Journal of Econometrics 167 (2012) 133–139

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Hahn–Hausman test as a specification test[☆]

Yoonseok Lee[a,*], Ryo Okui[b]

[a] Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109-1220, USA

[b] Institute of Economic Research, Kyoto University, Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan

ARTICLE INFO

Article history:
Received 1 December 2009
Received in revised form
2 May 2011
Accepted 21 October 2011
Available online 3 November 2011

JEL classification:
C12
C21

Keywords:
Hahn–Hausman test
Sargan test
Many instruments
Overidentifying restrictions test
Specification test

ABSTRACT

This paper develops a modified version of the Sargan [Sargan, J.D., 1958. The estimation of economic relationships using instrumental variables. Econometrica 26 (3), 393–415] restrictions, and shows that it is numerically equivalent to the test statistic of Hahn and Hausman [Hahn, J., Hausman, J., 2002. A new specification test for the validity of instrumental variables. Econometrica 70 (1), 163–189] up to a sign. The modified Sargan test is constructed such that its asymptotic distribution under the null hypothesis of correct specification is standard normal when the number of instruments increases with the sample size. The equivalence result is useful in understanding what the Hahn–Hausman test detects and its power properties.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The conventional asymptotic theory often provides a poor approximation of the finite sample distribution of instrumental variables estimators or test statistics for instrumental variables regressions. Examples are with weak instruments (e.g., Staiger and Stock, 1997 and Stock and Wright, 2000) or many instruments (e.g., Morimune, 1983; Bekker, 1994; Han and Phillips, 2006; Andrews and Stock, 2007; Hansen et al., 2008; Newey and Windmeijer, 2009; Chao et al., 2010; van Hasselt, 2010 and Anatolyev and Gospodinov, 2011). Hahn and Hausman (2002, HH hereafter) propose a test that examines the adequacy of the standard asymptotic result in linear instrumental variables regression models.

This paper contributes to the literature by developing a modified version of the Sargan test (1958) of overidentifying restrictions, and shows that it is numerically equivalent to the HH test statistic up to a sign. The modification is such that the asymptotic distribution of the test statistic is standard normal under the null hypothesis of correct specification, when the number of instruments increases with the sample size like Bekker (1994). Though these two tests are developed from very different motivations, the equivalence tells that they indeed examine the common hypothesis: the orthogonality between the instruments and the structural equation error.

This equivalence result provides many interesting implications. First of all, it explains why the HH test does not have power in detecting weak instruments (e.g., Hausman et al., 2005): it indeed tests for the exogeneity of the instruments. This finding also enables us to examine its power properties. The equivalence result is useful for overcoming several limitations of the original HH test. For example, we can easily handle cases with multiple endogenous regressors or with the LIML estimators in the modified Sargan test. Moreover, as the Sargan test is a special case of the J-test by Hansen (1982), the result provides a direction to extend the HH test to more general setup, such as moment-condition-based nonlinear

☆ Previous versions are circulated under the title "A Specification Test for Instrumental Variables Regression with Many Instruments." The authors acknowledge the valuable comments from Takeshi Amemiya, the Associate Editor, two anonymous referees, Donald Andrews, Mehmet Caner, Juan Carlos Escanciano, Nikolay Gospodinov, Jinyong Hahn, Han Hong, Simon Lee, Yukitoshi Matsushita, Whitney Newey, Taisuke Otsu, Peter Phillips, Mototsugu Shintani and Katsumi Simotsu, along with seminar participants at Columbia, Yale, the Kobe meeting of the Kansai Econometric Society, the 2009 SETA meeting, and the 2009 Far East and South Asia Meeting of the Econometric Society. Lee thanks the Cowles Foundation for Research in Economics at Yale University, where he was a visiting fellow while writing this paper. Okui appreciates the financial support from Japan Society of the Promotion of Science under KAKENHI22730176 and KAKENHI22330067. The usual disclaimer applies.

* Corresponding author. Tel.: +1 734 615 0177; fax: +1 734 764 2769.
E-mail addresses: yoolee@umich.edu (Y. Lee), okui@kier.kyoto-u.ac.jp (R. Okui).

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.10.005

models, whereas it is not clear how to generalize the idea of using reverse regression in HH for nonlinear models.

Several studies are closely related to the modified Sargan test developed in this paper. Andrews and Stock (2007) and Newey and Windmeijer (2009) consider testing problems with many weak instruments though the number of instruments is restricted such that it increases at a much slower rate than that of the sample size. Anatolyev and Gospodinov (2011) develop a modification of the critical values of the overidentifying restrictions test so that the test has correct size based on the chi-square approximation, when the number of instruments is proportional to the sample size. Chao et al. (2010) develop a specification test under heteroskedasticity for linear instrumental variables regressions with many instruments.

The remainder of the paper is organized as follows. Section 2 describes the basic framework and develops the modified Sargan test. Section 3 establishes the equivalence between the modified Sargan test and the HH test up to a sign. Section 4 discusses the implications of the equivalence results and concludes the paper. All the mathematical proofs are provided in Appendix.

2. Model and modified Sargan test

We consider a linear instrumental variables regression model given by

$$y_i = X_i'\beta + u_i$$

for $i = 1, 2, \ldots, n$, 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. Let Z_i be a $K \times 1$ vector of instruments, which we treat as deterministic, where $r < K < n$. 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

$$\alpha_n \equiv K/n \to \alpha \quad \text{as } n, K \to \infty$$

for some $0 \le \alpha < 1$. However, the number of regressors (r) is fixed and does not depend on n nor K. We exclude the fixed K case, but α can be zero when K diverges at a rate slower than n. We further assume that

$$X_i = \Pi' Z_i + V_i,$$

where Π is the $K \times r$ matrix of parameters whose value may depend on n as well as K. The unobservables $\varepsilon_i = (u_i, V_i')'$ are assumed to be independently and identically distributed (i.i.d.) and we define

$$\text{Var}(\varepsilon_i) \equiv \Sigma = \begin{pmatrix} \sigma_u^2 \sigma_{Vu}' \\ \sigma_{Vu} \Sigma_V \end{pmatrix}, \tag{1}$$

where $\sigma_{Vu} \neq 0$ so that X_i is correlated with u_i through the correlation between u_i and V_i. We make the following assumptions, where we let $P = Z\left(Z'Z\right)^{-1} Z'$ with $Z = (Z_1, \ldots, Z_n)'$.

Assumption 1. (i) $\alpha_n = \alpha + o(n^{-1/2})$ for some $0 \le \alpha < 1$ as $n, K \to \infty$. (ii) Z and Π are of full column rank. (iii) ε_i are i.i.d. for $i = 1, \ldots, n$ with mean zero and positive definite variance matrix Σ in (1); 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) $\sup_{1 \le i \le n} |Z_i' \pi_j| < \infty$ for all $j = 1, \ldots, r$, where π_j is the jth column of Π. (vi) $\sup_{n \ge 1} \sup_{1 \le j \le n} \sum_{i=1}^n |P_{ij}| / \sqrt{\alpha_n} < \infty$, 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 $n, K \to \infty$. (viii) X_i and u_i have finite eighth moments.

Assumption 1 is similar to the assumption in van Hasselt (2010), which is for the central limit theorem of the quadratic forms under many instrument framework. Note that ε_i is assumed homoskedastic though it does not need to be normal. See van Hasselt (2010) for more discussions about the assumptions. Condition (viii) is for the consistency of the asymptotic variance estimator of the modified Sargan test statistic defined below.

We let $\hat{\beta}_{2sls} = \left(X'PX\right)^{-1} X'Py$ be the two-stage least squares (2SLS) estimator, where $X = (X_1, \ldots, X_n)'$ and $y = (y_1, \ldots, y_n)'$. The standard Sargan test statistic (Sargan, 1958) is defined as

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

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 $\mathbb{E}(u_i Z_i) = 0$, the standard asymptotic theory (i.e., when K is fixed) gives $\hat{\beta}_{2sls} - \beta = O_p(n^{-1/2})$ and

$$S_n(\hat{\beta}_{2sls}) \to_d \chi^2_{K-r} \quad \text{as } n \to \infty. \tag{3}$$

When $K \to \infty$, however, the right hand side of (3) diverges and the asymptotic distribution of $S_n(\hat{\beta}_{2sls})$ is not well-defined. We need to modify the Sargan test statistic properly in order to analyze its asymptotic distribution.

More precisely, we let $\hat{B}$ be a consistent estimator of $\text{plim}_{n,K\to\infty}\, \hat{u}'P\hat{u}/n$ given by[1]

$$\hat{B} = \alpha_n \left(\hat{u}_b'\hat{u}_b/n\right) - \left(\hat{u}_b' PX/n\right) \left(X'PX/n\right)^{-1} \left(X'P\hat{u}_b/n\right),$$

where $\hat{u}_b = y - X\hat{\beta}_{b2sls}$ and

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

is the bias-corrected 2SLS estimator (e.g., Nagar, 1959) that satisfies $\hat{\beta}_{b2sls} - \beta = O_p(n^{-1/2})$ even when $n, K \to \infty$ with $\alpha \neq 0$. Then we can show that

$$\sqrt{n/\alpha_n}\left(\hat{u}'P\hat{u}/n - \hat{B}\right) \to_d \mathcal{N}(0, w) \quad \text{as } n, K \to \infty \tag{4}$$

under Assumption 1 (technical details are in Lemmas A.1 and A.2 in Appendix), where

$$w = 2(1-\alpha)\sigma_u^4 + \left(\lim_{n,K\to\infty} \sum_{i=1}^n (P_{ii}^2 - \alpha_n^2)/(n\alpha_n)\right) \times \left(\mathbb{E}u_i^4 - 3\sigma_u^4\right). \tag{5}$$

Note that, different from HH and Anatolyev and Gospodinov (2011), we allow for the case $\alpha = 0$. This difference necessitates us normalizing $\hat{u}'P\hat{u}/n$ by $\sqrt{n/\alpha_n}$ in (4) instead of $\sqrt{n}$, and thus the rate of convergence of $\hat{u}'P\hat{u}/n$ is faster than $\sqrt{n}$ when $\alpha = 0$. This asymptotic result can handle the cases with $\alpha > 0$ and with $\alpha = 0$ in a unified framework.[2]

From (4), a specification test can be obtained as the t-test statistic

$$T_n = \hat{d}_1/\sqrt{\hat{w}}, \tag{6}$$

where

$$\hat{d}_1 = \sqrt{n/\alpha_n}\left(\hat{u}'P\hat{u}/n - \hat{B}\right),$$

$$\hat{w} = 2\left(1-\alpha_n\right)\left(\hat{u}_b'\hat{u}_b/n\right)^2 + \left(\sum_{i=1}^n \left(P_{ii}^2 - \alpha_n^2\right)/(n\alpha_n)\right) \times \left(\sum_{i=1}^n \hat{u}_{b,i}^4/n - 3\left(\hat{u}_b'\hat{u}_b/n\right)^2\right)$$

[1] Note that $\hat{u}'P\hat{u}/n = u'Pu/n - u'PX/n\left(X'PX/n\right)^{-1} X'Pu/n \to_p \alpha\sigma_u^2 - \alpha^2\sigma_{Vu}'(\Theta + \alpha\Sigma_V)^{-1}\sigma_{Vu}$ as $n, K \to \infty$.

[2] Since we need to accommodate the possibility of $\alpha = 0$, the arguments made in the proof are slightly different from those of HH and Anatolyev and Gospodinov (2011).

with $\hat{u}_{b,i}$ being the ith element of $\hat{u}_b$. One remark is that

$$\hat{u}'P\hat{u}/n - \hat{B} = \hat{u}_b'(P - \alpha_n I)\hat{u}_b/n \tag{7}$$

holds with I being the n-dimensional identity matrix. See Appendix A.1 for the proof of (7). Therefore, T_n can be re-expressed as

$$T_n = \hat{d}_2/\sqrt{\hat{w}} \quad \text{with } \hat{d}_2 = \sqrt{n/\alpha_n}\left(\hat{u}_b'(P - \alpha_n I)\hat{u}_b/n\right). \tag{8}$$

This expression provides another interpretation of T_n: it is the standardized version of the minimized objective function for $\hat{\beta}_{b2sls} = \arg\min_\beta (y - X\beta)'(P - \alpha_n I)(y - X\beta)$. Because the 2SLS estimator $\hat{\beta}_{2sls}$ is biased in the presence of many instruments, bias correction is necessary when constructing overidentifying restrictions test statistics. This remark demonstrates that bias correction for the estimators (viz., (8)) is equivalent to bias correction for the test statistics (viz., (6)) in the linear instrumental variables regression, since (7) implies $\hat{d}_1 = \hat{d}_2$.

If we further assume that u_i is normally distributed as considered in Bekker (1994) and thus $\mathbb{E}u_i^4 = 3\sigma_u^4$, then the asymptotic variance w can be simply estimated by

$$\tilde{w} = 2(1 - \alpha_n)\left(\hat{u}_b'\hat{u}_b/n\right)^2.$$

In this case, we can develop a simpler test statistic $\tilde{T}_n$ given by

$$\tilde{T}_n = \hat{d}_2/\sqrt{\tilde{w}} = \left\{S_n(\hat{\beta}_{b2sls}) - K\right\}/\sqrt{2K(1 - \alpha_n)}, \tag{9}$$

where $S_n(\hat{\beta}_{b2sls}) = \hat{u}_b'P\hat{u}_b/(\hat{u}_b'\hat{u}_b/n)$ is the standard Sargan statistic in (2) using the bias-corrected 2SLS estimator $\hat{\beta}_{b2sls}$ instead of $\hat{\beta}_{2sls}$. The expression (9) motivates us to call $\tilde{T}_n$ and T_n as *modified Sargan test* statistics, in which the modification is to accommodate many instrument asymptotics.[3]

The following theorem gives the asymptotic null distribution of the modified Sargan test statistics under many instrument asymptotics.[4]

Theorem 1. *If Assumption 1 holds, $T_n \to_d \mathcal{N}(0, 1)$ as $n, K \to \infty$ under $\mathbb{E}(u_i Z_i) = 0$. In addition, when u_i is normally distributed, $\tilde{T}_n \to_d \mathcal{N}(0, 1)$.*

3. Equivalence results

The HH-test examines the adequacy of the standard asymptotic result in linear instrumental variables regression models, using the difference between the instrumental variables estimator and the inverse of that from the reverse regression. For the scalar X_i case, more precisely, they consider

$$\Delta_n = \frac{X'(P - \alpha_n I)y}{X'(P - \alpha_n I)X} - \frac{y'(P - \alpha_n I)y}{X'(P - \alpha_n I)y}, \tag{10}$$

which is formulated using the bias-corrected 2SLS estimators. When the standard asymptotic results are violated, these two estimators have different probability limits. Assuming normality, the HH-test statistic in this case is defined as[5]

$$m_2 = \sqrt{n}\Delta_n \left[\frac{2K}{n-K} \times \frac{\left\{(y - X\hat{\beta}_{b2sls})'(y - X\hat{\beta}_{b2sls})\right\}^2}{\hat{\beta}_{b2sls}^2 \{X'PX - (K/(n-K))X'(I-P)X\}^2}\right]^{-1/2}. \tag{11}$$

Noting that $(y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)X = 0$, however, the difference (10) can be rewritten as

$$\begin{aligned}\Delta_n &= \hat{\beta}_{b2sls} - \frac{(y - X\hat{\beta}_{b2sls} + X\hat{\beta}_{b2sls})'(P - \alpha_n I)y}{X'(P - \alpha_n I)y} \\ &= -\frac{(y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)y}{X'(P - \alpha_n I)y} \\ &= -\frac{(y - X\hat{\beta}_{b2sls})'(P - \alpha_n I)(y - X\hat{\beta}_{b2sls})}{X'(P - \alpha_n I)y}.\end{aligned} \tag{12}$$

Interestingly, this expression gives that the HH-test statistic m_2 in (11) is numerically equivalent to the modified Sargan test statistic $\tilde{T}_n$ in (9) up to a sign. Therefore, the HH-test can be regarded as a modification of Sargan's overidentifying restrictions test, where the modification is made to accommodate many instrument asymptotics.

Theorem 2. *It holds that $m_2 = \tilde{T}_n \cdot \mathrm{sgn}\left[-X'(P - \alpha_n I)y\right]$, where $\mathrm{sgn}[\cdot]$ gives the sign of its argument.*

Theorem 2 shows the equivalence between the two test statistics that are constructed under normality. (Note that the main results of HH are also developed assuming normality.) Though details are omitted, the equivalence result can be also obtained without normality between T_n in (6) and the HH-test statistic based on the variance expression in Theorem 4-4 of HH.

The equivalence result remains to hold under more general cases with multiple endogenous regressors. For example, we consider the case of two endogenous regressors $X = (x_1, x_2)$, where x_1 and x_2 are $n \times 1$ vectors. We let $\hat{\beta}_1$ and $\hat{\beta}_2$ be the bias-corrected 2SLS estimators of the coefficient on x_1 and x_2, respectively, using instruments Z. It appears that

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

We also consider the reverse regression of x_1 on y and x_2 using the same instruments Z, and 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 find that

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

[3] (9) does not exactly correspond to the normal-approximation of the chi-square random variable because of the additional factor $1 - \alpha_n$ in the denominator. See Anatolyev (accepted for publication) on this point.

[4] When we use T_n in practice, one problem is that we do not know whether to use the chi-square approximation or the standard normal approximation. As an anonymous referee notes, Anatolyev and Gospodinov (2011) propose a solution by adjusting the critical values so that the chi-square approximation works regardless of the choice of the asymptotic sequence.

[5] HH propose two test statistics, m_1 and m_2, where m_1 is based on the 2SLS estimator and m_2 is based on the bias-corrected 2SLS estimator (i.e., (11)). Theorem 4-3 in HH illustrates that, however, these two test statistics are equivalent (asymptotically). Since m_2 is relatively more tractable for our purpose, we only consider m_2 here. But Theorem 2 can be naturally extended to the equivalence between m_1 and $\tilde{T}_n$. Also note that there is a minor difference between m_2 here and that given in HH: HH use the LIML estimator to compute the standard error while we use the bias-corrected 2SLS estimator. However, the difference disappears at a rate faster than $n^{-1/2}$.

Under normality, the HH-test statistic for two endogenous variables is given by

$$m_3 = \sqrt{n}\check{w}^{-1/2}\left(\hat{\beta}_1 - \hat{\delta}_1^{-1}\right), \tag{13}$$

where

$$\check{w} = \frac{2K}{n-K} \times \frac{\left\{(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)\right\}^2}{\hat{\beta}_1^2\left[x_1'Px_1 - (K/(n-K))x_1'(I-P)x_1 - \frac{\{x_1'Px_2 - (K/(n-K))x_1'(I-P)x_2\}^2}{\{x_2'Px_2 - (K/(n-K))x_2'(I-P)x_2\}}\right]^2}.$$

The following theorem shows that m_3 in (13) is also numerically equivalent to $\tilde{T}_n$ up to a sign.

Theorem 3. *It holds that* $m_3 = \tilde{T}_n \cdot \mathrm{sgn}[-\tau_1]$, *where*

$$\tau_1 = x_1'(P - \alpha_n I)y - \frac{x_1'(P - \alpha_n I)x_2 \cdot x_2'(P - \alpha_n I)y}{x_2'(P - \alpha_n I)x_2}.$$

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 the sample covariance between $\hat{x}_1$ and y after $\hat{x}_2$ being 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 2.

4. Implications

The equivalence result between the HH-test and the modified Sargan test gives several interesting implications. First, HH state that the rejection of the null hypothesis implies an inadequacy of the conventional asymptotic results. The equivalence result in Section 3, however, implies that the HH-test is indeed a test for the moment condition $\mathbb{E}(u_i Z_i) = 0$. Therefore, the HH-test is supposed to have power toward the violation of the instrument-exogeneity or specification, not in detecting the presence of weak or irrelevant instruments. This finding also corresponds to what Hausman et al. (2005) find: the HH-test has very little power in detecting weak instruments.

Using the equivalence result, the power property of the HH-test in detecting the violation of the instrument-exogeneity can be readily analyzed by investigating that of the modified Sargan test.[6] We suppose that the data generating process is given by

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

where e_i is the error term that may be correlated with Z_i and γ is a $K \times 1$ parameter vector. Here $\gamma = 0$ corresponds to the exogeneity assumption of the instruments. We consider the following Pitman-type local alternative:

$$H_a : \gamma = \frac{\alpha_n^{1/4}}{n^{1/4}}\xi, \tag{14}$$

where ξ be a $K \times 1$ nonrandom vector, which does not depend on the sample size n. The following theorem shows that the modified Sargan test (and so does the HH-test from the equivalence result) consistently detects the same set of alternatives the standard Sargan test detects.

Theorem 4. *Suppose that as* $n, K \to \infty$, *both* $\xi'Z'Z\xi/n$ *and* $\xi'Z'Z\Pi/n$ *converge;* $\xi'Z'V/n = o_p(1)$ *and* $\xi'Z'u/\sqrt{n} = O_p(1)$.[7] *Under Assumption 1 and (14),* $T_n \to_d \mathcal{N}(C/\sqrt{w}, 1)$ *as* $n, K \to \infty$, *where*

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

The test thus 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; and $n^{-1/4}$ when K is proportional to the sample size (i.e., $K/n \to \alpha \in (0,1)$). This result illustrates the difficulty of detecting a violation of the instrument-exogeneity condition in the presence of many instruments. When $C = 0$, in addition, the test cannot detect this type of local alternative in (14). A leading example 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 (e.g., Newey, 1985). It thus shows that the overidentifying restrictions test cannot detect local alternatives with $C = 0$ regardless of whether K is fixed or increasing with n.[8]

Second, it is worth noting that the HH-test is two-sided because the test statistic is defined by the difference between two estimates and thus we do not know, a priori, whether a violation of the null hypothesis gives a large negative or positive value of the test statistic. On the other hand, the modified Sargan test is one-sided. When $\hat{u}'Z$ (or $\hat{u}_b'Z$) is close to zero and so is the standard Sargan test statistic (i.e., $\mathbb{E}(u_i Z_i) = 0$ likely holds), the modified Sargan test statistic has a large negative value (see, e.g., (9)) and it should be designed not to reject $\mathbb{E}(u_i Z_i) = 0$ in such a case. Moreover, C in Theorem 4 is non-negative for any ξ,[9] which implies that a violation of the null hypothesis $\mathbb{E}(u_i Z_i) = 0$ (if it can be detected) results in a large positive value of the test statistic T_n. If $\mathbb{E}(u_i Z_i) = 0$ is the null hypothesis, therefore, using the one-sided test based on T_n should achieve a higher power than the HH-test.

Third, HH rule out cases of $\beta = 0$ because an ingredient of the HH-test statistic is the inverse of an estimator. The equivalence result shows that, however, $\beta = 0$ does not cause any problem because we can write the HH-test statistic as T_n that can be well-defined even when $\beta = 0$.

Lastly, the equivalence result is useful when we consider the HH-test in more general settings. For example, the test statistic with two endogenous regressors in Section 5 of HH is very complicated and a larger number of endogenous regressors are difficult to be considered. However, it is straightforward to consider multiple endogenous regressors in the modified Sargan test framework as we discussed in the previous section. Furthermore, as the Sargan test is a special case of the J-test by Hansen (1982), we could consider nonlinear moment restriction models by developing a modified J-test in the presence of many moment conditions, whereas it is not clear how to extend the use of reverse regression equations to such general cases.[10] Another extension is to use the LIML estimator $\hat{\beta}_{\text{liml}} = \arg\min_\beta (y - X\beta)'P(y - X\beta)/(y - X\beta)'(y - X\beta)$,

[6] Here we consider the behavior of T_n but the essentially same result holds for $\tilde{T}_n$.

[7] These conditions are satisfied when $Z_i'\xi$, $\Pi'Z_i$, V_i and u_i have finite fourth-order moments, for example.

[8] Simulation results on comparing the size and power properties between the modified Sargan test and the HH-test can be found in Lee and Okui (2009).

[9] We note that $C = (1-\alpha)\lim_{n,K\to\infty} n^{-1}\xi'Z\left\{I - Z\Pi\left(\Pi'Z'Z\Pi\right)^{-1}\Pi'Z'\right\}Z\xi \geq 0$ since $1 - \alpha > 0$ and $I - Z\Pi\left(\Pi'Z'Z\Pi\right)^{-1}\Pi'Z'$ is idempotent.

[10] Apparently, it is not straightforward to consider the J-test in a general GMM setup particularly when K and n are proportional; we leave this extension as future research. 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.

which is shown to possess good properties in the presence of many instruments (e.g., Anderson et al., 2010). Note that we cannot extend the idea of HH directly using the LIML estimators because the LIML estimator is the optimal linear combination of the bias-corrected forward and reverse 2SLS estimators (HH, p. 169): the two LIML estimators become identical and thus the HH-test statistic based on LIML is zero. However, the modified Sargan test can use $\hat{\beta}_{\text{liml}}$ by obtaining the regression residual from it, and it still satisfies the asymptotic normality of Theorem 1 since $\hat{\beta}_{\text{liml}} - \beta = O_p(n^{-1/2})$ (e.g., van Hasselt, 2010).

Appendix. Mathematical proofs

Throughout the Appendix, van Hasselt (2010) is referred to as vH.

A.1. Proof of (7)

We note that

$$\begin{aligned}
&(y - X\hat{\beta}_{b2\text{sls}})'(P - \alpha_n I)(y - X\hat{\beta}_{b2\text{sls}}) \\
&\quad = (y - X\hat{\beta}_{2\text{sls}} - X\hat{\beta}_{b2\text{sls}} + X\hat{\beta}_{2\text{sls}})'P(y - X\hat{\beta}_{2\text{sls}} - X\hat{\beta}_{b2\text{sls}} \\
&\qquad + X\hat{\beta}_{2\text{sls}}) - \alpha_n(y - X\hat{\beta}_{b2\text{sls}})'(y - X\hat{\beta}_{b2\text{sls}}) \\
&\quad = (y - X\hat{\beta}_{2\text{sls}})'P(y - X\hat{\beta}_{2\text{sls}}) - \alpha_n(y - X\hat{\beta}_{b2\text{sls}})' \\
&\qquad \times (y - X\hat{\beta}_{b2\text{sls}}) + (\hat{\beta}_{2\text{sls}} - \hat{\beta}_{b2\text{sls}})'X'PX(\hat{\beta}_{2\text{sls}} - \hat{\beta}_{b2\text{sls}}), \qquad (A.1)
\end{aligned}$$

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

$$\begin{aligned}
\hat{\beta}_{2\text{sls}} - \hat{\beta}_{b2\text{sls}} &= (X'PX)^{-1}X'Py - \hat{\beta}_{b2\text{sls}} \\
&= (X'PX)^{-1}X'P(y - X\hat{\beta}_{b2\text{sls}}),
\end{aligned}$$

the last component (A.1) is $(y - X\hat{\beta}_{b2\text{sls}})'PX(X'PX)^{-1}X'P(y - X\hat{\beta}_{b2\text{sls}})$, which gives (7) for $n\hat{B} = \alpha_n(y - X\hat{\beta}_{b2\text{sls}})'(y - X\hat{\beta}_{b2\text{sls}}) - (y - X\hat{\beta}_{b2\text{sls}})'PX(X'PX)^{-1}X'P(y - X\hat{\beta}_{b2\text{sls}})$. □

A.2. Proof of Theorem 1

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

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

$$\frac{1}{\sqrt{n\alpha_n}} u'(P - \alpha_n I)u \to_d \mathcal{N}(0, w), \qquad (A.2)$$

$$\frac{1}{\sqrt{n}} u'(P - \alpha_n I)X = O_p(1), \qquad (A.3)$$

$$\frac{1}{n} X'(P - \alpha_n I)X \to_p (1 - \alpha)\Theta. \qquad (A.4)$$

Proof of Lemma A.1. We use Theorem 1 of vH to show (A.2).[11] The matrices U, M, V, C, Ω and a in Theorem 1 of vH are u, 0, u, $(P - \alpha_n I)/\sqrt{\alpha_n}$, σ_u^2 and 1 in this case, respectively. The conditions for Theorem 1 of vH are summarized in Assumption 1 in vH (vHA1 hereafter). A1 denotes Assumption 1 in this paper. A1(iii) corresponds to vHA1(a). In our case, $M = 0$ so vHA1(b) is automatically satisfied. We now consider vHA1(c). The first two conditions in vHA1(c) hold with $Q_{CM} = \mu_{CM} = 0$ because $M = 0$. (Non-positive definiteness of Q_{CM} is not an issue here.) Next, we have

$$\frac{1}{n}\text{tr}\left\{(P - \alpha_n I)/\sqrt{\alpha_n}\right\} = \frac{1}{n\sqrt{\alpha_n}}(K - K) = 0 \quad \text{and}$$

$$\frac{1}{n}\text{tr}\left\{(P - \alpha_n I)^2/\alpha_n\right\} = \frac{1}{n\alpha_n}\text{tr}\left\{(1 - 2\alpha_n)P + \alpha_n^2 I\right\} = 1 - \alpha_n,$$

so the third and the fourth conditions in vHA1(c) are satisfied with $\tau_C = 0$ and $\tau_{C^2} = 1 - \alpha$. The fifth condition in vHA1(c) is also satisfied because $\sum_{i=1}^n (P_{ii}^2 - \alpha_n^2)/(n\alpha_n)$ converges by A1(vii). Lastly, we have

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

by A1(i) and (vi) where $\kappa_{ij} = 1$ if $i = j$ and $= 0$ if $i \neq j$. Therefore, under A1, the conditions for Theorem 1 of vH are satisfied, which yields $u'(P - \alpha_n I)u/\sqrt{n\alpha_n} \to_d \mathcal{N}(0, w)$ as $n, K \to \infty$, where w is given as (5).

We also apply Theorem 1 of vH to show (A.3). The matrices U, M, V, C, Ω and a in Theorem 1 of vH are now (u, X), $(0, Z\Pi)$, (u, V), $(P - \alpha_n I)$, Σ and $(1, 0, \ldots, 0)'$ in this case, respectively. We verify that vHA1 is similarly satisfied as above. A1(iii) implies vHA1(a); A1(v) implies vHA1(b); and A1(iv), (vi) and (vii) imply vHA1(c). Therefore, under A1, Theorem 1 of vH yields (A.3) 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.4) follows. □

Lemma A.2. *Under Assumption 1, we have $(y - X\hat{\beta}_{b2\text{sls}})'(y - X\hat{\beta}_{b2\text{sls}})/n \to_p \sigma_u^2$ and $\sum_{i=1}^n (y_i - X_i'\hat{\beta}_{b2\text{sls}})^4/n \to_p \mathbb{E}(u_i^4)$ as n, $K \to \infty$.*

Proof of Lemma A.2. First, we have

$$\begin{aligned}
&\frac{1}{n}(y - X\hat{\beta}_{b2\text{sls}})'(y - X\hat{\beta}_{b2\text{sls}}) = \frac{1}{n}(\beta - \hat{\beta}_{b2\text{sls}})'X'X(\beta - \hat{\beta}_{b2\text{sls}}) \\
&\quad + \frac{2}{n}(\beta - \hat{\beta}_{b2\text{sls}})'X'u + \frac{1}{n}u'u \to_p \sigma_u^2
\end{aligned}$$

under Assumption 1 since $\beta - \hat{\beta}_{b2\text{sls}} \to_p 0$ as $n, K \to \infty$ by Theorem 3 of vH 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, we similarly have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n (y_i - X_i'\hat{\beta}_{b2\text{sls}})^4 = \frac{1}{n}\sum_{i=1}^n u_i^4 + \frac{4}{n}\sum_{i=1}^n \left\{X_i'(\beta - \hat{\beta}_{b2\text{sls}})\right\}u_i^3 \\
&\quad + \frac{6}{n}\sum_{i=1}^n \left\{X_i'(\beta - \hat{\beta}_{b2\text{sls}})\right\}^2 u_i^2 + \frac{4}{n}\sum_{i=1}^n \left\{X_i'(\beta - \hat{\beta}_{b2\text{sls}})\right\}^3 u_i \\
&\quad + \frac{1}{n}\sum_{i=1}^n \left\{X_i'(\beta - \hat{\beta}_{b2\text{sls}})\right\}^4 \\
&= \frac{1}{n}\sum_{i=1}^n u_i^4 + o_p(1) \to_p \mathbb{E}(u_i^4)
\end{aligned}$$

from Assumption 1. The last equality follows because

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^n \left\{X_i'(\beta - \hat{\beta}_{b2\text{sls}})\right\}^4\right| &\leq \frac{1}{n}\sum_{i=1}^n \|X_i\|^4\|\beta - \hat{\beta}_{b2\text{sls}}\|^4 \\
&= O_p(1)o_p(1) = o_p(1)
\end{aligned}$$

by the existence of the eighth-order moment of X_i and $\beta - \hat{\beta}_{b2\text{sls}} \to_p 0$, where $\|\cdot\|$ is the Euclidean norm. A similar argument can show that $\sum_{i=1}^n \{X_i'(\beta - \hat{\beta}_{b2\text{sls}})\}^3 u_i/n = o_p(1)$, $\sum_{i=1}^n \{X_i'(\beta - \hat{\beta}_{b2\text{sls}})\}^2 u_i^2/n = o_p(1)$ and $\sum_{i=1}^n \{X_i'(\beta - \hat{\beta}_{b2\text{sls}})\}u_i^3/n = o_p(1)$. □

[11] It should be noted that the CLT for quadratic forms by vH is closely related to the CLT by Kelejian and Prucha (2001, 2010). For example, (A.2) can be proved directly from Kelejian and Prucha (2001, 2010) when $K, n \to \infty$ but $K/n \to \alpha > 0$.

Proof of Theorem 1. We observe that

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

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

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

as $n, K \to \infty$. Furthermore, Lemma A.2 implies that $\hat{w} \to_p w$ as $\alpha_n = \alpha + o(n^{-1/2})$. It thus follows that $T_n \to_d \mathcal{N}(0, 1)$ as $n, K \to \infty$. Under the normality, it is easy to see that $\tilde{w} \to_p w$, so $\tilde{T}_n \to_p \mathcal{N}(0, 1)$ follows. □

A.3. Proof of Theorem 2

It is straightforward from (9) and (11) because (12) implies

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

A.4. Proof of Theorem 3

Note that

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

by the definition of the estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. It gives

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

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 $(y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2)'(P - \alpha_n I)x_2 = 0$. Thus, the difference between the two estimators can be written as

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

where $\hat{d}_2 = \sqrt{n/\alpha_n}(\hat{u}_b'(P - \alpha_n I)\hat{u}_b/n)$ with the regression residual $\hat{u}_b = y - x_1\hat{\beta}_1 - x_2\hat{\beta}_2$. Since

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

with $\tilde{w} = 2(1 - \alpha_n)(\hat{u}_b'\hat{u}_b/n)^2$ in this case, it follows that

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

where

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

A.5. Proof of Theorem 4

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

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

from Lemma A.1 since $X'Z\xi/n = \Pi'Z'Z\xi/n + V'Z\xi/n = O_p(1)$ is assumed. Thus, $\hat{\beta}_{b2sls}$ is consistent, even under the local alternative. Similarly to Lemma A.2, it follows 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 yield $\hat{w} \to_p w$ 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

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

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

$$\begin{aligned}
&\frac{1}{\sqrt{n\alpha_n}}\gamma'Z'Z\gamma = \frac{1}{n}\xi'Z'Z\xi, \\
&\frac{1}{\sqrt{n\alpha_n}}\gamma'Z'u = \frac{1}{n^{1/4}\alpha_n^{1/4}}\frac{1}{\sqrt{n}}\xi'Z'u = \frac{1}{K^{1/4}}O(1) = o_p(1), \\
&\frac{1}{n^{3/4}\alpha_n^{1/4}}\gamma'Z'X = \frac{1}{n}\xi'Z'(Z\Pi + V) = \frac{1}{n}\xi'Z'Z\Pi + o_p(1).
\end{aligned}$$

In addition, Lemma A.1 shows that

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

Lastly, Lemma A.1 gives that $u'(P - \alpha_n I)u/\sqrt{n\alpha_n} \to_d \mathcal{N}(0, w)$ and $X'(P - \alpha_n I)X/n \to_p (1 - \alpha) \lim_{n,K\to\infty} \Pi' Z' Z \Pi / n$. Therefore,

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

where C is given in Theorem 4. $\square$

References

Anatolyev, S., 2012. Inference in regression models with many regressors. Journal of Econometrics (accepted for publication).

Anatolyev, S., Gospodinov, N., 2011. Specification testing in models with many instruments. Econometric Theory 27, 427–441.

Anderson, T.W., Kunitomo, N., Matsushita, Y., 2010. On the asymptotic optimality of the LIML estimator with possibly many instruments. Journal of Econometrics 157, 191–204.

Andrews, D.W.K., 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.

Chao, J.C., Swanson, N.R., Hausman, J.A., Newey, W.K., Woutersen, T., 2010. Testing overidentifying restrictions with many instruments and heteroskedasticity. Mimeo.

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

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

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

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

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

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

Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67.

Lee, Y., Okui, R., 2009. A specification test for instrumental variables regression with many instruments. Cowles Foundation Discussion Papers No. 1741.

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

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

van Hasselt, M., 2010. Many instruments asymptotic approximations under nonnormal error distributions. Econometric Theory 26, 633–645.