamma(\theta_0) = 2E \left[ (z_t - \mu_z) \theta'_0 Y_{t+1} Y_{t+1}' G_2 \right] = 0_{H \times p}.
\]

(3.6)

We consider using both restrictions in (3.4) and (3.6) by stacking them:

\[
m(\theta_0) \equiv E \left[ m_t(\theta_0) \right] \equiv \begin{bmatrix} \rho(\theta_0) \\ g(\theta_0) \end{bmatrix} = 0_{(pH + H) \times 1}.
\]

(3.7)

As discussed in the previous section, the first order local identification of \( \theta_0 \) could be achieved in (3.7), if we could show that the following matrix has full column rank:

\[
\frac{\partial m(\theta_0)}{\partial \theta^0} = E \left[ \begin{bmatrix} \frac{\partial \psi_t(\theta_0)}{\partial \theta^0} \\ \frac{\partial g_t(\theta_0)}{\partial \theta^0} \end{bmatrix} \right] = \begin{bmatrix} 0_{H \times p} \\ 0_{H \times p} \end{bmatrix}.
\]

Lemma 3.2 Under Assumptions 3.1, 3.2, 3.3, 3.4, and 3.5, the matrix \( \mathbb{H} \) has full rank.

9
Lemma 3.3 shows that $\theta_0^*$ 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 $\theta_0^*$ is uniquely identified by (3.4), it is also uniquely identified by (3.6).

**Lemma 3.3** Under Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5, $\theta_0^*$ is uniquely identified by (3.6).

From Lemmas 3.2 and 3.3 we see that $\theta_0^*$ is not only (first-order) locally identified, but also globally identified by the moment conditions in (3.6). As a result, one may only use these moment conditions to estimate the common feature $\theta_0^*$. It is clear that the moment conditions in (3.6) are linear in $\theta$, 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 $\theta_0^*$ is high.

Using the sample average $\tilde{z}$ of $z_t$ as the estimator of $\mu_2$, we construct the feasible moment functions as

$$\tilde{m}_t (\theta) = \begin{bmatrix} \hat{v}_t (\theta) \\ \hat{g}_t (\theta) \end{bmatrix}.$$

The GMM estimator $\hat{\theta}_{m,T}$ is calculated using (2.4) by replacing $m_t (\theta)$ with $\tilde{m}_t (\theta)$ and using the weight matrix $W_{m,T}$ constructed by a first-step GMM estimator with identity weight matrix. From $\theta_0^* = G_2 \theta + l_n$, we can write

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

where we define

$$\mathbb{H}_T \equiv T^{-1} \sum_{t=1}^{T} (2 (z_t - \tilde{z}) \otimes I_p) G'_2 Y_{t+1} Y'_{t+1} G_2$$

and $S_T \equiv T^{-1} \sum_{t=1}^{T} (2 (z_t - \tilde{z}) \otimes I_p) G'_2 Y_{t+1} Y'_{t+1} l_n$. From Lemma 3.2, we see that

$$\hat{\theta}_{m,T} = - \left( \mathbb{H}_T^{'} W_{g,T} \mathbb{H}_T \right)^{-1} \mathbb{H}_T^{'} W_{g,T} S_T.$$  

(3.8)

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

$$\hat{\theta}_{g,T} = - \left( \mathbb{H}_T^{'} W_{g,T} \mathbb{H}_T \right)^{-1} \mathbb{H}_T^{'} W_{g,T} S_T.$$  

(3.9)

In the supplemental appendix (Lee and Liao, 2014), we show that under Assumptions 3.1, 3.2, 3.3, 3.4, 3.5 and Proposition 2.1 holds for the GMM estimators $\hat{\theta}_{m,T}$ and $\hat{\theta}_{g,T}$.

Let $\hat{\Omega}_{g,T}, \hat{\Omega}_{g_\psi,T}$ and $\hat{\Omega}_{\psi,T}$ be the estimators of $\Omega_g$, $\Omega_{g_\psi}$ and $\Omega_\psi$, respectively. The modified GMM estimator $\hat{\theta}_{g^*,T}$ as in (2.6) can be also obtained as

$$\hat{\theta}_{g^*,T} = - \left( \mathbb{H}_T^{'} W_{g^*,T} \mathbb{H}_T \right)^{-1} \mathbb{H}_T^{'} W_{g^*,T} (S_T - F_T).$$  

(3.10)

---

3One can use the first step estimator, for example, $\hat{\theta}_T = - \left( \mathbb{H}_T^{'} \mathbb{H}_T \right)^{-1} \mathbb{H}_T^{'} S_T$ to construct the weight matrices $W_{m,T}, W_{g,T}$ and $W_{g^*,T}$, and the estimators of $\tilde{\Omega}_g$, $\tilde{\Omega}_{g_\psi}$ and $\tilde{\Omega}_\psi$ in this Section.
where $W_{g^*, T} = \hat{\Omega}_{g, T} - \hat{\Omega}_{g^*, T} \hat{\Omega}_{g^*, T}^{-1} \hat{\Omega}_{g^*, T}$ as earlier, and

$$F_T = \hat{\Omega}_{g^*, T} \hat{\Omega}_{g^*, T}^{-1},$$

and

$$A_T = T^{-1} \sum_{t=1}^{T} \hat{\psi}_t (\hat{\theta}_{g, T}) = T^{-1} \sum_{t=1}^{T} (z_t - \bar{z}) \left( (G_2 \hat{\theta}_{g, T} + l_n)' Y_{t+1} Y_{t+1}' (G_2 \hat{\theta}_{g, T} + l_n) \right).$$

In the supplemental appendix, we also show that under Assumptions 3.1 3.2 3.3 3.4 3.5 and 3.6, Theorem 2.1 holds for the modified GMM estimator $\hat{\theta}_{g^*, T}$.

After the GMM estimators $\hat{\theta}_{m, T}$, $\hat{\theta}_{g, T}$ and $\hat{\theta}_{g^*, T}$ are obtained, we can use the J-test statistic defined in (2.7), (2.8) and (2.9) to conduct inference about the existence of common feature $\theta^0$. It is clear that the test based on $J_{g, T}$ is the easiest one to use in practice because $\hat{\theta}_{g, T}$ 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, again, when the dimension of parameter $\theta^0$ is high.

## 4 Simulation Studies

In this section, we investigate the finite sample performances of the proposed GMM estimators and J-tests using the Monte Carlo experiments D3, D4 and D5 from DR. Specifically, $Y_{t+1} = (Y_{1,t+1}, Y_{2,t+1}, Y_{3,t+1})'$ is generated from the following model:

$$Y_{t+1} = \Lambda_k F_{t+1} + u_{t+1}$$

for $k = 1, 2, 3$, where $\Lambda_k$ contains the factor loadings of asset return $Y_{j,t+1}$ in the $k$-th simulation design, $f_{l,t+1}$ ($l = 1, 2, 3$) is generated from a Gaussian generalized autoregressive conditional heteroskedastic model:

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

with

$$\begin{bmatrix}
\omega_1 & \alpha_1 & \beta_1 \\
\omega_2 & \alpha_2 & \beta_2 \\
\omega_3 & \alpha_3 & \beta_3
\end{bmatrix} = \begin{bmatrix}
0.2 & 0.2 & 0.6 \\
0.2 & 0.4 & 0.4 \\
0.1 & 0.1 & 0.8
\end{bmatrix}.$$
\[ \varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})' \] are independent with \( u_s \) for any \( t \) and \( s \) and are i.i.d. from \( N(0, I_3) \) and \( N(0, 0.5I_3) \) respectively.

The simulation designs D3, D4 and D5 are defined via their factor loadings:

\[
\Lambda_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & 0 & 0 \end{bmatrix}, \quad \Lambda_4 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1/2 & 1/2 & 0 \end{bmatrix} \quad \text{and} \quad \Lambda_5 = I_3
\]

respectively. For each simulated sample with sample size \( T \), we generate \( T + 1000 \) observations and drop the first 1000 observations to reduce the effect of the initial conditions of the data generating mechanism on the GMM estimation and inference.

We consider the portfolio \( \theta = (\theta_1, \theta_2, 1 - \theta_1 - \theta_2) \) where \( (\theta_1, \theta_2) \) is a real vector. There are two sets of moment conditions for estimating \( \theta \). The moment conditions proposed in DR are:

\[
E \left[ (z_{t+1} - \mu_z) | Y_{3,t+1} + \theta_1(Y_{1,t+1} - Y_{3,t+1}) + \theta_2(Y_{2,t+1} - Y_{3,t+1}) \right]^2 = 0, \tag{4.1}
\]

where \( z_{t+1} = (Y_{1,t+1}^2, Y_{2,t+1}^2, Y_{3,t+1}^2)' = (z_{1,t+1}, z_{2,t+1}, z_{3,t+1})' \), and the moment conditions defined using the Jacobian of the moment functions in (4.1) are:

\[
E \left[ (z_{t+1} - \mu_z) (Y_{j,t+1} - Y_{3,t+1}) | Y_{3,t+1} + \theta_1(Y_{1,t+1} - Y_{3,t+1}) + \theta_2(Y_{2,t+1} - Y_{3,t+1}) \right] = 0, \tag{4.2}
\]

for \( j = 1, 2 \). Four GMM estimators are studied in D4: (i) the GMM estimator \( \hat{\theta}_{\psi,T} \) based on (4.1); (ii) the GMM estimator \( \hat{\theta}_{m,T} \) based on (4.1) and (4.2); (iii) the efficient GMM estimator \( \hat{\theta}_{g^*,T} \) based on the modified moment conditions of (4.2); and (iv) the GMM estimator \( \hat{\theta}_{g,T} \) based on (4.2).

The finite sample properties of the four GMM estimators in D4 are summarized in Table C.1. In D4, both the moment conditions in (4.1) and (4.2) identify the unique common feature \( \theta_0' = (0, -1, 2) \). Hence, we can evaluate the bias, variance and MSE of the GMM estimators. From Table C.1, we see that: (i) with the growth of the sample size, the bias in \( \hat{\theta}_{m,T}, \hat{\theta}_{g^*,T} \) and \( \hat{\theta}_{g,T} \) goes to zero much faster than \( \hat{\theta}_{\psi,T} \); (ii) the variance of \( \hat{\theta}_{g^*,T} \) is smaller than that of \( \hat{\theta}_{g,T} \) which shows the gain in efficiency of using modified moment conditions with strong IVs; (iii) the finite sample properties of \( \hat{\theta}_{m,T} \) and \( \hat{\theta}_{g^*,T} \) are very similar when the sample size becomes large (e.g., \( T = 5,000 \)), and their variances are almost identical when the sample size is larger than 10,000; (iv) the MSE of \( \hat{\theta}_{\psi,T} \) goes to zero very slowly when compared with \( \hat{\theta}_{m,T}, \hat{\theta}_{g^*,T} \) and \( \hat{\theta}_{g,T} \).

We next investigate the properties of the J-tests in D3, D4 and D5. In addition to the J-tests \( J_{m,T}, J_{g,T} \) and \( J_{h,T} \) proposed in this paper, we also consider the J-test based on \( \hat{\theta}_{\psi,T} \) and the moment conditions in (4.1). Following DR, we consider two critical values: \( \chi_{1-\alpha}^2(H) \) and \( \chi_{1-\alpha}^2(H - p) \) for the last two J-tests at the nominal size \( \alpha \). The curves of the empirical rejection probabilities at each sample size of tests based on \( \chi_{1-\alpha}^2(H) \) and \( \chi_{1-\alpha}^2(H - p) \) are denoted as Ori-GMM1 and Ori-GMM2 respectively. The empirical rejection probabilities of the J-tests \( J_{m,T}, J_{g,T} \) and \( J_{h,T} \) are denoted as

---

\^ As noted in DR, there is no uniquely identified common feature in D3. As a result, the stacked moment conditions can not ensure a uniquely identified common feature either.
Eff-GMM, Jac-GMM and Sim-GMM respectively.

The empirical rejection probabilities of the J-tests in D3 and D4 are depicted in Figure C.1. In the simulation D3, we see that all the J-tests we considered are undersized. As noted in DR, this undersized phenomenon may be explained by the lack of unique identification, or the fact that the IVs used in constructing the moment conditions \((4.1)\) and \((4.2)\) are weak in finite samples. On the other hand, the empirical size properties of the J-tests are well illustrated in D4. From Figure C.1, we see that the over-identification tests based on \(\hat{\theta}_{g,T}\) have nice size control. The test based on \(\hat{\theta}_{m,T}\) has slight over-rejection for each nominal size we considered, and its size converges to the nominal level with the growth of the sample size. Moreover, it is clear that for the J-test statistic based on \(\hat{\theta}_{\psi,T}\), the test using \(\chi^2_{1-\alpha}(H)\) is conservative and undersized, and the test using \(\chi^2_{1-\alpha}(H)\) is over-sized.

From the simulation results in D3 and D4, we see that the J-tests based on \(\hat{\theta}_{g,T}\) have good size control. On the other hand, the tests based on \(\hat{\theta}_{\psi,T}\) with critical values from \(\chi^2_{1-\alpha}(H)\) is undersized. It is easy to see that the undersized test based on \(\chi^2_{1-\alpha}(H)\) suffers from poor power, while the test based on \(\chi^2_{1-\alpha}(H - p)\) leads to over-rejection. It is interesting to check (i) how much power the J-tests based on \(\hat{\theta}_{m,T}\) and \(\hat{\theta}_{g,T}\) gain when compared with the test based on \(\hat{\theta}_{\psi,T}\) and \(\chi^2_{1-\alpha}(H)\), and (ii) whether they are less powerful than the tests based on \(\hat{\theta}_{\psi,T}\) and \(\chi^2_{1-\alpha}(H - p)\).

The empirical rejection probabilities of the J-tests in D5 are depicted in Figure C.2. From Figure C.2, we see that the tests based on \(\hat{\theta}_{g,T}\) are much more powerful than the tests based on \(\hat{\theta}_{\psi,T}\) and \(\chi^2_{1-\alpha}(H)\). The J-test \(J_{h,T}\) is more powerful than the test \(J_{g,T}\) which only uses the moment conditions \((4.2)\). Moreover, the J-test \(J_{h,T}\) is as powerful as the test based on \(\hat{\theta}_{\psi,T}\) and \(\chi^2_{1-\alpha}(H - p)\), which has large size distortion in the finite samples as we have seen in Figure C.1.

## 5 Conclusion

This paper investigates the GMM estimation and inference when the Jacobian of the moment conditions is degenerate. We show that the zero Jacobian contains non-trivial 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. In particular, the GMM estimators using the Jacobian-based moment conditions show remarkably good finite sample properties. Moreover, the J-tests based on the Jacobian GMM estimator have good size control and better power than the commonly used GMM inference which ignores the information contained in the Jacobian moments.

## References

APPENDIX

A Proof of the Main Results in Section 2

Proof of Lemma 2.1. We first notice that by definition,

$$\mathbb{H} = \mathbb{H}(\theta_0) = \frac{\partial g(\theta_0)}{\partial \theta} = (A_1, \ldots, A_H)', \quad (A.1)$$
where $A_h = \frac{\partial^2 \rho_h}{\partial \theta \partial \theta} (\theta_0)$ for $h = 1, \ldots, H$. Because $\theta_0$ is an interior point in $\Theta$, the condition in (ii) is equivalent to

$$\forall x \in \mathbb{R}^p, (x' A_h x)_{1 \leq h \leq H} = 0 \text{ if and only if } x = 0.$$  \hspace{1cm} (A.2)

Now, suppose that $\text{Rank } (H) < p$. Then there exists a non-zero $\tilde{x} \in \mathbb{R}^p$ such that

$$\tilde{x}' H' = (\tilde{x}' A_1, \ldots, \tilde{x}' A_H) = 0_{1 \times pH},$$

which implies that $\tilde{x}' A_h \tilde{x} = 0$ for all $h$. This contradicts (A.2) and hence, we have $\text{rank } (H) = p$.

Assumption A.1 (i) $m(\theta) \equiv E [m_t(\theta)]$ is continuous in $\theta$; (ii) $\sup_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} [m_t(\theta) - m(\theta)] = O_p(T^{-\frac{1}{2}})$; (iii) $\sup_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \left[ \frac{\partial m_t(\theta)}{\partial \theta} - \frac{\partial m(\theta)}{\partial \theta} \right] = o_p(1)$; (iv) $\frac{\partial m(\theta)}{\partial \theta}$ is continuous in $\theta$.

The proof of Proposition 2.1 is standard (see, e.g., Newey and McFadden, 1994) and thus is omitted. We next present the proof of Theorem 2.1.

Proof of Theorem 2.1. By definition,

$$T^{-1} \sum_{t=1}^{T} \hat{g}_{\psi, t}(\theta) = T^{-1} \sum_{t=1}^{T} g_t(\theta) - \hat{\Omega}_{g\psi, T} \hat{\Omega}_{\psi, T}^{-1} T^{-1} \sum_{t=1}^{T} \psi_t(\hat{\theta}_{g, T}).$$  \hspace{1cm} (A.3)

Using the consistency of $\hat{\Omega}_{g\psi, T}$ and $\hat{\Omega}_{\psi, T}$, the $\sqrt{T}$-consistency of $\hat{\theta}_{g, T}$, Assumptions A.1 (i) and (ii), we deduce that

$$\hat{\Omega}_{g\psi, T} \hat{\Omega}_{\psi, T}^{-1} T^{-1} \sum_{t=1}^{T} \psi_t(\hat{\theta}_{g, T}) = o_p(T^{-\frac{1}{2}}).$$  \hspace{1cm} (A.4)

which together with Assumptions 2.1 (ii) and A.1 (ii) implies that

$$\left[ T^{-1} \sum_{t=1}^{T} \hat{g}_{\psi, t}(\theta) \right]' W_{\theta^*, T} \left[ T^{-1} \sum_{t=1}^{T} \hat{g}_{\psi, t}(\theta) \right] - E \left[ g'_t(\theta) \right] \Omega_{\theta, g}^{-1} E [g_t(\theta)] = o_p(1)$$  \hspace{1cm} (A.5)

uniformly over $\theta \in \Theta$, where $\Omega_{\theta, g} = \Omega_g - \Omega_{g\psi} \Omega_{\psi}^{-1} \Omega_{g\psi}$. It is clear that $E [g'_t(\theta) ] \Omega_{\theta, g}^{-1} E [g_t(\theta)]$ is uniquely minimized at $\theta_0$, because $\Omega_{\theta, g}$ is positive definite and $\theta_0$ is identified by $E [g_t(\theta)] = 0$. This together with the uniform convergence in (A.5) and the continuity of $E [g_t(\theta)]$ implies the consistency of $\hat{\theta}_{g^*, T}$.

Next, we note that $\hat{\theta}_{g^*, T}$ satisfies the first order condition

$$\left[ T^{-1} \sum_{t=1}^{T} \frac{\partial g_t(\hat{\theta}_{g^*, T})}{\partial \theta} \right]' W_{\theta^*, T} \left[ T^{-\frac{1}{2}} \sum_{t=1}^{T} \hat{g}_{\psi, t}(\hat{\theta}_{g^*, T}) \right] = 0.$$  \hspace{1cm} (A.6)
Applying the mean value theorem and using the consistency of $\hat{\Omega}_{g,T}$ and $\tilde{\Omega}_{\psi,T}$, Assumptions 2.1 and [A.1] we get

$$
T^{-\frac{1}{2}} \sum_{t=1}^{T} \tilde{g}_{g,t}(\tilde{\theta}_{g,T}) = T^{-\frac{1}{2}} \sum_{t=1}^{T} \left[ g_t(\theta_0) - \Omega_{g,T}^{-1} \psi_t(\theta_0) \right] + T^{-1} \sum_{t=1}^{T} \frac{\partial g_t(\tilde{\theta}_{g,T})}{\partial \theta_t} \left[ \sqrt{T}(\tilde{\theta}_{g,T} - \theta_0) \right] + o_p(1),
$$

(A.7)

where $\tilde{\theta}_{g,T}$ denotes a $p \times Hp$ matrix whose $j$-th ($j = 1, \ldots, Hp$) column represents the mean value (between $\theta_0$ and $\tilde{\theta}_{g,T}$) of the $j$-th moment function in $g_t(\theta)$. Under Assumption 2.1(ii),

$$
T^{-\frac{1}{2}} \sum_{t=1}^{T} \left[ g_t(\theta_0) - \Omega_{g,T}^{-1} \psi_t(\theta_0) \right] \rightarrow_d N \left( 0, \Omega_g - \Omega_{g,T}^{-1} \Omega_{\psi_g} \right).
$$

(A.8)

Using Assumptions [A.1](iii) and (iv), we have

$$
T^{-1} \sum_{t=1}^{T} \frac{\partial g_t(\tilde{\theta}_{g,T})}{\partial \theta_t} \rightarrow_p \mathbb{H} \text{ and } T^{-1} \sum_{t=1}^{T} \frac{\partial m_t(\tilde{\theta}_{g,T})}{\partial \theta_t} \rightarrow_p \mathbb{H}
$$

which together with (A.6), (A.7) and (A.8) proves the claimed result. ■

**Proof of Theorem 2.2.** Applying the mean value theorem, Assumptions 2.1 and A.1 we get

$$
T^{-\frac{1}{2}} \sum_{t=1}^{T} m_t(\tilde{\theta}_{g,T}) = T^{-\frac{1}{2}} \sum_{t=1}^{T} m_t(\theta_0) + T^{-\frac{1}{2}} \sum_{t=1}^{T} \frac{\partial m_t(\tilde{\theta}_{g,T})}{\partial \theta_t} (\tilde{\theta}_{g,T} - \theta_0)
$$

(A.9)

where $\tilde{\theta}_{g,T}$ denotes a $p \times H(p+1)$ matrix whose $j$-th ($j = 1, \ldots, H(p+1)$) column represents the mean value (between $\theta_0$ and $\tilde{\theta}_{g,T}$) of the $j$-th moment function in $m_t(\theta)$. Using Assumptions [A.1](iii) and (iv) and the $\sqrt{T}$-consistency of $\tilde{\theta}_{g,T}$, we get

$$
\left[ \sqrt{T}(\tilde{\theta}_{g,T} - \theta_0) \right] = -(\mathbb{H}'\Omega_g^{-1}\mathbb{H})^{-1}\mathbb{H}'\Omega_g^{-1} \left[ T^{-\frac{1}{2}} \sum_{t=1}^{T} g_t(\theta_0) \right] + o_p(1)
$$

(A.10)

Using the standard arguments of showing Proposition 2.1, we have

$$
T^{-\frac{1}{2}} \sum_{t=1}^{T} m_t(\tilde{\theta}_{g,T}) = \mathbb{P} \left[ T^{-\frac{1}{2}} \sum_{t=1}^{T} m_t(\theta_0) \right] + o_p(1).
$$

(A.11)
B Proof of the Main Results in Section 3

Proof of Lemma 3.1. First, we note that for any $\theta_*$ in (3.3), we can write

$$
\theta_* = G_2 \theta + l_n. \tag{B.1}
$$

Under Assumption 3.5, there exists a unique $\theta^0_*$ if and only if the linear equations $\Lambda' \theta_* = 0_{n \times 1}$ have one and only one solution. Using (B.1), we can write these linear equations as

$$
\Lambda' G_2 \theta = -\Lambda' l_n. \tag{B.2}
$$

It is clear that the above equations have a unique solution if and only if $\Lambda' G_2$ is invertible. Moreover, the unique solution is $\theta_0 = -(\Lambda' G_2)^{-1} \Lambda' l_n$. □

Proof of Lemma 3.2. For any square matrix $A$, we use $\text{Diag}(A)$ to denote the vector which contains all diagonal elements of $A$ started from upper-left to the lower-right. Note that $\text{Diag}(\cdot)$ is different from $\text{diag}(\cdot, \cdot)$ defined in the main text. First, we follow the arguments in Lemma B.4 of the supplemental appendix of DR to write

$$
\rho(\theta) = \text{Cov}(z_t, \text{Diag}(D_t)) * \text{Diag}(\Lambda' \theta_* \theta'_* \Lambda)
$$

$$
= G_1 * \text{Diag}(\Lambda' \theta_* \theta'_* \Lambda) \tag{B.3}
$$

where $G_1 \equiv \text{Cov}(z_t, \text{Diag}(D_t))$ and it is an $H \times p$ matrix with full rank by Assumption 3.3. Let $\Lambda = (\lambda_1, \ldots, \lambda_p)$, where $\lambda_j$’s ($j = 1, \ldots, p$) are $n \times 1$ real vectors. By definition,

$$
\theta'_* \Lambda = (\theta'_* \lambda_1, \ldots, \theta'_* \lambda_p)
$$

which implies that

$$
\text{Diag}(\Lambda' \theta_* \theta'_* \Lambda) = [(\theta'_* \lambda_1)^2, \ldots, (\theta'_* \lambda_p)^2]'.
$$

Hence, there is

$$
\Gamma(\theta) = 2G_1 \left[ (\theta'_* \lambda_1) \lambda_1, \ldots, (\theta'_* \lambda_p) \lambda_p \right]' G_2
$$

where $G_2$ is defined in the main text. Using the Kronecker product, we get

$$
g(\theta) \equiv \text{vec}(\Gamma(\theta)') = 2(G_1 \otimes G_2') \begin{pmatrix}
\lambda_1 \theta'_* \\
\vdots \\
\lambda_p \theta'_* \\
\end{pmatrix}
$$

$$
\tag{B.4}
$$
which further implies that
\[ \mathbb{H} = 2(G_1 \otimes G_2) \begin{pmatrix} \lambda_1 \lambda'_1 \\ \vdots \\ \lambda_p \lambda'_p \end{pmatrix} G_2. \] (B.5)

Let \( G_{1,hj} \) (\( h = 1, \ldots, H \) and \( j = 1, \ldots, p \)) be the \( h \)-th row and \( j \)-th column entry of \( G_1 \). Then we can rewrite the equation (B.5) as
\[ \mathbb{H} = 2 \begin{pmatrix} \sum_{j=1}^{p} G_{1,1j} G_2' \lambda_j \lambda'_j G_2 \\ \vdots \\ \sum_{j=1}^{p} G_{1,Hj} G_2' \lambda_j \lambda'_j G_2 \end{pmatrix}, \]
where \( \sum_{j=1}^{p} G_{1,hj} G_2' \lambda_j \lambda'_j G_2 \) is a \( p \times p \) matrix for any \( h = 1, \ldots, H \).

Suppose that there is an \( \tilde{x} \in \mathbb{R}^p \) such that
\[ \mathbb{H} \tilde{x} = 2 \begin{pmatrix} \sum_{j=1}^{p} G_{1,1j} G_2' \lambda_j \lambda'_j G_2 \tilde{x} \\ \vdots \\ \sum_{j=1}^{p} G_{1,Hj} G_2' \lambda_j \lambda'_j G_2 \tilde{x} \end{pmatrix} = 0_{H \times 1}. \] (B.6)
But \( G_1 \) has full column rank, which means that \( \mathbb{H} \tilde{x} = 0_{H \times 1} \) if and only if
\[ G_2' \lambda_j \lambda'_j G_2 \tilde{x} = 0_{p \times 1} \text{ for all } j. \] (B.7)

The condition in (B.7) implies that
\[ 0 = \sum_{j=1}^{p} G_2' \lambda_j \lambda'_j G_2 \tilde{x} = G_2' \Lambda' G_2 \tilde{x}. \]

We have shown in Lemma 3.1 that \( \Lambda' G_2 \) is invertible, which implies that \( G_2' \Lambda' G_2 \) is also invertible. Hence there must be \( \tilde{x} = 0 \). ■

**Proof of Lemma 3.3** Recall the equation (B.4) in the proof of Lemma 3.2
\[ g(\theta) = 2(G_1 \otimes G_2') \begin{pmatrix} \lambda_1 \lambda'_1 \theta_* \\ \vdots \\ \lambda_p \lambda'_p \theta_* \end{pmatrix}. \] (B.8)

As the common feature \( \theta_*^0 \) satisfies \( \Lambda' \theta_*^0 = 0_{p \times 1} \), we immediately have \( \lambda'_j \theta_*^0 = 0 \) for any \( j = 1, \ldots, p \), which implies that \( g(\theta_0) = 0_{pH \times 1} \). This shows that \( \theta_*^0 \) is one possible solution of the linear equations \( g(\theta) = 0_{pH \times 1} \). By the relation between \( \theta_* = G_2 \theta + l_n \), and the definition of the matrix \( \mathbb{H} \), we can
write the linear equations $g(\theta) = 0_{p \times 1}$ as

$$
\mathbb{H} \theta = -2(G_1 \otimes G_2') \begin{pmatrix} 
\lambda_1 \lambda_1' l_n \\
\vdots \\
\lambda_p \lambda_p' l_n
\end{pmatrix}
$$

which together with the fact that $\mathbb{H}$ is a full rank matrix implies that $\theta^0_*$ is the unique solution.

C Tables and Figures
Table C.1. Finite Sample Properties of the GMM Estimators in D4

<table>
<thead>
<tr>
<th></th>
<th>( \theta_{g,T}(1) )</th>
<th>( \theta_{g^*,T}(1) )</th>
<th>( \theta_{m,T}(1) )</th>
<th>( \theta_{\psi,T}(1) )</th>
<th>( \theta_{g,T}(2) )</th>
<th>( \theta_{g^*,T}(2) )</th>
<th>( \theta_{m,T}(2) )</th>
<th>( \theta_{\psi,T}(2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T=500</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.3912</td>
<td>0.3525</td>
<td>0.3737</td>
<td>0.2499</td>
<td>0.5954</td>
<td>0.5679</td>
<td>0.5914</td>
<td>0.6189</td>
</tr>
<tr>
<td>Variance</td>
<td>0.1686</td>
<td>0.1582</td>
<td>0.1794</td>
<td>0.2071</td>
<td>0.1811</td>
<td>0.1609</td>
<td>0.1772</td>
<td>0.1749</td>
</tr>
<tr>
<td>MSE</td>
<td>0.3216</td>
<td>0.2825</td>
<td>0.3190</td>
<td>0.2695</td>
<td>0.5356</td>
<td>0.4834</td>
<td>0.5269</td>
<td>0.5579</td>
</tr>
<tr>
<td><strong>T=1,000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.3582</td>
<td>0.2891</td>
<td>0.3279</td>
<td>0.2057</td>
<td>0.4544</td>
<td>0.3941</td>
<td>0.4353</td>
<td>0.4897</td>
</tr>
<tr>
<td>Variance</td>
<td>0.1362</td>
<td>0.1253</td>
<td>0.1483</td>
<td>0.1689</td>
<td>0.1431</td>
<td>0.1251</td>
<td>0.1427</td>
<td>0.1154</td>
</tr>
<tr>
<td>MSE</td>
<td>0.2645</td>
<td>0.2088</td>
<td>0.2558</td>
<td>0.2112</td>
<td>0.3495</td>
<td>0.2804</td>
<td>0.3322</td>
<td>0.3552</td>
</tr>
<tr>
<td><strong>T=5,000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.1093</td>
<td>0.0655</td>
<td>0.0797</td>
<td>0.1310</td>
<td>0.1186</td>
<td>0.0797</td>
<td>0.0968</td>
<td>0.3032</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0425</td>
<td>0.0296</td>
<td>0.0334</td>
<td>0.0630</td>
<td>0.0422</td>
<td>0.0274</td>
<td>0.0317</td>
<td>0.0374</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0544</td>
<td>0.0339</td>
<td>0.0397</td>
<td>0.0801</td>
<td>0.0563</td>
<td>0.0337</td>
<td>0.0411</td>
<td>0.1293</td>
</tr>
<tr>
<td><strong>T=10,000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.0513</td>
<td>0.0325</td>
<td>0.0346</td>
<td>0.1138</td>
<td>0.0558</td>
<td>0.0400</td>
<td>0.0450</td>
<td>0.2498</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0214</td>
<td>0.0126</td>
<td>0.0130</td>
<td>0.0370</td>
<td>0.0215</td>
<td>0.0120</td>
<td>0.0118</td>
<td>0.0236</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0240</td>
<td>0.0137</td>
<td>0.0142</td>
<td>0.0499</td>
<td>0.0246</td>
<td>0.0136</td>
<td>0.0138</td>
<td>0.0860</td>
</tr>
<tr>
<td><strong>T=25,000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.0160</td>
<td>0.0144</td>
<td>0.0121</td>
<td>0.0859</td>
<td>0.0168</td>
<td>0.0174</td>
<td>0.0161</td>
<td>0.1962</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0091</td>
<td>0.0047</td>
<td>0.0047</td>
<td>0.0205</td>
<td>0.0094</td>
<td>0.0046</td>
<td>0.0046</td>
<td>0.0137</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0094</td>
<td>0.0049</td>
<td>0.0049</td>
<td>0.0279</td>
<td>0.0097</td>
<td>0.0049</td>
<td>0.0048</td>
<td>0.0522</td>
</tr>
<tr>
<td><strong>T=50,000</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.0074</td>
<td>0.0087</td>
<td>0.0069</td>
<td>0.0741</td>
<td>0.0073</td>
<td>0.0100</td>
<td>0.0087</td>
<td>0.1656</td>
</tr>
<tr>
<td>Variance</td>
<td>0.0047</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.0134</td>
<td>0.0047</td>
<td>0.0023</td>
<td>0.0023</td>
<td>0.0093</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0047</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.0189</td>
<td>0.0047</td>
<td>0.0024</td>
<td>0.0024</td>
<td>0.0367</td>
</tr>
</tbody>
</table>

Notes: 1. The simulation results are based on 10,000 replications; 2. the probability limit of \( \hat{\theta}_{\psi,T} \), \( \hat{\theta}_{m,T} \), \( \hat{\theta}_{g,T} \) and \( \hat{\theta}_{g^*,T} \) are (0,-1); (iii) For the GMM estimators \( \hat{\theta}_{m,T} \), \( \hat{\theta}_{g,T} \) and \( \hat{\theta}_{g^*,T} \), the weight matrices are constructed using the equation (12) in DR and the GMM estimator defined in (3.10) with identity matrix; (iv) for the GMM estimator \( \hat{\theta}_{\psi,T} \), a first-step estimator based on the moment conditions in (4.1) and the identity matrix is calculated and then used to construct the efficient weight matrix.
Figure C.1. The Empirical Rejection Probabilities of the Over-identification Tests in D3 and D4

Notes: 1. The simulation results are based on 10,000 replications; 2. to estimate the empirical size of the tests in different sample sizes, we start with $T = 50$ and move to $T = 500$; we then add 500 more observations each time until $T = 6,000$. 
Figure C.2. The Empirical Rejection Probabilities of the Over-identification Tests in D5

Notes: 1. The simulation results are based on 10,000 replications; 2. to estimate the empirical size of the tests in different sample sizes, we start with $T = 50$ and move to $T = 500$; we then add 500 more observations each time until $T = 15,000$. 