A Specification Test for Instrumental Variables Regression with Many Instruments *

Yoonseok Lee† Ryo Okui‡

First draft: April 2009

This version: November 2009

Abstract

This paper considers specification testing for instrumental variables estimation in the presence of many instruments. The test proposed is a modified version of the Sargan (1958, *Econometrica* 26(3): 393-415) test of overidentifying restrictions. The test statistic asymptotically follows the standard normal distribution under the null hypothesis of correct specification when the number of instruments increases with the sample size. We find that the new test statistic is numerically equivalent up to a sign to the test statistic proposed by Hahn and Hausman (2002, *Econometrica* 70(1): 163-189). We also assess the size and power properties of the test.

Keywords and phrases: instrumental variables estimation; many instruments; overidentifying restrictions test; specification test.

JEL classification: C12; C21.

*The authors acknowledge the valuable comments of Jinyong Hahn, Han Hong, Simon Lee, Taisuke Otsu, Katsumi Simotsu, Mototsugu Shintani, Mehmet Caner, Juan Carlos Escanciano, and Yukitoshi Matsushita, along with participants at the Kobe meeting of the Kansai Econometric Society, the 2009 International Symposium on Econometric Theory and Applications (SETA09) held at Kyoto, and the 2009 Far East and South Asia Meeting of the Econometric Society held at Tokyo. The usual disclaimer applies.

†Department of Economics, University of Michigan, 611 Tappan Street, 365C Lorch Hall, Ann Arbor, MI 48109-1220, USA. Email: yoolee@umich.edu

‡Institute of Economic Research, Kyoto University, Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan. Email: okui@kier.kyoto-u.ac.jp

1 Introduction

It is often observed that conventional asymptotic theory provides a poor approximation of the finite sample distribution of instrumental variables estimators or test statistics for instrumental variables regression. Many studies document this problem in the presence of weak instruments (e.g., Staiger and Stock, 1997; Stock and Wright, 2000) and in the presence of many instruments (e.g., Kunitomo, 1980; Morimune, 1983; Bekker, 1994; Anatolyev and Gospodinov, 2009; van Hasselt, 2009). It is also well known that asymptotic approximation based on many (weak) instruments provides, in many situations, more accurate approximation.

This paper examines specification tests for the linear instrumental variables regression when the number of instruments increases with the sample size (e.g., Bekker, 1994) and the equation error could be nonnormal. The null hypothesis is the orthogonality between the instruments and the error term in the structural equation as for standard overidentifying restrictions tests. We particularly consider the Sargan test (Sargan, 1958) and derive its limiting behavior using many instrument asymptotics under which both the sample size and the number of instruments tend to infinity. Based on this result, a new test is constructed by modifying the Sargan test so that the asymptotic null distribution is standard normal under many instrument asymptotics. We also study the local power of the proposed test under many instrument asymptotics.

Most interestingly, we show that the proposed test is very much closely related with the test proposed by Hahn and Hausman (2002). Hahn and Hausman (2002) observe that under standard asymptotics (i.e., with a fixed number of instruments) an instrumental variables estimator has the same probability limit as the inverse of the instrumental variables estimator from the reverse regression. These two estimators, however, have different limits when standard asymptotic analysis is inadequate (e.g., the case with weak instruments). The Hahn–Hausman test is based on the difference between these estimators and can be used to check whether the standard asymptotic results are reliable. We show that the Hahn–Hausman test statistic is numerically equivalent up to a sign to the modified Sargan test developed in this paper, though these tests develop from very different motivations. Based on this equivalence result, it results that the null hypothesis of the Hahn–Hausman test corresponds to standard overidentifying restrictions tests, and therefore it is much easier to analyze the size and power properties of the Hahn–Hausman test. In particular, the power properties of the Hahn–Hausman test are not well known in the literature and our local power results should then facilitate the understanding of the behavior of the Hahn–Hausman test.

Our new interpretation of the Hahn–Hausman test is also useful for overcoming several limitations of the original Hahn–Hausman test and provides us with some guidelines on how to extend the Hahn–Hausman test to more general settings. For example, we can easily handle cases with multiple endogenous variables in our framework. While the Hahn–Hausman test involves the inverse of the estimate from the reverse regression it appears difficult to interpret the test when the regression parameter is zero. A zero parameter value should then not be a problem because the test is essentially an overidentifying restrictions test. Moreover, as the Sargan test is a special case of the J-test by Hansen (1982), it would be possible to extend our new test to a more general setup, such as moment condition-based nonlinear models, whereas the idea of using reverse regression in Hahn and Hausman (2002) appears to be limited to linear regression models and thus difficult to generalize further.

Several studies are closely related to the current discussion. For example, Kunitomo, Morimune and Tsukuda (1983) derive higher-order asymptotic approximations of the distribution of the overidentifying restrictions test but assume the number of instruments is fixed. Work by Andrews and Stock (2007) and Newey and Windmeijer (2009) also consider testing problems with many weak instruments. However, the number of instruments is restricted to grow much slower than the sample size. Independently of our work, Anatolyev and Gospodinov (2009) also consider the specification testing problem under many instrument asymptotics. They propose an alternative way to compute the critical values of the test so that the test has the correct size, even when the number of instruments is proportional to the sample size. While their work is closely related to the current analysis, our paper has a different scope and provides some novel results. In particular, we derive the equivalence (up to a sign) between the modified Sargan test and the Hahn–Hausman test; we obtain the local power of the overidentifying restriction test under many instrument asymptotics; and unlike Anatolyev and Gospodinov (2009), we consider the case where the fourth moment of the error term affects the asymptotic result.

The remainder of the paper is organized as follows. Section 2 describes the basic framework. Section 3 proposes new specification tests based on the Sargan test under many instrument asymptotics. Section 4 establishes the equivalence between our new test and the Hahn and Hausman (2002) test up to a sign. Some Monte Carlo simulation results are presented in Section 5. Section 6 concludes the paper with several important remarks. All of the mathematical proofs are provided in the Appendix.

2 Model

We consider a linear instrumental variables regression model given by

$$y_i = X_i'\beta + u_i \tag{1}$$

for $i = 1, 2, \cdots, n$, where y_i is the scalar outcome variable and X_i is the $r \times 1$ vector of regressors that is possibly correlated with an unobserved error u_i. We assume a $K \times 1$ vector of instruments, Z_i, which we treat as deterministic, where $r \leq K < n$. The results hold when Z_i is random, provided that all of the assumptions given below are stated conditional on $Z = (Z_1, \cdots, Z_n)'$. We also let $P = Z\left(Z'Z\right)^{-1}Z'$. Throughout the paper, we consider the asymptotic sequence under which both the sample size, n, and the number of instruments, K, tend to infinity with satisfying

$$\alpha_n \equiv \frac{K}{n} \to \alpha \quad \text{as } n, K \to \infty \tag{2}$$

for some $0 \leq \alpha < 1$. However, the number of endogenous regressors, r, is fixed and does not depend on n nor K. Note that we let $K \to \infty$ as $n \to \infty$ while we assume K to be smaller than n (i.e., $\alpha < 1$). However, we exclude the fixed K case; $\alpha = 0$ when K diverges at the slower rate than the rate of n. We further assume that

$$X_i = \Pi' Z_i + V_i, \tag{3}$$

where Π is the $K \times r$ matrix of nuisance parameters whose value may depend on n as well as K. For the independently and identically distributed vector of unobservables $\varepsilon_i = (u_i, V_i')'$, we define:

$$Var\left(\varepsilon_i\right) \equiv \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{Vu}' \\ \sigma_{Vu} & \Sigma_V \end{pmatrix} \tag{4}$$

conformably as $(u_i, V_i')'$, where $\sigma_{Vu} \neq 0$ in general so that X_i is correlated with u_i through the correlation between u_i and V_i. We make the following assumptions.

Assumption 1. ***(i)*** *$\alpha_n = \alpha + o(n^{-1/2})$ for some $0 \leq \alpha < 1$ as $n, K \to \infty$.* ***(ii)*** *Z and Π are of full column rank.* ***(iii)*** *$\varepsilon_i = (u_i, V_i')'$ are independently and identically distributed for $i = 1, 2, \cdots, n$, with mean zero and positive definite variance matrix Σ given in (4); the fourth moment of ε_i exists.* ***(iv)*** *$\Pi' Z' Z \Pi / n \to \Theta$ as $n, K \to \infty$, where Θ is positive definite and finite.* ***(v)*** *$\sup_{1 \leq i \leq n} |Z_i' \pi_j| < \infty$ for all $j = 1, \cdots, r$, where π_j is the jth column of Π.* ***(vi)*** *$\sup_{n \geq 1} \sup_{1 \leq j \leq n} \sum_{i=1}^n |P_{ij}| / \sqrt{\alpha_n} < \infty$, where P_{ij} is the (i,j)th element of P.* ***(vii)*** *$\sum_{i=1}^n (P_{ii}^2 - \alpha_n^2)/(n\alpha_n)$ converges as $n, K \to \infty$.*

Assumption 1 is similar to that imposed in van Hasselt (2009, Assumptions 1, 3 and 4). This assumption also implies the conditions for the central limit theorem of the quadratic forms in Kelejian and Prucha (2001).[1] This assumption guarantees that the modified Sargan test statistic in the next section has a well-defined asymptotic distribution under many instrument asymptotics. We note that the normality of the unobservables is not assumed here but we implicitly assume homoskedasticity. Given Z_i is nonrandom, the null hypothesis of instrument validity (i.e., $H_0 : \mathbb{E}(u_i Z_i) = 0$ for all i) holds automatically. Condition (iv) implies that the information accumulation by adding new instruments is limited and thus bounded even with $K \to \infty$. Note that this condition allows for moderately weak instruments though the full-rankness of Π in condition (ii) rules out underidentification. The following condition is employed to show the consistency of the asymptotic variance estimator of the modified Sargan test statistic.

Assumption 2. *X_i and u_i have finite eighth moments.*

3 Specification Tests with Many Instruments

Let $\hat{\beta}_{2sls}$ be the two-stage least squares (2SLS) estimator given by

$$\hat{\beta}_{2sls} = \left(X'PX\right)^{-1} X'Py,$$

where $X = (X_1, \cdots, X_n)'$ and $y = (y_1, \cdots, y_n)'$. The Sargan test statistic (Sargan, 1958; or the J test statistic, Hansen, 1982) is defined as

$$S_n(\hat{\beta}_{2sls}) = \frac{\hat{u}'P\hat{u}}{\hat{\sigma}_u^2}, \tag{5}$$

where $\hat{u} = y - X\hat{\beta}_{2sls}$ and $\hat{\sigma}_u^2 = \hat{u}'\hat{u}/n$. It is well known that under the null hypothesis $H_0 : \mathbb{E}(u_i Z_i) = 0$, standard asymptotic theory (i.e., when K is fixed) gives

$$S_n(\hat{\beta}_{2sls}) \to_d \chi^2_{K-r} \quad \text{as } n \to \infty. \tag{6}$$

Conventional (first-order) asymptotics, however, may not provide an accurate approximation of the finite sample distribution of $S_n(\hat{\beta}_{2sls})$, particularly when the number of instruments is large. In this

[1] Instead of condition (vi), we can assume a weaker condition, $\sup_{n\geq 1} \sup_{1\leq j\leq n} \sum_{i=1}^{n} |P_{ij} - \alpha_n \kappa_{ij}|/\sqrt{\alpha_n} < \infty$, which is implied by the conditions (i) and (vi), where $\kappa_{ij} = 1$ if $i = j$ and 0 otherwise. Note that this condition corresponds to Assumption 2 in Kelejian and Prucha (2001). To use the result by Kelejian and Prucha (2001, Theorem 1), however, we need a slightly stronger moment condition for u_i: $\sup_{n\geq 1} \sup_{1\leq i\leq n} \mathbb{E}|u_i|^{4+\eta} < \infty$ for some $\eta > 0$.

section, we instead consider higher-order approximation based on the many instrument asymptotics, which should provide more accurate finite sample results (e.g., Bekker, 1994). In particular, we develop specification tests similar to the Sargan test that are suitable under $n, K \to \infty$.

As $\hat{u} = (I - X(X'PX)^{-1}X'P)u$, where $u = (u_1, \ldots, u_n)'$ and I is the n-dimensional identity matrix, we can show that

$$\frac{\hat{u}'P\hat{u}}{n} = \frac{u'Pu}{n} - \frac{u'PX}{n}\left(\frac{X'PX}{n}\right)^{-1}\frac{X'Pu}{n} \to_p \alpha\sigma_u^2 - \alpha^2\sigma_{Vu}'(\Theta + \alpha\Sigma_V)^{-1}\sigma_{Vu} \tag{7}$$

as $n, K \to \infty$ under Assumption 1. Note that the probability limit (7) is simply zero when $\alpha = 0$. This is also the case of the standard asymptotics with $n \to \infty$ but when K is fixed. The probability limit (7) can be consistently estimated by

$$\begin{aligned}\hat{B} &= \alpha_n \left\{\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls})\right\} \\ &\quad - \left\{\frac{1}{n}(y - X\hat{\beta}_{b2sls})'PX\right\}\left(\frac{1}{n}X'PX\right)^{-1}\left\{\frac{1}{n}X'P(y - X\hat{\beta}_{b2sls})\right\},\end{aligned}$$

where

$$\hat{\beta}_{b2sls} = \left\{X'\left(P - \alpha_n I\right)X\right\}^{-1} X'\left(P - \alpha_n I\right)y \tag{8}$$

is the bias-corrected 2SLS estimator (e.g., Nagar, 1959; Donald and Newey, 2001; Hahn and Hausman, 2002). Using similar techniques as Bekker (1994), Kelejian and Prucha (2001), Hahn and Hausman (2002) and van Hasselt (2009), we can derive that

$$\sqrt{\frac{n}{\alpha_n}}\left(\frac{\hat{u}'P\hat{u}}{n} - \hat{B}\right) \to_d \mathcal{N}(0, w) \quad \text{as } n, K \to \infty \tag{9}$$

under Assumption 1, where

$$w = 2(1-\alpha)\sigma_u^4 + \left(\lim_{n,K\to\infty}\frac{1}{K}\sum_{i=1}^{n}P_{ii}^2 - \alpha^2\right)\left\{\mathbb{E}(u_i^4) - 3\sigma_u^4\right\}. \tag{10}$$

We note that in our setting the fourth moment of u_i possibly affects the asymptotic distribution (9); this is excluded from the analysis in Anatolyev and Gospodinov (2009). From (9), the t-test statistic is given by

$$T_{n,1} = \frac{\hat{d}_1}{\sqrt{\hat{w}}}, \tag{11}$$

where

$$\hat{d}_1 = \sqrt{\frac{n}{\alpha_n}}\left\{\frac{1}{n}(y - X\hat{\beta}_{2sls})'P(y - X\hat{\beta}_{2sls}) - \hat{B}\right\}$$

and

$$\hat{w} = 2(1-\alpha_n)\left\{\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls})\right\}^2 \qquad (12)$$
$$+ \left(\frac{1}{n}\sum_{i=1}^{n}\frac{P_{ii}^2 - \alpha_n^2}{\alpha_n}\right)\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - X_i\hat{\beta}_{b2sls})^4 - 3\left\{\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls})\right\}^2\right].$$

If we further assume that u_i is normally distributed and thus $\mathbb{E}(u_i^4) = 3\sigma_u^4$, then the asymptotic variance w can be simply estimated by

$$\tilde{w} = 2(1-\alpha_n)\left\{\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls})\right\}^2. \qquad (13)$$

All technical details are in the Appendix. We note that $T_{n,1}$ in (11) is nothing but a properly standardized quadratic form $\hat{u}'P\hat{u}$ and has a very similar structure as the standard Sargan test statistic (5). It can thus be considered a *modified Sargan test* statistic, where the modification is based on nonstandard (second-order) asymptotics with many instruments. The following theorem derives the asymptotic null distribution of the test statistic $T_{n,1}$.

Theorem 1. *If Assumptions 1 and 2 hold,* $T_{n,1} \to_d \mathcal{N}(0,1)$ *as* $n, K \to \infty$.

Under many instrument asymptotics, this result shows that the properly standardized quadratic form $\hat{u}'P\hat{u}$ follows an asymptotic normal distribution. Therefore, we can expect that the standard chi-square approximation in (6) performs poorly with Sargan's model specification test, particularly when K is large relative to n.[2]

Alternatively, we can consider the modified Sargan test based on the bias corrected 2SLS estimator, $\hat{\beta}_{b2sls}$. By construction, we have

$$\frac{1}{n}(y - X\hat{\beta}_{b2sls})'P(y - X\hat{\beta}_{b2sls}) = \frac{\hat{u}'P\hat{u}}{n} + \frac{1}{n}(y - X\hat{\beta}_{b2sls})'PX\left(X'PX\right)^{-1}X'P(y - X\hat{\beta}_{b2sls}) \to_p \alpha\sigma_u^2$$

as $n, K \to \infty$ (see the proof of Lemma 1 in the Appendix for the details), which implies that $(y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)(y - X\hat{\beta}_{b2sls})/n = \hat{u}'P\hat{u}/n - \hat{B}$ from the definition of $\hat{B}$. Therefore, similarly as (9), we can show that

$$\sqrt{\frac{n}{\alpha_n}}\left\{\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)(y - X\hat{\beta}_{b2sls})\right\} \to_d \mathcal{N}(0, w) \qquad (14)$$

[2]One problem may be that we do not know when we should use the chi-square approximation or a standard normal approximation. Similar results could be found in Donald, Imbens and Newey (2003) with moderately many instruments, or in Calhoun (2008) in the context of the F test in linear regressions with many regressors.

as $n, K \to \infty$, based on which we obtain the t-test statistic as

$$T_{n,2} = \frac{\hat{d}_2}{\sqrt{\hat{w}}}, \tag{15}$$

where

$$\hat{d}_2 = \sqrt{\frac{n}{\alpha_n}} \left\{ \frac{1}{n} (y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)(y - X\hat{\beta}_{b2sls}) \right\}.$$

The following lemma states that the test statistics, $T_{n,1}$ and $T_{n,2}$, are in fact numerically equivalent.

Lemma 1. *Under the linear specification (1) and (3), $T_{n,1} = T_{n,2}$.*

Because the 2SLS estimator is biased in the presence of many instruments, bias correction is necessary when constructing overidentifying restrictions test statistics. Lemma 1 demonstrates that for the 2SLS estimators and the overidentifying restrictions test statistics based on them, bias correction for the estimators is equivalent to bias correction for the test statistics in the linear instrumental variables regression. From Lemma 1 and Theorem 1, we can also conclude that $T_{n,2} \to_d \mathcal{N}(0,1)$ as $n, K \to \infty$ under Assumptions 1 and 2. The modified Sargan test could be constructed based on the limited information maximum likelihood (LIML) estimator[3]

$$\hat{\beta}_{liml} = \arg\min_{\beta} \frac{(y - X\beta)' P (y - X\beta)}{(y - X\beta)' (y - X\beta)},$$

whose asymptotic normality is obtained similarly.

Corollary 1. *If Assumptions 1 and 2 hold, $T_{n,2} \to_d \mathcal{N}(0,1)$ and,*

$$T_{n,3} = \frac{1}{\sqrt{\hat{w}_l}} \sqrt{\frac{n}{\alpha_n}} \left\{ \frac{1}{n} (y - X\hat{\beta}_{liml})'(P - \alpha_n I)(y - X\hat{\beta}_{liml}) \right\} \to_d \mathcal{N}(0,1)$$

as $n, K \to \infty$, where $\hat{w}_l$ is defined as (12) with replacing $\hat{\beta}_{b2sls}$ by $\hat{\beta}_{liml}$.

Note that the asymptotic variance of $\hat{\beta}_{liml}$ generally differs from that of $\hat{\beta}_{b2sls}$ under the second-order asymptotics (e.g., van Hasselt, 2009). However, this difference does not affect the asymptotic variance of the properly standardized quadratic form in the fitted residuals. Therefore, we can use the same formula (12) for $\hat{w}_l$ when constructing the specification test $T_{n,3}$.

[3] We note that it is difficult to extend the idea of Hahn and Hausman (2002) when the test is based on the LIML estimators. This is because the LIML estimator is the optimal linear combination of the bias-corrected forward 2SLS and reverse 2SLS estimators (Hahn and Hausman, 2002, p.169), and therefore the estimators become identical when we use LIML.

We finally discuss the power properties of the test by considering the following data generating process:

$$\begin{aligned} y_i &= X_i'\beta + Z_i'\gamma + u_i, \\ X_i &= \Pi' Z_i + V_i, \end{aligned}$$

where γ is a $K \times 1$ parameter vector and $\gamma = 0$ corresponds to the null hypothesis $H_0 : \mathbb{E}(u_i Z_i) = 0$. We let ξ be a $K \times 1$ vector, which does not depend on the sample size n, and consider the following Pitman-type local alternative:

$$H_a : \ \gamma = \frac{\alpha_n^{1/4}}{n^{1/4}} \xi. \tag{16}$$

We further assume the following conditions.

Assumption 3. *As $n, K \to \infty$, both $\xi' Z' Z \xi / n$ and $\xi' Z' X / n$ converge in probability; and $\xi' Z' u / \sqrt{n} = O_p(1)$.*

Note that Assumption 3 is satisfied when $Z_i'\xi$, X_i and u_i have finite fourth-order moments.

Theorem 2. *Suppose that Assumptions 1, 2 and 3 are satisfied. Then, under the local alternative (16), $T_{n,1} = T_{n,2} \to_d \mathcal{N}(C/\sqrt{w}, 1)$ as $n, K \to \infty$, where*

$$C = (1 - \alpha) \left\{ \left(\lim_{n,K\to\infty} \frac{1}{n} \xi' Z' Z \xi \right) - \left(\lim_{n,K\to\infty} \frac{1}{n} \xi' Z' X \right) \Theta^{-1} \left(\lim_{n,K\to\infty} \frac{1}{n} X' Z \xi \right) \right\}.$$

Theorem 2 shows that the modified Sargan test consistently detects the same set of alternatives the standard Sargan test detects. The modified Sargan test has a nontrivial power against local alternatives that contract to the null at the rate of $\alpha_n^{1/4} n^{-1/4}$. Note that this rate corresponds to $n^{-1/2}$ when K is fixed. When K is proportional to the sample size (i.e., α_n converges to a nonzero constant), the rate is $n^{-1/4}$. This result illustrates the difficulty of detecting a violation of the orthogonality condition in the presence of many instruments.

When $C = 0$, the test cannot detect this type of local alternative (16). A leading example of such alternatives is the case of $\gamma = \Pi$ when the dimension of X is one. This inconsistency of overidentifying restrictions tests is also observed when K is fixed, as discussed in Newey (1985). Thus, the test cannot detect local alternatives with $C = 0$, regardless of whether K is fixed or increases with n.

4 Comparison with the Hahn–Hausman Test

In this section, we show that the modified Sargan test statistics, $T_{n,1}$ and $T_{n,2}$, are numerically equivalent to the test statistics suggested by Hahn and Hausman (2002) up to a sign. Here, we consider the case where the dimension of X_i is one and the error term u_i is normally distributed. In the Appendix B, we show that even when there are two endogenous variables, the equivalence result (up to a sign) remains to hold. We, in fact, expect that the equivalence result holds in general, irrespective of the number of endogenous regressors and of the error distribution. Note that Hahn and Hausman (2002) do not discuss cases where there are more than three endogenous regressors.

The Hahn–Hausman test is based on the difference between the instrumental variables estimator of β and the inverse of the estimator that uses the same set of instruments, but where the roles of the dependent variable and the regressor are reversed. The basic idea is that the 2SLS estimator of X on y (i.e., the reverse regression) using Z as instruments is asymptotically equivalent to $1/\hat{\beta}_{2sls}$ when the standard first-order asymptotics (with K fixed) is adequate. However, when the conventional asymptotics do not provide a good approximation (e.g., $K \to \infty$), these estimators converge to two different limits.

To avoid the problem of the bias in the difference, we consider the bias-corrected 2SLS estimator of β in (8). We define the difference as

$$\frac{X'(P-\alpha_n I)y}{X'(P-\alpha_n I)X} - \frac{y'(P-\alpha_n I)y}{X'(P-\alpha_n I)y}. \tag{17}$$

Given $(y - X\hat{\beta}_{b2sls})'(P-\alpha_n I)X = 0$ by construction, however, the difference (17) can be rewritten as

$$\begin{aligned}\frac{X'(P-\alpha_n I)y}{X'(P-\alpha_n I)X} - \frac{y'(P-\alpha_n I)y}{X'(P-\alpha_n I)y} &= \hat{\beta}_{b2sls} - \frac{(y - X\hat{\beta}_{b2sls} + X\hat{\beta}_{b2sls})'(P-\alpha_n I)y}{X'(P-\alpha_n I)y} \\ &= -\frac{(y - X\hat{\beta}_{b2sls})'(P-\alpha_n I)y}{X'(P-\alpha_n I)y} \\ &= -\frac{(y - X\hat{\beta}_{b2sls})'(P-\alpha_n I)(y - X\hat{\beta}_{b2sls})}{X'(P-\alpha_n I)y}. \end{aligned} \tag{18}$$

Note that $(y - X\hat{\beta}_{b2sls})'P(y - X\hat{\beta}_{b2sls})$ is a quadratic form of the sample covariance between the regression residual and the instruments upon which Sargan's overidentifying restrictions test (5) is based. As shown in the previous section, we can see that $\alpha_n(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls})$ in the numerator of (18) is the term that demeans (or corrects the bias of) the Sargan statistic when K is proportional to n. This basic result shows that both the Hahn–Hausman test and the Sargan test are indeed based on the asymptotic behavior of the same quantity.

More precisely, the Hahn–Hausman test statistic based on the bias-corrected 2SLS estimator is defined as[4]

$$
\begin{aligned}
m_2 &= \left[\frac{2K}{n-K}\cdot\frac{\left\{(y-X\hat{\beta}_{b2sls})'(y-X\hat{\beta}_{b2sls})\right\}^2}{\hat{\beta}_{b2sls}^2\left\{X'PX-\frac{K}{n-K}X'\left(I-P\right)X\right\}^2}\right]^{-1/2} \\
&\quad\times\sqrt{n}\left\{\frac{X'(P-\alpha_n I)y}{X'(P-\alpha_n I)X}-\frac{y'(P-\alpha_n I)y}{X'(P-\alpha_n I)y}\right\} \\
&= \left[2\left(1-\alpha_n\right)\frac{\left\{(y-X\hat{\beta}_{b2sls})'(y-X\hat{\beta}_{b2sls})\right\}^2}{\hat{\beta}_{b2sls}^2\left\{X'(P-\alpha_n I)X\right\}^2}\right]^{-1/2} \qquad (19) \\
&\quad\times\sqrt{\frac{n}{\alpha_n}}\left\{\frac{X'(P-\alpha_n I)y}{X'(P-\alpha_n I)X}-\frac{y'(P-\alpha_n I)y}{X'(P-\alpha_n I)y}\right\}
\end{aligned}
$$

under normality, where the term in the square root is the standard error of the difference. In comparison, the modified Sargan test $T_{n,2}$ in (15) with $\tilde{w}$ in (13), which reflects the many instrument asymptotics, is given by

$$
\begin{aligned}
T_{n,2} &= \frac{1}{\sqrt{2\alpha_n\left(1-\alpha_n\right)n}}\left\{S_n(\hat{\beta}_{b2sls})-K\right\} \\
&= \left[2\left(1-\alpha_n\right)\left\{(y-X\hat{\beta}_{b2sls})'(y-X\hat{\beta}_{b2sls})\right\}^2\right]^{-1/2} \qquad (20) \\
&\quad\times\sqrt{\frac{n}{\alpha_n}}\left(y-X\hat{\beta}_{b2sls}\right)'\left(P-\alpha_n I\right)\left(y-X\hat{\beta}_{b2sls}\right),
\end{aligned}
$$

where $S_n(\hat{\beta}_{b2sls})=\{(y-X\hat{\beta}_{b2sls})'P(y-X\hat{\beta}_{b2sls})\}/\{(y-X\hat{\beta}_{b2sls})'(y-X\hat{\beta}_{b2sls})/n\}$ is the standard Sargan statistic based on the bias-corrected 2SLS estimator. The following theorem shows that the m_2 test of Hahn and Hausman (2002) is equivalent to the test based on $T_{n,2}$ in (20) up to a sign.

Theorem 3. $m_2 = T_{n,2}\cdot\mathrm{sgn}\left[-X'(P-\alpha_n I)y\right]$, *where* $\mathrm{sgn}[\cdot]$ *gives the sign of its argument.*

This result shows that the test of Hahn and Hausman (2002) can be regarded as a modification of Sargan's overidentifying restrictions test reflecting many instrument asymptotics (up to the scalar multiplication of $\mathrm{sgn}\left[-X'(P-\alpha_n I)y\right]$). Though omitted in this paper, the equivalence result could be also derived without the normality assumption if we use $\hat{w}$ in (12) and the variance expression in Theorem 4-4 of Hahn and Hausman (2002).

[4] In addition to m_2, Hahn and Hausman (2002) also discuss a different statistic m_1 based on the 2SLS estimator. Theorem 4-3 of Hahn and Hausman (2002) shows that, however, m_1 and m_2 are asymptotically equivalent. Therefore, in this discussion we focus our attention on m_2.

Several interesting implications are in order from Theorem 3. Hahn and Hausman (2002) document good finite sample properties of their tests as compared to the Sargan test. Theorem 3 shows that, however, this good performance comes from the many instrument asymptotics providing better (finite sample) approximation, not that the test statistics are fundamentally different from the standard overidentifying restrictions tests, including Sargan's. Another implication is that the possibility of the coefficient being zero in the reverse regression is no longer a problem when using the Hahn–Hausman test once we reformulate it as the modified Sargan test. For example, in Theorem 4-2 in Hahn and Hausman (2002), the coefficient β appears in the denominator of the asymptotic variance and it will make the test difficult to interpret when the coefficient is indeed zero though the test statistic itself is well-defined. Moreover, because of the equivalence, the results obtained in the previous section apply to the Hahn–Hausman test. In particular, the power properties of the Hahn–Hausman test (Theorem 2) are new to the literature.

It is important to note that the Hahn–Hausman test m_2 is two-sided because we do not know, a priori, whether a violation of the null hypothesis implies a large negative value of the test statistics or a large positive value of it. On the other hand, the modified Sargan test $T_{n,2}$ is one-sided. Therefore, we can achieve a higher power by using the new test statistic $T_{n,2}$ and making the test one-sided because we know that a violation of the null implies a large positive value of the test statistic $T_{n,2}$. For example, the Hahn–Hausman test rejects the null hypothesis at the 5% level when $|m_2| > 1.96$. From Theorem 3, however, it is equivalent to $|T_{n,2}| > 1.96$ since $|sgn[-X'(P - \alpha_n I)y]| = 1$ unless $X'(P - \alpha_n I)y = 0$. Note that the probability of $X'(P - \alpha_n I)y = 0$ is typically zero (even when $\beta = 0$ and n, K are large) because $X'(P - \alpha_n I)y$ is random.[5] One the other hand, the modified Sargan test rejects the null hypothesis at the 5% level when $T_{n,2} > 1.68$ since it is one-sided test. This difference makes the power properties of these tests different.

5 Monte Carlo Simulations

In this section, we employ Monte Carlo experiment results to consider the size and power performance of the modified Sargan test proposed in this paper. We note that the simulation results reported in Hahn and Hausman (2002) are also useful for this purpose because of the equivalence result in Theorem 3. In addition to the findings of Hahn and Hausman (2002), however, we examine

[5] When $\beta = 0$, it is easy to see that $X'(P - \alpha_n I)y/n \to_p -\alpha\sigma_{Vu}$, which is normally assumed not zero, and $X'(P - \alpha_n I)y/\sqrt{n} = O_p(1)$ as $n, K \to \infty$ under Assumption 1. (See Lemma A.2) So even when n, K are large, $X'(P - \alpha_n I)y = 0$ is very unlikely. One remark is that, under this case, $sgn[-X'(P - \alpha_n I)y]$ reflects the sign of σ_{Vu}.

the powers of the tests when a subset of the instruments violate orthogonality. The Monte Carlo experiments are conducted using Ox 5.10 (Doornik, 2007) for Linux.

The design of the experiment is similar to that considered by Hahn and Hausman (2002). The data generating process is given by

$$y_i = X_i\beta + Z_i'\gamma + u_i \quad \text{and} \quad X_i = \Pi' Z_i + V_i$$

for $i = 1, \ldots, n$, where $X_i \in \mathbb{R}$ and $Z_i \in \mathbb{R}^K$. We let $\beta = 0$ since the new test statistic is exactly invariant to the value of β; we define the parameters Π and γ later.[6] We consider two different distributions of the instruments and the error term:

[D-I] $Z_i \sim \mathcal{N}(0, I_K)$ and

$$\begin{pmatrix} u_i \\ V_i \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right),$$

where ρ is the parameter representing the degree of endogeneity;

[D-II] $Z_i = \eta_i\sqrt{3/5}$, where η_i is a $K \times 1$ vector of independent t-distributed random variables with 5 degrees of freedom and

$$\begin{pmatrix} u_i \\ V_i \end{pmatrix} \sim \zeta_i\sqrt{3/5} \times \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right),$$

where ζ_i is distributed as a t-distribution with 5 degrees of freedom.

Note that [D-I] is the same specification as Hahn and Hausman (2002); in [D-II], the error distribution is symmetric but has heavy tails. We also note that the variances of an element of Z_i, u_i and V_i are 1 for both cases. We consider $n = 250, 1000$ for the sample size, $K = 5, 10, 30$ for the number of instruments, and $\rho = 0, 0.5, 0.9$ for the degree of endogeneity, $R_f^2 = 0.01, 0.2$ for the theoretical R^2 of the first-stage regression.[7] $R_f^2 = 0.01$ reflects relatively weak instruments whereas $R_f^2 = 0.2$ reflects

[6] Even with finite samples, the simulation results does not vary much over the different values of β. More simulation results for different (nonzero) values of β (e.g., $\beta > 0$ or $\beta < 0$) is available upon request to the authors.

[7] Hahn and Hausman (2002) consider $\rho = -0.9, -0.5, 0.5, 0.9$. However, the results with $\rho = -0.9$ and -0.5 are very similar to those with $\rho = 0.9$ and 0.5 respectively, so are omitted. For the distributions of Z_i and $(u_i, V_i)'$, we also consider the case with $Z_i \sim \mathcal{N}(0, I_K)$ and and the error term is multivariate lognormal: $\binom{u_i}{V_i} = [(e-1)e]^{-1/2}\binom{e^{v_{1i}} - e^{1/2}}{e^{v_{2i}} - e^{1/2}}$ for $\binom{v_{1i}}{v_{2i}} \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$. Note that the lognormal error distribution is skewed and has heavy tails. However, the simulation result in this case still remains almost the same as the other designs.

relatively strong instruments. There are two specifications of the coefficients $\gamma = (\gamma_1, \cdots, \gamma_K)'$ and $\Pi = (\pi_1, \cdots, \pi_K)'$. In Model $\mathcal{M}_0$, which imposes the null hypothesis $H_0 : \mathbb{E}(u_i Z_i) = 0$, we let

$$\mathcal{M}_0 : \quad \gamma_k = 0 \text{ and } \pi_k = c(K) \text{ for all } k,$$

where $c(K)$ is chosen so that $R_f^2 = \Pi'\Pi/(\Pi'\Pi + 1)$ becomes the assigned value. $\gamma = 0$ implies that the instruments are exogenous, which is the null hypothesis of the overidentifying restriction tests. In Model $\mathcal{M}_1$, we let

$$\mathcal{M}_1 : \quad \gamma_1 = 0.1 \text{ and } \gamma_k = 0 \text{ for all } k \neq 1; \, \pi_k = c(K) \text{ for all } k.$$

This specification corresponds to an alternative hypothesis because the first instrument is not valid.

We compare the following eight tests: "`Sargan`" (Sargan test based on $\hat{\beta}_{2sls}$); "`SB`" (Sargan test based on $\hat{\beta}_{b2sls}$); "`SL`" (Sargan test based on $\hat{\beta}_{liml}$); "`HH`" (the Hahn-Hausman test); "`MSn`" (the modified Sargan test based on $\hat{\beta}_{b2sls}$ assuming normality, which is equivalent up to a sign to the Hahn–Hausman test); "`MSnL`" (the modified Sargan test based on $\hat{\beta}_{liml}$ assuming normality); "`MSnn`" (the modified Sargan test based on $\hat{\beta}_{b2sls}$ without assuming normality, which is equivalent up to a sign to the nonnormal version of the Hahn–Hausman test); and "`MSnnL`" (the modified Sargan test based on $\hat{\beta}_{liml}$ without assuming normality). We note that the difference between `HH` and `MSn` is that `HH` is a two-sided test while `MSn` is one-sided. The nominal size of the tests is 5%. For the first three Sargan tests, the critical value is obtained from the χ^2_{K-1} distribution; for the Hahn–Hausman and the latter four modified Sargan tests, the critical value is obtained from the standard normal distribution. For Model $\mathcal{M}_0$, we compute the empirical size of each test; for Model $\mathcal{M}_1$, we compute the size-adjusted rejection probabilities. The number of replications is 1000. The Monte Carlo results are summarized in Tables 1-4. Tables 1 and 3 report the actual sizes of the tests (in Model $\mathcal{M}_0$) and Tables 2 and 4 report the size-adjusted powers (in Model $\mathcal{M}_1$).

[Tables 1-4 about here.]

For Model $\mathcal{M}_0$, we obtain results similar to those in Hahn and Hausman (2002). The results show that the size of `Sargan` can be heavily distorted. The size distortion is particularly severe when R_f^2 is small, ρ is large and/or K is large. Although `SB` performs better than `Sargan` in terms of size, the size of `SB` also deviates from the nominal size when $R_f^2 = 0.01$ and $\rho = 0.9$. `HH` tends to be conservative, but can exhibit size distortion when $\rho = 0.9$. The magnitude of the size distortion of those tests is larger than that observed in Hahn and Hausman (2002) when $\rho = 0.9$ but smaller when $\rho = 0,\ 0.5$. This may be due to the fact that Hahn and Hausman (2002) use the LIML estimator to compute the standard error while we use the bias-corrected 2SLS estimator. The size properties of `MSn` and `MSnn` are similar to that of `SB`. We note that the sizes of `MSn` and `MSnn` are very similar, even under the nonnormal designs (Distribution [D-II]). The performances of `SL`, `MSnL` and `MSnnL` are similar to each other, even under the nonnormal designs. They do not exhibit size distortion, instead they tend to be conservative.

Next, we consider the size-adjusted powers of the tests in Model $\mathcal{M}_1$. When $R_f^2 = 0.2$ and $K = 5, 10$, all of the tests have similar power. The power of `HH` is worse than those of other tests when $R_f^2 = 0.02$, $n = 1000$ and $K = 30$ while other tests are equally powerful. This result is due to the fact that `HH` is two-sided while other tests are one-sided. The power properties of those one-sided tests are different between the tests using the bias-corrected 2SLS (`SB`, `MSn`, `MSnn`) and the LIML estimator (`SL`, `MSnL`, `MSnnL`) when $R_f^2 = 0.01$, however. We note that `SB` and `MSn` must have similar size-adjusted powers because `MSn` is obtained by a linear transformation of `SB`. The same comment applies to the relationship between `SL` and `MSnL`. On the other hand, it is notable that the powers of `MSn` and `MSnn` are similar (and so are the powers of `MSnL` and `MSnnL`), even under the nonnormal designs. We note that the power of `Sargan` is higher than the other tests. `SB`, `MSn` and `MSnn` are more powerful than `SL`, `MSnL` and `MSnnL`. Lastly, we observe that the powers of the tests decrease as K increases. This finding is consistent with the theoretical result in Section 3.

We may summarize the lessons from the Monte Carlo simulations in the following way. The standard Sargan test should be used with caution because it may suffer from a severe size distortion particularly when the number of instruments is moderately large comparing with the sample size. It is then advisable to use a test based on a bias-corrected estimator or a modified Sargan test via many instrument asymptotics. Moreover, it is better to make the test one-sided to improve the power. Tests based on the LIML estimator may be conservative. Thus, it is safe to use LIML-based tests because we can avoid the size distortion. However, these tests may have relatively weak power. Correcting the test statistics for possible nonnormality may then not be that crucial.

6 Discussion

This paper develops a new specification test for instrumental variables regression. To this end, we examine the asymptotic distribution of the Sargan test statistic when the number of instruments increases with the sample size and modify it such that it asymptotically follows a standard normal distribution under the null hypothesis of correct specification.

We also show that the new test statistic is numerically equivalent to the test statistic developed by Hahn and Hausman (2002) up to a sign. This implies that the Hahn–Hausman test is in fact a test for overidentifying restrictions, properly adjusted to reflect many instruments or to obtain better finite sample approximation. Our equivalence result is useful when we consider the extension of the Hahn–Hausman test to more general settings. For example, in our framework we can easily handle cases with multiple endogenous variables. Note that the test statistic with two endogenous regressors in Hahn and Hausman (2002, Section 5) is very complicated. Furthermore, as the Sargan test is a special case of the J-test by Hansen (1982), we could consider extensions to more general nonlinear moment restriction models and develop a modified J-test in the presence of many moment conditions, whereas it is difficult to extend the use of reverse regression equations to such general cases. We note that Newey and Windmeijer (2009, Theorem 5) provide an asymptotic result for the J-test under many weak moments asymptotics though they restrict the number of instruments to grow much slower than the sample size in order to achieve a standard chi-square asymptotic distribution.

It is also interesting to consider the properties of the Sargan test under alternative asymptotic sequences. For instance, Hausman, Stock and Yogo (2005) examine the performance of the Hahn–Hausman test in the presence of weak instruments. They find that the Hahn–Hausman test does not have a strong power in detecting the presence of weak or irrelevant instruments. This result also applies to our test because of the equivalence result. On the other hand, this finding is natural from our point of view: the Hahn–Hausman test statistic is numerically equivalent (up to a sign) to the overidentifying restrictions test statistic as the latter does not examine the strength of the instruments. Note that as we assume the concentration parameter grows at the same rate as the sample size (Assumption 1(iv)), the set of instruments in our case is stronger than what the many weak instrument asymptotics literature normally assumes (e.g., Chao and Swanson, 2005; Han and Phillips, 2006; Andrews and Stock, 2007; Hansen, Hausman and Newey, 2008; Newey and Windmeijer, 2009).

A Appendix: Mathematical Proofs

A.1 Proof of Lemma 1

Given the denominators of $T_{n,1}$ and $T_{n,2}$ are the same, it is sufficient to show that $d_1 = d_2$ to derive the equivalence between $T_{n,1}$ and $T_{n,2}$. We note that

$$\begin{aligned}
& (y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)(y - X\hat{\beta}_{b2sls}) \\
= \; & (y - X\hat{\beta}_{2sls} - X\hat{\beta}_{b2sls} + X\hat{\beta}_{2sls})'P(y - X\hat{\beta}_{2sls} - X\hat{\beta}_{b2sls} + X\hat{\beta}_{2sls}) \\
& -\alpha_n(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls}) \\
= \; & (y - X\hat{\beta}_{2sls})'P(y - X\hat{\beta}_{2sls}) - \alpha_n(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls}) \\
& +(\hat{\beta}_{2sls} - \hat{\beta}_{b2sls})'X'PX(\hat{\beta}_{2sls} - \hat{\beta}_{b2sls}),
\end{aligned}$$

where the last equality follows because $(y - X\hat{\beta}_{2sls})'PX = 0$. Given

$$\hat{\beta}_{2sls} - \hat{\beta}_{b2sls} = (X'PX)^{-1}X'Py - \hat{\beta}_{b2sls} = (X'PX)^{-1}X'P(y - X\hat{\beta}_{b2sls}),$$

we have

$$\begin{aligned}
& (y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)(y - X\hat{\beta}_{b2sls}) \\
= \; & (y - X\hat{\beta}_{2sls})'P(y - X\hat{\beta}_{2sls}) \\
& -\alpha_n(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls}) + (y - X\hat{\beta}_{b2sls})'PX(X'PX)^{-1}X'P(y - X\hat{\beta}_{b2sls}) \\
= \; & (y - X\hat{\beta}_{2sls})'P(y - X\hat{\beta}_{2sls}) - n\hat{B}.
\end{aligned}$$

It thus follows that $d_1 = d_2$. □

A.2 Proofs of Theorem 1 and Corollary 1

We first present two technical lemmas used to show the theorem.

Lemma A.2. *Under Assumption 1, as $n, K \to \infty$ we have*[8]

$$\frac{1}{\sqrt{n\alpha_n}}u'(P - \alpha_n I)u \to_d \mathcal{N}(0, w), \tag{A.1}$$

$$\frac{1}{\sqrt{n}}u'(P - \alpha_n I)X = O_p(1), \tag{A.2}$$

$$\frac{1}{n}X'(P - \alpha_n I)X \to_p (1 - \alpha)\Theta. \tag{A.3}$$

[8] It is important to note that $u'(P-\alpha_n I)u$ needs to be normalized by $\sqrt{n\alpha_n}$ not by $\sqrt{n}$. By doing so, the asymptotic distribution does not degenerate, even when $\alpha = 0$. This is because the asymptotic variance of $u'(P - \alpha_n I)u/\sqrt{n}$ is given by $2\alpha(1-\alpha)\sigma_u^4 + \left(\lim_{n,K\to\infty} n^{-1}\sum_{i=1}^n P_{ii}^2 - \alpha^2\right)\left\{\mathbb{E}(u_i^4) - 3\sigma_u^4\right\}$ and it is zero when $\alpha = 0$. In general, the rate of convergence of $u'(P - \alpha_n I)u/n$ is n if K is fixed; $\sqrt{n}$ if $K \to \infty$ but $K/n \to \alpha > 0$ using the CLT for quadratic forms (e.g., Kelejian and Prucha, 2001); somewhere between n and $\sqrt{n}$ if $K \to \infty$ but $K/n \to 0$. Normalizing with $\sqrt{n\alpha_n}$ reflects this irregular rate of convergence. Actually, if we exclude the case of $\alpha = 0$, then we simply derive the asymptotic distribution of $u'(P - \alpha_n I)u/\sqrt{n}$. However, even in this situation, the test statistics developed in this paper do not change. This is because the modification is simply moving the α_n (or α) term between the numerator and the denominator. Therefore, it would also not affect the equivalence result in Section 4. However, this degenerating asymptotic variance problem does not take place for (A.2).

Proof of Lemma A.2 We use Theorem 1 of van Hasselt (2009) to show (A.1). The matrices U, M, V, C, Ω and a in Theorem 1 of van Hasselt (2009) are u, 0, u, $(P - \alpha_n I)/\sqrt{\alpha_n}$, σ_u^2 and 1 in our case, respectively. The conditions for Theorem 1 of van Hasselt (2009) are summarized in Assumption 1 in van Hasselt (2009). Assumption 1(iii) in this paper corresponds to Assumption 1(a) in van Hasselt (2009). In our case, $M = 0$ so Assumption 1(b) in van Hasselt (2009) is automatically satisfied. We now consider Assumption 1(c) in van Hasselt (2009). The first two conditions in Assumption 1(c) hold with $Q_{CM} = \mu_{CM} = 0$ because $M = 0$ in our case.[9] Next, we have

$$\begin{aligned}\frac{1}{n} tr \left\{(P - \alpha_n I)/\sqrt{\alpha_n}\right\} &= \frac{1}{n\sqrt{\alpha_n}}(K - K) = 0 \quad \text{and} \\ \frac{1}{n} tr \left\{(P - \alpha_n I)^2/\alpha_n\right\} &= \frac{1}{n\alpha_n} tr \left\{(1 - 2\alpha_n) P + \alpha_n^2 I\right\} = 1 - \alpha_n,\end{aligned}$$

so the third and the fourth conditions in Assumption 1(c) are satisfied with $\tau_C = 0$ and $\tau_{C^2} = 1 - \alpha$. The fifth condition in Assumption 1(c) is also satisfied because $\sum_{i=1}^n (P_{ii}^2 - \alpha_n^2)/(n\alpha_n)$ converges by Assumption 1(vii). Lastly, we have

$$\sup_{n\geq 1} \sup_{1\leq j\leq n} \sum_{i=1}^n |P_{ij} - \alpha_n \kappa_{ij}|/\sqrt{\alpha_n} \leq \sup_{n\geq 1} \sup_{1\leq j\leq n} \sum_{i=1}^n |P_{ij}|/\sqrt{\alpha_n} + \sup_{n\geq 1} \sqrt{\alpha_n} < \infty$$

by Assumptions 1(i) and (vi) where $\kappa_{ij} = 1$ if $i = j$ and $= 0$ if $i \neq j$. Therefore, under Assumption 1, the conditions for Theorem 1 of van Hasselt (2009) are satisfied, which yields $u'(P - \alpha I)u/\sqrt{n\alpha_n} \to_d \mathcal{N}(0, w)$ as $n, K \to \infty$, where w is given as (10).

We also use Theorem 1 of van Hasselt (2009) to show (A.2). The matrices U, M, V, C, Ω and a in Theorem 1 of van Hasselt (2009) are now (u, X), $(0, Z\Pi)$, (u, V), $(P - \alpha_n I)$, Σ and $(1, 0, \cdots, 0)'$ in this case, respectively. We verify that Assumption 1 in van Hasselt (2009) is similarly satisfied as above. Assumption 1(iii) implies Assumption 1(a) in van Hasselt (2009); Assumption 1(v) implies Assumption 1(b) in van Hasselt (2009); and Assumptions 1(iv), 1(vi) and 1(vii) imply Assumption 1(c) in van Hasselt (2009). Therefore, under Assumption 1, Theorem 1 of van Hasselt (2009) yields (A.2) as $\mathbb{E}\{u'(P - \alpha_n I)X\} = 0$. Lastly, given $\mathbb{E}\{X'(P - \alpha_n I)X\} = (1 - \alpha_n)\Pi' Z' Z\Pi$, (A.3) follows.

Lemma A.3. *Under Assumptions 1 and 2, as $n, K \to \infty$ we have*

$$\begin{aligned}&2(1 - \alpha)\left\{\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls})\right\}^2 \\ &+ \left(\frac{1}{K}\sum_{i=1}^n P_{ii}^2 - \alpha^2\right)\left[\frac{1}{n}\sum_{i=1}^n (y_i - X_i'\hat{\beta}_{b2sls})^4 - 3\left\{\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls})\right\}^2\right] \\ &\to_p 2(1 - \alpha)\sigma_u^4 + \left(\lim_{n,K\to\infty} \frac{1}{K}\sum_{i=1}^n P_{ii}^2 - \alpha^2\right)\{\mathbb{E}(u_i^4) - 3\sigma_u^4\}.\end{aligned}$$

Proof of Lemma A.3 We only need to show

$$\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls}) \to_p \sigma_u^2 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n (y_i - X_i'\hat{\beta}_{b2sls})^4 \to_p \mathbb{E}(u_i^4)$$

as $n, K \to \infty$. First, we have

$$\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls}) = \frac{1}{n}(\beta - \hat{\beta}_{b2sls})'X'X(\beta - \hat{\beta}_{b2sls}) + \frac{2}{n}(\beta - \hat{\beta}_{b2sls})'X'u + \frac{1}{n}u'u \to_p \sigma_u^2$$

[9] Though Q_{CM} is zero in this case and thus it is no longer positive definite, it only affects the final expression of the variance.

given under Assumption 1, $\beta - \hat{\beta}_{b2sls} \to_p 0$ as $n, K \to \infty$ by Theorem 3 of van Hasselt (2009) and it can be easily verified that $X'X/n = O_p(1)$, $X'u/n = O_p(1)$ and $u'u/n \to_p \sigma_u^2$. Second, for the estimation of the fourth moment of u_i, we similarly have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(y_i - X_i'\hat{\beta}_{b2sls})^4 &= \frac{1}{n}\sum_{i=1}^{n}\left\{X_i'(\beta - \hat{\beta}_{b2sls})\right\}^4 + \frac{4}{n}\sum_{i=1}^{n}\left\{X_i'(\beta - \hat{\beta}_{b2sls})\right\}^3 u_i \\
&\quad + \frac{6}{n}\sum_{i=1}^{n}\left\{X_i'(\beta - \hat{\beta}_{b2sls})\right\}^2 u_i^2 \\
&\quad + \frac{4}{n}\sum_{i=1}^{n}\left\{X_i'(\beta - \hat{\beta}_{b2sls})\right\} u_i^3 + \frac{1}{n}\sum_{i=1}^{n} u_i^4 \\
&= \frac{1}{n}\sum_{i=1}^{n} u_i^4 + o_p(1) \to_p \mathbb{E}(u_i^4)
\end{aligned}$$

from Assumptions 1 and 2. The last equality follows because

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left\{X_i'(\beta - \hat{\beta}_{b2sls})\right\}^4\right| \leq \frac{1}{n}\sum_{i=1}^{n}\|X_i\|^4\|\beta - \hat{\beta}_{b2sls}\|^4 = O_p(1)o_p(1) = o_p(1)$$

by the existence of the eighth-order moment of X_i (Assumption 2) and $\beta - \hat{\beta}_{b2sls} \to_p 0$, where $\|\cdot\|$ is the Euclidean norm. A similar argument can show that $\sum_{i=1}^{n}\{X_i'(\beta - \hat{\beta}_{b2sls})\}^3 u_i/n = o_p(1)$, $\sum_{i=1}^{n}\{X_i'(\beta - \hat{\beta}_{b2sls})\}^2 u_i^2/n = o_p(1)$ and $\sum_{i=1}^{n}\{X_i'(\beta - \hat{\beta}_{b2sls})\} u_i^3/n = o_p(1)$.

Proof of Theorem 1 As $T_{n,1} = T_{n,2}$ from Lemma 1, we consider $T_{n,2}$ here. We observe that

$$\begin{aligned}
& (y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)(y - X\hat{\beta}_{b2sls}) \\
= \;& u'(P - \alpha_n I)u + (\hat{\beta}_{b2sls} - \beta)'X'(P - \alpha_n I)X(\hat{\beta}_{b2sls} - \beta) - 2(\hat{\beta}_{b2sls} - \beta)'X'(P - \alpha_n I)u \\
= \;& u'(P - \alpha_n I)u + u'(P - \alpha_n I)X\left\{X'(P - \alpha_n I)X\right\}^{-1}X'(P - \alpha_n I)u.
\end{aligned}$$

as $\hat{\beta}_{b2sls} - \beta = (X'(P - \alpha_n I)X)^{-1}X'(P - \alpha_n I)u$. Therefore, Lemma A.2 implies

$$\begin{aligned}
& \sqrt{\frac{n}{\alpha_n}}\left\{\frac{1}{n}(y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)(y - X\hat{\beta}_{b2sls})\right\} \\
= \;& \frac{1}{\sqrt{n\alpha_n}}u'(P - \alpha_n I)u + \frac{1}{\sqrt{K}}\left\{\frac{1}{\sqrt{n}}u'(P - \alpha_n I)X\right\}\left\{\frac{1}{n}X'(P - \alpha_n I)X\right\}^{-1}\left\{\frac{1}{\sqrt{n}}X'(P - \alpha_n I)u\right\} \\
= \;& \frac{1}{\sqrt{n\alpha_n}}u'(P - \alpha_n I)u + o_p(1) \to_d \mathcal{N}(0, w)
\end{aligned}$$

as $n, K \to \infty$. Furthermore, Lemma A.3 implies that $\hat{w} \to_p w$ as $\alpha_n = \alpha + o(n^{-1/2})$. It thus follows that $T_{n,2} \to_d \mathcal{N}(0, 1)$ as $n, K \to \infty$. □

Proof of Corollary 1 The result for $T_{n,2}$ is straightforward from the equivalence result in Lemma 1. For $T_{n,3}$, we first see that $\hat{\beta}_{liml} - \beta = O_p(n^{-1/2})$ even with $n, K \to \infty$ by Theorem 2 of van Hasselt (2009). Similarly as for the proof of Theorem 1, the $\sqrt{n}$-consistency of $\hat{\beta}_{liml}$ and Lemma

A.2 implies

$$\begin{aligned}
&\sqrt{\frac{n}{\alpha_n}}\left\{\frac{1}{n}(y-X\hat{\beta}_{liml})'(P-\alpha_n I)(y-X\hat{\beta}_{liml})\right\} \\
=\ & \frac{1}{\sqrt{n\alpha_n}}u'(P-\alpha_n I)u+\frac{1}{\sqrt{K}}\sqrt{n}(\beta-\hat{\beta}_{liml})'\left\{\frac{1}{n}X'(P-\alpha_n I)X\right\}\sqrt{n}(\beta-\hat{\beta}_{liml}) \\
&+\frac{2}{\sqrt{K}}\sqrt{n}(\beta-\hat{\beta}_{liml})'\left\{\frac{1}{\sqrt{n}}X'(P-\alpha_n I)u\right\} \\
=\ & \frac{1}{\sqrt{n\alpha_n}}u'(P-\alpha_n I)u+o_p(1)\to_d \mathcal{N}(0,w)
\end{aligned}$$

as $n,K\to\infty$. Noting that $\hat{\beta}_{liml}-\beta\to_p 0$, the essentially same argument as the proof of Lemma A.3 shows that $\hat{w}_l\to_p w$ and thus $T_{n,3}\to_d \mathcal{N}(0,1)$ as $n,K\to\infty$. □

A.3 Proof of Theorem 2

We observe that, for $y=X\beta+Z\gamma+u$ in this case,

$$\begin{aligned}
&\hat{\beta}_{b2sls}-\beta \\
=\ & \left\{\frac{1}{n}X'(P-\alpha_n I)X\right\}^{-1}\frac{1}{n}X'(P-\alpha_n I)(Z\gamma+u) \\
=\ & (1-\alpha_n)\frac{\alpha_n^{1/4}}{n^{1/4}}\left\{\frac{1}{n}X'(P-\alpha_n I)X\right\}^{-1}\frac{1}{n}X'Z\xi+\left\{\frac{1}{n}X'(P-\alpha_n I)X\right\}^{-1}\frac{1}{n}X'(P-\alpha_n I)u \\
=\ & o_p(1)
\end{aligned}$$

from Lemma A.2 and Assumption 3. Thus, $\hat{\beta}_{b2sls}$ is consistent, even under the local alternative. This observation and the fact that we consider the local alternative, indicate that $(y-X\hat{\beta}_{b2sls})'(y-X\hat{\beta}_{b2sls})/n\to_p\sigma_u^2$ and $\sum_{i=1}^n(y_i-X_i'\hat{\beta}_{b2sls})^4/n\to_p\mathbb{E}(u_i^4)$, which also yields $\hat{w}\to_p w$ as $n,K\to\infty$.

Next, we investigate the property of the numerator of the test statistic. Given

$$y-X\hat{\beta}_{b2sls}=[I-X\{X'(P-\alpha_n I)X\}^{-1}X'(P-\alpha_n I)](Z\gamma+u),$$

we obtain

$$\begin{aligned}
&(y-X\hat{\beta}_{b2sls})'(P-\alpha_n I)(y-X\hat{\beta}_{b2sls}) \\
=\ & (Z\gamma+u)'(P-\alpha_n I)(Z\gamma+u) \\
&-(Z\gamma+u)'(P-\alpha_n I)X\{X'(P-\alpha_n I)X\}^{-1}X'(P-\alpha_n I)(Z\gamma+u) \\
=\ & (1-\alpha_n)\gamma'Z'Z\gamma+2(1-\alpha_n)\gamma'Z'u+u'(P-\alpha_n I)u \\
&-\{(1-\alpha_n)\gamma'Z'X+u'(P-\alpha_n I)X\}\{X'(P-\alpha_n I)X\}^{-1}\{(1-\alpha_n)X'Z\gamma+X'(P-\alpha_n I)u\},
\end{aligned}$$

where the last equality is because $(P-\alpha_n I)Z=(1-\alpha_n)Z$. By Assumption 3 and using the local alternative $\gamma=(\alpha_n/n)^{1/4}\xi$, we have

$$\frac{1}{\sqrt{n\alpha_n}}\gamma'Z'Z\gamma=\frac{1}{n}\xi'Z'Z\xi=O_p(1),$$

$$\frac{1}{\sqrt{n\alpha_n}}\gamma'Z'u=\frac{1}{n^{1/4}\alpha_n^{1/4}}\frac{1}{\sqrt{n}}\xi'Z'u=\frac{1}{K^{1/4}}O_p(1)=o_p(1),$$

and

$$\frac{1}{n^{3/4}\alpha_n^{1/4}}\gamma' Z' X = \frac{1}{n}\xi' Z' X = O_p(1).$$

In addition, Lemma A.2 shows that

$$\frac{1}{n^{3/4}\alpha_n^{1/4}}u'(P-\alpha_n I)X = \frac{1}{n^{1/4}\alpha_n^{1/4}}\frac{1}{\sqrt{n}}u'(P-\alpha_n I)X = \frac{1}{K^{1/4}}O_p(1) = o_p(1).$$

Therefore, we have

$$\hat{d}_2 = \frac{1}{\sqrt{n\alpha_n}}(y - X\hat{\beta}_{b2sls})'(P-\alpha_n I)(y - X\hat{\beta}_{b2sls}) \to_d \mathcal{N}(C, w)$$

from Lemma A.2, where,

$$C = (1-\alpha)\lim_{n,K\to\infty}\frac{\xi' Z' Z\xi}{n} - (1-\alpha)\left(\lim_{n,K\to\infty}\frac{\xi' Z' X}{n}\right)\Theta^{-1}\left(\lim_{n,K\to\infty}\frac{X' Z\xi}{n}\right).$$

□

A.4 Proof of Theorem 3

It is straightforward from (19) and (20) because (18) implies

$$\begin{aligned} m_2 &= \left(2(1-\alpha_n)\frac{\left\{(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls})\right\}^2}{\left[\frac{X'(P-\alpha_n I)y}{X'(P-\alpha_n I)X}\{X'(P-\alpha_n I)X\}\right]^2}\right)^{-1/2} \\ &\quad \times\sqrt{\frac{n}{\alpha_n}}\left\{-\frac{(y - X\hat{\beta}_{b2sls})'(P-\alpha_n I)(y - X\hat{\beta}_{b2sls})}{X'(P-\alpha_n I)y}\right\} \\ &= \frac{\hat{d}_2}{\sqrt{\tilde{w}}}\times\left\{-\frac{|X'(P-\alpha_n I)y|}{X'(P-\alpha_n I)y}\right\}, \end{aligned}$$

where $\hat{\beta}_{b2sls} = \{X'(P-\alpha_n I)X\}^{-1}X'(P-\alpha_n I)y$. □

B Appendix: Equivalence Result with Two Endogenous Regressors

This appendix shows the equivalence between our modified Sargan test and the Hahn–Hausman test under normality when there are two endogenous variables. We consider the Hahn–Hausman test statistic in equation (5.5) of Hahn and Hausman (2002).

Let $X = (x_1, x_2)$, where x_1 and x_2 are $n \times 1$ vectors of endogenous regressors. Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the bias-corrected 2SLS estimators of the coefficient on x_1 and x_2, respectively. It appears that

$$\hat{\beta}_1 = \frac{x_2'(P-\alpha_n I)x_2 \cdot x_1'(P-\alpha_n I)y - x_1'(P-\alpha_n I)x_2 \cdot x_2'(P-\alpha_n I)y}{x_1'(P-\alpha_n I)x_1 \cdot x_2'(P-\alpha_n I)x_2 - \{x_1'(P-\alpha_n I)x_2\}^2}. \tag{A.4}$$

We also consider the reverse regression of x_1 on y and x_2 using the same instruments Z. Let $\hat{\delta}_1$ and $\hat{\delta}_2$ be the bias-corrected 2SLS estimators of the coefficient on y and x_2, respectively. We can see

$$\hat{\delta}_1 = \frac{x_2'(P-\alpha_n I)x_2 \cdot x_1'(P-\alpha_n I)y - x_1'(P-\alpha_n I)x_2 \cdot x_2'(P-\alpha_n I)y}{y'(P-\alpha_n I)y \cdot x_2'(P-\alpha_n I)x_2 - \{y'(P-\alpha_n I)x_2\}^2}. \tag{A.5}$$

The Hahn–Hausman test statistic for the two endogenous variables case is given by[10]

$$\sqrt{n}\breve{w}^{-1/2}\left(\hat{\beta}_1 - \frac{1}{\hat{\delta}_1}\right), \tag{A.6}$$

where

$$\begin{aligned}
\breve{w} &= \frac{2K}{n-K}\cdot\frac{\left\{(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)\right\}^2}{\hat{\beta}_1^2\left[x_1'Px_1 - \frac{K}{n-K}x_1'(I-P)x_1 - \frac{\left\{x_1'Px_2 - \frac{K}{n-K}x_1'(I-P)x_2\right\}^2}{\left\{x_2'Px_2 - \frac{K}{n-K}x_2'(I-P)x_2\right\}}\right]^2} \\
&= 2\alpha_n(1-\alpha_n)\frac{\left\{(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)\right\}^2}{\hat{\beta}_1^2\left[x_1'(P-\alpha_n I)x_1 - \frac{\{x_1'(P-\alpha_n I)x_2\}^2}{x_2'(P-\alpha_n I)x_2}\right]^2} \\
&= n^2\alpha_n\frac{\tilde{w}}{\hat{\beta}_1^2\left[x_1'(P-\alpha_n I)x_1 - \frac{\{x_1'(P-\alpha_n I)x_2\}^2}{x_2'(P-\alpha_n I)x_2}\right]^2}.
\end{aligned}$$

We now show that the test statistic (A.6) is numerically equivalent to $T_{n,2}$ up to a sign. First, (A.4) and (A.5) imply

$$\hat{\beta}_1 - \frac{1}{\hat{\delta}_1} = -\frac{(y - x_1\hat{\beta}_1)'(P-\alpha_n I)y\cdot x_2'(P-\alpha_n I)x_2 - (y - x_1\hat{\beta}_1)'(P-\alpha_n I)x_2\cdot y'(P-\alpha_n I)x_2}{x_2'(P-\alpha_n I)x_2\cdot x_1'(P-\alpha_n I)y - x_1'(P-\alpha_n I)x_2\cdot x_2'(P-\alpha_n I)y}. \tag{A.7}$$

by rewriting $y = y - x_1\hat{\beta}_1 + x_1\hat{\beta}_1$. Note that

$$\begin{aligned}
(y - x_1\hat{\beta}_1)'(P-\alpha_n I)x_2 &= (y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(P-\alpha_n I)x_2 + \hat{\beta}_2x_2'(P-\alpha_n I)x_2 \\
&= \hat{\beta}_2x_2'(P-\alpha_n I)x_2,
\end{aligned}$$

by the definition of the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. Therefore, the numerator of the ratio (A.7) becomes

$$\begin{aligned}
& (y - x_1\hat{\beta}_1)'(P-\alpha_n I)y\cdot x_2'(P-\alpha_n I)x_2 - \hat{\beta}_2x_2'(P-\alpha_n I)x_2\cdot y'(P-\alpha_n I)x_2 \\
=\ & \{x_2'(P-\alpha_n I)x_2\}\cdot(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(P-\alpha_n I)y \\
=\ & \{x_2'(P-\alpha_n I)x_2\}\cdot(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(P-\alpha_n I)(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2),
\end{aligned}$$

where the last equality follows from the fact that $(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(P-\alpha_n I)x_1 = 0$ and $(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(P-\alpha_n I)x_2 = 0$. To sum up, the difference between the two estimators can be written as

$$\begin{aligned}
\hat{\beta}_1 - \frac{1}{\hat{\delta}_1} &= -\frac{\{x_2'(P-\alpha_n I)x_2\}\cdot(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(P-\alpha_n I)(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)}{x_2'(P-\alpha_n I)x_2\cdot x_1'(P-\alpha_n I)y - x_1'(P-\alpha_n I)x_2\cdot x_2'(P-\alpha_n I)y} \\
&= -\frac{1}{\hat{\beta}_1}\times\left[x_1'(P-\alpha_n I)x_1 - \frac{\{x_1'(P-\alpha_n I)x_2\}^2}{x_2'(P-\alpha_n I)x_2}\right]^{-1}\times\sqrt{n\alpha_n}\hat{d}_2.
\end{aligned}$$

It follows that

$$\sqrt{n}\breve{w}^{-1/2}\left(\hat{\beta}_1 - \frac{1}{\hat{\delta}_1}\right) = -\mathrm{sgn}[\tau_1]\frac{\hat{d}_2}{\sqrt{\tilde{w}}} = T_{n,2}\cdot\mathrm{sgn}[-\tau_1],$$

[10] As in the case of a single endogenous variable, there is minor difference between the test statistic here and that given in equation (5.5) of Hahn and Hausman (2002). However, the difference disappears at a rate faster than $n^{-1/2}$.

where

$$\tau_1 = \hat{\beta}_1 \left[x_1'(P-\alpha_n I)x_1 - \frac{\{x_1'(P-\alpha_n I)x_2\}^2}{x_2'(P-\alpha_n I)x_2} \right]$$
$$= x_1'(P-\alpha_n I)y - \frac{x_1'(P-\alpha_n I)x_2 \cdot x_2'(P-\alpha_n I)y}{x_2'(P-\alpha_n I)x_2}.$$

Thus, we have established that the Hahn–Hausman test statistic is numerically equivalent to the modified Sargan test up to sign, even when there are two endogenous variables.

Note that if we let $\hat{x}_1 = (P-\alpha_n I)x_1$ and $\hat{x}_2 = (P-\alpha_n I)x_2$, which are the predicted x_1 and x_2 from the first-stage regression (with some modification to correct the bias), τ_1 reflects nothing but the sample covariance between $\hat{x}_1$ and y after $\hat{x}_2$ is projected out: $\tau_1 = \hat{x}_1'\{I - \hat{x}_2(\hat{x}_2'\hat{x}_2)^{-1}\hat{x}_2'\}y$. In comparison, τ_1 is simply $\hat{x}_1'y$ when there is only one endogenous regressor x_1 in Theorem 3. Therefore, even when there are more than three endogenous regressors (provided that the number of endogenous regressors is small and finite compared to the number of instruments), we can expect that the Hahn–Hausman test would be numerically equivalent to the modified Sargan test up to a sign, where the sign depends on the negative of the marginal sample covariance between the predicted endogenous regressor, which is used for the reverse regression, and the dependent variable.

References

Anatolyev, S. and Gospodinov, N. (2009). Specification testing in models with many instruments, *Econometric Theory*: forthcoming.

Andrews, D.W.K. and Stock, J. H. (2007). Testing with many weak instruments, *Journal of Econometrics* **138**(1): 24-46.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* **62**(3): 657–681.

Calhoun, G. (2008). Hypothesis testing in linear regression when k/n is large, unpublished manuscript, Iowa State University.

Chao, J. C. and Swanson, N. R. (2005). Consistent estimation with a large number of weak instruments, *Econometrica* **73**(5): 1673–1692.

Donald, S. G., Imbens, G. W. and Newey, W. K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions, *Journal of Econometrics* **117**: 55-93.

Donald, S. G. and Newey, W. K. (2001). Choosing the number of instruments, *Econometrica* **69**(5): 1161–1191.

Doornik, J. A. (2007). *Ox 5 - An Object-oriented Matrix Language*, Timberlake Consultant Ltd., London.

Hahn, J. and Hausman, J. (2002). A new specification test for the validity of instrumental variables, *Econometrica* **70**(1): 163–189.

Han, C. and Phillips, P. C. B. (2006). GMM with many moment conditions, *Econometrica* **74**(1), 147–192.

Hansen, C., Hausman, J. and Newey, W. K. (2008). Estimation with many instrumental variables, *Journal of Business & Economics Statistics* **26**(4): 398–422.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica* **50**(4): 1029–1053.

Hausman, J., Stock, J. H. and Yogo, M. (2005). Asymptotic properties of the Hahn–Hausman test for weak instruments, *Economics Letters* **89**: 333–342.

Kelejian, H. H. and Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications, *Journal of Econometrics* **104**: 219–257.

Kunitomo, N. (1980). Asymptotic expansions of the distributions of estimators in a linear functional relationship and simultaneous equations, *Journal of the American Statistical Association* **75**: 693–700.

Kunitomo, N., Morimune, K. and Tsukuda, Y. (1983). Asymptotic expansions of the distributions of the test statistics for overidentifying restrictions in a system of simultaneous equations, *International Economic Review* **24**(1): 199–215.

Morimune, K. (1983). Approximate distribution of the k-class estimators when the degree of overidentifiability is large compared with the sample size, *Econometrica* **51**(3): 821–841.

Nagar, A. L. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations, *Econometrica* **27**(4): 575–595.

Newey, W. K. (1985). Generalized method of moments specification testing, *Journal of Econometrics* **29**: 229–256.

Newey, W. K. and Windmeijer, F. (2009). GMM with many weak moment conditions, *Econometrica* **77**(3): 687-719.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables, *Econometrica* **26**(3): 393–415.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments, *Econometrica* **65**(3): 557–586.

Stock, J. H. and Wright, J. H. (2000). GMM with weak identification, *Econometrica* **68**(5): 1097–1126.

van Hasselt, M. (2009). Many instruments asymptotic approximations under nonnormal error distributions, *Econometric Theory*: forthcoming.

			Sargan	SB	SL	HH	MSn	MSnL	MSnn	MSnnL
$R_f^2 = 0.01$										
$n = 250$	$K = 5$	$\rho = 0.0$	0.037	0.028	0.011	0.015	0.021	0.006	0.022	0.006
		$\rho = 0.5$	0.065	0.044	0.010	0.022	0.035	0.007	0.035	0.007
		$\rho = 0.9$	0.229	0.205	0.038	0.138	0.183	0.027	0.183	0.027
	$K = 10$	$\rho = 0.0$	0.044	0.027	0.006	0.010	0.024	0.003	0.024	0.003
		$\rho = 0.5$	0.046	0.032	0.005	0.017	0.028	0.003	0.028	0.003
		$\rho = 0.9$	0.249	0.173	0.031	0.122	0.159	0.024	0.161	0.025
	$K = 30$	$\rho = 0.0$	0.028	0.011	0.002	0.012	0.013	0.002	0.014	0.002
		$\rho = 0.5$	0.028	0.019	0.002	0.013	0.021	0.002	0.021	0.002
		$\rho = 0.9$	0.184	0.092	0.015	0.071	0.100	0.015	0.100	0.015
$n = 1000$	$K = 5$	$\rho = 0.0$	0.044	0.038	0.031	0.020	0.028	0.026	0.028	0.026
		$\rho = 0.5$	0.062	0.049	0.036	0.031	0.040	0.029	0.040	0.029
		$\rho = 0.9$	0.182	0.088	0.051	0.061	0.077	0.043	0.078	0.044
	$K = 10$	$\rho = 0.0$	0.046	0.036	0.016	0.017	0.030	0.011	0.031	0.011
		$\rho = 0.5$	0.086	0.051	0.027	0.025	0.047	0.022	0.047	0.022
		$\rho = 0.9$	0.314	0.121	0.055	0.091	0.115	0.049	0.115	0.049
	$K = 30$	$\rho = 0.0$	0.045	0.026	0.011	0.021	0.026	0.011	0.026	0.011
		$\rho = 0.5$	0.070	0.043	0.017	0.025	0.043	0.017	0.043	0.018
		$\rho = 0.9$	0.591	0.155	0.043	0.119	0.155	0.043	0.155	0.043
$R_f^2 = 0.2$										
$n = 250$	$K = 5$	$\rho = 0.0$	0.046	0.046	0.046	0.025	0.035	0.035	0.035	0.035
		$\rho = 0.5$	0.055	0.051	0.049	0.027	0.039	0.034	0.039	0.035
		$\rho = 0.9$	0.071	0.057	0.050	0.031	0.042	0.036	0.042	0.036
	$K = 10$	$\rho = 0.0$	0.044	0.043	0.042	0.029	0.040	0.040	0.040	0.040
		$\rho = 0.5$	0.056	0.046	0.044	0.033	0.042	0.040	0.042	0.041
		$\rho = 0.9$	0.109	0.055	0.045	0.033	0.050	0.043	0.050	0.043
	$K = 30$	$\rho = 0.0$	0.035	0.030	0.029	0.034	0.034	0.034	0.034	0.034
		$\rho = 0.5$	0.063	0.038	0.032	0.034	0.042	0.036	0.043	0.036
		$\rho = 0.9$	0.268	0.057	0.040	0.044	0.063	0.044	0.063	0.044
$n = 1000$	$K = 5$	$\rho = 0.0$	0.055	0.055	0.055	0.033	0.046	0.046	0.046	0.046
		$\rho = 0.5$	0.057	0.056	0.055	0.034	0.046	0.045	0.046	0.045
		$\rho = 0.9$	0.058	0.056	0.055	0.036	0.046	0.045	0.046	0.045
	$K = 10$	$\rho = 0.0$	0.066	0.066	0.066	0.031	0.055	0.055	0.055	0.055
		$\rho = 0.5$	0.070	0.067	0.067	0.031	0.055	0.055	0.055	0.055
		$\rho = 0.9$	0.081	0.069	0.069	0.033	0.062	0.058	0.062	0.058
	$K = 30$	$\rho = 0.0$	0.048	0.048	0.048	0.040	0.048	0.048	0.048	0.048
		$\rho = 0.5$	0.070	0.050	0.048	0.042	0.050	0.048	0.050	0.048
		$\rho = 0.9$	0.111	0.052	0.048	0.042	0.052	0.048	0.052	0.048

Table 1: Sizes of overidentifying restriction tests under Distribution D-I

			Sargan	SB	SL	HH	MSn	MSnL	MSnn	MSnnL
$R_f^2 = 0.01$										
$n = 250$	$K = 5$	$\rho = 0.0$	0.159	0.132	0.130	0.107	0.132	0.130	0.132	0.130
		$\rho = 0.5$	0.124	0.125	0.118	0.108	0.125	0.118	0.125	0.118
		$\rho = 0.9$	0.173	0.140	0.157	0.140	0.140	0.157	0.141	0.157
	$K = 10$	$\rho = 0.0$	0.123	0.114	0.094	0.093	0.114	0.094	0.114	0.094
		$\rho = 0.5$	0.123	0.108	0.087	0.097	0.108	0.087	0.108	0.087
		$\rho = 0.9$	0.197	0.154	0.129	0.154	0.154	0.129	0.154	0.129
	$K = 30$	$\rho = 0.0$	0.096	0.086	0.078	0.068	0.086	0.078	0.086	0.078
		$\rho = 0.5$	0.099	0.083	0.086	0.066	0.083	0.086	0.083	0.086
		$\rho = 0.9$	0.207	0.142	0.122	0.142	0.142	0.122	0.142	0.122
$n = 1000$	$K = 5$	$\rho = 0.0$	0.494	0.439	0.405	0.394	0.439	0.405	0.439	0.405
		$\rho = 0.5$	0.563	0.534	0.475	0.509	0.534	0.475	0.534	0.476
		$\rho = 0.9$	0.719	0.639	0.555	0.639	0.639	0.555	0.639	0.555
	$K = 10$	$\rho = 0.0$	0.427	0.358	0.320	0.305	0.358	0.320	0.358	0.320
		$\rho = 0.5$	0.447	0.414	0.363	0.392	0.414	0.363	0.414	0.363
		$\rho = 0.9$	0.602	0.496	0.374	0.496	0.496	0.374	0.496	0.374
	$K = 30$	$\rho = 0.0$	0.274	0.245	0.181	0.160	0.245	0.181	0.245	0.181
		$\rho = 0.5$	0.334	0.240	0.189	0.211	0.240	0.189	0.240	0.189
		$\rho = 0.9$	0.629	0.323	0.203	0.323	0.323	0.203	0.323	0.203
$R_f^2 = 0.2$										
$n = 250$	$K = 5$	$\rho = 0.0$	0.164	0.162	0.162	0.15	0.162	0.162	0.162	0.162
		$\rho = 0.5$	0.172	0.165	0.165	0.162	0.165	0.165	0.165	0.165
		$\rho = 0.9$	0.183	0.186	0.185	0.186	0.186	0.185	0.186	0.185
	$K = 10$	$\rho = 0.0$	0.144	0.145	0.144	0.117	0.145	0.144	0.145	0.144
		$\rho = 0.5$	0.148	0.150	0.150	0.135	0.150	0.150	0.150	0.150
		$\rho = 0.9$	0.167	0.157	0.157	0.154	0.157	0.157	0.157	0.156
	$K = 30$	$\rho = 0.0$	0.103	0.100	0.102	0.070	0.100	0.102	0.100	0.102
		$\rho = 0.5$	0.095	0.107	0.104	0.082	0.107	0.104	0.107	0.104
		$\rho = 0.9$	0.103	0.105	0.100	0.104	0.105	0.100	0.105	0.100
$n = 1000$	$K = 5$	$\rho = 0.0$	0.575	0.575	0.575	0.572	0.575	0.575	0.575	0.575
		$\rho = 0.5$	0.637	0.636	0.636	0.633	0.636	0.636	0.636	0.636
		$\rho = 0.9$	0.666	0.667	0.668	0.667	0.667	0.668	0.668	0.668
	$K = 10$	$\rho = 0.0$	0.480	0.480	0.480	0.453	0.480	0.480	0.480	0.480
		$\rho = 0.5$	0.500	0.500	0.500	0.473	0.500	0.500	0.500	0.500
		$\rho = 0.9$	0.517	0.516	0.517	0.496	0.516	0.517	0.516	0.517
	$K = 30$	$\rho = 0.0$	0.310	0.310	0.310	0.244	0.310	0.310	0.310	0.310
		$\rho = 0.5$	0.315	0.321	0.321	0.255	0.321	0.321	0.321	0.321
		$\rho = 0.9$	0.314	0.321	0.318	0.271	0.321	0.318	0.321	0.318

Table 2: Size-adjusted powers of overidentifying restriction tests under Distribution D-I

			Sargan	SB	SL	HH	MSn	MSnL	MSnn	MSnnL
$R_f^2 = 0.01$										
$n = 250$	$K = 5$	$\rho = 0.0$	0.035	0.027	0.009	0.013	0.021	0.005	0.021	0.005
		$\rho = 0.5$	0.058	0.046	0.011	0.019	0.036	0.007	0.038	0.007
		$\rho = 0.9$	0.248	0.194	0.032	0.136	0.170	0.022	0.172	0.023
	$K = 10$	$\rho = 0.0$	0.026	0.014	0.001	0.007	0.011	0.001	0.011	0.001
		$\rho = 0.5$	0.045	0.025	0.001	0.014	0.023	0.001	0.023	0.001
		$\rho = 0.9$	0.253	0.161	0.016	0.108	0.152	0.014	0.155	0.015
	$K = 30$	$\rho = 0.0$	0.033	0.016	0.000	0.011	0.023	0.002	0.024	0.003
		$\rho = 0.5$	0.041	0.024	0.001	0.020	0.032	0.003	0.032	0.003
		$\rho = 0.9$	0.236	0.128	0.011	0.098	0.141	0.015	0.144	0.017
$n = 1000$	$K = 5$	$\rho = 0.0$	0.043	0.038	0.022	0.020	0.029	0.015	0.029	0.015
		$\rho = 0.5$	0.073	0.055	0.034	0.034	0.044	0.025	0.044	0.025
		$\rho = 0.9$	0.190	0.094	0.054	0.065	0.082	0.040	0.082	0.040
	$K = 10$	$\rho = 0.0$	0.043	0.026	0.017	0.016	0.024	0.017	0.024	0.017
		$\rho = 0.5$	0.070	0.043	0.024	0.025	0.039	0.023	0.039	0.023
		$\rho = 0.9$	0.321	0.111	0.045	0.086	0.103	0.039	0.103	0.039
	$K = 30$	$\rho = 0.0$	0.050	0.026	0.010	0.023	0.026	0.010	0.026	0.010
		$\rho = 0.5$	0.076	0.029	0.014	0.021	0.029	0.014	0.029	0.014
		$\rho = 0.9$	0.571	0.152	0.046	0.122	0.152	0.046	0.153	0.047
$R_f^2 = 0.2$										
$n = 250$	$K = 5$	$\rho = 0.0$	0.049	0.048	0.048	0.024	0.038	0.037	0.039	0.039
		$\rho = 0.5$	0.057	0.049	0.048	0.028	0.038	0.038	0.040	0.040
		$\rho = 0.9$	0.078	0.057	0.050	0.034	0.044	0.040	0.044	0.041
	$K = 10$	$\rho = 0.0$	0.039	0.036	0.036	0.019	0.033	0.033	0.034	0.033
		$\rho = 0.5$	0.057	0.039	0.038	0.023	0.036	0.034	0.038	0.034
		$\rho = 0.9$	0.109	0.052	0.041	0.030	0.046	0.036	0.048	0.038
	$K = 30$	$\rho = 0.0$	0.036	0.032	0.030	0.037	0.039	0.036	0.042	0.039
		$\rho = 0.5$	0.074	0.041	0.033	0.036	0.049	0.041	0.050	0.046
		$\rho = 0.9$	0.285	0.061	0.035	0.044	0.066	0.052	0.069	0.055
$n = 1000$	$K = 5$	$\rho = 0.0$	0.061	0.061	0.061	0.034	0.045	0.045	0.046	0.046
		$\rho = 0.5$	0.064	0.061	0.061	0.035	0.046	0.046	0.047	0.046
		$\rho = 0.9$	0.063	0.062	0.062	0.037	0.048	0.047	0.048	0.047
	$K = 10$	$\rho = 0.0$	0.046	0.046	0.046	0.032	0.042	0.042	0.042	0.042
		$\rho = 0.5$	0.052	0.048	0.046	0.033	0.044	0.043	0.044	0.044
		$\rho = 0.9$	0.066	0.049	0.049	0.033	0.046	0.045	0.046	0.045
	$K = 30$	$\rho = 0.0$	0.054	0.054	0.054	0.048	0.054	0.054	0.055	0.055
		$\rho = 0.5$	0.065	0.056	0.054	0.048	0.056	0.054	0.058	0.055
		$\rho = 0.9$	0.123	0.061	0.056	0.051	0.061	0.056	0.061	0.059

Table 3: Sizes of overidentifying restriction tests under Distribution D-II

			Sargan	SB	SL	HH	MSn	MSnL	MSnn	MSnnL
$R_f^2 = 0.01$										
$n = 250$	$K = 5$	$\rho = 0.0$	0.165	0.157	0.126	0.121	0.157	0.126	0.158	0.126
		$\rho = 0.5$	0.136	0.132	0.142	0.121	0.132	0.142	0.132	0.142
		$\rho = 0.9$	0.196	0.165	0.190	0.165	0.165	0.190	0.163	0.191
	$K = 10$	$\rho = 0.0$	0.136	0.126	0.078	0.096	0.126	0.078	0.127	0.078
		$\rho = 0.5$	0.098	0.100	0.084	0.084	0.100	0.084	0.100	0.085
		$\rho = 0.9$	0.202	0.164	0.123	0.164	0.164	0.123	0.163	0.123
	$K = 30$	$\rho = 0.0$	0.088	0.090	0.073	0.078	0.090	0.073	0.091	0.073
		$\rho = 0.5$	0.094	0.083	0.071	0.075	0.083	0.071	0.082	0.072
		$\rho = 0.9$	0.174	0.133	0.120	0.133	0.133	0.120	0.134	0.120
$n = 1000$	$K = 5$	$\rho = 0.0$	0.496	0.424	0.398	0.385	0.424	0.398	0.422	0.398
		$\rho = 0.5$	0.550	0.511	0.475	0.507	0.511	0.475	0.511	0.475
		$\rho = 0.9$	0.670	0.643	0.540	0.643	0.643	0.540	0.643	0.540
	$K = 10$	$\rho = 0.0$	0.422	0.358	0.336	0.288	0.358	0.336	0.359	0.336
		$\rho = 0.5$	0.471	0.413	0.360	0.368	0.413	0.360	0.413	0.360
		$\rho = 0.9$	0.558	0.470	0.380	0.470	0.470	0.380	0.470	0.380
	$K = 30$	$\rho = 0.0$	0.286	0.272	0.226	0.170	0.272	0.226	0.272	0.226
		$\rho = 0.5$	0.351	0.274	0.222	0.249	0.274	0.222	0.274	0.222
		$\rho = 0.9$	0.588	0.35	0.199	0.350	0.350	0.199	0.350	0.199
$R_f^2 = 0.2$										
$n = 250$	$K = 5$	$\rho = 0.0$	0.16	0.161	0.162	0.147	0.161	0.162	0.161	0.161
		$\rho = 0.5$	0.177	0.180	0.183	0.167	0.180	0.183	0.183	0.182
		$\rho = 0.9$	0.202	0.205	0.199	0.205	0.205	0.199	0.199	0.200
	$K = 10$	$\rho = 0.0$	0.154	0.152	0.153	0.131	0.152	0.153	0.152	0.153
		$\rho = 0.5$	0.159	0.16	0.162	0.141	0.160	0.162	0.160	0.162
		$\rho = 0.9$	0.165	0.164	0.167	0.163	0.164	0.167	0.164	0.168
	$K = 30$	$\rho = 0.0$	0.100	0.097	0.099	0.08	0.097	0.099	0.097	0.099
		$\rho = 0.5$	0.093	0.096	0.097	0.084	0.096	0.097	0.099	0.097
		$\rho = 0.9$	0.108	0.097	0.091	0.097	0.097	0.091	0.097	0.092
$n = 1000$	$K = 5$	$\rho = 0.0$	0.573	0.572	0.572	0.572	0.572	0.572	0.573	0.573
		$\rho = 0.5$	0.601	0.601	0.601	0.601	0.601	0.601	0.601	0.601
		$\rho = 0.9$	0.645	0.644	0.644	0.644	0.644	0.644	0.644	0.644
	$K = 10$	$\rho = 0.0$	0.476	0.477	0.479	0.415	0.477	0.479	0.477	0.477
		$\rho = 0.5$	0.503	0.510	0.510	0.454	0.510	0.510	0.506	0.509
		$\rho = 0.9$	0.541	0.542	0.541	0.488	0.542	0.541	0.542	0.541
	$K = 30$	$\rho = 0.0$	0.285	0.288	0.288	0.241	0.288	0.288	0.288	0.288
		$\rho = 0.5$	0.299	0.298	0.299	0.250	0.298	0.299	0.299	0.299
		$\rho = 0.9$	0.324	0.301	0.302	0.269	0.301	0.302	0.301	0.302

Table 4: Size-adjusted powers of overidentifying restriction tests under Distribution D-II