Alternative Over-identifying Restriction Test in GMM with Grouped Moment Conditions*

Kazuhiko Hayakawa
Department of Economics,
Hiroshima University

December 22, 2014

Abstract

This paper proposes a new over-identifying restriction test called the *diagonal J test* in the generalized method of moments (GMM) framework. Different from the conventional over-identifying restriction test, where the sample covariance matrix of moment conditions is used in the weighting matrix, the proposed test uses a block diagonal weighting matrix constructed from the efficient optimal weighting matrix. We show that the proposed test statistic asymptotically follows a weighted sum of chi-square distributions with one degree of freedom. Since we need to decompose the moment conditions into groups when implementing the proposed test, we propose two methods to split the moment conditions. The first is to use K-means method that is widely used in the cluster analysis. The second is to utilize the special structure of moment conditions where they are available sequentially. Such a case typically appears in the panel data models. We also conduct a local power analysis of the diagonal J test in the context of dynamic panel data models, and show that the proposed test has almost the same power as the standard J test in some specific cases. Monte Carlo simulation reveals that the proposed test has substantially better size property than the standard test does, while having almost no power loss in comparison to the standard test when it has no size distortions.

*The author is grateful to Hashem Pesaran, Eiji Kurozumi, Yohei Yamamoto, Naoya Katayama, the participants of SWET2014 in Otaru, Hitotsubashi-Sogang conference on Econometrics 2014 and the seminar at Osaka University for their helpful comments. He also acknowledges the financial support from the Grant-in-Aid for Scientific Research (KAKENHI 25780153) provided by the JSPS.

1 Introduction

Since the seminal work of Hansen (1982), the generalized method of moments (GMM) estimator has been widely used in empirical economic studies including macro- and micro-econometrics and finance[1]. However, since the early 1990s, the finite sample behavior of the GMM estimator has been called into question and a great number of studies have been conducted so far in this regard. Broadly speaking, there are two factors that affect the finite sample behavior of the GMM estimator: the strength of identification and the number of moment conditions. To overcome the poor finite-sample properties of the GMM estimator, alternative estimation procedures such as the generalized empirical likelihood estimator (e.g., Newey and Smith, 2004) have been proposed. There are also several studies that attempt to improve the GMM estimator. For instance, Windmeijer (2005) proposes a bias-corrected standard error for the two-step GMM estimator, while Stock and Wright (2000) and Kleibergen (2005) propose inferential methods that are robust to the strength of identification.

While many of the previous studies have focused on the estimation and inference of parameters, since it is common practice to check the validity of moment conditions, we discuss the behavior of the over-identifying restriction test, which is often called the J or Sargan/Hansen test[2]. Early contributions in this line are Tauchen (1986), Kocherlakota (1990), and Andersen and Sørensen (1996), who showed that the J test often suffers from substantial size distortions. More recently, Anatolyev and Gospodinov (2010), Lee and Okui (2012), and Chao, Hausman, Newey, Swanson and Woutersen (2013) proposed new J tests that are valid even when the number of instruments is large in cross-sectional regression models. Using a general model, Newey and Windmeijer (2009) demonstrated that the J test is valid even under many weak moments asymptotics. However, since they assumed $m^2/n \to 0$ or an even stronger restriction, where m is the number of moment conditions and n is the sample size, the J test might lose validity when m is comparable with n. In fact, Bowsher (2002) demonstrated that a size distortion problem occurs in the GMM estimation of dynamic panel data models with a large number of moment conditions. Roodman (2009) provided cautionary empirical examples associated with the GMM estimation of panel data models.

The purpose of this paper is to contribute to this literature. Specifically, we propose a new J test that has good size property in finite samples even when the number of moment conditions is large. Although the number of moment conditions could be large, the asymptotic framework used is the standard one: the sample size tends to infinity with the number of moment conditions being fixed. As noted in Windmeijer (2005), Roodman (2009), Bai and Shi (2011), and others, when the number of moment conditions is large, an important problem of the GMM is in the estimation of the optimal weighting matrix. In statistics literature, there are voluminous studies on the estimation of the covariance matrix in which the dimension is comparable to or even larger than the sample size[3]. One of the important consequences is that the large dimensional sample covariance matrix is a poor estimator for the true covariance matrix. Furthermore, as will be discussed later, using the inverse of such a large dimensional covariance matrix makes the problem much more serious. This implies that the J test, which involves the inverse of the covariance matrix, does not perform well when the number of moment conditions is large relative to the sample size.

[1] Hall (2005) is a useful reference for GMM.

[2] In what follows, we call this the J test.

[3] See, for example, Ledoit and Wolf (2004), Fan, Fan and Lv (2008), Bickel and Levina (2008), and Fan, Liao and Mincheva (2011).

To address the problem associated with inverting a large dimensional covariance matrix, we propose the use of a *block diagonal* matrix in which the dimension of each block is small compared to the sample size. Since the inverse of a block diagonal matrix only requires the inversion of each small block, we can avoid inverting the large dimensional matrix. By using such a block diagonal matrix instead of the conventional covariance matrix, we propose a new J test, which we call the *diagonal J test*. We show that, different from the standard J test, the diagonal J test asymptotically follows a weighted sum of chi-square distribution with one degree of freedom. When implementing the proposed test, we need to split the moment conditions into several groups. For this, two approaches are proposed. The first is to use so-called K-means method, which is often used in the literature on cluster analysis. The second is to utilize the special structure of moment conditions that are available sequentially[4]. This typically arises in the panel data models. We also conduct a local power analysis in the context of the dynamic panel data models, and show that the diagonal J test based on models in forward orthogonal deviations has almost the same power as the standard J test. Monte Carlo simulations are carried out to investigate the finite sample behavior of the new test. The simulation results show that while the proposed test dramatically improves the size property, it has almost the same power as the standard J test when the empirical size is correct.

The rest of this paper is organized as follows. In Section 2, we introduce the model and assumptions, and review the GMM. In Section 3, we propose a new test and derive its asymptotic distribution. In Section 4, we propose two methods to decompose the moment conditions. In Section 5, we conduct a local power analysis of the standard and diagonal J tests in the context of dynamic panel data models. In Section 6, we conduct Monte Carlo simulations to assess the performance of the proposed test. Finally, we conclude in Section 7.

2 Review of GMM estimators and the J test

GMM estimators

Let us consider moment conditions $E[\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta}_0)] = E[\mathbf{g}_i(\boldsymbol{\theta}_0)] = \mathbf{0}$, where $\mathbf{g}(\cdot,\cdot)$ is an $m \times 1$ known function, $\{\mathbf{x}_i\}_{i=1}^N$ are independent observations, and $\boldsymbol{\theta}_0$ is the true value of a $p \times 1$ vector of unknown parameters.

One-step, two-step, and continuous-updating (CU-) GMM estimators are defined as[5]

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}_{1step} &= \underset{\theta}{\operatorname{argmin}}\ \widehat{\mathbf{g}}(\boldsymbol{\theta})'\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{g}}(\boldsymbol{\theta}), \\
\widehat{\boldsymbol{\theta}}_{2step} &= \underset{\theta}{\operatorname{argmin}}\ \widehat{\mathbf{g}}(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\theta}})^{-1}\widehat{\mathbf{g}}(\boldsymbol{\theta}), \\
\widehat{\boldsymbol{\theta}}_{CU} &= \underset{\theta}{\operatorname{argmin}}\ \widehat{\mathbf{g}}(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}(\boldsymbol{\theta})^{-1}\widehat{\mathbf{g}}(\boldsymbol{\theta}),
\end{aligned} \tag{1}$$

where $\widehat{\mathbf{W}}$ is a positive-definite matrix, $\widetilde{\boldsymbol{\theta}}$ is a preliminary consistent estimator of $\boldsymbol{\theta}$, and

$$\widehat{\mathbf{g}}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^N \mathbf{g}_i(\boldsymbol{\theta}), \qquad \widehat{\boldsymbol{\Omega}}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^N \widetilde{\mathbf{g}}_i(\boldsymbol{\theta})\widetilde{\mathbf{g}}_i(\boldsymbol{\theta})', \qquad \widetilde{\mathbf{g}}_i(\boldsymbol{\theta}) = \mathbf{g}_i(\boldsymbol{\theta}) - \widehat{\mathbf{g}}(\boldsymbol{\theta}). \tag{2}$$

We make the following standard assumptions, which are cited from Arellano (2003).

[4] Chamberlain (1992) calls these the *sequential moment conditions.*
[5] The CU-GMM estimator was proposed by Hansen, Heaton and Yaron (1996).

Assumption 1. *(i) The parameter space* $\boldsymbol{\Theta}$ *for* $\boldsymbol{\theta}$ *is a compact subset of* $\mathbb{R}^p$ *such that* $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ *and* $\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})$ *is continuous in* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ *for each* $\mathbf{x}_i$*. (ii)* $\boldsymbol{\theta}_0$ *is in the interior of* $\boldsymbol{\Theta}$ *and* $\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})$ *is continuously differentiable in* $\boldsymbol{\Theta}$*.*

Assumption 2. $\widehat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$*, where* $\mathbf{W}$ *is positive-definite.*

Assumption 3. $E[\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})]$ *exists for all* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*, and* $\mathbf{W}E[\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})] = \mathbf{0}$ *only if* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$*.*

Assumption 4. $\widehat{\mathbf{g}}(\boldsymbol{\theta})$ *converges in probability uniformly in* $\boldsymbol{\theta}$ *to* $E[\mathbf{g}(\mathbf{x}_i, \boldsymbol{\theta})]$*.*

Assumption 5. $\widehat{\boldsymbol{G}}(\boldsymbol{\theta}) = \partial \widehat{\mathbf{g}}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'$ *converges in probability uniformly in* $\boldsymbol{\theta}$ *to a non-stochastic matrix* $\mathbf{G}(\boldsymbol{\theta})$*, and* $\mathbf{G}(\boldsymbol{\theta})$ *is continuous at* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$*.*

Assumption 6. $\sqrt{N}\widehat{\mathbf{g}}(\boldsymbol{\theta}_0)$ *satisfies the central limit theorem:*

$$\sqrt{N}\widehat{\mathbf{g}}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{g}_i(\boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}),$$

where the positive-definite matrix $\boldsymbol{\Omega}$ *is defined as*

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\theta}_0) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E[\mathbf{g}_i(\boldsymbol{\theta}_0)\mathbf{g}_i(\boldsymbol{\theta}_0)'].$$

Assumption 7. *For* $\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}_0)$*,* $\mathbf{G}'\mathbf{W}^{-1}\mathbf{G}$ *is non-singular.*

Under these assumptions, as required for the derivation of the distribution of the J test, the GMM estimators are consistent and asymptotically normal.

J test

In the empirical studies where GMM is used, we first check the validity of moment conditions through the J test prior to performing an inference of the parameters of interest. The J test statistic is typically computed as

$$\begin{aligned} \mathcal{J}(\widehat{\boldsymbol{\theta}}_{2step}) &= N \cdot \widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_{2step})'\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\theta}}_{1step})^{-1}\widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_{2step}), \quad (3) \\ \mathcal{J}(\widehat{\boldsymbol{\theta}}_{CU}) &= N \cdot \widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_{CU})'\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\theta}}_{CU})^{-1}\widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_{CU}). \quad (4) \end{aligned}$$

Note that using the centered weighting matrix is important when constructing the test statistic, since an uncentered weighting matrix $N^{-1}\sum_{i=1}^{N} \mathbf{g}_i(\boldsymbol{\theta})\mathbf{g}_i(\boldsymbol{\theta})'$ is not the variance of the moment conditions when they are invalid[6]. Then, under $H_0 : E[\mathbf{g}_i(\boldsymbol{\theta}_0)] = \mathbf{0}$, we have

$$\mathcal{J}(\widehat{\boldsymbol{\theta}}_{2step}), \mathcal{J}(\widehat{\boldsymbol{\theta}}_{CU}) \xrightarrow{d} \chi^2_{m-p}.$$

Problems with the J test

In the literature, the poor performance of the J test with many moment conditions has been demonstrated; see, for example, Bowsher (2002) and Roodman (2009). To illustrate why the standard J test does not work when the number of moment conditions is large, let us consider a simple example. Let $m \times 1$ vectors $\mathbf{x}_1, \cdots, \mathbf{x}_N$ be a random sample from $\mathcal{N}_m(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, and define

[6]See Hall (2000) and Andrews (1999, p.549).

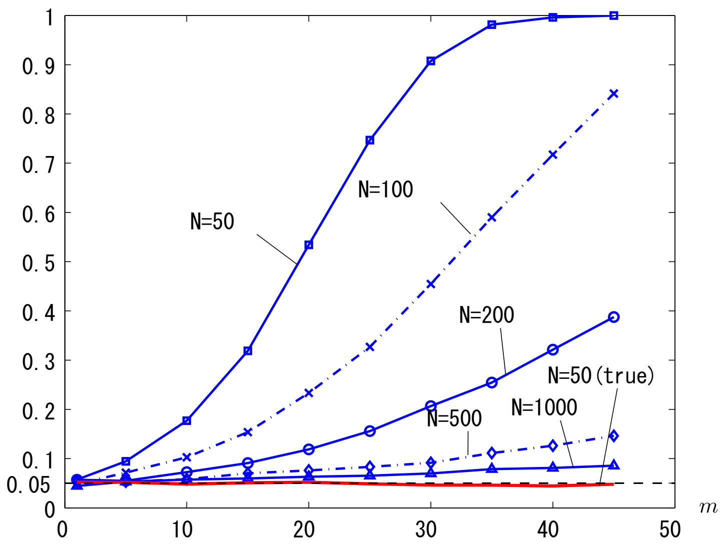

Figure 1: Empirical size of the test for mean vector

$\bar{\mathbf{x}} = N^{-1}\sum_{i=1}^{N}\mathbf{x}_i$ and $\mathbf{S} = (N-1)^{-1}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$. Further, assume that $m < N - 1$. Then, it is straightforward that $E(\mathbf{S}) = \mathbf{\Sigma}_0$, while[7]

$$E(\mathbf{S}^{-1}) = \frac{N}{N-m-2}\mathbf{\Sigma}_0^{-1}. \tag{5}$$

This implies that although the sample covariance $\mathbf{S}$ is an unbiased estimator of $\mathbf{\Sigma}_0$, $\mathbf{S}^{-1}$ is not an unbiased estimator of $\mathbf{\Sigma}_0^{-1}$. Importantly, the bias gets larger as m approaches N. To show how serious this is, let us consider a test of mean vector $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$. The test statistic is given by $H = N \cdot (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)$, which asymptotically follows χ^2_m with a fixed m and a large N[8]. Figure 1 shows the simulation results for the case of $m = 1, 5, 10, ..., 45$ with $N = 50, 100, 500, 1000$ based on 5,000 replications. The significance level is 5%. From the figure, we find that size distortion becomes larger as m increases for each N. Further, we find that as N increases, the size property improves. However, it should be noted that size distortions remain when $m = 45$, even for the case of $N = 500, 1000$. We also computed the size by using true covariance $\mathbf{\Sigma}_0$ instead of sample covariance $\mathbf{S}$ for the case of $N = 50$[9]. From the figure, we find that there are no size distortions for any value of m. This clearly shows that the estimation of the covariance matrix results in a significant problem.

While this example is obtained in a very restrictive situation, a similar problem could occur for the J test as well since the basic structure of the test statistic is very similar to this example. In fact, the simulation results in Section 6 show substantial size distortion of the J test when the number of moment conditions is large relative to the sample size. This motivates an alternative test that can address the size distortion problem.

[7] See Anderson (2003, p.273).

[8] The exact distribution of H is an F distribution.

[9] In this case, the exact distribution of H is a χ^2 distribution.

3 New J test

In this section, we propose a new J test that addresses the problems associated with the covariance matrix estimation. The above analysis shows that the size distortion problem of the J test could be associated with the estimation of the inverse of the large-dimensional covariance matrix. There are two possible approaches to address the problem. The first approach is to simply reduce the number of moment conditions so that the reduced dimension is sufficiently small compared to the sample size. This is easy and straightforward to implement in practice; moreover, at the cost of efficiency, we can expect the bias of estimates to be small. However, as will be demonstrated in Section 6, where the new test is applied to the GMM estimation of dynamic panel data models, this approach does not always work well in terms of power owing to the weak instruments problem. The second approach is to use an improved covariance matrix estimator that works well even when the dimension is large relative to the sample size. In recent statistics literature, there are many studies that have proposed improved estimations of a large-dimensional covariance matrix(e.g. Ledoit and Wolf, 2012). However, most of their analyses are valid under a strong assumption that observations are *iid* and normal, need a tuning parameter for implementation, or impose certain structures on the variance matrix for dimension reduction, such as sparsity or factor structures. Unfortunately, these are non-trivial for the GMM setting.

Therefore, we propose a different approach that has the following distinctive features: (i) it does not require the choice of tuning parameters or specific structure, (ii) it is simple to implement, and (iii) it is asymptotically equivalent to the standard J test, or is as powerful as the standard J test in certain cases. Since the problem lies in inverting a large-dimensional matrix, we propose the use of a block diagonal matrix for the weighting matrix in which the dimension of each block is not large. Since $\mathbf{A}^{-1} = \text{diag}(\mathbf{A}_1^{-1}, ...\mathbf{A}_q^{-1})$ for a block diagonal matrix $\mathbf{A} = \text{diag}(\mathbf{A}_1, ...\mathbf{A}_q)$, we can avoid the inversion of a large-dimensional matrix by using the block diagonal matrix.

For this, we first decompose the moment conditions $\mathbf{g}_i(\boldsymbol{\theta})$ into q groups:

$$\mathbf{g}_i(\boldsymbol{\theta}) = \left[\mathbf{g}_{i1}(\boldsymbol{\theta})', \mathbf{g}_{i2}(\boldsymbol{\theta})', ..., \mathbf{g}_{iq}(\boldsymbol{\theta})'\right]', \tag{6}$$

where $\mathbf{g}_{ij}(\boldsymbol{\theta}), (j = 1, 2, ..., q)$ is $m_j \times 1$ with $\sum_{j=1}^{q} m_j = m$. In practice, how to choose the groups is an important issue. This problem will be discussed in the next section.

Based on this decomposition, we consider the following weighting matrix:

$$\widehat{\boldsymbol{\Omega}}_{diag}(\boldsymbol{\theta}) = \text{diag}\left[\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{g}}_{i1}(\boldsymbol{\theta})\widetilde{\mathbf{g}}_{i1}(\boldsymbol{\theta})', \frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{g}}_{i2}(\boldsymbol{\theta})\widetilde{\mathbf{g}}_{i2}(\boldsymbol{\theta})', ..., \frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{g}}_{iq}(\boldsymbol{\theta})\widetilde{\mathbf{g}}_{iq}(\boldsymbol{\theta})'\right], \tag{7}$$

where $\widetilde{\mathbf{g}}_{ij}(\boldsymbol{\theta}) = \mathbf{g}_{ij}(\boldsymbol{\theta}) - N^{-1}\sum_{i=1}^{N}\mathbf{g}_{ij}(\boldsymbol{\theta})$, $(j = 1, ..., q)$. Note that this weighting matrix is constructed from $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\theta})$ by letting $\frac{1}{N}\sum_{i=1}^{N}\widetilde{\mathbf{g}}_{ik}(\boldsymbol{\theta})\widetilde{\mathbf{g}}_{il}(\boldsymbol{\theta})' = \mathbf{0}$ for $k \neq l$. Hence, this weighting matrix ignores the correlation between the different groups of moment conditions. If all the groups of moment conditions are mutually uncorrelated, that is, $E[\mathbf{g}_{ik}(\boldsymbol{\theta})\mathbf{g}_{il}(\boldsymbol{\theta})'] = \mathbf{0}$ for $k \neq l$, then $\widehat{\boldsymbol{\Omega}}_{diag}(\boldsymbol{\theta})$ is a consistent estimator of the optimal weighting matrix since $\boldsymbol{\Omega} = \boldsymbol{\Omega}_{diag} \equiv \text{plim}_{N\to\infty}\widehat{\boldsymbol{\Omega}}_{diag}(\boldsymbol{\theta}_0)$.

Since $\widehat{\boldsymbol{\Omega}}_{diag}$ is a block diagonal matrix, when computing an inverse, we need to invert an $m^* = \max_{1\leq j\leq q} m_j$-dimensional matrix at the maximum. This dimension can be substantially smaller than m. Thus, by using weighting matrix $\widehat{\boldsymbol{\Omega}}_{diag}$, we can address the problem of inverting a large-dimensional weighting matrix when constructing the J test. Using $\widehat{\boldsymbol{\Omega}}_{diag}$, we propose the following test statistic:

$$\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j) = N \cdot \widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j)' \widehat{\boldsymbol{\Omega}}_{diag}(\widehat{\boldsymbol{\theta}}_j)^{-1} \widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j), \qquad (j = 1step, 2step, CU). \tag{8}$$

Since this statistic uses a block diagonal matrix, we call it the *diagonal J test* or *diagonal Sargan/Hansen test*. Since $\widehat{\boldsymbol{\Omega}}_{diag}(\widehat{\boldsymbol{\theta}}_j)$ is not the estimated covariance matrix of the moment conditions $E[\mathbf{g}_i(\boldsymbol{\theta}_0)] = \mathbf{0}$ in general, $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j)$ does not follow the standard chi-square distribution. The asymptotic distribution of this test statistic is given in the following theorem. The proof is given in the appendix.

Theorem 1. *Let Assumptions 1 to 7 hold. Then, under* $H_0 : E[\mathbf{g}_i(\boldsymbol{\theta}_0)] = \mathbf{0}$*, as* $N \to \infty$*, we have*

$$\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j) \ \overset{d}{\to} \ \sum_{k=1}^{m-p} \lambda_{j,k} z_k^2, \qquad (j = 1step, 2step, CU)$$

where $z_k \sim iid\mathcal{N}(0,1)$ *and* $\lambda_{j,k}, (j = 1step, 2step, CU; k = 1, ..., m-p)$ *are non-zero eigenvalues of*

$$\mathbf{H}_j = \boldsymbol{\Omega}^{1/2} \left[\mathbf{I}_m - \mathbf{K}_j\right]' \boldsymbol{\Omega}_{diag}^{-1} \left[\mathbf{I}_m - \mathbf{K}_j\right] \boldsymbol{\Omega}^{1/2}, \qquad (j = 1step, 2step, CU) \tag{9}$$

with

$$\begin{aligned} \mathbf{K}_{1step} &= \mathbf{G}\left(\mathbf{G}'\mathbf{W}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{W}^{-1}, \\ \mathbf{K}_{2step} &= \mathbf{K}_{CU} = \mathbf{G}\left(\mathbf{G}'\boldsymbol{\Omega}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\boldsymbol{\Omega}^{-1}. \end{aligned}$$

Note that asymptotics is taken with a large N and a fixed m as the purpose of this paper is to propose a new test with accurate size in a finite sample. Investigating many-moments asymptotics is beyond the scope of the current paper.

Remark 1. Since the one-step GMM and the two-step and CU-GMM estimators have different asymptotic distributions, the distributions of the corresponding test statistics too differ.

Remark 2. A similar result to Theorem 1 is obtained in Jagannathan and Wang (1996) and Parker and Julliard (2005). However, there are several differences between these two papers and Theorem 1 as summarized in Table 1. First, while the two studies use the same weighting matrix in the estimation and construction of the J test statistic, our test uses a different weighting matrix in the estimation and construction of the test statistic. For estimation, we use $\widehat{\mathbf{W}}$ for one-step GMM and $\widehat{\boldsymbol{\Omega}}$ for two-step and CU-GMM, while $\widehat{\boldsymbol{\Omega}}_{diag}(\boldsymbol{\theta})$ is used only when the test statistic is constructed[10]. Second, in the two studies, the weighting matrix does not depend on unknown parameters, whereas in our case, the weighting matrix depends on unknown parameters since it is a subset of a conventional optimal weighting matrix.

Remark 3. If $\boldsymbol{\Omega} \neq \boldsymbol{\Omega}_{diag}$, the test statistic $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j)$ $(j = 1step, 2step, CU)$ always follows the nonstandard distribution as described above. However, in some special cases, $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j)$ $(j = 1step, 2step, CU)$ follows the standard χ^2 distribution. If $\boldsymbol{\Omega}_{diag} = \boldsymbol{\Omega}$ but $\mathbf{W} \neq \boldsymbol{\Omega}$, we have $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j) \overset{d}{\to} \chi^2_{m-p}$ for $j = 2step, CU$. Furthermore, if $\boldsymbol{\Omega}_{diag} = \boldsymbol{\Omega} = \mathbf{W}$, we have $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j) \overset{d}{\to} \chi^2_{m-p}$ for $j = 1step, 2step, CU$. In these special cases, the diagonal J test follows the χ^2 distribution as in the standard J test. Such cases appear in the analysis of panel data models (see Section 4.2).

[10] Although it is possible to use the block diagonal matrix in estimation and inference, such an approach is not pursued in this paper.

Table 1: Comparison of three tests

Test	Estimator	Test statistic and its distribution
Standard J test	$\widehat{\boldsymbol{\theta}}_{2step} = \underset{\theta}{\text{argmin}}\ \widehat{\mathbf{g}}(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\theta}}_{1step})^{-1}\widehat{\mathbf{g}}(\boldsymbol{\theta})$,	$\mathcal{J}_{2step} = N \cdot \widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_{2step})'\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\theta}}_{1step})^{-1}\widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_{2step}) \xrightarrow{d} \chi^2_{m-p}$
HJ test	$\widehat{\boldsymbol{\theta}}_{1step} = \underset{\theta}{\text{argmin}}\ \widehat{\mathbf{g}}(\boldsymbol{\theta})'\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{g}}(\boldsymbol{\theta})$	$\mathcal{J}_{HJ} = N \cdot \widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_{1step})'\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_{1step}) \xrightarrow{d} \sum_{k=1}^{m-p} \lambda_{HJ,k} z_k^2$ where $z_k \sim iid\mathcal{N}(0,1)$ and $\lambda_{HJ,k}$ are non-zero eigenvalues of $\mathbf{H}_{HJ} = \boldsymbol{\Omega}^{1/2}\left[\mathbf{I}_m - \mathbf{K}_{1step}\right]'\mathbf{W}^{-1}\left[\mathbf{I}_m - \mathbf{K}_{1step}\right]\boldsymbol{\Omega}^{1/2}$
Diagonal J test	$\widehat{\boldsymbol{\theta}}_{1step} = \underset{\theta}{\text{argmin}}\ \widehat{\mathbf{g}}(\boldsymbol{\theta})'\widehat{\mathbf{W}}^{-1}\widehat{\mathbf{g}}(\boldsymbol{\theta})$ $\widehat{\boldsymbol{\theta}}_{2step} = \underset{\theta}{\text{argmin}}\ \widehat{\mathbf{g}}(\boldsymbol{\theta})'\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\theta}}_{1step})^{-1}\widehat{\mathbf{g}}(\boldsymbol{\theta})$	$\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j) = N \cdot \widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j)'\widehat{\boldsymbol{\Omega}}_{diag}(\widehat{\boldsymbol{\theta}}_j)^{-1}\widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j) \xrightarrow{d} \sum_{k=1}^{m-p} \lambda_{j,k} z_k^2$. where $z_k \sim iid\mathcal{N}(0,1)$ and $\lambda_{j,k}$ are non-zero eigenvalues of $\mathbf{H}_j = \boldsymbol{\Omega}^{1/2}\left[\mathbf{I}_m - \mathbf{K}_j\right]'\boldsymbol{\Omega}_{diag}^{-1}\left[\mathbf{I}_m - \mathbf{K}_j\right]\boldsymbol{\Omega}^{1/2}$ for $j = 1step, 2step$

Note: "HJ test" stands for the Hansen–Jagannathan test. $\mathbf{K}_j, (j = 1step, 2step)$ is defined in Theorem 1. The results for the CU-GMM estimator for the standard and diagonal J tests are identical to those for the two-step GMM estimator.

Computation of critical values

Since the asymptotic distribution is a weighted sum of the chi-square distribution with one degree of freedom, the standard chi-square distribution cannot be used to compute the critical values. There are two approaches to compute the critical values for the current case. The first is to use a simulation method as outlined in Jagannathan and Wang (1996), and the second is to use analytical results. Since the details of the former approach are given in Appendix C of Jagannathan and Wang (1996), we consider the latter approach in detail. There are numerous studies that examine the computation of distribution functions of general quadratic forms of normal variables[11]. Among them, we employ the procedure by Imhof (1961) because it is almost exact. The following lemma, which is used to compute both the critical values and local powers in Section 5, summarizes the results of Imhof (1961).

Lemma 1. *Let* $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ *and* $\mathbf{z} = \boldsymbol{\Omega}^{-1/2}\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$*, where* $\mathbf{x}$ *and* $\mathbf{z}$ *are* $m \times 1$ *functions. Define a quadratic form* $Q = (\mathbf{x}+\mathbf{b})'\mathbf{A}(\mathbf{x}+\mathbf{b})$ *and let* $\mathbf{B} = \boldsymbol{\Omega}^{1/2}\mathbf{A}\boldsymbol{\Omega}^{1/2} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}'$*, where* $\boldsymbol{\Lambda}$ *is a diagonal matrix whose diagonal elements are eigenvalues of* $\mathbf{B}$ *and each column of* $\mathbf{P}$ *is the corresponding eigenvector with* $\mathbf{P}'\mathbf{P} = \mathbf{I}_m$*. Using these, the quadratic form* Q *can be written as*

$$Q = \left(\mathbf{z} + \boldsymbol{\Omega}^{-1/2}\mathbf{b}\right)'\mathbf{B}\left(\mathbf{z} + \boldsymbol{\Omega}^{-1/2}\mathbf{b}\right) \stackrel{d}{=} (\mathbf{z}+\boldsymbol{\delta})'\boldsymbol{\Lambda}(\mathbf{z}+\boldsymbol{\delta}) = \sum_{k=1}^{m} \lambda_k (z_k + \delta_k)^2 = Q(\boldsymbol{\lambda}, \boldsymbol{\delta}), \quad (10)$$

where $z_k \sim iid\mathcal{N}(0,1)$*,* $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_m)'$*, and* δ_k *is the k-th element of* $\boldsymbol{\delta} = \mathbf{P}'\boldsymbol{\Omega}^{-1/2}\mathbf{b}$*. Imhof (1961) showed the following result:*

$$P(Q \le x) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin\theta(v)}{v\rho(v)}dv, \quad (11)$$

where

$$\theta(v) = \frac{1}{2}\sum_{j=1}^{m} r_j \tan^{-1}(\lambda_j v) + \delta_j^2\lambda_j v(1+\lambda_j^2 v^2)^{-1} - \frac{vx}{2},$$

[11] See Ullah (2004, pp.52-56) for a brief survey

$$\rho(v) = \prod_{j=1}^{m} \left(1+\lambda_j^2 v^2\right)^{r_j/4} \exp\left\{ \frac{1}{2}\sum_{j=1}^{m} (\delta_j \lambda_j v)^2/(1+\lambda_j^2 v^2), \right\}$$

with r_j being the multiplicities of the non-zero distinct $\lambda_j s$.

The critical value at significance level α, denoted as cv_α, can be obtained from (10) and (11) such that $P(Q \leq cv_\alpha) = 1-\alpha$, where $\lambda_{j,k}, (j = 1step, 2step, CU)$ are now eigenvalues of $\mathbf{H}_j, (j = 1step, 2step, CU)$ given in (9), with $\boldsymbol{\delta} = \mathbf{b} = \mathbf{0}$[12].

Asymptotic distribution under local alternatives

Next, to investigate the local power properties of the diagonal J test, we derive the asymptotic distribution under local alternatives $H_1 : E[\mathbf{g}_i(\boldsymbol{\theta}_0)] = \mathbf{c}/\sqrt{N}$, where $\mathbf{c}$ is a finite constant vector.

Theorem 2. *Let Assumptions 1 to 7 hold. Then, under $H_1 : E[\mathbf{g}_i(\boldsymbol{\theta}_0)] = \mathbf{c}/\sqrt{N}$, where $\mathbf{c}$ is a finite constant vector, as $N \to \infty$, we have*

$$\mathcal{J}(\widehat{\boldsymbol{\theta}}_j) \overset{d}{\to} \sum_{k=1}^{m-p} (z_k + \delta_{j,k})^2 = Q\left(\mathbf{1}_{m-p}, \boldsymbol{\delta}_j\right), \qquad (j = 2step, CU) \tag{12}$$

$$\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j) \overset{d}{\to} \sum_{k=1}^{m-p} \lambda_{j,k}^{diag} \left(z_k + \delta_{j,k}^{diag}\right)^2 = Q\left(\boldsymbol{\lambda}_j^{diag}, \boldsymbol{\delta}_j^{diag}\right), \qquad (j = 1step, 2step, CU) \tag{13}$$

where $z_k \sim iid\mathcal{N}(0,1)$, $\mathbf{1}_{m-p} = (1,...,1)'$ $\delta_{j,k}$ is the k-th element of $\boldsymbol{\delta}_j = \mathbf{P}_j'\boldsymbol{\Omega}^{-1/2}\mathbf{c}$, where $\mathbf{P}_j$ is a matrix of eigenvectors of $\mathbf{I}_m - \boldsymbol{\Omega}^{-1/2}\mathbf{G}(\mathbf{G}'\boldsymbol{\Omega}^{-1}\mathbf{G})^{-1}\mathbf{G}'\boldsymbol{\Omega}^{-1/2}$ corresponding to non-zero eigenvalues. $\boldsymbol{\lambda}_j^{diag} = (\lambda_{j,1}^{diag}, ..., \lambda_{j,m-p}^{diag})$ is a vector of non-zero eigenvalues of $\mathbf{H}_j$ given by (9), and $\delta_{j,k}^{diag}$ is the k-th element of $\boldsymbol{\delta}_j^{diag} = \mathbf{P}_j^{diag\prime}\boldsymbol{\Omega}^{-1/2}\mathbf{c}$, where $\mathbf{P}_j^{diag}$ is a matrix of eigenvectors of $\mathbf{H}_j$ corresponding to $\lambda_{j,1}^{diag}, ..., \lambda_{j,m-p}^{diag}$.

Note that $\mathcal{J}(\widehat{\boldsymbol{\theta}}_j)$ follows the non-central chi-square distribution with non-centrality $\boldsymbol{\delta}_j'\boldsymbol{\delta}_j$, while $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j)$ follows the weighted sum of the non-central chi-square distribution with one degree of freedom and non-centrality $\left(\delta_{j,k}^{diag}\right)^2$. Using this result, the local power of the standard and diagonal J tests can be computed by applying the method by Imhof (1961) as described in Lemma 1.

Before we provide a detailed local power analysis in the context of dynamic panel data models in Section 5, we first consider how the moment conditions can be decomposed into groups in the next section.

4 Splitting the moment conditions

In the previous section, we have introduced a diagonal J test based on the grouped moment conditions. In this section, we discuss how to make groups. Specifically, we propose two approaches[13]. The first is to use a clustering algorithm called the K-means method first developed in cluster analysis literature. This approach is particularly useful when information that could

[12] In the Monte Carlo studies, we used the Matlab code written by Paul Ruud. The integration in (11) is carried out for $0 \leq v \leq 15$.

[13] Another approach is to use a pure diagonal (not *block* diagonal) weighting matrix $\widehat{\boldsymbol{\Omega}}_{diag}(\boldsymbol{\theta}) = \text{diag}\left[\frac{1}{N}\sum_{i=1}^{N}\widetilde{g}_{i1}(\boldsymbol{\theta})^2, \frac{1}{N}\sum_{i=1}^{N}\widetilde{g}_{i2}(\boldsymbol{\theta})^2, ..., \frac{1}{N}\sum_{i=1}^{N}\widetilde{g}_{im}(\boldsymbol{\theta})^2\right]$ where $q = m$. However, the unreported simulation results show that this approach suffers from substantial power loss. Hence, we do not consider this approach.

be used to make groups is not available. On the other hand, the second approach is to use a special structure of moment conditions where they are available sequentially. This situation typically appears in the analysis of panel data models. We now consider both approaches in detail.

4.1 K-means method

When outside information that could be used to make groups is not available, we need to make groups using a data-dependent method. The ideal approach is to construct a data-dependent criterion so that the power is maximized. However, even if such a method were to exist, it is unfortunately infeasible, because there are too many cases to be compared. For instance, even if the number of moment conditions is as small as 6 (i.e., $m = 6$), the total number of cases (the number of groups and selection of members in each group) then becomes 242. When $m = 10$, the total number of cases becomes 32,283. Thus, even if a data-dependent criterion associated with power were to exist, it is almost infeasible to compare all cases even for a moderate number of moment conditions[14]. Instead, we propose to use a clustering algorithm called the K-means method[15]. The K-means method was originally proposed in cluster analysis literature and is now widely used in many applications including pattern recognition, etc. Recently, the K-means method has also been used in econometrics for panel data analysis; as examples, see Lin and Ng (2012), Bonhomme and Manresa (2014), and Sarafidis and Weber (2015). While these papers mainly consider a regression analysis of heterogeneous panel data models, we apply the K-means method directly to the moment conditions $g_{ij}(\widehat{\boldsymbol{\theta}}), (i = 1, ..., N; j = 1, ..., m)$ where m moment conditions are grouped into q groups. Although the K-means method does not take the power into consideration, as shown in the Monte Carlo study in Section 6, it turns out to be quite useful since two tests where moment conditions are grouped by the K-means method and grouped by utilizing the sequentiality (which will be discussed below) have almost the same power.

We first give a brief introduction of the K-means method and then explain how it can be used for our case. Suppose that we have observations $x_{ij}, (i = 1, \cdots, n;\ j = 1, \cdots, p)$ where n is the sample size and p is the number of variables. In standard cluster analysis, the goal is to assign n observations to K clusters so that similar observations are assigned into the same cluster while dissimilar observations are assigned to different clusters. To be more specific, let $C_1, ..., C_K$ be the sets containing the indices of observations in each cluster. These sets satisfy $C_1 \cup C_2 \cup \cdots \cup C_K = \{1, 2, ..., n\}$ and $C_k \cap C_{k'} = \emptyset$ for $k \neq k'$. This means that each observation belongs to at least one of the K clusters and no observation belongs to more than one cluster.

Let $d_j(x_{ij}, x_{i'j})$ be the degree of *dissimilarity* between the i- and i'-th observations of the j-th variable, and define the total dissimilarity between i and i' over all variables $j = 1, ..., p$:

$$D(\mathbf{x}_i, \mathbf{x}_{i'}) = \sum_{j=1}^{p} d_j(x_{ij}, x_{i'j}) \tag{14}$$

where $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})'$ is a $p \times 1$ vector of the i-th observation. In many cases, the squared Euclidean distance $d_j(x_{ij}, x_{i'j}) = (x_{ij} - x_{i'j})^2$ is used:

$$D(\mathbf{x}_i, \mathbf{x}_{i'}) = \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 = \|\mathbf{x}_i - \mathbf{x}_{i'}\|^2. \tag{15}$$

[14]This situation is similar to the model selection problem where the number of regressors is large. If we try to find the best model with K regressors based on an information criterion, say, AIC, we need to consider $2^K - 1$ cases, which is almost infeasible even for a moderately large K.

[15]For a brief introduction of the K-means method, see Hastie, Tibshirani and Friedman (2009)

Alternatively, we could use the sample correlation coefficient:

$$D(\mathbf{x}_i, \mathbf{x}_{i'}) = 1 - \rho_{ii'}, \tag{16}$$

where $\rho_{ii'} = \sum_{j=1}^{p}(x_{ij} - \bar{x}_i)(x_{i'j} - \bar{x}_{i'})/\sqrt{\sum_{j=1}^{p}(x_{ij} - \bar{x}_i)^2 \sum_{j=1}^{p}(x_{i'j} - \bar{x}_{i'})^2}$, $\bar{x}_i = p^{-1}\sum_{j=1}^{p} x_{ij}$, and $\bar{x}_{i'} = p^{-1}\sum_{j=1}^{p} x_{i'j}$. When using the correlation coefficient (16), it should be noted that the degree of correlation for different *observations* i and i' is computed. This is in sharp contrast to the conventional use of the correlation coefficient where different *variables* are compared.

Using the dissimilarity measure $D(\mathbf{x}_i, \mathbf{x}_{i'})$, the *within*-cluster dissimilarity for a cluster k is defined as $\sum_{i \in C_k} \sum_{i' \in C_k} D(\mathbf{x}_i, \mathbf{x}_{i'})$[16]. Since the goal of cluster analysis is to assign the observations into clusters so that all observations within a cluster are similar while observations in different groups are not similar, the optimization problem we need to solve becomes

$$(C_1^*, C_2^*, \cdots, C_K^*) = \underset{C_1, C_2, \cdots, C_K}{\text{argmin}} \sum_{k=1}^{K} \sum_{i \in C_k} \sum_{i' \in C_k} D(\mathbf{x}_i, \mathbf{x}_{i'}). \tag{17}$$

This problem is quite hard to solve unless n and K are very small. Fortunately, a simple algorithm called the K-means method can provide a solution. The algorithm of the K-means method is given as follows.

1. Assign a number from 1 to K randomly to each observation.

2. Iterate the following steps (a) and (b) until the assignments do not change:

   (a) For each of K clusters, compute the center point (called the *centroid*)[17].

   (b) Assign each observation to the cluster whose centroid is the closest[18].

Since the final cluster assignment depends on the initial cluster assignment, we need to try several initial assignments in practice, which avoids the local minima.

This is a brief overview of the K-means method. We now explain how it can be applied to our purpose: splitting the m moment conditions $g_{ij}(\widehat{\boldsymbol{\theta}}), (i = 1, ..., N; j = 1, ..., m)$ into q ($q = K$ in the above example) groups. It should be stressed that we are interested in grouping the moment conditions, not observations as in the standard situation explained above. This is the main difference between the common practice and this paper. In our case, m and N correspond to n and p, respectively, (not p and n) in the above example. Hence, the number of clusters K should satisfy $K \leq m$, and not $K \leq N$. We need to solve the following problem using the K-means method:

$$(C_1^*, C_2^*, \cdots, C_K^*) = \underset{C_1, C_2, \cdots, C_K}{\text{argmin}} \sum_{k=1}^{K} \sum_{j \in C_k} \sum_{j' \in C_k} D\left(\underline{\mathbf{g}}_j(\widehat{\boldsymbol{\theta}}), \underline{\mathbf{g}}_{j'}(\widehat{\boldsymbol{\theta}})\right), \tag{18}$$

where $\underline{\mathbf{g}}_j(\widehat{\boldsymbol{\theta}}) = \left(g_{1j}(\widehat{\boldsymbol{\theta}}), g_{2j}(\widehat{\boldsymbol{\theta}}), \cdots, g_{Nj}(\widehat{\boldsymbol{\theta}})\right)'$ is an $N \times 1$ vector of jth moment condition.

Although the K-means method is a powerful tool to construct the groups, there are also some practical problems. First, when using the K-means method, we need to pre-specify the

[16]When $\mathbf{x}_i$ and $\mathbf{x}_{i'}$ are very similar, then $D(\mathbf{x}_i, \mathbf{x}_{i'})$ becomes small.

[17]Typically, the centroid of the k-th cluster is a p-dimensional vector of averages computed over observations in that cluster.

[18]The closeness can be measured by the squared Euclidean distance (15) or correlation coefficient (16).

number of clusters q. The choice of q is important since it is associated with the behavior of the diagonal J test in finite samples. The purpose of using the block diagonal matrix in the diagonal J test was to reduce the bias of the inverse of the sample covariance matrix. Hence, the dimension of each block, m_j, should be small compared with the sample size N for *all* j. Thus, when m is large, choosing a small value for q is not preferable since that could result in a moderately large ratio m_j/N and the empirical sizes would be distorted due to the bias of the covariance matrix (see (5)). On the other hand, choosing a large value of q too is not preferable since that choice could result in substantial power loss even if the sizes are correct[19]. Hence, in practice, we need to choose the value of q so that $m_j/N \leq \epsilon$ for all j where ϵ is a pre-specified positive small value such that the bias of the covariance matrix becomes small[20]. Consequently, we could choose a positive integer q such that $q \geq m/(\epsilon N)$[21].

Another possible pitfall of the K-means method is that it is difficult to control the size of each block m_j. For instance, when $m = 50$ and $q = 5$, the K-means method could provide a result that $m_1 = m_2 = m_3 = m_4 = m_5 = 10$. In this case, the diagonal J test is expected to work in a finite sample when N is as large as, say 200, since m_j/N is small enough for *all* j ($m_j/N = 0.05$). However, there is also a possibility that the K-means method results in clustering with $m_1 = m_2 = m_3 = m_4 = 1$ and $m_5 = 46$. In this case, the diagonal J test might not work in a finite sample since m_5/N is not small enough when N is, say, 200 ($m_5/N = 0.23$). Hence, when using the K-means method, we need to care about the maximum size of each block $m^* = \max_{1\leq j\leq q} m_j$. If we encounter such a case with a large m^*, we could apply the K-means method again to that large group and obtain further clustered moment conditions. Applying this procedure, we can reduce the maximum dimension m^* to some extent.

4.2 Sequential moment conditions

The K-means method is useful when a natural way to split the moment conditions is not available. However, in some cases, it is possible to decompose the moment conditions in a natural way by utilizing the special structure of moment conditions. Such a case appears when moment conditions are available sequentially, which typically arises in panel data models(Chamberlain, 1992). To be more specific, let us consider linear dynamic panel data models as a leading example since this type of model is extensively used in the literature[22][23]. Consider the following dynamic panel data model with fixed effects:

$$y_{it} = \alpha y_{i,t-1} + \boldsymbol{\beta}'\mathbf{x}_{it} + \eta_i + v_{it} = \mathbf{w}_{it}'\boldsymbol{\theta} + \eta_i + v_{it}, \qquad (i = 1, ..., N; t = 1, ..., T), \tag{19}$$

where $\boldsymbol{\beta}$ and $\mathbf{x}_{it}$ are $(p-1)\times 1$ vectors, and $\mathbf{w}_{it} = (y_{i,t-1}, \mathbf{x}_{it}')'$ and $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}')'$ are $p\times 1$ vectors. Assume that $|\alpha| < 1$, v_{it} is serially and cross-sectionally uncorrelated but can be heteroskedastic,

[19]Unreported simulation results show that the test with $q = m$, i.e., a pure diagonal weighting matrix, has substantially lower power as compared to other schemes in spite of the empirical sizes being correct.

[20]Unfortunately, deciding the optimal value for ϵ is very difficult and we need to decide empirically. In view of the simulation results in Section 6, when $\epsilon \leq 0.1$, the size distortions are reasonably small. A detailed analysis on the choice of ϵ is beyond the scope of the paper.

[21]This can be obtained as follows. Summing $m_j/N \leq \epsilon$ over $j = 1, ..., q$, we have $m/N \leq q\epsilon$ where $\sum_{j=1}^{q} m_j = m$ is used. Then, solving for q, we obtain $q \geq m/(\epsilon N)$.

[22]Other examples in which the diagonal J test might be useful are the panel count data models (Montalvo (1997), Blundell, Griffith and Windmeijer (2002), and Windmeijer (2008)) and the life-cycle models (Shapiro (1984), Zeldes (1989), Keane and Runkle (1992), and Wooldridge (2001, pp.434-435)). See also Wooldridge (1997) for other non-linear panel data models with sequential moment conditions.

[23]Bowsher (2002) and Roodman (2009) demonstrate the poor performance of the J test in the context of dynamic panel data models.

and $\mathbf{x}_{it}$ is assumed to be endogenous, that is, $E(\mathbf{x}_{is}v_{it}) = 0$ for $s < t$ and $E(\mathbf{x}_{is}v_{it}) \neq 0$ for $s \geq t$[24]. We now review the GMM estimators proposed in the literature.

4.2.1 Models in first-differences (DIF)

To remove the fixed effects, we transform the model through first-differences:

$$\Delta y_{it} = \Delta \mathbf{w}_{it}'\boldsymbol{\theta} + \Delta v_{it}, \qquad (i = 1, ..., N; t = 2, ..., T),$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta \mathbf{w}_{it} = \mathbf{w}_{it} - \mathbf{w}_{i,t-1}$, and $\Delta v_{it} = v_{it} - v_{i,t-1}$. Since y_{is} and $\mathbf{x}_{is}, (0 \leq s \leq t-2)$ are uncorrelated with Δv_{it}, we have the following moment conditions(Arellano and Bond, 1991):

$$E(y_{is}\Delta v_{it}) = \mathbf{0}, \qquad (s = 0, ..., t-2, t = 2, ..., T),$$
$$E(\mathbf{x}_{is}\Delta v_{it}) = \mathbf{0}, \qquad (s = 0, ..., t-2, t = 2, ..., T),$$

or in the matrix form,

$$E[\mathbf{g}_i^D(\boldsymbol{\theta})] \;=\; E(\mathbf{Z}_i^{D\prime}\Delta \mathbf{v}_i) = \left[E(\mathbf{z}_{i2}^D\Delta v_{i2})', E(\mathbf{z}_{i3}^D\Delta v_{i3})', ..., E(\mathbf{z}_{iT}^D\Delta v_{iT})'\right]' = \mathbf{0}, \tag{20}$$

where $\mathbf{z}_{it}^D = (y_{i0}, ..., y_{i,t-2}, \mathbf{x}_{i0}', ..., \mathbf{x}_{i,t-2}')'$, $\mathbf{Z}_i^D = \text{diag}(\mathbf{z}_{i2}^{D\prime}, ..., \mathbf{z}_{iT}^{D\prime})$, and $\Delta \mathbf{v}_i = (\Delta v_{i2}, ..., \Delta v_{iT})'$.

Note that these moment conditions can be seen as a decomposition into $q = T - 1$ groups:

$$\begin{aligned} E[\mathbf{g}_i^D(\boldsymbol{\theta})] &= \left[E(\mathbf{z}_{i2}^D\Delta v_{i2})', E(\mathbf{z}_{i3}^D\Delta v_{i3})', ..., E(\mathbf{z}_{iT}^D\Delta v_{iT})'\right]' \\ &= \left[E[\mathbf{g}_{i1}^D(\boldsymbol{\theta})]', E[\mathbf{g}_{i2}^D(\boldsymbol{\theta})]', ..., E[\mathbf{g}_{i,T-1}^D(\boldsymbol{\theta})]'\right]', \end{aligned} \tag{21}$$

where $E[\mathbf{g}_{i,t-1}^D(\boldsymbol{\theta})] = E(\mathbf{z}_{it}^D\Delta v_{it})$ for $t = 2, ..., T$ with $q = T - 1$. Note that each group is composed of moment conditions available at each period t. Thus, by focusing on the panel data model where moment conditions are available sequentially, we can decompose the moment conditions in a natural way, that is, regard the moment conditions available at period t as one group.

Note that since the number of moment conditions is $m = pT(T-1)/2$, it can be very large even for a moderate number of T and p. For example, when $T = 10$ and $p = 5$, we have $m = 225$ moment conditions. In practice, to mitigate the finite sample bias of the GMM estimator, empirical researchers often employ the strategy wherein only a subset of instruments, say of three lags, is used in each period[25]. In this case, $\mathbf{z}_{i2}^D = (y_{i0}, \mathbf{x}_{i0}')'$, $\mathbf{z}_{i3}^D = (y_{i0}, y_{i1}, \mathbf{x}_{i0}', \mathbf{x}_{i1}')'$, $\mathbf{z}_{is}^D = (y_{i,s-4}, y_{i,s-3}, y_{i,s-2}, \mathbf{x}_{i,s-4}', \mathbf{x}_{i,s-3}', \mathbf{x}_{i,s-2}')'$ for $s = 4, ..., T$, and the number of moment conditions is $m = 3p(T-2)$. Even in this case, we still have 120 moment conditions when $T = 10$ and $p = 5$. Hence, even when the number of moment conditions is reduced, the standard J test may not work well, because this number can remain large compared to the sample size N. However, for the diagonal J test, because the maximum dimension m^* is $p(T-1)$ when all available instruments are used and $m^* = 3p$ when only three lagged instruments are used, the dimension is much smaller. Since the choice of instruments affects the performance of the J test, we will discuss several schemes for the choice of instruments later.

Given the above decomposition of the moment conditions, we describe the testing procedure. First, the conventional one- and two-step GMM estimators are computed as

$$\widehat{\boldsymbol{\theta}}_{1step} \;=\; \left(\widehat{\mathbf{Q}}_{DIF}'\widehat{\mathbf{W}}_{DIF}^{-1}\widehat{\mathbf{Q}}_{DIF}\right)^{-1}\left(\widehat{\mathbf{Q}}_{DIF}'\widehat{\mathbf{W}}_{DIF}^{-1}\widehat{\mathbf{q}}_{DIF}\right), \tag{22}$$

[24] It is straightforward to allow for weakly or strictly exogenous variables.

[25] Okui (2009) proposes a procedure to select the number of moment conditions that minimize the mean-squared error of the GMM estimator in dynamic panel data models.

$$\widehat{\boldsymbol{\theta}}_{2step} = \left(\widehat{\mathbf{Q}}'_{DIF}\widehat{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\theta}})^{-1}\widehat{\mathbf{Q}}_{DIF}\right)^{-1}\left(\widehat{\mathbf{Q}}'_{DIF}\widehat{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\theta}})^{-1}\widehat{\mathbf{q}}_{DIF}\right), \tag{23}$$

where $\widehat{\mathbf{Q}}_{DIF} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i^{D\prime}\Delta\mathbf{W}_i$, $\widehat{\mathbf{q}}_{DIF} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i^{D\prime}\Delta\mathbf{y}_i$, $\Delta\mathbf{W}_i = (\Delta\mathbf{w}_{i2}, ..., \Delta\mathbf{w}_{iT})'$, $\Delta\mathbf{y}_i = (\Delta y_{i2}, ..., \Delta y_{iT})'$, and $\widehat{\mathbf{W}}_{DIF} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i^{D\prime}\mathbf{H}\mathbf{Z}_i^{D}$ in which $\mathbf{H}$ is a matrix with 2s on the main diagonal, -1s on the first-subdiagonal, and zeros otherwise. Note that when v_{it} is serially uncorrelated and homoskedastic with $var(v_{it}) = \sigma_v^2$, $\widehat{\mathbf{W}}_{DIF}$ is a consistent estimator of $E(\mathbf{Z}_i^{D\prime}\Delta\mathbf{v}_i\Delta\mathbf{v}_i'\mathbf{Z}_i^{D}) = \sigma_v^2 E(\mathbf{Z}_i^{D\prime}\mathbf{H}\mathbf{Z}_i^{D})$ up to a scalar. Hence, in such a case, the one-step estimator is asymptotically efficient. $\widehat{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\theta}})$ is computed as in (2) and $\widetilde{\boldsymbol{\theta}}$ is a preliminary consistent estimator of $\boldsymbol{\theta}$. The CU-GMM is obtained from (1) by numerical optimization. The test statistics $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j)$ can be obtained by plugging $\widehat{\boldsymbol{\theta}}_{1step}$ and $\widehat{\boldsymbol{\theta}}_{2step}$ into (8), and by putting (21) into (6). The critical values can be obtained as explained immediately after Lemma 1.

4.2.2 Models in forward orthogonal deviations (FOD)

Earlier, we used the first-differences to remove the fixed effects. Alternatively, we may transform the model by forward orthogonal deviations:

$$y_{it}^* = \mathbf{w}_{it}^{*\prime}\boldsymbol{\theta} + v_{it}^*, \qquad (i = 1, ..., N; t = 1, ..., T-1),$$

where $y_{it}^* = c_t[y_{it} - (y_{i,t+1} + \cdots + y_{iT})/(T-t)]$, $\mathbf{w}_{it}^* = c_t[\mathbf{w}_{it} - (\mathbf{w}_{i,t+1} + \cdots + \mathbf{w}_{iT})/(T-t)]$, and $v_{it}^* = c_t[v_{it} - (v_{i,t+1} + \cdots + v_{iT})/(T-t)]$ with $c_t^2 = (T-t)/(T-t+1)$. Since y_{is} and $\mathbf{x}_{is}, (0 \le s \le t-1)$ are uncorrelated with v_{it}^*, we have the following moment conditions:

$$E(y_{is}v_{it}^*) = \mathbf{0}, \qquad (s = 0, ..., t-1, t = 1, ..., T-1),$$
$$E(\mathbf{x}_{is}v_{it}^*) = \mathbf{0}, \qquad (s = 0, ..., t-1, t = 1, ..., T-1),$$

which can be written in the matrix form as

$$\begin{aligned} E[\mathbf{g}_i^F(\boldsymbol{\theta})] &= E(\mathbf{Z}_i^{F\prime}\mathbf{v}_i^*) = \left[E(\mathbf{z}_{i1}^F v_{i1}^*)', E(\mathbf{z}_{i2}^F v_{i2}^*)', ..., E(\mathbf{z}_{i,T-1}^F v_{i,T-1}^*)'\right]' \\ &= \left[E[\mathbf{g}_{i1}^F(\boldsymbol{\theta})]', E[\mathbf{g}_{i2}^F(\boldsymbol{\theta})]', ..., E[\mathbf{g}_{i,T-1}^F(\boldsymbol{\theta})]'\right]' = \mathbf{0}, \end{aligned} \tag{24}$$

where $\mathbf{z}_{it}^F = (y_{i0}, ..., y_{i,t-1}, \mathbf{x}_{i0}', ..., \mathbf{x}_{i,t-1}')'$, $\mathbf{Z}_i^F = \text{diag}(\mathbf{z}_{i1}^{F\prime}, ..., \mathbf{z}_{i,T-1}^{F\prime})$, and $\mathbf{v}_i^* = (v_{i1}^*, ..., v_{i,T-1}^*)'$. As in the models in first-differences, (24) represents the decomposition of moment conditions, that is, $E[\mathbf{g}_{it}^F(\boldsymbol{\theta})] = E(\mathbf{z}_{it}^F v_{it}^*), (t = 1, ..., T-1)$ constitute the $q = T-1$ groups in $E[\mathbf{g}_i^F(\boldsymbol{\theta})] = \mathbf{0}$. The diagonal J test statistic is obtained by substituting (24) into (6).

The one- and two-step GMM estimators are given by

$$\widehat{\boldsymbol{\theta}}_{1step} = \left(\widehat{\mathbf{Q}}'_{FOD}\widehat{\mathbf{W}}_{FOD}^{-1}\widehat{\mathbf{Q}}_{FOD}\right)^{-1}\left(\widehat{\mathbf{Q}}'_{FOD}\widehat{\mathbf{W}}_{FOD}^{-1}\widehat{\mathbf{q}}_{FOD}\right), \tag{25}$$

$$\widehat{\boldsymbol{\theta}}_{2step} = \left(\widehat{\mathbf{Q}}'_{FOD}\widehat{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\theta}})^{-1}\widehat{\mathbf{Q}}_{FOD}\right)^{-1}\left(\widehat{\mathbf{Q}}'_{FOD}\widehat{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\theta}})^{-1}\widehat{\mathbf{q}}_{FOD}\right), \tag{26}$$

where $\widehat{\mathbf{Q}}_{FOD} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i^{F\prime}\mathbf{W}_i^*$, $\widehat{\mathbf{q}}_{FOD} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i^{F\prime}\mathbf{y}_i^*$, $\mathbf{y}_i^* = (y_{i1}^*, ..., y_{i,T-1}^*)'$, $\mathbf{W}_i^* = (\mathbf{w}_{i1}^*, ..., \mathbf{w}_{i,T-1}^*)'$, and $\widehat{\mathbf{W}}_{FOD} = N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i^{F\prime}\mathbf{Z}_i^F$. $\widehat{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\theta}})$ is computed as in (2), and $\widetilde{\boldsymbol{\theta}}$ is a preliminary consistent estimator of $\boldsymbol{\theta}$. The CU-GMM is obtained from (1) by numerical optimization. Note that the one-step GMM estimator is asymptotically efficient when v_{it} is serially uncorrelated and homoskedastic with $var(v_{it}) = \sigma^2$ because $\widehat{\mathbf{W}}_{FOD}$ is a consistent estimator of the optimal weighting matrix $\boldsymbol{\Omega} = E(\mathbf{Z}_i^{F\prime}\mathbf{v}_i^*\mathbf{v}_i^{*\prime}\mathbf{Z}_i^F) = \sigma^2 E(\mathbf{Z}_i^{F\prime}\mathbf{Z}_i^F)$ up to a scalar[26]. However, if

[26]The difference in the form of the optimal weighting matrix for models in first-differences and in forward orthogonal deviations comes from the fact that the first-differenced errors are serially correlated by construction whereas the errors of models in forward orthogonal deviations are serially uncorrelated and homoskedastic when original errors are serially uncorrelated and homoskedastic.

there is cross-sectional and/or time-series heteroskedasticity, then $\sigma^2 E(\mathbf{Z}_i^{F\prime}\mathbf{Z}_i^F)$ is no longer the optimal weighting matrix and the two-step GMM estimator is more efficient.

4.2.3 System models

In the literature, it is known that the first-difference GMM estimators may suffer from the weak instruments problem when α in (19) is close to one and/or $var(\eta_i)/var(v_{it})$ is large (e.g., Blundell and Bond, 1998; Blundell, Bond and Windmeijer, 2000)[27]. To overcome this problem, it is now a common strategy to use the so-called system GMM estimator by Arellano and Bover (1995) and Blundell and Bond (1998)[28]. The system GMM estimator exploits the moment conditions for models in first-differences (or orthogonal deviations) and in levels, while the latter is given as

$$E[\mathbf{g}_{it}^L(\boldsymbol{\theta})] = E[\mathbf{z}_{it}^L u_{it}] = \mathbf{0}, \qquad (t = 2, ..., T), \tag{27}$$

with $u_{it} = \eta_i + v_{it}$, or in the matrix form

$$E[\mathbf{Z}_i^{L\prime}\mathbf{u}_i] = \mathbf{0}, \tag{28}$$

where $\mathbf{Z}_i^L = \text{diag}(\mathbf{z}_{i2}^{L\prime}, ..., \mathbf{z}_{iT}^{L\prime})$ with $\mathbf{z}_{it}^L = (\Delta y_{i,t-1}, \Delta\mathbf{x}_{i,t-1}')'$ and $\mathbf{u}_i = (u_{i2}, ..., u_{iT})'$. Note that for moment conditions (27) to be valid, we need to assume that $E(y_{it}\eta_i)$ and $E(\mathbf{x}_{it}\eta_i)$ are constant over time, while we do not need to impose this assumption for moment conditions associated with models in first-differences and in forward orthogonal deviations. The system GMM estimator is obtained from the combined moment conditions:

$$\begin{aligned} E[\mathbf{g}_i^S(\boldsymbol{\theta})] &= E[\mathbf{Z}_i^{S\prime}\mathbf{r}_i] = \big[E(\mathbf{z}_{i2}^D\Delta v_{i2})', E(\mathbf{z}_{i3}^D\Delta v_{i3})', ..., E(\mathbf{z}_{iT}^D\Delta v_{iT})', \\ &\qquad\qquad E(\mathbf{z}_{i2}^L u_{i2})', E(\mathbf{z}_{i3}^L u_{i3})', ..., E(\mathbf{z}_{iT}^L u_{iT})'\big]' \\ &= \big[E[\mathbf{g}_{i1}^D(\boldsymbol{\theta})]', E[\mathbf{g}_{i2}^D(\boldsymbol{\theta})]', ..., E[\mathbf{g}_{i,T-1}^D(\boldsymbol{\theta})]', \\ &\qquad E[\mathbf{g}_{i1}^L(\boldsymbol{\theta})]', E[\mathbf{g}_{i2}^L(\boldsymbol{\theta})]', ..., E[\mathbf{g}_{i,T-1}^L(\boldsymbol{\theta})]'\big]' = \mathbf{0}, \end{aligned} \tag{29}$$

where $\mathbf{r}_i = (\Delta\mathbf{v}_i', \mathbf{u}_i')'$, $\mathbf{Z}_i^S = \text{diag}(\mathbf{Z}_i^D, \mathbf{Z}_i^L)$, $\mathbf{g}_{it}^D(\boldsymbol{\theta}) = \mathbf{z}_{i,t+1}^D\Delta v_{i,t+1}$, and $\mathbf{g}_{i,t}^L = \mathbf{z}_{i,t+1}^L u_{i,t+1}$ for $t = 1, ..., T-1$ with $q = 2(T-1)$. If models in forward orthogonal deviations are employed instead of first-differences, we only replace $\mathbf{g}_{it}^D(\boldsymbol{\theta}) = \mathbf{z}_{i,t+1}^D\Delta v_{i,t+1}$ with $\mathbf{g}_{it}^F(\boldsymbol{\theta}) = \mathbf{z}_{it}^F v_{it}^*$. Note that $\mathbf{g}_{it}^D(\boldsymbol{\theta}), (t = 1, ..., T-1)$ (or $\mathbf{g}_{it}^F(\boldsymbol{\theta}), (t = 1, ..., T-1)$) and $\mathbf{g}_{i,t}^L, (t = 1, ..., T-1)$ constitute $q = 2(T-1)$ groups. The diagonal J test statistics based on the system GMM estimators are obtained by substituting (29) into (6).

4.2.4 Relationship between the standard and diagonal J tests

We now investigate the relationship between the standard and diagonal J tests. We demonstrate that the form of the model, that is, in first-differences or forward orthogonal deviations, and the heteroskedasticity of v_{it} affect the relationship between the standard and diagonal J tests.

First, we consider the models in first-differences and in forward orthogonal deviations. Although the choice of first-differences or forward orthogonal deviations for the elimination of

[27] Although this weak instruments problem is widely known, Hayakawa (2009) demonstrates that this problem does not always happen even if α is close to 1. He shows that if initial conditions do not follow the steady state distribution, the instruments can be strong even if α is close to 1; therefore, bias can be small.

[28] Bun and Windmeijer (2010) show that the system GMM estimator does suffer from the weak instruments problem if $var(\eta_i)/var(v_{it})$ is large. See also Hayakawa (2007) for the finite sample bias of the system GMM estimator.

fixed effects is a minor issue in view of the equivalence result of Arellano and Bover (1995), it is not so for the diagonal J test. For the models in first-differences, since the transformed error Δv_{it} is serially correlated, we have $E[\mathbf{g}_{it}^D(\boldsymbol{\theta})\mathbf{g}_{i,t+1}^{D\prime}(\boldsymbol{\theta})] = E(\Delta v_{i,t+1}\Delta v_{i,t+2}\mathbf{z}_{it}^D\mathbf{z}_{i,t+1}^{D\prime}) \neq \mathbf{0}$. This implies that $\boldsymbol{\Omega}_{diag}$ never coincides with $\boldsymbol{\Omega}$, which in turn indicates that $\mathcal{J}_{diag}$ always follows the nonstandard distribution as in Theorem 1. However, this is not the case for the model in forward orthogonal deviations. To see this, we first consider the structure of transformed error v_{it}^*. If v_{it} is temporally and cross-sectionally heteroskedastic, that is, $var(v_{it}) = \sigma_{it}^2$, we have

$$E(v_{it}^* v_{is}^*) = \begin{cases} \frac{T-t}{T-t+1}\left(\sigma_{it}^2 + \frac{\sigma_{i,t+1}^2+\cdots+\sigma_{iT}^2}{(T-t)^2}\right) & t=s \\ c_t c_s \left(\frac{-(T-t)\sigma_{it}^2+\sigma_{i,t+1}^2+\cdots+\sigma_{iT}^2}{(T-t)(T-s)}\right) & t>s \end{cases}.$$

Hence, if both time-series heteroskedasticity and cross-sectional heteroskedasticity are present, v_{it}^* is serially correlated and heteroskedastic. In this case, the optimal weighting matrix is $\boldsymbol{\Omega}$ with $\boldsymbol{\Omega} \neq \boldsymbol{\Omega}_{diag}$. This is also the case if only time-series heteroskedasticity is present. However, when v_{it} is homoskedastic over time but heteroskedastic over i, that is, $\sigma_{it}^2 = \sigma_i^2$ for $t = 1, ..., T$, we have

$$E(v_{it}^* v_{is}^*) = \begin{cases} \sigma_i^2 & t=s \\ 0 & t>s \end{cases}.$$

In this case, v_{it}^* is serially uncorrelated and the optimal weighting matrix is $\boldsymbol{\Omega}_{diag}$. Thus, when there is no time-series heteroskedasticity, we have $\boldsymbol{\Omega}_{diag} = \boldsymbol{\Omega}$, and hence, as noted in Remark 3, the diagonal J tests computed from the two-step and CU-GMM estimators, that is, $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j)$ $(j = 2step, CU)$, asymptotically follow the standard χ^2 distribution. Furthermore, under no time-series heteroskedasticity and cross-sectional heteroskedasticity, that is, if v_{it} is iid, we have $\boldsymbol{\Omega}_{diag} = \boldsymbol{\Omega} = \mathbf{W}$. Hence, as noted in Remark 3, the diagonal J tests computed from the one-step, two-step, and CU-GMM estimators, that is, $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j)$ $(j = 1step, 2step, CU)$, asymptotically follow the standard χ^2 distribution. Hence, in these special cases, the diagonal J tests follow the standard χ^2 distribution as in the standard J test. However, it should be noted that unlike the standard J test, the diagonal J tests do not require inversing a large-dimensional weighting matrix. For the case of the system GMM estimator, since the moment conditions for models in first-differences (or in forward orthogonal deviations) and in levels are correlated, we have $\boldsymbol{\Omega} \neq \boldsymbol{\Omega}_{diag}$, which implies that the diagonal J test always follows the non-standard distribution as in Theorem 1.

5 Local power analysis of the diagonal J test

In Section 3, we have derived the asymptotic distribution of the diagonal J test under local alternatives in a general framework. In this section, we conduct a detailed local power analysis of the standard and diagonal J tests for dynamic panel data models considered in the previous section. Specifically, (i) we compare the powers of the standard and diagonal J tests, and (ii) investigate the effects of estimator efficiency on power.

We consider the following dynamic panel data model:

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta x_{it} + \eta_i + v_{it}, \qquad (i = 1, ..., N; t = 1, ..., T), \\ x_{it} &= \rho x_{i,t-1} + \tau \eta_i + \theta v_{it} + e_{it}, \end{aligned}$$

where $\eta_i \sim iid(0, \sigma_\eta^2)$ and $e_{it} \sim iid(0, \sigma_e^2)$.

If v_{it} is serially uncorrelated, we have a series of moment conditions. Specifically, the moment conditions for models in first-differences are $E[\mathbf{g}_i^D(\boldsymbol{\theta})] = E(\mathbf{Z}_i^{D\prime}\Delta\mathbf{v}_i) = \mathbf{0}$. The moment conditions for models in forward orthogonal deviations are $E[\mathbf{g}_i^F(\boldsymbol{\theta})] = E(\mathbf{Z}_i^{F\prime}\mathbf{v}_i^*) = \mathbf{0}$. The moment conditions for the system combining models in first-differences and in levels are given by $E[\mathbf{g}_i^S(\boldsymbol{\theta})] = E[\mathbf{g}_i^{DL}(\boldsymbol{\theta})] = E(\mathbf{Z}_i^{DL\prime}\mathbf{r}_i^{DL}) = \mathbf{0}$, where $\mathbf{Z}_i^S = \mathbf{Z}_i^{DL} = \text{diag}(\mathbf{Z}_i^D, \mathbf{Z}_i^L)$ and $\mathbf{r}_i^S = \mathbf{r}_i^{DL} = (\Delta\mathbf{v}_i', \mathbf{u}_i')'$. The moment conditions for the system combining models in forward orthogonal deviations and in levels are given by $E[\mathbf{g}_i^S(\boldsymbol{\theta})] = E[\mathbf{g}_i^{FL}(\boldsymbol{\theta})] = E(\mathbf{Z}_i^{FL\prime}\mathbf{r}_i^{FL}) = \mathbf{0}$, where $\mathbf{Z}_i^{FL} = \text{diag}(\mathbf{Z}_i^F, \mathbf{Z}_i^L)$ and $\mathbf{u}_i^{FL} = (\mathbf{v}_i^{*\prime}, \mathbf{u}_i')'$.

To derive the local power, we need to construct the locally violated moment conditions. While there are several approaches for this, we consider the simple case where v_{it} is serially correlated. Specifically, we assume that v_{it} follows a first-order moving average (MA(1)) process given by

$$v_{it} = \varepsilon_{it} + \phi\varepsilon_{i,t-1},$$

where $\varepsilon_{it} \sim iid(0, \sigma_t^2)$[29]. In this case, not all of the moment conditions are invalid: only the moment conditions using the closest instruments are invalid. For instance, in $E[\mathbf{g}_i^D(\boldsymbol{\theta})] = \mathbf{0}$, only the moment conditions $E(y_{i,t-2}\Delta v_{it}) = 0$ and $E(x_{i,t-2}\Delta v_{it}) = 0, (t = 2, ..., T-1)$ are invalid and the remaining moment conditions are valid. Similarly, for the case of $E[\mathbf{g}_i^F(\boldsymbol{\theta})] = \mathbf{0}$, only the moment conditions $E(y_{i,t-1}v_{it}^*) = 0$ and $E(x_{i,t-1}v_{it}^*) = 0, (t = 1, ..., T-2)$ are invalid. For the moment conditions of the systems, in addition to the above cases, all the moment conditions associated with models in levels, $E(\Delta y_{i,t-1}u_{it}) = 0$ and $E(\Delta x_{i,t-1}u_{it}) = 0, (t = 2, ..., T)$, are invalid.

To specify the local alternative, consider a sequence $\phi = \phi_N = c_\phi/\sqrt{N}$, where c_ϕ is a finite constant. In this case, we have locally violated moment conditions $E[\mathbf{g}_i^l(\alpha)] = \mathbf{c}^l/\sqrt{N}$, $(l = D, F, DL, FL)$, where the specific forms of $\mathbf{c}^l$ are given by

$$\begin{aligned}
\mathbf{c}^D &= -c_\phi\left(\sigma_0^2, \theta\sigma_0^2, \sigma_1^2\mathbf{e}_2', \theta\sigma_1^2\mathbf{e}_2', ..., \sigma_{T-2}^2\mathbf{e}_{T-1}', \theta\sigma_{T-2}^2\mathbf{e}_{T-1}'\right)', \\
\mathbf{c}^F &= c_\phi\left(c_1\sigma_0^2, \theta c_1\sigma_0^2, c_2\sigma_1^2\mathbf{e}_2', \theta c_2\sigma_1^2\mathbf{e}_2', ..., c_{T-1}\sigma_{T-2}^2\mathbf{e}_{T-1}', \theta c_{T-1}\sigma_{T-2}^2\mathbf{e}_{T-1}'\right)', \\
\mathbf{c}^{DL} &= \left(\mathbf{c}^{D\prime}, c_\phi\sigma_1^2, \theta c_\phi\sigma_1^2, \cdots, c_\phi\sigma_{T-1}^2, \theta c_\phi\sigma_{T-1}^2\right)', \\
\mathbf{c}^{FL} &= \left(\mathbf{c}^{F\prime}, c_\phi\sigma_1^2, \theta c_\phi\sigma_1^2, \cdots, c_\phi\sigma_{T-1}^2, \theta c_\phi\sigma_{T-1}^2\right)',
\end{aligned}$$

where $\mathbf{e}_r = (0, ..., 0, 1)'$ is an $r \times 1$ vector where the last element is 1 and all other elements are zero, and $c_t = \sqrt{(T-t)/(T-t+1)}$.

For the sample sizes and parameters, we consider the following cases: $T = 5, 10$, $\alpha = 0.2, 0.8$, $\beta = 1$, $\tau = 0.25$, $\theta = -0.1$, $\sigma_e^2 = 0.16$, $\sigma_\eta^2 = 1, 5$, and $\sigma_t^2 = 0.5 + (t-1)/(T-1)$[30]. Initial conditions are assumed to follow stationary distribution. First, we compare the power properties of the standard and diagonal J tests that are based on models in first-differences and in forward orthogonal deviations as well as the systems where these models are combined with models in levels. Here, we consider the cases where the moment conditions are estimated by two-step estimators. The standard J test, $\mathcal{J}(\widehat{\boldsymbol{\theta}}_{2step})$, is denoted as "J(DIF) |J(FOD)" or "J(DIF&LEV)|J(FOD&LEV)"[31]. The diagonal J test, $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_{2step})$, is denoted as "$\text{J}_{diag}$(DIF)," "$\text{J}_{diag}$(FOD)," "$\text{J}_{diag}$(DIF&LEV)," and "$\text{J}_{diag}$(FOD&LEV)." The power plots for $\mathcal{J}(\widehat{\boldsymbol{\theta}}_{2step})$ and $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_{2step})$ are given in Figures 2 and 3. From the figures, we find that the diagonal J test for models in forward orthogonal deviations has almost the same power as the standard J test.

[29] In the Monte Carlo studies in Section 6, we consider temporally and cross-sectionally heteroskedastic errors.

[30] In the supplementary appendix, we also report the results of the following cases: $\sigma_t^2 = 1$ and $\sigma_t^2 = 0.5 + (t-1)$.

[31] Note that J(DIF) and J(FOD) yield the same power. This is also true for J(DIF&LEV) and J(FOD&LEV).

This implies that using a block diagonal matrix in the weighting matrix has almost no negative effect on the power property. However, the power of the diagonal J test based on models in first-differences tends to be lower. The degree of power loss depends mainly on α. When $\alpha = 0.2$, the power loss is not large, but when $\alpha = 0.8$, the power loss is not negligible, although it decreases as T increases. Similar results apply to system models. When models in forward orthogonal deviations are used in the system, the diagonal J test has almost the same power as the standard J test[32]. However, when models in first-differences are used in the system, the diagonal J test tends to have lower power than the standard J test with a few exceptions. For the effect of σ_η^2, by comparing Figures 2 and 3, we find that the magnitude of σ_η^2 has almost no effect on the relative performance between the standard J test and the diagonal J test based on models in forward orthogonal deviations. However, the diagonal J test based on the models in first-differences becomes less powerful when σ_η^2 is large.

Next, we investigate the effect of the efficiency of the estimator used to compute the test statistic. Note that the two-step GMM estimator is more efficient than the one-step GMM estimator since the errors have time-series heteroskedasticity. In Figures 4 and 5, we provide the power plots of $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_{1step})$ and $\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_{2step})$ for the same models as above[33]. From the power plots, we find that for models in first-differences and in forward orthogonal deviations, the difference between the powers of the one- and two-step estimators is very minor. However, for system models, tests based on two-step estimators have higher powers than those based on one-step estimators. This is because the efficiency loss of one-step estimators as compared to two-step estimators is not so large for the models in first-differences and in forward orthogonal deviations whereas this is not the case for system models. In system models, since the correlation between moment conditions for the models in first-differences or forward orthogonal deviations and those for models in levels is not weak (cf. Kiviet, 2007), using the two-step procedure results in a large efficiency gain, and becomes more powerful.

Summarizing the local power analysis, we may conclude that the diagonal J test based on models in forward orthogonal deviations and in systems utilizing them with two-step estimators has comparable power properties to the standard J test. However, as will be shown in the next section, the diagonal J test has substantially better size property than the standard J test in finite samples, which makes the diagonal J test more appealing.

6 Monte Carlo simulation

In this section, we conduct a Monte Carlo simulation to investigate the performance of the diagonal J test and compare it with the standard J test.

6.1 Design

We consider the following data generating process:

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta x_{it} + \eta_i + v_{it}, \quad (i = 1, ..., N; t = -49, -48, ..., -1, 0, 1, ..., T), \\ x_{it} &= \rho x_{i,t-1} + \tau \eta_i + \theta v_{it} + e_{it}, \\ v_{it} &= \varepsilon_{it} + \phi \varepsilon_{i,t-1}, \end{aligned}$$

[32]When $T = 10$, $\alpha = 0.8$, and $\sigma_\eta^2 = 1$, the diagonal J test shows slightly higher power than the standard J test. It is quite hard to identify the reason for this since the local power of the diagonal J test depends on eigenvalues and eigenvectors of matrix (9), which is quite hard to investigate analytically.

[33]Further results are provided in the supplementary appendix.

where $\eta_i \sim iid\mathcal{N}(0, \sigma_\eta^2)$, $e_{it} \sim iid\mathcal{N}(0, \sigma_e^2)$, and $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_{it}^2)$, where σ_{it}^2 is generated as $\sigma_{it}^2 = \delta_i \tau_t$, in which $\delta_i \sim \mathcal{U}[0.5, 1.5]$ and $\tau_t = 0.5 + (t-1)/(T-1)$ for $t = 1..., T$ and $\tau_t = 0.5$ for $t = -49, ..., -1, 0$. This model is identical to that used in the local power analysis except that the current error terms are temporally and cross-sectionally heteroskedastic. Note that the regressor x_{it} is an endogenous variable. The data are generated as

$$\begin{pmatrix} y_{it} \\ x_{it} \end{pmatrix} = \begin{pmatrix} \alpha & \beta\rho \\ 0 & \rho \end{pmatrix} \begin{pmatrix} y_{i,t-1} \\ x_{i,t-1} \end{pmatrix} + \begin{pmatrix} (1+\beta\tau)\eta_i \\ \tau\eta_i \end{pmatrix} + \begin{pmatrix} (1+\theta\beta)v_{it} + \beta e_{it} \\ \theta v_{it} + e_{it} \end{pmatrix}, \quad (i = 1, ..., N;\ t = -49, -48, ...0, 1, ., T)$$

with initial conditions

$$\begin{pmatrix} y_{i,-50} \\ x_{i,-50} \end{pmatrix} = \begin{pmatrix} 1-\alpha & -\beta\rho \\ 0 & 1-\rho \end{pmatrix}^{-1} \begin{pmatrix} (1+\beta\tau)\eta_i \\ \tau\eta_i \end{pmatrix} + \boldsymbol{\xi}_{i,-50}, \qquad (i = 1, ..., N)$$

$$\boldsymbol{\xi}_{i,-50} = \sum_{j=0}^{10} \begin{pmatrix} \alpha & \beta\rho \\ 0 & \rho \end{pmatrix}^j \begin{pmatrix} (1+\theta\beta)v_{ij} + \beta e_{ij} \\ \theta v_{ij} + e_{ij} \end{pmatrix}, \tag{30}$$

where $v_{ij} \sim \mathcal{N}(0, \bar{\sigma}_i^2)$ and $e_{ij} \sim iid\mathcal{N}(0, \sigma_e^2)$ $(j = 0, ..., 10)$, with $\bar{\sigma}_i^2 = (T+51)^{-1} \sum_{t=-50}^{T} \sigma_{it}^2$. The samples sizes we consider are $T = 4, 7, 10$, $N = 110, 250, 500$[34]. For the parameters, we set $\alpha = 0.2, 0.8$, $\beta = 1$, $\rho = 0.5$, $\tau = 0.25$, $\theta = -0.1$, $\sigma_\eta^2 = 1, 5$, and $\sigma_e^2 = 0.16$. For the value of ϕ, we consider $\phi = 0, 0.2, 0.5, 0.8$. $\phi = 0$ corresponds to size, and $\phi = 0.2, 0.5, 0.8$ corresponds to power. For the computation of the critical values of the diagonal J test, we consider two approaches: one is a simulation method based on 10,000 random draws, and the other is an analytical method as described in Section 3. In the tables below, we report the results where critical values are computed by the analytical method because the two results are almost identical. The significance level is 5% and the number of replications is 1,000.

6.2 Choice of models, estimation method, instruments, and groups

In finite samples, the choice of models, estimation method, instruments, and how moment conditions are grouped substantially affect the performance of the GMM estimators and related tests. Here, we give a brief overview of these choices.

6.2.1 Choice of models and estimation method

For the models, we consider the following four types: (i) models in first-differences (denoted as "DIF"), (ii) models in forward orthogonal deviations (denoted as "FOD"), (iii) a system composed of models in first-differences and in levels (denoted as "DIF& LEV"), (iv) a system composed of models in forward orthogonal deviations and in levels (denoted as "FOD& LEV").

For the estimation method, we consider the following three types: (i) one-step method (denoted as "1step"), (ii) two-step method (denoted as "2step"), and (iii) continuously updated method (denoted as "CUE").

[34]Since the number of moment conditions for the system GMM estimator using all instruments is 108 when $T = 10$, we set $N = 110$, instead of $N = 100$, to avoid a singular weighting matrix.

6.2.2 Choice of instruments

Next, we discuss the three schemes for the choice of instruments. Since we use only the closest instruments in the moment conditions for models in levels because of redundancy, in the following, we mainly consider moment conditions associated with models in first-differences or in forward orthogonal deviations.

All instruments If we use all available instruments, then the GMM estimators will be efficient. In this case, the instruments used in each period are $\mathbf{z}_{it}^D = (y_{i0}, ..., y_{i,t-2}, x_{i0}, ..., x_{i,t-2})', (t = 2, ..., T)$ and $\mathbf{z}_{it}^F = (y_{i0}, ..., y_{i,t-1}, x_{i0}, ..., x_{i,t-1})', (t = 1, ..., T-1)$. In this case, the number of moment conditions for models in first-differences or in forward orthogonal deviations is $m = T(T-1) = O(T^2)$, and that for system models is $m = (T+2)(T-1) = O(T^2)$. In the diagonal J test, we need to invert a matrix of size $m^* = 2(T-1)$ at the maximum. Hence, if $m^* = 2(T-1)$ is large as compared to N, the diagonal J test may not deliver accurate sizes in the finite sample.

Sargan's rule of thumb Although using many instruments leads to efficient estimation, it also causes a bias. Therefore, many empirical studies use only a subset of moment conditions by using a few lagged variables at each period. While there is no theoretical background, we investigate a rule of thumb by Sargan (1958) to determine the lag length of instruments. Sargan (1958) proposed a simple rule of thumb that the number of instruments should be smaller than $N/20$ in a context of cross-section or time-series model. We apply this rule to the choice of the lag length at each period. Specifically, in this case, the instruments used in each period are $\mathbf{z}_{it}^D = (y_{i,\max\{0,t-1-\ell\}}, ..., y_{i,t-2}, x_{i,\max\{0,t-1-\ell\}}, ..., x_{i,t-2})', (t = 2, ..., T)$ and $\mathbf{z}_{it}^F = (y_{i,\max\{0,t-\ell\}}, ..., y_{i,t-1}, x_{i,\max\{0,t-\ell\}}, ..., x_{i,t-1})', (t = 1, ..., T-1)$ with $\ell = int(N/(20 \times 2)) \geq 1$, where $int(A)$ denotes the integer part of A. In this case, 2ℓ instruments are used at the maximum in each period and the number of moment conditions for models in first-differences and in forward orthogonal deviations is $m = \ell(\ell-1) + 2(T-\ell)\ell = O(T)$, and that for system models is $m = \ell(\ell-1) + 2(T-\ell)\ell + 2(T-1) = O(T)$. In the current simulation design, $\ell = 2, 6, 12$ for $N = 110, 250, 500$, respectively. In the diagonal J test, we need to invert an $m^* = 2\ell$-dimensional matrix at the maximum, which is substantially smaller than m.

Collapsed instruments In an attempt to reduce the number of moment conditions further, several studies such as Bun and Kiviet (2006) propose the so-called "collapsed instruments." A distinctive feature of collapsed instruments is that the total number of moment conditions is $O(1)$, and hence, it does not increase even if T is large. We consider the following collapsed instruments for models in first-differences and in forward orthogonal deviations:

$$\mathbf{Z}_i^D = \mathbf{Z}_i^F = \begin{bmatrix} y_{i0} & x_{i0} & 0 & 0 \\ y_{i1} & x_{i1} & y_{i0} & x_{i0} \\ \vdots & \vdots & \vdots & \vdots \\ y_{i,T-2} & x_{i,T-2} & y_{i,T-3} & x_{i,T-3} \end{bmatrix}. \tag{31}$$

For the system model, we use the following collapsed instruments:

$$\mathbf{Z}_i^S = \text{diag}\left[\begin{pmatrix} y_{i0} & x_{i0} \\ \vdots & \vdots \\ y_{i,T-2} & x_{i,T-2} \end{pmatrix}, \begin{pmatrix} \Delta y_{i1} & \Delta x_{i1} \\ \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta x_{i,T-1} \end{pmatrix} \right]. \tag{32}$$

Note that the number of moment conditions is 4 for all cases and does not depend on T. We do not consider the diagonal J test for this case since the standard J test is expected to work since the number of moment conditions is small enough as compared to the sample size N.

In the following, we call the first two instruments *diagonal instruments*, as opposed to *collapsed instruments*, as they can be expressed as a block-diagonal matrix.

6.2.3 How to make groups

In the diagonal J test, how to make groups from the original moment conditions is an important issue. We consider two approaches. The first is to utilize the special structure of sequential moment conditions. The second is to use the K-means method outlined in Section 4, by assuming that the grouped structure is unknown. Hence, in this case, groups are determined by using the information from data only. For the choice of dissimilarity measure, we use the correlation coefficient (16) since it performs much better than the squared Euclidean distance (15). The corresponding diagonal J tests are denoted as J_{diag}^{seq} and J_{diag}^{K}, respectively. Further, to show how the K-means method constructs groups close to those used in J_{diag}^{seq}, we report the adjusted Rand index by Hubert and Arabie (1985)[35]. The adjusted Rand index measures the closeness between two outcomes of clustering. If the clustering assignment of J_{diag}^{K} is exactly the same as those used in J_{diag}^{seq}, then the adjusted Rand index becomes one. Hence, when the adjusted Rand index is close to one, it means that very similar grouped moment conditions are used in J_{diag}^{seq} and J_{diag}^{K}.

6.3 Results

We investigate all the combinations of the above choices of models and instruments except for the one-step method with collapsed instruments. This leads to 32 cases in total. However, to save space, we report the results associated with models in forward orthogonal deviations because the local power analysis exhibits better performance in this case. We also omit the results with $\phi = 0.8$ and $N = 500$. All the simulation results are provided in the supplementary appendix, which is available upon request.

Simulation results for the case of $\alpha = 0.2$ are given in Tables 2, 3, and 4. First, we see the results for the case of $\alpha = 0.2$. When $T = 4$, the sizes of the standard J test are slightly distorted. However, the distortion becomes smaller as N gets larger. When $N = 500$ as reported in the supplement, the sizes are close to the nominal level. However, when $T = 7$ and 10, the size distortion of the standard J test based on diagonal instruments is substantial. When $T = 10$ and $N = 110$, the empirical size is 100% when all instruments are used. Contrary to these standard J tests based on diagonal instruments, the standard J test based on collapsed instruments has empirical sizes close to the nominal level. Because of this size property, in the following discussion, we only consider the standard J test based on collapsed instruments.

For the diagonal J test, we first investigate the performance of the K-means method. From the table, we find that the adjusted Rand index is almost equal to one for models in forward orthogonal deviations with $T = 4$. This indicates that almost the same grouped moment conditions are used in J_{diag}^{seq} and J_{diag}^{K}. Although the adjusted Rand index becomes a bit smaller when the system model is considered, its value is still high and we can thus say that similar grouped moment conditions are used in J_{diag}^{seq} and J_{diag}^{K}. On the effect of the number of moment

[35] For a brief introduction of the adjusted Rand index, see the supplement of Yeung and Ruzzo (2001), which is available at the author's website.

conditions m, we find that as T, and hence m, gets larger, the adjusted Rand index decreases. However, even when m is as large as 90 as in the models in forward orthogonal deviations, the adjusted Rand index exceeds 0.9 in almost all cases. Hence, we may conclude that the K-means method is quite useful to construct groups and works well even when the number of moment conditions is large.

With regard to the empirical sizes of J_{diag}^{seq} and J_{diag}^{K}, we find that the sizes are close to the nominal level except for the case $T = 7, 10$ and $N = 110$, where all instruments are used. The reason for size distortion is that the dimension of the largest block, m^*, is not small enough relative to the sample size N. Further, we find that the diagonal J test based on the one-step GMM estimation of the system model becomes more size-distorted when σ_η^2 increases from 1 to 5. This is because the bias of the one-step GMM estimator for the system becomes larger as σ_η^2 increases from 1 to 5. With regard to power, the diagonal J test has similar power to the standard J test when it works well. We also find that the results between J_{diag}^{seq} and J_{diag}^{K} are very similar. This implies that the K-means based diagonal J test, J_{diag}^{K}, is useful when the grouped structure is unknown. For the consequences of the choice of models, from the table, we find that tests based on models in forward orthogonal deviations are more powerful than those based on system models for both standard and diagonal J tests. The power decline is more evident for the standard J test based on collapsed instruments. Further, we find that, when $T = 4$, the power of the system models is substantially low. Comparing the standard J test based on collapsed instruments and the diagonal J test, we find that the latter is more powerful when $T = 4$ and less powerful when $T = 7, 10$ if models in forward orthogonal deviations are used. However, for system models, we find that in many cases, the diagonal J test yields higher power than the standard J test based on collapsed instruments for $T = 4, 7, 10$.

Next, we consider the case $\alpha = 0.8$ (Tables 5,6, and 7). Most of the above implications apply to this case as well. As in the case of $\alpha = 0.2$, the standard J test based on diagonal instruments has large size distortions in almost all cases, and works well only when $T = 4$ and $N = 500$. However, as expected, the standard J test based on collapsed instruments has empirical sizes close to the nominal level. The diagonal J test, J_{diag}^{seq} and J_{diag}^{K}, tends to have correct empirical sizes in many cases. A few exceptions arise when $T = 10, N = 110$, where all instruments are used, and when $\sigma_\eta^2 = 5$. The reason for the former is that m^*, the largest dimension of blocks, is not sufficiently small as compared to the sample size N. The reason for the latter is considered to be associated with the poor performance of estimators due to a low signal-to-noise ratio. In terms of power, we find that the diagonal J test tends to have higher power than the standard J test based on collapsed instruments in almost all cases. This implies that as a solution to the size distortion problem of the standard J test, simply reducing the number of moment conditions is not always useful. Such a test exhibits very low power, especially for system models. For the choice of the model, we find that as in the case of $\alpha = 0.2$, the standard J test based on collapsed instruments for models in forward orthogonal deviations tends to be more powerful than that for the system model when $T = 7, 10$. However, for the diagonal J test, the superiority is mixed and hence inconclusive.

Summarizing the results, the standard J test based on diagonal instruments performs well only when the number of moment conditions is sufficiently as smaller to the sample size N. The standard J test based on collapsed instruments has empirical sizes close to the nominal level in almost all cases. For the diagonal J test, except for a few cases where the dimension of the largest block m^* is not small as compared to N, the sizes are close to the nominal level. In terms of power, the diagonal J test is as powerful as the standard J test when the latter has small size distortions. As compared to the standard J test based on collapsed instruments, the superiority of the diagonal J test depends on the time length T, the degree of persistency α, and models. When $T = 7, 10$, $\alpha = 0.2$ and the models in first differences or in forward orthogonal deviations are considered(excluding system models), the standard J test based on collapsed instruments is more powerful than the diagonal J test. However, for all other cases, including all cases with $\alpha = 0.8$, the diagonal J test is more powerful.

7 Conclusion

In this paper, we proposed a new J test called the diagonal J test to check the validity of moment conditions. The diagonal J test uses a block diagonal matrix in the weighting matrix to avoid inversing a large-dimensional covariance matrix. We derived the asymptotic distribution of the new test and showed that it follows a weighted some of the chi-square distribution with one degree of freedom. Since we need to split the moment conditions into several groups when implementing the new test, we proposed two methods to construct groups. The first is to use the so-called K-means method often used in cluster analysis literature. This method is particularly useful when outside information that could be used to construct groups is not available. The second method is to utilize a special structure where moment conditions are available sequentially. Such a situation typically appears in panel data models. We then conducted a local power analysis in the context of dynamic panel data models and showed that the diagonal J test utilizing models in forward orthogonal deviations with two-step estimators has comparable power to the standard J test. Monte Carlo simulation results showed that although the standard J test substantially suffers from size distortion when the number of moment conditions is large, the proposed test has sizes close to the nominal one, regardless of the number of moment conditions. With regard to power, both the standard and diagonal J tests perform similarly when there are no size distortions in the former, as expected from a local power analysis.

Appendix

Before providing the proofs of theorems, we give some lemmas.

Lemma A1. *Let* $\widehat{\boldsymbol{\theta}}_j = \operatorname{argmin}_\theta \widehat{\mathbf{g}}(\boldsymbol{\theta})'\widehat{\mathbf{V}}_j^{-1}\widehat{\mathbf{g}}(\boldsymbol{\theta}), (j = 1step, 2step, CU)$*, where* $\widehat{\mathbf{V}}_{1step} = \widehat{\mathbf{W}}$*,* $\widehat{\mathbf{V}}_{2step} = \widehat{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\theta}})$*, and* $\widehat{\mathbf{V}}_{CU} = \widehat{\boldsymbol{\Omega}}(\boldsymbol{\theta})$*. Also, assume that* $\widehat{\mathbf{V}}_j \xrightarrow{p} \mathbf{V}_j, (j = 1step, 2step, CU)$*, where* $\mathbf{V}_{1step} = \mathbf{W}$ *and* $\mathbf{V}_{2step} = \mathbf{V}_{CU} = \boldsymbol{\Omega}$*. Then, under Assumptions 1 to 7, we have*

$$\sqrt{N}\widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j) = \left[\mathbf{I}_m - \mathbf{G}\left(\mathbf{G}'\mathbf{V}_j^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{V}_j^{-1}\right]\sqrt{N}\widehat{\mathbf{g}}(\boldsymbol{\theta}_0) + o_p(1) = \left[\mathbf{I}_m - \mathbf{K}_j\right]\sqrt{N}\widehat{\mathbf{g}}(\boldsymbol{\theta}_0) + o_p(1),$$

where $\mathbf{K}_j = \mathbf{G}\left(\mathbf{G}'\mathbf{V}_j^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{V}_j^{-1}$.

Proof. Using the mean-value theorem $\widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j) = \widehat{\mathbf{g}}(\boldsymbol{\theta}_0) + \widehat{\mathbf{G}}(\bar{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0)$, where $\bar{\boldsymbol{\theta}}$ is a mean value lying between $\widehat{\boldsymbol{\theta}}_j$ and $\boldsymbol{\theta}_0$, and the first-order condition $\widehat{\mathbf{G}}(\widehat{\boldsymbol{\theta}})'\widehat{\mathbf{V}}_j^{-1}\widehat{\mathbf{g}}(\widehat{\boldsymbol{\theta}}_j) = \mathbf{0}$, we have

$$\widehat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0 = -\left[\widehat{\mathbf{G}}(\widehat{\boldsymbol{\theta}})'\widehat{\mathbf{V}}_j^{-1}\widehat{\mathbf{G}}(\bar{\boldsymbol{\theta}})\right]^{-1}\widehat{\mathbf{G}}(\widehat{\boldsymbol{\theta}})'\widehat{\mathbf{V}}_j^{-1}\widehat{\mathbf{g}}(\boldsymbol{\theta}_0). \qquad (j = 1step, 2step, CU)$$

Substituting this into the above mean-value expansion and using $\widehat{\mathbf{V}}_j \xrightarrow{p} \mathbf{V}_j$, the result follows. □

Lemma A2. *Let* $\mathbf{\Phi}$ *be an* $m \times m$ *positive-semidefinite weighting matrix. Then, under Assumption 6, and* $H_1 : E[\mathbf{g}_i(\boldsymbol{\theta}_0)] = \mathbf{c}/\sqrt{N}$*, where* $\mathbf{c}$ *is a vector of finite constant, we have*

$$N \cdot \widehat{\mathbf{g}}(\boldsymbol{\theta}_0)'\mathbf{\Phi}\widehat{\mathbf{g}}(\boldsymbol{\theta}_0) \xrightarrow{d} \sum_{k=1}^{m} \lambda_k (z_k + \delta_k)^2,$$

where $z_k \sim iid\mathcal{N}(0,1)$*,* $\lambda_1, ..., \lambda_m$ *are the eigenvalues of* $\mathbf{\Omega}^{1/2}\mathbf{\Phi}\mathbf{\Omega}^{1/2}$*,* δ_k *is the kth element of* $\boldsymbol{\delta} = \mathbf{P}'\mathbf{\Omega}^{-1/2}\mathbf{c}$*, with* $\mathbf{P}$ *being an eigenvector matrix corresponding to* $\lambda_1, ..., \lambda_m$*.*

Proof. From Assumption 6, we have $\sqrt{N}\mathbf{\Omega}^{-1/2}\widehat{\mathbf{g}}(\boldsymbol{\theta}_0) \to^d \mathcal{N}(\mathbf{\Omega}^{-1/2}\mathbf{c}, \mathbf{I}_m) = \mathbf{z} + \mathbf{\Omega}^{-1/2}\mathbf{c}$ where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$. Using this, we obtain

$$\begin{aligned} N \cdot \widehat{\mathbf{g}}(\boldsymbol{\theta}_0)'\mathbf{\Phi}\widehat{\mathbf{g}}(\boldsymbol{\theta}_0) &= N \cdot \widehat{\mathbf{g}}(\boldsymbol{\theta}_0)'\mathbf{\Omega}^{-1/2}\left(\mathbf{\Omega}^{1/2}\mathbf{\Phi}\mathbf{\Omega}^{1/2}\right)\mathbf{\Omega}^{-1/2}\widehat{\mathbf{g}}(\boldsymbol{\theta}_0) \\ &\xrightarrow{d} (\mathbf{z} + \mathbf{\Omega}^{-1/2}\mathbf{c})'\left(\mathbf{\Omega}^{1/2}\mathbf{\Phi}\mathbf{\Omega}^{1/2}\right)(\mathbf{z} + \mathbf{\Omega}^{-1/2}\mathbf{c}) = S. \end{aligned}$$

Using the spectral decomposition $\mathbf{\Omega}^{1/2}\mathbf{\Phi}\mathbf{\Omega}^{1/2} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$, where $\mathbf{\Lambda}$ is a diagonal matrix containing eigenvalues of $\mathbf{\Omega}^{1/2}\mathbf{\Phi}\mathbf{\Omega}^{1/2}$ and $\mathbf{P}$ is the matrix of corresponding eigenvectors, we have

$$\begin{aligned} S &= (\mathbf{z} + \mathbf{\Omega}^{-1/2}\mathbf{c})'\mathbf{P}\mathbf{\Lambda}\mathbf{P}'(\mathbf{z} + \mathbf{\Omega}^{-1/2}\mathbf{c}) = (\mathbf{P}'\mathbf{z} + \mathbf{P}'\mathbf{\Omega}^{-1/2}\mathbf{c})'\mathbf{\Lambda}(\mathbf{P}'\mathbf{z} + \mathbf{P}'\mathbf{\Omega}^{-1/2}\mathbf{c}) \\ &\stackrel{d}{=} (\mathbf{z} + \boldsymbol{\delta})'\mathbf{\Lambda}(\mathbf{z} + \boldsymbol{\delta}) \stackrel{d}{=} \sum_{k=1}^{m} \lambda_k (z_k + \delta_k)^2, \end{aligned}$$

where we used $\mathbf{P}'\mathbf{z} =^d \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_m)$. □

Proof of Theorem 1

Using Lemma A1 and substituting $\mathbf{\Phi}_j = [\mathbf{I}_m - \mathbf{K}_j]' \mathbf{\Omega}_{diag}^{-1} [\mathbf{I}_m - \mathbf{K}_j], (j = 1step, 2step, CU)$ into $\mathbf{\Phi}$ in Lemma A2 with $\mathbf{c} = \mathbf{0}$, we have[36]

$$\mathcal{J}_{diag}(\widehat{\boldsymbol{\theta}}_j) = N \cdot \widehat{\mathbf{g}}(\boldsymbol{\theta}_0)' [\mathbf{I}_m - \mathbf{K}_j]' \mathbf{\Omega}_{diag}^{-1} [\mathbf{I}_m - \mathbf{K}_j] \widehat{\mathbf{g}}(\boldsymbol{\theta}_0) + o_p(1) \xrightarrow{d} \sum_{k=1}^{m} \lambda_{j,k} z_k^2,$$

where $z_k \sim iid\mathcal{N}(0,1)$, $\lambda_{j,1}, ..., \lambda_{j,m}$ are the eigenvalues of $\mathbf{H}_j = \mathbf{\Omega}^{1/2} [\mathbf{I}_m - \mathbf{K}_j]' \mathbf{\Omega}_{diag}^{-1} [\mathbf{I}_m - \mathbf{K}_j] \mathbf{\Omega}^{1/2}, (j = 1step, 2step, CU)$. Also, by noting that $\mathbf{I}_m - \mathbf{K}_j$ is an idempotent but non-symmetric matrix of rank $m - p$ and using $\text{rank}(\mathbf{AB}) \leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\}$[37], we have $\text{rank}(\mathbf{H}_j) = m - p$, which indicates that there are $m - p$ nonzero eigenvalues in $\mathbf{H}_j$. □

Proof of Theorem 2

The result for the standard J test is derived by Newey (1985). The proof for the diagonal J test is similar to that of Theorem 1, except that we do not need to impose $\mathbf{c} = \mathbf{0}$.

References

Abadir, K. M. and J. R. Magnus (2005) *Matrix Algebra*: Cambridge University Press.

Anatolyev, S. and N. Gospodinov (2010) "Specification Testing in Models with Many Instruments," *Econometric Theory*, 27, p. 427.

[36] Note that the asymptotic distribution of Hansen-Jagannathan test derived by Parker and Julliard (2005) is obtained by letting $\mathbf{\Phi}_{HJ} = [\mathbf{I}_m - \mathbf{K}_{1step}]' \mathbf{W}^{-1} [\mathbf{I}_m - \mathbf{K}_{1step}]$.

[37] See Abadir and Magnus (2005, p.81)

Andersen, T. G. and B. E. Sørensen (1996) “GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study,” *Journal of Business and Economic Statistics*, 14, 328–352.

Anderson, T. W. (2003) *An Introduction to Multivariate Statistical Analysis*: Wiley, New York.

Andrews, D. W. K. (1999) “Consistent Moment Selection Procedures for Generalized Method of Moments Estimation,” *Econometrica*, 67, 543-564.

Arellano, M. (2003) *Panel Data Econometrics*, Oxford: Oxford University Press.

Arellano, M. and S. Bond (1991) “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *Review of Economic Studies*, 58, 277-297.

Arellano, M. and O. Bover (1995) “Another Look at the Instrumental Variable Estimation of Error-Components Models,” *Journal of Econometrics*, 68, 29-51.

Bai, J. and S. Shi (2011) “Estimating High Dimensional Covariance Matrices and its Applications,” *Annals of Economics and Finance*, 12, 199-215.

Bickel, P. J. and E. Levina (2008) “Regularized Estimation of Large Covariance Matrices,” *The Annals of Statistics*, 36, 199–227.

Blundell, R. and S. Bond (1998) “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models,” *Journal of Econometrics*, 87, 115-143.

Blundell, R., S. Bond, and F. Windmeijer (2000) “Estimation in Dynamic Panel Data Models: Improving on the Performance of the Standard GMM Estimator,” in Baltagi, B. H. ed. *Non-stationary Panels, Panel Cointegration and Dynamic Panels*, 15 of Advances in Econometrics, Amsterdam: JAI Press, 53-91.

Blundell, R., R. Griffith, and F. Windmeijer (2002) “Individual Effects and Dynamics in Count Data Models,” *Journal of Econometrics*, 108, 113–131.

Bonhomme, S. and E. Manresa (2014) “Grouped Patterns of Heterogeneity in Panel Data.” working paper.

Bowsher, C. G. (2002) “On Testing Overidentifying Restrictions in Dynamic Panel Data Models,” *Economics Letters*, 77, 211-220.

Bun, M. J. G. and J. F. Kiviet (2006) “The Effects of Dynamic Feedbacks on LS and MM Estimator Accuracy in Panel Data Models,” *Journal of Econometrics*, 132, 409-444.

Bun, M. J. G. and F. Windmeijer (2010) “The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models,” *Econometrics Journal*, 13, 95-126.

Chamberlain, G. (1992) “Comment: Sequential Moment Restrictions in Panel Data,” *Journal of Business and Economic Statistics*, 10, 20-26.

Chao, J. C., J. A. Hausman, W. K. Newey, N. R. Swanson, and T. Woutersen (2013) “Testing Overidentifying Restrictions with Many Instruments and Heteroskedasticity,” *Journal of Econometrics*.

Fan, J., Y. Fan, and J. Lv (2008) "High Dimensional Covariance Matrix Estimation Using a Factor Model," *Journal of Econometrics*, 147, 186–197.

Fan, J., Y. Liao, and M. Mincheva (2011) "High Dimensional Covariance Matrix Estimation in Approximate Factor Models," *Annals of Statistics*, 39, 3320–3356.

Hall, A. (2005) *Generalized Method of Moments*, Oxford: Oxford University Press.

Hall, A. R (2000) "Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test," *Econometrica*, 68, 1517–1527.

Hansen, L. P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

Hansen, L. P., J. Heaton, and A. Yaron (1996) "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-80.

Hastie, T., R. Tibshirani, and J. Friedman (2009) *The Elements of Statistical Learning*, 2: Springer.

Hayakawa, K. (2007) "Small Sample Bias Properties of the System GMM Estimator in Dynamic Panel Data Models," *Economics Letters*, 95, 32-38.

Hayakawa, K. (2009) "On the Effect of Mean-Nonstationarity in Dynamic Panel Data Models," *Journal of Econometrics*, 153, 133-135.

Hubert, L. and P. Arabie (1985) "Comparing Partitions," *Journal of Classification*, 2, 193–218.

Imhof, J. P. (1961) "Computing the Distribution of Quadratic Forms in Normal Variables," *Biometrika*, 48, 419-426.

Jagannathan, R. and Z. Wang (1996) "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3–53.

Keane, M. P. and D. E. Runkle (1992) "On the Estimation of Panel-Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous," *Journal of Business and Economic Statistics*, 10, 1-9.

Kiviet, J. F. (2007) "On the Optimal Weighting Matrix for the GMM System Estimator in Dynamic Panel Data Models." working paper.

Kleibergen, F. (2005) "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 73, 1103-1123.

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

Ledoit, O. and M. Wolf (2004) "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices," *Journal of Multivariate Analysis*, 88, 365–411.

Ledoit, O. and M. Wolf (2012) "Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices," *Annals of Statistics*, 40, 1024–1060.

Lee, Y and R. Okui (2012) "Hahn–Hausman Test as A Specification Test," *Journal of Econometrics*, 167, 133–139.

Lin, C. C. and S. Ng (2012) "Estimation of Panel Data Models with Parameter Heterogeneity when Group Membership is Unknown," *Journal of Econometric Methods*, 1, 42–55.

Montalvo, J. G. (1997) "GMM Estimation of Count-Panel-Data Models with Fixed Effects and Predetermined Instruments," *Journal of Business & Economic Statistics*, 15, 82–89.

Newey, W. K (1985) "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229–256.

Newey, W. K. and R. J. Smith (2004) "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

Newey, W. K. and F. Windmeijer (2009) "Generalized Method of Moments With Many Weak Moment Conditions," *Econometrica*, 77, 687–719.

Okui, R. (2009) "The Optimal Choice of Moments in Dynamic Panel Data Models," *Journal of Econometrics*, 151, 1-16.

Parker, J. A. and C. Julliard (2005) "Consumption Risk and the Cross Section of Expected Returns," *Journal of Political Economy*, 113, 185-222.

Roodman, D. (2009) "A Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics and Statistics*, 71, 135-158.

Sarafidis, V. and N. Weber (2015) "A Partially Heterogeneous Framework for Analyzing Panel Data," *Oxford Bulletin of Economics and Statistics*.

Sargan, J. D. (1958) "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

Shapiro, M. D. (1984) "The Permanent Income Hypothesis and the Real Interest Rate: Some Evidence from Panel Data," *Economics Letters*, 14, 93–100.

Stock, J. H. and J. Wright (2000) "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.

Tauchen, G. (1986) "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics*, 4, 397-416.

Ullah, A. (2004) *Finite Sample Econometrics*: Oxford University Press.

Windmeijer, F. (2005) "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators," *Journal of Econometrics*, 126, 25-51.

Windmeijer, F. (2008) "GMM for Panel Count Data Models," in *The Econometrics of Panel Data; Fundamentals and Recent Developments in Theory and Practice*: L. Matyas and P. Sevestre, Chap. 18.

Wooldridge, J. M. (1997) "Multiplicative Panel Data Models Without the Strict Exogeneity Assumption," *Econometric Theory*, 13, 667–678.

Wooldridge, J. M. (2001) *Econometric Analysis of Cross Section and Panel Data*: MIT Press.

Yeung, K. Y. and W. L. Ruzzo (2001) “Principal Component Analysis for Clustering Gene Expression Data,” *Bioinformatics*, 17, 763-774.

Zeldes, S. P. (1989) “Comsumption and Liquidity Constraints: An Empirical Investigation,” *Journal of Political Economy*, 97, 305-346.

Table 2: Simulation results with $T = 4$, $\alpha = 0.2$, $\beta = 1.0$

				Size			Power						Adjusted Rand Index		
				$\phi = 0$			$\phi = 0.2$			$\phi = 0.5$			ϕ		
Model/IV	Method	m	m^*	J	J_{diag}	J_{diag}^K	J	J_{diag}	J_{diag}^K	J	J_{diag}	J_{diag}^K	0	0.2	0.5
$N = 110,\ \sigma_\eta^2 = 1$															
FOD/All	1step	12	6	–	5.9	5.9	–	7.7	7.7	–	34.8	34.8	0.996	0.994	0.993
	2step	12	6	10.4	6.2	6.3	12.3	7.4	7.5	47.8	33.5	33.4	0.995	0.996	0.994
	CUE	12	6	6.4	5.0	5.0	6.8	5.2	5.2	29.1	25.6	25.5	0.994	0.994	0.991
FOD/Sargan	1step	10	4	–	5.5	5.5	–	6.9	6.9	–	30.4	30.4	0.999	0.999	1.000
	2step	10	4	8.3	5.6	5.6	9.5	6.6	6.6	38.1	28.6	28.6	0.999	0.999	0.999
	CUE	10	4	5.4	4.4	4.4	6.4	5.6	5.6	21.1	21.3	21.3	1.000	1.000	1.000
FOD/Collapsed	2step	4	–	4.6	–	–	5.4	–	–	23.1	–	–	–	–	–
	CUE	4	–	3.8	–	–	4.9	–	–	17.7	–	–	–	–	–
FOD & LEV/All	1step	18	6	–	3.3	3.6	–	3.7	3.7	–	19.1	17.2	0.888	0.905	0.928
	2step	18	6	16.2	3.7	4.1	17.5	3.8	4.2	48.5	22.2	19.0	0.885	0.901	0.921
	CUE	18	6	11.7	5.2	5.1	11.2	5.8	5.6	45.3	24.5	22.5	0.887	0.895	0.912
FOD & LEV/Sargan	1step	16	4	–	3.9	4.8	–	3.7	4.0	–	15.2	13.2	0.845	0.868	0.907
	2step	16	4	13.2	3.9	3.9	15.1	4.1	4.1	41.5	19.6	17.3	0.842	0.864	0.900
	CUE	16	4	9.1	5.1	4.6	10.3	5.8	4.7	37.8	22.4	18.5	0.843	0.853	0.885
FOD & LEV/Collapsed	2step	4	–	5.3	–	–	4.8	–	–	4.9	–	–	–	–	–
	CUE	4	–	4.7	–	–	4.6	–	–	4.1	–	–	–	–	–
$N = 110,\ \sigma_\eta^2 = 5$															
FOD/All	1step	12	6	–	6.0	6.0	–	8.1	8.2	–	33.9	33.8	0.996	0.996	0.994
	2step	12	6	10.6	5.8	5.8	12.8	7.6	7.7	47.0	32.3	32.3	0.995	0.995	0.994
	CUE	12	6	5.9	5.1	5.1	5.7	4.4	4.4	21.9	18.6	18.5	0.993	0.993	0.990
FOD/Sargan	1step	10	4	–	5.3	5.3	–	6.4	6.4	–	25.0	25.0	0.999	0.999	1.000
	2step	10	4	8.6	5.3	5.3	9.8	6.7	6.7	32.6	23.6	23.6	0.999	0.999	0.999
	CUE	10	4	4.9	5.1	5.1	4.9	3.8	3.8	9.3	9.2	9.2	1.000	1.000	1.000
FOD/Collapsed	2step	4	–	4.3	–	–	5.2	–	–	16.8	–	–	–	–	–
	CUE	4	–	2.5	–	–	3.7	–	–	7.3	–	–	–	–	–
FOD & LEV/All	1step	18	6	–	8.0	8.7	–	6.7	6.6	–	16.8	14.1	0.867	0.884	0.924
	2step	18	6	23.1	8.0	8.5	25.1	5.2	6.3	53.6	18.4	15.0	0.858	0.874	0.911
	CUE	18	6	12.0	5.6	6.6	12.3	5.2	5.5	42.1	18.5	13.5	0.842	0.843	0.864
FOD & LEV/Sargan	1step	16	4	–	9.7	9.3	–	6.7	7.0	–	12.7	11.2	0.815	0.841	0.897
	2step	16	4	23.3	8.3	8.6	23.4	5.6	6.2	47.0	16.0	13.3	0.805	0.829	0.875
	CUE	16	4	10.3	5.6	5.6	11.1	5.1	4.6	33.9	15.3	11.2	0.786	0.787	0.815
FOD & LEV/Collapsed	2step	4	–	6.1	–	–	5.2	–	–	4.2	–	–	–	–	–
	CUE	4	–	4.0	–	–	3.5	–	–	2.7	–	–	–	–	–
$N = 250,\ \sigma_\eta^2 = 1$															
FOD/All	1step	12	6	–	5.9	5.9	–	8.8	8.8	–	72.9	72.8	0.998	0.998	0.997
	2step	12	6	8.4	6.4	6.4	12.1	8.9	8.9	78.6	72.3	72.2	0.995	0.995	0.993
	CUE	12	6	6.9	5.2	5.2	10.2	8.4	8.4	72.9	70.6	70.6	0.998	0.996	0.995
FOD/Sargan	1step	12	6	–	5.9	5.9	–	8.8	8.8	–	72.9	72.8	0.998	0.997	0.995
	2step	12	6	8.4	6.4	6.4	12.1	8.9	8.8	78.6	72.3	72.3	0.995	0.993	0.993
	CUE	12	6	6.9	5.2	5.2	10.2	8.4	8.4	72.9	70.6	70.6	0.995	0.997	0.994
FOD/Collapsed	2step	4	–	6.4	–	–	8.3	–	–	52.4	–	–	–	–	–
	CUE	4	–	6.2	–	–	7.8	–	–	47.8	–	–	–	–	–
FOD & LEV/All	1step	18	6	–	5.1	5.1	–	6.3	6.5	–	53.9	52.1	0.912	0.932	0.968
	2step	18	6	10.9	4.9	5.0	14.6	6.9	6.9	78.8	64.7	60.0	0.902	0.926	0.957
	CUE	18	6	8.7	5.0	5.0	12.8	7.4	7.1	79.0	65.5	59.9	0.902	0.918	0.948
FOD & LEV/Sargan	1step	18	6	–	5.1	5.2	–	6.3	6.5	–	53.9	51.5	0.911	0.932	0.967
	2step	18	6	10.9	4.9	5.0	14.6	6.9	6.9	78.8	64.7	59.8	0.904	0.926	0.956
	CUE	18	6	8.7	5.0	4.9	12.8	7.4	6.9	79.0	65.5	59.7	0.902	0.919	0.944
FOD & LEV/Collapsed	2step	4	–	5.9	–	–	5.6	–	–	7.7	–	–	–	–	–
	CUE	4	–	5.8	–	–	5.2	–	–	7.0	–	–	–	–	–
$N = 250,\ \sigma_\eta^2 = 5$															
FOD/All	1step	12	6	–	6.5	6.5	–	7.6	7.6	–	68.6	68.5	0.997	0.998	0.996
	2step	12	6	8.2	6.1	6.1	10.6	8.8	8.8	74.0	67.5	67.5	0.995	0.995	0.993
	CUE	12	6	6.6	5.4	5.4	8.6	6.6	6.6	58.1	55.6	55.6	0.996	0.996	0.994
FOD/Sargan	1step	12	6	–	6.5	6.6	–	7.6	7.7	–	68.6	68.3	0.995	0.995	0.994
	2step	12	6	8.2	6.1	6.1	10.6	8.8	8.7	74.0	67.5	67.5	0.996	0.995	0.996
	CUE	12	6	6.6	5.4	5.4	8.6	6.6	6.5	58.1	55.6	55.4	0.997	0.996	0.993
FOD/Collapsed	2step	4	–	5.8	–	–	7.2	–	–	38.1	–	–	–	–	–
	CUE	4	–	5.6	–	–	6.4	–	–	27.8	–	–	–	–	–
FOD & LEV/All	1step	18	6	–	11.2	11.0	–	9.7	9.8	–	48.2	41.6	0.854	0.871	0.945
	2step	18	6	18.1	8.6	8.8	24.1	9.1	9.6	82.6	57.3	44.2	0.838	0.855	0.908
	CUE	18	6	9.2	5.0	5.9	12.9	6.7	6.6	76.2	57.5	34.8	0.821	0.812	0.845
FOD & LEV/Sargan	1step	18	6	–	11.2	11.1	–	9.7	9.9	–	48.2	41.9	0.855	0.872	0.944
	2step	18	6	18.1	8.6	8.8	24.1	9.1	9.8	82.6	57.3	43.8	0.838	0.854	0.908
	CUE	18	6	9.2	5.0	6.1	12.9	6.7	6.6	76.2	57.5	35.3	0.820	0.811	0.846
FOD & LEV/Collapsed	2step	4	–	5.5	–	–	5.4	–	–	5.4	–	–	–	–	–
	CUE	4	–	5.5	–	–	5.0	–	–	4.2	–	–	–	–	–

Note: The entries are in %. The significance level is 5%. J and J_{diag} are defined in (3), (4) and (8), respectively. The model "FOD" and "FOD & LEV" denote the models in forward orthogonal deviations and the system models combining models in forward orthogonal deviations and in levels, respectively. The instruments "All" uses the all past variables as instruments. "Sargan" uses the fixed number of instruments where lag length is determined by Sargan rule of thumb. "Collapsed" use instruments (31) or (32). m denotes the number of moment conditions and m^* denotes the dimension of the largest block in the diagonal J test.

Table 3: Simulation results with $T = 7$, $\alpha = 0.2$, $\beta = 1.0$

				Size			Power						Adjusted Rand Index		
				$\phi = 0$			$\phi = 0.2$			$\phi = 0.5$			ϕ		
Model/IV	Method	m	m^*	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	0	0.2	0.5
$N = 110,\ \sigma^2_\eta = 1$															
FOD/All	1step	42	12	–	9.5	9.3	–	11.4	11.0	–	65.8	64.1	0.958	0.956	0.946
	2step	42	12	70.9	9.6	9.4	74.5	10.8	10.3	98.7	62.2	60.9	0.951	0.951	0.943
	CUE	42	12	65.0	11.8	11.3	68.6	14.1	13.3	97.3	68.9	67.2	0.954	0.957	0.949
FOD/Sargan	1step	22	4	–	5.1	5.2	–	6.2	6.4	–	52.3	51.6	0.973	0.973	0.970
	2step	22	4	22.4	4.8	4.5	26.5	6.2	6.2	78.2	51.8	51.4	0.976	0.975	0.971
	CUE	22	4	18.4	5.2	5.0	22.6	7.2	7.1	72.3	52.4	51.6	0.975	0.976	0.971
FOD/Collapsed	2step	4	–	5.2	–	–	11.3	–	–	81.9	–	–	–	–	–
	CUE	4	–	4.8	–	–	11.1	–	–	80.9	–	–	–	–	–
FOD & LEV/All	1step	54	12	–	5.9	5.7	–	6.2	6.2	–	51.0	41.3	0.871	0.875	0.873
	2step	54	12	93.8	5.3	5.3	94.3	5.9	6.6	99.6	55.3	45.2	0.867	0.869	0.871
	CUE	54	12	91.4	11.3	9.8	92.8	12.9	12.0	99.8	64.7	50.4	0.871	0.864	0.856
FOD & LEV/Sargan	1step	34	4	–	4.3	4.4	–	4.4	4.2	–	32.3	27.7	0.836	0.858	0.892
	2step	34	4	55.2	4.0	4.7	64.5	4.0	4.8	96.9	48.6	38.4	0.832	0.855	0.880
	CUE	34	4	48.0	5.9	5.7	54.4	9.3	7.4	94.3	61.2	48.0	0.832	0.842	0.853
FOD & LEV/Collapsed	2step	4	–	5.5	–	–	5.8	–	–	28.4	–	–	–	–	–
	CUE	4	–	5.2	–	–	4.9	–	–	27.4	–	–	–	–	–
$N = 110,\ \sigma^2_\eta = 5$															
FOD/All	1step	42	12	–	9.7	10.0	–	11.8	11.5	–	65.9	65.1	0.957	0.957	0.952
	2step	42	12	70.2	9.4	9.3	74.0	11.2	10.6	98.5	61.8	60.7	0.954	0.954	0.951
	CUE	42	12	64.8	11.4	11.3	68.3	14.0	13.4	97.0	65.6	63.9	0.959	0.959	0.953
FOD/Sargan	1step	22	4	–	4.8	4.9	–	5.7	5.9	–	36.4	35.8	0.959	0.958	0.954
	2step	22	4	21.9	4.3	4.3	24.4	5.1	5.1	64.9	33.4	33.3	0.960	0.964	0.956
	CUE	22	4	18.3	4.9	4.7	20.9	6.3	6.2	53.2	29.8	29.4	0.957	0.959	0.959
FOD/Collapsed	2step	4	–	5.0	–	–	7.5	–	–	60.6	–	–	–	–	–
	CUE	4	–	4.5	–	–	7.5	–	–	52.5	–	–	–	–	–
FOD & LEV/All	1step	54	12	–	11.9	11.8	–	10.1	10.7	–	44.7	35.2	0.925	0.928	0.935
	2step	54	12	94.5	11.1	11.5	95.1	9.5	9.3	99.9	50.9	40.0	0.921	0.924	0.932
	CUE	54	12	91.8	11.0	12.5	92.3	11.8	10.6	99.7	56.1	40.1	0.917	0.911	0.908
FOD & LEV/Sargan	1step	34	4	–	14.2	14.7	–	8.4	8.6	–	22.2	18.8	0.806	0.833	0.894
	2step	34	4	70.3	10.3	10.4	77.5	6.9	7.9	98.0	38.9	30.1	0.796	0.823	0.879
	CUE	34	4	46.2	6.0	5.6	51.6	6.9	6.6	93.4	46.6	34.2	0.774	0.765	0.790
FOD & LEV/Collapsed	2step	4	–	5.0	–	–	5.6	–	–	15.8	–	–	–	–	–
	CUE	4	–	5.3	–	–	4.8	–	–	13.3	–	–	–	–	–
$N = 250,\ \sigma^2_\eta = 1$															
FOD/All	1step	42	12	–	6.6	6.8	–	11.6	11.4	–	99.0	98.9	0.955	0.962	0.946
	2step	42	12	27.9	6.5	6.3	38.9	11.4	11.2	100.0	98.7	98.3	0.962	0.967	0.956
	CUE	42	12	25.3	5.8	5.8	35.3	11.3	10.6	100.0	98.9	98.6	0.963	0.964	0.958
FOD/Sargan	1step	42	12	–	6.6	6.3	–	11.6	10.9	–	99.0	98.9	0.962	0.962	0.950
	2step	42	12	27.9	6.5	6.3	38.9	11.4	10.5	100.0	98.7	98.5	0.958	0.961	0.956
	CUE	42	12	25.3	5.8	5.7	35.3	11.3	10.6	100.0	98.9	98.8	0.958	0.965	0.959
FOD/Collapsed	2step	4	–	6.1	–	–	15.2	–	–	99.5	–	–	–	–	–
	CUE	4	–	6.1	–	–	15.3	–	–	99.4	–	–	–	–	–
FOD & LEV/All	1step	54	12	–	5.9	5.7	–	9.6	8.4	–	96.7	92.7	0.898	0.902	0.906
	2step	54	12	44.3	5.9	5.2	56.6	9.6	8.7	100.0	98.7	95.4	0.893	0.896	0.897
	CUE	54	12	42.3	6.5	6.1	54.4	11.8	10.4	100.0	98.9	95.1	0.896	0.889	0.881
FOD & LEV/Sargan	1step	54	12	–	5.9	5.2	–	9.6	8.5	–	96.7	92.5	0.896	0.898	0.905
	2step	54	12	44.3	5.9	5.5	56.6	9.6	9.0	100.0	98.7	95.9	0.897	0.897	0.892
	CUE	54	12	42.3	6.5	6.0	54.4	11.8	10.6	100.0	98.9	94.9	0.898	0.895	0.882
FOD & LEV/Collapsed	2step	4	–	5.6	–	–	7.8	–	–	52.0	–	–	–	–	–
	CUE	4	–	5.5	–	–	7.8	–	–	51.6	–	–	–	–	–
$N = 250,\ \sigma^2_\eta = 5$															
FOD/All	1step	42	12	–	6.5	7.1	–	11.7	11.6	–	98.6	98.6	0.959	0.962	0.953
	2step	42	12	28.3	6.1	6.2	38.6	10.4	10.3	99.9	98.2	97.5	0.964	0.970	0.962
	CUE	42	12	25.5	5.7	5.8	34.3	10.3	9.7	99.7	98.0	97.6	0.963	0.967	0.963
FOD/Sargan	1step	42	12	–	6.5	6.6	–	11.7	10.6	–	98.6	98.2	0.961	0.961	0.953
	2step	42	12	28.3	6.1	5.7	38.6	10.4	9.6	99.9	98.2	97.6	0.961	0.964	0.961
	CUE	42	12	25.5	5.7	5.7	34.3	10.3	10.0	99.7	98.0	97.6	0.961	0.966	0.967
FOD/Collapsed	2step	4	–	6.0	–	–	10.2	–	–	91.8	–	–	–	–	–
	CUE	4	–	6.0	–	–	10.3	–	–	91.0	–	–	–	–	–
FOD & LEV/All	1step	54	12	–	13.3	12.0	–	14.3	12.6	–	94.9	82.7	0.917	0.917	0.940
	2step	54	12	52.1	9.6	8.3	66.5	10.9	10.9	100.0	97.0	84.8	0.909	0.913	0.932
	CUE	54	12	42.4	6.1	6.5	53.0	9.2	9.1	100.0	97.6	78.6	0.904	0.895	0.902
FOD & LEV/Sargan	1step	54	12	–	13.3	11.8	–	14.3	12.5	–	94.9	83.3	0.916	0.916	0.941
	2step	54	12	52.1	9.6	8.4	66.5	10.9	11.0	100.0	97.0	84.5	0.910	0.911	0.932
	CUE	54	12	42.4	6.1	6.8	53.0	9.2	9.5	100.0	97.6	78.6	0.905	0.895	0.899
FOD & LEV/Collapsed	2step	4	–	5.2	–	–	5.8	–	–	31.6	–	–	–	–	–
	CUE	4	–	5.4	–	–	6.0	–	–	32.0	–	–	–	–	–

Table 4: Simulation results with $T = 10$, $\alpha = 0.2$, $\beta = 1.0$

				Size			Power						Adjusted Rand Index		
				$\phi = 0$			$\phi = 0.2$			$\phi = 0.5$			ϕ		
Model/IV	Method	m	m^*	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	0	0.2	0.5
$N = 110$, $\sigma^2_\eta = 1$															
FOD/All	1step	90	18	–	14.2	13.1	–	14.0	12.8	–	84.9	82.7	0.933	0.936	0.928
	2step	90	18	100.0	18.2	17.5	100.0	19.2	17.8	100.0	84.6	82.5	0.933	0.932	0.931
	CUE	90	18	100.0	33.1	31.1	100.0	32.4	30.7	100.0	91.8	89.7	0.929	0.933	0.929
FOD/Sargan	1step	34	4	–	5.3	5.7	–	6.3	7.1	–	70.3	69.8	0.936	0.933	0.930
	2step	34	4	48.6	5.7	5.6	52.4	7.1	7.4	97.5	71.2	70.6	0.940	0.935	0.933
	CUE	34	4	43.7	6.3	6.4	49.4	8.9	8.9	96.4	72.8	71.9	0.934	0.936	0.927
FOD/Collapsed	2step	4	–	6.5	–	–	16.3	–	–	97.1	–	–	–	–	–
	CUE	4	–	6.3	–	–	16.3	–	–	96.9	–	–	–	–	–
FOD & LEV/All	1step	108	18	–	6.0	5.1	–	8.2	6.6	–	77.2	59.1	0.808	0.811	0.817
	2step	108	18	100.0	39.0	34.5	100.0	39.9	33.6	100.0	88.8	77.7	0.816	0.814	0.809
	CUE	108	18	100.0	45.2	40.2	100.0	44.9	39.4	100.0	88.7	80.1	0.819	0.810	0.804
FOD & LEV/Sargan	1step	52	4	–	3.9	3.9	–	5.1	4.9	–	45.7	40.0	0.828	0.849	0.879
	2step	52	4	91.1	3.3	3.7	95.4	5.5	4.9	100.0	69.3	56.0	0.829	0.844	0.871
	CUE	52	4	87.8	6.8	7.0	91.4	11.8	10.3	100.0	84.9	73.0	0.826	0.825	0.829
FOD & LEV/Collapsed	2step	4	–	4.8	–	–	9.0	–	–	73.8	–	–	–	–	–
	CUE	4	–	4.5	–	–	8.7	–	–	73.5	–	–	–	–	–
$N = 110$, $\sigma^2_\eta = 5$															
FOD/All	1step	90	18	–	13.9	13.1	–	13.8	13.8	–	84.4	84.1	0.934	0.933	0.931
	2step	90	18	100.0	17.9	17.6	100.0	20.1	17.5	100.0	84.4	83.6	0.930	0.930	0.930
	CUE	90	18	100.0	32.6	30.8	100.0	31.2	30.9	100.0	90.7	88.5	0.928	0.931	0.929
FOD/Sargan	1step	34	4	–	5.7	5.5	–	6.2	6.6	–	50.6	49.8	0.906	0.906	0.895
	2step	34	4	47.8	5.8	5.2	51.6	6.3	6.0	92.3	49.8	48.8	0.910	0.909	0.904
	CUE	34	4	42.7	6.6	7.1	47.5	7.9	8.6	87.6	47.7	44.8	0.907	0.905	0.900
FOD/Collapsed	2step	4	–	6.7	–	–	11.6	–	–	86.1	–	–	–	–	–
	CUE	4	–	6.4	–	–	11.5	–	–	84.6	–	–	–	–	–
FOD & LEV/All	1step	108	18	–	16.0	12.8	–	13.9	12.0	–	74.8	58.1	0.924	0.919	0.918
	2step	108	18	100.0	38.9	36.6	100.0	41.3	35.8	100.0	84.4	75.4	0.915	0.913	0.910
	CUE	108	18	100.0	47.9	45.3	100.0	48.9	46.7	100.0	87.1	75.8	0.920	0.917	0.910
FOD & LEV/Sargan	1step	52	4	–	16.3	16.6	–	8.2	9.9	–	27.8	22.7	0.800	0.820	0.884
	2step	52	4	96.7	12.2	12.6	98.9	7.9	8.3	100.0	54.7	42.9	0.794	0.816	0.876
	CUE	52	4	87.2	6.9	6.4	91.8	9.0	8.6	100.0	71.4	50.6	0.771	0.758	0.781
FOD & LEV/Collapsed	2step	4	–	3.7	–	–	5.7	–	–	49.3	–	–	–	–	–
	CUE	4	–	3.7	–	–	5.4	–	–	48.2	–	–	–	–	–
$N = 250$, $\sigma^2_\eta = 1$															
FOD/All	1step	90	18	–	5.9	5.0	–	11.5	11.6	–	99.9	99.9	0.934	0.937	0.934
	2step	90	18	91.1	6.2	6.5	94.9	11.3	12.3	100.0	99.9	99.9	0.937	0.941	0.937
	CUE	90	18	89.3	6.2	5.8	94.2	12.3	12.3	100.0	100.0	99.9	0.938	0.942	0.942
FOD/Sargan	1step	78	12	–	5.3	4.6	–	10.5	10.4	–	100.0	99.9	0.948	0.949	0.943
	2step	78	12	76.8	4.8	5.3	85.9	10.8	10.5	100.0	100.0	99.8	0.947	0.949	0.944
	CUE	78	12	74.3	4.7	4.1	84.8	11.8	11.3	100.0	100.0	99.8	0.947	0.951	0.949
FOD/Collapsed	2step	4	–	4.6	–	–	27.9	–	–	100.0	–	–	–	–	–
	CUE	4	–	4.7	–	–	28.0	–	–	100.0	–	–	–	–	–
FOD & LEV/All	1step	108	18	–	4.0	4.4	–	8.9	5.8	–	99.9	99.4	0.831	0.831	0.843
	2step	108	18	98.6	4.8	3.6	99.3	9.3	7.7	100.0	100.0	99.3	0.832	0.827	0.833
	CUE	108	18	98.5	6.4	4.7	99.2	13.2	10.4	100.0	100.0	99.4	0.835	0.827	0.813
FOD & LEV/Sargan	1step	96	12	–	3.8	3.5	–	8.4	6.1	–	99.8	99.2	0.878	0.875	0.882
	2step	96	12	93.8	3.3	2.8	97.4	8.3	6.8	100.0	100.0	99.3	0.877	0.873	0.872
	CUE	96	12	93.4	4.8	5.1	97.1	11.1	8.8	100.0	100.0	99.3	0.881	0.871	0.854
FOD & LEV/Collapsed	2step	4	–	5.3	–	–	12.6	–	–	98.7	–	–	–	–	–
	CUE	4	–	5.3	–	–	12.7	–	–	98.7	–	–	–	–	–
$N = 250$, $\sigma^2_\eta = 5$															
FOD/All	1step	90	18	–	5.9	5.3	–	11.2	10.7	–	99.9	99.9	0.934	0.936	0.935
	2step	90	18	90.3	6.6	6.3	94.9	11.3	10.3	100.0	99.9	99.9	0.936	0.939	0.937
	CUE	90	18	88.4	6.6	6.0	93.8	11.8	11.2	100.0	99.9	99.8	0.936	0.939	0.938
FOD/Sargan	1step	78	12	–	5.1	5.1	–	10.8	10.7	–	100.0	100.0	0.942	0.943	0.938
	2step	78	12	76.5	4.7	4.7	84.7	10.2	10.5	100.0	100.0	99.9	0.942	0.941	0.937
	CUE	78	12	73.7	5.0	4.5	83.8	11.1	10.7	100.0	100.0	99.8	0.941	0.942	0.938
FOD/Collapsed	2step	4	–	4.8	–	–	14.6	–	–	99.6	–	–	–	–	–
	CUE	4	–	4.6	–	–	14.6	–	–	99.6	–	–	–	–	–
FOD & LEV/All	1step	108	18	–	10.8	8.4	–	13.5	11.3	–	99.7	97.3	0.934	0.927	0.931
	2step	108	18	98.9	7.0	7.0	99.7	10.9	9.0	100.0	99.7	96.7	0.929	0.924	0.926
	CUE	108	18	98.5	5.8	5.7	99.2	11.5	9.7	100.0	99.9	96.2	0.930	0.921	0.910
FOD & LEV/Sargan	1step	96	12	–	11.2	9.5	–	13.7	11.2	–	99.7	97.2	0.924	0.919	0.940
	2step	96	12	95.2	6.8	6.5	98.8	10.4	8.2	100.0	99.9	97.1	0.920	0.916	0.930
	CUE	96	12	93.0	4.3	5.2	97.0	10.4	8.5	100.0	99.8	98.1	0.919	0.908	0.902
FOD & LEV/Collapsed	2step	4	–	4.5	–	–	6.7	–	–	87.4	–	–	–	–	–
	CUE	4	–	4.8	–	–	7.0	–	–	87.3	–	–	–	–	–

Table 5: Simulation results with $T = 4$, $\alpha = 0.8$, $\beta = 1.0$

				Size			Power						Adjusted Rand Index		
				$\phi = 0$			$\phi = 0.2$			$\phi = 0.5$			ϕ		
Model/IV	Method	m	m^*	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	0	0.2	0.5
$N = 110,\ \sigma^2_\eta = 1$															
FOD/All	1step	12	6	–	6.8	6.8	–	13.7	13.9	–	31.4	30.7	0.989	0.969	0.875
	2step	12	6	11.6	6.4	6.5	17.9	11.3	11.3	36.2	25.7	24.7	0.982	0.953	0.828
	CUE	12	6	2.8	2.9	2.7	3.4	3.9	3.8	13.1	14.3	14.0	0.971	0.922	0.783
FOD/Sargan	1step	10	4	–	5.8	5.8	–	13.2	13.2	–	31.6	30.2	0.996	0.980	0.870
	2step	10	4	8.8	5.0	5.1	15.6	10.9	10.8	33.1	27.2	26.9	0.991	0.962	0.827
	CUE	10	4	2.1	2.2	2.2	2.4	3.6	3.4	7.7	11.8	11.1	0.986	0.940	0.759
FOD/Collapsed	2step	4	–	4.5	–	–	6.1	–	–	11.3	–	–	–	–	–
	CUE	4	–	1.5	–	–	2.0	–	–	3.1	–	–	–	–	–
FOD & LEV/All	1step	18	6	–	4.6	4.0	–	9.8	9.6	–	75.9	60.1	0.880	0.912	0.893
	2step	18	6	15.7	4.9	5.5	27.7	9.6	9.8	92.6	79.7	68.5	0.880	0.909	0.894
	CUE	18	6	9.4	4.9	5.2	22.1	10.2	9.6	89.1	70.1	55.0	0.857	0.897	0.876
FOD & LEV/Sargan	1step	16	4	–	4.9	4.4	–	9.9	9.2	–	76.0	49.0	0.835	0.872	0.847
	2step	16	4	13.4	5.2	4.8	22.4	10.4	9.2	87.5	79.9	59.5	0.830	0.869	0.849
	CUE	16	4	7.1	4.1	4.1	17.2	10.0	8.5	83.8	69.2	45.1	0.804	0.853	0.827
FOD & LEV/Collapsed	2step	4	–	4.8	–	–	4.8	–	–	8.4	–	–	–	–	–
	CUE	4	–	3.3	–	–	2.3	–	–	4.1	–	–	–	–	–
$N = 110,\ \sigma^2_\eta = 5$															
FOD/All	1step	12	6	–	6.4	6.4	–	11.6	11.6	–	24.4	24.4	1.000	1.000	0.999
	2step	12	6	11.0	5.5	5.5	16.7	10.2	10.3	29.6	19.5	19.5	0.999	0.999	0.998
	CUE	12	6	2.1	2.5	2.5	3.2	2.7	2.7	9.6	9.5	9.4	1.000	1.000	0.996
FOD/Sargan	1step	10	4	–	5.0	5.0	–	11.6	11.6	–	22.2	22.2	1.000	1.000	1.000
	2step	10	4	8.4	4.3	4.3	14.1	9.7	9.7	25.2	19.3	19.3	1.000	1.000	1.000
	CUE	10	4	1.8	1.5	1.5	2.1	2.8	2.8	4.9	7.1	7.1	1.000	1.000	1.000
FOD/Collapsed	2step	4	–	4.6	–	–	4.7	–	–	5.9	–	–	–	–	–
	CUE	4	–	1.1	–	–	1.4	–	–	1.7	–	–	–	–	–
FOD & LEV/All	1step	18	6	–	5.3	5.4	–	9.9	8.8	–	77.3	58.4	0.888	0.913	0.887
	2step	18	6	18.4	5.3	5.3	27.6	8.5	7.6	93.2	77.7	61.5	0.885	0.906	0.887
	CUE	18	6	9.1	4.0	4.9	20.9	8.6	7.8	88.5	67.3	49.2	0.846	0.887	0.866
FOD & LEV/Sargan	1step	16	4	–	5.8	5.4	–	10.3	8.6	–	77.6	49.2	0.840	0.873	0.836
	2step	16	4	16.5	5.1	5.3	22.8	8.9	7.4	89.0	78.4	54.5	0.834	0.861	0.837
	CUE	16	4	6.3	4.5	4.1	15.3	7.5	6.7	84.6	67.3	42.1	0.785	0.837	0.809
FOD & LEV/Collapsed	2step	4	–	5.9	–	–	5.5	–	–	10.2	–	–	–	–	–
	CUE	4	–	3.1	–	–	2.0	–	–	3.1	–	–	–	–	–
$N = 250,\ \sigma^2_\eta = 1$															
FOD/All	1step	12	6	–	7.1	7.1	–	31.5	31.4	–	54.0	52.7	0.998	0.988	0.903
	2step	12	6	8.7	7.0	7.1	34.0	29.5	29.1	53.2	46.8	44.7	0.996	0.977	0.871
	CUE	12	6	3.1	2.7	2.7	10.3	12.9	12.4	31.0	38.1	36.6	0.994	0.955	0.826
FOD/Sargan	1step	12	6	–	7.1	7.0	–	31.5	31.5	–	54.0	52.6	0.998	0.989	0.903
	2step	12	6	8.7	7.0	7.0	34.0	29.5	29.5	53.2	46.8	45.0	0.997	0.978	0.866
	CUE	12	6	3.1	2.7	2.7	10.3	12.9	12.6	31.0	38.1	36.6	0.994	0.957	0.833
FOD/Collapsed	2step	4	–	6.7	–	–	10.3	–	–	24.8	–	–	–	–	–
	CUE	4	–	3.5	–	–	4.6	–	–	10.7	–	–	–	–	–
FOD & LEV/All	1step	18	6	–	4.8	4.6	–	24.6	22.1	–	98.9	88.7	0.885	0.945	0.916
	2step	18	6	11.2	4.9	5.1	35.3	28.2	26.7	99.7	99.8	95.8	0.882	0.946	0.921
	CUE	18	6	7.4	4.4	5.0	33.3	24.3	21.6	100.0	99.2	89.0	0.854	0.940	0.903
FOD & LEV/Sargan	1step	18	6	–	4.8	4.5	–	24.6	22.1	–	98.9	88.6	0.885	0.945	0.914
	2step	18	6	11.2	4.9	5.1	35.3	28.2	26.8	99.7	99.8	96.1	0.881	0.945	0.921
	CUE	18	6	7.4	4.4	4.9	33.3	24.3	21.8	100.0	99.2	88.9	0.853	0.939	0.903
FOD & LEV/Collapsed	2step	4	–	5.2	–	–	6.5	–	–	7.7	–	–	–	–	–
	CUE	4	–	3.8	–	–	5.9	–	–	7.2	–	–	–	–	–
$N = 250,\ \sigma^2_\eta = 5$															
FOD/All	1step	12	6	–	6.8	6.8	–	27.8	27.8	–	38.3	38.3	1.000	1.000	1.000
	2step	12	6	8.4	7.2	7.2	31.1	24.8	24.8	37.2	30.6	30.5	1.000	1.000	0.999
	CUE	12	6	2.7	2.7	2.7	7.8	9.0	9.0	21.1	22.5	22.5	1.000	1.000	0.999
FOD/Sargan	1step	12	6	–	6.8	6.8	–	27.8	27.8	–	38.3	38.3	1.000	0.999	0.999
	2step	12	6	8.4	7.2	7.2	31.1	24.8	24.8	37.2	30.6	30.6	1.000	0.999	0.999
	CUE	12	6	2.7	2.7	2.7	7.8	9.0	9.0	21.1	22.5	22.5	1.000	1.000	1.000
FOD/Collapsed	2step	4	–	6.7	–	–	8.2	–	–	13.4	–	–	–	–	–
	CUE	4	–	2.8	–	–	2.2	–	–	3.6	–	–	–	–	–
FOD & LEV/All	1step	18	6	–	7.5	7.0	–	25.5	23.1	–	98.6	84.5	0.892	0.935	0.894
	2step	18	6	16.6	7.0	6.9	36.2	26.7	24.2	100.0	99.3	89.1	0.883	0.932	0.898
	CUE	18	6	6.5	3.3	4.3	31.7	21.3	17.1	100.0	99.0	82.5	0.815	0.917	0.874
FOD & LEV/Sargan	1step	18	6	–	7.5	7.2	–	25.5	23.4	–	98.6	84.8	0.891	0.934	0.895
	2step	18	6	16.6	7.0	7.1	36.2	26.7	24.2	100.0	99.3	89.4	0.884	0.931	0.898
	CUE	18	6	6.5	3.3	4.4	31.7	21.3	17.4	100.0	99.0	83.0	0.816	0.918	0.874
FOD & LEV/Collapsed	2step	4	–	7.2	–	–	6.1	–	–	10.8	–	–	–	–	–
	CUE	4	–	3.8	–	–	2.9	–	–	3.8	–	–	–	–	–

Table 6: Simulation results with $T = 7$, $\alpha = 0.8$, $\beta = 1.0$

				Size			Power						Adjusted Rand Index		
				$\phi = 0$			$\phi = 0.2$			$\phi = 0.5$			ϕ		
Model/IV	Method	m	m^*	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	0	0.2	0.5
$N = 110,\ \sigma^2_\eta = 1$															
FOD/All	1step	42	12	–	10.5	10.4	–	33.3	33.4	–	89.5	88.8	0.960	0.952	0.893
	2step	42	12	73.0	10.3	10.2	90.8	31.4	31.5	99.0	73.9	72.0	0.950	0.934	0.835
	CUE	42	12	59.0	9.7	9.1	75.1	24.5	23.0	89.6	55.8	52.7	0.956	0.915	0.734
FOD/Sargan	1step	22	4	–	5.6	5.6	–	34.8	34.4	–	88.2	86.3	0.990	0.982	0.883
	2step	22	4	22.8	5.0	5.5	58.2	31.4	31.3	94.6	79.2	77.1	0.984	0.979	0.827
	CUE	22	4	12.8	4.2	4.2	25.1	15.9	15.2	39.3	33.0	32.6	0.981	0.930	0.608
FOD/Collapsed	2step	4	–	4.8	–	–	13.7	–	–	51.0	–	–	–	–	–
	CUE	4	–	2.3	–	–	5.8	–	–	19.0	–	–	–	–	–
FOD & LEV/All	1step	54	12	–	6.6	6.9	–	24.6	22.5	–	99.3	97.4	0.904	0.924	0.902
	2step	54	12	95.1	6.2	6.6	97.4	25.6	23.6	100.0	99.7	98.4	0.904	0.924	0.903
	CUE	54	12	91.4	9.7	9.2	96.9	30.1	26.4	100.0	99.6	97.1	0.879	0.913	0.888
FOD & LEV/Sargan	1step	34	4	–	5.5	4.8	–	21.7	18.3	–	98.7	79.5	0.839	0.873	0.849
	2step	34	4	61.8	4.9	4.6	76.0	24.7	20.8	99.9	99.3	88.6	0.837	0.871	0.853
	CUE	34	4	42.3	4.6	4.6	66.8	27.0	20.6	100.0	99.1	86.2	0.784	0.847	0.838
FOD & LEV/Collapsed	2step	4	–	4.5	–	–	5.3	–	–	9.1	–	–	–	–	–
	CUE	4	–	3.4	–	–	3.4	–	–	8.7	–	–	–	–	–
$N = 110,\ \sigma^2_\eta = 5$															
FOD/All	1step	42	12	–	9.7	9.6	–	33.0	33.0	–	86.7	86.7	0.991	0.988	0.972
	2step	42	12	72.4	10.4	10.4	90.2	30.9	30.5	98.4	66.2	66.4	0.988	0.983	0.958
	CUE	42	12	58.0	9.2	9.0	73.4	22.8	22.8	86.9	47.4	47.5	0.988	0.977	0.948
FOD/Sargan	1step	22	4	–	5.2	5.2	–	33.3	33.3	–	82.1	82.3	0.999	0.998	0.989
	2step	22	4	22.6	4.7	4.7	58.6	29.8	29.8	92.2	70.4	70.7	0.998	0.997	0.987
	CUE	22	4	11.5	3.4	3.4	21.9	12.4	12.5	28.0	20.2	20.1	0.997	0.994	0.972
FOD/Collapsed	2step	4	–	4.1	–	–	7.4	–	–	22.2	–	–	–	–	–
	CUE	4	–	1.1	–	–	2.3	–	–	5.8	–	–	–	–	–
FOD & LEV/All	1step	54	12	–	8.8	7.6	–	23.4	21.1	–	99.6	97.0	0.958	0.965	0.955
	2step	54	12	96.1	8.8	8.3	97.7	23.8	22.0	100.0	99.5	97.4	0.955	0.964	0.956
	CUE	54	12	90.5	7.8	8.0	96.8	25.9	23.2	100.0	98.6	94.2	0.927	0.951	0.945
FOD & LEV/Sargan	1step	34	4	–	6.6	5.2	–	23.3	19.4	–	99.4	87.2	0.853	0.873	0.850
	2step	34	4	71.8	6.4	5.6	76.8	23.2	19.8	99.9	99.3	90.1	0.852	0.870	0.852
	CUE	34	4	40.7	3.9	3.5	64.2	22.2	15.9	100.0	98.8	83.7	0.773	0.832	0.826
FOD & LEV/Collapsed	2step	4	–	6.4	–	–	3.9	–	–	9.1	–	–	–	–	–
	CUE	4	–	3.5	–	–	1.2	–	–	5.7	–	–	–	–	–
$N = 250,\ \sigma^2_\eta = 1$															
FOD/All	1step	42	12	–	8.6	8.7	–	69.1	69.1	–	99.3	99.3	0.972	0.962	0.892
	2step	42	12	34.4	9.5	9.3	87.6	69.8	69.3	99.7	94.9	94.4	0.971	0.951	0.834
	CUE	42	12	23.7	6.8	6.5	73.0	48.7	47.4	91.5	88.7	87.1	0.980	0.935	0.768
FOD/Sargan	1step	42	12	–	8.6	8.3	–	69.1	69.3	–	99.3	99.2	0.974	0.960	0.893
	2step	42	12	34.4	9.5	9.2	87.6	69.8	68.7	99.7	94.9	94.2	0.969	0.953	0.831
	CUE	42	12	23.7	6.8	6.5	73.0	48.7	46.8	91.5	88.7	87.9	0.981	0.938	0.763
FOD/Collapsed	2step	4	–	5.0	–	–	18.9	–	–	82.4	–	–	–	–	–
	CUE	4	–	4.4	–	–	11.2	–	–	41.2	–	–	–	–	–
FOD & LEV/All	1step	54	12	–	6.6	6.6	–	51.2	49.3	–	100.0	100.0	0.913	0.947	0.920
	2step	54	12	49.7	6.6	6.4	88.2	62.5	60.3	100.0	100.0	100.0	0.915	0.949	0.927
	CUE	54	12	40.9	5.8	6.4	88.7	63.0	59.7	100.0	100.0	100.0	0.886	0.944	0.916
FOD & LEV/Sargan	1step	54	12	–	6.6	6.5	–	51.2	49.7	–	100.0	100.0	0.915	0.948	0.923
	2step	54	12	49.7	6.6	6.7	88.2	62.5	60.1	100.0	100.0	100.0	0.913	0.947	0.930
	CUE	54	12	40.9	5.8	6.2	88.7	63.0	59.4	100.0	100.0	100.0	0.886	0.944	0.915
FOD & LEV/Collapsed	2step	4	–	5.1	–	–	5.6	–	–	4.1	–	–	–	–	–
	CUE	4	–	5.1	–	–	5.7	–	–	8.2	–	–	–	–	–
$N = 250,\ \sigma^2_\eta = 5$															
FOD/All	1step	42	12	–	8.8	8.7	–	69.7	69.6	–	98.5	98.5	0.995	0.993	0.968
	2step	42	12	34.1	9.4	9.5	87.1	70.4	70.0	99.4	90.5	91.0	0.992	0.987	0.955
	CUE	42	12	22.9	6.2	6.1	69.8	45.6	45.3	87.4	81.9	81.1	0.996	0.983	0.949
FOD/Sargan	1step	42	12	–	8.8	8.9	–	69.7	69.5	–	98.5	98.5	0.993	0.990	0.969
	2step	42	12	34.1	9.4	9.4	87.1	70.4	70.3	99.4	90.5	90.8	0.992	0.986	0.953
	CUE	42	12	22.9	6.2	6.2	69.8	45.6	45.5	87.4	81.9	81.7	0.996	0.983	0.951
FOD/Collapsed	2step	4	–	4.2	–	–	8.2	–	–	42.5	–	–	–	–	–
	CUE	4	–	3.4	–	–	3.1	–	–	9.2	–	–	–	–	–
FOD & LEV/All	1step	54	12	–	10.8	9.9	–	55.1	53.2	–	100.0	99.2	0.953	0.972	0.960
	2step	54	12	66.1	10.5	10.1	88.6	60.5	58.6	100.0	100.0	99.1	0.951	0.972	0.960
	CUE	54	12	41.5	5.9	5.8	87.1	55.6	49.8	100.0	100.0	99.8	0.897	0.963	0.950
FOD & LEV/Sargan	1step	54	12	–	10.8	9.8	–	55.1	53.3	–	100.0	99.2	0.953	0.972	0.959
	2step	54	12	66.1	10.5	10.1	88.6	60.5	58.7	100.0	100.0	99.1	0.951	0.972	0.960
	CUE	54	12	41.5	5.9	5.9	87.1	55.6	49.5	100.0	100.0	99.8	0.898	0.963	0.950
FOD & LEV/Collapsed	2step	4	–	6.9	–	–	6.3	–	–	7.7	–	–	–	–	–
	CUE	4	–	5.1	–	–	3.9	–	–	7.1	–	–	–	–	–

Table 7: Simulation results with $T = 10$, $\alpha = 0.8$, $\beta = 1.0$

				Size			Power						Adjusted Rand Index		
				$\phi = 0$			$\phi = 0.2$			$\phi = 0.5$			ϕ		
Model/IV	Method	m	m^*	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	J	J_{diag}	J^K_{diag}	0	0.2	0.5
$N = 110$, $\sigma^2_\eta = 1$															
FOD/All	1step	90	18	–	14.7	14.4	–	48.4	44.9	–	99.6	99.5	0.951	0.945	0.918
	2step	90	18	100.0	19.3	18.7	100.0	49.8	48.0	100.0	96.8	96.7	0.950	0.942	0.884
	CUE	90	18	100.0	35.8	34.6	100.0	56.6	55.5	100.0	92.9	92.0	0.938	0.916	0.787
FOD/Sargan	1step	34	4	–	5.6	6.0	–	42.6	42.9	–	99.6	99.6	0.966	0.959	0.919
	2step	34	4	46.0	5.3	5.4	87.6	39.6	39.7	100.0	98.8	98.3	0.968	0.963	0.892
	CUE	34	4	37.9	6.0	5.8	61.1	25.2	24.5	83.4	56.1	54.3	0.960	0.924	0.583
FOD/Collapsed	2step	4	–	6.0	–	–	24.0	–	–	81.5	–	–	–	–	–
	CUE	4	–	4.1	–	–	13.8	–	–	35.0	–	–	–	–	–
FOD & LEV/All	1step	108	18	–	8.7	7.0	–	35.7	29.1	–	100.0	99.8	0.872	0.894	0.873
	2step	108	18	100.0	33.1	30.3	100.0	54.0	46.4	100.0	100.0	99.5	0.869	0.884	0.865
	CUE	108	18	100.0	27.5	22.6	100.0	44.9	35.9	100.0	99.2	97.6	0.858	0.870	0.834
FOD & LEV/Sargan	1step	52	4	–	4.5	3.6	–	27.0	22.9	–	99.8	91.6	0.839	0.868	0.844
	2step	52	4	95.2	4.3	4.4	98.9	32.1	26.2	100.0	99.9	96.8	0.837	0.868	0.844
	CUE	52	4	87.8	5.3	5.8	98.5	39.8	30.4	100.0	99.9	95.1	0.768	0.844	0.831
FOD & LEV/Collapsed	2step	4	–	4.7	–	–	8.5	–	–	8.8	–	–	–	–	–
	CUE	4	–	4.0	–	–	7.5	–	–	13.0	–	–	–	–	–
$N = 110$, $\sigma^2_\eta = 5$															
FOD/All	1step	90	18	–	14.0	14.3	–	47.2	45.5	–	99.5	99.3	0.978	0.974	0.961
	2step	90	18	100.0	19.5	19.2	100.0	48.8	47.7	100.0	95.3	94.8	0.976	0.973	0.952
	CUE	90	18	100.0	34.7	34.9	100.0	57.5	57.1	100.0	91.7	91.7	0.974	0.967	0.941
FOD/Sargan	1step	34	4	–	5.3	5.3	–	41.1	40.9	–	99.4	99.4	0.995	0.991	0.978
	2step	34	4	46.3	4.7	5.0	86.2	37.8	38.0	99.9	97.4	97.5	0.993	0.992	0.975
	CUE	34	4	37.7	5.1	5.1	56.2	17.3	17.5	74.7	35.9	36.2	0.992	0.985	0.936
FOD/Collapsed	2step	4	–	4.6	–	–	9.1	–	–	44.4	–	–	–	–	–
	CUE	4	–	3.0	–	–	4.1	–	–	10.4	–	–	–	–	–
FOD & LEV/All	1step	108	18	–	10.5	10.7	–	35.3	32.7	–	100.0	99.9	0.965	0.967	0.960
	2step	108	18	100.0	34.4	33.1	100.0	52.2	48.7	100.0	100.0	98.8	0.960	0.962	0.956
	CUE	108	18	100.0	22.9	21.3	100.0	38.0	33.4	100.0	99.5	95.5	0.954	0.955	0.947
FOD & LEV/Sargan	1step	52	4	–	6.1	6.3	–	30.7	26.9	–	100.0	94.5	0.855	0.873	0.853
	2step	52	4	98.5	6.5	5.6	99.0	30.9	26.8	100.0	99.9	97.0	0.854	0.870	0.855
	CUE	52	4	89.0	4.7	5.2	97.7	32.2	23.8	100.0	99.6	93.4	0.762	0.830	0.832
FOD & LEV/Collapsed	2step	4	–	6.3	–	–	5.7	–	–	8.8	–	–	–	–	–
	CUE	4	–	4.2	–	–	4.2	–	–	9.7	–	–	–	–	–
$N = 250$, $\sigma^2_\eta = 1$															
FOD/All	1step	90	18	–	8.9	8.3	–	81.7	80.4	–	100.0	100.0	0.958	0.948	0.921
	2step	90	18	90.7	9.0	8.0	99.9	80.3	79.3	100.0	100.0	100.0	0.958	0.950	0.886
	CUE	90	18	86.9	7.8	7.6	99.8	71.2	69.6	100.0	99.8	99.6	0.966	0.947	0.718
FOD/Sargan	1step	78	12	–	6.3	6.1	–	82.3	81.6	–	100.0	100.0	0.970	0.963	0.936
	2step	78	12	78.6	6.8	6.0	99.5	80.8	79.0	100.0	100.0	100.0	0.970	0.962	0.902
	CUE	78	12	73.9	6.3	5.6	98.9	72.1	71.4	100.0	99.8	99.8	0.974	0.954	0.696
FOD/Collapsed	2step	4	–	4.4	–	–	42.1	–	–	96.7	–	–	–	–	–
	CUE	4	–	4.4	–	–	31.6	–	–	68.7	–	–	–	–	–
FOD & LEV/All	1step	108	18	–	5.5	5.3	–	68.9	65.0	–	100.0	100.0	0.877	0.914	0.886
	2step	108	18	98.6	5.1	4.8	100.0	80.6	76.1	100.0	100.0	100.0	0.880	0.913	0.889
	CUE	108	18	97.9	6.4	5.1	100.0	81.5	75.4	100.0	100.0	100.0	0.857	0.906	0.880
FOD & LEV/Sargan	1step	96	12	–	4.9	5.2	–	69.3	66.5	–	100.0	100.0	0.898	0.939	0.913
	2step	96	12	94.7	4.8	4.4	100.0	81.2	78.4	100.0	100.0	100.0	0.898	0.935	0.917
	CUE	96	12	93.6	5.0	4.7	99.8	82.0	77.5	100.0	100.0	100.0	0.868	0.927	0.908
FOD & LEV/Collapsed	2step	4	–	4.9	–	–	9.5	–	–	9.2	–	–	–	–	–
	CUE	4	–	4.3	–	–	9.9	–	–	19.3	–	–	–	–	–
$N = 250$, $\sigma^2_\eta = 5$															
FOD/All	1step	90	18	–	8.4	8.4	–	82.0	81.2	–	100.0	100.0	0.983	0.978	0.961
	2step	90	18	91.1	9.1	8.9	99.9	80.0	79.6	100.0	100.0	100.0	0.983	0.977	0.945
	CUE	90	18	86.9	8.0	7.4	99.7	69.0	68.3	100.0	99.8	99.8	0.988	0.976	0.931
FOD/Sargan	1step	78	12	–	6.2	5.9	–	81.4	80.8	–	100.0	100.0	0.990	0.986	0.970
	2step	78	12	78.6	6.6	6.5	99.6	80.6	80.2	100.0	100.0	100.0	0.990	0.987	0.960
	CUE	78	12	73.4	6.2	6.2	98.4	70.1	70.2	100.0	99.8	99.8	0.993	0.983	0.943
FOD/Collapsed	2step	4	–	4.2	–	–	14.3	–	–	71.3	–	–	–	–	–
	CUE	4	–	4.0	–	–	7.9	–	–	18.9	–	–	–	–	–
FOD & LEV/All	1step	108	18	–	11.4	11.1	–	71.8	70.4	–	100.0	100.0	0.964	0.970	0.962
	2step	108	18	99.7	11.4	10.5	100.0	75.9	74.1	100.0	100.0	100.0	0.961	0.970	0.964
	CUE	108	18	98.1	5.7	5.7	100.0	70.8	63.2	100.0	100.0	100.0	0.925	0.966	0.958
FOD & LEV/Sargan	1step	96	12	–	10.6	9.4	–	73.0	72.0	–	100.0	100.0	0.958	0.972	0.964
	2step	96	12	98.8	10.1	9.0	100.0	76.1	75.0	100.0	100.0	100.0	0.956	0.971	0.965
	CUE	96	12	93.5	3.9	4.5	99.7	70.7	62.3	100.0	100.0	100.0	0.905	0.963	0.957
FOD & LEV/Collapsed	2step	4	–	4.2	–	–	5.9	–	–	5.8	–	–	–	–	–
	CUE	4	–	3.5	–	–	5.8	–	–	9.1	–	–	–	–	–

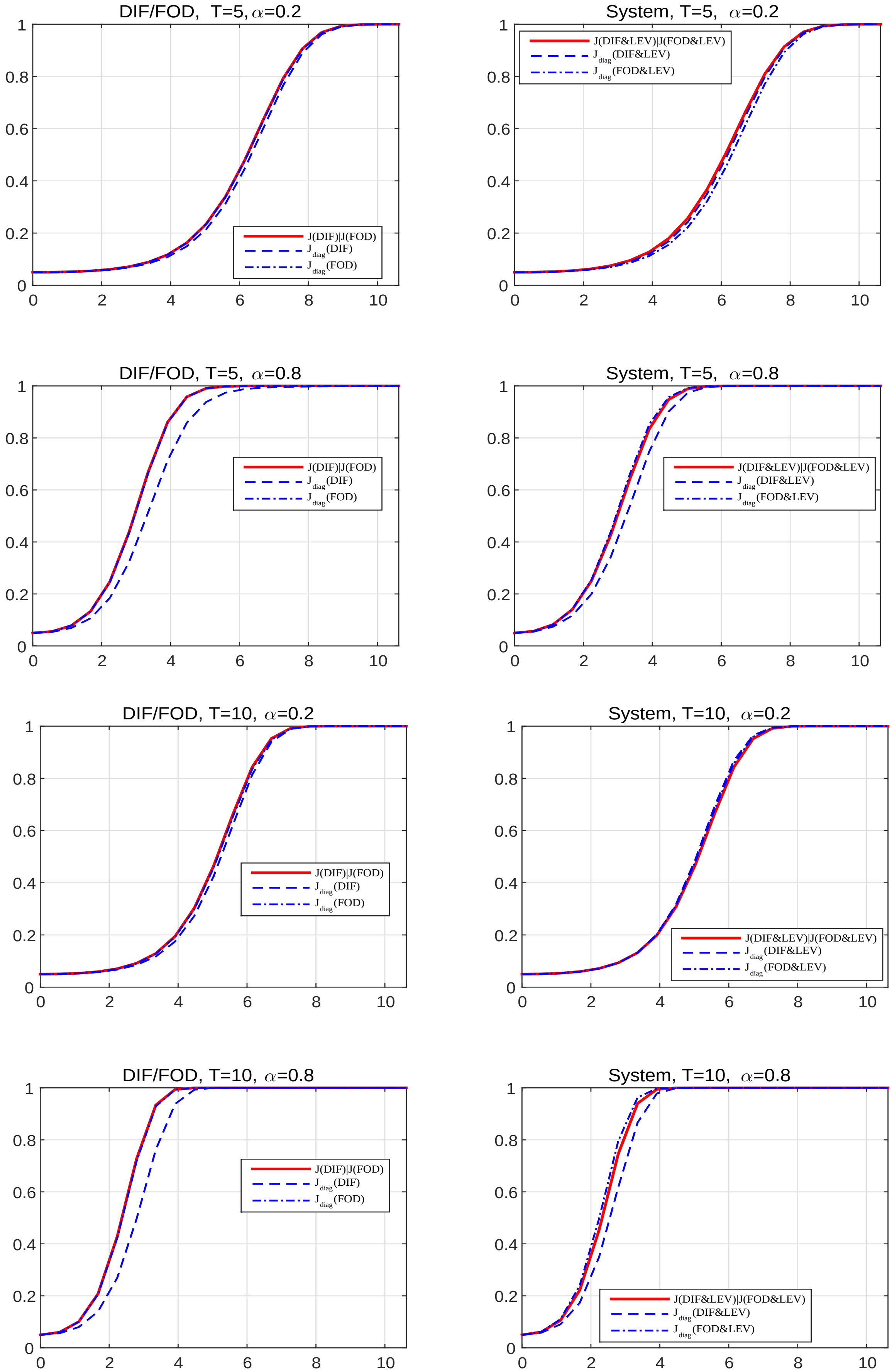

Figure 2: Local powers of the standard and diagonal J tests ($\sigma_t^2 = 0.5 + (t-1)/(T-1)$, $\sigma_\eta^2 = 1$)

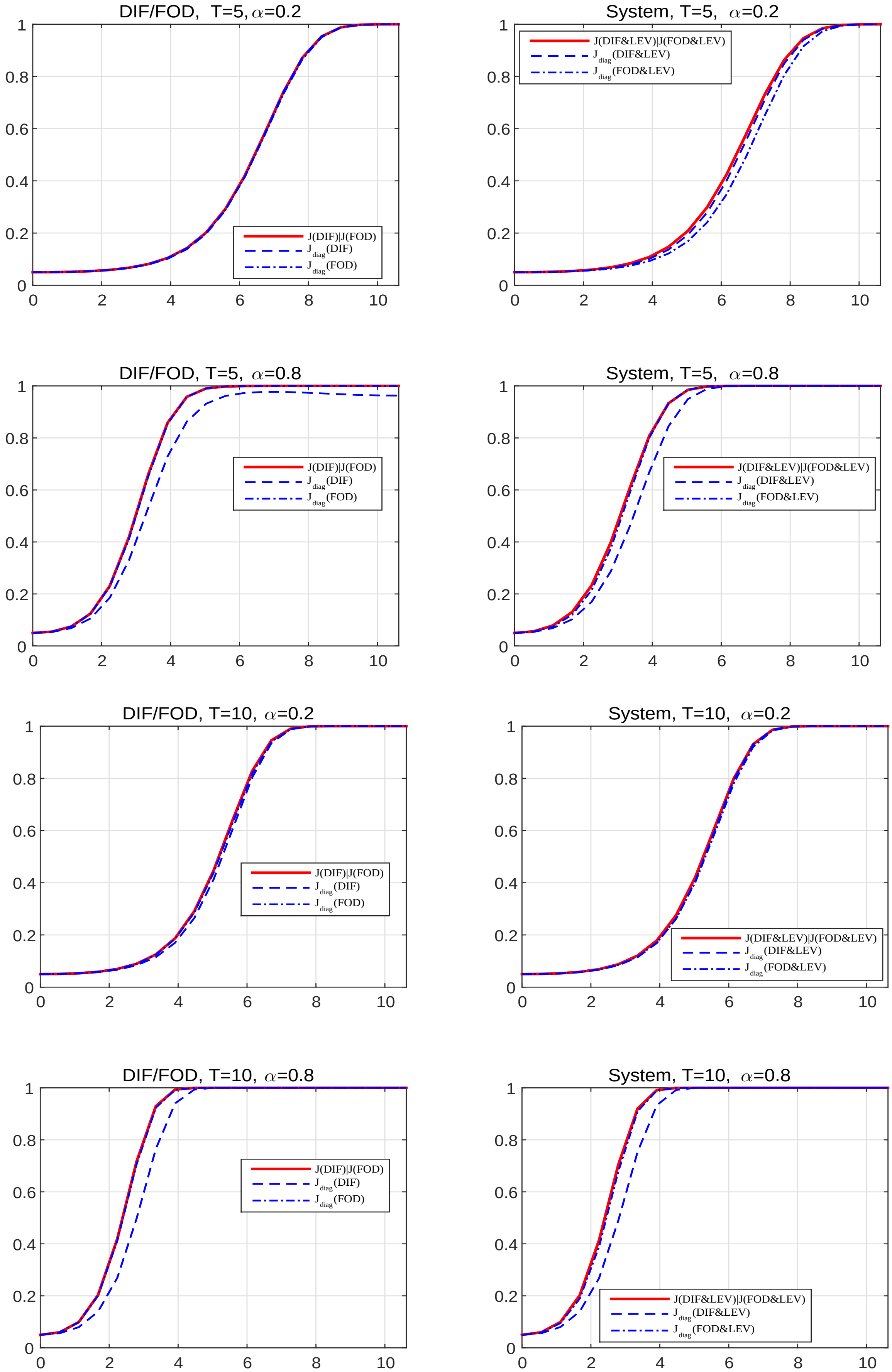

Figure 3: Local power of the standard and diagonal J tests ($\sigma_t^2 = 0.5 + (t-1)/(T-1)$, $\sigma_\eta^2 = 5$)

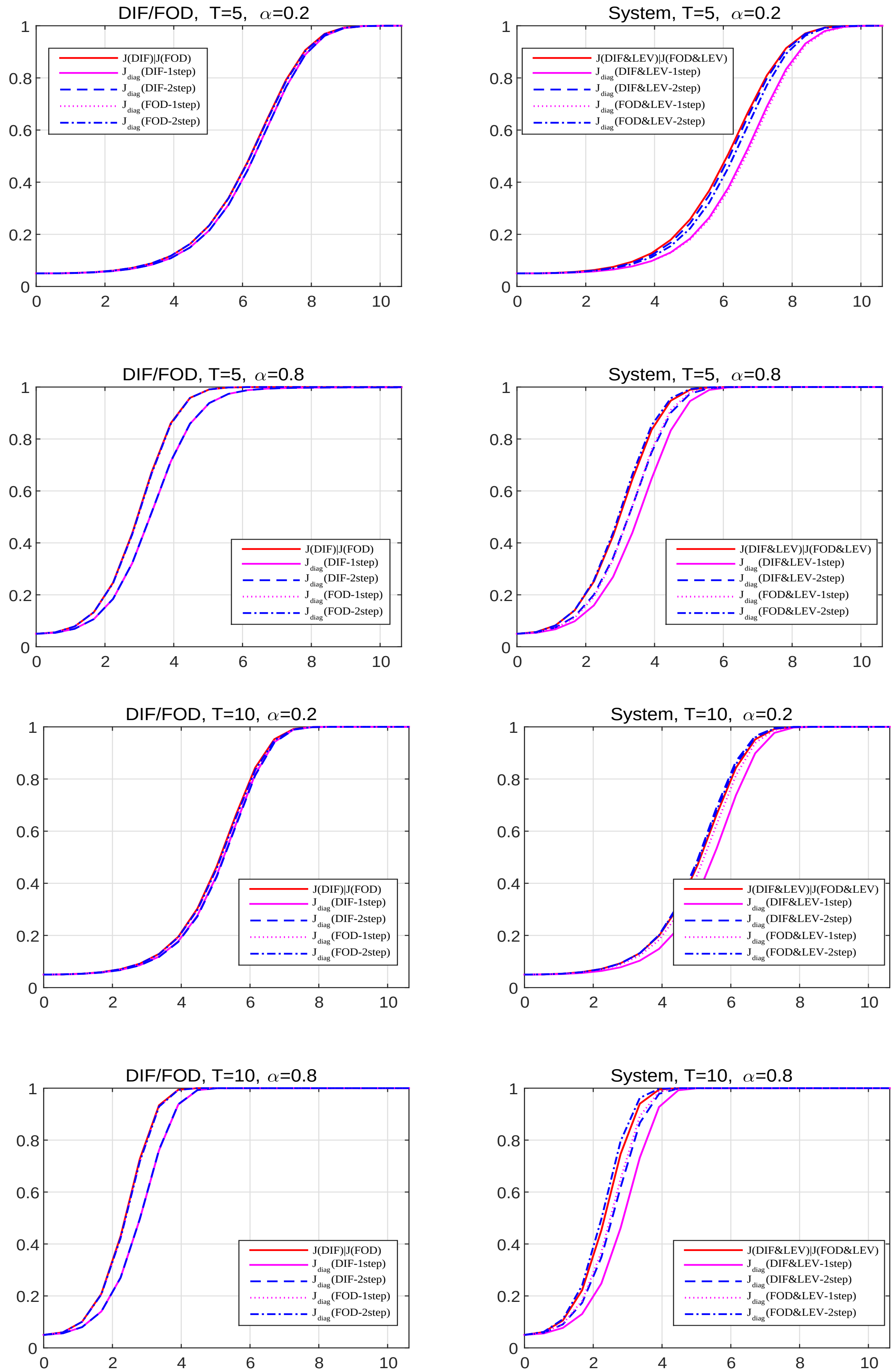

Figure 4: Local power of the diagonal J tests with one- and two-step estimators ($\sigma_t^2 = 0.5 + (t-1)/(T-1)$, $\sigma_\eta^2 = 1$)

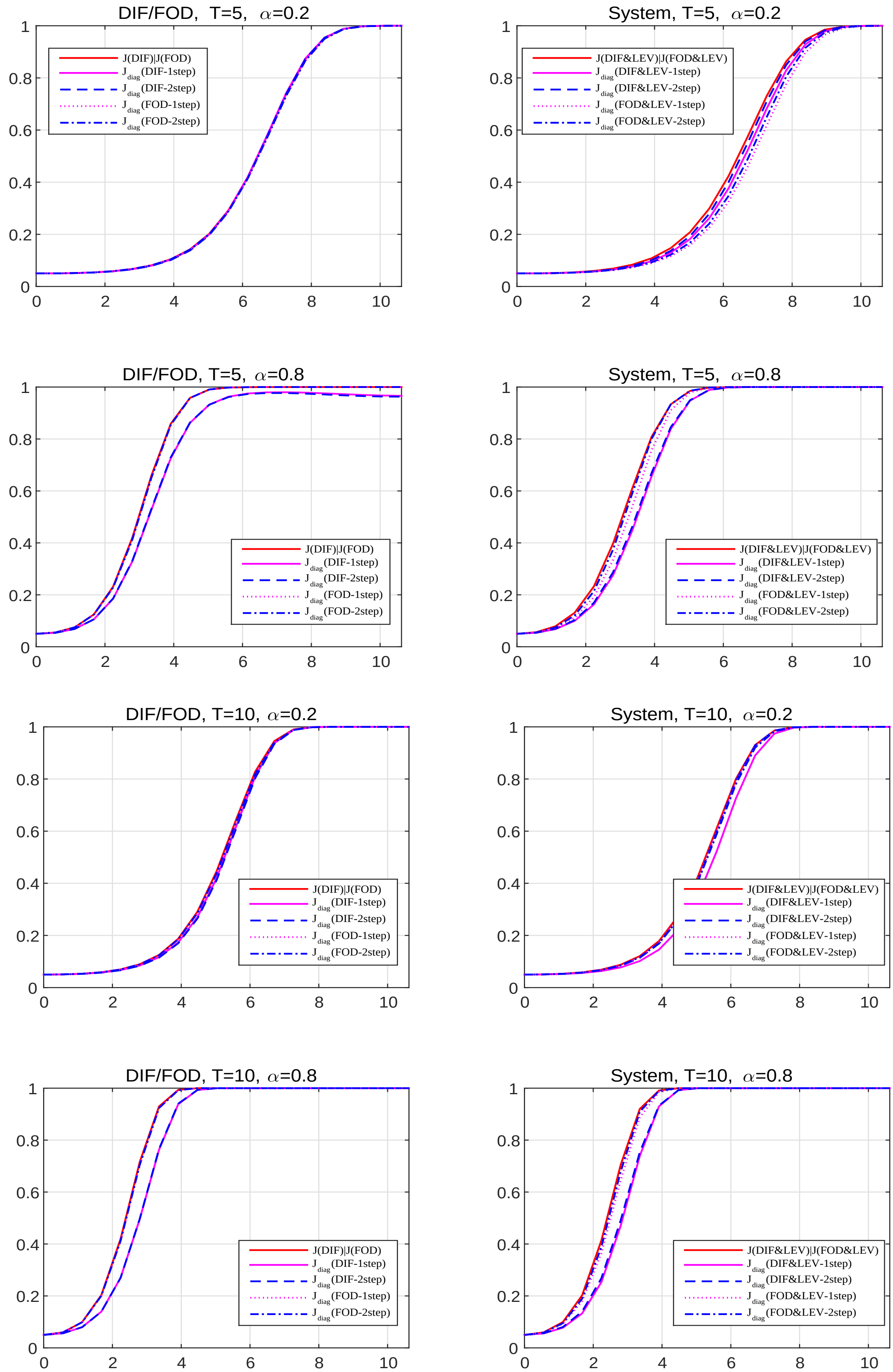

Figure 5: Local power of the diagonal J tests with one- and two-step estimators ($\sigma_t^2 = 0.5 + (t-1)/(T-1)$, $\sigma_\eta^2 = 5$)